Minimal Spline Fit: a model-free method for determining statistical noise of experimental data series

T. Nagy*[,1], T. Turányi[2]

[1]Institute of Materials and Environmental Chemistry, Research Centre for Natural Sciences, ELKH, Budapest, Hungary
[2]Institute of Chemistry, ELTE Eötvös Loránd University, Budapest, Hungary

Abstract
Robust and accurate chemical kinetics models of low uncertainty are required to aid the development of novel combustion devices using simulations. Parameter optimization against experimental data is a possible way to develop such models. A proper objective function that can handle reference data of different types and magnitudes is obtained by normalizing the deviations by the corresponding experimental error. We propose a novel model-free method and a corresponding code, called Minimal Spline Fit, to estimate the statistical noise of experimental data series and to predict its noise-free profile.

Introduction

Over the past decades, the efforts of the gas kinetics community, helped by the advancement of experimental and theoretical methods and computers, led to an immense increase in mechanistic and kinetic knowledge, which allowed the assembly of detailed kinetic reaction mechanisms in combustion chemistry [1]. Kinetic reaction mechanisms are models of the reacting system that contain not only the reactions of species but also their corresponding kinetic, thermodynamic and transport parameters. A possible error of the mechanisms is that certain mechanistic details are missing. Also, the parameters are usually uncertain, which leads to uncertainties in model predictions. The reliability of models can be increased by reducing their uncertainty via tuning and constraining their parameters using not yet considered or newly published experimental data and theoretical results. In direct experiments, the measured properties are exclusively determined by a single reaction step, thus they provide direct information on its rate parameters. Indirect measurements focus on global properties, like ignition delay time, laminar burning velocity and concentration data etc., which can be reproduced by simulations using complex reaction mechanisms only.

The first kinetic parameter optimization studies on combustion kinetic mechanisms were carried out by Frenklach et al. [2,3] and Sheen and Wang [4,5]. A similar method was recently applied by Pitsch and his coworkers [6]. The most comprehensive determination and uncertainty minimization of kinetic parameters in a detailed reaction mechanism can be achieved by fitting the model results to all available direct and indirect experimental data and theoretical results on the system [7]. This approach allowed the efficient global optimization of all Arrhenius parameters in their joint uncertainty domain of all important reactions [8–10]. This methodology has been used for the development of several detailed reaction mechanisms (see e.g. refs. [11,12]).

Mechanism optimization is always based on the minimization of an objective function that measures the deviation of model results from the experimental data. A proper objective function that can handle reference data of different types, dimensions and magnitudes is obtained by normalizing the deviations by the corresponding experimental error [7]. The following objective function, has been applied in several studies:

$$E(\mathbf{p}) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{\left(Y_{ij}^{\mathrm{sim}}(\mathbf{p}) - Y_{ij}^{\mathrm{exp}}\right)^2}{\sigma_i^2} \qquad (1)$$

Here, N is the number of data series, N_i is the number of data within the series, Y_{ij}^{sim} and Y_{ij}^{exp} are the properly transformed simulated and experimental data, respectively. The uncertainty of the experimental dataset (Y_{ij}^{exp}, $j = 1, \ldots, N_i$) is characterized by σ_i standard deviation. A safe upper estimate can be given to experimental errors based on systematic variations (σ_{sys}) and statistical errors (σ_{stat}) using the following formula [13], which assumes that they are uncorrelated:

$$\sigma = \sqrt{\sigma_{\mathrm{sys}}^2 + \sigma_{\mathrm{stat}}^2} \qquad (2)$$

Systematic errors are usually reported by the experimentalist or at least approximately known for most standard experimental devices and methods. The statistical noise can be assessed from measurements repeated at identical conditions. However, the measurements are typically not repeated at identical conditions and the data are usually measured as part of a series, where one of the experimental conditions (e.g. temperature (T), equivalence ratio (φ), time (t)) is systematically changed or changing. In the following, with the term data series, we mean a data set with such a dependence on a provided independent variable.

