Journal of Pharmaceutical and Biomedical Analysis 126 (2016) 156–162

View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Journal of Pharmaceutical and Biomedical Analysis

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

New frontiers in pharmaceutical analysis: A metabolomic approach to check batch compliance of complex products based on natural substances☆

L. Mattoli[a,*], M. Burico[a], G. Fodaroni[a], S. Tamimi[a], S. Bedont[a], P. Traldi[b], M. Stocchero[c]

[a] Ricerca Fitochimica, Aboca SpA Società Agricola, loc. Aboca 20, 52037 Sansepolcro (AR), Italy
[b] IENI, CNR, C.so Stati Uniti, 4, Padova, Italy
[c] S-IN, Soluzioni Informatiche, via G. Ferrari 14, 36100 Vicenza, Italy

ARTICLE INFO

Article history:
Received 31 December 2015
Received in revised form 8 April 2016
Accepted 9 April 2016
Available online 13 April 2016

Keywords:
Medicinal plants
Metabolomic
Batch quality comparison

ABSTRACT

Natural substances, particularly medicinal plants and their extracts, are still today intended as source for new Active Pharmaceutical Ingredients (**APIs**). Alternatively they can be validly employed to prepare medicines, food supplements or medical devices. The most adopted analytical approach used to verify quality of natural substances like medicinal plants is based still today on the traditional quantitative determination of marker compounds and/or active ingredients, besides the acquisition of a fingerprint by TLC, NIR, HPLC, GC.

Here a new analytical approach based on untargeted metabolomic fingerprinting by means of Mass Spectrometry (MS) to verify the quality of *grinTuss adulti syrup*, a complex products based on medicinal plants, is proposed. Recently, untargeted metabolomic has been successfully applied to assess quality of natural substances, plant extracts, as well as corresponding formulated products, being the complexity a resource but not necessarily a limit. The untargeted metabolomic fingerprinting includes the monitoring of the main constituents, giving weighted relevance to the most abundant ones, but also considering minor components, that might be notable in view of an integrated – often synergistic – effect on the biological system.

Two different years of production were investigated. The collected samples were analyzed by Flow Injection ElectroSpray Ionization Mass Spectrometry Analysis (FIA-ESI-MS) and a suitable data processing procedure was developed to transform the MS spectra into robust fingerprints. Multivariate Statistical Process Control (MSPC) was applied in order to obtain multivariate control charts that were validated to prove the effectiveness of the proposed method.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Natural substances as medicinal plants have been used since ancient times in the folk medicine for treating a broad range of diseases. Although the last decade has witnessed a marked growth in the market of plant-based natural products, their high complexity in terms of composition still makes a challenging task the guarantee of quality, efficacy and safety requirements [1,2]. In general, quality control of natural substances as medicinal plants is based on three important pharmacopoeias definitions: identity, purity and content or assay. To prove identity and purity, criteria such as type of preparation sensory properties, physical constants, contaminants, moisture, ash content and solvent residues have to be checked. Content or assay is the most difficult area of quality control to perform, since in most natural substances as medicinal plants and corresponding formulated products the active constituents are not known. When the active constituents (e.g. sennosides in Senna) or markers (e.g. alkylamides in Echinacea) are known, a vast array of modern chemical analytical methods such as Ultraviolet/Visible spectroscopy (UV/Vis), Thin Layer Chromatography (TLC), High-Performance Liquid Chromatography (HPLC), Gas Chromatography (GC), Mass Spectrometry (MS), or a combination of GC or LC and MS, can be employed [3]. In all other cases, where no active constituent or marker can be defined, it seems to be necessary to determine the most of the chemical constituents in order to ensure the reliability and repeatability of pharmacological and clinical research,

☆ Selected paper from conference of Recent Development in Pharmaceutical Analysis (RDPA2015), June 28–July 1, 2015, Perugia, Italy.

* Corresponding author.
E-mail address: lmattoli@aboca.it (L. Mattoli).

http://dx.doi.org/10.1016/j.jpba.2016.04.010
0731-7085/© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Table 1
Workflow for developing an analytical method based on metabolomics.

STEP	DEFINITION
Pre-processing	Generic term for methods to go from raw instrumental data to clean data for processing.
Pre-treatment	Transforming the clean data to make them ready for data processing.
Processing	Data analysis (PCA, PLS, etc).
Post-processing	Transforming the results from the processing for interpretation and visualization (e.g. antilog).
Validation	All the activities aimed at assuring the quality of the conclusion drawn from the data analysis.
Interpretation	Hypothesis generation, visualization of the data.

Table 2
Quali-quantitative composition of the 'bad' quality samples.

