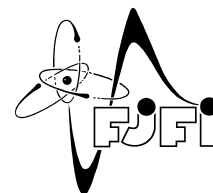


ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE  
Fakulta jaderná a fyzikálně inženýrská



# **Analýza úniku ruthenia-106 v roce 2017 pomocí atmosférického inverzního modelování**

## **Analysis of ruthenium-106 release in 2017 using atmospheric inverse modeling**

Diplomová práce

Autor: **Bc. Tomáš Brisučiak**  
Vedoucí práce: **Ing. Ondřej Tichý, Ph.D.**  
Akademický rok: 2021/2022

## ZADÁNÍ DIPLOMOVÉ PRÁCE

Student:	Bc. Tomáš Brisučiak
Studijní program:	Aplikace přírodních věd
Studijní obor:	Aplikované matematicko-stochastické metody
Název práce (česky):	Analýza úniku ruthenia-106 v roce 2017 pomocí atmosférického inverzního modelování
Název práce (anglicky):	Analysis of ruthenium-106 release in 2017 using atmospheric inverse modeling

### Pokyny pro vypracování:

- 1) Nastudujte a shromážděte dostupné informace o dosud neobjasněném úniku ruthenia-106 na přelomu září/října 2017.
- 2) S využitím dostupných měření a atmosférického transportního modelu sestavte inverzní úlohu.
- 3) Diskutujte možnost zahrnutí apriorní informace do sestavené inverzní úlohy, např. informaci o neurčitostech měření nebo informaci o prostorové a časové distribuci měření.
- 4) Na základě zvolené apriorní informace sestavte pravděpodobnostní model a odvoďte rovnice pro odhad jeho parametrů.
- 5) Dosažené výsledky validujte vzhledem k měřením a vzhledem k doposud publikovaným odhadům zdrojového členu.

Doporučená literatura:

- 1) O. Masson, et al., Airborne concentrations and chemical considerations of radioactive ruthenium from an undeclared major nuclear release in 2017. Proceedings of the National Academy of Sciences 116(34), 2019, 16750–16759.
- 2) P. Seibert, A. Frank, Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode. Atmospheric Chemistry and Physics 4(1), 2004, 51–63.
- 3) O. Tichý, V. Šmídl, R. Hofman, A. Stohl, LS-APC v1.0: a tuning-free method for the linear inverse problem and its application to source-term determination. Geoscientific Model Development 9, 2016, 4297–4311.
- 4) J.D. Le Brazidec, et al., MCMC methods applied to the reconstruction of the autumn 2017 Ruthenium 106 atmospheric contamination source. Atmospheric Environment: X (6), 2020, 100071.
- 5) O. Saunier, et al., Atmospheric modeling and source reconstruction of radioactive ruthenium from an undeclared major release in 2017. Proceedings of the National Academy of Sciences 116(50), 2019, 24991–25000.
- 6) L.M. Western, et al., Source estimation of an unexpected release of Ruthenium-106 in 2017 using an inverse modelling approach. Journal of Environmental Radioactivity 220–221, 2020, 106304.

Jméno a pracoviště vedoucího diplomové práce:

Ing. Ondřej Tichý, Ph.D.

Ústav teorie informace a automatizace, v.v.i., Pod Vodárenskou věží 4, 182 08, Praha 8

Jméno a pracoviště konzultanta:

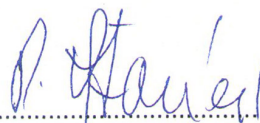
Datum zadání diplomové práce: 31.10.2020

Datum odevzdání diplomové práce: 3.5.2021

Doba platnosti zadání je dva roky od data zadání.

V Praze dne 26. října 2020

  
.....  
garant oboru

  
.....  
vedoucí katedry



  
.....  
děkan

*Poděkování:*

Rád bych poděkoval vedoucímu práce Ing. Ondřeji Tichému, Ph.D. za skvělé vedení a za neocenitelné rady a pomoc při tvorbě této práce.

Děkuji také rodičům za podporu během studia a tvorby práce.

*Čestné prohlášení:*

Prohlašuji, že jsem tuto práci vypracoval samostatně a uvedl jsem všechnu použitou literaturu.

V Praze dne 5. ledna 2022

Bc. Tomáš Brisučiak

*Název práce:*

**Analýza úniku ruthenia-106 v roce 2017 pomocí atmosférického inverzního modelování**

*Autor:* Bc. Tomáš Brisučiak

*Obor:* Aplikované matematicko-stochastické metody

*Druh práce:* Diplomová práce

*Vedoucí práce:* Ing. Ondřej Tichý, Ph.D., Ústav teorie informace a automatizace, v.v.i., Pod Vodárenskou věží 4, 182 08, Praha 8

*Abstrakt:* Práce se zabývá odhadem časového průběhu radiačního úniku ruthenia-106 do atmosféry v roce 2017 pomocí naměřených koncentrací. V teoretické části jsou uvedeny základy Bayesovské teorie pravděpodobnosti včetně možností aproximace odhadů s důrazem na metodu Variační Bayes. Metoda Variační Bayes je poté použita na zformulovaný pravděpodobnostní model a je provedeno odvození již existujícího algoritmu LS-APC. Ten je následně vylepšen o možnost uvažovat také korelaci jednotlivých složek v modelu šumu. Vylepšený algoritmus je verifikován na syntetických datech a datech z experimentu ETEX se známým průběhem úniku. V závěru práce je proveden samotný odhad časového průběhu úniku ruthenia-106. Získané výsledky jsou konzistentní s dosud publikovanou literaturou.

*Klíčová slova:* atmosférický únik, bayesovské metody, pravděpodobnostní model, řídká lineární regrese

*Title:*

**Analysis of ruthenium-106 release in 2017 using atmospheric inverse modeling**

*Author:* Bc. Tomáš Brisučiak

*Abstract:* We study the ruthenium-106 atmospheric release case from 2017 and its emission rate estimation from concentrations measurements. The basics of bayesian probability theory are covered in the theoretical part including possible approximations of estimation with emphasis on the Variational Bayes method. The Variational Bayes method is used to estimate parameters of the formulated probability model, and the derivation of an existing inversion algorithm (LS-APC) is carried out. This algorithm is then improved to allow inclusion of component correlation in the noise model. The improved algorithm is verified using synthetic data as well as data from the ETEX experiment with known emission profile. Finally, the estimation of the ruthenium-106 release is performed. Notably, obtained estimates are consistent with those published in literature.

*Keywords:* atmospheric release, bayesian inference, probability model, sparse linear regression

# Obsah

<b>Úvod</b>	<b>8</b>
<b>1 Inverzní modelování</b>	<b>11</b>
1.1 Lineární inverzní modelování . . . . .	11
1.2 Atmosférické transportní modely . . . . .	12
1.2.1 Gaussovský model . . . . .	12
1.2.2 Eulerovský model . . . . .	12
1.2.3 Lagrangeovský částicový model . . . . .	13
1.3 Řešení inverzní úlohy . . . . .	14
1.3.1 Optimalizační přístup . . . . .	14
1.3.2 Pravděpodobnostní přístup . . . . .	15
<b>2 Bayesovské odhady a jejich aproximace</b>	<b>16</b>
2.1 Základy Bayesovské teorie . . . . .	16
2.1.1 Volba apriorního rozdělení . . . . .	17
2.2 Aproximace . . . . .	17
2.2.1 MAP odhad . . . . .	18
2.2.2 Expectation-maximization (EM) algoritmus . . . . .	18
2.2.3 Laplaceova aproximace . . . . .	18
2.2.4 Markov Chain Monte Carlo . . . . .	19
2.2.5 Variační Bayesova aproximace . . . . .	19
<b>3 Odvození použitých algoritmů</b>	<b>22</b>
3.1 LS-APC . . . . .	22
3.1.1 Pravděpodobnostní model . . . . .	22
3.1.2 Odvození algoritmu pomocí VB aproximace . . . . .	24
3.2 Modifikace pro model s plně rozvolněným rozptylem šumu . . . . .	27
3.3 Modifikace pro model se rozptylem šumu podle kategorií . . . . .	27
3.4 Modifikace pro model s korelovaným šumem . . . . .	28
<b>4 Výsledky</b>	<b>30</b>
4.1 Syntetická data . . . . .	31
4.1.1 Vliv lokalizace na stabilitu . . . . .	31
4.1.2 Vliv lokalizace na kvalitu odhadu . . . . .	31
4.2 Data z experimentu ETEX . . . . .	36

4.2.1	Volba lokalizační masky podle vzdálenosti měřicích stanic . . . . .	37
4.2.2	Nebinární lokalizační maska . . . . .	37
4.2.3	Volba lokalizační masky podle času měření . . . . .	41
4.3	Únik ruthenia-106 v roce 2017 . . . . .	44
4.3.1	Dataset IAEA . . . . .	44
4.3.2	Dataset Ro5 . . . . .	52
4.3.3	Shrnutí a srovnání výsledků . . . . .	59
4.3.4	Shrnutí výsledků z dostupné literatury . . . . .	60
<b>Závěr</b>		<b>64</b>
<b>A Použitá rozdělení pravděpodobnosti</b>		<b>70</b>
A.1	Vícerozměrné normální rozdělení . . . . .	70
A.2	Ořezané normální rozdělení . . . . .	70
A.3	Ořezané vícerozměrné normální rozdělení . . . . .	71
A.4	Gamma rozdělení . . . . .	71
A.5	Wishartovo rozdělení . . . . .	71

# Úvod

Koncem září a začátkem října 2017 byly v Evropě poprvé od Černobylské havárie pozorovány detekovatelné koncentrace ruthenia-106[22, 18, 5]. Naměřené koncentrace se pohybovaly od jednotek  $\mu\text{Bq}/\text{m}^3$  až kolem  $150 \text{ mBq}/\text{m}^3$ . Tato množství sice pro člověka ani prostředí nepředstavují žádné nebezpečí, ale stále se jedná o radionuklid, který nevzniká přirozeně a ve vzduchu se normálně nevyskytuje. Navíc detekce ve většině evropských států znamená, že nemohlo jít o zanedbatelný únik.

Všechny evropské země mají síť měřících stanic monitorujících jak přirozenou, tak uměle vzniklou radioaktivitu v prostředí, zejména atmosféře. Ve většině případů jsou tato měření prováděna vzduchovými filtry, kterými projdou desítky až stovky  $\text{m}^3/\text{h}$ . Tyto filtry jsou měněny ručně v pravidelných intervalech (typicky denně až týdně) a posílány do laboratoří, kde je určena aktivita ruthenia-106 gama spektroskopii. V některých případech je intenzita ruthenia měřena ze vzorků dešťové vody ( $\text{Bq}/\text{l}$ ), trávy ( $\text{Bq}/\text{kg}$ ) nebo výtěry z dané oblasti ( $\text{Bq}/\text{m}^2$ ).

Kvůli značným rozdílům v délce získávání vzorků není možné přímo srovnat naměřené intenzity v různých zemích. Mapa na obrázku 1 převzatá ze zprávy francouzského Institutu pro ochranu před radiací a nukleární bezpečnost IRSN (Institut de Radioprotection et de Sûreté Nucléaire)[18] se snaží o korekci a měření pro každou stanici průměruje přes dobu, po kterou bylo ruthenium-106 v této stanici pozorováno.

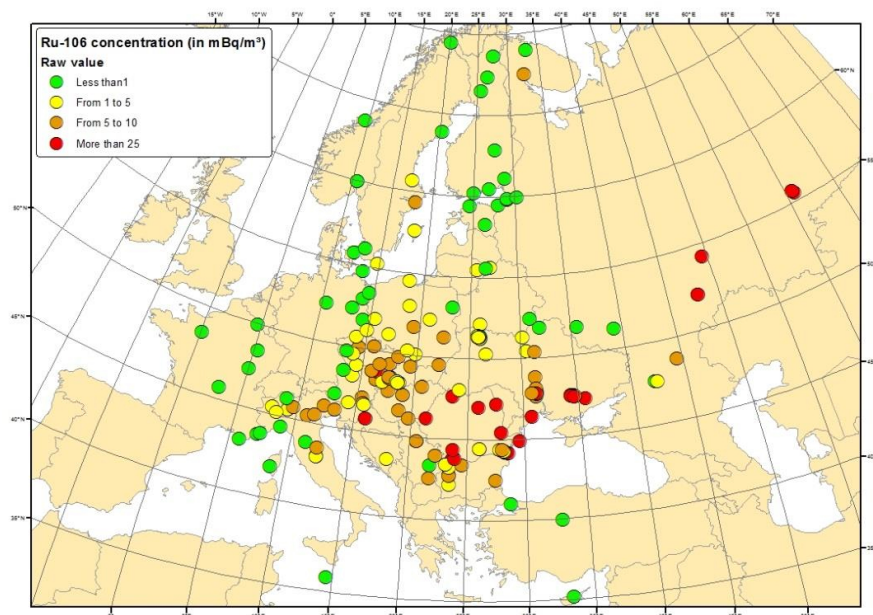
Na mapě je patrný klesající trend z východu na západ. Největší intenzity sice byly naměřeny v Rumunsku, ale detekce ruthenia-106 v několika geograficky vzdálených stanicích zároveň únik v Rumunsku vylučuje. Zdroj úniku musel být mnohem vzdálenější, aby se mrak s rutheniem-106 stihl takto rozprostřít.

Už ve zprávě IRSN[18] se jeví jako pravděpodobný únik z oblasti jižního Uralu v Ruské federaci. Od té doby bylo publikováno mnoho pokusů o objasnění tohoto úniku různými metodami. Tyto výsledky budou diskutovány v sekci 4.3.4. Veskrze se shodují na úniku z oblasti jižního Uralu a jako pravděpodobný zdroj označují kombinát na zpracování radioaktivních materiálů Majak nacházející se v Čeljabinské oblasti Ruské federace, viz mapa na obrázku 2.

Literatura se shoduje, že únik nesouvisí s produkcí plutonia pro vojenské účely, ale s civilním využitím pro přepracování jaderného paliva[15, 7]. Dokonce je přímo identifikován reaktor typu VVER-440[15], jehož palivo se rutinně zpracovává v kombinátu Majak.

Ruská meteorologická služba Roshydromet přiznává zvýšené koncentrace ruthenia-106 v této oblasti na konci září 2017, vylučuje však možnost, že by zdrojem byl právě kombinát Majak[22]. Ruská akademie věd svolala mezinárodní komisi, jejíž oficiální stanovisko znělo, že neexistuje dostatek důkazů pro žádnou z navržených teorií.





Obrázek 1: Mapa koncentrací ruthenia-106 po provedení korekcí. Převzato z [18].



Obrázek 2: Poloha kombinátu Majak.

Cílem této práce bude se pokusit tento únik objasnit a určit časový průběh úniku. Odhadem lokace se tato práce nebude zabývat, protože považujeme za dostatečně prokázané, že k úniku došlo v kombinátu Majak nebo alespoň oblasti jižního Uralu[18, 22, 12, 31, 27, 29, 35]. Odhadem lokace jsme se také již dostatečně zabývali ve výzkumném úkolu. Tento problém odhadu časového průběhu úniku zformulujeme jako inverzní úlohu, kterou budeme řešit Bayesovským přístupem. Ukáže se však nutnost použití aproximací, které proto prozkoumáme se zvláštním důrazem na metodu Variační Bayes. Právě na této aproximaci je založen již existující algoritmus LS-APC, který se pokusíme modifikovat pro možnost modelovat i korelace mezi složkami šumu na základě např. vzdálenosti měřících stanic. Modifikovaný algoritmus se pokusíme použít na odhad časového průběhu úniku ruthenia-106 v roce 2017.

# Kapitola 1

## Inverzní modelování

V této kapitole popíšeme metodologii tvorby tzv. source-receptor sensitivity (SRS) matice a provedeme formulaci inverzní úlohy. Také popíšeme základní typy atmosférických transportních modelů. Nakonec ukážeme možné přístupy k řešení inverzní úlohy.

### 1.1 Lineární inverzní modelování

Při formulaci inverzní úlohy budeme postupovat podle [28]. Tzv. source-receptor (s-r) vztah mezi členem zdroje  $x$  a receptoru  $y$  popisuje citlivost receptoru na zdroj. Tento koncept lze aplikovat na různé typy zdrojů a receptorů. Receptor může být například průměrná koncentrace nějaké látky v dané oblasti a čase. Nebo to může být měření v konkrétní měřicí stanici. Zdroj může být bodový, oblast nebo daný objem, který je aktivní v nějakém čase.

Tento s-r vztah je obecně nelineární  $\frac{\partial y}{\partial x}$ . Nás však bude zajímat lineární případ hlavně z důvodu, že lagrangeovský transportní model nedokáže simulovat nelineární chemické reakce. To však nepředstavuje velký problém, protože námi studované procesy během transportu vzduchem jsou lineární. V lineárním případě přechází s-r vztah na jednoduchý poměr  $\frac{y}{x}$ .

V případě úniku do atmosféry typicky nemáme pouze jeden receptor a zdrojů také může být více. Předpokládáme-li únik  $x_i$  v čase  $i$ , lze pomocí atmosférických transportních modelů vypočítat koncentraci  $y_j$  v čase  $j$ . Z nich můžeme definovat tzv. SRS (source-receptor sensitivity) matici  $M$  s prvky

$$m_{ij} = \frac{y_j}{x_i}. \quad (1.1)$$

Výpočet SRS matice  $M$  je sice výpočetně velmi náročný, ale stačí jej provést pouze jednou. Poté již nemusíme pokaždé spouštět atmosférický transportní model, ale stačí provést jednoduché maticové násobení.

Pokud uspořádáme zdroje ve všech možných časech úniku do vektoru  $x \in \mathbb{R}^{n \times 1}$  a všechna měření do vektoru  $y \in \mathbb{R}^{p \times 1}$ , získáme lineární model

$$y = Mx + e, \quad (1.2)$$

kde  $\mathbf{y} \in \mathbb{R}^{p \times 1}$  je vektor naměřených dat,  $\mathbf{x} \in \mathbb{R}^{n \times 1}$  neznámý zdrojový člen a  $\mathbf{M} \in \mathbb{R}^{p \times n}$  je SRS matice. Člen  $e \in \mathbb{R}^{p \times 1}$  představuje šum vzniklý z chyb jak měření, tak samotného transportního modelu.

V této práci budeme SRS matice získávat ze zpětného chodu lagrangeovského atmosférického transportního modelu HYSPLIT[11].

## 1.2 Atmosférické transportní modely

Radionuklidy se z místa úniku šíří různými cestami, ale vzduchem zasáhnou největší oblast a šíří se nejrychleji. Proto je důležité předpovědět disperzi radionuklidů ve vzduchu. Mrak radionuklidů se hýbe s větrem, dále se oblak promíchává turbulentní difúzí a probíhá v něm mnoho dalších procesů jako např. absorpce do okolí, radioaktivní rozpad, chemické reakce a jiné změny. Tyto procesy popisuje rovnice atmosférického šíření[20]

$$\frac{\partial c}{\partial t} = -\mathbf{v}\nabla c + \nabla\mathbf{K}\nabla c + E + R + D, \quad (1.3)$$

kde  $c$  je koncentrace,  $\mathbf{v}$  vektor větru,  $\mathbf{K}$  matice koeficientů turbulentní difúze, členy  $E$ ,  $R$ ,  $D$  popisují změnu koncentrace způsobenou zářením, chemickými reakcemi a absorpcí.

Největší roli při výběru vhodného modelu hraje velikost uvažovaného prostoru. Na lokální úrovni (1–10 km) rovnici zjednoduší předpoklad homogenního a stacionárního pole větru. Na regionální úrovni (100–1000 km) už je třeba započítat měnící se meteorologickou situaci. Na ještě větších vzdálenostech již jsou nutné kompromisy mezi přesností a výpočetní náročností. Překvapivě náročné jsou i simulace na malé úrovni (10–1000 m) kvůli nutnosti detailní simulace kolem budov a jiných překážek.

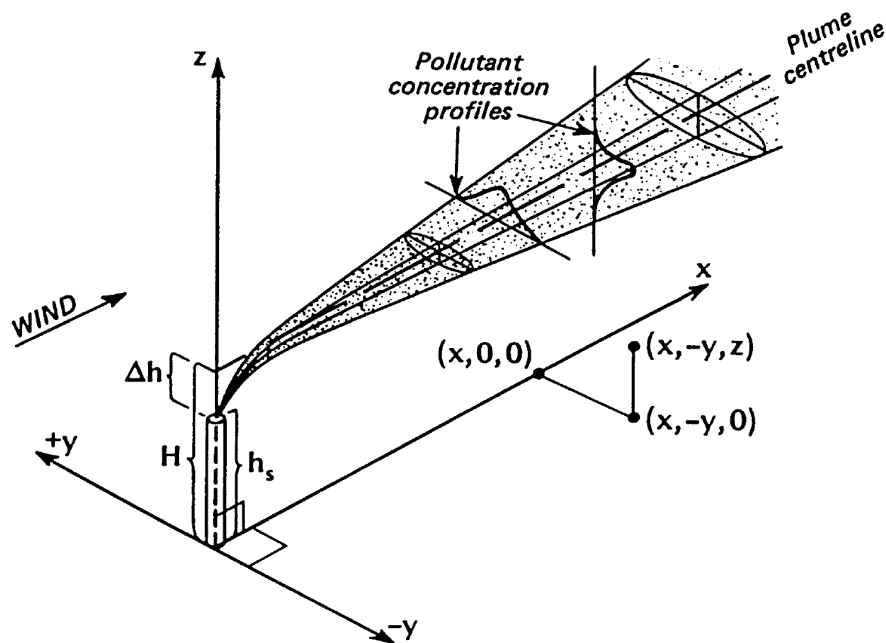
Složitější modely jsou také potřeba v případech složitého terénu, rychle se měnícího počasí nebo pokud vypuštěný materiál ovlivňuje tok vzduchu. Na základě přístupu k řešení rovnice (1.3) lze modely rozdělit na gaussovské, eulerovské a lagrangeovské.

### 1.2.1 Gaussovský model

Analytické řešení[10] rovnice (1.3) je možné za předpokladu jediného zdroje a stálého homogenního větru s neměnnou intenzitou turbulence. Tím získáme pole normálně rozdělených koncentrací ve směru větru, viz obrázek 1.1. Protože pro vyhodnocení je potřeba pouze vyčíslit získané analytické řešení pro každý receptor místo numerického řešení diferenciálních rovnic, je použití gaussovských modelů velice rychlé a zvládne je i obyčejný osobní počítač. Díky předpokladu jediného zdroje také nepotřebují mnoho vstupních dat. Na lokální úrovni (1–10 km) fungují dobře, selhávají ale zejména při nízkých rychlostech větru a složitém terénu. Na delší vzdálenosti velmi rychle ztrácí na přesnosti.

### 1.2.2 Eulerovský model

Přesnější, ale výpočetně mnohem náročnější, řešení[10] rovnice (1.3) získáme budeme-li na ni nahlížet jako na systém parciálních diferenciálních rovnic druhého řádu tří prostorových a jedné časové souřadnice s příslušnými počátečními podmínkami. Tyto rovnice obecně nelze řešit analyticky.



Obrázek 1.1: Schématický obrázek gaussovského oblaku. Převzato z [4].

Obecně se metody na řešení parciálních diferenciálních rovnic skládají ze dvou kroků. Prvním je prostorová diskretizace (většinou na mřížce), kdy se parciální diferenciální rovnice převede na systém obyčejných diferenciálních rovnic, kde každá závisí pouze na jediné časové proměnné. Druhý krok je samotné vyřešení získaných obyčejných diferenciálních rovnic s počáteční podmínkou, kde existuje dobře prozkoumaný matematický aparát.

Tento přístup lze dále vylepšit proměnlivou diskretizací, kdy je možné použít jemnější síť v oblastech, kde dochází k velké numerické chybě. V každém případě se jedná o výpočetně velmi náročné řešení.

### 1.2.3 Lagrangeovský částicový model

Rovnici (1.3) lze řešit také stochasticky[10]. Celková hmota nebo intenzita uniklé látky je rozdělena na malé „částice“, kterými můžou být myšleny opravdu částice aerosolu nebo dále nedělitelné části objemu vypuštěné látky. Lagrangeovský částicový model pak simuluje trajektorii mnoha takových částic řešením obyčejných diferenciálních rovnic prvního řádu. Rychlost částice se pak skládá z rychlosti pohybu prostředí (nejčastěji vzduchu) získané z meteorologických dat a z turbulentní difúze, která je modelována náhodnou procházkou.

Výpočetní náročnost závisí na počtu simulovaných částic, ale už malé množství trajektorií dává představu o směru pohybu látky. Výhodou modelu pro inverzní úlohu je možnost spuštění ve zpětném chodu, kdy částice vypouštíme z měřicích stanic.

## 1.3 Řešení inverzní úlohy

Inverzní úlohou nazveme odhad zdrojového členu  $x$  když známe naměřená data  $y$  a předpokládáme platnost vztahu (3.1). Jedná se tedy o problém lineární regrese, která se standardně řeší metodou nejmenších čtverců, kterou bychom získali odhad ve tvaru

$$x_{LS} = (M^T M)^{-1} M^T y. \quad (1.4)$$

Takový postup je ovšem nevhodný pokud je matice  $M$  špatně podmíněná, což v kontextu práce bude téměř vždy. Navíc do takto získaného odhadu nemáme jak zanést následující požadavky vyplývající z fyzikální podstaty úlohy:

- Pozitivita – složky zdrojového členu  $x$  požadujeme nezáporné, protože radiace uniká do vzduchu a ne naopak.
- Řídkost – radiační únik není normální jev, proto pokud data neposkytují dostatečnou informaci o opaku, preferujeme nulové složky zdrojového členu.
- Hladkost – prudké změny složek zdrojového členu jsou sice možné, ale ne příliš časté.

Všechny uvedené problémy vyřešíme regularizací – přidáním další informace, která dále omezí možná řešení. Pro řešení celé inverzní úlohy pak existují dva hlavní přístupy a to optimalizační a pravděpodobnostní [16]. Oba přístupy spočívají v nalezení takových parametrů úniku, aby se naměřená data  $y$  lišila od jejich rekonstrukce  $Mx$  co nejméně.

