Dimension reduction of non-equilibrium plasma kinetic models using principal component analysis

Dimension reduction of non-equilibrium plasma kinetic models using principal component analysis

Kim Peerenboom,[1,2] Alessandro Parente,[1] Tomáš Kozák,[3] Annemie Bogaerts,[3] and Gérard Degrez[1]

[1)] *Aero-Thermo-Mechanics Department, Université Libre de Bruxelles, Brussels, Belgium*

[2)] *Department of Applied Physics, Eindhoven University of Technology, Eindhoven, The Netherlands*

[3)] *Department of Chemistry, University of Antwerp, Antwerp, Belgium*

The chemical complexity of non-equilibrium plasmas poses a challenge for plasma modeling because of the computational load. This paper presents a dimension reduction method for such chemically complex plasmas based on principal component analysis. Principal component analysis (PCA) is used to identify a low-dimensional manifold in chemical state space that is described by a small number of parameters: the principal components. Reduction is obtained since continuity equations only need to be solved for these principal components and not for all the species. Application of the presented method to a CO_2 plasma model including state-to-state vibrational kinetics of CO_2 and CO demonstrates the potential of the PCA method for dimension reduction. A manifold described by only two principal components is able to predict the CO_2 to CO conversion at varying ionization degrees very accurately.

I. INTRODUCTION

Many plasma applications exploit the rich chemistry that non-equilibrium plasmas provide. Examples include biomedical plasmas[1–3] and plasmas for environmental applications[4–6]. Non-equilibrium plasmas are very effective in creating chemically reactive species such as radicals and electronically and vibrationally excited states, even at low temperatures.

This chemical complexity of non-equilibrium plasmas poses a challenge for plasma modeling. Not only do many species have to be taken into account, often also the internal states have to be dealt with by a state-to-state approach because of the non-equilibrium character of the plasma. Furthermore, many different time scales are involved in the chemical reactions, resulting in a stiff system of equations. These issues cause the computational load to become prohibitive for spatially resolved simulations of chemically complex plasmas.

Clearly, there is a need for dimension reduction techniques to keep simulations manageable in the plasma physics community. Roughly speaking, there are two main strategies for reducing the dimensionality: mechanism reduction and reparametrization of the chemical state space. An example of mechanism reduction is grouping of energy levels in state to state kinetic models[7,8]. Although developed as an analysis tool, the algorithm of Lehmann to determine principal reaction pathways[9,10] can also be viewed as an example of mechanism reduction.

Despite the impressive results obtained in combustion simulations[11–13], the second approach of reparametrization of the chemical state space has not yet found its way to plasma physics modeling. It is the most common approach in combustion and is based on the fact that the composition lies on a so-called low dimensional manifold. Due to the existence of this low dimensional manifold, it is possible to describe the composition with only a small number of parameters. Successful methods using this approach are for example Flamelet Generated Manifolds[11] and Principal Component Analysis (PCA)[14–22]. Other approaches have been proposed to deal with highly non-linear systems, based on non-linear component analysis in combination with diffusion maps[23].

This article presents the application of PCA to non-equilibrium plasma models. PCA is a statistical procedure to reduce the dimensionality of datasets by transforming a large number of correlated variables into a smaller number of uncorrelated variables called principal components (PCs). As a case study, a 0D state-to-state kinetic model of CO_2 including

vibrational levels of CO_2 and CO is used. The original variables (species densities) are projected onto the principal component basis. Reduction is achieved by solving continuity equations for only a few principal components. To our knowledge it is the first time that this approach is applied to a plasma model. Only very recently[12], this approach was demonstrated for a combustion simulation. To tailor the approach of solving principal component continuity equations to plasmas, a log transformation is introduced to deal with the exponential behavior of the vibrationally excited states, which has never been considered in prior applications of the method. Another novel contribution is the use of tabulation within PCA, to improve the accuracy of the reconstruction.

This article is organized as follows. The CO_2 plasma kinetics model is presented in section II. Subsequently, the theory of principal component analysis is introduced in section III. Also the principal component continuity equations are derived in this section. In section IV, results of the application of PCA to the CO_2 plasma model are presented. Firstly, the influence of log transformation, scaling and retrieving the original variables from a lookup table is studied through an a priori analysis of the dataset. We will also briefly discuss the physical interpretation of the principal components. Secondly, the principal component continuity equations are solved to demonstrate the ability to reduce non-equilibrium plasma simulations. Finally, conclusions are drawn in section V.

II. CO_2 PLASMA MODEL

A model was developed to simulate the state-to-state vibrational kinetics of CO_2 in a non-equilibrium plasma, with special attention to the dissociation of CO_2 due to the excitation of its vibrational levels. The model can be characterized as a zero-dimensional model, which means that the time evolution of the densities of the plasma species in a fixed volume of plasma is calculated. Mathematically, it is expressed by a set of coupled ordinary differential equations of the form

$$\frac{\mathrm{d}n_i}{\mathrm{d}t} = \sum_j \left[(a_{ij}^{\mathrm{R}} - a_{ij}^{\mathrm{L}}) k_j \prod_l n_l^{\mathrm{L}} \right], \tag{1}$$

where n_i is the density of species i, a_{ij}^{R} and a_{ij}^{L} are the right-hand side and left-hand side stoichiometric coefficients of species i in the reaction j, k_j is the reaction rate constant and n_l^{L} is the density of the lth reactant of reactions j. These equations are also coupled with the

Neutral ground states	CO_2, CO, O_2, O_3, O, C_2O, C, C_2
Vibrational levels	$CO_2(25)$, CO(10), $O_2(4)$
Electronic states	$CO_2(2)$, CO(4), $O_2(2)$
Charged species	CO_2^+, e

TABLE I. Summary of all species included in the model. The numbers between brackets indicate the number of excited levels taken into account.

Boltzmann equation which is solved for the electron energy distribution function (EEDF). Energy is supplied to the plasma electrons by an external electric field (specified by the reduced electric field E/N and by the reduced frequency ω/N, in case of an AC field, where $N = \sum n_i$ is the total gas density). We use an existing code ZDPlasKin[24] which features an interface for description of the plasma species and reactions, a solver for the set of differential equations and a Boltzmann equation solver BOLSIG+[25].

The species and reactions used in this work are based on our previous work[6]. We have found that the dissociation of CO_2 is strongly influenced by the population of highly excited vibrational levels, which are determined mainly by electron impact vibrational excitation and by the VT and VV relaxation of the CO_2 vibrational levels. In the present work we focus in particular on the kinetics of CO_2 vibrational levels and on the applicability of PCA for reduction of the computation load. Compared to the previous model[6], we have reduced the reaction chemistry by omitting all charged species (except for electrons and CO_2^+ ions), which play a minor role in the kinetics of the neutral species and we simply set the electron density to a given value.

The species included in the model are summarized in table I. For CO_2, CO and O_2, we have included several vibrational levels and electronically excited levels. The vibrational levels of CO_2 used in the model are shown in figure 1. We use four lowest effective symmetric mode levels (denoted by letters) and 21 asymmetric mode levels (denoted by numbers) up to the dissociation limit of the molecule. Moreover, we use 10 vibrational levels of CO and 4 vibrational levels of O_2.

The reactions used in the model are of three types: electron impact reactions, heavy particle reactions involving change of vibrational energy and heavy particle reactions leading to breaking and formation of new species, see tables IV, V and VI, respectively. In the next paragraphs, we will have a closer look at the reactions used, however, for even more detailed

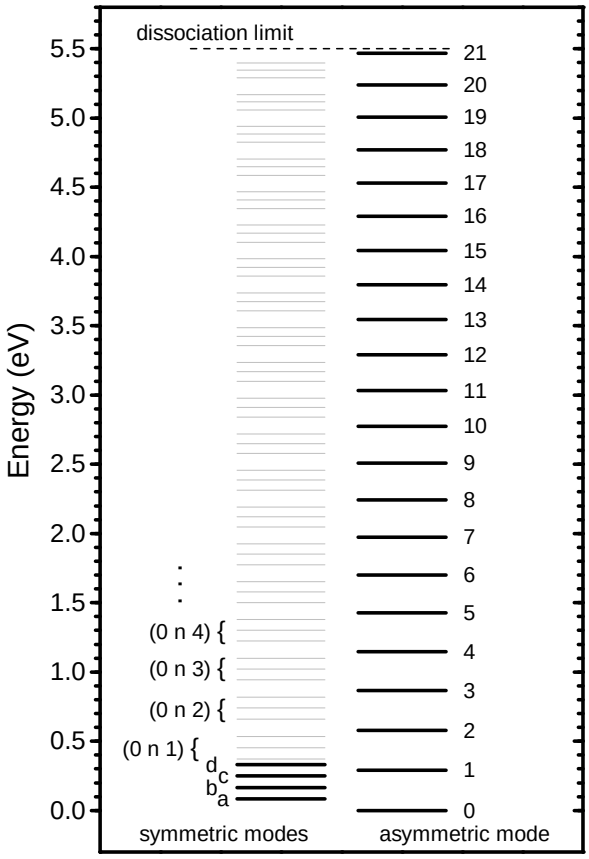


FIG. 1. Effective energy levels of CO_2 included in the model (black lines). The asymmetric mode levels $(0,0^0,v_3)$ where $v_3 = 0 \ldots 21$ are denoted by the numbers v_3. The levels $(v_1,v_2^{l_2},v_3)$ satisfying $n = v_2 + 2v_1 > 0$ for a given v_3 are grouped to effective symmetric mode levels and labeled $(0,n,v_3)$ in the figure. The four lowest symmetric mode levels (with $v_3 = 0$) are included in the model and denoted by letters $a \ldots d$, corresponding to $n = 1 \ldots 4$, respectively.

information about the reactions and the rate constants used, see[6].

