An Efficient Chemical Kinetics Solver Using High Dimensional Model Representation

Jeffrey A. Shorter[†] **and Precila C. Ip**

Mission Research Corporation, 1 Tara Blvd., Suite 302, Nashua, New Hampshire 03062

Herschel A. Rabitz*

Department of Chemistry, Princeton University, Princeton, New Jersey 08544-1009

Received: November 6, 1998; In Final Form: June 18, 1999

A high dimensional model representation (HDMR) technique is introduced to capture the input−output behavior of chemical kinetic models. The HDMR expresses the output chemical species concentrations as a rapidly convergent hierarchical correlated function expansion in the input variables. In this paper, the input variables are taken as the species concentrations at time t_i and the output is the concentrations at time $t_i + \delta$, where δ can be much larger than conventional integration time steps. A specially designed set of model runs is performed to determine the correlated functions making up the HDMR. The resultant HDMR can be used to (i) identify the key input variables acting independently or cooperatively on the output, and (ii) create a high speed fully equivalent operational model (FEOM) serving to replace the original kinetic model and its differential equation solver. A demonstration of the HDMR technique is presented for stratospheric chemical kinetics. The FEOM proved to give accurate and stable chemical concentrations out to long times of many years. In addition, the FEOM was found to be orders of magnitude faster than a conventional stiff equation solver. This computational acceleration should have significance in many chemical kinetic applications.

1. Introduction

Chemical kinetics models are important for analysis and design in many areas of chemistry and for industrial processes. Two goals for model development are improving prediction quality and reducing run times. These goals are typically linked in an inverse manner such that increased model prediction quality leads to slower models, and conversely, to reduce model run times, quality is sometimes sacrificed. This paper presents a new analysis tool that uses a high dimensional model representation (HDMR) to (i) identify key model input variables (and sets of cooperating variables) that have significant influence on the model output and should be the focus of future research to improve the model and (ii) produce an extremely fast and accurate fully equivalent operational model (FEOM) serving as an efficient chemical kinetic solver. The FEOM can directly replace the original chemical kinetic equation integrator to significantly reduce computational costs for following the chemical evolution. This paper applies the HDMR technique to a stratospheric chemistry model for analysis and illustrates the capability of creating a fast FEOM chemical kinetics solver.

The first focus of the paper is on the identification of key model variables. Previous attempts at identifying key physical variables in kinetic systems have generally relied on traditional sensitivity analysis to provide insights into chemical mechanisms. Sensitivity analysis quantifies the effects that single parameter variations have on the model output.[1−5] These investigations have generally been local around an operating point in the variable space. Global analyses have traditionally focused on output uncertainties and not key parameter identification.[6−9] In this paper, we will show that the HDMR procedure can give global coverage for identifying key input variables and their cooperative effects on the model output.

The second focus of this paper is on introducing a new representation as a basis for creating an alternative to traditional chemical kinetic ordinary differential equation solvers. This representation does not solve the chemical equations directly; rather, it relies on a special precalculated database capturing the chemical model's input−output relationships. From these quantitative input−output relationships, a fully equivalent operational model (FEOM) can be constructed to directly calculate species concentrations and related chemical properties from the inputs of the initial photochemical state. For example, the equivalent representation can take species concentrations at time t_1 and directly calculate concentrations at time t_2 based on knowledge of the functional input−output relationships. This evaluation can be extremely fast because often only a limited set of additions and multiplications is involved. In contrast, a traditional approach would numerically integrate the set of coupled kinetic differential equations, often taking many operationally complex steps to march forward from t_1 to t_2.

Several approaches have been taken to create model representations to act as a kinetic equation solver. Output quality from different representation approaches can vary greatly depending on (i) the type of representation and (ii) the representation construction methods. Spivakovsky et al.[10] used a curve fitting method to express the input−output chemical model response, Turányi[11] extended the approach of Spivakovsky et al. by expressing the model input−output relation as an expansion in orthogonal polynomials, and Tatung[3] and Georgopoulos[12] use a direct decoupled method to represent the model with the focus on uncertainty analyses. The HDMR approach used here employs an ***exact*** finite order hierarchical function expansion to capture the model input−output relationships. The functions in the expansion are optimal for the particular kinetic model, and the encapsulated input−output

[†] Present address: TASC, 55 Walkers Brook Drive, Reading, MA 01867.

10.1021/jp9843398 CCC: $18.00
Published on Web 08/19/1999

information is used to generate the FEOM. All of the representation approaches above share the common feature of first performing a set of model runs to capture the temporal input−output behavior over some window in time; the differences reside in how this task is done and how the information is used. In addition to representation approaches, there are numerous efforts to directly speed up chemical integrators.[13,14]

The structure of the paper is as follows. Section 2 describes the HDMR technique. An illustration of HDMR applied to a stratospheric chemistry mechanism is presented in section 3. The key input variables of the chemical mechanism are determined and a fast chemical solver is created. Some brief conclusions are given in section 4.

2. Technique

2.1. High Dimensional Model Representation (HDMR). The HDMR technique addresses the perceived difficult problem of mapping the input−output relationship behavior of complex systems. A traditional approach to mapping the behavior of a system with n input variables $x_1, ..., x_n$ would consist of sampling each variable at s points to assemble an interpolated lookup table of computational effort scaling as $\sim s^n$. Realistically, one may expect s to be approximately 10−20 and n to be $10-10^2$ or larger (e.g., $n = 46$ in the model in section 3.1). Viewed from this perspective, attempts at creating a lookup table would be prohibitive. Furthermore, the evaluation of a new point by interpolation in an n-dimensional space would be exceedingly difficult. However, this analysis implicitly assumes that all n variables are important and, most significantly, that there are correlations among variables to all orders (i.e., independently, in pairs, ..., up to all n variables acting in a tightly correlated fashion).

