Improving GEOS-Chem Model Tropospheric Ozone through Assimilation of Pseudo Tropospheric Emission Spectrometer Profile Retrievals

Kumaresh Singh[1], Paul Eller[1], Adrian Sandu[1], Kevin Bowman[3], Dylan Jones[2], and Meemong Lee[3]

[1] Department of Computer Science, Virginia Tech, Blacksburg, VA 24060
{kumaresh,peller,sandu}@cs.vt.edu
[2] Department of Physics, University of Toronto, Toronto, ON, Canada
dbj@atmosp.physics.utoronto.ca
[3] NASA Jet Propulsion Laboratory, Padadena, CA 91109, USA
{kevin.bowman,Meemong.Lee}@jpl.nasa.gov

Abstract. 4D-variational or adjoint-based data assimilation provides a powerful means for integrating observations with models to estimate an optimal atmospheric state and to characterize the sensitivity of that state to the processes controlling it. In this paper we present the improvement of 2006 summer time distribution of global tropospheric ozone through assimilation of pseudo profile retrievals from the Tropospheric Emission Spectrometer (TES) into the GEOS-Chem global chemical transport model based on a recently-developed adjoint model of GEOS-Chem v7. We are the first to construct an adjoint of the linearized ozone parameterization (linoz) scheme that can be of very high importance in quantifying the amount of tropospheric ozone due to upper boundary exchanges. Tests conducted at various geographical levels show that the mismatch between adjoint values and their finite difference approximations could be up to 87% if linoz module adjoint is not used, leading to a divergence in the quasi-Newton approximation algorithm (L-BFGS) during data assimilation. We also present performance improvements in this adjoint model in terms of memory usage and speed. With the parallelization of each science process adjoint subroutine and sub-optimal combination of checkpoints and recalculations, the improved adjoint model is as efficient as the forward GEOS-Chem model.

Keywords: Global chemistry and transport, Adjoints, Inverse modeling, Data assimilation, Linoz scheme.

1 Introduction

GEOS–Chem is a global 3-D model of atmospheric composition driven by assimilated meteorological observations from the Goddard Earth Observing System (GEOS) of the NASA Global Modeling and Assimilation Office. It is applied by research groups around the world to study a wide range of atmospheric composition problems such as assessing intercontinental transport of pollution, evaluating consequences of regulations and climate change on air quality, comparison of model estimates to

satellite observations and field measurements, and fundamental investigations of tropospheric chemistry.

Adjoint models are powerful tools widely used in meteorology and oceanography for applications such as data assimilation, model tuning, sensitivity analysis, and the determination of singular vectors. The adjoint model computes the gradient of a response function with respect to control variables. Generation of adjoint code may be seen as the special case of differentiation of algorithms in reverse mode, where the dependent function is a scalar. Developing a complete adjoint of global atmospheric models involves rigorous work of constructing and testing adjoints of each of the complex science processes individually, and integrating those into a consistent adjoint model. It is well accepted that adjoint construction is an extremely challenging task.

Original work on the adjoint of GEOS-Chem began in 2003, focusing on the adjoint of the offline aerosol simulation. By 2005, the adjoint was expanded to include a tagged CO simulation and a full chemistry simulation as well as observational operators for MOPITT (CO) and IMPROVE network (aerosols). Henze et. al(2007) [1] discusses the construction of adjoint for GEOS-3 v6 of GEOS-Chem. We implemented a standard adjoint model for GEOS-4 v7 of GEOS-Chem and made it more user friendly, with the wider goal of making this adjoint publicly available as part of the standard GEOS-Chem code. This will provide the community of GEOS-Chem users with the tools needed for performing sensitivity analysis and data assimilation.

KPP chemistry was first interfaced with GEOS-Chem and its adjoint in Henze et al. (2007), see Appendices therein. We improved upon this implementation in terms of automation, performance, benchmarking, and documentation. GEOS-4 convection and advection processes are based on entirely different algorithms as compared to GEOS-3. Convection and wet-deposition adjoints are discrete adjoints and were constructed in a hybrid fashion using the automatic differentiation software TAMC [2] together with manual coding. Advection adjoint on the other hand is continuous and was obtained by calling the forward subroutine with reverse wind fields. We also provided a way of calculating the scaled emission and dry-deposition adjoints by modifying the KPP chemistry integrator. All these pieces are tested extensively and integrated together to build the full adjoint GEOS-Chem model.

