



Cite this: *Phys. Chem. Chem. Phys.*,
2026, **28**, 2932

PyMESpec: a Python toolbox for automated modulation excitation spectroscopic data analysis and transient experiments

Alfred Worrad,^{ab} Quentin Kim,^{ab} Sagar Sourav,^{cd} Himaghna Bhattacharjee^{ab} and Dionisios G. Vlachos^{id} *^{abc}

Recent advances in laboratory automation have created a growing need for efficient and scalable tools to process experimental data without human intervention. Of great interest to the catalysis community is the analysis of time-resolved spectroscopic data, which can capture transients and identify highly active minority species. One such technique is modulation excitation spectroscopy (MES). Modulation produces large datasets unsuitable for manual processing. We introduce the python modulation excitation spectroscopy (PyMESpec) toolkit, an open-source library for analyzing MES experiments. PyMESpec offers a fast and flexible baseline correction, phase-sensitive detection (PSD), chemometric deconvolution, and automated reaction rate extraction. PyMESpec can be used through a command-line interface with configuration files or *via* a graphical user interface (GUI). This facilitates high-throughput, reproducible processing of spectroscopic datasets in automated and adaptive experimentation. PyMESpec applies to large datasets from spectroscopies and transient experiments in general. We demonstrate it using spectra from diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) over CeO₂ and modulation excitation ultraviolet-visible (ME UV-vis) and near-ambient pressure x-ray photoemission spectroscopy (NAP-XPS) on vanadia/titania catalyst for oxidative propane dehydrogenation.

Received 10th November 2025,
Accepted 7th January 2026

DOI: 10.1039/d5cp04322k

rsc.li/pccp

1. Introduction

Automated laboratory experiments, often referred to as the self-driving lab, are increasingly sought after as they enable efficient use of equipment and higher data throughput handling compared to individual researchers manually handling data and tasks. They improve productivity, accelerate discovery, and reduce human error in routine tasks.^{1–4} Automation also brings significant challenges, including the planning of sequential experiments to explore the parameter space efficiently, data processing, and data interpretation.^{3,5} Software toolkits, such as NEX Torch,⁶ have been developed to design experiments and guide catalyst optimization using Bayesian optimization. While human in the loop is essential for high-level decision-making, creating data analysis pipelines for self-driving labs is a necessity, but often complex, potentially non-uniform, and situationally dependent.

One challenge in heterogeneous catalysis research is that the reactive sites often exist in small populations that are hard to distinguish from less catalytically active, abundant species.^{7,8} Advanced spectroscopic techniques and data analysis tools can discern subtle signals, and with transient perturbations, identify catalytically active sites and elucidate reaction mechanisms *in situ* or under *operando* conditions.

Modulation excitation spectroscopy (MES) is a powerful and sensitive technique that enhances the signal-to-noise ratio, enabling kinetic and mechanistic analysis.^{9–11} Through periodic modulations, it allows phase-sensitive detection to isolate dynamic spectral responses from static background signals and noise. Temperature, pressure, pH, isotopic labels, electrical potential, and light flux, are often modulated⁸ and data is collected using a host of spectroscopic techniques, such as ATR,^{7,12} DRIFTS,^{13–16} Raman,¹⁷ XRD,¹⁸ and UV-vis.^{15,17} Time-resolved techniques,⁹ such as temporal analysis of products (TAP)¹⁹ and steady-state isotopic transient kinetic analysis (SSITKA),²⁰ also leverage controlled transients to extract mechanistic insights by generating large datasets.

These techniques generate large datasets of hundreds to thousands of spectra. Processing such datasets manually is time-consuming and prone to subjective bias and inconsistent interpretation. Automated data analysis software, capable of

^a Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE, 19716, USA. E-mail: vlachos@udel.edu

^b Center for Catalytic Science and Technology, University of Delaware, Newark, DE, 19716, USA

^c Delaware Energy Institute, University of Delaware, 221 Academy St., Newark, DE, 19716, USA

^d Department of Chemical Engineering, Indian Institute of Technology Madras, Chennai, Tamil Nadu, 600036, India



processing large datasets, could overcome this challenge and provide real-time feedback for adaptive experimentation and expedited exploration. Orange-spectroscopy, an extension of the orange data mining suite,²¹ can perform preprocessing and visualization of spectroscopic datasets but is not tailored for phase-sensitive detection and kinetic rate extraction in MES experiments.

