

# MODELY ŠÍŘENÍ ZNEČISTĚNÍ ŽIVOTNÍM PROSTŘEDÍM – OD DETERMINISTICKÝCH ODHADŮ K PRAVDĚPODOBNOSTNÍMU HODNOCENÍ

Petr PECHA, Ústav teorie informace a automatizace, AV ČR, Praha  
Luboš HOUSA, absolvent FJFI ČVUT Praha, KM, softwarové inženýrství

## 1 Úvod

Práce na vývoji a implementaci pravděpodobnostního přístupu k hodnocení následků mimořádných úniků škodlivin do životního prostředí byly zahájeny v rámci projektu 6/2003 (poskytovatel SÚJB) za spolupráce autorů s kolegy z ÚJV Řež, divize EGP. Výsledky projektu v této oblasti jsou popsány v [18], konkrétním výsledkem je pak verze výpočetního kódu HAVAR RP (Reliability Prediction). V současné době pokračuje vývoj v ÚTIA AV ČR ve spolupráci s absolventy KM FJFI, výhledově s cílem vytvořit metodologický a SW základ pro rozvoj moderních bayesovských metod pro podporu rozhodování v krizových situacích vyvolaných nehodami jaderné či nejaderné povahy.

Aby nedošlo k nedorozumění ohledně zde zmiňované problematiky, vymezíme nejprve její roli z hlediska celkového pohledu na bezpečnostní analýzy označované termíny PSA (Probabilistic Safety Assessment) či PRA (Probabilistic Risk Assessment). Metody poskytují komplexní pohled na bezpečnost jaderného zařízení, kdy se musí brát v úvahu pravděpodobnost selhání různých komponent a následující vývoj na základě stromu poruch se specifickými přechodovými procesy. Cílem je pak určování konzistentních měr bezpečnosti formulovaných z numerických odhadů prováděných na pravděpodobnostním základě. Podle typu analyzované oblasti a následků uvedeme obvyklou klasifikaci užívanou v jaderné oblasti:

- „*PSA-úroveň 1*“ - identifikuje a kvantifikuje sekvence událostí vedoucích ke ztrátě integrity struktur aktivní zóny a k masivnímu poškození paliva;
- „*PSA-úroveň 2*“ - vychází z výsledků analýz v předcházející úrovni 1 a pokračuje v šetření následků v oblasti kontainmentu, přičemž poskytuje odhady izotopického složení a dynamiky úniku aktivity do životního prostředí;
- „*PSA-úroveň 3*“ - na základě předchozích analýz popisuje transport radionuklidů při únicích vně jaderného zařízení, určuje distribuci aktivity v komponentách životního prostředí a v konečné fázi odhaduje následky radiačních nehod na lidské zdraví.

Ve fázi povolovacího řízení analýzy PSA a zvláště pak jejich úroveň 3 musejí demonstrovat splnění bezpečnostních kritérií daných vyhláškami a prokázat tak způsobilost nového technického díla bezpečně plnit svoji funkci. Jedním z postupů (například u kódu PC COSYMA) takové pravděpodobnostní analýzy je volba jedné konkrétní sady všech vstupních parametrů v konzervativním směru s výjimkou meteorologie, kdy výpočty v celém prostoru kolem zdroje radioaktivního znečištění jsou mnohonásobně opakovány pro různé meteorologické sekvence (určované například na základě výběru z dlouhodobých povětrnostních statistik). V libovolném bodě prostoru jsou pak konstruovány distribuční funkce sledovaných výstupních veličin, které například umožní stanovit, s jakou pravděpodobností nebude přestoupen daný limit.

Ve světě bylo vypracováno několik výpočetních kódů, které se snaží provést úplnou analýzu neurčitostí všech vstupních parametrů nebo alespoň dílčí analýzu neurčitostí některých submodelů (atmosférický nebo depoziční, potravní řetězce, dosimetrický model) [3, 4, 9, 10, 11]. Vzhledem ke složitosti úlohy a její silné nelinearitě se jedná o mnohonásobné Monte

Carlo výpočty s následujícím statistickým zpracováním náhodných realizací výstupu. Je třeba si tedy uvědomit, že principem takových procedur je vlastně opakované vyvolávání deterministického jádra environmentálního modelu vždy s konkrétní realizací náhodných vstupů modelu (i když se to děje automaticky).

Naše aspirace se netýkají celkového posuzování ani zpětného prokazování způsobilosti provozu již pracujícího jaderného zařízení. Problém zúžíme na pravděpodobnostní analýzu vztaženou k jednomu zcela konkrétnímu okamžiku, který ztotožníme s počátkem mimořádného úniku radioaktivity do atmosféry. Necht' je k tomuto okamžiku k dispozici popis konkrétní meteorologické situace a současně i její krátkodobá předpověď na 48 hodin. Například pro hodinové sekvence bodových předpovědí (t. zn. vztažených k místu jaderného zařízení) generované ČHMÚ jsou pro každý hodinový interval předpovídány (střední) hodnoty rychlosti větru v referenční výšce (obvykle 10m nad zemí), směru větru, kategorie stability atmosféry, výšky směšovací vrstvy, atmosférické srážky (mm/hod), ... apod. Tyto hodnoty považujeme za nejlepší odhad („best estimate“, dále označované též jako nominální) dodávané poskytovatelem dat, nicméně stochastický charakter jevů má za následek, že tyto odhady jsou zatíženy pozorovacími, měřicími a modelovými chybami a takovým odhadům skutečné (neznámé) hodnoty musí být přiřazen náhodný charakter.

Na základě předchozích úvah lze rozlišit dva základní přístupy k modelování, a sice:

- tradiční **deterministický přístup**, kdy vstupním parametrům modelu jsou přiřazeny jejich očekávané (případně konzervativní „nejhorší“) hodnoty, přičemž model následně generuje vždy jedinou očekávanou hodnotu sledovaného výstupu (případně její jedinou nejhorší hodnotu – v případě tzv. „worst-case“ analýz),
- moderní **pravděpodobnostní přístup**, kdy většině nejdůležitějších vstupních parametrů je přiřazen náhodný charakter včetně meteorologických charakteristik (u nich nejde o časovou variabilitu, nýbrž o stochastický charakter ve výše zmíněném smyslu).

Pravděpodobnostní přístup ve svém důsledku umožňuje statisticky popsat šíření vstupních neurčitostí modelem, analyzovat citlivost výstupu na fluktuace jednotlivých vstupních parametrů modelu a kvantifikovat prostorovou závislost chyb modelu, což souhrnně vyjadřují termíny:

1. **Analýza neurčitostí (AN)**, která ve svém důsledku umožní generovat pravděpodobnostní odpovědi na otázky hodnocení následků mimořádných úniků.
2. **Citlivostní studie (CS)**, umožňující definovat relevantní grupu náhodných vstupních parametrů modelu, tzn. omezit se na ty parametry, které mají nezanedbatelný vliv na fluktuace sledovaných výstupů.
3. **Kovarianční matice (KM)** chyb modelu, která definuje prostorové šíření fluktuací modelovaných hodnot.

Z předchozího textu by již mělo být zřejmé, na co jsme problém zúžili a na co se soustředíme. Kromě dílčích analýz AN, CS a KM konečným cílem a záměrem je možnost zpřesnit matematické modelování na základě korekcí se skutečnými pozorováními (měřeními) v terénu, prováděných časově konsistentně s časovými kroky matematického modelu. Tato procedura se nazývá **asimilace modelu s měřeními** a obecně je dnes považována za nezbytný krok ke **zvýšení spolehlivosti** matematických odhadů a k přechodu od izolovaného modelování k věrnějšímu popisu reality. Současně definovanou úlohu považujeme za jednu z významných procedur z širokého spektra metod analýz PSA úrovně 3.

## 2 Matematicko-fyzikální model šíření znečištění životním prostředím

Deterministický model zavádí konceptuální schémata pro vyjádření přírodních procesů ve formě soustav diferenciálních rovnic s příslušnými počátečními a okrajovými podmínkami. Při dostatečné výpočetní kapacitě a aplikaci nových znalostí je možno tyto modely v určitých směrech zpřesňovat a rozšiřovat jejich složitost, nicméně jakkoliv sofistikované deterministické modely nemohou zcela do důsledků popsat přírodní procesy, které jsou ve své podstatě fraktální a chaotické.

V dalším textu zaměříme pozornost na analýzu mimořádných úniků škodlivin do atmosféry z přízemních nebo vyvýšených zdrojů v bližších a středních vzdálenostech do 100 km od zdroje znečištění. Pozornost bude soustředěna především na následující veličiny:

- 3-D rozložení koncentrace škodlivin ve vzduchu, umožňující například kontrolu překročení prahových koncentrací chemických substancí,
- 2-D rozložení časových integrálů koncentrace škodlivin v přízemní vrstvě vzduchu, které jsou důležité při krizovém zvládnání jaderných událostí při rozhodování o vyhlášení neodkladných opatření (jodová profylaxe, ukrytí či evakuace obyvatel) v zasažených oblastech,
- 2-D rozložení plošných koncentrací škodlivin deponovaných na zemském povrchu v okolí do 100 km od zdroje znečištění, což bude důležité hlavně pro případ dalšího terestrického transportu radionuklidů a na něm závislých různých typů dávek záření a jejich úvazků, užívaných v krizovém řízení pro klasifikaci závažnosti nehody a rozhodování o aplikaci následných protipatření na ochranu obyvatel. Může být důležitá i depozice některých chemických látek.

Je důležité si uvědomit, že vyjmenované 3 typy výstupů modelu jsou **hlavními řídicími veličinami** pro následné hodnocení důsledků nehod, pomocí nichž jsou potom dopočítávány nejrůznější typy dávek záření a prováděna kvantifikace zdravotního rizika. Současně je třeba za tím vidět složitost celého problému, kdy například pro radioaktivní úniky je nutné řešit úlohu prostorově, pro každý nuklid z unikajícího spektra zvlášť a v závěru provést superpozici následků přes celou grupu radionuklidů.