Entry	Composition
1	Honey (81%), deionized water (19%)
2	Cane sugar (50%), honey (25%), deionized water (25%)
3	Deionized water (81%), honey (15.6%), Curry plant (*Elichrysum Italicum*) freeze dried extract (0.5%), Plantain (*Plantago Lanceolata*) freeze dried extract (0.2%), Gumweed (*Grindelia Robusta*) freeze dried extract (0.2%), plantain mother tincture (0.7%).

and to enhance product quality control. By definition, a chromatographic fingerprint of natural substances as medicinal plants and corresponding formulated products refers to the profiles which can illustrate the specific properties of raw materials, semi-finished products and finished products after appropriate processing, and can be obtained by suitable analytical techniques [4–6].

In recent years, new analytical platforms based on Nuclear Magnetic Resonance (NMR) or Mass Spectrometry (MS) have been successfully applied to determine the composition of natural complex products based on medicinal plants in terms of structurally different compounds such as primary and secondary metabolites [7]. By means of the metabolomic fingerprinting is therefore possible to bridge the lack of characterization of natural substances such as medicinal plants, extracts, fractions and formulations comprising them, that so can be characterized by no more than a reduced set of components but from a very large set of metabolites (hundreds) typical for each natural product.

Two complementary approaches are currently available for the investigation of natural complex products by metabolomics: the targeted and the untargeted approach [8]. Both of them are required to achieve the complete characterization of natural products. In particular, while targeted metabolomic is aimed to the quantitative determination of specific sets of metabolites (usually 5–10 metabolites for each class under investigation), the untargeted approach is aimed to obtain a broad picture of the whole metabolome by detecting as many metabolites as possible (usually hundreds of metabolites), without necessarily identifying nor quantifying specific compounds [9].

By targeted metabolomics methods it is possible to perform quality control quantifying known metabolites. By untargeted methods quality control is evaluated taking a product fingerprint including both known and unknown metabolites and classifying it as 'good' or 'poor' quality product, based on the comparison of the obtained fingerprint with those of 'good' samples. In our study the good samples are approved batches prepared according to the Good Manufacturing Practice (GMP), using controlled raw materials, through a validated production process constantly verified by conventional approaches.

Untargeted metabolomic fingerprinting is a high throughput method to perform global analysis of complex mixtures and provides a fast and accurate method for sample classification.

Starting from the general assumption that several metabolites may contribute to the activity of natural complex products, in this study we investigated the possibility to extract useful information concerning the biological uniformity or diversity among different batches of the production supply chain, considering the whole formulation rather than few and well-selected compounds of the formulated natural complex products. In this context, the Flow Injection ElectroSpray Ionization Mass Spectrometry Analysis (FIA-ESI-MS) of the whole formulated natural complex products represents a valuable alternative to the traditional reductionist approach, providing big data generation in high-throughput fashion by avoiding any prior chromatographic steps [6] although suffering of instrumental signal fluctuation and ion suppression issues.

Based on these considerations, this study was focused to overcome the problems related to untargeted mass spectrometry-based metabolomic analysis such as reproducibility in order to obtain a validated quality control protocol for monitoring the batch-quality of *grinTuss adulti syrup* by Multivariate Statistical Process Control (MSPC). More specifically, the adopted workflow was based on the preliminary sample preparation followed by its direct infusion through flow injection into a mass spectrometer endowed with an ESI interface and an ion trap as analyzer, to generate a fingerprint useful for assessing the quality of the formulated natural complex product. About this, a specific data processing procedure was developed to allow the calculation of a robust fingerprint useful for building multivariate control charts. The key points of the proposed approach were the use of the so called Golden Standard (GS) sample to guarantee the comparison of samples analyzed in different experimental sessions and the introduction of a robust method for peak list matching and data normalization.

Actually, regulatory guidelines for the development and validation of methods based on untargeted metabolomics are not available [10], but researchers members of the Metabolome Standard Initiative (MSI) proposed the workflow reported in Table 1 as good practice for developing analytical methods based on metabolomics [11]. We decided to follow the reported workflow in the development of our method for quality control.

2. Experimental

2.1. Experimental design

Two different years of production were monitored collecting 24 samples, each one corresponding to a different batch of production. In particular, the last 6 batches produced as test set were used while the remaining 18 batches were employed to build the control charts. Moreover, for the test set 3 samples were expressly produced in laboratory by using a deliberately incorrect composition of *grinTuss adulti syrup* in order to reproduce the profile of possible batches deriving from production errors and were therefore used

