Atmospheric Pollution Research 5 (2014) 219-225

Atmospheric Pollution Research

www.atmospolres.com

Application of direct regularization techniques and bounded–variable least squares for inverse modeling of an urban emissions inventory

Ana Y. Vanoye, Alberto Mendoza

Department of Chemical Engineering, Tecnológico de Monterrey, Ave. Eugenio Garza Sada 2501, Monterrey 64849, Mexico

ABSTRACT

Inverse modeling, coupled with comprehensive air quality models, is being increasingly used for improving spatially and temporally resolved emissions inventories. Of the techniques available to solve the corresponding inverse problem, regularization techniques can provide stable solutions. However, in many instances, it is not clear which regularization parameter selection method should be used in conjunction with a particular regularization technique to get the best results. In this work, three regularization techniques (Tikhonov regularization, truncated singular–value decomposition, and damped singular–value decomposition) and three regularization parameter selection methods (generalized cross validation, the L–curve method (LC), and normalized cumulative periodograms) were applied in conjunction with an air quality model with the aim of identifying the best combination of regularization technique and parameter selection method when using inverse modeling to identify possible flaws in an urban–scale emissions inventory. The bounded–variable least–squares method (BVLS), which is not usually considered a regularization method, was also investigated. The results indicate that the choice of the regularization parameter explains most of the differences between the regularization techniques used, with the LC method exhibiting the best performance for the application described here. The results also show that the BVLS scheme provides the best agreement between the observed and modeled concentrations among the mathematical techniques tested.

Keywords: *Air quality model, photochemical modeling, emissions evaluation, inverse problem*

Corresponding Author:
Alberto Mendoza
☎ : *+52-81-8158-2091*
🖷 : *+52-81-8328-4250*
✉ : *mendoza.alberto@itesm.mx*

Article History:
Received: 08 October 2013
Revised: 19 December 2013
Accepted: 20 December 2013

doi: 10.5094/APR.2014.027

1. Introduction

Three–dimensional comprehensive air quality models (AQMs) describe the atmospheric transport and transformation of trace species and are routinely used in the development of pollution–reduction strategies and other air quality management policies. However, to produce accurate, trustworthy results, AQMs rely on the use of detailed emission inventories that, even today, convey a great degree of uncertainty (Miller et al., 2006; Russell, 2008). This translates into model applications in which discrepancies between model–derived concentrations and observations of air pollutants can be quite large. In this description, we are assuming that the AQM is perfect and the emissions are one of the most uncertain input parameters in the modeling effort (e.g. Russell and Dennis, 2000; Tian et al., 2010).

One way of reducing emission inventory uncertainty is to use inverse modeling or data assimilation techniques to identify possible flaws in the construction of such emission inventories. Applications of inverse modeling range from global (e.g., Petron et al., 2002) or continental (e.g., Elbern et al., 2007) to regional or urban scales (e.g., Quelo et al., 2005) for a variety of atmospheric species such as stratospheric ozone depletion substances (e.g., Xiao et al., 2010), greenhouse gases (e.g., Stohl et al., 2009), radioactive material (e.g., Winiarek et al., 2012), and tropospheric ozone and aerosol precursors (e.g., Gilliland et al., 2006; Napelenok et al., 2008; Henze et al., 2009). In this context, several mathematical techniques have been used to find solutions for the corresponding inverse problems, including four–dimensional data assimilation (e.g., Meirink et al., 2008), Kalman or ensemble Kalman filtering (e.g., Napelenok et al., 2008), and the use of adjoint models (e.g., Hakami et al., 2005). Li et al. (2010) used genetic algorithms for optimizing inventories, but their application was limited because of the necessary computational requirements.

One approach for performing this top–down emissions inventory evaluation is to first use a forward model (the AQM) to compute both the simulated concentration fields of pollutants and their responses to changes in emissions (sensitivity fields). With this, a linear model of the form $Gm=d$ can be constructed, where d is a vector containing the difference between modeled and observed concentrations, G is a matrix containing the sensitivity coefficients of all pollutant species to changes in the emission strengths, and m is a vector of emission strength changes that brings the observations and model–derived concentrations into agreement. Then, if an over–determined least–squares problem is solved, the corresponding inverse model can be represented as $m^{est}=(G^TG)^{-1}G^Td$, where G^T is the transpose of G. However, inverse problems are typically ill conditioned, and this is an inconvenience because in practice, observations often possess a certain degree of error or noise (Aster et al., 2005).

Several mathematical techniques based on the incorporation of known properties about the solution have been developed with the aim of improving the conditioning of direct inverse problems, including regularization. However, there are few examples of the use of regularization techniques to obtain inverse–derived emissions inventories. In particular, few formal methods have been applied to obtain the value of the regularization parameters. This paper addresses the issue of performing inverse modeling of an urban air–pollutant emission inventory by applying direct regularization techniques. Three regularization techniques and

three regularization parameter choice methods were assessed. An additional technique investigated, which is not usually considered a regularization method, was the bounded–variable least–squares (BVLS) method.

2. Methods

2.1. Regularization methods

Three regularization methods were explored in this work: Tikhonov regularization (TIKH), truncated singular–value decomposition (TSVD), and damped singular–value decomposition (DSVD). Tikhonov's method consists of substituting the least–squares problem for a problem of the form (Neumaier, 1998):

$$m^{est} = \min\left\{\|Gm-d\|_2 + \|Lm\|_2\right\} \quad (1)$$

where $L=\lambda I$, I is typically the identity matrix, and $\lambda \in \mathbb{R}$ is a regularization parameter that controls the weight given to the minimization of the additional restriction relative to the minimization of the residual norm. Thus, TIKH seeks a solution that minimizes a criterion made up of the sum of two components: a weighted least–square term and a quadratic penalty term on the solution.

