Perspective on Chemometric Data Analysis Processes

John H. Kalivas[1] and Robert C. Spiers[2]

[1]*Department of Chemistry, Idaho State University, 921 South 8th Ave., Stop 8023, Pocatello Idaho 83209, USA*
[2]*Department of Physics and Astronomy, 217 Sharp Lab, University of Delaware, Newark, Delaware 19716, USA*

Correspondence
John H. Kalivas, Department of Chemistry, Idaho State University, 921 South 8th Ave., Stop 8023, Pocatello Idaho 83209, USA
Email: kalijohn@isu.edu

Summary
Chemometrics encompasses many facets such as experimental design, data collection and analysis, and many others. This paper, in honor of Paul Geladi, provides our perspective on alternative data analysis approaches that embrace the scientific intuition of the theory of analytical chemistry (TAC) by Booksh and Kowalski. Since the terms chemometrics, machine learning, and artificial intelligence are often used interchangeably without clear definition, we briefly offer our viewpoint on these three terms and their usage. Next, we launch into the significance of matrix effects (the Achilles heel of chemical analysis) followed by its influence on model selection and interpretation (explanation), figures of merit, and sample similarity assessment for model reliability, outlier detection, and classification. The motivation for this perspective is to express our opinion on expanding the role of the TAC multivariate ideologies by integrating a more comprehensive matrix effect viewpoint with the Rashomon effect to improve conventionally practiced data analysis and modeling processes. This expository discussion revolves around spectroscopic data such as near infrared, fluorescence, etc., but the concepts are relevant to other chemometric data structures such as quantitative structure activity relationship (QSAR).

KEYWORDS

Chemometrics, machine learning, artificial intelligence, model interpretation, matrix effect, sample similarity, Rashomon effect

1 INTRODUCTION

This perspective emphasizes our views on key chemometric data analysis processes that we think are static and fragmented. In particular, upholding the importance of the theory of analytical chemistry (TAC) multivariate principles beyond application to rudimentary multi-way methods and respective figures of merit (FOM).[1,2] We consider these topics as appropriate to honor Paul Geladi's vast chemometric work. Paul was a ground-breaking creative thinker, and we highlight one of his progressive projects named regression model comparison plot (REMOCOP) in section 5 on the Rashomon effect.[3,4] Due to the often and inappropriate interchange of the terms chemometrics, machine learning (ML), and artificial intelligence (AI) in this journal and others, the perspective begins with a brief narrative on terminology differences. Making the clarification is important for the reader to understand what we mean by chemometrics. Basically, when it is stated that ML or AI is being used with chemical data, it should be recognized that because chemical data is involved, it is chemometrics.

A substantial part of this perspective demonstrates the wide influence of sample matrix effects underlying chemometric data analysis. By expanding the matrix effect viewpoint, the multivariate position of the TAC can be integrated with the Rashomon effect[5-7] to improve conventionally practiced data analysis and modeling processes. Briefly, assuming a measured sample response is due to the amount of pure analyte present, then we think of a matrix effect as anything that can alter the pure analyte response such as sample interferents, physicochemical and physiochemical effects, environmental and instrument measurement conditions, sample preparations, etc. The TAC is briefly overviewed in section 3. Following, matrix effect details are provided in section 4 to lay the foundation for succeeding sections detailing the relevance of

the TAC and Rashomon effect on model selection and interpretation, FOM, and sample similarity assessment. Whether the goal is quantitative analysis using regression or classification (including outlier detection) using complex deep learning or other ML methods, matrix effects play a central role in making it difficult to obtain accurate and precise analyses as well as any chance of model interpretation.

We are not claiming our viewpoint is the final say on alternative chemometric processes and there are other opinions on directional advancements. Our goal is to present some different chemometric thinking compared to what seems to be mainstream. Tweaking a partial least squares (PLS) algorithm or applying and/or modifying a popular off-the-shelf ML do not count as evolving the paradigm in our opinion. We hope the paper initiates active discussion and fresh ponderings on data analysis schemes and, at the same time, clarifies some misconceptions in chemometrics. The reader should keep in mind that while the examples and much of the discussion relies on spectroscopic data such as near infrared, Raman, etc., all concepts are pertinent to other chemometric dataset types such as quantitative structure activity relationship (QSAR) and other activity relationship constructions or a collection of features profiling each sample in a dataset.

2 CHEMOMETRICS, MACHINE LEARNING, AND ARTIFICIAL INTELLIGENCE

The terms chemometrics, ML, and AI are often used interchangeably and, to us, they have distinct meanings. Following is a brief presentation of the uniqueness of each of these words to better understand the fundamentals of matrix effects in chemometrics.

2.1 Chemometrics

An excellent review with historical documentation of the broad chemometric field is available.[8] For a definition of chemometrics, we rely on the one provided by Svante Wold in his insightful 1995 paper (which is always a great re-read for insight and motivation) where chemometrics is defined as: '*How to get chemically relevant information out of measured chemical data, how to represent and display this information, and how to get such information into data*'.[9] This perspective is concerned with the first part, how to get chemically relevant information out of measured chemical data. As for the second part, concerning the representation and display of data, our laboratory is using immersive analytic approaches focusing on virtual reality as the next generation display interface to visualize data.[10-13] The third part of the definition we categorize as mostly pertaining to measurement processes including experimental design and is not part of this perspective. Many of our points are in step with his themes that we refer to throughout the perspective. Some specifically related quotes from his 1995 paper are:

- "If we let statistics become too influential, its present fashions … , might convert chemometrics into a useless appendix of mathematical theory. … In essence, we must remain chemists and adapt statistics to chemistry instead of vice versa."
- "Here the population concept is at best an abstract model of a real problem, in my view lacking so many properties of a real population that is not very useful in chemistry."
- "And chemometrics must continue to be motivated by chemical problem solving, not by method development."
- "However, a successful future for chemometrics depends, on continued problem orientation, with the important task solving chemical information problems in an efficient

and appropriate way. Certainly, some method development is needed, but this should be motivated by solving real chemical problems, not by mathematical and statistical sophistication. To quote Einstein: 'make everything as simple as possible – but not simpler'."

- "Hence we can safely conclude that any chemical data we put our hands on are 'contaminated' … ."
- "The study of complicated systems is by necessity multivariate."
- "Multivariate analysis is the indirect observation of intrinsic, 'latent', variables."

2.2 Machine Learning

This area of study stems from computer science focusing on building algorithms that enable a computer to learn from the data and make predictions--a data driven approach. Machine learning algorithms are typically agnostic to the discipline they are applied to, e.g., deep learning, principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), uniform manifold approximation and projection (UMAP), feature selection, transfer learning, etc. are generally applicable to a variety of applications. It is when a ML algorithm is applied to or developed with the intention to a solve a chemical problem that it becomes chemometrics. Chemometrics is its own entity qualified to handle the nuances of chemical data. Not all ML approaches are necessarily the best for chemometrics since some are not explicitly intended to solve chemical problems. Other fields that tune ML methods to suit their discipline's dataset structures and needs also label their respective areas with field-dependent names. These names and field include predictive analytics for business, econometrics for economic forecasting,

pedometric for soil modeling, physics-informed machine learning for physics, among others. Partial least squares (PLS) can be considered a ML method but when it is applied to solve a chemical problem, then it becomes chemometrics. If PLS is used to solve a business problem, then it would be a predictive analytics study.

2.3 Artificial Intelligence

We find that it is increasingly common for AI to be used as an umbrella term encompassing advanced learning techniques like deep neural networks and transformers, all the way to simple machine learning tasks such as PLS. It is challenging to pinpoint a strict definition of AI because it does not yet really exist and hence, it is more an understanding of a machine thinking and reasoning like humans. The characterization from Ref. 14 is rather general which states that AI, in its broadest sense, is intelligence exhibited by machines, particularly computer systems. It is a computer science field that develops and studies software methods and hardware to enable machines to *perceive and comprehend their environment and use intelligent learning and reasoning* to take actions that maximize their chances of achieving defined goals. By this definition, chemometric and ML algorithms are not considered AI. While they have built in decisions to appear intelligent, these autonomous decisions are typically based on pre-set algorithm numerical thresholds. Thus, such autonomous algorithms are written by humans with human cognitive biases to perform the programed task. For example, PLS is not AI and even autonomous versions that determine the optimal number of latent variables (LVs) to form the model are not AI. While there have been recent glimpses of AI for simple tasks, to date, true AI does not exist. What is currently named AI is really algorithmic as are ML and chemometrics and

if chemical data is involved, what is often mistaken as AI and ML should be named chemometrics.

3 The Theory of Analytical Chemistry (TAC)

Close to 30 years have passed since the TAC by Booksh and Kowalski was published where multivariate instruments and corresponding data orders with respective multivariate figures of merit (FOM) were described. [1,2] The TAC advanced conventional univariate chemical analysis thinking to multivariate concepts with two main themes. In the first, analytical instruments are labeled by the tensorial order (number of dimensions) of the data they produce, e.g., a spectrum from a spectrometer is composed of multiple measurements from a single sample assembled into a vector and is first-order for one dimension of data. The second broadened basic analytical chemistry FOM from univariate to multivariate. Increasing instrument and data orders extends the multivariate nature of the data and FOM. Correspondingly, the information content and net analyte signal (NAS) increase and hence, analyte selectively also increases permitting optimal designs of laboratory analysis procedures. An emphasis of this perspective is to apply the higher order multivariate principle of the TAC to other common data analysis processes currently operating in lower orders.

4. MATRIX EFFECTS

In a university general chemistry course, students learn about intra- and inter-atomic and molecular interactions and their effects on respective physicochemical properties, e.g., boiling points of a polar liquids are greater than nonpolar liquids or the strengths of some intra- and inter-atomic interactions are correlated to physicochemical property trends on the periodic table.

Later in analytical chemistry courses, these atomic and molecular interaction effects on chemical analysis are presented under the umbrella term matrix effects.[15] Because of matrix effects, early analyses typically involved developing extraction and chromatographic methods to isolate the analyte into a relatively matrix effect free environment with perfect selectivity. [16] Following extractions, the final univariate analysis was commonly gravimetric, volumetric, electroanalytical, or spectral. Alternatives to extraction were the standard addition method (essentially matrix matching the calibration standards to the sample matrix) or identifying an analyte sensor that is interference free (or nearly free with a small selectivity coefficient for interferences) such as a spectral wavelength or an ion selective electrode.[15]

Before proceeding, we want to first define what constitutes a pure analyte. Whether atomic or molecular, there is a quantum mechanical structure to the analyte and this assumed configuration is perturbed by its environment and measurement process that we name matrix effects. Thus, we think of the pure analyte as that in a gas phase at infinitely low pressure where the analyte is least likely to be perturbed. Svante Wold noted that this idealized state lacking matrix effects is experimentally inaccessible and impractical for real-world use since samples are imperfect.[9] Even for pure liquid or solid forms of the analyte, there is a perturbed quantum mechanical relationship compared to the isolated gas phase. Thus, each sample has its own unique matrix effects that can be homogeneous throughout the sample or heterogeneous, as is common with solid samples. It should be noted that for some chemical systems, sample-wise matrix effects trivially vary between samples due to little change in sample compositions and in this case, matrix effected pure component spectra can probably be taken as respective pure component spectra.