The electron impact reactions (table IV) are given in the form of a cross section database that is passed to the Boltzmann equation solver. Because we have omitted the charged species, the reactions which result in the formation of ions are not taken into account in the kinetic part of the model, nevertheless, they have to be included in the cross section database in order to calculate the EEDF correctly.

The reactions in tables V and VI are given in the form of temperature-dependent rate constants. The vibrational energy exchange reactions in table V include three types of reactions: VT relaxation (V1 – V4), intermode VV relaxation which represents vibrational

energy exchange among the symmetric and asymmetric mode of CO_2 (V5) and VV relaxation between CO_2 and CO molecules (V6 – V8). VT relaxation is the main process responsible for the decrease of vibrational energy stored in the molecules. The VV relaxation among CO_2 levels (V6) is almost resonant, i.e. little vibrational energy is lost to translation, and is responsible for redistribution of the vibrational energy in the asymmetric mode. This reaction leads to the population of high-energy levels (Treanor distribution)[26]. For all reactions, only single-quantum transitions are taken into account.

Vibrational excitation can also have a significant effect on the chemical reactions summarized in table VI. This is modeled by the Fridman–Macheret α-model[27,28] which assumes that the activation energy of a reaction is lowered by the vibrational energy of the reacting species multiplied by an effectiveness parameter α. This parameter is assumed to be 1 for strongly endothermic reactions, such as (N1) and (N9), and 0.5 for thermoneutral reactions such as (N3) and (N4). Therefore, an excited molecule of CO_2 with vibrational energy close to the dissociation limit is easily dissociated through reactions (N1) and (N3), which are the most efficient channels of CO_2 dissociation in non-equilibrium plasmas[29]. Therefore, as we have already mentioned earlier, an accurate calculation of dissociation requires a calculation of the vibrational distribution function of CO_2, which is a result of many interactions between the individual vibrational levels (table V).

Altogether the model includes 57 species and 1683 individual reactions, most of the reactions come from the state-to-state interactions of vibrational levels. It is clear that this amount of species and reactions in a spatially resolved simulation can not be handled in a reasonable time.

The calculations performed to obtain data for the PCA were done in the following way. We assumed constant gas pressure, gas temperature, electron density and electric field intensity during the whole simulation time. This represents a gas of constant temperature flowing through a homogeneous plasma, which can be used, for example, as a first approximation of a microwave surfaguide discharge[30]. These simplifications do not at all limit the applicability of the PCA which could be used in the same way even when the electric field, electron density or gas temperature changes.

In this case study, the values of the plasma parameters are also not critically important. We used values lying in the typical range of parameters used in the microwave discharge experiments[28], i.e. a pressure of 100 Torr, a reduced electric field of 50 Td with a frequency

of 2.45 GHz and a gas temperature of 300 K. We performed several runs with different values of the ionization degree (electron density) which ranged from 10^{-7} to 10^{-5}. The simulation time varied with the ionization degree so that the specific energy input was 1 eV/molecule in all calculations. It was found out in our calculations that this value is optimal for obtaining high energy efficiency of the CO_2 conversion under the above mentioned conditions.

III. PRINCIPAL COMPONENT ANALYSIS

At any point in time and space, the composition in a plasma is fully determined by a set of chemical state space variables. Throughout this article, the species densities are used as the chemical state space variables, but the mole or mass fractions can be used just as well. The chemical state space is the space spanning all possible compositions and its dimension is equal to the number of species, hereafter denoted as Q. Due to convection, diffusion and chemical reactions, the composition changes in time and space. These changes can be seen as trajectories in the chemical state space. Reduction techniques rely on the fact that the composition relaxes quickly to a low-dimensional manifold due to equilibration driven by fast processes. This low-dimensional manifold is a lower dimensional space (with dimension $< Q$) inside the chemical state space. The task of PCA now is to search for a new set of variables that optimally describe the manifold.

In figure 2, a schematic overview of the PCA method can be seen. The method starts with generation of training data. The idea of PCA is that the training data can be generated with representative, inexpensive 0D or 1D models. The results of these inexpensive models are used to define the low-dimensional manifold, which can then be used to reduce expensive 2D or 3D models. This approach was successfully demonstrated for a 2D syngas jet flame[12].

The training data is stored in a matrix $\mathbf{X_0}$ $(n \times Q)$, containing n observations of Q variables. In this work, the variables are the species densities and the observations correspond to different time steps of the CO_2 plasma kinetics model as presented in section II. As a next step, the data must be scaled and centered[31] to improve the performance of PCA:

$$\mathbf{X} = \left(\mathbf{X_0} - \bar{\mathbf{X}}_\mathbf{0}\right) \mathbf{D}^{-1}, \tag{2}$$

where $\mathbf{X}$ contains the scaled data, $\bar{\mathbf{X}}_\mathbf{0}$ contains the mean of each variable and $\mathbf{D}$ is a diagonal matrix that contains the scaling coefficients of all the variables. Different scaling methods

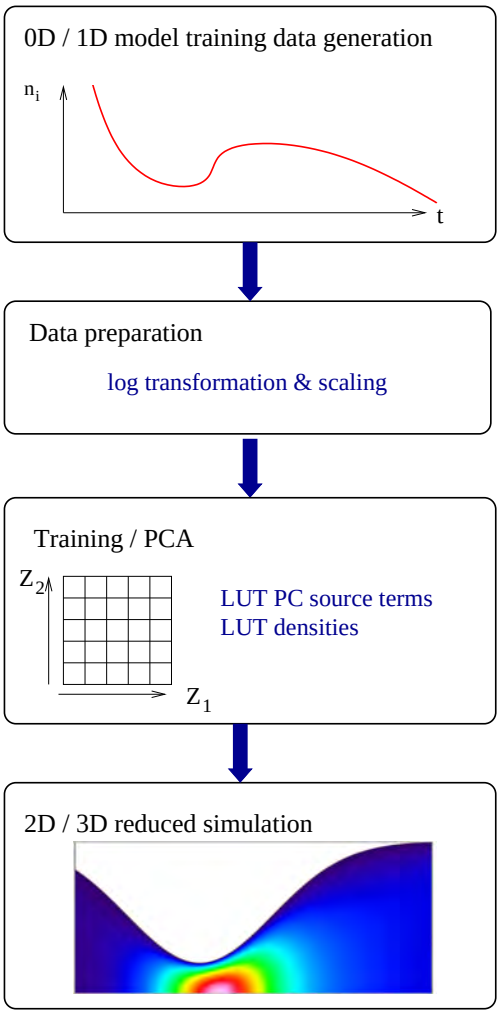


FIG. 2. Schematic overview of dimension reduction using PCA. In the figure 'LUT' is used as an abbreviation for lookup table.

can be used, as detailed in the table below. The best scaling method is the one giving the smoothest manifold. This can be judged visually looking at the smoothness of the manifold after PCA has been performed. Another option is to calculate a so-called smoothness parameter which gives a quantification of the non-linearity of the manifold[12].

After the centering and scaling the covariance matrix $\mathbf{C_X}$ is calculated:

$$\mathbf{C_X} = \frac{1}{n-1}\mathbf{X}^T\mathbf{X}. \tag{3}$$

The covariance matrix measures the correlation between the different species. The principal components, which are the parametrization variables of the manifold, are calculated as the

Method	Scaling coefficient
auto	$D_{ii} = s_i$
pareto	$D_{ii} = \sqrt{s_i}$
vast	$D_{ii} = \frac{s_i^2}{\bar{X}_{0,i}}$
range	$D_{ii} = \max\left(X_{0,i} - \bar{X}_{0,i}\right) - \min\left(X_{0,i} - \bar{X}_{0,i}\right)$
level	$D_{ii} = \bar{X}_{0,i}$

TABLE II. Different scaling methods[31]. The symbol s_i refers to the standard deviation of variable i.

eigenvectors $\mathbf{A}$ of the covariance matrix:

$$\mathbf{C_X} = \mathbf{ALA}^T. \tag{4}$$

Note that the principal components are linear combinations of all the original variables. Also remark that the eigenvalues in the diagonal matrix $\mathbf{L}$, which are a measure of the variance, are ordered in decreasing order. As a result, the first principal components account for most of the variation in the dataset. The original dataset can now be projected onto the principal component basis:

$$\mathbf{Z} = \mathbf{XA}, \tag{5}$$

where $\mathbf{Z}$ contains the so called principal component scores. Likewise, the original dataset can also be recovered from the principal component scores:

$$\mathbf{X} = \mathbf{ZA}^{-1}, \tag{6}$$

where $\mathbf{A}^{-1} = \mathbf{A}^T$. The fact that the first principal components account for most of the variation can be used for dimension reduction. This reduction can be achieved by approximating $\mathbf{X}$ with the first q ($q < Q$) principal components:

$$\mathbf{X} \approx \mathbf{X_q} = \mathbf{Z_q}{\mathbf{A_q}}^T, \tag{7}$$

where $\mathbf{A_q}$ is the truncated version of $\mathbf{A}$.

