

UC Irvine

Faculty Publications

Title

A numerical method for parameterization of atmospheric chemistry: Computation of tropospheric OH

Permalink

<https://escholarship.org/uc/item/2bh07365>

Journal

Journal of Geophysical Research, 95(D11)

ISSN

0148-0227

Authors

Spivakovsky, C. M
Wofsy, S. C
Prather, M. J

Publication Date

1990

DOI

10.1029/JD095iD11p18433

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

A Numerical Method for Parameterization of Atmospheric Chemistry: Computation of Tropospheric OH

C. M. SPIVAKOVSKY AND S. C. WOFSY

Division of Applied Sciences and Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts

M. J. PRATHER

NASA Goddard Space Flight Center, Institute for Space Studies, New York

An efficient and stable computational scheme for parameterization of atmospheric chemistry is described. The 24-hour-average concentration of OH is represented as a set of high-order polynomials in variables such as temperature, densities of H₂O, CO, O₃, and NO_x (defined as NO + NO₂ + NO₃ + 2N₂O₅ + HNO₂ + HNO₄) as well as variables determining solar irradiance: cloud cover, density of the overhead ozone column, surface albedo, latitude, and solar declination. This parameterization of OH chemistry was used in the three-dimensional study of global distribution of CH₃CCl₃ (Spivakovsky et al., this issue). The proposed computational scheme can be used for parameterization of rates of chemical production and loss or of any other output of a full chemical model. Coefficients for the polynomials are computed to provide the least squares fit to results of the full chemical model. Highly overdetermined systems are used with the sets of independent variables selected randomly in accordance with the distributions expected in the atmosphere. The least squares problem is solved using the Householder method of triangularization (by orthogonal transformations). The method allows detection and rectification of ill-defined conditions (i.e., linear dependence among terms), as well as evaluation of the individual contribution of each term of the polynomial in reducing the residual vector. On the basis of that information the terms that have little bearing on the residual norm are discarded. Once the domain and the statistical distributions of independent variables are chosen, the entire parameterization procedure is implemented as a complete sequence of computer programs requiring no subjective analysis. The output of the procedure includes estimates of accuracy of the approximation against an independent sample of points, and computer written FORTRAN subroutines to compute the polynomials.

1. INTRODUCTION

Efficient computation of rates of chemical production and loss is essential in the three-dimensional chemical transport models intended to study distributions of reactive atmospheric tracers. Direct calculation of these rates requires solution of a nonlinear system of kinetic equations relating the unknown rates to tracer concentrations. The number of calls to a chemical module in a three-dimensional simulation is typically counted in millions, and even if advanced numerical algorithms are used, it is extremely expensive to solve the kinetic system directly within the chemical tracer model (CTM). An alternative approach, parameterization, involves replacing the nonlinear system with a set of explicit expressions describing the functional relationships between the unknown chemical rates (or concentrations of radicals, e.g., OH) and input parameters, such as tracer concentrations, temperature, pressure, and photodissociation constants (*J* values).

Dunker [1986] proposed to represent the *N*-dimensional domain of independent variables as a set of rectangular cells and to approximate the outputs of a photochemical model in each cell with a second-order Taylor expansion about the center of the cell. Marsden et al. [1987] replaced their photochemical model with quadratic polynomials obtained as a least squares fit (LSF) to the model results. These low-order approximations, when successful, have an advantage of algorithmic simplicity and low cost. The LSF polynomials reproduce the original function better in the integral sense than the Taylor expansions, which give an exact

value at the center of the cell and larger errors toward the boundaries.

In more complicated cases, however, higher-order approximations are preferable to avoid using an excessively fine grid, which in the *N*-dimensional space may lead to astronomical numbers of grid cells and thus to impractical cost in computer time and storage. However, the LSF procedure for high-order polynomials is often regarded as unsuitable, unstable, or too risky. The purpose of this paper is to describe an efficient and stable parameterization procedure based on computing high-order LSF polynomials. The procedure relies heavily on algorithms described by Lawson and Hanson [1974]. It was applied to parameterization of global OH used in the three-dimensional study of the distribution of CH₃CCl₃ [Spivakovsky et al., this issue], and it is being presently applied to parameterization of tropospheric chemistry for studies of regional pollution [Jacob et al., 1989].

2. DESCRIPTION OF THE PARAMETERIZATION PROCEDURE

The concentration of OH (or any other output of a photochemical model) can be regarded as a function of *N* independent variables (parameters) such as concentrations of trace gases, temperature, and radiation conditions. The task of parameterization is to find an analytical expression that mimics the functional relationship between the input parameters and the output of the full chemical model (FCM). As the first step of the parameterization procedure, we thoroughly sample the function using the FCM to obtain a set of points designated below as true points. An ordinary regular scheme of sampling, with points in the parameter space forming a rectangular grid, leads to enormous number of the FCM calculations and results in a sample that, despite its size, still does not define the most useful higher-order terms. For example, if

Copyright 1990 by the American Geophysical Union.