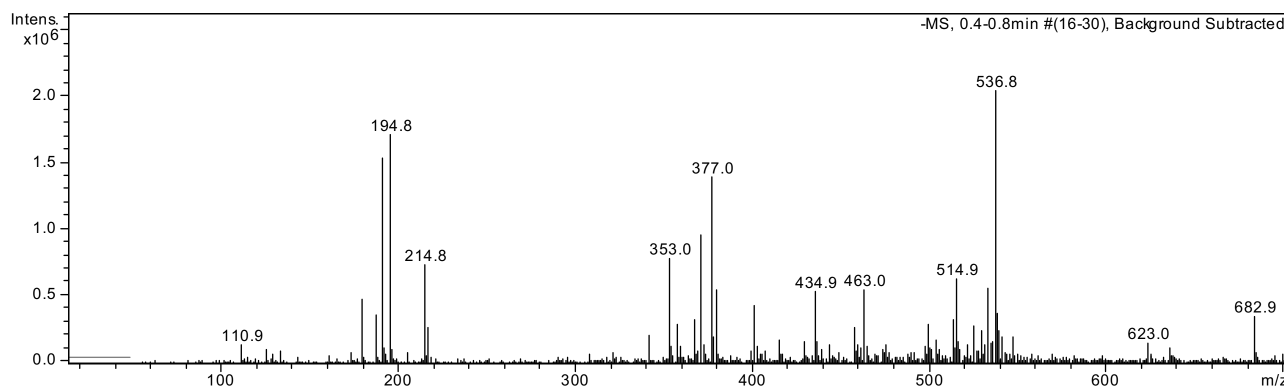

Fig. 1. *GrinTuss adulti syrup* fingerprint mass spectrum background subtracted.

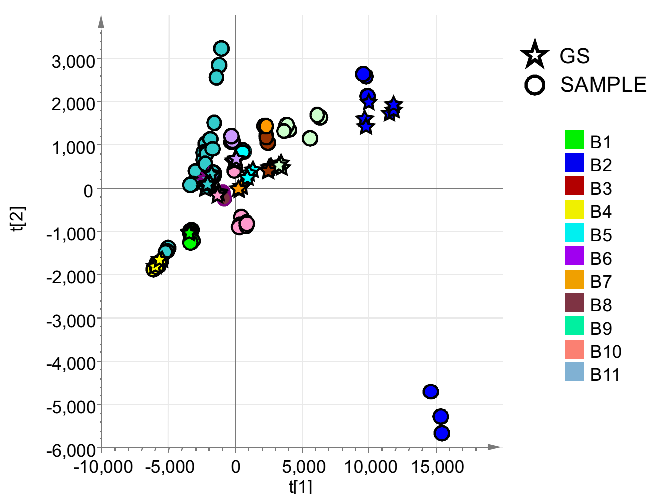

Fig. 2. PCA score scatter plot of the collected samples coloured according to the experimental session (R^2 = 0.75); stars were used to indicate the GS while circles the samples.

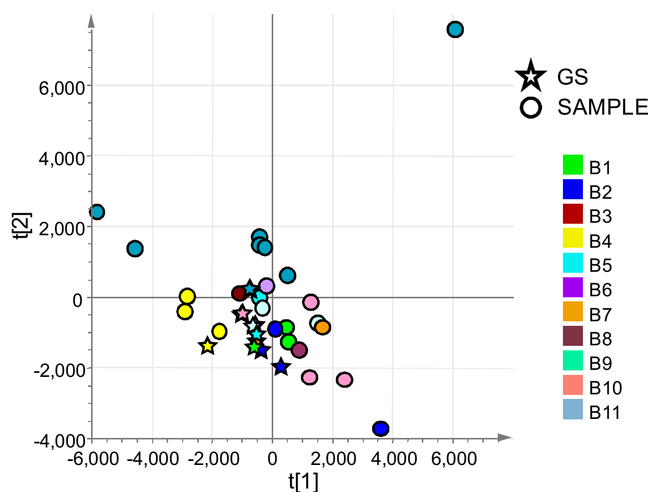

Fig. 3. PCA score scatter plot of the collected samples after data processing coloured according to the experimental session (R^2 = 0.52); stars were used to indicate the GS while circles the samples.

as model of 'bad' quality batches (Table 2). In particular, the 'bad' quality samples were obtained by preparing *grinTuss adulti syrup* sugary compositions without freeze-dried extracts (Table 2, entry 1 and 2) and a *grinTuss adulti syrup* formulation with similar relative ratios of sugary components and freeze-dried extracts (Table 2, entry 3).

The 18 samples used as training set to build the control charts represented the profile of 'good' quality batches since they were prepared according to the optimized manufacturing protocol (according to GMP) and submitted to the standardized conventional quality control operations. Furthermore, GS was selected as an approved batch of the production supply chain exerting the same quali-quantitative profile of the 'good' quality samples. A preliminary data analysis of the training set based on Principal Component Analysis (PCA) allowed us to identify the GS to use in the experimental sessions. Thus the GS was selected as the closed sample to the centre of the PCA score space in order to be the most representative sample of the training set. All the samples were stored at −80 °C prior to perform the analytical session to avoid bias due to the conservation of the GS. Our previous studies on the stability of the product proved that no relevant modifications in the metabolic content must be expected during the conservation. However, modifications were highlighted between fresh and frozen samples and, as a consequence, all the samples were frozen before the analysis.

