# User Manual

## Methods Validation App

Version: 1.0

Date: 2026-01-12

Giovanni Solarino

giovanni.solarino@unito.it

# Contents

# 1 Introduction

This User Manual describes the functionalities, workflow, and usage of the Web-App version of the Methods Validation App. The aim is to provide clear, and practical instructions for end-users.

The application is freely available at https://mva.databloom.eu/

## 1.1 Purpose of the Application

The Methods Validation App (MVA) is designed to simplify and standardize the evaluation of analytical methods by automating the statistical procedures required during method validation. Its purpose is to provide analysts with a reliable, user-friendly environment in which calibration models, accuracy, precision, detection limits, and other key validation parameters can be computed consistently and reproducibly.

MVA supports validation strategies aligned with the main international standards, including the ICH Q2(R2) guidelines. Its workflow and statistical framework make it suitable for applications that must comply with regulatory expectations in pharmaceutical, clinical, forensic, environmental, and other analytical contexts.

By integrating data handling, statistical testing, and graphical inspection tools in a single interface, MVA reduces manual calculation errors, improves reproducibility, and helps laboratories maintain high analytical quality standards.

## 1.2 Target Audience

This manual is intended for analysts, laboratory technicians, quality assurance personnel, and researchers who perform or supervise the validation of analytical methods. The Methods Validation App (MVA) is suitable for users with varying levels of statistical expertise, from beginners who need guided workflows to experienced analysts who require a fast and reliable tool to support validation

## 1.3 How To Cite

If you use this software, cite it as: [CITE TO ADD]

# 2 Architecture

MVA is structured as a browser-based application that separates the user interface from the statistical computation layer. The graphical interface, built with NiceGUI, manages the interaction with the user and orchestrates the exchange of data with the

underlying Python engine. Users load datasets, select validation parameters, and inspect outputs through the frontend, while all analytical operations—including regression modelling, variance assessment, residual analysis, and evaluation of precision and accuracy are executed exclusively in the backend.
The backend consists of a collection of Python modules that implement the statistical framework described in the validation protocol. The file **stat_test.py** contains the core computational logic. It receives the normalized dataset prepared by the interface and performs the entire regression-selection routine. The workflow begins with the grouping of calibration levels and the evaluation of homoscedasticity using both the Levene test and an F-test. The outcome of these tests determines whether the backend will construct models using ordinary least squares or weighted least squares. When heteroscedasticity is detected, the backend computes three candidate weighting schemes—no weighting, $1/x$, and $1/x^2$—and selects the optimal one by comparing the variance of normalized residuals across calibration levels.

Once the variance structure has been established, the backend fits both linear and quadratic regression models. The implementation uses the Ordinary Least Squares (OLS) and Weighted Least Squares (WLS) interfaces from Statsmodels, generating paired models for the averaged calibration points as well as for the raw data. Model selection relies on a sequence of diagnostic tests: Mandel's test verifies whether the quadratic term provides a statistically significant improvement, and a secondary check compares paired residuals using a t-test and an F-test. Only when both tests confirm the absence of meaningful differences does the backend override Mandel's result in favour of the simpler model. Residuals from the selected model undergo the Shapiro–Wilk test, and a kernel density estimate is optionally computed for visualization purposes.

The backend also implements the Hubaux and Vos approach for the estimation of limits of detection and quantification. The algorithm uses the fitted linear model, the relevant subset of calibration levels, and the chosen weighting scheme to calculate the prediction interval at zero concentration and derive the statistical decision limit. The implementation supports internal-standard-normalized data and dynamically adapts weighting computations accordingly.

Precision and accuracy calculations follow the same computational route as the calibration workflow. The backend reorganizes the dataset by days and replicates, generates every admissible combination of calibration curves, constructs the corresponding model, and predicts the unknown sequence. Precision is derived from the coefficient of variation across predictions, and accuracy from the mean back-calculated concen-

trations relative to the nominal values. Both intra-day and inter-day routines reuse the same modelling functions and statistical tests, ensuring full methodological consistency throughout the validation process.

Data exchange between the interface and the backend occurs in the form of structured Python objects, which the frontend translates into tables, or plots. The frontend never performs statistical operations; it merely triggers backend routines and displays their outputs.

# 3 User Interface Guide

## 3.1 About page

The About page (Figure 1) serves as the application's homepage and is the first interface displayed upon launching MVA. It provides a concise yet informative overview of the software's purpose, guiding principles, and core functionalities. The page introduces users to the rationale behind the implemented validation workflow, outlining how MVA structures and automates key analytical steps such as calibration, precision, accuracy, and sensitivity assessments.

In addition to the introductory text, the About page includes a visual timeline summarizing the sequence of validation tasks supported by the application. This helps users quickly understand the overall workflow and navigate to the corresponding modules. A dedicated bibliography section is also provided, featuring key references that form the scientific and methodological foundation of MVA, allowing users to explore the theoretical background in more detail.