In this perspective, sample-wise matrix effects are regarded as a larger-scale problem than just the molecular-scale interactions creating quantum mechanical perturbations. For example, recent developments in bioanalysis require adding the cellular-scale microenvironment as a matrix effect where the interaction between physiochemical and chemical processes alter the measured properties. Also contributing to sample-wise matrix effects are all aspects of processes involved in collecting and preparing a sample and the measurement conditions such as the environment and instrument. The reader is referred to Refs. 17-19 for excellent discussions of these and other extensive sources of matrix effects.

Using spectroscopy and basic chemical theory with no matrix effects, a measured spectrum for the ith sample $\mathbf{x}_i$ is represented as a linear combination of pure component spectra for all responding constituents $\mathbf{k}_v$ and their matching amounts y_{iv}. Recently, a mathematical characterization of matrix effects was derived leading to the expression

$$\mathbf{X} = (\mathbf{Y}\ \mathbf{A})\mathbf{F} + \mathbf{E} \tag{1}$$

for a training sample set $\mathbf{X}$ used to form regression or classification models where $\mathbf{F}$ represents all possible pure component spectra $\mathbf{k}$ and basis set pure component matrix effects $\mathbf{r}$ with $\mathbf{F} = \begin{pmatrix} \mathbf{K} \\ \mathbf{R} \end{pmatrix}$ per corresponding amounts y (in $\mathbf{Y}$) and α (in $\mathbf{A}$) and $\mathbf{E}$ symbolizes randomly distributed gaussian error that is henceforth removed for a simpler discussion.[20] If a component or matrix effect is not present in the sample, the matching amounts y and α are zero. In essence, a spectrum is a snapshot of the sample structure that can be expressed as a linear combination of pure component spectra and matrix effect basis vectors. Ideally, $\mathbf{X}$ represents an independent and identically distributed (iid) training set. However, as noted by Svante Wold's quote in section 2.1, that to speak of the training sample set population is abstract and lacks many properties of a real

population. Hence, due to the breadth of sample-wise matrix effects, every new sample from the real population to be predicted for analyte amount or class membership has the potential to be an outlier (out of distribution (ood)) to the training set.

It may be that a handful of vectors from **F** dominate an **x** with larger α values than other vectors and these leading vectors are probably particular to the measurement system. For example, with spectral data, major matrix effect vectors would likely be different between NIR, FTIR, NMR, fluorescence, etc. From eq. 1, it is apparent that a multitude of linear combinations of **R** exists to form **x** with the particular sample non-zero **K** and distinguishing pure component matrix effects enfolded to **x** is not practicable.

While attempts with quantum mechanics have been made to interpret NIR spectra relative to wavelength peaks, these endeavors have been for simple single component chemical systems in a solid or liquid states as well as in a solvent.[21,22] Following these spectral interpretations are often interpretations of model basis vectors and regression vectors such as those for PLS. As noted in the next section, such interpretations are not reasonable.

Chemical data is unique in that the basic underlying cause of most matrix effects are the sample's complexities of atoms and molecules and teasing out the exact matrix effects is generally not possible. Simple examples of matrix effected spectra from **F** were recently presented and discussed.[20] Chemometricians should develop and apply models with matrix effects in mind, either to alleviate the unfortunate influence of matrix effects on their results or to leverage the appearance of similar matrix effects to make more accurate predictions. Chemical data is similar to other discipline datasets, e.g., econometrics, psychometrics, pedometrics, or biometrics to name a few, where the respective field dataset underlying variance structure are incredibly complex.

4.1 Non-Uniqueness of Basis Vectors for X and the Regression Vector Model $\hat{\mathbf{b}}$

Numerous schemes exist that endeavor to isolate and characterize matrix effects. Consider the simple PCA of a data matrix **X** where the PCs or LVs are extracted to form a vector basis set spanning **X**. Attempts have been made to interpret the structure of each PC relative to the measured variables making up **X**, such as wavelengths with spectral data. However, PCA is just one of the many matrix decompositions that can be used with each creating a separate basis set with unique sturcutre.[23] Furthermore, because spectra are non-unique in that many possible linear combinations of the matrix effects exists, a model regression vector estimated from training set **X** and **y** by

$$\hat{\mathbf{b}} = \mathbf{X}^{+}\mathbf{y} = \left(\left(\mathbf{Y}\,\mathbf{A}\right)\mathbf{F}\right)^{+}\mathbf{y} \tag{2}$$

is just one of the many feasible models where the superscript plus sign denotes a generalized inverse. This multiplicity of regression vectors is further characterized as the Rashomon effect in section 5.3.[5-7] Briefly, the Rashomon effect stems from Akira Kurosawa's 1950 film 'Rashomon' where a murder is described in four contradictory ways by four witnesses.[24] The Rashomon effect means that each person (model) does not provide a full objective accounting of all the information.

Using the singular value decomposition (SVD) of $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{T}$ as an example basis set, the above points are further supported by substituting the SVD into eq. 2 to form

$$\hat{\mathbf{b}} = \mathbf{X}^{+}\mathbf{y} = \left(\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^{T}\right)\mathbf{y} = \mathbf{V}\boldsymbol{\beta} \tag{3}$$

showing that any regression vector can be expressed as a linear combination of the eigenvectors in **V**.[25] Another basis set can be similarly used such as PLS LVs to form the same regression vector[23] and hence, the same $\hat{\mathbf{b}}$ can be obtained from each basis set.

A complementary manifestation of eq. 3 is

$$\hat{\mathbf{b}} = \left(\mathbf{V}\mathbf{P}\boldsymbol{\Sigma}^{-1}\mathbf{U}^T\right)\mathbf{y} = \mathbf{V}\mathbf{P}\boldsymbol{\alpha} \tag{4}$$

where **P** represents a diagonal matrix with filter factors on the diagonal.[26,27] Example filter values are: PCR (p_i =1 or 0), ridge regression (RR) ($0 \leq p_i \leq 1$), and PLS ($0 \leq p_i < \infty$). Other filter factor values can be used to form a multitude of models (e.g., generalized RR) that all predict with equal accuracy. If new prediction samples are iid to the calibration set, then which ML algorithm used to form the prediction model does not matter (deep learning, PLS, etc.)[28] demonstrating the Rashomon effect described in section 5.3.

5 MODEL INTERPRETATION (EXPLANATION)

In light of the above discussions, one may be asking how then can a model such as a PLS regression vector be interpreted? Before diving into model interpretation and answering this question, we want to define what we mean by interpretable versus explainable models. These two terms are often used interchangeably but there is a difference and unfortunately, there are no agreed upon distinct meanings.[29-33] After our clarification, we describe the Rashomon effect[5-7] and use it to further justify why we do not think interpretation (explanation) of a chemometric model is possible under normal conditions with today's knowledge base. We also discuss model diversity, multiplicity, generalizability relative to the bias-variance trade-off, and selection, important to understanding models and the Rashomon effect. We conclude this section with steps that can be taken to perhaps provide limited model interpretability (explainabiltiy).

5.1 Model Interpretation versus Explanation

The convolution of trying to unravel the differences between interpretability and explainability is too complex for this discussion and we will accept a simple version. Interpretability is seen as the mechanics of how model predictions form from the model without necessarily knowing why, i.e., the model cause and effect. Explainability is being able to explain in human terms the internal mechanics of why the model formed by the machine or deep learning system. As an example, the coefficients of a PLS regression vector $\hat{\mathbf{b}}$ dictate its shape and direction that weight and linearly combine the respective sample spectrum wavelengths by $\hat{y} = \mathbf{x}^T\hat{\mathbf{b}}$. It is easy to see *how* the predicted analysis amount is formed from the vector product sum. However, *why* the particular wavelength $\hat{\mathbf{b}}$ coefficient values is not apparent, i.e., what did the model learn to become that shape and direction? Because model coefficient values stem from the particular training set sample-wise matrix effects, it is clear that coefficient values themselves are dynamic and not fixed to certain values. Thus, with chemical data, we postulate the *why* is not obtainable. For example, trying to explain a PLS regression vector relative to the LVs is not realistic as another basis set (see section 4.1) could be used to make that same regression vector giving a completely different model explanation. Using one basis set over others limits the chemical viewpoint. Additionally, as shown in section 5.3 on the Rashomon effect, if the training samples change by a cross validation, then different regression vectors result with different explanations. This unobtainability of model explanation is further elaborated in all the following subsections of section 5. Lastly, it needs to be emphasized that we use explanation to refer to the common use of model interpretation in the chemical and chemometric literature.

It is often stated that complex models (a model with large number of parameters) are less explainable than simpler models.[28] This analysis stems from Occam's Razor that among equally accurate models, the simpler ones are better. However, in context of the discussion on matrix effects and models, simpler models are not necessarily more explainable for chemical data and the simpler model may not be capturing as much of the matrix effects as a more complex model. For example, the weights of a linear model (e.g., regression coefficients) might seem explainable, but they are dependent on features used, pre-processing, etc. Case in point, the relation between COVID risk and vaccination can be positive or negative depending on whether the features include measures of immunodeficiency, old age, or infancy.[34] Thus, we strongly agree with the statement that neither linear models, rule-based systems, nor decision trees are fundamentally explainable.[29]

Some approaches to interpreting or explaining a model rely on a post-hoc approach. For example, a sample molecule is classified into a category because its spectrum or set of molecular descriptors *look like* those in the category; a "this looks like that" approach.[35-40] However, due to the underlying intricacies of matrix effects, even post-hoc approaches can be sketchy with chemical data.

5.1.1 Simple versus Complex Models

Further muddying the waters of model explainability relative to simple versus complex models are recent bias-variance tradeoff observations with deep learning networks.[41-46] Classical theory dictates that models balancing the ML bias-variance trade-off is located between underfitted and overfitted models. Shown in Figure 1 is a characterization of this classical view where a model from the lower part of the U-shaped curve denotes a useful model. Past chemometric work

suggested that the classical U-curve is more skewed due to a slower increase in bias as model complexity increases.[47] The possibility of having both bias and variance decrease with increasing model complexity was also pointed out.

Current work with deep learning networks are showing a dissimilar pattern from the classical understanding where large overparametrized networks maintain small generalization errors in spite of the models being overfitted.[41-46] Displayed in Figure 1 is a depiction of the situation and similar plots were obtained with real data. Note that at first the classical U-curve forms followed by the second drop for a 'double descent' curve. Recent work details some reasoning behind the double descent observed in real data.[45,46] Such complex models are difficult to make sense of and, in spite of the general acceptance of Occam's Razor, the simple model may not be best and more intricate models can apparently be more accurate with better generalizability. With chemometric data, in order to form a model orthogonal to the full complement of the matrix effects, the 'global' NAS model requires $\mathbf{Fb} = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}^T$ and this NAS model would be too overfitted (bumpy or jagged in appearance) to explain[20,48] and may reside in the double descent.