The fundamental principle underlying the HDMR is that, from the perspective of the output, the order of the correlations between the independent variables will die off rapidly. This assertion does not eliminate strong variable dependence or even the possibility that all the variables are important. Various sources of information support this point of there being limited high-order correlations. First, the variables in most physical models are chosen to enter as independent entities. This kinematic simplicity tends to survive in the output, although often scrambled in a complex fashion. Second, traditional statistical analyses of model behavior has revealed that a variance and covariance analysis of the output in relation to the input variables often adequately describes the physics (i.e., only low-order correlations describe the dynamics). These general observations lead to a dramatically reduced computational scaling when one seeks to map input−output relationships of complex systems. Considering this analysis, one may now show that the labor involved to learn the input−output behavior scales as only $\sim (sn)^l/l!$ for $l \ll n$. Here, l is the highest order of significant variable correlation, and typically, $l \leq 3$ has been found to be quite adequate. Such polynomial effort poses a far more tractable algorithm than the often-accepted view of exponential growth, $\sim s^n$, indicated above with traditional algorithms.

Evaluating the input−output response of the model generates a HDMR. This is achieved by expressing each model output variable as a hierarchical, correlated function expansion of a special mathematical structure and evaluating each term of the expansion independently. One may show that a model output that is a function of the input variables, $g(\mathbf{x}) \equiv g(x_1, x_2, ..., x_n)$, can be decomposed into summands of different dimensions:

$$g(x_1, x_2, ..., x_n) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1 \leq i < j \leq n} f_{ij}(x_i, x_j) + ... + f_{1,2,3,...,n}(x_1, x_2, ..., x_n) \quad (1)$$

where f_0 is a constant, the function $f_i(x_i)$ describes the independent action of the variable x_i upon the output, while $f_{ij}(x_i, x_j)$ gives the pair correlated impact of x_i and x_j upon the output, etc. Finally, the last term $f_{1,2,3,...,n}(x_1, x_2, x_3, ..., x_n)$ contains any residual correlated behavior over all of the system variables.[15−17] It is important to note that this expansion is of finite order and is always an exact representation of the model output.

The expansion is evaluated relative to a nominal point $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_n)$ in the overall input variable phase space. The f_0 term is the model output evaluated at the nominal point. The higher order terms are evaluated as cuts in the input variable phase space through the nominal point. Each first-order function $f_i(x_i)$ is evaluated along its variable axis through the nominal point. Each second-order function $f_{ij}(x_i, x_j)$ is evaluated in a plane defined by each binary set of input variables through the nominal point, etc. The functions are prescribed as

$$f_0 = g(\bar{\mathbf{x}})$$

$$f_i(x_i) = g(\bar{\mathbf{x}}^i, x_i) - f_0$$

$$f_{ij}(x_i, x_j) = g(\bar{\mathbf{x}}^{i,j}, x_i, x_j) - f_i(x_i) - f_j(x_j) - f_0 \quad (2)$$

where the terminology of $\bar{\mathbf{x}}^i$ means that all the variables are at the nominal value except x_i, etc. The process of subtracting off the lower order functions removes their dependence to ensure a unique contribution from the new expansion function. As a result, the expansion functions only contain the information of the specified level of interaction, and they satisfy the null point criteria:

$$f_{ij...l}(x_i, x_j, ..., x_l)|_{x_p = \bar{x}_p} = 0 \qquad \text{for} \qquad p \in (i, j, ..., l) \quad (3)$$

This criterion ensures that the functions in eq 1 are orthogonal using a special inner product defined with respect to the nominal point.[17] Application of the property in eq 3 to $g(x_i, ..., x_n)$ in eq 1 will yield the relations in eq 2. The functions in eqs 1 and 2 yield exact information about $g(x_i, ..., x_n)$ along the cut lines, surfaces, subvolumes, etc. through the nominal point. The function $g(\mathbf{x})$ evaluated at a point $\mathbf{x}$ off of these cuts can be obtained by low-order interpolation of the functions on the right-hand side of eq 1. Any residual sensitivity to the choice of $\bar{\mathbf{x}}$ will rapidly disappear as sufficient terms for convergence are utilized.

The physical content and compact form of the HDMR expansion can be especially valuable in systems with large numbers of variables ($n \gg 10$) when convergence is achieved at low order. A benefit of this HDMR structure is that each expansion function uniquely describes the physical cooperation between the input model variables (either individually or collectively) upon the model output of interest. Thus, through an analysis of the magnitude and behavior of the expansion functions, one can determine (i) the input variables that have the largest impact on the model output, (ii) the nonlinear extent of the input−output relationships, and (iii) the input variables that are working cooperatively.

The individual functions are calculated separately by running the model a number of times with a series of judiciously chosen input variable sets that are consistent with eq 2 and the null point criteria in eq 3. Special care is taken to ensure that the

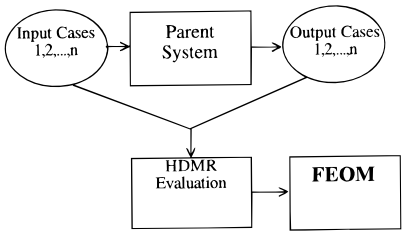

Figure 1. Schematic for evaluating HDMR expansion functions. A series of specially designed input sets are generated and run through the model. The resulting input−output information is transformed into a HDMR using eq 1. The HDMR is then encapsulated into the FEOM.

relevant portion of the model input phase space (i.e., as indicated from a typical full year simulation with the original model) is covered when generating the HDMR. The HDMR generation process is shown schematically in Figure 1. Importantly, no regression is employed and no constraining forms are applied to the expansion functions; they are each numerically represented as an interpolated look-up table. Other attempts at casting analogous model look-up table replacements largely have not been successful, as they have either (1) run up against the exponential growth problem $\sim s^n$, (2) relied on fitting high-order functions, or (3) employed complete sets of mathematical functions. The HDMR procedure introduces the critical concept of hierarchical variable correlations in natural compatibility with system behavior and then admits arbitrary correlated behavior.

The HDMR is generated to produce output at a particular time. Importantly, the number of full model runs performed to determine the HDMR expansion functions only depends on the number of input variables and not on the number of output species. The expansion functions are calculated in sequence from the zeroth order to the highest desired order utilizing the formulation in eq 2. The f_0 term is determined with a single model run with all input variables set to their nominal values.