In this work, we introduce the python modulation excitation spectroscopy (PyMESpec), a python-based package for automating data analysis of MES and transient spectroscopic experiments. PyMESpec facilitates the processing of large-scale spectroscopic datasets by systematic baseline correction, noise filtering, phase-sensitive analysis, chemometric deconvolution, and kinetic rate extraction with minimal user oversight. It can increase throughput and reproducibility, enabling autonomous transient experiments in self-driving laboratories. We demonstrate its use with experimental data of CeO₂ diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) and modulation excitation ultraviolet-visible (ME UV-vis) and near-ambient pressure X-ray photoemission spectroscopy (NAP-XPS) on vanadia/titania catalyst for oxidative propane dehydrogenation.

2. Methods

2.1. Case study tutorials

To demonstrate the capabilities of the PyMESpec suite, we present two experimental datasets involving CeO₂ diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) and modulation excitation ultraviolet-visible (ME UV-vis) and near-ambient pressure X-ray photoemission spectroscopy (NAP-XPS) on vanadia/titania catalyst for oxidative propane dehydrogenation. These examples demonstrate the toolbox's ability to process large, time-resolved spectroscopic data to extract chemical insights *via* automated workflows. A flow diagram showcasing a standard subset of possible workflows is shown in Fig. S1. All data processing performed within this work can be found as part of a series of Jupyter notebook tutorials on the Github page <https://github.com/worradal/PyMESpec> under the src/example_files/jupyter_notebook_tutorials/folder. PyMESpec natively supports CSV files across all package operations; however, custom data processing classes can be implemented to handle any file type (if converting to CSV is undesirable). Example screenshots showcasing the GUI layout can be found in the SI, Fig. S3 and S4.

2.1.1. 4V/TiO₂ catalyst and ME-UV-vis data. Supported 4V/TiO₂ catalyst was synthesized in an Ar atmosphere in a glove box using incipient wetness impregnation. Specifically, a calculated amount of Vanadium isopropoxide was added to a calculated amount of anhydrous methanol and added dropwise to the TiO₂ support. The sample was allowed to dry overnight inside the glove box at room temperature. Then, it was quickly transferred to a muffle furnace for further drying at 393 K for 2 h and calcination at 723 K for 4 h in static air.

ME-UV-vis spectra were collected for the propane (C₃) oxidative dehydrogenation (ODH) reaction over this catalyst. An

in-house-built low-void volume environmental cell,²² an Avantes ULS2048CL-EVO UV-vis spectrometer equipped with an AvaLight-XE light source, and an adjustable UV/VIS/NIR high-temperature probe were used for data collection. BaSO₄ was used as a reference for the background spectrum of the sample absorbance spectra. The 4V/TiO₂ sample was calcined in 10% O₂/Ar for 30 min at 400 °C. Following this, the sample was conditioned in the C₃ ODH reaction mixture (C₃:O₂:Ar, 1:1:8) for 1 h to achieve a stable structure for the reaction. Following this, O₂ was switched periodically with He with a full cycle period of 60 s, where O₂ was admitted into the environmental cell for 30 s and replaced by He for 30 s. The data was collected at 100 ms per spectrum with averaging over 10 spectra (1 s per spectrum), leading to 60 spectra per cycle. Spectra were collected for a total of 10 periods.

2.1.2. 4V/TiO₂ UV-vis and NAP-XPS. The 4V/TiO₂ sample was calcined in 10% O₂/Ar for 30 min at 400 °C, then cooled to 130 °C before purging with pure Ar. The sample was reduced at 130 °C by introducing Ar saturated with ethanol in a bubbler. The ethanol bubbler was kept at 20 °C for 60 mbar partial pressure of ethanol in the Ar flow. The sample was exposed to ethanol/Ar flow for 30 min while collecting UV-vis spectra.

Near ambient pressure X-ray photoelectron spectroscopy (NAP-XPS) spectra were collected at Lehigh University using a SPECS DeviSim NAP reactor cell coupled with a PHOIBOS 150 NAP electron energy analyzer equipped with an Al K α source. 4V/TiO₂ was heated in 1 mbar of O₂ to 400 °C before cooling down to 130 °C and pulling vacuum. 1 mbar of ethanol was then introduced to the chamber for 140 min. Spectra of the O 1s and V 2p XPS region (540–505 eV) were taken every 14 minutes during the reduction by ethanol. The NAP-XPS data was analyzed using CasaXPS²³ and used to compute pure component UV-vis spectra for V⁵⁺, V⁴⁺ and V³⁺ species, for chemometric deconvolution. A tutorial for obtaining the pure component UV-vis spectra from known fractions in XPS can be found in the Jupyter notebook tutorial “computing_pure_components_tutorial.ipynb.” Additional details on the matching of XPS and UV-vis can be found in Sections S1, S2, and in Fig. S2 of the SI.