In[15], the idea was proposed to derive continuity equations for the principal components from the species continuity equations. Although the final goal is to use continuity equations for the principal components in 2D or 3D simulations as illustrated in figure 2, we will stick to

0D training data and 0D principal component continuity equations in this initial study. This is necessary to demonstrate the feasibility and reduction potential of the proposed approach for plasma simulations. Furthermore, it allows to focus on the effect of the combination of the log transformation and the tabulation. To derive the 0D continuity equations for the PCs, we start from the coupled set of ODE's describing the time evolution of the species:

$$\frac{\partial n_i}{\partial t} = S_{n_i}, \tag{8}$$

where S_{n_i} is the source term for the densities. After subtracting the species mean $\overline{n}_i$ and dividing by the scaling factor D_{ii}, we get:

$$\frac{\partial \left(\frac{n_i - \overline{n}_i}{D_{ii}} \right)}{\partial t} = \frac{S_{n_i}}{D_{ii}}. \tag{9}$$

Multiplication of the equations with the ith component of the kth PC a_{ik} and summation over all the species gives:

$$\frac{\partial \sum_i a_{ik} \left(\frac{n_i - \overline{n}_i}{D_{ii}} \right)}{\partial t} = \sum_i a_{ik} \frac{S_{n_i}}{D_{ii}}. \tag{10}$$

As $\sum_i a_{ik} \left(\frac{n_i - \overline{n}_i}{D_{ii}} \right)$ is the definition of the PC score Z_k, equation (10), can be rewritten as:

$$\frac{\partial Z_k}{\partial t} = S_{Z_k}, \tag{11}$$

where S_{Z_k} is the source term for Z_k, defined as:

$$S_{Z_k} = \sum_i a_{ik} \frac{S_{n_i}}{D_{ii}}. \tag{12}$$

Alternatively, when a log transformation is applied before carrying out PCA, there is an extra factor n_i in the denominator and the principal component source terms read:

$$S_{Z_k} = \sum_i a_{ik} \frac{S_{n_i}}{D_{ii} n_i}. \tag{13}$$

In principle, one could use these expressions to calculate the source terms for the principle components. However, it has been shown[21] in combustion simulations that the reconstruction errors of the densities strongly affect the accuracy of the source terms. Due to the approximation of the densities by a limited number of principal components using equation (7), a small error is made. However, this small error will propagate and increase because of the strong non-linearity of the source terms. As a result, the source terms soon become too

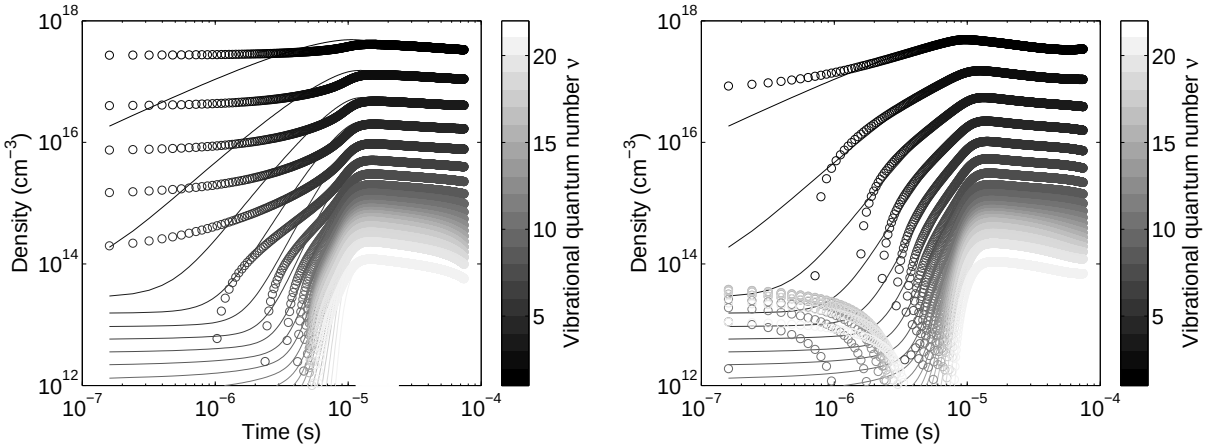
inaccurate when the number of retained PCs is decreased. Therefore, we use linear interpolation from a lookup table to recover the principal component source terms and species densities, as will be further discussed in section IV A 2.

IV. RESULTS

A. A priori PCA

1. Log transformation

A characteristic of non-equilibrium plasmas is that they can enhance chemical reactivity by electronic or vibrational excitation of the reacting species. In equilibrium, the population of such excited states is distributed according to a Boltzmann distribution where the ratio between the populations of two states is proportional to the exponent of the energy difference between the two levels. Due to this exponential behavior, different excited states can have very different populations. Also outside equilibrium, the population of the different energy levels varies over orders of magnitude. Moreover, the population of one excited state can also vary wildly in space and time due to gradients in for example the electron temperature. These variations of orders of magnitude pose a serious challenge to the performance of principal component analysis as can be seen in figure 3. This figure shows the a priori

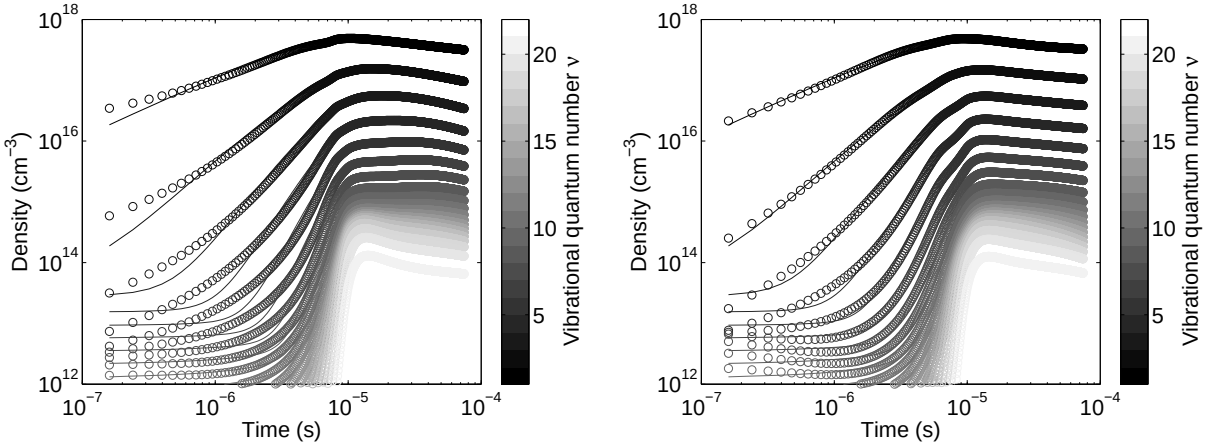


(a)5 PCs (b)10 PCs

FIG. 3. A priori reconstruction of the time evolution of the CO_2 asymmetric stretch mode levels with 'range' scaling and without log transformation.

reduction of the training dataset by equation (7) using 5 and 10 principal components. The solid lines represent the original training data, whereas the circles represent the a priori reconstruction using equation (7). The color indicates the vibrational level. Whereas the reconstruction is accurate at later times, it fails for the initial phase and even generates negative densities. This can be seen as circles running off the figure. Especially the lower species concentrations are affected by the errors. It can be seen that increasing the number of retained principal components improves the result, but still the errors are unsatisfactory.

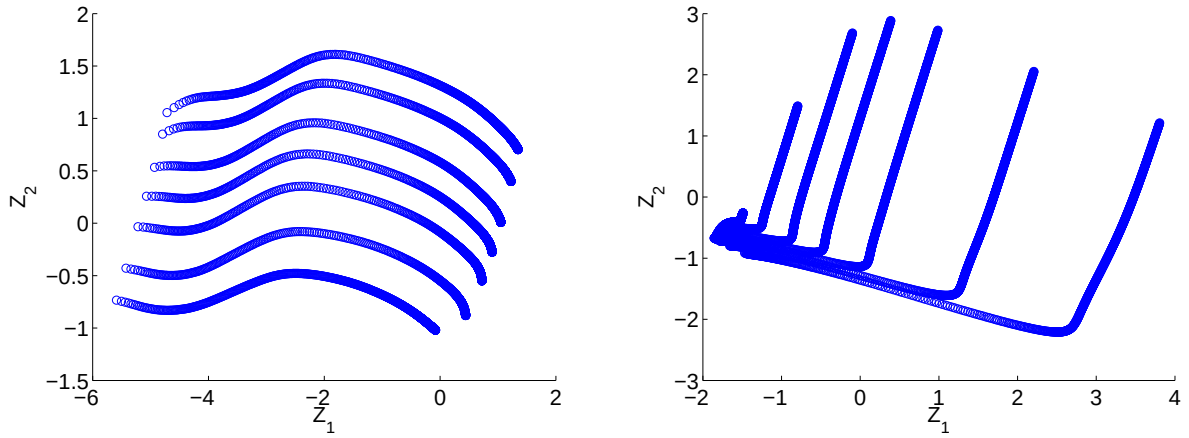
To decrease the dynamic range of the species concentrations, a log transformation is performed on the original dataset. The result can be seen in figure 4. The decrease in the



(a)5 PCs log (b)10 PCs log

FIG. 4. A priori reconstruction of the time evolution of the CO_2 asymmetric stretch mode levels with 'range' scaling and log transformation.

dynamic range greatly improves the performance of principal components analysis. The results remain physical and the reconstruction captures much better the lower species concentrations. The log transformation also has a beneficial effect on the manifold shape as can be seen in figure 5. The manifold shows the time dependent reaction trajectories of the plasma kinetics model at seven different ionization degrees. The smoothness and uniqueness of the manifold are crucial for the accuracy of the data reconstruction. Uniqueness is needed to be able to assign a single value of the species densities to each set of values of the first two principal components. Furthermore, a small change in the principal component values should ideally lead to a small change in the species densities for accurate data reconstruction. The less gradients in the manifold, the better. When looking at the manifold, the



(a) (b)

FIG. 5. Comparison of the manifold shape with (a) and without (b) log transformation. The 'range' scaling method is used. The manifold shows the time dependent reaction trajectories of the plasma kinetics model at seven different ionization degrees.

reaction trajectories should therefore be well separated and non-overlapping. With the log transformation all the points on the manifold can be uniquely described as a function of the two principal components Z_1 and Z_2. Without the log transformation this is not the case and multiple observations can be associated with single values of Z_1 and Z_2. As a result, without the log transformation more principal components are needed to uniquely describe the chemical state space.