### 1.3.1 Optimalizační přístup

Optimalizační přístup spočívá v nalezení parametrů úniku, které minimalizují zvolenou ztrátovou funkci  $J$  [16]. Ta může být různá, většinou se však jedná o nějakou normu rozdílu naměřených dat a jejich rekonstrukce. Předpokládá se, že parametry, které tuto ztrátovou funkci minimalizují, poté tvoří také optimální odhad zdrojového členu. Většina optimalizačních technik spočívá v iterativním procesu, kde v každém kroku je ztrátová funkce minimalizována různými pravidly pro změnu parametrů a jsou získány vylepšené parametry odhadu.

Vyjděme tedy z modelu (3.1) a jako ztrátovou funkci uvažujme kvadratickou normu jeho rezidua

$$J = \omega_0^{-1} (y - Mx)^T (y - Mx) \quad (1.5)$$

s hyperparametrem  $\omega_0$ . Přímá minimalizace této ztrátové funkce by znovu vedla na odhad ve smyslu nejmenších čtverců (1.5), který jsme již zamítli. Proto do ztrátové funkce přidáme regularizační člen  $g(x)$ , který bude penalizovat velká  $x$ . Do tohoto členu se navíc může začlenit apriorní informace jak bylo v kontextu úniků do atmosféry provedeno např. v článku [13].

Ztrátová funkce pak má tvar

$$J = \omega_0^{-2} (y - Mx)^T (y - Mx) + \alpha g(x), \quad (1.6)$$

kde  $\alpha > 0$  je další hyperparametr úlohy. Regularizační člen  $g(x)$  lze volit různě [3]. Nejčastější volbou je tzv. ridge (Tichonova) regularizace, kdy  $g(x) = \|x\|_2 = \sum x_i^2$ . Pro nás zajímavější je

volba  $g(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum |x_i|$  zvaná LASSO (Least Absolute Shrinkage and Selection Operator), která vede na řídké řešení[3].

Hlavní nevýhodou optimalizačního přístupu je nutnost ručního ladění hyperparametrů, kdy sice v tomto příkladu máme pouze dva, ale ve složitějších ztrátových funkcích jejich počet narůstá.

### 1.3.2 Pravděpodobnostní přístup

Ekvivalentně lze minimalizaci ztrátové funkce formulovat jako maximalizaci pravděpodobnosti

$$f(\mathbf{y} | \mathbf{x}) = \mathcal{N}(\mathbf{M}\mathbf{x}, \omega_0^{-1} \mathbf{I}_p) \propto \exp\left(-\frac{1}{2} \omega_0 (\mathbf{y} - \mathbf{M}\mathbf{x})^\top (\mathbf{y} - \mathbf{M}\mathbf{x})\right), \quad (1.7)$$

kde  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  je vícerozměrné normální rozdělení se středem  $\boldsymbol{\mu}$  a kovarianční maticí  $\boldsymbol{\Sigma}$  (viz příloha A),  $\mathbf{I}_p$  je jednotková matice rozměru  $p \times p$  a symbol  $\propto$  značí proporcionálně rovno, tedy rovno až na konstantu.

Toto pravděpodobnostní rozdělení doplníme o námi zvolené apriorní rozdělení parametrů a aplikací Bayesovy věty získáme aposteriorní rozdělení všech parametrů. Volbou apriorních rozdělení můžeme přirozeně začlenit nejistotu měření a modelu a také požadované předpoklady. Výhodou je, že nezískáváme pouze bodový odhad, ale rovnou celé jeho pravděpodobnostní rozdělení. Tento postup je podrobněji rozebrán v kapitolách 2 a 3.

# Kapitola 2

## Bayesovské odhady a jejich aproximace

V této kapitole shrneme základní poznatky Bayesovské teorie pravděpodobnosti [1, 14], která nám umožňuje pracovat s nejistotou a zahrnout naše předpoklady, které vychází z fyzikální podstaty úlohy. Získané odhady však mnohdy nelze vyjádřit analyticky a numerické řešení není v takto vysokých dimenzích vhodné. Přistoupíme tedy k možnostem aproximace a zvláštní důraz budeme klást na metodu Variční Bayes [30], která se již v bakalářské práci [6] osvědčila na odhad reálného úniku. Známější a používanější metoda Markov Chain Monte Carlo poskytovala ještě přesnější odhady, ale za cenu přílišné výpočetní náročnosti.

### 2.1 Základy Bayesovské teorie

Máme k dispozici naměřená data  $\mathbf{y}$  a chceme se dozvědět co nejvíce o systému nebo zdroji, ze kterého vznikla. Necht' vektor  $\boldsymbol{\theta}$  je sada parametrů, které popisují proces vzniku dat  $\mathbf{y}$ . Místo deterministické závislosti  $\mathbf{y} = g(\boldsymbol{\theta})$  pracujeme v Bayesovské teorii s pravděpodobnostmi k vyjádření našich předpokladů.

Hustotou pravděpodobnosti náhodného spojitého vektoru  $\boldsymbol{\theta}$  z prostoru parametrů  $\Theta$  nazýváme  $f(\boldsymbol{\theta})$  splňující

$$f(\boldsymbol{\theta}) \geq 0 \quad \forall \boldsymbol{\theta}, \quad (2.1)$$

$$\int_{\Theta} f(\boldsymbol{\theta}) d\boldsymbol{\theta} = 1. \quad (2.2)$$

Data  $\mathbf{y}$  můžeme vysvětlit parametrickým pravděpodobnostním *modelem pozorování*  $f(\mathbf{y} | \boldsymbol{\theta})$ . Kompletní model dat získáme přidáním apriorní hustoty pravděpodobnosti  $f(\boldsymbol{\theta})$ , která vyjadřuje co víme nebo alespoň předpokládáme o parametrech  $\boldsymbol{\theta}$ .

V praxi však řešíme úlohu, kdy máme k dispozici data  $\mathbf{y}$ , která chceme použít k zjištění neznámých parametrů  $\boldsymbol{\theta}$  jejich zdroje. Tato úloha je řešena sestavením aposteriorní hustoty  $f(\boldsymbol{\theta} | \mathbf{y})$ . Bayesova věta nám lehce umožní provést toto prohození podmíněnosti

$$f(\boldsymbol{\theta} | \mathbf{y}) = \frac{f(\boldsymbol{\theta}, \mathbf{y})}{f(\mathbf{y})} = \frac{f(\mathbf{y} | \boldsymbol{\theta}) f(\boldsymbol{\theta})}{f(\mathbf{y})}, \quad (2.3)$$

kde  $f(\mathbf{y}) = \int_{\Theta} f(\mathbf{y} | \boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}$  je normalizační konstanta. Použitím symbolu  $\propto$  značícího *proporcionálně rovno* můžeme zkráceně psát

$$f(\boldsymbol{\theta} | \mathbf{y}) \propto f(\mathbf{y} | \boldsymbol{\theta}) f(\boldsymbol{\theta}). \quad (2.4)$$



Tímto získáme kompletní popis  $f(\boldsymbol{\theta} | \mathbf{y})$ , ale může nás zajímat např. pouze aposteriorní rozdělení podmnožiny  $\theta_1$ , kterou získáme marginalizací přes její doplněk  $\theta_2$

$$f(\theta_1 | \mathbf{y}) \propto \int_{\Theta_2} f(\boldsymbol{\theta} | \mathbf{y}) d\theta_2. \quad (2.5)$$

Moment funkce  $g(\boldsymbol{\theta})$  definovaný jako

$$\mathbb{E}_{f(\boldsymbol{\theta} | \mathbf{y})}[g(\boldsymbol{\theta})] = \int_{\Theta} g(\boldsymbol{\theta}) f(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}, \quad (2.6)$$

budeme značit prostě  $\widehat{g(\boldsymbol{\theta})}$ , nebude-li moct dojít k omylu.

### 2.1.1 Volba apriorního rozdělení

Volba apriorního rozdělení  $f(\boldsymbol{\theta})$  je subjektivní volbou návrháře modelu a na výsledný odhad má značný vliv[30]. Obecně pokud předpokládáme, že data jsou dostatečně informativní, volíme málo informativní apriorní rozdělení, abychom omezili jeho vliv na aposteriorní rozdělení  $f(\boldsymbol{\theta} | \mathbf{y})$ . V této práci bude apriorní rozdělení plnit několik úloh

- Regularizace, tedy doplnění dat v případech, kdy nejsou dostatečná.
- Zanesení požadavků a omezení na parametry  $\boldsymbol{\theta}$  do modelu.
- Vyjádření apriorní neznalosti  $\boldsymbol{\theta}$ . V takovém případě se budeme snažit volit neinformativní[1] apriorní rozdělení.

S výhodou budeme používat tzv. konjugovaná apriorní rozdělení  $f_{\text{conj}}(\boldsymbol{\theta})$  [30, 1], která splňují

$$f_{\text{conj}}(\boldsymbol{\theta} | \mathbf{y}) \propto f(\mathbf{y} | \boldsymbol{\theta}) f_{\text{conj}}(\boldsymbol{\theta}), \quad (2.7)$$

tedy po aplikaci Bayesovy věty přebírá aposteriorní rozdělení funkční tvar apriorního.

## 2.2 Aproximace

Bayesovský přístup, tedy aplikace Bayesovy věty (2.3), normalizace, marginalizace (2.5) a výpočet momentů aposteriorního rozdělení (2.6), je proveditelný pouze pro úzkou třídu modelů. Numerické řešení je sice možné, ale zejména ve vyšších dimenzích výpočetně příliš náročné a tudíž nevhodné.

Tomuto problému se vyhneme nalezením vhodné aproximace skutečného aposteriorního rozdělení pomocí rozdělení, pro které už tyto výpočty půjdou provést. Těch existuje mnoho a několik z nich v této sekci představíme. Zvláštní důraz pak budeme klást na aproximaci Variační Bayes.

### 2.2.1 MAP odhad

Existují případy, kdy nepotřebujeme celé aposteriorní rozdělení, ale stačí nám bodový odhad  $\hat{\theta}$ . Tento bodový odhad  $\hat{\theta} = \hat{\theta}(\mathbf{y})$  můžeme interpretovat jako extrémní případ aproximace aposteriorní hustoty, kdy  $f(\theta | \mathbf{y})$  je nahrazena vhodně položenou Diracovou delta funkcí:

$$f(\theta | \mathbf{y}) \approx \tilde{f}(\theta | \mathbf{y}) = \delta(\theta - \hat{\theta}(\mathbf{y})). \quad (2.8)$$

Zbývá otázka jak tento bodový odhad zvolit. Jednou z používaných možností je MAP (maximum aposteriority) odhad[30], který bere v úvahu apriorní rozdělení parametrů  $\theta$  a je definován

$$\hat{\theta}_{\text{MAP}} = \underset{\theta}{\operatorname{argmax}} f(\mathbf{y} | \theta) f(\theta). \quad (2.9)$$

### 2.2.2 Expectation-maximization (EM) algoritmus

Integrály typu (2.6) můžeme aproximovat Expectation-maximization algoritmem [9, 30], který ukážeme pro případ  $\theta = [\theta_1, \theta_2]$ :

**E-krok** V n-té iteraci aproximujeme marginální rozdělení  $\theta_1$ :

$$\tilde{f}^{[n]}(\theta_1 | \mathbf{y}) \approx f(\theta_1 | \mathbf{y}, \hat{\theta}_2^{[n-1]}). \quad (2.10)$$

**M-krok** Marginálu z E-kroku použijeme pro zlepšení odhadu  $\hat{\theta}_2$ :

$$\hat{\theta}_2^{[n]} = \underset{\theta_2}{\operatorname{argmax}} \int_{\Theta_1} \tilde{f}^{[n]}(\theta_1 | \mathbf{y}) \ln f(\theta_1, \theta_2, \mathbf{y}) d\theta_1. \quad (2.11)$$

### 2.2.3 Laplaceova aproximace

Laplaceova metoda [23] spočívá v nahrazení aposteriorního rozdělení  $f(\theta | \mathbf{y})$  normálním rozdělením se stejnou hodnotou, první a druhou derivací v bodě  $\hat{\theta}_{\text{MAP}}$ . Konkrétně

$$f(\theta | \mathbf{y}) \approx \mathcal{N}(\hat{\theta}_{\text{MAP}}, \mathbf{H}^{-1}), \quad (2.12)$$

kde  $\mathbf{H} \in \mathbb{R}^{p \times p}$  je záporně vzatá Hessova matice logaritmu sdruženého rozdělení  $f(\theta, \mathbf{y})$  vzhledem k  $\theta \in \mathbb{R}^p$  a vyčíslená v  $\theta = \hat{\theta}_{\text{MAP}}$ :

$$\mathbf{H} = - \left( \frac{\partial^2 \ln f(\theta, \mathbf{y})}{\partial \theta_i \partial \theta_j} \right)_{\theta = \hat{\theta}_{\text{MAP}}}, \quad i, j = 1, \dots, p. \quad (2.13)$$

Nevýhodou této metody je přesnost pouze na úzkém okolí  $\hat{\theta}_{\text{MAP}}$ .

## 2.2.4 Markov Chain Monte Carlo

Integrál (2.6) můžeme aproximovat z principu Monte Carlo [3] konečnou sumou

$$\widehat{g(\boldsymbol{\theta})} = \frac{1}{n} \sum_{i=1}^n g(\boldsymbol{\theta}_i), \quad (2.14)$$

kde  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n$  je posloupnost  $n$  nezávislých vzorků z pravděpodobnostního rozdělení  $f(\boldsymbol{\theta} | \mathbf{y})$ .

Tento odhad je nestranný a jeho rozptyl nezávisí na dimenzi  $\boldsymbol{\theta}$ . Teoreticky je tedy možné získat velmi přesný odhad už z malého množství vzorků [3]. Problém je, že vzorky nemusí být nezávislé a tedy skutečný počet vzorků je daleko menší. Dále pokud v některých oblastech je  $g(\boldsymbol{\theta})$  malá a  $f(\boldsymbol{\theta} | \mathbf{y})$  velká (nebo naopak), tak střední hodnota může být z velké části tvořena oblastmi s malou pravděpodobností. Pro potřebnou přesnost tedy bude potřeba daleko více vzorků.

Pokud nelze nebo není výhodné získávat vzorky přímo z rozdělení  $f(\boldsymbol{\theta} | \mathbf{y})$ , sestrojí se Markovův řetězec, který konverguje k danému rozdělení a setrvává v něm. Teorie okolo Markovových řetězců je rozebrána např. v [24].

*Posloupnost  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n$  nazveme Markovovým řetězcem, pokud*

$$f(\boldsymbol{\theta}_{n+1} | \boldsymbol{\theta}_n, \dots, \boldsymbol{\theta}_1) = f(\boldsymbol{\theta}_{n+1} | \boldsymbol{\theta}_n), \quad (2.15)$$

*tedy následující stav závisí pouze na stavu současném, nikoli na minulých.*

Vzorkováním z vhodného Markovského řetězce získáme nestranný odhad integrálu (2.6).

## 2.2.5 Variační Bayesova aproximace

Variační Bayesova (VB) aproximace [30] hledá vhodnou aproximaci aposterioriálního rozdělení v prostoru funkcí

$$\mathbb{F}_c = \left\{ f(\boldsymbol{\theta} | \mathbf{y}) \mid f(\boldsymbol{\theta} | \mathbf{y}) = \prod_{i=1}^q f(\theta_i | \mathbf{y}) \right\}, \quad (2.16)$$

kde  $q > 1$  je počet parametrů.

Chceme zjistit jak moc je aproximace  $\tilde{f}(\boldsymbol{\theta} | \mathbf{y}) \in \mathbb{F}_c$  blízká původnímu aposterioriálnímu rozdělení  $f(\boldsymbol{\theta} | \mathbf{y})$ . Bylo ukázáno [2], že pro získání maximální informace z dat je optimální logaritmická ztrátová funkce, která vede na Kullback-Leiblerovu divergenci [19] definovanou

$$\text{KLD}(\tilde{f}(\boldsymbol{\theta} | \mathbf{y}) \parallel f(\boldsymbol{\theta} | \mathbf{y})) = \int_{\Theta} \tilde{f}(\boldsymbol{\theta} | \mathbf{y}) \ln \frac{\tilde{f}(\boldsymbol{\theta} | \mathbf{y})}{f(\boldsymbol{\theta} | \mathbf{y})} d\boldsymbol{\theta} \quad (2.17)$$

s vlastnostmi  $\text{KLD}(\cdot, \cdot) \geq 0$ ,  $\text{KLD}(\cdot, \cdot) = 0$  právě tehdy když jsou si argumenty rovny skoro všude a obecně  $\text{KLD}(f \parallel \tilde{f}) \neq \text{KLD}(\tilde{f} \parallel f)$ .

Prostor funkcí  $\mathbb{F}_c$  (2.16) byl zvolen tak, aby funkce z něj splňovaly předpoklady Variační Bayesovy věty [30].

**Variační Bayesova věta.** *Nechť  $f(\boldsymbol{\theta} | \mathbf{y})$  je posteriorní rozdělení parametru  $\boldsymbol{\theta} = [\theta'_1, \dots, \theta'_q]$ .*

*Dále necht'  $\tilde{f}(\boldsymbol{\theta} | \mathbf{y})$  je aproximativní rozdělení, které lze zapsat jako produkt vzájemně nezávislých rozdělení pro  $\theta_1, \dots, \theta_q$*

$$\tilde{f}(\boldsymbol{\theta} | \mathbf{y}) = \prod_{i=1}^q \tilde{f}(\theta_i | \mathbf{y}). \quad (2.18)$$

*Pak minimum Kullback-Leiblerovy divergence*

$$\tilde{f}(\boldsymbol{\theta} | \mathbf{y}) = \underset{\tilde{f}}{\operatorname{argmin}} \operatorname{KLD}(\tilde{f}(\boldsymbol{\theta} | \mathbf{y}) \| f(\boldsymbol{\theta} | \mathbf{y})) \quad (2.19)$$

*nastává pro*

$$\tilde{f}(\theta_i | \mathbf{y}) \propto \exp\left(\mathbb{E}_{\tilde{f}(\boldsymbol{\theta}_i | \mathbf{y})} [\ln f(\boldsymbol{\theta}, \mathbf{y})]\right), \quad i = 1, \dots, q, \quad (2.20)$$

*kde  $\boldsymbol{\theta}_{/i}$  značí doplněk  $\theta_i$  do  $\boldsymbol{\theta}$  a  $\tilde{f}(\boldsymbol{\theta}_{/i} | \mathbf{y}) = \prod_{j=1, j \neq i}^q \tilde{f}(\theta_j | \mathbf{y})$ . Funkci  $\tilde{f}(\boldsymbol{\theta} | \mathbf{y})$  budeme nazývat VB–aproximace a funkce  $\tilde{f}(\theta_i | \mathbf{y})$  VB–marginály.*

Důkaz Variační Bayesovy věty lze nalézt např. v [30]. Výsledná aproximace je deterministická a apriori ji není předepsána žádná funkční forma. Aproximace (2.20) tvoří sadu implicitních rovnic, které přirozeně vedou na iterativní algoritmus IVB, který bude popsán později.

## Aplikace VB aproximace

Praktické použití VB metody zahrnuje následujících osm kroků [30] jejichž formulace je převzata z výzkumného úkolu:

**Krok 1: Formulace modelu** Sestrojení sdružené hustoty pravděpodobnosti  $f(\boldsymbol{\theta}, \mathbf{y})$  z modelu dat  $f(\mathbf{y} | \boldsymbol{\theta})$  a apriorního rozdělení  $f(\boldsymbol{\theta})$ .

**Krok 2: Rozdělení parametrů** Rozdělení vektoru parametrů  $\boldsymbol{\theta}$  na  $q$  pod-vektorů, které splňují (zde pro přehlednost  $q = 2$ )

$$\ln f(\theta_1, \theta_2, \mathbf{y}) = g(\theta_1, \mathbf{y})' h(\theta_2, \mathbf{y}), \quad (2.21)$$

kde  $g(\theta_1, \mathbf{y})$  a  $h(\theta_2, \mathbf{y})$  jsou vektory kompatibilní dimenze. Pokud logaritmus sdruženého rozdělení nelze zapsat jako skalární součin ve tvaru (2.21), pak VB metodu na daný model nelze použít.

**Krok 3: VB–marginály** Přímou aplikací Variační Bayesovy věty (2.20) získáme VB–marginály  $\tilde{f}(\theta_1), \dots, \tilde{f}(\theta_q)$ .

**Krok 4: Rozpoznání standardních forem** Rozpoznáme funkční tvary VB–marginál jako formu některého standardního parametrického rozdělení

$$\tilde{f}(\theta_i | \mathbf{y}) \equiv f_{\text{std}}(\theta_i | \mathbf{p}_i), \quad (2.22)$$

kde  $\mathbf{p}_i$  jsou vektory parametrů, tzv. tvarovací parametry, daného standardního rozdělení. Tento krok může značně zkomplikovat nevhodná volba apriorního rozdělení a právě proto je tak výhodné volit konjugovaná apriorní rozdělení.

**Krok 5: VB–momenty** Momenty parametrických rozdělení z předchozího kroku nazýváme VB–momenty. Ty jsou funkcemi tvarovacích parametrů a jsou většinou snadno dohledatelné nebo vypočítatelné.

**Krok 6: Zjednodušení rovnic** Tento krok je uveden pouze pro konzistenci s [30] a v samotné práci nebude použit. Pokud je to možné, maximálně zjednodušíme rovnice pro VB–momenty a tvarovací parametry.

**Krok 7: IVB algoritmus** Je-li soustava rovnic z kroků 4 a 5 analyticky neřešitelná, lze použít iterační algoritmus IVB konvergující ke skutečným VB–marginálám [26]. Algoritmus ukážeme pro  $\theta = [\theta_1, \theta_2]$ , ale lze jej použít pro obecný počet parametrů  $q \geq 1$ .

Pro  $n = 2, 3, \dots$  provedeme cyklickou iteraci následujících kroků, která monotónně snižuje KLD (2.17):

1. Vypočteme hodnotu VB–marginály  $\theta_2$  v  $n$ -té iteraci:

$$\tilde{f}^{[n]}(\theta_2 | \mathbf{y}) \propto \exp\left(\int_{\Theta_1} \tilde{f}^{[n-1]}(\theta_1 | \mathbf{y}) \ln f(\theta_1, \theta_2, \mathbf{y}) d\theta_1\right). \quad (2.23)$$

2. Využitím nové hodnoty VB–marginály  $\theta_2$  spočítáme VB–marginálu  $\theta_1$  v  $n$ -té iteraci:

$$\tilde{f}^{[n]}(\theta_1 | \mathbf{y}) \propto \exp\left(\int_{\Theta_2} \tilde{f}^{[n-1]}(\theta_2 | \mathbf{y}) \ln f(\theta_1, \theta_2, \mathbf{y}) d\theta_2\right). \quad (2.24)$$

Algoritmus může dosáhnout lokálního minima, vhodnou volbou počátečních hodnot a pořadí výpočtu však lze dosáhnout rychlejší konvergence ke smysluplnějšímu výsledku.

**Krok 8** IVB algoritmus poskytuje výsledky ve formě tvarovacích parametrů standardních rozdělení a jejich VB–momentů.

Metodu aproximace VB budeme v této práci používat, protože se dobře osvědčila v bakalářské práci [6] a výzkumném úkolu. Pravděpodobnostní model i přes omezení, která jsou pro použití aproximace nutná, dovoluje pohodlně vložit požadovanou apriorní informaci. Použití této aproximace přirozeně vede na jednoduchý iterativní algoritmus IVB, který velmi rychle konverguje k výslednému odhadu. Celková výpočetní náročnost je malá, zejména ve srovnání se vzorkovacími metodami.

# Kapitola 3

## Odvození použitých algoritmů

V této kapitole budeme řešit inverzní úlohu, tedy hledání vektoru  $\mathbf{x} \in \mathbb{R}^{n \times 1}$  za předpokladu, že vektor  $\mathbf{y} \in \mathbb{R}^{p \times 1}$  lze vysvětlit lineárním modelem

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{e}, \quad (3.1)$$

kde  $\mathbf{M} \in \mathbb{R}^{p \times n}$  a  $\mathbf{e} \in \mathbb{R}^{p \times 1}$ .

Uvedeme formulaci již existujícího algoritmu LS-APC (Least Squares with Adaptive Prior Covariance)[32], který tuto úlohu řeší, a ukážeme jeho odvození. Jedním z předpokladů tohoto algoritmu je, že šum  $\mathbf{e}$  má stejně rozdělené nezávislé složky a jeho rozptyl lze tedy modelovat jediným skalárním parametrem  $\omega$ . Tento předpoklad zkusíme rozvolnit a příslušným způsobem modifikujeme algoritmus.

První variantou bude případ, kdy každá složka šumu má vlastní parametr rozptylu. Ve druhé variantě bude vektor  $\mathbf{y}$  rozdělen do kategorií (v kontextu radiačního úniku např. podle země měření nebo vzdálenosti od určitého bodu), kdy každá kategorie bude mít vlastní parametr rozptylu šumu. Nakonec upustíme od předpokladu, že složky šumu jsou nekorelované.

### 3.1 LS-APC

#### 3.1.1 Pravděpodobnostní model

Předpokládáme, že složky vektoru šumu  $\mathbf{e}$  jsou nezávislé a stejně rozdělené s normálním rozdělením a tedy můžeme psát

$$f(\mathbf{e} | \omega) = \mathcal{N}(\mathbf{0}, \omega^{-1} \mathbf{I}_p), \quad (3.2)$$

kde  $\mathbf{I}_p$  je jednotková matice rozměrů  $p \times p$ . Parametr  $\omega^{-1}$  vyjadřuje skalární rozptyl šumu a v této variantě modelu předpokládáme, že je pro všechny realizace společný.  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  je vícerozměrné normální rozdělení, viz příloha A.

