

Causality, Machine Learning and Human Insight

Harald Martens^{1,2}

Bio-chemometrician

¹Prof. emerit. Dept. Engineering Cybernetics, Norw. U. of Sci. & Technol. NTNU, Trondheim Norway

²Senior consultant, Idletechs AS

Abstract:

Modern instruments generate Big Data that require information extraction before they can be used. A hybrid modelling framework is presented and illustrated. Its purpose is to convert meaningless data to meaningful information and to contribute to a theoretical, practical, and democratic basis for tomorrow's handling of Big Data in science and technology.

A call for causality

Writing this essay in *Analytica Chimica Acta* feels like a homecoming for me, since my first data modelling paper, with the title '*Factor analysis of chemical mixtures. Non-negative factor solutions for spectra of cereal amino acids*', was published here in ACA, in 1979 (REF01). As biochemist, I combined slow chromatography, others' agronomic knowledge and my amateur multivariate data modelling.

Today, many instruments in chemistry generate information-rich BIG DATA, fast and cheaply. But high-speed measurements are often quite "dirty", even in chemistry. Hyperspectral imaging of natural products is one example: Measurement noise is not the problem. We have to unmix overlapping absorbance signals from mixed chemical constituents and distinguish these from the effects of physical light scattering and measurement artifacts, e.g. due to illumination changes (REF02).

Purely data-driven modelling is in AI lingo called Machine Learning (ML). For BIG DATA from multi-channel spectrometers and imagers in systems with limited causal complexity, the use of complex black box AI-based deep learning is alienating, needlessly risky and a computational overkill. It is better to combine our prior knowledge and observation-based discoveries into improved causal insight.

A generic machine learning framework for handling both known and unknown causalities in technical BIG DATA will now be outlined and illustrated. The presentation is based on my personal experience, and thus highly subjective.

Machine learning with an eye for causalities

Prior knowledge and empirical data may be combined in different ways, in multivariate hybrid modelling. A generic R&D framework is shown in Figure 1 for a continuous stream of high-dimensional data, e.g. from hyperspectral or thermal video cameras. It has a deductive, an inductive and an abductive stage:

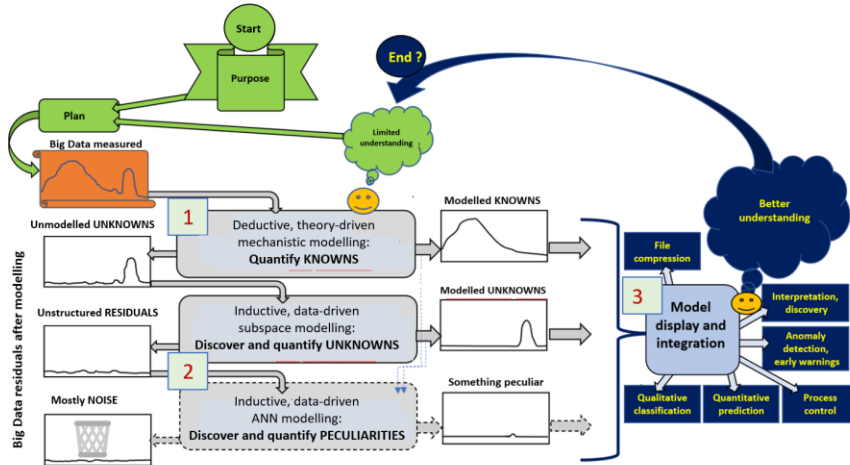


Figure 1: A framework for modelling technical BIG DATA on a small computer, for real-world problem solving and knowledge enhancement (REF03): 1) Deductive "expert system": Causal modelling of KNOWNs. 2) Inductive machine learning: Data driven modelling of systematic and – if needed, peculiar - UNKNOWNs (REF04). 3) Abductive deep learning: Machines learning to behave intelligently (REF05).

A statistically validated use of this hybrid modelling framework may help us generate value and maintain a chain of trust in R&D and industrial operation: Modelling our new Big Data measurements in terms of established causal theory, prior experience and available external descriptors reduces the cost of training data and the risk for spurious conclusions. And by only extracting new variations in our unmodelled residuals if they are clear and systematic, we continue learning, creatively and critically. By combining the known and unknown variation patterns, we can deliver reliable results and improve our causal understanding (REF06).

