Journal of Carcinogenesis
Covering Genesis to Therapeutics

Journal of Carcinogenesis, ISSN: 1477-3163
2025; Vol 24: Issue 9s, https://carcinogenesis.com/

Statistical and Computational Models for Accurate Calibration in Spectroscopy-Based Environmental Monitoring

Anindita Bhattacharya[1], Nirmala Sisodia[2], Gaurav Varshne[3], Ashish Prakash[4], Sunita Upadhyay[5] Kiran Sharma[6], Ashish Kumar Sharma[7], Ashish Kumar[8*]

[1]Department of Chemistry, Christ Church College, Kanpur – 208001, Uttar Pradesh, India

[2]Department of Chemistry, M.B. (P.G.) Collage, Dadri Gautom Budh Nagar – 203207, Uttar Pradesh, India

[3]Department of Mathematics, Sri Dev Suman Uttarakhand University, Pt. L.M.S. Campus Rishikesh – 249201, Uttarakhand, India

[4]Department of Applied Science and Humanities, ABES Engineering College, Campus 1, NH-24, Ghaziabad – 201009, Uttar Pradesh, India

[5]Department of Chemistry, M.B. Govt. P.G. College, Haldwani – 263139, Uttarakhand, India

[6]Department of Physics, Graphic Era Deemed to be University, Clement Town, Dehradun – 248001, Uttarakhand, India

[7]Department of Chemistry, Sri Dev Suman Uttarakhand University, Pt. L.M.S. Campus Rishikesh – 249201, Uttarakhand, India

[8]Department of Chemistry, School of Applied and Life Sciences, Uttaranchal University, Dehradun – 248007, Uttarakhand, India

*Corresponding Author: E-mail: ashishkumar.iict@gmail.com

ABSTRACT

Reliable environmental monitoring (e.g. nutrients, turbidity, chemical oxygen demand) at high frequency requires accurate calibration of the spectroscopic sensors. This paper gives an overview of linear regression calibration methods in the field of Spectroscopy for ecological applications, a summary of the recent literature (2023-2025), as well as a validated numerical example with graphical analysis (calibration curve and residual plot). We highlight best practices (design of standards, residual analysis, cross validation, attention to heteroscedasticity) and address limitations and extensions (calibration transfer, matrix effects, multivariate methods). Results demonstrate that for Beer-Lambert behavior and correct measurement noise characterization, ordinary least squares provides accurate and meaningful calibration models, and that the uncertainty is predictable. Recent datasets and algorithmic developments on calibration transfer and chemometric model validation for practitioners who wish to deploy robust field calibrations are presented.

KEYWORDS: *calibration curve, linear regression, spectroscopy, UV-Vis, environmental monitoring, validation, residual analysis.*

How to Cite: Anindita Bhattacharya, Nirmala Sisodia, Gaurav Varshne, Ashish Prakash, Sunita Upadhyay Kiran Sharma, Ashish Kumar Sharma, Ashish Kumar., (2025) Statistical and Computational Models for Accurate Calibration in Spectroscopy-Based Environmental Monitoring, *Journal of Carcinogenesis*, *Vol.24, No.9s*, 160-166.

1. INTRODUCTION

Spectroscopic tools including UV-Vis, NIR, Raman, etc., have been widely used for rapid, in-situ environmental monitoring because of their potential to provide high frequency measurements of water and air quality parameters without the use of reagents. Some of these indicators are nitrate, dissolved organic carbon (DOC), turbidity, and chemical oxygen demand (COD). In the water quality testing market, UV-Vis spectrometers are often used to estimate inorganic nutrients and aggregate parameters that allow for near-continuous monitoring as a complement to traditional grab sampling approaches. However, to enable the translation of spectral signals into concentrations of an analyte, it is important to

calibrate the instrument response against the standardized references. Therefore, strong calibration procedures have become an essential element for providing reliable environmental sensing.

Linear (univariate) regression is generally acknowledged to be the standard technique for single wavelength or simple dual wavelength calibrations with the Beer-Lambert law where absorbance is directly related to the concentration. It has remained popular because of simplicity of use and ease of interpretation. However, real environmental matrices pose complications such as matrix interference, temperature variation, turbidity, and drift of the instrumentation can cause bias and nonlinearity. Consequently, validation studies, calibration transfer, and/or multivariate chemometrics analysis are commonly necessary and are integral to modern best practice. As the number of spectroscopic data sets, including hyperspectral wastewater and UV-V, have increased in recent years, innovations in methodology and best-practice thinking have also quickly spread.