2.2. Chemical and reagents

Absolute ethanol 99.8% and pure methanol 99.9% were purchased from Sigma-Aldrich. Ultrahigh purified water used in this study was prepared in a PurelabUltra water purification system (ELGA, UK). The batches of *grinTuss adulti syrup* were produced by Aboca SpA (Sansepolcro, Italy). *GrinTuss adulti syrup* is a natural complex product consisting mainly of fraction extracted from Elichrysum *Italicum* (top flowered head), Plantago *Lanceolata* (leaf), Grindelia *Robusta* (top flowered head) and honey. Three batches were appositely produced in laboratory according to the composition reported in Table 2.

2.3. Sample preparation

0.5 g of each sample are extracted with 100 mL of ethanol/water 50:50 (w/w) in ultrasonic bath for 20 min. The solution obtained, after filtration on 0.45 μm cellulose acetate syringe filter, was transferred in a vial and analyzed. To achieve a signal intensity of 10^5–10^6, the obtained solution was analyzed without further dilution.

The sample, treated with ethanol/water 50:50 (w/w), resulted soluble. The use of other solvents, including water, methanol or their mixtures, led to abnormal peak shape not easy to integrate in the data pre-processing step.

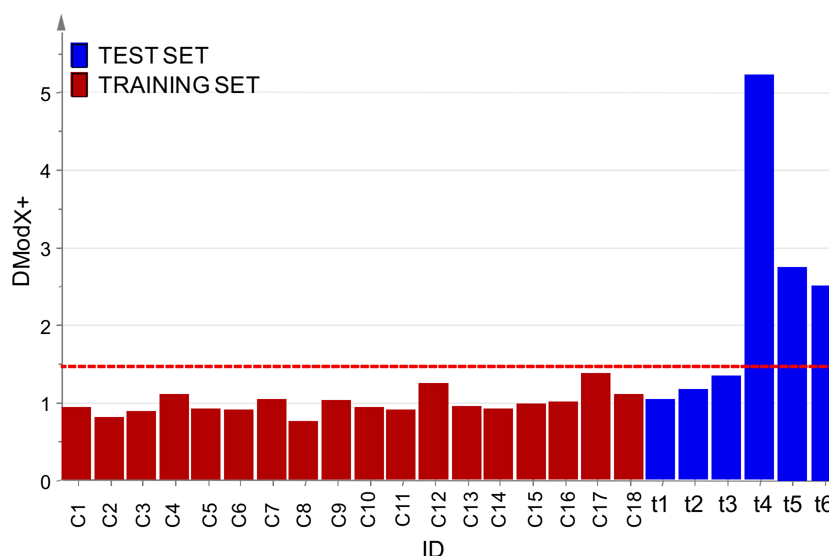

Fig. 4. DModX+ control chart (control limit equal to 95%); in red are reported the DModX+ of the samples of the training set (from C1 to C18) while in blue the DModX+ of the predicted samples (last 6 batches of production from t1 to t6).

2.4. FIA-ESI-MS conditions

FIA-ESI-MS analysis was carried out in a HPLC 1100 Series (Agilent Technologies INC., Santa Clara, CA) system equipped with a vacuum degasser, a binary pump, a Peltier thermostated autosampler at 10 °C and the effluent was analyzed by an ion trap mass spectrometer SL series equipped with an ESI interface (Agilent Technologies INC., Santa Clara, CA) operating in negative ion mode. Without any chromatographic separation (being the column replaced with an adapter connecting the capillaries between the injector and the interface) the flow rate was 0.2 mL/min and an injector volume of 5 μL. The instrumental parameters were optimized in wide mode, with a width from 50 to 1500 m/z. The ion trap mass spectrometer optimized parameters were set as follows: Entrance Capillary Voltage +4500 V; End Plate Offset −500 V; Dry Gas flow 8 L/min; Dry Temperature 350 °C; Skimmer −33.1 V; Capillary Exit −111.5 V; Oct 1 DC −12.00 V; Oct RF 300.0 Vpp; Oct 2 DC −1.70 V; Lens 1 5.5 V; Lens 2 57.2 V; Trap Drive 45.2. The ESI interface nebulizer gas pressure was set at 20 psi.

The mobile phase consisted of pure methanol (A) and ultrapure water (B) according to the following elution: 0–5 min A/B 50:50 (v/v). Before injecting a new sample, a system equilibration time of 20 min was required, using as eluting solvent A/B 50:50 (v/v). Different mobile phases were tested but the mixture methanol/water 50:50 (v/v) was finally chosen as led to the best analytical performance. In absence of the column, the introduction of the sample using the HPLC pump by flow injection create a peak that eluted between 0.05 min and 5.0 min, the time required to the signal to return to the baseline.

