Bayesian Estimation of Prior Variance in Source Term Determination

Václav Šmídl and Radek Hofman

Institute of Information Theory and Automation of the CAS, Prague, Czech Republic

[smidl,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 solution of a liner system $y = Mx$ with vector of observations y and source-receptor-sensitivity (SRS) matrix M. In reality, this is usually not possible since the system is ill-conditioned and an optimisation of a regularised cost function must be used:

$$J(x) = (y - Mx)^\top R^{-1}(y - Mx) + x^\top(\alpha B)^{-1}x, \qquad (1)$$

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.

Typically, B is assumed to be known, and α is optimized manually by a trial and error procedure, e.g. [1, 3]. We propose to relax the assumption of known α and B and use an **objective Bayesian approach** for their estimation from data. The general problem is not analytically tractable and approximate estimation techniques has to be used. We present Variational Bayesian (VB) solution [2] of two special cases based on different assumptions:

(i) structure of B is known and only the α is estimated. VB method yields in this case an iterative estimation algorithm.

(ii) B is diagonal with all elements unknown and $\alpha = 1$. This prior favors sparse solutions of the inverse problem.

The method is demonstrated on data from ETEX tracer experiment.

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

$$p(x|y, M) = \frac{p(y|x, M)p(x)}{\int p(y|x, M)p(x)dx} \propto p(y|x, M)p(x), \qquad (2)$$

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

$$p(y|x, M) = \mathcal{N}\left(Mx, R^{-1}\right), \quad p(x) = \mathcal{N}\left(0, (\alpha B)^{-1}\right), \qquad (3)$$

the result of the Bayes rule (2) 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 $\Sigma_x = \left((\alpha B)^{-1} + M^T R^{-1} M\right)^{-1}$. 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. Hierarchical model for unknown α

The variance of the prior distribution is typically unknown. In Bayesian formalism, it can be estimated from the data. This requires formulation of prior distribution of the unknown parameters $p(\alpha)$. A common choice of the prior is the gamma distribution:

$$p(\alpha) = \mathcal{G}(a_0, b_0) = \frac{b_0^{a_0}}{\Gamma(a_0)}\alpha^{a_0-1}\exp(-b_0\alpha), \alpha > 0. \qquad (4)$$

The formal solution of the estimate is difficult to evaluate:

$$p(x|y, M) = \int p(y|x, M)p(x|\alpha)d\alpha. \qquad (5)$$

4. Hierarchical model of diagonal B

A more complex model arise when we consider the matrix B to be also unknown. As a first step, we assume that B is diagonal modelled by vector diag(d).
The prior distribution of all elements of d_i may be chosen as

$$p(d_i) = \mathcal{G}(a_0, b_0) = \frac{b_0^{a_0}}{\Gamma(a_0)}d_i^{a_0-1}\exp(-b_0 d_i), d_i > 0. \qquad (6)$$

Following the Variational Bayes approximation, the posterior estimate can be obtained by solving the following equations:

$$\tilde{p}(x|y, M) \propto \exp\left(\int \tilde{p}(d|y)\log p(x, d, y|M)d\,d\right),$$
$$\tilde{p}(d|y, M) \propto \exp\left(\int \tilde{p}(x|y)\log p(x, d, y|M)dx\right),$$

yielding iterative Algorithm 1.

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

$$p(x) = t\mathcal{N}(0, \mathrm{diag}(d)^{-1}, \langle 0, \infty\rangle).$$

In this case the Variational Bayes solution requires to evaluate moments of the truncated Normal distribution which are

$$\widehat{x} = \mu - \sqrt{\sigma}\frac{\sqrt{2}[\exp(-\beta^2) - \exp(-\alpha^2)]}{\sqrt{\pi}(erf(\beta) - erf(\alpha))}, \qquad (7)$$

$$\widehat{x^2} = \sigma + \mu\widehat{x} - \sqrt{\sigma}\frac{\sqrt{2}[b\exp(-\beta^2)]}{\sqrt{\pi}(erf(\beta) - erf(\alpha))}. \qquad (8)$$

Algorithm 1

1. choose initial guess of $\hat{d}$,
2. solve least squares problem
$$J = (Mx - y)^T R^{-1}(Mx - y) + x^T \mathrm{diag}(\hat{d})x,$$
including the covariance matrix Σ_x.
3. compute moments on positive support, $\widehat{x}, \widehat{x^2}$
4. update estimate of d: $d_i = (a_0 + 0.5)/(b + 0.5(\widehat{x^2}))$,
5. if not converged GOTO 2.

