NON-LINEAR REGRESSION OF THE THERMOCHEMICAL STATE-SPACE ONTO A REDUCED NUMBER OF PRINCIPAL COMPONENTS

M.R. Malik*, B. Isaac, A. Coussement*, A. Parente***
Rafi.Malik@ulb.ac.be
* Université Libre de Bruxelles, Ecole polytechnique de Bruxelles, Aero-Thermo-Mechanics Department, Bruxelles, Belgium
** University of Utah, Department of Chemical Engineering, Salt Lake City, UT, USA

Abstract
Large kinetic mechanisms are required in order to model accurately combustion systems. If no parameterization of the thermo-chemical state-space is used, solution of the species transport equations can become computationally prohibitive as the resulting system contains a wide range of time and length scales. Parameterization of the thermo-chemical state-space without an a priori prescription of the dimension of the underlying manifold would lead to a reduced yet accurate description. To this end, the potential offered by Principal Component Analysis (PCA) in identifying low-dimensional manifolds is very appealing. The present work seeks to advance the understanding and application of the PC-transport approach by analyzing the ability to parameterize the thermo-chemical state with the PCA basis using non-linear regression. In order to demonstrate the accuracy of the method within a numerical solver, unsteady perfectly stirred reactor (PSR) calculations are shown using the PC-transport approach. The PSR analysis extends previous investigations by the Authors to more complex fuels (e.g. methane, propane), showing the ability of the approach to deal with relatively large kinetic mechanisms. The ability to achieve highly accurate mapping through Gaussian Process based nonlinear regression is also shown. In addition, a novel method based on local regression of the PC source terms is also investigated.

Introduction
The numerical modelling of turbulent combustion is a very challenging task as it combines the complex phenomena of turbulence and chemical reactions. This study becomes even more challenging when large detailed kinetic mechanisms are used in order to understand some special features such as pollutant formation. The coupling of the kinetic equations with the physical features of the problem results in a set of partial differential equations which is often very large to solve and not adequate for the current means. In a CFD calculation, the number of species tracked impacts the memory usage and CPU time. It is thus important to

minimize this number by the use of a simpler but representative set of variables. Therefore, there is a need for methods allowing to efficiently parameterizing the thermo-chemical state of a reacting system with a reduced number of optimal reaction variables. Among those, Principal Component Analysis (PCA) appears as an ideal candidate to fulfil the purpose. PCA offers the possibility of automatically reducing the dimensionality of data sets consisting of a large number of correlated variables, while retaining most of the variation present in the original data. After reduction, the new set of variables, called principal components (PCs), are orthogonal, uncorrelated and linear combinations of the original variables (T, p *and* Y_i). By retaining the PCs containing most of the variance and transporting them in a numerical simulation, the dimensionality of the system can be highly reduced. Another advantage of PCA reside in the fact that the PCs can be obtained through data sets based on simple systems (such as canonical reactors) and then applied to a similar, more complex system.
The present work seeks to advance the understanding and application of the PC-transport approach by applying this method to more complex fuels such as methane and propane. This analysis is carried out on the solution of an unsteady PSR calculation using the PC-transport approach for large kinetic mechanisms. Next, the PC-transport approach is coupled with non-linear regression (GPR) in order to increase the size reduction potential of PCA. Finally, the PC-transport approach is enhanced by the use of local regression (GPR).

PC-transport Approach

In the work of Sutherland and Parente [1], a combustion model is proposed where conservation equations for the PCs are derived from the general species transport equation:

$$\frac{\partial}{\partial t}\left(\rho Z_q\right) + \frac{\partial}{\partial x_i}\left(\rho u_i Z_q\right) = \frac{\partial}{\partial x_i}\left(D_{Z_q}\frac{\partial}{\partial x_i}\left(Z_q\right)\right) + s_{Z_q}$$

$$s_{Z_q} = \frac{1}{\rho}\sum_{k=1}^{Q}\frac{R_k}{d_k}A_{kq}$$

In PCA-based combustions models, one of the major weaknesses is that a linear model is trying to describe a highly non-linear process. In order to take full advantage of the PC analysis, a nonlinear mapping to the linear underlying surface by using nonlinear regression was proposed [2]. This allows to fully utilize the underlying manifold identified by the principal component analysis. The linear basis derived from the PC analysis allows deriving simple transport equations, while using nonlinear functions within this basis allows capturing the nonlinearities which are always present in combustion systems.