2. Reconstruction and interpolation

The log transformation greatly improves the results of the reconstruction of the species densities. Still, errors are visible when ten principal components are retained. It is known from combustion simulations that these reconstruction errors strongly affect the accuracy of the calculation of the source terms needed in the principal component continuity equations. The reason is that PCA is a linear method while the non-equilibrium processes in plasmas are highly non-linear. In the current plasma study, the problem is even bigger since the log transformation requires to take the exponent to recover the species densities. This causes a blow up of the error of the principal component source terms when not all principal components are retained. Only very recently, a solution for this problem has been proposed

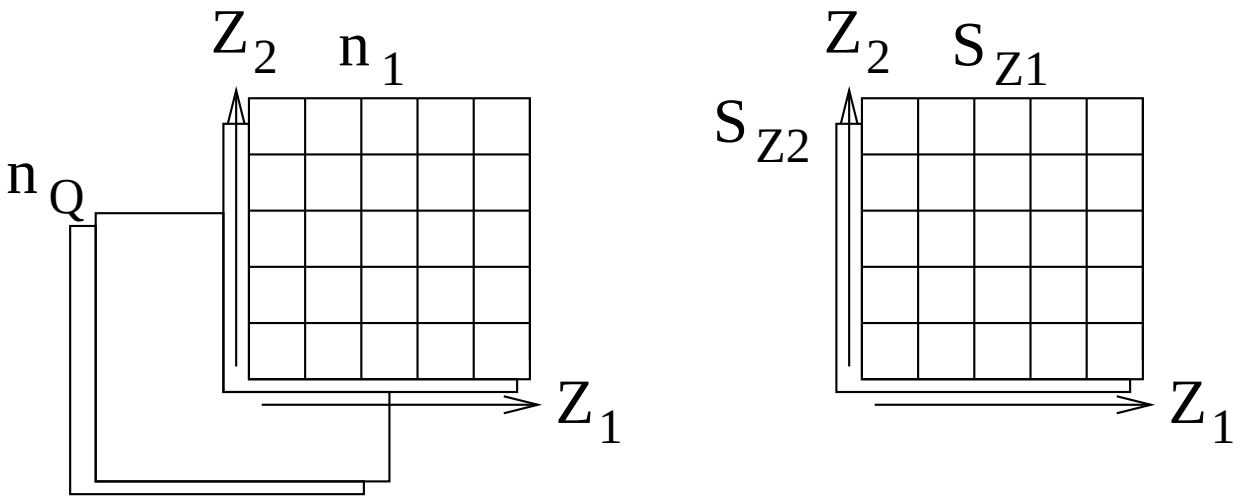


FIG. 6. 2D lookup tables for the densities and the principal component source terms. The lookup tables for the principal component source terms are only used in the a posteriori study.

in the form of non-linear regression[12,21]. In this case the densities are not recovered using relation (7) but non-linear regression techniques are used to recover all the state space variables including the principal component source terms. As an alternative to non-linear regression we use linear interpolation from 2D lookup tables to recover all the state space variables (densities in our case) and principal component source terms in the current study. In these lookup tables the densities and principal components source terms are stored as a function of Z_1 and Z_2, as depicted in figure 6.

A comparison between the reconstruction using relation (7) and interpolation can be seen in figure 7. Interpolation clearly outperforms the reconstruction using relation (7), when 2 principal components are retained. This is also quantified in table III. Figure 8 displays the eigenvalue magnitudes and the variance explained by the first two principal components. From this figure, it could indeed be seen that two principal components are sufficient to explain more than 95% of the variance of the data.

3. Scaling

The importance of scaling of the data for combustion simulations has been outlined in[31]. Scaling of the data also has a large influence on the shape of the manifold in the case of plasmas, as can be seen from figure 9. Except for the 'vast' scaling method, all scaling

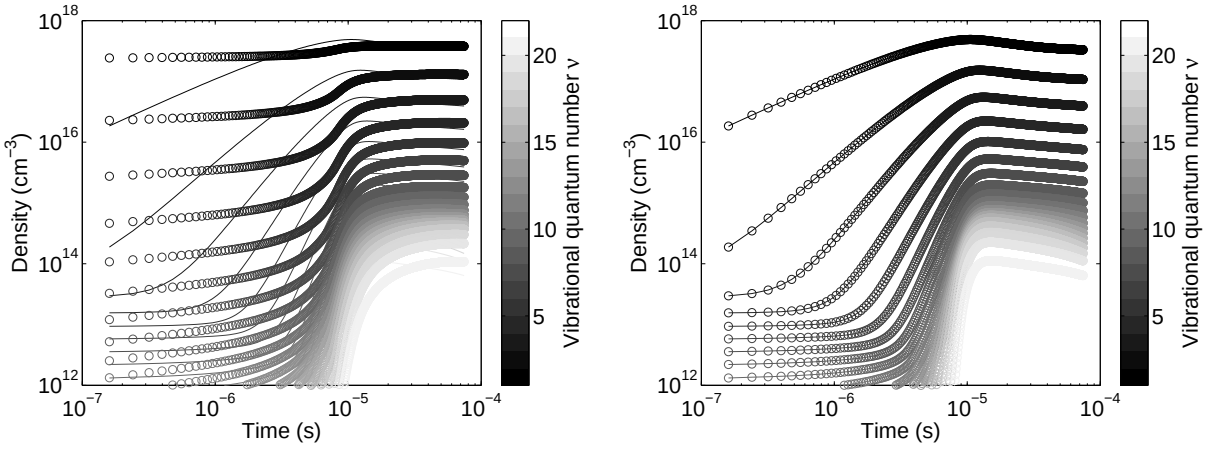


FIG. 7. A priori recovering of the time evolution of the CO_2 asymmetric stretch mode levels using two principal components using a) relation (7) and b) interpolation from a lookup table. Before performing PCA a log transformation is applied to the original data. Scaling method: 'range'.

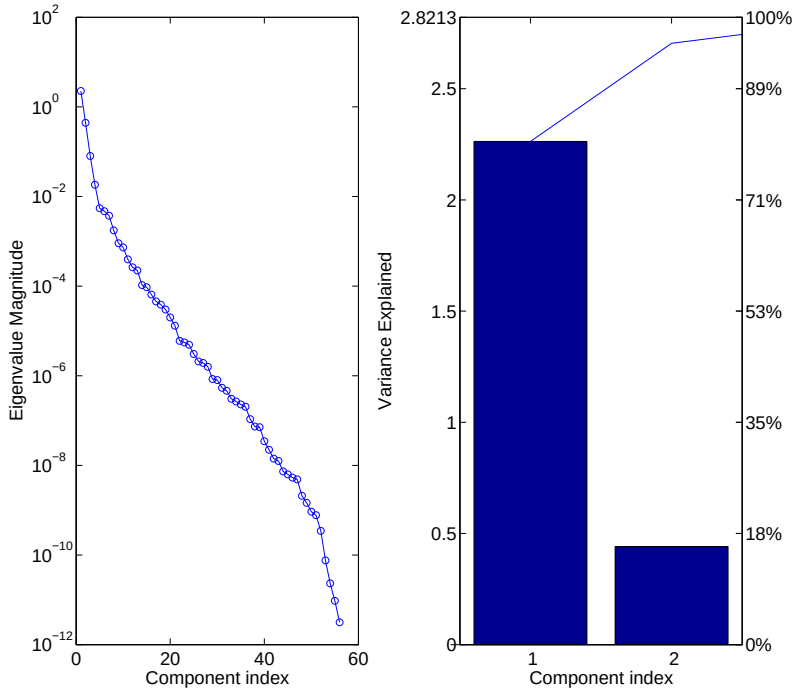


FIG. 8. Eigenvalue magnitudes and variance. Log transformation and the 'range' scaling method are used.

TABLE III. R^2 statistics and normalized root mean square (NRMS) error for the log of the species densities comparing reconstruction and interpolation from a lookup table. The 'range' scaling method and a log transformation are used.

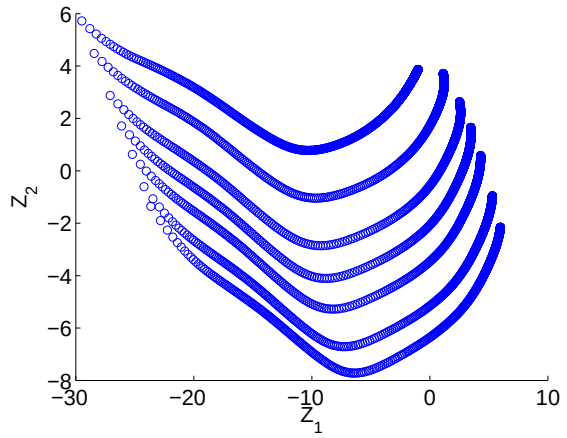
species	R^2 reconstruction	NRMS error reconstruction	R^2 interpolation	NRMS error interpolation
CO_2	0.9255	0.0001	1	3.177e-18
CO	0.9889	0.0012	1	3.220e-18
O_2	0.9984	0.0008	1	3.189e-18
O_3	0.9596	0.0073	1	3.074e-18

methods give a unique mapping using two principal components and a log transformation. It seems that for the dataset used the 'range' scaling method performs best. The reaction trajectories of the training set are most smoothly projected onto the principal component basis. The 'range' scaling gives the clearest interpretation of the manifold. Looking at the manifold in the case of 'range' scaling, it seems that Z_2 corresponds to the ionization degree (different reaction trajectories of the training set) and Z_1 to the progress of the reaction within one reaction trajectory. To check this physical interpretation, scatter plots of Z_1 and Z_2 against the log of the CO ground state density and the electron density are presented in figure 10. It can be seen that Z_1 is strongly correlated with the ground state CO density. The amount of CO produced is a measure of the reaction progress in the CO_2 to CO conversion process. Component Z_2 also has a smaller but still clear correlation with the electron density.