Matematicky formulováno, modelované výstupní veličiny lze obecně vyjádřit vektorem  $Y$ :

$$Y = \mathfrak{R}(Z), \text{ kde } Y \equiv [Y_1, Y_2, Y_3, \dots]; \quad Z \equiv [Z_1, Z_2, Z_3, \dots, Z_M]; \quad (1a)$$

Velká písmena vyjadřují fakt, že příslušné proměnné mají náhodný charakter, jejich konkrétní náhodné realizace budou označovány malými písmeny. Vektor  $Z$  vstupních parametrů modelu  $\mathfrak{R}$  má dimenzi  $M$ . Pokud se omezíme na jedinou výstupní proměnnou představovanou jednou složkou  $Y_i$  vektoru  $Y$  (například na jednu ze shora explicitně vyjmenovaných na začátku tohoto odstavce), bude se jednat o analýzu jediné skalární proměnné (v dalším textu se bude jednat většinou o prostorové rozložení depozice určitého jednoho testovaného nuklidu) podle schématu:

$$Y_i = Y = \mathfrak{R}^{\text{GPM}}(Z) \quad (1b)$$

V dalším používáme fyzikální model  $\mathfrak{R}^{\text{GPM}}$  advekce a disperze škodlivin ve vlečce unášené větrem (a příslušných navazujících submodelů), který je formulován v přiblížení Gaussova modelu v jeho jednoduché variantě GPM (Gaussian Plume Model) pro přímočaré šíření (vyhovuje v blízkosti zdroje) a alternativně ve složitější verzi segmentovaného Gaussova modelu  $\text{GPM}^{\text{seg}}$ , který umožňuje respektovat časově proměnné sekvence krátkodobých meteorologických předpovědí a modelovat situaci ve větších vzdálenostech. V tomto případě

jsou vstupní parametry  $Z_m$ ,  $m=1, \dots, M$  vyjádřeny jejich nominálními (očekávanými) hodnotami  $z_m^{\text{nom}}$ . Pravděpodobnostní verze obou modelů, kdy vstupní parametry jsou považovány za náhodné, umožňují modelovat šíření jejich neurčitostí modelem směrem ke sledovaným výstupním veličinám. V současné době je algoritmus výpočetně náročného pravděpodobnostního modelu GPM<sup>seg</sup> testován na jeho použití v procesu mnohonásobně opakovaných cyklů.

### 3 Variabilita a neurčitost

Transportní procesy přenosu znečištění jsou ve své podstatě stochastické a příslušné neurčitosti neumožňují dosáhnout přesné predikce (turbulentní mechanismy v atmosféře, molekulární procesy suché a mokré depozice apod.). Jiné vstupní parametry mohou vykazovat tak zvanou variabilitu, pokud se mění s časem či v prostoru nebo se jejich hodnoty mohou lišit pro jednotlivé kategorie v rámci celé populace (dynamika úniku znečištění ze zdroje, suchá depozice pro jednotlivé kategorie typu zemského povrchu, spotřeba potravin pro jednotlivé věkové kategorie atd.). Z tohoto hlediska je třeba rozlišovat mezi **neurčitostí** a **variabilitou**.

**Variabilita** obecně vyjadřuje difference vstupní proměnné způsobené časovými, prostorovými nebo mezikategoriálními závislostmi. Na druhé straně stochastické vlastnosti vedou k **neurčitostem** vstupů v důsledku neúplné znalosti a neúplného popisu složitého systému, parametrizace modelu nebo v důsledku chyb v určování datových konstant.

Ponechme stranou další podrobnosti a uveďme hlavní myšlenku, jakým způsobem popisované programy ošetřují variabilitu některých proměnných. I když je fakt, že v určitých případech lze variabilitu interpretovat jako (diskrétní) veličinu s určitou empirickou frekvenční funkcí, v produktu HAVAR RP je důsledně použit princip nikoliv jediné veličiny s určitou distribucí, nýbrž rozlišení na kategorie je provedeno již v samotné vstupní části. Tak například výstupy (dávek, dávkových příkonů, zdravotního rizika) jsou generovány paralelně pro všechny věkové kategorie. Stejně tak primární vstup celého prostorového rozložení typu zemského povrchu kolem zdroje znečištění (kategorizován do pěti skupin: 1- vodní plocha, 2- tráva, 3- zemědělské kultury, 4- lesy, 5- zástavba) umožní implicitně rozlišit skutečný lokálně závislý typ povrchu a zahrnout jeho vliv na prostorové rozložení výstupů.

Ovšem v rámci konkrétní kategorie jsou uvažovány neurčitosti každé dílčí kategorie. S časovými variacemi určité proměnné se obvykle vypořádáváme tak, že se provede její středování v rámci určitého časového intervalu, přičemž se určují neurčitosti této střední hodnoty. Tak například skutečná dynamika úniku škodlivin je nahrazována počástech konstantními ekvivalentními (hodinovými) úniky s následným odhadem neurčitosti těchto středních hodnot. Stejný princip může být použit i v případě prostorového středování prováděný s ohledem na použitou polární výpočetní síť. U některých složitých konceptuálně zaváděných charakteristik není hodnota pro jednotlivé kategorie považována za náhodnou, nicméně podle její hodnoty jsou generovány neurčitosti dalších asociovaných proměnných. Tak například pro konkrétní kategorii stability atmosféry podle Pasquillova třídění jsou odvozovány neurčitosti koeficientů disperze nebo stochastické fluktuační středních hodnot směru větru.

Metodami kvantifikace variability a neurčitostí vstupů se zabývá statistická analýza náhodných simulací (pozorování). Jednou z možností jsou techniky vycházející z bayesovské inference se zahrnutím určité apriorní informace o náhodných charakteristikách analyzovaných dat. S rozvojem výpočetní techniky nabyly na významu postupy klasické statistiky, kdy z analyzované populace nezávislých pozorování jsou postupně vybírány  $m$ -tice prvků ( $s$  vrácením), přičemž pro každý takový výběr se vypočítává požadovaná výběrová distribuční statistika (například průměr, standardní odchylka, vychýlení). Pro velký počet

provedených výběrů se tak dostane stejný počet odhadů požadované hledané statistiky, z nichž lze stanovit neurčitost distribuční statistiky celé populace. Naznačený postup popisuje princip tzv. **bootstrap** simulace, která je možnou alternativou k další často používané metodě **maximální věrohodnosti** pro odhad hledaných statistik. Výsledky analýz jsou následně uplatněny při elicitačních procedurách expertů [2, 3, 4, 22] vyúsťujících v doporučení pro volbu charakteristik variability nebo neurčitostí.

#### 4 Hlavní příčiny existence neurčitostí vstupních parametrů

Neurčitosti vstupující do modelů šíření znečištění lze klasifikovat jako charakteristiky týkající se přirozené podstaty věci, pojmového (konceptuálního) vymezení (omezení) modelu a datové (parametrické) oblasti:

- Do první skupiny patří náhodné proměnné, jejichž neurčitosti vedou k nevyhnutelnému faktu, že systém má stochastický charakter. Většina takových proměnných je náhodná již ve svém principu, další mohou být modelovány jako náhodné přesto, že při vynaložení velkého úsilí a prostředků by mohly být stanovovány (měřeny) přesně. Ze znalosti vlivu šíření neurčitostí modelem můžeme určit oblasti, kde další alternativní výzkum nebo případná další zpřesňování měřících technik mohou vést k redukci chyb a tím ke zvyšování spolehlivosti predikce modelu.
- Každý matematický model obsahuje jistá zjednodušení reality a zavádí tak další neurčitosti do modelování. Již v samotném počátku při volbě výpočetního modelu je nutné udělat kompromis mezi detailností popisu (Gaussův model, Lagrangeova částicová schémata, podrobný eulerovský přístup) a časovou náročností a složitostí výpočtu. Potom následuje celá řada dalších vynucených kroků týkajících se volby alternativních submodelů transportu znečištění, detailnosti výpočetní mříže časového a prostorového rozlišení, volby počátečních a okrajových podmínek atd. Ukazuje se, že nedokonalá znalost fyzikální podstaty věci vedoucí k použití zjednodušených dílčích poloempirických formulí představuje klíčový zdroj neurčitostí výsledné predikce.
- Neurčitosti odhadu parametrů (konstant) modelu pramení z mnoha příčin. Některé neurčitosti vyplývají z chyb měření (náhodné chyby analytických měřících zařízení, systematická zkreslení v důsledku chybné kalibrace přístrojů), další například z nepřesných transformací hodnot z nepřímých pozorování nebo z vyvozování závěrů o náhodných charakteristikách parametrů modelu na základě experimentů provedených s nereprezentativními vzorky.

Vzhledem ke složitosti modelování transportu aktivity životním prostředím a jejich zřejmé nelinearitě ve vztahu ke vstupním parametrům jsou používány numerické metody založené na Monte-Carlo postupech spočívajících v mnohonásobném opakování výpočtu se sekvencemi předem vygenerovaných náhodných realizací vstupů. První fází výpočtů je tedy generování náhodných realizací grupy vstupních parametrů, které předpokládá pečlivou analýzu v několika krocích:

1. Identifikace množiny  $M$  vstupních parametrů, které nezanedbatelně přispívají k neurčitosti predikce.
2. Definice distribuční funkce případně odhady dolní a horní hranice fluktuací jednotlivých vstupních parametrů.
3. Odhady a zahrnutí míry korelace mezi jednotlivými vstupními parametry.
4. Volba vhodné metodiky generování realizací vstupních náhodných vektorů.

5. Provedení co největšího počtu  $K$  výběrů (který připouští kapacita výpočetního zařízení) realizací náhodného vstupního vektoru.

Prvním krokem je tedy správný odhad rozsahu grupy vstupních parametrů modelu a jejich náhodných charakteristik. Z literatury vyplývá, že model obvykle obsahuje velké množství vstupních parametrů, z nichž jen určitá část svými fluktuacemi významněji ovlivňuje výstup. To umožňuje na základě studií citlivosti provádět určitou redukci počtu parametrů a tak definovat optimální rozměr  $M$  náhodného vstupního vektoru  $\mathbf{Z} \equiv [Z_1, Z_2, Z_3, \dots, Z_M]$ .