2.5. FIA-ESI-MS analysis

The collected samples were analyzed in 11 different analytical sessions, the test set corresponding to the last session. In all the sessions, experiments were run in triple randomizing the run order: 3 golden standards were run at the beginning, in the middle and at the end of each session, while a blank run was performed after each sample. All the sessions and runs were evaluated, also, to check intra-laboratory repeatability of the method between two different analysts.

2.6. Data processing

Several sources of variation could affect the data sets produced by FIA-ESI-MS with the result to cover the biological diversity of the collected samples. The main sources of variability were identified in the sample collection and preparation, analytical session and biological diversity of the samples. So it was applied specific procedures for reducing the effect of these source of variability in order to produce fingerprints and, then, multivariate control charts assessing only to the biological variability. More specifically, the variability due to sample collection and preparation was reduced by training the operator while variations due to the experimental platform were limited using randomized run sequences with GS and blank samples, calibration procedures for the detector and the application of suitable tools for data pre-processing and pretreatment.

2.6.1. Data pre-processing

The data acquired were peak picked and, within each experimental session, the background was subtracted to the corresponding sample using *Subtract View Spectrum as Background* routine of the Data Analysis software (trap version 3.3, Bruker Daltonik GmbH). A typical *grinTuss adulti syrup* fingerprint mass spectrum bachground subtracted in Fig. 1 is reported.

As a result, for each sample a peak list with m/z values and the corresponding peak intensity was generated.

2.6.2. Data pre-treatment: peak list matching and data normalization

Since the analysis of the collected samples can produce peak lists where the same metabolite could be present with a different value of m/z depending on the mass accuracy and resolution of the mass spectrometer, a data matching procedure was applied to match corresponding peaks in the obtained peak lists. The result of the peak list matching was a data table having the sample representation in terms of peak intensity in rows and columns corresponding to the m/z of the detected metabolites. Peaks detected in different samples were matched through the calculation of a suitable match score (Eq. (1)) that was defined as:

$$\text{score}_{ij} = 1 - \frac{|\text{m/z}_i - \text{m/z}_{\text{master}j}|}{d} \quad (1)$$

where m/z_i is the m/z value for the peak *i* to be matched, $m/z_{masterj}$ is the m/z value of the peak *j* of the master peak list and *d* the m/z tolerance. The master peak list was defined as the list containing all the detected peaks represented by the mean m/z values calculated for the samples showing that peak. Due to the m/z resolution of the mass detector used in this study, $d = 0.8$ was considered. The algorithm used for matching the peaks takes one peak from a peak list at a time and matches the peak to either the best matching peak of the master peak list or appends a new peak to the master peak list, if matching is not found for the peak. The best matching is calculated on the basis of the match score while the m/z value of the master peak list is updated by mean calculation if matching is found. The algorithm can be summarized as the following 4 step procedure:

1 the master peak list is initially set to the peak list of the first sample;
2 the peak list of the second sample is compared with the master peak list; for each peak of the peak list the match score is calculated and, on the basis of the best score, each peak is matched to a peak of the master peak list or added as new peak of the master peak list if the score is zero;
3 the m/z values of the master peak list are updated on the basis of the matched peaks;
4 iterate steps 2 and 3 including the other samples one at time.

The final result of the procedure is the master peak list and the data table containing all matched samples. In case of new samples to be matched, the step 2 of the proposed algorithm is iteratively used without the updating step of the master peak list. The peaks showing match score equal to zero are excluded from the data table and singularly investigated to highlight potential contaminants.

The master peak list could reflect the order of the samples used during the peak list matching procedure. For this reason, it was applied the following strategy for defining the order of the samples to use. A first data matching was performed by using a m/z tolerance equal to 2. The obtained data table was centred on the mean and submitted to PCA considering 3 principal components. In this way a ranking for the samples based on the increasing distance to the centre of the PCA score space was defined. The sample closest to the centre was used as starting point for peak list matching (step 1) while the obtained ranking defined the order to use. The procedure was repeated until convergence of the list corresponding to the ranking.

Also, the resulting data table may contain empty gaps, in case a matching peak was not found in some of the peak lists. To reduce the number of gaps in the resulting data table and then to limit the possible influence of missing data in the control charts, only the metabolites detected in more than 90% of the samples in the calculation of the master peak list were selected.

The master peak list was built considering only the GS included in the training set. As a result, a master list composed of 529 peaks was obtained.