2.1.3. CeO₂ DRIFTS. DRIFTS spectra were acquired in a Praying Mantis high-temperature reaction chamber (Harrick) using a Nicolet 6700 spectrometer (Thermo Fisher) equipped with a DTGS detector. CeO₂ nanopowder (US Nano, 10 nm) was loaded into the cell, then calcined in 10% O₂/Ar for 30 min at 450 °C, and then held in vacuum for 30 min at 450 °C before cooling to 250 °C. While remaining under vacuum, the sample was cooled and heated between 250 °C and 25 °C for a total of four cycles. Spectra were acquired every 25 °C during both cooling and heating cycles. All heating and cooling rates were 10 °C min⁻¹.

2.2. MES data structure

We first outline the structure of the data. An experimental variable, such as a reactant concentration, is periodically modulated, commonly *via* a step waveform, and the system's response is monitored using a spectroscopic technique sensitive to chemical or structural changes. The time-resolved



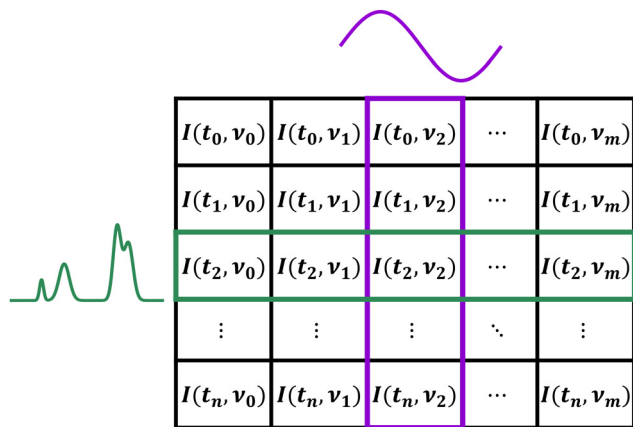


Fig. 1 Data structure $I(t, \nu)$ of MES experiments, with each row representing a spectrum collected at a point in time, and each column the temporal progression for a select spectral coordinate value.

spectroscopic data is a signal $I(t, \nu)$, where t is the time of collection and ν is the spectral coordinate within each spectrum (e.g., wavenumber, frequency, or energy). The $I(t_i, \nu_j)$ data forms a matrix where each row corresponds to a spectrum collected at time t_i , and each column corresponds to the spectral coordinate ν_j , with each element containing the signal $I(t_i, \nu_j)$; a pictorial representation is shown in Fig. 1. This data structure allows automated analysis in PyMESpec.

3. Results and discussion

3.1. Baseline correction

Nearly all spectroscopic techniques suffer from background fluctuations due to environmental, instrumental, or sample effects.²⁴ Our library includes a comprehensive baseline correction module, allowing the use of standard correction methods in the form of linear or polynomial baseline subtraction for simple baseline drifts. We also implement asymmetric reweighted penalized least squares (arPLS)²⁵ to effectively preserve true spectral features, even for more complex baseline corrections. Correction methods are tunable through PyMESpec package scripting, with simplified versions available in GUI. The baseline correction module is modular and

extensible. Customized correction methods can be implemented by defining classes that specify how the correction is applied to a 1D spectrum.

Baseline correction schemes can be applied to individual spectra, full $I(t, \nu)$ matrices, or select ranges of the spectral coordinate. We showcase the arPLS method, which corrects for baseline drift in a set of CeO_2 FTIR spectra at different temperatures in Fig. 2A. In the original FTIR data, a temperature-dependent baseline drift is present that we wish to remove. We highlight the extracted baseline in relation to the original and baseline corrected version of a single spectrum in Fig. 2B for clear emphasis on the change in the spectrum. Upon applying arPLS baseline correction, the change in the hydroxyl region ($\sim 3000\text{--}3800\text{ cm}^{-1}$) is more clearly visible and is less subject to thermal noise. For convenience and reproducibility, the exact code used for this baseline correction can be found in the example tutorial Jupyter notebook “ $\text{CeO}_2\text{_DRIFTS_tutorial.ipynb}$ ” on the GitHub page, which walks the user through the impact that adjusting parameters has on the shape of the baseline. For an example tutorial on how to define a custom baseline correction class and implemented it within a workflow, we would like to direct the reader to the tutorial located in “ $\text{CeO}_2\text{_DRIFTS_custom_baseline_tutorial.ipynb}$.”