Paper number 90JD01300.
0148-0227/90/90JD-01300\$05 00

three values were used for each of the N parameters in a regular sampling scheme, then, for $N=10$ as much as 59,049 FCM calculations have to be performed, but only terms up to the second order in each variable can be used (the maximum order for each variable is determined by the number of projections onto the corresponding axis).

We chose to generate true points by selecting values for parameters randomly in accordance with the distributions expected in the atmosphere for each of the parameters. The random generation scheme provides the following advantages: (1) it allows us to achieve the best accuracy for the most prevalent conditions; (2) the number of FCM calculations is not dictated by a particular design of the sampling scheme; (3) the number of points projected onto each axis in the parameter space is roughly equal to the total number of points and in practice does not present a limitation for using terms of any reasonable order in a given variable. To ensure that the resulting polynomial reflects the general properties of the original function rather than a particular alignment of points in the sample, we use highly overdetermined LSF systems with the number of data points exceeding the number of unknowns by at least a factor of 10.

In solving the least squares problem we apply the method of Householder triangularization [Lawson and Hanson, 1974] to an overdetermined linear system $A_{m \times n} \times X_n = B_m$, where $A_{m \times n}$ is a matrix of m rows (with each row consisted of terms of the polynomial evaluated at the associated true point in the parameter space) and n columns (corresponding to n terms of the polynomial); unknown polynomial coefficients form the n -dimensional vector X_n ; ordinates of the true points (i.e., results of the FCM calculations) define the m -dimensional vector B_m . The Householder method of solution for LSF problems is superior in stability to the use of normal equations $(A^T A)_{n \times n} \times X_n = A^T \times B_m$, commonly recommended in handbooks. As shown by Lawson and Hanson [1974] precision v^2 has to be used to form and solve normal equations for the same class of problems that can be solved by the Householder method with precision v .

The algorithm HFTI by Lawson and Hanson [1974] takes advantage of orthogonal transformations, which do not change euclidean length and scalar product of vectors. Combining orthogonal transformations with interchange of columns, the algorithm transforms matrix $A_{m \times n}$ into an upper triangular matrix with diagonal elements nonincreasing in magnitude. Columns with diagonal elements smaller than a certain threshold (and associated terms of the polynomial) can be discarded, eliminating ill conditions. Inherent to this method is an immediate evaluation of the role that each column of matrix A (and each term of the polynomial) plays in reducing the residual vector.

The algorithm HFTI can be outlined as follows. At step 1 columns of $A_{m \times n}$ are interchanged to assign position 1 to a column of maximum euclidean length; the Householder transformation Q_1 is applied to $A_{m \times n}$ so that components 2 through m of the first column in $Q_1 A$ are zeroed; the magnitude of the first element of the first column r_1 then is equal to the euclidean length of the column. Note that if any of the columns of A are "almost parallel" to its first column, they will be transformed into columns "almost parallel" to the first column of $Q_1 A$ (because Q_1 is orthogonal). Columns of this sort are easily detectable in $Q_1 A$ as those with small magnitudes of the sum of squared components 2 through m . At step 2 the column for which this sum is the greatest is interchanged with the second column of $Q_1 A$; the orthogonal transformation

$$Q_2 = \begin{bmatrix} 1 & 0 \\ 0 & Q_H \end{bmatrix}$$

where Q_H is in turn the $(m-1)$ -dimensional Householder transformation that results in zeroing components 3 through m of the new column 2, is then applied to matrix $Q_1 A$; the magnitude of the diagonal element r_2 is then equal to the euclidean length of the vector comprising components 3 through m of column 2 in $Q_1 A$. Again, if any of the columns of $Q_1 A$ are "almost parallel" to the plane defined by the two first columns of $Q_1 A$, they will be transformed by Q_2 into columns whose sum of squared components 3 through m is small. Therefore at the next step, when candidates for position 3 in $Q_2 Q_1 A$ are considered, these columns will be left behind. Note, that Q_2 affects only columns 2 through n of $Q_1 A$ and rows 2 through m of $Q_1 A$ and $Q_1 B$, as will transformations Q_j ($j=3, \dots, n$) at subsequent steps affect only columns j through n of $Q^{(j-1)} A$ and rows j through m of $Q^{(j-1)} A$ and $Q^{(j-1)} B$. Here $Q^{(j-1)}$ denotes $Q_{j-1} \times \dots \times Q_1$.

