A Priori Calibration of Transient Kinetics Data via Machine Learning

M. Ross Kunz[1], Adam Yonge[2], Rakesh Batchu[1], Zongtang Fang[1], Yixiao Wang[1], Gregory Yablonsky[4], Andrew J. Medford[2], and Rebecca Fushimi [*1]

[1]Energy & Environment Science and Technology, Idaho National Laboratory
[2]School of Chemical & Biomolecular Engineering, Georgia Institute of Technology
[3]Department of Electronics & Information Systems ELIS, Ghent University
[4]Department of Energy, Environmental & Chemical Engineering, Washington University in St. Louis

October 1, 2021

Abstract

The temporal analysis of products reactor provides a vast amount of transient kinetic information that may be used to describe a variety of chemical features including the residence time distribution, kinetic coefficients, number of active sites, and the reaction mechanism. However, as with any measurement device, the TAP reactor signal is convoluted with noise. To reduce the uncertainty of the kinetic measurement and any derived parameters or mechanisms, proper preprocessing must be performed prior to any advanced analysis. This preprocessing consists of baseline correction, i.e., a shift in the voltage response, and calibration, i.e., a scaling of the flux response based on prior experiments. The current methodology of preprocessing requires significant user discretion and reliance on previous experiments that may drift over time. Herein we use machine learning techniques combined with physical constraints to convert the raw instrument signal to chemical information. As such, the proposed methodology demonstrates clear benefits over the traditional preprocessing in the calibration of the inert and feed mixture products without need of prior calibration experiments or heuristic input from the user.

Keywords: Transient Kinetics, TAP, Cheminformatics, Calibration, and Convex Optimization

*Corresponding author: Rebecca Fushimi, rebecca.fushimi@inl.gov

1 Introduction

Transient kinetic investigations of catalytic materials are ideal for discriminating subtleties that may be unobservable within a steady-state experiment, leading to potentially more information about the underlying reaction mechanism [1]. This manuscript will focus on the kinetic information obtained by the Temporal Analysis of Products (TAP) reactor due to its well-defined operational regimes. The generated highly resolute experimental data leads to the observed evolution of the intermediate products, thereby facilitating the expression of multiple elementary processes simultaneously [2–5]. A TAP reactor experiment consists of a series of transient responses per species that is ideal for repeatable and reproducible kinetic measurements. However, the experimental data is imperfect in measuring the chemical signal as it may include various types of noise. These may consist of intrinsic noise within a pulse (e.g., noise due to electrical signals, heater oscillations, and pulse valve drifts), noise between pulse responses (e.g., gauge error and mass spectrometer drift), or noise due to experimental setup and catalyst preparation (e.g., loading and pressure variations) [6,7]. This noise directly relates to the signals within the mass spectrometer such that each measurement must be calibrated due to unique differences in baseline voltage and scaling factors for each species. It is crucial to calibrate the outlet response appropriately to obtain accurate information about the kinetics, e.g., residence time, apparent kinetic constants, reactivities, gas concentrations, reaction rates, number of active sites, and intrinsic kinetic coefficients [8–14].

The calibration of each TAP flux response differs from typical standardization techniques, i.e., subtracting the mean (μ) and dividing by the standard deviation (σ), as this would result in a negative value for the number of molecules measured [15–17]. As such, traditional methods for calibration require each response to be baseline corrected followed by scaling by a pre-calculated mass spectrometer calibration coefficient (see Figure 1 as an example) [18,19]. Current methods manually detect the baseline through visual inspection leading to potential user biases. Additionally, significant time can be devoted to the mass spectrometer calibration by running multiple inert experiments to determine the correct scaling relationships. However, this does not guarantee that the mass spectrometer will not drift between the inert experiment and the reaction experiment, not to mention that the mass spectrometer can drift within an experiment. As such, misleading kinetic information may be obtained if the calibration is not properly accounted for.

A novel method is developed for treating and calibrating the data without reliance on previous inert experiments while utilizing the relationships of the reactor physics, e.g., the expected distribution of the outlet flux and statistics. More specifically, machine learning via convex optimization with physics-based constraints is used to automatically preprocess the raw data into chemical quantities. The baseline correction and calibration are performed through examining the relationships between the transient flux responses rather
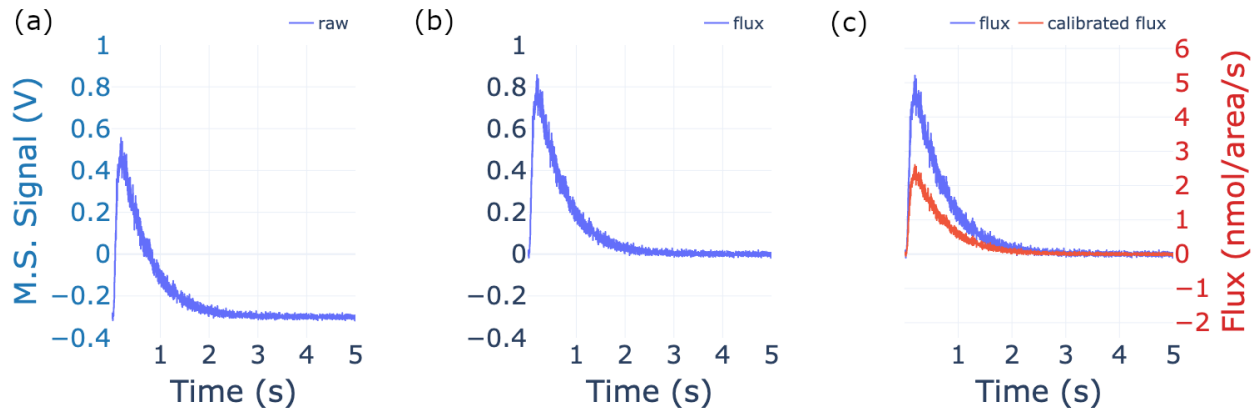
Figure 1: Example of the preprocessing of TAP data. From left to right, the figure starts with a raw voltage measurement which must be baseline corrected by shifting the flux in voltage followed by calibration through scaling the flux (in red).

than averages or summary information, i.e., moments. As such, it will be shown that this method allows for automatic preprocessing of TAP data for consistent kinetic estimates while saving instrument time as there is no reliance on inert experiments.

2 Methodology

2.1 The TAP reactor and standardization

It is vital to understand the application domain prior to any analysis or preprocessing, especially if physical constraints are to be followed. Briefly, the TAP experiment consists of an inert and reactant blend injected into the tubular micro-reactor housing a catalyst sandwiched between two inert materials in thin zone TAP reactor configuration. Each of the products, reactants and inert species diffuse through the TAP micro-reactor and are measured by the mass spectrometer with respect to time. The inert response acts as a control variable describing the transport within the reactor as well as the total number of molecules per volt. The number of molecules within a pulse injection is typically small and is within a low-pressure regime such that only Knudsen diffusion is present in the transport. As such, when the reactant and inert gas are injected in a one-to-one ratio, the total area of the inert response describes the number of molecules injected while the area of the reactant describes the net amount of consumption/reaction by the catalyst [4]. The conversion (χ) is defined as

$$\chi = \frac{\int F_{inert}(t)dt - \int F_{reactant}(t)dt}{\int F_{inert}(t)dt} = 1 - \frac{m^0_{reactant}}{m^0_{inert}} \tag{1}$$

3

where F represents the flux response, t is time, and m^0 is the zeroth moment. Further moments, e.g., m^1, m^2, and m^3, complete the description of the pulse by examining the mean residence time, dispersion and the kurtosis.