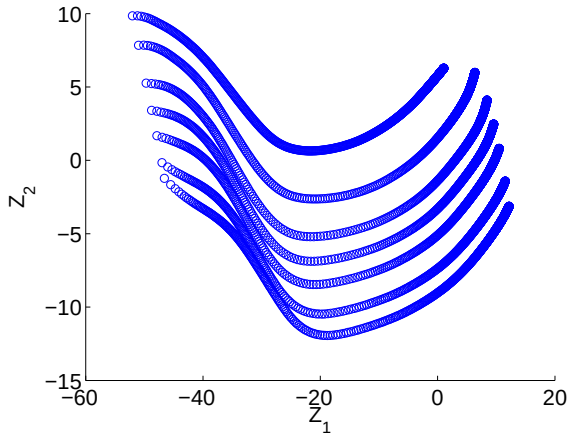
B. A posteriori PCA

1. Verification

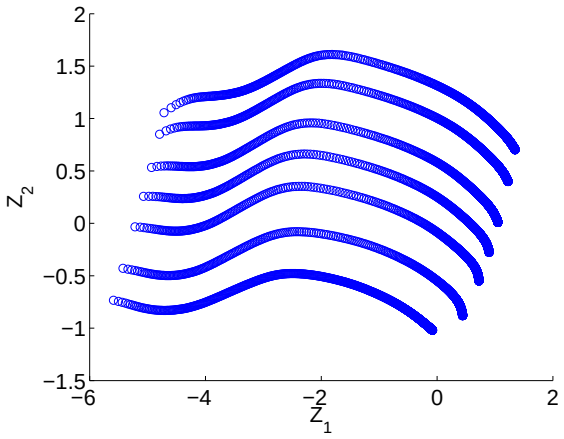
To demonstrate the suitability of PCA to carry out reduced simulations, time integration of the principal component continuity equations is performed at an ionization degree of 1e-6, which is one of the training data sets. Log transformation, tabulation and 'range' scaling are used for the reduced simulation. Both the principal component source terms and the densities are recovered by linear interpolation from a lookup table parametrized by Z_1 and Z_2. The principal component source terms are recovered every time step, while the species



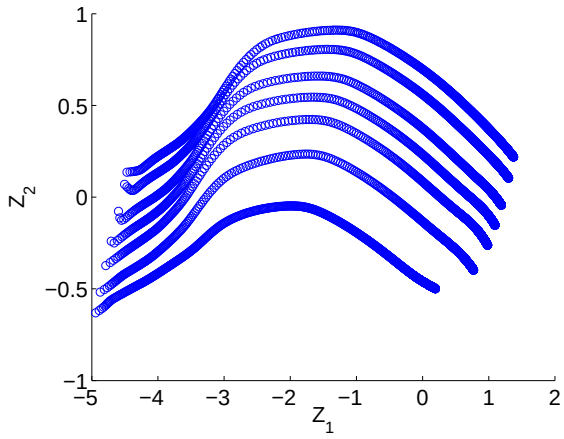
(a)auto



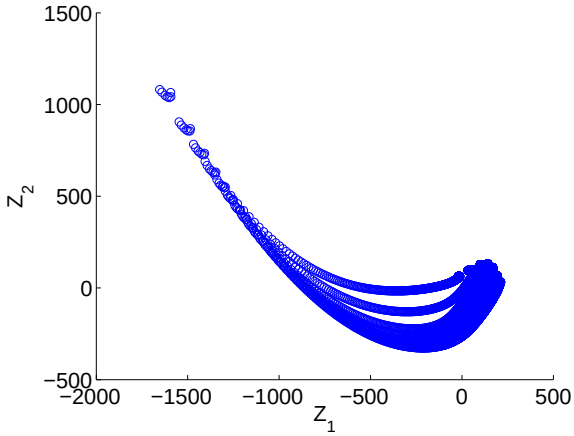
(b)pareto



(c)range

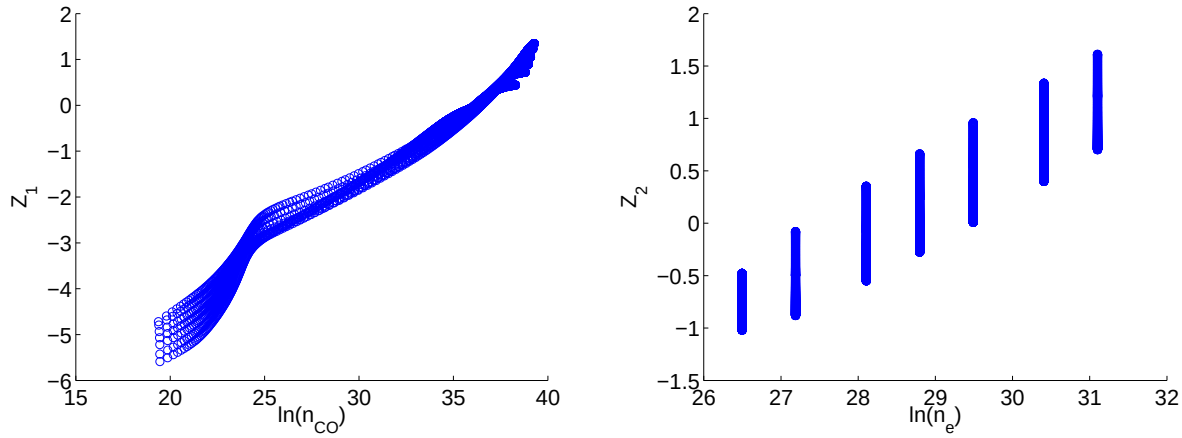


(d)level



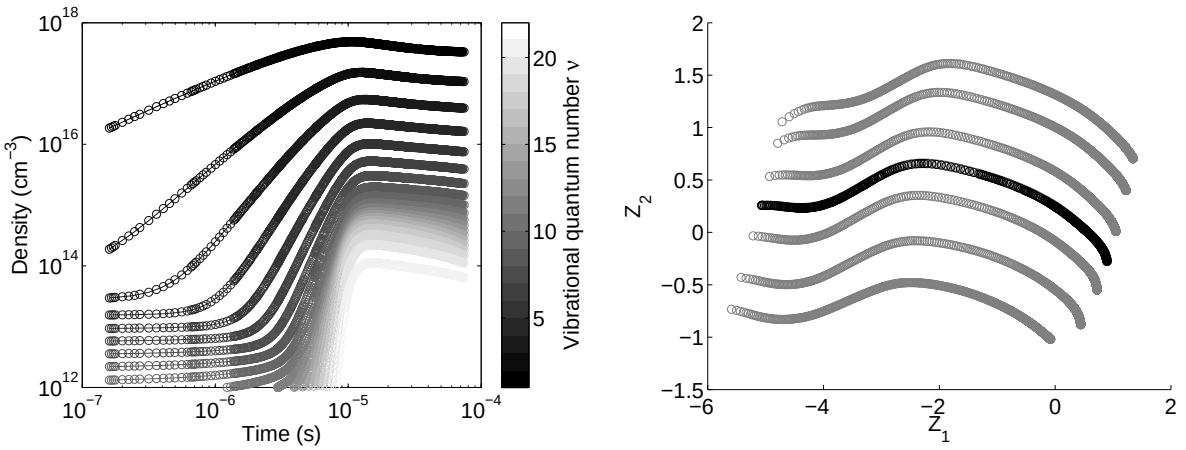
(e)vast

FIG. 9. Influence of the scaling method on the shape of the manifold.



(a) (b)

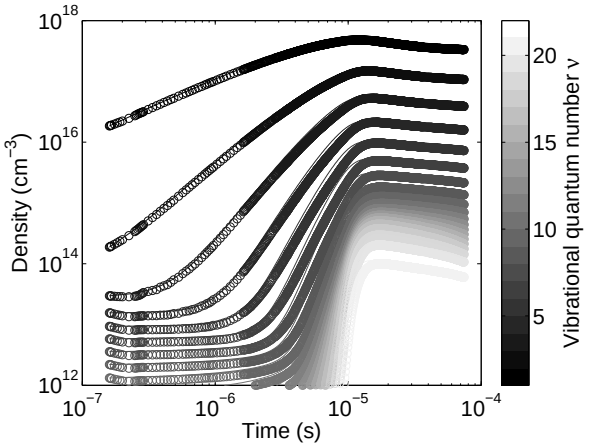
FIG. 10. Scatter plot of the log of the CO density vs Z_1 (a) and of the log of the electron density vs Z_2 (b).



