Comparative Study of Optimization Methods in a PM$_{2.5}$ Transport Adjoint Model

Ning Li[1,2], Daosheng Wang[2] and Xianqing Lv[2,*]

[1]School of Science, Dalian Jiaotong University, Dalian 116028, China
[2]Key Laboratory of Physical Oceanography, Ocean University of China, Qingdao, China
*Corresponding author

Abstract—**This paper focuses on the practical performances of the limited-memory BFGS (L-BFGS) method and the steepest descent method (GDM-S) by an adjoint data assimilation approach. The optimization procedure of the L-BFGS method in the ideal experiments clearly shows that the parameters should be scaled to similar magnitudes on the order of unity to improve the convergence efficiency. As compared with the GDM-S, the L-BFGS method really uses much fewer steps to reach a satisfactory solution, but the performances are almost the same in the parameters inversions with the two optimization algorithms. In practical experiments, simulation results show good agreement with the observations of the period when the 21th APEC summit took place.**

Keywords-PM$_{2.5}$ transport model; adjoint method; the L-BFGS method; the steepest descent method; parameters estimation

I. INTRODUCTION

PM$_{2.5}$ pollution has gained widespread concern due to its adverse influence on human and ecosystem health. Traditionally, they can be obtained from analysis of the PM$_{2.5}$ on a continental scale which depends on the sophisticate knowledge of the distribution. However, with incomplete observations in the spatial or temporal range, model estimates demonstrate significant uncertainties. Therefore, both further utilization of PM$_{2.5}$ measurements and improvement of model estimates are identified as the important components for the continued analysis of PM$_{2.5}$ sources.

Data assimilation methods provide a configuration for combining observations and models to form an optimal estimate of the PM$_{2.5}$ sources. In this method, observations are used to constrain estimates of model parameters that are both influential and uncertain. Among all data assimilation methods, four-dimensional variational (4D-Var) data assimilation is regarded as one of the most effective and powerful approaches developed over the past two decades. Efforts have been made with regard to the applications of the 4D-Var data assimilation in the study of PM$_{2.5}$ emissions[1-3].

The essential part of the adjoint assimilation method is to find model control variables that minimize the cost function. Large-scale optimization techniques are required in the previous minimization problem. The steepest descent method has been recognized as one of the simplest minimization algorithms so that it attracted many researchers' attention since it was put forward. With the negative gradient direction as the search direction, the steepest descent method is particularly useful when the dimension of the problem is very large. As an adaptation of the standard BFGS method, the L-BFGS method has been proved to be globally convergent for convex objective functions and provides a fast convergence rate with requirement of minimal storage.

The present work seeks to compare the performances of the L-BFGS method and the steepest descent method (GDM-S) via application of data assimilation technique using the adjoint of the PM$_{2.5}$ transport adjoint model [4].

This paper is organized as follows. In Section II, a series of ideal experiments are carried out for the comparison of the performances of optimization methods, including the L-BFGS method and the steepest descent method. In Section III, PM$_{2.5}$ concentration distributions are simulated a domain encompassing China. Finally, a summary and some conclusions are made in Section IV.

II. TWIN EXPERIMENT

In this study, the computing area is a domain encompassing China. The horizontal resolution is $0.5° \times 0.5°$, totally 141 $\times$ 71 grids in the area. The inverse integral time step is 600s and simulation period is 168 hours. The background value is fixed as 15.0 μg/m^3. Inflow boundary values are set equal to the background values. The wind data are derived from the National Centers for Environmental Prediction (NCEP) winds daily averaged for each $2.5°$ latitude by $2.5°$ longitude region. As shown in Figure I , seventy-four major Chinese cities, with the sign of "assimilated cities", are picked out and treated as the observation positions. The observations of these cities are assimilated in the optimization procedure. The PM$_{2.5}$ concentrations records of these cities are selected as the observations at two hours interval. Considering that actual in situ observations definitely contain noises, the artificial random error is added to the "observations". The maximum percentage error is 5%.