The flux response determines the probability distribution of measuring a molecule at a specific time where the moments of the distribution are used to determine apparent kinetic parameters [4, 14]. More specifically, the TAP outlet response can be approximated by the Gamma distribution:

$$f(t; \alpha, \beta) = \frac{1}{\gamma(\alpha)\beta^\alpha} t^{\alpha-1} e^{-t/\beta} \tag{2}$$

where t is time, α is the shape, and β is the scale [12]. When only diffusion is present within the reactor, an α of 1.5 and β of $1/3$ corresponds to the dimensionless standard diffusion curve. Using the properties of the Gamma distribution, the mean residence time (τ) is measured as $\alpha\beta = m^1/m^0$, variance residence time (τ_{σ^2}) as $\alpha\beta^2 = m^2/m^0 - \tau^2$, peak residence time ($\tau_p$) as $(\alpha - 1)\beta = argmax(flux)$, and the area normalizing coefficient (τ_A) as $\Gamma(\alpha)\beta^\alpha$. These properties do not assume a specific reaction mechanism and when combined provide consistent explanations of the reaction with respect to each kinetic measurement. Therefore, this information can be leveraged in the standardization of the experimental data.

The TAP experiment consists of a series of transient flux responses for each of the diffused gasses from the mass spectrometer. There are two different types of errors that can occur: the error associated within a pulse response and the error between pulse responses. The errors within a pulse response, caused by heater oscillations, electrical signal, etc., contribute to a baseline shift in the total number of molecules, i.e., the zeroth moment of the flux (m^0) [6]. On the other hand, errors between pulse responses, caused by drift in mass spectrometer detector sensitivities etc., affect the measurements of the total number of active sites and apparent kinetics. To account for these errors, the traditional method of calibration is done through manual centering a flux by the baseline mean (μ_b) while scaling each gas pulse series by their respective mean area (μ) of an inert experiment:

$$\boldsymbol{m}_C^0 = \left(\int F_i(t) - \mu_{b,i} dt \right) / \mu \tag{3}$$

where i denotes each flux response in a pulse series and $\boldsymbol{m}_C^0$ denotes a vector of calibrated areas. This is similar to standard normalization practices to account for different scales and centers amongst a variety of variables. Mathematically, normalization is the process of subtracting the mean (μ) and dividing by the standard deviation σ of a vector of responses:

$$\boldsymbol{m}_N^0 = \left(\boldsymbol{m}^0 - \mu \right) / \sigma \tag{4}$$

4

where $\boldsymbol{m}_N^0$ is a vector of normalized areas. Equations 4 and 3 both have the same goal of standardizing the data, but the shift in Equation 3 is constrained such that the number of molecules measured is not negative. Additionally, the scaling factor differs by using a first or a second order moment. However, if the distribution of the $\boldsymbol{m}^0$ is skewed, the first and second order moment have a linear relationship and hence can be used interchangeably [20]. As such, it is the goal to show the relation between these two equations such that advanced data analysis may be applied to the traditional calibration. This will be done through smoothing the flux to account for the noise within a flux, baseline correction to account for the mean shift or centering, and calibration to account for the scaling factor between species.

2.2 Smoothing the flux

First, to separate the chemical signal from the random noise within a flux, a univariate cubic spline approximation is applied [21]. This enables robust estimation of the distributional parameters, such as the peak residence time, while ensuring that the flux is neither under/over-represented. While this step is minor in terms of preprocessing, it ensures that the calibration is only determined by the chemical signal and not the noise. This reduction in noise within a flux is critical in obtaining accurate estimates of the residence time distributional properties as the heteroscedastic nature of the noise will affect the measurement of the moments [6].

2.3 Baseline correction

In accordance with the probability distribution and the calibration requirements, the baseline of the flux response must be zero. However, in an experimental setting, the baseline may shift due to within pulse variation and the baseline may not be reached for chemical species that have longer elution time than the collection time. To address this issue, the baseline can be estimated using the shape and scale parameters of the Gamma distribution of a flux. The shape of the Gamma distribution affects the outlet flux by time, i.e., time of position in the reactor, while the scale affects the kinetic rates [12]. As such, the same shape parameter should be used in either case of pure diffusion or reaction occurring in the outlet flux. The peak residence time, i.e., the mode, property of the distribution is used for the estimation of the distribution parameters as it will not be affected by shifts in the baseline. Assuming a shape parameter of an ideal reactor at 1.5, then the scale of a flux may be estimated, using the mode property of the Gamma distribution, as two times the peak residence time, i.e., $\hat{\beta} = 2\tau_p$. Once the scale is estimated, the probability of the distribution at the end of the experiment time may be determined by Equation 2. The tail of the flux can then be shifted to match the probability of the Gamma distribution providing a similar shift to the normalization procedure

in Equation 4. Algorithm 1 describes the steps required to perform baseline correction based on the Gamma distribution.

Algorithm 1 Baseline correction via distributional assumptions.

Require: Let F represent the gas flux, t_{end} denote the time at the end of the flux, $f(\cdot)$ denotes the Gamma distribution defined in Equation 2, and $F[t_{end}]$ denotes the voltage of the flux at the end of the experiment time.

1: $\tau_p = argmax(F)$
2: $\Gamma_{tail} = f(t_{end}; 1.5, 2\tau_p)$
3: $\bar{F} = F - F[t_{end}] + \Gamma_{tail}$

2.4 A priori calibration between gas species using moments

Ideally, the scale of the voltage response given by the mass spectrometer would measure the exact number of molecules for each response. However, the differences in voltage scale between gases is dictated by multiple attributes which may include the ionization efficiency, gain of the electron multiplier, and mass differences based on instrument drift [22]. It is necessary to determine the appropriate scale, based on the known feed mixture blend amounts, to measure the kinetic coefficients, gas concentrations, and reaction rates. These scaling coefficients, i.e., calibration coefficients (ζ), are traditionally calculated within an inert experiment where the mean $\bar{m}_g^0$ of the reactants/products are scaled to the mean inert $\bar{m}_I^0$, i.e., $\bar{m}_g^0 = \zeta \bar{m}_I^0$, where $\bar{m}$ denotes the mean moment. If the baseline relationship between the two gases is exactly zero, i.e., the baseline is appropriately determined for both species, the calibration coefficient ζ can be directly solved through the use of an Ordinary Least Squares (OLS) model where the scaling relationship is due to the standard deviation as in Equation 4. Note that the scaling factor (μ or σ) in Equations 3 or 4 can be used interchangeably due to the linear relationship between the mean and variance of the Gamma distribution, i.e., $\tau_{\sigma^2}/\tau = \beta$. The use of a linear model, rather than the mean estimate, has additional benefits as the intercept can be verified if it is exactly zero and robust modifications may be used to account for outgassing within the series, i.e., an additional pulsed amount of gas within the reactor at an unexpected time [23].

