

Přednáška 4: Bayesovská sparse regression, aplikace zobecněných lineárních regresních modelů

Ondřej Tichý

Bayesovské metody v machine learningu (BML)

March 9, 2018

Obsah přednášky

- ▶ Krátké opakování
- ▶ Bayesovská formulace problému
 - ▶ sparse regression
 - ▶ další možnosti (pouze náznak)
- ▶ Odhad tzv. zdrojového členu při radioaktivním úniku do ovzduší
- ▶ Odhad neznámé lokace úniku

Opakování

Zobecněný lineární regresní model

Předpoklad vysvětlení každé naměřené hodnoty:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + e_i. \quad (1)$$

My budeme využívat maticový zápis:

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & & x_{2n} \\ \vdots & & \ddots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}}_{\boldsymbol{\beta}} + \mathbf{e}. \quad (2)$$

Opakování

Optimalizační přístup

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (3)$$

- ▶ nejmenší čtverce

$$\hat{\boldsymbol{\beta}}_{OLS} = \arg \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (4)$$

- ▶ regularizace - Tichonovova

$$\hat{\boldsymbol{\beta}}_{Tikhonov} = \arg \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \alpha \|\boldsymbol{\beta}\|_2^2 = (\mathbf{X}^T \mathbf{X} + \alpha I)^{-1} \mathbf{X}^T \mathbf{y}. \quad (5)$$

- ▶ regularizace - LASSO

$$\hat{\boldsymbol{\beta}}_{LASSO} = \arg \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \alpha \|\boldsymbol{\beta}\|_1, \quad (6)$$

Opakování

Bayesovská lineární regrese

- aposteriorno ve tvaru $\tilde{f}(\beta) = \mathcal{N}(\mu_\beta, \Sigma_\beta)$ a $\tilde{f}(\omega) = \mathcal{G}(c, d)$ s tvarovacími parametry:

$$\Sigma_\beta = \left(\hat{\omega} X^T X + I_n \right)^{-1} \quad (7)$$

$$\mu_\beta = \Sigma_\beta \hat{\omega} X^T \mathbf{y} = \left(\hat{\omega} X^T X + I_n \right)^{-1} \hat{\omega} X^T \mathbf{y} \quad (8)$$

$$c = c_0 + \frac{p}{2} \quad (9)$$

$$d = d_0 + \frac{1}{2} \left(\mathbf{y} \mathbf{y}^T - 2 \mathbf{y}^T X \hat{\beta} + \widehat{\beta^T X^T X \beta} \right) \quad (10)$$

s momenty (viz tabulky, Wikipedia, ...)

$$\hat{\beta} = \mu_\beta \quad (11)$$

$$\widehat{\beta \beta^T} = \mu_\beta \mu_\beta^T + \Sigma_\beta \quad (12)$$

$$\hat{\omega} = c/d \quad (13)$$

Opakování

Variační Bayesova proximace

- ▶ máme model dat $f(\mathbf{y}|\theta)$,
- ▶ máme q podmíněně nezávislých proměnných, tzn. pro jejich arpiorna platí

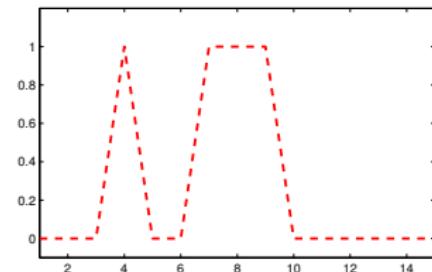
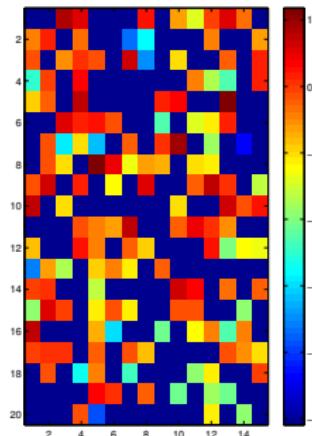
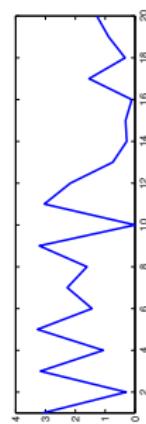
$$f(\theta_1, \theta_2, \dots, \theta_q) = \prod_{i=1}^q f(\theta_i), \quad (14)$$