In this paper we present 4-D variational data assimilation [3] results of forecast improvements through pseudo profile retrievals from Tropospheric Emission Spectrometer (TES, JPL NASA) [4]. The designed interface to run 4-D Var data assimilation using GEOS-Chem allows users to tweak all the limited memory BFGS (L-BFGS) method [5] and GEOS-Chem related parameters through a single driver file in addition a list of plug-n-play response function calculation subroutines including TES observation operator based forcing. We also present the first validation results of the adjoint of linearized ozone chemistry (linoz) scheme [6]. The construction of linoz scheme adjoint is a major step towards computing the sensitivity of tropospheric ozone with respect to upper boundary layer exchanges. The standard version of GEOS-Chem adjoint is completely parallel. The continuous advection adjoint inherits the parallelism of the forward subroutine. For rest of the science process adjoints, we have implemented shared memory system based parallel version. Some discrete adjoint calculations require intermediate variable values from forward calculations. In such cases, either these variables are recomputed in the adjoint mode or are written to checkpoint files [7] during the forward calculation and read in the adjoint mode.

We recalculated variable values in the adjoint mode wherever this had a minimal performance penalty.

This paper is organized as follows. Section 2 discusses the construction of adjoints for various science processes. It also summarizes software engineering aspects of the standard GEOS-Chem adjoint model and provides speedup results from shared memory based parallelization. Sections 3 exhibits validation results of the linoz scheme adjoint (verification of accuracy against finite difference approximations). Section 4 describes the framework for using the adjoint model in scientific applications such as 4-D variational data assimilation and displays improvement results from TES profile retrievals. Section 5 draws conclusions and points to future work.

2   Standard Adjoint Model of GEOS-Chem

The mathematical formulation for calculating gradients of a model output using the adjoint method can be derived from the equations governing the forward model or from the forward model code. The former approach leads to the continuous adjoint, while the latter leads to the discrete adjoint [8]. Continuous adjoint gradients may differ from the actual numerical gradients of cost function J, and continuous adjoint equations (and requisite boundary/initial conditions) for some systems are not always readily derivable; however, solutions to continuous adjoint equations can be more useful for interpreting the significance of the adjoint values. Many previous studies have also described the derivation of discrete adjoints of such systems [9][10]. An advantage of the discrete adjoint model is that the resulting gradients of the numerical cost function are exact, even for nonlinear or iterative algorithms, making them easier to validate. Furthermore, portions of the discrete adjoint code can often be generated directly from the forward code with the aid of automatic differentiation tools.

We have implemented the chemistry simulations in GEOS-Chem using the Kinetic Pre-Processor (KPP) [11]. KPP provides a library of several chemical solvers together with their tangent linear and adjoint integrators in addition to the native SMVGEARII solver [12][13]. It generates chemistry adjoint files in a similar fashion as it generates the forward chemistry. The generated files are interfaced with the GEOS-Chem adjoint code, updating the KPP global variables, parameters and initialization files. The adjoint of GEOS-4 advection subroutine is continuous and is derived by calling the forward subroutine with reversed wind fields. The GEOS-4 convection, planetary boundary mixing (pbl_mix) and wet deposition adjoints are discrete and have been constructed using the tangent linear and adjoint model complier (TAMC). Intermediate parameters in the forward run are checkpointed every dynamic time step and are read in during the adjoint run. The linoz scheme is linear and is self adjoint. In GEOS-Chem emission and dry deposition are handled through chemistry via fake equations. The rates for these processes are calculated separately and then attached to the chemistry reaction rates. The adjoints of these subroutines are scaled and are calculated using the adjoint integrator. The adjoint integrator provides adjoints with respect to the rates which are then multiplied with the individual rates and accumulated over time.