Can we see how it feels to be a photon?

Figure 2 shows a visual example of human "deep learning": A herring-shape sugar candy, in the producer's identity-colors red, white and blue. Seen in reflectance, the middle part is the brightest, and in transmission it is the darkest. Assuming a spatially constant light source and detector sensitivity across the image, the RGB values are, more or less, proportional to Reflectance and transmittance data.

Red, white and blue: Why ?



Figure 2 How does it feel to be a photon here? A herring-shaped sugar candy, seen in reflection and transmission REF07.

I was told by the factory that the red and blue colors were caused by light absorption of dyes from raspberries and blueberries, respectively, while the whiteness was caused by building in lots of tiny air bubbles, just like in whipped egg white.

Conclusion: Because light absorption and light scattering affect photons very differently, we can understand what causes color- and whiteness variation, - if we so wish.

Demo: Beer-Lambert's law hold also for NIR transmittance through powders

With only 3 color channels – R, G and B, there are limits to what our cameras can detect, and to what we can distinguish by modelling. And having to do both reflectance and transmittance measurements is cumbersome. What if we measure *only transmitted* light, but with many more wavelength channels instead?

Figure 3 illustrates the advantage of multichannel measurements and compares mathematical modelling based on a purely inductive method (top) and a combined inductive/ deductive method (bottom). It concerns high-speed near infrared spectroscopy through intact white-looking powders (mixtures of pure wheat protein and wheat starch). They are expected to differ in both absorbance and scattering. But to generalize, we assume their pure spectra to be unknown.

There are five mixture ratios of protein and starch powders. Each mixture was measured in transmission at 10 different conditions to simulate real-world troubles: five powder thicknesses × two powder compressions, in two technical replicates (REF08, REF09, REF10). The expected linearity of the spectral response to the analyte (protein) variation was optimized by converting the diffuse transmittance T into absorbance ($A = \log(1/T)$). The absorbance spectra show several broad peaks. Their partial overlap makes them unsuited for classical single-wavelength calibration.

But multivariate calibration "machine learning" (REF06) has no problem with partially overlapping peaks. By cross-validated Partial Least Squares Regression based on three of the five mixtures yields reasonable prediction ability (C), even for the remaining independent test mixtures:

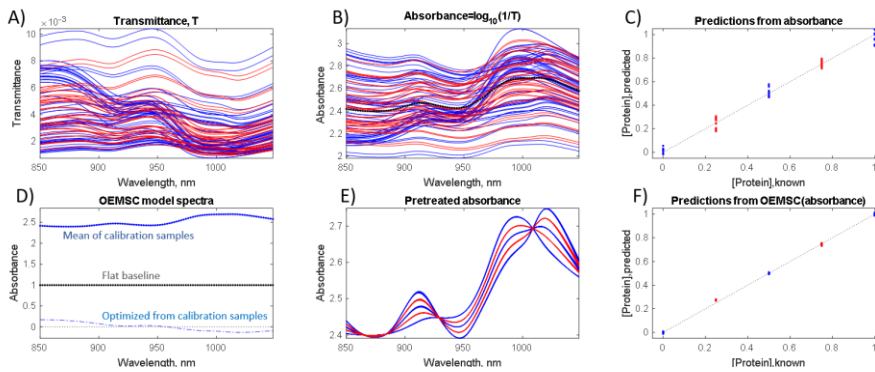


Figure 3 Demo experiment of linear hybrid subspace modelling: *Top: Conventional calibration.* A) Transmittance spectra T of powder mixtures (wheat protein: wheat starch in ratios 0:100, 25:75, 50:50, 75:25 and 100:0) measured under different conditions. B) Absorbance $A = \log_{10}(1/T)$. C) Conventional multivariate calibration by PLSR.

Bottom: Abductive calibration: D) Spectral OEMSC model, E) The 100 absorbance spectra in B) after OEMSC pre-processing. F) Multivariate calibration based on E). Blue: calibration mixtures. Red: test mixtures.

But something was wrong: the machine learning model required *three* inductive model dimensions (PLS Components, PCs) in the calibration model, even though we only had *one* type of composition variation differing in *two* respects (absorption and scattering).

