An iterative four-dimensional data assimilation method for photochemical air quality modeling based on inverse modeling and direct sensitivity analysis

Alberto Mendoza-Dominguez* and Armistead G. Russell
School of Civil and Environmental Engineering, Georgia Institute of Technology, 200 Bobby Dodd Way, Atlanta, GA 30332-0512

* Corresponding author; phone: 404-385-0570, fax: 404-894-8266, e-mail: albert@themis.ce.gatech.edu.

1. Details on the coupling of DDM-3D with the CIT Photochemical Airshed Model

The CIT model, as other Eulerian models, simulates the formation and transport of photochemical air pollution by solving the atmospheric diffusion equation (I):

$$\frac{\partial c_i}{\partial t} = - \nabla \cdot (u c_i) + \nabla \cdot (K \nabla c_i) + R_i(c, ..., c_n; T, t) + S_i \quad i = 1, ..., I$$  \hspace{1cm} (S1)

where $c_i(x,t)$ is the ensemble average concentration of species $i$, $K(x,t)$ is the second-order turbulent diffusion tensor, $u(x,t)$ is the advective field, $R_i$ is the gas phase reaction rate of species $i$, $S_i$ is the elevated point-source rate of emissions of species $i$, $T(x,t)$ is the temperature, $t$ is time, and $x$ is the position vector. The ground-level emissions are entered in the formulation in the surface boundary condition:
\[ v'_k c_i - K_{zz} \frac{\partial c_i}{\partial z} = E_i \]  

where \( v'_k \) is the dry deposition velocity and \( E_i \) is the ground-level emission rate. The complete formulation can be found elsewhere (I-3). The CIT model, extended with DDM-3D (4), can calculate local sensitivity coefficients of model outputs to model parameters and inputs. The sensitivity coefficients are computed through the derivative of the model output (i.e. species concentration, \( c_i \)) with respect to a semi-normalized model input or parameter (i.e. \( \varepsilon_j = p_j/p_{\rho} \)), where \( \varepsilon_j \) is a scaling variable with a nominal value of one; \( p_j \) is the scaled model parameter [e.g. rate constant] or model input [e.g. area source emissions]; and \( p_{\rho} \) is the nominal value of the model parameter or model input). This results in an auxiliary equation that is very similar to the atmospheric diffusion equation (4):

\[ \frac{\partial s_{ij}}{\partial t} = -\nabla \cdot (u s_{ij}) + \nabla \cdot (K \nabla s_{ij}) + J_{ik} s_{kj} + \frac{\partial R_i}{\partial \varepsilon_j} + \frac{\partial S_i}{\partial \varepsilon_j} - \nabla \cdot (u c_i) s_{ij} + \nabla \cdot (K \nabla c_i) s_{ij} \]  

where \( s_{ij} \) is the semi-normalized sensitivity of species \( i \) to parameter \( j \), \( J \) is the Jacobian matrix \( (J_{ik} = \partial R_j/\partial c_k) \). The last two terms are production due to wind speed \( (u) \) and diffusivity \( (K) \), and exist only if sensitivity to wind speed and diffusivity are applied. Of interest is the auxiliary equation for the surface boundary condition since it includes the term for production due to ground-level emissions. The auxiliary equation for the surface boundary condition is (4):

\[ v'_k s_{ij} - K_{zz} \frac{\partial s_{ij}}{\partial z} = -v'_k c_i s_{ij} + K_{zz} \frac{\partial c_i}{\partial z} s_{ij} + E_i s_{ij} \]  

The right-hand side terms represent production due to dry deposition and ground-level emissions, and exist only if sensitivity to these terms is applied.
2. Definition of weighting factors for the observations

The linear system of interest is given by:

\[ m = \left( G^T W_e G + W_m \right)^{-1} G^T W_e d \]  

(S5)

Here we give further details on the form of the elements in the \( W_e \) and \( G \) matrices.