A first-order function, $f_i(x_i)$, is calculated from $g(\bar{\mathbf{x}}^i, x_i)$ by setting all the input variables except x_i to their nominal input values and then performing a series of model runs with the input value of x_i varied over a specified range. The f_0 term is subtracted off from each model output to produce the function $f_i(x_i)$ as shown in eq 2. The points sampling the function can be picked as appropriate for the variation of $g(\bar{\mathbf{x}}^i, x_i)$. Here we assume that s values are used for each input variable. Thus $s - 1$ model runs specify each first-order expansion function on a well-resolved grid. The model run at the nominal point is not required since the value of the first-order function is zero at that point by virtue of eq 3. If there are n input variables in the model, then there are n first-order expansion functions. Thus, determination of all the first-order functions requires $n(s - 1)$ model runs.

A second-order function, $f_{ij}(x_i, x_j)$, is calculated by setting all the input variables, except x_i and x_j, to their nominal values and performing a series of runs with the values of x_i and x_j varied to cover the binary surface input phase space. Each variable is represented with s points, thus the binary input surface requires $(s - 1)^2$ model runs. The points along the two nominal cuts $x_i = \bar{x}_i$ and $x_j = \bar{x}_j$ of the surface are zero (viz. eq 3) and need not be recalculated. There are $n(n - 1)/2$ s-order expansion functions, and they are evaluated with $n(s - 1)^2(n - 1)/2$ model runs. The higher order expansion functions would be calculated in an analogous manner, with the expectation of rapid convergence of the correlated function expansion at relatively low order. In practice, a large number of expansion functions at any order may be insignificant. These can be identified and eliminated by selectively sampling each function

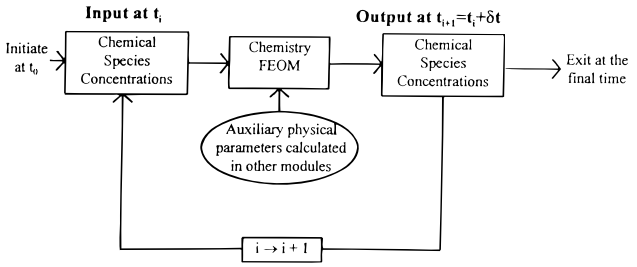

Figure 2. Schematic of chemistry FEOM implementation.

to determine its magnitude. After the overhead for determining the expansion functions is paid, the accurate and rapidly evaluated HDMR expression may be used for all subsequent model numerical analyses.

An inherent property of a closed chemical system is conservation of mass. Although the chemical species can change in quantity, the total amount of each element must remain constant. This fact places a very strong demand on chemical kinetics solvers, and the requirement is automatically fulfilled by each expansion function of the chemical kinetic HDMR. Hence, the HDMR-based FEOM replacement of the original model is guaranteed to conserve mass regardless of the order of truncation of the expansion. This result implies that the HDMR at any level of truncation corresponds to a physically acceptable mechanism, even if it differs somewhat from the original model mechanism (e.g., it might be missing a small high-order correlation term).

Satisfaction of the mass conservation property can be understood conceptually from the fact that each HDMR expansion function describes the impact on the output through a set of coupled kinetic interactions (i.e., unimolecular interactions, bimolecular interactions, etc.) over a fixed number of model variables. Thus, if an interaction produces a chemical species, then that same interaction must also appropriately destroy the species with the elements that form the new species. For example, an increase in NO concentration, in the illustration below, will increase the $ClONO_2$ concentration and also decrease the ClO concentration. This maintains the mass balance of the Cl atoms, and this property is preserved by the HDMR expansion.

2.2. Fully Equivalent Operational Model. The FEOM is a specific use of the HDMR where the model representation is encapsulated and used to directly calculate equivalent model output from model input sets. The benefit of the FEOM approach is that the equivalent model output is calculated in a fraction of the time required by the original model. The generation of a FEOM for evaluating chemical species concentration evolution starts by defining the input as the species concentrations and any other auxiliary variables (e.g., the temperature and photochemical rate functions in the illustration in section 3). The output is then the species concentrations at a time δ later. The latter output concentrations, along with an upgrade of the auxiliary variables, are then used as input again (see Figure 2). The high-speed operation of the FEOM rests on two factors: (1) the interpolation of the modest number of low dimensional functions in eq 1 is very rapid, and (2) the time step δ can be taken as much larger than the standard stiff integration time step. In the illustration below, the FEOM time step was 1 day ($\delta = 1$ day) while the Gear solver required up to $\sim$2000 time steps for the integration over 1 day. Note that even though the FEOM takes one-day time steps, it still produces a response consistent with the full diurnal chemistry model.

TABLE 1: Table of the Species Included in the Chemical Mechanism, and the Number of First- and Second-Order Terms That Were Retained for Calculating Each Species

species	no. of 1st order terms	no. of 2nd order terms	species	no. of 1st order terms	no. of 2nd order terms
O	5	5	CH_3O	9	6
$O(^1D)$	2	1	CH_2O	15	15
O_2	0	0	HCO	13	17
O_3	18	12	HOCl	26	74
NO	23	52	CH_3OOH	25	75
NO_2	21	54	Cl	16	34
NO_3	17	13	ClO	24	76
N_2O_5	16	24	HCl	20	30
HNO_3	23	52	$ClONO_2$	26	74
N_2O	5	5	HO_2NO_2	23	52
H	19	21	CO_2	2	1
OH	13	7	BrO	22	38
HO_2	14	6	Br	22	38
H_2O	15	15	HBr	20	40
H_2O_2	25	50	$BrONO_2$	22	28
H_2	15	25	BrCl	25	50
CH_4	7	3	Cl_2O_2	20	55
CO	30	170	N_2	0	0
CH_3	20	55	OClO	10	55
CH_3O_2	30	170			

