Combustion and Flame 177 (2017) 89–97

Contents lists available at ScienceDirect

Combustion and Flame

journal homepage: www.elsevier.com/locate/combustflame

Outlier analysis for a silicon nanoparticle population balance model

Sebastian Mosbach[a], William J. Menz[a], Markus Kraft[a,b,*]

[a] Department of Chemical Engineering and Biotechnology, University of Cambridge, New Museums Site, Pembroke Street, Cambridge CB2 3RA, United Kingdom

[b] School of Chemical and Biomedical Engineering, Nanyang Technological University, 62 Nanyang Drive, Singapore 637459, Singapore

ARTICLE INFO

Article history:
Received 26 August 2016
Revised 19 September 2016
Accepted 2 December 2016
Available online 5 February 2017

Keywords:
Silicon
Nanoparticles
Population balance
Regression influence diagnostics

ABSTRACT

We assess the impact of individual experimental observations on a multivariate population balance model for the formation of silicon nanoparticles from the thermal decomposition of silane by means of basic regression influence diagnostics. The nanoparticle model is closely related to one which has been used to simulate soot formation in flames and includes morphological and compositional details which allow representation of primary particles within aggregates, and of coagulation, surface growth, and sintering processes. Predicted particle size distributions are optimised against 19 experiments across ranges of initial temperature, pressure, residence time, and initial silane mass fraction. The influence of each experimental observation on the model parameter estimates is then quantified using the Cook distance and DFBETA measures. Seven model parameters are included in the analysis, with five Arrhenius pre-exponential factors in the gas-phase kinetic rate expressions, and two kinetic rate constants in the population balance model. The analysis highlights certain experimental conditions and kinetic parameters which warrant closer inspection due to large influence, thus providing clues as to which aspects of the model require improvement. We find the insights provided can be useful for future model development and planning of experiments.

1. Introduction

Gas-phase synthesis in hot-wall reactors is a common way in which silicon nanoparticles are manufactured. Shock-tubes are another set-up in which especially the early phase of formation of these particles can be studied. Typically, these synthesis processes begin with silane (SiH_4) as a precursor, which is transformed into the eventual nanoparticle product at high temperatures. A variety of models have been proposed to describe this transformation [1]. These models usually contain unknown or low-confidence (kinetic) parameters with large uncertainties associated to them. Systematic parameter estimation techniques can then be employed to arrive at better values for these quantities, based on available experimental data. One of the most elementary parameter estimation methods is least-squares optimisation, *i.e.* minimising the distance between experimental observations and model prediction as measured by a sum-of-squares objective function. The result of such an optimisation is a set of values, called ('best') estimates, for the selected model parameters. Not all experimental data points may equally inform the optimal value of the parameters, though – different parameters may be determined to a varying extent by different observations. In order to assess which experiments are the most relevant in the optimisation, one can conduct what may be called an omission-based regression influence analysis [2]. Firstly, optimise the model against the full data set, and then repeat the optimisation with one of the data points removed, for each of the data points. Based on the difference between the parameter estimates of the full optimisation and the optimisations with an omitted data point, it is then possible to quantify the influence of individual observations on the model overall or on individual parameters. Several such measures have been proposed [3,4], the most widely-used one being Cook's distance [5], and applied to detect influential data points, high-leverage points, and statistical outliers [6,7].

An alternative approach to quantifying influence of experimental observations is uncertainty propagation [8], part of which is concerned with how experimental measurement errors propagate into model parameters and responses. Some of these methods allow calculating the relative contribution of each data point (and its error bar) to the uncertainty in each of the parameters. In particular, the Data Collaboration framework [9] exploits the pairwise consistency of data set units to identify outliers.

Yet another approach, called perturbation of the optimum, has been developed for constrained optimisation [10, p. 34] and

* Corresponding author at: Department of Chemical Engineering and Biotechnology, University of Cambridge, New Museums Site, Pembroke Street, Cambridge CB2 3RA, United Kingdom.
E-mail address: mk306@cam.ac.uk (M. Kraft).

http://dx.doi.org/10.1016/j.combustflame.2016.12.006
0010-2180/

Table 1
The gas-phase kinetic mechanism. Values in bold correspond to parameters chosen for the influence analysis. Units for the Arrhenius pre-exponential factors are cm, mol, and s.