A detailed GEOS-Chem adjoint function call flow is presented in Figure 1. It provides a visualization of the order in which individual science processes are called in the forward and adjoint mode, and the way checkpoint files are written and read.
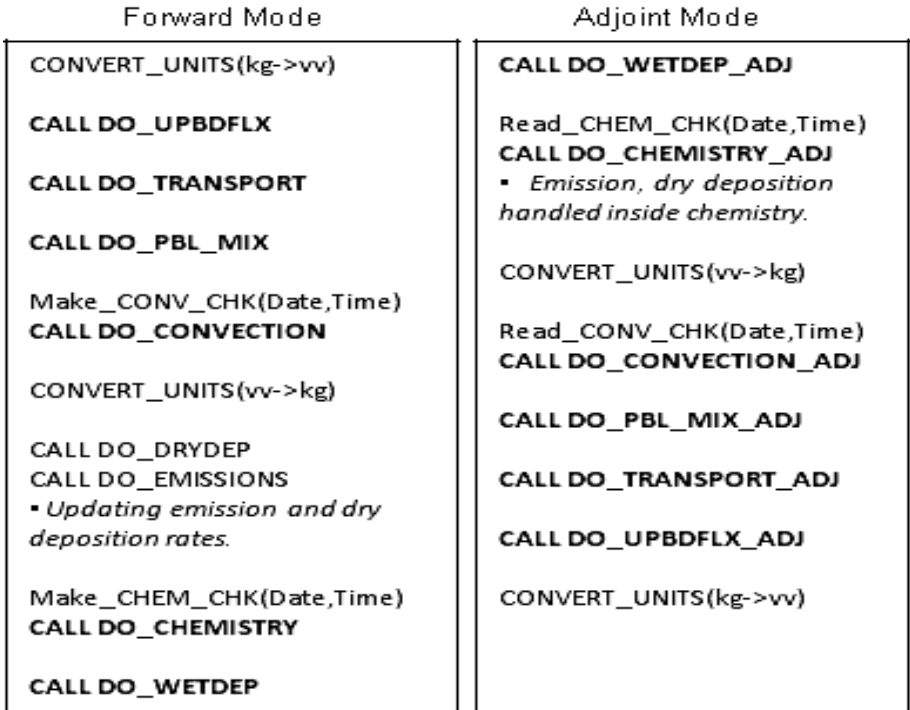
Forward Mode	Adjoint Mode
CONVERT_UNITS(kg->vv)	**CALL DO_WETDEP_ADJ**
CALL DO_UPBDFLX	Read_CHEM_CHK(Date,Time) **CALL DO_CHEMISTRY_ADJ** ▪ *Emission, dry deposition handled inside chemistry.*
CALL DO_TRANSPORT	
CALL DO_PBL_MIX	CONVERT_UNITS(vv->kg)
Make_CONV_CHK(Date,Time) **CALL DO_CONVECTION**	Read_CONV_CHK(Date,Time) **CALL DO_CONVECTION_ADJ**
CONVERT_UNITS(vv->kg)	**CALL DO_PBL_MIX_ADJ**
CALL DO_DRYDEP CALL DO_EMISSIONS ▪ *Updating emission and dry deposition rates.*	**CALL DO_TRANSPORT_ADJ** **CALL DO_UPBDFLX_ADJ**
Make_CHEM_CHK(Date,Time) **CALL DO_CHEMISTRY**	CONVERT_UNITS(kg->vv)
CALL DO_WETDEP	

Fig. 1. GEOS-Chem forward and adjoint function call flows. The adjoint of science processes are called in reverse order in the adjoint mode. CONVERT_UNITS() converts unit for tracer variable in the forward mode, and adjoint variable in the adjoint mode. Make_*_CHK() are subroutines to make checkpoint files and Read_*_CHK() to read the checkpoint files as per input date and time.

2.1   Code Structure Overview

The newly developed GEOS-Chem adjoint model is well structured and follows the coding style provided in the GEOS-Chem users' manual [14]. For one science process, all the forward and related adjoint subroutines are kept in the same module file for the ease of users to look into only one file. To handle checkpointing, the additional module file CHECKPOINT_MOD.F has been provided. In addition, various subroutines to perform observation and background cost function calculations, define adjoint variables, include satellite observations are provided through separate files.

The standard GEOS-Chem adjoint package (GCv7_ADJ) is available for download from our project website [15]. Users have been provided with five modes of application built on top of the full adjoint model. The source files for each of these modes are in separate directories inside the main code directory. FWD_SMV mode is the forward

GEOS-Chem code that is available from the Harvard's website. This mode uses the SMVGEAR integrator for chemistry calculations. FWD_KPP is equivalent to FWD_SMV except it uses KPP for chemistry, providing users a suite of fast and highly accurate integrators to choose from. ADJ_FD is the finite difference testing module which users can choose to validate newly built adjoint subroutines. ADJ_SENST and ADJ_4DVAR are selected to perform sensitivity analysis and 4-D variational data assimilation respectively. Users can choose one of these options by simply (un)commenting the mode option in the compilation script (v7-04-10.cmp).

