

What IS and what is NOT chemometrics

A.M.C. Davies

Norwich Near Infrared Consultancy, 10 Aspen Way, Cringleford, Norwich NR4 6UA, UK. E-mail: td@nnirc.co.uk

Introduction

From time to time papers are published that may not be untrue but are dangerous!

There was a new example recently and it seems a good time to repeat previous advice concerning what chemometrics is suppose to be about. There are varying descriptions, but the essential fact is that it should be a blend of chemistry and mathematics. If there is no chemistry input then it is NOT chemometrics.

What is chemometrics?

The most straight-forward explanation of "Chemometrics" (a word coined by Svente Wold in 1972 so it is now 40 years old!) that I have employed for many years is: "Chemometrics is concerned with the application of mathematical and statistical techniques to extract chemical and physical information from complex data". Paul Geladi added "the application of computer science" to this list and the advance of computer science has been a very important aspect of the development of chemometrics. Ian Cowe offered a different perspective which is reproduced in the box from an earlier column.¹ He said "What we (chemometricians) do is mainly to look at pictures". My favourite example of this is a recollection from Harald Martens.² In 1981 he visited Karl Norris' Beltsville laboratory and was showing his principal component analysis of ground wheat spectra to Karl and mathematician Bill Hruschka. Harald said "I still remember Karl's explosive eagerness and childish joy when he, all of a sudden, discovered the beautiful spectrum of water vapour in the tenth principal component loading plot... I had thought it was random instrument noise!"

In commenting on developments in chemometrics in the fore-runner of this column in 1991,³ I said "If you take chemistry out of chemometrics it is no longer CHEMometrics" and Paul Geladi wrote a column for me in 1997 on that theme, the importance of "CHEMO" (now available from the *Spectroscopy Europe* website via <http://bit.ly/P9ihHE>).⁴

What is NOT chemometrics?

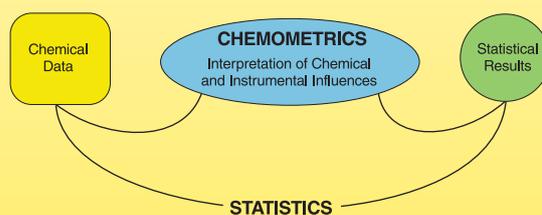
The catalyst for this article was a paper by Jim Reeves in a recent issue of the *Journal of Near Infrared Spectroscopy (JNIRS)* accessing a program called "Eureqa"⁵ (but pronounced "eureka"). I would like to make it clear that this is not a criticism of Jim; it was a useful thing to do and his verdict is not particular favourable. However, I do want to say that this is NOT

chemometrics. The program is designed to model complex data by testing multiple data operators in combination with genetic algorithms. There have been similar programs developed in the past. The main problem is that there is such scope for over-using data that no solution it produces should be used without further independent testing with new data but there is nothing to stop unskilled operators from missing this vital step. The second point is that there is almost no input of chemical knowledge,

continued on page 36

Ian Cowe's explanation for the difference between statistics and chemometrics

The aim of a Statistician is to reduce the information in a set of data to a small number of statistical variables, which summarise the relationship between various sets of data. He may have no knowledge of chemistry or instrumentation.



A Chemometrician brings knowledge of the chemical and sometimes the instrumental influences, which affect the data. The aim here is often to display the data in ways that allow chemical interpretation of the system. This may involve transforming the data in ways which "bring out" features which were not evident from the raw data or deriving new variables which are functions of the original data. Chemometricians make their living by "adding value" to the process of statistical analysis by bringing in skills which Statisticians lack.

Ian Cowe is the Chemometric Projects Coordinator at Foss Electric Development (UK) Ltd. Ian is particularly known for his work in introducing Principal Components Analysis into NIR spectroscopy. (Ian is now retired.)

continued from page 33

so from the previous discussion it is not chemometrics. I am pleased to learn that, at present, the program probably requires a super-computer to produce its suspect results in a reasonable time! Of course desktop computers will continue to get faster but I hope the message will be accepted that this approach is a mistake and is not needed.

Chemometric education

One of the reasons that people consider that there is a need for “Eureka-like” programs is that education in chemometrics has been slow to develop and is still very unevenly available. There are books (see www.impublications.com/shop/NIR-Spectroscopy-books/) but you need more than books to get

started. Training courses are often run at NIR conferences but you need to get there. On-line would appear to be the obvious solution. ICNIRS (the international body for NIR spectroscopy, www.icnirs.org) has been trying to develop a programme of lectures for several years. Well-researched data have been made available for home study and in May 2012 they setup a joint venture with the University of Córdoba⁶ for a virtual training programme on NIR technology. Another recent addition to this area of on-line education is from Eigenvector Research which is a well-respected chemometric company who have many years of experience in chemometric training in addition to their development of chemometric software. I have had a brief look at it and liked what I saw. It is available now.

Conclusions

From time to time calibration programs that attempt to remove the human input and replace it with automation will be developed but their aim is remove the chemo from chemometrics and that must be avoided! The answer is better training and (more importantly) more readily available training and now there are answers on the horizon which should be welcomed and utilised.

References

1. A.M.C. Davies, *Spectroscopy Europe* **10(4)**, 28 (1998).
2. H. Marten, *J. Near Infrared Spectrosc.* **4**, 12 (1996).
3. A.M.C. Davies, *Spectroscopy World* **3(2)**, 25 (1991).
4. P. Geladi, *Spectroscopy Europe* **9(4)**, 31 (1997).
5. J.B. Reeves, *J. Near Infrared Spectrosc.* **20**, 317 (2012).
6. *NIR news* **23(5)**, 3 (2012).