Lifecycle of multivariate methods according to United States Pharmacopeia Chapter <1039> Chemometrics

Alexander Kadenkin

Chemometrics is a powerful tool widely used for method development in the pharmaceutical industry. This whitepaper describes the lifecycle of multivariate models and summarizes the workflow of the development of chemometrical models according to the new USP chapter <1039>.

Metrohm International Headquarters; Herisau, Switzerland

Metrohm NIRSystems

Introduction

Chemometrics is originally defined as a tool for gathering chemical information from physical or chemical measurements [1]. This definition is very broad and can be related to all types of chemical and physical measurements e.g., starting from a simple pH measurement. In the modern narrower sense this term is related to extracting chemical information from multivariate data. The term «multivariate» describes the amount of data used for the information extraction. In classical univariate methods, chemical information is comprised in the intensity of a single peak that is specific for the analyte. In contrast, multivariate analysis uses the whole measured response for its analysis, e.g. a spectrum or chromatogram.

Because of the use of all of the available information, multivariate methods enable new analytical possibilities for various analytical techniques. Methods of multivariate analysis can be «fed» with the complete measured spectrum and at the same time provide solutions for multiple analytical tasks, e.g., identification, classification, or collecting quantitative information. When using multivariate analysis, one analyzer can provide the same amount of information after single measurment as dozens of reference analyzers if it is calibrated for these tasks. The methods of multivariate analysis can be applied for the analysis of data measured with various analytical techniques such as Near-Infrared Spectroscopy (NIRS) [2], Raman Spectroscopy [3], chromatographic techniques such as ion chromatography [4] or other techniques, such as titration [5] or electrochemical analysis [6].

The potential of multivariate methods in the analysis of pharmaceutical products was also recognized by the Unated States Pharmacopea and resulted in the new USP chapter <1039> Chemometrics, which was published on August 1, 2017 and which is valid since the December 1, 2017 [7]. This chapter focuses on the analysis of multidimensional data from analytical instruments like spectroscopic or chromatographic data. Generally, this chapter summarizes different stages of the development of multivariate analysis methods, as shown in **Figure 1**. Additionally, it introduces the «Analytical Procedure Lifecycle» approach for multivariate methods. Each stage of method development is briefly summarized in the present white paper.