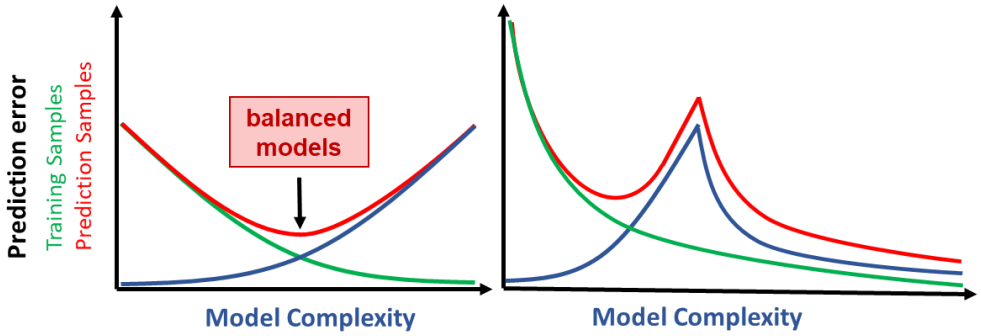

Figure 1. Graphical characterization of the classical bias-variance trade-off perception (left) and the proposed view with the 'double descent' (right).

5.2 Making Explainable Models

A general goal in ML is create "more explainable" mdoels.[39] However, making such models is not without many challenges and a recent paper highlights the problem of making explainable models.[41] Presented in this section is one aspect of creating explainable models by placing constraints into the model building process, a common tactic in chemometrics.

By constraining a model to satisfy specific penalties,[49-51] explanation of the resultant model may improve towards the particular phenomena constrained. Penalties on known pure component analyte ($\mathbf{k}_a$), interferences ($\mathbf{k}_i$), and matrix effects ($\mathbf{r}_m$) spectra, i.e., $\mathbf{F} = \begin{pmatrix} \mathbf{k}_a^T \\ \mathbf{k}_i^T \\ \mathbf{r}_m^T \end{pmatrix}$ have been used where the constraints strive for collinearity between the model and the analyte pure component spectrum ($\mathbf{k}_a^T\hat{\mathbf{b}} = 1$) and orthogonality with interferents and matrix effects ($\mathbf{k}_i^T\hat{\mathbf{b}} = 0$ and $\mathbf{r}_m^T\hat{\mathbf{b}} = 0$) or $\mathbf{Fb} = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}^T$. The final selected model is supposedly explained as nulling interferents and matrix effects but in reality, it is just one of the many models that can null these effects and yet each model has a different shape and magnitude and is hence, non-explainable. Only CLS provides an orthogonal NAS model and ILS methods form oblique NAS models.[20,52-54] It is important to note that the CLS model only captures a local NAS view corresponding to the particular training set used.[20] The more penalties included in forming models, the more each penalty is compromised relative to the other penalties in the group, i.e., the optimizer does not focus on one penalty but compromises across all penalties. Even with dominant constraints, it may be difficult to attach meaning to the actual effect of the major phenomena constrained.

Perhaps explainability requires using inputs that themselves are individually explainable such as wavelengths. However, relative to our discussion on sample matrix effects and the Rashomon effect (including wavelength selection) in the next section, identifying analyte specific features is difficult short of chromatographically or chemically extracting the analyte from its sample matrix.

5.3 Rashomon Effect

It was becoming apparent to many researchers that many different and contradictory models can be formed from one dataset (regression, decision trees, neural networks, etc.) that all accurately predict new samples. This modeling observation was named the Rashomon effect[5] where each model (person) does not provide a full objective accounting of all the possible matrix effects.[5,24] A worthwhile quick read is a recent viewpoint on the Rashomon effect in ML characterizing the circumstances reasonably well.[6]

Svante Wold noted 'Importantly, the domain, the multidimensional volume, containing the 'best' solution to given objective has a nonzero volume. This necessitates experiments to find the best point(s), the best set(s) of conditions inside this domain (our parenthetical pluralization).'[9] Recently, a detailed paper was published that mathematically describes the Rashomon effect[7] including the Rashomon volume, set, and ratio all which parallel Wold's statement. Since the original Rashomon effect paper, several updated versions have been posted. Briefly, the Rashomon volume is the number of models in the hypothesis space, the Rashomon set corresponds to those models with a risk (prediction error) less than or equal to a minimum risk plus an offset of θ where θ is the Rashomon parameter with $\theta \geq 0$ (the set of models with a loss below a threshold). Depicted in Figure 2 are various Rashomon sets depending on the θ value.

The Rashomon ratio is obtained from the Rashomon set ratioed to the Rashomon volume. From a plot of the Rashomon ratio against the risk (loss), the Rashomon curve is obtained with a Γ-shape. Figure 3 illustrates the plot and the reader is referred to it for a qualitative description of the Rashomon curve. A large Rashomon ratio at the elbow correlates to machine learning algorithms with similar accurate performances and represents an area to select models from and which ML algorithm used to train and predict with does not matter (deep learning, PLS, etc.).[7,28]

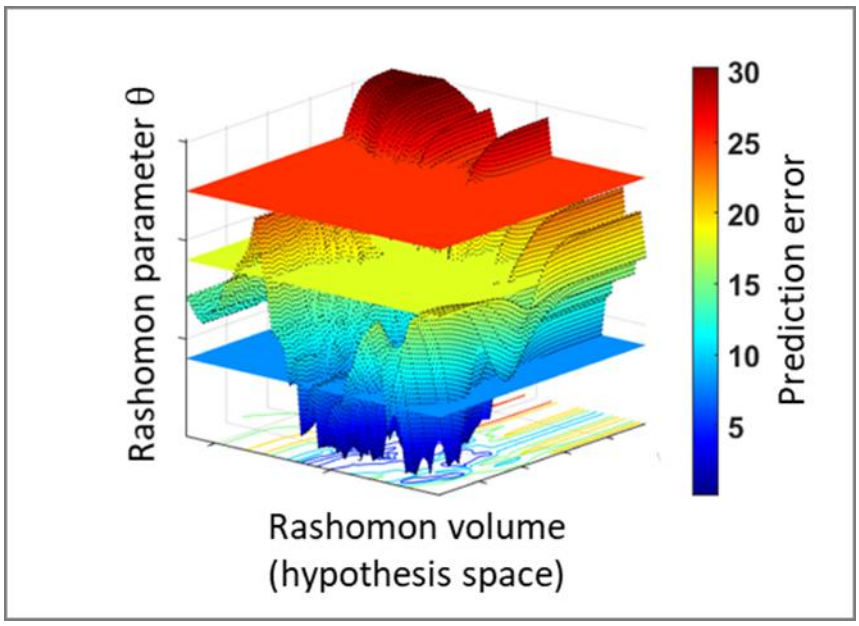

Figure 2. A representation of a Rashomon set in 3-D

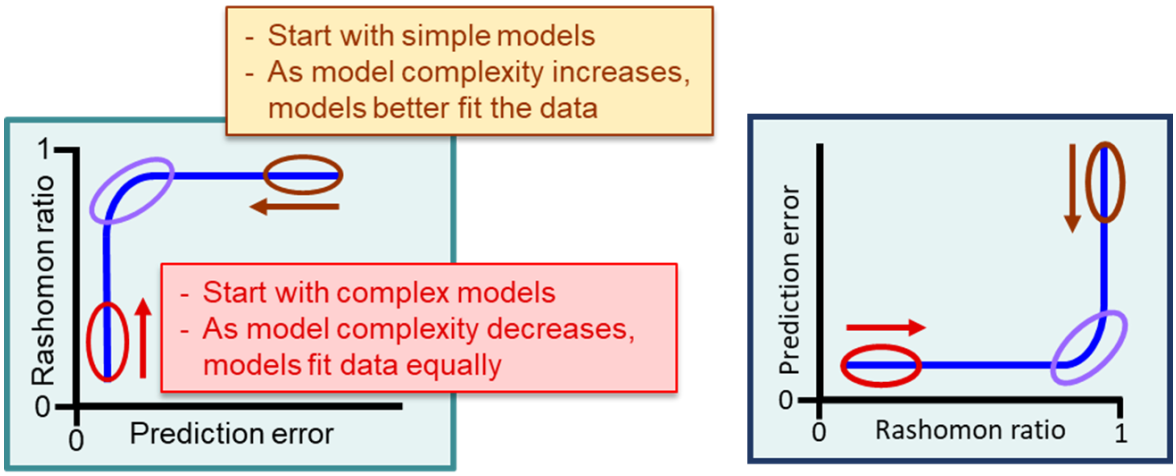

Figure 3. Depiction of the Rashomon curve in the original sense (left) and as chemometricians may view it (right). In both figures the purple circled region corresponds to the Rashomon elbow.

5.3.1 Chemometric Examples of the Rashomon Effect

Following are multiple chemometric examples of the Rashomon effect ending with Paul Geladi's regression model comparison plot (REMOCOP).[3,4]

The easiest Rashomon effect to recognize is wavelength selection with too many selection algorithms to list. Regardless of the algorithm, each time it is run with a different pipeline including samples used, different wavelengths are identified as 'best' and each best set accurately predicts samples. Displayed in Figure 4 are prediction errors plotted against the multiple linear regression (MLR) vector 2-norms for three subset sizes using 10,000 random selections each for a NIR data set. From Figure 4, it is observed that multiple good predicting models exists with different sample and wavelength combinations. The histogram in Figure 4 indicates that only a few wavelengths are perhaps favored. It may be possible to mine out the important wavelengths using the variable importance cloud (VIC).[55] The VIC reveals wavelength importance relative to

the significance of other good model wavelengths in the Rashomon set. For example, the VIC can disclose that a wavelength may be important only when another wavelength is not selected. Regardless, in context of our sample-wise matrix effects discussions, it would be difficult to provide any explanation to why selected wavelengths are important relative the complex matrix effects.

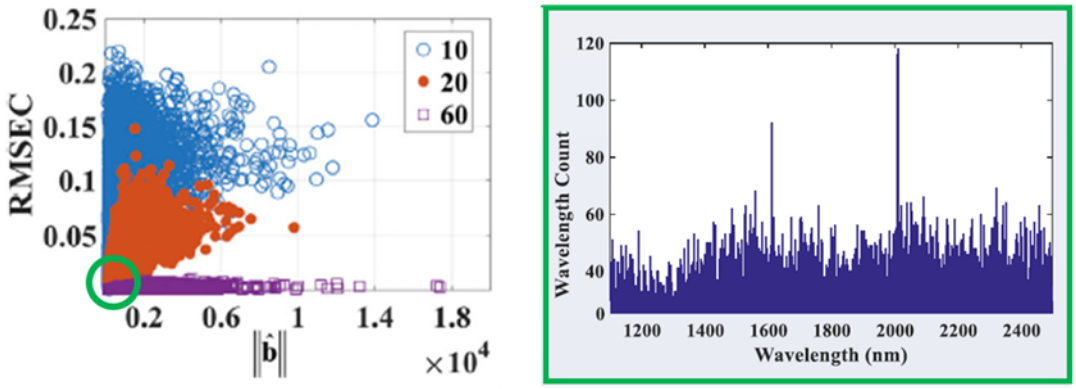

Figure 4. Prediction and MLR regression vector 2-norms for 10,000 random combinations of wavelengths for three subset sizes noted by the symbols using NIR spectra (left). A histogram of the wavelengths selected (right) in the green circled region for the better models

The Rashomon effect is apparent when regularization tuning parameters and/or pre-processing methods are involved forming diverse regression models predicting similarly. Plotted in Figure 5 are accurately predicting RR model vectors over a range of ridge parameters. Another set of regression vectors are plotted in Figure 6 for a modeling updating method with two adjustable tunning parameters where model selection concentrates on model diversity predicting target samples similarly.[56]