In application to a reaction experiment, changes in the mass spectrometer measurement are assumed to be fixed when moving from the previous inert experiment. However, this is rarely the case as the instrument will drift in time. Constant inert calibration is impractical as this would require a complete work stop to exchange the catalytic material to an inert bed in the micro-reactor and hence potentially change the experimental conditions. However, the calibration coefficient can be extended to a reaction setting using the information gained by the Gamma distribution. More specifically, the distribution normalizing coefficient τ_A estimates the effect of conversion based on the residence time distribution properties. As such, the $\boldsymbol{m}_g^0$ is a linear combination of the area due to the conversion and the instrument drift:

$$\boldsymbol{m}_g^0 = \mu + \zeta_1 \boldsymbol{\tau}_A + \zeta_2 \boldsymbol{m}_I^0 \tag{5}$$

where μ, ζ_1, ζ_2 are estimated through application of an OLS model. The value of μ is the intercept and hence determines the shift between each species. If the baseline was determined prior to Equation 5, the the value of μ should be zero. The value of ζ_1 indicates the scaling relation to the m_g^0 and the distribution normalizing coefficient. The size of ζ_1 is based on the amount of conversion. The last term ζ_2, is the regression coefficient describing the scale from the inert to the gas species, i.e., accounting for the mass spectrometer differences between each gas species. Thus, a calibration coefficient from the inert to the reactant/product m_g^0 can be obtained even in the presence of reaction.

After calibration, the following m^0 relationships should hold between the reactants (r), products (p), and the inert (I):

If the reaction is reversible:

$$m_r^0 = m_I^0 \text{ and } \tilde{m}_r^1 > \tilde{m}_I^1$$

If the reaction is irreversible:

$$m_r^0 < m_I^0, \tilde{m}_r^1 < \tilde{m}_I^1 \text{ and } m_r^0 + m_p^0 \leq m_I^0$$

where $\tilde{m}^1$ is the area normalized first moment. This information assumes an 50/50 blend pulse injection between the inert and the reactant. However, this still applies to unequal blends, but a scaling coefficient based on the blend amounts must be applied, e.g., if reversible, then the m^0 amounts per inert and reactant correspond to the blend amount. The first relationship states that if the normalized first moment of the reactant is greater than the inert, then the reaction is reversible, and the calibration coefficient can be set such that the area of the reactant equals the area of the inert. The second relationship states that the area of the reactant must be less than the area of the inert based on the amount of conversion. If the normalized first moments are equal, then no reaction has occurred. The final relationship describes the balance between the number of molecules injected and the combination of the reactants and products. The less than or equal to sign allows flexibility in the event there is accumulation of a species on the surface of the catalyst. Algorithm 2 describes the steps required for calibration within a reaction experiment using momentary information.

Algorithm 2 Calibration scaling via distributional assumptions.

Require: Let $\bar{F}$ represent the gas flux that is baseline corrected using Algorithm 1, I denote the inert gas, and g denote a reactant/product gas.

1: **for each** i^{th} pulse **do**
2:     $m_{I,i}^0 = \int \bar{F}_{I,i}(t)dt$
3:     $m_{g,i}^0 = \int \bar{F}_{g,i}(t)dt$
4:     $\tau_{A,i} = \Gamma(\alpha_i)\beta_i^{\alpha_i}$ of $\bar{F}_{g,i}$
5: **end for**
6: Solve $\boldsymbol{m}_g^0 = \mu + \zeta_1\boldsymbol{\tau}_A + \zeta_2\boldsymbol{m}_I^0$ via OLS or robust OLS.

2.5 A priori calibration between gas species using transient information and machine learning

Given the moment-based relationships between the gas species, the relationship between the transient responses may be leveraged in a similar fashion to Equation 5. The goal then is to compare the gas to the inert species per pulse to determine the appropriate calibration coefficient for an irreversible reactant and each product within a reaction experiment. First, consider an irreversible reactant and the reactant flux has been transformed to have the same velocity as the inert flux via Graham's Law. Based on the m^0 relationship, the reactant flux must have a positive scaling relationship to the inert flux, the difference between the calibrated reactant flux and the inert flux must be greater than or equal to zero when a product/surface species is formed, and the area of the reactant flux must be less than the inert flux. Numerically, the calibration coefficient can be solved as a linear model between the inert and the reactant with the following constraints:

$$\min_b \sum_{i=1}^{N} \left(F_{I,i} - F_{r,i}b\right)^2 \tag{6}$$

$$\text{such that: } b \geq 0, F_{I,i} - F_{r,i}b \geq 0, \text{ and } m_I^0 \geq m_r^0 \tag{7}$$

where b is the calibration coefficient from the reactant species to the inert species. The objective function and the first constraint consist of a variation of OLS called non-negative least squares, the second constraint ensures there is not a negative number of molecules measured in the residuals, and the last constraint ensures that the area of the reactant flux is less than the inert flux area. The combination of the objective function and the constraints will be denoted as Temporal Convex Constrained Optimization (TCCO) for ease of referencing throughout the paper.

The application of TCCO does not have to be restricted to the relationship between the inert and a reactant gas species. For example, a series of inert flux may be calibrated to a single inert reference sample such that all fluxes are calibrated to one pulse injection. In general, TCCO can be applied to several reactants

and products or even mass spectrometer fragmentations but requires the alteration of Equations 6-7 to a multiple linear regression form. Let $\boldsymbol{y}$ be a vector of the flux in which to calibrate to, i.e., the Dependent Variable (DV) typically the inert flux, $\boldsymbol{X}$ be a matrix of flux responses, i.e., the set of Independent Variables (IV), and $\boldsymbol{res}$ denote the residuals, then

$$\min_b ||\boldsymbol{y} - \boldsymbol{Xb}||_2^2 \tag{8}$$

$$\text{such that: } \boldsymbol{b} \geq 0, \boldsymbol{res} \geq 0, \text{ and } m_y^0 \geq m_X^0 \tag{9}$$

where m_X^0 is the sum of the zeroth moments for each variable in $\boldsymbol{X}$. Equations 8-9 are convex in optimization and can be solved using open-source solvers [24, 25].

2.6 Transient Error Analysis for Kinetic measurements

The smoothing, baseline correction, and calibration coefficient calculation are combined to form the Transient Error Analysis for Kinetic (TEAK) measurements algorithm. For ease of use, the complete workflow is given in Algorithm 3. Software implementation of TEAK as well as commonly used TAP analysis techniques can be found at `https://github.com/IdahoLabResearch/tapsap`.

Algorithm 3 Transient Error Analysis for Kinetic measurements (TEAK).

Require: Let F represent the gas flux, *smooth* represent the application of the spline approximation in Section 2.2, I denote the inert gas, and g denote a reactant/product gas.
1: **for each** i^{th} pulse **do**
2:     $\hat{F}_{g,i} = smooth(F_i)$
3:     $\bar{F}_{g,i} = baseline(\hat{F}_{g,i})$
4:     $\tilde{F}_{g,i} = tcco(\hat{F}_{I,i}, \hat{F}_{g,i})$
5: **end for**

---

3 Experimental Data

3.1 Inert experiment with varying temperature

A series pulse response data of argon was obtained from the Temporal Analysis of Products reactor, TAP-3E. The tubular micro-reactor housing in TAP-3E is of $0.00385m$ internal diameter and $0.0564m$ in length. In Knudsen conditions, argon is pulsed over an inert quartz bed $(250 - 300\mu m)$ at five different temperatures between 100 to $500°C$ with $100°C$ interval. At each of the temperature, two identical experiments are conducted for reproducibility. Each experiment is recorded for 100 pulses with a collection time of $5s$ and

pulse spacing of $5.1s$. Argon is monitored at $m/z = 40$ Atomic Mass Unit (AMU) from a quadrupole mass spectrometer.