Points of such data series contain statistical noise, which could be extracted if we knew their noise-free dependence on the varied condition variable. Here, we propose a novel method and a code to find such function

* Corresponding author: nagy.tibor@ttk.hu

based exclusively on the data. Using this function, we can estimate the statistical noise of the data series. Note, that this function will not only capture systematic changes, but systematic errors, too, which could be determined only by using more accurate measurement techniques, thus cannot be decomposed by data processing.

Experimental data and its transformation

Let us assume, that we have a noisy data series, composed of N data pairs:

$$\left(x_j, y_j\right) \qquad j = 1, \dots, N. \tag{3}$$

Here, x_j denotes the j^{th} value of the independent variable, which is changing or is systematically changed in the experiment (e.g. time, temperature, equivalence ratio, etc.), and y_j denotes the corresponding measured data. We assume that the x_j values are known without error, whereas values y_j have random noise of similar origin.

Often there is an underlying knowledge on the expected form of the true $y(x)$ function. For example, in the case of the rate coefficient of an elementary reaction, the extended Arrhenius expression ($k(T) = AT^n \exp(-E/RT)$) is often used as a good approximation for its temperature dependence. In this case either a variable or an equivalent axis transformation (e.g. $Y_j = f_y\left(y_j\right)$ and $X_j = f_x(x_j)$) is applied to resolve data variations of different orders of magnitude (e.g. $Y_j = \ln y_j$, where $y_j = k\left(T_j\right)$) or uneven distribution (e.g. $X_j = 1/x_j$, where $x_j = T_j$), but most importantly to obtain the same noise distributions implying equal standard deviations for each $f_y(y_j)$ transformed data point:

$$Y_j = \begin{cases} y_j & \text{if } \sigma\left(y_j\right)\text{- s} \quad \text{are the same} \\ \ln y_j & \text{if } \sigma\left(\ln y_j\right)\text{- s} \quad \text{are the same} \\ f_y(y_j) & \text{if } \sigma\left(f_y(y_j)\right)\text{- s are the same.} \end{cases} \tag{4}$$

In the following, the common standard deviation of the transformed dataset, $\{Y_j, j = 1, \dots, N\}$ will be denoted by σ. The transformed experimental values with their standard deviations and corresponding simulated values are used in the objective function (see Eq. (1)).

Regarding the indirect combustion measurements: ignition delays are controlled by the rate of one or few elementary reactions in not too wide temperature intervals, thus an Arrhenius-type transformation can also be applied to them. In the case of concentrations measured in shock tubes, flow reactors, perfectly stirred reactors, depending on the measurement technique, logarithmic or no transformation should be applied. If the signal is proportional to the concentration then errors are constant and no transformation is needed. If the signal is linear function of the logarithm of the concentration (e.g. in absorbance measurements), then logarithmic transformation of the concentrations is needed to have same error characteristics.

An often used error type is the relative error, which is defined mainly for positive quantities, thus its scale is bounded from below by –1 (i.e. –100%), and apart from small changes it behaves asymmetrically. Instead, we recommend the use of error expressed on natural logarithmic scale, as it mimics relative error at small values (e.g. 0.2) due to the following property:

$$\Delta \ln y_j \approx \frac{\Delta y_j}{y_j} \qquad \text{if } \left|\Delta \ln y_j\right| \ll 1, \tag{5}$$

Furthermore, it can describe even large changes symmetrically and unlimitedly in both positive and negative directions.