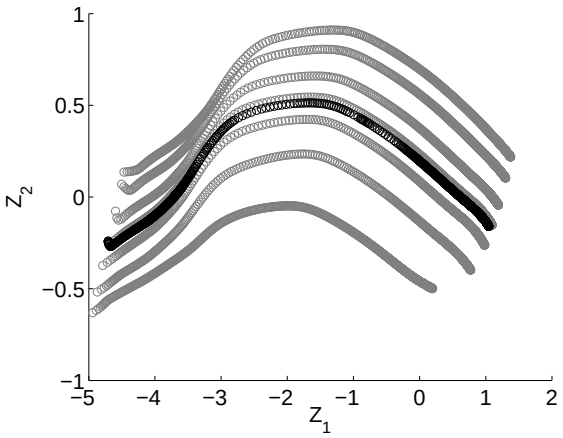
(a)CO_2 vibrational levels (b)manifold

FIG. 11. Comparison of the full and the reduced calculation for the 'range' scaling method. In figure b, the gray circles are the training data, the black ones the reduced simulation.

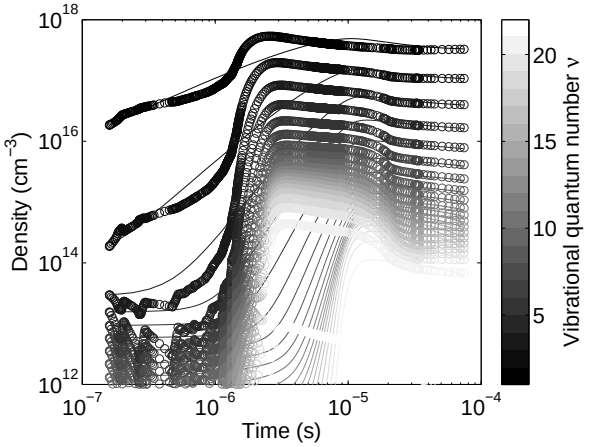
densities are only needed when the reduced simulation has finished. In figure 11 it can be seen that the reduced simulation accurately describes the time evolution of the vibrational levels of CO_2. The reduced simulation also nicely follows the trajectory of the training set in principal component space. One can argue that when tabulation is used, PCA is not strictly necessary as the state space variables and source terms can be tabulated as a function of any variable. That this is not the case, can be seen in figure 12, where different scaling



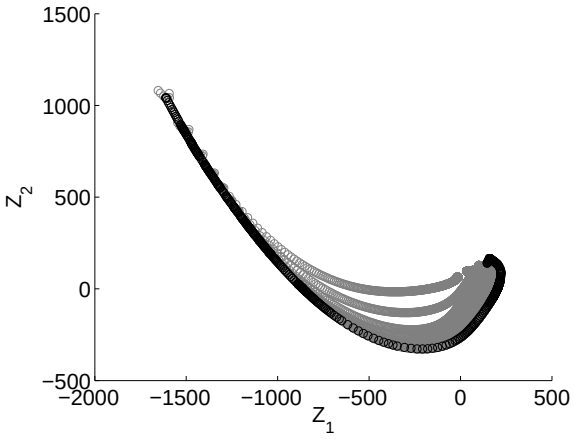
(a)CO_2 vibrational levels, 'level' scaling



(b)manifold, 'level' scaling



(c)CO_2 vibrational levels, 'vast' scaling

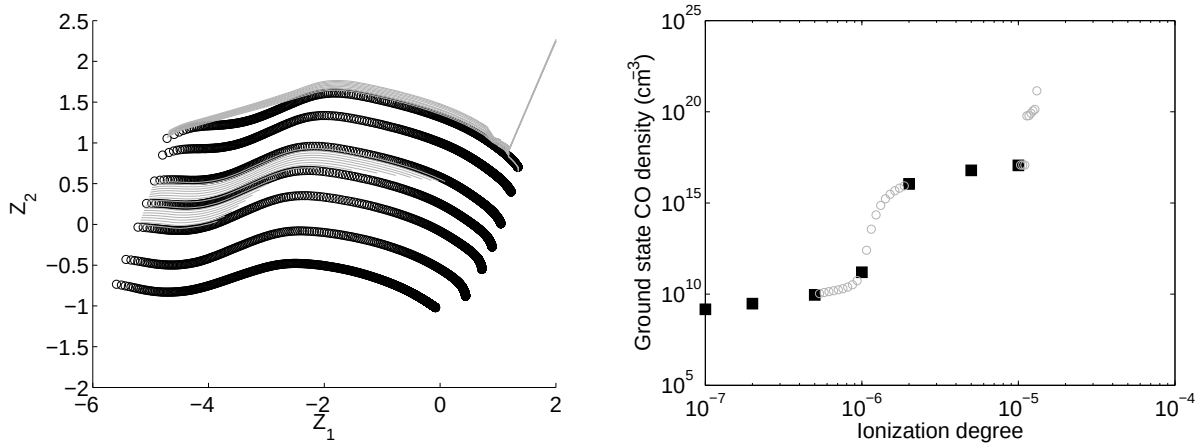


(d)manifold, 'vast' scaling

FIG. 12. Comparison of the full and the reduced calculation for different scaling methods.

methods are used. The shape of the manifold is different when different parametrization variables are used and this influences the accuracy of the reduced simulations. The 'level' scaling method gives a less smooth manifold than the 'range' scaling method. As a result the reduced simulation is less accurate. Usage of the 'vast' scaling method really ruins the reduced calculation. The usage of PCA and proper scaling is necessary to find an optimal parametrization of the manifold.

The speedup obtained in the reduced simulations is quite substantial. Since the full model is implemented in Fortran and the reduced model in Matlab, it is difficult to compare the calculation times directly. However the big difference in calculation time between the full model (ca 30 s) and the reduced model (ca 0.5 s) supports the potential of PCA to



(a) (b)

FIG. 13. Comparison of training data in black and reduced simulations in gray at different ionization degrees. The end time of the reduced simulations is 7.38e-6 s.

reduce the calculation time of non-equilibrium plasma simulations. Speedups of one to two orders of magnitude should be attainable. The speedup is obtained because the method not only reduces the number of variables to be solved for, but also reduces the stiffness of the equations and avoids run-time calculation of the chemistry via the tabulation of the principal component source terms. Note that although the reduced simulation only solves for two variables, it gives as much physical information as the original simulation: the time evolution of all the species densities.

2. Prediction

For real simulations it is not sufficient that the method can accurately redo the training simulations. The method also has to have predictive power. To test the predictive power of PCA, time integration of the principal component continuity equations is performed here at different ionization degrees that were not in the original training data set. The principal component continuity equations were integrated until 7.38e-6 s. The result can be seen in figure 13. Both the training data and the reduced simulations show a clear transition in CO_2 conversion once a certain ionization degree has been reached, as is clear from the higher CO ground state density. This transition can be understood as follows. As the reduced simulations are run for a fixed time, the specific energy input (eV/molecule) is

lower for a lower ionization degree. As was shown in[6] above a critical specific energy input there is enough vibrational excitation to stimulate the dissociation from high vibrational levels, thus increasing the CO_2 to CO conversion. When the ionization degree is inside the range of the original data set (interpolation in the lookup table) the reduced simulations give an accurate prediction of the amount of CO produced after 7.38e-6 s as a function of the ionization degree. Also the trajectories of the reduced simulation follow nicely the manifold shape of the original training data. When the ionization degree is outside the training range, but close to the boundaries of the training set, the method still gives a reasonable prediction of the amount of CO produced. However, when the ionization degree lies too far from the original training set, the method fails. This is inherent to the use of a lookup table. It is required that the values of the principal components obtained in the reduced simulation stay within the range of the table. If they run out of this range, linear extrapolation based on the points on the boundary of the table will soon become inaccurate for highly non-linear processes like plasma kinetics. When using regression this is not necessarily the case. Regression takes into account non-linearity. Furthermore, all the training points (instead of only a few on the boundary) are used when doing a prediction based on regression. Therefore, when the principal components run out of the range of the original training data, non-linear regression might be better than interpolation from a lookup table.

V. CONCLUSIONS

A reduction method based on principal component analysis has been implemented and tested on a state-to-state kinetics model of CO_2. Principal component analysis was used to identify the directions with most variance in a training dataset. In the a priori analysis it was shown that for our CO_2 plasma model, a log transformation is crucial for an accurate mapping of the original data set onto the principal components and back. Further improvement of the accuracy was achieved by using tabulation of the state space variables as a function of the principal components. Also scaling of the training data was shown to be important for the identification of the manifold.

A posteriori validation was provided by solving, for the first time to the authors knowledge, the principal component continuity equations projected into a log feature space and using tabulation to improve the accuracy of source term reconstruction. With only two

principal components, the reduced model predicts very accurately the CO_2 to CO conversion at varying ionization degrees. Furthermore, the reduced model needs only a fraction of the computational time of the detailed kinetic model. This is not so much of an issue in the present example, where the original model is 0D, but it will become crucial when implementing large chemistry sets in 2D or 3D models. The speedup is obtained because the method not only reduces the number of variables to be solved for, but also reduces the stiffness of the equations and avoids run-time calculation of the chemistry via the tabulation of the principal component source terms.

The use of principal component analysis for the reduction of plasma models looks very promising. The significant reduction observed in the present study could open up the way for spatially resolved models with detailed state-to-state kinetics. Future work will investigate the application of PCA to spatially resolved simulations.

ACKNOWLEDGMENTS

This research has been funded by the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office, Project IAP-VII P7/34.

REFERENCES

[1]G Y Park, S J Park, M Y Choi, I G Koo, J H Byun, J W Hong, J Y Sim, G J Collins, and J K Lee. Atmospheric-pressure plasma sources for biomedical applications. *Plasma Sources Science and Technology*, 21(4):043001, 2012.