For data normalization, the following expression (Eq. (2)) for the intensity $\hat{I}_{ikJ}$ in the matched peak list was considered:

$$\hat{I}_{ikJ} = \alpha_k I_{iJ} + \delta_{ikJ} \quad (2)$$

where I_{iJ} is the intensity of the metabolite *i* in the sample *J*, α_k is a factor depending on the experimental run *k* and δ_{ikJ} is the background random noise. The effects of the random noise were reduced by calculating the median of the 3 replicates of each sample within the same experimental session while the effects of α_k were taken into account applying Median Fold Change normalization (MFC normalization) on the GS injections. Since the GS was run in all the different experimental sessions and in three different times within the same experimental session, it was possible to estimate α_k and then to normalize the data following a two steps procedure: chosen a run of the GS as reference, firstly the fold change (i.e. the ratio between the intensity measured for a specific run of the GS and that of the reference) between GS and reference is calculated for each variable of the master peak list and secondly the median of the all fold changes is calculated. The median is then used to normalise all the samples nearest to the considered GS. The procedure is repeated for each run of the GS.

It should be noted that the choice of the reference does not affect the performance of the obtained control charts. Indeed, different choices produce data sets with a different factor scale without modification of the topology of the latent space.

2.7. Data analysis

After peak list matching, median of the replicates and MFC normalization each sample resulted to be described by the same set of variables. The obtained representation was submitted to multivariate data analysis. PCA is a well-known multivariate technique useful for exploratory data analysis and for data compression [12]. For quality control purposes, PCA can be successfully applied to summarize the features characterising batches of 'good' quality in order to discover batches of 'poor' quality when compared to the former by the PCA model. About this, latent variables derived by PCA, the residuals of the model or suitable combinations of scores and residuals such as DModX+ (i.e. the Distance to the Model augmented with a term measuring how far outside the acceptable model domain the projection of the observation falls) can be displayed by control charts [13–15]. This enables the specification of multivariate control limits, rather than control limits connected to univariate control charts of the single measured variables. The use of such multivariate control charts helps ensure that the product is manufactured with good quality attributes, according to all the monitored variables. In this study, the control charts based on the DModX+ parameter showed the most interesting behaviour, discovering 'bad' batches and correctly classifying "good" batches.

Data were mean centred and Pareto scaled prior to perform data analysis.

Multivariate statistical data analysis was performed by SIMCA 13 (Umetrics, Umea, Sweden) while data pre-treatment was applied by R-functions developed within the platform R 3.0.2 (R Foundation for Statistical Computing).

3. Results

The data set resulted to be composed of 105 analytical runs (24 samples × 3 replicates + 33 GS runs) and 529 m/z variables. To highlight the effects of the proposed procedure for data pre-treatment, we firstly report the investigation of the data set prior to calculate the median of the replicates and to apply data normalization. The obtained PCA model showed 2 principal components, $R^2 = 0.75$ and $Q^2_{CV7\text{-}folds} = 0.68$. The score scatter plot is reported in Fig. 2. It is possible to observe how the GS sample runs span a large space in the plot if compared with the region occupied by the other samples (the variance of the X-block related to GS is 0.85 times that of the other samples). In other words, since the GS sample runs correspond to the same sample it can be affirmed that the variability produced in the data by the different experimental sessions has the same magnitude of the variability arising from the biological diversity of the samples.

To reduce the effects of the experimental session, the pre-treatment procedure described above was applied. After the calculation of the median of the replicates and MFC normalization on the GS, the PCA model showed 2 principal components, $R^2 = 0.52$

and $Q^2_{CV7\text{-}folds}$ = 0.25. The score scatter plot is shown in Fig. 3. In this case the GS runs are more closed to each other than in the previous case and the effects due to the experimental session are reduced (the variance of the X-block related to GS is 0.15 times that of the other samples).

The control charts for MSPC where built on the basis of the PCA model of the data set composed of the 18 samples representing batches of 'good' quality. The PCA model showed 3 principal components, R^2 = 0.63 and $Q^2_{CV7\text{-}folds}$ = 0.26. The test set composed of the last 6 produced batches were projected on the PCA model in order to obtain the parameters useful to evaluate the compliance with the 'good' batches. On the basis of the DModX+ control chart with a control limit equal to 95%, 3 batches resulted to be compliant while 3 batches resulted very different. The batches not compliant were those batches expressly produced in laboratory as batches of 'bad' quality and then, the control chart resulted to correctly highlight 'bad' batches while the batches resulted to be compliant were batches of 'good' quality. Fig. 4 reports the DModX+ control chart.

4. Discussion

Analytical variability in automated Direct Infusion Mass Spectrometry (DIMS) metabolomic analysis, as FIA-ESI-MS, is a well-known issues [16]. DIMS-based untargeted metabolomics measures many hundred of metabolites in a single experiment and many researcher are still working to reduce within-experiment analytical variation, particularly true for large scale multi batches analysis. For this reason the use of a metabolomic analysis for quality evaluation of complex natural products batches through different years of production is a challenging task. Many studies are present in the field of metabolomics and some 'best practice' procedures are reported. Among these the constancy of the ion numbers between different batches can be evaluated. Also a robust quality control (CQ) sample, replicate measurement, blank samples can be introduced in the metabolomic workflow to evaluate the instrumental performance [17].