Determination of the random noise of a dataset

If the noise-free $Y(X)$ dependence between variables X and Y were known (i.e. the expected value of Y_j: $\mu_j = Y(X_j)$), then the population variance of the statistical noise (i.e. σ^2) superposed on the data could be estimated by the mean squared deviations between data Y_j and its expected value at the corresponding X_j:

$$\sigma^2 \approx \frac{1}{N}\sum_{j=1}^{N}\left(Y_j - Y\left(X_j\right)\right)^2. \tag{6}$$

However, the exact $Y(X)$ function is not known, thus usually an appropriate flexible model function $Y_{\text{fit}}(X; \mathbf{P})$, which contains p number of unknown parameters ($\mathbf{P} = (P_1, \dots, P_p)$), is fitted to the data by minimizing the sum of squared deviations. Let $\mathbf{P}_{\text{opt}}$ denote the vector of parameter values that minimize the deviation. If this minimum value is divided by the number of degrees of freedom, $\nu = N - p$, it serves as an unbiased estimator of the variance of the noise population, assuming that $Y\left(X; \mathbf{P}_{\text{opt}}\right)$ is a good approximation to $Y(X)$:

$$\text{var}_{\text{fit}} = \sigma_{\text{fit}}^2 \approx \frac{1}{\nu}\sum_{i=1}^{N}\left(Y_j - Y\left(X_j; \mathbf{P}_{\text{opt}}\right)\right)^2. \tag{7}$$

The obtained variance can be used in the error function to normalize the squared deviation of the simulated and experimentally determined Y_j values. The square root of the variance estimator is only a biased estimator for the standard deviation (σ) of the noise population, as no generally applicable unbiased estimator exists for σ. Nevertheless, this σ value can be used to estimate the statistical error σ_{stat} of the data points. The estimate of the variance depends on the goodness of the fitting function. If it doesn't capture the systematic evolution of the data, then it will overestimate the noise variance. If it captures its systematic evolution and also some of the noise, then it will underestimate the noise variance. To minimize such errors, a class of functions of increasing flexibility should be used for fitting and the fitting should be stopped where all systematic behaviour is described, but no noise is picked up.

Fitting functions - Polynomials

Polynomials are standard functions used for approximating non-periodic analytic functions:

$$Y_{\text{poly}-n}(X; \mathbf{P}) = a_0 + a_1 X + \dots + a_n X^n \tag{8}$$

Here, n is the order of the polynomial, which is the power of X in the highest-order term with non-zero coefficient $(a_n \neq 0)$. The parameters of the polynomial are the coefficients of the monomial terms, that is $\mathbf{P} = (a_0, \dots, a_n)$, thus $p = n + 1$. All analytic functions can be approximated with its Taylor series, thus with increasing the order, a more and more flexible functional form is obtained, and in theory most data evolutions can be captured with a polynomial of sufficient high order. A n-th order polynomial can go through any $n + 1$ data points with different X_j values, thus can be used for interpolation, too. Furthermore, coefficients of the best fitting polynomial of a given order can be calculated by solving a system of linear equations, and thus can be directly calculated for any order. However, for very high orders, such as $n = 10$, the coefficients and the evaluated values can have high numerical errors even at double precision accuracy, which limits their applicability. Furthermore, polynomial coefficients and thus polynomial values depend on all data points considered during fitting, which can result in unexpected oscillations of high-order polynomials between data points of similarly magnitude if at other values steep changes are captured. This is called the Runge's phenomenon in polynomial interpolation [14].

Splines with optimized control points

We propose a novel type of fitting function based on an unusual application of splines. Splines are piecewise polynomials, which are used for interpolation and smoothing of data points. Interpolating splines go through all of their defining points, called knots or control points. The spline curves between adjacent knots are determined by at most only a few neighbouring data points, thus are insensitive to steep changes more than a few points away. Consequently, they are much more flexible functions than polynomials and can adapt to large variations in steepness.

However, splines cannot be used directly for fitting as noisy data points cannot serve as control points. Instead, independently from the data points, let us assume a spline function defined by n control points, which lie in the same space as the investigated series of data pairs:

$$(X_{cj}, Y_{cj}) \qquad j = 1, \dots, n. \tag{9}$$

We assume that the X_{cj} values are different and are in ascending order, and the first and the last of them are fixed to the minimum and the maximum values of the data series, respectively.

$$\min_{j=1,\dots,N} X_j =: X_{c1} < \dots < X_{cn} := \max_{j=1,\dots,N} X_j \tag{10}$$