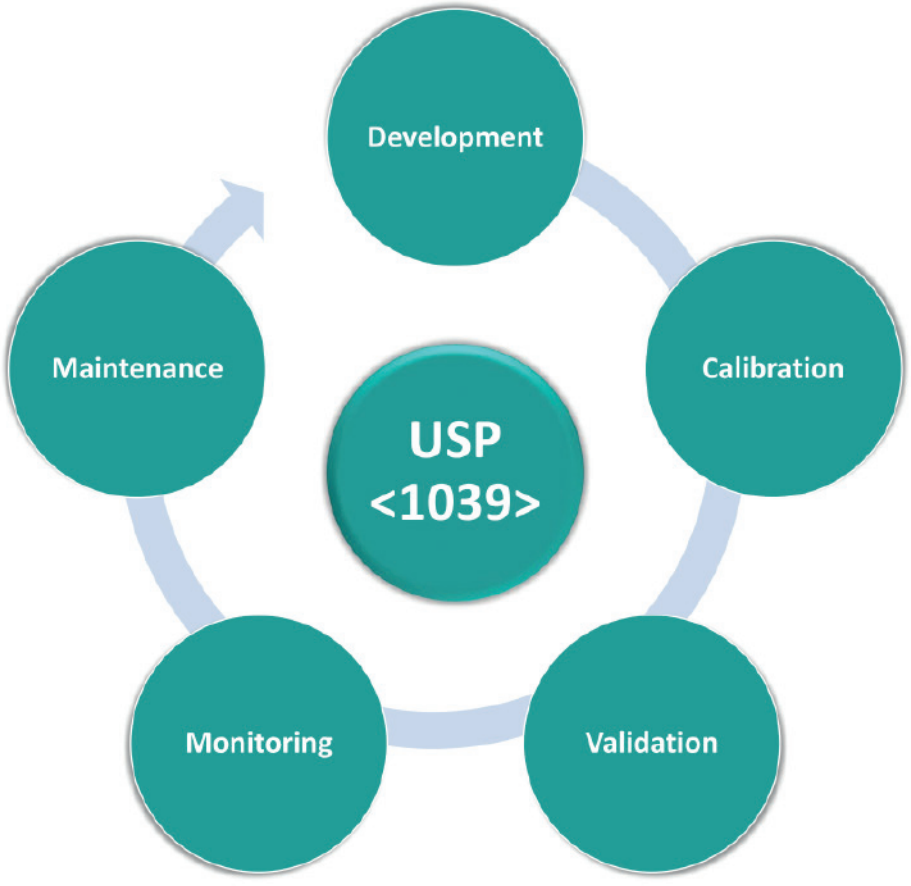


Figure 1. Costs for the method development with and without method transfer.

Development

The chemometric model as a part of an analytical procedure must fulfill a predefined, intended purpose, which should be in accordance with the previously defined analytical target profile (ATP). In this phase the requirements for the model development and the performance criteria for the key characteristics of the analytical procedure should be specified. This phase involves different steps, namely the selection of the sample set followed by exploratory data analysis, algorithm selection, and risk analysis. These aspects are briefly described.

Sample selection

The analytical performance of the application in routine analysis depends on the properties of samples used for the calibration development. The range or type of variability as well the number of samples must be selected based on scientifically sound principles. The range of the application should cover the range of the final routine application and should also include samples outside specification. Furthermore, sample variation expected in the final routine analysis should be included in the calibration set. Such variability can be particle size variation of, e.g., cellulose, different lots of raw materials, or variations under intermediate conditions such as day-to-day variations or multiple operators. The estimation of influence factors and types of variability can be realized using a risk based approach such as failure mode and effect analysis (FMEA) or a feasibility study.

The USP chapter does not provide exact information about the number of samples needed for the development of a calibration. However, it is mentioned that the number of samples increases with the increasing complexity of the application. Furthermore, a high number of samples can improve analytical figures of merit of the final application. In the ideal case, the samples should be distributed uniformly, which is not always possible. In this case, the sample selection must be explained using scientifically sound principles, such as Design of Experiments (DoE) or historical database approach.

Prior to the model development, data obtained from samples should be evaluated statistically. Approaches like exploratory data analysis can be used in order to understand the structure of the data, identify outliers, or select representative samples. Outliers can be identified, e.g., using Hotelling's T^2 statistics or residual distance.

Preprocessing

Various preprocessing techniques can be applied on the spectra in order to remove redundant information, e.g., background. Such pretreatments can dramatically improve the analytical performance of the model. Preprocessing algorithms should be selected based on the understanding of the data and the analytical techniques that were used. For example, some algorithms used in chromatography for the alignment of the x-axis are quite often unnecessary for spectroscopic data especially when using well-calibrated dispersive NIR instruments. Preprocessing techniques should be applied iteratively and their outputs should be compared with each other. This can lead to a better understanding of the data and improve the analytical figures of merit of the final model. On the other hand, the analytical chemist must be aware of the fact that overprocessing of the data can reduce the signal and increase undesirable noise, which can significantly affect the analytical figures of merit (**Example 1**).

Example 1

Estimated analytical figures of merit (in mg) for the determination of ibuprofen (0–800 mg) in tablets by NIRS, using different preprocessing algorithms. The simplest preprocessing (2^{nd} derivative) results in low and similar analytical figures of merit. The quite often used combination of 2^{nd} derivative and standard normal variate (SNV) leads to overprocessing indicated by increased analytical figures of merit. Furthermore, the error of cross-validation is more than 50% higher than the error of calibration, which clearly indicates overfitting in this case.