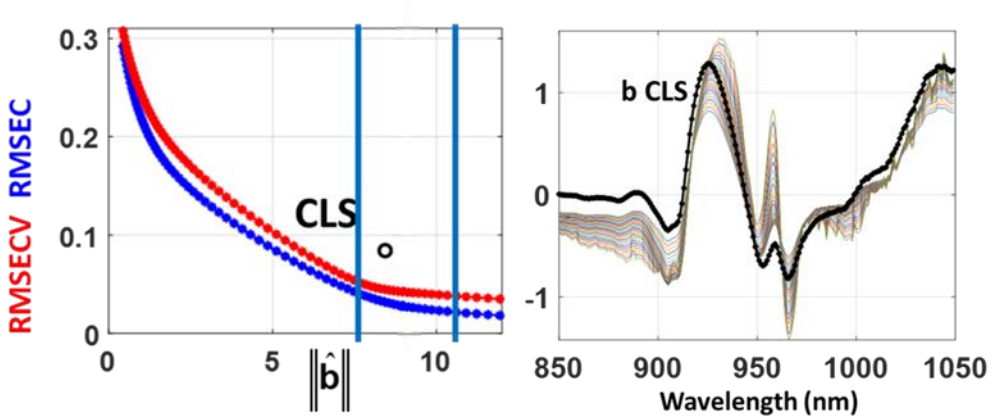

Figure 5. Plots of prediction error against RR model vector 2-norms (left) for a range of ridge parameters showing a Rashomon set of models marked by the blue vertical lines and the local orthogonal CLS model (black circle). Respective RR and CLS regression vectors are plotted (right) with the CLS regression vector marked by the black line showing near orthogonality of the RR models.

Another multiple tuning parameter example involves varying the degree of spectral orthogonal projection for three pre-processing methods producing different regression vector solutions.[57] The three projection approaches are named generalized Tikhonov regularization (GTR), generalized net analyte signal (GNAS), and generalized least squares (GLS) with respective tuning parameters β, γ, and λ varying the degree of orthogonality. Shown in Figure 7 is the GLS prediction error heat map for various levels of pre-processing and number of PLS LVs. From the heatmap, is seen that GLS makes sets of models predicting equivalently and accurately. It is doubtful any of the diverse models are explainable to the true underlying matrix effects. The GTR and GNAS heat maps are similar with multiple accurately predicting models.

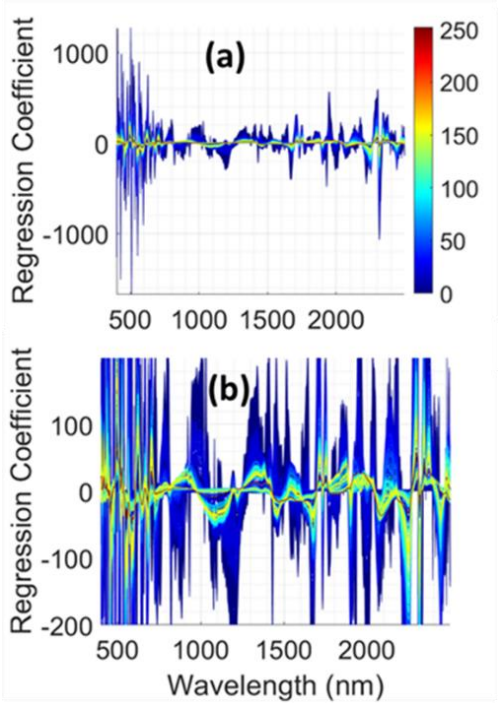

Figure 6. A Rashomon set of PLS regression vectors accurately predict target samples. Reprinted with permission from Spiers RC, Kalivas JH. Reliable model selection without reference values by utilizing model diversity with prediction similarity. J. Chem. Inf. Model. 2021;61:2220-2230. Copyright 2021 American Chemical Society. https://pubs.acs.org/doi/10.1021/acs.jcim.0c01493

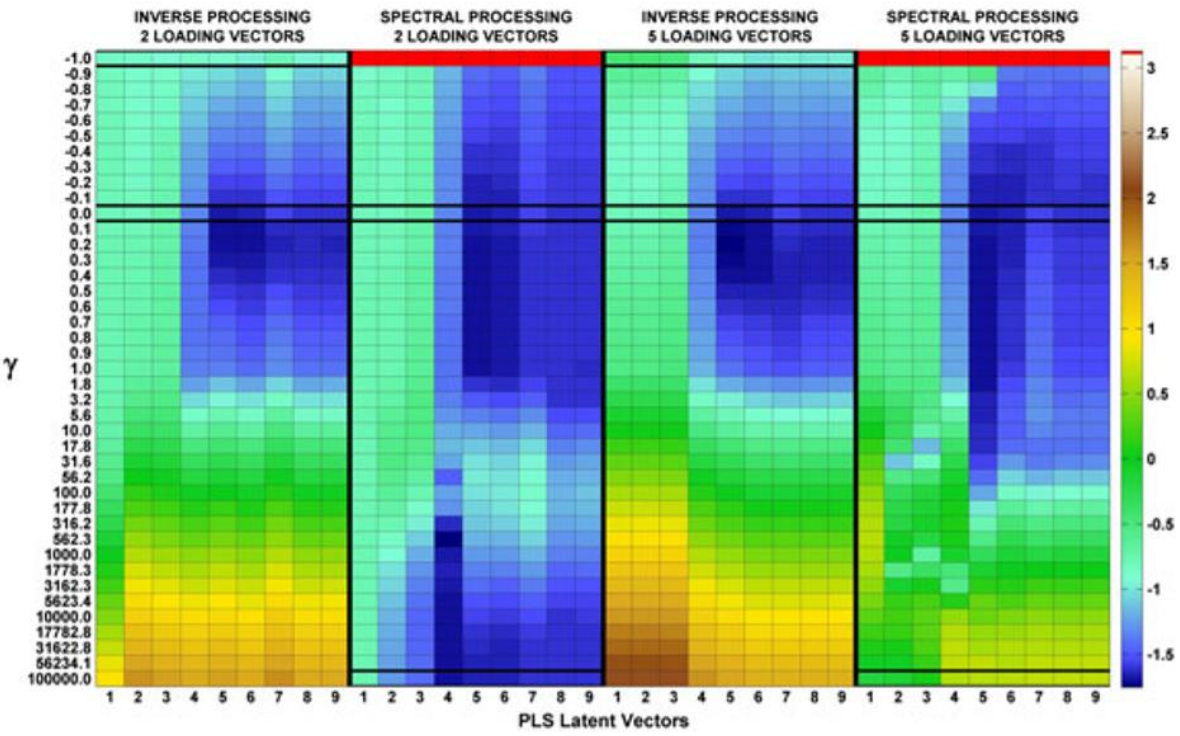

Figure 7. Heat map of log10(RMSEV) across GLS projection regularization parameter γ, the degree of orthogonal projection. Models without projections occur at γ = 0 and the orthogonal NAS models are found at the black horizontal lines. Reprinted from E. Andries, J.H. Kalivas. Interrelationships between generalized Tikhonov regularization, generalized net analyte signal, and generalized least squares for desensitizing a multivariate calibration to interferences. *J. Chemom.* 2013, 27: 126-140. https://doi.org/10.1002/cem.2501

Paul Geladi Example

Highlighted here is Paul's modeling approach named regression model comparison plot (REMOCOP).[3,4] He proposed that stronger predicting models will cluster together in a PC score plot from the PCA of a collection of models spanning respective tuning parameter ranges, e.g.,

PLS, principal component regression (PCR), RR models. The under- and overfitted models will also correspondingly cluster.

Shown in Figure 8 are PCA score plots of the combined set of full tuning parameter ranged PLS and RR regression vectors for QSAR and NIR spectral datasets.[56] Labeled are the better predicting selected models, suitable models but not selected, and under- and over-fitted models demonstrating the corresponding clusters and Rashomon effect, i.e., numerous regression models with different shapes and magnitudes are formed to effectively predict samples.

In Paul's REMOCOP work, it was suggested that score values of a given model from a region of better predicting models can be modified to produce other good predicting models. This tactic essentially generates additional models in the Rashomon set. A similar approach was used where different groupings of score vectors were combined to create a collection of models from the Rashomon set with different shapes and magnitudes accurately predicting samples.[58]

5.3.2 Economic Example

Using structural and flexible functional modeling approaches, economic scientists arrived at the Rashomon situation. The existence of a single optimal model is questionable because many models fulfilled a data set with extraction of contradictory inferences.[59-61] To assist in the analysis, it was advised that data should be allowed to speak for itself, but economic theory should be used whenever possible. Specifically, the perceptive Griliches quote from 1967, "Although it is admirable to use deep statistical theory with much stochastic structure in analysis, this is no substitute for economic theory."[62] This statement is akin to Wold's emphasis on the

chemistry and chemical theory. Besides this economic example, numerous other examples exist in other fields.

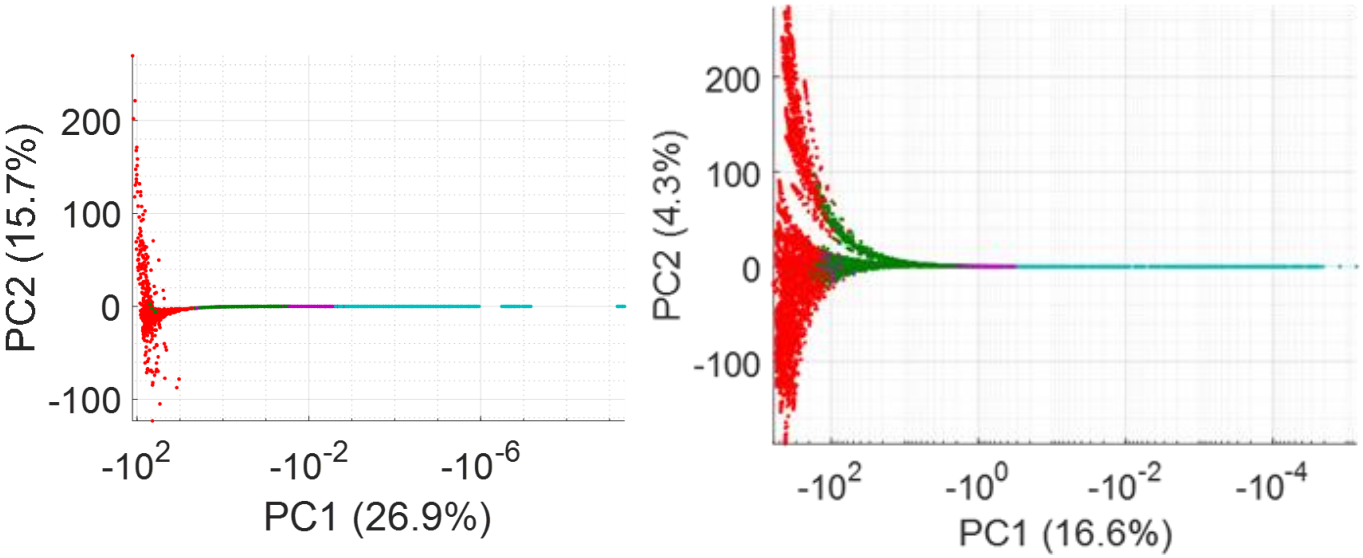

Figure 8. Score plots (logarithmic PC1) of RR and PLS models across tuning parameter ranges from multiple data splits for datasets QSAR (left) and NIR (right) showing RR and PLS selected models (green), over-fitted (red), under-fitted (cyan), and acceptable non-selected models (magenta). Reprinted with permission from Spiers RC, Kalivas JH. Reliable model selection without reference values by utilizing model diversity with prediction similarity. J. Chem. Inf. Model. 2021;61:2220-2230. Copyright 2021 American Chemical Society. https://pubs.acs.org/doi/10.1021/acs.jcim.0c01493