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 approximated as 12 hours. Temporal resolution of x is 1h.

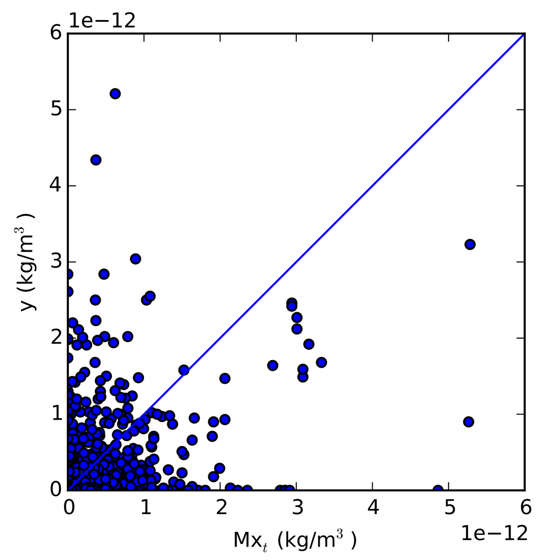

Figure 1: Scatter plot of model vs. measurements (x_t is true source term).

7.1 Result of linear regression

To demonstrate bad conditioning of the problem we present a linear regression solution $y = Mx$:

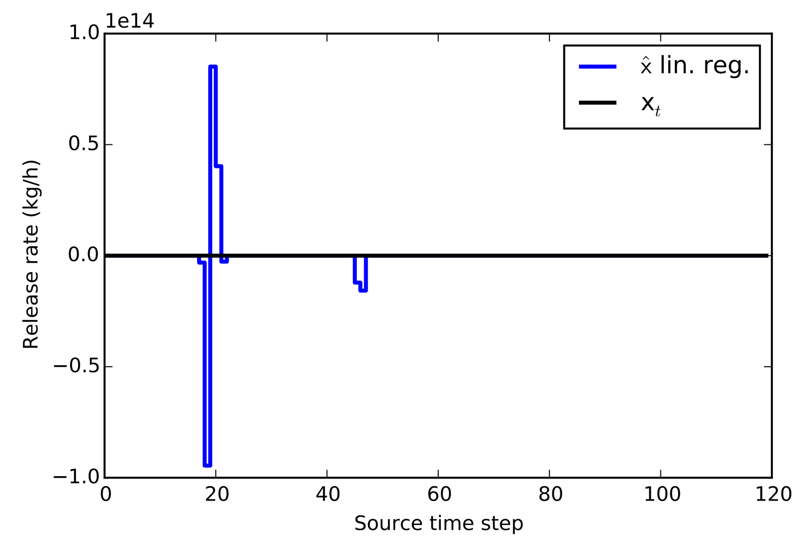

Figure 2: Liner regression solution of the problem.

Such an estimate of x is of no use and we have to employ more advanced methods.

7.2 Results of subjective method

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

$$\alpha = 1, \quad R = r^2 I, \quad B = b^2 I$$

being diagonal matrices (I is an identity matrix). When analytically minimised, optimal x and its posterior error covariance P are given by LSE

$$(M^\top R^{-1}M + B^{-1})x = M^\top R^{-1}y, \quad P = (M^\top R^{-1}M + B^{-1})^{-1}. \qquad (9)$$

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 x is positive, here 99.9%. The method is linear and robust but its drawback is that coefficients r and b 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 b we evaluate x for $r = 1 \times 10^{-13}$ and $b \in \{1, 20, 500\}$:

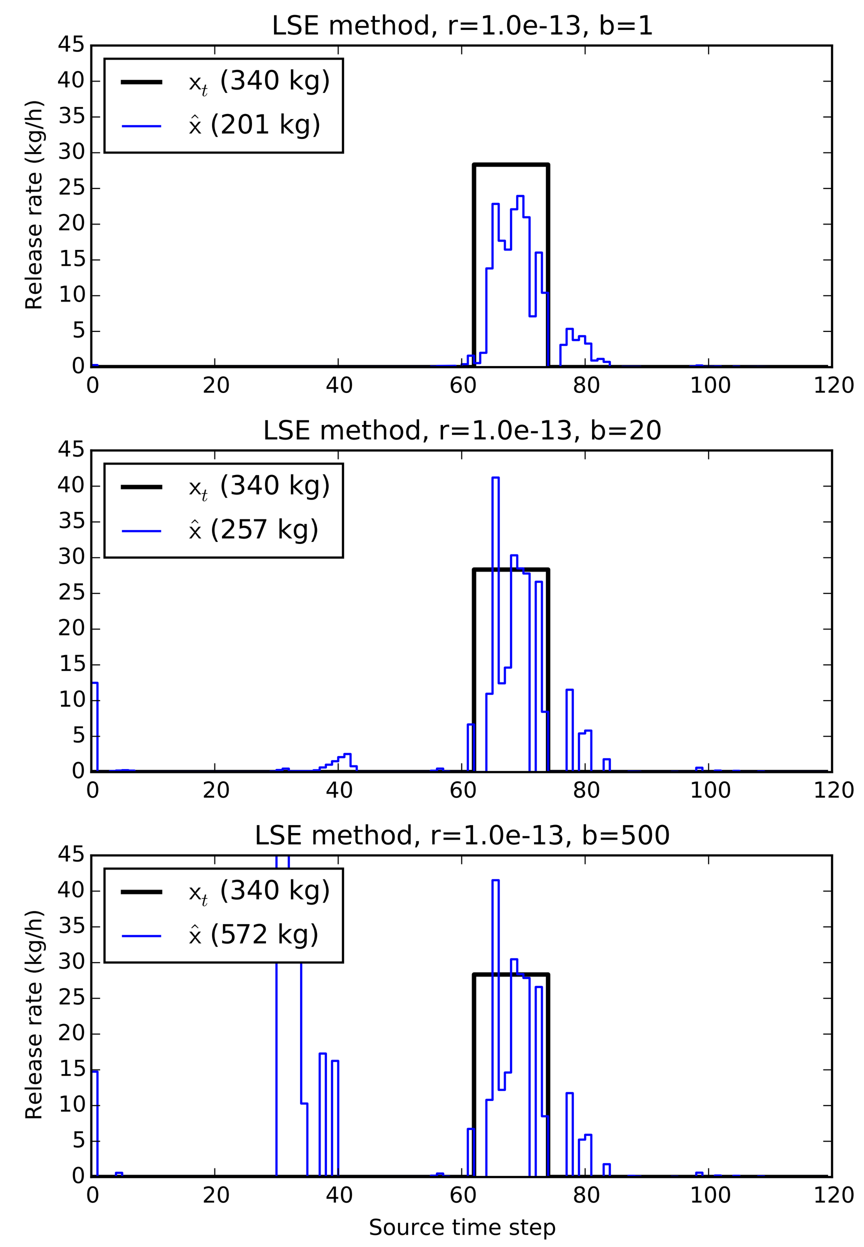

Figure 3: LSE solution for different values of b (prior error).

7.3 Results of adaptive method

Results in this section were obtained using different adaptive methods where we estimate α for some fixed B or a diagonal B for fixed α.