Pretreatment	RMSEC	RMSECV
2^{nd} derivative	5.8	7.4
2^{nd} derivative, SNV	8.0	12.0
SNV, 2^{nd} derivative	8.2	13.1

Algorithm selection

USP mentions that it is not possible to predict the performance of different algorithms for a particular application. The combination of all steps of the model development, namely sample and variable selection, preprocessing, algorithm selection as well as optimization of algorithm specific parameters can have an impact on the analytical performance of the model.

Generally, the selection of the algorithm depends on the aims of the previously defined ATP. Furthermore, this selection can be limited by the functionality of the software that is used. This should not be seen as a disadvantage. Quite often different algorithms provide comparable results and the availability of only one or two algorithms saves time during model development (**Example 2**).

Example 2

Estimated analytical figures of merit (in %) using different algorithms and same pretreatments for quality control of moisture in lactose solutions. The analytical figures of merit differ only slightly indicating similar analytical possibilities of the specific application.

Algorithm	RMSEC	RMSECV
Principal Component Regression (PCR)	0.29	0.32
Partial Least Squares Regression (PLS)	0.27	0.29
Support Vector Machines Regression (SVMR)	0.24	0.27

Variable selection

The selection of the subset of the original data can dramatically improve the analytical figures of merit of the model such as accuracy, precision, and robustness. Through the reduction of the data, the influence of irrelevant variables in the data can be minimized by the variable selection. The selection of the data should be based on experience and knowledge. For example, determination of moisture content using NIR spectroscopy results in better analytical figures of merit when only sensitive water bands at 1450 and 1950 nm are included (**Example 3**).

Example 3

Estimated analytical figures of merit (in %) for the determination of moisture (60–85%) in skin creams using NIR spectroscopy using different wavelength regions. Selection of the specific wavelength region reduces dramatically calibration and cross-validation errors. Furthermore, the estimated error of cross-validation for the full wavelength region is two-times higher than the error of calibration when using the full spectral range. This is a clear indication for overfitting.

Spectral range	RMSEC	RMSECV
Full, 400–2 500 nm	0.83%	1.73%
1 350–1 550 and 1 800–2 000 nm	0.57%	0.75%

A further reason for variable selection is improved robustness. The modelling of noise (overfitting) can occur when using the complete data set. This can have an influence on the predictive properties of the model and therefore on the robustness of the model.

Summary

The mentioned steps are usually supported by dedicated software provided together with the analyzer. Examples of such software are Vision Air Complete software for Metrohm Vis-NIR analyzers or MiraCal software for portable Metrohm Raman analyzers, which guide the operator through the different steps of the model development and support their users also in further steps.

Calibration

General information

Each chemometric algorithm used for data analysis has its specific settings. Additionally, the implementation of the algorithms and the visualization of results differs from vendor to vendor. Therefore, it is not possible to provide an overview, which is valid for all algorithms, vendors, and techniques.

However, the development of multivariate models has one step in common for all algorithms, vendors, and techniques. This step is an estimation of the model performance and the fine-tuning of algorithm specific settings based on the performance of cross-validation. The cross-validation is usually automatically performed by the used software during model development. The software removes a subset of the data from the calculation and builds the model using the remaining set. This model is used for the prediction of the removed subset. The procedure is repeated until all subsets are predicted. Finally, the software calculates the analytical figures of merit for the cross-validation based on the predictions. The best approach for cross-validation is a so called «leave-one-out cross-validation», which means that only one sample is removed at a time.

The most common analytical figure of merit estimated during cross-validation is the root-mean-squared error of cross-validation (RMSECV). A further important parameter is the root-mean-squared error of calibration (RMSEC), which is estimated using the whole data set. The RMSEC decreases continuously because the model describes additional variance of the data with increasing complexity. In contrast, RMSECV decreases only until the model describes the last meaningful information. Addition of further factor results in an increase of RMSECV: the model starts to describe irrelevant information, which leads to overfitting. This results in a strong difference between RMSEC and RMSECV values.