Continuing in this way, one achieves the desired triangularization of $A_{m \times n}$ as $R = Q^{(n)} A$, with the diagonal elements r_1, r_2, \dots, r_n which possess two important characteristics: they are non increasing in magnitude and each element r_i provides a measure of linear independence for the first i columns of $A_{m \times n}$. Consequently, if it is found that $|r_i|$ is smaller than a certain linear dependency tolerance τ , the process of triangularization should be stopped at a step i and terms i through n of the polynomial discarded.

As noted above, columns of $A_{m \times n}$ can also be rated by their contributions to reducing the euclidean length of the residual vector $\Delta = A_{m \times n} X_m - B_m$. Conveniently, these contributions arise as an additional benefit of the algorithm, in the form of components of vector $Q^{(m)} B$. After step j , i.e., after triangularization of j columns is completed (and, hence, j terms of the polynomial are defined), the norm of the residual vector $\Delta^{(j)}$ satisfies the equation

$$\|\Delta^{(j)}\|^2 = \|Q^{(j)}(A_{m \times j} \times X_j - B_m)\|^2 = \sum_{i=j+1}^m b_i^{(j)2}$$

where $b_i^{(j)}$ denote components of $Q^{(j)} B$. Since transformation $Q^{(j+1)}$ does not change the sum of squared components $j+1$ through m of vector $Q^{(j)} B$,

$$\|\Delta^{(j)}\|^2 = \sum_{i=j+1}^m b_i^{(j+1)2} = b_{j+1}^{(j+1)2} + \|\Delta^{(j+1)}\|^2$$

i.e., adding a term $j+1$ reduces the squared residual norm by $b_{j+1}^{(j+1)2}$. Hence a measure of effectiveness of a term i can be given by $|b_i^{(j)}| / \|B_m\|$. We therefore have a procedure for discarding the terms that are ineffective in improving the fit.

Selection of appropriate terms for the polynomial is an essential part of the parameterization procedure. The goal is to compose a set that results in a well-conditioned matrix A and contains those and only those terms that significantly reduce the residual vector. This process is carried out automatically, without relying on intuition or a priori knowledge of the behavior of the function. We begin the process of selection by using a set of about 300 to 400 lower-order terms. Prior to computation of the matrix $A_{m \times n}$, we rescale parameters to approximately $[-1, 1]$ range. Using HFTI, we dispose of the terms that would cause ill conditions and of those that have little bearing on the residual norm. We then redo the LSF calculation with the remaining terms or, if the residual vector is not sufficiently small, with the set expanded by a new portion of higher-order terms. On the basis of our overall experience with parameterization of functions arising from chemical models, we exclude the crossterms of more than four parameters, thus speeding up considerably the selection process. Although this process usually leads to an acceptable approximation, it does not necessarily result in the optimal choice of terms, because the individual contributions of terms depend on contributions of other terms included in the LSF system. An alternative approach leading to mutually independent contributions (e.g., singular value

decomposition) would involve a linear change of variables, i.e., the original function would be approximated by a linear combination of polynomials, rather than monomials as in the present scheme.

The HFTI algorithm provides solutions to the LSF problem for multiple right-hand-side vectors (e.g., various output quantities of the FCM) at little additional cost. We adjusted the HFTI algorithm for sequential accumulation of rows of $A_{m \times n}$, also described by *Lawson and Hanson* [1974]. This modification is important for processing vast systems originating from problems that involve polynomials in a large number of independent variables (e.g., a typical size for systems we used was approximately $3,000 \times 300$). The number of computer operations (defined as a multiplication or division and an addition or subtraction) for the Householder method of solving LSF problem is approximately equal to $mn^2(k+1)/k$, where k is the number of rows in the portion of the sequentially added data [Lawson and Hanson, 1974]. Thus the sequential processing of true points requires little additional time for solving the LSF problem. (In the extreme case of $k=1$ it causes an increase in the number of operations only by a factor of 2).

For computers with physical memory large enough to accommodate the complete matrix $A_{m \times n}$ the selection process can be organized more efficiently by discarding less useful terms as soon as the associated column is chosen for the next position in the upper triangular matrix R . At a step i , as soon as the column i in $R_{m \times i} = Q^{(i)}A_{m \times i}$ is established, one can compute $b_i^{(i)}$ and, depending on the magnitude of $|b_i^{(i)}| / \|B_m\|$, either keep this column (and term) or discard it and resume the search for column i among columns $i+1$ through n of $Q^{(i-1)}A$.

