Reduced-order kinetic plasma models using principal component analysis: Model formulation and manifold sensitivity

Aurélie Bellemans* and Thierry Magin†
Aeronautics and Aerospace Department, von Karman Institute for Fluid Dynamics, 72 Chaussée de Waterloo, 1640 Rhode-Saint-Genèse, Belgium

Axel Coussement‡ and Alessandro Parente§
Service d'Aéro-Thermo-Mécanique, Université libre de Bruxelles, 50 avenue F.D. Roosevelt, 1050 Bruxelles, Belgium

(Received 31 January 2017; published 24 July 2017)

Plasma flows involve hundreds of species and thousands of reactions at different time scales, resulting in a very large set of governing equations to solve. Simulating large reacting systems in nonequilibrium plasma mixtures remains a challenge with the currently available computational resources. Principal component analysis (PCA) offers a general and rather simple and automated method to reduce large kinetic mechanisms by principal variable selection. This work shows how to adapt and apply the PCA-scores technique, which has its origin in the combustion field, to a collisional-radiative model. We have successfully applied this technique to argon plasmas, reducing the set of governing equations by more than 90%, leading to an important speed-up of the calculation and a reduction of computational cost.

DOI: 10.1103/PhysRevFluids.2.073201

I. INTRODUCTION

Accurate predictive simulations are important to understand and solve many of today's technological challenges to nonequilibrium plasma applications, such as the optimization of electric propulsion thrusters operating on electronegative plasmas [1,2], the reduction of the ignition delay time and ignition temperature in plasma-assisted combustion [3,4], the design of lightweight carbon-composite materials used as thermal protection systems for spacecraft during atmospheric entries [5,6], as well as to understand intricate phenomena such as sunspot formation and magnetic reconnection in solar physics [7,8]. Nonequilibrium effects in plasmas can be described by means of physicochemical models of various complexity and fidelity levels. The most accurate description lies in the state-to-state (STS) models which provide the populations of the internal energy levels of all molecules and atoms of a plasma. The inner states of each species are solved separately by means of a detailed kinetic mechanism for the electronic mode of atoms and for the rotational, vibrational and electronic modes of molecules. Collisional-radiative (CR) models describe both collisional and radiative elementary processes. Full STS models have been developed mostly for atomic plasmas [9–11] for which hundreds of inner states can be involved. One ends up with a massive system of governing equations which is very expensive to solve numerically. With the current computational resources, the detailed chemistry of nonequilibrium mixtures is often calculated through zero-dimensional or one-dimensional computational fluid dynamics solvers.

To perform three-dimensional simulations, the computational cost inherent to STS models must be alleviated. For instance, the use of a simplified chemistry obtained by degrading its level of detail while conserving some of the information about the internal energy level populations allows us to

*Present address: ATM, Université libre de Bruxelles; aurelie.bellemans@vki.ac.be
†thierry.magin@vki.ac.be
‡axel.coussement@ulb.ac.be
§alessandro.parente@ulb.ac.be

preserve the solution accuracy. For molecular plasmas, thousands of inner states are present. Many reduction techniques have been developed in the literature. Multitemperature (MT) models rely on the assumption of a Maxwell-Boltzmann distribution for the population of each internal energy mode of the species, provided that the system is close to equilibrium. This most probable distribution function, according to statistical mechanics, is described by means of specific rotational, vibrational, and electronic temperatures. In strong nonequilibrium situations, this assumption is wrong and the populations deviate from a Maxwell-Boltzmann distribution. Alternative descriptions, such as hybrid models, combine both STS and MT approaches [12–17]. Coarse-grain models have also been proposed based on a so-called binning approach: The inner energy states of the species are lumped into several bins after some suitable averaging. After lumping the levels, the macroscopic rates are extracted for each group [18–20]. However, many of these techniques require strong physical insight to finely tune the reduced model. For instance, additional energy equations are used to close the governing equations by providing the temperatures used in the distributions for the MT and coarse-grain models [21–23].

Recently, a new generation of reduction techniques has been developed in the combustion community using an empirical method called principal component analysis (PCA) [24]. The advantage of this method lies in its simplicity as the main parameters for the reduction are selected in an automated way after solving an eigenvalue problem [25,26]. In this contribution, accurate physical models for plasmas are combined to PCA reduction techniques. In previous papers, we have pioneered the development of PCA models for plasma applications [27,28]. Although the results reported in previous work were very encouraging, the reduced model using the so-called manifold generated principal component analysis (MG-PCA) technique remained expensive. The goal of the present paper is to examine how an optimized PCA technique can perform in terms of system reduction compared to MG-PCA. The paper focuses on the development of a global reduction technique based on the PCA-score method. A score or principal component is a new variable which is a linear combination of the original variables. The PCA-score technique is applied to reduce the mass fractions of a 34-species argon model for the simulation of shock tube experiments. The first argon test case chosen has been validated by Kapper and Cambier [11] based on experiments performed at the University of Toronto's Institute of Aerospace Studies (UTIAS) [29]. Anticipating the application in a future work of the reduction technique to CR models for air, we propose to verify the robustness of the reduced model by exploring a broader range of free-stream conditions representative of re-entry flight.

The paper has the following structure: The physical model for collisional-radiative argon chemistry is presented in Sec. II. It describes the governing equations for modeling one-dimensional shocks to simulate argon shock tube experiments. Section III shows how PCA scores are derived from principal component analysis after applying some preprocessing techniques, such as outlier removal, data centering, and scaling. The PCA-score technique is then applied to the argon shock tube simulations for a large span of free-stream conditions in Sec. IV, followed by the conclusion.

II. PHYSICAL MODELING

In this section, one briefly describes the collisional-radiative model for argon plasmas before applying the PCA-score technique to reduce it. The same chemical mechanism has already been reduced in a previous work [28] based on the MG-PCA reduction method.

A. Detailed chemistry modeling

The electronic-specific CR model for argon used here has been originally developed by Vlcek [30] and Bultel *et al.* [31], and later adapted and validated by Kapper and Cambier [10,11] against experimental data acquired in the UTIAS shock tube facility. The gas mixture consists of Ar, and its ionization products, Ar^+ and e^-. The influence of additional components, such as Ar^{++} and Ar_2^+, has been neglected [31]. The number of electronic energy levels retained for Ar and Ar^+ are 31 and

TABLE I. Energy data for the electronic inner states of Ar.

i	E_i [eV]	a_i	$j_{c,i}$	i	E_i [eV]	a_i	$j_{c,i}$
1	0	1	1.5	17	13.864	3	1.5
2	11.548	5	1.5	18	13.903	5	1.5
3	11.624	3	1.5	19	13.979	9	1.5
4	11.723	1	0.5	20	14.013	7	1.5
5	11.828	3	0.5	21	14.063	5	1.5
6	12.907	3	1.5	22	14.068	5	1.5
7	13.076	7	1.5	23	14.090	3	1.5
8	13.095	5	1.5	24	14.099	7	1.5
9	13.153	3	1.5	25	14.153	3	1.5
10	13.172	5	1.5	26	14.214	5	0.5
11	13.273	1	1.5	27	14.234	5	0.5
12	13.283	3	0.5	28	14.236	7	0.5
13	13.302	5	0.5	29	14.241	1	0.5
14	13.328	3	0.5	30	14.255	3	0.5
15	13.480	1	0.5	31	14.304	3	0.5
16	13.845	1	1.5				

2, respectively. The mixture contains in total $N = 34$ species when adding up the free electrons. Table I provides the excitation energy (E_i), degeneracy (g_i), and core angular momentum ($j_{c,i}$) for the 31 electronic energy levels Ar(i) of argon. From this table, one can observe there are two possible values for the angular momentum: $1/2$ and $3/2$. These correspond to two possible ionization potentials when considering the ionization reactions starting from the excited states. For example, Ar(2) will ionize to $Ar^+(1)$ and Ar(4) to $Ar^+(2)$. Table II shows the energy data for those two ionized states. To account for thermal nonequilibrium between heavy particles and free electrons, separate translational temperatures are denoted in what follows by symbols T_h and T_e, respectively.