Thus, the remaining free, $p = 2n - 2$ coordinates of the control points serve as the parameters of the $Y(X;\mathbf{P})$ spline function:

$$\mathbf{P} = (X_{c2}, \dots, X_{c,n-1}, Y_{c1}, \dots, Y_{cn}). \tag{11}$$

Depending on the order of polynomials between the adjacent knots and their connection properties at the control points (e.g. continuous differentiability), one can define various splines. Linear splines connect control points with straight lines. Fitting a continuous, piecewise linear function (i.e. linear splines) to dense noise data is useful if a smoothed data set is needed as an input in another program, which reconstructs the original profile from it by linear interpolation. A natural cubic spline employs cubic polynomials between control points and is twice continuously differentiable at the knots. A special type of cubic spline is the Akima spline [15], which is especially cheap to evaluate and do not lead to unphysical wiggles in the resulting curve opposed to natural splines and higher-order polynomials. Akima splines, however, only once continuously differentiable at the knots, which may limit their use for certain purposes.

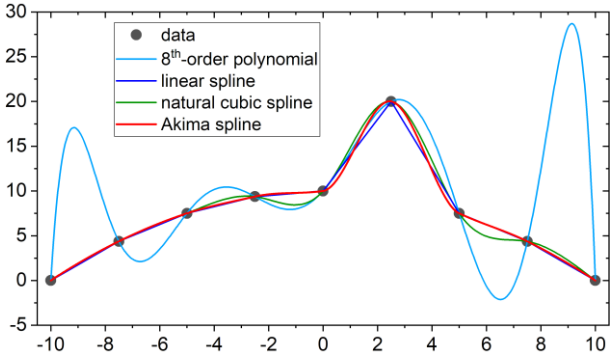

Figure 1 Comparison of interpolating functions.

Figure 1 demonstrates the advantageous properties of Akima splines on an interpolation example between 9 points that follow each other with largely varying steepness. The 8th-order polynomial and the natural cubic spline show unexpected, large-amplitude oscillations between points (i.e. Runge phenomenon), whereas Akima splines follow the evolutions of control points naturally, without wiggles. In the next chapter, we discuss measures for the prediction error of the fit that penalize both under- and overfitting and thereby help us to find polynomials of optimal order, and splines with optimal number of control points.

Measures for the prediction error of the fit

A standard measure for quantifying the scatter of data points around a fitting curve is the root-mean-square deviation ($\mathrm{RMSD_{fit}}$), which is proportional to the square of the variance of the fit:

$$\mathrm{RMSD_{fit}} \approx \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - Y(X_i;\mathbf{P}_{\mathrm{opt}})\right)^2} \tag{12}$$

This quantity takes into account underfitting only and decreases monotonically as the flexibility of the fitting curve is systematically increased.

A better measure, which is also used to estimate the standard deviation of the noise sample in the data series, is the square root of the variance, as it takes into account the decrease of degrees of freedom with increasing parameter number:

$$\sigma_{\mathrm{data}} \approx \sigma_{\mathrm{fit}} \approx \sqrt{\mathrm{var}_{\mathrm{fit}}} = \sqrt{N/\nu}\,\mathrm{RMSD_{fit}}. \tag{13}$$

It becomes non-applicable when the degrees of freedom ν becomes zero, which happens when the number of parameters reaches the number of data points. This corresponds to a maximum polynomial order of $n = N - 1$, and maximum number of spline control points of $n = [N/2] + 1$, where square bracket denotes integer part.