Gaussian Process Regression

In this study, the state-space variables $T, p \; and \; Y_i$ and the PC source terms s_{Z_q} are mapped to the PC basis using nonlinear regression:

$$\phi \approx f_\phi(Z_q)$$

where f_ϕ is the nonlinear regression function and ϕ represent the dependant variables. Gaussian Processes (GPs) [3, 4] does not assume a specific model for the regression function. By doing so, GPs are less parametric and let the data speak for themselves. A Gaussian process generates data in the domain of interest such that any finite subset of the range follows a multivariate Gaussian distribution. The dependant variables can thus be described by a Gaussian distribution:

$$\phi \approx GP(m(x), K(x, x'))$$

where m is a mean function and K is a covariance function (or kernel). The mean function is often assumed to be zero. The covariance function used here is the Squared Exponential:

$$K(x, x') = {\sigma_f}^2 exp\left[\frac{-(x - x')^2}{2l^2}\right]$$

with ${\sigma_f}^2$ being the signal variance and l the characteristic length scale. These two parameters of the covariance function are called hyper-parameters. After an initial guess, those hyper-parameters are optimized using a Gaussian likelihood function.

Local Regression

In order to improve further the accuracy of the regression and increase PCA's potential for size reduction, a novel approach will be tested where the PC-score approach is coupled with locally regressed state-space variables (PC-L-GPR). The idea is to divide the PC state-space into bins or clusters, and to perform a GP regression separately in each of these bins. As a consequence, a better regression would be obtained (if each bin is chosen appropriately) and the computational time required for GPR will also be reduced. In order to define such bins, a conditioning variable has to be chosen. This variable should well characterize the state-space. Possible candidates are the PCs source terms, as the latter are highly non-linear over the PC space. Clustering the source terms manifolds such as they can be approximated by quasi-linear functions in each bin would simplify and accelerate the regression algorithm.

Results

The proposed method is demonstrated in a PSR, comparing the calculations using the full set of equations to the standard PC-transport approach, and the PC-transport approach using nonlinear regression, and this for two different fuels: a simple one, methane (CH_4), and a more complex, propane (C_3H_8).

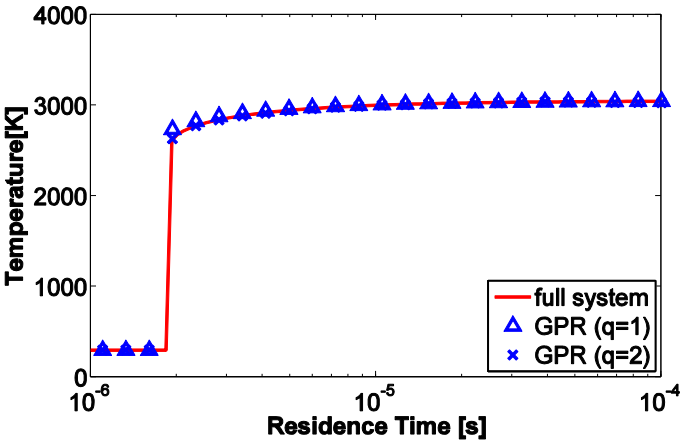

Figure 1: PSR temperature as a function of the residence time (methane fuel, GRI-3.0 mechanism)

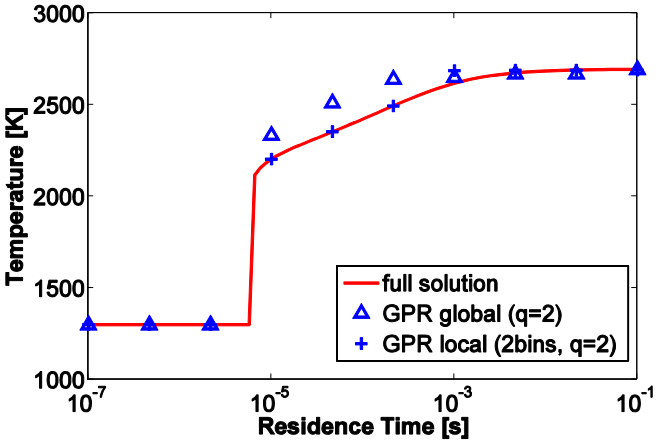

Figure 2: PSR temperature as a function of the residence time (propane fuel, San Diego mechanism)