- ▶ pak aposteriorní rozdělení $\tilde{f}(\theta_i|\mathbf{y})$ pro i ty parametr získáme VB approximací jako

$$\tilde{f}(\theta_i|\mathbf{y}) \propto \exp \left[E_{\tilde{f}(\theta_{\setminus i}|\mathbf{y})} (\ln f(\theta_1, \theta_2, \dots, \theta_q, \mathbf{y})) \right], \quad \forall i. \quad (15)$$

Opakování

Cvičný příklad - pěkná data



y

$=$

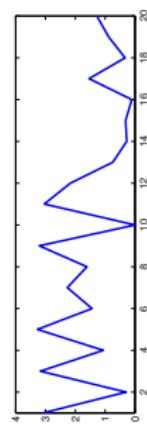
X

β

+šum

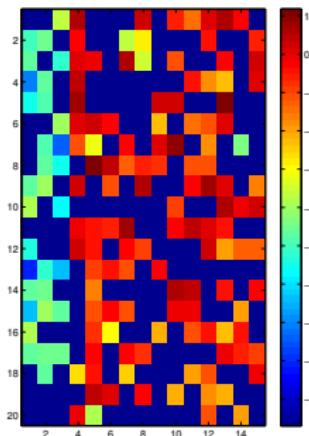
Opakování

Cvičný příklad - těžší data

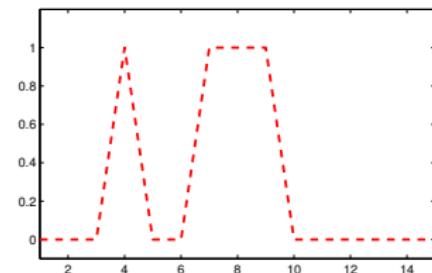


y

$=$



X

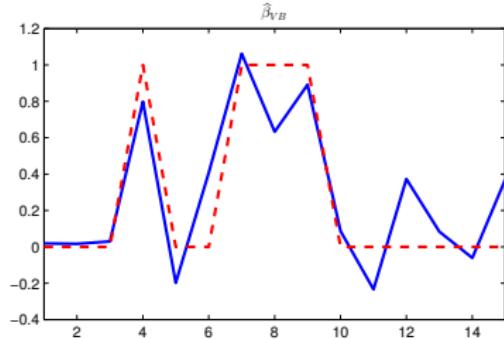
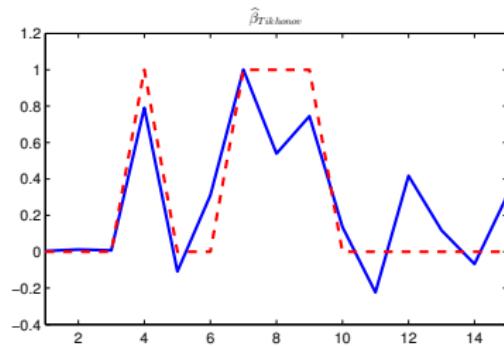
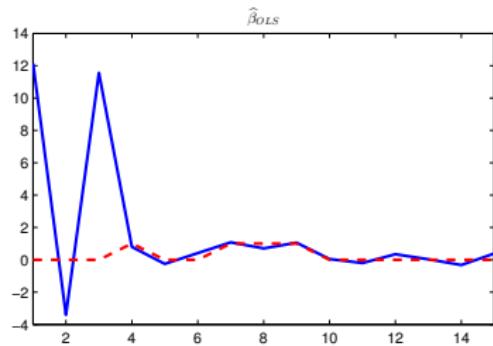


β

+šum

Opakování

Výsledky



Bayesovská sparse regression

Formulace

- ▶ též “Sparse bayesian learning” nebo “Relevance vector machine”.
- ▶ model dat a šumu zachováme,

$$f(\mathbf{y}|\boldsymbol{\beta}, \omega) = \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \omega^{-1}I_p), \quad (16)$$

$$f(\omega) = \mathcal{G}(c_0, d_0). \quad (17)$$

- ▶ budeme modelovat rozptyl každého prvku vektoru $\boldsymbol{\beta}$ pomocí hyperparametru $v_j \in \mathbb{R}$:

$$f(\boldsymbol{\beta}) = \mathcal{N}(\mathbf{0}, I_n) \quad \dashrightarrow \quad f(\boldsymbol{\beta}|\mathbf{v}) = \mathcal{N}\left(\mathbf{0}, \text{diag}(v_1, \dots, v_n)^{-1}\right) \quad (18)$$