The bottom of Figure 3 shows results for the same 100 spectra after semi-causal pre-processing by Optimized EMSC (OEMSC, REF09), which optimizes a Beer-Lambert-like bilinear model (Extended Multiplicative Signal Correction, EMSC REF11, REF08) wrt predictive ability. The NIR spectra pre-treated by OEMSC gave excellent predictions, both for calibration and test samples (F). And only *two* inductive model dimensions were needed – *one* i OEMSC and *one* i PLSR.

Figure 4 explains why the semi-causal OEMSC pre-processing was developed: Contrary to e.g. ANN, CNN and traditional statistical regressions, the PLSR provides *windows* into the high-dimensional space of the spectra:

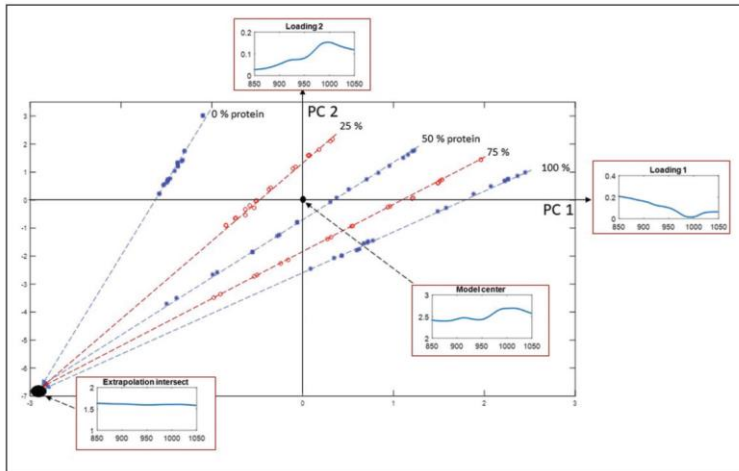


Figure 4 A 2D window into the high-dimensional space of the powder absorbance spectra in Figure 3B): The PLS scores for 60 calibration samples (blue) and 40 test samples (red), plotted for PCs # 1 and # 2, along with the corresponding model center (the calibration mean spectrum), the PC loading spectra and the common extrapolation spectrum.

Here, the first two PLS PCs show that different sample thicknesses and sample compactions for the five chemical compositions form five straight lines, pointing towards a common, flat baseline (a multiplicative effect) but unevenly spaced (an additive curvature effect?). Like a curved 3D banana can be approximated by a flat 2D boomerang, the linear OEMSC model handled the curvature (REF06). And following Beer-Lambert's law mathematically, it also modelled both multiplicative and additive effects simultaneously.

Conclusion: Multichannel diffuse NIR spectra with purely data driven PLSR machine learning worked reasonably, even for these "dirty" powder mixture spectra. But it worked far better after semi-causal OEMSC pre-processing.

Light absorbances from mixed chemical are approximately linear and additive, while effective optical path length changes, due to light scattering, is multiplicative. Combining multichannel spectra and simple linear algebra, we can therefore quantify what causes "color"- and "whiteness" changes, even in NIR spectra. So, Beer'-Lambert's law appears to hold even for diffuse transmission through powders.

Beer-Lambert's law can also model hyperspectral video of changes in complex biological material

Hyperspectral VNIR video with kinetic modelling: Reflection is usually easier to measure than transmission. Today's hyperspectral remote-sensing imaging give massive streams of spatiotemporal diffuse Reflectance (R) spectra, and hyperspectral video even more so. How to handle technical BIG DATA like this? Just storing and transmitting them is a problem. How to interpret them, in terms of known and unknown chemical, physical and instrumental variations?

The top of Figure 5 shows a piece of wood (spruce) that was monitored over 21 hrs (REF12, REF03) from its wet to its dried state by hyperspectral VNIR (Visible and NIR) "video" (500- 1005 nm). Images with > 2.3 million pixels were taken repeatedly overnight. This resulted in > 350 million

reflectance spectra with 159 wavelength channels in each, for a single wood sample. This is how the hyperspectral video was modelled:

Following the steps in Figure 1:

1. Modelling KNOWNNS by deduction: Each absorbance spectrum $\log_{10}(1/R)$ was submitted to a conventional EMSC with 5 spectral inputs (the mean, tap water, wood pigment, a flat and a slanted baseline). Each spectrum was dividing by the relative optical path length estimated from the mean spectrum, while the other relative contributions were estimated and subtracted. This semi-causal pre-processing accounted for 98.9 % of the variance in the input absorbances.
2. Modelling UNKNOWNNS by induction: The pre-processed absorbance spectra were submitted to a principal component analysis (PCA). This revealed two unexpected but clear variation patterns, which explained another 0.7% of the variance. Hence, with 5 known and 2 unknown spectral components, 99.6% of the variance in the >350 million spectra was modelled. Since the unmodelled residuals appeared to represent mostly random noise, and no serious outliers had been found, there was no need for a final inductive, data-driven ANN clean-up stage.
3. New causal insight by abduction: When the 7 model components were averaged in each image and modelled as functions of drying time, 4 of the 5 known and 1 of the 2 unknown components were found to follow first-order reaction kinetics.

Hyperspectral SWIR video with two-domain IDLE modelling:

Vitale et al. (REF13) applied the same modelling framework to another hyperspectral video of drying wood, in the SWIR wavelength range (930-2200 nm), again modelling its light absorption and light scattering developments by the hybrid modelling, using EMSC & PCA.

However, the drying also caused the wood to *shrink, gradually*. There was also some camera/object motion. This spatiotemporal motion was quantified by optical flow estimation and summarized into a PCA *motion map* by so-called dual-domain IDLE-modelling (REF14), illustrated at the bottom of Figure 5, and used for motion compensation.

Hence, the BIG hyperspectral video file was *compressed* into two SMALL subspace models with respect to all interesting changes in chemical composition, light scattering and spatial shrinkage, in the wavelength-, time- and space-domains, simultaneously.

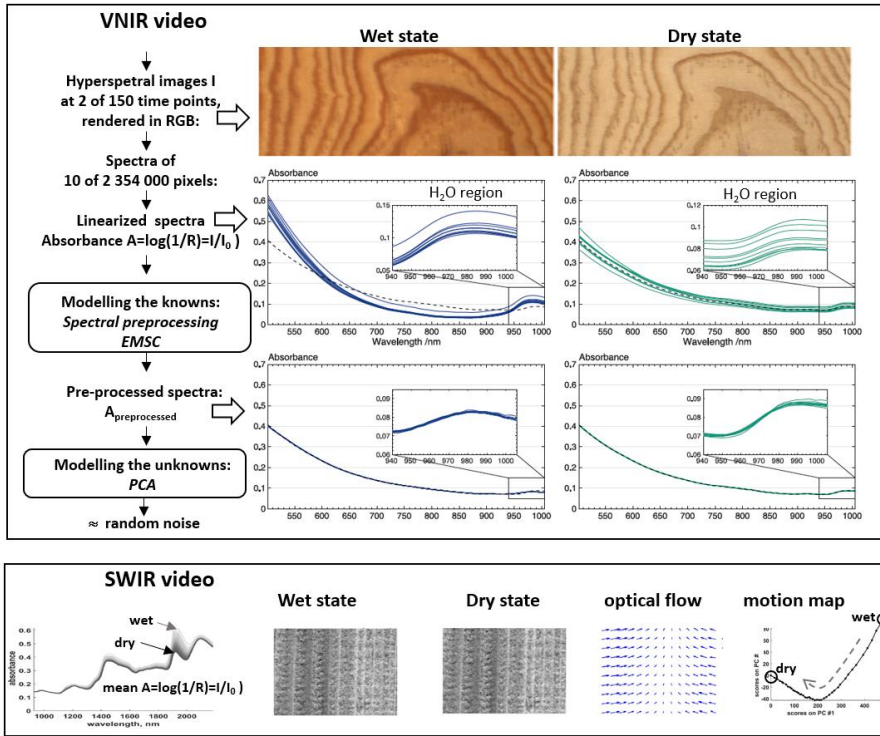


Figure 5 Hyperspectral video monitoring of drying wood:

Top: A piece of spruce monitored by VNIR video with EMSC & PCA:

Summary of how the input absorbance of its hyperspectral VNIR video was modelled, Absorbance before and after conventional knowledge-driven EMSC pre-processing, followed by data-driven PCA modelling,

Bottom: Another piece of spruce monitored by SWIR video with EMSC, IDLE and PCA:

The SWIR absorbance spectra (mean over each image in video) show variation due to light absorption (water) and light scattering (air replacing water). But comparing the first and last image in video also shows considerable optical flow (spatial shrinkage). The main motion map (from PCA of all optical flow images) reveals two phases in the shrinkage process.

