

Cutting the data cake — having your cake and sampling it

Tony Davies (Chemometrics Column Editor)

This month I want to tackle one of the most crucial concepts of chemometrics. The proposition is: **the way in which data was collected may not be the best form for analysing it.**

Many chemists and some spectroscopists find this difficult but it is essential that we overcome our inhibitions of only regarding data from this conventional point of view. This article will try to present a pictorial view of data manipulations which I hope will be helpful. If Professor Stephen Hawking prefers pictures to mathematics,¹ I don't see why anyone should be embarrassed about using a "naive" approach.

Figure 1 is our data cake. It is a layer cake; each layer being a spectrum. For example in our cake the x axis is the wavenumber (or wavelength) data, the y axis is the sample number and the z axis is the transmission (or absorption) data.

If we have a rectangular cake the social custom is to offer slices cut at right angles to the longest slice. This is the same for our data cake because such a slice would compare the transmission for each sample at the same wavenumber — the customary way of analysing spectroscopic data. In a simple case, where there was a

direct relationship between the sample number and its transmission at a particular wavenumber, our slice would be a useful graph, as shown in Figure 1.

Figure 2 indicates how we could obtain two imaginary slices of cake. If we averaged the data in the y direction, this would be the average spectrum for the data set. We can also average, or just sum, the data in the x direction; this would represent the total transmission of each sample over the wavenumber range. We would not expect such data to be very interesting but I want you to remember that it is an option.

Now I want to cut a different cake. I need to introduce the idea of a graphic scaling factor. This is shown in Figure 3(a); the idea is that the top half of the graph represents a scaling factor between 0.0 and 1.0. If we cut a slice off the scaling factor in the x direction and then re-calculate the spectrum, the whole spectrum would be reduced by the same factor. Instead of a knife we could use a cutter to remove part of the scaling factor, as shown in Figure 3(b). When the spectrum is re-calculated the product would be a modified spectrum as shown in Figure 3(c); some peaks have been

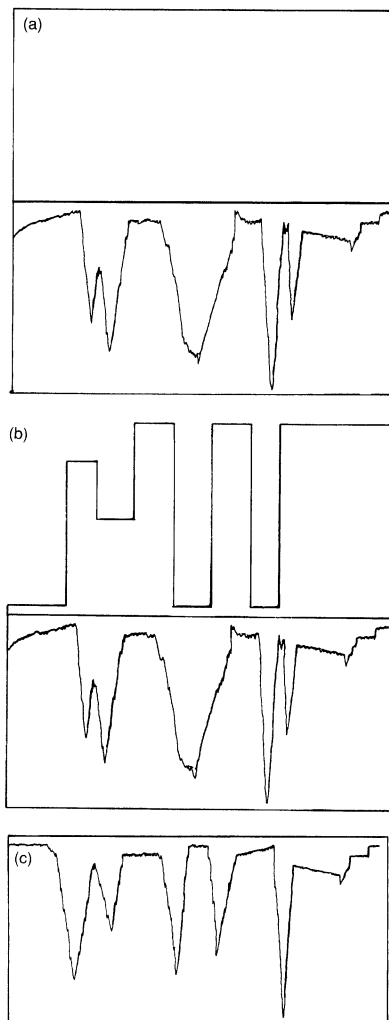


Figure 3. Graphic scaling factor.

enhanced while others have been removed. We will now transfer the idea to the cake.

