NOVÁ GENERACE METODOLOGIE GLUE (ODHADU NEJISTOT NA ZÁKLADĚ GENERALIZOVANÉ VĚROHODNOSTI)

Šárka D. Blažková[1], Keith J. Beven[2]

Abstrakt

Příspěvek se zabývá predikcemi průtoků a případně dalších hydrologických veličin s odhadem nejistot pro případ, že existují (alespoň nějaká) data, v úvodu je však uveden postup výběru metody predikcí a odhadu nejistot i pro jiné hydrologické úlohy (předpověď v reálném čase, předpovědi na povodích bez pozorování).

V řadě článků o práci s nejistotami v environmentálním modelování byla již použita nová generace metodologie GLUE (Generalised Likelihood Uncertainty Estimation). V této verzi jsou limity pro přijatelné simulace stanoveny předem, tj. před začátkem modelování. Může se stát, že žádná simulace nebude splňovat tyto podmínky, popř. vyhovujících simulací je tak málo, že z nich není možno sestrojit predikční meze. Potom existuje několik cest, jak postupovat při takové úloze, od zařazení dalších procesů do simulačního modelu, až po zmírnění našich požadavků na shodu modelu s měřenými daty. Jestliže se rozhodneme pro tuto cestu, můžeme z výhodou použít tzv. Pareto optimalizaci, kde všechna kriteria jsou normalizována a mohou být současně zmírňována ve stejné míře.

Klíčová slova: nejistoty, GLUE, hydrologické predikce, Pareto optimalizace

1. Úvod

V evropské a světové odborné hydrologické literatuře se v současné době vedou spory o výběr metody odhadu nejistot. V rámci projektu Risk and Uncertainty (viz informaci ve formě Wiki stránek), kterého se účastní konsorcium výzkumných a aplikačních hydrologických pracovišť ve Velké Británii (UK Flood Risk Management Research Consortium), byl sestaven postup rozhodování o vhodné metodě predikcí s odhadem nejistot v závislosti na typu úlohy, důležitosti zamýšleného projektu a množství dat (včetně jejich neexistence). V úvodu internetové stránky projektu [1] je tento postup přehledně znázorněn v blokovém schématu (Obr.1), přičemž jak celý úvod, tak schéma odkazuje na vysvětlení postupů a termínů a na další prameny.

Toto schéma je rovněž využito v první kapitole nové knihy prof. Bevena Environmental Modelling: An Uncertain Future [2] nedávno vydané v nakladatelství Routledge. V knize je velký prostor věnován přístupům k důležitým rozhodnutím (typu bezpečnosti přehrad nebo úložišť radioaktivních odpadů), jak ze strany výzkumníků, tak ze strany rozhodovatelů (decision-makerů). Beven konstatuje, že environmentální modeláři nebyli zatím příliš dobří v odhadu nejistot svých predikcí a rozhodovatelé moc nevěděli, co si počít s vědeckými nejistotami, pokud je modeláři předložili. Beven považuje představu, že rozhodovatelé nechtějí slyšet o vědeckých nejistotách, za mýtus a je toho názoru, že oni pouze chtějí dostat nejlepší možný vědecký odhad jako základ pro své rozhodnutí. Poskytovat takové podklady je jistě naší povinností (povinností výzkumníků). Myslím, že obě strany cítí, že je tu řada nedorozumění (jak již zmíněné rozdílné názory mezi samotnými výzkumníky), tak mezi výzkumníky a rozhodovateli, tak v komunikaci s dalšími partnery – žádné závažné rozhodnutí nezávisí pouze na výsledcích výzkumu.

Doména rozhodovatelů – analýza rizik, je, jak známo, také předmětem vědeckého bádání. Obtížné případy těžko kvantifikovatelných gnoseologických (epistémických, Knightovských) nejistot jsou v Bevenově knize také zmíněny s odvoláním zejména na knihu Ben-Heima [3].

2. Typy (zdroje) chyb

Aby si čtenář uvědomil, v čem vlastně problém výzkumníka při odhadu nejistot spočívá, je nutno dokreslit situaci presentací zdrojů chyb a nejistot v procesu modelování (e.g. [4]). V libovolném bodě v prostoru nebo čase (x,t) je

$$O(x,t) + \varepsilon_O(x,t) + \varepsilon_C(\Delta x,\Delta t,x,t) = M(\theta,\varepsilon_\theta,I,\varepsilon_I,x,t) + \varepsilon_M(\theta,\varepsilon_\theta,I,\varepsilon_I,x,t) + \varepsilon_r \quad (1)$$

kde O je a vektor pozorování, M je vektor modelovaných hodnot, θ je vektor parametrů modelu, I je vektor vstupů a okrajových podmínek, ε_O je chyba pozorování, $\varepsilon_C(\Delta x,\Delta t,x,t)$ je chyba z nesouměřitelnosti pozorovaných proměnných ve vztahu k modelovaným proměnným závislá na časových a prostorových měřítcích ($\Delta x,\Delta t$), ε_θ je matice chyb v parametrech, ε_I je matice chyb ve vstupech a okrajových podmínkách, $\varepsilon_M(\theta,\varepsilon_\theta,I,\varepsilon_I,x,t)$ je chyba modelu, ε_r je náhodná chyba.