The governing model is run with the prescribed parameters. The model-generated PM$_{2.5}$ concentrations at grid points of the observation positions (See Figure I) are taken as the "observations". A initial guess value of control parameters, SS (or IC), is fixed as 0 in this work (or the background value for IC) and is chosen to run the governing model, then the simulation values are obtained. The adjoint model is driven by the discrepancy between simulated values and observations. The gradients with respect to control parameters are calculated by the adjoint variables obtained by backward integration of the adjoint equations. The control parameters at the

independent points should be updated with a certain optimization algorithm until the convergence criterion is met. The parameters at other points are determined by the interpolation of the values at the independent points. With the procedure repeated, the parameters will be optimized continuously and the difference between simulated values and "observations" will be diminished. So does the difference between the prescribed and the inverted parameters. Once a convergence criterion is met, the iteration will be terminated. In this work, the criterion is that the number of iteration steps is equal to 100 exactly in both ideal experiments and practical experiments.

For model performance evaluation, the values of normalized cost function, the mean absolute errors (MAEs) of the inversion results and the MAEs between the "observations" and simulated values in "assimilated cities" are selected and exhibited in Table 1. As the key for the simulation of $PM_{2.5}$, the inversions of the parameters , IC and SS, are the vital part of the whole model. These two parameters will be inverted together based on the actual $PM_{2.5}$ distributions and the emission. The optimization procedure in which the $PM_{2.5}$ observations could be assimilated to improve the simulated results.
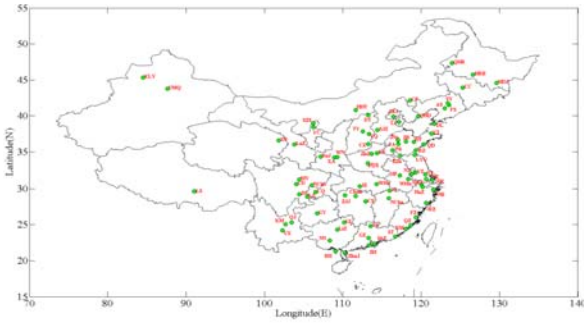

FIGURE I. THE COMPUTING AREA AND LOCATIONS OF THE 82 CITIES IN WHICH OBSERVATIONS ARE AVAILABLE

A. Numerical Implementation of the L-BFGS Method

In this work, we employ the L-BFGS version of Liu and Nocedal [5]. The number of corrections used in the L-BFGS update is taken as 5 (usually between 3 and 7, see Alekseev et al. [6]).

TE1 is carried out using the L-BFGS. The results in Table 1 indicate that the normalized cost functions are not converged absolutely, which leads to the failure inversions of the two parameters and terrible simulations. The similar results had been obtained in Yang and Hamrick [7]. In this work, emissions and deposition are both expressed by SS which combines the primary and secondary $PM_{2.5}$. According to the dimensional analysis and the model debugging, the magnitude of SS is set to be 10^{-5}. And the small magnitude of SS is regarded as the reason of the failed convergence during the assimilation procedure. Therefore, L-BFGS with an appropriate scaling are performed in TE2. The variation of the normalized cost function, and that of the error statistics for the efficiency testing are also presented in Table 1.

With the scaling, the case that the minimization process hardly converged in TE1 have been improved significantly. The results in Table1 illustrate that the values of the normalized cost functions have been decreased by three order of magnitude in TE2. Therefore, the scaling should be carefully made, especially when different kinds of parameters are intended to optimize by means of the L-BFGS.

TABLE I. THE MAE STATISTICS BEFORE AND AFTER ASSIMILATION

Experiment	$J_{100}/J1$	K_1 (µg/m³)		K_2 (10⁻⁶µg/m³/s)	
		Before	After	Before	After
TE1	2882	6.33	6.24	3.16	1411
TE2	0.00046	6.33	1.30	3.16	0.77
TE3	0.00053	6.33	1.50	3.16	1.01

J_{100} is the final value of cost function and J1 is the initial value of cost function. K_1 and K_2 are the MAEs between prescribed and inverted IC and SS, respectively.

B. Comparison with the Steepest Descent Method

TE3 is carried out using GDM-S. As the simplest minimization algorithm, GDM-S has been proved to be particularly useful for the large dimension problems. The performance of the steepest descent method will be examined to compare to L-BFGS with scaling. The variation of the normalized cost function, that of the inversion error, and that of the inversion results are plotted in Figure II and Figure III.