Za předpokladu lineárního modelu (3.1) pak můžeme rovnici pozorování zapsat pravděpodobnostně

$$f(\mathbf{y} | \omega, \mathbf{M}, \mathbf{x}) = \mathcal{N}(\mathbf{M}\mathbf{x}, \omega^{-1} \mathbf{I}_p) \propto \omega^{p/2} \exp\left(-\frac{1}{2} \omega (\mathbf{y} - \mathbf{M}\mathbf{x})^\top (\mathbf{y} - \mathbf{M}\mathbf{x})\right). \quad (3.3)$$

Parametr  $\omega$  je neznámý. Odhadován bude podobně jako v [34]. Přidáme tedy jeho apriorní rozdělení, které zvolíme jako konjugované k normálnímu – tedy gamma rozdělení

$$f(\omega) = \mathcal{G}(\vartheta_0, \rho_0) \propto \omega^{\vartheta_0-1} \exp(-\rho_0\omega), \quad (3.4)$$

kde skalární apriorní parametry  $\vartheta_0$  a  $\rho_0$  se neodhadují, ale jsou předem zvoleny. Chceme získat co nejméně informativní apriorní rozdělení a proto bychom je měli nastavit co nejmenší, ideálně nulové což by vedlo na neinformativní Jeffreysovu apriorní hustotu[1]. Pro lepší numerickou stabilitu však volíme velmi malá čísla – konkrétně  $10^{-10}$ .

Apriorní rozdělení vektoru  $\mathbf{x}$  je navrženo aby podporovalo vlastnosti požadované z fyzikální podstaty úlohy

1. Pozitivita – složky zdrojového členu požadujeme nezáporné.
2. Řídkost – složky jsou nulové, pokud data neposkytují dostatečnou informaci že by neměly být.
3. Hladkost – prudké změny složek jsou sice možné, ale ne příliš časté.

Volíme jej tedy jako ořezané normální rozdělení s neznámou kovarianční maticí, kterou budeme dále modelovat

$$f(\mathbf{x} | \mathbf{v}, \mathbf{l}, \psi_1, \dots, \psi_{n-1}) = \mathcal{N}^{\text{tr}}(\mathbf{0}, \Sigma_x^{-1}; [0, +\infty)) \propto \left( \prod_{j=1}^n v_j \right)^{1/2} \exp\left(-\frac{1}{2} \mathbf{x}^\top \mathbf{L} \mathbf{Y} \mathbf{L}^\top \mathbf{x}\right), \quad (3.5)$$

kde  $\mathcal{N}^{\text{tr}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; [a, b])$  je ořezané normální rozdělení se středem  $\boldsymbol{\mu}$  a kovarianční maticí  $\boldsymbol{\Sigma}$  na nosiči  $[a, b]$ , vynucuje pozitivitu řešení.

Kovarianční matice  $\Sigma_x \in \mathbb{R}^{n \times n}$  má tridiagonální strukturu ve tvaru

$$\Sigma_x = \mathbf{L} \mathbf{Y} \mathbf{L}^\top. \quad (3.6)$$

Skládá se z matice  $\mathbf{Y} = \text{diag}(\mathbf{v})$  s neznámým vektorem  $\mathbf{v} = [v_1, \dots, v_n]$  na diagonále a zbytkem prvků nulových.  $\mathbf{L} \in \mathbb{R}^{n \times n}$  je dolní bidiagonální matice s tvarem

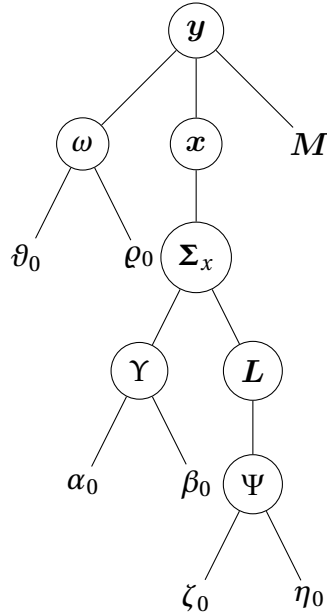
$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ l_1 & 1 & 0 & 0 \\ 0 & \ddots & 1 & 0 \\ 0 & 0 & l_{n-1} & 1 \end{pmatrix}, \quad (3.7)$$

s neznámými prvky mimo diagonálu tvořícími vektor  $\mathbf{l} = [l_1, \dots, l_{n-1}]$ . Tím je zachována tridiagonální struktura kovarianční matice  $\Sigma_x$  a můžeme každou diagonálu modelovat zvlášť.

Zbývá zvolit apriorní rozdělení pro vektory  $\mathbf{v}$  a  $\mathbf{l}$  jejichž odhad pak plně určí rozklad (3.6). Složky vektoru  $\mathbf{v}$  jsou stejně rozdělené

$$f(v_j) = \mathcal{G}(\alpha_0, \beta_0) \propto v_j^{\alpha_0-1} \exp(-\beta_0 v_j) \quad \forall j \in 1, \dots, n, \quad (3.8)$$

kde skalární parametry  $\alpha_0$  a  $\beta_0$  jsou voleny, aby vedly na málo informativní rozdělení, stejně jako  $\vartheta_0$  a  $\rho_0$ . Každé  $v_j$  funguje jako váha pro příslušnou složku  $x_j$ . Malá hodnota  $v_j$  znamená,



Obrázek 3.1: Grafické znázornění struktury pravděpodobnostního modelu.

že  $x_j$  má velkou oporu v datech. Naopak velká hodnota  $v_j$  tlačí  $x_j$  k nule. To vynucuje námi požadované řídké řešení[33].

Složky  $l_j$  má význam korelace sousedních hodnot  $x_j$  a  $x_{j+1}$ , kdy pro  $l_j = 0$  jsou nekorelované a pro  $l_j = -1$  plně korelované. Abychom pokryli tyto krajní hodnoty, volíme apriorní rozdělení

$$f(l_j | \psi_j) = \mathcal{N}(l_0, \psi_j^{-1}) \propto \psi_j^{1/2} \exp\left(-\frac{1}{2}\psi_j(l_j - l_0)^2\right), \quad (3.9)$$

$$f(\psi_j) = \mathcal{G}(\zeta_0, \eta_0) \propto \psi_j^{\zeta_0-1} \exp(-\eta_0\psi_j) \quad \forall j \in 1, \dots, n-1, \quad (3.10)$$

kde volíme  $l_0 = -1$  a parametr přesnosti  $\psi_j$ , který je dále parametrizován skalárními parametry  $\zeta_0 = \eta_0 = 10^{-2}$ . Parametr  $l_j$  se tedy bude pohybovat v rozmezí přibližně  $-1 \pm 100$ .

Grafické znázornění pravděpodobnostního modelu je na obrázku 3.1.

### 3.1.2 Odvození algoritmu pomocí VB aproximace

Nyní na zformulovaný model aplikujeme Variační Bayesovu aproximaci (2.2.5). Toto odvození je přímo převzato z výzkumného úkolu.

#### Krok 1

První krok byl proveden už samotnou formulací pravděpodobnostního modelu.



## Krok 2

Přirozený logaritmus sdruženého rozdělení  $f(\mathbf{y}, \mathbf{x}, \mathbf{v}, \mathbf{l}, \psi_1, \dots, \psi_{n-1}, \omega)$

$$\begin{aligned} \ln f(\mathbf{y}, \mathbf{x}, \mathbf{v}, \mathbf{l}, \psi_1, \dots, \psi_{n-1}, \omega) &= \frac{p}{2} \ln \omega - \frac{1}{2} \omega (\mathbf{y} - \mathbf{M}\mathbf{x})^\top (\mathbf{y} - \mathbf{M}\mathbf{x}) + \\ &+ (\vartheta_0 - 1) \ln \omega - \varrho_0 \omega + \frac{1}{2} \sum_{j=1}^n \ln v_j - \frac{1}{2} \mathbf{x}^\top \mathbf{L} \mathbf{Y} \mathbf{L}^\top \mathbf{x} + (\alpha_0 - 1) \sum_{j=1}^n \ln v_j - \beta_0 \sum_{j=1}^n v_j + \\ &+ \frac{1}{2} \sum_{j=1}^{n-1} \ln \psi_j - \frac{1}{2} \sum_{j=1}^{n-1} \psi_j (l_j - l_0)^2 + (\zeta_0 - 1) \sum_{j=1}^{n-1} \ln \psi_j - \eta_0 \sum_{j=1}^{n-1} \psi_j + \gamma, \end{aligned} \quad (3.11)$$

kde  $\gamma$  představuje členy, které nezávisí na žádném z parametrů  $\mathbf{y}, \mathbf{x}, \mathbf{v}, \mathbf{l}, \psi_1, \dots, \psi_{n-1}, \omega$ .

## Krok 3

Aplikací Variační Bayesovy věty (2.20) získáme následující VB-marginály

$$\tilde{f}(\mathbf{x} | \mathbf{y}) \propto \exp \left[ -\frac{1}{2} \widehat{\omega} (\mathbf{y} - \mathbf{M}\mathbf{x})^\top (\mathbf{y} - \mathbf{M}\mathbf{x}) - \frac{1}{2} \mathbf{x}^\top \widehat{\mathbf{L} \mathbf{Y} \mathbf{L}^\top} \mathbf{x} \right], \quad (3.12)$$

$$\tilde{f}(\omega | \mathbf{y}) \propto \exp \left[ \left( \vartheta_0 + \frac{p}{2} - 1 \right) \ln \omega - \omega \left( \varrho_0 + \frac{1}{2} (\mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{M} \widehat{\mathbf{x}} - \widehat{\mathbf{x}}^\top \mathbf{M}^\top \mathbf{y} + \mathbf{x}^\top \widehat{\mathbf{M}^\top \mathbf{M}} \mathbf{x}) \right) \right], \quad (3.13)$$

$$\tilde{f}(v_j | \mathbf{y}) \propto \exp \left[ \left( \alpha_0 + \frac{1}{2} - 1 \right) \ln v_j - v_j \left( \beta_0 + \frac{1}{2} (\widehat{\mathbf{L}^\top \mathbf{x} \mathbf{x}^\top \mathbf{L}})_{j,j} \right) \right], \quad (3.14)$$

$$\tilde{f}(l_j | \mathbf{y}) \propto \exp \left[ -\frac{1}{2} \widehat{\psi}_j (l_j - l_0)^2 - \frac{1}{2} l_j^2 \widehat{x_{j+1}^2} \widehat{v}_j \right], \quad (3.15)$$

$$\tilde{f}(\psi_j | \mathbf{y}) \propto \exp \left[ \left( \zeta_0 + \frac{1}{2} - 1 \right) \ln \psi_j - \psi_j \left( \eta_0 + \frac{1}{2} (\widehat{l_j - l_0})^2 \right) \right]. \quad (3.16)$$

## Krok 4

Díky vhodné volbě apriorních rozdělení můžeme předpokládat, že VB-marginály z kroku 3 mají následující tvary

$$\tilde{f}(\mathbf{x} | \mathbf{y}) = \mathcal{N}^{\text{tr}}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x; [0, +\infty)), \quad (3.17)$$

$$\tilde{f}(\omega | \mathbf{y}) = \mathcal{G}(\vartheta, \varrho), \quad (3.18)$$

$$\tilde{f}(v_j | \mathbf{y}) = \mathcal{G}(\alpha, \beta_j), \quad (3.19)$$

$$\tilde{f}(l_j | \mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}_{l_j}, \boldsymbol{\Sigma}_{l_j}), \quad (3.20)$$

$$\tilde{f}(\psi_j | \mathbf{y}) = \mathcal{G}(\zeta_j, \eta_j). \quad (3.21)$$

Tvarovací parametry rozdělení ihned vidíme

$$\vartheta = \vartheta_0 + \frac{p}{2}, \quad (3.22)$$

$$\varrho = \varrho_0 + \frac{1}{2}(\mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{M} \hat{\mathbf{x}} - \hat{\mathbf{x}}^\top \mathbf{M}^\top \mathbf{y} + \mathbf{x}^\top \widehat{\mathbf{M}^\top \mathbf{M}} \mathbf{x}) = \varrho_0 + \frac{1}{2} \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{M} \hat{\mathbf{x}} + \frac{1}{2} \text{tr}(\widehat{\mathbf{x}} \mathbf{x}^\top \mathbf{M}^\top \mathbf{M}), \quad (3.23)$$

$$\alpha = \alpha_0 + \frac{1}{2}, \quad \beta_j = \beta_0 + \frac{1}{2} \left( \widehat{\mathbf{L}^\top \mathbf{x} \mathbf{x}^\top \mathbf{L}} \right)_{j,j}, \quad (3.24)$$

$$\zeta_j = \zeta_0 + \frac{1}{2}, \quad \eta_j = \eta_0 + \frac{1}{2} \widehat{(l_j - l_0)^2}. \quad (3.25)$$

Dále srovnáním lineárních

$$\omega \mathbf{y}^\top \mathbf{M} = \boldsymbol{\mu}_x^\top \boldsymbol{\Sigma}_x^{-1} \quad (3.26)$$

a kvadratických

$$\omega \mathbf{M}^\top \mathbf{M} + \mathbf{L} \boldsymbol{\Upsilon} \mathbf{L}^\top = \boldsymbol{\Sigma}_x^{-1} \quad (3.27)$$

členů  $\mathbf{x}$  získáme

$$\boldsymbol{\Sigma}_x = \left( \widehat{\omega} \mathbf{M}^\top \mathbf{M} + \widehat{\mathbf{L} \boldsymbol{\Upsilon} \mathbf{L}^\top} \right)^{-1}, \quad (3.28)$$

$$\boldsymbol{\mu}_x = \boldsymbol{\Sigma}_x \widehat{\omega} \mathbf{M}^\top \mathbf{y}. \quad (3.29)$$

Poslední dva parametry získáme obdobně srovnáním lineárních

$$\mathbf{x}^\top \mathbf{L} \boldsymbol{\Upsilon} \mathbf{L}^\top \mathbf{x} = \sum_{j=1}^{n-1} x_{j+1} v_j x_j l_j + c, \quad (3.30)$$

a kvadratických

$$\mathbf{x}^\top \mathbf{L} \boldsymbol{\Upsilon} \mathbf{L}^\top \mathbf{x} = \sum_{j=1}^{n-1} x_{j+1}^2 v_j l_j^2 + d, \quad (3.31)$$

členů, kde  $c$  a  $d$  neobsahují člen  $l_j$  resp.  $l_j^2$ . Tudíž

$$\boldsymbol{\Sigma}_{l_j}^{-1} = \widehat{v_j x_{j+1}^2} + \widehat{\psi_j}, \quad (3.32)$$

$$\mu_{l_j} = \boldsymbol{\Sigma}_{l_j} \left( l_0 \widehat{\psi_j} - \widehat{\bar{x}_j x_{j+1} v_j} \right) \quad (3.33)$$

## Krok 5

VB–momenty spočteme podle vzorců uvedených v příloze A.

## Krok 7

Zvolíme následující počáteční hodnoty

- $\alpha_0 = \beta_0 = \vartheta_0 = \varrho_0 = 10^{-10}$ ,
- $\eta_0 = \zeta_0 = 10^{-2}$ ,
- $\widehat{\omega}^{[0]} = \frac{1}{\max \mathbf{M}^\top \mathbf{M}}$ ,
- $\widehat{\boldsymbol{\Upsilon}}^{[0]}$  a  $\widehat{\mathbf{L}}^{[0]}$  jednotkové matice příslušných rozměrů.

V každém kroku IVB algoritmu spočteme postupně nové odhady  $\boldsymbol{\Sigma}_x$ ,  $\hat{\mathbf{x}}$ ,  $\widehat{\boldsymbol{\Upsilon}}$ ,  $\widehat{\mathbf{L}}$  a nakonec  $\widehat{\omega}$  použitím hodnot z předchozího kroku.

## Krok 8

Po dostatečném počtu kroků IVB algoritmu získáváme hledaný odhad zdrojového členu  $\hat{x}$  a jeho kovarianční matici  $\Sigma_x$ .

## 3.2 Modifikace pro model s plně rozvolněným rozptylem šumu

Předpoklad jednotného rozptylu pro všechny složky šumu je značně nerealistický i když alespoň jako hrubá aproximace funguje. Můžeme tento předpoklad rozvolnit a rovnice (3.2) pak přejde na

$$f(e | \omega) = \mathcal{N}(0, \mathbf{\Omega}^{-1}), \quad (3.34)$$

kde  $\mathbf{\Omega} = \text{diag}(\omega_1, \dots, \omega_n)$  a  $\omega_i$  je skalární rozptyl  $i$ -té složky šumu.

Lineární model (3.1) pak nabývá tvaru

$$f(\mathbf{y} | \omega, \mathbf{M}, \mathbf{x}) = \mathcal{N}(\mathbf{M}\mathbf{x}, \mathbf{\Omega}^{-1}) \propto \left( \prod_{j=1}^p \omega_j \right)^{1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{M}\mathbf{x})^\top \mathbf{\Omega}(\mathbf{y} - \mathbf{M}\mathbf{x})\right). \quad (3.35)$$

Každý parametr  $\omega_j$  bude mít vlastní apriorní rozdělení

$$f(\omega_j) = \mathcal{G}(\vartheta_0, \varrho_0) \propto \omega_j^{\vartheta_0-1} \exp(-\varrho_0 \omega_j). \quad (3.36)$$

Princip odvození algoritmu se těmito úpravami nijak nezmění. Změní se pouze aposteriorní rozdělení  $\tilde{f}(\mathbf{x} | \mathbf{y})$ ,  $\tilde{f}(\omega_j | \mathbf{y})$  a jejich parametry. Rozdělení  $\tilde{f}(\mathbf{x} | \mathbf{y})$  si zachovává tvar, ale jeho parametry se změni na

$$\Sigma_x = \left( \mathbf{M}^\top \widehat{\mathbf{\Omega}} \mathbf{M} + \widehat{\mathbf{L}} \mathbf{Y} \mathbf{L}^\top \right)^{-1}, \quad (3.37)$$

$$\boldsymbol{\mu}_x = \Sigma_x \mathbf{M}^\top \widehat{\mathbf{\Omega}} \mathbf{y}, \quad (3.38)$$

dále  $\tilde{f}(\omega_j | \mathbf{y}) = \mathcal{G}(\vartheta, \varrho_j)$  s parametry

$$\vartheta = \vartheta_0 + \frac{1}{2}, \quad (3.39)$$

$$\varrho_j = \varrho_0 + \frac{1}{2} \left[ \mathbf{y} \mathbf{y}^\top - \mathbf{y} \widehat{\mathbf{x}}^\top \mathbf{M}^\top - \mathbf{M} \widehat{\mathbf{x}} \mathbf{y}^\top + \mathbf{M} \widehat{\mathbf{x}} \widehat{\mathbf{x}}^\top \mathbf{M}^\top \right]_{j,j}. \quad (3.40)$$

Zbylá aposteriorní rozdělení se nijak nezmění. Počáteční hodnoty volíme stejné, místo  $\widehat{\omega}$  nyní máme  $\widehat{\mathbf{\Omega}}$ , kterou inicializujeme na  $\widehat{\omega}^{[0]} \mathbf{I}_p$ .

## 3.3 Modifikace pro model se rozptylem šumu podle kategorií

Další variantou je rozdělit měření do kategorií např. podle toho, ve které zemi k měření došlo nebo jak je měřicí stanice vzdálena od určitého geografického bodu nebo jakkoliv jinak. Místo toho aby se odhadoval rozptyl šumu každého měření tak algoritmus modifikujeme, aby dokázal modelovat rozptyl šumu pro každou kategorii zvlášť.

Necht' jsou tedy složky vektoru  $\mathbf{y}$  rozděleny do  $K$  kategorií, kde počet složek v  $i$ -té kategorii označíme  $c_i$ . Stále platí rovnice (3.34), patří-li však složky  $y_i$  a  $y_j$  (a tedy i jim příslušné  $e_i$  a  $e_j$ ) do stejné kategorie, budeme požadovat aby příslušné parametry rozptylu šumu měly stejnou hodnotu, tedy  $\omega_i = \omega_j$ .

Tímto získává lineární model (3.1) tvar

$$f(\mathbf{y} | \boldsymbol{\omega}, \mathbf{M}, \mathbf{x}) = \mathcal{N}(\mathbf{M}\mathbf{x}, \boldsymbol{\Omega}^{-1}) \propto \left( \prod_{j=1}^K \omega_j^{c_j} \right)^{1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{M}\mathbf{x})^\top \boldsymbol{\Omega}(\mathbf{y} - \mathbf{M}\mathbf{x})\right). \quad (3.41)$$

Náš požadavek se tedy promítne pouze do tvaru matice  $\boldsymbol{\Omega}$ , která zůstává diagonální, ale prvky na diagonále patřící do stejné kategorie mají stejnou hodnotu. Matice  $\boldsymbol{\Omega}$  je tedy tvaru

$$\boldsymbol{\Omega} = \begin{pmatrix} \omega_1 \mathbf{I}_{c_1} & 0 & \dots & 0 \\ 0 & \omega_2 \mathbf{I}_{c_2} & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \omega_K \mathbf{I}_{c_K} \end{pmatrix}. \quad (3.42)$$

Změní se pouze aposteriorní rozdělení  $\tilde{f}(\omega_j | \mathbf{y}) = \mathcal{G}(\vartheta_j, \varrho_j)$  s hodnotami parametrů

$$\vartheta_j = \vartheta_0 + \frac{c_j}{2} \quad (3.43)$$

$$\varrho_j = \varrho_0 + \frac{1}{2} [(\mathbf{y}_c - \widehat{\mathbf{M}_c \mathbf{x}})(\mathbf{y}_c - \widehat{\mathbf{M}_c \mathbf{x}})^\top]_{j,j} = \varrho_0 + \frac{1}{2} [\mathbf{y}_c \mathbf{y}_c^\top - \mathbf{y}_c \widehat{\mathbf{x}}^\top \mathbf{M}_c^\top - \mathbf{M}_c \widehat{\mathbf{x}} \mathbf{y}_c^\top + \mathbf{M}_c \widehat{\mathbf{x}} \widehat{\mathbf{x}}^\top \mathbf{M}_c^\top]_{j,j}, \quad (3.44)$$

kde  $\mathbf{y}_c$  je vektor obsahující pouze takové složky vektoru  $\mathbf{y}$ , které patří do  $j$ -té kategorie. Podobně matice  $\mathbf{M}_c$  obsahuje pouze řádky matice  $\mathbf{M}$ , které patří do  $j$ -té kategorie.

### 3.4 Modifikace pro model s korelovaným šumem

I plně rozvolněná diagonální matice  $\boldsymbol{\Omega}$  stále modeluje pouze rozptyl šumu jednotlivých měření a neuvažuje korelaci. Tu můžeme modelovat upuštěním od diagonální struktury matice  $\boldsymbol{\Omega}$ .

Vyjdeme z modelu s plně rozvolněným rozptylem šumu v sekci 3.2. Rovnice (3.34) tvarově zůstává, jen matice  $\boldsymbol{\Omega}$  nyní není diagonální. Její prvky  $\Omega_{i,i}$  na diagonále stále mají význam rozptylu šumu, prvky mimo diagonálu  $\Omega_{i,j}$ ,  $i \neq j$  mají význam korelace mezi  $i$ -tou a  $j$ -tou složkou šumu.

Lineární model (3.1) pak zůstává stejný, jenom jej zapíšeme v trochu jiném tvaru, se kterým se nám bude později lépe pracovat

$$f(\mathbf{y} | \mathbf{x}, \boldsymbol{\Omega}) = \mathcal{N}(\mathbf{M}\mathbf{x}, \boldsymbol{\Omega}^{-1}) \propto |\boldsymbol{\Omega}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{M}\mathbf{x})^\top \boldsymbol{\Omega}(\mathbf{y} - \mathbf{M}\mathbf{x})\right) \propto |\boldsymbol{\Omega}|^{\frac{1}{2}} \exp\left(-\frac{1}{2} \text{tr}(\boldsymbol{\Omega}(\mathbf{y} - \mathbf{M}\mathbf{x})(\mathbf{y} - \mathbf{M}\mathbf{x})^\top)\right). \quad (3.45)$$

Největší změna nastane v apriorním rozdělení matice  $\mathbf{\Omega}$ , které již nemůžeme zapsat jako součin apriorních rozdělení jejich diagonálních prvků. Místo toho jako apriorní rozdělení celé matice zvolíme Wishartovo rozdělení, které je v podstatě zobecnění gamma rozdělení

$$f(\mathbf{\Omega}) = \mathcal{W}(\varrho_0 \mathbf{I}_p, \vartheta_0) \propto |\mathbf{\Omega}|^{\frac{\vartheta_0 - p - 1}{2}} \exp\left(-\frac{1}{2} \text{tr}(\varrho_0^{-1} \mathbf{I}_p \mathbf{\Omega})\right). \quad (3.46)$$

Změní se pouze VB–marginála

$$\tilde{f}(\mathbf{\Omega} | \mathbf{y}) = \mathcal{W}(\mathbf{\Sigma}, \nu) \quad (3.47)$$

a její parametry snadno zjistíme srovnáním příslušných členů

$$|\mathbf{\Omega}|^{\frac{1}{2}} |\mathbf{\Omega}|^{\frac{\vartheta_0 - p - 1}{2}} = |\mathbf{\Omega}|^{\frac{\vartheta_0 - p}{2}} = |\mathbf{\Omega}|^{\frac{(\vartheta_0 + 1) - p - 1}{2}} = |\mathbf{\Omega}|^{\frac{\nu - p - 1}{2}}, \quad (3.48)$$

$$\begin{aligned} \text{tr}(\varrho_0^{-1} \mathbf{I}_p \mathbf{\Omega} + \mathbf{\Omega}(\mathbf{y} - \mathbf{M}\mathbf{x})(\mathbf{y} - \mathbf{M}\mathbf{x})^\top) &= \text{tr}(\mathbf{\Omega}(\varrho_0^{-1} \mathbf{I}_p + (\mathbf{y} - \mathbf{M}\mathbf{x})(\mathbf{y} - \mathbf{M}\mathbf{x})^\top)) = \\ &= \text{tr}((\varrho_0^{-1} \mathbf{I}_p + (\mathbf{y} - \mathbf{M}\mathbf{x})(\mathbf{y} - \mathbf{M}\mathbf{x})^\top) \mathbf{\Omega}) = \text{tr}(\mathbf{\Sigma}^{-1} \mathbf{\Omega}). \end{aligned} \quad (3.49)$$

Hodnoty parametrů jsou tedy

$$\nu = \vartheta_0 + 1, \quad (3.50)$$

$$\mathbf{\Sigma} = (\varrho_0^{-1} \mathbf{I}_p + (\mathbf{y} - \widehat{\mathbf{M}\mathbf{x}})(\mathbf{y} - \widehat{\mathbf{M}\mathbf{x}})^\top)^{-1}. \quad (3.51)$$

Ostatní aposteriorní rozdělení se nijak nezmění. Počáteční hodnoty volíme stejné kromě parametru  $\varrho_0$ , který nyní volíme  $\varrho_0 = 10^{10}$ .