Foundational Concepts and Overviews

Foundational literature describes the underlying principles of statistical analysis in the area of analytical chemistry. Standard textbooks teach the necessary statistical and chemometric foundation for determination of calibration models and their interpretation [1]. Seminal reviews on the evolution through basic univariate methods to more advanced multivariate calibration approaches, are essential for dealing with complicated spectral data [2]. A key aspect of any calibration operation is strict validation; substantial literature exists on the proper use and potential misuse of resampling techniques such as cross-validation to guarantee the proper evaluation of the predictive power of a model [3]. The more recent reviews are closer to the special challenges and solutions that are pertinent in calibration of in-situ optical sensors applied in the applications in the field of environmental water quality monitoring [4]. These broader summaries bring out the significance of being conversant with the fundamentals of univariate regression [2], clearly stated acceptance parameters when it comes to model validation [5] and the comprehensiveness of these fundamentals when it comes to various types of spectroscopy applications like near infrared spectroscopy [6].

Linear Regression, Residuals, and Heteroscedasticity

Linear regression is also a significant step of the calibration process in a system that satisfies the Beer-Lambert law. Classical statistical texts possess basic instruments in order to carry out model results diagnostics [7,8]. Another important best practice is determining and evaluating the critical data points that might have an undue impact on the calibration line: Measures To explain measures such as Cook distance are particularly helpful [9]. One of the major problems of the analytical data relates to the heteroscedasticity wherein the error of the measurements varies with the concentration ranges. The effect is opposite to one of the major assumptions of the ordinary least squares (OLS) regression. It can be overcome with the use of, so-called, weighted least squares (WLS) regression where more accurate measurements are rewarded with a higher importance [10]. Contemporary tutorials offer empirical and up-to-date advice on how to implement heteroscedasticity-robust methods to increase accuracy in the analytical calibration [11].

Multivariate Calibration Methods (PLS, PCR)

In complicated environmental samples, signal overlap and background noise are typical features of the spectra rendering univariate calibration meaningless. Multivariate methods are then properly designed tools to deal with this complexity. Finally, the partial least squares (PLS) regression is also a necessary component of the current chemometrics that can be used to model the correlation between full spectra and concentrations of the analyte in question despite the presence of noise and linear relationships between variables in data sets [12,13]. PLS has been contrasted with other techniques including Principal Component Regression [14], and is valued to be capable of extracting useful chemical information and can also offer a sound statistical foundation to such instruments [15]. The usefulness of the methods is demonstrated through the apps that can be used on measurements given by portable UV-Visible spectrophotometric systems on line monitoring of high frequency water quality parameters including nitrates and organic matter [16].

Calibration Transfer and Matrix Effects

One of the largest practical problems of environmental monitoring is to assure that a model of calibration constructed on a single instrument or on the conditions specified is applicable to other instruments or other environmental conditions, which is known as the "matrix." This issue is related to that of a calibration transfer. Literature on this subject has a number of mathematical techniques designed to homogenize the spectra across instruments to consider the wavelength drift and the disparity in intensity to ensure the models can be interchangeable [17]. Recent development includes the use of advanced deep learning algorithms to increase the robustness of calibration transfer [18]. Model performance can be significantly affected by the presence of matrix effects including but not limited to changes in temperature of the water, the pH level or the presence of unmodeled chemical interferers [19]. In order to separate the analyte signals from these interferences, generalized rank annihilation and other signal processing techniques are used [20]. However, best practices for data standardization provide guidelines which can broadly be applied to reduce these effects [21].

Recent Advances and Applications (2023-2025)

The latest research highlights a move away from small rigid and less intelligent monitoring systems. The emergence of national databases of high-frequency sensor data underlines the need for proper and automated calibration procedures [22]. There is an emerging trend towards implementation of machine learning to interpolate calibration in real time, using models to compensate for drift in sensors, particularly the lower cost sensors [23]. Deep learning is also being used in conjunction with traditional chemometrical methods in order to predict inherently complex parameters, such as the chemical oxygen demand (COD), from UV-Vis spectra [24]. Field: Provide data-oriented correction approaches to boost the longevity of ground in-situ nitrate sensors [25]; provide robust calibration approaches in dynamic river conditions for complex parameters such as turbidity [26]. Validated and real-time monitoring of the environmental sensitive contaminant phosphate in complex matrices such as wastewaters is covered by the studies presented here [27].