- ▶ v_j se stávají parametry modelu, zvolíme jejich apriorní rozdělení

$$f(v_j) = \mathcal{G}(a_0, b_0), \quad \forall j. \quad (19)$$

Bayesovská sparse regression

VB aproximace

- aposteriorno $\tilde{f}(\beta|v) = \mathcal{N}(\mu_\beta, \Sigma_\beta)$, $\tilde{f}(v_j) = \mathcal{G}(a_j, b_j)$, $\forall j$, a $\tilde{f}(\omega) = \mathcal{G}(c, d)$ má (po odvození) následující tvarovací parametry:

$$\Sigma_\beta = \left(\hat{\omega} X^T X + \text{diag}(\hat{\mathbf{v}}) \right)^{-1} \quad (20)$$

$$\mu_\beta = \Sigma_\beta \hat{\omega} X^T \mathbf{y} = \left(\hat{\omega} X^T X + \text{diag}(\hat{\mathbf{v}}) \right)^{-1} \hat{\omega} X^T \mathbf{y} \quad (21)$$

$$a_j = a_0 + \frac{1}{2} \quad (22)$$

$$b_j = b_0 + \frac{1}{2} \left(\widehat{\beta \beta'} \right)_{j,j} \quad (23)$$

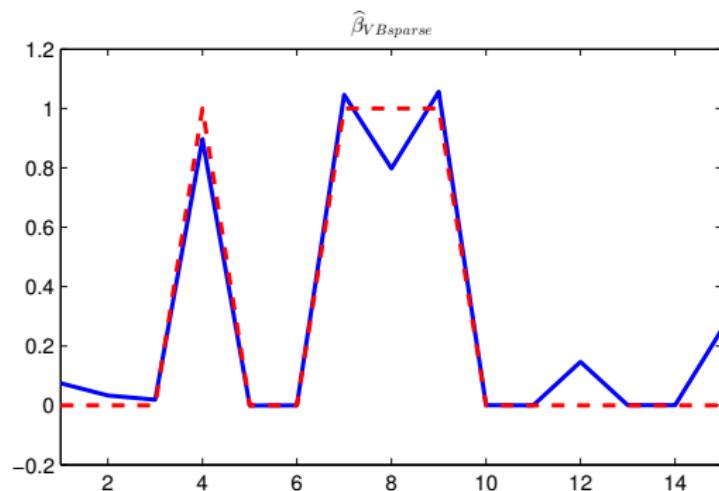
$$c = c_0 + \frac{p}{2} \quad (24)$$

$$d = d_0 + \frac{1}{2} \left(\mathbf{y} \mathbf{y}^T - 2 \mathbf{y}^T X \widehat{\beta} + \widehat{\beta^T X^T X \beta} \right) \quad (25)$$

...a příslušné momenty.

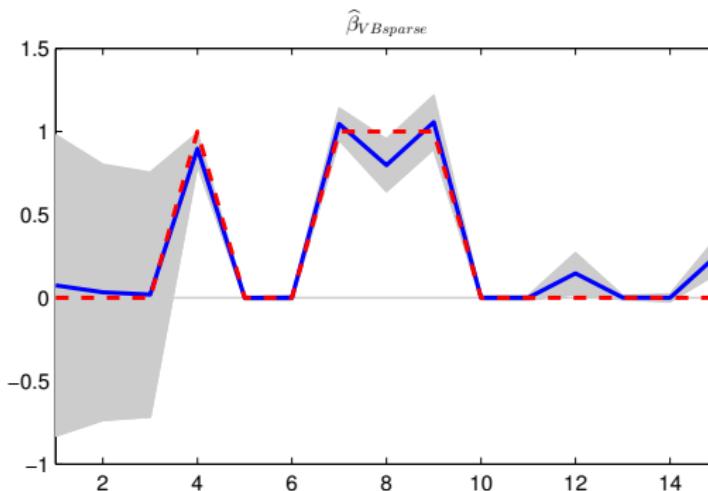
Bayesovská sparse regression

Řešení



Výhody bayesovského přístupu

- ▶ kvantifikace neurčitosti



- ▶ potlačení "ladicích" knoflíků, i když...
- ▶ nepřeberné modelovací možnosti

Dodatek

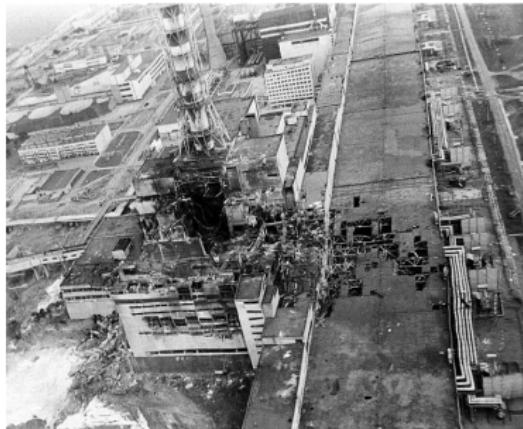
Co jsme neřešili a možná bychom měli...