Při praktické aplikaci jediné, co je k disposici, jsou residua:

$$\varepsilon(x,t) = O(x,t) - M(\theta, \varepsilon_\theta, I, \varepsilon_I, x, t) \quad (2)$$

Z (1) a (2) vyplývá, že residua ε(x,t) mohou být vyjádřena jako

$$\varepsilon(x,t) = \varepsilon_M(\theta,\varepsilon_\theta,I,\varepsilon_I,x,t) - \varepsilon_C(\Delta x,\Delta t,x,t) - \varepsilon_O(x,t) + \varepsilon_r \quad (3)$$

Je nepravděpodobné, že první tři členy na pravé straně rovnice (3) by byly svou podstatou náhodné. Budou zahrnovat různé gnoseologické chyby. Je zřejmé, že tyto různé členy na pravé straně nemohou být desagregovány z jediné pozorované řady residuí ε(x,t), neučiníme-li velmi silné předpoklady (např. předpoklad aditivity a Gaussovského rozdělení).

3. Metoda GLUE

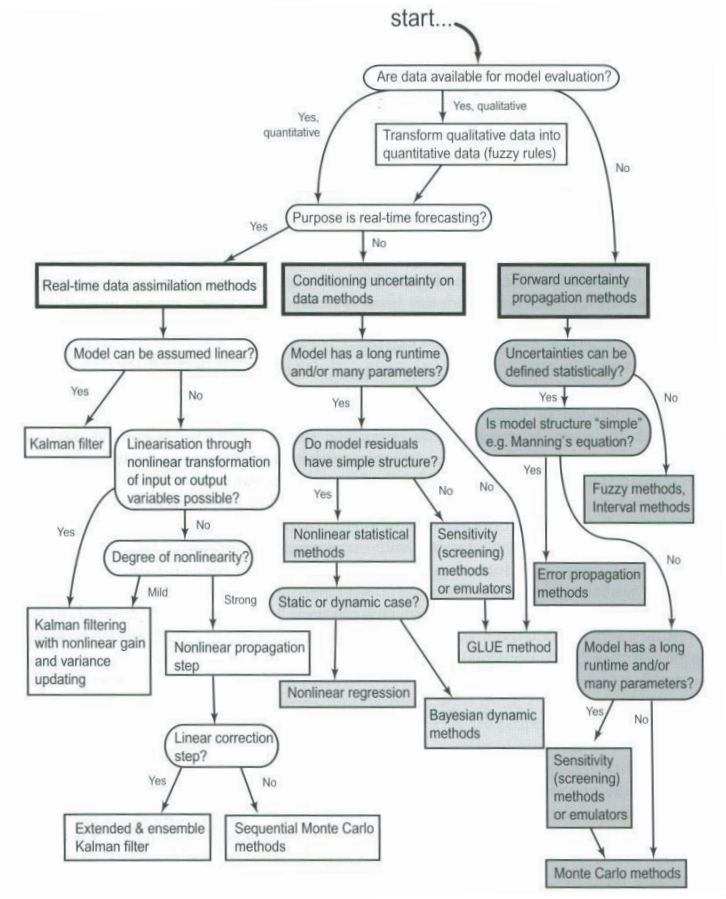
Metoda odhadu nejistot na základě generalizované věrohodnosti (např. [5], [6]) je, jak je zřejmé z Obr.1, vhodná pro případy, kdy výše zmíněné silné předpoklady o chybách při modelování neplatí, naším cílem je poznání fungování hydrologických procesů (základní výzkum) nebo se jedná o projekt pro praxi s případnými dalekosáhlými důsledky a kdy je k disposici výpočetní technika, která tento přístup umožňuje (tj. např. paralelní systém PC VÚV TGM nebo University v Lancasteru s desítkami procesorů).

Parametry modelu jsou vzorkovány z fyzikálně rozumných rozmezí, nepředpokládá se existence jediného optima každého parametru, důležitá je sada parametrů jako celek. Prozkoumává se tedy co nejpodrobněji (v závislosti na možnostech výpočetní techniky) parametrický prostor – hledají se pokud možno všechny simulace, které jsou v rozumné shodě s pozorovanými daty (v Obr.1 termín *conditioning uncertainty on data*). Neshodující se simulace jsou odmítnuty – dále se s nimi nepracuje. V současné době úlohy GLUE ve vědecké hydrologické literatuře mají tisíce až milióny simulací.