3.2. MES phase-sensitive detection (PSD)

The first major MES-specific feature of the PyMESpec package is PSD, which isolates signals that oscillate at a chosen modulation frequency at different phase offsets from the experimental modulation. PSD can be performed either on an average modulation cycle or the full experimental time series, depending on the signal quality and analysis objectives. The average-cycle approach, which is a common practice but not a strict requirement, involves averaging individual modulation periods to form a representative cycle. The conventional formulation of PSD⁷ is provided in eqn (1).

$$I_k(\phi_k^{\text{PSD}}, \nu) = \frac{2}{T} \int_0^T I(t, \nu) \sin(k\omega_s t + \phi_k^{\text{PSD}}) dt \quad (1)$$

Here ϕ_k^{PSD} is the demodulation phase angle, T is the period length, ω_s is the stimulation angular frequency, and k is the demodulation index (most commonly taken to be 1). This

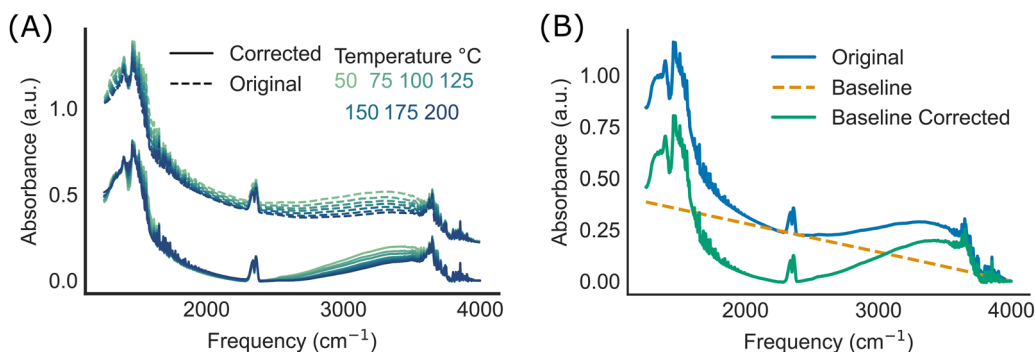


Fig. 2 (A) Example usage of arPLS baseline correction for baseline-drifted FTIR spectra for CeO_2 collected from 50–200 °C after calcination. The original spectra are all uniformly shifted for visualization purposes. (B) Before and after, including the baseline for a sample spectrum at 75 °C.



approach typically involves an iterative approach to find the phase ϕ_k^{PSD} angles, by varying ϕ_k^{PSD} from 0 to 360 degrees. An efficient alternative is to use the Fourier transform,¹⁰ which for computer codes is implemented as the fast Fourier transform (FFT) in numerical libraries, such as NumPy.²⁶ The FFT approach directly accesses the amplitude and phase information for all harmonics in a non-iterative way. The FFT approach is also advantageous since it does not require quasi-steady. The FFT computes the Fourier transform of the time-resolved signal as shown in eqn (2), yielding a frequency-domain representation suitable for phase-resolved analysis.

$$I(\omega_{\text{fft}}, \nu) = \int_{-\infty}^{\infty} I(t, \nu) e^{-i\omega_{\text{fft}} t} dt \quad (2)$$

Here, ω_{fft} is the angular frequency in the Fourier domain (not to be confused with the spectral coordinate ν). For a comparable value to ϕ_k^{PSD} for the phase offset, we apply eqn (3) to obtain $\phi_{k,\text{fft}}^{\text{PSD}}$.

$$\phi_{k,\text{fft}}^{\text{PSD}} = \arctan \frac{\text{Im} I(\omega_{k,\text{fft}}, \nu)}{\text{Real} I(\omega_{k,\text{fft}}, \nu)} \quad (3)$$

Since the target frequencies are typically low but nonzero, we can remove noise and filter out spectator species. The spectator species have a constant signal in the original spectra, leading to their entire contribution only at the 0 frequency, while noise only populates high frequencies. To obtain the time-resolved phase domain spectra, the inverse Fourier transform in eqn (4) is applied to the signal after first filtering all undesired contributions by setting $I(\omega_{\text{fft}} \neq \omega_{k,\text{fft}}, \nu) = 0$.

$$I(t, \nu) = \frac{1}{2\pi} \int_{-\infty}^{\infty} I(\omega_{\text{fft}}, \nu) e^{i\omega_{\text{fft}} t} d\omega_{\text{fft}} \quad (4)$$

Since PSD commonly uses $k = 1$ for the demodulation index (the fundamental frequency), we employ this for our vanadia on titania case study tutorial. The supported 4V/TiO₂ MES experiment conducted at 400 °C switching on and off O₂, while keeping propane flow constant, is shown in Fig. 3. For a direct comparison between the time index (Fig. 3A) and the phase angle (Fig. 3B) in this example, use $\frac{\pi * i}{30}$, where i is the time index.