4-D variational data assimilation requires multiple iterations of forward and backward model runs before it converges to a suitable initial condition. Users have been provided the option to stop at any iteration and restart from the same point at a later time. The number of iterations to run is handled via the same run script (v7-04-10.run) which is used to initiate the geos executable.

2.2 Speedup Results: Shared Memory System Based Parallelization

Harvard's GEOS-Chem code is programmed parallel for shared memory systems. One of the challenges in developing the adjoint model was to parallelize this model completely. For chemistry adjoint we used THREADPRIVATE variables to allow multiple threads to execute the KPP chemistry routines for different grid cells in parallel. Emission and dry deposition adjoints are handled through chemistry. Advection adjoint being continuous leverages forward parallelization. For convection, planetary boundary mixing and wet deposition adjoints, we created OpenMP parallel versions of the subroutines taking care of the thread shared and private variables. The new GEOS-Chem adjoint code is completely parallel and has been tested for consistency against the serial version.

Discrete adjoints of non-linear (in terms of dependent input variable) subroutines require intermediate variable values from forward calculations. In such cases, either these variables are recomputed in the adjoint mode or are written to checkpoint files during the forward calculation and read in the adjoint mode. Both these options add performance penalty to the adjoint calculations. Checkpointing was used whenever recalculation involved lots of parameters and the required intermediate variables were few. The benefit of recalculations however is that these code portions could be parallelized in a similar way as the forward mode.
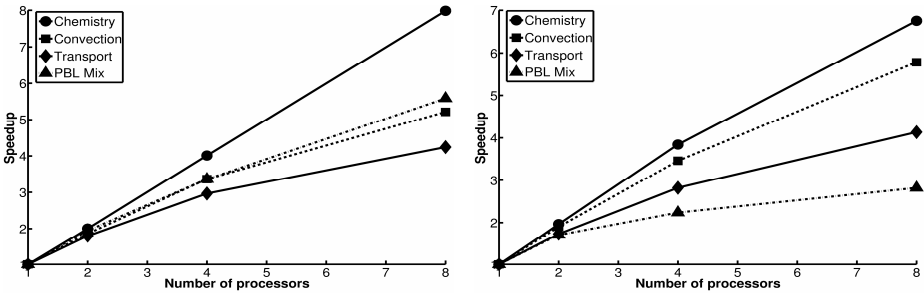

Fig. 2. Speedup graphs for chemistry, convection, advection and planetary boundary mixing subroutines in forward(*left*) and adjoint(*right*) mode on 1, 2, 4 and 8 processors. The simulation window for this analysis was 24 hours performed on July 2001 GEOS-Chem data.

Presented in Figure 2 are the speedup graphs for all the science process subroutines in forward and adjoint mode that are most time consuming. These graphs show that the adjoint mode scales well considering the performance penalties due to checkpointing and recalculations.

3   Linearized Ozone Chemistry Adjoint

The linearized ozone (linoz) scheme [6] is a stratospheric ozone chemistry mechanism for atmospheric models that focus on the troposphere and was developed with primary goals of (1) accurate calculation of the cross tropopause flux, and (2) reasonable representation of the ozone gradients near the tropopause. Earlier, in tropospheric CTMs, stratospheric ozone were handled by specifying climatology-based mixing ratios for ozone in the lower stratosphere and then allowing the transport to determine the cross-tropopause flux. The stratospheric level at which the ozone mixing ratios were specified, as well as other factors affecting the flux, varied from model to model. The fluxes produced by models using variants of this technique ranged from over 1400 Teragram(Tg)/yr, to less than 400 Tg/yr. Such discrepancies were important to resolve since they gave almost opposite interpretations of the role of tropospheric photochemistry: the large fluxes resulted in net photochemical loss of ozone throughout most of the troposphere, whereas the much smaller fluxes required a net production in order to balance with near-surface losses.