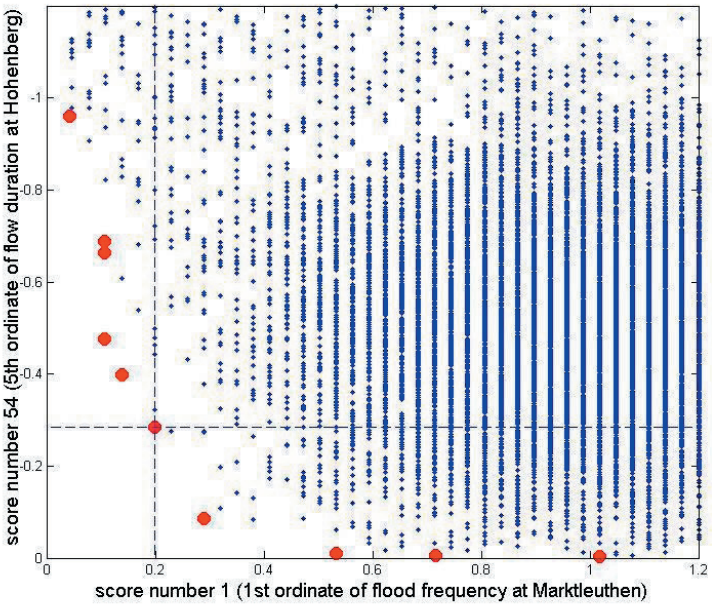
V nové generaci metodologie GLUE [7] jsou limity pro přijatelné simulace stanoveny předem, tj. před začátkem modelování. Za vyhovující simulace považujeme takové, které jsou v rámci mezí chyby pozorování. V případě průtoků jsme pracovali s chybou stanovení konsumpční křivky [8] v několika profilech, s chybou stanovení maximální roční vodní hodnoty sněhu a s odhadem chyb výskytu ročního maxima průtoků v určité sezóně. Celkem jsme pracovali se 114 kriterii (sledovali jsme zvlášť horní a dolní meze jednotlivých kvantilů čáry překročení maximálních průtoků a čar trvání průtoků). Vyhovujících simulací bylo tak málo (39), že z nich nebylo možno sestrojit predikční meze.

V takovém případě existuje několik cest, jak postupovat od zařazení dalších procesů do simulačního modelu (to jsme činili několikrát), přes hustší vzorkování parametrického prostoru (tato cesta již byla rovněž použita), přes definování a použití modelu chyb (error model – což v tomto případě jsme nepovažovali za schůdné vzhledem ke složitosti úlohy) až po zmírnění našich požadavků na shodu modelu s měřenými daty. Tato cesta byla použita ve spojení s optimalizací typu Pareto, kde všechna kriteria jsou normalizována a mohou být

současně zmírňována ve stejné míře (Obr. 2). Nakonec jsme pracovali s 4192 simulacemi, z nichž byly vypočteny predikční meze.



Obr.1: Blokové schéma (decision tree) výběru metody odhadu nejistot.
Podle http://www.floodrisknet.org.uk/methods/Introduction; převzato z publikace [2]



Obr. 2: Zjednodušený případ Pareto fronty konstruované ze dvou kriterií, která jsou v rozporu. Tečky jsou jednotlivé simulace, velké kroužky jsou simulace ležící na Pareto frontě. Dřívější verse obrázku z článku [8]

4. Závěr

Nová verze metody GLUE byla použita v rámci základního výzkumu postupu pro důležité případy odvození povodní dlouhé doby opakování (1000 nebo 10 tisíc let na důležitých vodních dílech). Délka simulací v článku [8] byla 10 tisíc let, což považujeme za vyhovující pro dobu opakování 1000 let. V současné době ve výzkumu pokračujeme modelováním řad o délce 100 tisíc let a vývojem metody výběru návrhových hydrogramů z dlouhých simulací.

Poděkování

Výzkum byl podpořen z projektu Labe V (SP/2e7/229/07) a z předcházejících grantů MŽP.

Literatura

[1] Floodrisknet, 2008. http://www.floodrisknet.org.uk/methods/Introduction

[2] BEVEN, K. J., 2009. Environmental Modelling: An Uncertain Future?, Routledge: London.

[3] BEN-HAIM, Y., 2006. Info-Gap Decision Theory. 2nd Edition, Academic Press: Amsterdam.

[4] BEVEN, K. J., 2005. On the concept of model structural error. Water Science & Technology, 52, pp165–175.

[5] BEVEN, K. J. – BINLEY, A. M., 1992. The future of distributed models: model calibration and uncertainty prediction, Hydrol. Process., 6: pp 279–298.

[6] BEVEN, K. J., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling, Adv. Water Resource :16, pp 41–51.

[7] BEVEN, K. J., 2006. A manifesto for the equifinality thesis. J. Hydrology, 320, pp 18–36.

[8] BLAZKOVA, S. –BEVEN, K., 2009. A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic, Water Resour. Res., 45, W00B16, doi: 10.1029/2007WR006726

Kontakty na autora

1 Výzkumný ústav vodohospodářský T.G. Masaryka v.v.i., e-mail: sarka_blazkova@vuv.cz

2 The Lancaster Environment Centre, Lancaster University, Uppsala University, e-mail: k.beven@lancaster.ac.uk