Pro vektor  $\mathbf{Z}$  lze zavést některé jeho číselné (maticové) charakteristiky. K základním patří vektorová střední hodnota a kovarianční matice. Střední hodnota náhodného vektoru se definuje jako vektor středních hodnot jeho jednotlivých složek, tedy:

$$E[\mathbf{Z}] = \{ E[Z_1], E[Z_2], \dots, E[Z_M] \} \quad (2a)$$

Pro dvě náhodné veličiny  $Z_i$  a  $Z_k$  je kovariance  $cov(Z_i, Z_k)$  definována podle:

$$cov(Z_i, Z_k) = E[ (Z_i - E[Z_i]) \cdot (Z_k - E[Z_k]) ] \quad (2b)$$

přičemž pro  $i=k$  přechází  $cov(Z_i, Z_i)$  na rozptyl  $D(Z_i)$ .

Kovarianční matici  $C(\mathbf{Z})$  vstupního náhodného vektoru  $\mathbf{Z} \equiv [Z_1, Z_2, Z_3, \dots, Z_M]$  lze zapsat schématem:

$$C(\mathbf{Z}) = \begin{bmatrix} D(Z_1) & cov(Z_1, Z_2) & \dots & cov(Z_1, Z_M) \\ cov(Z_2, Z_1) & D(Z_2) & \dots & cov(Z_2, Z_M) \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ cov(Z_M, Z_1) & cov(Z_M, Z_2) & \dots & D(Z_M) \end{bmatrix} \quad (2c)$$

Rozsah grupy (to znamená počet  $M$  relevantních vstupních parametrů) se odhaduje na základě analýzy významnosti vlivu parametru na výstupní veličiny (předběžné třídění a citlivostní studie, přičemž statisticky nevýznamné parametry jsou dále charakterizovány pouze deterministicky jejich nominálními hodnotami).

## 5 Charakteristiky neurčitostí vstupních parametrů modelů šíření znečištění

Existuje několik přístupů pro vyjádření neurčitostí v závislosti na doménách aplikace. V oblasti transportních modelů připadají v úvahu metody vyplývající z teorie intervalové matematiky, teorie „fuzzy sets“ a pravděpodobnostní analýzy [6, 7, 10]. Cílem intervalové analýzy je odhadovat rozsahy hodnot sledovaných cílových veličin modelu v závislosti na intervalových rozsazích vstupních parametrů. Aplikace teorie fuzzy množin umožňuje vyjádřit takovou neurčitost, která místo klasicky chápané náhodnosti má spíše charakter jisté vágnosti nebo nejasnosti pojmu. „*Pojmy s neurčitě vymezenými hranicemi prostupují lidský jazyk a myšlení, a hrají tak důležitou roli při výměně informací. Profesor Lofti Zadeh si této okolnosti všimnul a učinil z ní základ nové ideje, kvantifikace neurčitosti. Výsledkem jeho práce je koncept t. zv. fuzzy množin, množin, u kterých je pro každou věc z nějakého (vhodně zvoleného) universa určen stupeň příslušnosti, kterým tato věc náleží množině.*“ [19]. Fuzzy logika našla uplatnění při řízení složitých systémů, například v oblasti konstrukce rozhraní pro

komunikaci s roboty v přirozené řeči jak v psané tak v mluvené formě, kdy expertní znalosti operátora jsou převáděny do formy lingvistických pravidel.

V naší analýze zatím používáme výhradně **pravděpodobnostní přiblížení**, které je adekvátní při použití ve fyzikálních systémech pro kvantitativní popis obsažených neurčitostí. Při určování distribuční funkce náhodného rozložení vstupních parametrů se používají buď statistické metody odhadu anebo distribuce je konstruována empiricky (a tím i subjektivně) na základě expertních odhadů a doporučení. Použití statistických metod je podmíněno dostatečným množstvím reprezentativních dat (vzorků), kdy určité události je přiřazována jistá pravděpodobnost interpretovaná v termínech frekvence výskytu události. Při dostatečně velkém počtu pokusů je tato pravděpodobnost definována v klasickém smyslu jako poměr počtu výskytu události k celkovému počtu pokusů. V případě nedostatečného množství dat mohou být použity zmíněné subjektivní odhady distribucí doporučené experty v dané oblasti.

Na základě obecně přijatého souhlasu je nutné deklarovat následující vlastnosti náhodných vstupních parametrů modelů:

- Myslitelný rozsah jejich hodnot (např. horní a dolní meze možných hodnot)
- Typ rozdělení náhodného parametru
- Vzájemnou závislost mezi náhodnými hodnotami parametrů

K určování typu rozdělení poznamenejme, že je nutné respektovat obecně publikovaná expertní doporučení (pokud jsou pro daný případ k dispozici). Je obecným pravidlem, že jen zřídka lze experimentálně podpořit konkrétně provedený závěr. Pokud jsme nuceni přijmout vlastní rozhodnutí, je třeba si uvědomit možný subjektivní prvek, který by při povrchním postupu mohl vést ke zkreslení výsledků. Obecně doporučovaný postup pro tento případ říká, že pokud máme minimální znalosti o možných náhodných charakteristikách, je rozumné se omezit na rovnoměrné rozdělení v rámci definovaného rozsahu.

Jestliže bude k dispozici další expertní znalost směrem k podrobnějšímu určení typu rozdělení, lze akceptovat příslušný lépe odpovídající typ jako je unimodální a symetrické rozdělení (např. trojúhelníkového typu) nebo rozdělení respektující předpokládané zkosení k dolní či horní mezi. Zkušenosti ukazují, že mnoho kladných fyzikálních, chemických, biologických i toxikologických náhodných proměnných sleduje logaritmický trend. Proto pro fluktuační většího rozsahu by měla být dáována přednost pracovat s logaritmy příslušných hodnot a přizpůsobit uniformní, trojúhelníkové nebo normální rozdělení na logaritmická (LogUniform, LogTriangular, LogNormal). Je třeba mít na paměti, že často i experti obtížně formulují průběhy hustoty pravděpodobnosti v limitních dolních a horních oblastech. Je proto lepší nadefinovat určité rozumné hraniční body a uvažovat omezená (truncated) rozdělení. Pravděpodobnostní odhady následků úniků budou reprezentativní jen tehdy, pokud předchozí proces přiřazení náhodných charakteristik všem důležitým vstupním parametrům bude korektní.

Pravděpodobnostní verze modelu GPM jsou implementovány do systému HAVAR-RP vyvíjeného v rámci projektu 6/2003 [18]. Jedním z jeho subsystémů je interaktivní aplikace GENERATOR vyvinutá a popsána v [5], která implementuje metodu LHS (Latin Hypercube Sampling) pro vzorkování náhodných vstupů. V první fázi uživatel interaktivně volí pro každý parametr typ rozdělení a jeho charakteristiky včetně způsobu zadávání. Panely nabízejí pro typ rozdělení následující volby:

- normální, normální omezené
- logaritmicko normální, logaritmicko normální omezené (LogNorm)
- rovnoměrné

- logaritmicky rovnoměrné (LogUni)
- diskrétní rovnoměrné
- trojúhelníkové

Způsoby zadání charakteristik (středních hodnot a rozptylů případně mezí) lze provádět z nabídky:

- standardní zadání (absolutní hodnoty)
- standardní zadání (relativní hodnoty)
- kvantily (0.05 a 0.95) - absolutně
- kvantily (0.10 a 0.90) - absolutně
- kvantily (0.05 a 0.95) - relativně
- kvantily (0.10 a 0.90) - relativně

$m$ -tá složka vstupního vektoru (1b) představuje  $m$ -tý náhodný vstupní parametr  $Z_m$ . Necht' má například normální rozdělení se střední hodnotou  $z_m^{\text{nom}}$  a rozptylem  $\sigma_m^2$ . Při absolutním standardním formátu zadávání je myšleno přímé generování z rozdělení  $N(z_m^{\text{nom}}, \sigma_m)$ . Pokud je specifikováno relativní zadávání, parametr je vyjádřen separací

$$Z_m = z_m^{\text{nom}} \cdot C_m \quad \text{resp.} \quad (Z_m)^a = (z_m^{\text{nom}})^a \cdot C_m \quad (3)$$

Zde  $C$  je normalizovaná náhodná proměnná s mediánem 1.0 (0.5-kvantil) a mezemi či disperzemi transformovanými z příslušných hodnot absolutní proměnné případně jejího rozsahu (ten může být zadán i v libovolné škále, automaticky se provede převod do normalizovaného tvaru). Index  $a$  v alternativním vztahu značí příslušnou kategorii uvažovanou v rámci variability vstupní proměnné  $m$  (rozlišována na kategorie  $a$ ). Celá situace je pro jednotlivá rozdělení podrobně rozebrána v [18].

Aby náš výklad nebyl příliš obecný, ukazuje obrázek 2 konkrétně zvolené náhodné grupy vstupních parametrů atmosférického a depozičního submodelu. Tabulka 1 popisuje grupu submodelu potravních řetězců a tabulka 2 grupu dozimetrického submodelu. Tyto hodnoty byly použity při analýze radiologického rizika mimořádných úniků radioaktivity [18] pomocí kódu HAVAR RP. Vstupní grupy všech submodelů je třeba adaptovat na dostupná data v ČR.

## 6 Metody vzorkování

Jak již bylo řečeno, složitost problému a nelinearita jsou hlavním důvodem pro použití numerického postupu, který umožňuje dosáhnout realistických výsledků. Proto je nezbytné v prvním kroku zvolit vhodnou metodu vzorkování spočívající v generování  $M$ -tic realizací  $\{z_1^k, z_2^k, z_3^k, \dots, z_M^k\}_{k=1, \dots, K}$  vstupního náhodného vektoru  $Z$ , a to jak s ohledem na dostatečně velký počet realizací  $K$  tak vzhledem ke korektnímu zahrnutí skutečných korelací mezi vstupními parametry.

