Review

Journal of CHEMOMETRICS

Received: 1 December 2011, Revised: 26 February 2012, Accepted: 28 February 2012, Published online in Wiley Online Library: 2012

(wileyonlinelibrary.com) DOI: 10.1002/cem.2441

Advantages of orthogonal inspection in chemometrics

Rui Climaco Pinto[a], Johan Trygg[a]* and Johan Gottfries[b]

The demand for chemometrics tools and concepts to study complex problems in modern biology and medicine has prompted chemometricians to shift their focus away from a traditional emphasis on model predictive capacity toward optimizing information exchange via model interpretation for biological validation. The interpretation of projection-based latent variable models is not straightforward because of its confounding of different systematic variations in the model components. Over the last 15 years, this has spurred the development of orthogonal-based methods that are capable of separating the correlated variation (to Y) from the noncorrelated (orthogonal to Y) variations in a single model. Here, we aim to provide a conceptual explanation of the advantages of orthogonal variation inspection in the context of Partial Least Squares (PLS) in multivariate classification and calibration. We propose that by inspecting the orthogonal variation, both model interpretation and information quality are improved by enhancement of the resulting level of knowledge. Although the predictive capacity of PLS using orthogonal methods may be identical to that of PLS alone, the combined result can be superior when it comes to the model interpretation. By discussing theory and examples, several new advantages revealed by inspection of orthogonal variation are highlighted. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: OSC; PLS; OPLS; OPLS-DA; orthogonal variation; predictive variation

1. INTRODUCTION

Over the last 10 years, orthogonal methods have been developed and used in spectroscopic data analysis for calibration [1], classification [2], and calibration transfer [3]. They have also been extensively used in "omics" applications, for example, to analyze data from nuclear magnetic resonance (NMR) or gas chromatography/mass spectroscopy (GC-MS)-based technologies (proteomics and metabolomics) [4,5]. Orthogonal variation, which has occasionally been referred to as "structured noise" [2], or "irrelevant information" [6,7], has been shown to be present in most experimental data, often representing a larger proportion than the information correlated to the response variables [8]. It may be the result of different types of physical, chemical, biological, and instrumental factors [9]. Examples include drifts in the results due to changes in operating or environmental conditions, e.g., pressure, temperature, humidity, and instrument-specific factors [10,11]; physical factors, such as particle size, homogeneity and light scattering, chemical and molecular properties [12]; biological sample variation due to different animals, diets and time of analysis [13,14]; instrumental factors, such as nonlinear instrument responses, unknown baseline effects, and other foreseen or unforeseen anomalies [15–17].

Orthogonal Signal Correction (OSC) by Wold *et al*. [18] started exploring the idea of obtaining better and more easily interpretable models by filtering the information not related to the dependent variable (**y**) prior to the application of a calibration method like PLS. Several other algorithms followed: as a filter, with different implementations and slightly different results e.g. Sjöblom *et al*. [19], Andersson [7], Fearn [8], Wise and Gallagher [20], Westerhuis *et al*. [21], and Höskuldsson [22]), or as an integrated model including both predictive and orthogonal variation (as in Trygg and Wold [23]), and focusing on multiblock (e.g., as in Trygg [24], Trygg [25], and Löfstedt and Trygg [26]). Alternative integrated solutions by using post-PLS transformations (e.g. of Kvalheim *et al*. [27], Ergon [28], and Kemsley and Tapp [29]) or by using experimental variability in replicates (see Zhu *et al*. [30]) have also been developed. These methods have the capacity to further understand experimental variability and detect and quantify unforeseen phenomena by characterization of anomalies. Here, we provide an overview of examples where knowledge has been gained from inspection of orthogonal variation.

2. ORTHOGONAL METHODS AND PLS

The rationale behind most of the orthogonal methods used in conjunction with PLS has been to use it as a preprocessing filter for mainly calibration and classification. Because of inherent algorithmic differences, similar but not identical models were derived. As an example, the integrated Orthogonal Projections to Latent Structures (OPLS) method [23] together with the post-PLS transformations methods [27–29] models the same total variation in X for single-**y** variable models, but this can differ in the way they structure the two types of variations and how the algorithm is implemented [27,31].