3.2 Inert experiment with calibration

The inert experiments were performed in TAP by pulsing carbon dioxide and argon (volume ratio with 50:50) over quartz at room temperature. An electronic pulse width of $135\mu s$ was sent to the valve every $5.1s$ for a total of 1000 pulses. At each pulse a different atomic mass unit (AMU) was monitored cycling between 40, 44, 28 16 and 12 to record the pulse response for argon, carbon dioxide, the fragmentation of carbon dioxide on mass 16 and 12, respectively.

3.3 Reaction experiment

The Strong Electrostatic Adsorption (SEA) method was used for the synthesis of $1.0wt\%$ Pt/SiO_2 catalyst for the oxidation experiments. A commercial silica (AEROSIL OX50, $50m^2/g$) from EVONIK was chosen as the support and a precursor of tetraamine platinum (II) hydroxide ($Pt(NH_3)4(OH)2$, 99%, from Aldrich) was employed to deposit the metal. The precursor was dissolved in deionized (DI) water and the initial pH was adjusted to 11.5 with NaOH. Silica was added to the solution and the contents were shaken for one hour. The resulting mixture was washed with DI water, filtered, and dried overnight under vacuum. The material was pressed and sieved, retaining the $250 - 300\mu m$ fraction. Next, the catalyst was pretreated ex-situ in 50% oxygen and argon flow ($30mL/min$) at $400°C$ for $30min$ followed by reduction in 4% hydrogen and argon flow ($50mL/min$) at $400°C$ for one hour. The ex-situ oxidation and reduction was performed with three cycles.

Approximately $15.6mg$ of pretreated catalyst with the particle size of $250 - 300\mu m$ was loaded between two zones of the same particle size quartz sand (Sigma Aldrich). The total length of the reactor was $0.0564m$, with a catalyst zone of $0.002m$, and a cross sectional area of $1.256 \times 10^{-5}m^2$. The TAP reactor was evacuated at $300°C$ to a pressure of $1 \times 10^{-7}torr$ and the catalyst was subjected to at least three cycles of alternating pulses of 200 pulses of carbon monoxide and argon and 200 pulses of oxygen and argon to activate the platinum and reach a reproducible starting point for pulsing experiments. Prior to oxygen adsorption, the catalyst was again reduced at $300°C$ by introducing a sequence of 50% carbon monoxide and argon pulses until no carbon dioxide formation was detected. The TAP reactor was subsequently heated to $500°C$ and kept for $30min$ to remove adsorbed carbon monoxide and then cooled to the desired temperature for testing oxidation. The adsorption of oxygen on the catalyst was recorded in separate experiments by pulsing a 1:1 oxygen and argon mixture at $300°C$ with different pulsing intervals of 2.0, 2.5, 3.0, 3.5, and $4.0s$. The time

evolution of three mass fragments was followed, namely argon (AMU 40), oxygen (AMU 32), and carbon dioxide (AMU 44). There was no carbon dioxide production detected at the beginning of each oxidation experiment.

4   Results and Discussion

To thoroughly test the TEAK preprocessing methodology, several different cases are applied: simulated data with a known outlet flux calibration coefficient, an inert experiment where argon is exposed to multiple temperatures, an inert experiment where two different gases are calibrated to one another, and finally a reaction experiment where oxygen interacts with a platinum catalyst. Each simulation/experiment is selected to show the benefits of TEAK in an incremental fashion. The simulated data shows the benefits of calibration from flux to flux using TCCO, the inert experiment with temperature variation gives an example of TCCO over multiple data sets of a similar nature, the inert experiment with calibration shows how TEAK deals with non-linear drift, and the reaction experiment shows how TEAK can achieve consistent measurements of the conversion without prior calibration experiments.

4.1   Application to simulated data

Simulated data was generated to produce a chemical signal with arbitrary scaling values. The simulated reactor consists of a thin-zone setup with the total length equal to $1cm$, a bed porosity of 0.5, and a diffusion coefficient of $0.5cm^2/s$. A total of $1mol$ was injected into the reactor and measured over three seconds. The reaction consists of a single elementary step where a reactant gas irreversibly interacts with a catalyst at the rate of $1.15mol/s$. After which, the reactant outlet flux was scaled arbitrarily by values of 2.3 and 0.23 to simulate the mass spectrometer measuring the reactant at a non-ideal scale as shown in Figure 2.

The data generated is ideal for validating potential calibration coefficients for TEAK as it has a known calibration coefficient and no smoothing nor baseline correction is required. Only irreversible reaction is considered in the simulation as the reversible reaction would return the same number of molecules as the inert flux. Application of TCCO assumes the DV is the inert flux while the IV only consists of the scaled reactant flux (2.3 or 0.23). TCCO applied to the inert and scaled reactant flux determined the calibration coefficient within 1×10^{-7} of the true calibration coefficients (2.3 or 0.23).

4.2   Application to the inert experiment with varying temperature

The inert experiment with varying temperature consists of measuring pulsed argon within an inert experiment over several temperatures. The objective of this example is to verify that given clean data, the TEAK
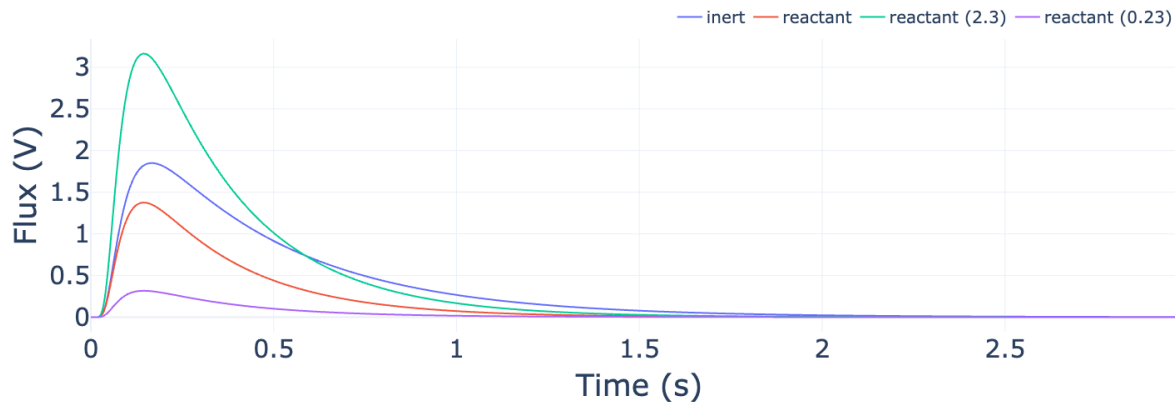
11

Figure 2: Simulated reaction data that is scaled arbitrarily to test the constrained convex optimization in Section 2.5.

methodology applied from flux to flux is consistent with the traditional series of moment-based calibration. First, each individual temperature was pre-processed independently where the traditional method baseline corrected each flux by the mean of the flux tail (from 4.5 to 5s). The zeroth moments produced by the traditional method were then compared to the data being baseline corrected using Algorithm 1 followed by using TCCO within the flux series. More explicitly, a single pulse index was selected arbitrarily (3rd pulse) as the DV and TCCO was applied to each other flux in the series as the IV independently.

