Data Article

Datasets for high hydrogen content syngas fuel variability effect on combustion physicochemical properties

Kai Zhang\textsuperscript{a}, Xi Jiang\textsuperscript{b,*}

\textsuperscript{a} Linné Flow Center, Department of Mechanics, Royal Institute of Technology (KTH), Osquars Backe 18, SE-10044, Stockholm, Sweden
\textsuperscript{b} School of Engineering & Materials Science, Queen Mary University of London, Mile End Road, London, E1 4NS, UK

\textbf{Article info}

\textbf{Article history:}
Received 29 November 2019
Accepted 3 January 2020
Available online 11 January 2020

\textbf{Keywords:}
UQ
Combustion
Syngas
Sensitivity analysis
Fuel variability

\textbf{Abstract}

The dataset presented in this article is related to the uncertainty quantification of fuel variability effect on high hydrogen content syngas combustion physicochemical properties. The 1D flame data included in this dataset are collected using PREMIX module available in Chemkin-Pro. Inputs to and outputs collected from the PREMIX module are generated and post-processed using UQTk-3.0.4, an open-access uncertainty quantification (UQ) toolkit developed at Sandia National Laboratories. The 1D flame data here refers to the calculation of flame speed, flame temperature, NO emission, etc. using three detailed chemical mechanisms: the GRI-Mech 3.0, the San Diego, and the NUI Galway Mechanism. The main analysis performed using UQTk-3.0.4 focuses on obtaining main and joint sensitivity effects (Sobol Indices) of uniformly distributed fuel uncertainty on 1D premixed physicochemical property. Other parameters such as the resulted probability density function or fluctuation of these properties are also explored. This new and original dataset is suitable for further analyzing fuel variability effect on other significant flame controlling parameters such as Karlovitz number, flame thickness, etc. in the discipline of turbulent combustion simulation.

© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

* Corresponding author.
E-mail address: xi.jiang@qmul.ac.uk (X. Jiang).

https://doi.org/10.1016/j.dib.2020.105116
2352-3409/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Here we describe part of the results generated with our datasets to help clarify the data structure. Some of the pre- and post-processing techniques are also described for ease to understand how data are distributed in each data consisting directory. The example figures produced from our datasets deal with that obtained using the GRI-Mech 3.0 mechanism. The data structure is the same for those obtained with other mechanism.

Four high hydrogen content syngas fuel mixtures are assumed namely H60CO30, H50CO40, H40CO50, and H30CO60, giving an example that H60CO30 represents a syngas-hydrogen mixture with 60%H2, 30%CO in addition to the other two fixed compositions of 5.5%CO2 and 4.5%CH4. Each composition except CH4 is assumed to contain a small variation of 1.5%, e.g., CO2 composition can vary from 4% to 7% in H60CO30, H2 can vary from 58.5% to 61.5% with their actual value sampled from a uniformly distributed PDF. The effect of these changes on flame physicochemical properties is of great importance, and hence uncertainty quantification method is employed for the investigation. Details of the UQ are described in the following section. The sampled data from assumed uniform PDFs are pre-

---

### Specifications Table

| Subject | Energy | Energy Engineering and Power Technology; Renewable Energy, Sustainability and the Environment |
|---------|--------|------------------------------------------------------------------------------------------|
| Specific subject area |        |                                                                                           |
| Type of data | Table and Figures | Numerical calculation of 1D premixed flame in PREMIX module available in Chemkin-Pro [1], pre- and post-process flame data in UQTk-3.0.4 [2], an open-access uncertainty quantification toolkit developed at Sandia National Laboratories. Three detailed chemical mechanisms: the GRI-Mech 3.0 [3], the San Diego [4], and the NUI Galway Mechanism [5] are used for calculating 1D flame physicochemical properties at various conditions. Other Python features are added to UQTk scripts to generate figures reported in this article. |
| How data were acquired |        |                                                                                           |
| Data format | Raw and analysis | Four cases are analyzed namely HXXCOXX, e.g. H60CO30 represents a syngas mixture with 60%H2, 30%CO in addition to the other two species with their mole fraction always set to constant: 5.5%CO2 and 4.5%CH4. Flame physicochemical properties at 6 equivalence ratios are calculated using PREMIX: 0.45, 0.5, 0.55, 0.6, 0.8, 1.0. Unburnt gas temperature and pressure are set at 400K, 1 atm. |
| Parameters for data collection |        | Description of data collection | Numerical calculation of 1D flame properties invoking pre- and post- statistical analysis. |
| Data source location | Institution: Queen Mary University of London | City/Town/Region: London | Country: United Kingdom |
| Data accessibility | Repository name: Mendeley data | Direct URL to data: https://doi.org/10.17632/67854cyf78.1 |
| Related research article | Zhang, K. and Jiang, X., 2019. Uncertainty quantification of fuel variability effects on high hydrogen content syngas combustion. Fuel, 257, p.116,111. https://doi.org/10.1016/j.fuel.2019.116111 |