Conclusion: The hybrid modelling sequence in Figure 1 can give fast, simple, understandable description of spatiotemporal processes from technical BIG DATA measurements, as shown for these two examples of Beer-Lambert modelling of hyperspectral video, at strongly reduced file size.

Ahead, not headless

Society is now at a crossroad wrt Big Data and Artificial Intelligence. Current AI developments for very complex problems are powerful and fascinating. But we don't see their societal consequences yet. So, we should tread with extreme care:

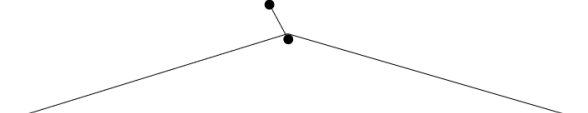
I don't need to understand how the breaks in my car's work, as long as I know that the breaks work and *somebody else* knows how it works. But I don't want to use a black box AI solution if I cannot know when it doesn't work, and *nobody* knows how it works.

In this paper I have tried to demonstrate that faster, simpler, less data-demanding and more interpretable "deep learning" solutions are possible, at least for multichannel BIG DATA from chemical, physical or technical systems with limited causal complexity, known or unknown.

But the present hybrid deductive/inductive/abductive modelling framework needs further method improvements (REF24). That could include finding "unmixing end members" by "simplex intersect" (REF01, REF14, REF15), relevant-subspace modelling (REF16) in sparse versions (REF17), adaptive handling of strong heterogeneities (REF18, REF19) and simplified linear modelling of mechanistic causality equations by multivariate metamodeling (REF20).

On the other hand, there are fundamental mathematical modelling limitations: The paradox of duality in linear mixture modelling (REF09), the subjectivity involved in choice of causal model (REF21) and the mathematical "sloppiness" of many non-linear causal models (REF22, REF23). And we must never give up our strive to reduce the Math Gap in science and society (REF24, REF25).

In summary, I believe ANN-based black box machine learning represents needless, headless, alienating overkill for many types of chemical, physical, and technical Big Data. The explainable hybrid subspace framework (Figure 1) might be tried first, at least as pre-processing, under an old man's motto:



No causal interpretation without predictive validity!
And no prediction without attempted causal interpretation!

References:

- REF01 Martens H (1979) Factor analysis of chemical mixtures. Non-negative factor solutions for spectra of cereal amino acids. *Analytica Chimica Acta* **112**, 423-442.
- REF02 J.F. Fortuna and H. Martens, H (2017) Multivariate data modelling for de-shadowing of airborne hyperspectral imaging *J. Spectral Imaging* **6**, a2.
- REF03 Martens H (2021) Interpretable machine learning with an eye for the physics: Hyperspectral Vis/NIR "video" of drying wood analyzed by hybrid subspace modeling. *NIR News Vol. 32(7-8)* 24-32.
- REF04 Vitale R, Zhyrova A, Fortuna JF, de Noord OE, Ferrer A, Martens H (2017) On-The-Fly Processing of continuous high-dimensional data streams. *Chemometrics and Intelligent Laboratory Systems* **161** 118-129.
- REF05 Strümke I (2023) Maskiner som tenker. Algoritmenes hemmeligheter og veien til kunstig intelligens. Kagge Forlag, Oslo. ISBN: 978-82-489-3250-5 (in Norwegian).
- REF06 Martens H and Naes T (1989) *Multivariate Calibration*. J. Wiley & Sons Ltd Chichester UK. ISBN: 978-0-471-93047-1 .
- REF07 Martens (2021) Enn om vi var gjennomsiktige? *Kjemi* **6** (2021) pp 8-14 (in Norwegian). The candy and its colors are from Schleswig-Holstein, Germany.

REF08 Martens H, Pram Nielsen J. and Balling Engelsen S (2003) Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures. *Anal. Chem.*; 75 (3) 394 – 404.

REF09 Martens H (2011) The informative converse paradox: Windows into the unknown. *Chemometrics and Intelligent Laboratory Systems* 107, 124–138.

REF10 Martens H (2021) Understanding the root cause(s) of nonlinearities in near infrared spectroscopy *NIR News* Vol. 32(1–2) 20–26.

REF11 Martens H and Stark E (1991) Extended multiplicative signal correction and spectral interference subtraction: New pre-processing methods for near infrared spectroscopy. *J.Pharmaceutical & Biomedical Analysis* 9(8),625-635.

REF12 Stefansson P, Fortuna J, Rahmati H, Burud I, Konevskikh T and Martens H (2020) Hyperspectral time series analysis: hyperspectral image data streams interpreted by modeling known and unknown variations. In : *Hyperspectral Imaging* (Jose Manuel Amigo, ed.). Pp 305 – 331. <https://doi.org/10.1016/B978-0-444-63977-6.00014-6>.

Field Code Changed

REF13 Vitale R, Stefansson P, Marini F, Ruckebusch C, Burud I, Martens H (2020) *Fast analysis, processing and modeling of hyperspectral videos: challenges and possible solutions*, in: *Comprehensive Chemometrics* (2nd Edition, ISBN: 9780124095472), Elsevier, doi:10.1016/B978-0-12-409547-2.14605-0.

REF14 Vitale R, Ruckebusch C, Burud I and Martens H (2022) Hyperspectral Video Analysis by Motion and Intensity Pre-processing and Subspace Autoencoding. *Frontiers in Chemistry*, doi 10.3389/fchem.2022.818974.

REF15 Martens H and Kohler A (2009) Mathematics and Measurements for High-throughput Quantitative Biology. *Biological Theory*, Special issue on Quantifying Biology. *Biological Theory* 4(1), 29-43.

REF16 Wold S, Martens H and Wold H (1983) The Multivariate Calibration Problem in Chemistry solved by the PLS Method. *Proc. Conf. Matrix Pencils*, (A. Ruhe and B. Kågström, eds.), March 1982, *Lecture Notes in Mathematics*, Springer Verlag, Heidelberg, 286-293.

REF17 Hovde Liland K, Høy M, Martens H and Sæbø S (2013) Distribution based truncation for variable selection in subspace methods for multivariate regression. *Chemometrics and Intelligent Laboratory Systems* 122 103–111.

REF18 Tøndel K, Indahl UG, Gjuvland AB, Jon Olav Vik JO, Peter Hunter, Stig W Omholt SW and Martens H (2011) Hierarchical Cluster-based Partial Least Squares Regression (HC-PLSR) is an efficient tool for metamodeling of nonlinear dynamic models. *Systems Biology*, 5:90 <http://www.biomedcentral.com/1752-0509/5/90>.

REF19 Martens H (2009) Non-linear multivariate dynamics modelled by PLSR. *Proceedings of the 6th International Conference on Partial Least Squares and Related Methods*, Beijing 4-7 2009 (V.E.Vinzi, M.Tenenhaus and R.Guan, eds), Publishing House of Electronics Industry, <http://www.phei.com.cn>, p. 139-144.

REF20 Tøndel K and Martens H (2014) Analyzing complex mathematical model behavior by PLSR-based multivariate metamodeling. *WIREs Computational Statistics*, [Volume 6, Issue 6](#), pages 440–475, November/December 2014. DOI: 10.1002/wics.1325.

Field Code Changed

REF21 Isaeva J., Martens M, Sæbø, Wyller JA & Martens H (2012) The modelome of line curvature: Many nonlinear models approximated by a single bi-linear metamodel with verbal profiling. *Physica D: Nonlinear Phenomena* **241**, 877–889.

REF22 Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP (2007) Universally sloppy parameter sensitivities in systems biology models. *BMC Syst. Biol.*; **3**: 1871–1878.

REF23 Tafintseva V, Tøndel K, Ponosov A, Martens H (2014) Global structure of sloppiness in a nonlinear model. *J. Chemometrics*, [Volume 28, Issue 8](#), pages 645–655.

REF24 Martens H (2015) Quantitative Big Data: Where Chemometrics can contribute. *J. Chemometrics* **29** (11) 563–581 <http://onlinelibrary.wiley.com/doi/10.1002/cem.2740/epdf>.

Field Code Changed

REF25 Skjærvold NK, Tøndel K, Cedersund G, Brovold H, Rahmati H, Munck LM, Martens H (2017) Multivariate analyses and the bridging of biology's "Math-Gap". *Encyclopedia of Analytical Chemistry*, Online © 2006–2017 John Wiley & Sons, Ltd. DOI: 10.1002/9780470027318.a9616.