Prior to the application of TEAK, each of the flux were smoothed independently via the cubic spline given in Section 2.2. This allows for a better estimate of the peak residence time used in baseline correction and calibration to the signal rather than the noise in TCCO. For example, Figure 3 displays the smoothed flux where the peak height and flux shape does not significantly differ from that in the flux response. Additionally, the residuals of the smoothing produces a Gaussian distribution consistent with previous studies [6].

The comparison between the m^0 measurements after calibration can be seen in Figure 4. The zeroth moments of the traditional method exhibits a minor slope over the pulse number while there is no drift within the TEAK results. It would at first appear that TEAK exhibits poor measurements in select pulse numbers, e.g., 9th, 16th, 80th, etc. However, upon closer inspection of the individual flux, each of the respective flux contains outgassing. These moments with outgassing are easily identified using TEAK where a user may exclude these points from the kinetic analysis. Comparison of the mean zeroth moments show relatively similar values of 0.518 and 0.525 for the traditional and TEAK methods. However, TEAK significantly reduces the variance of the zeroth moments, by a factor of 32 times, after accounting for the outgassing effects.
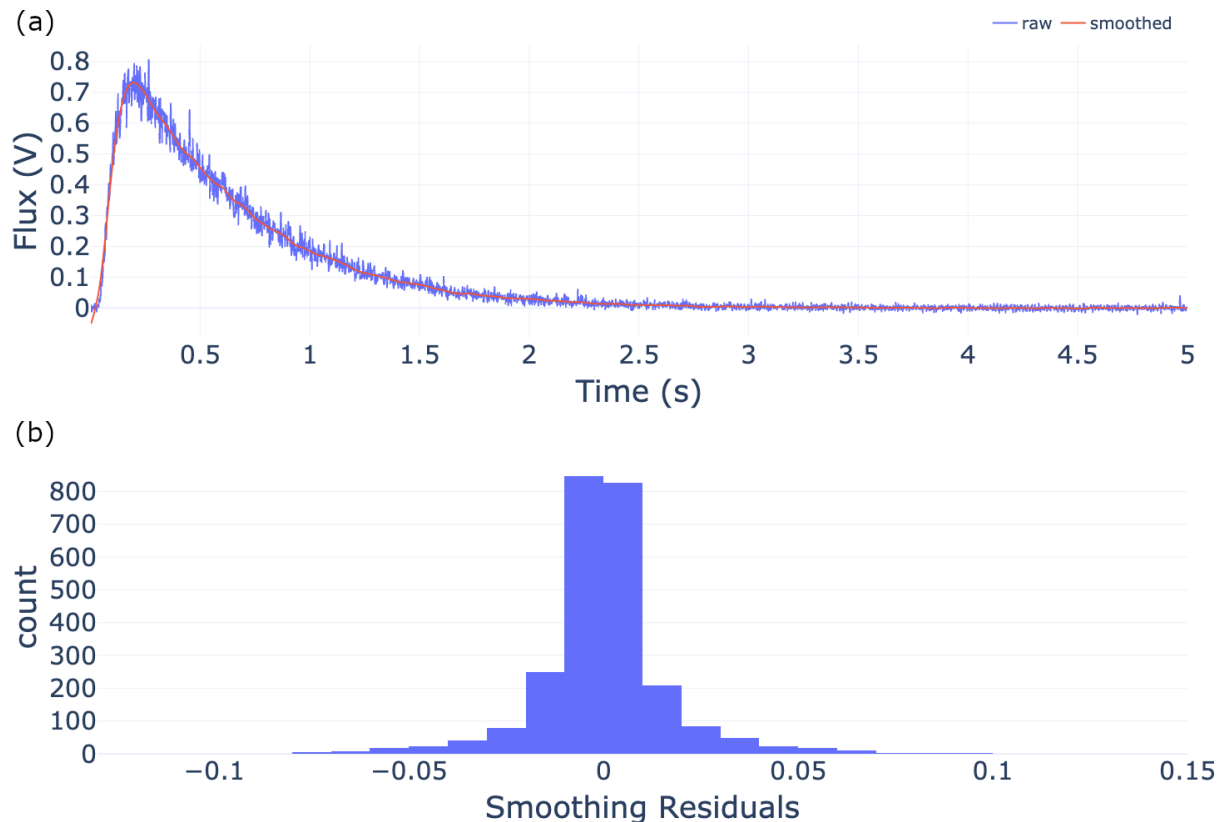
12

Figure 3: A single spline smoothed argon response from the inert data with varying temperature at $100°C$. Subfigure (a) describes the relationship between the noisy outlet response and the estimated trend. Subfigure (b) describes the obtained residual noise.

4.3 Application to the inert experiment with calibration

The inert experiment with calibration shows how TEAK can account for non-linear drift between pulse measurements and how to calibrate two different gas species (argon and carbon dioxide). Additional fragmentations of the carbon dioxide were inspected to show evidence of the Gamma distribution relationship between the mean and variance residence time.