5.3.3 Changing the Training Sample Set Forms Diverse Models

In a noteworthy paper on the NAS and model explanation, mathematical CLS and ILS expressions were chemically and mathematically derived showing the effects of calibration samples and noise on determining respective models.[63] Briefly, $\mathbf{B}_{CLS} = \mathbf{\Sigma}^{-1}\mathbf{K}^T\left(\mathbf{K}\mathbf{\Sigma}^{-1}\mathbf{K}^T\right)^{-1}$ and

$\mathbf{B}_{ILS} = \left(\mathbf{K}^T \Psi \mathbf{K} + \mathbf{\Sigma}\right)^{-1} \mathbf{K}^T \Psi$ where $\mathbf{\Sigma}$ denotes the covariance matrix of measurement error and $\mathbf{\Psi}$ is the sample amount covariance estimated by $\mathbf{Y}^T\mathbf{Y}$. For both CLS and ILS, **K** represents the same matrix affected pure component spectra included in **F**. The CLS model formed is the local NAS model relative to the matrix effects in the calibration set used to first estimate the local **K**, and then $\mathbf{B}_{CLS}$, i.e., $\mathbf{KB}_{CLS} = \mathbf{I}$ only for the matched set. Depending on the degree of matrix effects differences when changing calibration samples for CLS, the estimated **K** and $\mathbf{B}_{CLS}$, will locally change. The ILS models maintain $\mathbf{KB}_{ILS} = \mathbf{L}$ where **L** contains relaxation values from orthogonality (see section 5.3.5).[52] As a reminder, in order to form a model orthogonal to the full complement of the matrix effects, the 'global' NAS model requires $\mathbf{FB} = \mathbf{I}$ and this NAS model would be too overfitted (bumpy or jagged in appearance) to explain.[20,48]

It was shown with chemical data that by changing the calibration (training) set through random selection, the resultant ILS regression vector directions and magnitudes correspondingly change for a collection of diverse models and yet, the models accurately predict samples left out of the calibration.[63,64] Changing the training samples alters matrix effects and by necessity, diverse regression vectors are expected from the different linear combinations of respective matrix effects and model explanation is not likely. Other examples of diverse models predicting samples well and similarly have been demonstrated and leveraged for model selection.[65]

5.3.4 Model Selection: Diverse Models are Better than One Model

In the previous subsection, it was mathematically documented and verified with chemical data that by selecting different calibration samples from a global iid sample pool, accurate predicting diverse ILS model regression vectors are formed with unique directions and magnitudes.[63,64]

Each model captures a different view of the particular underlying sample-wise matrix effects. Also previously noted was that changing model tuning parameters, accurate predicting diverse models are formed. Hence, it is judicious to use a family of models capturing the full calibration sample-wise matrix effect possibilities for reliable predictions of new samples with their particular matrix effects. Dissimilar models predicting a sample similarly improves the chances that the prediction is accurate. Trusting one model viewpoint does not seem prudent, e.g., expecting the one selected model that nulls non-analyte information in the training set to similarly null the non-analyte information present in new prediction samples. Instead, it is advocated here to exploit matrix effects to identify a collection of models that agree on the new sample predicted amounts.

Model diversity is not new to chemometrics. Model ensemble, consensus modeling, the local modeling algorithm LOCAL, and others use multiple prediction models. However, using model diversity in chemometrics seems underutilized. For example, strong chemical evidence of the usefulness of diverse models is stacked regression. This approach was used with PLS where it was shown that using a diversity of models allowed successful prediction of an outlier in the prediction set.[66] A recent paper showed that by varying the model prediction error criterion for selecting the best model, diverse models are obtained resulting in composite predctions.[67]

Model diversity is similar to using multiple wavelengths (multivariate data) to obtain better answers in the analysis of spectral data. Requiring diversity also parallels fusion of multiple instruments and the multi-way method parallel factor analysis (PARAFAC) where each instrument and PARAFAC order provide alternate assessments of underlying sample-wise matrix effects for expanded data characterizations. Leveraging model diversity fits in with Svante Wold's statement that "The study of complicated systems is by necessity multivariate."[9]

We end this section with a Carl Jung quote, “Meaningful coincidences are unthinkable as pure chance, the more they multiply, the greater and more exact the correspondence is … they can no longer be regarded as pure chance, but for the lack of a causal explanation, have to be thought of as meaningful arrangements.”[68]

5.3.5 Figures of Merit

A Rashomon set of models will by default generate corresponding Rashomon sets of multivariate figures of merit (FOM) such as net analyte signal (NAS) with different orthogonality/angles to the matrix effects, selectivity, sensitivity, limit of detection, etc.[2,20,51] Shown in Figure 9 are heat maps of relaxed CLS (rCLS) FOM selectivity and sensitivity values. Axis values are calculated by varying the relaxation l_{aa} and l_{ia} values for the analyte and the interferent respectively calculated by $l_{aa} = \mathbf{k}_a^T \hat{\mathbf{b}}$ and $l_{ia} = \mathbf{k}_i^T \hat{\mathbf{b}}$ where the local $\mathbf{K}$ was estimated from local CLS for the particular training set.[51] The FOM values for the local orthogonal CLS models with relaxation values with $l_{aa} = 1.0$ and $l_{ia} = 0.0$ are at the heat map centers. Clearly there are many useful models with acceptable FOM values. If FOM are being used to optimize a chemical analysis, then it would be expected to select from the Rashomon set, a process with the best FOM values. The optimized system relative to the FOM is one of many from a Rashomon set and many other FOM values are possible for the same chemical analysis with respective equally accurate models.

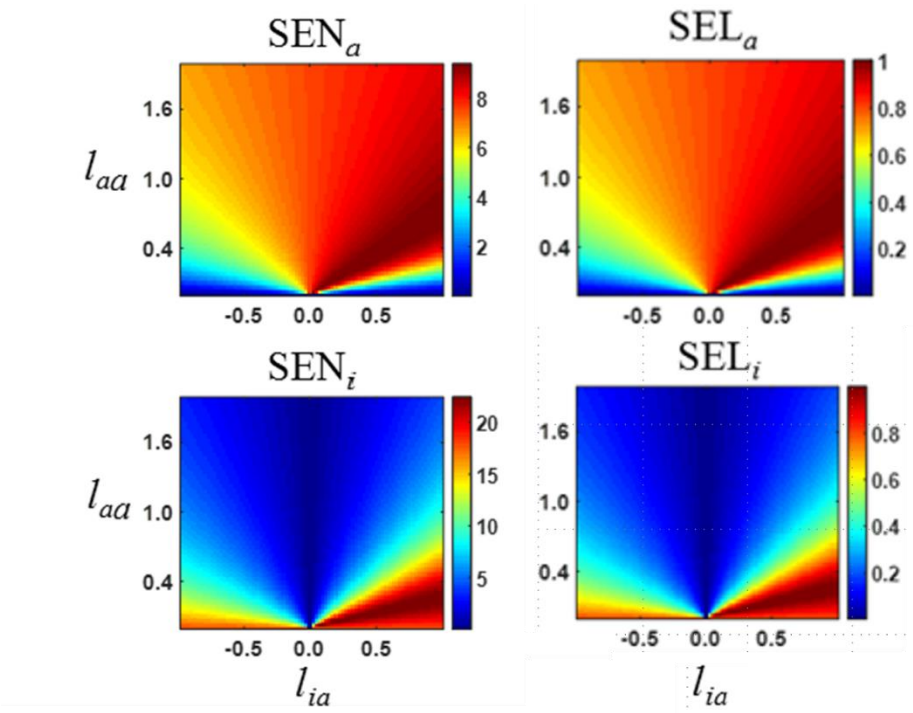

Figure 9. Heat maps of FOM sensitivity and selectivity for the analyte methanol and interferent water. Reprinted from J.H. Kalivas, J. Ferré, A.J. Tencate. "Selectivity-Relaxed Classical and Inverse Least Squares Calibration and Selectivity Measures with a Unified Selectivity Coefficient. J. Chemom. 2017, 31: 2925. https://doi.org/10.1002/cem.2925

5.4 Model Summary

Predicting with multiple useful diverse models is a direct ramification of the TAC multivariate guideline. Each TAC data order affords a different sample matrix effect perspective and likewise, individual models do the same. The TAC and the concept of predicting with diverse models also fit Svante Wold's statement "The study of complicated systems is by necessity multivariate."[9]

The Rashomon effect was explored and we hope the examples presented among the many possible chemometric situations encourage discussion and usage of it. In each of the examples presented as well as other Rashomon effects observed in the chemometric literature, the TAC is central to the multivariate advantages obtained. In view of the Rashomon effect, Sir George Box's famous quote is indeed insightful, "Essentially, all models are wrong, but some are useful." following up with "… the practical question is, how wrong do they have to be to be useful."[69] A similar less familiar quote is from Peter Truran "... seemingly incompatible models may be used to make predictions about the same phenomenon. ... For each model we may believe that its predictive power is an indication of its being at least approximately true. But if both models are successful in making predictions, and yet mutually inconsistent, how can they both be true?"[70]

Regardless of model constraints (penalties), degree of orthogonality, preprocessing (SNV, MSC, derivatives, etc.),[71] what is obtained are 'useful' models. At best, a useful model is perhaps 'rationalizable' and may have some inferences. Correspondingly, useful FOM are obtained and if one model is being selected, it should have the best compromise set of FOM values.

We hope the reader is swayed that model explanation and interpretation are two unique concepts. Model interpretation is *how* a model affects the prediction and what chemometricians are more interested in is explaining the *why* of a model. Just because we view models as unexplainable due to the convolution of the matrix effects, it does not mean that models are not useful. The usefulness of training and prediction is obvious from the decades of research and applications. Additionally, we strongly agree with the quote from the famous statistician Sir David Cox, "Construction of idealized representations that capture important stable aspects of

such systems is, however, a vital part of general scientific analysis … .”[72] Lastly, a generalizable model does not mean the model is explainable.

6 SIMILARITY ASSESSMENT

Determining if a sample is similar to another sample or to a domain space spanned by a collection of samples is not as simple as one may think. A basic innate human characterization of the similarity between two objects (samples) is the closeness relative to the degree of agreement between respective features. However, establishing a mathematical understanding of this deep-rooted human similarity notion is difficult if not impossible. Even with a mathematical construct C, it is still vague to state object “A is similar to object B with respect to C”[73,74] because the similarity is strictly relative to C and another mathematical measure can result in a completely different assessment. Unfortunately, many approaches to quantifying similarity have been developed and each is distinct in the way it encapsulates particular sample-wise matrix effects yielding different similarity results per measure including ensembled measures.[73-84] Thus, the current approach to sample similarity is fragmented with each measure only providing a partial view of matrix effect wholeness. Said another way, considering all possible matrix effects as an unbroken whole, then a similarity measure is conceptualized as one particular aspect of the whole.