The kinetic mechanism accounts for the following collisional and radiative processes:

(1) Excitation and de-excitation by electron and atom impact:

$$\mathrm{Ar}(i) + e^- \underset{Fji}{\overset{Cij}{\rightleftharpoons}} \mathrm{Ar}(j) + e^-, \tag{1}$$

$$\mathrm{Ar}(i) + \mathrm{Ar}(1) \underset{Lji}{\overset{Kij}{\rightleftharpoons}} \mathrm{Ar}(j) + \mathrm{Ar}(1), \tag{2}$$

(2) Ionization and recombination by electron and atom impact:

$$\mathrm{Ar}(i) + e^- \underset{Oi}{\overset{Si}{\rightleftharpoons}} \mathrm{Ar}^+ + 2e^-, \tag{3}$$

$$\mathrm{Ar}(i) + \mathrm{Ar}(1) \underset{Wi}{\overset{Vi}{\rightleftharpoons}} \mathrm{Ar}^+ + e^- + \mathrm{Ar}(1), \tag{4}$$

TABLE II. Energy data for the electronic inner states of Ar.

i	E_i [eV]	a_i	$j_{c,i}$
1	15.760	4	1.5
2	15.937	2	0.5

TABLE III. Collisional and radiative elementary processes for argon: forward rate coefficient k_f, backward rate coefficient k_b, and collisional-radiative process.

k_f	k_b	Process
Cij	Fij	Excitation by electron impact
Kij	Lij	Excitation by impact with the ground electronic state
Si	Oi	Ionization by electron impact
Vi	Wi	Ionization by impact with the ground electronic state
$(1\text{-}\Lambda)Aji$	ΛAji	Radiative excitation and de-excitation
$(1\text{-}\Lambda)Ri$	ΛRi	Radiative ionization and recombination

(3) Spontaneous emission and absorption (bound-bound):

$$\mathrm{Ar}(i) + h_P \nu \underset{\Lambda Aji}{\overset{(1-\Lambda)Aji}{\rightleftharpoons}} \mathrm{Ar}(j), \tag{5}$$

(4) Photo-ionization and radiative recombination (bound-free and *free-bound*)

$$\mathrm{Ar}(i) + h_P \nu \underset{\Lambda Ri}{\overset{(1-\Lambda)Ri}{\rightleftharpoons}} \mathrm{Ar}^+ + e^-, \tag{6}$$

(5) Bremsstrahlung emission (free-free)

$$e^- + \mathrm{Ar}^+ \rightarrow e^- + \mathrm{Ar}^+ + h_P \nu, \tag{7}$$

where the indices $i = \{1, \ldots, 31\}$ and $j > i$ denote the electronic energy levels of Ar involved in the transitions. The indices for the two levels of Ar^+ are omitted here for brevity of the notation. Superelastic collisions have been neglected; i.e., only the ground state, Ar(1), participates in atomic collisions, due to the low density of excited states in the regime where these dominate over electron-impact collisions. Quantities h_P and ν in Eqs. (5) and (6) stand for Planck's constant and the radiation frequency. Table III gives an overview of all the rate coefficients used for each chemical process. In the model [10], radiation absorption is taken into account via the use of escape factors Λ that can take values between 0 and 1, bounds which correspond, respectively, to optically thick and thin plasmas. An optically thin plasma is considered here. The total number of elementary processes (both collisional and radiative) taken into account in the CR model is equal to 962. Each of these processes intervenes at a specific location after the shock wave. The initial production of electrons after the shock front starts with atom-impact collisions. The lowest excited states of argon are populated and then ionized to produce the very first electrons. Once enough electrons are produced, the chemistry dynamics changes, and electron-impact processes take over to excite additional internal energy states and produce more electrons.

B. Governing equations

The Euler equations for two-temperature plasmas represent the conservation of mass for the N species, mixture momentum, mixture energy, and electron energy. For a one-dimensional (1D) steady flow in the x direction, they take the general differential form:

$$\frac{d}{dx} F = \dot{\Omega}. \tag{8}$$

In this expression, the vector F contains the fluxes, and the vector $\dot{\Omega}$ has the source terms:

$$F = \begin{pmatrix} \rho u y_1 \\ \vdots \\ \rho u y_N \\ \rho u^2 + p \\ \rho u\left(h + \frac{1}{2}u^2\right) \\ \rho u y_e e^e \end{pmatrix}, \quad \dot{\Omega} = \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_N \\ 0 \\ -Q_{\text{radI}} \\ \Omega^{ET} + \Omega^{Ch} - p_e \frac{d}{dx}u - Q_{\text{radII}} \end{pmatrix}.$$

Quantity ρ is the total mass density; u, the hydrodynamic velocity; y_i, the mass fraction of species i; ω_i, its chemical production term; p, the mixture pressure; and h, the mixture enthalpy. The variables only related to free electrons are the electron specific energy e^e, the electron energy production due to thermal relaxation with heavy particles

$$\Omega^{ET} = \frac{3}{2} n_e k_B (T_h - T_e) \frac{1}{\tau^{ET}}, \tag{9}$$

the electron energy production due chemical reactions induced by electron impact Ω^{Ch}, the electron partial pressure p_e, and the electron number density n_e. The expression for the relaxation time is obtained from kinetic theory

$$\frac{1}{\tau^{ET}} = \sum_{j \neq e} \frac{8}{3} \frac{m_e}{m_j} \sqrt{\frac{8 k_B T_e}{\pi m_e}} n_j \bar{Q}_{ej}^{(1,1)},$$

where quantity m_j stands for the mass of species j; n_j, its number density; k_B, the Boltzmann constant; and $\bar{Q}_{ej}^{(1,1)}$, the reduced momentum collision integral.

Radiative source terms are added to the mixture and electron energy conservation equations [10],

$$Q_{\text{radI}} = \sum (E_j - E_i) n_j A_{ji}, \tag{10}$$

$$Q_{\text{radII}} = \sum (E_{Ar^+} - E_i) n_j R_{ji} + Q_{\text{Brem}}, \tag{11}$$

$$Q_{\text{Brem}} = -1.42 \times 10^{-40} Z_{\text{eff}}^2 \sqrt{T_e} n_{\text{Ar}^+} n_e [\text{W/m}^3]. \tag{12}$$

Quantity Q_{radI} represents the radiative power due to bound-bound transitions driven by the Einstein transition coefficients, Aji [s^{-1}]. Another radiative power term, Q_{radII}, has to be added for the conservation of electron energy. It regroups the bound-free and free-free transitions. Quantity $E_{\text{Ar+}}$ stands for the energy of the ionized argon levels. For Bremsstrahlung emission, a value of 1.67 has been taken for the effective charge Z_{eff}^2 to better match the experimental data, even though double ionization is neglected in this work.

C. Detailing the shock structure

The shock is treated as a discontinuity located at the position $x = 0$. Starting from the free stream velocity, pressure, and temperature, the postshock conditions are obtained based on the Rankine-Hugoniot jump relations under the assumption of frozen kinetics within the shock. The electron temperature is also assumed to be frozen across the shock and precursor ionization is disregarded. Table IV reviews the free-stream parameters for the test cases considered in a velocity and pressure grid. The first test case corresponds to one of the shock tube experiments of UTIAS [29]. For the other test cases, as shown in Fig. 1, the free-stream pressure and velocity have been changed to more severe conditions to observe how the method can catch the complex system dynamics in the nonequilibrium regime. For the intermediate cases 2 and 3, the pressure has been decreased while conserving the same free-stream velocity and temperature. The free stream conditions of case 4

TABLE IV. Shock-tube test cases considered in the present work. Case 1, UTIAS experiments; case 2, intermediate nonequilibrium conditions; and case 3, free-stream conditions representative of flight.