First, each flux was baseline corrected individually, from the flux tail using the traditional method and from Algorithm 1 in TEAK. The zeroth moments of each mass were measured for determining the calibration coefficient in the traditional method. Figure 5 (a) displays the zeroth moments prior to calibration. A sinusoidal trend over the m^0 responses can be detected through visual inspection. This trend exposes a weakness in the traditional method of calibration as different means of the m^0 may be extracted depending on when the experiment was conducted with respect to the pulse index. Figure 5 (b) compares the mean and standard deviation of the zeroth moments. There exists a linear trend of the mean and standard deviation with a slope of 0.2. This linear relationship indicates a skewed distribution, such as a Gamma distribution,
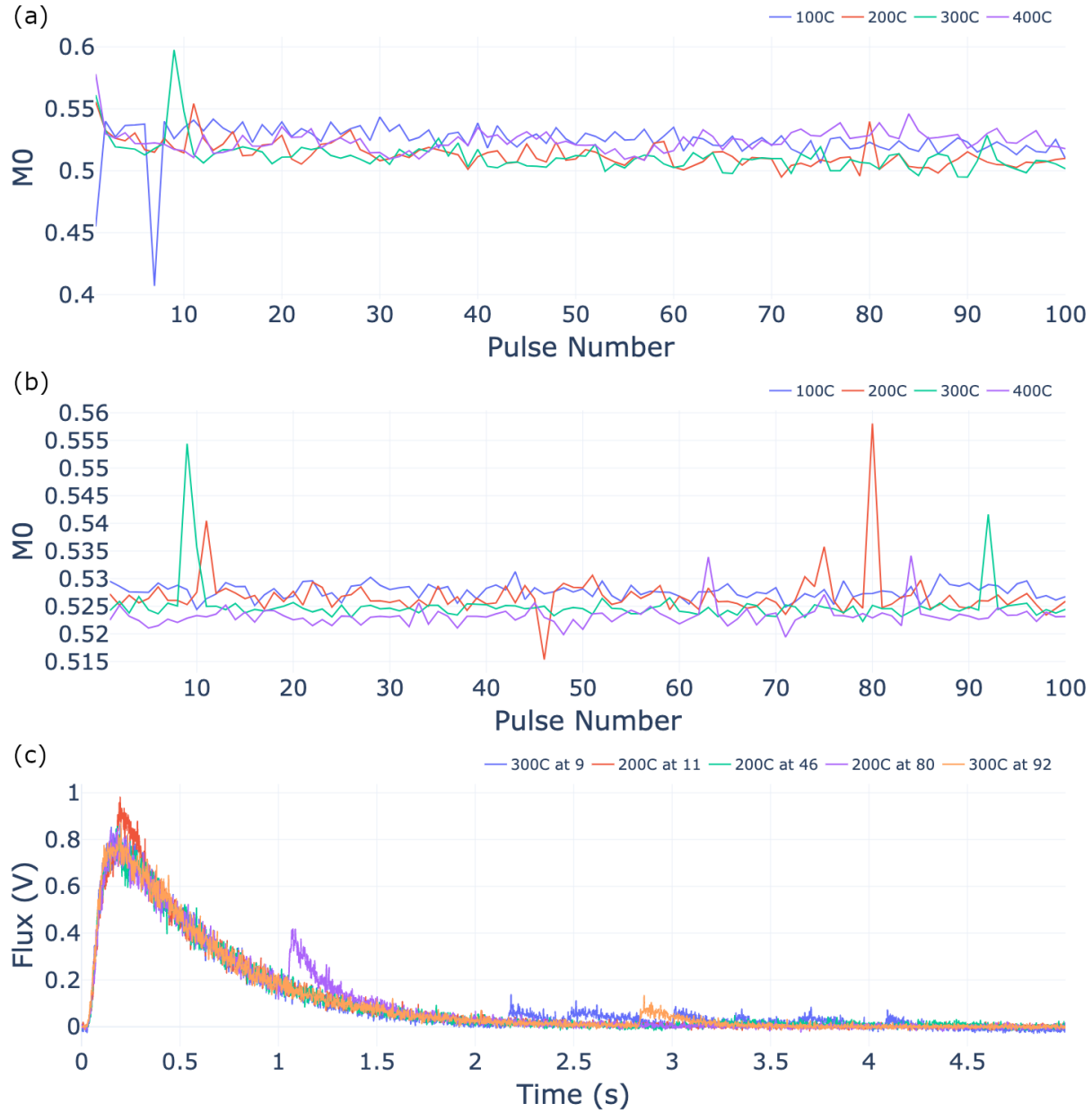
Figure 4: Comparison of the traditional calibration in subfigure (a) and TEAK in subfigure (b) applied to the inert temperature varying data. Subfigure (c) gives examples of outgassing which affects the zeroth moment measurements.

within the mass spectrometer as indicated in Section 2.4.

After baseline correction, the calibration coefficient for the traditional method was calculated as the ratio of the mean zeroth moments over the whole series. The TEAK method applies two forms of calibration. First, TCCO was applied to each argon flux using the 5th argon pulse as the DV. The choice of the DV pulse number will change the overall m^0 of each calibrated flux, but the relationship between the DVs and IVs after calibration is consistent. Additionally, the choice of a DV should not include any outgassing effects.

Second, TCCO was applied to each corresponding pulse number of carbon dioxide as the IV where the argon flux was the DV. Figure 5 (c) compares the m^0 measurements for argon and carbon dioxide obtained by the traditional method and TEAK. The argon TEAK pulse is difficult to discern as it almost exactly corresponds to the carbon dioxide series. TEAK removes the non-linear drift as well as reduces the mean square error of the zeroth moment difference between argon and carbon dioxide by nearly 3 orders of magnitude (9.5×10^{-3} compared to 1.1×10^{-5}).

4.4   Application to the reaction experiment

The final example consists of an irreversible reaction of oxygen interacting with a platinum catalyst. This experiment is particularly difficult for the traditional calibration as there exists mass spectrometer drift from pulse to pulse, the calibration coefficient from the inert experiment is not an exact description of the mass relation, and there are many instances of outgassing. Additionally, with reaction, the m^0 response will change the mean and variance of each mass in the reaction depending on the amount of conversion. In the traditional calibration method, it is assumed that the mass spectrometer variation within an inert experiment may be used to dictate the drift in the reaction experiment. TEAK avoids this assumption by performing the calibration with respect to each individual flux response using TCCO.

In preparation of the zeroth moment values, the traditional method preprocessing consisted of baseline correcting the flux via the tail of the flux and scaling oxygen to argon through a calibration coefficient of a prior inert experiment. The TEAK preprocessing workflow consisted of smoothing the data, baseline correction via Algorithm 1, application of TCCO to argon using a reference argon flux, and application of TCCO from the argon flux (the DV) to the oxygen flux (the IV) on a pulse-by-pulse basis.