Poslední změna nijak neplyne z odvození algoritmu, ale je to přidaná heuristika. V kapitole 4 se ukáže, že odhadovat úplně všechny prvky matice  $\mathbf{\Omega}$  má negativní vliv na konvergenci algoritmu. Proto v každém kroku IVB algoritmu aplikujeme na odhad matice  $\mathbf{\Omega}$  tzv. lokalizační masku  $\mathbf{\Lambda}$

$$\widehat{\mathbf{\Omega}} = \widehat{\mathbf{\Omega}} \odot \mathbf{\Lambda}, \quad (3.52)$$

kde  $\odot$  je násobení matic po prvcích a lokalizační maska  $\mathbf{\Lambda}$  je matice stejného rozměru jako  $\mathbf{\Omega}$  obsahující prvky z intervalu  $[0, 1]$ .

Lokalizační maska tedy zeslabí nebo úplně odstraní odhad korelace mezi šumem některých měření. Lokalizační maska není odhadována v modelu, ale zvolíme ji ještě před spuštěním IVB algoritmu.

# Kapitola 4

## Výsledky

V této kapitole nejdříve ukážeme nutnost použití lokalizační masky u algoritmu LS-APC-w představeného v kapitole 3. Následovat bude analýza vlivu lokalizační masky na kvalitu odhadu zdrojového členu.

Jako metriky kvality odhadu budeme používat střední absolutní odchylku odhadnutého zdrojového členu od skutečného ( $MAE_x$ ), střední absolutní odchylku rekonstrukce od měření ( $MAE_y$ ) a koeficient determinace rekonstrukce ( $R^2$ ), které jsou definovány následovně:

$$MAE_x = \frac{1}{n} \sum_{i=1}^n |x_i - x_{\text{true}_i}|, \quad (4.1)$$

$$MAE_y = \frac{1}{n} \sum_{i=1}^n |y_i - (M\mathbf{x})_i|, \quad (4.2)$$

$$R^2 = \frac{\sum_{i=1}^n ((M\mathbf{x})_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (4.3)$$

kde  $n$  je délka srovnávaných vektorů,  $x_i$  složka odhadu,  $x_{\text{true}_i}$  složka skutečného zdrojového členu,  $y_i$  složka měření a  $\bar{y}$  střední hodnota měření.

Následně algoritmy představené v kapitole 3 vyzkoušíme na datech z reálného experimentálního úniku ETEX (European Tracer Experiment)[25]. Pro tento experiment je znám skutečný zdrojový člen, což umožňuje přímé srovnání odhadů.

Nakonec algoritmy aplikujeme na data z úniku ruthenia-106 do atmosféry na podzim roku 2017, který je popsán v úvodu práce. Výsledné odhady srovnáme s dostupnou literaturou.

Pro použité algoritmy budeme zejména v legendách grafů a v tabulkách používat zkrácené názvy uvedené v závorkách:

- LS-APC bez modifikací.
- LS-APC s diagonálním rozvolněným rozptylem šumu (LS-APC-Ovolne).
- LS-APC s kategoriemi podle země, kde dané měření proběhlo (LS-APC-Ozeme).
- LS-APC s kategoriemi podle vzdálenosti vzdušnou čarou od Majaku rozdělených po 100 km (LS-APC-O100).

- LS-APC s plně rozvolněným modelem šumu používající Wishartovo rozdělení (LS-APC-w).

## 4.1 Syntetická data

LS-APC-w nejdříve vyzkoušíme na syntetických datech, která budou simulovat radiační únik, a ukážeme nutnost použití lokalizace. Poté zanalyzujeme vliv lokalizační masky na kvalitu odhadu zdrojového členu.

### 4.1.1 Vliv lokalizace na stabilitu

Vhodnou volbou lokalizační masky (3.52) můžeme ovlivnit, které prvky matice  $\mathbf{\Omega}$  se ponechají do další iterace algoritmu. Touto volbou tedy přímo říkáme, mezi kterými měřeními se má modelovat korelace.

Jako první na syntetických datech ukážeme vliv lokalizační masky na stabilitu odhadu zdrojového členu. K tomuto účelu si sestrojíme syntetický dataset, který bude simulovat radiační únik. Prvky matice  $M \in \mathbb{R}^{20 \times 10}$  vezmeme z rovnoměrného rozdělení mezi 0 a 1. Prvky menší než 0.5 navíc nahradíme nulami pro zhoršení podmíněnosti matice. Skutečný zdrojový člen  $\mathbf{x}_{\text{true}}$  bude sloupcový vektor

$$\mathbf{x}_{\text{true}} = [0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0]^T. \quad (4.4)$$

Vektor „naměřených“ dat  $\mathbf{y}$  se pak získá podle předpokládaného modelu bez přidaného šumu

$$\mathbf{y} = M\mathbf{x}_{\text{true}}. \quad (4.5)$$

Případné záporné složky vektoru  $\mathbf{y}$  se nahradí nulami.

Odhadnutý zdrojový člen pomocí algoritmu LS-APC-w bez použití lokalizační masky je na obrázku 4.1. Vidíme, že po 100 iteracích se odhad ještě příliš neliší od  $\mathbf{x}_{\text{true}}$  – sice je mírně podhodnocen, ale zachovává správně prudký začátek a konec úniku. Mimo tento úsek je odhad správně řídký. Po 1000 iteracích však vidíme, že odhad začne divergovat k naprosto nesprávným hodnotám. Je to způsobeno tím, že se algoritmus snaží odhadnout příliš velké množství parametrů bez patřičné opory v datech.

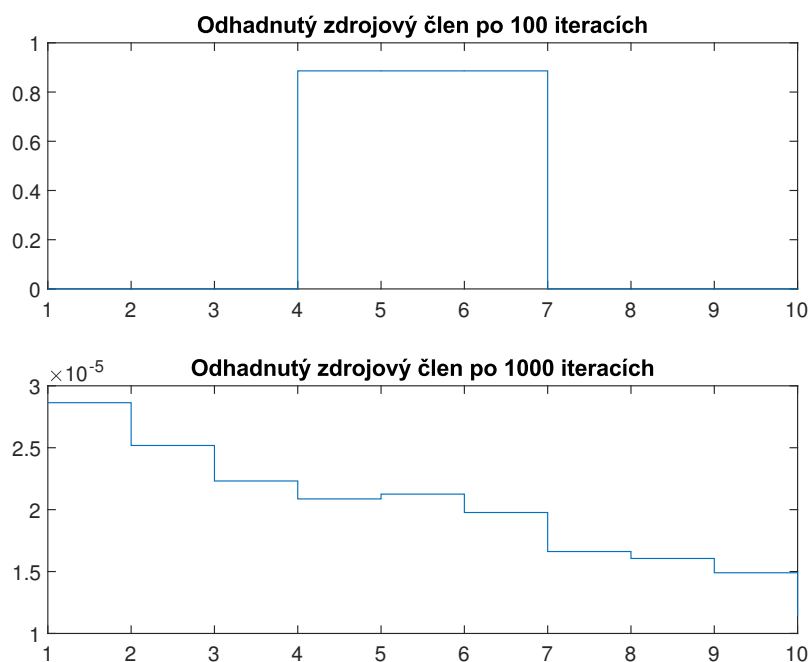
Na obrázku 4.2 dále vidíme pomalou konvergenci hodnot  $\sum \hat{x}_i$  a  $\sum \hat{\Omega}_{i,j}$ , což podporuje závěr, že bez použití lokalizace není algoritmus LS-APC-w stabilní.

Počet odhadovaných parametrů snížíme použitím lokalizační masky. Jako první se nabízí zvolit diagonální matici patřičného rozměru se samými jedničkami. Jak vidíme na obrázku 4.3 už tato jednoduchá volba vede k perfektní rekonstrukci zdrojového členu.

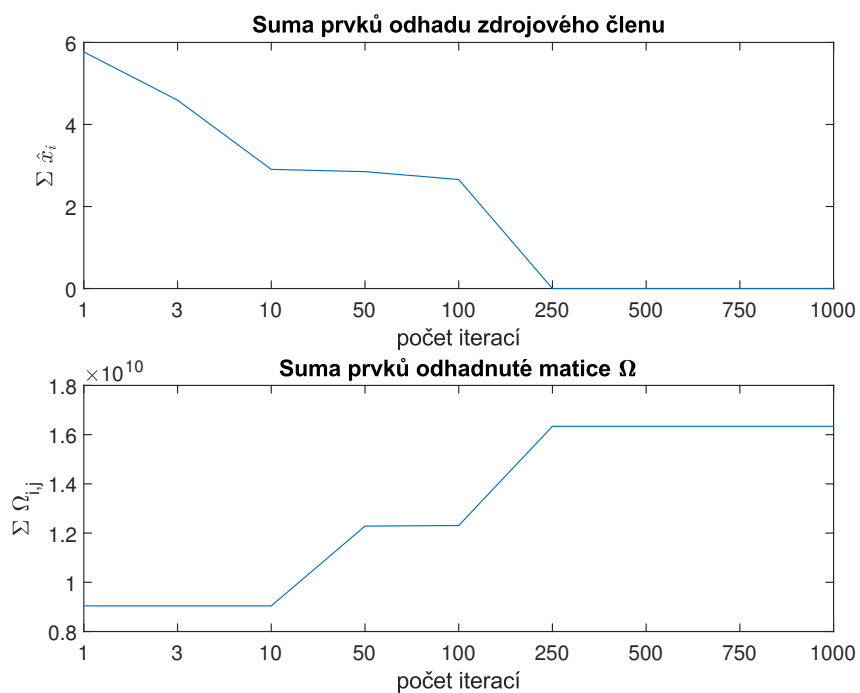
Ke konvergenci hodnot  $\sum \hat{x}_i$  a  $\sum \hat{\Omega}_{i,j}$  při použití diagonální lokalizační masky na obrázku 4.4 dochází už po několika iteracích.

### 4.1.2 Vliv lokalizace na kvalitu odhadu

Dále na syntetických datech ukážeme schopnost algoritmu LS-APC-w vypořádat se s zašumělými daty a vliv počtu nenulových prvků lokalizační masky na kvalitu odhadu.

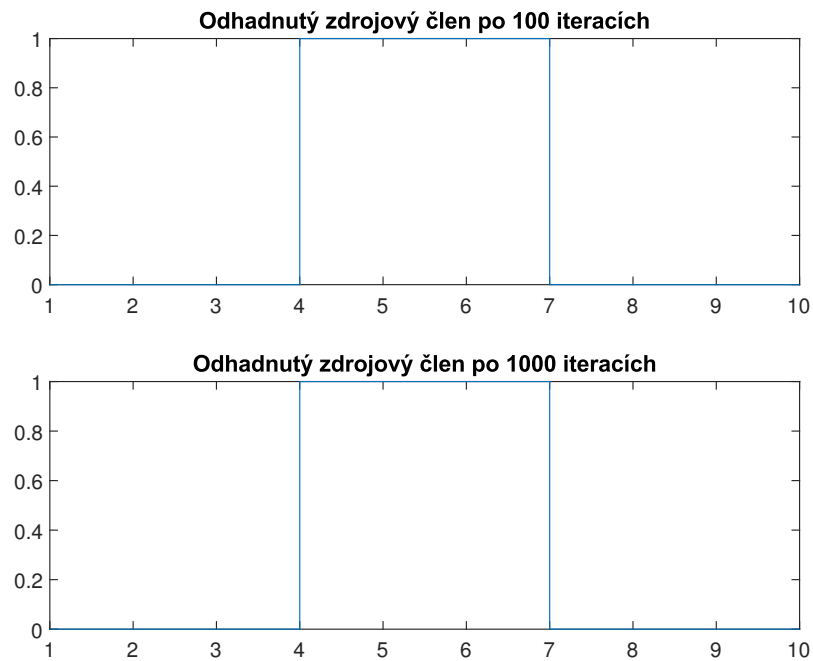


Obrázek 4.1: Odhadnutý zdrojový člen po 100 a 1000 iteracích LS-APC-w bez lokalizace.

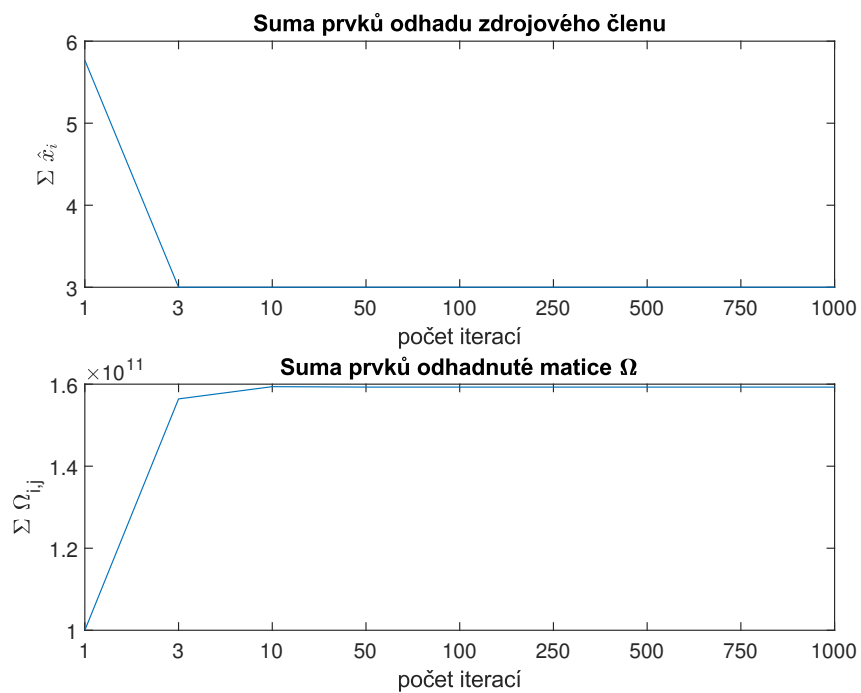


Obrázek 4.2: Vliv počtu iterací na stabilitu odhadu algoritmem LS-APC-w bez lokalizace.





Obrázek 4.3: Odhadnutý zdrojový člen po 100 a 1000 iteracích LS-APC-w s diagonální lokalizační maskou.



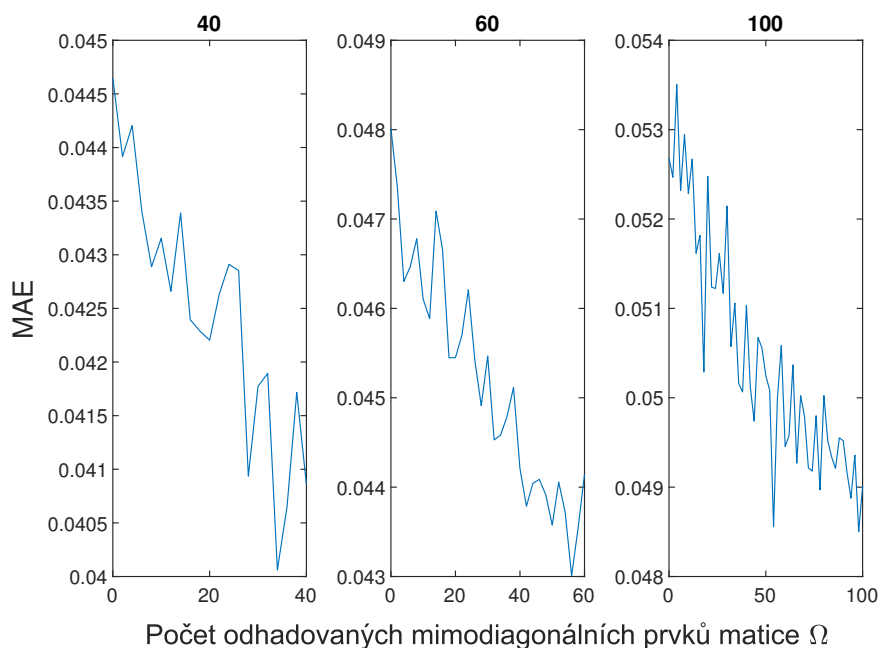
Obrázek 4.4: Vliv počtu iterací na stabilitu odhadu algoritmem LS-APC-w s diagonální lokalizační maskou.

Pro každý použitý počet nenulových mimodiagonálních prvků lokalizační masky (tj. navíc k diagonálním)  $0, 1, 2, \dots, k$  zopakujeme 1000x následující:

- Vygenerujeme prvky matice  $M \in \mathbb{R}^{20 \times 10}$  z rovnoměrného rozdělení mezi 0 a 1. Prvky menší než 0.5 navíc nahradíme nulami pro zhoršení podmíněnosti matice.
- Jako skutečný zdrojový člen vezmeme opět vektor  $\mathbf{x}_{\text{true}} = [0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0]^T$ .
- Vygenerujeme symetrickou pozitivně semidefinitní matici  $\Sigma \in \mathbb{R}^{20 \times 20}$  s právě  $k$  nenulovými mimodiagonálními prvky.
- Vytvoříme lokalizační masku s požadovaným počtem nenulových mimodiagonálních prvků.
- Vygenerujeme vektor šumu  $\mathbf{e} \in \mathbb{R}^{20 \times 1}$  z vícerozměrného normálního rozdělení  $\mathcal{N}(\mathbf{0}, \Sigma)$ .
- Vektor  $\mathbf{y}$  získáme jako  $\mathbf{y} = M\mathbf{x}_{\text{true}} + c\mathbf{e}$ , kde skalár  $c$  vyjadřuje míru zašumění a jeho hodnotu budeme měnit. Případné záporné prvky se nahradí nulami.
- Nakonec získáme odhad zdrojového členu  $\hat{\mathbf{x}}$  a můžeme jej porovnat s  $\mathbf{x}_{\text{true}}$  pomocí hodnoty  $\text{MAE}_x$ , kterou úplně nakonec zprůměrujeme přes všech 1000 iterací.

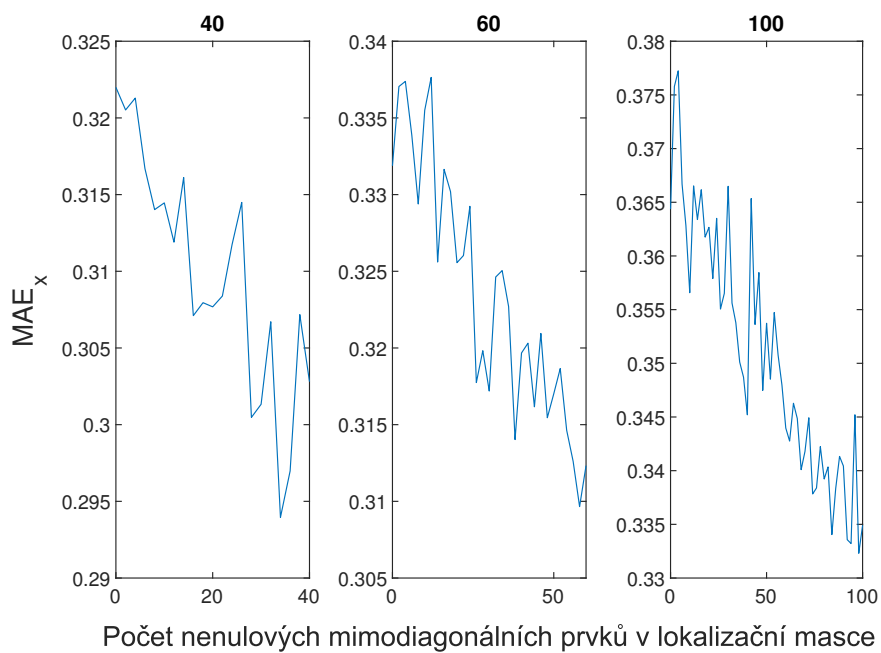
Na obrázku 4.5 je závislost průměrné kvality odhadu ( $\text{MAE}_x$ ) zdrojového členu na počtu nenulových mimodiagonálních prvků lokalizační masky pro různý počet nenulových mimodiagonálních prvků v kovarianční matici šumu  $\mathbf{e}$ . Použitá míra zašumění  $c$  je 0.1, tedy relativně nízká. I přes menší oscilace lze jednoznačně vidět trend snižování  $\text{MAE}_x$ . Stejný trend lze vidět i při mnohem vyšší míře zašumění  $c = 0.8$  na obrázku 4.6.

Počet nenulových mimodiagonálních prvků v kovarianční matici šumu



Obrázek 4.5:  $MAE_x$  v závislosti na počtu nenulových mimodiagonálních prvků lokalizační masky. Míra zašumění  $c$  je 0.1.

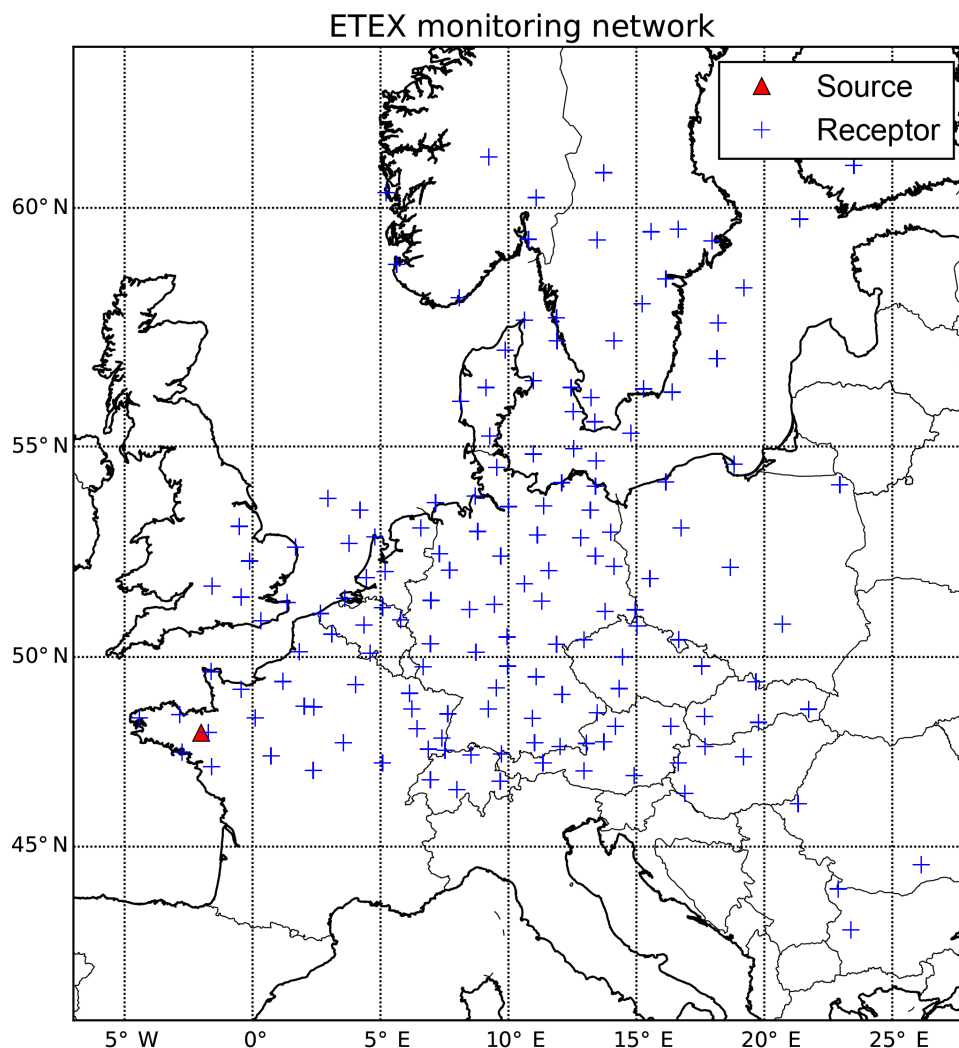
Počet nenulových mimodiagonálních prvků v kovarianční matici šumu



Obrázek 4.6:  $MAE_x$  v závislosti na počtu nenulových mimodiagonálních prvků lokalizační masky. Míra zašumění  $c$  je 0.8.

## 4.2 Data z experimentu ETEX

V rámci experimentu ETEX (European Tracer Experiment) byly během října a listopadu 1994 provedeny dva testovací úniky[25]. Látka PMCH (perfluoromethylcyclohexane) byla vypuštěna do atmosféry v Bretani a po 72 hodin po úniku byla měřena její přítomnost na 168 měřicích stanicích v 17 zemích Evropy. Mapa se zdroji úniku a měřicími stanicemi je na obrázku 4.7.



Obrázek 4.7: Zdroj úniku (červený trojúhelník) a měřicí stanice (modré kříže) v experimentu ETEX. Převzato z [32].

Kromě toho, že ETEX je jedním z mála takto velkých kontrolovaných úniků, je k němu také k dispozici velké množství informací. Pro nás je hlavní, že známe přesně kolik látky uniklo, kdy a kde k úniku došlo a jak dlouho trval. Odhadnutý zdrojový člen pak můžeme srovnat s realitou.

Pro získání SRS matice pro ETEX dataset byl použit atmosferický transportní model HYSPLIT[11] puštěného v dopředném režimu. Jako meteorologická reanalýza byla použita data ERA-Interim[8]

od Evropského centra pro střednědobé předpovědi počasí (ECMWF) s prostorovým rozlišením 0.5 stupně. Prostorová doména:  $-5^\circ - 30^\circ \text{ E}$ ,  $40^\circ - 65^\circ \text{ N}$ . Časová doména: 22. října 00:00 – 28. října 00:00 s rozlišením 1 hodina. Celkem je tedy sledováno 144 hodin.