Case	v [m/s]	v [Ma]	p [Pa]	T [K]
1	5075	15.9	685.3	293.6
2	5075	15.9	300	293.6
3	5075	15.9	50	293.6
4	8000	25.1	50	293.6
5	8000	25.1	685.3	293.6

match typical re-entry velocity and pressure: 8 km/s and 50 Pa, respectively. Case 5 completes the test grid simulating high pressures and speeds.

Following Refs. [32,33], the conservative form given in Eq. (8) is transformed into a system of ordinary differential equations easily solved by means of the LSODE [34] package. The resulting SHOCKING code has been verified against literature results [10] and validated by using the UTIAS shock-tube experiments [29]. For instance, Figure 2(a) shows the evolution of the translational temperature of the heavy species and free electrons after the shock. We recall briefly some results obtained by means of the CR model [10] that will be reproduced in Sec. III based on the reduction method. The free electron temperature describes a peak right after the shock, before decreasing and rising again until thermal equilibrium is reached. This peak is triggered by heavy impact ionization as it is the reaction that creates the first electrons. Chemistry changes afterwards, and electron impact processes take over the dynamics of the system, as indicated by a dip in the electron temperature. Experimental data have been plotted in Fig. 2(b) against the calculated electron number density and total density. Fair agreement is found between the simulation results and experimental data. Kapper and Cambier [11] have shown that the fluctuations of the shock structure observed experimentally can be reproduced by means of unsteady numerical simulations and explained on the basis of the coupling of the nonlinear kinetics of the CR model with wave propagation within the induction zone. Steady simulations are selected here for their simplicity to develop the reduction method.

The temperature, electron density, and mixture density profiles after the shock are shown in Fig. 3(a) for the cases 1, 2, and 4. The free-stream velocity drives the postshock heavy-species temperature, as clearly observed for case 4 in Fig. 3(a). The thermal relaxation distance is inversely proportional to the quantity $nT_e^{1/2}$. At constant free stream velocity, the relaxation distance for case 2 is thus larger for case 1 due to a lower free stream pressure. The higher free stream velocity of case 2 has the opposite effect due to a higher postshock temperature in spite of a lower value of pressure. In Fig. 3(b), the mixture density, proportional to p/T_h, decreases from case 1 to case 4, while the electron number density proportional to p/T_e decreases from case 1 to case 2, and then slightly increases from case 2 to case 4, due to the rise in electron temperature.

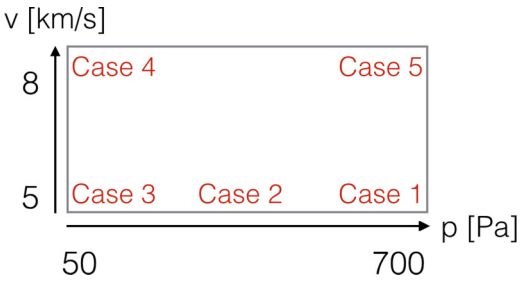


FIG. 1. Free stream velocity and pressure conditions.

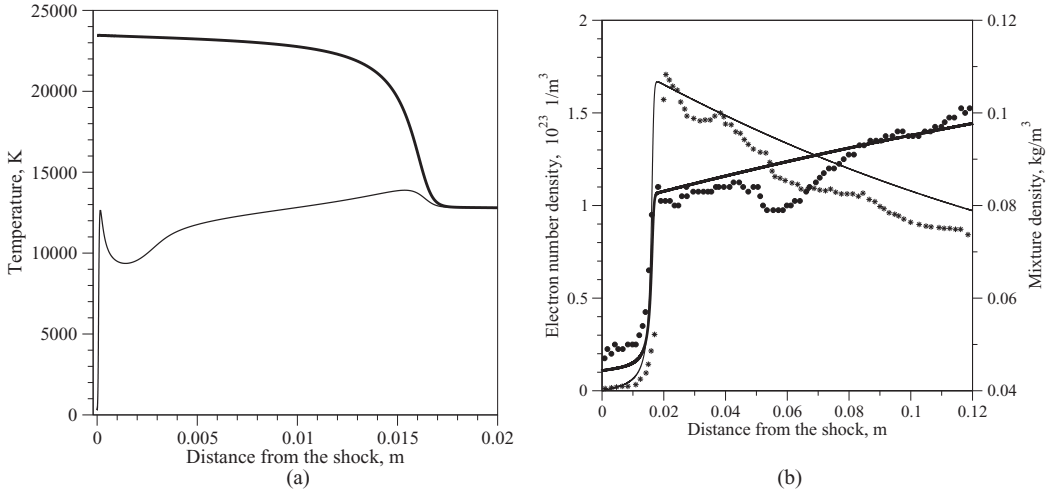


FIG. 2. Postshock simulation, case 1. (a) Translational temperatures: thick line, T_h; thin line, T_e. (b) Electron number density n_e: thin line, simulation; $*$, experimental data. Mixture density ρ: thick line, simulation; $\bullet$, experimental data.

The collisional-radiative model allows us to calculate the populations of the inner states of argon in an accurate way through the full simulation domain. When plotting these populations over their degeneracy, n_i/g_i, against their corresponding inner energy in a semilogarithmic scale, a Boltzmann plot is obtained as shown in Fig. 4(a). For a Boltzmann population at equilibrium, the observations fit to a straight line. To quantify the degree of nonequilibrium, the coefficient of determination $R(x)$ is introduced as follows, $R^2(x) = 1 - \sum_i [z_i(x) - f_i(x)]^2 / \sum_i [z_i(x) - \bar{z}(x)]^2$, based on the populations $z_i(x) = \log[n_i(x)/g_i]$ for the energy level i of argon and the value of a straight-line approximation $f_i(x)$, where the average value is computed as $\bar{z} = \sum_i z_i$. At a distance from the shock larger than 0.001 m, the equilibrium condition is satisfied for the argon electronic energy level

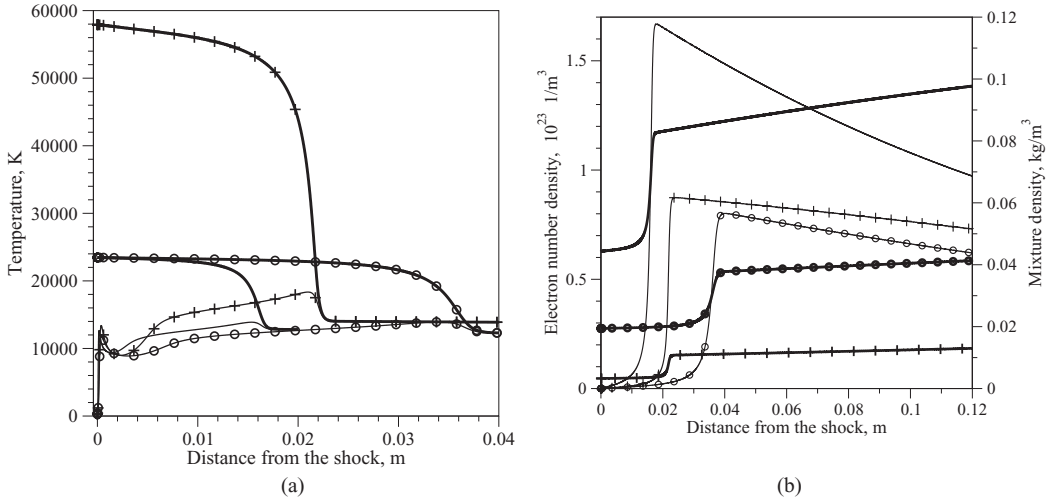
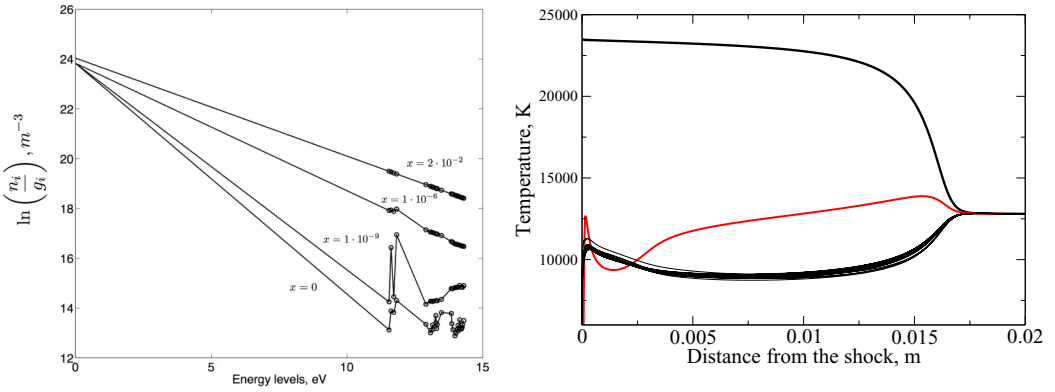


FIG. 3. Postshock simulation. Lines without symbols, case 1; lines with $\circ$, case 2; lines with $+$, case 4. (a) Translational temperatures: thick line, T_h; thin line, T_e. (b) Electron number density n_e, thin line. Mixture density ρ, thick line.