In our study we propose a new method for reducing the effect of the different analytical sessions on the data in order to obtain control charts where the biological source of variation can be easily highlighted. For this reason, we introduced the use of GS samples in each experimental session and a suitable data pre-treatment based on the information collected through the analysis of the GS samples in the different analytical sessions. Specifically, locked master peak list was used for matching the samples to reduce the effect of the decay of the performance of the ion source during the experiment, median of the replicates was applied to remove the effect of the random noise while data normalization based on Median Fold Change was performed on the GS to estimate the factor α_k specifying the response of the detector for each run. As a result, the mode of action of the proposed method can be clearly understood taken into account their relationships with the expression used for the intensity of the detected peaks. On the other hand, other methods capable to remove the effects of the experimental session on the collected data could be used. We recognized two main families of techniques: techniques based on multivariate approaches and techniques based on the analysis of each single detected metabolite. The former use PLS-DA or PCA to model the variance due to the experimental session and removes the related score-loading product from the collected data while the latter use ANOVA or its variations to remove the effects due to the different analytical session. We are investigating these alternative approaches in order to test their performance in a production environment since they were developed only for research aims. Artefacts and over-fitting could be produced by these techniques and their mode of action on the detected signals are often difficult to understand. Then, in this study we preferred to use the described approach instead of other methods to avoid over-fitting and to have a robust method to implement in quality control.

The multivariate control charts used in this study investigated the correlation structures underlying the measured variables. This allowed us to discover complex structures within the samples that cannot be highlighted by classical univariate approach. Anyway, univariate data analysis based on classical control charts was implemented in order to disclose significant variations in each single variable. Indeed, multivariate and univariate data analysis must be considered as complementary approaches. In this study, the univariate control charts confirmed the conclusion of the multivariate control charts. However, it should be noted that in the multivariate case the DModX+ control chart is able to summarize the effects included in more than 500 univariate control charts. Also, the part of the peak list of the new samples to be investigated not included in the master peak list must be analyzed to evaluate the risk of poor quality batches. Usually, the number of these peak is very small and a manual check is not time consuming. In our study the peaks being metabolites that were not included in the master peak list were not more than 15, corresponding to the 2% of the total number of recorded signals in the worst case while in more than the 80% of the samples we did not find metabolites that were not included in the master peak list.

In this study, we used a multinormal distribution to describe the distribution of the 'good' batches in the latent space of the PCA model. In case of more complex distributions, it is possible to apply the kernel density estimation method to evaluate the distribution of the training set samples in the latent space and the distribution of the residuals.

The proposed method, based on the GS and on the master peak list used for peak matching, depends on the sample selected as GS. However, since the collected samples correspond to the same product, in principle it is possible to choose a new GS to replace the old one having approximately the same number of peaks of the master peak list. The master peak list should be updated by excluding the peaks not found in the new GS, if one want to re-use all the samples collected in the past while data normalization can be recalculated on the new GS, once the new GS is run together the old GS.

In a routine quality control metabolomic analysis, the following workflow can be used as Standard Operating Procedure (SOP) for monitoring batches quality:

- ○ acquisition of the new sample in triple with related GS and blank samples randomizing the experimental runs;
- ○ matching of the replicates to the master peak list;
- ○ normalize the data;
- ○ calculation of the median of the replicates;
- ○ projection of the sample on the control charts;
- ○ evaluation of the compliance on the basis of the control charts (PCA scores, DModX+);
- ○ update of the control charts with the compliant samples; if the sample is not compliant to the good quality batches, it is necessary to proceed to checkout and only after more insights if the batch meets the quality requirements the control charts can be updated.

5. Conclusions

Nowadays, metabolomic fingerprinting can be considered as the evolution of the consolidated approaches for analyzing well-selected marker compounds for quality evaluation of complex products, based on natural substances as medicinal plants. It is quick scan, selective, specific, and it has been used to discriminate different plant samples [18] as well as multi-component finished

products [19]. By means of untargeted metabolomics a fingerprint of the product can be obtained; therefore it provides tools for quality control more accurate than traditional methods.

The metabolomic approach is widely employed in medicine and biology, due to the high amount of molecular information that can be gain by it. However the application of metabolomics based-validated methods to quality control of complex natural products is relatively unexplored. Following the workflow of MSI it was possible to reach the goal of having a multivariate statistical method, useful to assess batch compliance as well as to monitor the production chain.

The robust analytical protocol developed for evaluating the complex product composition, offers a criterion based on the application of multivariate control charts useful to evaluate globally the complex products quality compliance. By this way the maintenance of the biological properties batch to batch can be assured and, consequently, their efficacy and safety.

References

[1] J.B. Calixto, Efficacy safety, quality control, marketing and regulatory guidelines for herbal medicines (phytotherapeutic agents), Braz. J. Med. Biol. Res. 33 (2000) 179–189.