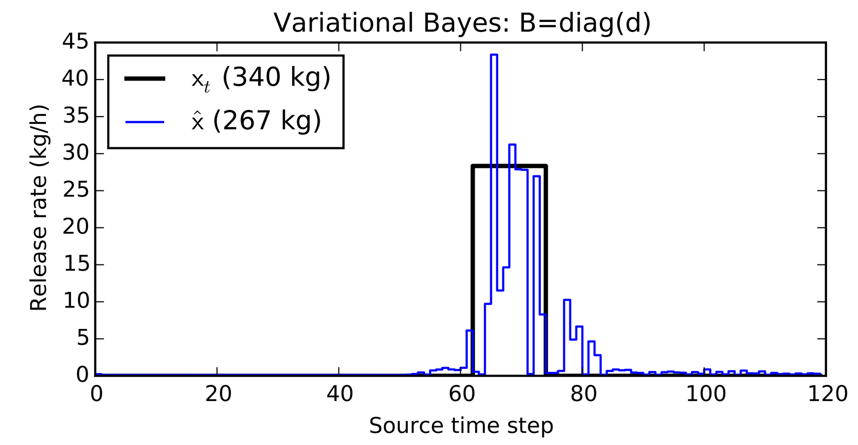

Figure 4: VB solution when $\alpha = 1$ and diagonal elements of B are estimated. This solution resemble solution of LSE for $b = 20$.

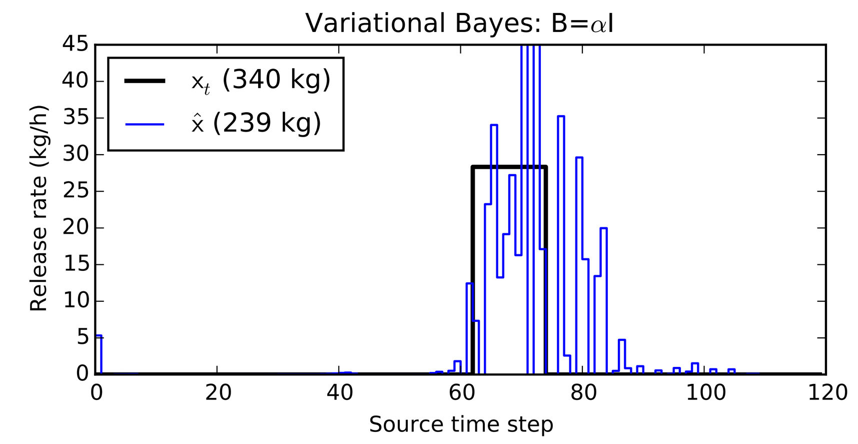

Figure 5: VB solution when α is estimated and $B = I$.

Lastly, we examine solution for $B = \alpha A^\top A$, where α is estimated and A is a difference operator introducing covariance between adjacent elements of x:

$$A = \begin{bmatrix} 1 & -1 & \\ & 1 & -1 \\ & & 1 \end{bmatrix}, \quad B = A^\top A = \begin{bmatrix} 1 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{bmatrix}$$

This settings produces a smooth solution:

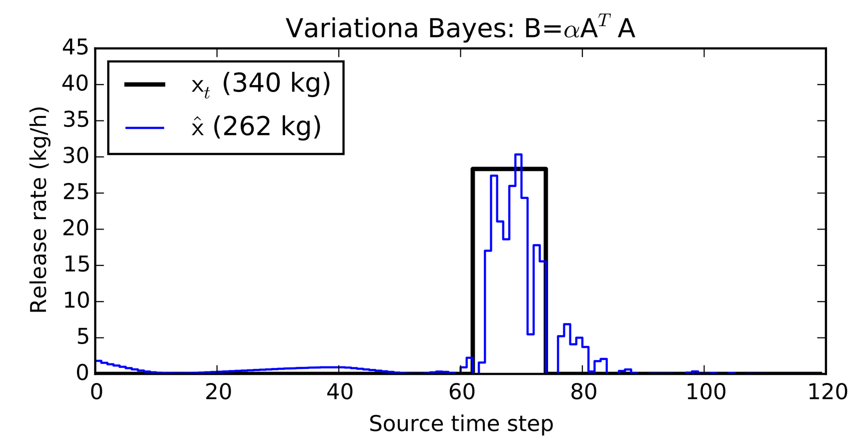

Figure 6: VB solution when α is estimated and $B = A^\top A$

Conclusion

Adaptive methods seem to be promising because they eliminate subjective factor from initialisation of source inversion setup. Our future work will be focused on application of the method to more complex data and its extension to estimate both α and B (even in non-diagonal forms).

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/*