Použijeme data z prvního úniku, kdy bylo vypuštěno 340 kg PMCH rovnoměrně necelých 12 hodin. Měření byla prováděna na 168 stanicích co tři hodiny po dobu 72 hodin. Máme tedy k dispozici vektor naměřených dat  $\mathbf{y} \in \mathbb{R}^{5040 \times 1}$ , SRS matici  $\mathbf{M} \in \mathbb{R}^{5040 \times 144}$  a skutečný zdrojový člen  $\mathbf{x}_{\text{true}} \in \mathbb{R}^{144 \times 1}$ .

Pro zlepšení numerické stability výpočtu normalizujeme vektor  $\mathbf{y}$ , který obsahuje řádově rozdílné hodnoty. Normalizaci provedeme vydělením všech jeho složek největší složkou. Stejnou hodnotou pak vydělíme všechny prvky matice  $\mathbf{M}$ . Tato úprava nemá žádný vliv na hledaný odhad zdrojového členu  $\mathbf{x}$ . Dále vektor  $\mathbf{y}$  obsahuje záporné hodnoty, které z fyzikálního hlediska jako koncentrace nedávají smysl. Tyto hodnoty byly spolu s příslušnými řádky matice  $\mathbf{M}$  odstraněny. Zůstane nám tedy vektor naměřených dat  $\mathbf{y} \in \mathbb{R}^{3104 \times 1}$  a SRS matice  $\mathbf{M} \in \mathbb{R}^{3104 \times 144}$ .

#### 4.2.1 Volba lokalizační masky podle vzdálenosti měřicích stanic

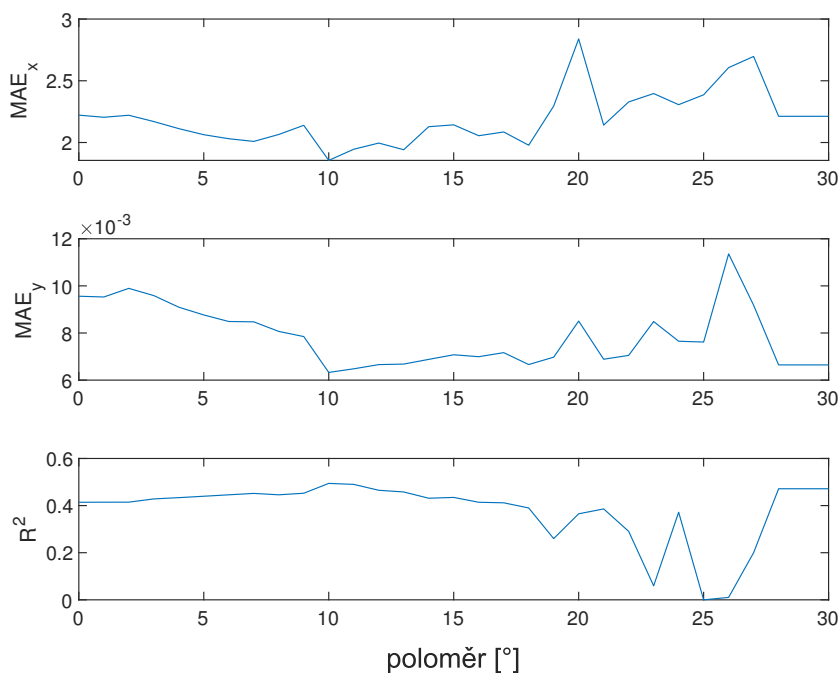
Jak jsme již ukázali na syntetických datech, použití lokalizační masky a její struktura má zásadní vliv na výsledný odhad zdrojového členu. Nyní zkusíme využít informace o polohách jednotlivých měřicích stanic. Lokalizační masku sestojíme tak, že budeme uvažovat korelace pouze mezi měřeními ze stanic vzdálených méně nebo rovno danému poloměru (ve stupních). Prozatím bude maska binární, tedy modelovat nebo nemodelovat korelaci.

Samozřejmě čím více má lokalizační maska nenulových prvků, tím více se blížíme případu, kdy ji vůbec nepoužíváme. To, jak jsme ukázali na syntetických datech, vede k nestabilitě algoritmu a špatnému odhadu zdrojového členu. Z obrázku 4.8 je však patrné, že nelze volit poloměr ani příliš malý.

Podle všech metrik vychází jednoznačně nejlépe poloměr  $10^\circ$ . Odhadnutý zdrojový člen algoritmem LS-APC-w s poloměrem  $10^\circ$  v porovnání s odhadem z klasického LS-APC a skutečným zdrojovým členem je na obrázku 4.9. Odhad úniku je koncentrován do správné časové oblasti až na menší peak kolem 110. hodiny. Hodnoty metrik a celková suma úniku ve srovnání s algoritmy z výzkumného úkolu jsou v tabulce 4.1. Podle všech metrik je odhad pomocí LS-APC-w s poloměrem  $10^\circ$  přesnější než klasické LS-APC i všechny ostatní algoritmy.. Celková suma úniku je však ještě více podhodnocena, což je částečně způsobeno tím, že klasické LS-APC kompenzovalo celkovou sumu vytvořením peaků se skoro až dvojnásobnou hodnotou oproti skutečnosti.

#### 4.2.2 Nebinární lokalizační maska

Dosud jsme lokalizační masku volili binární – tj. obsahující pouze dvě hodnoty a to 0 pokud se příslušná korelace nemá modelovat a 1 pokud má. Nyní však připustíme veškeré hodnoty z uzavřeného intervalu  $[0, 1]$ . Stále budeme uvažovat korelace pouze mezi měřeními ze stanic vzdálených méně nebo rovno danému poloměru, prvky lokalizační masky ale budou následujícími funkcemi vzdálenosti.



Obrázek 4.8: Graf závislosti metrik kvality odhadu na poloměru.

V lineárním případě

$$\Lambda_{i,j} = \begin{cases} 1 - \frac{1}{r}x_{i,j} & \forall x_{i,j} \leq r, \\ 0 & \text{jinak.} \end{cases} \quad (4.6)$$

V kvadratickém případě

$$\Lambda_{i,j} = \begin{cases} \frac{1}{r^2}x_{i,j}^2 - \frac{2}{r}x_{i,j} + 1 & \forall x_{i,j} \leq r, \\ 0 & \text{jinak,} \end{cases} \quad (4.7)$$

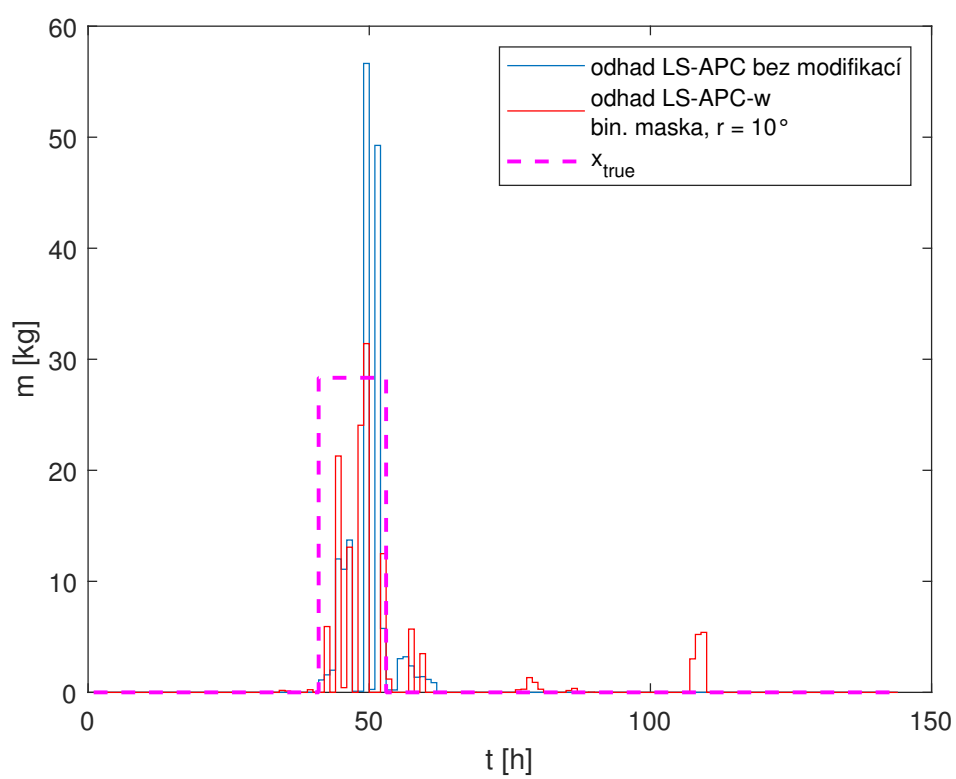
kde koeficienty funkce byly zvoleny, aby funkce protínala body  $[0, 1]$  a  $[r, 0]$ .

A v exponenciálním případě

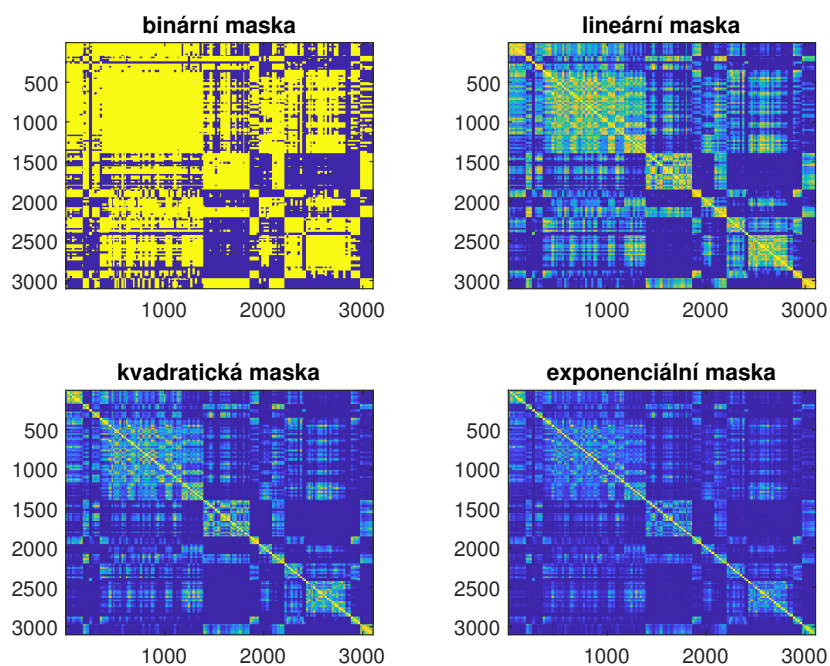
$$\Lambda_{i,j} = \begin{cases} \exp\left(-\frac{4}{r}x_{i,j}\right) & \forall x_{i,j} \leq r, \\ 0 & \text{jinak.} \end{cases} \quad (4.8)$$

Ve všech třech funkcích je  $r$  zvolený poloměr ve stupních a  $x_{i,j}$  je vzdálenost měřicích stanic  $i$ -tého a  $j$ -tého měření. Pro představu jak konkrétně vypadají různé druhy lokalizačních masek pro konkrétní poloměr  $10^\circ$ , je na obrázku 4.10 jejich grafické znázornění.

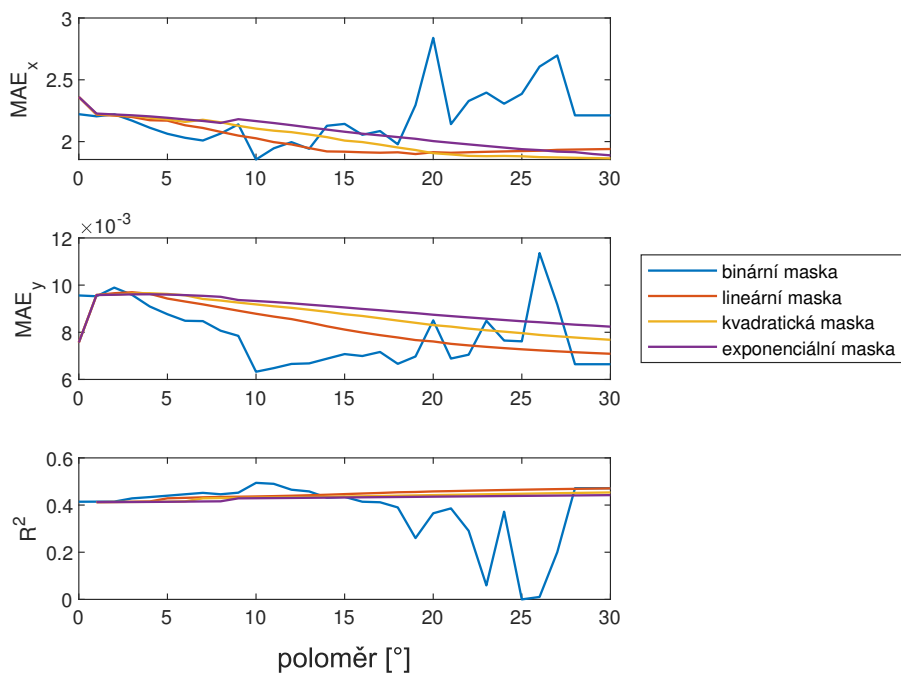
Metriky kvality odhadu při použití různých druhů lokalizačních masek v závislosti na poloměru jsou na obrázku 4.11. Lze vidět, že mezi jednotlivými nebinárními maskami není až tak velký rozdíl, nejlépe však dopadla maska lineární. Při jejich použití je lepší zvolit co největší poloměr, ale vidíme že na jeho volbu nejsou zdaleka tak citlivé jako binární maska. Hodnoty metrik a celková suma úniku pro různé volby lokalizačních masek jsou v tabulce 4.2. Použití binární masky s poloměrem  $10^\circ$  stále dosahuje lehce lepších výsledků. Nebinární masky ale stále mají využití pokud bychom např. neznali polohy měřicích stanic.



Obrázek 4.9: Odhad zdrojového členu pomocí LS-APC-w s poloměrem  $10^\circ$  v porovnání s odhadem z klasického LS-APC a skutečným zdrojovým členem.



Obrázek 4.10: Grafické znázornění lokalizačních masek pro  $r = 10^\circ$ .



Obrázek 4.11: Graf závislosti metrik kvality odhadu na poloměru pro různé druhy lokalizačních masek.



algoritmus	MAE <sub>x</sub>	MAE <sub>y</sub>	R <sup>2</sup>	suma úniku [kg]
LS-APC bez modifikací	2.0744	0.0072286	0.38403	167.5234
LS-APC-Ovolne	2.2087	0.0066647	0.4436	37.0178
LS-APC-O250	2.2318	0.0070697	0.35339	18.6276
LS-APC-Ozeme	2.2681	0.0068527	0.30527	43.6721
LS-APC-w bin. maska, r=10°	1.8445	0.0063201	0.50081	136.8139

Tabulka 4.1: Hodnoty metrik kvality odhadu a celková suma úniku pro LS-APC-w s poloměrem 10 ° ve srovnání s ostatními algoritmy.

algoritmus	MAE <sub>x</sub>	MAE <sub>y</sub>	R <sup>2</sup>	suma úniku [kg]
LS-APC bez modifikací	2.0744	0.0072286	0.38403	167.5234
LS-APC-w bin. maska, r=10°	1.8445	0.0063201	0.50081	136.8139
LS-APC-w lin. maska, r=19°	1.8993	0.0076696	0.45543	227.9523
LS-APC-w kvad. maska, r=30°	1.8641	0.0076816	0.45343	229.3428
LS-APC-w exp. maska, r=30°	1.8889	0.0082379	0.44208	254.8208

Tabulka 4.2: Hodnoty metrik a celková suma úniku pro různé volby lokalizačních masek.

### 4.2.3 Volba lokalizační masky podle času měření

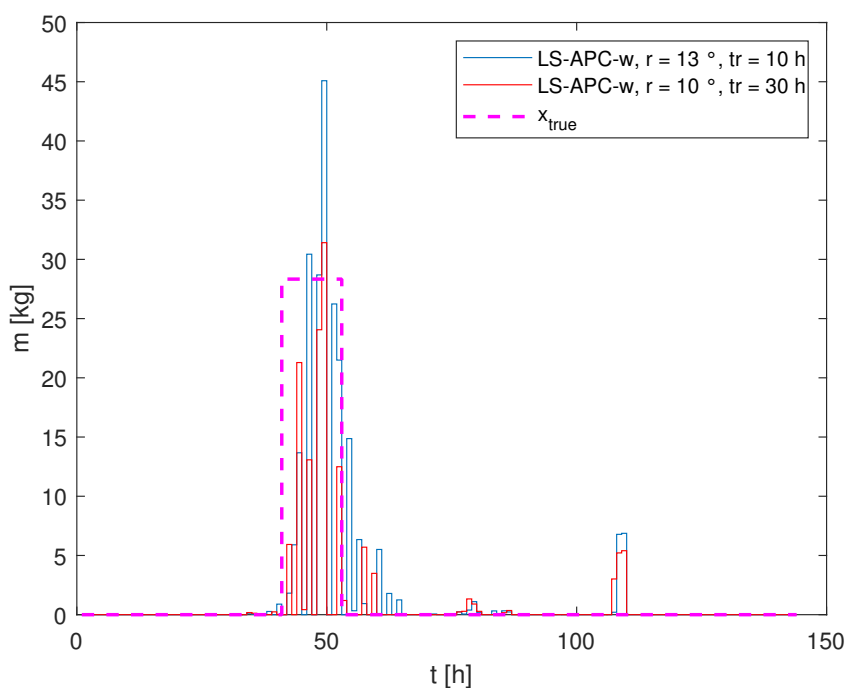
Při sestrojování lokalizační masky  $\Lambda$  můžeme využít nejen informace o polohách měřících stanic, ale máme k dispozici i čas (s přesností na hodiny), kdy k měření došlo. Při tvorbě masky stále budeme uvažovat korelace pouze mezi měřeními ze stanic vzdálených méně nebo rovno danému poloměru (ve stupních). Navíc se ale omezíme na měření, která proběhla méně než daný čas v hodinách od sebe. Tomuto maximálnímu časovému rozdílu budeme říkat *časový poloměr*. Výsledná maska bude opět pouze binární.

Na obrázku 4.13 jsou metriky kvality odhadu zdrojového členu v závislosti na poloměru a časovém poloměru při použití algoritmu LS-APC-w. Podle MAE<sub>y</sub> a R<sup>2</sup> bychom čas neměli vůbec uvažovat a brát všechna měření, která jsou ze stanic vzdálených méně než 10 ° stejně jako v podsekcí 4.2.1. Podle MAE<sub>x</sub> bychom měli vzít lehce větší poloměr 13 ° a časový poloměr zvolit 10 hodin. Za povšimnutí stojí, že poměrně špatný odhad dostáváme při použití velkého poloměru i časového poloměru.

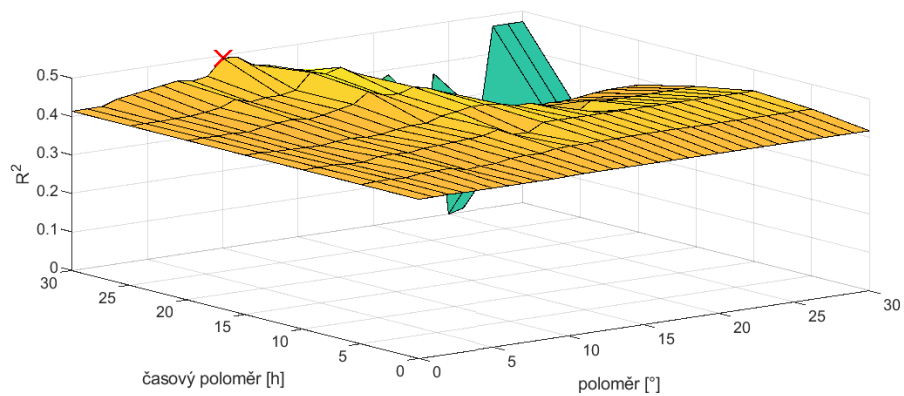
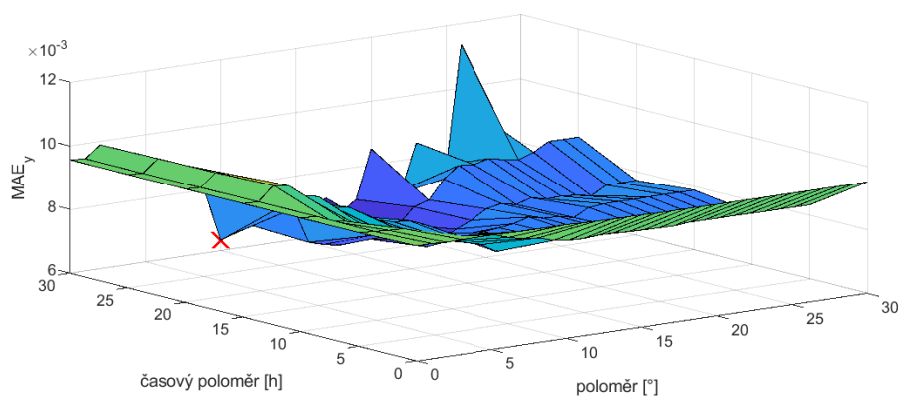
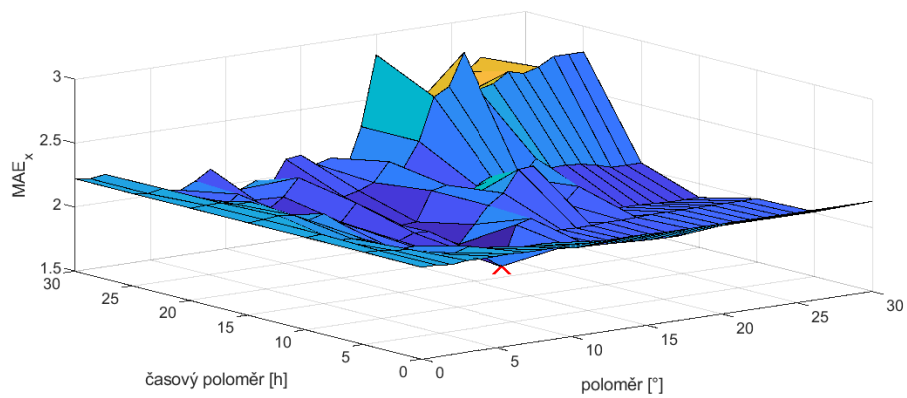
Odhady získané algoritmem LS-APC-w za použití těchto dvou navržených nastavení pro lokalizační masku srovnáme na obrázku 4.12. Oba odhady jsou koncentrovány do správné časové oblasti až na menší peak kolem 110. hodiny. Hodnoty metrik kvality odhadu a odhad celkové sumy úniku jsou uvedeny v tabulce 4.3. Odhad při použití masky s omezeným časovým poloměrem je nejbližší skutečné celkové sumě úniku 340 kg a dosahuje nejlepší hodnoty MAE<sub>x</sub> ze všech dosud zkoušených nastavení. Hodnoty MAE<sub>y</sub> a R<sup>2</sup> sice nejsou nejlepší, ale stále srovnatelné s ostatními zkoušenými nastaveními lokalizační masky.

algoritmus	MAE <sub>x</sub>	MAE <sub>y</sub>	R <sup>2</sup>	suma úniku [kg]
LS-APC bez modifikací	2.0744	0.0072286	0.38403	167.5234
LS-APC-w r=10°, tr=30h	1.8445	0.0063201	0.50081	136.8139
LS-APC-w r=13°, tr=10h	1.7659	0.0075910	0.44290	222.5300

Tabulka 4.3: Hodnoty metrik kvality odhadu a odhad celkové sumy úniku za použití nastavení pro lokalizační masku získaných v podsekcí 4.2.3.



Obrázek 4.12: Odhady zdrojového členu získané algoritmem LS-APC-w za použití nastavení pro lokalizační masku získaných v podsekcí 4.2.3.



Obrázek 4.13: Metriky kvality odhadu v závislosti na poloměru a časovém poloměru. Červeným křížkem jsou vyznačena globální minima, resp. maximum u  $R^2$ .

## 4.3 Únik ruthenia-106 v roce 2017

Pro únik ruthenia-106 do atmosféry na podzim roku 2017, který je blíže popsán v úvodu práce, máme k dispozici dva datasety. Použijeme oba a poté provedeme jejich srovnání mezi sebou a s dostupnou literaturou.

### 4.3.1 Dataset IAEA

Do tohoto datasetu jsou zahrnuta měření koncentrací provedená členskými státy organizace IAEA[17], která jsme doplnili o měření z článku Shershakov et al.[29]. Dataset má podobu tabulek obsahující polohu stanice kde bylo prováděno měření, čas začátku a konce měření a naměřené koncentrace. Některá měření mají také uvedenou nejistotu v procentech. Celkem dataset obsahuje 451 naměřených koncentrací. Tyto tabulky bylo nutno poloautomaticky převést do formátu atmosférického modelu HYSPLIT.

Atmosférický model HYSPLIT byl spuštěn ve zpětném chodu s následujícími parametry

- Prostorová doména:  $5^\circ - 115^\circ$  E,  $25^\circ - 65^\circ$  N.
- Horizontální diskretizace domény:  $0.5 \times 0.5$  stupňů. Vertikální diskretizace je brána automaticky z použitých meteorologických dat.
- Výsledné koncentrace jsou brány z vrstvy 0 – 300 metrů nad zemí.
- Časová doména je 20. 9. 2017 0:00 až 10. 10. 2017 18:00 diskretizována po 6 hodinách.
- Jako meteorologická reanalýza jsou použita data GFS (Global Forecast System) s rozlišením  $0.5 \times 0.5$  stupňů s časovou diskretizací 6 hodin. Vertikálně 26 vrstev pokrývající nerovnoměrně celou troposféru.