Model validation

It should be mentioned that the validation of the model mentioned here is not the same as the validation of the method. The aim of model validation is to demonstrate that it is suitable for its intended purpose. The validation of the model should be performed using a representative independent sample set. It must fulfill the requirements of the USP chapter <1225> *Validation of Compendial Methods* according to the method specific category **[9]**.

Typical figures of merit estimated during validation of quantitative models are specificity, accuracy, precision, linearity, range, and robustness, which are described in USP chapter <1125> **[9]**. Additional figures of merit for quantitative applications such as limit of detection (LOD), limit of quantification (LOQ), sensitivity etc. may not be required but could provide additional application related knowledge relevant for the lifecycle of the analytical procedure. Furthermore, it may be mandatory to determine these analytical figures of merit according ICH Q2 **[8]**. Qualitative models require the determination of robustness and specificity.

Accuracy

The accuracy of quantitative models can be estimated by a statistical comparison of predicted values versus reference values. Analytical figures of merit such as root mean-squared error of prediction (RMSEP), standard error of prediction (SEP), and bias can be used as indicators of accuracy. Very important is RMSEP, which should be comparable with RMSEC and RMSECV, and which should meet the requirements of ATP. For qualitative models, accuracy can be expressed as rate for positive or negative classification of validation samples.

Precision

The true precision can be estimated by the measurement of the same sample under intermediate conditions such as variation of the measurement time, operator, instruments, or laboratories involved.

Specificity

Scientific meaning of the chemometric model, e.g., selection of the wavelength range should be demonstrated and validated wherever possible. The exact determination of specificity depends on the application and algorithm, e.g., qualitative models should be capable to identify or classify the samples correctly.

Linearity

The direct estimation of linearity is not always possible since some of chemometric algorithms are nonlinear. However, it can be sufficient to demonstrate the performance of the model using the correlation coefficient, slope, intercept, and residual sum of squares of the plot between predicted versus reference values. Furthermore, the residues should not show specific patterns when plotted against concentration.

Range

The range of the calculated model is defined by the range of calibration samples and should meet the requirements of the ATP.

Robustness

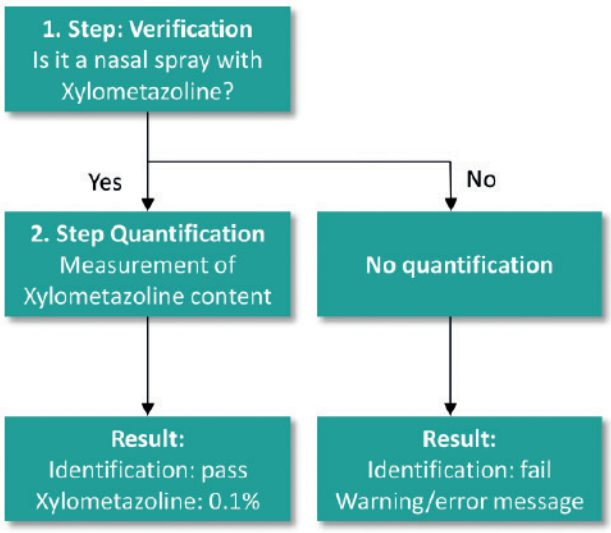
Robustness can be influenced by the calibration strategy that is used. Including expected variation (e.g., various operators, lots, days) into the calibration set has a positive impact on the robustness of the model.

Important for the model validation is the selection of the representative and independent sample set based on the requirements of the ATP. Independency means that the selected samples were not used for the calibration development or model optimization. Being representative implies that the validation samples include all types of variance expected in routine analysis (e.g., particle size variations, manufacturer, geographical origin).

Precondition for a successful validation is the definition of the acceptance criteria before performing the model validation. These criteria can be specific for each technique, e.g., for secondary techniques like NIR, the accuracy is limited by the accuracy of the reference technique, but the precision can be better due to improved sample handling. In addition, it is indispensable to set the diagnostic limits of the model and implement an outlier detection. This outlier detection can be based on leverage statistics and should identify the out-of-model-space samples. These can be for example samples out of the calibration range or completely different materials than those which were used in the calibration set (**Example 4**). Possible outliers should be included into the validation set in order to test the capability of the model to detect possible outliers in the final routine application.