(a) (b)

FIG. 4. Postshock simulation of case 1. (a) Argon inner-state populations in a Boltzmann plot at various distances after the shock. (b) Comparison of the excitation temperature of the levels T_x (thin black lines) against T_e (red line) and T_h (thick line).

populations. Another way to visualize the departure with respect to equilibrium of the various energy level populations is to extract the excitation temperature of each excited level to the temperature of the free electrons,

$$T_x(i) = \frac{E(1) - E(i)}{k_B \ln \frac{n(1)g(i)}{n(i)g(1)}}, \tag{13}$$

with the index 1 referring to the ground state of argon. These temperatures have been compared for the first excited state Ar(2) and last excited state Ar(31) in Table V. The comparison can be visualized in Fig. 4(b). The excitation temperatures are close to each other and clearly distinct from the electron and heavy-particle temperatures, until thermal equilibrium is reached.

III. PRINCIPAL COMPONENT ANALYSIS FOR CHEMISTRY REDUCTION

Principal component analysis is a statistical method based on an orthogonal transformation projecting the primitive variables of a system on its principal components or scores. These components are the directions with the largest variance within the data sample and are a linear combination of the original variables. The global idea is to reduce the system by eliminating the less important variables and to build a new low-dimensional manifold based on those principal

TABLE V. Coefficient of determination for case 1.

Distance [m]	R^2	T_e [K]	T_2 [K]	T_{31} [K]
0	0.2295	293.6	5439.7	6880.2
10^{-9}	0.2107	293.6	5439.7	6937.9
10^{-8}	0.0375	293.7	5440.5	7195.5
10^{-7}	0.0431	294.3	5480.2	7578.5
10^{-6}	0.5764	300.4	6074.7	8078.9
10^{-5}	0.9218	396.4	7673.4	9005.9
10^{-4}	0.9921	10321.5	9963.4	10545.8
10^{-3}	1.0000	9493.9	10334.0	10250.2
10^{-2}	1.0000	12807.8	12807.8	12807.8

components or scores. The computational cost decreases considerably as only a smaller number of variables, the scores, are taken into account to solve the set of governing equations.

An eigenvalue problem is solved on a training data set made up of the mass fractions of the system. This eigenvalue problem gives the base of principal components. By truncating the base with the component corresponding to the highest eigenvalue, one can define a new, reduced, state-space. Two main methods have been developed based on principal component analysis: PCA score by Sutherland *et al.* [35] and manifold generated principal component analysis (MG-PCA) by Coussement *et al.* [36]. The PCA-score method allows for transporting the principal components directly, as PCA is a linear transformation of the original state space. If a set of governing equations exists in the original space, a set of equations can be found in the space given by the scores or principal components. The MG-PCA method, on the contrary, transports a set of variables of the original state space and reconstructs the others at every iteration. The main difference between both methods lies in the working space. When working with scores, the governing equations are rewritten accordingly in terms of principal components, while MG-PCA uses the original equations in the original state space. This section of the paper describes the PCA-score technique in more detail and shows how it can be coupled to a rotation method, such as the VARIMAX method [37], for retrieving a more stable formulation of the source terms and increasing the robustness of the code.

A. PCA

PCA starts with a training data set containing the value of all conserved variables for several observations of the system state space. These conserved variables correspond to mass fractions, temperatures, and velocity. However, previous work [38] has shown the reduction works best when using only mass fractions when carrying out PCA assembled in a matrix $\mathbf{Y}$,

$$\mathbf{Y} = \begin{bmatrix} y_{11} & \cdots & y_{1Q} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nQ} \end{bmatrix}, \tag{14}$$

with n being the number of observations in the PCA sample and Q being the number of original variables, in this case the 34 mass fractions of the argon system.

Preprocessing techniques are applied to prepare the data for PCA. These preprocessing techniques consist in centering and scaling. Centering allows focusing on fluctuations around mean values. Choosing a good scaling method is essential as it can affect the size and the accuracy of the PCA reduction. An overview of different scaling techniques (variable stability, Pareto, max, etc.) are given in the work of Parente and Sutherland [39]. In previous work on the reduction of collisional-radiative chemistry, Pareto scaling has been determined as the best scaling method.

To find the weighting factors, we define the principal components as a linear combination of the original variables with the covariance matrix $\mathbf{S}$. The element (i, j) of the covariance matrix is the covariance between the original variables:

$$\mathbf{S} = \frac{1}{(n-1)} \mathbf{Y}^T \mathbf{Y}. \tag{15}$$

Recalling the eigenvector decomposition of a symmetric, nonsingular matrix, $\mathbf{S}$ can be decomposed as

$$\mathbf{S} = \mathbf{A} \boldsymbol{\lambda} \mathbf{A}^T, \tag{16}$$

where $\boldsymbol{\lambda}$ is a $(Q \times Q)$ diagonal matrix containing the eigenvalues of $\mathbf{S}$ in descending order. The linear transformation given in Eq. (16) simply recasts the original variables into a set of new uncorrelated variables, whose coordinate axes are described by $\mathbf{A}$. The projection of the original state-space, $\mathbf{Y}$, onto the matrix $\mathbf{A}$ defines a new set of variables $\mathbf{Z}$ referred to as principal component scores,

$$\mathbf{Z} = \mathbf{Y}\mathbf{A}. \tag{17}$$

The matrix $\mathbf{A}$ can be truncated to keep only the first q columns, associated to the q largest eigenvalues to reduce the dimensionality of the state space. The corresponding $(Q \times q)$ matrix can be used to obtain a reduced set of principal component scores, $\mathbf{Z}_q$, carrying most of the variance originally contained in $\mathbf{Y}$,

$$\mathbf{Z}_q = \mathbf{Y}\mathbf{A}_q. \tag{18}$$

This equation can be inverted to obtain an approximation of the state space $\mathbf{Y}$, based on the q most energetic PCs,

$$\tilde{\mathbf{Y}}_q = \mathbf{Z}_q \mathbf{A}_q^T. \tag{19}$$

The implementation of the PC-score approach requires the solution of transport equations which are formally very similar to the ones of classic nonconserved scalars, i.e., reacting species. More generally, if a set of transport equations exists under the following conservative form,

$$\frac{\partial}{\partial t}\rho\mathbf{y} + \boldsymbol{\nabla}\cdot(\rho\mathbf{u}\otimes\mathbf{y}) = \omega_y, \tag{20}$$

then it can be rewritten in the score space as follows [39]:

$$\frac{\partial}{\partial t}\rho\mathbf{z} + \boldsymbol{\nabla}\cdot(\rho\mathbf{u}\otimes\mathbf{z}) = \omega_z. \tag{21}$$

In these relations, $\mathbf{y}$ is a mass fraction for a single species and $\mathbf{z}$ is a single principal component, which are each individual realizations of the vectors $\mathbf{Y}$ and $\mathbf{Z}$ respectively. The species source terms should be transformed by using the truncated matrix of eigenvectors $\mathbf{A}_q$,

$$\omega_Z = \omega_Y \mathbf{A}_{\boldsymbol{q}}. \tag{22}$$

Notice that the transformed source terms can be nonlinear in the $\mathbf{Z}$ variables, since they are calculated by projecting the matrix of eigenvectors $\mathbf{A}_q$ onto the original source terms that can be nonlinear functions in the reconstructed variables.

