Spectroscopic techniques applied to the real time monitoring and control of the production processes of the pharmaceutical and food industry

Spectroscopic techniques applied to the real time monitoring and control of the production processes of the pharmaceutical and food industry

Gentian Gavoci
Supervisor: Prof. Francesco Geobaldo
Co-Supervisor: Prof. Francesco Savorani

The enclosed dissertation is about the use of spectroscopic techniques, mainly near infrared spectroscopy (NIRS), in combination with multivariate data analysis for the real time monitoring and control of the production processes of the pharmaceutical and food industry. The expected final product quality is the main task to accomplish for the industrial organizations taking also into account the costs and environmental impact. Thus far, the quality assessment of manufactured products is performed primarily on a post-production based testing, using an off-line laboratory strategy. This approach may result in products that do not meet the requested quality characteristics and should be disposed or reworked. The implementation of the strategy based on the Process Analytical Technology (PAT) concept, accepted as an effective tool for process monitoring and control, represents an innovative solution to avoid unwanted consequences stemming from the abovementioned quality control approach. The PAT paradigm involves an active process control, starting from the quality control of incoming raw materials and through the continuous process control, leading to semi-finished or final products within specifications having little variation with respect to the critical quality parameters. Spectroscopic sensors, in combination with computational analysis (multivariate data analysis), are regarded as the most advantageous process analysers for PAT successful implementation, moving the quality measurements closer to the process via at-line, on-line and in-line strategies. PAT, allowing real time collection, analysis and sharing of production process data, becomes a powerful tool in the manufacturing step included in the management of the value chain over the life cycle of pharmaceutical and food products as addressed by Industry 4.0.

The research activity is focused on three industrial research case studies.

The first case study, concerns a pharmaceutical industry which mainly deals with the manufacturing of dietary supplements, has been focused on the “real-time” monitoring of raw materials and quality control of semi-finished and/or end products.
Raw materials. In order to assure the quality of the end products a series of analysis need to be performed on the incoming raw materials, then on the semi-finished products and lastly on the end products. With respect to the solid raw materials, the NIR spectra have been first acquired and then subjected to multivariate data analysis (pre-processing and modelling) in order to develop chemometric models for identity confirmation and classification purposes. The raw materials have been purchased by the company over a period of three years and consist of botanical dry extracts, pure substances and other solids. For the raw materials with a sufficient number of lots have been developed pattern recognition models based on the PCA (principal component analysis) algorithm and classification models taking advantage of the PLS-DA (partial least squares-discriminant analysis) method. The qualitative models developed showed the potential for identity confirmation or classification of even more complex ingredients such as dry extracts. One of the developed classification models related to Ginkgo biloba dry extracts has been successfully implemented within the company enabling in this way the qualitative prediction of new dry extracts.
Semi-finished product. Concerning the semi-finished and/or end products, a quantitative model has been developed according to a design made of 11 concentration levels ranging from 5% to 10% of the active constituent (docosahexaenoic acid) taking into account the production process and specification limits. To accomplish this objective, the NIR spectra of 22 standard mixtures (considering the replicates) have been collected in diffuse reflectance modality, pre-processed by using several combinations of techniques and lastly subjected to the PLS (partial least squares) linear regression algorithm. The best and the simplest model, derived from SNV (standard normal variate)

+ MC (mean centring) pre-processed NIR spectral data, resulted in a RMSECV (root mean square error of cross-validation) of 0.31 and in a R^2 (determination coefficient) of 0.96. The quantity of DHA in two end products have been predicted showing the allocation of samples within the specification limits.

The second case study, involved a company of vegetable oils, has been focused on four objectives: 1. The assessment of the shelf life of vegetable oils; 2. Quality evaluation of extra virgin olive oils as a function of the storage time; 3. The effect of two cold pressing systems on five vegetable oils produced starting from their seeds; 4. The evaluation of cold-pressed linseed oil oxidative stability when subjected to accelerated oxidation.
Shelf-life assessment. Hemp, linseed and sunflower oils have been yielded via mechanical extraction using a cold pressing system (expeller press). The oils stored in dark glass bottles have been subjected to artificial radiation under NEON and LED light having the same light power and colour, switching between light and dark. At time intervals of two months, the vegetable oils have been first analysed with the UV-VIS spectroscopic technique and NIRS and afterwards the spectral data collected have been subjected to qualitative modelling. The PCA model build on the visible spectral region provided a clear separation of the hemp oil after four months of exposure to the two light sources with a more drastic effect of NEON light on the chlorophylls decomposition.
Extra virgin olive oil. Two extra virgin olive oils have been stored in two tanks with a six months lapse of storage time between the oils. At intervals of 15 days the oils have been subjected to NIR analysis and chemometric modelling, showing a storage effect on the PCA scores plot according to the trend of oil temperature variation over the storage period.
Comparison between two cold pressing systems. Five seeds (hemp, linseed, sunflower, pumpkin and walnut) have been subjected to two cold pressing systems (expeller press and hydraulic press), each of them working under two pressing conditions. The PCA models built on the NIR spectral data displayed a separation related only to walnut oil due to the two different of cold pressing systems.
Cold-pressed linseed oil oxidative stability. The linseed oil yielded from the cold-pressed seeds using an expeller press at two pressing speeds (20% and 80%) has been subjected to accelerated oxidation. The NIR spectra of the linseed oil before and after the oxidative treatment have been collected and subjected to PCA modelling accordingly. The effect of the two pressing speeds is more apparent after the accelerated oxidation ad shown by the different scores on the PCA model.

The third case study concerns the coffee industry and has, as its main aim, the classification of green coffee beans according to their geographical origin using NIRS and chemometrics.
The set of green coffee beans considered came from several countries located in two continents. The samples as such have been subjected to NIR testing in two laboratories employing the same spectrometer model. The PCA algorithm and the classification PLS-DA algorithm have been applied on the pre-processed spectral data. Both the PCA models calculated on the spectral data collected in the two laboratories, and pre-processed using SNV + MC, displayed a rather good separation of the coffee samples according to both the continent and country of origin. The performance of the PLS-DA models developed has been evaluated based on the prediction of the external test sets showing good predictions ranging from 90% to 98% concerning the country-based classification models. Moreover, a variable selection on the whole spectrum using iPLS-DA (intervalPLS-DA) algorithm has been performed. The performance of the iPLS-DA models resulted better, either for the classification based on continent or for the country-based classification, showing in the second case a prediction ranging from 95% to 100%. In both the cases the higher prediction was achieved using SNV + MC as pre-processing methods.