[2] P. Goldman, Herbal medicines today and the roots of modern pharmacology, Ann. Intern. Med. 135 (2001) 594–600.

[3] K.S. Booksh, B.R. Kowalski, Theory of analytical chemistry, Anal. Chem. 66 (1994) 782A–791A.

[4] S. Yongyu, D. Shujun, W. Jianye, C. Wenyu, W. Huijuan, in: Yukihiro Shoyama (Ed.), Quality Control Method for Herbal Medicine—Chemical Fingerprint Analysis, Quality Control of Herbal Medicines and Related Areas, Intech, Rijeka, 2011, pp. 171–194.

[5] A. Kamboj, in: Omboon Vallisuta (Ed.), Analytical Evaluation of Herbal Drugs, Drug Discovery Research in Pharmacognosy, Intech, Rijeka, 2012, pp. 23–60.

[6] L. Mattoli, M. Burico, A. Maidecchi, E. Bianchi, E. Ragazzi, L. Bellotto, R. Seraglia, P. Traldi, The fingerprinting of Sedivitax, a commercial botanical dietary supplement: the classical LC-MS approach vs direct metabolite mapping, Nat. Prod. An Indian J. 8 (2012) 129–151.

[7] K. Saito, R.A. Dixon, L. Willmitzer, Biotechnology in agriculture and forestry, in: Plant Metabolomics, Springer, Berlin, 2006.

[8] M. Commisso, P. Strazzer, K. Toffali, M. Stocchero, F. Guzzo, Untargeted metabolomics: an emerging approach to determine the composition of herbal products, Comput. Struct. Biotechnol. J. 4 (2013) e201301007.

[9] A. Alonso, S. Marsal, A. Julià, Analytical methods in untargeted metabolomics: state of the art in 2015, Front. Bioeng. Biotechnol. 3 (2015) 1–23.

[10] S. Naz, M. Vallejo, A. García, C. Barbas, Method validation strategies involved in non-targeted metabolomics, J. Chromatogr. A 1353 (2014) 99–105.

[11] R. Goodacre, D. Broadhurst, A.K. Smilde, B.S. Kristal, J.D. Baker, R. Beger, C. Bessant, S. Connor, G. Calmani, A. Craig, T. Ebbels, D.B. Kell, C. Manetti, J. Newton, G. Paternostro, R. Somorjai, M. Sjostrom, J. Trygg, F. Wulfert, Proposed minimum reporting standards for data analysis in metabolomics, Metabolomics 3 (2007) 231–241.

[12] J.E. Jackson, A Users Guide to Principal Components, John Wiley, New York, 1991.

[13] J.F. MacGregor, T. Kourti, Statistical process control of multivariate processes, Control Eng. Practice 3 (1995) 403–414.

[14] C. Wikström, C. Albano, L. Eriksson, H. Fridén, E. Johansson, Å. Nordahl, S. Rånnar, M. Sandberg, N. Kettaneh-Wold, S. Wold, Multivariate process and quality monitoring applied to an electrolysis process −Part II. multivariate time-Series analysis of lagged latent variables, Chemom. Intell. Lab. Syst. 42 (1998) 233–240.

[15] E. Eriksson, J. Kettaneh-Wold, C. Wikström, S. Wold, Multi- and Megavariate Data Analysis, Part II: Method Extensions and Advanced Applications, Appendix II, MKS Umetrics AB, Umeå, 2006.

[16] J.A. Kirwan, D.I. Broadhurst, R.L. Davidson, M.R. Viant, Characterizing and correcting batch variation in automated direct infusion mass spectrometry (DIMS) metabolomics workflow, Anal. Bioanal. Chem. 405 (2013) 5147–5157.

[17] J.A. Kirwan, R.J.F. Weber, D.I. Broadhurst, M.R. Viant, Direct infusion mass spectrometry (DIMS) metabolomics dataset: a benchmark for data processing and quality control, Sci. Data 1 (140012) (2012), http://dx.doi.org/10.1038/sdata.2014.12.

[18] (a) L. Mattoli, F. Cangi, A. Maidecchi, C. Ghiara, E. Ragazzi, M. Tubaro, L. Stella, F. Tisato, P. Traldi, Metabolomic fingerprinting of plant extracts, J. Mass. Spectrom 41 (2006) 1534–1545;
(b) C. Daniel, T. Kersten, S. Kehraus, G.M. König, W. Knöß, Metabolic fingerprinting for the identification and classification of medicinal plants, Zeitschrift fur Phytotherapie 29 (2008) 270–274.

[19] L. Mattoli, F. Cangi, C. Ghiara, M. Burico, A. Maidecchi, E. Bianchi, E. Ragazzi, L. Bellotto, R. Seraglia, P. Traldi, A metabolite fingerprint for the characterization of commercial botanical dietary supplements, Metabolomics, Metabolomics 7 (2011) 437–445.