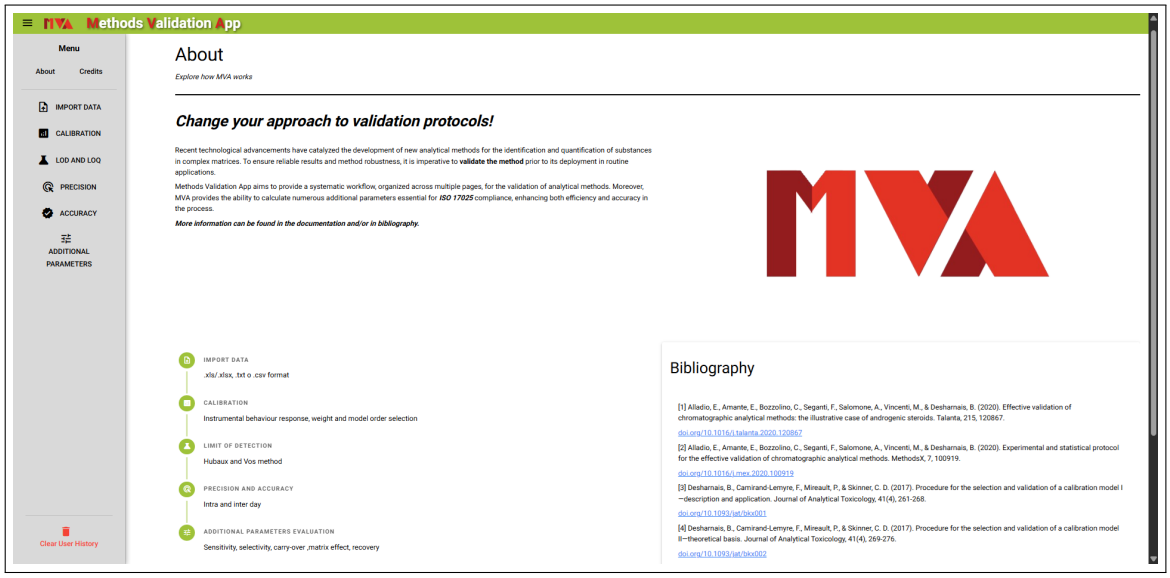


Figure 1: About page of the Methods Validation App.

### 3.2 Navigation Menu

The closable left drawer menu (Figure 2) provides access to all sections of the application and allows users to navigate seamlessly between the different MVA modules. Each calculation page (Calibration, LOD and LOQ, Precision, Accuracy, and Additional Parameters) operates independently, meaning that analyses performed in one module do not influence or modify the results of another.

Data uploaded through the *Import Data* page serve as the foundational dataset for every calculation module with the exception of *Additional Parameters*. This latter section requires its own dedicated input files, as the computations performed there involve parameters and experimental layouts that differ from those used in the main validation workflow.

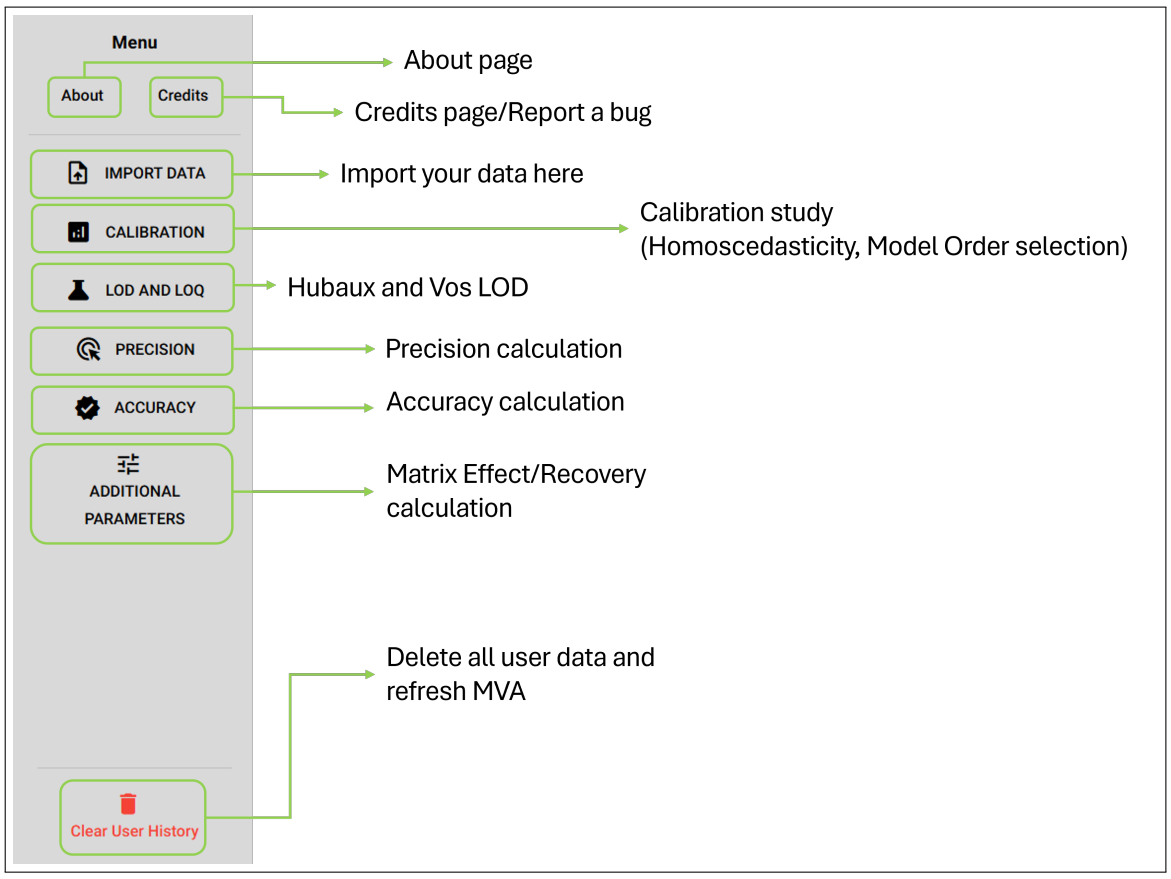


Figure 2: Navigation Menu sections.