The static sections of the spectra are removed, isolating only the contributions from active species. The region above 800 nm in the original signal, dominated by noise, is also smoothed considerably. The phase-sensitive plot indicates that regions around 400 nm and 600 nm appear ~180 degrees out of phase with each other. This is consistent with the regions ~425 nm and ~550–650 nm being associated with the oxidized (V⁵⁺) and reduced (V⁴⁺ and V³⁺) species, respectively. This out-of-phase behavior was severely obscured in the original spectra, whereas upon PSD, the reduction and reoxidation cycles between the oxidized and reduced species are clearly seen. For additional considerations on the kinetic analysis of MES-PSD, we refer the readers to the comprehensive works by Bravo and co-workers.^{14,27} A tutorial Jupyter notebook is available for the analysis of V/TiO₂, with the PSD section in the first part of the “V_TiO2_tutorial.ipynb.” We note that, for other non-MES applications, general noise filtering methods utilizing Savitzky-Golay²⁸ and low-pass filtering are also implemented.

3.3. Chemometric analysis

Another prominent feature of our software package is the chemometric analysis module. Under the assumption that linear mixing applies, we can quantify the relative contributions of chemical species or catalytic sites. This is achieved using a library of reference spectra for each pure component. At each time step, the measured spectrum is deconvoluted using non-negative linear least squares (NNLS)²⁹ to ensure physically meaningful (*i.e.*, non-negative) contributions from each reference. The deconvolution solves the optimization problem given in eqn (5):

$$\min_{\beta \geq 0} \|y - X\beta\|_2^2 \quad (5)$$

Here, y is the observed spectrum, X is the matrix of reference spectra, and β is the fitted species concentrations or weights. To assess the uncertainty of the extracted species fractions, we compute confidence intervals (CI) using eqn (6).

$$\beta_i \pm z_{\alpha/2} \cdot \sqrt{\hat{\sigma}^2 [(X^T X)^{-1}]_{ii}} \quad (6)$$

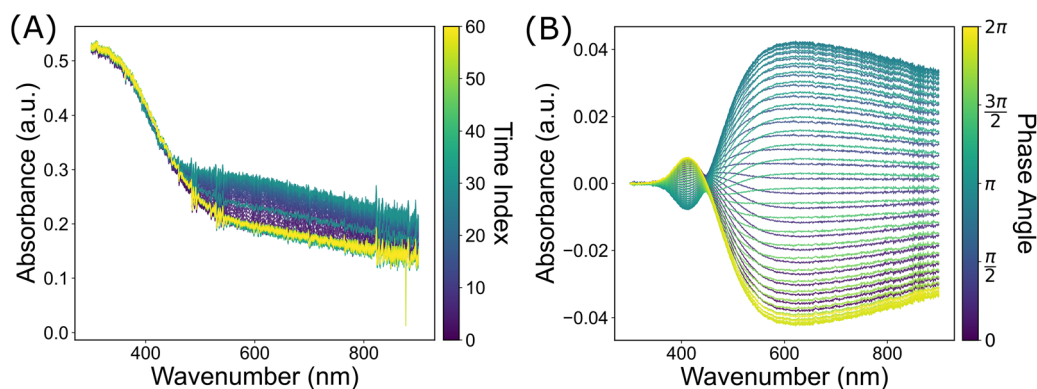


Fig. 3 (A) Original UV-vis spectra of 4V/TiO₂ while switching on and off O₂, at constant propane flow. (B) Phase-sensitive detection using the fundamental frequency. To map from the time index, i , to the phase angle, use $\frac{\pi * i}{30}$.



Here α defines the confidence interval, $z_{\alpha/2}$ is the corresponding critical value from the normal distribution, $\hat{\sigma}^2$ is the estimated error of the residuals, and the ii refers to the i -th diagonal element of the covariance matrix $(X^T X)^{-1}$. To enforce the non-negativity constraint on the resulting confidence intervals, we bound the lower limit at zero. Additionally, when interpreting results in terms of fractional species contributions, we cap the upper bound at 1 to reflect a closed composition space. This chemometric approach is applied to the same 4V/TiO₂ system used in the PSD analysis. The results are illustrated in Fig. 4.