### Value of the Data

- Uncertainty of syngas fuel variability often induces unstable operation issues of gas turbines. A quantified impact could help reduce such unstable problems by minimizing the uncertainty of dominating species from upstream gasification process.
- Power plants, gas turbine engineers can benefit from these data.
- Probability density functions (PDFs) for flame quantities are constructed via pushing 1,000,000 sample points into a constructed surrogate model. These data points can be further processed to understand the risky conditions of operating gas turbines, e.g. to calculate Karlovitz number which has an indication of flame stability in Borghi diagram [6].
- Many controlling parameters have been used in the presented datasets such as equivalence ratio, hydrogen content, etc. to construct surrogate models that can replace ODEs in 1D premixed code. The massive quantified impact of uncertainty on flame physicochemical properties such as CO2, CO, NO, etc. is of great importance in practical gas turbine operation.

### 1. Data description

Here we describe part of the results generated with our datasets to help clarify the data structure. Some of the pre- and post-processing techniques are also described for ease to understand how data are distributed in each data consisting directory. The example figures produced from our datasets deal with that obtained using the GRI-Mech 3.0 mechanism. The data structure is the same for those obtained with other mechanism.

Four high hydrogen content syngas fuel mixtures are assumed namely H60CO30, H50CO40, H40CO50, and H30CO60, giving an example that H60CO30 represents a syngas-hydrogen mixture with 60%H2, 30%CO in addition to the other two fixed compositions of 5.5%CO2 and 4.5%CH4. Each composition except CH4 is assumed to contain a small variation of 1.5%, e.g., CO2 composition can vary from 4% to 7% in H60CO30, H2 can vary from 58.5% to 61.5% with their actual value sampled from a uniformly distributed PDF. The effect of these changes on flame physicochemical properties is of great importance, and hence uncertainty quantification method is employed for the investigation. Details of the UQ are described in the following section. The sampled data from assumed uniform PDFs are pre-
processed based on a deterministic sampling method that instead of random sampling over the entire uncertain space, only typical points are sampled.

The sampled (or pre-processed) data are stored with explicit file name given as: 'Xch4_sample.txt', 'Xco2_sample.txt', 'Xco_sample.txt', and 'Xh2_sample.txt'. These files can be found in any of the sub-folders of the directory 'Data/Datasets/1atm_GRI-Mech 3.0_UQ/HXXCOXX/Eq = XXX' where XX represents the H₂ and CO mole fraction as described above, and XXX represents the equivalence ratios. For a typical HXXCOXX, though the four sampled data files can always be found in any of its sub-folders of 'Eq = XXX', the file with the same name always share the same content.

These sampled points are then passed through 1D premixed code in Chemkin to output flame physicochemical property data stored in the directory 'Data/Datasets/1atm_GRI-Mech 3.0_PREMIX/HXXCOXX/'. The data are named in the form 'Y_sample_YY.txt', where Y refers to quantity such as adiabatic flame temperature (T), flame speed (S), NO, CO₂, and NO₂ emissions (NO, CO₂, NO₂). YY represents the equivalence ratio at which these flame quantities are calculated, e.g. YY = 0.55, 0.6 and 10 represent equivalence ratio of 0.55, 0.6 and 1.0 respectively. These flame quantities are then passed back to the directory which stores sampled (or pre-processed) data according to the given subfolders name defined and are used for post-processing in UQTk 3.0.4. For instance, the NO_sample_055.txt is passed to the subfolder 'NO' in the directory 'Data/Datasets/1atm_GRI-Mech 3.0_UQ/HXXCOXX/Eq = 0.55'. All post-processed data, though stored in the same directory, are re-collected into other top-level directories for plotting useful figures with examples given as follows.