- ▶ Pracovali jsme pouze s nagenerovaným příkladem bez fyzikální podstaty
- ▶ Co jsme vynechali?
 - ▶ pozitivita
 - ▶ další modely vektoru β
 - ▶ např. spike and slab prior, horseshoe prior
 - ▶ informace o měřeních y

Aplikace: odhad tzv. zdrojového členu

Úvod

- ▶ Proč se problémem určování zdrojů zabývat?

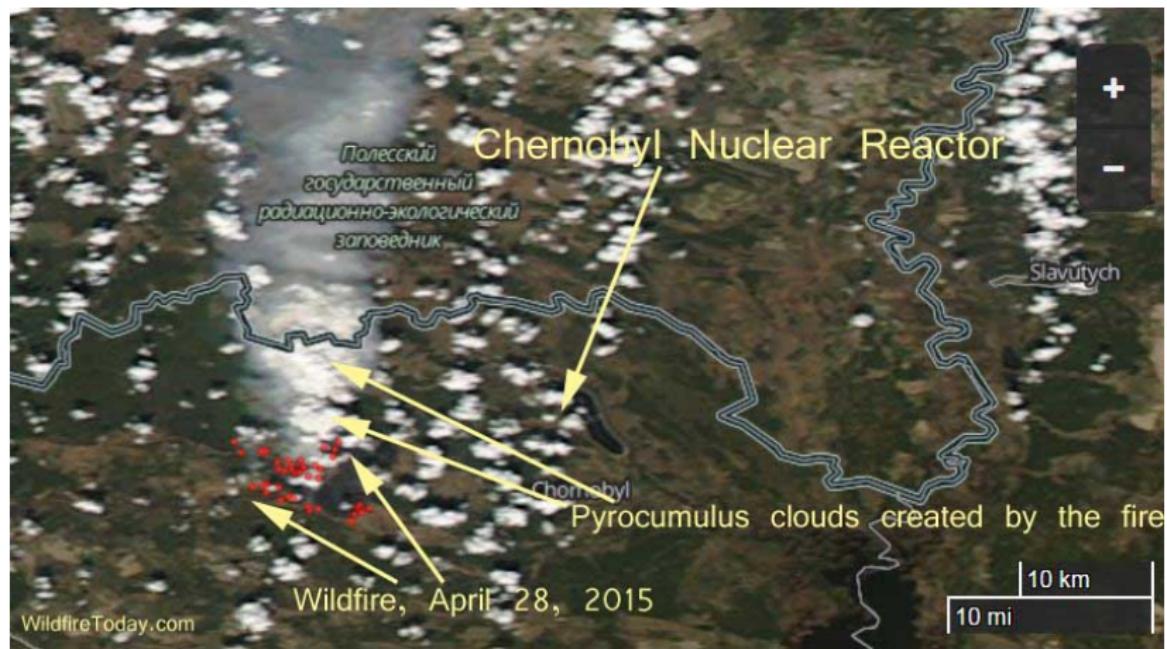


- ▶ Mnoho drobnějších úniků (lékařství atd.), mnohé z nich neobjasněny.

Aplikace: odhad tzv. zdrojového členu

Úvod

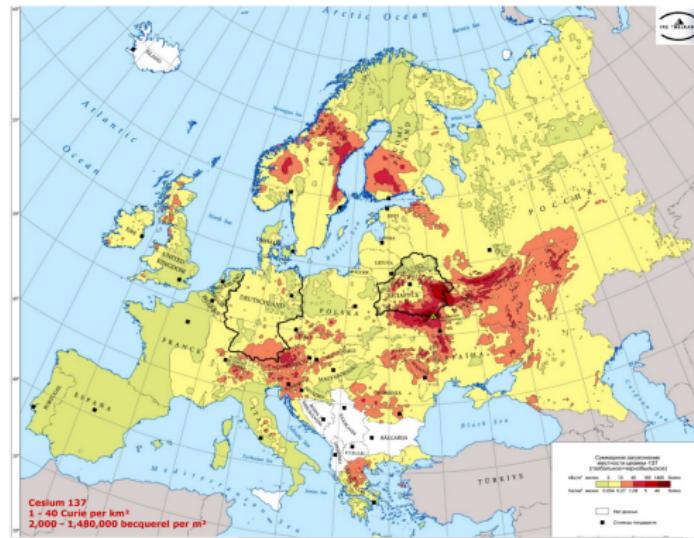
- ▶ Příklad: požáry v oblasti Černobylu v roce 2015.