Through chemometrics analysis, we obtain fractions for vanadium in each oxidation state, allowing us to track the working state of the catalyst over time. In this example, the fraction of V⁵⁺ decreases as it reduces to V⁴⁺ and V³⁺ species during propane flow in the absence of O₂. When O₂ is switched on, the reduced species quickly reoxidize to V⁵⁺. The system does not completely reoxidize to 100% V⁵⁺, as there exists a baseline of reduced vanadium species under O₂ and propane flow. In previous computational work, we provided evidence that a small set of highly active V⁵⁺ isolated species may form³⁰ that reduce faster than most other V⁵⁺ species, which may explain why the working state of the catalyst is not being fully oxidized. The chemometrics analysis is available in the tutorial notebook “V_TiO₂_tutorial.ipynb.”

3.4. Automated rate analysis

Another key feature of our software package is the automated extraction of kinetic rates. Rate analysis can be performed either on a cycle-averaged pulse, which enhances the signal-to-noise ratio, or on individual pulses, enabling the user to track trends over multiple pulses, such as in catalyst deactivation. Numerous commercial instrument packages include basic rate estimation tools; they are typically designed for small datasets, require user interaction to select data points, or are not easily incorporated into modular workflows. Our implementation is focused on large datasets requiring minimal

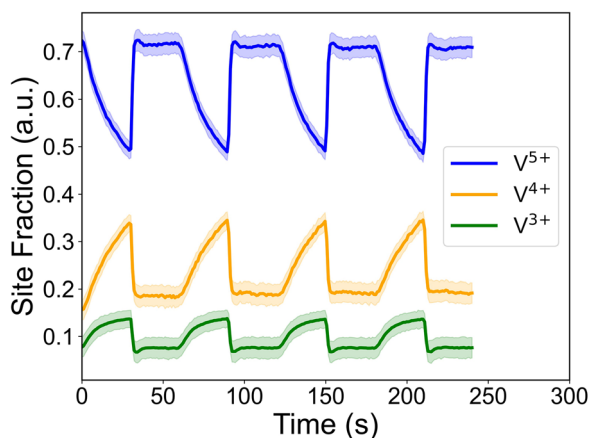


Fig. 4 Chemometric analysis to determine the fractional coverage of three different oxidation states of vanadium, with the shaded regions indicating the 95% confidence intervals, based on NNLS fitting of reference spectra obtained from UV-vis and NAP-XPS experiments. Ascending V⁵⁺ sections occur when O₂ flow is on, while descending when O₂ flow is off.

human intervention. Additionally, our package can compute rates directly from the raw spectra or from the species concentrations obtained from the chemometric analysis shown in the previous section.

Regardless of the type of data selected, for rate extraction on a dataset containing multiple pulses, it will first identify the maxima and minima of the $I(t, \nu)$ signal (or concentration) using SciPy's find_peaks method.³¹ The maxima and minima are then used as the start and end points for each kinetic window. To extract rates from each kinetic window, a kinetic equation chosen by the user is fit to the data, with an exponential fit being the default. Users can implement their own fitting function beyond the exponential function, most commonly used for adsorption, desorption, or redox processes. Optionally, for increased accuracy, the initial peak detection can be followed by adaptive interval refinement, in which the fitting boundaries are adjusted iteratively to minimize the root mean square error (RMSE) of the fitted model.

While the PyMESpec comes with a built-in peak detection scheme, the framework is modular and can accommodate custom peak detection algorithms, including machine learning approaches such as convolutional neural networks (CNNs), for handling complex or low-signal-to-noise-ratio systems. We apply the automated rate calculation to the 4V/TiO₂ system to extract the redox rates for the V⁵⁺ species in Fig. 5.