B. Rotated scores

The objective of using a rotation method with PCA is to maximize the variance expressed by the eigenvectors. Using such a rotation method simplifies the interpretation of the PCA results as the score loadings after rotation will be either very large or close to zero. Consequently, it clearly shows which variables are expressed the most by the scores, and thus which are the dominating ones, for a given size. It is important to note that the total expressed variance does not change before and after rotation. Besides a better interpretation of the PCA results, a rotation method also ensures additional stability within the CFD code. After rotation, the data flatten out and contain fewer peaks, which is beneficial for the computation of the source terms, for example, as stated by Coussement *et al.* [40]. It ensures robustness within the code.

The VARIMAX rotation developed by Kaiser [37] has generally been accepted as the most accurate orthogonal rotation and has been used widely in combination with PCA [41]. The VARIMAX rotation criterion maximizes the sum of the variances of the squared coefficients within each eigenvector. The axes in the new system are rotated to maximize the rotation criterion given by the following expression:

$$V = \frac{Q\sum_{i=1}^{Q}\left(b_{ij}^2\right)^2 - \left(\sum_{i=1}^{Q} b_{ij}^2\right)^2}{Q^2}. \tag{23}$$

In this expression, b stands for the principal component loadings and Q stands for the number of variables, with $i \in \{1,\ldots,Q\}$ and $j \in \{1,\ldots,q\}$ with q being the number of scores or principal components.

A normalized variant of this criterion exists. The b_{ij} terms in expression (23) should be normalized by the square root of the communalities h_i. The communalities are defined as the sum of the squares of the ith row of the loading matrix. This normalized variant has been used in the present work:

$$V_{\text{scaled}} = \frac{Q\sum_{i=1}^{Q}\left(b_{ij}^2/h_i^2\right)^2 - \gamma\left(\sum_{i=1}^{Q} b_{ij}^2/h_i^2\right)^2}{Q^2}. \tag{24}$$

For working with rotated scores, the rotation matrix $\mathbf{T}$ should be taken into account when transforming the conserved variables and source terms to the new score space:

$$\mathbf{Z}_q = \mathbf{YTA}_q, \tag{25}$$

$$\omega_Z = \omega_Y \mathbf{TA}_q. \tag{26}$$

IV. RESULTS

The set of three free-stream conditions presented in Table IV will allow us to demonstrate the applicability of the PCA-score technique to plasma flows. The SHOCKING code has been modified to solve the equations in the score space. Only the species transport equations need to be projected onto the eigenvector space:

$$\frac{\partial}{\partial x}[\rho u \mathbf{Z}_i] = \omega_{Zi}, \tag{27}$$

with $i \in \{1, \ldots, q\}$. It is important to notice we need all species mass fractions to calculate the thermodynamic and chemical production terms of the plasma. After each iteration of the solver, the scores are reverted to mass fractions by using Eq. (19).

A. PCA scores for re-entry conditions

In a next step, the necessary amount of scores is determined for each test case. Case 1 corresponds to the validating UTIAS shock tube experiments. Case 2 uses half a lower pressure, and test case 4 approaches re-entry conditions. Ideally, the case should run using only one score. The main criterion for retaining the right amount of scores is the following: They should represent the original manifold in a perfect way; i.e., all the populations should be reconstructed with high accuracy. This can be important for reproducing radiative spectra. The ideal number of scores required to reconstruct with sufficient accuracy the original state space can be determined with an *a priori* study comparing the reconstruction error of each variable after PCA. Of course, this is not a sufficient condition as the actual solution of the score transport equations in a CFD code leads to error propagation, resulting in a higher number of scores needed. Figure 5 presents the results of the PCA-score reduction for case 1. The left picture shows the temperature evolution after the shock. The results based on three scores show major differences in the shape of the electron temperature profile and in the relaxation time. Moreover, the equilibrium temperature is wrong because the composition is not accurately computed. The discrepancies are less pronounced when using four scores. The results obtained with five scores or more are in excellent agreement with the full CR model. Figure 5(b) shows the ionization degree of the plasma. A very good agreement is also observed starting with five scores. The dimensionality reduction is impressive as the species equations have been reduced from 34 to only five.

The reduction of case 2 shown in Fig. 6 is very similar to the previous one as only the pressure has been lowered. Notice that equilibrium temperature is already reached with only three scores, as opposed to case 1. The results for case 4 are shown in Fig. 7. In the case, the speed has been increased and the pressure lowered to match typical atmospheric re-entry situations. As the free-stream conditions are more severe, the solver requires more scores to retrieve an accurate

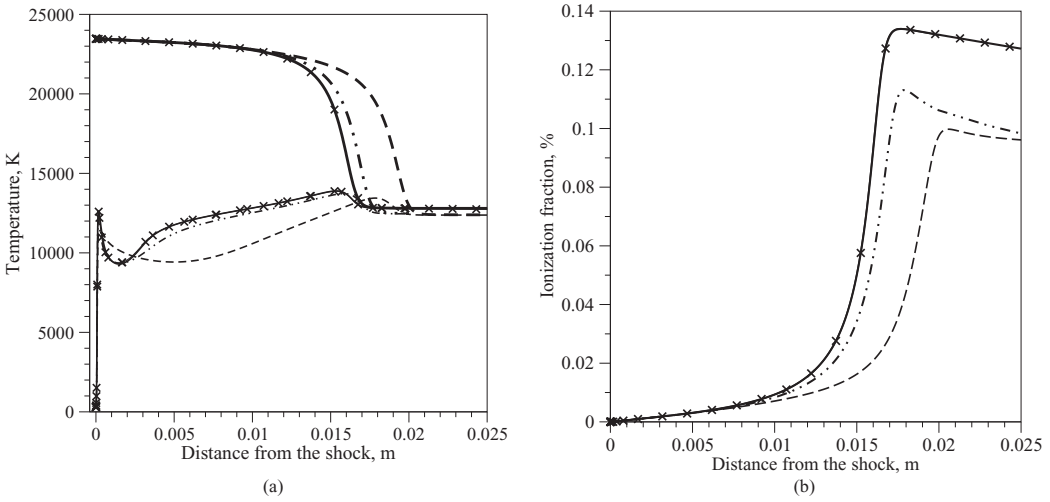


FIG. 5. Postshock simulation case 1: −−, 3 scores; − · ·−, 4 scores; ——, 5 scores; ×, full simulation. (a) Temperatures: thick lines T_h, thin lines T_e. (b) Ionization degree.

solution. The PCA-scores solution is identical to the full CR model with six scores for the temperature field and 13 scores for the ionization degree.