The resulting zeroth moments of each method are given in Figure 6 (a) and (b) after removal of the outgassing pulses. The removal of the outgassing effects can be accomplished automatically using a simple moving window t-test or through time series anomaly detection methods [26]. However, it is recommended to confirm the outgassing through visual inspection prior to removal. The argon zeroth moment of the traditional method exhibits drift from pulse to pulse while the oxygen, after approximately 300 pulses, appears to follow the same drift pattern. It is difficult to determine at what point the variation over a series of pulses were due to drift or conversion. On the contrary, the TEAK argon zeroth moments are consistent across the pulse number while the oxygen values converge to the argon series as the pulse number increases. Significant differences in the measured zeroth moments exist early in the pulse index (less than 150) which will affect the interpretation of the characterization. Figure 6 (c) displays the conversion values for each method, i.e., $1 - m_{O_2}^0/m_{Ar}^0$. The TEAK method was able to capture the conversion values early in the
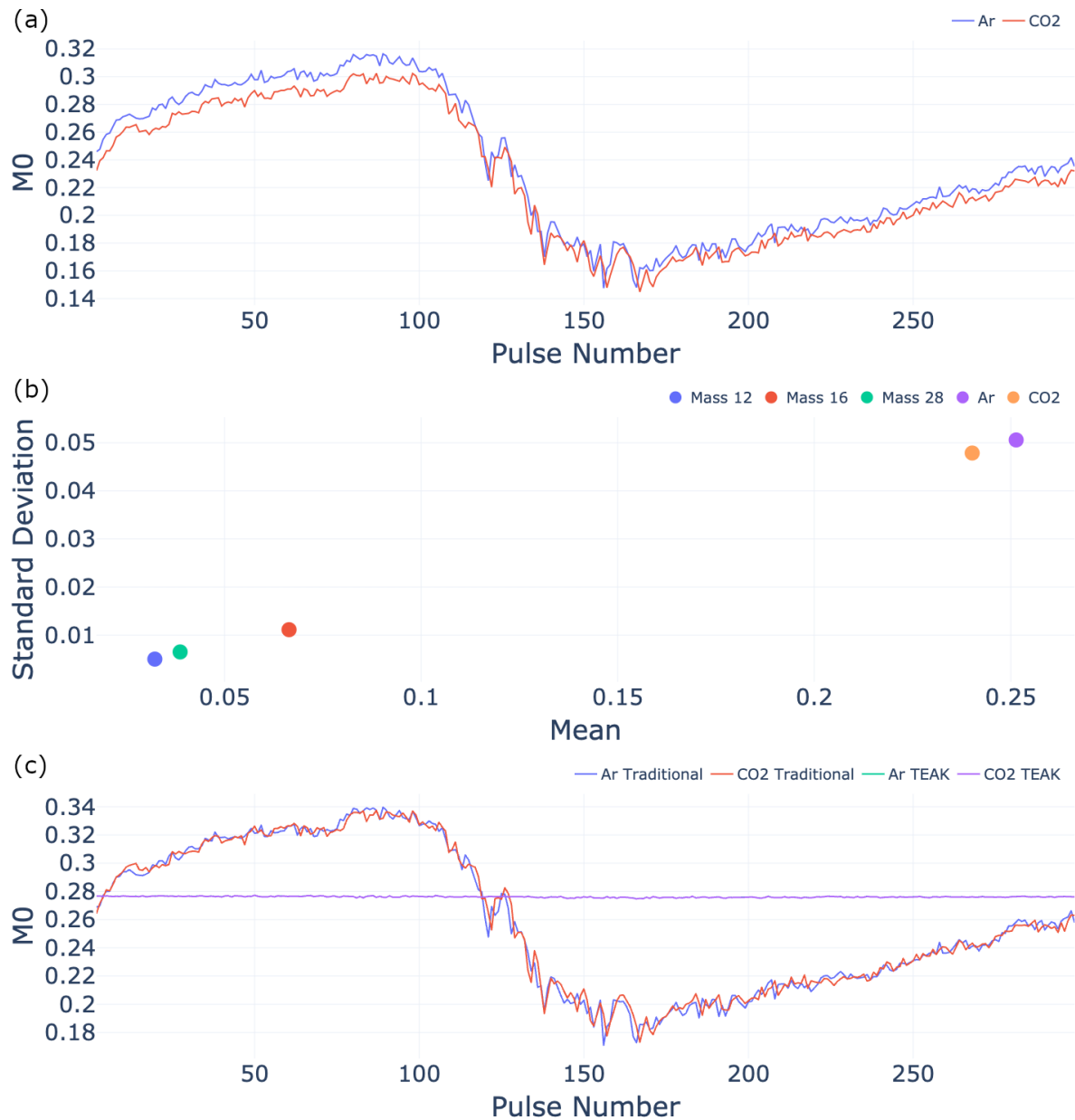
Figure 5: The raw inert reactor information for each mass collected by the mass spectrometer. Subfigure (a) displays each uncorrected m^0 value of argon and carbon dioxide with respect to the pulse index number. Subfigure (b) denotes the relationship between the mean and standard deviation of the argon, carbon dioxide, and fragmentations. Subfigure (c) displays the corrected m^0 for the traditional calibration and TEAK. The argon and carbon dioxide values of TEAK are visually indiscernible when compared to the same scale of the traditional calibration.

pulse number where the noise dominates the signal. Additionally, TEAK showed that the oxygen is no longer being consumed at the 300th pulse number while the traditional method suggests that reaction may still be occurring. The zeroth moment measurement can be considered the foundational characterization of the experiment as it relates to the conversion, apparent kinetics, Damköhler number, and measurement

16

of the total number of active sites [14, 27]. Furthermore, when examining transient kinetic responses, the reaction rate and gas concentration measurements will be affected by the scale of the flux. More specifically, the reaction rate is determined as a function of the difference of the inert diffusion and the reactant flux response, i.e., $R = f(I - F)$ [8, 9, 12, 18]. When the inert is improperly scaled, the transient rate will be over/underestimated with respect to the concentration and may lead to incorrect conclusions regarding the chemical kinetics.
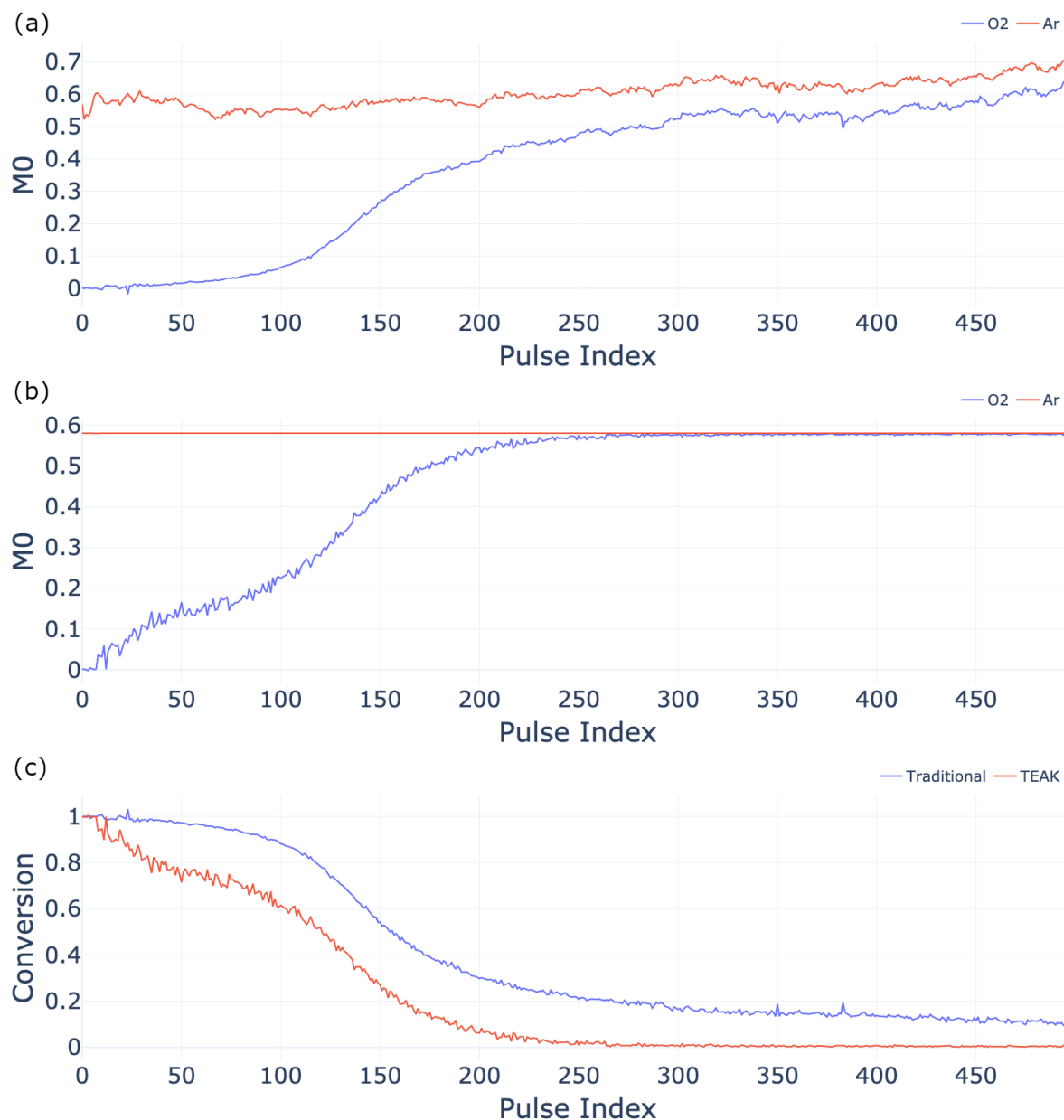

Figure 6: Zeroth moment values obtained from the traditional calibration in subfigure (a) compared to TEAK in subfigure (b). Subfigure (c) compares the two different conversion values obtained for each method.