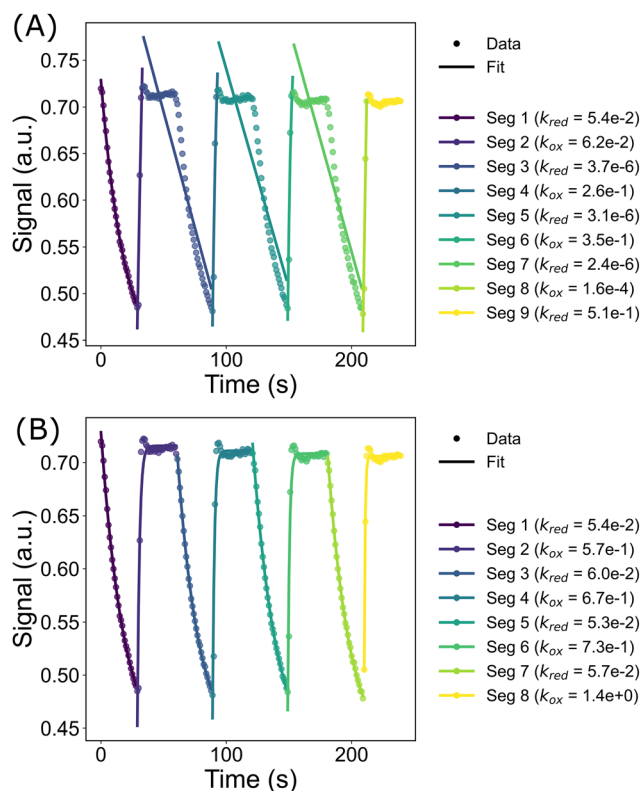


Fig. 5 Automated peak detection and rate calculation for each pulse in the MES experiment based on the V⁵⁺ species as calculated from the populations computed in the chemometrics section (Fig. 4). (A) Fit pulsed rates from absolute maxima and minima before adaptive peak finding, and (B) sections and calculated rates from adaptive peak finding.



This automated rate detection scheme can compute the reduction and reoxidation rates based on the chemometric deconvolutions for the 4V/TiO₂ system. Our adaptive fitting scheme drastically improves the rate-fitting and detection of regions with a change in conditions. The reoxidation process is faster than the reduction process by approximately an order of magnitude. This information is unattainable in steady state experiments, further illustrating the importance of transient experiments. While assigning the start and end points directly to the times when a modulation occurs may seem trivial, this is not feasible in sequential reactions (A → B → C) when the delay of intermediate species is not explicitly known. This analysis is available in the tutorial notebook “V_TiO2_tutorial.ipynb.”

4. Conclusions

High-throughput data analysis workflows are required for spectral analysis of highly complex transient spectral techniques. We have outlined PyMESpec, a flexible data analysis software package and a pipeline that requires minimal human oversight. In its most complete form, it takes raw transient spectra, employs sophisticated baseline correction algorithms to get comparable spectra, which, using phase-sensitive detection, can filter for signals of interest. Additionally, deconvolutions are applied to obtain active species balances based on reference spectra enabling the calculations of reaction rates. These individual components can be flexibly utilized for individual spectrum purposes as well.

Author contributions

A. W. wrote the PyMESpec software. Q. K. and S. S. conducted experiments. H. B. helped with initial code development. D. G. V. provided supervision. A. W. wrote the manuscript with input from S. S. and D. G. V. All authors contributed to the analysis and discussion of the results.

Conflicts of interest

There are no conflicts of interest to declare.

Data availability

PyMESpec is freely available on its GitHub project page (<https://github.com/worradal/PyMESpec>) along with detailed user documentation, Jupyter notebook tutorials for the python package, and example configuration files for the GUI and command line implementation. Details can be found on the GitHub page.

Supplementary information is available. See DOI: <https://doi.org/10.1039/d5cp04322k>.

Acknowledgements

This work was supported by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, under

Award Number DE-SC0024085. The authors would like to thank Dr Ryan Thorpe for his assistance with XPS experiments performed at the Institute for Functional Materials and Devices at Lehigh University and for his help with the analysis. We acknowledge the use of Visual Studio Code for the expedited writing of unit tests and documentation.