The singular–value decomposition of matrix G with r=rank(G), as in the following equation:

$$G = USV^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T \quad (2)$$

can be used to obtain the Moore–Penrose pseudo–inverse of G. However, the generalized inverse solution can become unstable when some of the singular values, σ_i, are small (Aster et al., 2005). Therefore, in TSVD, it is assumed that it is possible to recover a useful model by truncating the sum in Equation (2) in an upper–bound $k<r$ before the smallest singular values start dominating (Hansen, 1990). When $k=r$, the solution of TSVD is identical to the solution obtained by ordinary least–square methods. However, a solution obtained from TSVD with $k<r$ will tend to be more stable (Aster et al., 2005).

In equation (2), U is an orthogonal matrix with columns that are unit basis vectors (u_i) spanning the data space, V is an orthogonal matrix with columns that are basis vectors (v_i) spanning the model space, V^T is the transpose of V, v_i^T is the transpose of v_i, and S is a diagonal matrix with nonnegative diagonal elements (σ_i, the singular values).

Finally, DSVD (Ekstrom and Rhoads, 1974) may be regarded as a regularization method that follows Tikhonov in terms of its TSVD, with the difference being that DSVD introduces a smoother cutoff by means of filter factors that decay slower than the Tikhonov method, overall requiring less filtering (Hansen, 1998; Lin et al., 2011).

2.2. Regularization parameter choice methods

Although the proper choice of the regularization parameter (either the continuous parameter λ or the discrete parameter k) is essential for the effectiveness of the regularization methods applied (Hansen, 1998), the optimal determination of this parameter remains an open issue (Krawczyk–Stando and Rudnicki, 2007; Lin et al., 2011). Some previous studies where inverse modeling has been applied to evaluate emissions inventories have used formal methods to obtain the value of the regularization parameter (Fan et al., 1999; Mendoza–Dominguez and Russell, 2001; Krakauer et al., 2004; Chai et al., 2009; Henze et al., 2009; Saide et al., 2009). However, the strategy of assigning values to the regularization parameter subjectively, or empirically, prevails (e.g., Eckhardt et al., 2008).

In this study, we explore three methods that do not require a good estimate of the noise variance: generalized cross validation (GCV) (Golub et al., 1979; Haber and Oldenburg, 2000), which is a parameter–choice method based on ordinary cross validation (Allen, 1974); the L–curve method (LC), which uses a plot of the valid regularization parameters of the (semi) norm of the regularized solution versus the corresponding residual norm (the best regularization parameter must be located in the corner of the L–curve) (Hansen and O'Leary, 1993); and normalized cumulative periodograms (NCP) (Rust, 2000; Rust and O'Leary, 2008), which chooses the regularization parameter for which the residual becomes closer to behave as white Gaussian noise.

2.3. Regularization with restrictions

Regardless of whether regularization methods tend to yield more stable, precise solutions, these solutions will often lack physical sense or violate some of the restrictions imposed by the nature of the problem. In our case we need to guarantee positive emissions. Several techniques incorporate additional restrictions to impose boundaries for the solution or add additional information about the solution. One such technique is the BVLS (Stark and Parker, 1995), which solves linear least–squares problems with upper and lower bounds on the variables.

BVLS uses an active set strategy in which the unconstrained least–squares problems for each candidate set of free variables are solved using the QR decomposition. The method also includes a "warm–start" feature that accelerates the solution by allowing for some of the variables to initialize at their upper or lower bounds. Stark and Parker's BVLS algorithm is based on the non–negative least squares method (Lawson and Hanson, 1974).

In this study, we use this additional technique in our application and compare it with the solutions obtained by regularization.

2.4. Application to the GMA emissions inventory

The base case, reported by Mendoza and Garcia (2009), for the modeling of photochemical pollutants in the Guadalajara Metropolitan Area (GMA; 20° 40' 25" N, 103° 20' 38" W) was used as case study. The GMA is the second largest urban center in Mexico, emissions are rather concentrated around the urban core (~600 km^2), and no important emission sources are located around this core. In the application described here, the same AQM, spatial configuration of the modeling domain, as well as the same meteorological fields, emissions inventory, and initial and boundary conditions were used. In that study, the California/Carnegie Institute of Technology (CIT) model extended with the capacity of estimating first–order sensitivity coefficients through the direct–decoupled method for three–dimensional models (Yang et al., 1997) was applied for the simulation of a three–day high–ozone concentration episode, occurring from May 16 to May 18, 2001. May was selected for the modeling exercise because it is the month when O_3 concentrations peak in the GMA (Zuk et al., 2007). The modeling domain was a computational matrix composed of 40×40 cells (horizontal resolution), with each cell being 4×4 km and geographically centered in the GMA. In addition, the domain included six vertical levels topping at 3 100 m.

The emission inventory used by Mendoza and Garcia (2009) was based on the 1995 Official Emissions Inventory for the GMA. This inventory had to be extended to provide coverage for the additional municipalities that were included in the modeling domain and that were not part of the GMA. In addition, this inventory had to be scaled from the base year (1995) to the modeled year (2001) and had to be spatially segregated based on