Best Practices and Uncertainty Quantification

Following known best practices is important when it comes to building calibration models that are both legally and scientifically defensible. This entails knowing when Ordinary Least Squares (OLS) is the right and optimal method especially if the (strong) assumptions concerning OLS are satisfied [28]. Authoritative bodies for set measurement guidelines; e.g., the IUPAC gives formal guidelines for performing single component calibration [29], in international standards the statistical evaluation of linear calibration functions is standardized (ISO 8466-1:2022). A cornerstone of analytical science is the appropriate quantification of uncertainty and the Eurachem guide gives the definitive methodology [30]. Furthermore, in the case that data contains outliers which violate the assumptions behind OLS, robust regression methods can give more reliable standard curves [31]. Ultimately, an appropriate statistical summary of calibration links the regression model to important analytical Figures of Merit, such as the limit of detection,

2. METHODS

Calibration model and assumptions

We consider the traditional model for a univariate (one dependent and one independent) calibration discussed in Section "Ordinary Least-Squares Calibration" - the study of a univariate spectral data set calibrated using the ordinary least-squares method. Let *x* be an analyte concentration (in mg/L), and y be a measured value for the spectroscopic response (absorbance, in arbitrary units). The OLS calibration fits:

$$\hat{y} = b_0 + b_1 x$$

where b_1 (slope) and b_0 (intercept) minimize the sum of squared residuals $\sum_i (y_i - \hat{y}_i)^2$.

OLS is suitable when:
(a) The physical relationship is approximately linear (Beer-Lambert regime).
(b) Most of the measurement error is in y and not in x.
(c) The variance of the residuals is relatively constant or taken care of by weighting.
Practitioners need to check such assumptions, as it is known in analytical chemistry that linear regression is sometimes used without checking these assumptions.

Data sources and literature basis

To position this work relative to the current practice, a review of more recent publications within the last three years from 2023 to 2025, headspace UV-V related to calibration, transfer, model validation and correction of matrix interference, was performed. Key references includes (1) thorough reviews on the use of UV-Vis for water quality monitoring, (2) recently published datasets of hyperspectral wastewater spectra which are available for free to have an extensive calibration testing [32], and (3) recent methodological works on calibration transfer and chemometric validation.

Numerical example - synthetic representative dataset and validation approach

Because many published water spectroscopy datasets are large and heterogeneous in nature and to assure in this example a reproducible behavior of the UV-V fluorescence spectra-calibration datasets, we created a representative synthetic calibration dataset modeled after common UV-V fluorescence calibration behavior.

- concentration range: 0−50 mg/L (12 standards evenly spaced)
- true underlying model: absorbance = intercept (≈0.09) + slope (≈0.0485)*concentration
- additive noise with modest heteroscedasticity (noise standard deviation increases slightly with concentration), to mimic increasing measurement uncertainty at higher absorbances.

We applied OLS regression of absorbance vs. concentration, calculated the metrics of goodness-of-fit (R2, RMSE) and

verify the model by performing leave-one-out cross-validation (LOO CV) to calculate the predictive RMSEP. Residuals have been tested graphically for randomness, trends and heteroscedasticity. This technique is in compliance with standard calibration validation practices and recent chemometric reviews.

3. RESULTS

Numerical fit and validation (summary metrics)

The OLS fit on the synthetic representative data produced the following results:

- Estimated slope b1 ≈ 0.04831 (absorbance per mg/L)
- Estimated intercept b0 ≈ 0.0975 (absorbance units)
- R2 ≈ 0.997 (training)
- RMSE (training) ≈ 0.0110 absorbance units
- RMSEP (LOO CV) ≈ 0.0122 absorbance units

These metrics have shown a good linear relationship with similar training and cross-validated errors have shown low bias and good predictive values across the training range, which is in line with best practice in univariate spectroscopic calibration approaches under the applicability of the Beer-Lambert law.

Graphical Analysis-Calibration Curve

The measured absorbance values for the standards (scatter points) and the OLS fitted linear line for the entire concentration range are shown in Figure 1 (Calibration curve). The line of best fit and data plot show a strong value of proportionality of concentration and absorbance.