The developed linear model for ozone (Linoz) chemistry was simple and computationally efficient. In this method the ozone chemical tendency is expressed as a linear function of ozone, temperature, and the overhead ozone column. The linearizations are performed about an observed climatological state for a standard set of latitudes, months, and altitudes. The 24-hour, zonal-mean ozone photochemical tendency is calculated at each time step for each stratospheric grid box from these monthly varying coefficients. The change in ozone with time due to local chemistry is given by,

$$d\mathrm{f}/d\mathrm{t} = (\text{P-L})[\ \mathrm{f},\ \mathrm{T},\ \mathrm{C}_{\mathrm{O}_3}]\ . \tag{1}$$

where (P - L) represents the ozone tendency (in units of ppmv/s), the square brackets denote a functional dependence, f is the ozone mixing ratio, T is temperature, and $\mathrm{C}_{\mathrm{O}_3}$ is the column ozone above the point under consideration.

3.1   Linoz Adjoint Validation Results

The linoz scheme is linear and it is self adjoint, i.e. the adjoint variable follows the same dynamics as the forward tracer concentrations. Presented below in Figure 3 are the plots of mismatch between the adjoint values and their finite difference approximations with and without using linoz adjoint subroutine. These are generated using July 2006 summertime GEOS-Chem data run over a 1 day period at various geographical levels. The results show clearly that the consistency could be compromised by as large as 87% if the linoz adjoint subroutine is not used.
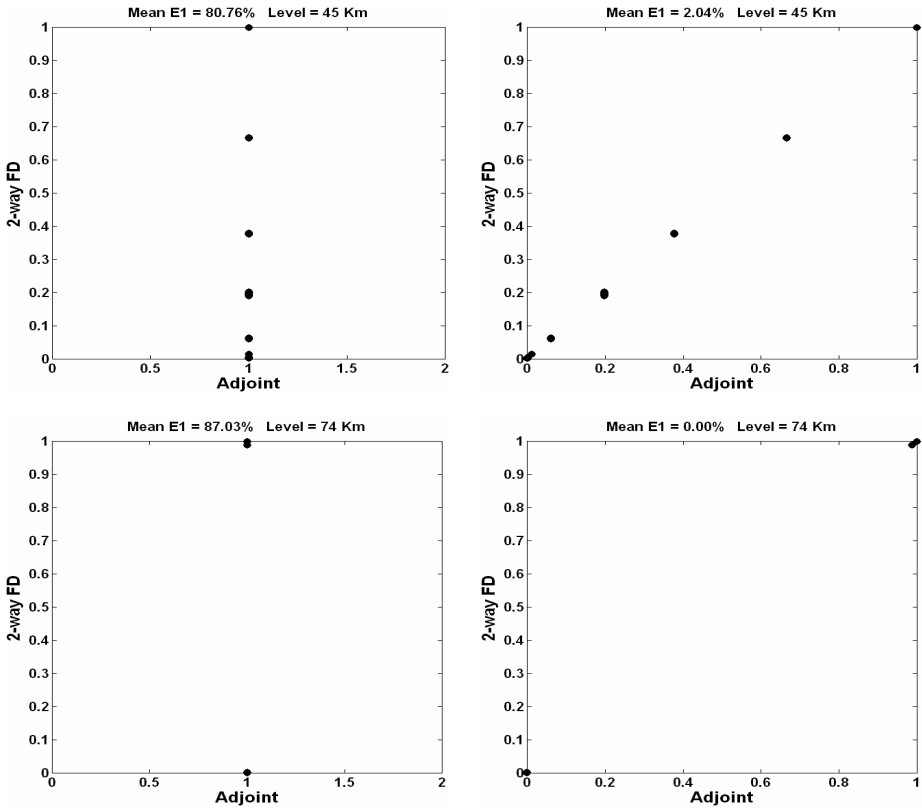
Fig. 3. Validation results of adjoint values with respect to their finite-difference approximations with (*right*) and without (*left*) using linoz adjoint subroutine for levels 45 Km (*top*) and 74 Km (*bottom*) for 2006 summertime GEOS-Chem data, 1 day simulation

4   Data Assimilation through TES Profile Retrievals

4D-Var data assimilation allows the optimal combination of three sources of information: an a priori (background) estimate of the state of the atmosphere, knowledge about the physical and chemical processes that govern the evolution of pollutant fields as captured in the chemistry transport model GEOS-Chem, and observations of some of the state variables.