Accessing sample spectral similarity is further compounded by eq. 1 which specifies that different linear combinations of the basis vectors from **F** can create what appear to be indistinguishable spectra in shape and magnitude with widely varying analyte and other constituent amounts. Thus, similar spectra do not mean equivalent analyte amounts, as well as anything else physically or chemically similar between visually matching spectra. Additionally, it

seems unreasonable to expect that using one or two spectral similarity measures can act as the only viewing portals to capture all the diverse sample-wise matrix effects. Because a spectrum is a complex collection of matrix effects, it only makes sense to apply the TAC and use a multivariate approach for determining similarity and in so doing, leverage the multivariate Rashomon effect. Using a large collection of similarity measures provides a more reliable similarity assessment because many different aspects of the hidden matrix effects can be characterized for a consensus viewpoint of sample similarity. By using an assortment of measures, matrix effects are exploited providing a thorough view of sample heterogeneity improving the chances of accurately assessing how similar two samples are to each other. Applying multiple similarity measures is akin to a person using all their senses to combine multiple information streams ultimately producing contextual generalizations. Using the TAC to advance similarity assessment to be multivariate with multiple measures also fits Svante Wold's point that "The study of complicated systems is by necessity multivariate."[9] However, while multiple similarity views of matrix effects are possible, currently it is difficult to attribute which particular matrix effects are assessed by each similarity measure.

Nonetheless, we have successfully applied the Rashomon effect by implementing hundreds of similarity measures to improve numerous common chemometric similarity applications. The rundown consists of using mid-infrared, NIR, UV, and VIS full spectra and thermogravimetric data for beer product authentication,[85] ICP-MS for validating fava beans from Santorini,[86] mid-infrared to identify adulterated strawberry puree,[87] heavy metal contamination of clams,[86] verification of meat samples as either turkey or chicken or pork,[86] and microplastic classification,[88] thirteen chemical measurements to classify three Italian wine cultivars,[85] restoration of defaced serial numbers using lock-in infrared thermography,[89,90] and outlier

detection.[91] Displayed in Figure 10 are heat maps for outlier detection and multi-class classification.[85,91] The outlier detection situation is determining if any samples in a training set are outliers and the classification example is deciding which of three classes a target sample belongs to. Note that not every similarity measure identifies samples as outliers and vice versa. The zoom in Figure 10 clarifies this observation and the two samples with large similarity values were deemed outliers. Equally, not every similarity measure correctly determines the target sample to be a class 1 member and vice versa in Figure 10. If only one similarity measure is used, samples could be incorrectly assessed. The particular set of similarity measures used in these applications are just one set of many possibilities. The key suggestion here is that matrix effects as a whole are not assessed by one similarity view but it is better realized by multiple views for a more complete portrayal. The success of using multiple similarity measures is a direct application of using the TAC with the Rashomon effect where each fragmented measure presents its objective accounting of the matrix effects.

In addition to using a large collection of measures to assess sample similarity, an unconventional concept we have found extremely useful and included in our studies and Figure 10 is to not optimize tuning based measures. For example, many similarity measures are based on an optimized number of PCs. However, it has been shown that altering calibration samples changes the dataset matrix effect complexity resulting in different regression models. Correspondingly, depending on the training sample-wise matrix effects, different numbers of PCs will be optimized for classification or outlier detection due to the respective covariance structure variants. The classic PC-based measure Mahalanobis distance (MD) can be used to describe the process. Instead of optimizing the number of PCs for one MD value, the MD is obtained for

consecutively increased number of PCs. The stopping point is not critical and to date we have used the 99% rule. Pictured in Figure 11 is a depiction of this windowing process.

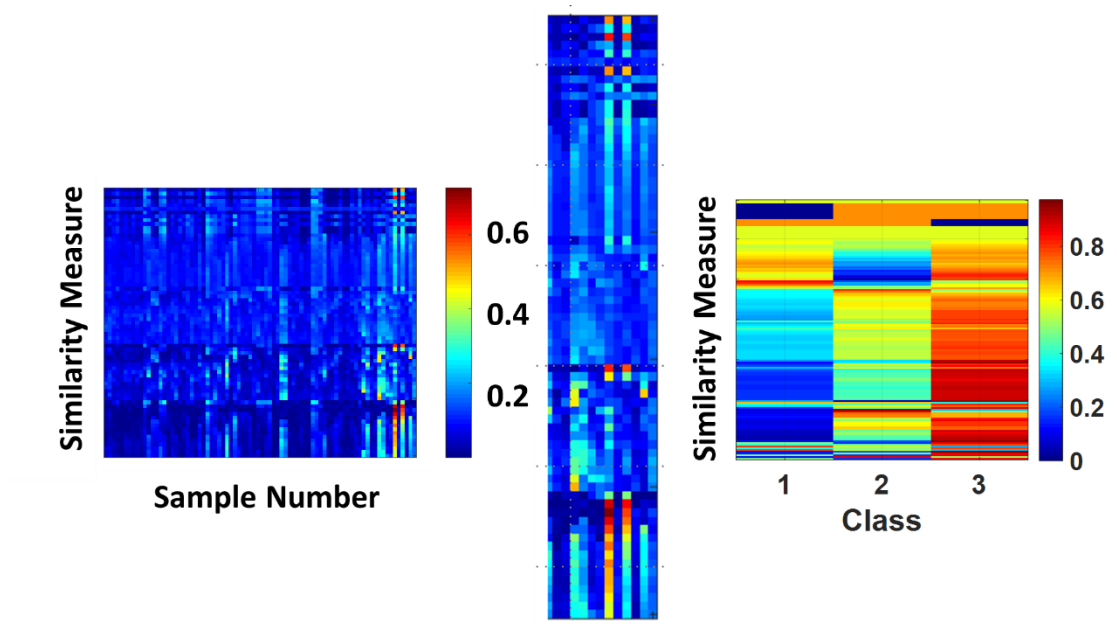

Figure 10. Heat maps of similarity values for outlier detection of each sample in a training set (left), a zoom of potential outlier samples (middle), and similarity values for a target sample being assessed for class membership to one of three classes (right). For both heat maps, color bar values correspond to a fused normalized similarity measures that can be converted to PRISM z-score values. Reprinted with permission from Brownfield B, Kalivas JH. Consensus outlier detection using sum of ranking differences of common and new outlier measures without tuning parameter selections. Anal. Chem. 2017; 89:5087-5094 Copyright 2017 American Chemical Society https://pubs.acs.org/doi/10.1021/acs.analchem.7b00637 and Brownfield B, Lemos T, Kalivas JH Consensus classification using non-optimized classifiers. *Anal. Chem*. 2018; 90:4429-4437. Copyright 2018 American Chemical Society. https://pubs.acs.org/doi/10.1021/acs.analchem.7b04399

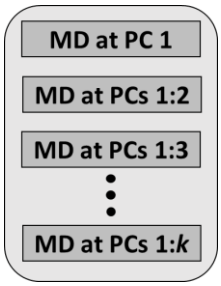

Figure 11. Characterization of a window of MD values for one sample. The value for *k* used to date has been the 99% rule.

Using tuning parameter windows provide a more thorough view of sample similarly relative to the matrix effects captured by the MD. The windowing process is just another multivariate tactic, as are using collections of diverse models and similarity measures, multiple wavelengths, etc. When PCs are optimized to one particular value, only one viewpoint of the MD is gained, i.e., a limited sample matrix effect view. It does not seem reasonable to expect one opinion to fully characterize the intricacies of the target sample relationship to the source sample PC base, let alone allow generalization possibilities.

Complementing the idea of PC windows, we have found it useful to not optimize the number of nearest neighbors nor PLS LVs in PLS discriminant analysis in our classification studies. Instead, we use a family of nearest neighbors and a collection of LVs.[85-90] The classification heat map in Figure 10 contains the respective windows. Equally, we observed it is not useful to optimize the number of PLS LVs for outlier detection by studentized residuals and instead, a window of PLS LV studentized residual values is used.[91] Included in the Figure 10 outlier

detection heat map are PLS LV windows. In another study, the window concept was successfully used in synchronous fluorescence spectra where rather than optimizing the wavelength interval between the excitation and emission wavelengths ($\Delta\lambda$), a collection of $\Delta\lambda$ intervals were used thereby removing the need for an experienced user to select.[92] Not optimizing tuning parameters or pipelines is a direct impact of the TAC with the Rashomon effect where each tuning parameter value such as a fixed number of PCs or LVs only provides a fragmented objective accounting of the information.

The Rashomon effect of multiple similarity measures was recently advanced to form a quantitative composite sample-wise z-score value named the physicochemical responsive integrated similarity measure (PRISM).[20] Two PRISM values are used to establish each final sample PRISM z-score. One is based on directly leveraging spectral matrix effects forming $PRISM_X$ that can be used for numerous sample similarity purposes. The second can be used when regression is involved and, in this case, the matrix effects are indirectly exploited to form the $PRISM_y$ similarity measure by comparing analyte prediction amounts and other prediction reliability measures. Sample PRISM z-scores were used to evaluate model generalizability (prediction reliability) and outlier detection. Recently, $PRISM_X$ was used for one-class classification producing a z-score and conformal prediction values allowing probability of classification to be assessed.[93]

Important to the success of PRISM is that sample-wise differences for respective similarity measures are used, another non-typical data analysis approach. Using sample-wise differences of similarities relative to a base, such as a mean spectrum, permits a more detailed comparison of respective matrix effects between two particular objects.[20,94-97]

Another distinctive aspect of PRISM is that some of its similarity measures can be thought of as assessing how much work is being performed in the comparison operation. For example, rather than using extended inverted signal correction (EISC) to preprocess spectra, we use the difference between magnitudes of respective sample transfer coefficients to assess how hard it is to transform a spectrum to appear similar to another spectrum.[98] The same work concept can be applied to Procrustes analysis similarity measures. For example, if large difference exists between respective coefficients regardless of the fact that the transfer fit error is small, then the two samples would be considered poorly matrix matched because of the difficulty in the mathematical transformation to obtain the small residual fit error.

The $PRISM_X$ and $PRISM_y$ values are instrumental in a new autonomous local modeling approach named local adaptive fusion regression (LAFR).[99] This approach is able to mine a library of spectra with analyte reference values and ultimately select a local calibration set that not only closely span a target sample spectrum but the calibration reference values also closely span the unknown target sample analyte amount. Shown in Figure 12 are boxplots of the LAFR calibration reference values for 30 randomly selected target samples from a library with 1172 samples. Samples are cattle feces from North Australia over a 10-year collection with 3 sampling methods measured in the NIR for analyte percent crude protein content. Note the small interval of LAFR calibration reference analyte amounts spanning each target sample analyte amount. Prediction errors (RMSEPs) for the 30 target samples are respectively 1.26 and 0.73 using the full library (global, protein ranging from 0 to 26%) and LAFR local models showing significant improvement. The corresponding R^2 values are 0.87 and 0.94. The largest library tested with LAFR is library of nearly 100,000 NIR spectra on soil samples covering a diverse set of sail samples across the United States.[99]

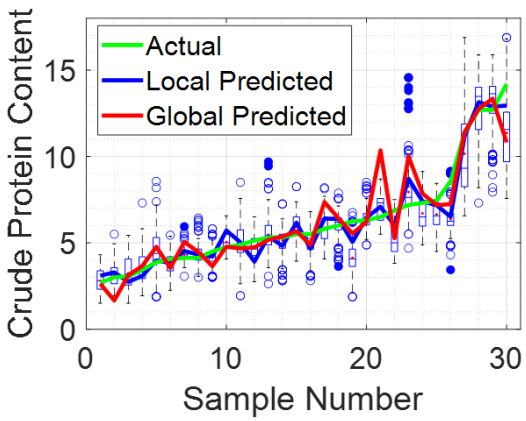

Figure 12. Boxplots of local modeling method LAFR calibration reference samples for respective target sample analyte values (green) and predicted amounts by LAFR (blue) and the global model (red).

