This was originally published in Spectroscopy, 14(3), p. 12-13 (1999)

Statistics in Spectroscopy Part 32 - Linearity in Calibration - Act II Scene IV

by H. Mark and J. Workman

We continue our discussion started by the responses received to our column “Linearity in Calibration”(1). So far our discussion has extended over three previous columns(2-4).

In our last column(4) we stated: “we are not fighting a "holy war" against PCA/PLS etc.” and then went on to discuss what our original column was really about.

However, if there is a "holy war" being fought at all, then from our point of view it is against the practice of simply accepting the results of the computer's cogitations without attempting to understand the underlying phenomena that affect the behavior of the calibration models, regardless of the algorithm used. This has been our fight since the beginning - which can be verified by going back and rereading our very first column ever (5).

Jerry and I don’t always agree, but we do agree on the following: it is incomprehensible how a person calling himself a scientist can fail to wonder WHY calibration models behave the way they do, and try to relate their behavior to the properties of the data giving rise to them. There are reasons for everything that happens, whether we know what those reasons are or not, and the goal of science is to determine what those underlying reasons or principles are. At least that is the goal of every other field of scientific endeavor that we are aware of - why is Chemometrics exempt?

Real data, as we have seen, is far too complicated to work with to try to obtain fundamental understanding, just as the physical world is often too complicated to study directly in toto. Therefore work such as was presented in the "Linearity in Calibration" column is needed, creating a simplified system where the characteristic of interest can be isolated and studied - just as physical experiments often work with a simplified portion of the physical world for the same reason. This might be categorized as "Experimental Chemometrics": controlling the nature of the data in a way that allows us to relate the properties of the data to the behavior of the model. Does this mimic the "real world"? No, but it does provide a window into the inner workings of the calibration calculations, and we need as many such windows as we can get.

We will go so far as to make an analogy with Chemistry itself. The alchemists of old had an enormous empirical knowledge base, and from that could do all manner of useful things. But we do not consider alchemy a science, and it did not become a science until the underlying principles and phenomena were discovered and codified in a way that all could use. The current state of Chemometrics is more nearly akin to alchemy than Chemistry: we can do all manner of useful things with it, but it is all empirical and there are still many areas where even the most expert and prominent practitioners treat it as a "black box" and make no attempt to understand the inner workings of that black box. Empiricism is important and even necessary, but hardly

sufficient. The ultimate test of whether something is scientific is its ability to predict - and that does NOT mean SEP!!

The irony of the situation is that a good deal of basic knowledge is available. The field of Chemometrics bypasses all the Statistical basics and jumps right into the heavy-duty sophisticated algorithms: everybody just wants to start running before they can even crawl. We commented on this situation in an earlier column(6), and what response we received was on the order of “Why was so much space wasted before getting to the important part?” It is certainly unfortunate that the portion of the discussion that was perceived as “wasted space” was the important part, but was not recognized as such.

The early foundations of Statistics go back to the 1600's or so, to the time when probability theory was recognized as a distinct branch of mathematics. The current problem is that nobody currently seems to apply the knowledge gained over the intervening span of time, or to be interested in applying that knowledge, or to do fundamental investigations at all. The chemometric community completely ignores the previous mathematical basis underlying its structure. The science of Statistics does, in fact, form a firm foundation that Chemometrics is built on. It is almost shameful that he modern Chemometrics community seems to be content to build ever higher and fancier superstructures on a foundation that is solid enough, but to which it is hardly connected.

Worse, there seems to be an active antipathy to such investigations: just look at the firestorm we aroused by publishing a very small and innocuous study of the fundamental behavior of a particular data system! In fact, from the response, you'd almost think we committed heresy or attacked religious beliefs, in daring to suggest that PCR/PLS was not always the best way to go, much less do some serious research on the subject.

Everybody gives lip service to the concept of "fundamental research is good for the long run", but nobody seems interested in putting that concept into practice, even with the possibility of fairly short-term returns. Let us look at a couple of examples.

We have recently had the pleasure of reading Richard Kramer’s new book(7) and found the following passage:

> “But, it would be dangerous to assume that we can routinely get away with extrapolation of that kind. Sometimes it can be done, sometimes it can’t. There is no simple rule that can tell us which situation we might be facing.” (see page 129 in (7)).

and that passage seems to sum up the current state of affairs. Theoretically, a good straight line should be extrapolatable almost indefinitely, yet we all know how risky it is to extrapolate even a little bit beyond the range of our data. Why doesn’t practice conform to theory? The obvious answer is that something is non-linear. But why can’t we detect this? As Rich says, we don’t have any simple rules. Well, OK, so we don’t have simple rules. Maybe no simple rules exist. But then, why don’t we at least have complicated rules to help us make such important decisions? At least then we’d have *a*

way to predict (in the scientific sense) something that is worthwhile knowing. As it stands we have nothing, and nobody seems interested in finding out why.