The adjoint method has been used widely for variational data assimilations and inverse modeling. In these applications, a cost function is defined as

$$J = J_{Observations} + J_{Background}$$
$$= (y - F(x))^T S_e^{-1} (y - F(x)) + (E - E^b)^T B^{-1} (E - E^b) . \qquad (2)$$

where E is the state vector at the initial time, E^b is an *a priori* (background or initial guess) estimate of inputs, and B is background error covariance matrix, $y \in R^m$ is the observation vector whose vectors are retrievals of Ozone, m is the total number of

observations used in the inversion analysis, $S_e \in R^{m \times m}$ is the block diagonal error co-variance matrix of the observation vector, $x \in R^n$ is the state vector whose elements are the strengths of O_3 and $F(x) \in R^m$ is the TES observation operator applied to the state vector according the following equation,

$$F(x) = y_a + A_{yy} (\ln[H(x)] - y_a) . \tag{3}$$

where $H : R^n \to R^m$ represents the interpolation operator to convert GEOS-Chem vector into TES observation vector, A_{yy} is the averaging kernel and y_a is the *a priori* TES observation vector.

The cost function consists of two parts: the first part is a measure of model prediction errors, and the second part is a penalty for deviation from a priori estimates of model inputs. In typical applications the adjoint method is used to calculate the gradient of the cost function with respect to initial concentrations (for data assimilation applications) or model parameters such as emissions (for inverse modeling applications). This gradient is then used in an iterative optimization algorithm in order to minimize the cost function, reducing the mismatch of model predictions and observations by adjusting the inputs (e.g. emissions) within a reasonable range. Variational methods provide an important approach for constraining emissions of various species on a spatially resolved basis.

In order to carry out data assimilation using GEOS-Chem, an interface has been developed to call the optimization routine L-BFGS [16] through a driver file named 4dvar_driver.f. The framework provides users the ability to perform 4-D variational data assimilation with respect to both tracer and emission species. Below we present details of an experiment based on synthetic observations (*twin experiment framework*).

Framework Details. An observation grid is defined over the computational domain and synthetic observations are produced by performing a forward GEOS-Chem run on the initial reference concentration field c_0^0. A perturbation is then added to c_0^0 to get a perturbed initial condition c_p^0 (which is considered the background state)

$$c_p^0 = c_0^0 + \text{perturbation}$$

This perturbed concentration is then transferred to the optimization subroutine in order to obtain the best estimate c_{op}^0 of the original concentration c_0^0 after several iterations.

At iteration 0, $\qquad\qquad\qquad x_0 = c_p^0$

At each subsequent iteration k (k≥1),

$$x_{k+1} \leftarrow \text{L-BFGS} (x_k, f, g)$$
$$c_{op}^0 \leftarrow x_{k+1}$$
$$(f, g) \leftarrow \text{reverse\_mode} (c_{op}^0, \text{Observation\_Chk})$$

where f is the cost function and g is the gradient of the cost function.

4.1 Results

To validate the developed data assimilation test-bed, we conducted a twin experiment using ozone concentrations at 00:00 hrs on July 2006 as the initial condition with TES profiles generated from observations as twice the initial condition. Figure 4 represents

the correction in the initial condition after 17 iterations of the optimization routine, verified by the 32% decrease in the cost function within 9 model runs. (Intermediate iterations are of negligible cost as compared to the model runs that heavily dominate the cost of running l-BFGS optimization routine). The plots validate that the designed data assimilation framework works nicely.
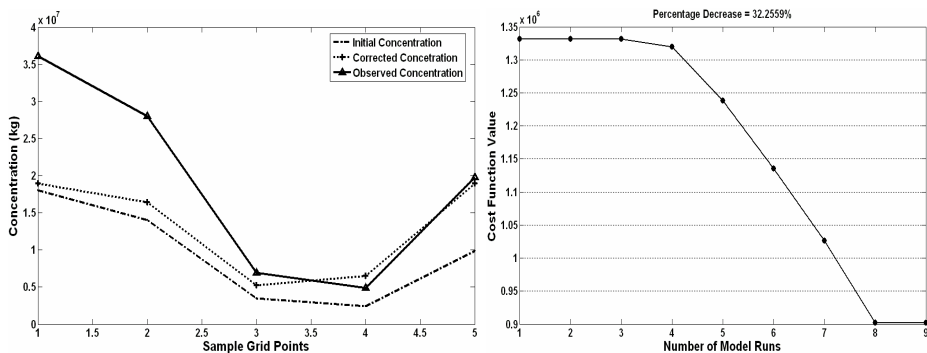

Fig. 4. 4-D variational data assimilation results for twin experiment run over 2006 summertime GEOS-Chem data for 3 days with TES profile retrievals generated synthetically. Plots include correction in the initial concentration (*left*) and decrease in the cost function with respect to model runs (*right*).