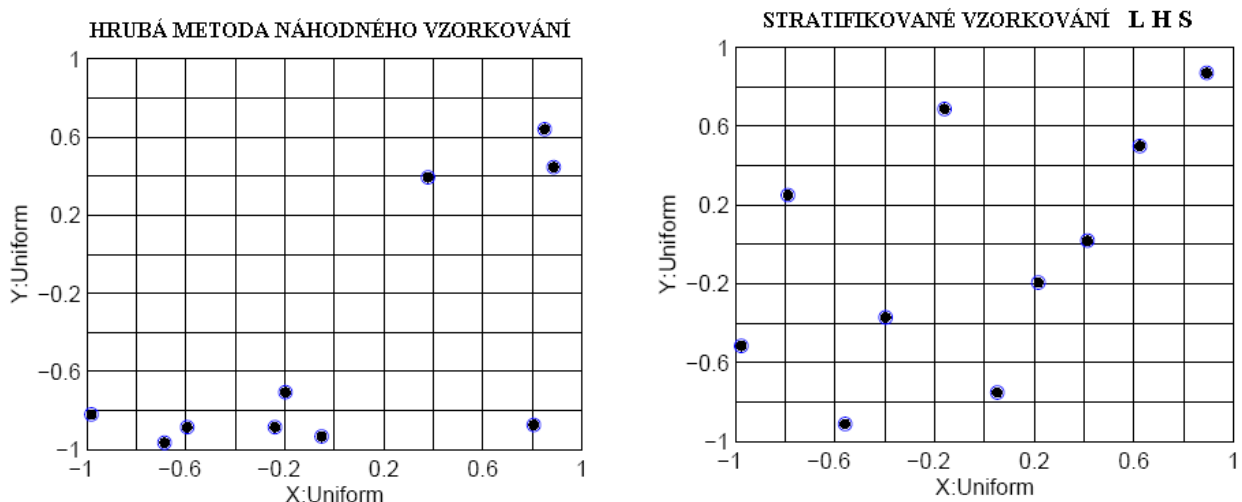
V prvním případě jde o volbu optimálního výběrového algoritmu. Naznačíme hlavní důvod, proč nahrazujeme základní metodu hrubého Monte Carlo vzorkování metodou stratifikovaného vzorkování LHS. I když hrubá metoda se snadno implementuje, její základní nevýhodou je možná tvorba nepravidelných shluků realizací ve výběrovém prostoru, zvláště pak při malém počtu pokusů. To si vynucuje generovat poměrně rozsáhlé vzorky (tisíce až desetitisíce realizací), což v případě následující aplikace ve složitém environmentálním modelu může vést k neúnosným dobám výpočtu resp. při zmenšené velikosti vzorku k nestabilním výsledkům (myšleno ve smyslu srovnání výsledných hodnot s opakovaným procesem). Ale ani při velkém počtu vzorků  $K$  nezaručuje hrubé MC vzorkování úplné pokrytí všech oblastí výběrového prostoru. Je tedy zřejmé, že tato metoda není příliš vhodná například pro určování extrémních hodnot kvantilů výstupních veličin simulovaných



matematickým modelem. Proto bylo třeba přejít k metodám stratifikovaného vzorkování, v našem případě konkrétně k proceduře LHS, která rozvíjí myšlenku „vzorkování na základě důležitosti“ (Importance Sampling).

Zjednodušeně řečeno, metoda LHS předem klade určitá deterministická pravidla na mechanismus generování vzorků tak, aby vykazovaly lepší statistické vlastnosti. Představme si případ, kdy rozsah každé náhodné veličiny  $m$  ( $m=1, \dots, M$ ) je rozdělen na  $K$  disjunkčních intervalů, přičemž intervalové pravděpodobnosti necht' jsou stejné (jejich suma musí být 1.0). Bez dalších podrobností konstatujeme, že takto byl kartézský součin rozsahů všech  $K$  náhodných veličin rozdělen na  $K^M$  částí (hyperkostek), přičemž z nich provádíme výběr  $K$  M-tic, který je uspořádan do výběrové matice  $G^{K,M}$ . A právě pravidla tohoto výběru jsou jádrem procedury LHS, kdy vždy do  $m$ -tého sloupce matice  $G$  (odpovídá  $m$ -té náhodné veličině,  $m=1, M$ ) jsou vybrány prvky tak, aby matice měla v každém tomto sloupci permutaci čísel  $1, \dots, K$ . Tak se zajistí, že každý dílčí interval každé náhodné veličiny bude použit ke generování právě jednou. Výběry potom rovnoměrně pokrývají možné rozsahy náhodných vstupů a s nimi prováděná analýza šíření neurčitostí environmentálním modelem je robustní i pro menší rozsahy vzorků, což je u těchto složitých modelů naprosto podstatné.

K demonstraci principů vzorkování uvádíme obrázek 1, kde je v 2-D uveden případ dvou náhodných proměnných  $X$  a  $Y$  (tedy  $M=2$ ) rovnoměrně rozdělených vždy na intervalu  $\langle -1; 1 \rangle$ , přičemž kroužky znázorňují vygenerované hodnoty deseti dvojic  $(x_k; y_k)$  (tedy počet realizací dvojic je  $K=10$ , stejně jako počet dílčích subintervalů se stejnou pravděpodobností výběru). První obrázek hrubé MC metody ukazuje možné náhodné shluky realizací. Naopak realizace pro LHS vykazují „rovnoměrnější“ zaplnění výběrového prostoru a lze si docela dobře představit, jak při zvětšujícím se počtu realizací  $K$  (a tím i při zhušťování pravidelné čtvercové mřížky) rovnoměrné pokrytí celého výběrového prostoru se nejen zachová, ale i zvýrazní. Obrázek 1 se vztahuje k případu pouze dvou proměnných ( $M=2$ ), a tudíž jde o znázornění v 2-D. Čtenář si jistě dovede představit případ  $M=3$ , kdy výběrový prostor by byl znázorněn krychlí rozdělenou na  $K^3$  základních kostek. Induktivně lze pak pro velká  $M$  vytušit princip rozdělení  $M$ -dimensionální nadkrychle výběrového prostoru na  $K^M$  již zmíněných hyperkostek.



Obr. 1. Náhodné výběry (páry realizací dvou náhodných veličin  $X$  a  $Y$ , každá je rovnoměrně rozdělená na  $\langle -1; +1 \rangle$ ) generované hrubou MC metodou a procedurou Latin Hypercube Sampling ( $K=10, M=2$ )

Zvolenou grupu vstupních parametrů je nutné analyzovat také z hlediska možných vzájemných korelací. Intuitivně lze například vycítit, že fluktuace horizontální a vertikální disperze by mohly vykazovat určitou lineární závislost. Proto součástí expertních odhadů jsou též doporučení na zavádění párových korelací. Na druhé straně je třeba potlačit falešné korelace, které se mohou při procesu generování náhodných vzorků vyskytnout jako důsledek použitého algoritmu (např. při párování náhodných výběrů u metody LHS). V literatuře [12] je nabízeno řešení spočívající v možnosti předepsat algoritmu generování náhodných vstupů předem zkonstruovanou korelační matici podle vztahu (2c). Výsledkem je zachování daných pořadových korelací ve výsledné transformaci. Implementace tohoto postupu do aplikace GENERATOR systému HAVAR-RP je blíže popsána v [5].

Již jsme zmínili, že jednou z cest zvyšování spolehlivosti predikce je zpřesňování odhadu náhodných charakteristik neboli jejich redukovatelnost. Redukci však nelze provádět u přirozených neurčitostí stochastického charakteru inherentně obsažených v systému. Na druhé straně některé neurčitosti modelu a jeho parametrů mohou být až k určité mezi snižovány při použití přesnějších měřících instrumentů a metod, zpřesněném scénáři úniku a při zlepšeních ve formulaci modelu. Poslední zmínka odráží fakt, že v rámci existující koncepce modelu některé neurčitosti jsou neredukovatelné, nicméně mohou být perspektivně snižovány v souladu se zpřesňováním teorie a s rozvojem konceptuálního schématu modelu.

## 7 Návrh implicitních grup neurčitostí vstupních parametrů

Aplikaci předchozích závěrů budeme demonstrovat na simulaci transportu radionuklidů do životního prostředí po jejich mimořádném úniku. Musí být uvažován řetězec jednotlivých vzájemně provázaných submodelů podle schématu:

*únik* → *atmosférický a* → *transport aktivity* → *dozimetrický model*  
*radionuklidů* → *depoziční model* → *potravními řetězci* → *pro odhady dávek*  
 {M1 neurčitostí} {M2 neurčitostí} {M3 neurčitostí}

Zde M1, M2 a M3 představují počet vstupních parametrů, jimž byl přiřazen náhodný charakter. Jedná se o minimální návrh vytipovaný v [18] na základě obsáhlé rešerše podložené publikovanými závěry o důležitosti jednotlivých vstupů. Bohužel je třeba poznamenat, že doporučované závěry se poměrně dost liší a uplatňuje se zde subjektivní názor jednotlivých autorů. Navíc důležitost konkrétního parametru se může podstatně měnit s typem scénářů úniku. Z tohoto úhlu je třeba se dívat i na náš výběr uvedený na obrázku 2 a v tabulkách 2 a 3. Počet neurčitostí je zde volen M1=12, M2=16, M3=9, nicméně vyvinuté softwarové prostředky jsou obecné a umožňují tedy snadné rozšíření. Pro zajímavost uvedme, že analýza neurčitostí (pouze základní statistiky) byla prováděna v systému COSYMA [3] pro M1=24, M2=35 a M3=100 (toto číslo je tak vysoké hlavně proto, že variabilita je zahrnuta jako neurčitosti). Nicméně naše cíle jsou širší (směrem k asimilaci modelu s reálnými měřeními) než je jen základní AN a tudíž i rozsah úlohy musí být optimalizován vzhledem k omezení výpočetní náročnosti.