3. Illustration

3.1. Chemical Mechanism. The HDMR technique was applied to a 0-D stratospheric chemical kinetics model for an air parcel at 45° N latitude and 20 km altitude. The heterogeneous chemistry model has been adapted from the NASA GSFC 2-D atmospheric chemistry model[18,19] to run in 0-D. The model contains 39 species and 106 reactions. The chemical species and reactions are shown in Tables 1 and 2, respectively. The diurnal cycle was approximated with a square wave, and the seasonally dependent number of daylight hours was calculated based on the 45° N latitude. For a given 24-hour day, the photolysis rates (during the day light period) and the temperature are held constant. However, the photolysis rates and temperature vary on a seasonal basis. The seasonally varying photolysis rates for the 26 photolysis reactions were taken from the NASA GSFC 2-D model. The photolysis reactions were separated into five groups based on their seasonally varying rate's functional form (not shown here). This allowed five seasonally varying photolysis rate functions to prescribe all 26 photolysis rates at any time of the year. The photolysis reactions in each of the groups A, B, C, D, and E are identified in Table 2. The seasonally varying temperature was also taken from the NASA 2-D model. The temperature and photolysis rate functions are treated as HDMR variables discretized over their respective range of values. The total number of input variables is 46. This includes 39 chemical species, 5 photolysis rate functions, number of daylight hours, and temperature. The model produces a periodic annual seasonal cycle for most species. Chemical species without chemical sources (such as N_2O and CH_4) decay away and species without loss channels (such as HNO_3) build up. It must be noted that since this model does not include transport processes, the results differ from those of the NASA 2-D model. The extent of the input variable phase space was based on the extreme minimum and maximum values calculated in a multiyear simulation of the original model.

This model is utilized here to illustrate the HDMR technique. The HDMR will be evaluated for a simulation time step of 1 day. The model time step of 24 h is initialized at the day/night solar terminator and ends at the same point the next day. This allowed the evaluation of the chemical mechanism on a realistic time scale in which long-term atmospheric dynamics would not likely have a significant impact. Additionally, upon conversion of the HDMR into a FEOM, the 1 day time step would be ideal for incorporation into an integrated chemistry-transport climate model where seasonal variations and multiyear simulations are important. For investigations of local effects on short time scales (several hours to days) in the stratosphere or for investigations of tropospheric chemistry, shorter time steps would be required (viz., 15 min to 1 h). Although this has not been done, a similar procedure for generating a HDMR and subsequent FEOM could be followed. We expect that significant computational savings would also be achieved.

3.2. Identification of Key Species Variables. The stratospheric chemistry model described in section 3.1 was analyzed using the HDMR approach. Each of the first- and second-order expansion functions of eq 1 as evaluated for each output species to identify the key input variables and sets of variables that drive the chemistry and determine the model output concentrations. Although all 39 species were expressed in terms of their own HDMR expansions, for this illustration we will focus on analyzing O_3. A one-day time step was used for the HDMR as described in Section 3.1, and the results below are particular to the latter constraint.

There are 46 first-order expansion functions for each output species and two typical first-order expansion functions are shown in Figure 3. The functions describe the impact on O_3 due to NO_2 input and the impact on CH_2O due to the number of daylight hours. The vast majority of the expansion functions have modest curvature. This is consistent with the results of Chen, et al.[2] and Dubey et al.[5] There are some expansion functions with significant curvature, and this variation of behavior exemplifies the importance of the HDMR allowing for arbitrary nonlinearities, as well as the benefit of numerically representing the functions. However, in the present case, a low-order set of polynomials would suffice to represent and interpolate these functions and provide even further computational savings in the chemical solver described in the next subsection. The discrete values of the functions were found to all form smooth curves, except for some involving very small values. The noise in the latter cases was due to lack of sufficient precision from the Gear solver used to generate the functions.

The absolute magnitude of each of the first-order expansion functions was evaluated to identify which variables dominate the model output response of the O_3 concentration. The input variables are rank ordered in Table 3 by the absolute magnitude of the first-order expansion functions. Also shown in Table 3 is the rank ordering of the expansion functions for each of the input species that are important for predicting O_3 (i.e., chemical species listed in the first column of Table 3). Although the magnitude of some of the expansion functions suggests ignoring them, the small daily variation of some species can make these apparently small terms significant. The key species for determining the temporal evolution of O_3 in order of significance are O_3, NO_2, H_2O, $ClONO_2$, NO, BrO, N_2O_5, and HNO_3. Additionally, the photolysis rate functions A and B of Table 2 are important.

The input variables that operate cooperatively to impact the output are described by the second-order functions. The four largest magnitude functions that represent pairwise synergistic impact on the output O_3 concentrations are (1) the number of daylight hours coupled with the photolysis rate function A, (2) O_3 coupled with the photolysis rate function B, (3) O_3 coupled with number of daylight hours, and (4) the number of daylight hours coupled with the photolysis rate function *B*. As an illustration, the surface representing the impact on the O_3 output

TABLE 2: Reactions Included in the Chemical Mechanism[a]