Each element of the diagonal matrix \( W_e \) (\( \omega_{ik} \)) is computed as:

\[ \omega_{ik} = \omega_{ik}^{<N>} \cdot \omega_{ik}^{<>} \cdot \omega_{ik}^{<co>} \cdot \omega_{ik}^{<yp>} \cdot \omega_{ik}^{>l} \]  

(S6)

We define \( \omega_{ik}^{<N>} \) as:

\[ \omega_{ik}^{<N>} = \frac{1}{N_i} \]  

(S7)

where \( N_i \) is the total (valid) number of observations for species \( i \). This allows each species to have equal weight in the solution. The \( \omega_{ik}^{<>} \) factor weights according to the inverse variance of the measurements (5):

\[ \omega_{ik}^{<>} = \frac{1}{\sigma_{ik}^{<co>}} \]  

(S8)

Thus, the prediction error of the more accurate observations will have a greater weight in the minimization process than the inaccurate observations. The \( \omega_{ik}^{<co>} \) factor weights the difference between observations and predictions such that larger differences will have larger weight:

\[ \omega_{ik}^{<co>} = \left| \frac{O_{ik} - \bar{P}_{ik}^{<base>}}{\bar{O}_{ik}} \right| \]  

(S9)

The \( \omega_{ik}^{<yp>} \) factor is derived such that a observation of species \( i \) at one location and time is weighted according to the average observed concentration domain-wide, of the same species, at the same time. With the \( \omega_{ik}^{<>} \) factor, the observation is weighted according to the average of the observations at the same location over the whole temporal window of assimilation. The last two
factors give more weight to large values in the observations, a property that can be useful in accelerating convergence (as investigated here) or when large observed values are of concern more than small values.

A scaling coefficient is added to each sensitivity coefficient ($s_{ikj}$) to give the components of the $G$ matrix (the $g_{ikj}$’s):

$$g_{ikj} = \omega_{ij} <E_{i} > \cdot s_{ikj}$$  \hspace{1cm} (S10)

and,

$$\omega_{ij} <E_{i} > = \frac{1}{\sigma_{g_{ij}}^{<E_{i} >}}$$  \hspace{1cm} (S11)

where $\sigma_{g_{ij}}^{<E_{i} >}$ is the geometric standard deviation of the $i^{th}$ emission source (we assume that the original emission estimates come from a lognormal distribution [6]). The reason for scaling the sensitivities is numerical in nature rather than statistical, even though the scaling factors are computed using a measure of the a priori uncertainty in emissions. What it is intended with this weighting scheme is to over-relax the rate of change of the emission adjustment estimates for those sources that are suspect of being highly uncertain, and thus, likely to change considerably from their base level estimates. Effectively, the approach allows the uncertain sources to change more rapidly in search for stable adjustments.

3. Estimation of the penalty function

Mathematically, the way the penalty function works is to add a (small) positive quantity to the diagonal of the linearized normal equations at each iteration, in order that the parameter correction vector can be confined to a region where the linearization yields “meaningful” approximations to the nonlinear function.
Here we define the procedure to calculate the elements of the diagonal matrix $W_m$ given in equation (S5). First, an iterative procedure is followed to generate a estimate of the ridge parameter, $\lambda$ (7). At the $i^{th}$ step in the iteration procedure, a $\lambda_i$ is estimated using (8):

$$
\lambda_i = \frac{\hat{e}_i^2}{m_{i-1}^T G^T W_e G m_{i-1}}
$$

(S12)

and

$$
\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{\hat{e}_i^2}{(K \cdot I - J)}
$$

where $m_{i-1}$ is calculated from employing the $\lambda_{i-1}$ estimate ($W_m = \lambda_{i-1} I$) in equation (S5). To start the process, $\lambda_i$ is estimated using the $m_o$ from an ordinary least squares regression (i.e., $\lambda_o = 0$). To terminate the iteration process, the following expression is used (7), for $i > 1$:

$$
\frac{(\lambda_i - \lambda_{i-1})}{\lambda_{i-1}} \leq 20 \left( \frac{\text{Tr} \left( (G^T W_e G)^{-1} \right) - I}{J} \right)^{-1.3}
$$

(S13)

where $\text{Tr}[M]$ denotes the trace of the square matrix $M$.