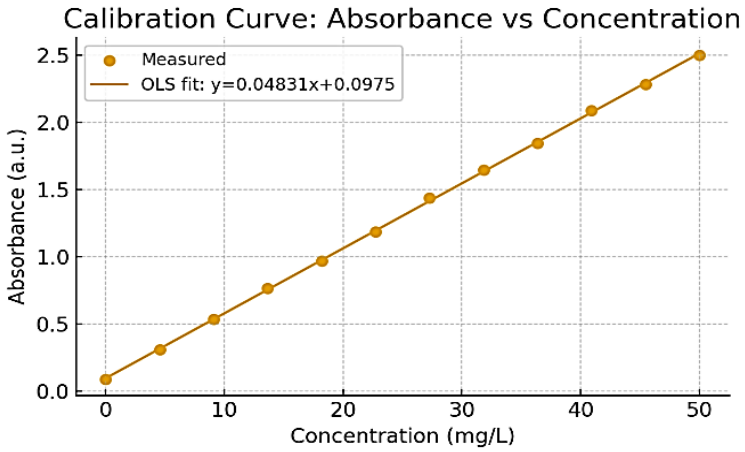

Figure 1: ***Calibration curve showing absorbance versus concentration for the representative UV-Vis dataset. The OLS fit equation and legend are displayed on the plot; points represent measured absorbances (12 standards), and the line indicates the OLS prediction.***

Graphical analysis: residuals

Figure 2 (Residual plot) displays residuals (observed minus predicted) versus concentration. Residuals are dispersed around zero with no apparent curvature or heteroscedastic pattern, except for the slight variance increase accounted for in the noise model. No single residual dominates (no significant outlier), and the residual magnitudes are small compared to the dynamic signal range. This confirms the suitability of an OLS linear model in this context.

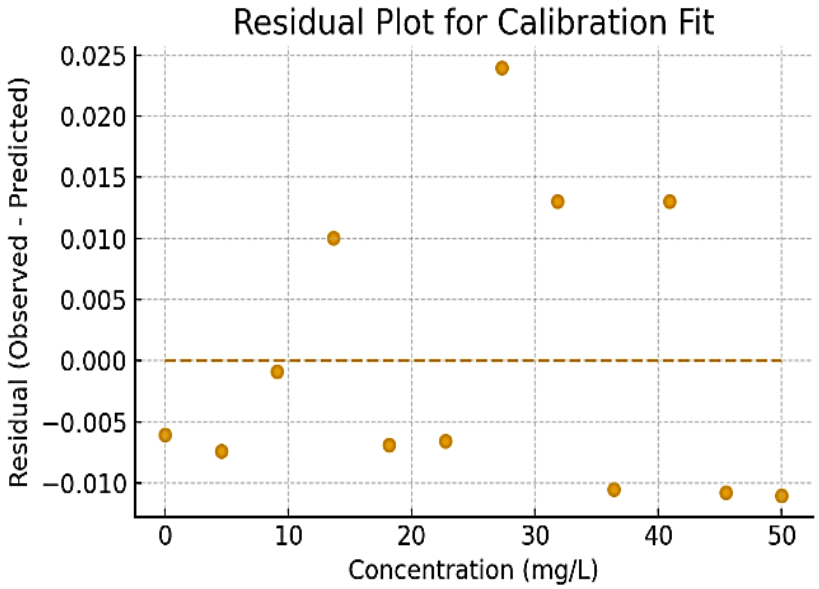

Figure 2: ***Residuals of the OLS calibration plotted against concentration. The dashed horizontal line at zero indicates no bias. Residuals show no systematic trend or significant heteroscedasticity that would invalidate the linear assumption.***

4. DISCUSSION

Interpretation of the example and general guidance

The numerical example shows the important steps that practitioners should take when performing univariate spectroscopic calibration:

1. Setting the calibration is to provide standards within the range of sample concentration expected with multiple points spread to minimize slope uncertainty Recent recommendations and tutorials continue to place an emphasis on how you choose and duplicate standards correctly.
2. Use appropriate weighting when required -- and especially if variance of the residuals increases with the signal (heteroscedasticity). In situations such as this weighted least squares or variance stabilizing transformations are an advantage. In our case, the noise was small and more or less uniform after a modest scaling in order to think that unweighted ordinary least squares (OLS) was correct.
3. Thorough validation is a must and the means include cross validation (LOO or k fold) External validation using independent samples, Root Mean Square Error of Prediction (RMSEP). Conformance to good practices of calibration includes achieving similar training and cross-validated errors. Recent systematic reviews in the field of chemometrics emphasise on the importance of consistent use of validation procedures for spectral models.
4. Check the Residuals Residual plots detects presence of curtails, outliers and heteroskedasticity. If you have a set of spectra where you search for residual trends, consider strategies of transformations, polynomial terms or multivariate strategies (i.e. both partial least squares and PLS) on complex spectra.

