Hierarchical Prior for Source Term Determination and Its VB Estimation

Václav Šmídl, Ondřej Tichý and Radek Hofman
Institute of Information Theory and Automation of the CAS, Prague, Czech Republic
[smidl,otichy,hofman]@utia.cas.cz

1. Introduction

We are concerned with the problem of determination of a source term of an atmospheric release. In its simplest form, the problem can be viewed as a liner system $y = Mx$ with vector of observations y and source-receptor-sensitivity (SRS) matrix M, where solution x can be found using least-square method. In reality, this is usually not possible since the system is ill-conditioned and an optimisation of a regularised cost function must be used:

$$J(x) = (y - Mx)^\top R^{-1}(y - Mx) + x^\top Bx, \quad (1)$$

where the first term minimizes the error of the measurements with covariance matrix R, and the second term is the regularization with weight α. Various types of regularization arise for different choices of matrix B. For example, Tikhonov regularization arises for B in the form of identity matrix, and smoothing regularization for Laplacian operator. In this contribution, we interpret B probabilistically as a covariance matrix of the prior distribution of x.

Traditionally, B is chosen in a fixed form, e.g.

$$B = \alpha I + \beta\Delta^\top\Delta, \quad (2)$$

where Δ is a differentiating operator and α, β are chosen scalar factors. The choise is often done manually by a trial and error procedure, e.g. [1, 3]. We propose to relax the assumption of known structure of B and use an **objective Bayesian approach** for their estimation from data.

2. Bayesian approach

In Bayesian approach, the task of inferring data from observations is formalized as an update of prior belief about the parameter values updated by the available observations. Formally, our knowledge about the parameter vector x is described by the probability density function

$$p(x|y, M) = \frac{p(y|x, M)p(x)}{\int p(y|x, M)p(x)dx} \propto p(y|x, M)p(x), \quad (3)$$

where $p(x)$ is the prior distribution, $p(y|x, M)$ is the likelihood of the measurements. For the choice of Gaussian models

$$p(y|x, M) = \mathcal{N}\left(Mx, R^{-1}\right), \quad p(x|B) = \mathcal{N}\left(0, B^{-1}\right), \quad (4)$$

the result of the Bayes rule (3) is again a Gaussian distribution $p(x|y) = \mathcal{N}(\hat{x}, \Sigma_x)$, where $\hat{x}$ coincides with the optimizer of (1) and the covariance matrix is $\Sigma_x = \left(B + M^T R^{-1} M\right)^{-1}$. Hence, the standard approach to the source term estimation is equivalent to maximum aposteriori estimate of the outlined Bayesian solution.

In common with the standard approach, the precision matrices R, B are assumed to be known. A key advantage of the of Bayesian approach is its ability to infer even these parameters from the data.

3. Prior model for unknown covariance

Consider the following structure of matrix B:

$$B = L^\top \Upsilon L, \quad L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ l_1 & 1 & 0 & 0 \\ 0 & \ddots & 1 & 0 \\ 0 & 0 & l_{n-1} & 1 \end{pmatrix}, \quad \Upsilon = \begin{pmatrix} \upsilon_1 & 0 & 0 & 0 \\ 0 & \upsilon_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \upsilon_n \end{pmatrix}.$$

where the vectors of unknowns are $\mathbf{l} = [l_1, \ldots, l_{n-1}], \boldsymbol{\upsilon} = [\upsilon_1, \ldots, \upsilon_n]$. The Bayesian formalism requires to define prior distribution on all unknowns. We choose

$$p(\upsilon_i) = \mathcal{G}(a_0, b_0) \qquad p(l_i) = \mathcal{N}(-1, \sigma_0) \quad (5)$$

The formal solution of the estimate is difficult to evaluate:

$$p(x|y, M, \boldsymbol{\upsilon}, \mathbf{l}) = \int p(y|x, M)p(x|\mathbf{l}, \boldsymbol{\upsilon})p(\mathbf{l}, \boldsymbol{\upsilon})d\mathbf{l}, \boldsymbol{\upsilon}. \quad (6)$$