Once the ranges and the statistical distributions of the parameters are chosen, the entire parameterization procedure is carried out by a complete sequence of computer programs, requiring no subjective analysis. The output of the procedure consists of coefficients for the polynomials and computer-generated FORTRAN subroutines for efficient computation of the polynomials (e.g., within the CTM). The FORTRAN program that generates these subroutines is designed to ensure that polynomials are calculated with a minimum number of multiplications. Figure 1 gives an example of the computer-generated subroutine for calculation of the OH concentration as a function of five parameters $x(i)$ which

```

FUNCTION F111 (X,C)
DIMENSION X(5),C(146)
F111=0.
F111=F111+
* C( 1)
* +X(1)*(C( 3)+X(1)*(C( 9)+X(1)*(C( 45)+X(1)*(C( 50)+X(1)*
* C( 20)+X(1)*(C( 51)+X(1)*(C( 15)+X(1)*(C( 24)+X(1)*(C( 37)
* +X(1)*C( 42)))))+X(2)*C( 69)+X(3)*C( 53)+X(4)*C( 83)+X(5)*
* C( 77))+X(2)*C( 59)+X(3)*C( 44)+X(4)*C( 57)+X(5)*C( 65))+X(2)*
* C( 97)+X(3)*C( 76)+X(2)*C( 79)+X(3)*C( 64))+X(4)*C( 54)+X(2)*
* C(101)+X(3)*C( 89)+X(4)*C(116))+X(5)*C(109)+X(2)*C(107)+X(3)*
* C( 87)+X(4)*C(137)+X(5)*C(120))+X(2)*C( 39)+X(3)*C( 38)+X(2)*
* C( 91))+X(4)*C( 49)+X(2)*C( 74)+X(3)*C( 94))+X(5)*C( 60)
* +X(3)*C( 85)+X(4)*C( 90)))
F111=F111
* +X(2)*C( 16)+X(1)*C( 31)+X(2)*C( 18)+X(1)*C( 55)+X(2)*C( 47)
* +X(1)*C(125)+X(2)*C( 10)+X(1)*C( 88)+X(2)*C( 25)+X(2)*C( 7)
* +X(2)*C( 26)+X(2)*C( 13)+X(2)*C( 52)+X(2)*C( 33)))))+X(4)*
* C( 99))+X(3)*C(134)+X(4)*C(105)+X(5)*C(117))+X(3)*C( 92)+X(4)*
* C( 62)+X(5)*C( 82)))
F111=F111
* +X(3)*C( 12)+X(1)*C( 35)+X(2)*C( 63)+X(1)*C( 73))+X(3)*
* C( 48)+X(1)*C( 61)+X(1)*C( 68)+X(4)*C(122))+X(2)*C(129)
* +X(1)*C(136))+X(3)*C( 32)+X(1)*C(127)+X(1)*C(126))+X(2)*
* C(132)+X(1)*C(145))+X(3)*C( 14)+X(1)*C(141)+X(3)*C( 17)+X(1)*
* C(144)+X(2)*C(143)+X(3)*C( 8)+X(3)*C( 21)+X(3)*C( 11)+X(3)*
* C( 43)+X(3)*C( 30)))))+X(4)*C( 86))+X(4)*C( 81)+X(5)*C(135)
* +X(1)*C(146))+X(4)*C(113)+X(1)*C(111))+X(5)*C(138)+X(1)*
* C(139)))
F111=F111
* +X(4)*C( 6)+X(1)*C( 40)+X(2)*C( 34)+X(1)*C( 78))+X(3)*
* C( 23)+X(1)*C(128)+X(2)*C( 70))+X(4)*C( 22)+X(1)*C( 46)+X(1)*
* C( 71))+X(2)*X(2)*C(130)+X(3)*C( 96)+X(4)*C( 36)+X(1)*C(106)
* +X(2)*C( 98)+X(3)*C( 95)+X(4)*C( 5)+X(1)*C( 93)+X(4)*C( 27)
* +X(4)*C( 2)+X(4)*C( 29)+X(4)*C( 4)+X(4)*C( 56)+X(4)*
* C( 19)))))))+X(5)*C(131)))
F111=F111
* +X(5)*C( 28)+X(1)*C( 41)+X(2)*C( 58)+X(1)*C( 72))+X(3)*
* C(108)+X(1)*C(133)+X(2)*C(142))+X(4)*C( 67)+X(1)*C(104)+X(2)*
* C(100))+X(5)*C(118)+X(1)*C(115)+X(1)*C(123))+X(2)*C(124)
* +X(4)*C(119)+X(1)*C(121))+X(5)*C( 75)+X(1)*C(114)+X(4)*C(140)
* +X(5)*C(103)+X(5)*C( 66)+X(5)*C(102)+X(5)*C( 80)+X(5)*
* C( 84)+X(5)*C(112)+X(5)*C(110)))))))))
RETURN
END

```

Fig. 1. The computer-written FORTRAN function to compute the concentration of OH expressed as a polynomial in five variables: NO, O₃, CO, water vapor, and temperature.

correspond to concentrations of NO_x (defined as $\text{NO} + \text{NO}_2 + \text{NO}_3 + 2\text{N}_2\text{O}_5 + \text{HNO}_2 + \text{HNO}_4$), CO , O_3 , water vapor, and temperature, respectively. In addition to these subroutines and associated polynomial coefficients, the output of the parameterization procedure contains estimates for accuracy of the approximation, (namely, results of tests performed using a separate sample of true points that does not overlap with the sample used to form the LSF system), and a list of the individual contributions of different terms to reducing the residual norm.