As shown in Figure II, L-BFGS with scaling ensures a quick convergence to approach its optimal solution at about the 25th iteration step, and then the values of normalized cost function weakly decrease with fluctuation, as the iteration goes on. Different from L-BFGS, GDM-S makes the normalized cost function to decline consistently on the whole. In spite of the advantages at the beginning stage of the iteration, the convergence rate with the steepest descent method enters a slowly decline stage. And this leads to the final values of the cost function and the MAEs between simulated values and "observations" obtained with GDM-S are still slightly smaller than those obtained with the L-BFGS. But the L-BFGS method has a better performance than the other one in terms of the inversion errors of the parameters, and the inverted parameters show a extremely similar pattern with the prescribed one. The results demonstrate that L-BFGS really uses much fewer steps to reach a satisfactory result compared with GDM-S, while the values of the minimized cost function and inverted parameters are almost the same when convergence criterion is met.

In conclusion, the performances of L-BFGS and GDM-S are both satisfactory in terms of the inversion results. Each of the two methods has distinct advantages and disadvantages in the process of the optimization procedure. L-BFGS method occupies less computer storage and converges fast, but requires detailed operation, especially when different kinds of parameters are intended to optimize. Despite of a slow speed of convergence, the small magnitude of SS has no effect on the convergence process with GDM-S, which means GDM-S indeed has great applicability in large-scale optimization and should be given more attention as an important choice
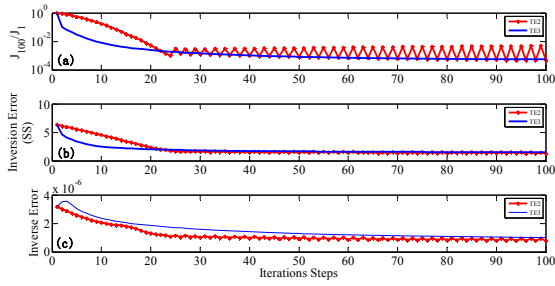
FIGURE II. (A) THE VALUES OF THE NORMALIZED COST FUNCTION IN TE2 ANDTE3, (B) THE MAES BETWEEN PRESCRIBED AND INVERTED IC TE2 ANDTE3, (C) THE MAES BETWEEN PRESCRIBED AND INVERTED SS TE2 ANDTE3.
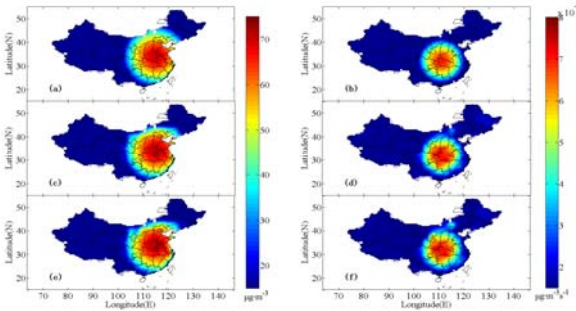

FIGURE III. (A) THE PRESCRIBED IC, (B) THE PRESCRIBED SS, (C) THE INVERTED IC IN TE2, (D) THE INVERTED SS IN TE2, (E) THE INVERTED IC IN TE3 AND (F) THE INVERTED SS IN TE3.

III. PRACTICAL EXPERIMENTS

The practical experiment and data assimilation are carried out for a period of one week from 5 November to 11 November 2014. This period of time was chosen because the time frame coincides with the date of the 21th APEC summit taking place in Beijing and the heating period in northeast China. In this study, the daily (24-h) $PM_{2.5}$ concentrations at the AQS sites were calculated from midnight to midnight local time of the next day on the basis of hourly $PM_{2.5}$ observations following the protocol of the IMPROVE network.

According to the conclusion of Section II，in the practical experiment, IC and SS are inverted simultaneously. The model-produced $PM_{2.5}$ concentrations are researched as the main target. L-BFGS with scaling and GDM-S are used in PE1 and PE2, respectively, aiming to make a comparison of the performance. The most important difference from TE is that both the position and the values of observations are used in this section. The results at the 100th step are analyzed.