5   Conclusion

We have improved upon the standard adjoint model of GEOS-Chem for faster, more accurate and ready for real data experiments. The adjoint code has been fully parallelized and is as efficient as the forward mode considering the performance penalties added due to checkpointing and intermediate variable recalculations in the adjoint mode. It will allow world-wide community of GEOS-Chem users to perform powerful analytic and scientific applications such as sensitivity analysis and 4-D variational data assimilation in real time.

An adjoint of linoz chemistry will help scientists worldwide to better quantify the tropospheric ozone coming from upper boundary layer exchanges. We have also successfully implemented the TES observation operator based cost function calculations, ready to perform 4-D variational data assimilation and sensitivity calculations with real data. These improvements could add ample scientific value to the way we analyze physical and chemical dynamics of the earth's atmosphere.

References

1. Henze, D.K., Hakami, A., Seinfeld, J.H.: Development of the Adjoint of GEOS Chem. Atmos. Chem. Phys. 7, 2413–2433 (2007)
2. Giering, R., Kaminski, T.: Recipes for Adjoint code Construction. ACM Trans. Math. Softw. 24, 437–474 (1998)

3. Elbern, H., Schmidt, H.: A four-dimensional variational chemistry data assimilations scheme for Eulerian chemistry transport modeling. J. Geophys. Res. 104, 18583–18598 (1999)
4. Tropospheric Emission Spectrometer (TES), JPL NASA,
   `http://tes.jpl.nasa.gov/`
5. Byrd, R.H., Lu, P., Nocedal, J., Zhu, C.: A limited memory algorithm for bound constrained optimization. Scientific Computing 16, 1190–1208 (1995)
6. McLinden, C.A., Olsen, S.C., Hannegan, B., Wild, O., Prather, M.J., Sundet, J.: Stratospheric ozone in 3-D models: A simple chemistry and the cross-tropopause flux. J. Geophys. Res. 105(D11), 14653–14666 (2000)
7. Griewank, A., Walther, A.: Algorithm 799: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation. ACM Trans. Math. Softw. 26, 19–45 (2000)
8. Giles, M., Pierce, N.: An introduction to the adjoint approach to design. Flow, Turbulence and Control 65, 393–415 (2000)
9. Sandu, A., Daescu, D., Carmichael, G.R., Chai, T.: Adjoint sensitivity analysis of regional air quality models. J. Comput. Phys. 204, 222–252 (2005a)
10. Muller, J.F., Stavrakou, T.: Inversion of CO and NOx emissions using the adjoint of the IMAGES model. Atmos. Chem. Phys. 5, 1157–1186 (2005)
11. Sandu, A., Daescu, D.N., Carmichael, G.R.: Direct and adjoint sensitivity analysis of chemical kinetic systems with KPP: Part I – theory and software tools. Atmos. Environ. 37, 5083–5096 (2003)
12. Jacobson, M.Z., Turco, R.: SMVGEAR: A Sparse-Matrix, Vectorized Gear Code For Atmospheric Models. Atmos. Environ. 28, 273–284 (1994)
13. Jacobson, M.Z.: Technical Note: Improvement of SMVGEAR II on Vector and Scalar Machines through Absolute Error Tolerance Control. Atmos. Environ. 32, 791–796 (1998)
14. GEOS-Chem users manual,
    `http://www-as.harvard.edu/chemistry/trop/geos/doc/man`
15. GEOS-Chem adjoint project webpage,
    `http://people.cs.vt.edu/~asandu/Public/GCv7_ADJ`
16. Zhu, C., Byrd, R.H., Lu, P., Nocedal, J.: L-BFGS-B: a limited memory FORTRAN code for solving bound constrained optimization problems. Tech. rep., Northwestern University (1994)