Figure 4(a) shows the two part cake; reminiscent of a "Battenburgh". In Figure 4(b) the scaling factor has been cut by a cutter so that each spectrum in the cake is equally affected. The re-calculated cake is shown in Figure 4(c) together with

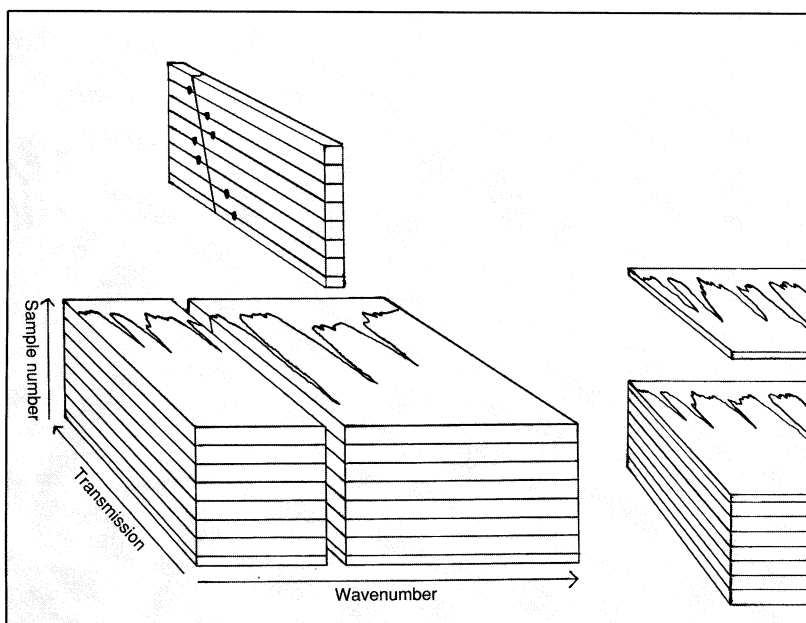


Figure 1. The data cake.

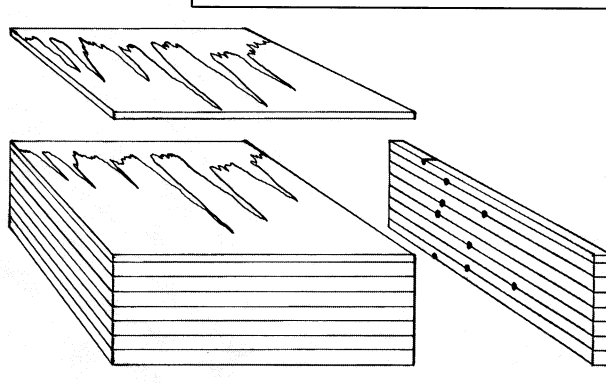


Figure 2. Computed slices from the data cake.