* Correspondence to: Johan Trygg, Computational Life Science Cluster, Department of Chemistry, Umeå University, SE-90187 Umeå, Sweden.
E-mail: johan.trygg@chem.umu.se

a R. C. Pinto, J. Trygg
Computational Life Science Cluster, Department of Chemistry, Umeå University, SE-90187 Umeå, Sweden

b J. Gottfries
Department of Chemistry and Molecular Biology, Gothenburg University, SE-40530 Gothenburg, Sweden

Generally speaking, orthogonal methods separate data variation into three distinct types: correlated (predictive) variation, uncorrelated (orthogonal) variation, and model residuals. For single **y**, these methods model the matrix **X** in the following way:

$$\mathbf{X} = \mathbf{t}_p.\mathbf{p}_p + \mathbf{T}_o.\mathbf{P}_o + \mathbf{E} \quad (1)$$

in which **T** and **P** are the scores and loadings, respectively, and the subscripts indicate whether they explain the predictive (p) or the orthogonal (o) variation. The matrix **E** accounts for the residuals or noise [13]. Interpretation of PLS models is done by the loadings **P**, loading weights **W**, or the regression coefficients **B**. The interpretation of the coefficient profile (**b**), as derived from linear models $\mathbf{y} = \mathbf{X}.\mathbf{b} + \mathbf{f}$, may disagree with interpretations based on predictive loadings and t-scores. This is mainly because the multicomponent PLS regression coefficients needs to be rotated to compensate for the orthogonal variation in **X**, which is captured by the PLS model components. Much trust has been placed in the assumption of inherent interpretation from regression coefficient profiles (**b**) [23,32] in indirect multivariate calibrations without understanding the influence of the orthogonal variation.

It has been shown that predictions by PLS and orthogonal methods for single-**y** problems can perform equally well provided that identical model complexity and cross-validations are compared [23,33]. Nonetheless, the quantitative and qualitative interpretation and multiple subjective decisions required for constructing a valid prediction model has been shown to be critical for the outcome. This includes, for example, selection of observations, variables, scaling, preprocessing methods, transformations, and quality control. We argue that interpretation of orthogonal variation in combination with the predictive variation may improve decision-making, that is, OPLS can perform superior to PLS. This was exemplified in a recent large international study coordinated by the Food and Drug Administration and published in *Nature Biotechnology* [34]. The objective was to create validated prediction models for 13 clinical endpoints.

3. EXAMPLES

In its seminal Orthogonal Signal Correction paper, Wold *et al*. [18] understood already that the orthogonal variation loadings could “provide interesting hints for the causes of the undesired variation”. They made an effort to explain the orthogonal variation in their near-infrared data, attributing it to a baseline, scatter effects and to nonlinear relationships between **X** and **Y**.

Several studies have demonstrated the benefits of separating predictive and orthogonal information by providing a detailed understanding of the factors underlying both types of variability in the respective experiments. However, the use of PLS alone and other regression methods still dominate peer-reviewed publications. We firmly believe that orthogonal methods improve multivariate models because of the increased knowledge acquired by using orthogonal inspection. A number of illustrative examples are described here.

3.1. Unfolding spectral compartments

Berntsson *et al*. [35] performed a calibration transfer between Near Infrared (NIR) instruments used for the characterization of raw materials used in tablet production. Because it is well established that water induces large responses in parts of the NIR spectra, levels of water in the training set samples were varied. Thus, a reasonable level of diversity in the training set data with regard to alterations in moisture over time was obtained, which allowed successful application of PLS-discriminant analysis (PLS-DA) models without loss of predictive accuracy. However, Stenlund *et al*. [15] re-examined the data generated by OPLS-DA and suggested additional interpretations. One observed difference between instrument output performances was attributed to Wood’s anomaly [36], which was successfully represented by the predictive component. Meanwhile, the water content-related variation was contained in the orthogonal component. Thus, OPLS enabled separation of the instrument-related predictive information from the uncorrelated structure in the data; that is, water and its spectral regions were clearly visible in the orthogonal loadings. In summary, Wood’s anomaly effects were clearly modeled in the predictive component, which is logical because Wood’s anomaly is a true instrumental difference. However, water variation was contained in the orthogonal data modeling component, suggesting that even though the instruments possessed differences regarding Wood’s anomaly, they handled water signals similarly. Interestingly, the original analysis by PLS-DA mainly modeled water content in the first PLS component and Wood’s anomaly-related variation in PLS components 2–4. Interpreting both predictive and orthogonal **p**-loading profiles, together with raw spectra and the inherent absorbance patterns, provided a more complete understanding of the calibration transfer.

3.2. Improving experimental protocols

In a metabolomics study, genetically modified and wild-type (WT) poplar tree GC-MS data were analyzed by OPLS [37]. Discrimination between the genotypes was possible along the predictive component. In addition, the results indicated that there were two subgroups of samples in the orthogonal component for the WT. Analysis of the orthogonal loading profiles indicated that sucrose was the main metabolite responsible for within-group separation. Interestingly the subclustering pattern suggested that an uncontrolled sampling procedure had been used, that is, by scraping, which was deduced by orthogonal interpretation. Thus, the use of OPLS led to new knowledge about a potentially important variability factor related to the experimental procedure, which could be a target for improvement in subsequent investigations.

3.3. Keeping track of daily moisture variability

In this work, mixtures of carrageenans were analyzed using NIR, Mid Infrared (MIR), and Raman spectroscopy [38]. The aim of the work was to improve the analysis by modeling all the data simultaneously. This was achieved using a hierarchical PLS approach, which involved building three separate PLS models by performing regression on individual sets against a matrix containing the carrageenans proportions. All the significant scores vectors were then concatenated and regressed against the same **Y** matrix, producing an upper-level PLS model with enhanced prediction capability. This model still contained a mix of predictive and orthogonal data variation and, although good calibration results were obtained, the interpretability was not as clear as using an OPLS model.

By using OPLS (the same way as the aforementioned procedure for PLS), the eventual model separated pure predictive and pure orthogonal latent variables for each of the individual models. The selected predictive scores of each of the models were concatenated together to create a predictive OPLS upper-level model. In parallel, the cross-validated significant orthogonal scores were concatenated and modeled, this time by using Principal Components Analysis (PCA) for further interpretation. Although the predictive capabilities of PLS and OPLS methods were coinciding, the interpretation of the latent variables of OPLS was clearer, because it contained all predictive information separated from orthogonal variation. Furthermore, the orthogonal information was interpreted by looking at the **Y**-orthogonal PCA components, and scores clusters related to sampling days were found. Inspection of the orthogonal low-level models revealed that this time trend was caused by the NIR and MIR but not the Raman spectra. The loadings of the lower levels suggested that the key factor was moisture variation over time, and therefore, it made sense that the Raman data were not affected because of its low sensitivity to water content. Thus, OPLS modeling generated more commensurable information from the multiple data sets than PLS.

3.4. Checking the origin of uncorrelated information

Andersson [7] pointed at opportunities arising from inspecting the orthogonal variation. The direct orthogonalization (DO) method was applied to model the degree of esterification in dried pectin powders using mid-infrared spectroscopy. The orthogonal variation was inspected regarding the importance of variance among the extracted fractions. The orthogonal variation was decomposed using PCA, where scores and loadings were inspected to relate to the phenomena involved. The authors acknowledge the existence of a first orthogonal component without strong features and attributed it to the offset of the scatter. Then, resemblance between the second component and the general curvatures of the measured spectra was identified, which led to the conclusion that the inspected first component explained the dominating scatter signal. Furthermore, one peak identified in the third component was found to be important for description of the orthogonal variation. Thus, it was concluded that DO could improve the qualitative information by interpretation of the global modeling, providing better means of detecting outliers (like abnormal spectra containing low or high levels of uncorrelated spectral features), as well as insight into the origin of the observed background information. The tangible targets for suggested improvements included baseline drift such that identified scatter effects could be removed from the spectral data.

Beckwith-Hall *et al*. [39] applied OSC to ^{1}H-NMR spectral data allowing orthogonal variation inspection in combination with PLS, by which confounding in the original NMR data was explained. The entailing interpretation included several possible factors, for example instrumental, environmental, and physiological variations in the applied rat plasma samples. In the initial experiment, parallel analyses using two different NMR instruments were compared to evaluate interspectrometer variation and robustness of the NMR analysis. Using OSC, a large fraction of the global spectral data variation was removed, which was identified as interspectrometer variation. In the second experiment, on toxicological screening, the initial model data were identified as a drift and were subsequently removed, because the derived orthogonal scores correlated with the NMR experimental run order. Next, metabolic changes in rat urine after drug administration were studied, where the orthogonal component was shown to model within-group interanimal variation in response to the drug. Hence, a more accurate pharmacokinetic to pharmacological response to the drug could be defined.

Dumarey *et al*. [14] evaluated the antioxidant capacity of green tea with chromatographic fingerprints and multivariate analysis. By separating the predictive and orthogonal variability and analyzing the loadings, it was possible to identify peaks correlating with the antioxidant capacity of the tea, which was compared with a parallel analysis using PLS with and without an orthogonal filter. While inspecting the classic PLS regression coefficients, they noticed that the caffeine and other peaks were contributing to the model. However, the orthogonal decomposition PLS method indicated that these peaks were mainly present in the orthogonal fraction, but not in the predictive one, and thus not contributing to the antioxidant capacity of green tea. It was concluded that orthogonal-based PLS performed better because of the straightforward interpretation regarding reproducibility, interpretability, and predictive information.

3.5. Investigating common and specific variation in multiple data sets

Gabrielsson *et al*. [10] used a bi-directional orthogonal method to combine a matrix with itself after the first derivatives on UV spectra of the conversion of nitrobenzene to aniline. By inspecting the orthogonal variation, they could model an offset. Furthermore, it was shown that the patterns of the orthogonal scores were related to the rate of product formation, which was instrumental for future selections of preprocessing. Rantalainen *et al*. [40] used the OPLS framework to analyze multiblock data from a metabolomics–proteomics study in mice. Their first strategy was to use OPLS-DA to find the important variables for discrimination on each of the data blocks. Next, these variables were used in an OPLS model with the other block to investigate their relationship. The second strategy was a straightforward use of O2PLS to estimate how much of the variance was shared, as well as to find correlation patterns in the two blocks. After they separated the orthogonal variation (specific to each data block), they checked that there was no remaining discriminant power.

3.6. Inspecting nonlinear effects

Svensson *et al*. [33] investigated six orthogonal signal correction algorithms and their characteristics. Using a simulated data set with nonlinear features, they concluded the advantage of analyzing orthogonal variation, for example, to reveal nonlinear relations. They refer too that although the use of orthogonal methods may not result in lower prediction errors compared with the classic PLS, the advantage with those methods is its enhanced interpretability by concentrating that variation into one or few components.

3.7. Other examples

Further reports describe or investigate how to profit from orthogonal inspection. Bylesjö *et al*. [13] used orthogonal variability to help classification in a hybrid OPLS-DA–SIMCA (soft independent modeling of class analogy) model. Stenlund *et al*. [41] studied the Fourier-Transform Infrared (FT-IR) spectroscopy images of liver tissue and examined the origin of some of the orthogonal information. Kvalheim *et al*. [27] compared a postprocessing target projection method with OPLS, obtaining similar results. Bylesjö *et al*. [42] used OPLS for microarray normalization, identifying the origin of both predictive and orthogonal sources of variation. Gabrielsson *et al.* [11] applied an orthogonal methodology to multiblock batch process data, being able to understand the origin of the orthogonal information. A recently published orthogonal method by Löfstedt and Trygg [26] allows one to find the joint variation and the unique orthogonal variation in multiple (>2) block modeling. Trygg and Wold [23] and Gottfries *et al.* [9] highlight the importance of analyzing the orthogonal information to identify and understand it, aiming at future process improvements. The orthogonal framework has also been evaluated by different authors [43–47] and used in different contexts [48–57].

4. DISCUSSION

Modern biology and medicine have created fantastic new opportunities for the field of chemometrics to make a real impact. The overwhelming size and complexity of the "omics" technologies (genomics, proteomics, metabolomics) have turned biology and medicine to focus more on the adoption of chemometrics. At the same time, this development has prompted chemometricians to pay more attention to model interpretation and transparency. The concept of orthogonal variation, introduced in the landmark OSC (1998) [18] and OPLS (2002) [23] papers, has been an important contributor. Biological and medical applications are key areas for chemometrics' strong development and impact.

We have provided a conceptual explanation of the advantages of orthogonal variation inspection in the context of PLS multivariate classification and calibration. In several examples, advantages revealed by inspection of orthogonal variation were highlighted, which were not apparent otherwise. We have proposed that by inspecting both predictive and orthogonal variation, model interpretation and information quality are improved. As mentioned, a large comparative study published by the MicroArray Quality Control Consortium [34] concluded that the only orthogonal-based method in that study (OPLS) was also the best single method for the prediction of 13 different clinical endpoints, whereas the performance of PLS was substantially worse. This increased performance might very well have been the result of parallel increased knowledge gained by using OPLS, which guided the subjective decision-making through the modeling task.

Philosophically speaking, new knowledge often emanates from empirical sources and therefore is not usually widely accepted without rigorous scrutiny and corroboration by many research articles. Thus, anomalous results may be regarded as questionable or as exceptions rather than being seriously appreciated. Thomas Kuhn [58] devoted much attention to the importance of proper analysis of anomalies in the quantity and quality of revolutionary scientific steps. We propose that orthogonal modeling is a useful discovery analysis tool that reveals the structure of such information. The present paper describes three examples of analyses of data, all of them including "anomalous" variation, ranging from the effect of water content in calibration samples to the influence of daily variations in ambient humidity. In all cases, inspection of orthogonal components allowed the "anomaly" to be distinguished by a structured approach based on orthogonal loadings and scores. Thus, orthogonal methods reveal additional properties compared with previous correlation methods because they can constructively present unexpected phenomena captured by the description matrices. By using orthogonal methods, we can start to unravel why and which data structures end up being uncorrelated and thus find suspected anomalies in the data or data generation procedure. Once there is such a new tool, to view nature with improved potential for understanding of discovery, there is little excuse to its absence. We therefore foresee that the introduction of orthogonal methods will have a similar impact and influence to the field of chemometrics as the development of the PLS method once had.

REFERENCES

1. Talebpour Z, Tavallaie R, Ahmadi SH, Abdollahpour A. Simultaneous determination of penicillin G salts by infrared spectroscopy: evaluation of combining orthogonal signal correction with radial basis function-partial least squares regression. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 2010; **76**(5): 452–457.
2. Kang J, Choi M, Kang S, Kwon H, Wen H, Lee CH, Park M, Wiklund S, Kim HJ, Kwon SW, Park S. Application of a (1)H nuclear magnetic resonance (NMR) metabolomics approach combined with orthogonal projections to latent structure-discriminant analysis as an efficient tool for discriminating between Korean and Chinese herbal medicines. *J. Agric. Food Chem.* 2008; **56**(24): 11589–11595.
3. Igne B, Roger JM, Roussel S, Bellon-Maurel V, Hurburgh CR. Improving the transfer of near infrared prediction models by orthogonal methods. *Chemom. Intell. Lab. Syst.* 2009; **99**(1): 57–65.
4. Eriksson L, Antti H, Gottfries J, Holmes E, Johansson E, Lindgren F, Long I, Lundstedt T, Trygg J, Wold S. Using chemometrics for navigating in the large data sets of genomics, proteomics, and metabonomics (gpm). *Anal. Bioanal. Chem.* 2004; **380**(3): 419–429.
5. Akesson L, Gabrielsson J, Pefley R, Trygg J, Bruce S, Fuller JM, Lundstedt T, Stenlund H, Lernmark A, Madsen R, Tupling T, Moritz T. Serum metabolite signature predicts the acute onset of diabetes in spontaneously diabetic congenic BB rats. *Metabolomics* 2011; 1–11.
6. Woody NA, Feudale RN, Myles AJ, Brown SD. Transfer of multivariate calibrations between four near-infrared spectrometers using orthogonal signal correction. *Anal. Chem.* 2004; **76**(9): 2595–2600.
7. Andersson CA. Direct orthogonalization. *Chemom. Intell. Lab. Syst.* 1999; **47**(1): 51–63.
8. Fearn T. On orthogonal signal correction. *Chemom. Intell. Lab. Syst.* 2000; **50**(1): 47–52.
9. Gottfries J, Johansson E, Trygg J. On the impact of uncorrelated variation in regression mathematics. *J. Chemom.* 2008; **22**(11–12): 565–570.
10. Gabrielsson J, Jonsson H, Airiau C, Schmidt B, Escott R, Trygg J. OPLS methodology for analysis of pre-processing effects on spectroscopic data. *Chemom. Intell. Lab. Syst.* 2006; **84**(1–2): 153–158.
11. Gabrielsson J, Jonsson H, Airiau C, Schmidt B, Escott R, Trygg J. The OPLS methodology for analysis of multi-block batch process data. *J. Chemom.* 2006; **20**(8–10): 362–369.
12. Oprea TI, Gottfries J. Chemography: the art of navigating in chemical space. *J. Comb. Chem.* 2001; **3**(2): 157–166.
13. Bylesjö M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemom.* 2006; **20**(8–10): 341–351.
14. Dumarey M, van Nederkassel AM, Deconinck E, Heyden YV. Exploration of linear multivariate calibration techniques to predict the total antioxidant capacity of green tea from chromatographic fingerprints. *J. Chromatogr. A* 2008; **1192**(1): 81–88.
15. Stenlund H, Johansson E, Gottfries J, Trygg J. Unlocking interpretation in near infrared multivariate calibrations by orthogonal partial least squares. *Anal. Chem.* 2009; **81**(1): 203–209.
16. Malinowski ER, Mccue M. Qualitative and quantitative—determination of suspected components in mixtures by target transformation factor-analysis of their mass-spectra. *Anal. Chem.* 1977; **49**(2): 284–287.
17. Christie OHJ. Data laundering by target rotation in chemistry-based oil exploration. *J. Chemom.* 1996; **10**(5–6): 453–461.
18. Wold S, Antti H, Lindgren F, Ohman J. Orthogonal signal correction of near-infrared spectra. *Chemom. Intell. Lab. Syst.* 1998; **44**(1–2): 175–185.
19. Sjöblom J, Svensson O, Josefson M, Kullberg H, Wold S. An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra. *Chemom. Intell. Lab. Syst.* 1998; **44**(1–2): 229–244.
20. Wise BM, Gallagher NB. Orthogonal signal correction. http://www.eigenvector.com/MATLAB/OSC.html (accessed 28 November 2011).

21. Westerhuis JA, de Jong S, Smilde AK. Direct orthogonal signal correction. *Chemom. Intell. Lab. Syst.* 2001; **56**(1): 13–25.
22. Höskuldsson A. Variable and subset selection in PLS regression. *Chemom. Intell. Lab. Syst.* 2001; **55**(1–2): 23–38.
23. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). *J. Chemom.* 2002; **16**(3): 119–128.
24. Trygg J, O2-PLS for qualitative and quantitative analysis in multivariate calibration. *J. Chemom.* 2002; **16**(6): 283–293.
25. Trygg J, Wold S. O2-PLS, a two-block (X-Y) latent variable regression (LVR) method with an integral OSC filter. *J. Chemom.* 2003; **17**(1): 53–64.
26. Löfstedt T, Trygg J. OnPLS—a novel multiblock method for the modelling of predictive and orthogonal variation. *J. Chemom.* 2011; **25**(8): 441–455.
27. Kvalheim OM, Rajalahti T, Arneberg R. X-tended target projection (XTP)-comparison with orthogonal partial least squares (OPLS) and PLS post-processing by similarity transformation (PLS plus ST). *J. Chemom.* 2009; **23**(1–2): 49–55.
28. Ergon R. PLS post-processing by similarity transformation (PLS plus ST): a simple alternative to OPLS. *J. Chemom.* 2005; **19**(1): 1–4.
29. Kemsley EK, Tapp HS. OPLS filtered data can be obtained directly from non-orthogonalized PLS1. *J. Chemom.* 2009; **23**(5–6): 263–264.
30. Zhu Y, Fearn T, Samuel D, Dhar A, Hameed O, Bown SG, Lovat LB. Error removal by orthogonal subtraction (EROS): a customised pretreatment for spectroscopic data. *J. Chemom.* 2008; **22**(1–2): 130–134.
31. Trygg J. Prediction and spectral profile estimation in multivariate calibration. *J. Chemom.* 2004; **18**(3–4): 166–172; Biagioni DJ, Astling DP, Graf P, Davis MF. Orthogonal projection to latent structures solution properties for chemometrics and systems biology data. *J. Chemom.* 2011; **25**(9): 514–525.
32. Eriksson L, Johansson E, Kettaneh-Wold N, Trygg J, Wikström C, Wold S. *Multi- and Megavariate Data Analysis. Part I, Basic Principles and Applications*. Umetrics AB: Umeå, Sweden, 2006.
33. Svensson O, Kourti T, MacGregor JF. An investigation of orthogonal signal correction algorithms and their characteristics. *J. Chemom.* 2002; **16**(4): 176–188.
34. Shi LM, Campbell G, Jones WD, Campagne F, Wen ZN, Walker SJ, Su ZQ, Chu TM, Goodsaid FM, Pusztai L, Shaughnessy JD, Oberthuer A, Thomas RS, Paules RS, Fielden M, Barlogie B, Chen WJ, Du P, Fischer M, Furlanello C, Gallas BD, Ge XJ, Megherbi DB, Symmans WF, Wang MD, Zhang J, Bitter H, Brors B, Bushel PR, Bylesjö M, Chen MJ, Cheng J, Cheng J, Chou J, Davison TS, Delorenzi M, Deng YP, Devanarayan V, Dix DJ, Dopazo J, Dorff KC, Elloumi F, Fan JQ, Fan SC, Fan XH, Fang H, Gonzaludo N, Hess KR, Hong H, Huan J, Irizarry RA, Judson R, Juraeva D, Lababidi S, Lambert CG, Li L, Li Y, Li Z, Lin SM, Liu GZ, Lobenhofer EK, Luo J, Luo W, McCall MN, Nikolsky Y, Pennello GA, Perkins RG, Philip R, Popovici V, Price ND, Qian F, Scherer A, Shi TL, Shi WW, Sung JY, Thierry-Mieg D, Thierry-Mieg J, Thodima V, Trygg J, Vishnuvajjala L, Wang SJ, Wu JP, Wu YC, Xie Q, Yousef WA, Zhang L, Zhang XG, Zhong S, Zhou YM, Zhu S, Arasappan D, Bao WJ, Lucas AB, Berthold F, Brennan RJ, Buness A, Catalano JG, Chang C, Chen R, Cheng YY, Cui J, Czika W, Demichelis F, Deng XT, Dosymbekov D, Eils R, Feng Y, Fostel J, Fulmer-Smentek S, Fuscoe JC, Gatto L, Ge WG, Goldstein DR, Guo L, Halbert DN, Han J, Harris SC, Hatzis C, Herman D, Huang JP, Jensen RV, Jiang R, Johnson CD, Jurman G, Kahlert Y, Khuder SA, Kohl M, Li JY, Li L, Li ML, Li QZ, Li S, Li ZG, Liu J, Liu Y, Liu ZC, Meng L, Madera M, Martinez-Murillo F, Medina I, Meehan J, Miclaus K, Moffitt RA, Montaner D, Mukherjee P, Mulligan GJ, Neville P, Nikolskaya T, Ning BT, Page GP, Parker J, Parry RM, Peng XJ, Peterson RL, Phan JH, Quanz B, Ren Y, Riccadonna S, Roter AH, Samuelson FW, Schumacher MM, Shambaugh JD, Shi QA, Shippy R, Si SZ, Smalter A, Sotiriou C, Soukup M, Staedtler F, Steiner G, Stokes TH, Sun QL, Tan PY, Tang R, Tezak Z, Thorn B, Tsyganova M, Turpaz Y, Vega SC, Visintainer R, von Frese J, Wang C, Wang E, Wang JW, Wang W, Westermann F, Willey JC, Woods M, Wu SJ, Xiao NQ, Xu J, Xu L, Yang L, Zeng X, Zhang JL, Zhang L, Zhang M, Zhao C, Puri RK, Scherf U, Tong W, Wolfinger RD, Consortium M. The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat. Biotechnol.* 2010; **28**(8): 827–U109.
35. Berntsson O, Danielsson LG, Lagerholm B, Folestad S. Quantitative in-line monitoring of powder blending by near infrared reflection spectroscopy. *Powder Technol.* 2002; **123**(2–3): 185–193.
36. Wood RW. On a remarkable case of uneven distribution of light in a diffraction grating spectrum. *Philos. Mag.* 1902; **4**: 396–402.
37. Wiklund S, Johansson E, Sjostrom L, Mellerowicz EJ, Edlund U, Shockcor JP, Gottfries J, Moritz T, Trygg J. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Anal. Chem.* 2008; **80**(1): 115–122.
38. Eriksson L, Toft M, Johansson E, Wold S, Trygg J. Separating Y-predictive and Y-orthogonal variation in multi-block spectral data. *J. Chemom.* 2006; **20**(8–10): 352–361.
39. Beckwith-Hall BM, Brindle JT, Barton RH, Coen M, Holmes E, Nicholson JK, Antti H. Application of orthogonal signal correction to minimise the effects of physical and biological variation in high resolution H-1 NMR spectra of biofluids. *Analyst* 2002; **127**(10): 1283–1288.
40. Rantalainen M, Cloarec O, Beckonert O, Wilson ID, Jackson D, Tonge R, Rowlinson R, Rayner S, Nickson J, Wilkinson RW, Mills JD, Trygg J, Nicholson JK, Holmes E. Statistically integrated metabonomic-proteomic studies on a human prostate cancer xenograft model in mice. *J. Proteome Res.* 2006; **5**(10): 2642–2655.
41. Stenlund H, Gorzsas A, Persson P, Sundberg B, Trygg J. Orthogonal projections to latent structures discriminant analysis modeling on in situ FT-IR spectral imaging of liver tissue for identifying sources of variability. *Anal. Chem.* 2008; **80**(18): 6898–6906.
42. Bylesjö M, Eriksson D, Sjodin A, Jansson S, Moritz T, Trygg J. Orthogonal projections to latent structures as a strategy for microarray data normalization. *BMC Bioinformatics* 2007; **8**: 207.
43. Cloarec O, Dumas ME, Craig A, Barton RH, Trygg J, Hudson J, Blancher C, Gauguier D, Lindon JC, Holmes E, Nicholson J. Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic H-1 NMR data sets. *Anal. Chem.* 2005; **77**(5): 1282–1289.
44. Esteban-Diez I, Gonzalez-Saiz JM, Pizarro C. An evaluation of orthogonal signal correction methods for the characterisation of arabica and robusta coffee varieties by NIRS. *Anal. Chim. Acta* 2004; **514**(1): 57–67.
45. Yu HL, MacGregor JF. Post processing methods (PLS-CCA): simple alternatives to preprocessing methods (OSC-PLS). *Chemom. Intell. Lab. Syst.* 2004; **73**(2): 199–205.
46. Ni WD, Brown SD, Man RL. The relationship between net analyte signal/preprocessing and orthogonal signal correction algorithms. *Chemom. Intell. Lab. Syst.* 2009; **98**(2): 97–107.
47. Goicoechea HC, Olivieri AC. A comparison of orthogonal signal correction and net analyte preprocessing methods. Theoretical and experimental study. *Chemom. Intell. Lab. Syst.* 2001; **56**(2): 73–81.
48. Esteban-Diez I, Gonzalez-Saiz JM, Pizarro C. OWAVEC: a combination of wavelet analysis and an orthogonalization algorithm as a preprocessing step in multivariate calibration. *Anal. Chim. Acta* 2004; **515**(1): 31–41.
49. Esteban-Diez I, Gonzalez-Saiz JM, Pizarro C, Generalization of OWAVEC method for simultaneous noise suppression, data compression and orthogonal signal correction. *Anal. Chim. Acta* 2005; **544**(1–2): 89–99.
50. Yee NG, Coghill GG. Orthogonal signal correction of potato crisp near infrared spectra. *J. Chin. Inst. Chem. Eng.* 2003; **34**(5): 563–574.
51. Yee NG, Coghill GG. Factor selection strategies for orthogonal signal correction applied to calibration of near-infrared spectra. *Chemom. Intell. Lab. Syst.* 2003; **67**(2): 145–156.
52. Shen Q, Jiang JH, Shen GL, Yu RQ. Ridge estimated orthogonal signal correction for data preprocessing prior to PLS modeling: QSAR studies of cyclooxygenase-2 inhibitors. *Chemom. Intell. Lab. Syst.* 2006; **82**(1–2): 44–49.
53. Kim K, Lee JM, Lee IB. A novel multivariate regression approach based on kernel partial least squares with orthogonal signal correction. *Chemom. Intell. Lab. Syst.* 2005; **79**(1–2): 22–30.
54. Whelehan OP, Earll ME, Johansson E, Toft M, Eriksson L. Detection of ovarian cancer using chemometric analysis of proteomic profiles. *Chemom. Intell. Lab. Syst.* 2006; **84**(1–2): 82–87.
55. Feudale RN, Liu Y, Woody NA, Tan HW, Brown SD. Wavelet orthogonal signal correction. *J. Chemom.* 2005; **19**(1): 55–63.
56. Goodarzi M, Shahbazikhah P, Sohrabi MR, Fathabadi M, Nouri SH. Direct orthogonal signal correction-partial least squares for simultaneous spectrophotometric determination of sulfamethoxazole and trimethoprim in pharmaceutical formulation and synthetic samples. *J. Chil. Chem. Soc.* 2009; **54**(3): 309–313.
57. Niazi A, Goodarzi M. Orthogonal signal correction-partial least squares method for simultaneous spectrophotometric determination of cypermethrin and tetramethrin. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 2008; **69**(4): 1165–1169.
58. Kuhn TS. *The Structure of Revolutions*. UC Press: Chicago, 1962.