Počet náhodných veličin :  Záhloví panelu

Počet vygenerovaných n-tic  HAVAR-RP: Aplikace GENERATOR: neurčitosti modelu ADM : demo pro DP - 14.3.2006

Vlastnosti náhodných veličin

škálový faktor intenzity úniku aktivit	nom. hodn. <input type="text" value="1"/>	0.05-kvantil <input type="text" value="1"/>	ořezání <input type="text" value="2,5-sigma"/>
	rozdělení <input type="text" value="ořezané lognormální"/>	0.95-kvantil <input type="text" value="10"/>	počet hodnot <input type="text" value="11"/>
škálový faktor horizont. disperze	nom. hodn. <input type="text" value="1"/>	0.05-kvantil <input type="text" value="1"/>	ořezání <input type="text" value="2-sigma"/>
	rozdělení <input type="text" value="ořezané lognormální"/>	0.95-kvantil <input type="text" value="4"/>	počet hodnot <input type="text" value="11"/>
horizont. flukvace směru větru	nom. hodn. <input type="text" value="1"/>	$\mu$ <input type="text" value="0"/>	ořezání <input type="text" value="3,5-sigma"/>
	rozdělení <input type="text" value="ořezané normální"/>	sigma <input type="text" value="1"/>	počet hodnot <input type="text" value="3"/>
škál.fakt.rychl. suché depoz. - elem	nom. hodn. <input type="text" value="1"/>	levá mez <input type="text" value="2"/>	Standardní zadání <input checked="" type="checkbox"/> Standardní zadání - rel. Kvantily - 0,05 a 0,95 - abs. Kvantily - 0,10 a 0,90 - abs. Kvantily - 0,05 a 0,95 - rel. Kvantily - 0,10 a 0,90 - rel.
	rozdělení <input type="text" value="log-rovnorné"/>	pravá mez <input type="text" value="518"/>	
škál.fakt.rychl. suché depoz. - aer.	nom. hodn. <input type="text" value="1"/>	levá mez <input type="text" value="4"/>	počet hodnot <input type="text" value="11"/>
	rozdělení <input type="text" value="log-rovnorné"/>	pravá mez <input type="text" value="1040"/>	

Nastavení korelací Uložit panel Původní

Vygenerovat a uložit výsledky Načíst panel Zrušit

Obr. 2. Interaktivní zadávání náhodných charakteristik vstupních parametrů atmosférického a depozičního modelu – 1. část. Dalších 7 parametrů se volí na pokračovacím menu, stručně:

6. faktor vymývání - elem. jód: log-rovnorné ; meze standardně relativně: <1; 100>
7. faktor vymývání - aerosoly: log-rovnorné ; meze standardně relativně: <2.3; 600>
8. faktor advekční rychl. vlečky: - rovnorné ; meze standardně relativně: <-1; 1>
9. korekce exponentu profilu větru: normální  $3\sigma$  omez. ;  $N(\mu=1.0, \sigma=0.15)$
10. faktor vertikální disperze: log-normální; zadání kvantily, relativně <5%=1.0, 95%=4.0>
11. korekce směšovací výšky: trojúhelníkové symetrické; meze: <0.5; 1.5>
12. korekce tepel. vzhosu : log-normální; zadání kvantily, relativně <5%=1.0, 95%=10.0>

Tab.1

### Implicitní grupa neurčitostí dynamického ingesčního modelu systému HAVAR-RP

proměnná	ninimum	střední hodnota	naximum	rozdělení	rozměr
FCM1: Intercepce na listové části-frakce		1.0		Normální, $\sigma=0.15$ ( $1\sigma$ omez.)	-
FCM2: Škálový faktor pro $CR_{s-d}$ půda-rostlina Cs i Sr	0.4		2.5	LogUni (relativně)	-
FCM3: Vyluhování – relativní frakce	0.4		2.5	LogUni (relativně)	-
FCM4: Poločas odstranění - weathering	15.0	25.0	35.0	trojúhelníkové	den <sup>-1</sup>
FCM5: Plošná hust. kg/m <sup>2</sup> kořenové zóny(škálově)	0.86	1.0	1.14	rovnorné	-
FCM6: Transf. Cs do mléka ( $d.l^{-1}$ ) – škála fluktuací	0.66		3.00	LogUni (relativně)	-

FCM7: Transf. Šr do mléka ( $d.l^{-1}$ ) – škála fluktuační <sup>(*)</sup>	0.5		1.25	LogUni (relativně)	-
FCM8: Transf. I do mléka ( $d.l^{-1}$ ) – škála fluktuační <sup>(*)</sup>	0.66		3.0	LogUni (relativně)	-
FCM9: Transf. Cs do masa ( $d.kg^{-1}$ ) – škála fluktuační <sup>(*)</sup>	0.50	1.0	1.5	rovnoměrné	-
FCM10: Plošný výnos plodin - korekce	0.8	1.0	1.2	rovnoměrné	-
FCM11: Časový posun vegetace (posun sklizně)	-15	0	+15	rovnoměrné	den
FCM12: Faktor zpracování pro mouku	0.4	0.5	0.6	rovnoměrné	-
FCM13: Faktor zpracování pro zeleninu	0.6	0.8	1.0	rovnoměrné	-
FCM14: Faktory spotřeby potravin	0.7	1.0	1.3	trojúhelníkové	-
FCM15: Krmné dávky skotu – stájový výkrm	0.85	1.0	1.15	rovnoměrné	-
FCM16: Doba zdržení ke konzumaci mléka <sup>(**)</sup>		2.0		normální $\sigma=0.86$ , ( $2\sigma$ mez.)	den

(\*) ...normalizovaný bezrozměrný náhodný faktor podle schématu (3)

(\*\*) ... podle doporučení kódu OSCAAR [4]

Tab. 2

### Implicitní grupa vstupních neurčitostí dosimetrického submodelu systému HAVAR-RP

proměnná	minimum	střední hodnota	maximum	rozdělení	rozměr
DOS1: $F_{lok}^{(1)}$ – lokační faktor (frakce času setrvání v místě)	0.75	0.875	1.00	rovnoměrné	-
DOS2: SF- faktor stínění budov	0.10	0.20	0.30	rovnoměrné	-
DOS3: $d_{fast}$ – depozice Cs137, koeficient formule ENV(t)	0.40	0.52	0.71	rovnoměrné	-
DOS4: $T1/2^{fast}$ – depozice Cs137, poločas odstranění (rychlá složka)	0.41	1.1	1.4	rovnoměrné	rok
DOS5: $T1/2^{slow}$ – depozice Cs137, poločas odstranění (pomalá složka)	24.3	28.0	29.4	rovnoměrné	rok
DOS6: $FF_{filtr}$ – frakce interní filtrace v budovách (cesta inhalace)	0.43	0.64	0.84	rovnoměrné	-
DOS7: $k_{fast}$ - koef. rychlé složky formule koeficientu resuspenze	3.6 E-9	2.63E-8	4.9 E-8	rovnoměrné	$m^{-1}$
DOS8: $T1/2_{RES}^{fast}$ – poločas snížení rychlé složky koef. resuspenze	0.50	1.35	2.2	rovnoměrné	rok
DOS9: $Fu_{inh}^a$ <sup>(2)</sup>	0.5	1.0	1.5	rovnoměrné	-

(1) ... faktor setrvání (odhadnuto z COSYMA, RODOS)

(2) ... multiplikační faktor pro intenzitu dýchání (odhadnuto z MARC-2A)

## 8 Šíření neurčitostí modelem

Pro výběr vhodné metody AN a CS je podstatné, že fyzikální model  $\mathcal{R}^{GPM}$  je silně nelineární. Je možné si sice představit rozvoj výstupní veličiny Y v okolí referenčního (nominálního) vstupního vektoru  $Z^{nom}$  a zavést omezení jen na lineární členy (pro blízké okolí – bližší diskuse je prováděna v [18]), nicméně pro praktické zvládnutí celé úlohy připadá pouze numerické řešení založené na vzorkovacích metodách (sampling-based) spočívajících v mnohonásobně opakovaném modelování transportu znečištění ve sledovaných komponentách environmentálního modelu s jednotlivými realizacemi grupy vstupních parametrů a jejich promítnutí do náhodných charakteristik sledovaných výstupů. Konkrétně, každá jednotlivá realizace  $\underline{k}$  vstupního náhodného vektoru  $\{k_{Z_1}, k_{Z_2}, k_{Z_3}, \dots, k_{Z_M}\}$  je dosazena do modelu (1b) a je proveden samostatný výpočet příslušné realizace náhodného

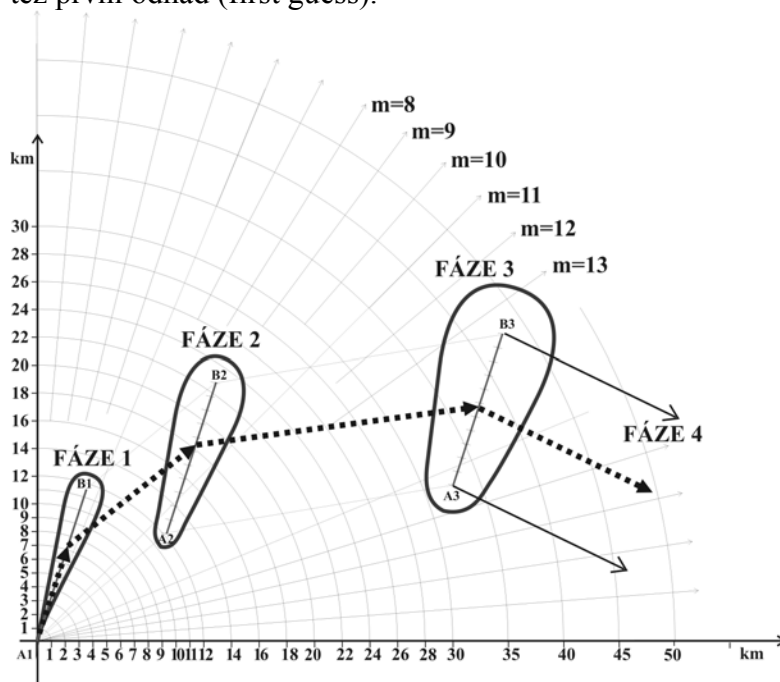
výstupu  $y^k$ . Znovu zdůrazněme, že jsme se sice omezili na jedinou výstupní náhodnou proměnnou  $Y$ , nicméně model vždy simuluje její prostorové rozložení kolem zdroje (podrobněji dále). Mnohonásobným opakováním této sekvence dostáváme páry

$$\{y^k; z_1^k, z_2^k, z_3^k, \dots, z_M^k\}_{k=1,2,3,\dots,K} \quad (4)$$

kteří jsou základem navazující kvantitativní analýzy šíření neurčitostí modelem. Celkový počet realizací  $K$  musí nabývat velkých hodnot, řádově několika tisíc. Nicméně na základě předchozích úvah lze předpokládat, že použití LHS vzorkování bude dávat robustní výsledky i pro menší  $K$  (několik stovek).