Maybe a new approach is needed. Maybe this is where Fred Cahn's work is pertinent: if you can approximate the non-linearity with a Taylor series, then maybe the quality of the fit can provide a diagnostic to form the foundation of a rule on which to base a decision. Maybe something else will work. We don't know, but it's a possible starting point. Fred, you're in the ideal position to pursue this, how about it - will you accept this challenge?

The above example, of course, is relatively abstract and "academic", and as such perhaps not of too much interest to the majority. Another example, with more practical application, is transfer of calibration models from one instrument to another. This is an endeavor of enormous current practical importance. Witness that hardly a month passes without at least one article on that topic in one or more of the analytical or spectroscopic journals. Yet all those reports are the same: "Effect of Data Treatment ABC Combined with Algorithm XYZ Compared to Algorithm UVW" or some such: they are all completely empirical studies. In themselves there is nothing wrong with that. The problem is that there is nothing else. There are no critical reviews summarizing all this work and extracting those aspects that are common and beneficial (or common and harmful, for that matter).

Even worse, there are no fundamental studies dealing with the relationship of the algorithm's behavior to the underlying physics, chemistry, mathematics or instrumental effects. It is not difficult to see that the calibration transfer problem breaks down into two pieces:

A) The effect of instrumental variation on the data
B) The effect of variations of the data on the model

Studying the effects of instrumental performance should be the province of the manufacturers. Unfortunately, the perception is that it is to their benefit to release such results only if they turn out to be "good", and there is little incentive for them to perform studies whose only purpose is to increase scientific knowledge. Thus it is up to academia to pick up this particular ball, if there is any interest in it at all.

Fundamental studies in those areas will eventually give rise to real knowledge about how and when calibrations can be transferred, and provide us with trustworthy recipes for doing the transfer. Such knowledge will also provide us with the confidence of knowing that the underlying science is sound, and thus take us beyond the "my algorithm is better than your algorithm" stage that we are now at.

Furthermore, true fundamental understanding could also be applied in reverse. Then instrument manufacturers could concentrate on those aspects of construction and operation that affect the transferability situation, and be able to verify their capabilities in an unambiguous, scientifically valid and agreed-on manner.

This is just one other example of a current problem that COULD be attacked with fundamental studies, with both short and long-term benefits that are obvious to all.

Connecting to the statistical foundations, as described above, can have other benefits. For example, computing an SEP on a validation set of data is considered the be-all and end-all of calibration diagnostics. This is an important calculation, to be sure, but it has its limitations, as well. For example, the SEP alone has no diagnostic capability: it tells you nothing about what you need to do in order to improve a calibration model. For another, even when you compare SEPs from different models and choose the model with the smallest SEP, that does not necessarily mean you are choosing the best model. We often see “robustness” bandied about in discussions of calibration models, but what diagnostics do we have to quantify “robustness”? Without such a diagnostic, how can we expect to evaluate “robustness” either in isolation or to compare with SEP?

By putting all our attention on the SEP we have also lost the ability to evaluate calibrations on their own. When calibrating spectrometers to do quantitative analysis, where samples are cheap and easy to come by, this loss is not too serious, but what do you do when a project requires calibration runs that cost a million (or ten million) dollars per run, and minimizing the number of runs is the absolute top priority? In such a case, you will not only not have validation data, you will likely not even have enough calibration data to do a leave-one-out calculation, and then being able to evaluate models from calibration diagnostics alone will be critical. Statisticians have, in fact, developed diagnostic tests that provide information about such characteristics, but the Chemometric community, in our arrogance, think we know better and ignore all this prior work. The statistical community has also developed many local and semi-local diagnostic tools to help understand and improve calibration models; we really need to get back to the roots on this, as well.

There are innumerable unsolved problems in Chemometrics that need to be addressed: real, scientific problems, not just new ways to throw numbers around. Anyone who needs help in thinking of some can check out our recent column where we presented a few; maybe those will jog the innovative juices.

REFERENCES

1. Mark, H., Workman, J.; Spectroscopy; 13 (6), p.19-21 (1998)

2. Mark, H., Workman, J.; Spectroscopy; (1998)

3. Mark, H., Workman, J.; Spectroscopy; (1998)

4. Mark, H., Workman, J.; Spectroscopy; (1998)

5. Mark, H., Workman, J.; Spectroscopy; 2 (1), p.38-39 (1987)

6. Mark, H., Workman, J.; Spectroscopy; 13 (4), p.26-29 (1998)

7. Kramer, R.; "Chemometric Techniques for Quantitative Analysis"; Marcel Dekker; New York (1998)