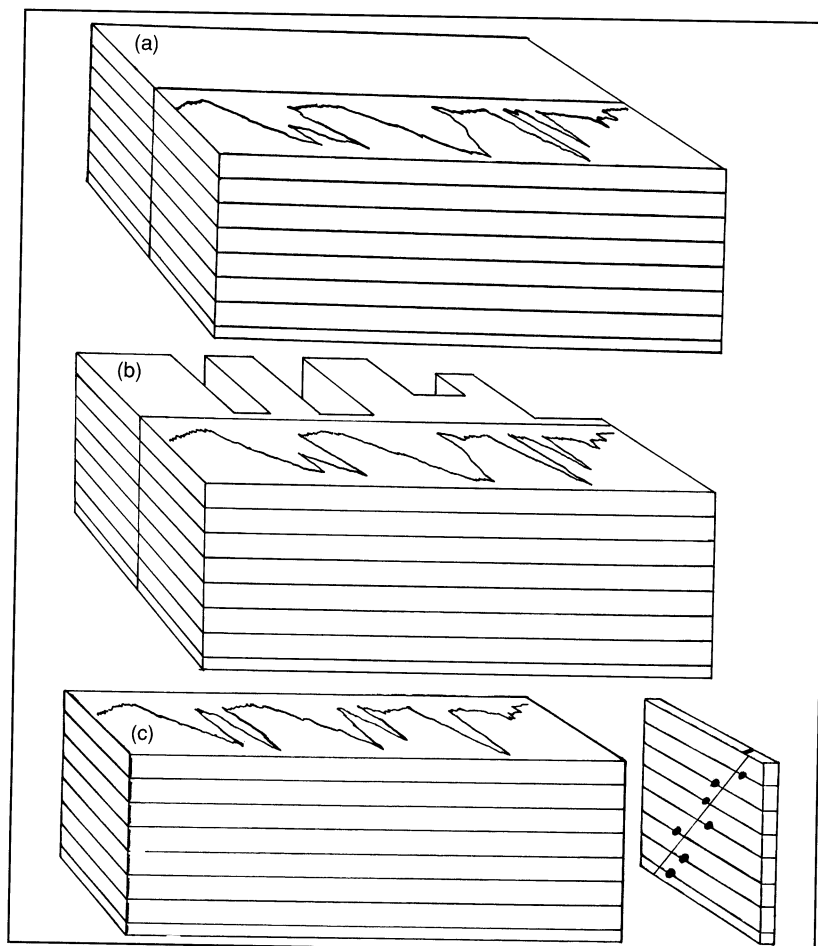


Figure 4. Cutting the scaled cake.

the new computed slice for the total transmission in the x direction. If we have been very clever with the choice of our cutter then these values might show a relationship with the sample number (or some other variable).

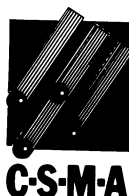
Now we come to the difficult part! We still have our cake (this is all happening in computer memory) and we can save this new computed slice, return to the original cake, select a new cutter and compute another new slice. I leave it to you to imagine forming new slices into a NEW data cake. Thus proving that you can have your cake and sample it!

Why do we want to do this? This may seem like a bit of fun with no real purpose. There is a purpose. Chemometrics has some very clever ways of choosing the data cutter, which are powerful tools for analysing complex data. In the next issue we shall meet one of them — *factor analysis*.

Reference

1. S.W. Hawking, "A Brief History of a Brief History", *Popular Science (New York)* **235**, 70-72 (1989)

The Centre for Surface and Materials Analysis



The CSMA provides one of the most comprehensive surface and materials analysis laboratories in the world. An unrivalled range of techniques is now available to clients for the complete answer to surface/interface problems using a single or multitechnique approach.

- | | |
|--|--|
| Chemical Information
• ESCA
• Auger
• SIMS
• FT-IR
• Electron Microprobe | Physical Information
• SEM
• Low Voltage SEM
• STEM
• XRD
• Optical Microscopy |
|--|--|

From semiconductors to catalysts or polymers to metals, there aren't many materials that we can't characterise. Whether you are monitoring quality, problem solving or characterising novel materials you should know more about us.

For a full information pack contact Dr Andrew Swift (061 200 4441).

THE CENTRE OF SURFACE AND MATERIALS ANALYSIS
 UMIST PO BOX 88
 MANCHESTER M60 1QD
 UNITED KINGDOM



Circle 020 for further information

UNIVERSITY OF YORK DEPARTMENT OF CHEMISTRY 1990 SHORT COURSES For Industrial Chemists 9 - 13 APRIL 1990



1 New Approaches in Organic Synthesis

An intensive five-day introduction or refresher for research and development chemists or managers in the chemical and pharmaceutical industries who are interested in recent trends in organic synthesis.

- Design and applications of new reagents and catalysts
- Routes to high-value speciality chemicals, including organofluorine compounds
- Retro-synthetic analysis
- The search for cheap effective oxidants
- Workshops and Case studies
- Problem solving exercises

2 Modern Analytical Techniques

A series of complementary but self-contained two-day modules for industrial scientists and management level professionals. Each will feature lectures with key invited speakers and problem-solving workshops and will focus on recent developments and industrial applications.

- High Resolution solids NMR Spectroscopy, including MAS
- Fourier Transform Infrared and Laser Raman Spectroscopy with Microprobe Detection
- High Resolution Electron Microscopy with X-ray Analysis

Modules may be taken individually or linked together for a discounted rate.

Further details from: Dr Carys Calvert, Department of Chemistry, University of York, York YO1 5DD. Telephone (0904) 432576 (direct) or 432511 (general office). Telex 57933 YORKUL. Fax (0904) 433433.

Circle 021 for further information