Stručně zmíníme numerický algoritmus simulačního modelu. Úloha je řešena numericky na kartézské nebo polární výpočetní mříži s počátkem v místě úniku. V našem případě je pro řešení zvolena polární mříž, kdy prostor kolem zdroje znečištění je rozdělen na 80 pravidelných úhlových segmentů a 35 radiálních (neekvidistantních) pásem (další zjemnění sítě je plánováno). Dynamika úniku s trváním až několik desítek hodin je aproximována ekvivalentními hodinovými segmenty (ve smyslu bilance uvolňované aktivity radionuklidů). Pohyb každého hodinového segmentu je modelován v jeho dalších hodinových fázích s využitím krátkodobých hodinových meteorologických předpovědí. Na obrázku 3 je tato situace znázorněna formou šíření „gaussovské kapky“ (schématicky zobrazující například koncentraci aktivity v ose mraku) v časových okamžicích 1, 2, 3, .. hodin po počátku úniku segmentu (tzn. ve fázích 1, 2, 3, ...). Obdobně si lze snadno představit spojitou stáčeující se stopu vytvořenou suchou popřípadě i mokrou depozicí škodlivin pokud by šlo o modelování depozice na zemském povrchu.

Matematický model provádí simultánní výpočty pro všechny hodinové segmenty úniku a jejich následné fáze posuvů odpovídajících hodinovým meteorologickým předpovědím. Rozložení modelovaných výstupních veličin je potom dáno superpozicí pro všechny segmenty ve všech fázích a je tedy na polární síti vyjádřeno diskrétně vektorem dimenze  $N = 80 \times 35 = 2\,800$ . Tato rozměrnost indikuje náročnost uvažovaného problému. Poznamenejme, že z hlediska asimilačních metod se tomuto řešení říká pole modelu (background field) nebo též první odhad (first guess).



Obr. 3. Schéma k modelování postupu segmentované gaussovské vlečky znečištění nad terénem na polární výpočetní síti kolem zdroje ( 80 úhlových směrů (paprsků); 35 neekvidistančních radiálních pásem do 100 km kolem zdroje). Zobrazeno modelování pohybu hodinového segmentu úniku v jeho dalších hodinových fázích.

Uvažujme jedinou náhodnou závisle proměnnou  $Y$  z řídících proměnných jmenovaných v odstavci 2. Pokud nebude uvedeno jinak, výsledky vztahujeme k depozici určitého radionuklidu na zemském povrchu okolo zdroje. Veličina  $Y$  se modeluje podle obecného schématu (1b). Nicméně při konkrétním výpočtu na polární mříži je odpovídající odhad veličiny  $Y$  modelem vyjádřen diskretními hodnotami v bodech výpočetní mříže tvořících vektor  $\mathbf{x}$  a vztah (1b) pak formálně dává:

$$\mathfrak{R}^{\text{GPM}}(k_{z_1}, k_{z_2}, k_{z_3}, \dots, k_{z_M}) \Rightarrow {}^k\mathbf{x} \quad (5)$$

Zde  $k_{z_1}, k_{z_2}, k_{z_3}, \dots, k_{z_M}$  je jedna (konkrétně  $k$ -tá) realizace grupy vstupních náhodných parametrů Gaussova modelu šíření vlečky škodlivin. Obecně vyjádřeno,  $N$  hodnot odhadu analyzované výstupní veličiny  $Y$  v diskretních bodech výpočtové polární sítě je tedy vyjádřeno vektorem  $\mathbf{x} \equiv [X_1, X_2, \dots, X_N]$  dimenze  $N$ , kde  $N=2800$  odpovídá počtu výpočtových bodů. Poznamenejme, že hodnoty jsou řazeny do vektoru po paprscích, to znamená vždy za sebou po 35ti hodnotách postupně od paprsku 1 do 80.

## 9 Statistické metody zpracování náhodných výstupů modelu

Jedná se o různé statistické metody zpracovávající výstupy mnohonásobného modelování párů dat generovaných do schématu (4).

### 9.1 Analýza neurčitostí (AN)

Každý scénář úniku je popsán specifickými charakteristikami a příslušnými současně se vyskytujícími podmínkami. Proto i v případě pravděpodobnostního přístupu je třeba získané výsledky vztahovat k tomuto scénáři a neprovádět neopodstatněné generalizace. Výstupy z analýzy neurčitostí mohou mít různou formu:

- Vyhodnocování základních výběrových statistik.
- Modelování frekvenčních a distribučních funkcí cílových veličin, vykreslování histogramů a rozptylových diagramů.
- Provádění kvantifikace vlivu neurčitostí pomocí percentilů (5%, 95% apod.) či různě definovaných koeficientů neurčitosti.
- Analýza korelací mezi výsledky a zvolenými vstupními parametry.

Základními výběrovými statistikami jsou například střední hodnota, medián, minimum a maximum a standardní odchylka. Při korelačních testech jsou pro zvolenou závisle proměnnou zkoumány korelace mezi dvěma nebo více vstupními parametry.

Konkrétně v našem případě mnohonásobně opakované modelové odhady  ${}^k\mathbf{x}$  pro  $k=1$  až  $K$  jsou použity jak k odhadu neznámých hledaných skutečných hodnot  $\mathbf{x}^s$  tak ke konstrukci kovarianční matice chyb modelového odhadu. Pro odhad střední hodnoty náhodné veličiny v bodě polární sítě  $\underline{n}$  se používá statistický odhad (výběrový průměr)

$$E[X_n] = \frac{1}{K} \cdot \sum_{k=1}^K {}^k x_n \quad (6)$$

Pro rozptyl v bodě  $\underline{n}$  je použit odhad  $D(X_n)$  podle vztahu:

$$D(X_n) = \frac{1}{K-1} \cdot \sum_{i=1}^K ({}^k x_n - E[X_n])^2 \quad (7)$$

Běžnými vzorci mohou být vyjádřeny momenty vyššího řádu (šikmost, špičatost) a korelační koeficienty. Distribuční funkci CDF (Cumulative Density Function) odhadujeme empirickou distribuční funkcí v bodě  $\underline{n}$ :

$${}^n CDF_K(x) = \frac{1}{K} \sum_{k=1}^K \delta({}^k x_n < x) \quad (8)$$

interpretovanou jako pravděpodobnost nepřekročení hodnoty  $x$ . Většinou se v odhadech ptáme spíše na pravděpodobnost přestoupení určitého limitu  $x_{lim}$ , vyjadřovanou termínem výběrové statistiky CCDF (Complementary Cumulative Density Function) podle vztahu  ${}^n CCDF_K(x_{lim}) = 1 - {}^n CDF_K(x_{lim})$

Pro spojitou a monotónní distribuční funkci  $F(x)$  náhodné veličiny  $X$  se definuje kvantilová funkce  $F^{-1}(\alpha)$ . Pro  $0 < \alpha < 1$  se hodnota  $F^{-1}(\alpha) = x_\alpha$  nazývá  $\alpha$ -kvantilem. Je zřejmá interpretace  $\alpha$ -kvantilu  $x_\alpha$  jako pravděpodobnosti  $\alpha$ , se kterou nebude překročena hodnota  $x_\alpha$ . Používá se též vyjádření pomocí percentilů, kde  $k$ -tý percentil je definován v souladu s vyjádřením  $F^{-1}(k/100)$ . Obvykle mluvíme o  $k$ -tém percentilu, který nebude překročen s pravděpodobností  $k/100$ . V našem případě pro odhad  $\alpha$ -kvantilu  ${}^n x_\alpha$  rozdělení výstupní náhodné veličiny  $X_n$  v uzlu polární sítě  $\underline{n}$  použijeme vztah:

$${}^n x_\alpha = \min \{ {}^k x_n : {}^n CDF_K({}^k x_n) \geq \alpha \}_{(k=1, \dots, K)} \quad (9)$$

Dalšími důležitými statistikami jsou intervaly spolehlivosti (závisající na počtu náhodných realizací  $K$ ), a to jak pro výběrový průměr, tak pro kvantily. K jejich odhadu se užívá aplikace Studentova rozdělení.

## 9.2 Citlivostní studie (CS)

Citlivostní studie se zabývají odhadem důležitosti parciálního vlivu jednotlivých vstupních parametrů (složek vstupního vektoru) na sledované výstupy. Důležitost efektů je v první fázi CS posuzována na základě *třídění a rozptylových diagramů*, v dalších fázích jsou používány přesnější techniky *regresní a korelační analýzy* včetně využití parciálních pořadových korelací. Cílem je identifikovat ty vstupní parametry modelu, které mají nejdůležitější příspěvek ke změnám v důsledcích, a tak určit relevantní grupu vstupních neurčitostí. Obvykle se ukáže, že jen poměrně omezený počet z celkového množství neurčitých parametrů je významný a fluktuace dalších mají zanedbatelný efekt na výstup.

Matematický přístup k CS popisujeme v [5,18] a zde vzhledem k rozsahu nebude uváděn. Důležitou mírou vlivu plynoucí z prováděné regresní analýzy je *koeficient determinace*, který při zkoumání dílčího vlivu vstupního parametru indikuje frakci rozptylu v pozorování závisle proměnné vysvětlovanou regresí. Další mírou jsou *parciální korelační koeficienty*, poskytující odhad lineární závislosti mezi závisle proměnnou a dílčím parametrem vstupu s odstraněním vlivu ostatních parametrů. Obě kritéria jsou použita v interaktivním preprocesoru výstupů systému HAVAR RP a aktivují se stisknutím tlačítka „*citlivost*“ na funkční liště (viz obr. 4).