(1) $O_2 \rightarrow O + O$	A	(35) $NO + O_3 \rightarrow NO_2 + O_2$
(2) $H_2O \rightarrow H + OH$	A	(36) $NO_2 + O_3 \rightarrow NO_3 + O_2$
(3) $CO_2 \rightarrow CO + O$	A	(37) $H + O_3 \rightarrow OH + O_2$
(4) $N_2O \rightarrow N_2 + O$	A	(38) $OH + OH + M \rightarrow H_2O_2 + M$
(5) $HCl \rightarrow H + Cl$	A	(39) $OH + ClONO_2 \rightarrow HOCl + NO_3$
(6) $HO_2 \rightarrow O + OH$	A	(40) $CH_4 + OH \rightarrow CH_3 + H_2O$
		(41) $CH_3O_2 + NO \rightarrow CH_3O + NO_2$
(7) $O_3 \rightarrow O_2 + O(^1D)$	B	(42) $CH_3O + O_2 \rightarrow CH_2O + HO_2$
(8) $HNO_3 \rightarrow NO_2 + OH$	B	(43) $OH + NO_2 + M \rightarrow HNO_3 + M$
(9) $HO_2NO_2 \rightarrow OH + NO_3$	B	(44) $HO_2 + HO_2 \rightarrow H_2O_2 + O_2$
(10) $HO_2NO_2 \rightarrow HO_2 + NO_2$	B	(45) $CH_2O + O \rightarrow HCO + OH$
		(46) $CH_3O_2 + HO_2 \rightarrow CH_3OOH + O_2$
(11) $NO_3 \rightarrow NO_2 + O$	C	(47) $Cl + H_2 \rightarrow HCl + H$
(12) $NO_3 \rightarrow NO + O_2$	C	(48) $Cl + O_3 \rightarrow ClO + O_2$
		(49) $ClO + O \rightarrow Cl + O_2$
(13) $H_2O_2 \rightarrow OH + OH$	D	(50) $Cl + CH_4 \rightarrow HCl + CH_3$
(14) $N_2O_5 \rightarrow NO_2 + NO_3$	D	(51) $HCl + OH \rightarrow Cl + H_2O$
(15) $CH_2O \rightarrow HCO + H$	D	(52) $ClO + NO \rightarrow Cl + NO_2$
(16) $CH_2O \rightarrow CO + H_2$	D	(53) $OH + H_2O_2 \rightarrow H_2O + HO_2$
(17) $CH_3OOH \rightarrow CH_3O + OH$	D	(54) $H_2 + OH \rightarrow H_2O + H$
(18) $ClONO_2 \rightarrow Cl + NO_3$	D	(55) $N_2O_5 + M \rightarrow NO_2 + NO_3 + M$
(19) $HOCl \rightarrow OH + Cl$	D	(56) $ClO + NO_2 + M \rightarrow ClONO_2 + M$
(20) $BrO \rightarrow Br + O$	D	(57) $O + H_2O_2 \rightarrow OH + HO_2$
(21) $BrONO_2 \rightarrow Br + NO_3$	D	(58) $HO_2 + NO_2 + M \rightarrow HO_2NO_2 + M$
(22) $Cl_2O_2 \rightarrow Cl + OClO$	D	(59) $O + ClONO_2 \rightarrow ClO + NO_3$
(23) $BrCl \rightarrow Br + Cl$	D	(60) $CO + OH \rightarrow CO_2 + H$
(24) $O_3 \rightarrow O_2 + O$	E	(61) $HNO_3 + OH \rightarrow NO_3 + H_2O$
(25) $NO_2 \rightarrow NO + O$	E	(62) $NO + HO_2 \rightarrow OH + NO_2$
(26) $OClO \rightarrow Cl + O_2$	E	(63) $H_2O + O(^1D) \rightarrow OH + OH$
(27) $O + O_2 + M \rightarrow O_3 + M$		(64) $OH + HO_2 \rightarrow H_2O + O_2$
(28) $O + O_3 \rightarrow O_2 + O_2$		(65) $OH + O \rightarrow H + O_2$
(29) $H + O_2 + M \rightarrow HO_2 + M$		(66) $HO_2 + O \rightarrow OH + O_2$
(30) $OH + O_3 \rightarrow HO_2 + O_2$		(67) $NO_2 + O \rightarrow NO + O_2$
(31) $HO_2 + O_3 \rightarrow OH + O_2 + O_2$		(68) $NO_2 + O + M \rightarrow NO_3 + M$
(32) $ClO + HO_2 \rightarrow HOCl + O_2$		(69) $N_2O + O(^1D) \rightarrow NO + NO$
(33) $Cl + H_2O_2 \rightarrow HCl + HO_2$		(70) $NO_2 + NO_3 + M \rightarrow N_2O_5 + M$
(34) $O(^1D) + M \rightarrow O + M$		(71) $H_2 + O(^1D) \rightarrow OH + H$
		(72) $CH_4 + O(^1D) \rightarrow CH_3 + OH$
		(73) $CH_3 + O_2 + M \rightarrow CH_3O_2 + M$
		(74) $CH_2O + OH \rightarrow H_2O + HCO$
		(75) $HCO + O_2 \rightarrow CO + HO_2$
		(76) $Cl + HO_2 \rightarrow HCl + O_2$
		(77) $OH + HO_2NO_2 \rightarrow H_2O + O_2 + NO_2$
		(78) $CH_4 + O(^1D) \rightarrow H_2 + CH_2O$
		(79) $OH + CH_3OOH \rightarrow H_2O + CH_3O_2$
		(80) $OH + OH \rightarrow H_2O + O$
		(81) $ClO + OH \rightarrow Cl + HO_2$
		(82) $HOCl + OH \rightarrow H_2O + ClO$
		(83) $Cl + CH_2O \rightarrow HCl + HCO$
		(84) $HO_2 + HO_2 + M \rightarrow H_2O_2 + O_2 + M$
		(85) $Cl + HO_2 \rightarrow OH + ClO$
		(86) $HO_2NO_2 + M \rightarrow HO_2 + NO_2 + M$
		(87) $H + HO_2 \rightarrow H_2 + O_2$
		(88) $H + HO_2 \rightarrow H_2O + O$
		(89) $H + HO_2 \rightarrow OH + OH$
		(90) $NO + NO_3 \rightarrow NO_2 + NO_2$
		(91) $NO + O + M \rightarrow NO_2 + M$
		(92) $N_2O + O(^1D) \rightarrow N_2 + O_2$
		(93) $N_2O_5 \rightarrow HNO_3 + HNO_3$
		(94) $O(^1D) + N_2 + M \rightarrow N_2O + M$
		(95) $O(^1D) \rightarrow O$
		(96) $O + O + M \rightarrow O_2 + M$
		(97) $Br + O_3 \rightarrow BrO + O_2$
		(98) $Br + HO_2 \rightarrow HBr + O_2$
		(99) $BrO + ClO \rightarrow Br + Cl + O_2$
		(100) $BrO + BrO \rightarrow Br + Br + O_2$
		(101) $OH + HBr \rightarrow H_2O + Br$
		(102) $BrO + NO_2 + M \rightarrow BrONO_2 + M$
		(103) $ClO + ClO + M \rightarrow Cl_2O_2 + M$
		(104) $BrO + ClO \rightarrow Br + OClO$
		(105) $BrO + ClO \rightarrow BrCl + O_2$
		(106) $Cl_2O_2 + M \rightarrow ClO + ClO$

[a] The photolysis rate functions are identified as A−E.

due to the input O_3 concentration coupled with the photolysis rate function B is shown in Figure 4.