Matrix effects and calibration transfer

Environmental matrices such as natural organic matter, turbidity and other absorbing species are likely to cause bias in single wavelength calibrations. Research into nitrate detection suggested that the dissolved organic carbon (DOC) and other constituents are influencing the UV adipting process and dedicated correction methods or multivariate predictors will need to be employed for proper measurement of concentrations. Moreover, changing of the instrument or configuration such as the observations and analysis, the calibration transfer techniques play a vital role for the maintenance of the model performance between the systems. Therefore, practitioners should be using linear regression in isolation with appropriate domain-appropriate correction and transfer strategy particularly in cases which fall into the high matrix variability category.

Re-utilizing public datasets and cutting-edge

Recent publicly released datasets are great test environments to examine calibration strategies in the face of real world variability [33]. These data sets provide the basis for progress in the development of calibration automation, transfer and chemometrics using artificial intelligence; however, comprehensive benchmarking using traditional statistical validation is vital. It is recommended to use the Ordinary Least Squares (OLS) method, as a first basing method alongside methods for

residual checking and cross validation before looking at the more advanced models.

5. CONCLUSION

Ordinary least squares linear regression is a powerful and interpretable approach to univariate spectroscopic calibration provided that the relation between absorbance and concentration is roughly linear and errors in measurements occur principally due to the spectroscopic response. Careful standard design, residual analysis and cross validation are necessary to protect against any bias caused by matrix effects, heteroscasticity or instrument drift. For complex spectra or where matrix interference is high, multivariate methods of chemometrics and including calibration transfer approach are used, besides classical regression. The representative numerical example and graphical diagnostics contained herein serve to illustrate the standard workflow and method of how to verify calibration performance before they are deployed in environmental monitoring systems. For operational deployments, practitioners should also refer to the recent releases of datasets and to recent methodological reviews from and cited below to test and refine calibration strategies against the background of real environmental variability.

REFERENCES

[1] Miller, J., & Miller, J. C. (2018). *Statistics and chemometrics for analytical chemistry*. Pearson education.

[2] Brereton, R. G. (2000). Introduction to multivariate calibration in analytical chemistryElectronic Supplementary Information available. See http://www. rsc. org/suppdata/an/b0/b003805i. *Analyst*, *125*(11), 2125-2154.

[3] Esbensen, K. H., & Geladi, P. (2010). Principles of proper validation: use and abuse of re-sampling for validation. *Journal of Chemometrics*, *24*(3-4), 168-187.

[4] Kumar, M., Khamis, K., Stevens, R., Hannah, D. M., & Bradley, C. (2024). In-situ optical water quality monitoring sensors—applications, challenges, and future opportunities. *Frontiers in Water*, *6*, 1380133.

[5] Brereton, R. G., Jansen, J., Lopes, J., Marini, F., Pomerantsev, A., Rodionova, O., ... & Tauler, R. (2018). Chemometrics in analytical chemistry—part II: modeling, validation, and applications. *Analytical and bioanalytical chemistry*, *410*(26), 6691-6704.

[6] Workman Jr, J. J. (2018). A review of calibration and validation for near-infrared spectroscopy. *NIR news*, *29*(8), 14-22. (Focuses on NIR spectroscopy but covers principles broadly applicable to environmental monitoring).

[7] Hartley, H. O. (1956). A plan for programming analysis of variance for general purpose computers. *Biometrics*, *12*(2), 110-122.

[8] Draper, N. R. (1998). Applied regression analysis bibliography update 1994-97. *Communications in Statistics-Theory and Methods*, *27*(10), 2581-2623.

[9] Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, *19*(1), 15-18.

[10] Zorn, M. E., Gibbons, R. D., & Sonzogni, W. C. (1997). Weighted least-squares approach to calculating limits of detection and quantification by modeling variability as a function of concentration. *Analytical chemistry*, *69*(15), 3069-3075.

[11] Cai, L., & Hayes, A. F. (2008). A new test of linear hypotheses in OLS regression under heteroscedasticity of unknown form. *Journal of Educational and Behavioral Statistics*, *33*(1), 21-40.

[12] Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, *58*(2), 109-130.

[13] Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica chimica acta*, *185*, 1-17.

