1

WHAT IS CHEMOMETRICS?

Preceding chapters	
Dependent chapters	
Matrix skills	Low
Statistical skills	Low
Excel skills	Low
Chemometric skills	Low
Chemometrics Add-In	Not used
Accompanying workbook	None

Chemometrics is what chemometricians do

—a popular wisdom

Chemometricians are the people who drink beer and steal ideas from statisticians

—Svante Wold

1.1 SUBJECT OF CHEMOMETRICS

What is chemometrics and what it does can be explained in different ways using more or less sophisticated words. At present, we have no generally accepted definition and, apparently, there never will be. The most popular is designation by D. Massart that chemometrics is *the chemical discipline that uses mathematical, statistical and other methods employing formal logic to design or select optimal measurement procedures and experiments, and to provide maximum relevant chemical information by analyzing chemical data*. This definition gathered a lot of criticism; therefore, new delineations have been suggested. For instance, S. Wold suggested the following formula: *Chemometrics solves the following*

Chemometrics in Excel, First Edition. Alexey L. Pomerantsev.

tasks in the field of chemistry: 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.

This also was not adopted and everybody agreed that the best way to explain the chemometric essence is an old method, proven in other equally obscure areas. It was declared that *chemometrics is what chemometricians do*. Actually, it was a clear rip-off because a similar idea has already been published,[1] but nothing better has been thought up since.

Well, really, what are these chemometricians doing? Here is a small collection of subjects found in the papers published over the past 5 years. Substantially, chemometricians perform the following activities:

- manage the production of semiconductors, aspirin, beer, and vodka;
- investigate the causes of destruction of documents written by ancient Gallic ink;
- conduct doping control in sport;
- determine the composition of ancient Egyptian makeup;
- localize gold deposit in Sweden;
- control the state of forests in Canada;
- diagnose arthritis and cancer in the early stages;
- investigate the organics in comets;
- select pigs' diet;
- check how a diet affects the mental capacity;
- find traces of cocaine on banknotes collected in the British Parliament;
- detect counterfeit drugs;
- decide on the origin of wines, oils, and pigments.

As long as everybody has clearly apprehended *what* chemometricians do, it remains to explain *how* they do this. However, initially, we have to be acquainted with the basic principles of their activity. Those are not numerous, just three rules, and they are as simple as ABC. The first principle states that *no data are redundant*, that is, a lot is better than nothing. In practice, this means that if you record a spectrum, it would be stupid to throw out all readings except a few characteristic wavelengths. Scientifically speaking, this is the multivariate methodology in the experimental design and data analysis.

Any data contains undesirable component called *noise*. The nature of noise can be different and, in many cases, a noise is just a part of data that does not contain relevant information. Which data component should be considered as noise and which component as information is always decided considering the goals and the methods used to achieve them. This is the second chemometric principle: *noise is what we do not need*.

However, the noise and redundancy necessarily provoke nonsystematic (i.e., noncausal) but random (i.e., correlation) relations between the variables. The difference between causality and correlation has been illustrated humorously in a book by Box and Hunters.[2] There is an example of a high positive correlation between the number of inhabitants and the number of storks in Oldenburg (Germany) for the period from 1930 to 1936. This is good news for those who believe that the stork is a key factor in baby boom!

[1] P.W. Bridgman. On scientific method, in *Reflections of a Physicist*, Philosophical Library, NY, 1955.
[2] G.E.P. Box, W.G. Hunter, J.S. Hunter. *Statistics for Experimenters*, John Wiley & Sons Inc., NY, 1978.

However, the reason that these two variables are correlated is very simple. There was a somewhat hidden third variable, which had pair-wise causal relationship with both variables.[3] Therefore, the third principle of chemometrics states as follows: *seek for hidden variables*.

1.2 HISTORICAL DIGRESSION

Strangely enough, when can somebody claim beyond a doubt that a science was born on a certain date, in a certain place, and under specific circumstances? Who dares, for example, to specify where and when chemistry appeared first? It is only clear that it was long ago and very likely it did not happen under very nice circumstances, and it was related to an urgent necessity to quietly send a sixth priest of the Anubis temple to Kingdom Come or something. But we are quite certain that chemometrics was born on the evening, June 10, 1974, in a small Tex-Mex restaurant in Seattle, as the result of a pretty noisy revelry, arranged on the occasion of Swante's return home in Umea, after a long training under the renowned statistician George Box.

The American side was represented by the chemist Bruce Kowalski with his disciples, who had been intensively developing a software package for chemical data analysis. In fact, the word "chemometrics" has been used since the early 1970s by Svante Wold and his team from the University of Umea, in order to identify, briefly, what they were doing. Actually, they were engaged in an intensive implementation of the interesting ideas produced by Svante's father, Herman Wold, who was a famous statistician, developing many techniques (including the now famous method of the projection to latent structures – PLS) for data analysis in economic and social sciences.