[2]Tomoyuki Murakami, Kari Niemi, Timo Gans, Deborah O'Connell, and William G Graham. Afterglow chemistry of atmospheric-pressure heliumoxygen plasmas with humid air impurity. *Plasma Sources Science and Technology*, 23(2):025005, 2014.

[3]W Van Gaens and A Bogaerts. Reaction pathways of biomedically active species in an Ar plasma jet. *Plasma Sources Science and Technology*, 23(3):035015, 2014.

[4]L. M. Zhou, B. Xue, U. Kogelschatz, and B. Eliasson. Nonequilibrium plasma reforming of greenhouse gases to synthesis gas. *Energy & Fuels*, 12(6):1191–1199, 1998.

[5]T Nunnally, K Gutsol, A Rabinovich, A Fridman, A Gutsol, and A Kemoun. Dissociation

of CO_2 in a low current gliding arc plasmatron. *Journal of Physics D: Applied Physics*, 44(27):274009, 2011.

[6]Tomáš Kozák and Annemie Bogaerts. Splitting of CO_2 by vibrational excitation in non-equilibrium plasmas: a reaction kinetics model. *Plasma Sources Science and Technology*, 23(4):045004, 2014.

[7]Hai P. Le, Ann R. Karagozian, and Jean-Luc Cambier. Complexity reduction of collisional-radiative kinetics for atomic plasma. *Physics of Plasmas*, 20(12):123304, 2013.

[8]Thierry E. Magin, Marco Panesi, Anne Bourdon, Richard L. Jaffe, and David W. Schwenke. Coarse-grain model for internal energy excitation and dissociation of molecular nitrogen. *Chemical Physics*, 398:90–95, 2012.

[9]Ralph Lehmann. An algorithm for the determination of all significant pathways in chemical reaction systems. *Journal of Atmospheric Chemistry*, 47(1):45–78, 2004.

[10]A.H. Markosyan, A. Luque, F.J. Gordillo-Vzquez, and U. Ebert. Pumpkin: A tool to find principal pathways in plasma chemical models. *Computer Physics Communications*, 185(10):2697 – 2702, 2014.

[11]J A van Oijen and L P H de Goey. Modelling of premixed counterflow flames using the flamelet-generated manifold method. *Combustion Theory and Modelling*, 6(3):463–478, 2002.

[12]Benjamin J. Isaac, Jeremy N. Thornock, James Sutherland, Philip J. Smith, and Alessandro Parente. Modelling combustion systems with principal component analysis transport equations. *Combustion and Flame*, In review.

[13]Stephen B. Pope. Small scales, many species and the manifold challenges of turbulent combustion. *Proceedings of the Combustion Institute*, 34(1):1 – 31, 2013.

[14]A. Parente, J.C. Sutherland, L. Tognotti, and P.J. Smith. Identification of low-dimensional manifolds in turbulent flames. *Proceedings of the Combustion Institute*, 32(1):1579 – 1586, 2009.

[15]James C. Sutherland and Alessandro Parente. Combustion modeling using principal component analysis. *Proceedings of the Combustion Institute*, 32(1):1563 – 1570, 2009.

[16]Benjamin J. Isaac, Axel Coussement, Olivier Gicquel, Philip J. Smith, and Alessandro Parente. Reduced-order PCA models for chemical reacting flows. *Combustion and Flame*, In press, 2014.

[17]Hessam Mirgolbabaei. *Low-Dimensional Manifold Simulation of Turbulent Reacting Flows*

Using Linear and Nonlinear Principal Components Analysis. PhD thesis, North Carolina State University, 2014.

[18]Hessam Mirgolbabaei and Tarek Echekki. Nonlinear reduction of combustion composition space with kernel principal component analysis. *Combustion and Flame*, 161(1):118 – 126, 2014.

[19]Hessam Mirgolbabaei, Tarek Echekki, and Nejib Smaoui. A nonlinear principal component analysis approach for turbulent combustion composition space. *International Journal of Hydrogen Energy*, 39(9):4622 – 4633, 2014.

[20]Hessam Mirgolbabaei and Tarek Echekki. A novel principal component analysis-based acceleration scheme for LESODT: An a priori study. *Combustion and Flame*, 160(5):898 – 908, 2013.

[21]Amir Biglari and James C. Sutherland. A filter-independent model identification technique for turbulent combustion modeling. *Combustion and Flame*, 159(5):1960 – 1970, 2012.

[22]Yue Yang, Stephen B. Pope, and Jacqueline H. Chen. Empirical low-dimensional manifolds in composition space. *Combustion and Flame*, 160(10):1967 – 1980, 2013.

[23]Amit Singer, Radek Erban, Ioannis G. Kevrekidis, and Ronald R. Coifman. Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps. *Proceedings of the National Academy of Sciences*, 106(38):16090–16095, 2009.

[24]S. Pancheshnyi, B. Eismann, G.J.M. Hagelaar, and L.C. Pitchford. Computer code ZDPlasKin.

[25]G J M Hagelaar and L C Pitchford. Solving the Boltzmann equation to obtain electron transport coefficients and rate coefficients for fluid models. *Plasma Sources Sci. Technol.*, 14(4):722, 2005.

[26]C. E. Treanor, J. W. Rich, and R. G. Rehm. Vibrational relaxation of anharmonic oscillators with exchange dominated collisions. *J. Chem. Phys.*, 48(4):1798–1807, 1968.

[27]R.D. Levine and R.B. Bernstein. Thermodynamic approach to collision processes. In WilliamH. Miller, editor, *Dynamics of Molecular Collisions*, volume 2 of *Modern Theoretical Chemistry*, pages 323–364. Springer US, 1976.

[28]Alexander Fridman. *Plasma Chemistry*. Cambridge University Press, 2008.

[29]V D Rusanov, A A Fridman, and G V Sholin. The physics of a chemically active plasma with nonequilibrium vibrational excitation of molecules. *Soviet Physics Uspekhi*, 24(6):447, 1981.

[30]T. Silva, N. Britun, T. Godfroid, and R. Snyders. Optical characterization of a microwave pulsed discharge used for dissociation of CO_2. *Plasma Sources Sci. Technol.*, 23:025009, 2014.

[31]Alessandro Parente and James C. Sutherland. Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity. *Combustion and Flame*, 160(2):340 – 350, 2013.

[32]Yukikazu Itikawa. Cross sections for electron collisions with carbon dioxide. *J. Phys. Chem. Ref. Data*, 31(3):749–767, 2002.

[33]Makoto Hayashi. Electron collision cross-sections determined from beam and swarm data by boltzmann analysis. In Mario Capitelli and J.Norman Bardsley, editors, *Nonequilibrium Processes in Partially Ionized Gases*, volume 220 of *NATO ASI Series*, pages 333–340. Springer US, 1990.

[34]M. Hayashi. Bibliography of electron and photon cross sections with atoms and molecules published in the 20th century – carbon dioxide. *NIFS-DATA*, 74, 2003.

[35]James E. Land. Electron scattering cross sections for momentum transfer and inelastic excitation in carbon monoxide. *J. Appl. Phys.*, 49(12):5716–5721, 1978.

[36]M A Mangan, B G Lindsay, and R F Stebbings. Absolute partial cross sections for electron-impact ionization of CO from threshold to 1000 ev. *J. Phys. B: At., Mol. Opt. Phys.*, 33(17):3225, 2000.

[37]R. K. Janev and D. Reiter. Collision processes of hydrocarbon species in hydrogen plasmas. part 1. the methane family. *ChemInform*, 34(25), 2003.

[38]R.K. Janev, J.G. Wang, I. Murakami, and T. Kato. Cross sections and rate coefficients for electron-impact ionization of hydrocarbon molecules. *NIFS-DATA*, 68, 2001.

[39]A V Phelps. Tabulations of collision cross sections and calculated transport and reaction coefficients for electron collisions with O_2, technical report 28. Technical report, JILA information center, 1985.

[40]E. Krishnakumar and S.K. Srivastava. Cross-sections for electron impact ionization of O_2. *Int. J. Mass Spectrom. Ion Processes*, 113(1):1–12, 1992.

[41]Changjun Liu, Baldur Eliasson, Bingzhang Xue, Yang Li, and Yaquan Wang. Zeolite-enhanced plasma methane conversion directly to higher hydrocarbons using dielectric-barrier discharges. *React. Kinet. Catal. Lett.*, 74(1):71–77, 2001.

[42]B. Eliasson and U. Kogelschatz. Basic data for modelling of electrical discharges in gases:

Oxygen. Technical report, Brown Boveri Research Report KLR 86-11C, 1986.

[43]S Matejcik, A Kiendler, P Cicman, J Skalny, P Stampfli, E Illenberger, Y Chu, A Stamatovic, and T D Märk. Electron attachment to molecules and clusters of atmospheric relevance: oxygen and ozone. *Plasma Sources Sci. Technol.*, 6(2):140, 1997.

[44]Y. Itikawa and A. Ichimura. Cross sections for collisions of electrons and photons with atomic oxygen. *J. Phys. Chem. Ref. Data*, 19(3):637–651, 1990.

[45]Russ R. Laher and Forrest R. Gilmore. Updated excitation and ionization cross sections for electron impact on atomic oxygen. *J. Phys. Chem. Ref. Data*, 19(1):277–305, 1990.

[46]Jay A. Blauer and Gary R. Nickerson. A survey of vibrational relaxation rate data for processes important to CO_2-N_2-H_2O infrared plume radiation. Technical report AFRPL-TR-73-57, Ultrasystems, Inc., October 1973.

[47]M Capitelli, C M Ferreira, B F Gordiets, and A I Osipov. *Plasma Kinetics in Atmospheric Gases*. Springer, 2000.