Effectively, the parameter $\lambda$ rotates the solution vector towards the direction of larger sample spread (9). Once the ridge parameter $\lambda$ is estimated, a length parameter, $l$, that expands or shrinks the RR solution vector can be computed if needed. The length parameter modifies the estimates from the RR calculation as (10):

$$
m^{<\text{LMRR}>} = l m^{<\text{RR}>}
$$

(S14)

This method is known as length modified ridge regression (LMRR). The version of length parameter $l$ used here is the $\text{LMRR}_{p_{2}}$ obtained by Aldrin (10). It has to be added that as $G^T W_e G$ becomes better conditioned the usefulness of the ridge parameter (and hence of the length parameter) declines. The method selected here to compute the ridge parameter takes this in
account such that \( \lambda \) tends to zero as the condition number of \( G^T W_e G \) decreases. However, even though \( G^T W_e G \) might be well posed, one might want to constrain the amount the emissions are corrected and the penalty function can still be used.

If the solution given by the computation of purely stochastic parameters (\( \lambda \) and \( I \)) lies outside the bounds of the uncertainty limits, a further penalization can be applied where each diagonal element of \( W_m \), \( w_{jj} \), becomes:

\[
    w_{jj} = w_{jj}^{\text{unc}} \cdot w_{jj}^{\text{bounds}} \cdot w_{jj}^{\text{spread}} + \lambda \quad \text{(S15)}
\]

and

\[
    w_{jj}^{\text{unc}} = \frac{1}{\sigma_{R,j}^2} \ln\left( \frac{\sigma_{R,j}}{\sigma_{R,j}} \right)
\]

\[
    w_{jj}^{\text{bounds}} = \begin{cases} 
        \ln\left( \frac{\ln(\sigma_{R,j}^{<\ell_{j>}}) + \ln(1 + m_{j}^{<p>})}{\ln(\sigma_{R,j}^{<\ell_{j>}})} \right) & \text{if } m_{j}^{<p>} < 0 \\
        \ln\left( \frac{\ln(\sigma_{R,j}^{<\ell_{j>}}) - \ln(1 + m_{j}^{<p>})}{\ln(\sigma_{R,j}^{<\ell_{j>}})} \right) & \text{if } m_{j}^{<p>} > 0 
    \end{cases}
\]

\[
    w_{jj}^{\text{spread}} = \frac{(G^T W_e G + W_m)_{jj} \cdot (G^T W_e G + W_m)_{jj}}{\sum_{i=1}^{j} (G^T W_e G + W_m)_{jj} \cdot (G^T W_e G + W_m)_{ji}}
\]

where \( m_{j}^{<p>} \) is the estimated fractional change from the previous assimilation iteration. The \( w_{jj}^{\text{unc}} \) factor accounts for the uncertainty in the \textit{a priori} emission distribution, \( w_{jj}^{\text{bounds}} \) penalizes heavily estimates that approach the uncertainty limits imposed to the sources, and \( w_{jj}^{\text{spread}} \) takes in account the relative value of the diagonal elements of the \((G^T W_e G + W_m)\) matrix with respect to the off-diagonal terms. If off-diagonal terms dominate the diagonal terms, then the inverse of \((G^T W_e G + W_m)\) will have ill-condition problems. If after applying the new penalties the solution of source \( j \) is still out of the specified bounds, the value of \( w_{jj} \) increases iteratively until a solution inside the uncertainty bounds is obtained.
4. Standardization of the linear system

Before an attempt is made to calculate the penalty function, $W_m$, the usual practice is to standardize the matrix $G^T W_e G$ (11). The standardized form of $G^T W_e G$ resembles a correlation matrix, though the elements are not true correlation coefficients (5). Since the diagonal elements of the standardized version are equal to one, a given value of the penalty function (the $w_{ij}$ components) represents a percentage of perturbation (biasing) in the natural units for the regressors. This allows an easier interpretation of the values computed or assigned to $w_{ij}$. Standardization usually takes place by centering and scaling the regressor variables, though it depends on the specific application if centering is required. Here only scaling is applied.