3.3. FEOM: A Fast Chemistry Solver. The expansion functions calculated to identify the key input species have been assembled in eq 1 for each model output chemical species. The input variables to the chemistry FEOM are the 39 species concentrations, the number of daylight hours, temperature, and the five photolysis rate functions. The assemblage forms a FEOM chemical kinetics solver with a 1 day time step. The FEOM took species concentrations at time t_i and calculated the concentration of all species at $t_{i+1} = t_i + 1$ day. Then the FEOM was reapplied using the concentrations calculated for t_{i+1} to obtain values for $t_{i+2} = t_i + 2$ days, etc. By repeated application of the FEOM with updating of the ancillary parameters defined in section 2.2, the chemical evolution over the course of 30 years was evaluated (10 800 sequential FEOM applications). Implementation of the FEOM solver is shown schematically in Figure 2.

The FEOM output consists of 39 expansions, one for each output species. Each expansion in eq 1 was truncated after second order, as this was found to give quantitative results. Substantial numbers of the first- and second-order functions in the expansions were small in magnitude relative to f_0. This is due to the lack of importance of particular input variables, acting individually or cooperatively upon the final output. Since these functions would not significantly contribute to the final results, they were removed from the FEOM expansion. Their removal leads to an overall increase in the FEOM prediction speed, since fewer function operations are required. The number of first- and second-order terms retained for each output species is shown in Table 1. Generally, the variables significant at first order were also associated with the important second-order terms, but some exceptions did occur.

The quality and stability of the FEOM was assessed, respectively, by comparing its output with the original Gear-based model over 6 years and calculating a longtime FEOM simulation covering 30 model years. The 1-year temporal evolution of six of the 39 output species concentrations, calculated with 360 sequential FEOM operations and with a Gear-based solution, is shown in Figure 5. The species were chosen since they are important to the chemistry and also have a range of photochemical lifetimes. The quality of the results is typical of all the species. The agreement between the FEOM solver and the Gear integrator is very good. Both long- and short-time-scale species were accurately calculated. The agreement between the FEOM solver and Gear integrator for the rest of the 6 years is similar to the 1-year results. The ability of the FEOM solver to track small perturbations in the atmosphere (such as aircraft emissions or lightning-generated NO_x) was also investigated. It was found that both the FEOM solver and the Gear-based solver responded similarly to a test run in which a small constant flux of NO_2 added to the box model.

The temporal evolution of NO_2, O_3, and N_2O concentrations for 5 years is shown in Figure 6. This involved reapplying the FEOM 1800 times. The figure shows that the solution continues to be smooth and stable. The FEOM was continued to 30 years of simulation time, and the same degree of stability was found.

The computational savings with the FEOM was determined by observing the execution time of 1000 simulations, each for 1 day in length, and comparing the timing of the FEOM and

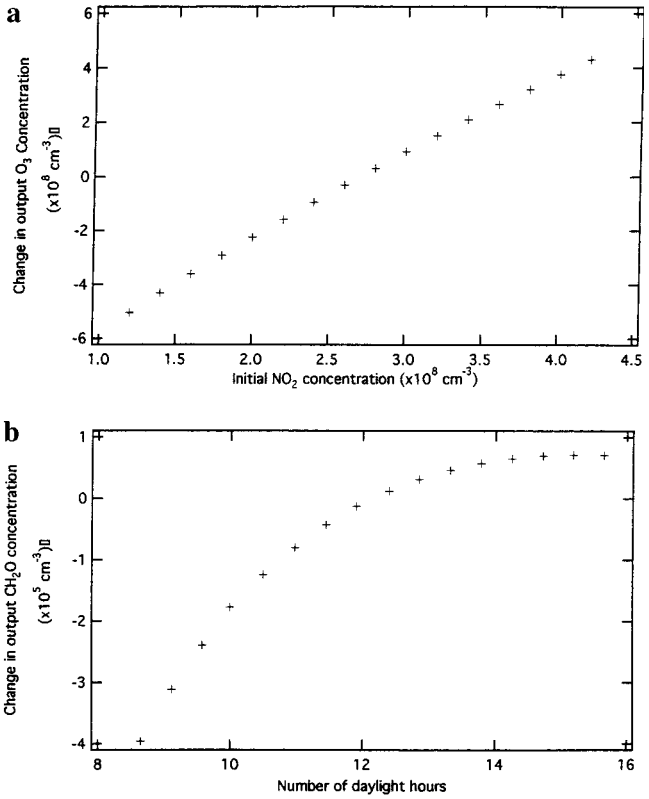

Figure 3. First-order expansion functions for (a) output O_3 with respect to input NO_2 and (b) output CH_2O with respect to input number of daylight hours.

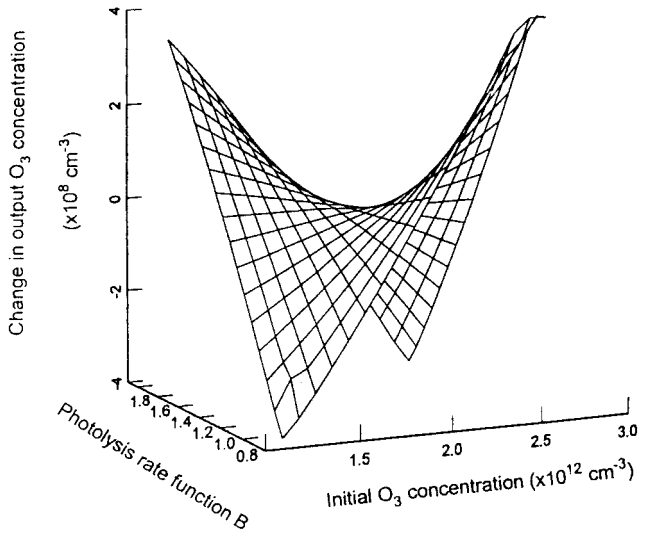

Figure 4. Second-order expansion functions describing the impact on output O_3 due to photolysis rate function *B* and initial O_3 concentration.

