

A General Model-Free Algorithm for Identifying Inconsistent Data Points in Indirect Combustion Datasets

Éva Valkó^{1,2}, Tibor Nagy³, András Gy. Szanthoffer^{2,4}, Máté Papp², Tamás Turányi^{2,*}

* lead presenter: tamas.turanyi@ttk.elte.hu

1 Institute of Mathematics, ELTE Eötvös Loránd University, Budapest, Hungary

2 Institute of Chemistry, ELTE Eötvös Loránd University, Budapest, Hungary

3 Institute of Materials and Environmental Chemistry,

HUN-REN Research Centre for Natural Sciences, Budapest, Hungary

4 ELTE Hevesy György PhD School of Chemistry, Budapest, Hungary

The design of high-efficiency and low-emission combustion systems relies heavily on kinetic reaction mechanisms. The development and validation of combustion kinetic models must be based on large amounts of reliable experimental data. For this reason, identifying inconsistent (i.e., probably wrong) data points in a data collection is a fundamental initial step in such studies. Ideally, consistency analysis of a data set should be based on physical considerations. However, this requires an in-depth understanding of the investigated experiments, which is not always possible, especially when the data is collected as a function of multiple condition variables.

This work proposes a generally applicable, model-free mathematical method for identifying inconsistent experimental data points within a data collection. The algorithm fits a series of flexible, multivariate functions (i.e., surfaces) to the collected indirect experimental data of a given type (e.g. laminar burning velocity, ignition delay time, concentrations) as a function of multiple condition variables (temperature, pressure, equivalence ratio, dilution, fuel blending ratio, etc.). In an iterative procedure, inconsistent data points are identified based on whether they fall within or outside the prediction confidence intervals of the fitted surfaces.

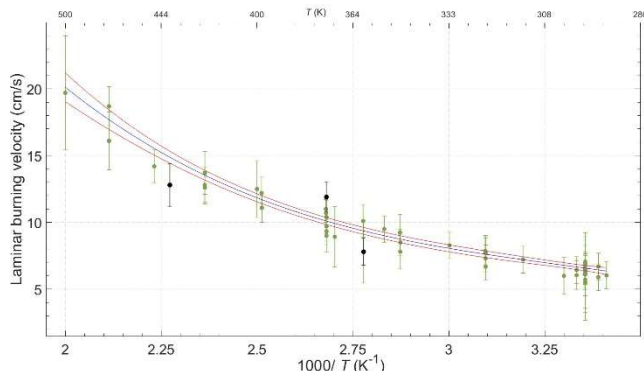


Figure 1. Measured NH_3/air LBV data and the fitted (multivariate) polynomial at $p = 1$ atm and $\varphi = 1$.

The algorithm was tested on a large collection of NH_3/air laminar burning velocity (LBV) measurements. Of the 348 measured data points ($T = 293\text{--}500$ K, $p = 0.3\text{--}10$ atm, $\varphi = 0.7\text{--}1.6$), 14 were eliminated and labelled as inconsistent. Figure 1 shows the consistent (green) and the inconsistent (black) measurements at $p = 1$ atm and $\varphi = 1$ with varying T , along with their respective experimental uncertainties, as well

as the fitted function (blue) and its uncertainty band (red). This automated, strictly mathematical approach also identified the inconsistent data points previously found by Szanthoffer et al. [1] via a simpler algorithm; moreover, it detected additional inconsistent data points.

References

[1] Szanthoffer AG, Papp M, Turányi T, Identification of well-parameterised reaction steps in detailed combustion mechanisms – A case study of ammonia/air flames. *Fuel*. 2025;380:132938.