For parameterization of the concentration of OH, the regions of different regimes were identified prior to the LSF computations (see section 3 and Table 1). Then for each region a single polynomial was obtained. However, when we extended the parameterization for a study of regional O_3 pollution [Jacob *et al.*, 1989], which required adequate representation of chemistry in both remote and highly polluted environments, and thus the accommodation of a much wider range for free parameters, at a certain

stage of the selection process the augmentation of additional terms was no longer reducing the residual norm at a satisfactory rate. Therefore the parameter domain had to be subdivided and a separate polynomial obtained for each regime. In general, the manual search for suitable subdivisions in the multidimensional space might be difficult. In the future we plan to implement a computerized search for optimal subdivisions based on the procedure outlined below.

The N -dimensional domain can be divided into two parts by a plane

$$\sum_{i=1}^N a_i x_i + a_{N+1} = 0$$

where x_i denote parameters. Coefficients a_i (for $i=1, \dots, N+1$) can be obtained as a solution to the following optimization problem. Define function $\rho(a_1, a_2, \dots, a_{N+1})$ as a root-mean-square error (r.m.s.) for a piece-polynomial function composed of two second-

TABLE 1. Ranges for Independent Variables Used in the Parameterization of Chemistry

Independent Variable (Parameter)	Tropics Year Round	Middle Latitudes Summer	Middle Latitudes Spring/Fall	Middle Latitudes Winter
NO _t , pptv	5-1000	5-1000	5-1000	5-1000
O ₃ , 10 ¹¹ mol cm ⁻³	1-15	1-25	1-25	1-25
CO, ppbv	40-300	50-600	50-600	50-600
H ₂ O ^a	0.1-4	0.1-4	0.1-4	0.1-4
ΔTemperature ^a , K	-15 to 15	-20 to 20	-20 to 20	-20 to 20
O ₃ column, DU	195-295	250-430	250-430	250-430
Latitude	30°S-30°N	30°-90°	30°-90°	30°-90°
Solar declination ^b	-23° to 23°	10° to 23°	-10° to 10°	-23° to -10°
Ground albedo	0.05-0.3	0.05-0.45	0.05-0.45	0.05-0.45
Cloudiness ^c				
below 800 mb	0-0.7	0-0.7	0-0.7	0-0.7
800-500 mbar	0-0.3	0-0.3	0-0.3	0-0.3
500-200 mbar	0-0.6	0-0.6	0-0.6	0-0.6
200-150 mbar	0-0.3	-	-	-

^a Ranges for H₂O and temperature are given in factors and increments, respectively, with respect to the mean seasonal vertical profiles [Oort, 1983].

^b Ranges for solar declination are given for the northern hemisphere, with the appropriate change of sign for the southern hemisphere.

^c Cloud albedo is assumed to be 50%, 40%, 30%, and 20% for clouds below 800, at 800-500, at 500-200, and at 200-150 mbar, respectively.

order polynomials with a separate LSF polynomial for each of the two parts of the domain: (1) $\sum_{i=1}^N a_i x_i + a_{N+1} \geq 0$ and (2) $\sum_{i=1}^N a_i x_i + a_{N+1} < 0$. Determine coefficients a_i (for $i=1, 2, \dots, N+1$) as those that define a minimum of $\rho(a_1, a_2, \dots, a_{N+1})$, using, for example, the method of steepest descent. Multiple calls to the LSF procedure will be necessary for solving this minimization problem. However, for a second-order polynomial, these calls are relatively inexpensive. An appropriate choice of the starting point for the descent may considerably speed up the minimization process. The same method, but with the first-order LSF polynomials (for which the LSF calculation is trivial) can be used to determine the starting point. After the plane of optimal subdivision is determined the full procedure with the automatic selection of appropriate higher-order terms would be carried out for each subdivision, and, if necessary, further subdivision performed.

3. PARAMETERIZATION OF GLOBAL OH

The 24-hour-average concentration of OH was expressed as a polynomial in 13 independent variables including temperature, radiation conditions, and concentrations of O_3 , NO_x , CO and water vapor. Table 1 summarizes the range of applicable values for the independent variables. Separate polynomials were obtained for different pressure layers (see Table 2) and for four different regimes: for latitudes $30^\circ - 90^\circ$ in summer; in winter; in spring and autumn; and for the tropics ($30^\circ S - 30^\circ N$) year around.