5 Conclusion

This manuscript has demonstrated the use of residence time distribution properties and machine learning through convex optimization with physical constraints to deconvolute a noisy signal into a chemical signature. This type of analysis was enabled through the repeatable and reproducible measurements of the TAP reactor between each gas/surface interaction and over a series of pulse responses. The developed methodology allows for separation of the noise due to external environmental factors, detrending the series of outlet responses due to drift within the system, and mass spectrometer calibration coefficient estimates without prior inert experiments.

The proposed methodology was verified through analysis of inert and reaction experiments with a range of conditions. As a result, more consistent estimates of kinetic measurements are obtained over the traditional calibration methodology. Additionally, the proposed methodology does not rely on user input for standardization of the data and hence has less potential for user biases while creating a streamlined process to convert instrument measurements to kinetic information.

References

[1] C. Bennet, "Understanding heterogeneous catalysis through the transient method," *Catalysis Under Transient Conditions*, vol. 178, 1982.

[2] K. Morgan, N. Maguire, R. Fushimi, J. Gleaves, A. Goguet, M. Harold, E. Kondratenko, U. Menon, Y. Schuurman, and G. Yablonsky, "Forty years of temporal analysis of products," *Catalysis Science & Technology*, vol. 7, no. 12, pp. 2416–2439, 2017.

[3] J. T. Gleaves, J. Ebner, and T. Kuechler, "Temporal analysis of products (tap)—a unique catalyst evaluation system with submillisecond time resolution," *Catalysis Reviews Science and Engineering*, vol. 30, no. 1, pp. 49–116, 1988.

[4] J. T. Gleaves, G. S. Yablonskii, P. Phanawadee, and Y. Schuurman, "Tap-2: An interrogative kinetics approach," *Applied Catalysis A: General*, vol. 160, no. 1, pp. 55–88, 1997.

[5] J. T. Gleaves, G. Yablonsky, X. Zheng, R. Fushimi, and P. L. Mills, "Temporal analysis of products (tap)—recent advances in technology for kinetic analysis of multi-component catalysts," *Journal of Molecular Catalysis A: Chemical*, vol. 315, no. 2, pp. 108–134, 2010.

[6] R. Roelant, D. Constales, G. S. Yablonsky, R. Van Keer, M. A. Rude, and G. B. Marin, "Noise in temporal analysis of products (tap) pulse responses," *Catalysis Today*, vol. 121, no. 3-4, pp. 269–281, 2007.

[7] D. Constales, G. S. Yablonsky, D. R. D'hooge, J. W. Thybaut, and G. B. Marin, *Advanced data analysis and modelling in chemical engineering*. Elsevier, 2016.

[8] G. Yablonsky, D. Constales, S. Shekhtman, and J. Gleaves, "The y-procedure: How to extract the chemical transformation rate from reaction–diffusion data with no assumption on the kinetic model," *Chemical Engineering Science*, vol. 62, no. 23, pp. 6754–6767, 2007.

[9] E. A. Redekop, G. S. Yablonsky, D. Constales, P. A. Ramachandran, C. Pherigo, and J. T. Gleaves, "The y-procedure methodology for the interpretation of transient kinetic data: Analysis of irreversible adsorption," *Chemical engineering science*, vol. 66, no. 24, pp. 6441–6452, 2011.

[10] E. A. Redekop, G. S. Yablonsky, D. Constales, P. A. Ramachandran, J. T. Gleaves, and G. B. Marin, "Elucidating complex catalytic mechanisms based on transient pulse-response kinetic data," *Chemical Engineering Science*, vol. 110, pp. 20–30, 2014.

[11] G. S. Yablonsky, E. A. Redekop, D. Constales, J. T. Gleaves, and G. B. Marin, "Rate-reactivity model: A new theoretical basis for systematic kinetic characterization of heterogeneous catalysts," *International Journal of Chemical Kinetics*, vol. 48, no. 6, pp. 304–317, 2016.

[12] M. R. Kunz, R. Batchu, Y. Wang, Z. Fang, G. Yablonsky, D. Constales, J. Pittman, and R. Fushimi, "Probability theory for inverse diffusion: Extracting the transport/kinetic time-dependence from transient experiments," *Chemical Engineering Journal*, p. 125985, 2020.

[13] M. R. Kunz, A. Yonge, Z. Fang, R. Batchu, A. J. Medford, D. Constales, G. Yablonsky, and R. Fushimi, "Data driven reaction mechanism estimation via transient kinetics and machine learning," *Chemical Engineering Journal*, vol. 420, p. 129610, 2021.

[14] D. Constales, G. S. Yablonsky, L. Wang, W. Diao, V. V. Galvita, and R. Fushimi, "Precise non-steady-state characterization of solid active materials with no preliminary mechanistic assumptions," *Catalysis Today*, vol. 298, pp. 203–208, 2017.

[15] J. Miller and J. C. Miller, *Statistics and chemometrics for analytical chemistry*. Pearson education, 2018.

[16] P. Gemperline, *Practical guide to chemometrics*. CRC press, 2006.

[17] R. G. Brereton, *Applied chemometrics for scientists*. John Wiley & Sons, 2007.

[18] M. R. Kunz, T. Borders, E. Redekop, G. S. Yablonsky, D. Constales, L. Wang, and R. Fushimi, "Pulse response analysis using the y-procedure: A data science approach," *Chemical Engineering Science*, vol. 192, pp. 46–60, 2018.

[19] C. Reece, *Kinetic Analysis and Modelling in Heterogeneous Catalysis*. PhD thesis, Cardiff University, 2017.

[20] G. Casella and R. L. Berger, *Statistical inference*, vol. 2. Duxbury Pacific Grove, CA, 2002.

[21] C. De Boor and C. De Boor, *A practical guide to splines*, vol. 27. springer-verlag New York, 1978.

[22] E. Ko, J. Benziger, and R. Madix, "Reactions of methanol on w (100) and w (100)-(5× 1) c surfaces," *Journal of Catalysis*, vol. 62, no. 2, pp. 264–274, 1980.

[23] D. F. Andrews, "A robust method for multiple linear regression," *Technometrics*, vol. 16, no. 4, pp. 523–531, 1974.

[24] A. Agrawal, R. Verschueren, S. Diamond, and S. Boyd, "A rewriting system for convex optimization problems," *Journal of Control and Decision*, vol. 5, no. 1, pp. 42–60, 2018.

[25] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.

[26] G. M. Ljung, "On outlier detection in time series," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 55, no. 2, pp. 559–567, 1993.

[27] D. Constales, Z. Fang, M. R. Kunz, G. Yablonsky, and R. Fushimi, "Methods for determining the intrinsic kinetic characteristics of irreversible adsorption processes," *Chemical Engineering Science*, vol. 207, pp. 344–351, 2019.