Summarizing, we hope we have provided the reader with a new perspective on assessing similarity to consider the next time samples or datasets are being compared. Four key points to consider are: the more similarity measures used, the greater the comprehensive viewpoint obtained (TAC and Rashomon effect); sample-wise differences of similarity measures are more informative than similarity values; more consistent comparisons are obtained by using windows of tuning parameters values or pipelines rather than optimizing values (TAC and Rashomon effect); and lastly, when performing a similarity comparison, think about assessing similarity by how hard it is transform a sample to match another sample even though the fit error is small after the transformation. Applications of these four points are the reasons for the success of new methods described above. We think taking this multivariate approach to sample similarity promotes assurance in obtaining chemically relevant information.[9]

7 CONCLUSIONS

We have presented some interesting chemometric directions to consider in fundamental and applied studies advocating for a deeper thinking on matrix effect intricacies relative to what is actually being captured by a model or similarity measure. Arguments are offered to apply the TAC with the Rashomon effect by expanding multivariate data analysis beyond basic multi-way processes to include higher order approaches founded on using multiple diverse models, pipelines, and similarity measures. Each data view acts as its own microscope to probe samples for hidden information shedding light on the underlying matrix effects.

Our key take home messages are: Occam's razor may not always be applicable (depends on the evidence from the chosen point of view); the curse of dimensionality is not necessarily a curse; the TAC and Rashomon effect for different data viewpoints are advantageous; chemometric models cannot be explained; sample-wise matrix effects should be leveraged in a positive way but are too complex to individually isolate.

The TAC and Rashomon effect were recently paralleled to physicist David Bohm's concepts explicate and implicate orders by relating their communalities with the two orders.[100] These orders are used by David Bohm to explain quantum mechanics and time in addition to linking the orders to neuropsychology and the philosophy of the mind.

Lastly, Svante Wold noted "… in chemometrics the main issue is to structure the chemical problem to a form that can be expressed as a mathematical relation." [9] We interpret this statement to imply that one needs to first think about a problem or question and how to solve or answer the situation and then develop the mathematics. This approach is akin to these Einstein quotes "I believe in intuitions and inspirations. I sometimes feel that I am right. I do not know that I am…

[but] I would have been surprised if I had been wrong." and "I am enough of the artist to draw freely upon my imagination. Imagination is more important than knowledge. Knowledge is limited. Imagination encircles the world."[101]

REFERENCES

1. Booksh KS, Kowalski BR. Theory of analytical chemistry." *Anal. Chem.* 1994;15:782A-791A.
2. Karimvanda SK, Kalivas JH, Abdollahia H. “Perspective on the Theory of Analytical Chemistry.” Submitted to *J. Chemom*. 2025.
3. Geladi P, Swerts J, Lindgren F. Multiwavelength microscopic image analysis of a piece of painted chinaware: classification and regression. *Chemom. Intell. Lab. Syst*. 1994;24:145-167.
4. Geladi P. The regression model comparison plot (REMOCOP). In Frontiers in Analytical Spectroscopy, Andrews DL, Davies AMC (Eds.), Proceedings of “Spectroscopy Across the Spectrum IV”, Norwich, July 11–14 1994, Royal Society of Chemistry, 1995.
5. Breiman L. Statistical modeling: two cultures. *Stat. Sci.* 2001:**16**:199-231.
6. D’Amour A. Revisiting Rashomon: a comment on the “The Two Cultures”. *Obs. Studies.* 2021;**7.1**:59-63.
7. Semenova L, Rudin C. A study in Rashomon curves and volume: a new perspective on generalization and model simplicity in machine learning. 2021. https://arxiv.org/abs/1908.01755v3 (assessed January 2025).
8. Brereton RG, Jansen J, Lopes J, Marini F, Pomerantsev A, Rodionova O, Roger JM, Walczak B, Tauler R. Chemometrics in analytical chemistry – part I: history, experimental design and data analysis tools. *Anal. Bioanal. Chem.* 2017;409:5891-5899.

9. Wold S. Chemometrics: what do we mean with it, and what do we want from it? *Chemom. Intell. Lab. Sys.* 1995;**30**:109-115.
10. Peper JMJ, Kalivas JH. Redefining spectral data analysis with immersive analytics: exploring domain shifted model spaces for optimal model selection. *Appl. Spectrosc.* 2024;in press.
11. Kalivas JH. Model regression vectors v2 voice, https://youtu.be/6OcYhc3iI7Y (assessed 2024).
12. Kalivas JH. MDPS VR v3 voice, https://youtu.be/jI9F4CSUSxo (assessed January 2025).
13. Kalivas JH. Outlier detection movie voice. https://www.youtube.com/watch?v=qKBKqvmPsnU (assessed January 2025).
14. Russel SJ, Norvig P. Artificial Intelligence: A Modern Approach (4th ed.). Hoboken: Pearson, 2021.
15. Harris DC, Lucy CA. Quantitative Chemical Analysis (10th ed.). W.H. Freem Macmillan Learning: New York, 2020.
16. Noad HM. Chemical Analysis, Qualitative and Quantitative. Lindsay and Blakiston, Philadelphia, 1849.
17. Miller CE. Chemical principles of near-infrared technology. In Near-Infrared Technology in the Agriculture and Food Industries 2nd ed., P. Williams and K. Norris (Eds.), 2001.
18. Williams PC, Norris K. Variables affecting near-infrared spectroscopic analysis. In Near-Infrared Technology in the Agriculture and Food Industries 2nd ed., P. Williams and K. Norris (Eds.), 2001.
19. Bissantz C, Kuhn B, Stahl M. A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* 2010; **53**:5061-5084.

20. Spiers, RC, Norby C, Kalivas JH. Physicochemical responsive integrated similarity measure (PRISM) for a comprehensive quantitative perspective of sample similarity dynamically assessed with NIR Spectra. *Anal. Chem.* 2023; **95**:12776-12784.

21. Krzysztof BB, Grabska J, Kirchler CG, Huck CW. NIR spectra simulation of thymol for better understanding of the spectra forming factor, phase, and concentration effects and PLS regression features. *J. Mol. Liq.* 2018;**268**:895-902.

22. Grabska J, Krzysztof BB, Mayr S, Huck CW. Theoretical simulation of near-infrared spectrum of Piperine: insight into band origins and the features of regression models. *Appl. Spectrosc.* 2021;**75**:1022-1032.

23. Kalivas JH. Basis sets for multivariate regression. *Anal. Chim. Acta*. 2001;**428**:31-40.

24. Kurosawa A.Rashomon. Tokyo, Daici, 1950.

25. Kalivas JH. Interrelationships of multivariate regression methods using eigenvectors basis sets. *J. Chemom.* 1999;**13**:111-132.

26. Hansen PC. Rank-deficient and Discrete Ill-posed Problems: Numerical Aspects of a Linear Inversion. SIAM Press: Philadelphia, PA, 1998.

27. Andries E, Kalivas JH. Multivariate calibration leverages and spectral F-ratios via filter factor representation. *J. Chemom.* 2010;**24**:249-260.

28. Liu R, Wang H, Glover KP, Feasel MG, Wallqvist A. Dissecting Machine-learning prediction of molecular activity: is an applicability domain needed for quantitative-structure-activity relationship models based on deep neural networks? *J. Chem. Info. Model.* 2019;**59**:117-126.

29. Lipton ZC. The mythos of model interpretability. 2017. https://arxiv.org/abs/1606.03490 (accessed January 2025).

30. Rudin C, Carlson D. The secrets of machine learning: ten things you wish you had known earlier to be more effective at data analysis. 2019. https://doi.org/10.48550/arXiv.1906.01998 (assessed January 2025).
31. Gall R. Explainability vs Interpretability: Two concepts that could help restore trust in AI. 2018. https://www.kdnuggets.com/2018/12/machine-learning-explainability-interpretability-ai.html (assessed January 2025).
32. Ferrini M. Shall we build transparent models right away? Towards Data Science. 2019. https://towardsdatascience.com/shall-we-build-transparent-models-right-away-196db0eeba6c (assessed January 2025).
33. Lador SM. Navigating the sea of explainability. Towards Data Science. 2019. https://towardsdatascience.com/navigating-the-sea-of-explainability-649672aa7bdd
34. Link-Gelles R, Ciesla AA, Mak J, Miller JD, Silk BJ, Lambrou AS, Paden CR, Shirk P, Britton A, Smith ZR, Fleming-Dutra KE. *MMWR*. 2024;**173(4)**,77-83. https://www.cdc.gov/mmwr/volumes/73/wr/mm7304a2.htm?s_cid=mm7304a2_w (assessed January 2025).
35. Chen C, Li O, Tao C, Barnett AJ, Su J, Rudin C. This looks like that: deep learning for interpretable image recognition. 2019. https://doi.org/10.48550/arXiv.1806.10574 (assessed January 2025).
36. Zhou B, Kholsa A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2921-2929. IEEE Computer Society. 2016.

37. Selvaraju R, Cogswell M, Das Abhishek D, Vedantam R, Parikh D, Barta D. Grad-CAM: visual explanations form deep networks via gradient-based localization. *Int. J. Comp. Vis.* 2020;**128**,336-359.

38. Liu Z, Li J, Li J, Yang T, Zhang Z, Wu H, Xu H, Meng J, Li F. Explainable deep-learning-assisted sweat assessment via a programable colorimetric chip. *Anal. Chem.* 2022;**94**:15864-15872.

39. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Mach. Intell.* 2019;**1**:206-215.

40. Rudin C, Chen C, Chen Z, Huang H, Semenova L, Zhong C. Interpretable machine learning: fundamental principles and 10 grand challenges. *Stat. Survey.* 2022;**16**:1-85. https://arxiv.org/abs/2103.11251 (assessed January 2025).

41. Zhang C, Bengio S, Hardt M, Recht B, Vinyals O. Understanding deep learning requires rethinking generalization. 2017. https://doi.org/10.48550/arXiv.1611.03530 (assessed January 2025).

42. T. Poggio T, Kawaguchi K, Liaoy Q, Miranda B, Rosasco L, Boix X, Hidary J, Mhaskar H. Theory of deep learning III: explaining the non-overfitting puzzle. 2018. https://doi.org/10.48550/arXiv.1801.00173 (assessed January 2025).

43. Belkin M, Hsu D, Ma S, Mandal S. Reconciling modern machine learning practice and the bias-variance trade-off. *PNAS*. 2019;**16**:15849-15854.