From Table 2, the value of the normalized cost function has been decreased by 2 orders of magnitude, roughly the same in PE1 and PE2 after assimilation. The MAEs between model-produced and observations in "assimilated cities" in PE2 are in close proximity to that of PE1, but slightly smaller.

TABLE II. ERROR STATISTICS OF PRACTICAL EXPERIMENTS BEFORE AND AFTER ASSIMILATION

Experiment	J_{100}/J_1	K_1 (µg/m³)		K_2 (µg/m³)	
		Before	After	Before	After
PE1	0.09282	50.50	14.11	57.26	23.18
PE2	0.08558	50.50	13.59	57.26	23.55

J_{100} is the final value of cost function and J1 is the initial value of cost function. K_1 is the MAE between simulated values and observations in "assimilated cities". K_2 is the MAE between the observations and simulated results in all the cities at the initial time

As shown in Figure IV, it's similar in forms of the simulations between L-BFGS and GDM-S. And they are both satisfactory that the simulations follow the variation trend of the observations. Although the model simulations after data assimilation are significantly approved, model still misses some features of observations. In detail, the best simulations come from Fuzhou, which is the "assimilated city" that the observations are used in the whole assimilation procedure. Obviously, it is persuasive to evaluate the simulation results with observations that were not used in the assimilation. As the major concern, the simulations in Beijing are almost the same with the observations, except underestimation in the first few hours. And the model overestimated the observations of Quanzhou and Jinan in the middle and final period, respectively. This may be caused by the ill-posedness of the inverse model. One can draw out the conclusion that the inverse modeling catches the variation characters in the assimilation time whether L-BFGS or GDM-S is used.

Spatially, the mean values of $PM_{2.5}$ concentrations during assimilation window across China are employed for the evaluation. As illustrated in Figure V, the mean values are smaller in southeast China where precipitation is rich. Also, the values are slightly large in middle and east coast of China, which is in conformity with industrial development. The values of northeast China is significantly large that the mean field of $PM_{2.5}$ concentrations are about 150-180 µg/m³. As mentioned above, the assimilation window exactly coincides with heating period in Northeast China. The domestic use of coal and biomass as a household fuel can also cause serious indoor air quality problems in poorly ventilated dwellings [9]. Therefore, $PM_{2.5}$ concentrations present a typical seasonal variation with high values in wintertime of Northeast China. Additionally, Beijing and its peripheral areas deserve special attention. As the capital city of China, Beijing is confronted with the serious problems of air quality raised by the emission of increased energy consumption and vehicle quantity. The amount of vehicle has exceeded 5,317, 000 in 2015 in Beijing. And during heating period (from November to March), $PM_{2.5}$ concentrations were 120-140 µg/m³, which were about 50% and 46% higher than those of non-heating period [10]. As shown in Figure IV, the mean $PM_{2.5}$ concentrations of Beijing are less than 75 µg/m³ from Nov. 5 to Nov. 11, which is significant lower than that in the same period of previous years. This gives the credit to the series of control measures, construction supervision and vehicle restrictions, which are implemented to improve the air quality in Beijing and host 21th APEC meeting successfully. As the results shown, the simulated $PM_{2.5}$ concentrations are reasonable and believable in this model.
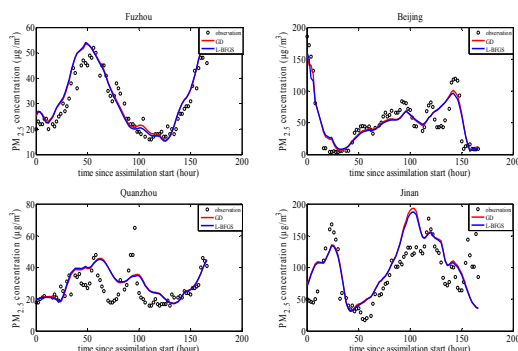
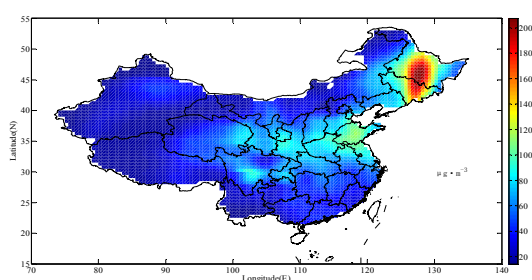FIGURE IV. THE TIME-VARYING PM$_{2.5}$ CONCENTRATIONS OF FOUR REPRESENTATIVE CITIES.