Aplikace: odhad tzv. zdrojového členu

Úvod

- ▶ Pro odhad důsledků úniku a přijetí příslušných opatření potřebujeme znát:
 - ▶ Co a kolik toho uniklo?
 - ▶ Kde to uniklo?
 - ▶ Které oblasti a jak moc jsou zasažené?



Odhad zdrojového členu při radiačním úniku

Matematický popis

- ▶ Monitorovací síť (p měření) \longmapsto vektor \mathbf{y} .
- ▶ Reanalýza počasí + transportní model \longmapsto koeficienty $m_{i,j}$.
- ▶ Časový průběh úniku \longmapsto vektor $\boldsymbol{\beta}$ (časový průběh úniku).

$$y_1 = m_{1,1}\beta_1 + m_{1,2}\beta_2 + \cdots + m_{1,n}\beta_n,$$

$$y_2 = m_{2,1}\beta_1 + m_{2,2}\beta_2 + \cdots + m_{2,n}\beta_n,$$

:::

$$y_p = m_{p,1}\beta_1 + m_{p,2}\beta_2 + \cdots + m_{p,n}\beta_n,$$

- ▶ Maticový zápis: $\mathbf{y} = M\boldsymbol{\beta}$.

Odhad zdrojového členu při radiačním úniku

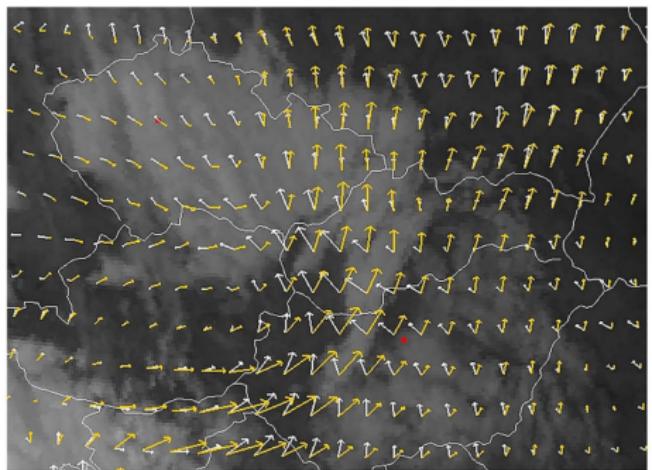
Monitorovací síť

- ▶ Měření koncentrace radionuklidů v ovzduší (SÚRO):



Odhad zdrojového členu při radiačním úniku

Reanalýzy počasí a transportní modely

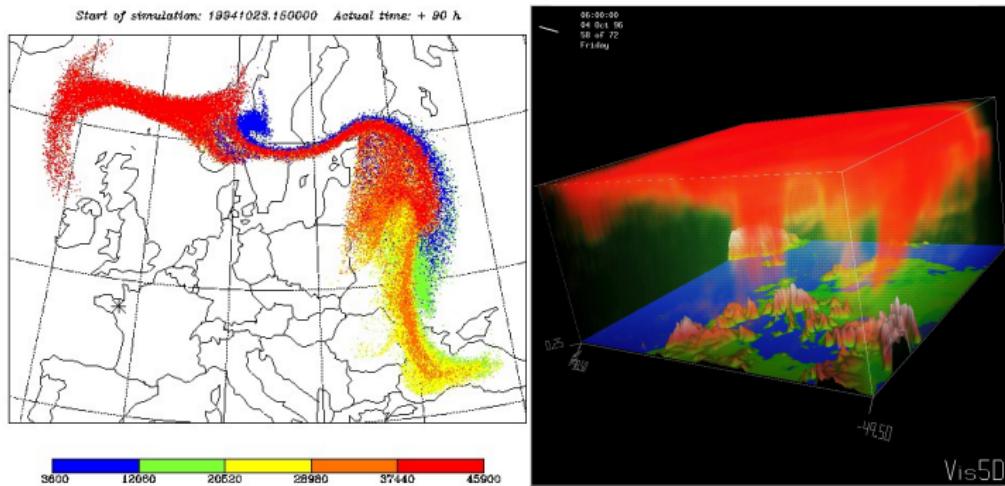


[Lee losy _ 2017]