## 4 Workflow and Functionalities

### 4.1 Data Import

The *Import Data* page (Figure 3) is the starting point of the analytical workflow and provides all tools required to upload, inspect, and prepare datasets before running

any calculation module. The page supports files in `.csv`, `.txt`, `.xls`, and `.xlsx` formats, which are validated on upload to ensure compatibility with the processing routines.

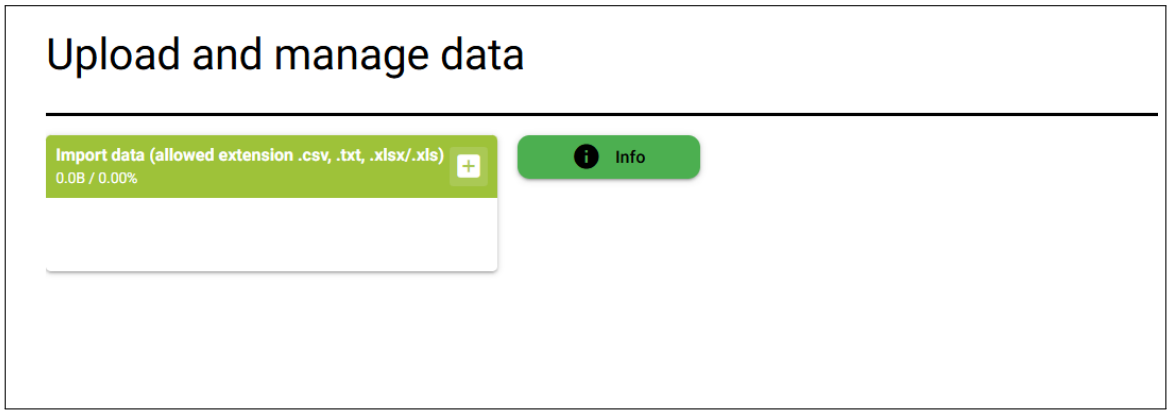


Figure 3: Import data page.

Once a file is uploaded, the system automatically parses the content and displays the available column names. Please note that column names must be in the first row of the imported dataset. Users are then prompted to select three essential fields:

- the concentration column,
- the analyte signal column,
- the internal standard (ISTD) signal column.

These selections (Figure 4) determine how the dataset is normalized and structured before being stored. The app performs an internal preprocessing step that includes signal normalization (when an internal standard concentration is provided), consistency checks, and error handling to prevent misalignment between selected fields. If duplicate selections or incompatible columns are detected, a notification alerts the user and prevents further processing.

After column selection, the user is required to provide additional metadata, including the analyte name, measurement unit, and the concentration of the internal standard. These values are stored internally and reused across all relevant modules.

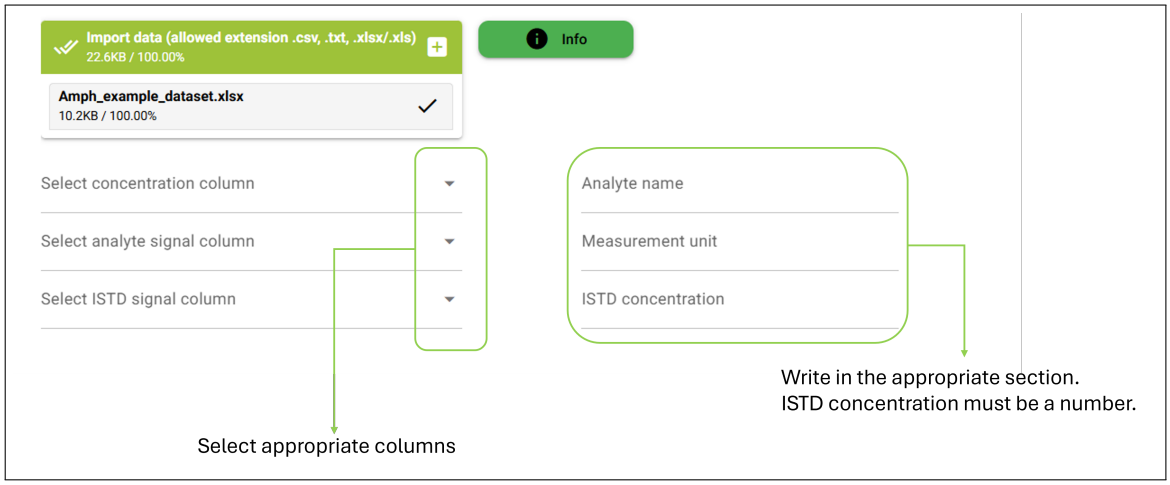


Figure 4: Metadata selection.