## 9.3 Výpočet kovarianční matice (KM) chyb modelu

V úvodu článku jsme naznačili dva hlavní důvody, proč je nutné přejít k stochastickým modelům. Prvním z nich je možnost formulovat odpovědi na otázky hodnocení následků na

pravděpodobnostním základě, což mnohem lépe vystihuje charakter úlohy. Druhý důvod spočívá v nezbytnosti zvýšit věrohodnost predikce matematického modelu na základě asimilace s daty měřeními v terénu s tím podstatným faktem, že do procedur objektivní analýzy optimální asimilace musejí vstupovat informace jak o chybách modelu tak o chybách měření. Ponechme zatím stranou rozbor pozorovacích chyb a zaměříme se na chyby modelu, které budou reprezentovány odpovídající kovarianční maticí chyb. Vycházíme z popisu výsledků v diskrétních bodech polární výpočtové sítě podle vztahu (4). Mnohonásobně opakované modelové odhady  ${}^k\mathbf{x}$  pro  $k=1$  až  $K$  jsou použity jak k odhadu neznámých skutečných hodnot  $\mathbf{x}^s$  tak ke konstrukci kovarianční matice chyb modelového odhadu, která je jednou ze základních vstupů do procedur asimilace a určuje prostorové šíření chyb. Vektor chyb modelu dimenze  $N$  se konstruuje podle:

$${}^k\boldsymbol{\varepsilon} = {}^k\mathbf{x} - \mathbf{x}^s; \quad \text{obecně: } \boldsymbol{\varepsilon} = \mathbf{x} - \mathbf{x}^s \quad (10)$$

Složky vektoru  $\boldsymbol{\varepsilon}$  označíme  $e_i$ ,  $i=1, \dots, N$ . Základním statistickým předpokladem je nevychýlenost našich odhadů, tedy  $E(\boldsymbol{\varepsilon}) \equiv \mathbf{0}$ . Bez dalších podrobností uvedeme schéma výpočtu kovarianční matice chyb modelu jako střední (očekávané) hodnoty součinu tohoto vektoru a jeho transpozice, tedy

$$B = \overline{\boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon}^T} = \begin{bmatrix} \overline{e_1 e_1} & \overline{e_1 e_2} & \cdots & \overline{e_1 e_N} \\ \overline{e_2 e_1} & \overline{e_2 e_2} & \cdots & \overline{e_2 e_N} \\ \cdots & \cdots & \cdots & \cdots \\ \overline{e_N e_1} & \overline{e_N e_2} & \cdots & \overline{e_N e_N} \end{bmatrix} \quad (11)$$

K výpočtu očekávaných hodnot prvků  $b_{ij}$  kovarianční matice  $B$  označených horním pruhem lze za určitých předpokladů použít znalost velkého počtu  $K$  provedených realizací podle

$$b_{ij} = \overline{e_i e_j} = \frac{1}{K} \cdot \sum_{k=1}^{k=K} ({}^k e_i \cdot {}^k e_j) \quad (12)$$

## 10 Ukázka aplikace pravděpodobnostního přístupu k hodnocení následků úniku radionuklidů

V předchozích odstavcích je zmíněn způsob výpočtu výběrových charakteristik (vztahy (6, 7, 8) : střední hodnoty, rozptyly, frekvenční a distribuční funkce, komplementární distribuční funkce) modelovaného výstupu z datových bloků (4). Je zřejmé, že znalost průběhu CDF resp. CCDF rozšiřuje dimenzi informace ve srovnání s dřívější jedinou deterministickou (očekávanou) hodnotou.

Hodnocení může probíhat na základě kvantilů (percentilů), jejichž výběrové hodnoty se určují podle vztahu (9). V aktuálních literárních pramenech jsou běžná doporučení uvažovat limitní hodnotu dávek pro kritickou skupinu ve formě 95% percentilu (0.95 kvantil). Tato hodnota je vyšší než odpovídající limit pro deterministický přístup, přičemž je brán v úvahu skutečný charakter úlohy vyplývající z existence variability a neurčitostí. Další možnosti volby kritéria hodnocení na pravděpodobnostním základě mohou být volena s přihlédnutím k rozsahu neurčitosti predikce na základě různě zaváděných faktorů, jako například:

- faktor neurčitosti definovaný poměrem: (95% percentil) / (5% percentil)
- referenční koeficient neurčitosti podle : (95% percentil) / (nominální hodnota)



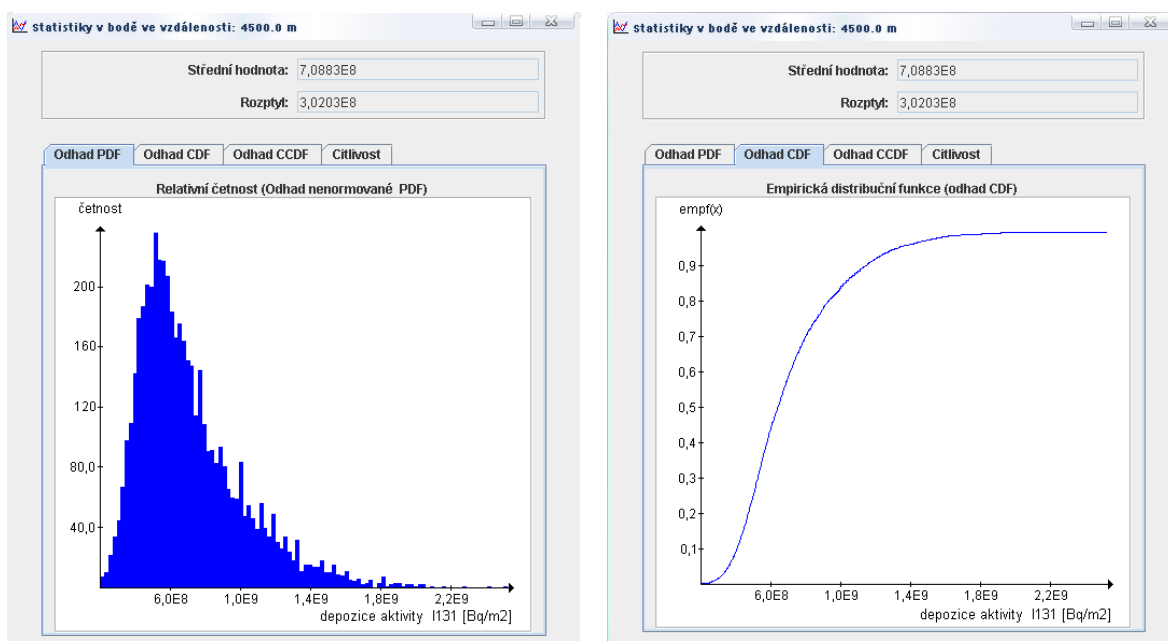
Statistické vyhodnocení výsledků stochastického modelování uloženého podle schématu (4) provádí interaktivní výstupní subsystém produktu HAVAR RP. Jak ukazuje obrázek 4 hlavní charakteristiky náhodného výstupu může uživatel volit přímo z nabídky na liště a informaci dostává v grafické formě.

Celý tok výpočtů realizujících možnost pravděpodobnostního hodnocení je popsán řetězcem

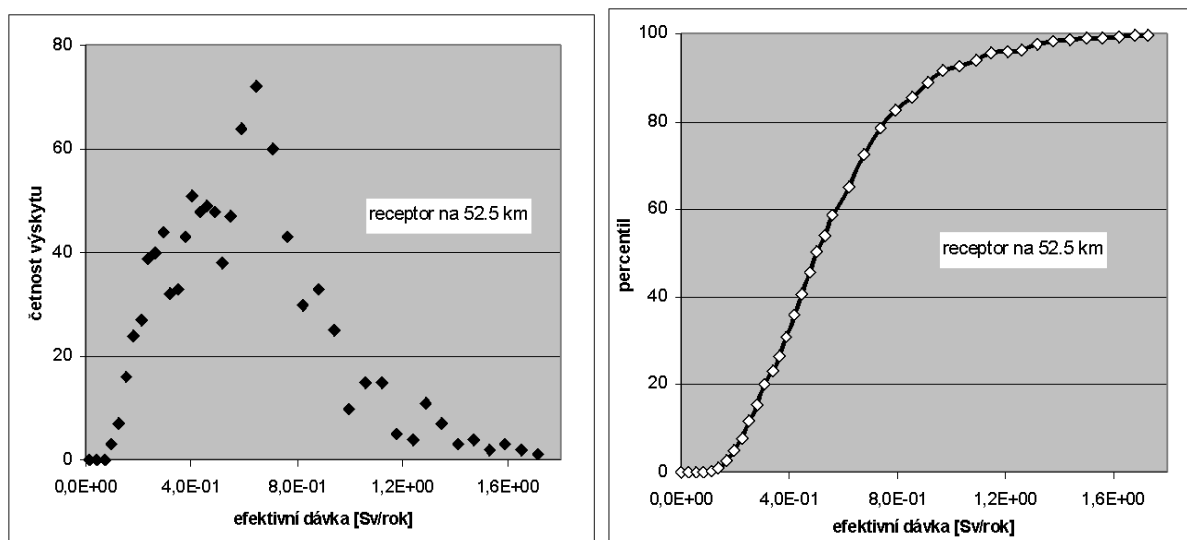
*interakt. preprocesor*                      *mnohonásobné vyvolání*                      *výstupní statistický*  
*grup neurčitostí*                      → *environmentálního modelu*                      → *preprocesor párů (4)*  
*(generování K realizací)*                      *( generování párů podle (4) )*                      *(interaktivní volby)*

Závěrem ukážeme některé výsledky aplikace předchozího řetězce výpočtů na jednoduchém scénáři úniku radionuklidů. Výška úniku byla 45 metrů, meteorologické podmínky byly neměnné po celou dobu úniku (přímočaré šíření), byly použity SCK/CEN formule pro odhad disperzních koeficientů pro kategorii stability atmosféry F. Byly modelovány náhodné hodnoty depozice radionuklidu I131 v důsledku totálního úniku aktivity 2.2 E+16 Bq. Užití statistického preprocesoru pro případ dílčí analýzy neurčitostí submodelu atmosférické disperze a depozice je na obrázku 4. Jedná se o hodnoty pod osou mraku v místě vzdáleném 4.5 km od zdroje úniku. Byla uvažována grupa neurčitostí pro 12 náhodných vstupních parametrů podle obrázku 2, bylo vygenerováno K=1000 (též testy s K=5000) náhodných realizací (dvanácti hodnot).