TABLE 2. The r.m.s. Errors of the Polynomial Fit

Pressure, mbar	$30^\circ S - 30^\circ N$	$30^\circ - 90^\circ$ Summer	$30^\circ - 90^\circ$ Spring/Fall	$30^\circ - 90^\circ$ Winter
1000	4	6	8	22
900	4	6	8	22
800	4	6	8	24
700	4	6	8	24
500	4	7	8	26
300	5	11	10	20
200	6	13	10	21
150	6			
100	6			

Root-mean-square errors are given for values of OH computed using analytic expressions compared with values calculated from the full photochemical model, for each region and altitude, in percent of the mean OH for the sample of 3,000 points (chosen randomly). The percentage of annual loss for CH_3CCl_3 (by reaction with OH) was 76, 17, 6, and 1% in tropics year around, and in mid-latitudes in summer, spring and winter, respectively.

Table 2 summarizes results of a test based on extensive sampling of the OH surface in the 13-dimensional parameter space. For each pressure level in the table we conducted 3,000 simulations (in addition to those used in the LSF system) using the FCM and the polynomial expressions. To ensure adequate representation of various atmospheric conditions, the input parameters for the test sample (as for the sample used in the LSF system) were chosen randomly in accordance with the distributions expected in the atmosphere. The r.m.s. given in percent of the mean OH for the sample is used as a measure of accuracy. The r.m.s. is significantly less than 10% for the tropics and for the low and middle troposphere at higher latitudes in all seasons but winter. A higher r.m.s. for winter (about 26%) and for the upper troposphere in other seasons (less than 13%) could have been reduced if required, but the effort was unwarranted since associated rates of reaction with CH_3CCl_3 , CO, CH_4 , i.e., tracers that we intended to study using the present parameterization of OH, are inconsequential. The mean error of the fit, a measure of systematic error for a given test sample, is less than 1% for all regions of parameter space.

Figure 2 illustrates the accuracy of the approximation. The concentration of OH is shown as a function of two parameters, O_3 and NO_x , with other parameters fixed at intermediate values. The surface was obtained using the polynomial expression. Also shown is a surface defined by the difference between OH concentrations computed using the polynomial fit and the FCM. This difference is negligibly small; largest errors are associated with the edges of the domain.

Figure 3 shows examples of vertical profiles for a range of conditions giving high, low, and intermediate OH values. The curves computed using the polynomials closely track results from the

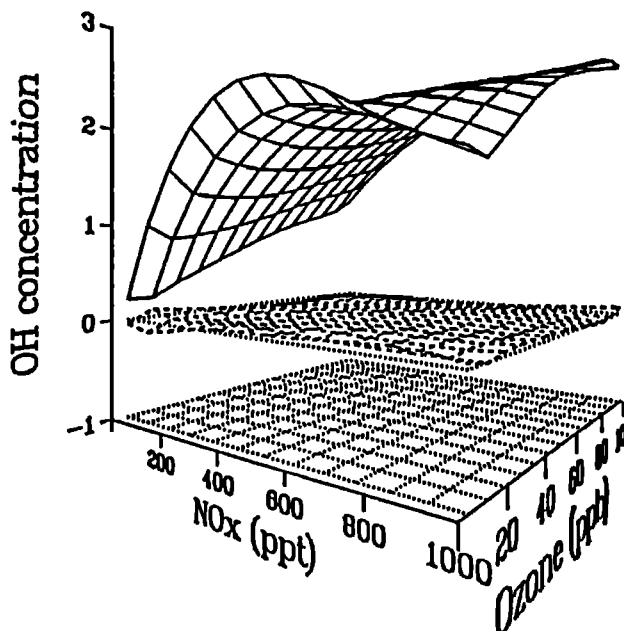


Fig. 2. Concentration of OH calculated using the polynomial (surface in solid lines). The surface given in dashed lines represents the difference between OH concentrations computed using the full scale chemical model and the parameterization. The concentration of OH is given as a function of the concentrations of O_3 and NO_x at the surface at 45° at the summer solstice. Other parameters were fixed at typical values, $H_2O = 2.37 \times 10^{-2}$ volume mixing ratio, CO = 90 ppbv, temperature = 302.6 K, ozone column = 245 DU, surface albedo = 0.09, cloud cover of 20%, 8%, 12%, and 6% at 800, 500, 200 and 150 mbar, respectively.