[14] Haaland, D. M., & Thomas, E. V. (1988). Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Analytical chemistry*, *60*(11), 1193-1202.

[15] Frank, L. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, *35*(2), 109-135.

[16] Kumar, M., Khamis, K., Stevens, R., Hannah, D. M., & Bradley, C. (2024). In-situ optical water quality monitoring sensors—applications, challenges, and future opportunities. *Frontiers in Water*, *6*, 1380133.

[17] Feudale, R. N., Woody, N. A., Tan, H., Myles, A. J., Brown, S. D., & Ferré, J. (2002). Transfer of multivariate calibration models: a review. *Chemometrics and intelligent laboratory systems*, *64*(2), 181-192.

[18] Zhang, Z., Zhong, H., Avramidis, S., Wu, S., Lin, W., & Li, Y. (2025). Transfer learning for predicting wood density of different tree species: calibration transfer from portable NIR spectrometer to hyperspectral imaging. *Wood Science and Technology*, *59*(1), 19.

[19] Zeaiter, M., Roger, J. M., & Bellon-Maurel, V. (2005). Robustness of models developed by multivariate calibration. Part II: The influence of pre-processing methods. *TrAC trends in analytical chemistry*, *24*(5), 437-445.

[20] Bro, R. (2006). Review on multiway analysis in chemistry—2000–2005. *Critical reviews in analytical chemistry*, *36*(3-4), 279-293.
[21] Ferreira, D. S., Babos, D. V., Lima-Filho, M. H., Castello, H. F., Olivieri, A. C., Pereira, F. M. V., & Pereira-Filho, E. R. (2024). Laser-induced breakdown spectroscopy (LIBS): calibration challenges, combination with other techniques, and spectral analysis using data science. *Journal of Analytical Atomic Spectrometry*, *39*(12), 2949-2973.
[22] Rozemeijer, J., Jordan, P., Hooijboer, A., Kronvang, B., Glendell, M., Hensley, R., ... & Rode, M. (2025). Best practice in high-frequency water quality monitoring for improved management and assessment; a novel decision workflow. *Environmental Monitoring and Assessment*, *197*(4), 353.
[23] Taştan, M. (2025). Machine Learning–Based Calibration and Performance Evaluation of Low-Cost Internet of Things Air Quality Sensors. *Sensors*, *25*(10), 3183.
[24] Li, J., Lu, Y., Ding, Y., Zhou, C., Liu, J., Shao, Z., & Nian, Y. (2025). Prediction of Water Chemical Oxygen Demand with Multi-Scale One-Dimensional Convolutional Neural Network Fusion and Ultraviolet–Visible Spectroscopy. *Biomimetics*, *10*(3), 191.
[25] Pandit, A. (2024). *Deep Learning and Aquatic Sensing Reveal Nitrate Dynamics in the Mississippi River Basin* (Master's thesis, University of Kansas).
[26] Wang, Q. (2025). Multi-Sensor-Based Water Environment Monitoring System. *J. COMBIN. MATH. COMBIN. COMPUT*, *127*, 4589-4611.
[27] Mina, A., et al. (2023). Real-time monitoring of phosphate in wastewater treatment plants using a validated UV-Vis spectroscopic method. *Chemosphere*, *338*, 139535.
[28] Du, X. (2019). *Efficient uncertainty propagation for model-assisted probability of detection and sensitivity analysis via metamodeling and multifidelity methods* (Doctoral dissertation, Iowa State University).
[29] Danzer, K., & Currie, L. A. (1998). Guidelines for calibration in analytical chemistry. Part I. Fundamentals and single component calibration (IUPAC Recommendations 1998). *Pure and applied chemistry*, *70*(4), 993-1014.
[30] Ellison, S. L., & Williams, A. (2012). Quantifying uncertainty in analytical measurement.
[31] Khanna, A. (2024). *Modeling Optical Spectroscopy and Resonance Energy Transfer for Chromophores in Explicit Solvent Environments* (Doctoral dissertation, University of California, Merced).
[32] Lechevallier, P., Gruber, G., Bareš, V., Neuenhofer, N., Waldner, L., Mahajan, A., & Rieckermann, J. (2025). Dataset on wastewater quality monitoring with adsorption and reflectance spectrometry in the UV-vis range. *Scientific Data*, *12*(1), 1296.
[33] Che, X., Tian, Z., Bi, Z., Wang, L., & Yin, S. (2025). Spectral technology for water quality detection: a review. *Applied Spectroscopy Reviews*, 1-26.