Example 4

Quantitative models calculate the content of analytes (e.g., active ingredient or excipient) from the measured chromatogram or spectrum without identification of the material. Here, it is indispensable to implement identity testing prior to quantification, since the operator runs a risk of accidentally analyzing the wrong material. Dedicated software products like Vision Air for Vis-NIR spectroscopy enable the combination of both methods in a single operating procedure. The software performs verification of the material (e.g., nasal spray or not) and in case of a pass-result it automatically performs a quantitative analysis (e.g., xylometazoline as active ingredient). In case of wrong material (e.g., only water) it provides a fail result and mark the sample with a warning or error, if this is defined in the operating procedure for routine analysis.



Model monitoring

An essential part of the lifecycle of the analytical procedure based on chemometric methods is a continuous performance monitoring of the model, which should be conducted and documented as a part of the calibration and validation procedure. The control strategy should define the required elements and plan for on-going monitoring and evaluation of the model performance. Ideally, the used analytical instrument should be qualified and undergo specific performance verification strategies, defined in related general USP chapters. The used control strategy should define intervals and events that trigger the review of the model. Exemplarily, events such as changes in samples or instrumental response, out-of-specification and out-of-trend results are mentioned in the chapter.

Based on results of the performance monitoring, it is necessary to perform risk assessment, which can lead to the implementation of changes in the operating procedure and trigger a predefined procedure for the model maintenance or update.

Model maintenance

According to the proposed analytical procedure lifecycle approach, the knowledge gathered during procedure performance qualification and continued performance verification should be used for continuous improvement of the analytical procedure (**Figure 2**, **[10]**). Therefore, it may be necessary to update the model after a certain period of use, based on the observations made during routine analysis. Such updates, especially adjustments and corresponding activities must be well documented. However, it is important to understand the reason for the updates prior their implementation, because it dramatically facilitates the selection of appropriate update procedures.