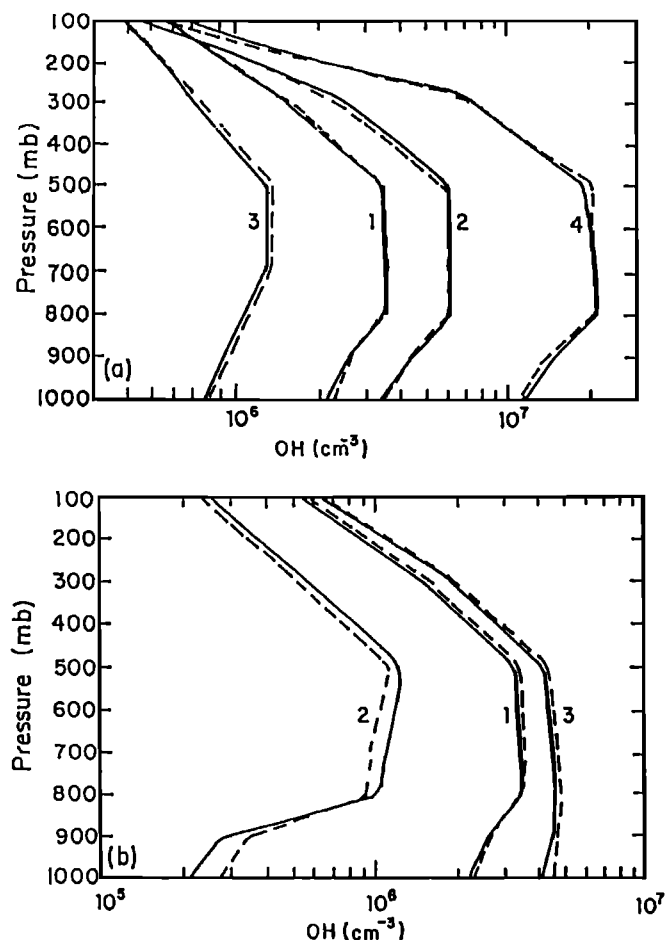
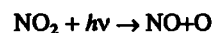
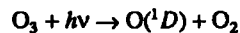


Fig. 3. Examples of OH profiles calculated using the FCM (dashed lines) and the parameterization (solid lines). (a) Calculations were performed for various background concentrations and temperatures, with insolation appropriate for the equator at equinox. Case 1 (Figure 3a) represents most probable conditions within the prescribed distribution: $\text{NO}_t = 100$ pptv, $\text{CO} = 90$ ppbv, $\text{O}_3 = 6 \times 10^{11}$ molecules cm^{-3} , with water vapor and temperature from mean tropical profiles [Oort, 1983]. Case 2 (Figure 3a) represents higher values than in case 1: $\text{NO}_t = 300$ pptv, $\text{CO} = 180$ ppbv, $\text{O}_3 = 9 \times 10^{11}$ molecules cm^{-3} , water vapor equal to twice the mean, with temperature 5° higher than the mean. Case 3 (Figure 3a) corresponds to lower values than in case 1: $\text{NO}_t = 33$ pptv, $\text{CO} = 60$ ppbv, $\text{O}_3 = 3 \times 10^{11}$ molecules cm^{-3} , water vapor three times lower than the mean and temperature 5° lower than the mean. Case 4 (Figure 3a) represents less probable conditions that result in especially high values for OH: $\text{NO}_t = 1000$ pptv, $\text{CO} = 40$ ppbv, $\text{O}_3 = 1.5 \times 10^{12}$ molecules cm^{-3} , water vapor equal to four times the mean. (b) Case 1 is the same as case 1 in Figure 3a and is given for comparison with calculations that assumed reduced and elevated levels of radiation. In all cases in Figure 3b, NO_t , CO , O_3 , water vapor, and temperature are fixed as in case 1 Figure 3a. Case 2 corresponds to reduced insolation for 30°N and declination -23° (January) and ozone column of 295 DU. The surface albedo was set to 0.05 and cloud cover to 70%, 30%, 60% and 30%. Case 3 represents full insolation at the equator at equinox, with ozone column 195 DU, surface albedo of 0.3, and no clouds.

FCM in all cases, including those with extreme values for insolation and background concentrations. In summary, results of extensive testing show that the polynomials derived here accurately represent the FCM for OH. At the same time, the computational cost was reduced by more than a factor of 600.