Odhad zdrojového členu při radiačním úniku

Reanalýzy počasí a transportní modely

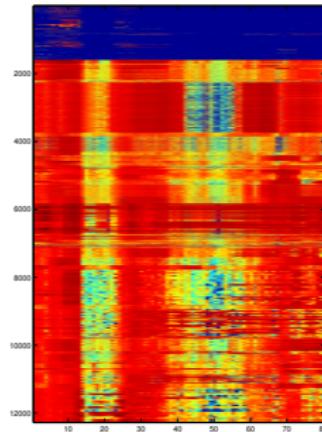
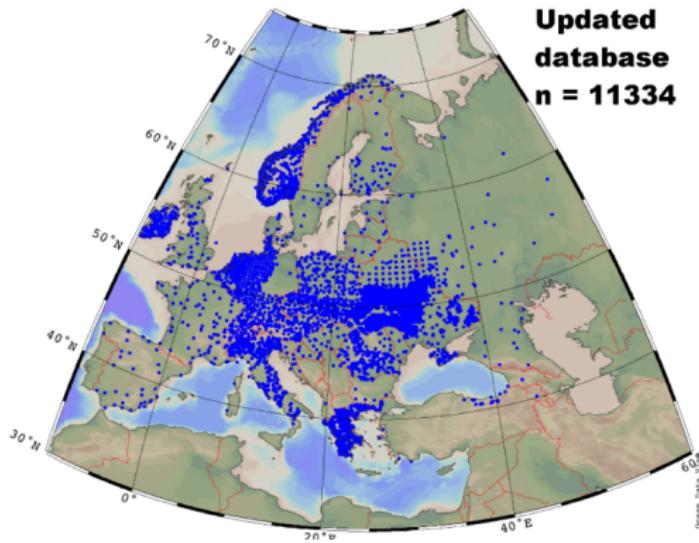
- Modely transportu částic a plynů.



Případ 1: Chernobyl

Cs-137

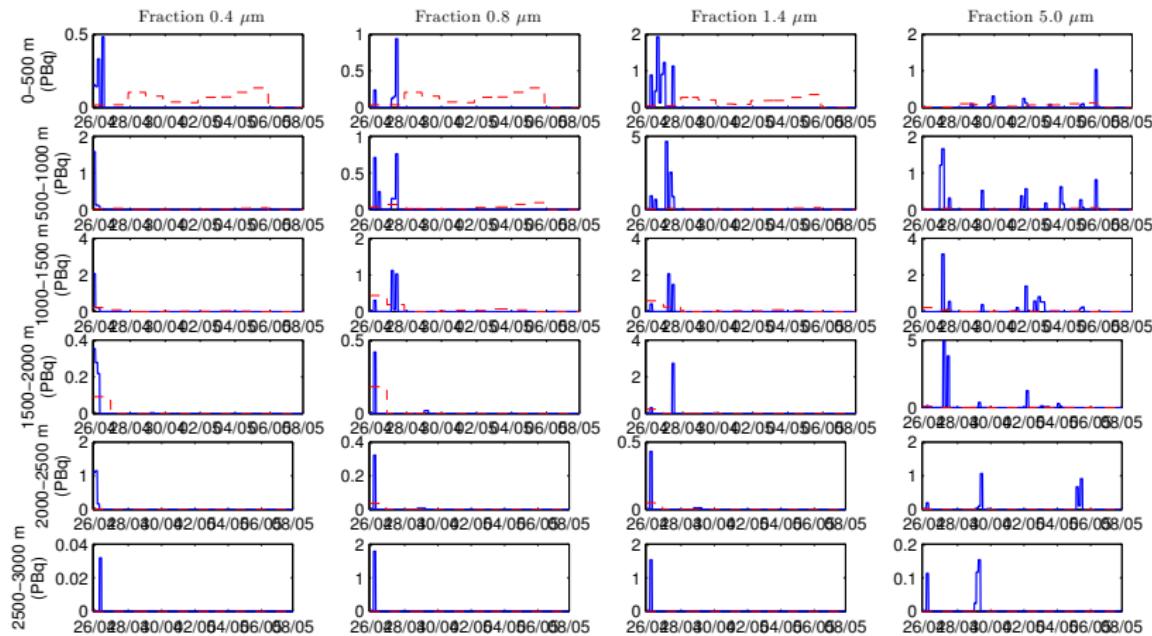
- ▶ časové rozlišení 3h (celkem 10 dnů)
- ▶ simulace z 6 výškových hladin (0 - 3000m)
- ▶ simulace 4 velikostí částic (0.4, 0.8, 1.4, 5.0 μm)