FIGURE V. THE MEAN FIELD OF PM2.5 CONCENTRATIONS.

IV. SUMMARY AND CONCLUSIONS

In this paper, a PM$_{2.5}$ transport adjoint model is applied to the comparison of the performance of L-BFGS and GDM-S. IC and SS are estimated by assimilating the observations with the adjoint method. The normalized cost function, the inversion error and the inversion results are chosen as the basis for the evaluation.

In twin experiment, the addition of a scaling to L-BFGS effectively improve the situation that the minimization process hardly converge due to the magnitude of SS which is much smaller than unity. From the comparison of the inversion errors and the values of the normalized cost function, L-BFGS with scaling has a faster convergent rate than GDM-S. However, the final values of the normalized cost function and inverted results are almost the same.

In the practical experiment, mean values of the PM$_{2.5}$ concentrations coincide with the observations fairly well during the assimilation window, which shows that adjoint inverse modeling with appropriate optimization method is a powerful tool for getting insight into constraining various underlying inputs for PM$_{2.5}$. The conclusions that can be drawn from inverse modeling of this region is limited for the scarcity of available and applicable observations. Nevertheless, the results from inverse modeling contain important information (in particular about the emissions inventory) that can help better understand the PM$_{2.5}$ distributions in deed. Furthermore, it is well worth investigating the advantage of some other simplified optimization algorithms. And the parameter sensitivity of the model will be investigated in the future works.

REFERENCES

[1] Henze D K, Seinfeld J H, and Shindell D T, "Inverse modeling and mapping US air quality influences of inorganic PM2.5 precursor emissions using the adjoint of GEOS-Chem," Atmospheric Chemistry and Physics, vol.9, no.16, pp. 5877-5903, 2009

[2] Zhang L, Liu L, and Zhao Y, "Source attribution of PM2.5 pollution over North China using the adjoint method," AGU Fall Meeting Abstracts, 2014, 1: 04.

[3] Capps S, Henze D K, and Russell A G, "Quantifying relative contributions of global emissions to PM2.5 air quality attainment in the US," AGU Fall Meeting Abstracts, 2011, 1: 07.

[4] D. S. Wang, N. Li, Y. L. Shen and X. Q. Lv, "The parameters estimation for a PM2.5 transport model with the adjoint method," Advances in Meteorology, in press.

[5] D.C. Liu, and J. Nocedal, "On the limited memory method for large scale optimization," Math Programming ,vol.45, no.3, pp. 503‑528, 1989.

[6] A. K. Alekseev, I. M. Navon, and J. L. Steward, "Comparison of advanced large-scale minimization algorithms for the solution of inverse ill-posed problems," Optimization Methods & Software, vol. 24, no. 1, pp. 63‑87, 2009.

[7] Z. Yang, and J.M. Hamrick, "Optimal control of salinity boundary condition in a tidal model using a variational inverse method," Estuarine Coastal & Shelf Science , vol.62, pp. 13‑24, 2005.

[8] Navon, I. M., X. Zou, M. Berger, P. K. H. Phua, T. Schlick, and F. X. Le Dimet, "Testing for reliability and robustness of optimization codes for large scale optimization problems," Optimization Techniques and Applications, vol.1, pp. 445‑ 480, 1992.

[9] Kebin Hea, Fumo Yanga, Yongliang Maa, Qiang Zhanga, Xiaohong Yaob, Chak K. Chanb, Steven Cadlec, Tai Chanc, and Patricia Mulawa, "The characteristics of PM2.5 in Beijing, China," Atmospheric Environment, vol. 35, no. 29, pp. 4959‑4970, 2001.

[10] F.K. Duana, K.B. He,T, Y.L. Ma, F.M. Yang, X.C.Yu, S.H. Cadle, T. Chan, and P.A. Mulawa,"Concentration and chemical characteristics of PM2.5 in Beijing, China: 2001–2002," Science of the Total Environment, vol. 355, no. 1-3, pp. 264–275, 2006.