The Akaike information criterion (AIC) estimates the out-of-sample prediction error and it assesses the relative quality of statistical models for a given data set [16,17]. It is calculated as:

$$\mathrm{AIC} = 2p - 2\ln \mathcal{L}_{\max}, \tag{14}$$

where $\mathcal{L}_{\max}$ denotes the maximum value of the likelihood function. For N normally distributed residuals, it simplifies to:

$$\mathrm{AIC} = 2p + N + N\ln\frac{2\pi \cdot \mathrm{RMSD}_{\mathrm{fit}}^2}{N}. \tag{15}$$

It estimates the relative amount of information lost by a model, which is used to represent the process that generated the noisy data. Better models lose less information, thus they have lower AIC value. When we have relatively low number of data points, comparable to the square of the number of parameters, AIC will overfit, which can be avoided by using a corrected AIC quantity (AICc) for model comparison [18–20]:

$$\mathrm{AICc} = \mathrm{AIC} + 2\frac{p^2 + p}{N - p - 1} \tag{16}$$

For large number of data, where $N \gg 2p^2$, it converges to AIC, thus it is safe to use AICc in all cases. AIC and AICc penalizes both underfitting and overfitting, and can be used to estimate the relative likelihood of two models (e.g. for model A and model B):

$$\frac{P_{\mathrm{A}}}{P_{\mathrm{B}}} = \exp\left(-\frac{\mathrm{AIC(c)_A} - \mathrm{AIC(c)_B}}{2}\right). \tag{17}$$

For example, if $\mathrm{AIC(c)_A} = \mathrm{AIC(c)_B} - 6$, then model A is roughly 20 times more probably than model B, or with other words: we can state with roughly 95% statistical confidence that model A is better than model B.

Optimal fitting functions and the noise of the data set

Optimal polynomials and Akima splines are determined here in a combined function selection and parameter optimization procedure, in which functions of increasing flexibility are fitted to the data series. In the case of polynomials, this is achieved by increasing the polynomial order gradually (i.e. n=0,1,2,…), whereas for splines it can be done by taking more and more control points and optimizing their positions together. In the case of Akima spline at least 3 control points are needed (i.e. n=3,4, ...). If the global minimum is found at each investigated number of control points then the $\mathrm{RMSD}_{\mathrm{fit}}$ decreases monotonically with the number of control points.

Increasing the number of degrees of freedom of the fitting curve, the function captures more accurately the noise-free, systematic evolution of the data series and the variance of the fit decreases steeply initially. At some complexity of the fitting function, the fit variance (see Eq. (6)) starts to stagnate with the number of parameters. The corresponding fitted curve is the best and simplest noise-free empirical model ("minimal spline" or minimal-order polynomial) and the corresponding root-mean-square error estimates the statistical noise of the data series. Increasing the complexity of the fitting function (i.e. the number of parameters) further, overfitting occurs as the noise profile of the data series are being captured gradually. However, noise cannot be fitted efficiently due to its random nature, thus the fit variance reduces much less steeply and this is the reason for the stagnation.

Akaike information criterion (AIC or AICc) can also be used to identify the optimal model: the first local minimum in the value of AICc as a function of the number of parameter identifies the optimal model. However, all models which have an AICc value less than the corresponding value of the optimal model plus 6, are not significantly (with less than approx. 95% statistical confidence) worse than the optimal model.

The optimal $Y(X; \boldsymbol{P}_{\mathrm{opt}})$ model can provide good estimates for the expected values of Y when evaluated at X_j values:

$$\mu_j = Y(X_j) \approx Y(X_j; \boldsymbol{P}_{\mathrm{opt}}), \tag{18}$$

Thus the variance of the noise on the data set can be estimated using Eq. (7). In summary, in the case of a noisy data series, the optimal polynomial or spline curve provides an approximation to the noise-free model and the empirical variance provides a reliable estimate of the variance of the noise on the Y_j points of the dataset.

Computer code Minimal Spline Fit

The discussed method has been implemented in a Fortran code called Minimal Spline Fit, which can be downloaded from the respecth.hu webpage [21]. Fitting Akima splines to data via tuning the position of their control points is a challenging global optimization problem, thus the particle swarm algorithm (zxmwd.f) is used. However, this cannot always find the global optimum in reasonable time, therefore additional heuristics is employed in the search. Optimized positions of control points for the best spline at given number of knots are stored. If improvement in the fit is observed after adding a new control point then the code always takes back a single control point (tries each except for the first and last one) and optimizes the positions of the rest. If a smaller fitting error is achieved for this simpler spline than the one which has been previously stored, then further control points are taken back until no improvement is observed at a given number of knots. If improvement was observed for a given reduced number of control points, but not for fewer, then adding of control points is restarted from this size. This iterative adding and removal of control points can achieve strictly decreasing and small $\mathrm{RMSD}_{\mathrm{fit}}$ values much more efficiently than direct global optimization. The increase in flexibility of

polynomials and splines is stopped when AICc increases in two subsequent steps.