References

- 1 C. Scheurer and K. Reuter, *Nat. Catal.*, 2025, **8**, 13–19.
- 2 Y. Xie, K. Sattari, C. Zhang and J. Lin, *Prog. Mater. Sci.*, 2023, **132**, 101043.
- 3 M. Seifrid, R. Pollice, A. Aguilar-Granda, Z. Morgan Chan, K. Hotta, C. T. Ser, J. Vestfrid, T. C. Wu and A. Aspuru-Guzik, *Acc. Chem. Res.*, 2022, **55**, 2454–2466.
- 4 Y. Shi, P. L. Prieto, T. Zepel, S. Grunert and J. E. Hein, *Acc. Chem. Res.*, 2021, **54**, 546–555.
- 5 A. Trunschke, *Catal. Sci. Technol.*, 2022, **12**, 3650–3669.
- 6 Y. Wang, T.-Y. Chen and D. G. Vlachos, *J. Chem. Inf. Model.*, 2021, **61**, 5312–5319.
- 7 A. Urakawa, T. Bürgi and A. Baiker, *Chem. Eng. Sci.*, 2008, **63**, 4902–4909.
- 8 C. Hess, *ACS Catal.*, 2025, **15**, 8005–8013.
- 9 P. Müller and I. Hermans, *Ind. Eng. Chem. Res.*, 2017, **56**, 1123–1136.
- 10 D. Baurecht and U. P. Fringeli, *Rev. Sci. Instrum.*, 2001, **72**, 3782–3792.
- 11 M. Chen, N. Maeda, A. Baiker and J. Huang, *ACS Catal.*, 2012, **2**, 2007–2013.
- 12 A. Urakawa, R. Wirz, T. Bürgi and A. Baiker, *J. Phys. Chem. B*, 2003, **107**, 13061–13068.
- 13 J. Weyel, L. Schumacher, M. Ziemba, M. Pfeiffer and C. Hess, *Acc. Chem. Res.*, 2024, **57**, 2643–2652.
- 14 P. D. Srinivasan, B. S. Patil, H. Zhu and J. J. Bravo-Suárez, *React. Chem. Eng.*, 2019, **4**, 862–883.
- 15 K. Yu, M. Scarpelli, S. Sourav, A. Worrada, J. A. Boscoboinik, L. Ma, S. N. Ehrlich, N. Marinkovic, W. Zheng and D. G. Vlachos, *Catal. Sci. Technol.*, 2024, **14**, 4948–4957.
- 16 I. C. t Have, J. J. G. Kromwijk, M. Monai, D. Ferri, E. B. Sterk, F. Meirer and B. M. Weckhuysen, *Nat. Commun.*, 2022, **13**, 324.
- 17 H. Hoyer and C. Hess, *J. Phys. Chem. C*, 2024, **128**, 21377–21386.
- 18 D. Ferri, M. A. Newton, M. Di Michiel, G. L. Chiarello, S. Yoon, Y. Lu and J. Andrieux, *Angew. Chem., Int. Ed.*, 2014, **53**, 8890–8894.
- 19 K. Morgan, N. Maguire, R. Fushimi, J. T. Gleaves, A. Goguet, M. P. Harold, E. V. Kondratenko, U. Menon, Y. Schuurman and G. S. Yablonsky, *Catal. Sci. Technol.*, 2017, **7**, 2416–2439.
- 20 C. Ledesma, J. Yang, D. Chen and A. Holmen, *ACS Catal.*, 2014, **4**, 4527–4547.
- 21 J. Demšar, T. Curk, A. Erjavec, C. Gorup, T. Hočvar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, M. Štajdohar, L. Umek, L. Žagar, J. Žbontar, M. Žitnik and B. Zupan, *J. Mach. Learn. Res.*, 2013, **14**, 2349–2353.



- 22 B. S. Patil, P. D. Srinivasan, E. Atchison, H. Zhu and J. J. Bravo-Suárez, *React. Chem. Eng.*, 2019, **4**, 667–678.
- 23 N. Fairley, V. Fernandez, M. Richard-Plouet, C. Guillot-Deudon, J. Walton, E. Smith, D. Flahaut, M. Greiner, M. Biesinger, S. Tougaard, D. Morgan and J. Baltrusaitis, *Appl. Surf. Sci. Adv.*, 2021, **5**, 100112.
- 24 C. R. Mittermayr, H. W. Tan and S. D. Brown, *Appl. Spectrosc.*, 2001, **55**, 827–833.
- 25 S.-J. Baek, A. Park, Y.-J. Ahn and J. Choo, *Analyst*, 2015, **140**, 250–257.
- 26 C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke and T. E. Oliphant, *Nature*, 2020, **585**, 357–362.
- 27 A. Torres-Velasco, B. S. Patil, H. Zhu, Y. Qi, S. G. Podkolzin and J. J. Bravo-Suárez, *J. Catal.*, 2024, **434**, 115508.
- 28 A. Savitzky and M. J. E. Golay, *Anal. Chem.*, 1964, **36**, 1627–1639.
- 29 C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Society for Industrial and Applied Mathematics, 1995.
- 30 A. Worrad, S. A. Khan, D. G. Vlachos and S. Caratzoulas, *J. Phys. Chem. C*, 2024, **128**, 18933–18941.
- 31 P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. P. Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G.-L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko, Y. Vázquez-Baeza and C. SciPy, *Nat. Methods*, 2020, **17**, 261–272.