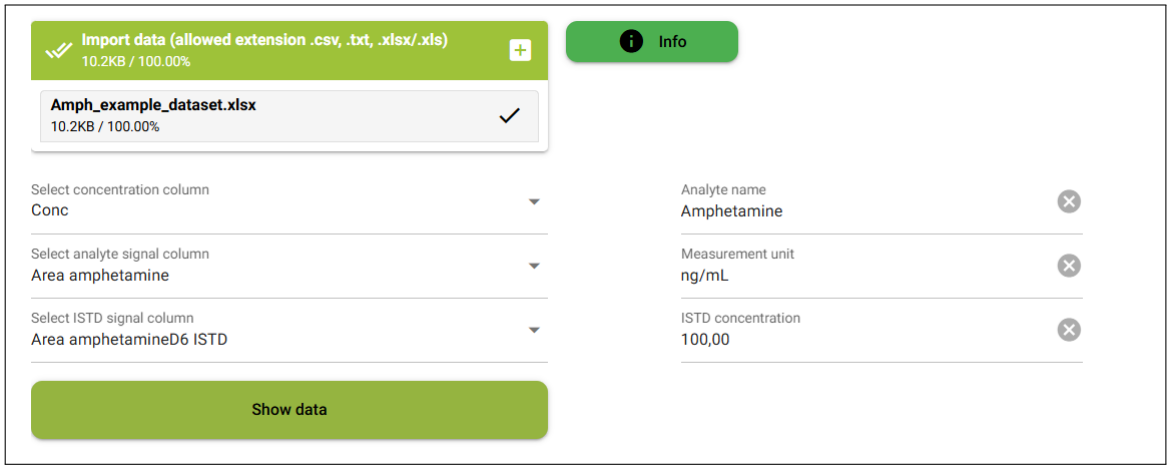


Figure 5: Metadata used in example dataset.

When all fields are correctly defined, the *Show data* button becomes available (Figure 5). This triggers the creation of a normalized dataframe preview (Figure 6), which is displayed in table format, along with the *Clear Data* switch button and the *GO TO CALIBRATION* button. The processed dataset is then saved in memory and made accessible throughout the application. A message in the status panel indicates whether a dataset is currently loaded and offers options to change the analyte mapping or clear memory entirely.

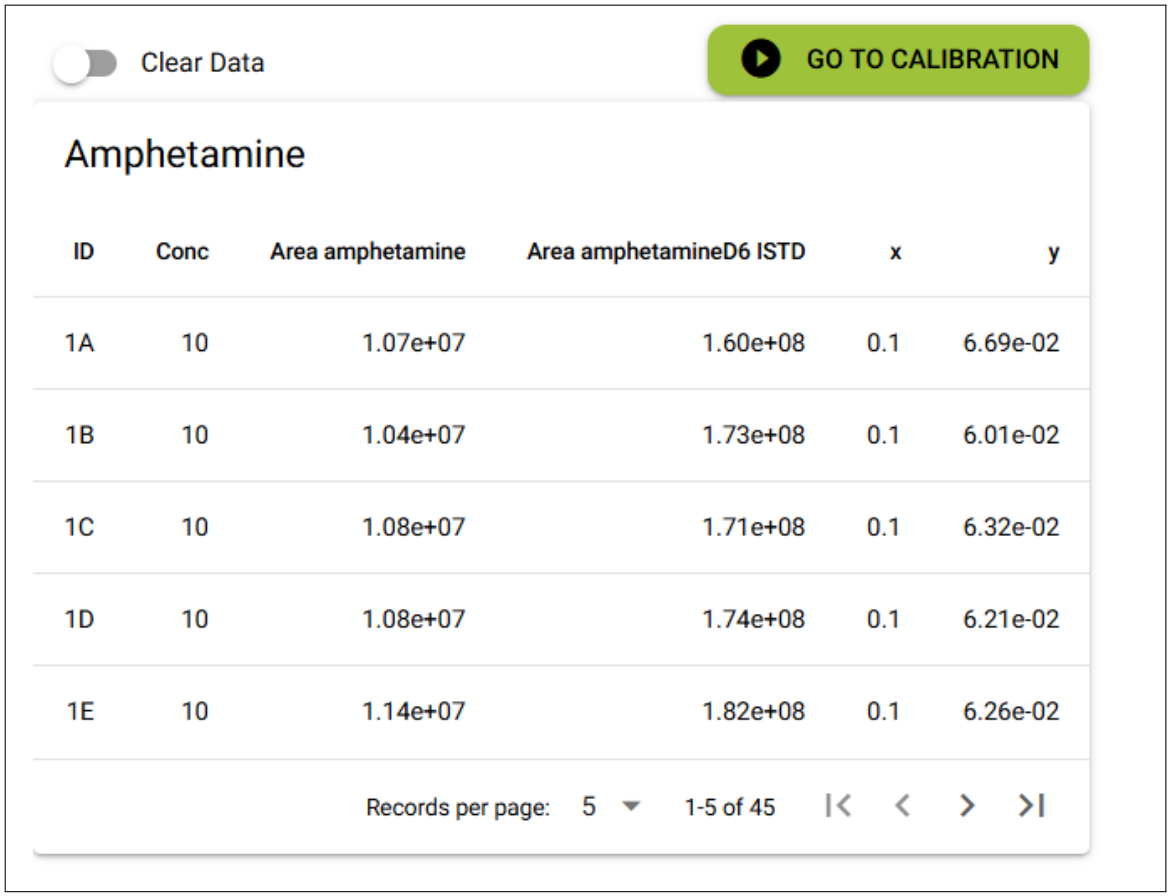


Figure 6: Normalized data frame preview.

A template dataset is also available through the *Info* button. This template illustrates the expected data structure, including how calibrators should be organized across multiple days and curves. The template can be downloaded directly for reference or for preparing new datasets.

## 4.2 Running an Analysis

After clicking the *GO TO CALIBRATION* button, the user is redirected to the *Calibration* page, where the calibration study is performed. The imported data are first displayed both in tabular format and as a scatter plot (Signal Intensity vs. Concentration) showing all data points and their means (Figure 7). Most figures are interactive and can be downloaded.

Scrolling down, the user will find tab panels dedicated to heteroscedasticity testing. The first two panels (Levene test and F-test ULOQ–LLOQ) present the corresponding statistical results along with their interpretations. Additionally, for the F-test, a graphical representation of data dispersion at the Upper Limit of Quantification (ULOQ) and the Lower Limit of Quantification (LLOQ) is provided (Figure 8).