Let $W_e = W_s^T W_s$, so that equation (S5) becomes:

$$m = \left[ (W_s G)^T W_s G + W_m \right]^{-1} (W_s G)^T W_s d = \left( X^T X + W_m \right)^{-1} X^T y$$  \hspace{1cm} (S16)

Scaling is accomplished by substituting each $x_{ik,j}$ by $x^{<s>}_{ik,j} = x_{ik,j} / s_j$, where $s_j = \sqrt{\sum_{k=1}^{j} x_{ik,j}^2}$. The scaled version of equation (S16) is then:

$$m^{<s>} = \left( X^{<s>}^T X^{<s>} + W_m^{<s>} \right)^{-1} X^{<s>}^T y$$  \hspace{1cm} (S17)

where $m^{<s>}$ is the scaled version of $m$. To recover the fractional changes in the natural units, each $m_{j}^{<s>}$ is multiplied by its corresponding $s_j$.

5. Convergence characteristics of Ridge Regression

It has been noted that the parameter $\lambda$ rotates the solution vector towards the direction of larger sample spread. This reorientation (and shrinkage) is done not because there is any prior belief
that the true regression vector is likely to align with the high spread directions of predictor design but to control the variance of the estimates. However, Marquardt (12) proved that ridge regression can help convergence in iterative least-squares estimation of nonlinear parameters. Effectively, ridge regression applied to the iterative solution of nonlinear problems defaults into a maximum neighborhood method. Marquardt proved that the ridge regression estimate \( \mathbf{m}_r \) is the minimum around the sphere of radius \( ||\mathbf{m}_r|| \) (where \( ||\cdot|| \) is the Euclidean norm), that \( ||\mathbf{m}_r|| \) is a continuous monotone decreasing function of the ridge parameter, and that the angle between \( \mathbf{m}_r \) and \( \mathbf{G}^T\mathbf{W}_e\mathbf{d} \) is a continuous monotone decreasing function of the ridge parameter. These conditions provided convergence, at least for local minima. Graphically, this implies that the resulting correction vectors lie within 90 degrees of the negative gradient of the objective function that minimizes \( e^T\mathbf{W}_e\mathbf{e} \). Actually, in some sense this procedure interpolates between the direction of advance of the Gauss-Newton method (which not infrequently suffers from divergence of the successive iterates) and steepest-descent method (which can suffer of slow convergence). So, as one iterates using the penalty function, there is increasing assurance that one is moving in the direction of the local minima.

If the penalty factors expressed in equation (S15) are added to the penalty function the regression vector shrinks and rotates towards the direction of highest negative gradient of \( e^T\mathbf{W}_e\mathbf{e} \). Thus, rate of convergence might decrease but as iterations proceed the vector will be re-oriented again.

Another condition that partially ensures convergence, at least in the application being explored here, is the fact that the AQM itself provides physical and chemical constraints. That is, the corrections cannot increase or decrease indefinitely. Given that the AQM provides a continuous solution space, the iteration will work towards a convergent point close to the base
emission estimates. Here another point needs to be made: it can be assumed that the original emission estimates provide good guesses of the “true” emissions such that the starting point of the iterations are not far away from the global minimum.

6. Model outputs and code implementation

Several statistical measures of the emission estimates are computed in order to better appreciate their behavior and to aid in the decision of which weighting combinations perform better in different assimilation cases. The coefficient of determination ($R^2$), variance-covariance matrix, total variance, variance inflation factors, condition number and variance proportions are all computed as a means to quantify the ill-posedness of the inverse problem (5).

The numerical implementation of the technique described here has been done using available linear algebra routines. Routines from LAPACK (Linear Algebra Package) were used to perform eigenvalue decompositions (e.g. singular value decomposition [SVD]) and matrix inversion operations. BLAS (Basic Linear Algebra Subprograms) routines were used for matrix-vector operations.
7. Hourly observations for August 9 and 10, 1992

TABLE S1. Number of hourly observations from August 9 and 10, 1992 used in the assimilation process.