[48]Thomas G. Kreutz, James A. O'Neill, and George W. Flynn. Diode laser absorption probe of vibration-vibration energy transfer in carbon dioxide. *The Journal of Physical Chemistry*, 91(22):5540–5543, 1987.

[49]R. D. Sharma. Near-resonant vibrational energy transfer among isotopes of CO_2. *Phys. Rev.*, 177:102–107, Jan 1969.

[50]A. Cenian, A. Chernukho, V. Borodin, and G. Śliwiński. Modeling of plasma-chemical reactions in gas mixture of CO_2 lasers I. gas decomposition in pure CO_2 glow discharge. *Contrib. Plasma Phys.*, 34(1):25–37, 1994.

[51]A. Cenian, A. Chernukho, and V. Borodin. Modeling of plasma-chemical reactions in gas mixture of CO_2 lasers. II. theoretical model and its verification. *Contrib. to Plasma Phys.*, 35(3):273–296, 1995.

[52]Thomas G. Beuthe and Jen-Shih Chang. Chemical kinetic modelling of non-equilibrium Ar-$\mathbf{CO_2}$ thermal plasmas. *Japanese Journal of Applied Physics*, 36(Part 1, No. 7B):4997–5002, 1997.

[53]Hans-Jürgen Mick, Michael Burmeister, and Paul Roth. Atomic resonance absorption spectroscopy measurements on high-temperature CO dissociation kinetics. *AIAA Journal*, 31(4):671–676, April 1993.

[54]I A Kossyi, A Yu Kostinsky, A A Matveyev, and V P Silakov. Kinetic scheme of the non-equilibrium discharge in nitrogen-oxygen mixtures. *Plasma Sources Sci. Technol.*,

1(3):207, 1992.

[55]S Hadj-Ziane, B Held, P Pignolet, R Peyrous, and C Coste. Ozone generation in an oxygen-fed wire-to-cylinder ozonizer at atmospheric pressure. *J. Phys. D: Appl. Phys.*, 25(4):677, 1992.

Appendix A: List of chemical reactions

No.	Reaction	Ref.	Note
(X1)	$e + CO_2 \longrightarrow e + CO_2$	[32]	a
(X2)	$e + CO_2 \longrightarrow e + e + {CO_2}^+$	[32]	a
(X3)	$e + CO_2 \longrightarrow e + e + CO^+ + O$	[32]	b
(X4)	$e + CO_2 \longrightarrow e + e + C^+ + O_2$	[32]	b
(X5)	$e + CO_2 \longrightarrow e + e + O^+ + CO$	[32]	b
(X6)	$e + CO_2 \longrightarrow O^- + CO$	[32]	b
(X7)	$e + CO_2 \longrightarrow e + CO + O$	[32]	b
(X8)	$e + CO_2 \longrightarrow e + CO_2e_1$	[33]	a
(X9)	$e + CO_2 \longrightarrow e + CO_2e_2$	[33]	a
(X10)	$e + CO_2 \longrightarrow e + CO_2v_a$	[34]	
(X11)	$e + CO_2 \longrightarrow e + CO_2v_b$	[34]	
(X12)	$e + CO_2 \longrightarrow e + CO_2v_c$	[34]	
(X13)	$e + CO_2 \longrightarrow e + CO_2v_d$	[34]	
(X14)	$e + CO_2v_i \longrightarrow e + CO_2v_j$	[34]	c
(X15)	$e + CO \longrightarrow e + CO$	[35]	a
(X16)	$e + CO \longrightarrow e + e + CO^+$	[35,36]	a
(X17)	$e + CO \longrightarrow e + e + C^+ + O$	[35,36]	b
(X18)	$e + CO \longrightarrow e + e + C + O^+$	[35,36]	b
(X19)	$e + CO \longrightarrow C + O^-$	[35]	b
(X20)	$e + CO \longrightarrow e + COe_1$	[35]	a
(X21)	$e + CO \longrightarrow e + COe_2$	[35]	a
(X22)	$e + CO \longrightarrow e + COe_3$	[35]	a
(X23)	$e + CO \longrightarrow e + COe_4$	[35]	a
(X24)	$e + COv_i \longrightarrow e + COv_j$	[35]	d
(X25)	$e + C \longrightarrow e + C$	[37]	
(X26)	$e + C \longrightarrow e + e + C^+$	[37]	
(X27)	$e + C_2 \longrightarrow e + C_2$	[38]	
(X28)	$e + C_2 \longrightarrow e + C + C$	[38]	
(X29)	$e + C_2 \longrightarrow e + e + {C_2}^+$	[38]	
(X30)	$e + O_2 \longrightarrow e + O_2$	[39]	a
(X31)	$e + O_2 \longrightarrow e + O + O$	[39]	b
(X32)	$e + O_2 \longrightarrow e + e + {O_2}^+$	[39]	a
(X33)	$e + O_2 \longrightarrow e + e + O + O^+$	[40]	b
(X34)	$e + O_2 \longrightarrow O^- + O$	[39]	b
(X35)	$e + O_2 \longrightarrow e + O_2v_1$	[39]	
(X36)	$e + O_2 \longrightarrow e + O_2v_2$	[39]	
(X37)	$e + O_2 \longrightarrow e + O_2v_3$	[39]	
(X38)	$e + O_2 \longrightarrow e + O_2e_1$	[39]	
(X39)	$e + O_2 \longrightarrow e + O_2e_2$	[39]	
(X40)	$e + O_3 \longrightarrow e + O_3$	[41]	
(X41)	$e + O_3 \longrightarrow e + O_2 + O$	[42]	
(X42)	$e + O_3 \longrightarrow e + e + {O_2}^+ + O$	[42]	

TABLE V. Vibrational energy transfer reactions of CO_2, CO and O_2.

No.	Reaction	Ref.	Note
(V1)	$CO_2v_x + M \longleftrightarrow CO_2 + M$	[46]	x=a,b,c,d
(V2a)	$CO_2v_i + M \longleftrightarrow CO_2v_{i\text{-}1(a)} + M$	[46]	
(V2b)	$CO_2v_i + M \longleftrightarrow CO_2v_{i\text{-}1(b)} + M$	[46]	
(V2c)	$CO_2v_i + M \longleftrightarrow CO_2v_{i\text{-}1(c)} + M$	[46]	
(V3)	$COv_i + M \longleftrightarrow COv_{i\text{-}1} + M$	[47]	
(V4)	$O_2v_i + M \longleftrightarrow O_2v_{i\text{-}1} + M$	[46]	
(V5)	$CO_2v_i + CO_2 \longleftrightarrow CO_2v_{i\text{-}1} + CO_2v_x$	[46]	x=a,b
(V6)	$CO_2v_i + CO_{2j} \longleftrightarrow CO_2v_{i\text{-}1} + CO_2v_{j+1}$	[48,49]	
(V7)	$COv_i + COv_j \longleftrightarrow COv_{i\text{-}1} + COv_{j+1}$	[47]	
(V8)	$CO_2v_i + COv_j \longleftrightarrow CO_2v_{i\text{-}1} + COv_{j+1}$	[46]	

(X43)	$e + O_3 \longrightarrow e + O^+ + O^- + O$	[42]
(X44)	$e + O_3 \longrightarrow O^- + O_2$	[43]
(X45)	$e + O_3 \longrightarrow O + O_2^-$	[43]
(X46)	$e + O \longrightarrow e + O$	[44]
(X47)	$e + O \longrightarrow e + e + O^+$	[45]

TABLE IV: Electron impact reactions described by collision cross sections.

(a) Same cross section used for reactions of CO_2v_i, and analogously for COv_i or O_2v_i.

(b) Cross section modified by lowering the energy threshold by the excited state energy used for reactions of CO_2v_i, and analogously for COv_i and O_2v_i.

(c) Cross section $0 \rightarrow 1$ scaled and shifted using Fridman's approximation[6].

(d) Cross sections $0 \rightarrow j$ $(j = 1 \ldots 10)$ scaled and shifted using Fridman's approximation[6].

TABLE VI. Reactions of neutrals included in the model. The vibrational levels of CO_2, CO and O_2 follow the same reactions as the ground states, but the corresponding reaction rates are modified according to the Fridman-Macheret α model (the activation energy is lowered by a fraction of the vibrational energy, see[6,28] for more details).

No.	Reaction	Ref.
(N1)	$CO_2 + M \longrightarrow CO + O + M$	[28]
(N2)	$CO + O + M \longrightarrow CO_2 + M$	[50]
(N3)	$CO_2 + O \longrightarrow CO + O_2$	[28]
(N4)	$O_2 + CO \longrightarrow CO_2 + O$	[28]
(N5)	$CO_2 + C \longrightarrow CO + CO$	[51]
(N6)	$CO + O_3 \longrightarrow CO_2 + O_2$	[52]
(N7)	$CO + C + M \longrightarrow C_2O + M$	[50]
(N8)	$O_2 + C \longrightarrow CO + O$	[50]
(N9)	$CO + M \longrightarrow O + C + M$	[53]
(N10)	$O + C + M \longrightarrow CO + M$	[52]
(N11)	$O + C_2O \longrightarrow CO + CO$	[51]
(N12)	$O_2 + C_2O \longrightarrow CO_2 + CO$	[50]
(N13)	$O + O_3 \longrightarrow O_2 + O_2$	[50]
(N14)	$O_3 + M \longrightarrow O_2 + O + M$	[52]
(N15)	$O + O_2 + M \longrightarrow O_3 + M$	[54]
(N16)	$O + O + M \longrightarrow O_2 + M$	[55]