The last tab panel displays the results of the weighting factor selection, which is performed only if heteroscedasticity is detected according to the Levene test.

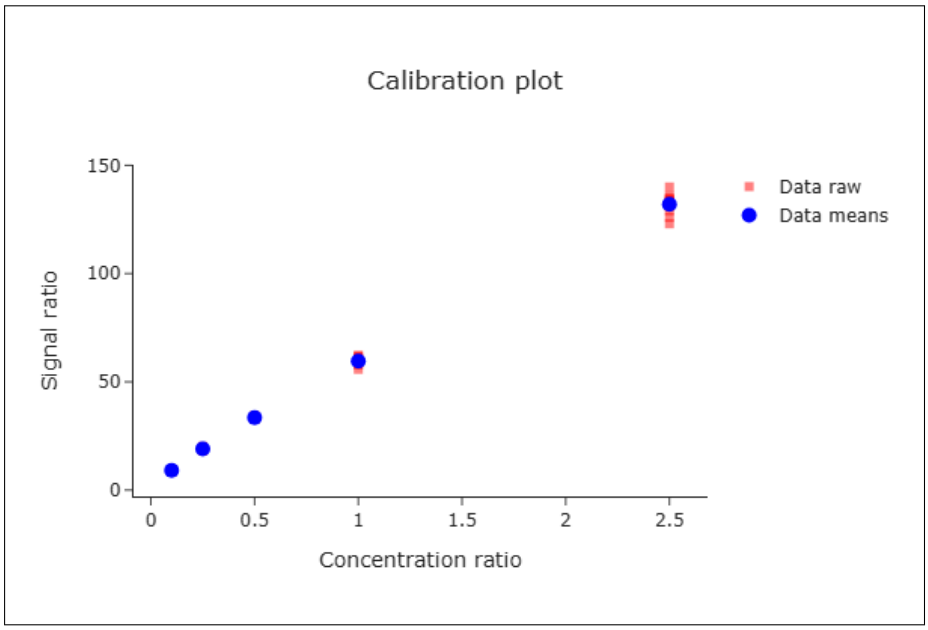


Figure 7: Scatter plot of imported data.

Finally, the most suitable calibration model is presented in the final section of the *Calibration* page. This section includes key statistical parameters and the calibration plot with the fitted model equation (Figure 9). By clicking the *Best Calibration Model* dropdown button, users can explore detailed information about the available calibration models (Linear and Quadratic). This includes graphical analyses of model residuals (QQ plot, Residuals vs. Concentration, KDE plot) and the results of the Shapiro–Wilk test for assessing residual normality (Figure 10).

The evaluation of the limit of detection (LOD) and the limit of quantification (LOQ) is performed on a separate page. The algorithm used requires the specification of several parameters, such as the number of calibrators and the level of statistical sig-

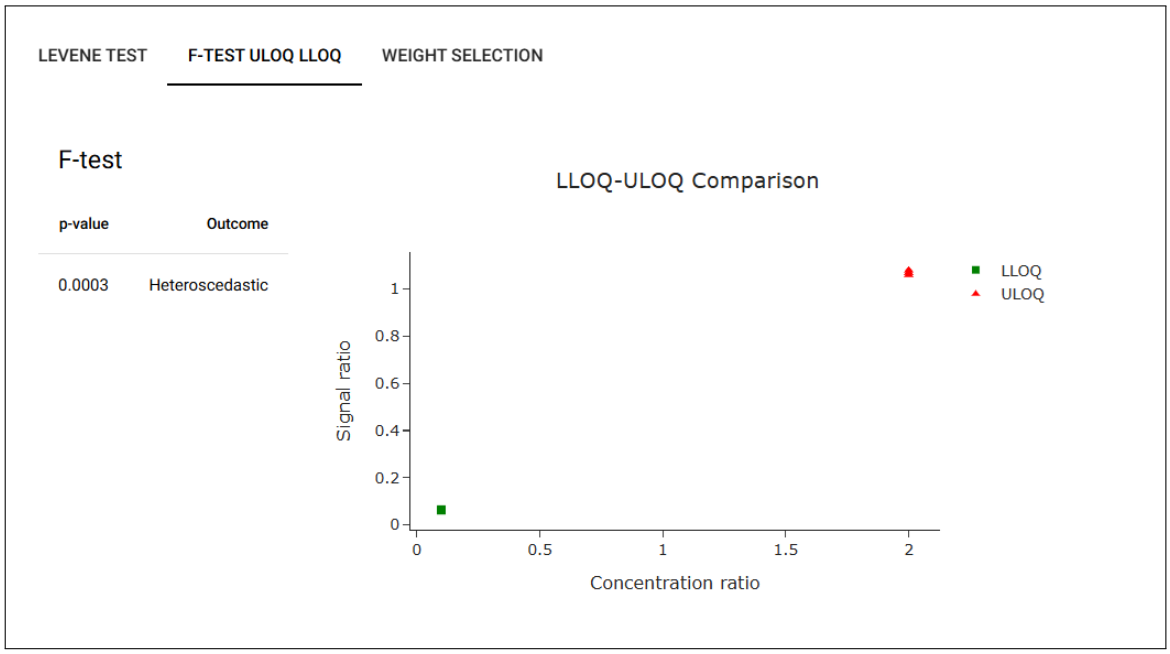


Figure 8: F-test tab panel.