44. Kawaguchi K, Kaelbling LP, Bengio Y. Generalization in deep learning. In Mathematical Aspects of Deep Learning. Cambridge University Press, 2020. https://doi.org/10.1017/9781009025096.003 (assessed January 2025).

45. Schaeer R, Khona M, Robertson Z, Boopathy A, Pistunova K, Rocks JW, Fiete IR, Koyejo O. Double descent demystied: identifying, interpreting & ablating the sources of a deep learning puzzle.2023. https://doi.org/10.48550/arXiv.2303.14151 (assessed January 2025).

46. Curth A. Classical statistical (in-sample) intuitions don't generalize well: a note in bias-variance tradeoffs, overfitting and moving from fixed to random designs. https://arxiv.org/abs/2409.18842 (assessed January 2025).

47. Farber NKM. A closer look at the bias-variance trade-off in multivariate calibration. *J. Chemom.* 1999;**13**:185-199.

48. Marbach R. On Wiener filtering and the physics behind statistical modeling. *J. Biomed. Opt.* 2002;**7**:130-147.

49. Kalivas JH. Overview of two-norm (L2) and one-norm (L1) Tikhonov regularization variants for full wavelength or sparse spectral multivariate calibration models or maintenance. *J Chemom.* 2012;**26**:218-230.

50. Gurden SP, Westerhuis JA, Bijlsma S, Smilde A. Modelling of spectroscopic batch process data using grey models to incorporate external information. *J. Chemom.* 2001;**15**:101-121.

51. Kiers HAL, Timmerman ME. Some views on multi-criteria methods for data analysis. *J. Chemom.* 2024;0:e3597.

52. Kalivas JH, Ferré J, Tencate AJ. Selectivity-relaxed classical and inverse least squares calibration and selectivity measures with a unified selectivity coefficient. *J. Chemom.* 2017;**31**:2925.

53. Palmer J, Kalivas JH. Characterizing multivariate calibration tradeoffs (bias, variance, selectivity, and sensitivity) to select model tuning parameters. *J. Chemom.* 2014;**28**:347-357.

54. Skou PB, Hosseini E, Ghasemi JB, Smilde AK, Eskildsen CE. Orthogonality constrained inverse regression to improve model selectivity and analyte predictions from vibrational spectroscopic measurements. *Anal. Chim. Acta.* 2021;**1185**:339073.

55. Dong J, Rudin C. Exploring the cloud of variable importance for the set of all good models. *Nature Mach. Intell.* 2020;**2**:810-824.

56. Spiers RC, Kalivas JH. Reliable model selection without reference values by utilizing model diversity with prediction similarity. *J. Chem. Inf. Model*. 2021;**61**:2220-2230.

57. Andries E, Kalivas JH. Interrelationships between generalized Tikhonov regularization, generalized net analyte signal, generalized least squares for desensitizing a multivariate calibration to interferences. *J. Chemom.* 2013;**27**:126-140.

58. Shahbazikhah P, Kalvas JH, Andries E, O'Loughlin T. Using the L_1 norm to select basis vectors for multivariate calibration and calibration updating. *J. Chemom.* 2016;**30**:109-120.

59. Leamer E, Leonard H. Reporting the fragility of regression analysis. *Rev. Econ. Stat.*1983;**65**:306-317.

60. Thompson GD. Choice of flexable functional forms: review and appraisal. *West. J. Agricul. Econ.* 1988;**13**:169-183.

61. Mountain D, Hsiao C. A combined structural and flexable functional approach for modeling energy substitution. *J. Am. Stat. Assoc.*1989;**84**:76-87.

62. Griliches Z. Distributed lags: a survey. *Econom.* 1967;**35**:16-49.

63. Brown CD. Discordance between net analyte signal theory and practical multivariate calibration. *Anal. Chem.* 2004;**76**:4364-4373.

64. Brown CD, Green RL. Critical factors limiting the interpretation of regression vectors in multivariate calibration. *TRAC*. 2009;**28**:506-514.

65. Marx CT, Calmon F du PC, Ustun B. Predictive multiplicity in classification. 2020. https://doi.org/10.48550/arXiv.1909.06677 (assessed January 2025).

66. Ni W, Brown SD, Man R. Stacked partial least squares regression for spectral calibration and prediction. *J. Chemom.* 2009;**23**:505-517.

67. Miranda-Quintana RA, Kim TD, Heider-Zadeh F, Ayers PW. On the impossibility of unambiguously selecting the best model for fitting data. *J. Math. Chem.* 2019;**57**:1755-1769.

68. Peat FD. Synchronicity: The Bridge Between Matter and Mind. Bantam, New York, NY, 1987.

69. Box D (Sir). Comment on "Model uncertainty, data mining and statistical inference". *J. Roy. Stat. Soc., Seriea A*. 1995;**158**:455–456.

70. Truran P. Models: useful but not true in practical applications of the philosophy of science, Springer Briefs in Philosophy, Springer, 2013 pp. 61–67. https://link.springer.com/chapter/10.1007/978-3-319-00452-5_10

71. Zeaiter M, Roger J-M, Bellon-Maurel V. Robustness of models developed by multivariate calibration. Part II: the influence of pre-processing methods. *TRAC*. 2005;**24**:437-445.

72. Cox DR. Comment on "Model uncertainty, data mining and statistical inference *J. Roy. Stat. Soc., Seriea A*. 1995;**158**:455–456.

73. Goodman N. Seven Structures on Similarity in Problems and Projects. Bobbs-Merri: New York, 1972.

74. Nikolova N, Jaworska J. Approaches to measure chemical similarity-a review. *QSAR Comb. Sci*. 2003;**22**:1006-1026.

75. Todeschini R, Ballabio D, Consonni V, Grisoni F. A new concept of higher-order similarity and the role of distance/similarity measures in local classification methods. *Chemom. Intell. Lab. Syst*. 2016;**157**:50-57.

76. Todeschini R, Ballabio D, Consonni V. Distances and Similarity Measures in Chemometrics and Chemoinformatics. In Encyclopedia of Analytical Chemistry, R.A. Meyers (Ed.), 2020.

77. Brereton RG, Lloyd GR. Re-evaluating the role of the Mahalanobis distance measure. *J. Chemom*. 2016;**30**:134-143.

78. Joun-Rimbaus D, Massart DL, Saby CA, Puel C. Determination of the representativity between two multidimensional data sets by a comparison of their structure. *Chemom. Intell. Lab. Syst*. 1998;**40**:129-144.

79. Joun-Rimbaus D, Massart DL, Saby CA, Puel C. Characterization of the representivity of selected sets of samples in multivariate calibration and pattern recognition. *Anal. Chim. Acta*. 1997;**350**:149-161.

80. Wold S, Søstrøm M. SIMCA. A Method for Analyzing Chemical Data in Terms of Similarity and Analogy. In Chemometrics: Theory and Publications, B.R. Kowalski (Ed.), American Chemical Society: Washington DC, USA, 1977; 243–282.

81. Samuel AZ, Mukojima R, Horii S, Ando M, Egashira S, Nakashima T, Iwatsuki M, Takyeama H. On selecting a suitable spectral matching method for automated analytical applications of Raman spectroscopy. *ACS Omega*. 2021;**6**:2060-2065.

82. Wale N, Watson IA, Karypis G. Indirect similarity based methods for effective scaffold - hopping in chemical compounds. *J. Chem. Inf. Model*. 2008;**48**:730-741.

83. Maggiora GM, Shanmugasundaram V. Molecularity similarity measures in Bajorth J. (ed.) Chemoinformatics and Computational Chemical Biology, Methods in Molecular Biology, vol. 6782, Springer Science+Business Media, 2011.

84. Medina-Franco JL, Maggiora GM, Giulianotti MA, Pinilla C, Houghton RA. A similarity-based data-fusion and comparison of compound databases. *Chem. Biol. Drug Des.* 2007;70:393-412.

85. Brownfield B, Lemos T, Kalivas JH Consensus classification using non-optimized classifiers. *Anal. Chem*. 2018; **90**:4429-4437.

86. Kalivas JH, Lemos T. Automatic food and beverage authentication and adulteration detection by classification hybrid fusion. *J. Chemom*. 2023;**37**:e3371.

87. Lemos T, Kalivas JH. Self-optimized one-class classification using sum of ranking differences combined with a receiver operator characteristic curve. *Anal. Chem*. 2020; **92**:5354-5361.

88. Chabuka BK, Kalivas JH. Application of hybrid fusion classification process for identification of microplastics based on Fourier transform infrared spectroscopy. *Appl. Spectrosc.* 2020;**74**:1167-1183.

89. Unobe I, Lau L, Kalivas J, Rodriguez R, Sorensen A, Restoration of defaced serial numbers using lock-in infrared thermography (Part I). *J. Spectr. Imaging*. 2019;**8**:Article ID a19.

90. Unobe I, Lau L, Kalivas J, Rodriguez R, Sorensen A. Restoration of defaced serial numbers using lock-in infrared thermography (Part II). *J. Spectr. Imaging*. 2019;**8**:Article ID a20.

91. Brownfield B, Kalivas JH. Consensus outlier detection using sum of ranking differences of common and new outlier measures without tuning parameter selections. *Anal. Chem.* 2017;**89**:5087-5094.

92. Stokes TD, Foteini M, Brownfield B, Kalivas JH, Mousdis G, Amine A, Georgiou C. Feasibility assessment of synchronous fluorescence spectral fusion by application to Argan oil for adulteration analysis. *Appl. Spectrosc.* 2018;**72**:432-441.

93. Redd HJ, Kalivas JH. Assessment of conformal prediction and standard normal distribution for autonomous consensus one-class classification. in press *J. Chemom.* 2024.

94. Sheridan RP, Hunt P, Culberson JC. Molecular transformation as a way of finding and exploiting consistent local QSAR. *J. Chem. Inf. Model*. 2006;**46**:180-192.

95. Dossetter AG, Griffen J, Leach A, Matched molecular pair analysis in drug discovery. *Drug Discovery Today.* 2013;**18**:724-731.

96. Kramer C, Fuchs JE, Whitebread S, Gedek P, Liedl KR. Matched molecular pair analysis: significance and impact of experimental uncertainty. *J. Med. Chem*. 2014; **57**:3786-3802.

97. Tynes M, Gao W, Burrill DJ, Batista ER, Perez D, Yang P, Lubbers N. Pairwise difference regression: a machine learning meta-algorithm for improved prediction and uncertainty quantification in chemical search. *J. Chem. Inf. Model*., 2021;**61**:3846-3857.

98. J. Ottaway J, Kalivas JH. Feasibility study to transform spectral and instrumental artifacts for multivariate calibration maintenance. *Appl. Spectrosc*. 2015;**69**:407-416.

99. Spiers R, Kalivas JH. Local adaptive fusion regression (LAFR) for local linear multivariate calibration: Application to large datasets. in press *Appl. Spectrosc.* https://doi.org/10.1177/00037028241308538;2025.

100. Kalivas JH. Two-fold perspective: chemometrics, ML, and AI are not synonyms and gaining greater multivariate advantages in modeling and data analysis. *Appl. Spectrosc.* 2025;https://doi.org/10.1177/00037028251330324

101. Viereck, GS. Saturday Evening Post, 1929.