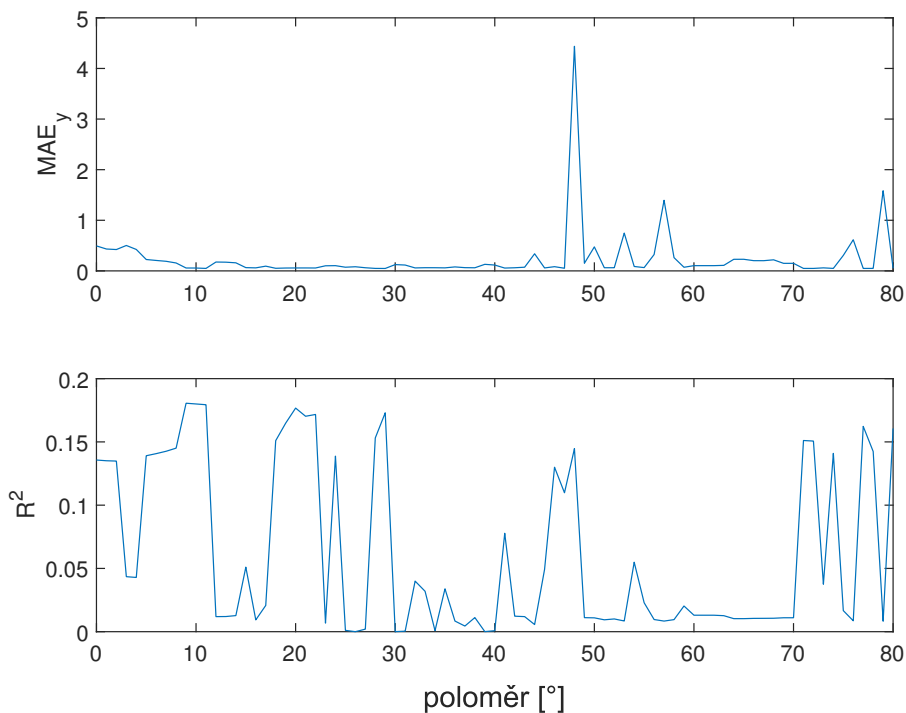
Prostorová doména je horizontálně rozdělena na 17901 buněk velikosti  $0.5 \times 0.5$  stupňů. Výstupem z atmosférického modelu HYSPLIT je tedy 17901 SRS matic  $M$ . Veškeré odhady zdrojového členu budeme provádět za předpokladu, že k úniku došlo z kombinátu Majak. Tato hypotéza je dobře podpořena literaturou (viz podsekcce 4.3.4) a ke stejnému závěru jsme došli ve výzkumném úkolu. Používat tedy budeme SRS matici  $M$  příslušnou buňce, ve které se Majak nachází.

#### 4.3.1.1 Volba lokalizační masky podle vzdálenosti měřicích stanic

Máme tedy připravená data a představu o jejich struktuře. Stejně jako v sekci 4.2.1 využijeme znalost poloh jednotlivých měřicích stanic. Lokalizační masku opět sestojíme binární tak, že budeme uvažovat korelace pouze mezi měřeními ze stanic vzdálených méně nebo rovno danému poloměru (ve stupních).

Na obrázku 4.14 jsou metriky kvality odhadu v závislosti na zvoleném poloměru. Protože nemáme k dispozici skutečný zdrojový člen, nemáme jak spočítat metriku  $MAE_x$ . Metriky celkem dost oscilují, což může ukazovat na nestabilitu algoritmu na těchto datech. Vykreslíme si tedy graf zdrojového členu při použití LS-APC-w s poloměrem dosahujícím nejlepších

metrik (29 °) na obrázku 4.15. Vidíme, že výsledný odhad se absolutně neshoduje s předchozími výsledky 4.3.4 a to ani umístěním v čase ani magnitudou úniku. Tento problém je snadné opravit a učiníme tak v následující podpodsekcí.



Obrázek 4.14: Graf závislosti metrik kvality odhadu na poloměru.

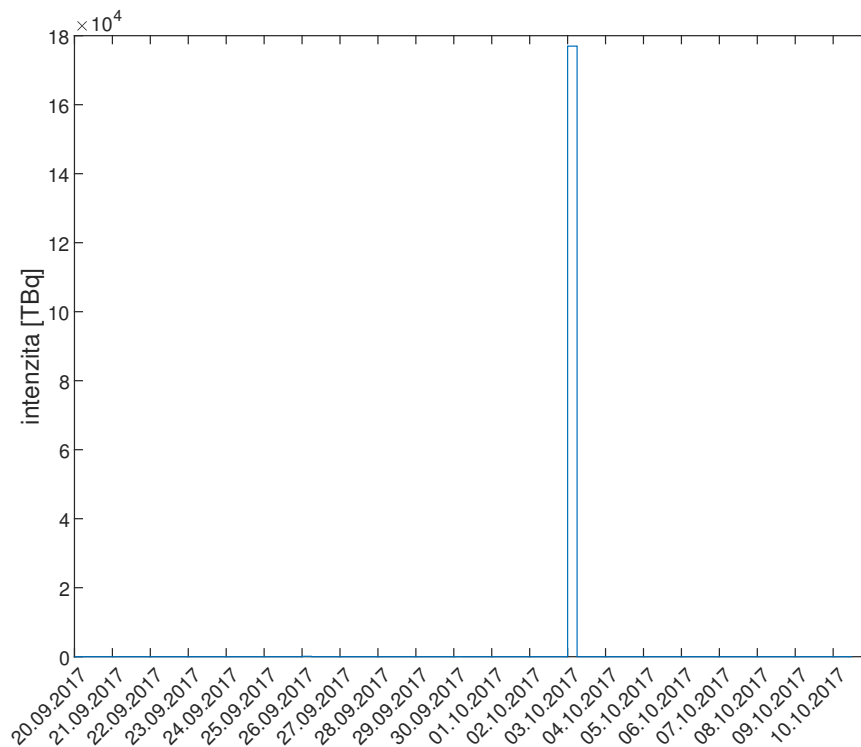
#### 4.3.1.2 Zlepšení odhadu pomocí volby parametrů

Jak bylo ukázáno v (3.46), apriorní rozdělení matice  $\mathbf{\Omega}$  má tvar  $f(\mathbf{\Omega}) = \mathcal{W}(\rho_0 \mathbf{I}_p, \vartheta_0)$ . Parametry  $\vartheta_0$  a  $\rho_0$  jsou předem zvoleny tak, abychom získali co nejméně informativní rozdělení a zároveň zachovali numerickou stabilitu. Konkrétní zvolená hodnota je  $\vartheta_0 = \rho_0^{-1} = 10^{-10}$ .

Experimentálně jsme zjistili, že změna hodnoty těchto dvou parametrů má pozitivní vliv na výsledný odhad algoritmem LS-APC-w. V textu budeme mluvit o změně hodnoty parametru  $\vartheta_0$ , ale s parametrem  $\vartheta_0$  vždy měníme i  $\rho_0$  se zachováním rovnosti  $\vartheta_0 = \rho_0^{-1}$ .

Celkem jsme vyzkoušeli hodnoty  $\vartheta_0 \in \{10^{-10}, 10^{-9}, \dots, 10^{-2}, 10^{-1}\}$ , ale pro přehlednost jsou na obrázku 4.16 pouze metriky pro volby  $10^{-3}$  a  $10^{-2}$ , které dopadly nejlépe. Zbavili jsme se oscilací hodnot metrik a v průměru se zlepšily.

Vykreslíme si odhadnutý zdrojový člen pro volby  $10^{-3}$  a  $10^{-2}$  za použití poloměru dosahujícího nejlepších metrik na obrázku 4.17. Získané odhady jsou nyní mnohem blíže publikovaným výsledkům 4.3.4. Hodnoty metrik kvality odhadu a celková suma úniku pro různé volby parametru  $\vartheta_0$  jsou v tabulce 4.4. Pozorujeme sice mírné zhoršení metrik, ale odhady zdrojových členů jsou mnohem smysluplnější. Zkontrolujeme ještě konvergenci na obrázku



Obrázek 4.15: Odhad zdrojového členu pomocí LS-APC-w s poloměrem 29 °.

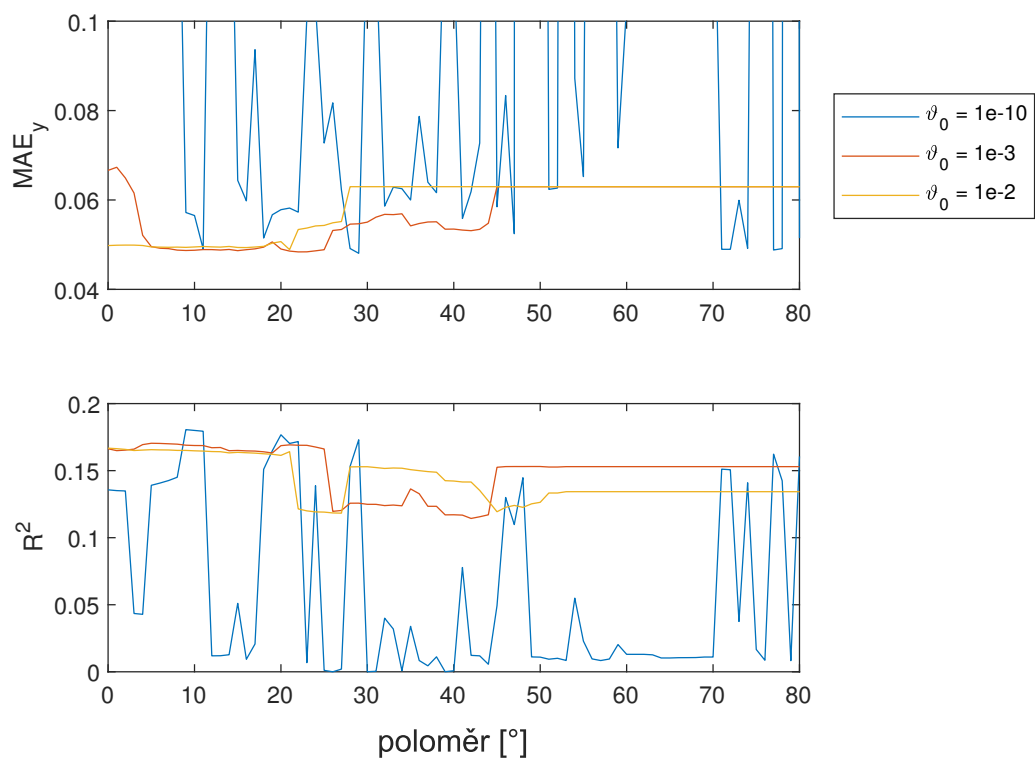
4.18. Vidíme pomalejší a horší konvergenci při volbě  $\vartheta_0 = 10^{-2}$ . Nakonec tedy volíme  $\vartheta_0 = \varrho_0^{-1} = 10^{-3}$ .

algoritmus	MAE <sub>y</sub>	R <sup>2</sup>	suma úniku [TBq]
LS-APC-w r=29°, $\vartheta_0=1e-10$	0.048075	0.17301	177102.7068
LS-APC-w r=22°, $\vartheta_0=1e-3$	0.048383	0.16894	185.9156
LS-APC-w r=21°, $\vartheta_0=1e-2$	0.048895	0.16413	122.8141

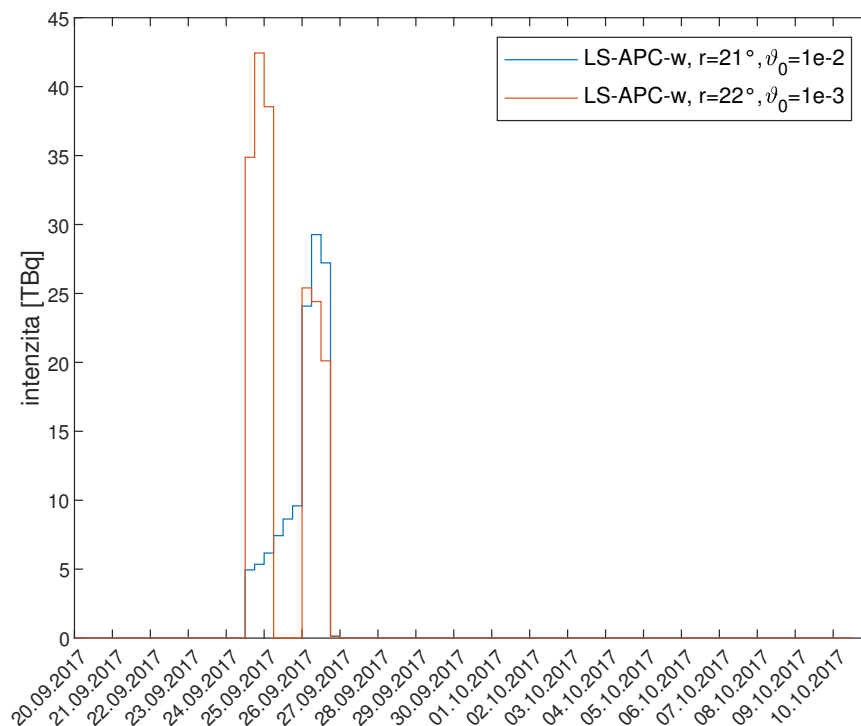
Tabulka 4.4: Metriky kvality odhadu a celková suma úniku pro různé volby parametru  $\vartheta_0$  a pro ně optimální poloměr.

#### 4.3.1.3 Nebinární lokalizační maska

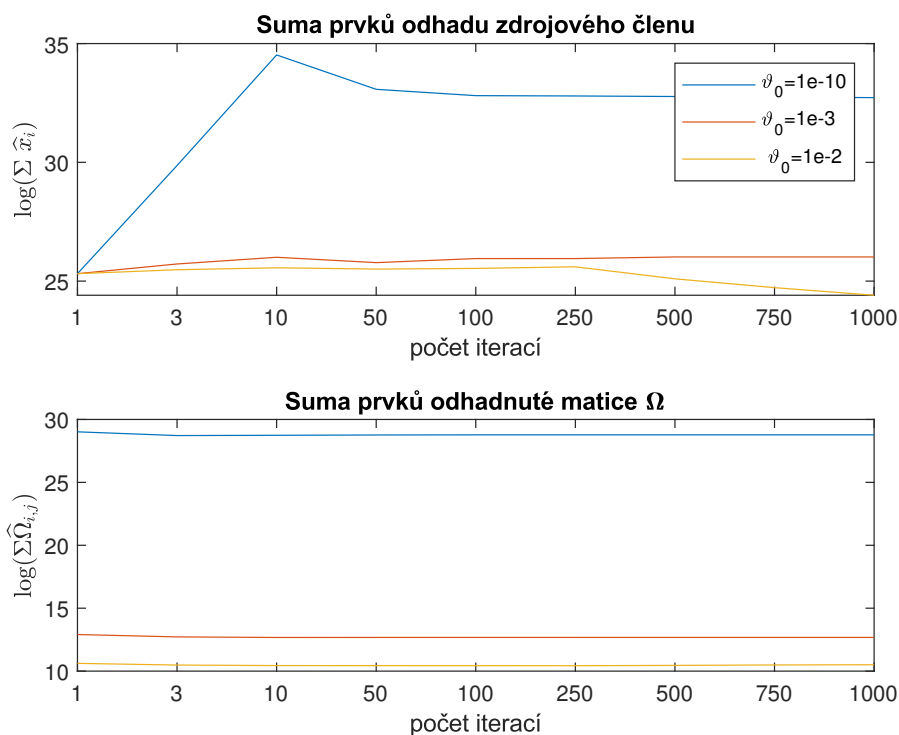
Nebinární lokalizační masky vytvoříme a budeme používat stejným způsobem jako v podsekcí 4.2.2. V algoritmu LS-APC-w použijeme hodnotu parametru  $\vartheta_0 = 10^{-3}$ . Metriky kvality odhadu při použití různých druhů lokalizačních masek v závislosti na poloměru jsou na obrázku 4.19 a v tabulce 4.5. Všechny nebinární masky se chovají podobně a při vyšších poloměrech mezi nimi dokonce skoro není rozdíl. Oproti binární masce nejsou nebinární tak citlivé na poloměr. Binární maska s poloměrem 22 ° a  $\vartheta_0 = 10^{-3}$  stále dosahuje lepší hodnoty MAE<sub>y</sub> a hodnoty R<sup>2</sup> jsou srovnatelné.



Obrázek 4.16: Graf závislosti metrik kvality odhadu na poloměru pro různé volby parametru  $\vartheta_0$ .

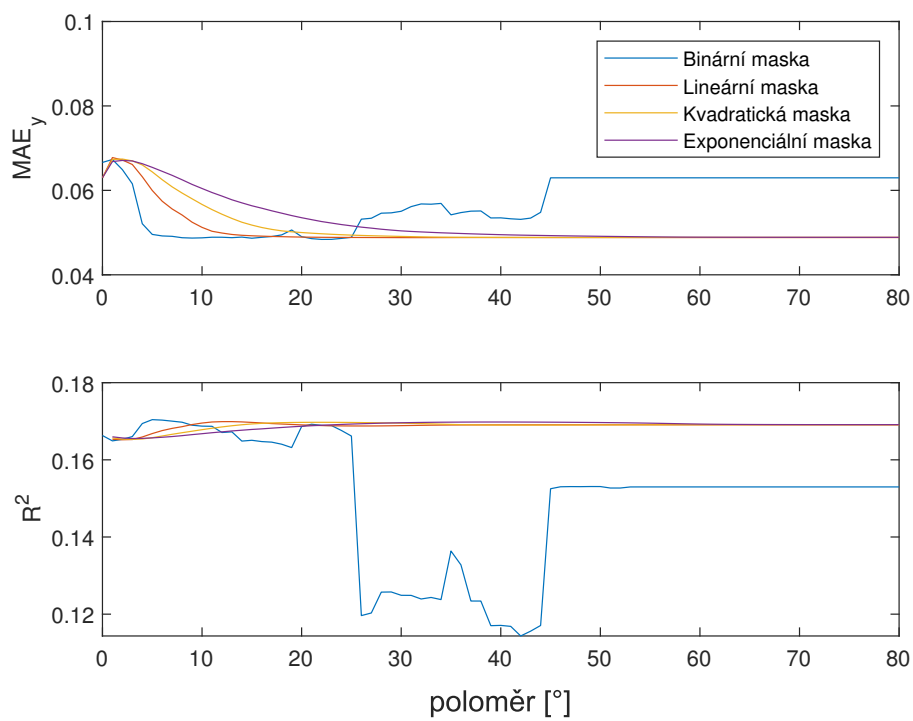


Obrázek 4.17: Odhad zdrojového členu pomocí LS-APC-w pro různé volby parametru  $\vartheta_0$ .



Obrázek 4.18: Graf konvergence algoritmu pro různé volby parametru  $\vartheta_0$ .





Obrázek 4.19: Graf závislosti metrik kvality odhadu na poloměru pro různé druhy lokalizačních masek.

algoritmus	MAE <sub>y</sub>	R <sup>2</sup>	suma úniku [TBq]
LS-APC-w bin. maska, r=22°	0.048383	0.16894	185.9156
LS-APC-w lin. maska, r=31°	0.048805	0.1689	232.8744
LS-APC-w kvad. maska, r=48°	0.048845	0.16903	235.2572
LS-APC-w exp. maska, r=80°	0.048875	0.16912	236.1884

Tabulka 4.5: Hodnoty metrik a celková suma úniku pro různé volby lokalizačních masek. Parametr  $\vartheta_0$  byl ve všech případech nastaven na  $10^{-3}$ .

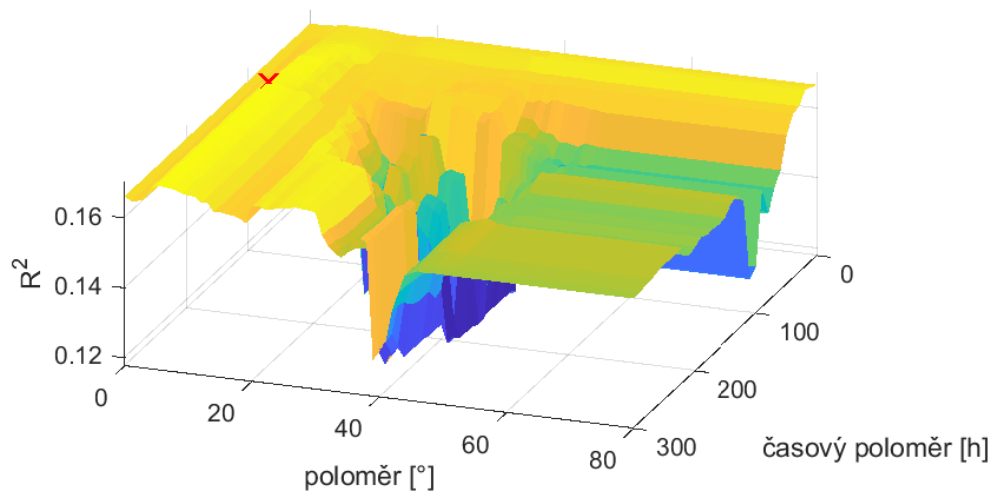
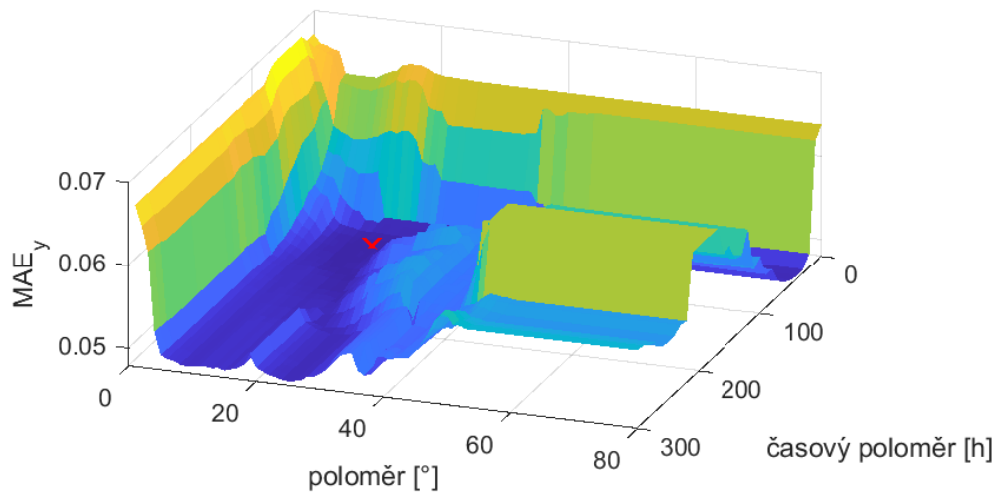
#### 4.3.1.4 Volba lokalizační masky podle času měření

Stejně jako v sekcích 4.2.3 využijeme znalosti časů, kdy k jednotlivým měřením došlo. Lokalizační masku dále omezíme na měření, která proběhla méně než daný čas v hodinách od sebe. Výsledná maska bude opět binární. Ve srovnání s daty z ETEX zde máme mnohem větší časovou doménu – skoro 900 hodin oproti 30.

Graf metrik kvality odhadu zdrojového členu v závislosti na poloměru a časovém poloměru při použití algoritmu LS-APC-w s  $\vartheta_0 = 10^{-3}$  je na obrázku 4.20. Nejlepší hodnoty MAE<sub>y</sub> dosahuje nastavení s poloměrem 16° a časovým poloměrem 70 hodin. Nejlepší hodnoty R<sup>2</sup> dosahuje nastavení s poloměrem 5° a časovým poloměrem 120 hodin. Hodnoty metrik kvality odhadu a odhad celkové sumy úniku jsou uvedeny v tabulce 4.6. Nelze optimalizovat obě metriky současně, ale nastavení  $r = 16^\circ$  a  $t_r = 70$  h zlepšuje obě metriky oproti odhadu, kde se časová vzdálenost neuvažuje.

algoritmus	MAE <sub>y</sub>	R <sup>2</sup>	suma úniku [TBq]
LS-APC-w r=22°	0.048383	0.16894	185.9156
LS-APC-w r=16°, tr=70h	0.047884	0.16571	188.9687
LS-APC-w r=5°, tr=120h	0.049378	0.17042	258.6762

Tabulka 4.6: Hodnoty metrik kvality odhadu a odhad celkové sumy úniku za použití lokalizační masky získaných v podsekcí 4.3.1.4.



Obrázek 4.20: Metriky kvality odhadu v závislosti na poloměru a časovém poloměru. Červeným křížkem je vyznačeno globální minimum u  $MAE_y$ , resp. maximum u  $R^2$ .

### 4.3.2 Dataset Ro5

Druhý dataset tvoří převážně měření ze stanic zapojených do Ro5 (Ring of Five) – neformální sdružení měřicích stanic a laboratoří, která navzájem spolupracují. Měření z dálného východu byla odstraněna, protože by neúměrně zvětšila nutnou prostorovou doménu. V datasetu tak zbude 899 měření.

Atmosférický model HYSPLIT byl spuštěn se stejnými parametry jako pro IAEA dataset až na prostorovou doménu, která byla  $-10^\circ - 100^\circ$  E,  $30^\circ - 75^\circ$  N. Z atmosférického modelu máme 20111 SRS matic  $M$ , každá odpovídající jedné  $0.5 \times 0.5$  stupňů buňce. Stejně jako u IAEA datasetu ale budeme předpokládat, že k úniku došlo z kombinátu Majak a pro veškeré odhady používat SRS matici příslušnou buňce s Majakem.

#### 4.3.2.1 Volba lokalizační masky podle vzdálenosti měřicích stanic

Máme tedy připravená data z Ro5 datasetu a představu o jejich struktuře. Stejně jako v sekci 4.2.1 využijeme znalost poloh jednotlivých měřicích stanic. Lokalizační masku opět sestrojíme binárně tak, že budeme uvažovat korelace pouze mezi měřeními ze stanic vzdálených méně nebo rovno danému poloměru (ve stupních).

Na obrázku 4.21 jsou metriky kvality odhadu v závislosti na zvoleném poloměru. Protože nemáme k dispozici skutečný zdrojový člen, nemáme jak spočítat metriku  $MAE_x$ . Vykreslíme si graf zdrojového členu při použití LS-APC-w s poloměrem dosahujícím nejlepších metrik ( $75^\circ$ ) na obrázku 4.22. Ten sice vypadá ve shodě s předchozími výsledky 4.3.4, ale na obrázku 4.23 vidíme velmi špatnou konvergenci odhadu. Tento problém opravíme v následující podsekci.