B. PCA scores with VARIMAX rotation

A VARIMAX rotation on the eigenvectors of the data has been carried on all the cases as it increases the robustness of the code. Moreover, rotated eigenvectors and loadings are easier to interpret as the variance expressed by them has been maximized. To illustrate this, the loadings for the first two scores have been compared against their rotated ones for case 2 in Fig. 8 for score 1 and in Fig. 9 for score 2. As a reminder, the loadings correspond to the weights of the original variables expressed by the principal components or scores. Figure 8 shows that score 1 is mainly composed

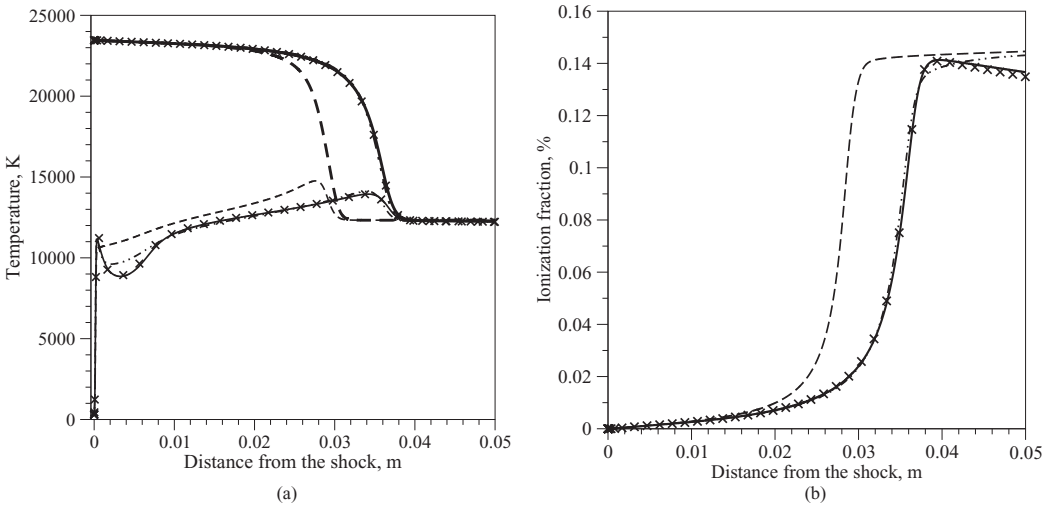


FIG. 6. Postshock simulation case 2: −−, 3 scores; − · ·−, 4 scores; ——, 5 scores; ×, full simulation. (a) Temperatures: thick lines, T_h; thin lines, T_e. (b) Ionization degree.

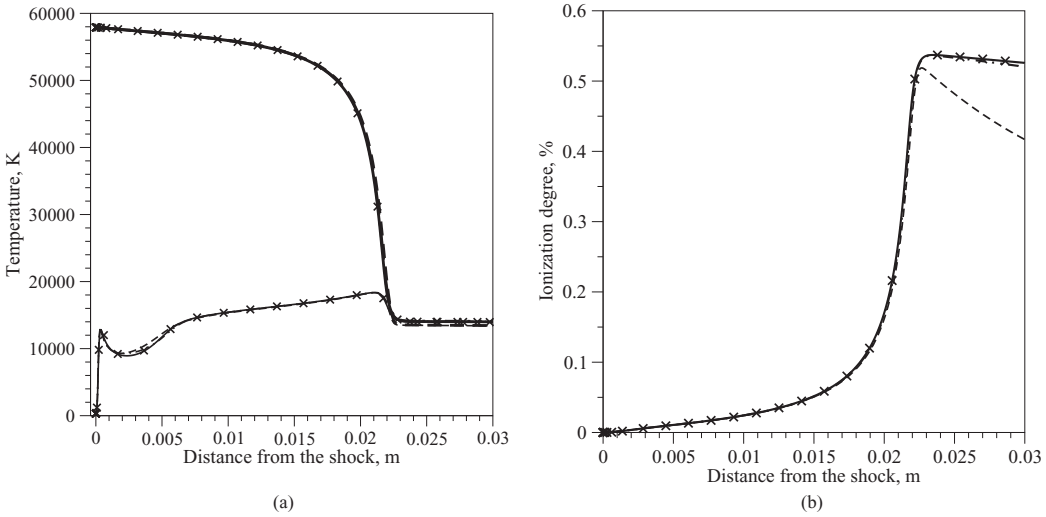


FIG. 7. Postshock simulation case 4: −−, 4 scores; − · ·−, 6 scores; ——, 13 scores; ×, full simulation. (a) Temperatures: thick lines, T_h; thin lines, T_e. (b) Ionization degree.

of the ground state of argon and the ionized species. This information was concealed within the unrotated scores. Figure 9 shows that all the species contribute to score 2.

As mentioned in Sec. III B, the rotation operation flattens the new source terms for calculating the mass conservation of the scores. Figure 10 represents the source terms for score 1 before and after rotation. The red curve, corresponding to the rotated scores, shows a significantly less noisy behavior than the nonrotated ones. This is an important results as it shows how VARIMAX rotation can improve the stability of the method within CFD codes.

C. Manifold sensitivity study

Previous tests only covered test cases 1, 2, and 4. To complete the testing grid, two additional cases have been reduced in the (p,v) parameter study: test case 3 and test case 5, as shown in Table IV. The preshock settings of case 3 match those of case 2. The results are similar. The best reduced model uses only four scores out of 34, which corresponds to a model reduction of 88%. Test case 5 has a free-stream velocity of 8 km/s combined with a high pressure of 685 Pa (matching

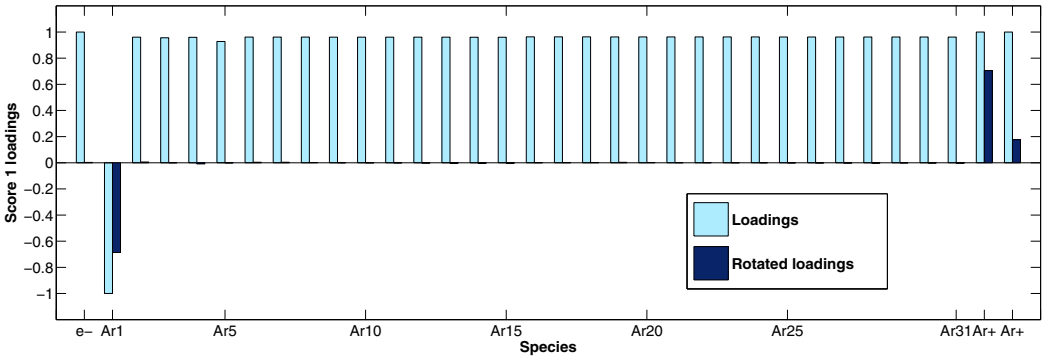


FIG. 8. Score 1 loadings for the original variables before and after VARIMAX rotation, case 2. Light blue bars, original loadings. Dark blue bars, rotated loadings.

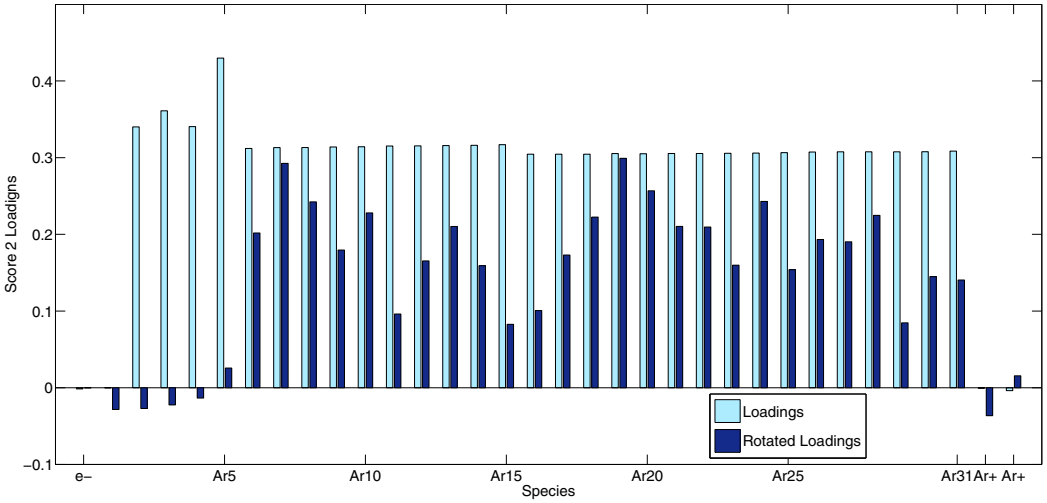


FIG. 9. Score 2 loadings for the original variables before and after VARIMAX rotation, case 2. Light blue bars, original loadings. Dark blue bars, rotated loadings.

UTIAS cases). Also in this case, six scores were needed to represent the full model, corresponding to a model reduction of 82%.