In the following, the ‘Fig. 1. Data_Figures_GRI-Mech _CO2_HCOratio’ refers to the figures created using post-processed data in the following directory: 'Data/Figures/GRI-Mech/CO2_HCOratio/'. The folder name ‘CO2_HCOratio’ means the data contained in the folder is for CO₂ emission (y-axis) with respect to H₂:CO ratio (x-axis). The file name in that directory is formatted as ‘XX_CO2_X.txt’ where XX is the mean or variance of CO₂, X can be 0, 1, 2, or 3 representing H₆₀CO₃₀, H₅₀CO₄₀, H₄₀CO₅₀, and H₃₀CO₆₀ in sequential order. Each file contains 6 rows of data representing the flame quantity at 6 equivalence ratios: 0.45, 0.5, 0.55, 0.6, 0.8 to 1.0.

The ‘Fig. 2. Data_Figures_GRI-Mech_CO2_PDF’ refers to the figure created using data in the following directory: 'Data/Figures/GRI-Mech/CO2_PDF/'. The folder name ‘CO2_PDF’ means the data contained in the folder is for CO₂ emission (x-axis) versus PDF (y-axis). The file name in that directory is formatted as PDF_X_axis_XX_XXX.txt where X represents x or y (data) of the figure, XX represents equivalence ratio, XXX indicates typical syngas-hydrogen mixture.
Hereafter we only describe figures without referring to the relevant directories unless it is exceptionally necessary. Fig. 3 shows the main sensitivity of CO₂ emission with respect to different mixture compositions. Three rows of data are available in each file, representing the main sensitivity of CO₂ emission with respect to H₂, CO, and CO₂ variability in the syngas-hydrogen mixture. Fig. 4 shows mean, standard deviation, and fluctuation of CO₂ emission. The fluctuation is defined as mean divided by standard deviation (square root of variance). Each row in the data file associates with a typical equivalence ratio.

The above-mentioned data structure also applies to those generated with NUI_Galway and San_Diego mechanism. Next, there is a special folder called ‘Cross_comparison’ in the directory ‘Data/Figures/’. Folder names are formatted in a similar way as that given above, the only exception is that the file names given in the folder are added with either ‘_N’, ‘_G’ or ‘_S’, each representing data collected from ‘NUI_Galway’, ‘GRI-Mech’, and ‘San_Diego’ folder in the directory ‘Data/Figures/’. Fig. 5 shows the main sensitivity comparison of adiabatic flame temperature data obtained using different chemical mechanisms.

2. Experimental design, materials, and methods

A PCE based surrogate model established based on a deterministic sampling (DM) method, is chosen to quantify the effect of high hydrogen content syngas fuel variability effect on its combustion thermophysical properties. The DM is recognized to converge faster than Monte-Carlo sampling (MCS) or Latin-Hypercube-Sampling (LHS) method [7]. The PCE is built for the convenience of performing variance-based sensitivity analysis (Sobol Indices [8]), that is, to answer the question of how the output uncertainty is affected by the input uncertainty on a variance base. Details of the PCE-DM are introduced below.
A general PCE representing a one-dimensional (1D) random variable (RV) \( u \) can be written as,

\[
u = \sum_{k=0}^{P} C_k H_{ek}(\xi) \quad \text{(1)}
\]

where \( C_k \) is the PC coefficient, \( H_{ek}(\xi) \) is 1D Hermite polynomial of order \( P \), and \( \xi \) is Gaussian RV in the range \([-\infty, \infty]\). The equation expresses a single variable as a set of Hermite functions \( H_{ek}(\xi) = (-1)^k e^{\xi^2/2} \frac{(d^k e^{\xi^2/2})}{d\xi^k} \) multiplied with deterministic PC coefficients \( C_k \) (refer to ‘c_k.txt’ in the directory where post-processing is performed). The Hermite polynomials are orthogonal with respect to the weight function \( w(\xi) = (1/\sqrt{2\pi})e^{-\xi^2/2} \), giving,