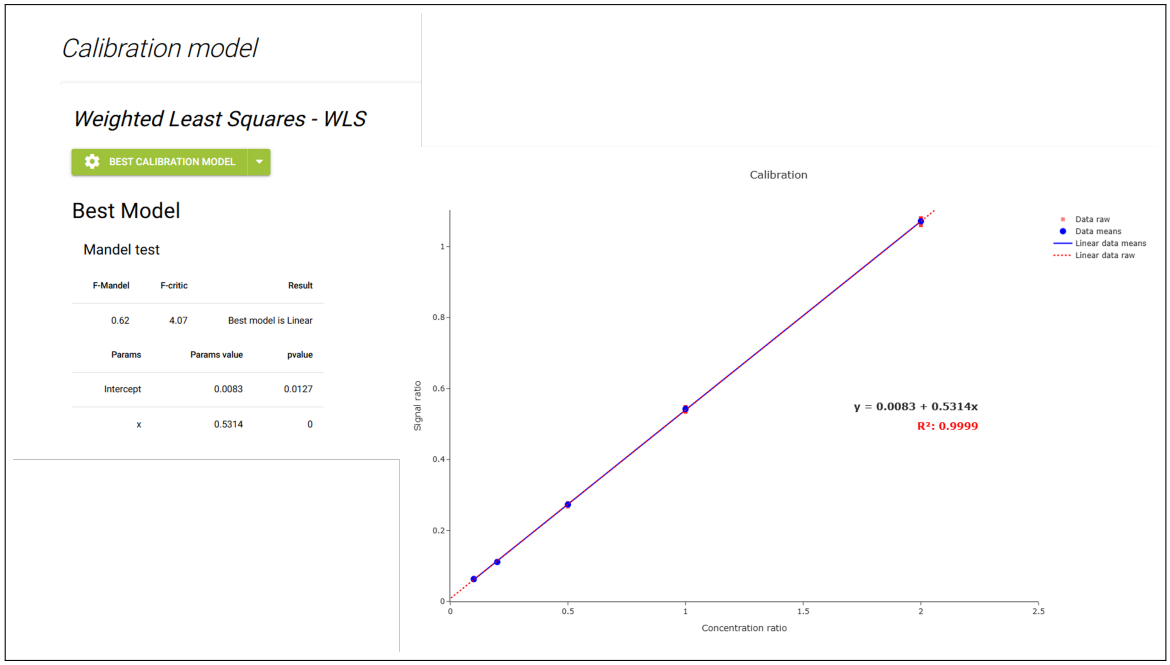


Figure 9: Best Calibration Model.

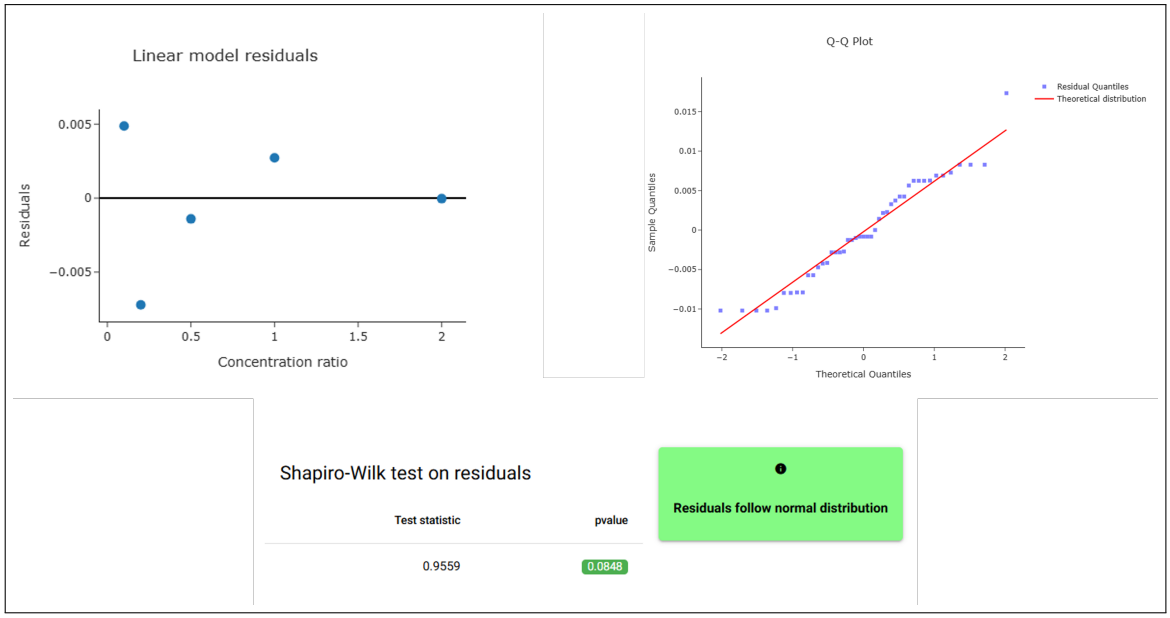


Figure 10: Linear Model residuals.

nificance. Although these parameters are set by default to generally optimal values, they can be adjusted to modify the resulting LOD (Figure 11). For instance, reducing the number of calibrators leads to a lower LOD, while increasing the statistical significance level (from 0.05 to 0.10) can also decrease the LOD. In addition, we recommend using the first calibrator as the LOQ, provided that its precision and accuracy meet the acceptable criteria.