Gear solvers. The analysis was performed with random initial conditions on a PentiumPro based workstation. The Gear solver required 4.5 s/simulation, while the FEOM required 5×10^{-4} s/simulation. The 4.5 s/simulation with the Gear solver reflects the need for the FEOM to be generated with good quality input data. This acceleration is at the expense of some nominal one-time computational overhead to calculate the FEOM. The latter overhead is minimal if the chemistry FEOM was incorporated as part of a full chemistry-transport package.

At first glance, the computational scaling above suggests a FEOM savings of $\sim 10^4$ per run after generating the FEOM. However, there are several caveats that must be appreciated when models that calculate the same output using different

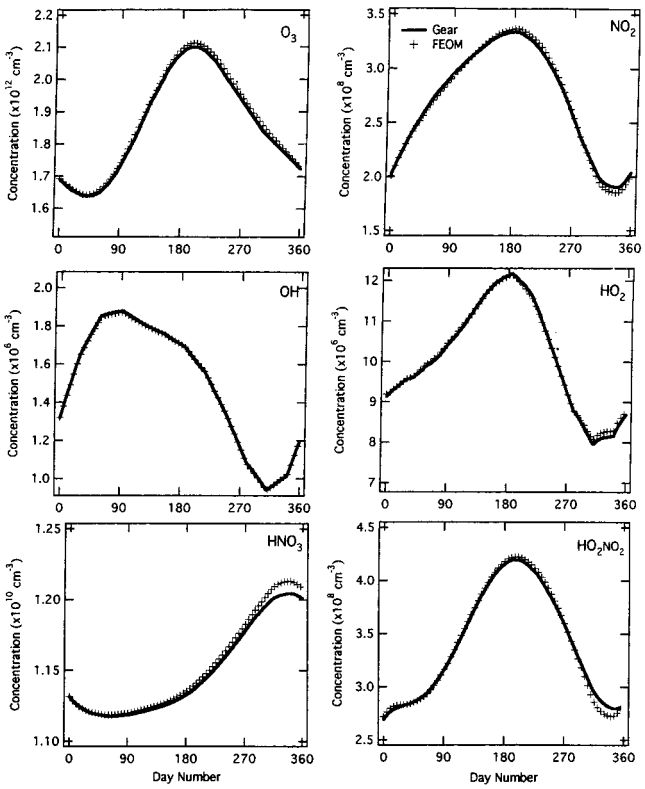

Figure 5. Comparison of six species in a 1-year chemistry calculation with a Gear-based solver (line) and a FEOM solver (+). The FEOM was constructed with first- and second-order expansion functions. The evident discrete aspects of some of the curves arise from switching between photolysis rates during the year. The FEOM calculation used 360 sequential 1-day calculations to make up the year simulation. Since the FEOM is implicitly mass conserving, the slight deviations did not cause the system to diverge over longer periods of time. This is demonstrated in Figure 5 with a 5-year simulation.

methods are compared. First is the selection of parameters for running the models. The Gear integrator has a number of tolerance parameters for output accuracy, and by varying the parameters, one can trade accuracy for speed. A tight tolerance in the Gear solver was used to generate the FEOM. This forced Gear to take ~2000 time steps to integrate the kinetics for 1 day. By diminishing the tolerances on Gear to the edge of safe and stable integration, the number of time steps was reduced to ~400 and the run time was reduced by a factor of 4. Even under these conditions, the FEOM was still $\sim 2 \times 10^3$ times faster than the Gear integrator. Note that 60% of the Gear time steps were taken to calculate the chemical evolution at the solar terminators (i.e., transition from light to dark and dark to light). Even if the number of steps in a modified Gear solver could be reduced to 20, which is not probable for safe numerical integration, and assuming that the run time scales with the number of steps, then the FEOM would still maintain a computational savings of $\sim 10^2$.

4. Conclusions

This paper describes the new high-dimensional model representation (HDMR) technique that (i) learns the model's nonlinear input−output relationships, (ii) identifies the key model input variables, and (iii) encapsulates the input−output relationship into an ultrafast fully equivalent operational model (FEOM) that can directly replace the original model. All of these capabilities result from expressing the HDMR as a rapidly convergent expansion in input variable cooperativity.

TABLE 3: First-Order Functions for O_3 and the Species That Contribute to O_3 Rank Ordered by the Largest Absolute Magnitude[a]

O_3	NO_2	$ClONO_2$	NO	BrO	N_2O_5	HNO_3
0.7486 O_3	0.3373 O_3	0.5847 $ClONO_2$	1.1394 O_3	0.4495 BrO	0.4251 N_2O_5	0.7145 HNO_3
0.0010 A	0.3063 NO_2	0.3266 O_3	0.3379 NO_2	0.2841 $BrONO_2$	0.3049 O_3	0.0127 N_2O_5
0.0004 NO_2	0.2592 NO	0.1008 ClO	0.2859 NO	0.1541 NO_2	0.2307 T	0.0115 NO_2
0.0003 H_2O	0.1528 N_2O_5	0.0434 NO_2	0.1673 N_2O_5	0.1307 O_3	0.2005 NO_2	0.0110 O_3
0.0003 B	0.0852 H_2O	0.0406 CH_4	0.0966 H_2O	0.1296 NO	0.1701 NO	0.0098 NO
0.0002 $ClONO_2$	0.0471 $ClONO_2$	0.0403 HCl	0.0497 B	0.0771 N_2O_5	0.1208 D	0.0076 H_2O
0.0002 NO	0.0445 HNO_3	0.0351 NO	0.0482 ClO	0.0398 H_2O	0.0362 $ClONO_2$	0.0033 CH_4
0.0001 BrO	0.0418 B	0.0293 HOCl	0.0464 HNO_3	0.0243 $ClONO_2$	0.0278 Day	0.0026 B
0.0001 N_2O_5	0.0397 T	0.0225 O_2	0.0441 T	0.0238 D	0.0187 ClO	0.0025 T
0.0001 HNO_3	0.0389 D	0.0192 N_2O_5	0.0419 E	0.0213 HNO_3	0.0087 BrO	0 .0020 $ClONO_2$

[a] The numbers in the columns are the relative magnitude of the first-order functions to the zeroth-order function.