In spite of the extensive search for the optimal fitting function, sometimes visually incorrect fits are generated, especially when the X_j data values of the series are rather unevenly distributed. In such cases, the data set should be split and variances should be determined separately for its parts, which should be averaged with proper weighting to estimate the variance of the whole data set. Due to such and similar issues, it is strongly recommended to check the fits visually by plotting them. For this purpose, programs that allow direct plotting of data files from command line are recommended (e.g. Gnuplot).

Examples

A 6th-order polynomial was used to generate 101 points equidistant in X and a stochastic noise following Gaussian distribution (μ=0, σ=3) was added to the points.

$$Y = 5X(X+1)^2(X-1)(X^2-4) \qquad (19)$$

Figure 2 shows the data, the noise-free polynomial and fit by 6th-order polynomial and the fits by Akima splines with 7 and 8 control points. The splines simulate closely the original noise-free polynomial similarly to the fitted 6th-order polynomial. Figure 3 shows the evolution of the σ_{fit}, AICc values and relative model probabilities with the number of parameters for both the polynomials and the Akima splines. The standard deviation of the Gaussian distribution used for sampling the noise (3.00), and the standard deviation of the noise sample on the data (2.84) are plotted with green solid and dashed lines, respectively. Estimated standard deviations for both series of fitting functions start to stagnate close to the latter value, when their AICc values reach a minimum: for polynomials from 6th order, for splines between 7-9 control points. The example demonstrates that Akima splines can mimic efficiently high-order polynomials. For several data series, especially for those with low number of points, polynomials often fail to model the trend of the noisy data smoothly.

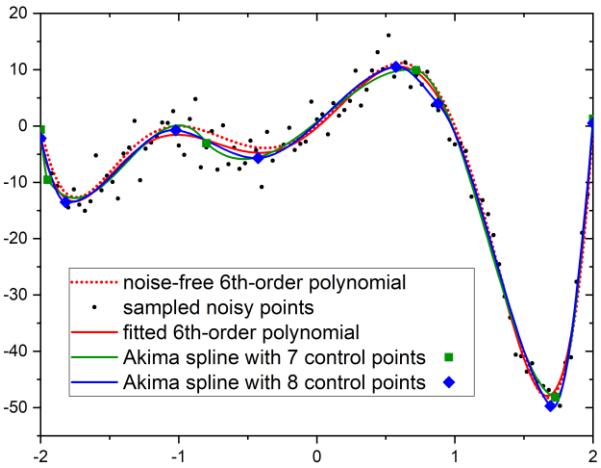

Figure 2. Fitting of a 6th-order polynomial and Akima-splines with 7 and 8 control points to data ($N = 101$) obtained from a noisy 6th-order polynomial.