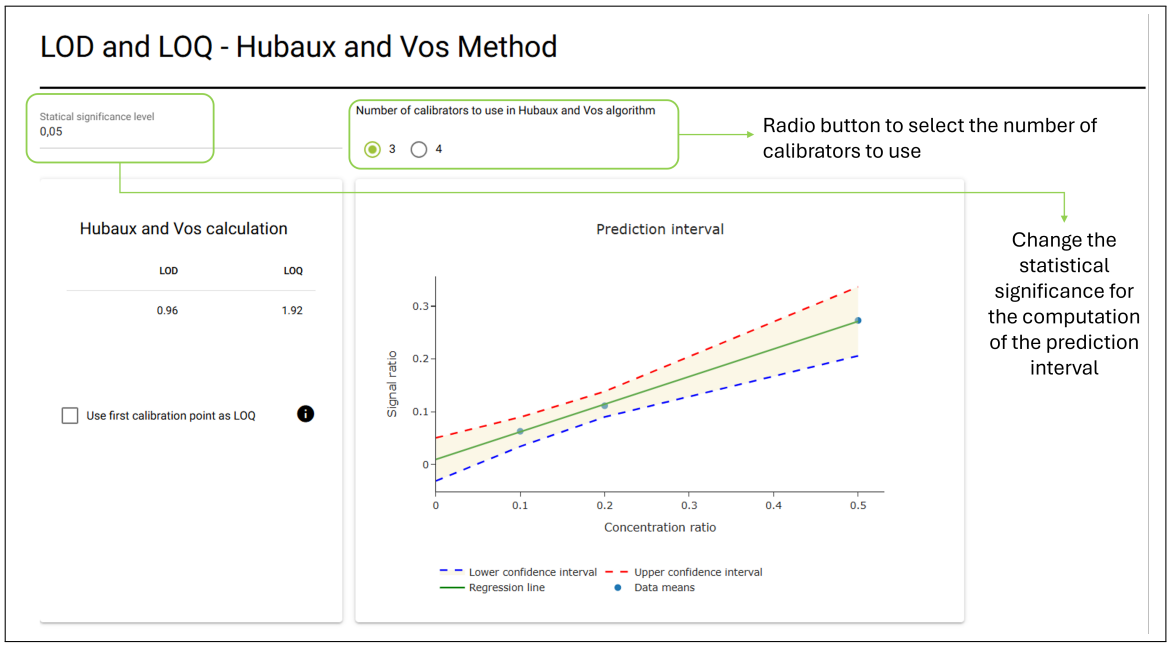


Figure 11: LOD and LOQ page.

Precision and accuracy calculations are available on dedicated pages. To begin the evaluation, the user must enter the number of calibration curves prepared per

day (the same number must be used for all days) and the number of validation days. For full compliance with the statistical framework implemented in the software, the recommended setup is nine calibration curves collected over three days (three replicates per day). A faster alternative is to use six calibration curves over two days (three replicates per day). If all calibration curves are generated on a single day, the software calculates precision and accuracy using intra-day data only. Results are automatically displayed in a table format with a color-coded scheme that highlights values meeting the user-selected acceptability criteria (Figure 12). By default, the target value is set to 20% and the acceptable limit to 25%.

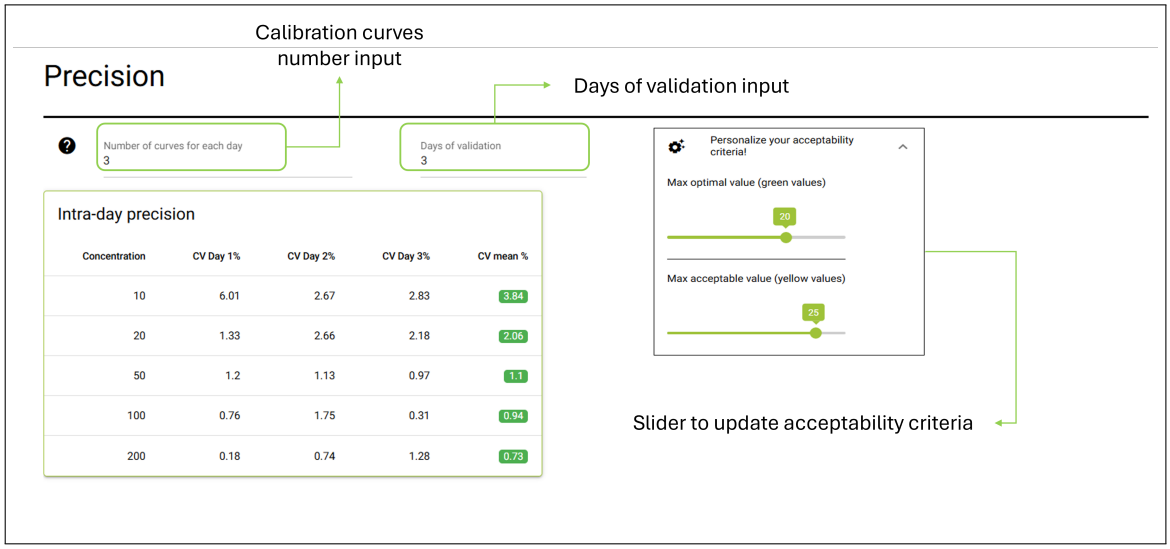


Figure 12: Precision page, the same interface is showed in Accuracy page.

# 5 Guidelines for Designing a Validation Study in LC-MS

This section provides practical recommendations for planning a robust method validation study. The principles outlined here complement the automated capabilities of MVA and help analysts design validation experiments that meet current expectations in LC–MS quantification, including matrix considerations, calibration strategy selection, variance structure assessment, and accuracy and precision evaluation.

## 5.1 General Recommendations for Calibration Curve Preparation

Reliable validation begins with well-constructed calibration curves. Calibration standards should be prepared in a matrix that closely reproduces the chemical composition of the study samples, as matrix effects remain one of the largest sources of variability. Matrix composition influences extraction efficiency, ionization behaviour, and overall instrument response; therefore, whenever possible, the authentic matrix

should be used to ensure consistent behavior between calibration standards and real samples .