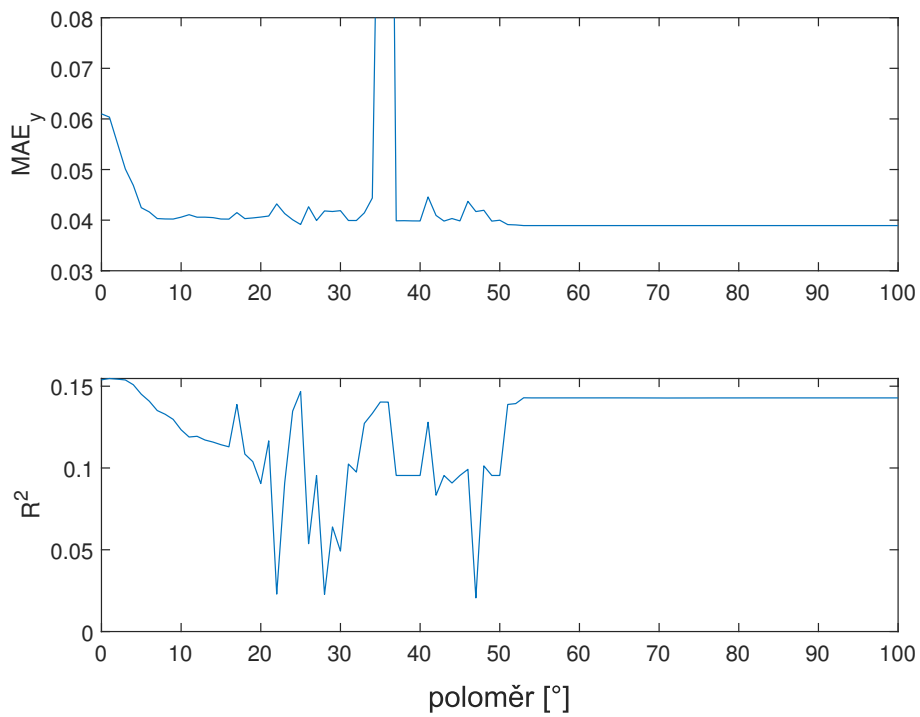
#### 4.3.2.2 Zlepšení odhadu pomocí volby parametrů

Stejně jako u IAEA datasetu v podsekci 4.3.1.2 budeme ladit parametry  $\vartheta_0$  a  $\varrho_0$  apriorního rozdělení matice  $\mathbf{\Omega}$ , které má tvar  $f(\mathbf{\Omega}) = \mathcal{W}(\varrho_0 \mathbf{I}_p, \vartheta_0)$ . Parametry opět budeme měnit oba společně při zachování rovnosti  $\vartheta_0 = \varrho_0^{-1}$ , ale v celém textu budeme pro přehlednost a stručnost mluvit pouze o optimalizaci parametru  $\vartheta_0$ .

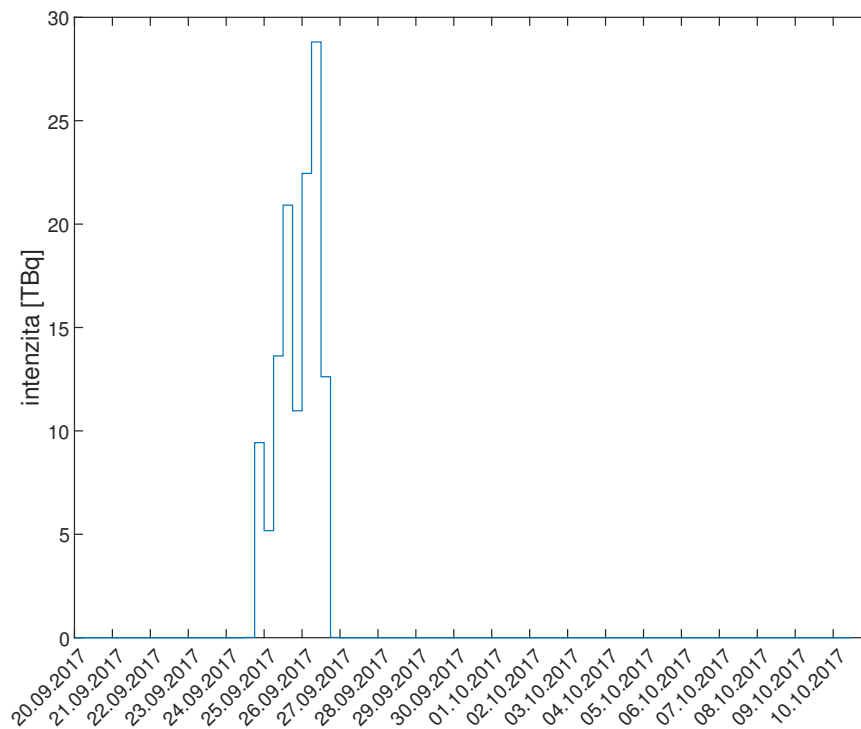
Vyzkoušeli jsme hodnoty  $\vartheta_0 \in \{10^{-10}, 10^{-9}, \dots, 10^{-2}, 10^{-1}\}$  a pro každou našli poloměr vedoucí na nejmenší  $MAE_y$ . Cílem je zlepšit konvergenci algoritmu. Pro přehlednost jsou na obrázku 4.24 pouze hodnoty parametrů, které vedou na dobrou konvergenci. Zásadní rozdíly mezi těmito odhady nejsou jak lze vidět na obrázku 4.25 a tabulce 4.7. Nakonec zvolíme  $\vartheta_0 = \varrho_0^{-1} = 10^{-5}$ , protože vede na odhad s nejrychlejší konvergencí a dosahuje nejmenší  $MAE_y$ .

algoritmus	$MAE_y$	$R^2$	suma úniku [TBq]
LS-APC-w $r=30^\circ, \vartheta_0=1e-3$	0.039137	0.14680	107.0667
LS-APC-w $r=32^\circ, \vartheta_0=1e-4$	0.039000	0.14570	94.6732
LS-APC-w $r=30^\circ, \vartheta_0=1e-5$	0.038946	0.14480	90.5274

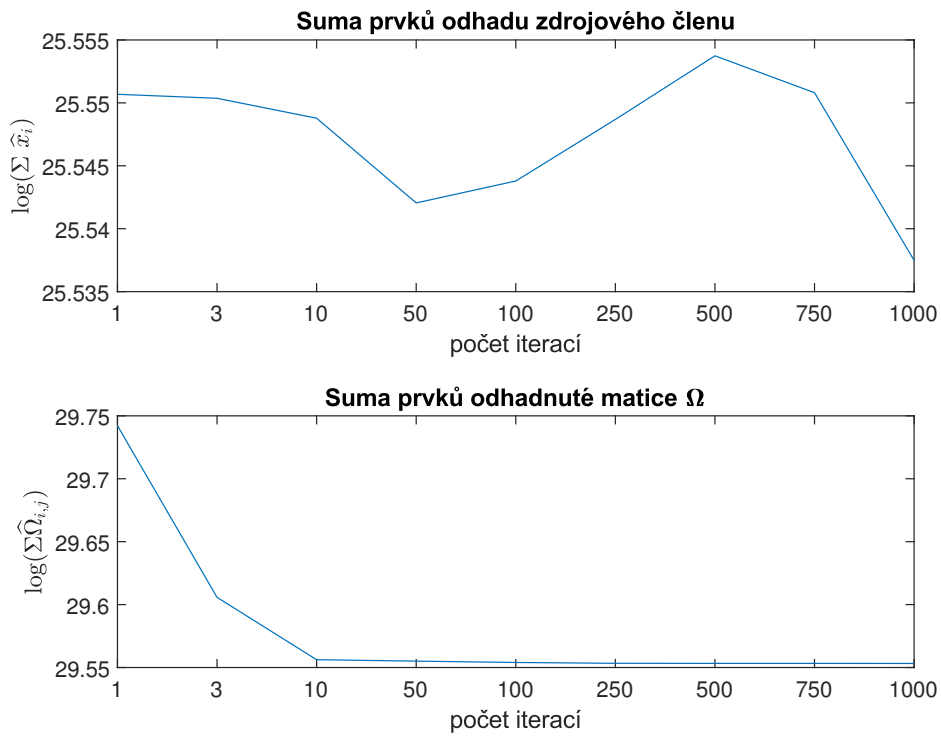
Tabulka 4.7: Metriky kvality odhadu a celková suma úniku pro různé volby parametru  $\vartheta_0$  a pro ně optimální poloměr.



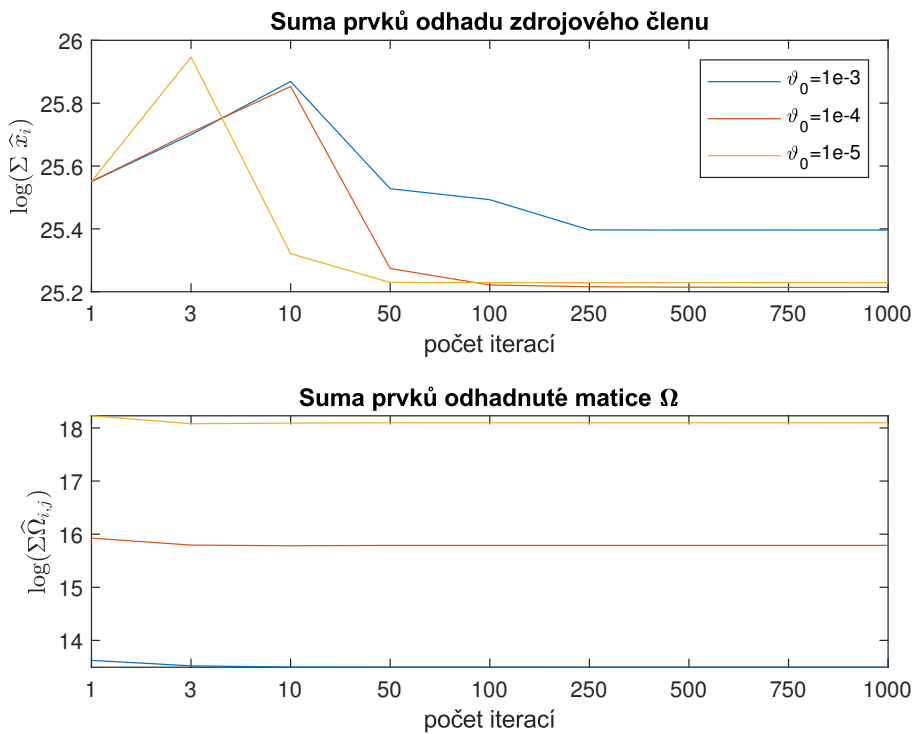
Obrázek 4.21: Graf závislosti metrik kvality odhadu na poloměru.



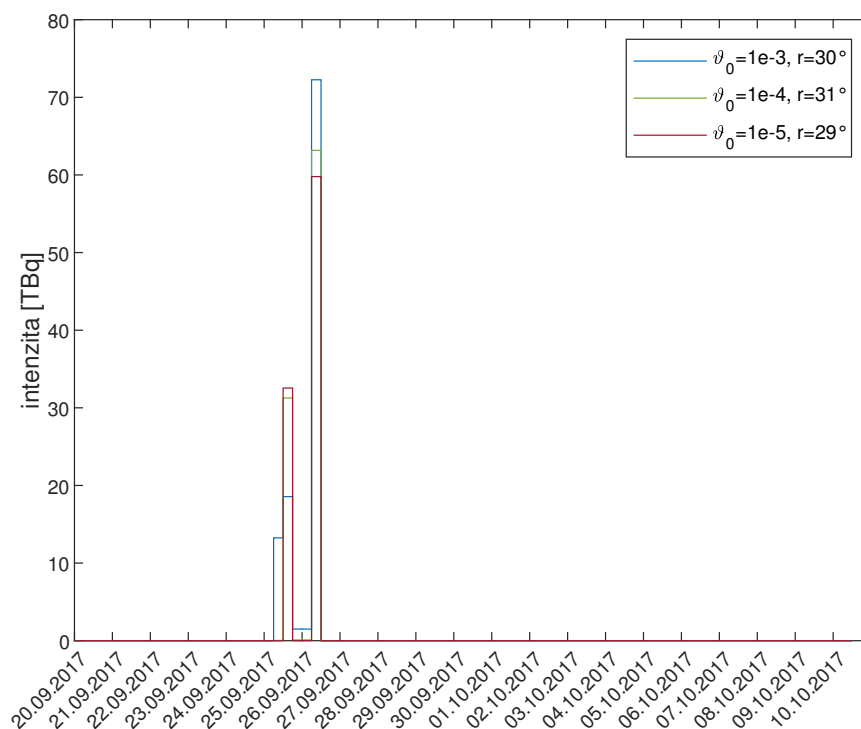
Obrázek 4.22: Odhad zdrojového členu pomocí LS-APC-w s poloměrem 75 °.



Obrázek 4.23: Konvergence parametrů při použití LS-APC-w s poloměrem 75 °.



Obrázek 4.24: Konvergence parametrů při použití LS-APC-w s různými hodnotami  $\vartheta_0$ .



Obrázek 4.25: Odhad zdrojového členu pomocí LS-APC-w pro různé volby parametru  $\vartheta_0$ .

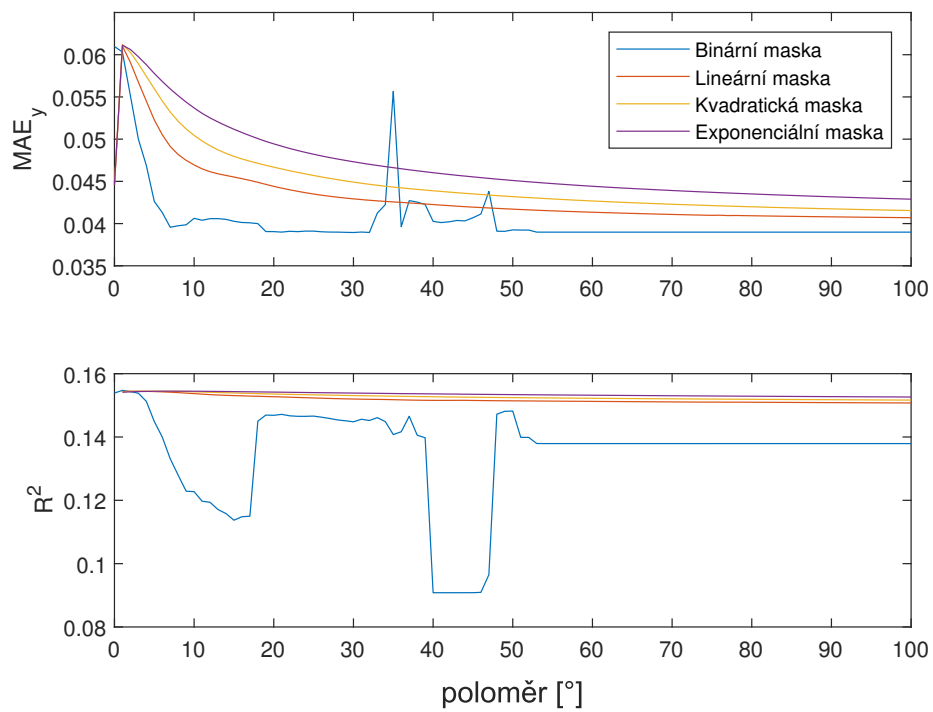
#### 4.3.2.3 Nebinární lokalizační maska

Nebinární lokalizační masky vytvoříme a budeme používat stejným způsobem jako v podsekcí 4.2.2. V algoritmu LS-APC-w použijeme hodnotu parametru  $\vartheta_0 = 10^{-5}$ . Metriky kvality odhadu při použití různých druhů lokalizačních masek v závislosti na poloměru jsou na obrázku 4.26 a v tabulce 4.8.

Stejně jako u datasetu IAEA se všechny nebinární masky chovají podobně a při vyšších poloměrech mezi nimi dokonce skoro není rozdíl. Oproti binární masce nejsou nebinární tak citlivé na poloměr. Na rozdíl od IAEA datasetu zde nejsou hodnoty metrik srovnatelné. Podle  $MAE_y$  vychází nejlépe binární maska, ale má nejhorší hodnotu  $R^2$ . Odhadnutá suma úniku je také méně než poloviční oproti odhadům z nebinárních masek.

algoritmus	$MAE_y$	$R^2$	suma úniku [TBq]
LS-APC-w bin. maska, $r=30^\circ$	0.038946	0.14480	90.5274
LS-APC-w lin. maska, $r=100^\circ$	0.040698	0.15074	206.0525
LS-APC-w kvad. maska, $r=100^\circ$	0.041543	0.15165	229.4011
LS-APC-w exp. maska, $r=100^\circ$	0.042888	0.15263	263.2220

Tabulka 4.8: Hodnoty metrik a celková suma úniku pro různé volby lokalizačních masek. Parametr  $\vartheta_0$  byl ve všech případech nastaven na  $10^{-5}$ .



Obrázek 4.26: Graf závislosti metrik kvality odhadu na poloměru pro různé druhy lokalizačních masek.



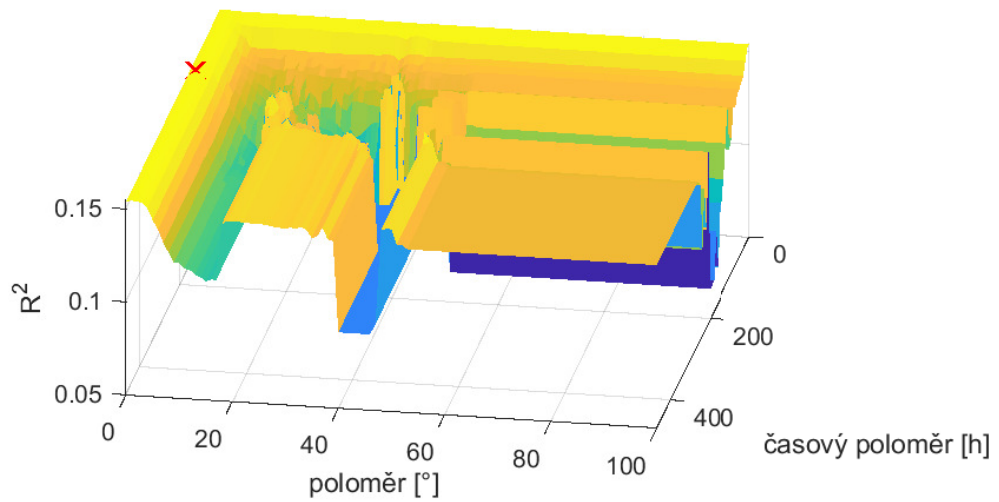
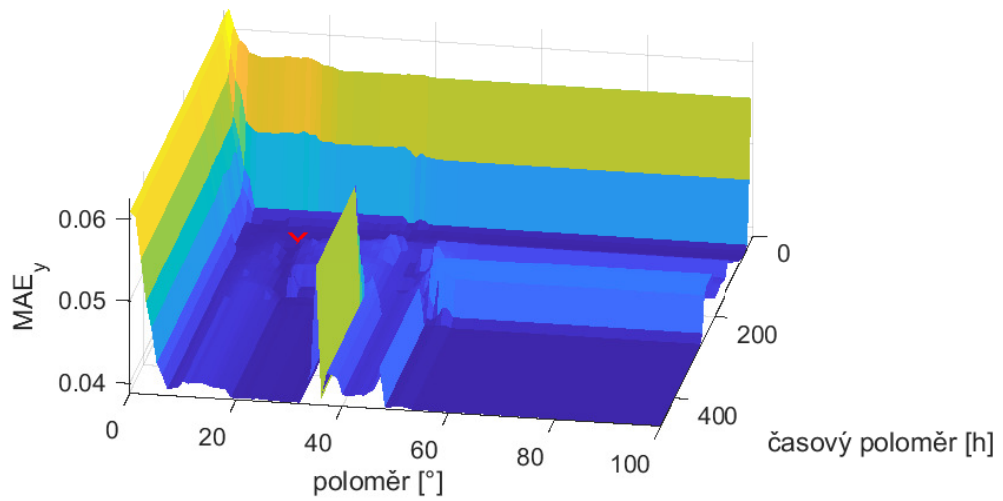
#### 4.3.2.4 Volba lokalizační masky podle času měření

Stejně jako v sekcích 4.2.3 a 4.3.1.4 využijeme znalosti časů, kdy k jednotlivým měřením došlo. Lokalizační masku dále omezíme na měření, která proběhla méně než daný čas v hodinách od sebe. Výsledná maska bude opět binární.

Graf metrik kvality odhadu zdrojového členu v závislosti na poloměru a časovém poloměru při použití algoritmu LS-APC-w s  $\vartheta_0 = 10^{-5}$  je na obrázku 4.27. Nejlepší hodnoty  $MAE_y$  dosahuje nastavení s poloměrem  $17^\circ$  a časovým poloměrem 80 hodin. Nejlepší hodnoty  $R^2$  dosahuje nastavení s poloměrem  $1^\circ$  a časovým poloměrem 150 hodin. Jak lze ale vidět v tabulce 4.9, přílišná optimalizace jedné metriky vede na špatné hodnoty druhé metriky. Zejména je to znatelné při nastavení vedoucí k maximální  $R^2$ , kdy se  $MAE_y$  skoro zdvojnásobila. Několikrát větší celková suma úniku oproti dosud získaným odhadům také značí, že takováto optimalizace není vhodná.

algoritmus	$MAE_y$	$R^2$	suma úniku [TBq]
LS-APC-w $r=30^\circ$	0.038946	0.14480	90.5274
LS-APC-w $r=17^\circ$ , $tr=80h$	0.038861	0.13545	147.0677
LS-APC-w $r=1^\circ$ , $tr=150h$	0.060294	0.15471	584.3919

Tabulka 4.9: Hodnoty metrik kvality odhadu a odhad celkové sumy úniku za použití lokalizační masky získaných v podsekcí 4.3.2.4.



Obrázek 4.27: Metriky kvality odhadu v závislosti na poloměru a časovém poloměru. Červeným křížkem je vyznačeno globální minimum u  $MAE_y$ , resp. maximum u  $R^2$ .

### 4.3.3 Shrnutí a srovnání výsledků

Nyní provedeme shrnutí nově získaných výsledků a srovnáme je s výsledky z výzkumného úkolu a s dostupnou literaturou.

#### 4.3.3.1 Shrnutí výsledků práce

Tabulka 4.10 shrnuje hodnoty metrik, celkovou sumu a datum odhadu úniku pomocí algoritmu LS-APC-w s různými volbami lokalizační masky. Parametr  $\vartheta_0$  byl v případě datasetu IAEA nastaven na  $10^{-3}$ , v případě Ro5 datasetu na  $10^{-5}$ . Tučně jsou zvýrazněny nejlepší hodnoty metrik pro daný dataset.

Z tabulky lze vidět, že nelze optimalizovat obě metriky zároveň. Pokud chceme co nejmenší  $MAE_y$ , zvolíme lokalizační masku, která uvažuje i časový rozestup mezi měřeními, s nižším poloměrem a celkem vysokým časovým poloměrem. Optimalizovat čistě  $R^2$  také lze a výsledné odhady nejsou úplně nerealistické. Celková suma odhadnutého úniku však vůbec neodpovídá ostatním.

Jako dobrý kompromis vychází vůbec lokalizační masku neomezovat podle času měření a pouze zvolit vhodný poloměr. Je ale možné že to vede na podhodnocení úniku. Použití nebinárních lokalizačních masek se příliš neosvědčilo – je složitější a nevede k lepšímu odhadu. Jediná jejich výhoda je, že od jisté hodnoty nejsou tak citlivé na zvolený poloměr.

Co se týče celkové sumy úniku, jsou odhady používající IAEA dataset konzistentnější a pohybují se od 186 TBq do 259 TBq. Časově jsou všechny úniku umístěny mezi 24. a 26. zářím. Odhady používající Ro5 dataset se pohybují mezi 91 a 206 TBq (resp. 584 TBq, pokud započítáme i extrémní odhad) a časově umístěny mezi 25. a 26. zářím.

dataset	algoritmus	$MAE_y$	$R^2$	suma úniku [TBq]	datum úniku
IAEA	LS-APC-w $r=22^\circ$	0.048383	0.16894	185.9156	24.-25.9 a 26.9.
	LS-APC-w lin. maska, $r=31^\circ$	0.048805	0.16890	232.8744	24.-25.9 a 26.9.
	LS-APC-w $r=16^\circ$ , $tr=70h$	<b>0.047884</b>	0.16571	188.9687	24.-25.9 a 26.9.
	LS-APC-w $r=5^\circ$ , $tr=120h$	0.049378	<b>0.17042</b>	258.6762	24.-25.9 a 26.9.
Ro5	LS-APC-w $r=30^\circ$	0.038946	0.14480	90.5274	25.9. a 26.9.
	LS-APC-w lin. maska, $r=100^\circ$	0.040698	0.15074	206.0525	25.9. a 26.9.
	LS-APC-w $r=17^\circ$ , $tr=80h$	<b>0.038861</b>	0.13545	147.0677	25.9. a 26.9.
	LS-APC-w $r=1^\circ$ , $tr=150h$	0.060294	<b>0.15471</b>	584.3919	25.9. a 26.9.

Tabulka 4.10: Shrnutí výsledných odhadů pro různé volby lokalizačních masek. Parametr  $\vartheta_0$  byl v případě datasetu IAEA nastaven na  $10^{-3}$ , v případě Ro5 datasetu na  $10^{-5}$ .

#### 4.3.3.2 Shrnutí výsledků z výzkumného úkolu a srovnání se získanými výsledky

Varianty algoritmu LS-APC použité ve výzkumném úkolu jsou popsány v kapitole 3 a samotné výsledky jsou shrnuty v tabulce 4.11. Použité datasety byly shodné jako v této práci. Získané odhady celkové sumy se pohybují od 105 TBq do 230 TBq. Časově byl únik umístěn

mezi 24. a 26. září. I když v této práci se odhadem lokace nezabýváme, je dobré zmínit, že ve všech případech byla lokace úniku odhadnuta v oblasti jižního Uralu velmi blízko Majaku.