The reduction potential of the PC-score approach has been demonstrated. In a next step, we want to investigate the applicability of a reduced model in different free-stream conditions. The reduced model is built on sample data form a 1D calculation. This simulation was performed using certain free-stream conditions. The objective is to change these preshock conditions and to assess if the PCA-based model still converges with high accuracy. The limits of the manifold are represented in Fig. 1, where the pressure varies from 50 to 700 Pa and the velocity varies from 5 to 8 km/s. The objective of this study is to investigate if a reduced model, based on particular training data given by the cases 1, 2, or 4, is still accurate using other free-stream parameters. To this purpose, the coefficient of determination R is evaluated between the mass fractions obtained with the full CR model and the reduced model. For $R^2 = 1$, the reduced model is in perfect accordance with the full one.

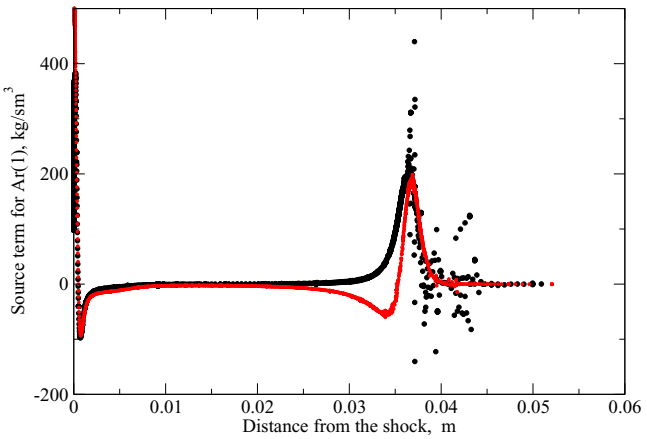


FIG. 10. Comparison of mass production source term for the first score with and without VARIMAX rotation. Black dots, original loadings; red dots, rotated loadings.

TABLE VI. Coefficient of determination for comparing the reduced models for cases 1, 2, and 4 in the (p,v) parameter space.

	Case 1	Case 2				Case 4			
p [Pa]	5 [km/s]	5 [km/s]	6 [km/s]	7 [km/s]	8 [km/s]	5 [km/s]	6 [km/s]	7 [km/s]	8 [km/s]
150		0.999	0.999	0.995		0.999	0.998	0.957	
200	0.869	0.991	0.999	0.999	0.969	0.981	0.999	0.997	0.869
250	0.976	0.851	0.991	0.999	0.999	0.920	0.996	0.999	0.975
300	0.990	0.532	0.950	0.999	0.999	0.877	0.981	0.999	0.990
350	0.986	0.212	0.855	0.998	0.999		0.961	0.999	0.986
400	0.991		0.720	0.992	0.999		0.941	0.999	0.991
450	0.994		0.569	0.981	0.999		0.924	0.997	0.994
500	0.996		0.423	0.960	0.999		0.910	0.993	0.996
550	0.997		0.290	0.931	0.999		0.898	0.989	0.997
600	0.998		0.172	0.893	0.999			0.984	0.998
650	0.998		0.0689	0.849	0.999			0.979	0.998
700	0.999			0.800	0.999			0.973	0.999

Table VI represents the R^2 error for all the cases in the (p,v) space. The general trend is that a reduced model can be used outside its training conditions, as long as the new free-stream parameters are less severe than the original ones. For instance, the reduced model derived from case 4 with 8 km/s preshock velocity can be used for all pressures up to 200 Pa with high accuracy. This accuracy also applies for conditions with similar preshock speeds.

Next, we investigate the possibility of generating reduced models from a finite number of cases spanning the complete domain of interest, in terms of speed and pressure, and to use them for conditions not originally included in the training dataset. Cases 1 and 3 are used to generate a reduced model that is later used to simulate the conditions of case 2. The PCA-score method has been applied to the combined data sets for cases 1 and 3 to find a global reduced model. The approach is found to be very effective, as it provides excellent results for case 2 as shown in Fig. 11. This is

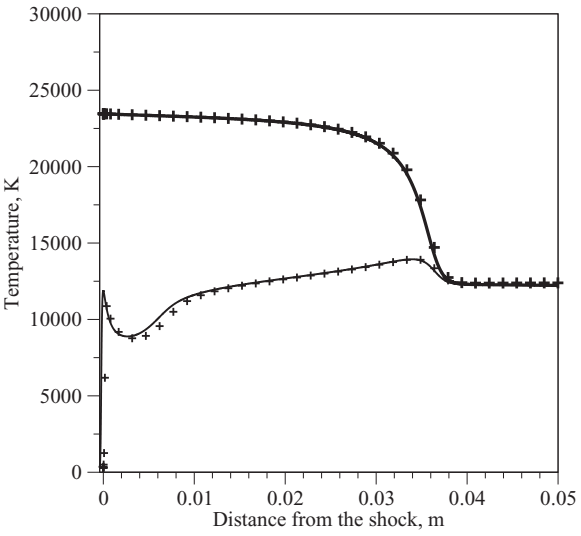


FIG. 11. Translational temperatures for case 2. Thick lines, T_h; thin lines, T_e; +, CR results for case 2; ——, interpolated score model based on the cases 1 and 3.

a good strategy for generating reduced models for CFD applications, since the operating conditions might vary in terms of pressure and velocity within a single simulation, and therefore a case built from a span of conditions is needed.

V. CONCLUSIONS

The present work shows how principal component analysis can be used to reduce large and complex chemistry models, such as this 34-species collisional-radiative argon plasma model. Moreover, it shows how the reduction techniques developed for combustion applications can easily be transferred to the plasma community. PCA score is able to provide a reduced model reproducing complex postshock properties such as the temperatures, populations, and ionization fraction in an accurate way. This is not always the case when using alternative reduction techniques which are based on coarse-grain models and multitemperature approximations or on time-scale-based reductions. PCA does not require in-depth knowledge of the physics of the reacting system to be able to find a reduced model. The technique is straightforward and automatic, and therefore user-friendly. PCA can also be used as a tool to study the physics of the problem, as it gives insight into the dynamics of the reacting system. A summary of the important conclusions drawn in this paper are presented:

(1) The 34 original variables representing the mass fractions of the species could be reduced to three scores using PCA. These three scores represent a linear combination of the original variables and maximize the variance in the system. This is an encouraging result as the set of governing species equations has been reduced by more than 90%, which leads to an important speed-up of the calculation and a reduction of computational cost. When changing the free-stream conditions to more severe parameters, i.e., low pressures and high speeds, the number of scores to represent detailed physics increases.

(2) PCA score has proven to be robust and more stable in combination with a rotation technique, such as the VARIMAX criterion. Another advantage of rotating the transformation matrix is the simplification of the interpretation of the principal components or scores.

(3) The sensitivity study has shown that a given reduced model can be used in a large frame of free-stream conditions outside the initial ones while conserving good accuracy. This means only one training data sample is needed for building a reduced model valid in a various free-stream conditions.

(4) Besides extrapolating the model, it can also be built by interpolating data. By combining two data sets, one can build an interpolated model for intermediate conditions. This property is very useful when dealing with CFD.

(5) As a concluding remark, we would like to point out PCA can also be used for more complex reacting schemes which include a higher level of nonlinearity. As only one linear system may not be sufficient to represent the entire manifold, the method can be applied locally in clusters to combine several linear representations. Each cluster has its local linear representation in an individual PCA reduced model.

ACKNOWLEDGMENTS

The research of A. Bellemans has been sponsored by the Belgian research fund F.R.S.-FNRS through a FRIA fellowship. The last author would like to acknowledge the support of Fédération Wallonie-Bruxelles, via Les Actions de Recherche Concertée (ARC) call for 2014–2019 to support fundamental research.

[1] D. M. Goebel and I. Katz, *Fundamentals of Electric Propulsion: Ion and Hall Thrusters* (Wiley, New York, 2008).