When dealing with bio-fluids, if the analyte is endogenous and an analyte-free matrix cannot be obtained, pooled matrices or background subtraction may be employed, though the analyst must be aware of their limitations, particularly the influence of endogenous concentrations on achievable LLOQs and potential inter-individual variability in matrix composition.

An internal standard should be incorporated in all calibration samples. Stable-isotope labeled internal standards provide the most effective correction for matrix effects and extraction-related variability, owing to their nearly identical physicochemical properties to the analyte. Structural analogues can be used when isotopically labeled standards are unavailable, but they require additional verification of response behavior across the calibration range, as differences in ionization efficiency or chromatographic retention may impair their ability to compensate for matrix-related bias.

A typical calibration curve contains at least five or six levels that span the expected working range of the method. Both blank and zero samples are recommended: the blank allows assessment of instrumental noise and potential interferences, while the zero sample (matrix + internal standard only) helps verify that the internal standard behaves consistently across the run. Calibration levels may be spaced equidistantly or non-equidistantly. Non-uniform spacing may be advantageous when addressing heteroscedastic behavior, which is common in LC–MS methods as variance tends to increase with analyte concentration. However, it is noteworthy that using uneven-spaced calibrators at low concentrations can inflate Type II errors when Hubaux and Vos method is used for LOD calculation (see Section 5.4).

The number of replicate injections should be selected in line with the expected variability of the matrix and the required precision. Although regulatory guidelines differ, replicates must be sufficient to evaluate variance structure reliably and support statistical tests used for model selection. MVA does not impose a fixed design, but it relies on adequate replication to assess homoscedasticity, weighting requirements, and residual behavior.

Finally, analysts should prepare calibration curves across several analytical days whenever feasible. Multi-day calibration enables the detection of day-to-day variations, long-term drift, and matrix instability—all of which contribute to a more realistic assessment of method robustness.

## 5.2 Selection of Calibration Strategy

The choice of calibration strategy - matrix-matched external calibration, surrogate matrices, surrogate analytes, or in-sample calibration — depends on the availability

of authentic matrices, the nature of the analyte, and the analytical context. No universal strategy is suitable for all situations, and the analyst must balance practicality, metrological requirements, and regulatory expectations.

When an analyte-free authentic matrix exists, matrix-matched external calibration offers the best alignment between calibration standards and study samples. It provides the closest match in extraction recovery, matrix effects, and ionization conditions, making it the preferred strategy for exogenous analytes. However, for endogenous analytes, background concentrations may limit achievable limits of quantification and introduce variability. In such cases, background subtraction or construction of pooled matrices may be considered, though these approaches also introduce their own limitations.

Surrogate matrices constitute a suitable alternative when authentic matrices are scarce, difficult to obtain, or naturally contaminated with the target analyte. Neat solutions, artificial matrices, or charcoal-stripped matrices may be used to construct calibration curves, provided that their behaviour is demonstrated to be equivalent to that of the study matrix. This equivalence is typically established by comparing calibration slopes or performing parallelism studies, as recommended in regulatory guidance.

Surrogate analytes may be employed when authentic standards are not available. Stable-isotope labelled analogues represent the best option due to their close chemical similarity to the target analyte. Structural analogues may be used when isotopically labelled standards are unavailable, but the analyst must demonstrate that response factors remain consistent across the concentration range and that no cross-talk, isotopic overlap, or matrix-dependent shifts impair their use.

## 5.3 Recommendations for Precision and Accuracy Experiments

Precision and accuracy studies form the core of method validation. These experiments evaluate the reliability of the back-calculated concentrations derived from the chosen calibration model.

Precision studies should examine both repeatability (intra-day) and intermediate precision (inter-day). Analysts should prepare multiple analytical series on different days, ideally using fresh calibrations each day or, when appropriate, combining multiple calibration sequences. The variability associated with the analytical workflow — sample preparation, extraction, instrumental conditions — should be representative of routine practice.

Accuracy is evaluated by comparing back-calculated concentrations with their nominal values. The accuracy assessment requires appropriate selection of concentration levels, typically covering the lower, mid-range, and upper portions of the calibration range.

When planning precision and accuracy experiments, analysts should also consider the nature of heteroscedasticity. Variance often increases with concentration in LC–MS systems, and this can influence the interpretation of precision at different levels of the calibration curve. Weighted regression strategies may be needed to stabilize variance across the range. MVA automatically evaluates homoscedasticity and selects the appropriate regression model based on statistical tests, but the experimental design must support reliable estimation of this variance.

## 5.4 Recommendations for LOD and LOQ Determination

The accurate determination of limits of detection and quantification depends heavily on a calibration design that provides stable estimation of residual variance, particularly at the lower concentration range. The Hubaux and Vos approach, implemented within MVA, requires adequately spaced low-level calibration points and reliable assessment of variance across these points. In some cases the calculated LOD can be higher than the first calibration point; this situation corresponds to Type II errors (i.e., false negatives). These errors increase when the uncertainty at the lower end of the calibration curve is overestimated. Sparse or poorly chosen low-concentration calibrators, inadequate modelling of heteroscedasticity, noisy blanks, and suboptimal internal standard performance all contribute to inflated prediction intervals and consequently to an elevated Type II error rate, yielding an artificially high LOD. Calibration levels near the lower end of the range should be selected carefully, avoiding concentrations so close to the noise threshold that reproducibility becomes unfeasible.