Apparently, with an analogy to psychometrics, biometrics, and other related disciplines, the term *chemometrics* was coined. The circumstances and the place of birth have left an indelible life mark for the science itself and for the people who are involved in chemometrics. Much later, Svante Wold, being already a very venerable scholar, joked at a conference "chemometricians are the people who drink beer and steal ideas from statisticians." On the other hand, perhaps, the reason is that the first chemometrician, William Gosset, better known under the pseudonym Student, was employed as an analyst at the Guinness brewery. Anyway, beer and chemometric have gone together for many years. I remember a story told by another famous chemometrician, Agnar Höskuldsson at a Winter Symposium on Chemometrics held near Moscow, in February 2005. The atmosphere at the ceremony of awarding a young Russian chemometrician was semi-official, and Agnar, who presented the award, delivered a fitting address. This remarkable history, almost a parable, is worth presenting word by word. Thus spoke Agnar.

Chemometricians are the cheerful people who like to drink beer and sing merry songs. In the mid-1970s, when I (i.e., Agnar) first met Svante, he had only two students. And so, we, four men, were sitting in an Umea pub in the evening, drank beer, and sang merry songs. After 2 hours, I asked Svante:

"It may already be enough, and is it time to go home?"

"No," Svante said, "still not enough, we will drink beer and sing merry songs."

[3]The subsequent investigations have revealed the nature of a hidden variable, which was proved to be the area of the fields raiseing the cabbage in the Oldenburg neighborhood.

Another hour passed, and I asked Svante again

"It may already be enough, and is it time to go home?"
"No," he answered, "we still have to drink beer and sing more songs."

Meanwhile, a few students went down the street, near the pub where we were. They heard we had fun, and stopped and began to ask of the passersby:

"Who are these cheerful people who are singing these merry songs?"

No one could answer them, so the students came in the pub and asked Svante:

"Who are you guys, and why are you singing these merry songs?"

And Svante replied:

"We are chemometricians, the cheerful people, who like to drink beer and sing merry songs. Sit down, drink beer, and sing along with us."

That was the beginning of the famous Scandinavian school of chemometrics, which gave us many great scientists and cheerful people.

The appearance of chemometrics in Russia was also marked by a funny story. Soon after the establishment of the Russian Chemometrics Society in 1997, my colleagues decided that it was a proper time to announce our existence *urbi et orbi*. Just that time, we received the first call to the All-Russian conference "Mathematical Methods in Chemistry," which was to be held in the summer of 1998, in Vladimir city. This was a very popular scientific forum established in 1972 by the Karpov Institute of Physical Chemistry. The conference was a place where many scientists reported on their results; the hottest discussions of the latest mathematical methods applied in chemistry were conducted. We had a hope that using this high rostrum, we could tell fellows about this wonderful "chemometrics" developed by our friends around the world and how interesting and promising this science is. The abstract was written in the name of the Russian Chemometrics Society and timely delivered to the organizing committee.

Time had passed, but there was no response. The deadline was approaching, the program and the lists of speakers were already published, but our contribution was not included. Our bewildered requests went unanswered so we decided that there should be some misunderstanding. As Vladimir is rather close to Moscow, it was decided that we would take a chance and visit the conference without invitation in order to sort things out on the spot. So we did, appearing on the first conference day for the participant registration. This raised a big stink accompanied with an explicit fright, screaming, loud statements against provocation that finally resulted in the appearance of very responsible *tovarishchi*.

Things were heading toward us getting out of the university, facing the tough administrative sanctions, arrest, and finally imprisonment. Fortunately, in an hour, through the efforts of our old friends from previous conferences, the situation was resolved and we got the casus clarified. The reason was that the unsophisticated Vladimir scientists believed that chemometrics is something like Scientology and our society is a kind of a harmful sect that tries to use the science conference as a chance to embarrass the unstable minds of Vladimir's people and turn their comprehension away from true scientific values. At the

end, we were saved by the fact that in the University library, a book titled "Chemometrics,"[4] published in 1987 was found. Final reconciliation occurred during the closing banquet, but our lecture, *de bene esse*, was not permitted.

It is possible that the committee members were guided by the sound judgment of Auguste Comte, who in 1825 warned[5] "Every attempt to employ mathematical methods in the study of chemical questions must be considered profoundly irrational and contrary to the spirit of chemistry. If mathematical analysis should ever hold a prominent place in chemistry – an aberration which is happily almost impossible – it would occasion a rapid and widespread degeneration of that science."

[4]M.A. Sharaf, D.L. Illman, B.R. Kowalski. *Chemometrics*, Wiley, New York, 1986 (Russian translation: Mir, 1987).
[5]A. Comte. *Cours de Philosophie Positive*, Bachelier, Paris, 1830.