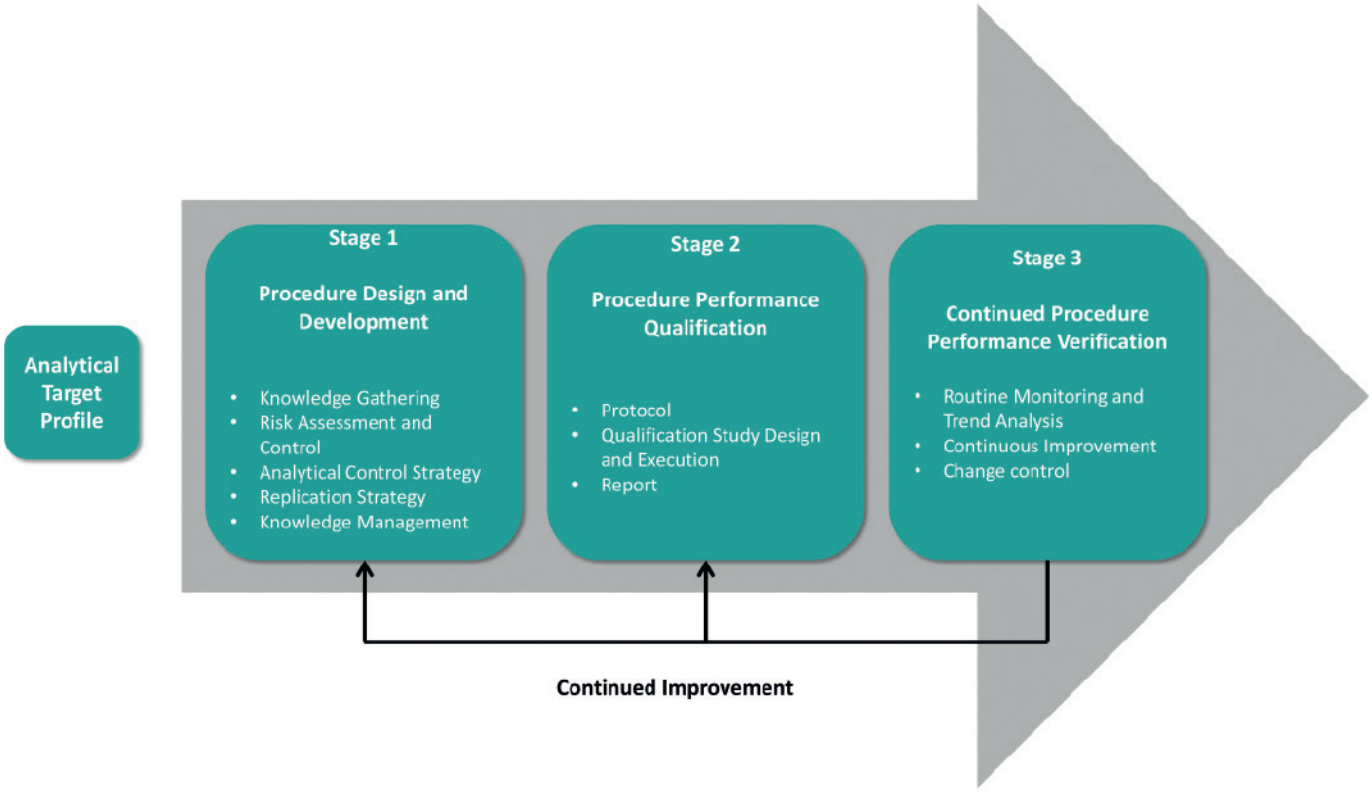


Figure 2. The analytical procedure lifecycle, adopted from **[10]**.

Changes in the analytical procedure. In this simple case, the calibration needs to be expanded, which can be caused by different reasons such as:

- Expansion of the calibration range (e.g., from 0–600 mg ibuprofen to 0–800 mg)
- New supplier for raw materials
- Previously unseen variations (particle size distribution)

Important for the expansion of the calibration range is the number of new samples that should be added as well as the understanding of the impact of this addition on the calibration and validation. In general, the operator can add all new samples or select a subset of new samples based on different applicable scientific methods.

Changes on the instrumental part

The change of the instrumental response can occur as a result of replacement of spare parts (e.g., lamps, columns), replacement of the defect components (e.g., pumps), or simple model transfer between different instruments. However, when using a well calibrated and standardized instrument, changes of spare parts or method transfer to a similar instruments do not require any model update (**Example 5**).

Example 5

Metrohm NIR Application Note NIR-011 demonstrates the impact of the spare part replacement and method transfer **[11]**. Here, the NIR model for the determination of caffeine content was transferred to 3 different units and validated using independent samples set. Additionally, in one unit, 4 different spare parts (lamps) were used. This work showed that RMSEP after the model transfer or replacement of the spare part is similar to the RMSEP of the original model due to excellent quality of the dispersive Vis-NIR instruments and dedicated procedures for instrument standardization.

On the other hand, changes in the analytical procedure (e.g., different sample preparation between the calibration and the validation step) may lead to the necessity of a model update. Here, simple adjustment methods such as slope and bias correction for quantitative models should be considered first.

USP <1039> mentions the possibility of calibration transfer, which can be applied when changing the instrument (e.g., different vendors) or measurement conditions (e.g., transfer from the laboratory analyzer to a process analyzer). In this case, it is necessary to apply instrument standardization and calibration transfer methods using stable transfer samples.

Finally, each step of model maintenance must be well documented. However, prior to the use of the updated model it must be revalidated using the acceptance criteria of the original validation protocol. The process of revalidation as well as the results muss be both documented.

Summary

The present white paper describes the lifecycle of analytical procedures based on multivariate models and summarizes the workflow of the development of chemometric models according to the new USP chapter <1039>. The simplified workflow for application and intervention is summarized in **Figure 3**. Additional information is available in the USP monograph itself and in technique specific USP chapters. The development of multivariate methods according to USP <1039> can be supported by a local Metrohm representative.