Případ 1: Chernobyl

Cs-137

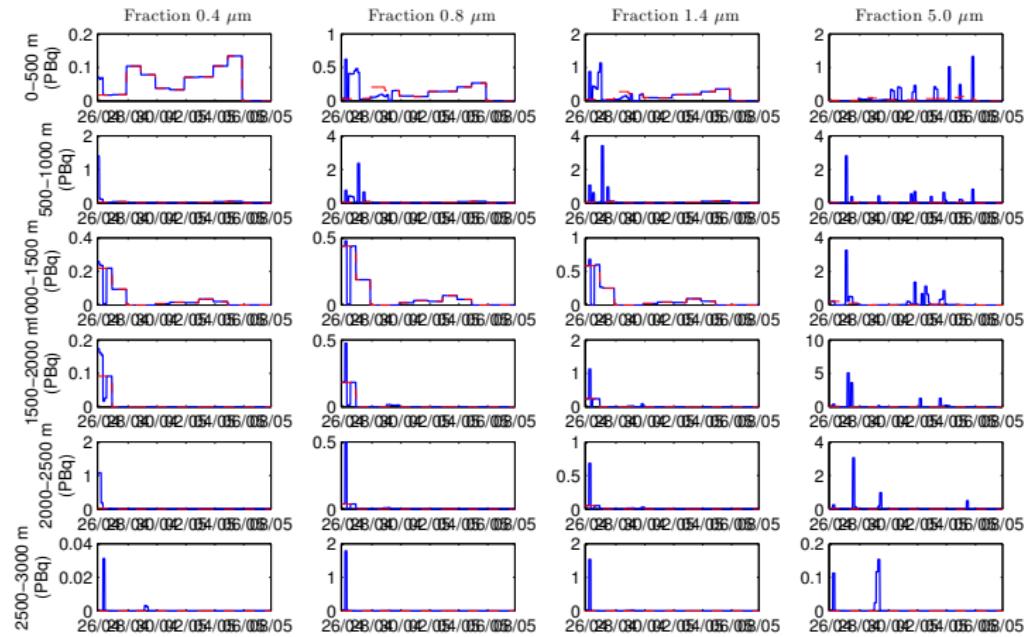
- ▶ výpočet se spike-and-slab apriornem



Případ 1: Chernobyl

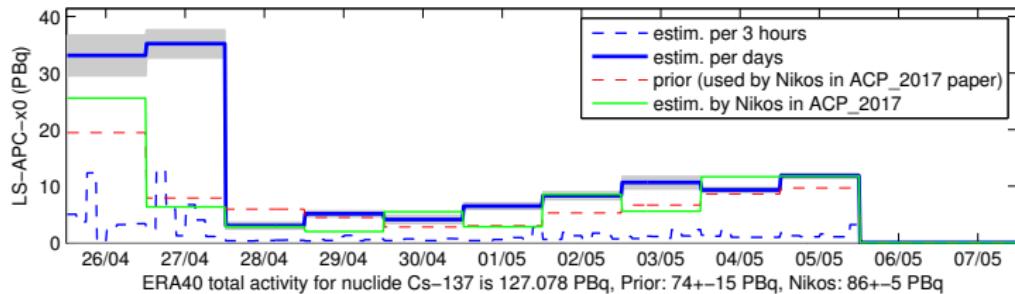
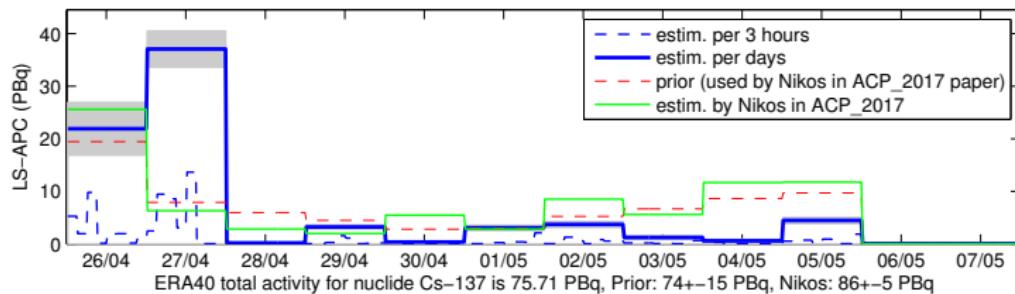
Cs-137