Srovnáním tabulek 4.10 a 4.11 vidíme, že nové odhady jsou umístěny časově stejně mezi 24. a 26. září. Celkovou sumou úniku jsou také řádově stejné v nižších stovkách TBq. Co se týče metrik došlo u nových odhadů k mírnému zlepšení až na metriku  $MAE_y$  u Ro5 datasetu, kde i nejlepší dosažená hodnota je mírně horší než nejlepší z výzkumného úkolu.

dataset	algoritmus	$MAE_y$	$R^2$	suma úniku [TBq]	datum úniku
IAEA	LS-APC	0.050045	0.16591	167.1727	25.–26.9.
	LS-APC-Ovolne	0.048857	0.15567	104.6278	24.–26.9.
	LS-APC-Ozeme	<b>0.048536</b>	<b>0.16657</b>	199.7519	25.9. a 26.9.
	LS-APC-O100	0.052951	0.15740	209.3739	25.–26.9.
Ro5	LS-APC	0.041443	0.15266	229.7351	25.–26.9.
	LS-APC-Ovolne	0.038906	0.14038	145.0820	25.9. a 26.9.
	LS-APC-Ozeme	<b>0.038776</b>	0.14193	167.1921	25.9. a 26.9.
	LS-APC-O100	0.039442	<b>0.15281</b>	229.0757	25.–26.9.

Tabulka 4.11: Stručný přehled odhadů získaných ve výzkumném úkolu.

#### 4.3.4 Shrnutí výsledků z dostupné literatury

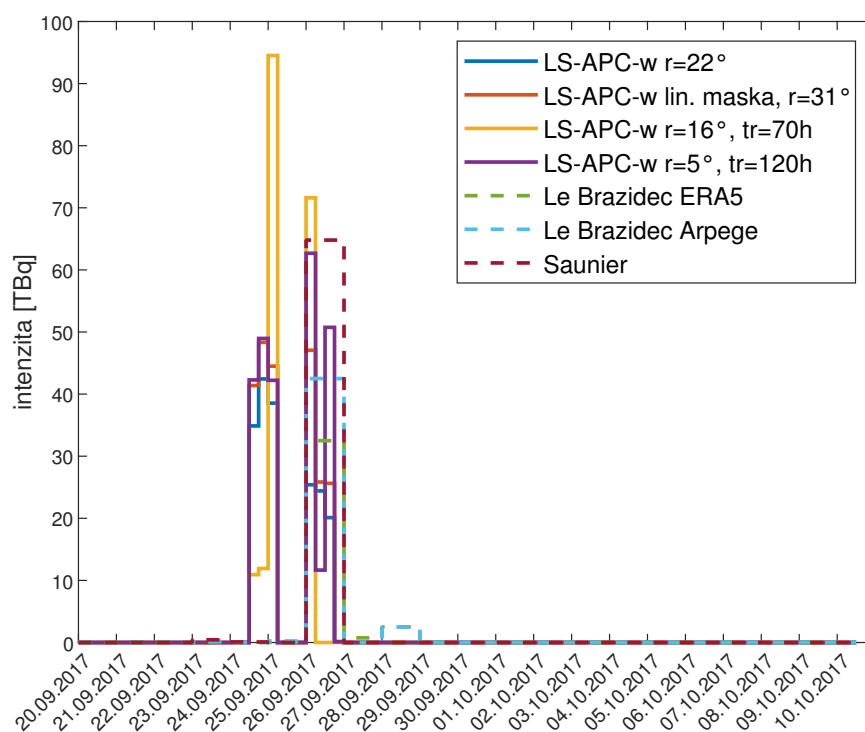
O objasnění úniku ruthenia-106 v roce 2017 se dosud pokoušelo několik článků různými metodami. I přes značně rozdílný formát dosažených výsledků se pokusíme o stručné shrnutí v tabulce 4.12. Literatura se vesměs shoduje, že k úniku o nižších stovkách TBq došlo kolem 26. září 2017, což je v souladu s námi získanými výsledky. Srovnání grafů zdrojových členů s odhady z článků (pokud bylo možno získat) je na obrázku 4.28 pro IAEA dataset, resp. obrázku 4.29 pro Ro5.

článek	lokace	suma intenzity [TBq]	datum úniku
Le Brazidec ERA5	56.325° N; 58.614° E	125	26. 9.
Le Brazidec ARPEGE	55.72° N; 58.968° E	171	26. 9.
Sørensen	pás od Permu a Jekatěrinburgu k Oděse	–	26. nebo 27. 9.
Saunier	jižní Ural	252	26. 9.
Shershakov	jižní a střední Ural	–	25.–26. 9.
Maffezzoli	oblast mezi Ukrajinou a Uralem	–	–
Western	jižní Ural	441	24. 9.

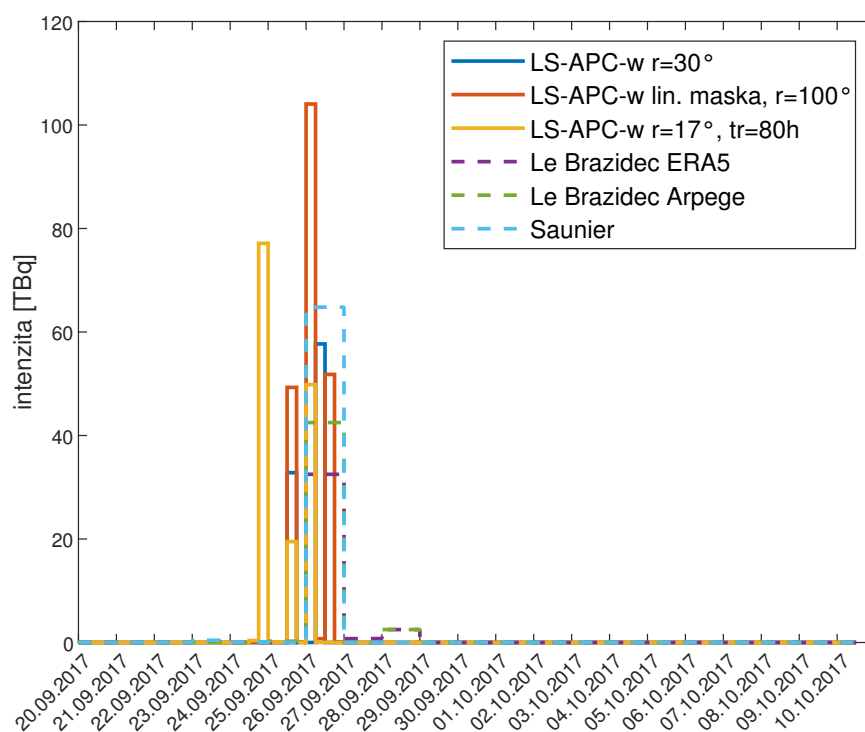
Tabulka 4.12: Stručný přehled dosud publikovaných výsledků.

Pro zmíněné články provedeme i krátké shrnutí použitých postupů a jak se liší od našeho.

**Le Brazidec et. al.**[12] provádí odhad jak lokace tak zdrojového členu pomocí metody Markov Chain Monte Carlo a klasickým Metropolis-Hastings algoritmem vylepšeným pomocí tzv. paralelního žihání. Dataset přebírá z Masson et. al.[22] jehož základem jsou také



Obrázek 4.28: Srovnání grafů zdrojových členů z této práce při použití IAEA datasetu s odhady z článků.



Obrázek 4.29: Srovnání grafů zdrojových členů z této práce při použití Ro5 datasetu s odhady z článků.

měření z Ro5. Použity jsou dva zdroje meteorologických dat – ARPEGE Météo France a ECMWF ERA5. Stejně jako my předpokládá, že měření lze vysvětlit lineárním vztahem a pravděpodobnostní rozdělení zdrojového členu získá aplikací Bayesovy věty. Liší se hlavně zvolenými apriorními hustotami – rozdíl mezi měřeními a rekonstrukcí modeluje pomocí log-normálního rozdělení, které by se mělo lépe vypořádat s měřeními velmi různých řádů. Struktura kovarianční matice šumu má jednodušší strukturu a předpokládá homoskedasticitu. Lokace úniku (koordináty  $x_1$  a  $x_2$ ) jsou přímo modelovány s uniformní apriorní hustotou z prostorové domény.

**Sørensen**[31] Pro každé měření poskytnuté organizací IAEA pustí atmosférický model DERMA ve zpětném chodu. Výsledné koncentrace jsou korelovány jak v čase tak prostoru a jsou identifikovány překryvy. Jako možné zdroje bere institut pro výzkum jaderných reaktorů NIIAR a kombinát Majak. V případě NIIAR odhaduje únik mezi 5. a 13. hodinou UTC 27. září 2017 a v případě Majaku ve stejný čas už 26. září 2017. Celkovou sumu úniku neodhaduje, ale volí ji aby výsledný odhad odpovídal měření s největší koncentrací. Pro NIIAR používá 460 TBq, pro Majak 1100 Tbq.

**Saunier et. al.**[27] odhaduje zdrojový člen minimalizací ztrátové funkce pomocí řešiče založeném na kvazi-Newtonově metodě. Minimalizace ztrátové funkce maximalizuje aposteriorní rozdělení získané z Bayesovy věty. Šum je modelován log-normálním rozdělením a je předpokládána homoskedasticita. Použitý dataset vychází z Ro5 rozšířeného o další měření. Pravděpodobné lokace úniku jsou stejně jako ve výzkumném úkolu získány srovnáním měření a jeho rekonstrukce, jen jsou použity jiné metriky.

**Shershakov et. al.**[29] Data organizace IAEA doplněná o měření ruského úřadu Roshyromet. Článek se zabývá čistě odhadem lokace. Simuluje různé hypotetické scénáře úniku a srovnává jak dobře sedí na naměřená data.

**Maffezzoli et. al.**[21] Měření z Milána doplněná o vybraná měření od organizace IAEA, která byla geograficky rozlehlá, prováděna dostatečně často a jejich maximum vysoko nad detekčním limitem. Postup je podobný jako Sørensen[31], jsou navíc přidány váhy, které posilují lokace daleko od všech zdrojů. To by mělo pomoci s tendencí hromadění výsledné koncentrace blízko měřicích stanic.

**Western et. al.**[35] Data z Masson et. al.[22]. Stejně jako my předpokládá, že měření lze vysvětlit lineárním vztahem a pravděpodobnostní rozdělení zdrojového členu získá aplikací Bayesovy věty. Pro zdrojový člen však používá Jeffreysovo apriorní rozdělení, které sice nezaráží do řešení žádnou informaci, ale není omezeno pouze na kladné hodnoty. Parametry úniku zjišťuje kombinací metod Markov Chain Monte Carlo a Expectation–Maximisation. Jako lokaci úniku předpokládá kombinát Majak.

# Závěr

Cílem práce bylo provést odhad časového průběhu úniku ruthenia-106 do atmosféry na podzim roku 2017 popsaného v úvodu práce. Umístění úniku do kombinátu na zpracování jaderných materiálů Majak již předpokládáme za dostatečně prokázané jak ve výzkumném úkolu, tak uvedené literatuře. Zaměřili jsme se proto především na modely pro odhad emise ze známé lokace.

V kapitole 1 byl formulován lineární model pro popis úniku radiace do atmosféry a z něj vycházející inverzní úloha pro odhad zdrojového členu včetně popisu základních přístupů pro její řešení. Také byly popsány atmosférické transportní modely nutné pro formulaci modelu a inverzní úlohy.

V kapitole 2 byly shrnuty základy Bayesovské teorie pravděpodobnosti a z ní vycházející odhady. Pravděpodobnostní přístup umožňuje pohodlné začlenění požadavků a předpokladů o řešení. Často však není možné odhady vyjádřit analyticky a proto bylo nutné prozkoumat možnosti aproximace. Jako vhodná aproximace byla po zkušenostech z bakalářské práce[6] a výzkumného úkolu metoda Variační Bayes.

V kapitole 3 bylo provedeno odvození již existujícího algoritmu založeného na této aproximaci – LS-APC[32]. Toto odvození bylo použito jako základ pro další úpravy pro možnost modelovat i korelaci mezi složkami šumu měření např. na základě vzdáleností jednotlivých měřících stanic.

V kapitole 4 byl výsledný algoritmus otestován nejdříve na syntetických datech, kde se ověřilo, že modelovat korelaci složek šumu vede na zlepšení odhadu zdrojového členu. Poté byl aplikován na data z experimentu ETEX, která jsou na verifikaci algoritmu vhodná, protože se jedná o reálný únik, u kterého známe skutečný zdrojový člen a můžeme jej porovnat se získaným odhadem.

Ukázalo se, že modelovat celou příslušnou kovarianční matici  $\Omega$  v rovnici pozorování vede k nestabilitě algoritmu a bylo nutné odhadovat pouze některé prvky. Toho bylo docíleno sestavením lokalizační masky  $\Lambda$ , která v každém kroku algoritmu buď utlumovala nebo rovnou vynulovala vybrané prvky kovarianční matice. Tato lokalizační matice byla sestavena heuristicky na základě vzdálenosti jednotlivých měřících stanic a časové prodlevě mezi měřeními.

Tyto úpravy vedly na zvýšenou citlivost celkového odhadu na parametry apriorního rozdělení kovarianční matice  $\Omega$ , které původně byly nastaveny, aby výsledné apriorní rozdělení bylo co nejméně informativní bez ztráty numerické stability. Získané odhady však byly očividně nesmyslné a proto bylo nutné ladit také tyto parametry v rámci hodnot, které malou informativnost zachovávají.



Nakonec byl algoritmus aplikován na únik ruthenia-106 popsáný v úvodu práce. Odhady byly provedeny za předpokladu, že k úniku došlo v kombinátu Majak. Časově jsou veškeré získané odhady umístěny mezi 24. až 26. září s celkovou sumou mezi 91 až 259 TBq. Získané výsledky jsou konzistentní s dostupnou literaturou.

# Literatura

- [1] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, 1985. ISBN: 0-387-96098-8.
- [2] Jose M. Bernardo. Expected information as expected utility. *Ann. Statist.*, 7(3):686–690, 05 1979.
- [3] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006. ISBN: 0-387-31073-8.
- [4] Jeff Bluett, Neil Gimson, Gavin Fisher, Clive Heydenrych, Tracy Freeman, and Jenny Godfrey. Good practice guide for atmospheric dispersion modelling. <http://tools.envirolink.govt.nz/assets/Uploads/Good-Practice-Guide-MFE-atmospheric-dispersion-modelling-jun04.pdf>, 2004. [Online; citováno 16. prosince 2021].
- [5] P Bossew, F Gering, Eric Petermann, T Hamburger, Christian Katzlberger, Miguel Ángel Hernández Ceballos, Marc De Cort, Krzysztof Gorzkiewicz, Renata Kierepko, and Jerzy Mietelski. An episode of ru-106 in air over Europe, September–October 2017 – Geographical distribution of inhalation dose over Europe. *Journal of Environmental Radioactivity*, 205-206:79–92, 05 2019.
- [6] T. Brisučiak. Lineární inverzní modely a jejich využití při určování zdroje radiačního úniku do atmosféry. bakalářská práce, České vysoké učení technické v Praze, 2019.
- [7] Michael W. Cooke, Adrian Botti, Dorian Zok, Georg Steinhauser, and Kurt R. Ungar. Identification of a chemical fingerprint linking the undeclared 2017 release of 106Ru to advanced nuclear fuel reprocessing. *Proceedings of the National Academy of Sciences*, 117(26):14703–14711, 2020.
- [8] D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kállberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597, 2011.

- [9] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39:1–22, 1977.
- [10] Ádám Leelőssy, Ferenc Molnár, Ferenc Izsák, Ágnes Havasi, István Lagzi, and Róbert Mészáros. Dispersion modeling of air pollutants in the atmosphere: a review. *Open Geosciences*, 6(3):257–278, 2014.
- [11] Roland Draxler and G. Hess. Description of the HYSPLIT\_4 modelling system. *NOAA Tech. Mem. ERL ARL-224*, 01 1997.
- [12] Joffrey Dumont Le Brazidec, Marc Bocquet, Olivier Saunier, and Yelva Roustan. MCMC methods applied to the reconstruction of the autumn 2017 Ruthenium-106 atmospheric contamination source. *Atmospheric Environment: X*, 6:100071, 2020.
- [13] S. Eckhardt, A. J. Prata, P. Seibert, K. Stebel, and A. Stohl. Estimation of the vertical profile of sulfur dioxide injection into the atmosphere by a volcanic eruption using satellite column measurements and inverse transport modeling. *Atmospheric Chemistry and Physics Discussions*, 8(1):3761–3805, February 2008.
- [14] P.D. Hoff. *A First Course in Bayesian Statistical Methods*. Springer Texts in Statistics. Springer New York, 2009. ISBN: 9780387924076.
- [15] Timo Hopp, Dorian Zok, Thorsten Kleine, and Georg Steinhauser. Non-natural ruthenium isotope ratios of the undeclared 2017 atmospheric release consistent with civilian nuclear activities. *Nature Communications*, 11:2744, 06 2020.
- [16] M. Hutchinson, H. Oh, and W. Chen. A review of source term estimation methods for atmospheric dispersion events using static or mobile sensors. *Information Fusion*, 36:130–148, 2017.
- [17] IAEA. Technical attachment: Status of measurements of Ru-106 in Europe. [https://www.criirad.org/accident-et-pollutions/aiea\\_Ru-106\\_Technical\\_Attachment-2.pdf](https://www.criirad.org/accident-et-pollutions/aiea_Ru-106_Technical_Attachment-2.pdf), 2017. [Online; citováno 18. prosince 2020].
- [18] IRSN. Report on the IRSN’s investigations following the widespread detection of 106Ru in Europe early October 2017. [https://www.irsn.fr/FR/Actualites\\_presse/Actualites/Documents/IRSN\\_Report-on-IRSN-investigations-of-Ru-106-in-Europe-in-october-2017.pdf](https://www.irsn.fr/FR/Actualites_presse/Actualites/Documents/IRSN_Report-on-IRSN-investigations-of-Ru-106-in-Europe-in-october-2017.pdf), 2018. [Online; citováno 18. prosince 2021].
- [19] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22:79–86, 1951.
- [20] A. Leelőssy, I. Lagzi, A. Kovacs, and R. Meszaros. A review of numerical models to predict the atmospheric dispersion of radionuclides. *Journal of environmental radioactivity*, 182:20–33, 2018.

- [21] Niccolò Maffezzoli, Giovanni Baccolo, Elena Di Stefano, and Massimiliano Clemenza. The Ruthenium-106 plume over Europe in 2017: a source-receptor model to estimate the source region. *Atmospheric Environment*, 212:239 – 249, 2019.
- [22] O. Masson et al. Airborne concentrations and chemical considerations of radioactive ruthenium from an undeclared major nuclear release in 2017. *Proceedings of the National Academy of Sciences*, 116(34):16750–16759, 2019.
- [23] T. Minka. *A family of algorithms for approximate Bayesian inference*. PhD thesis, Massachusetts Institute of Technology, 2001.
- [24] R. M. Neal. Probabilistic inference using Markov Chain Monte Carlo methods. Technical report, Dept. of Computer Science, University of Toronto, 1993.
- [25] K. Nodop, R. Connolly, and F. Girardi. The field campaigns of the European Tracer Experiment (ETEX): Overview and results. *Atmospheric Environment*, 32(24):4095–4108, 1998.
- [26] M. Sato. Online model selection based on the Variational Bayes. *Neural Computation*, 13:1649–1681, 2001.
- [27] O. Saunier, D. Didier, A. Mathieu, O. Masson, and J. Dumont Le Brazidec. Atmospheric modeling and source reconstruction of radioactive ruthenium from an undeclared major release in 2017. *Proceedings of the National Academy of Sciences*, 116(50):24991–25000, 2019.
- [28] P. Seibert and A. Frank. Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode. *Atmos. Chem. Phys.*, 4(1):51–63, 2004.
- [29] Vjacheslav M. Shershakov, R. V. Borodin, and Yu. S. Tsaturov. Assessment of possible location Ru-106 source in russia in September–October 2017. *Russian Meteorology and Hydrology*, 44:196–202, 2019.
- [30] V. Šmídl and A. Quinn. *The Variational Bayes Method in Signal Processing*. Springer, 2006. ISBN: 3-540-28819-8.
- [31] Jens Havskov Sørensen. Method for source localization proposed and applied to the October 2017 case of atmospheric dispersion of Ru-106. *Journal of Environmental Radioactivity*, 189:221 – 226, 2018.
- [32] Ondřej Tichý, Václav Šmídl, Radek Hofman, and Andreas Stohl. LS-APC v1.0: A tuning-free method for the linear inverse problem and its application to source-term determination. *Geoscientific Model Development*, 9:4297–4311, 11 2016.
- [33] M. E. Tipping. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244, 2001.
- [34] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 61(3):611–622, 1999.

- [35] Luke M. Western, Sarah C. Millington, Anastasia Benfield-Dexter, and Claire S. Witham. Source estimation of an unexpected release of Ruthenium-106 in 2017 using an inverse modelling approach. *Journal of Environmental Radioactivity*, 220-221:106304, 2020.

# Příloha A

## Použitá rozdělení pravděpodobnosti

### A.1 Vícerozměrné normální rozdělení

Nechť  $\mathbf{x} \in \mathbb{R}^{n \times 1}$ . Pak jeho vícerozměrné normální rozdělení se středem  $\boldsymbol{\mu} \in \mathbb{R}^{n \times 1}$  a kovarianční maticí  $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$  je

$$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}, \quad (\text{A.1})$$

s momenty

$$\widehat{\mathbf{x}} = \boldsymbol{\mu}, \quad (\text{A.2})$$

$$\widehat{\mathbf{xx}^\top} = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top. \quad (\text{A.3})$$

### A.2 Ořezané normální rozdělení

Ořezané normální rozdělení skaláru  $x$  na omezeném nosiči  $[a, b]$  je definováno jako

$$\mathcal{N}^{\text{tr}}(\mu, \sigma; [a, b]) = \frac{\sqrt{2} \exp\left(-\frac{1}{2\sigma}(x - \mu)^2\right)}{\sqrt{\pi\sigma}(\text{erf}(\beta) - \text{erf}(\alpha))} \chi_{[a,b]}(x), \quad (\text{A.4})$$

kde  $\alpha = \frac{a-\mu}{\sqrt{2\sigma}}$ ,  $\beta = \frac{b-\mu}{\sqrt{2\sigma}}$ ,  $\chi_{[a,b]}(x)$  je charakteristická funkce intervalu  $[a, b]$  a  $\text{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t \exp(-u^2) du$ .

Momenty se pak spočtou

$$\widehat{x} = \mu - \sqrt{\sigma} \frac{\sqrt{2}[\exp(-\beta^2) - \exp(-\alpha^2)]}{\sqrt{\pi}[\text{erf}(\beta) - \text{erf}(\alpha)]}, \quad (\text{A.5})$$

$$\widehat{x^2} = \sigma + \mu\widehat{x} - \sqrt{\sigma} \frac{\sqrt{2}[b \exp(-\beta^2) - a \exp(-\alpha^2)]}{\sqrt{\pi}[\text{erf}(\beta) - \text{erf}(\alpha)]}. \quad (\text{A.6})$$

### A.3 Ořezané vícerozměrné normální rozdělení

Chceme-li analyticky spočítat momenty ořezaného vícerozměrného normálního rozdělení  $\mathcal{N}^{\text{tr}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; [a, b])$ , musíme je aproximovat momenty  $\mathcal{N}^{\text{tr}}(\boldsymbol{\mu}, \text{diag}(\boldsymbol{\sigma}); [a, b])$ , kde  $\boldsymbol{\sigma}$  je vektor diagonálních prvků matice  $\boldsymbol{\Sigma}$ . Prvky momentu  $\widehat{\mathbf{x}}$  jsou pak dány rovnicí (A.5) a

$$\widehat{\mathbf{xx}}^{\text{T}} = \widehat{\mathbf{xx}}^{\text{T}} + \text{diag}(\widehat{\boldsymbol{\sigma}}), \quad (\text{A.7})$$

kde  $\widehat{\sigma}_i = \widehat{x}_i^2 - \widehat{x}_i \widehat{x}_i$ .

Tato aproximace může být příliš hrubá. V tom případě lze aproximovat

$$\widehat{\mathbf{xx}}^{\text{T}} = \widehat{\mathbf{xx}}^{\text{T}} + \text{diag}(\mathbf{o}) \boldsymbol{\Sigma} \text{diag}(\mathbf{o}), \quad (\text{A.8})$$

kde  $o_i = \widehat{\sigma}_i^{1/2} \sigma_i^{-1/2}$ .

### A.4 Gamma rozdělení

Gamma rozdělení skaláru  $x$  je definováno jako

$$\mathcal{G}(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), \quad (\text{A.9})$$

pro  $x, \alpha, \beta > 0$  a kde  $\Gamma(x)$  je gamma funkce.

Momenty gamma rozdělení jsou

$$\widehat{x} = \frac{\alpha}{\beta}, \quad (\text{A.10})$$

$$\widehat{x^2} = \frac{\alpha}{\beta^2}. \quad (\text{A.11})$$

### A.5 Wishartovo rozdělení

Wishartovo rozdělení je zobecnění gamma rozdělení do více dimenzí. Pro symetrickou a pozitivně definitní matici  $\mathbf{X}$  rozměru  $p \times p$  je jeho hustota definována jako

$$\mathcal{W}(\mathbf{V}, n) = \frac{|\mathbf{X}|^{(n-p-1)/2} \exp\left(-\frac{1}{2} \text{tr}(\mathbf{V}^{-1} \mathbf{X})\right)}{2^{np/2} |\mathbf{V}|^{n/2} \Gamma_p\left(\frac{n}{2}\right)}, \quad (\text{A.12})$$

pro  $n > p - 1$ ,  $n \in \mathbb{R}$  a  $\mathbf{V}$  je symetrická pozitivně definitní matice rozměru  $p \times p$ .  $\Gamma_p$  značí vícerozměrnou gamma funkci.

Momenty Wishartova rozdělení jsou

$$\widehat{\mathbf{X}} = n\mathbf{V}. \quad (\text{A.13})$$