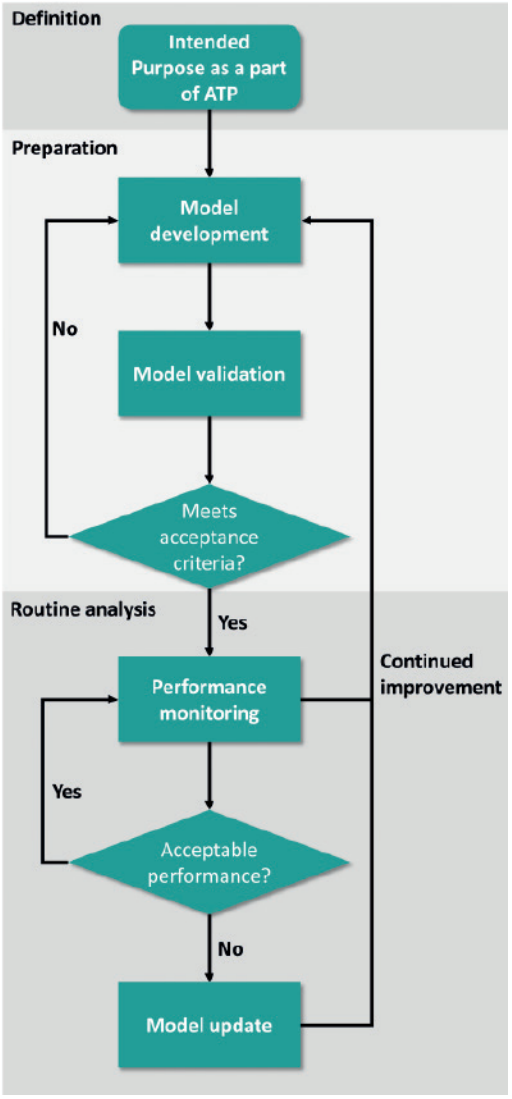


Figure 3. Simplified workflow for the model development.

References

[1] R.G. Brereton, Applied Chemometrics for Scientists, John Wiley and Sons, 2007.

[2] D.A. Burns, E.W. Ciurczak, Handbook of Near-Infrared Analysis, CRC Press, 3rd Edition, 2007.

[3] I.R. Lewis, H. Edwards, Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line, Marcel Dekker Inc., 2001.

[4] J. Zirojevic, Z. Jovic, A. Djurdjevic, A. Ciric, P. Djurdjevic, Chemometric-Assisted Determination of some Bisphosphonates and Their Related Substances in Pharmaceutical Forms by Ion Chromatography with Inverse UV Detection, Acta Chromatographica, 2015 (27) 1-23.

[5] M. Akhond, J. Tashkhourian, B. Hemmateenejad, Simultaneous determination of ascrobic, citric, and tartaric acids by potentiometric titration with PLS calibration, Journal of Analytical Chemistry, 2006 (61), 804-808.

[6] W. Henao-Escobar W, O. Domínguez-Renedo, M.A. Alonso-Lomillo, M.J. Arcos-Martínez, Resolution of quaternary mixtures of cadaverine, histamine, putrescine and tyramine by the square wave voltammetry and partial least squares method, Talanta, 2015 (143), 97-100.

[7] Chemometrics, <1039>, USP 40 (2017).

[8] ICH Q2(R1) Validation of Analytical Procedures: Text and Methodology, 1994.

[9] Validation of Compendial Methods, <1225>, USP 40 (2017).

[10] Stimuli article, Proposed New USP General Chapter: The Analytical Procedure Lifecycle <1220>, USP, 2016, **www.uspnf.com**

[11] Calibration model transfer of caffeine on the NIRS XDS Rapid Content Analyzer, Metrohm NIR Application Note NIR-011, **www.metrohm.com**

WP-031EN, published June, 2018