4. Variational solution of unknown B

Analytical solution of (6) is not available and a suitable approximation has to be found. Following the Variational Bayes approximation [2], the posterior estimate can be obtained by solving the following equations:

$$\tilde{p}(x|y, M) \propto \exp(\int \tilde{p}(\boldsymbol{\upsilon}|y)\tilde{p}(\mathbf{l}|y) \log p(x, y, \mathbf{l}, \boldsymbol{\upsilon}|M)d\boldsymbol{\upsilon}\, d\mathbf{l}),$$

$$\tilde{p}(\mathbf{l}|y, M) \propto \exp(\int \tilde{p}(x|y)\tilde{p}(\boldsymbol{\upsilon}|y) \log p(x, y, \mathbf{l}, \boldsymbol{\upsilon}|M)dx\, d\boldsymbol{\upsilon}),$$

$$\tilde{p}(\boldsymbol{\upsilon}|y, M) \propto \exp(\int \tilde{p}(x|y)\tilde{p}(\mathbf{l}|y) \log p(x, y, \mathbf{l}, \boldsymbol{\upsilon}|M)dx\, d\mathbf{l}),$$

which can be solved iteratively [2].

5. Positivity enforcement

Since the source term can not be negative, we seek only for positive solutions of the problem. Hence, we may restrict the support of prior $p(x)$ to positive domain only using truncated normal distribution

$$p(x_j) = t\mathcal{N}(0, \sigma_{x_j}^{-1}, \langle 0, \infty)),$$

which moments are non-trivial but available as

$$\hat{x} = \mu - \sqrt{\sigma}\frac{\sqrt{2}[\exp(-\beta^2) - \exp(-\alpha^2)]}{\sqrt{\pi}(erf(\beta) - erf(\alpha))}. \quad (7)$$

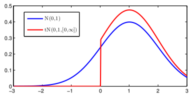


Figure: Example of the normal distribution $\mathcal{N}(1, 1)$, blue line, and the truncated normal distribution $t\mathcal{N}(1, 1, <0, \infty>)$, red line.

6. Numerical experiment

For testing we use data from ETEX-1. We have 3012 concentration samples from 168 stations with resolution of 3 hours. SRS matrix was calculated by LDPM FLEXPART. Release was vertically homogeneously distributed between 0 and 50 m. 340 kg of tracer was released between 23 Oct 16:00 UTC - 24 Oct 3:50 UTC, 1994, i.e. 11:50 h duration, here approximates as 12 hours. Temporal resolution of x is 1h.

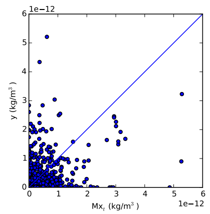
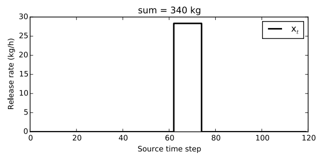


Figure 1: Scatter plot of model vs. measurements (x_t is true source term).

Source receptor sensitivities were calculated for three vertical layers: 0–50m, 50–300m and 300–1000m. Shape of the true source term follows:



7.1 Subjective method - single vertical layer

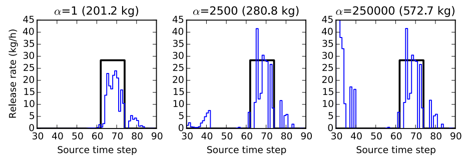
Performance of presented **adaptive** method will be compared to a well established method based on minimisation of (1) with

$$R = 1.3 \times 10^{25} I, \quad B = \alpha I$$

When analytically minimised, optimal x and its posterior error covariance P are given by LSE

$$(M^\top R^{-1} M + B)x = M^\top R^{-1} y, \quad P = (M^\top R^{-1} M + B)^{-1}. \quad (8)$$

Negative parts of the solution are iteratively suppressed by reduction of prior error variance of those elements where negative solutions occurred. Such elements are then forced to stay close to its prior value (here 0). Iterations are stopped when the majority of release is positive, here 99.9%. The method is linear and robust but its drawback is that coefficients r and α must be selected **subjectively** or tuned by a try and error approach, e.g. [1, 3]. To show that the solution is highly dependent on α we evaluate x $\sqrt{\alpha} \in \{1, 50, 500\}$:



7.2 Adaptive method - single vertical layer

Results in this section were using adaptive method where along x we **adaptively estimate** also both R and B. Here, three versions of B is studied, (2) with selected weights α and β and its generalization $B = L\Upsilon L^\top$.