[2] M. B. Thomas, D. Rafalskyi, T. Lafleur, and A. Aanesland, Experimental investigation of electron transport across a magnetic field barrier in electropositive and electronegative plasmas, Plasma Sources Sci. Technol. **25**, 045018 (2016).

[3] S. V. Pancheshnyi, D. A. Lacoste, A. Bourdon, and C. O. Laux, Ignition of propane-air mixtures by a repetitively pulsed nanosecond discharge, IEEE Trans. Plasma Sci. **34**, 2478 (2006).
[4] V. Adamovich and W. R. Lempert, Challenges in understanding and predictive model development of plasma-assisted combustion, Plasma Phys. Control. Fusion **57**, 014001 (2015).
[5] C. Park, *Nonequilibrium Hypersonic Aerothermodynamics* (Wiley, New York, 1990).
[6] B. Helber, A. Turchi, J. B. Scoggins, A. Hubin, and T. E. Magin, Experimental investigation of ablation and pyrolysis processes of carbon-phenolic ablators in atmospheric entry plasmas, Int. J. Heat Mass Trans. **100**, 810 (2016).
[7] H. Goedbloed and S. Poedts, *Principles of Magnetohydrodynamics* (Cambridge University Press, Cambridge, UK, 2004).
[8] A. Alvarez Laguna, A. Lani, H. Deconinck, N. N. Mansour, and S. Poedts, A fully implicit finite-volume method for multi-fluid reactive and collisional magnetized plasmas on unstructured meshes, J. Comput. Phys. **318**, 252 (2016).
[9] A. Bultel, B. G. Chéron, A. Bourdon, O. Motapon, and I. F. Schneider, Collisional-radiative model in air for earth re-entry problems, Phys. Plasmas **13**, 043502 (2006).
[10] M. G. Kapper and J.-L. Cambier, Ionizing shocks in argon, part 1: Collisional-radiative model and steady-state structure, J. Appl. Phys. **109**, 113308 (2011).
[11] M. G. Kapper and J.-L. Cambier, Ionizing shocks in argon, part 2: Transient and multi-dimensional effects, J. Appl. Phys. **109**, 113309 (2011).
[12] A. Bogaerts, R. Gijbels, and J. Vlcek, Collisional-radiative model for an argon glow discharge, J. Appl. Phys. **84**, 121 (1998).
[13] M. Panesi, T. E. Magin, A. Bourdon, A. Bultel, and O. Chazot, Fire II flight experiment analysis by means of a collisional-radiative model, J. Thermophys. Heat Transfer **23**, 236 (2009).
[14] J. Annaloro and A. Bultel, Elaboration of collisional-radiative models for flows related to planetary entries into the Earth and Mars atmospheres, Plasma Sources Sci. Technol. **22**, 025008 (2013).
[15] A. Guy, A. Bourdon, and M.-Y. Perrin, Consistent multi-internal-temperatures models for nonequilibrium nozzle flows, Chem. Phys. **420**, 15 (2013).
[16] M. Capitelli, G. Colonna, G. D'Ammando, K. Hassouni, A. Laricchiuta, and L. D. Pietanza, Coupling of plasma chemistry, vibrational kinetics, collisional-radiative models, and electron energy distribution function under non-equilibrium conditions, Plasma Process. Polymers **14**, 1600109 (2017).
[17] R. L. Macdonald, A. Munafo, C. O. Johnston, and M. Panesi, Nonequilibrium radiation and dissociation of CO molecules in shock-heated flows, Phys. Rev. Fluids **1**, 043401 (2016).
[18] T. E. Magin, M. Panesi, A. Bourdon, R. L. Jaffe, and D. W. Schwenke, Coarse-grain model for internal energy excitation and dissociation of molecular nitrogen, Chem. Phys. **398**, 90 (2012).
[19] M. Panesi and A. Lani, Collisional radiative coarse-grain model for ionization in air, Phys. Fluids **25**, 057101 (2013).
[20] A. Munafò and T. E. Magin, Modeling of stagnation-line nonequilibrium flows by means of quantum-based collisional models, Phys. Fluids **26**, 097102 (2014).
[21] H. P. Le, A. R. Karagozian, and J. L. Cambier, Complexity reduction of collisional-radiative kinetics for atomic plasma, Phys. Plasmas **20**, 123304 (2013).
[22] A. Munafò, M. Panesi, and T. E. Magin, Boltzmann rovibrational collisional coarse-grained model for internal energy excitation and dissociation in hypersonic flows, Phys. Rev. E **89**, 023001 (2014).
[23] G. Colonna, G. DAmmandob, L. D. Pietanza, and M. Capitelli, Radiation transfer, level and free electron kinetics in nonequilibrium atomic hydrogen plasma, in *Proceedings of the 27th International Symposium on Rarefied Gas Dynamics*, edited by Deborah A. Levin, Ingrid J. Wysong, and Alejandro L. Garcia, Vol. 1333 (AIP, Pacific Grove, CA, 2010).
[24] J. Shlens, A tutorial on principal component analysis, arXiv:1404.1100v1.
[25] A. Parente, J. C. Sutherland, L. Tognotti, and P. J. Smith, Identification of low-dimensional manifolds in turbulent flames, Proc. Combust. Inst. **32**, 1579 (2009).
[26] A. Coussement, O. Gicquel, and A. Parente, Kernel density weighted principal component analysis of combustion processes, Combust. Flame **159**, 2844 (2012).

[27] K. Peerenboom, A. Parente, T. Kozak, A. Bogaerts, and G. Degrez, Dimension reduction of non-equilibrium plasma kinetic models using principal component analysis, Plasma Sources Sci. Technol. **24**, 025004 (2014).
[28] A. Bellemans, A. Munafo, T. Magin, G. Degrez, and A. Parente, Reduction of a collisional-radiative mechanism for argon plasma based on principal component analysis, Phys. Plasmas **22**, 062108 (2015).
[29] I. I. Glass, *Over forty years of continuous research at UTIAS on nonstationary flows and shock waves* (Springer, New York, 1991).
[30] J. Vlcek, A collisional-radiative model applicable to argon discharges over a wide range of conditions, I: Formulation and basic data, J. Phys. D **22**, 623 (1989).
[31] A. Bultel, B. van Ootegem, A. Bourdon, and P. Vervisch, Influence of Ar_2^+ in an argon collisional-radiative model, Phys. Rev. E **65**, 046406 (2002).
[32] T. E. Magin, L. Caillault, A. Bourdon, and C. O. Laux, Nonequilibrium radiative heat flux modeling for the huygens entry probe, J. Geophys. Res. **111**, E07S12 (2006).
[33] A. Munafò, Multi-scale model and computational methods for aerothermodynamics, Ph.D. thesis, Ecole Centrale Paris, Paris, 2014.
[34] K. Radhakrishnan and A. C. Hindmarsh, Description and use of LSODE, the Livermore solver for ordinary differential equations, NASA Report 1327, NASA Lewis Research Center; Cleveland, OH, United States, 1993 (unpublished).
[35] J. Sutherland and A. Parente, Combustion modeling using principal component analysis, Proc. Combust. Inst. **32**, 1563 (2009).
[36] A. Coussement, A. Parente, and O. Gicquel, Mg-local-PCA method for reduced order combustion modeling, Proceed. Combust. Inst. **34**, 1117 (2013).
[37] H. F. Kaiser, The VARIMAX criterion for analytic rotation in factor analysis, Psychometrika **23**, 187 (1958).
[38] B. Isaac, Reduced-order modeling for reacting flows based on principal component analysis, Ph.D. thesis, University of Utah, Salt Lake City, 2014.
[39] A. Parente and J. Sutherland, Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity, Combust. Flame **160**, 340 (2013).
[40] A. Coussement, B. J. Isaac, O. Gicquel, and A. Parente, Assessment of different chemistry reduction methods based on principal component analysis: Comparison of the MG-PCA and score-PCA approaches, Combust. Flame **168**, 83 (2016).
[41] M. Richman, Rotation of principal components, J. Climatol. **6**, 293 (1986).