\[
\langle H_{ej} H_{ek} \rangle = \int_{-\infty}^{\infty} H_{ej}(\xi) H_{ek}(\xi) w(\xi) d\xi = \delta_{jk} \langle H_{e}^2 \rangle \quad \text{(2)}
\]
where \( \delta_{jk} \) is the Kronecker delta, and \(< >\) represents statistical expectation. For all \( j \neq k \), the expectation of the product of any two polynomials equals to zero. For \( j = k \) and \( u \approx \sum_{k=0}^{p} C_k H_{ek} \), the following expression for PC coefficient can be obtained from \( \langle H_{ek} u \rangle = \sum_{k=0}^{p} C_k \langle H_{ek} H_{ek} \rangle = C_k \langle H_{ek}^2 \rangle \) and \( \langle u H_{ek} \rangle = \int_{-\infty}^{\infty} u H_{ek}(\xi) w(\xi) d\xi \) (re-arrange equation (2)), giving,

\[
C_k = \frac{\langle u H_{ek} \rangle}{\langle H_{ek}^2 \rangle} = \frac{1}{\langle H_{ek}^2 \rangle} \int_{-\infty}^{\infty} u H_{ek}(\xi) w(\xi) d\xi
\]

where the denominator can be pre-calculated, and the integral is evaluated using numerical quadrature rules (referred to as DM method), written as,

\[
\int_{-\infty}^{\infty} u H_{ek}(\xi) w(\xi) d\xi = \sum_{i=1}^{N_q} q^i u(\xi^i) H_{ek}(\xi^i)
\]

where \( q^i \) is the corresponding weights, and \( N_q \) is the number of quadrature points needed for accurate integration. For multi-dimensional (D) Hermite polynomial of order \( P \), number of quadrature points \( N_q = (p + 1)^D \).

Assuming there exist three input uncertainties giving \( D = 3 \) which are associated with Gaussian PDF, and assuming a 5th order Hermite polynomial is required to replace a typical numerical model, equation (1) can be written as,

\[
Q(\lambda_1, \lambda_2, \lambda_3) = \sum_{k=0}^{5} C_k H_{ek}(\xi_1, \xi_2, \xi_3) = C_0 H_{e0}(\xi_1, \xi_2, \xi_3) + C_1 H_{e1}(\xi_1, \xi_2, \xi_3) + \ldots + C_5 H_{e5}(\xi_1, \xi_2, \xi_3)
\]

where the \( H_{ek}(\xi_1, \xi_2, \xi_3) \) is a multi-dimensional Hermite polynomial, the \( Q \) represents the interested quantity such as thermophysical and transport properties of the high hydrogen content syngas fuel at certain conditions. The \( \lambda_1, \lambda_2, \lambda_3 \) (refers to sampled or pre-processed data) represent the three uncertain model parameters which can also be expanded individually using PCE as,
\[ \lambda(\xi) = \sum_{k=0}^{P} C_k \mathcal{H}_k(\xi) \]  

(6)

Because it is assumed that the three input parameters have uncertainty that is normally distributed, first-order Hermite polynomial gives the following expression:

\[ \lambda(\xi) = C_0 \mathcal{H}_0(\xi) + C_1 \mathcal{H}_1(\xi) = C_0 \times 1 + C_1 \times \xi = \mu + \sigma \xi \]  

(7)

where \( C_0 = \mu \) (expectation) and \( C_1 = \sigma \) (standard deviation). The germ samples \( \xi \) are randomly sampled from \([-\infty, \infty]\) (refer to ‘qdpts.txt’ in the directory where post-processing is performed). Substitute equation (7) into (5), the quantity \( Q \) is decomposed to the form of PCE.

Using PCE based surrogate model provides ease of achieving sensitivity information (effect of input uncertainty on that of output) from decomposed terms of polynomial chaos expansion. For example, equation (5) can be further expanded as,

\[
Q(\mu_1 + \sigma_1 \xi_1, \mu_2 + \sigma_2 \xi_2, \mu_3 + \sigma_3 \xi_3) = C_0 + C_1 \mathcal{H}_1(\xi_1) + C_2 \mathcal{H}_2(\xi_2) + C_3 \mathcal{H}_1(\xi_3) + C_4 \mathcal{H}_2(\xi_1) + C_5 \mathcal{H}_1(\xi_1) \mathcal{H}_1(\xi_2) + C_6 \mathcal{H}_1(\xi_1) \mathcal{H}_1(\xi_3) + C_7 \mathcal{H}_2(\xi_2) + C_8 \mathcal{H}_1(\xi_2) \mathcal{H}_1(\xi_3) + C_9 \mathcal{H}_2(\xi_3) + \ldots
\]