Idx.	Reaction	A	β [–]	E [kcal/mol]	Ref.
1	SiH_4 (+M)⇌SiH_2 + H_2 (+M)	3.12×10^9	1.7	54.71	[14]
	Low pressure limit:	**3.96×10^{12}**	0	45.10	[1,15][a]
2	Si_2H_6 (+M)⇌SiH_4 + SiH_2 (+M)	1.81×10^{10}	1.7	50.20	[14]
	Low pressure limit:	**5.09×10^{53}**	−10.37	56.03	[14]
3	Si_2H_6 (+M)⇌Si_2H_4B + H_2 (+M)	9.09×10^9	1.8	54.20	[14]
	Low pressure limit:	**7.79×10^{40}**	−7.77	59.02	[1,14][b]
4	Si_3H_8 (+M)⇌SiH_2 + Si_2H_6 (+M)	6.97×10^{12}	1.0	52.68	[14]
	Low pressure limit:	1.73×10^{69}	−15.07	60.49	[14]
5	Si_3H_8 (+M)⇌Si_2H_4B + SiH_4 (+M)	3.73×10^{12}	1.0	50.85	[14]
	Low pressure limit:	**4.36×10^{76}**	−17.26	59.30	[14]
6	Si_2H_4B (+M)⇌Si_2H_4A (+M)	2.54×10^{13}	−0.2	5.38	[14]
	Low pressure limit:	1.10×10^{33}	−5.76	9.15	[14]
7	Si_2H_4B + H_2⇌SiH_4 + SiH_2	9.41×10^{13}	0	4.09	[14]
	Reverse coefficients:	9.43×10^{10}	1.1	5.79	[14]
8	Si_2H_4B + SiH_4⇌Si_2H_6 + SiH_2	1.73×10^{14}	0.4	8.90	[14]
	Reverse coefficients:	**2.65×10^{15}**	0.1	8.47	[14]

[a] A is from [1], β and E are from [15].
[b] A is from [1], β and E are from [14].

unconstrained least-squares optimisation [11], which has found application in chemical kinetics [2,12,13]. These methods allow calculating sensitivities of parameter estimates with respect to any other quantity in the objective function (or constraints), including in particular experimental data.

The purpose of this paper is to conduct an omission-based outlier analysis of a selection of experimental data for silicon nanoparticles produced from a silane precursor in hot-wall flow reactors and shock tubes which are modelled using a detailed population balance model. A main aim is to identify those experimental conditions which are the most challenging for the model. We apply a technique established in the field of regression influence diagnostics to quantify the influence of individual experimental observations on kinetic parameter estimates for this purpose. We determine the influence of the measurements on estimates of some Arrhenius pre-exponential factors in the gas-phase kinetic mechanism as well as the population balance model for the particle phase. Using a threshold for the influence values, specific measurements are then highlighted for further analysis, providing further insight into the model and potential improvements, as well as suggestions for future experiments.

2. Background

We firstly describe the model, provide some background on omission-based regression influence diagnostics, and how it can be used to identify outliers.

2.1. Population balance model for silicon nanoparticle formation

We briefly summarise the main features of the model here. Full details can be found in [1], and further in [16–20], noting that a closely related model has been applied to soot formation in flames (see for example [21] and references therein). It consists of two main parts, a gas-phase model, and a particulate phase model.

2.1.1. Gas phase

The gas-phase chemical kinetic reaction mechanism used is a modified version of the one proposed by [14], and is summarised in Table 1. Two isomers of Si_2H_4 are included: silene, *i.e.* H_2SiSiH_2, denoted by the suffix “A”, and silylene, *i.e.* $HSiSiH_3$, denoted by the suffix “B”. The first six reactions are third-body reactions whose pressure-dependence is given in Lindemann fall-off form. More details can be found in [1].

2.1.2. Particulate phase

The particle phase is described by a detailed, high-dimensional population balance model [1] covering aggregate morphology and chemical composition. In this model, each nanoparticle is represented as a list of primary particles, together with a (triangular) matrix, called connectivity matrix, each entry of which represents the common surface area for the corresponding pair of primary particles. For each primary particle, the number of silicon and the number of hydrogen atoms are stored. From this particle representation, beyond elementary properties like mass and chemical composition, several quantities of interest can be derived. These include for example, with some additional assumptions, collision and mobility diameter of aggregates, surface area, and sintering level.

The following processes which create or transform particles, or account for interaction of the particles with the gas phase, are represented in the model:

Inception: Any two molecules of any of the three species SiH_2, Si_2H_4A, and Si_2H_4B can collide to (irreversibly) form a new particle, which is assumed to consist of a single, spherical primary whose diameter follows directly from its mass, *i.e.* numbers of atoms. The rate at which this happens is assumed to be non-zero only if the diameter of the resulting particle exceeds a temperature- and pressure-dependent critical nucleus diameter. If the latter is the case, the inception rate is proportional to the product of the concentrations of the collision partners and the transition regime coagulation kernel. More details can be found in [1] and [16].

Condensation: An existing particle can grow through (barrier-free) deposition of SiH_2, Si_2H_4A, or Si_2H_4B molecules from the gas phase onto its surface. It is assumed that the collision efficiency, *i.e.* the probability of sticking, is unity. The rate is given by a free-molecular collision kernel.

Surface reaction: Apart from simply condensing, gas-phase species can also react heterogeneously on the particle surface. Specifically, silanes (SiH_4, Si_2H_6, and Si_3H_8) can be integrated into the particle, with each step releasing one, two, and three molecules of hydrogen, respectively. The rate is proportional to the particle surface area and an Arrhenius expression with non-zero activation energy. Rounding of adjacent primary particles caused by this process is also taken into account.

Hydrogen release: In order to attain a stable crystal structure, particles need to release some of the hydrogen acquired through each of the above processes. The rate of desorption is proportional to an Arrhenius expression and the coverage of hydrogen on the