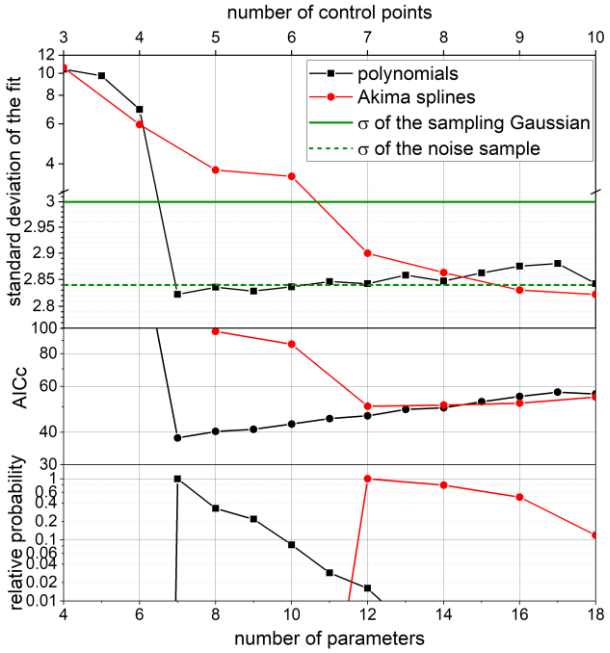

Figure 3. Evolution of σ_{fit}, AICc and relative probability of the fitted models functions.

Fig. 4 demonstrates that Akima splines are much more flexible than polynomials as they can fit a 5000-point noisy non-reactive pressure trace and store it in a noise-free form using merely 15 control points.

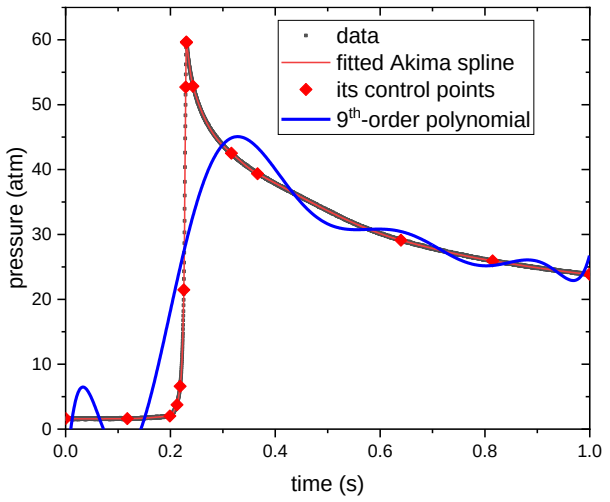

Figure 4. Generating a noise-free profile for non-reactive pressure traces in RCM experiments.

Conclusions

A method based on polynomials and wiggles-free Akima splines and the corrected Akaike information criterion was developed for the efficient estimation of the statistical noise of data series obtained in combustion or other experiments. A corresponding code was also developed and made available online [22]. The method has been used several times to assess statistical noise of data series in all published mechanism comparison and optimization studies by Turányi and coworkers (see e.g. refs. [7,11–13,23]). The ReSpecTh database [21] stores large amount of combustion experimental data collected from the literature in RKD format, the latest 2.3 version

of it allows storing experimental error information obtained a posteriori using code Minimal Spline Fit. The RKD files can be directly read by code Optima++ [24] to set up simulations for unbiased mechanism comparison and optimization studies which rely on the error function in Eq. (1), thus incorporate noise information on the experimental data. In this work, the method is validated on artificially generated noisy data sets with known noise profiles. A possible further use of the code is to fit or smooth and store complicated data series like non-reactive pressure traces in rapid compression machine measurements using Akima spline.

Acknowledgements

This work was supported by the Hungarian National Research, Development and Innovation Office [NKFIH grants K132109 and FK134332]. The authors are grateful to Dr. Carsten Olm for helpful discussions.

References

[1] H.J. Curran, Developing detailed chemical kinetic mechanisms for fuel combustion, Proc. Combust. Inst. 37 (2019) 57–81.

[2] D. Miller, M. Frenklach, Sensitivity analysis and parameter estimation in dynamic modeling of chemical kinetics, Int. J. Chem. Kinet. 15 (1983) 677–696.

[3] M. Frenklach, Transforming data into knowledge—Process Informatics for combustion chemistry, Proc. Combust. Inst. 31 (2007) 125–140.

[4] H. Wang, D.A. Sheen, Combustion kinetic model uncertainty quantification, propagation and minimization, Prog. Energy Combust. Sci. 47 (2015) 1–31.

[5] D.A. Sheen, H. Wang, The method of uncertainty quantification and minimization using polynomial chaos expansions, Combust. Flame. 158 (2011) 2358–2374.