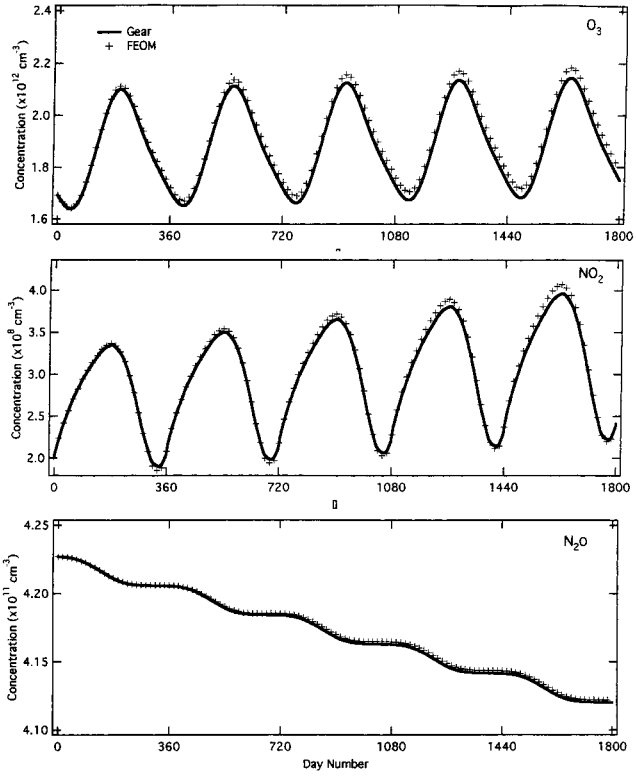

Figure 6. FEOM output concentrations for O_3, NO_2, and N_2O. The solid line is from the original solver, and the FEOM results are denoted by +.

The HDMR technique was demonstrated using a stratospheric chemistry box model. The input−output model relationships were revealed, and they quantitatively determined the key species driving the ozone concentrations on a 24 h basis. Also identified were the input variables that have important cooperative behavior to impact on the model output.

A fast fully equivalent operational model (FEOM) chemical kinetics solver was developed using the identified nonlinear independent and cooperative input−output model responses. This FEOM calculated chemical output concentrations comparable to a Gear integrator, but at least $\sim 10^2$ times faster for a stratospheric kinetics model. The FEOM solver was demonstrated to have sufficient sensitivity to accurately calculate the output effects when the chemical system is slightly perturbed (viz., aircraft exhaust in the atmosphere). Due to the unique mathematical structure of the underlying HDMR theory, the FEOM is implicitly mass conserving. This contributes to making FEOM chemical kinetic solvers inherently stable. The acceleration in the model run times attainable by a FEOM are dramatic and of considerable significance for kinetic modeling. Although the logic behind the kinetic FEOM is generic, its full scope of applicability needs further testing, including for demanding unstable and chaotic systems.

This work demonstrates the potential power of the HDMR technique for identifying key variables (and cooperating sets of variables) in large complex chemical mechanisms and for producing fast chemical kinetics solvers. It is also noted that the technique is generic in its capabilities and it can be applied to many types of models.

Acknowledgment. The DOE Atmospheric Chemistry Program supported this work. H.R. would also like to thank NASA for support.

References and Notes

(1) Rabitz, H.; Kramer, M.; Dacol, D. *Annu. Rev. Phys. Chem.* **1983**, *34*, 419.
(2) Chen, L.; Rabitz, H. A.; Considine, D.; Jackman, C. H.; Shorter, J. A. *J. Geophys. Res.* **1997**, *102*, 16201.
(3) Tatang, M. A. Direct Incorporation of Uncertainty in Chemical and Environmental Engineering Systems. Ph.D. Thesis, Massachusetts Institute of Technology, 1997.
(4) Vuilleumier, L.; Harley, R. A.; Brown, N. J. *Environ. Sci. Technol.* **1997**, *31*, 1206.
(5) Dubey, M. K.; Smith, G. P.; Hartley, W. S.; Kinnison, D. E.; Connell, P. S. *Geophys. Res. Lett.* **1997**, *24*, 2737.
(6) Gao, D.; Stockwell, W. R.; Milford, J. B. *J. Geophys. Res.* **1996**, *101*, 9107.
(7) Stolarski, R. S.; Douglas, A. R. *J. Geophys. Res.* **1986**, *91*, 7853.
(8) Ehhalt, D. H.; Chang, J. S.; Bulter, D. M. *J. Geophys. Res.* **1979**, *84*, 7889.
(9) Stewart, R. W.; Thompson, A. M. *J. Geophys. Res.* **1996**, *101*, 20953.
(10) Spivakovsky, C. M.; Wofsy, S. C.; Prather, M. J. *J. Geophys. Res.* **1990**, *95*, 18433.
(11) Turányi, T. *Comput. Chem.* **1994**, *18*, 45.
(12) Isukapalli, S.; Roy, A.; Georgopoulos, P. to be published.
(13) Sandilands, J. W.; McConnell, J. C. *J. Geophys. Res.* **1997**, *102*, 19073.
(14) Sandu, A.; Carmichael, G. R.; Potra, F. A. *Atmos. Environ.* **1997**, *31*, 475.
(15) Sobol, I. M. *Math. Model. Computational Exp.* **1993**, *1*, 403.
(16) Saltelli, A.; Sobol, I. M. *Reliab. Eng. Syst. Safety* **1995**, *50*, 225.
(17) Further mathematical details on HDMR are presented in this reference: Rabitz, H. A.; Alis, Ö. F.; Shim, K.; Shorter, J. A. *Comput. Phys. Commun.* **1999**, *117*, 11.
(18) Douglass, A. R.; Jackman, C. H.; Stolarski, R. S. *J. Geophys. Res.* **1989**, *94*, 9862.
(19) Jackman, C. H.; Douglass, A. R.; Rood, R. B.; McPeters, R. D.; Meade, P. E. *J. Geophys. Res.* **1990**, *95*, 7414.