The Reproducibility Crisis, a Comprehensive Set of Guides on XPS, and Better Data Fitting/Chemometrics of XPS Data

Tahereh G. Avval[1], George H. Major[1], Matthew R. Linford[1]

[1]Brigham Young University, Department of Chemistry and Biochemistry, Provo, UT, USA.

There is a reproducibility crisis in science that includes material characterization; the data and data work-up of multiple characterization methods are often reported poorly or incompletely in the scientific literature. For example, some of us recently found that approximately one-third of the X-ray photoelectron spectroscopy (XPS) data reported in the literature are seriously flawed.[1, 2] This value is in line with findings for other material characterization methods. In this talk, we will discuss the magnitude of the reproducibility problem, including some of the common errors that are made in XPS data analysis/peak fitting. In response to the reproducibility crisis, a group of XPS experts (including us) recently wrote a series of guides on various aspects of the technique.[3-5] These guides have been extensively downloaded and are being cited well. We will discuss these particular guides, including their structure, scope, and topics, as well as another set of guides that is currently being written.

Peak fitting is an important part of XPS data analysis. Peak fitting is the way in which chemical states are typically identified and quantified in XPS. Indeed, peak fitting is an integral part of much XPS data analysis because the widths of the signals in XPS are often comparable to their energy shifts (chemical shifts), i.e., signal overlap is very common in XPS. In this talk, we will touch on some of our latest results on peak fitting XPS data. For example, we will discuss the importance of applying appropriate constraints in XPS peak fitting.
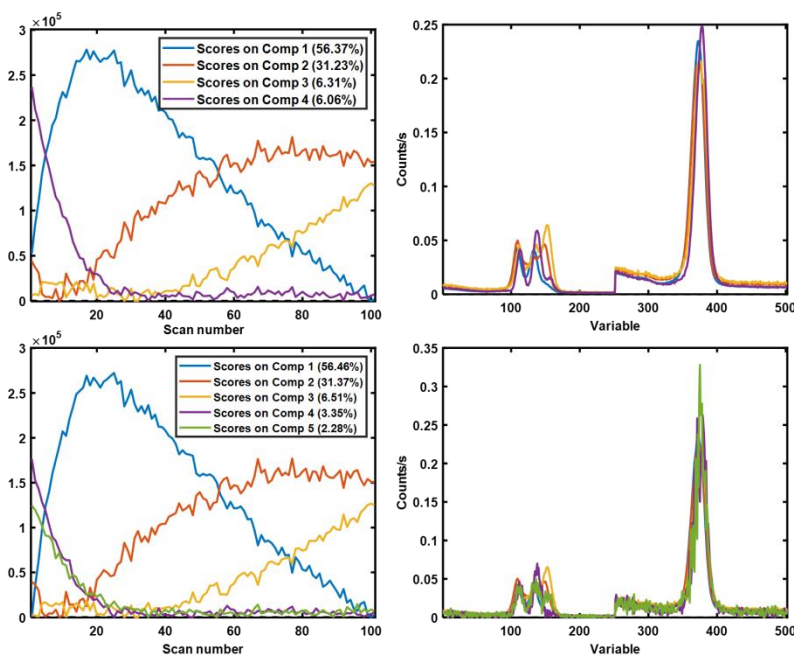

Figure 1. MCR scores and loadings (components) of four- and five-component models of the concatenated data (combined C 1s and O 1s spectra) from tartaric acid.

XPS instruments increasingly collect large numbers of related spectra, which may be generated in, for example, *operando* studies, degradation studies, or XPS imaging. The standard problem associated with dealing with large numbers of spectra is in interpreting/analyzing all of them. That is, it is not too hard to visually inspect and compare a handful of spectra, but visual inspection and comparison become quite difficult as the number of spectra rises. Chemometrics/informatics methods have been widely used to simultaneously analyze large numbers of spectra. We believe that these methods will be increasingly important in correct, informed XPS data analysis. That is, if the XPS peak

fitting and data analysis in the literature are already quite problematic, what will happen as even larger data sets are collected? As a possible answer to this question, we will describe in this talk the chemometrics analyses of two moderately large data sets obtained during the XPS analyses of cellulose and tartaric acid, where the XPS analyses themselves caused the materials to degrade. Indeed, while many materials do not degrade to a significant extent during XPS analysis, e.g., many inorganic materials, some do. Summary statistics were used to reveal the overall structures of the data sets, e.g., the presence of outliers or a discontinuity in the spectra. PCA was used to determine the number of underlying factors in them. MCR was particularly helpful in revealing the chemical evolution of the materials, where MCR nicely decomposed the data into factors that could be assigned to initial states, final states, and intermediates in the degradation processes. In both cases, degradation of the oxygenated hydrocarbons appeared to cause them to carbonize. MCR factors/components were peak fit, which further revealed their underlying chemical compositions. Cluster analysis was also performed on these data sets. It showed clustering of consecutive, neighboring spectra. The average spectra corresponding to these clusters provide a nice, abbreviated summary of the chemical evolution of the materials. Figures 1 and 2 show representative MCR and cluster analyses of the XPS data sets we analyzed.
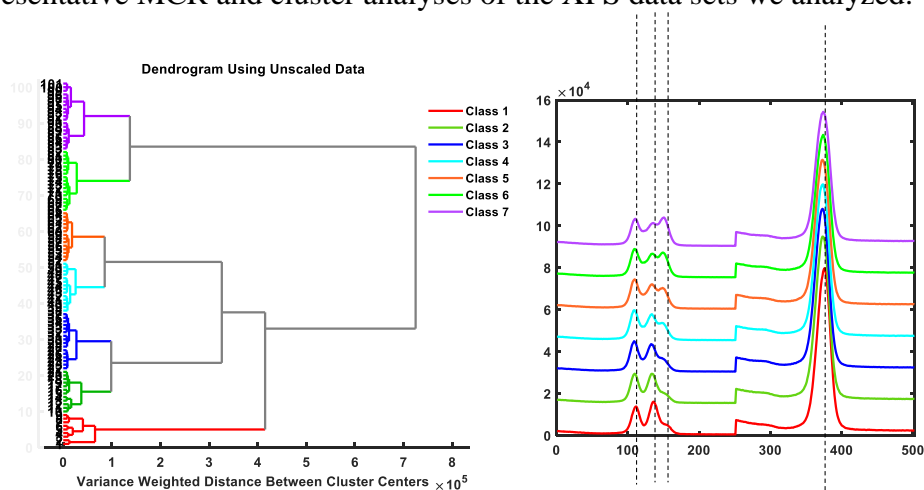

Figure 2. Cluster analysis of the concatenated data (combined C 1s and O 1s spectra) from tartaric acid.

References:

[1] Major GH et al., Journal of Vacuum Science & Technology A. 2020; 38:061204.
[2] Linford MR et al., Microscopy and Microanalysis. 2020; 26:1-2.
[3] Gengenbach TR et al., Journal of Vacuum Science & Technology A. 2021; 39:013204.
[4] Baer DR et al., Journal of Vacuum Science & Technology A. 2019; 37:031401.
[5] Major GH et al., Journal of Vacuum Science & Technology A. 2020; 38:061203.