[6] L. Cai, H. Pitsch, Mechanism optimization based on reaction rate rules, Combust. Flame. 161 (2014) 405–415.

[7] T. Turányi, T. Nagy, I.G.I. \relax G. Zsély, M. Cserháti, T. Varga, B.T.T. Szabó, I. Sedyó, P.T.T. Kiss, A. Zempléni, H.J.H.J. Curran, Determination of rate parameters based on both direct and indirect measurements, Int. J. Chem. Kinet. 44 (2012) 284–302.

[8] T. Nagy, T. Turányi, Uncertainty of Arrhenius parameters, Int. J. Chem. Kinet. 43 (2011) 359–378.

[9] T. Nagy, T. Turányi, Determination of the uncertainty domain of the Arrhenius parameters needed for the investigation of combustion kinetic models, Reliab. Eng. Syst. Saf. 107 (2012) 29–34.

[10] T. Nagy, É. Valkó, I. Sedyó, I.G. Zsély, M.J. Pilling, T. Turányi, Uncertainty of the rate parameters of several important elementary reactions of the H2 and syngas combustion systems, Combust. Flame. 162 (2015) 2059–2076.

[11] T. Varga, T. Nagy, C. Olm, I. \relax G.I.G. Zsély, R. Pálvölgyi, É. Valkó, G. Vincze, M. Cserháti, H.J.H.J. Curran, T. Turányi, Valkó, G. Vincze, M. Cserháti, H.J.H.J. Curran, T. Turányi, É. Valkó, G. Vincze, M. Cserháti, H.J.H.J. Curran, T. Turányi, Optimization of a hydrogen combustion mechanism using both direct and indirect measurements, Proc. Combust. Inst. 35 (2015) 589–596.

[12] T. Varga, C. Olm, T. Nagy, I.G. Zsély, É. Valkó, R. Pálvölgyi, H.J. Curran, T. Turányi, I.G. Zsély, É. Valkó, R. Pálvölgyi, H.J. Curran, T. Turányi, Development of a Joint Hydrogen and Syngas Combustion Mechanism Based on an Optimization Approach, 2016.

[13] C. Olm, T. Varga, É. Valkó, H.J. Curran, T. Turányi, Uncertainty quantification of a newly optimized methanol and formaldehyde combustion mechanism, 186 (2017) 45–64.

[14] C. Runge, Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten, Zeitschrift Für Math. Und Phys. 46 (1901) 224–243.

[15] H. Akima, A New Method of Interpolation and Smooth Curve Fitting Based on Local Procedures, J. ACM. 17 (1970) 589–602.

[16] H. Akaike, Information theory and the maximum likelihood principle, in: 2nd Int. Symp. Inf. Theory, 1973.

[17] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Contr. 19 (1974) 716–723.

[18] C.M. Hurvich, C.L. Tsai, Regression and time series model selection in small samples, Biometrika. (1989).

[19] E.J. Bedrick, C.-L. Tsai, Model Selection for Multivariate Regression in Small Samples, Biometrics. (1994).

[20] J.E. Cavanaugh, Unifying the derivations for the Akaike and corrected Akaike information criteria, Stat. Probab. Lett. (1997).

[21] ELTE Chemical Kinetics Laboratory and MTA-ELTE Complex Chemical Systems Research Group, ELTE Institute of Chemistry, ReSpecTh webpage, (n.d.).

[22] T. Nagy, Minimal Spline Fit Introducing Root-Mean-Square Fitting of Data Series with Akima Splines, (n.d.).

[23] C. Olm, I.G. Zsély, R. Pálvölgyi, T. Varga, T. Nagy, H.J. Curran, T. Turányi, Comparison of the performance of several recent hydrogen combustion mechanisms, Combust. Flame. 161 (2014).

[24] T. Varga, M. Papp, Á. Busai, I.G. Zsély, T. Nagy, T. Turányi, Optima++ package v2.0.0: A general C++ framework for performing combustion simulations and mechanism optimization, (2020).