Solar irradiance as a function of wavelength is one of the factors that determine the OH concentration and rates for chemical reactions in the atmosphere. Solar irradiance in turn is determined by solar zenith angle, cloudiness, surface albedo, and density of the overhead ozone column. Clearly, it is in the interest of both expediency of parameterization procedure and efficiency of the CTM computations to separate the two functional dependencies: (1) chemical reactivity on the level of solar irradiance and (2) solar irradiance on the conditions determining the number of photons available. Photolysis rates for reactions



correspond to two most important wavelength bands for the tropospheric chemistry and can be chosen as the parameters that describe radiation conditions. We applied this approach successfully to later work. For the present parameterization, a sufficiently accurate (although somewhat less elegant) approach was used. For each of the regions a polynomial in such independent variables as temperature and concentrations of NO_t , O_3 , CO , and water vapor was obtained for radiation conditions fixed at an intermediate level; a separate polynomial in such variables as latitude, solar declination, four layers of cloudiness, density of ozone column, surface albedo, and concentrations of O_3 and NO_t was defined as a correction factor to represent the full range of radiation conditions for the region. We also found in subsequent work that the detailed subdivision of the domain by pressure used in the present parameterization was unnecessary. Pressure can be treated as an additional parameter for three or fewer regimes, e.g., lower, middle, and upper troposphere.

4. SUMMARY AND CONCLUSIONS

The computational scheme described here allows us to represent a model output as a high order polynomial in the input parameters of the model. The following points are essential for the success of the procedure. (1) Highly overdetermined systems ($m \geq 10n$) should be used. (2) Regular rectangular schemes of sampling the domain should be avoided. Random generation of parameters in accordance with the distributions expected in the atmosphere provides an excellent scheme. (3) The LSF system in the form $A_{m \times n} \times X_n = B_m$ rather than in the form of normal equations should be used. (4) The Householder method for solving the LSF system should be used. It provides a mechanism for sorting out columns of the original matrix A (and hence the associated terms of the polynomial) to select those that do not cause ill conditions; in addition, it makes readily available the individual contribution of each column (and term) in reducing the residual norm, enabling further selection of the terms on the merits of effectiveness.

Terms involving more than four parameters do not prove to be particularly useful in reducing the residual norm. When the number of parameters N is large, crossterms represent extensive sets and a priori elimination of this group may speed the selection process considerably.

Rates for photodissociation of $\text{O}_3 \rightarrow \text{O}(^1D)$ and NO_2 can be used to characterize the level of solar irradiance. These J values can be used as input parameters in parameterization of chemical rates (or other components of the solution of the kinetic equations). The J values, in turn, can be computed using separate polynomials describing the functional dependence on cloudiness, density of the ozone column, solar zenith angle, and surface albedo.

The present parameterization of 24-hour-average global OH in the form of computer written FORTRAN subroutines and coeffi-

cients for evaluating the polynomials can be obtained from the authors. It was used in the three-dimensional study of the distribution of CH_3CCl_3 described by Spivakovsky *et al.* [this issue], and it is presently being used for studies of global distributions of CO and CH_4 at Harvard and at the Goddard Institute for Space Studies.

Acknowledgments. We thank J. Logan at Harvard for help in determining ranges and distributions for quantities used as independent variables in parameterization of OH. We also thank B. Mirman at Digital Equipment Corporation and J. Yatteau, R. Yevich, and D. Jacob at Harvard for useful discussions. We are grateful to C. Demore and J. Rollet for editorial assistance. This work was supported by NASA grant NAG5-719, NSF grant ATM84-13153, and EPA grant R814535010 to Harvard University, and by Division of Applied Sciences, Harvard University.

REFERENCES

- Dunker, A. M., The reduction and parameterization of chemical mechanisms for inclusion in atmospheric reaction-transport models, *Atmos. Environ.*, **20**, 479-486, 1986.
- Jacob, D. J., S. Sillman, J. A. Logan, and S. C. Wofsy, Least independent variables method for simulation of tropospheric ozone, *J. Geophys. Res.*, **94**, 8497-8509, 1989.
- Lawson, C. L. and R. J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, N. J., 1974.
- Marsden, A. R., Jr., M. Frenklach, and D. D. Reible, Increasing the computational feasibility of urban air quality models that employ complex mechanisms, *JAPCA*, **37**, 370-376, 1987.
- Oort, A. H., Global atmospheric circulation statistics, 1958-1973. *Natl. Oceanic Atmos. Admin. Prof. Pap.* **14**, 1983.
- Spivakovsky, C. M., R. Yevich, J. A. Logan, S. C. Wofsy, M. B. McElroy, and M. J. Prather, Tropospheric OH in a three-dimensional chemical tracer model: An assessment based on observations of CH_3CCl_3 . *J. Geophys. Res.*, this issue.
- M.J. Prather, NASA Goddard Space Flight Center, Institute for Space Studies, 2880 Broadway, New York, NY 10025.
- C.M. Spivakovsky and S.C. Wofsy, Division of Applied Sciences and Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138.

(Received September 15, 1989;
revised June 11, 1989;
accepted June 11, 1990.)