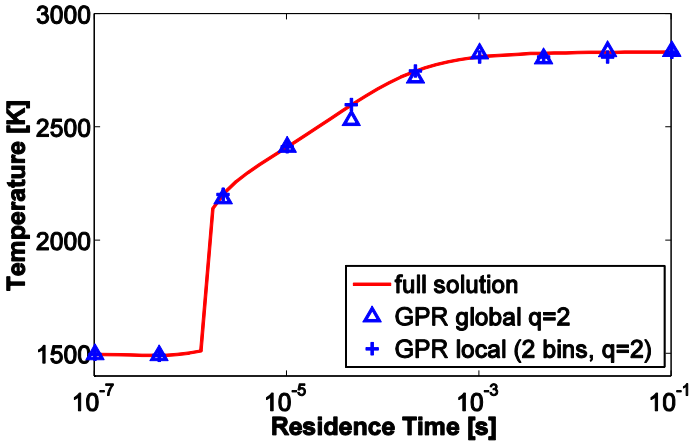

Figure 3: PSR temperature as a function of the residence time (propane fuel, Polimi mechanism)

Figure 1 show the solution using the PC-score model together with GPR (PC-GPR) for the methane case (with GRI-3.0 mechanism) while using 1 and 2 PCs. The result shows remarkable accuracy for the model with regression using only 2 PCs over the range of residences times for the predicted temperature. This conclusion also holds for both major and minor species.

Figures 2 and 3 show the solution using GPR for the propane cases with the San Diego and the Polimi mechanisms, respectively, comparing both the global and local regression approaches. Local clustering in these cases was performed based on the minima/maxima of the first PC source term only. It can be seen that using the local approach improves significantly the accuracy of the model, especially in the ignition/extinction region, leading to an almost perfect match, especially for the Polimi case. For the San Diego scheme, results using local regression are better than the global regression, but the ignition/extinction zone is still challenging, even with locally regressed manifolds. To further improve this, the clustering method was extended to the second PC source term as well, leading to a better regression for that conditioning variable too. Figure 4 show the solution for the San Diego case using local regression based on the clustering of the first PC source term only, and the solution based on the clustering of the second source term as well. It can be seen that when also using the second source term as a conditioning variable, which contains some of the system's remaining variance not included in the first source term, improves significantly the performance of the model, in all regions of the solution.

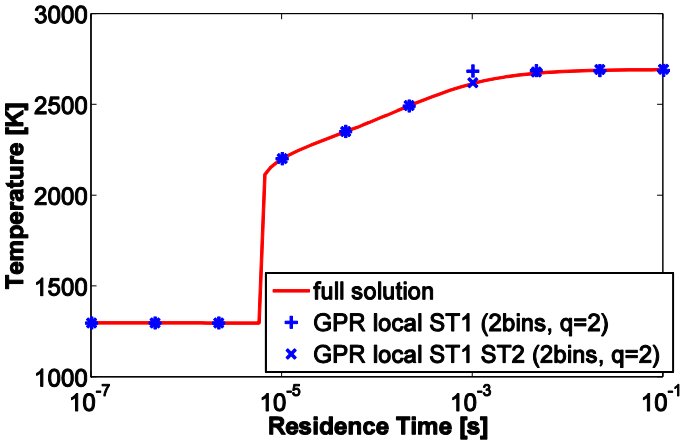

Figure 4: PSR temperature as a function of the residence time (propane fuel, San Diego mechanism)

References

[1] J. C. Sutherland, A. Parente, Proceedings of the Combustion Institute 32 (2009) 1563–1570.

[2] A. Biglari, J. C. Sutherland, Combustion and Flame (2012).

[3] C. Rasmussen, Gaussian processes for machine learning, 2006.

[4] D. Nguyen-Tuong, M. Seeger, J. Peters, Advanced Robotics (2009).

doi: 10.4405/39proci2016.VI3