Náhodné charakteristiky pro roční efektivní dávku (v roce spadu) pro kojence na kilometru 52.5 pod osou mraku jsou znázorněny na obrázku 5. Ingesční příjem je uvažován konzervativně (schéma lokální produkce – lokální spotřeba). Spektrum uniklých radionuklidů je tvořeno třemi nuklidy s celkovými uniklými aktivitami: Sr90: 1.0E+14 Bq; I131: 2.2E+15 Bq; Cs137: 3.3E+16 Bq. V tomto případě šlo o simultánní modelování se zahrnutím náhodných parametrů atmosférického modelu podle obrázku 2 s návazností na transport v potravních řetězcích s 16-ti neurčitostmi podle tabulky 1, vždy pro K=1000. Neurčitosti dosimetrického modelu nebyly uvažovány.



Obr. 4. Statistický preprocesor výsledků z analýzy neurčitosti daných schématem (4)



Obr. 5. Statistické rozdělení roční efektivní dávky pro kojence v místě pod osou mraku ve vzdálenosti 52.5 km od zdroje znečištění

## 11 Témata pro další metodický rozvoj

Shora diskutovaná problematika zahrnuje ve směru ke konkrétní aplikaci celou řadu zjednodušení a různých omezení, která je třeba podrobněji analyzovat, dopracovat, zkompletovat a případně nahradit novými znalostmi. Nicméně dosažené výsledky a vyvinuté programové prostředky umožňují otevřít diskusi o zavedení pravděpodobnostního přístupu k hodnocení následků nehod do regulačních předpisů. Z pohledu metodického rozvoje do popředí vystupují tři závažnější témata, k jejichž řešení by bylo možno nyní přistoupit, konkrétně:

1. Použití **rovníc chaosu** v jejich polynomiálním vyjádření. Při rozvoji přesnosti algoritmu (a tím i složitosti) matematického modelu transportu znečištění se za použití čistě numerických vzorkovacích metod Monte Carlo modelování narazí na progresivní nárůst nutných objemů výpočtu. Problémům vyplývajícím z rostoucí složitosti úlohy by bylo možno se vyhnout aplikací metody SRSM (Stochastic Respond Surface Method) v oblasti AN, která je hybridní metodou analytického přístupu a stratifikovaného vzorkování a je založena na polynomiálním vyjádření chaosu (Wiener, 1947). Vstupní proměnné jsou transformovány do standardizovaného tvaru (obvykle normovaného normálního rozdělení  $N(0,1)$ ) a hledaný výstup se pomocí nich vyjádří ve formě Hermitových polynomů s neznámými koeficienty. Tyto koeficienty se určí na základě regresního modelu z omezeného počtu realizací náhodného výstupu s původními netransformovanými vstupními parametry. V této oblasti semianalytických metod konstrukce „plochy (hyperplochy) odezvy“ by bylo možno navázat na dřívější práce prováděné na pracovišti ÚTIA v metodologicky blízké oblasti modelování odezvy náhodných teplotních polí v palivové kazetě rychlého reaktoru [13, 14].
2. Aplikace **bayesovských metod inference** [1] k aposteriornímu odhadu distribuce výstupu výpočetně složitého deterministického environmentálního modelu, která vychází jak z expertní znalosti náhodných charakteristik vstupů tak z jakýchkoliv dalších existujících údajů o vstupech a výstupech modelu. Právě v oblasti bayesovských odhadů existují v oddělení adaptivních systémů v ÚTIA dlouholeté

zkušenosti, které by mohly být úspěšně využity i v oblasti tvorby moderních softwarových nástrojů pro podporu krizového řízení.

3. Použití statistických metod objektivní analýzy v procesu **asimilace** reálných měření s výsledky modelu. Jak již bylo řečeno v úvodu, jedním z hlavních cílů a záměrem je možnost zpřesnit matematické modelování na základě korekcí se skutečnými měřeními v terénu, prováděných časově konsistentně s časovými kroky matematického modelu. Taková asimilace výsledků modelových odhadů s měřeními představuje nezbytný krok ke zvýšení věrohodnosti matematického modelování. Ze softwarového hlediska je třeba vyvinout systém tvořený dvěma vzájemně spolupracujícími komponentami, a to subsystémem analýzy neurčitostí a asimilačním subsystémem, jehož statistické metody musí při optimálních procedurách brát v úvahu jak chyby modelu tak chyby měření. Vzhledem k tomu, že i v této problematice byly ve spolupráci ÚTIA s FJFI dosaženy první výsledky, autoři předpokládají v dalším článku navázat na zde prezentované téma.

*Poděkování:* Práce byly iniciovány díky podpoře v rámci projektu 6/2003 (poskytovatel SÚJB) s tématem „*Vývoj programového vybavení pro hodnocení radiologických důsledků vážných havárií*“ (2003 – 2005).

## Literatura

- [1] Bates S.C., Cullen A., Raftery A.E.: Bayesian uncertainty assessment in multicompartment deterministic simulation models for environmental risk assessment. *Environmetrics* 2003; 14, pp 355-371.
- [2] Goossens L.H.J., Harper F.T.: Joint EC/USNRC Expert Judgement Driven Radiological Protection Uncertainty Analysis. *J. Radiolog. Prot.*, Vol 18, 1998.
- [3] Goossens L.H.J., Kraan B.C.P., Cooke R.M. (TUD), Jones J.A.(NRPB), Ehrhardt J., Fischer F., Hasemann I., (FZK): Overall Uncertainty Analysis, EUR 18826 EN, 2001.
- [4] Homma T., Inoue Y., Tomita K.: OSCAAR Calculations for the Hanford Dose Reconstruction Scenario of BIOMASS Theme 2. JAERI – Research 2000-049, 2000.
- [5] Housa L. (školitel Pecha P.) : Pravděpodobnostní přístup posuzování závažnosti radioaktivních úniků do atmosféry. Diplomová práce, KM FJFI-softwarové inženýrství, (květen 2006).
- [6] Irwin J.S., Hanna S.R. : Characterizing Uncertainty in Plume Dispersion Models. 9<sup>th</sup> Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Vol.1, Garmisch-Partenkirchen, 1-4 June, 2004.
- [7] Isukapalli S.S. : Uncertainty analysis of transport-transformation models. PhD thesis, Rutgers, The State University of New Jersey, New Brunswick, New Jersey, U.S.A., 1999.
- [8] Janoušek M., Vondráčková H., Pecha P., Hanuš J.: Výzkum a implementace algoritmů pro zpracování meteorologických dat v reálném čase. Zpráva k projektu 6/2002-E04, spolupráce ČHMÚ- ÚTIA AV ČR, 2004, 29 stran, deponováno na SÚJB.
- [9] Jones J.A., Brown J.,(NRPB), Goossens L., Kraan B., Cooke R., (TUD), Ehrhardt J., Hasemann I., Fischer F., (FZK) : Uncertainty Analysis on the Probabilistic Accident Consequence Code COSYMA. *Radiological Prot. Bulletin* No. 232, Dec. 2001.
- [10] Kok Y. S., Eleveld H. : Sensitivity and Uncertainty Analysis of the Atmospheric Dispersion Model NPK-PUFF. HARMON9 - Int. Conf. On Harmonisation within

Atmospheric Dispersion Modelling for Regulatory Purposes, Vol.1, Garmisch-Partenkirchen, 1-4 June, 2004.

[11] Leigh C.D., Thompson B.M., Campbell D.E., Longsine D.E., Kennedy R.A., Napier B.A.: User's Guide for GENII-S : A code for Statistical and Deterministic Simulations of Radiation Doses to Humans from Radionuclides in the Environment. SANDIA REPORT, SAND91-0561, UC-721, 1993

[12] McKay M. D. : Latin Hypercube Sampling as a Tool in Uncertainty Analysis of Computer Models. LAUR 92-2338, 1992.

[13] Pecha P.: Reliability Analysis of Fuel Assemblies Based on Random Temperature Fields Modelling. In: Proceedings 4.International Topical Meeting on Nuclear Reactor Thermal-Hydraulics. (Müller U., Rehme K., Rust K. eds.). G. Braun, Karlsruhe 1989, pp. 1325-1331.

[14] Pecha P.: Statistical treatment of thermal response values in fuel rods region. Jaderná energie, 35 (1989), 9, 332-342.

[15] Pecha P., Kuča P., Pechová E.: Sensitivity study of influence of input parameters variations for removal processes calculations on activity depletion in the radioactive plume and deposition on the ground. In: 7<sup>th</sup> Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling .... EC JRC-EI, Ispra 2001, pp. 489-493.

[16] Pecha P., Pechová E.: Application of the COSYMA code for comparative analysis of a certain accidental releases of radioactivity. In: Proceedings of the 4th International Conference IMUG2002. Brookhaven National Laboratory, Upton 2002, pp. 5-15.

[17] Pecha P., Pechová E.: Modelling of random activity concentration fields in air for purposes of probabilistic estimation of radiological burden. 10<sup>th</sup> Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling. Ioannina, Greece, 2005, pp. 540-544.

[18] Pecha P., Pechová E.: Pravděpodobnostní přístup k predikci důsledků úniku radioaktivity do životního prostředí. Etapa E 02-c projektu 6/2003, 119 stran, deponováno na SÚJB, 2005.

[19] Punčochář M. : Nedaleko nekonečna. Academia, Praha, 2004. ISBN 80-200-1205-6.

[20] Saltelli A., Chan K., Scott E. M. : Sensitivity Analysis. J. Wiley & Sons, LTD, ISBN 0-471- 99892-3, 2000.

[21] Vondráčková H., Janoušek M. (ČHMÚ), Pecha P., Hanuš J. (ÚTIA): Asimilace matematického modelování úniku radionuklidů s aktuálními meteorologickými vstupy. Zpráva k projektu 6/2002-E03-1, spolupráce ČHMÚ- ÚTIA AV ČR, 2003, 139 stran, deponováno na SÚJB.

[22] Young M.: Joint EC/USNRC Uncertainty Study: Dispersion and Deposition Panel. SANDIA NL, (1994).