<table>
<thead>
<tr>
<th>Station</th>
<th>O$_3$</th>
<th>NO</th>
<th>NO$_2$</th>
<th>NO$_x$</th>
<th>PAN</th>
<th>carbonyls</th>
<th>isoprene</th>
<th>toluene</th>
<th>ethene</th>
<th>OVOC</th>
<th>CO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 BARN$^a$</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 BFCK$^a$</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 BLCK$^c$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 BRMT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>5 CLOV</td>
<td>42</td>
<td>35</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 CONF$^c$</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 CONY$^e$</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 DALL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>47</td>
</tr>
<tr>
<td>9 DAWS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>48</td>
</tr>
<tr>
<td>10 FERN$^e$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 FORT</td>
<td>32</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 GAST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46</td>
</tr>
<tr>
<td>13 GWIN$^e$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 HOLC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>48</td>
</tr>
<tr>
<td>15 LOST</td>
<td>3$^i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 PANO</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 SCSP</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 SDEK</td>
<td>46</td>
<td>48</td>
<td>48</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 SHEN$^f$</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 STMT</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 TECH</td>
<td>37</td>
<td>10</td>
<td>10</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22 TUCK</td>
<td>48</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 YORK$^e$</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$x$
8. Behavior of the penalty function in the assimilation of noisy pseudo-observations

The values of λ and l at each iteration for the test with noisy pseudo-observations (for the combination of weighting factors that results in minimum total variance of the emission estimates) are shown in Table S2. In all cases, at the first iteration the extra components of the \( w_{ij} \) calculation (\( w_{ij}^{\text{unc}} \), \( w_{ij}^{\text{bounds}} \) and \( w_{ij}^{\text{spread}} \)) had to be used, and in three of those four cases the \( w_{ij} \) of some sources had to be increased iteratively to get an in-bounds solution. In the following iterations, \( \lambda \) and \( l \) were sufficient to constrain the solution to the assigned bounds in eight out of ten assimilation computations. The length parameter in particular deviates little from a base value of one (with \( l=1 \) the RR estimate is unchanged), and in all but one case it changes the value of the estimates by no more than 7%.
TABLE S2. Values of $\lambda$ and $l$ at each assimilation step for the noisy pseudo-data runs.

<table>
<thead>
<tr>
<th>Iteration</th>
<th>N1 Run</th>
<th></th>
<th>N2 Run</th>
<th></th>
<th>N3 Run</th>
<th></th>
<th>N4 Run</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda$</td>
<td>$l$</td>
<td>$\lambda$</td>
<td>$l$</td>
<td>$\lambda$</td>
<td>$l$</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>1</td>
<td>0.001</td>
<td>0.999$^b$</td>
<td>0.002</td>
<td>1.004$^a$</td>
<td>0.001</td>
<td>1.000$^b$</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>0.001</td>
<td>1.001</td>
<td>0.003</td>
<td>1.008$^a$</td>
<td>0.003</td>
<td>1.010</td>
<td>0.003</td>
</tr>
<tr>
<td>3</td>
<td>0.093</td>
<td>1.003</td>
<td>0.004</td>
<td>1.004$^a$</td>
<td>0.040</td>
<td>1.041</td>
<td>0.068</td>
</tr>
<tr>
<td>4</td>
<td>0.326</td>
<td>0.669</td>
<td>0.014</td>
<td>1.041</td>
<td>nn</td>
<td>nn</td>
<td>nn</td>
</tr>
</tbody>
</table>

$^a$ extra components ($w_{j<unc>}$, $w_{j<bounds>}$ and $w_{j<spread>}$) needed. $^b$ Iterative increasing in $w_{jj}$ needed. $^c$ Assimilation not needed.

Literature cited


(8) Gruber, M.H.J. *Improving efficiency by shrinkage: the James-Stein and ridge regression estimators* (Statistics: textbooks and monographs, v. 156); Marcel Dekker, Inc.: New York, **1998**.