- ▶ výpočet se spike-and-slab apriornem a předpokladem známého β_0



Případ 1: Chernobyl

Cs-137



Případ 2: únik I-131 v roce 2011

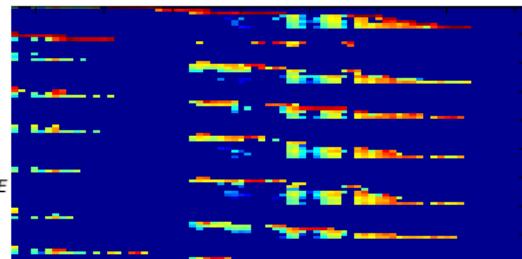
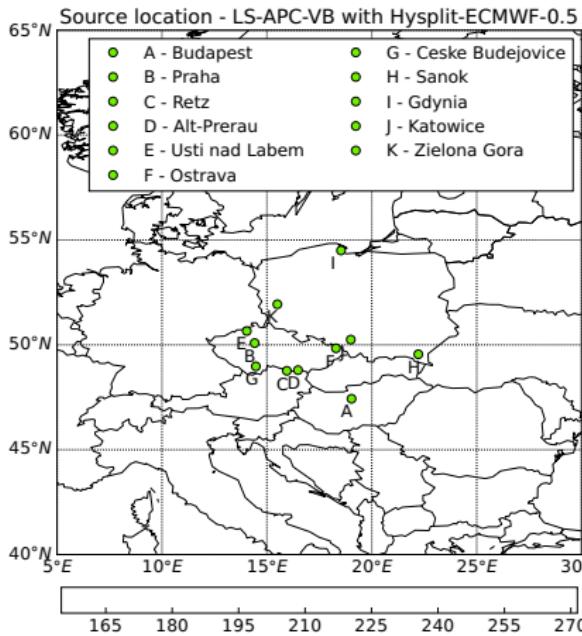
I-131

- ▶ V říjnu 2011 byl zaznamenán zvýšený výskyt jódu I-131 v ovzduší ve střední Evropě.
- ▶ 11. listopadu byla informována IAEA, která spustila vyšetřování.
- ▶ 17. listopadu byla IAEA informována o zdroji úniku jódu, maďarské továrni na výrobu izotopů (pro lékařské účely).
- ▶ Později bylo zjištěno:
 - ▶ únik trval od 8. září do 16. listopadu,
 - ▶ celkově unikl jód o aktivitě 342 GBq,
 - ▶ maximum úniku bylo mezi 12. a 14. říjnem (108 GBq),
 - ▶ únik nepřekročil roční povolenou dávku pro danou továrnu.
- ▶ Náš cíl: na základě měření tato zjištění potvrdit/vyvrátit.

Případ 2: únik I-131 v roce 2011

I-131

Měření z:



Případ 2: únik I-131 v roce 2011

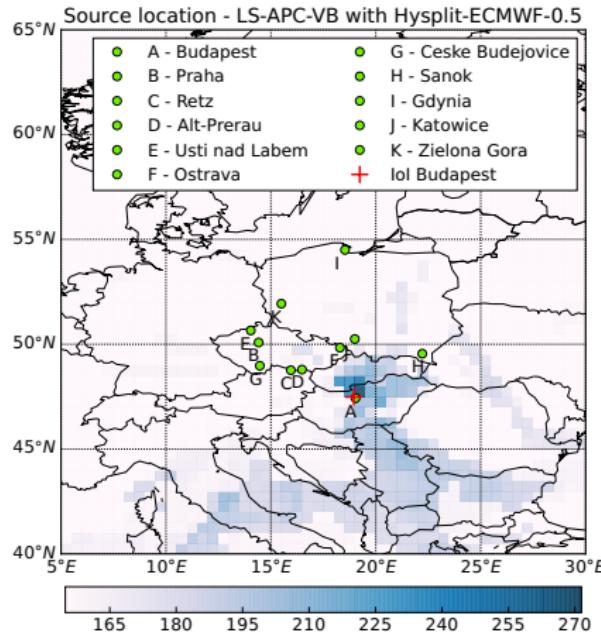
Určení lokace

- ▶ matici M mohu napočítat pro každý bod mapy,
- ▶ odhadu zdrojového členu β pro tyto body
- ▶ vyhodnocení “nejlepšího” bodu, např.
 - ▶ scatter plot
 - ▶ variační lower bound

$$f(M_k | \mathbf{y}) \propto f(M_k) \exp(E[f(\mathbf{y}, \theta_1, \dots, \theta_q)] - E[\theta_1] - \dots - E[\theta_q])$$

Případ 2: únik I-131 v roce 2011

I-131



Případ 2: únik I-131 v roce 2011

I-131

- ▶ Odhad zdrojového členu z nejpravděpodobnější lokace:

