Analytica Chimica Acta 885 (2015) 17–32

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Review

Review of validation and reporting of non-targeted fingerprinting approaches for food authentication

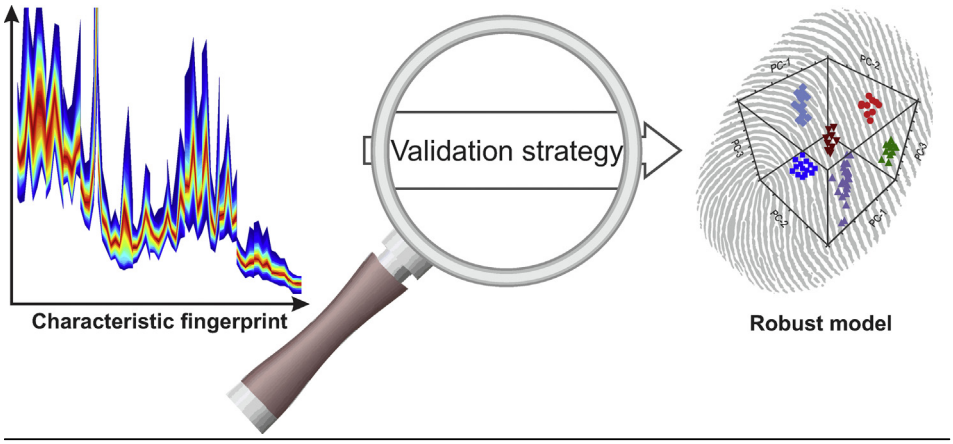
Janet Riedl, Susanne Esslinger, Carsten Fauhl-Hassek*

BfR Federal Institute for Risk Assessment, Department of Safety in the Food Chain, Max-Dohrn-Str. 8-10, D-10589 Berlin, Germany

HIGHLIGHTS

- The reliability of food fingerprinting studies has been critically reviewed.
- A validation scheme for multivariate statistical models is proposed.
- Recommendations to improve reporting quality are provided.
- Gaps of current validation practice are identified.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 13 February 2015
Received in revised form 13 May 2015
Accepted 2 June 2015
Available online 11 June 2015

Keywords:
Non-targeted analysis
Food fingerprinting
Authenticity

ABSTRACT

Food fingerprinting approaches are expected to become a very potent tool in authentication processes aiming at a comprehensive characterization of complex food matrices. By non-targeted spectrometric or spectroscopic chemical analysis with a subsequent (multivariate) statistical evaluation of acquired data, food matrices can be investigated in terms of their geographical origin, species variety or possible adulterations. Although many successful research projects have already demonstrated the feasibility of non-targeted fingerprinting approaches, their uptake and implementation into routine analysis and food surveillance is still limited. In many proof-of-principle studies, the prediction ability of only one data set was explored, measured within a limited period of time using one instrument within one laboratory. Thorough validation strategies that guarantee reliability of the respective data basis and that allow

Abbreviations: ANN, artificial neural networks; BP, back-propagation; CART, classification and regression trees; CC, correct classification; CP, counter-propagation; CV, cross validation; DA, discriminant analysis; DART, direct analysis in real time; DP, discrimination power; GA, genetic algorithm; HSI, hyperspectral imaging; kNN, *k*-nearest neighbors; LIT, linear ion trap; LDA, linear discriminant analysis; LOO, leave-one-out; LS, least-squares; NIST, National Institute of Standards and Technology; MC, Monte Carlo simulation; MD, Mahalanobis distance; MLP, multilayer perceptrons; MS, mass spectrometry; NMC, number of misclassification; NPV, negative predicted values; OPLS, orthogonal PLS; PCA, principal component analysis; PCR, principal component regression; PLS, partial least squares analysis; PLSCM, PLS class model; PPV, positive predicted values; PRESS, predicted residuals error sum of squares; PTR, proton transfer reaction; $Q^2(X/Y)$, prediction ability for *X*/*Y*-data; QqQ, triple quadrupole; *R*/*r*, correlation coefficient; R^2, coefficient of determination; $R^2(X/Y)$, recognition ability or explained variation in *X*/*Y*-data; RDA, regularized discriminant analysis; RDC/CV/P, relative deviation of calibration/cross validation/prediction; RESS/RSS, residual sum of squares; RMSEC/CV/P, root mean squared error of calibration/cross validation/prediction; RPD, residual predictive deviation or ratio of prediction to deviation; SEC/P/V, standard error of calibration/prediction/validation; SESI, secondary electrospray ionization; SIMCA, soft independent modelling of class analogy; SLDA, stepwise linear discriminant analysis; SSQ*X*/*Y*, cumulative sum of squares for *X*/*Y*-data; SVDD, support vector data description; SVM, support vector machines; UNEQ, unequal dispersed class; WPTER, wavelet packet transform for efficient pattern recognition.

* Corresponding author.
E-mail address: carsten.fauhl-hassek@bfr.bund.de (C. Fauhl-Hassek).

http://dx.doi.org/10.1016/j.aca.2015.06.003

Statistical model validation
System challenge
Analytical method validation

conclusion on the applicability of the respective approaches for its fit-for-purpose have not yet been proposed. Within this review, critical steps of the fingerprinting workflow were explored to develop a generic scheme for multivariate model validation. As a result, a proposed scheme for "good practice" shall guide users through validation and reporting of non-targeted fingerprinting results. Furthermore, food fingerprinting studies were selected by a systematic search approach and reviewed with regard to (a) transparency of data processing and (b) validity of study results. Subsequently, the studies were inspected for measures of statistical model validation, analytical method validation and quality assurance measures. In this context, issues and recommendations were found that might be considered as an actual starting point for developing validation standards of non-targeted metabolomics approaches for food authentication in the future. Hence, this review intends to contribute to the harmonization and standardization of food fingerprinting, both required as a prior condition for the authentication of food in routine analysis and official control.

Contents

1. Introduction . . . 19
2. Validation framework from raw data to fit-for-purpose models . . . 19
 2.1. Data preparation . . . 19
 2.2. Multivariate statistical modelling . . . 20
 2.3. Validation scheme for multivariate models . . . 20
3. Validation practice in food authentication studies . . . 21
 3.1. Model input information . . . 24
  3.1.1. Variables . . . 24
  3.1.2. Objects . . . 24
  3.1.3. Model performance and model input characterization . . . 25
 3.2. Statistical model validation and diagnostics . . . 25
  3.2.1. Performance indicators for multivariate regression models . . . 25
  3.2.2. Performance indicators for classification models . . . 26
 3.3. System challenge approaches . . . 27
 3.4. Method validation parameter and quality assurance measures . . . 27
4. Recommendations and future work . . . 28
 4.1. Reporting quality . . . 28
 4.2. Validation strategies . . . 29
 Supplementary data . . . 30
 References . . . 30

Dr Janet Riedl graduated in Chemistry and Toxicology at the University of Leipzig (Germany) and received her PhD for the application of metabolomics in ecotoxicology from the Technical University Bergakademie Freiberg (Germany). From 2008 to 2012 she investigated the metabolome of higher aquatic plants with GC–MS at the Helmholtz Centre of Environmental Research in Leipzig. Currently she is a post-doctoral researcher at the Federal Institute for Risk Assessment in Berlin (Germany) involved in the fingerprinting analysis of wine and spices with NMR and FT-IR. Her focus is on plant systems and pattern-based data analysis in ecotoxicology and food authenticity.

Dr Carsten Fauhl-Hassek is a food chemist and works at the department for Safety in the Food Chain based within the Federal Institute for Risk Assessment (BfR) in Germany. He is head of the unit Product Identify, Supply Chains and Traceability and has special expertise in wine analysis and appreciation, method development and validation (ring trials), authentication of food and feed, NMR and stable isotope analysis.

Dr Susanne Esslinger graduated in Food Chemistry and earned a PhD in Analytical Chemistry from the Humboldt-University of Berlin (Germany). From 2007 to 2010 she has been working at the BAM Federal Institute for Materials Research and Testing (Berlin), where she has been involved in studies of polybrominated flame retardants. Thereafter, Susanne Esslinger has been engaged at the Federal Institute for Risk Assessment (BfR) in scientific research. Her work involves the development/establishment of new techniques to verify food and feed authenticity as well as traceability using diverse analytical techniques. She has been involved in several projects funded by European agencies.