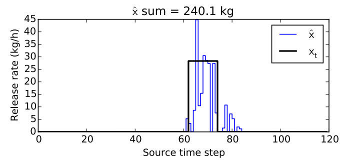


Figure 2: Sparse solution: $l_i = 0, \forall i$.

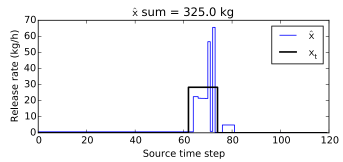


Figure 3: Sparse differences: $l_i = -1, \forall i$.

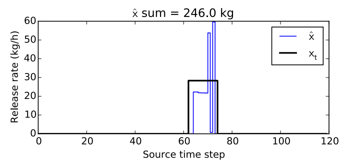
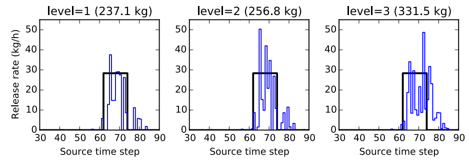


Figure 4: Nondiagonal Adaptive Covariance (NAC): l_i estimated.

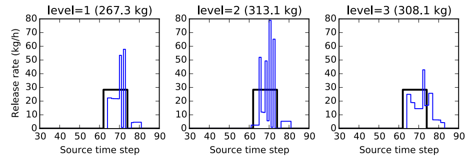
8.1 Subjective method - vertical profile

Following source term for all three vertical layers were obtained using LSE method with $R = 1.3 \times 10^{25} I$ and $B = \alpha I$, $\alpha = 25$. Since the structure of SRS matrix M is quite similar for all vertical layers, source terms given by the inversion are quite similar and it is hard to determine in which vertical layer the release occurred.



8.2 Adaptive method - vertical profile

Source term for all three vertical layers were obtained using NAC.



Adaptive method also provides a similar estimates for all vertical layers. However, as a Bayesian method, it gives us also probabilities of releases at each layer using Bayesian model selection:

$$p(M_i|y) \propto \int p(y|x, M_i, B)p(x|B)p(B)dxdB, \quad (9)$$

which is also approximated by the Variational Bayes method.

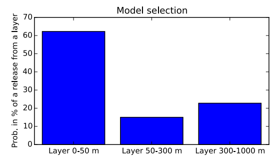


Figure 5: Probabilities of release occurrence at difference vertical layers.

Conclusion

This methodology is very flexible and can be also used for mutual probabilistic evaluation of different SRS matrices, which can be used for source localization.

References

- Olivier Saunier, Anne Mathieu, Damien Didier, Marilyne Tombette, Denis Quélo, Victor Winiarek, and Marc Bocquet.
  An inverse modeling method to assess the source term of the fukushima nuclear power plant accident using gamma dose rate observations.
  Atmospheric Chemistry and Physics, 13(22):11403–11421, 2013.
- Václav Šmídl and Anthony Quinn.
  The variational Bayes method in signal processing.
  Springer Science & Business Media, 2006.
- A Stohl, P Seibert, G Wotawa, D Arnold, JF Burkhart, S Eckhardt, C Tapia, A Vargas, and TJ Yasunari.
  Xenon-133 and caesium-137 releases into the atmosphere from the fukushima dai-ichi nuclear power plant: determination of the source term, atmospheric dispersion, and deposition.
  Atmospheric Chemistry and Physics, 12(5):2313–2343, 2012.

This research is supported by EEA/Norwegian Financial Mechanism under project 7F14287 STRADI: Source-Term Determination of Radionuclide Releases by Inverse Atmospheric Dispersion Modelling. Poster created with beamerposter.

WWW: *http://stradi.utia.cas.cz/*