(8)

where each term corresponds to the decomposed effect of a germ \( \xi \) on \( Q \). Moreover, global sensitivity information can be obtained because the total variance of a targeted quantity equals to the sum of the variance contribution from decomposed terms, i.e.,

\[ \text{Var}(Q) = \sum_{k>0} C_k^2 \langle \mathcal{H}_k^2 \rangle, \]

or,

\[
\text{Var}[Q(\mu_1 + \sigma_1 \xi_1, \mu_2 + \sigma_2 \xi_2, \mu_3 + \sigma_3 \xi_3)] = 0 + C_1^2 \langle \mathcal{H}_1^2(\xi_1) \rangle + C_2^2 \langle \mathcal{H}_1^2(\xi_2) \rangle + C_3^2 \langle \mathcal{H}_1^2(\xi_3) \rangle + C_4^2 \langle \mathcal{H}_2^2(\xi_1) \rangle + C_5^2 \langle \mathcal{H}_1^2(\xi_1) \mathcal{H}_1^2(\xi_2) \rangle + C_6^2 \langle \mathcal{H}_1^2(\xi_1) \mathcal{H}_1^2(\xi_3) \rangle + C_7^2 \langle \mathcal{H}_2^2(\xi_2) \rangle + C_8^2 \langle \mathcal{H}_1^2(\xi_2) \mathcal{H}_1^2(\xi_3) \rangle + \ldots
\]

(9)

The second and the fifth term of equation (9) represents the variance contribution of \( \xi_1 \) to the variance of the quantity \( Q \). In other words, the uncertainty effect of input on total variance of output can be written as,

\[
S_i = \frac{\sum_{k>0} C_i^2 \langle \varphi_k^2 \rangle}{\sum_{k>0} C_k^2 \langle \varphi_k^2 \rangle}
\]

(10)

where \( i \) refers to one of the three input parameters. In the given datasets, a total of \( N_q \) quadrature points are employed for PCE surrogate model construction, 1,000,000 points are pushed through the PCE model to ensure a high representation of PCE characters and to construct PDFs described before.

**Acknowledgments**

The authors would like to thank the UK Engineering and Physical Sciences Research Council (EPSRC) for supporting this research under project grant EP/K036750/2 “Clean Energy Utilisation from Biogas and Biomass Gasification”. We are also grateful to the colleagues in WP5 of the EU H2020 project “HPC4E (High Performance Computing for Energy)” for useful discussions.

**Conflict of Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
References

[1] Chemkin | Reaction Design. http://www.reactiondesign.com/products/chemkin/, 2019. (Accessed 27 November 2019).
[2] B.J. Debusschere, N.N. Habib, P.P. Philippe, M.K. Omar, G.G. Roger, P. Olivier, Numerical challenges in the use of polynomial chaos representations for stochastic processes, SIAM J. Sci. Comput. 26 (2004) 698–719, https://doi.org/10.1137/S1064827503427741.
[3] GRI Mech 3.0. http://combustion.berkeley.edu/gri-mech/version30/text30.html, 2019. (Accessed 27 November 2019).
[4] Chemical-Kinetic Mechanisms for Combustion Applications, San Diego Mechanism, Mechanical and Aerospace Engineering (Combustion Research), University of California at San Diego, 2019. http://combustion.ucsd.edu. (Accessed 27 November 2019).
[5] Y.J. Zhang, M. Olivier, L.P. Eric, B. Gilles, J.C. Henry, Assessing the predictions of a NOx kinetic mechanism on recent hydrogen and syngas experimental data, Combust. Flame 182 (2017) 122–141, https://doi.org/10.1016/j.combustflame.2017.03.019.
[6] R. Borghi, On the structure and morphology of turbulent premixed flames, in: Recent Advances in the Aerospace Sciences, 1985, pp. 117–138, https://doi.org/10.1007/978-1-4684-4298-4_7.
[7] W.L. Loh, On Latin hypercube sampling, Ann. Stat. 24 (1996) 2058–2080.
[8] I.M. Sobol, On sensitivity estimation for nonlinear mathematical models, Matematicheskoe Model. 2 (1990) 112–118.