

ImpLiMet – user manual

ImpLiMet provides an easy to utilize site for imputation of any dataset providing user with selection of previously proven imputation methods as well as possible information about the optimal method for their dataset for distinct missingness causes. Specific steps are provided below.

The screenshot shows the ImpLiMet v1.0 web application interface. The left sidebar contains navigation links: Getting start, Download sample data, Analyze, Troubleshoot, Authors and citing, and Return to CompLiMet. The main content area is titled "Imputation for Lipidomics and Metabolomics" and has two tabs: "Imputation" (active) and "Visualization".

Step 1

This box must be selected prior to data upload if input includes information about multiple feature measurement groups (see download sample data for information about the required input format).

Upload a file for imputation (*.csv).

Browse... Input_data_without_group_info.csv

Upload complete

Total number of samples: 45
Total number of features: 40
Total number of measurement groups: 1
Total number of missing values in the dataset: 21

Step 2

Remove samples with the selected % of missing values

Don't remove any samples

Remove features with the selected % of missing values

Don't remove any features

The sample(s) left : 45
and the feature(s) left: 40

Cleaned Data

Step 3

Select imputation method (Note: a minimum of 6 samples or 3 features without missing values is required for the full optimization option)

Select imputation method:

Optimization

full parameter search

Run IMPLIMET

Select If the dataset includes multiple feature measurement groups

Input file with features in columns and samples in rows (see Download sample data for details)

Samples with more than selected % of missing values will be removed from imputation

Select type of imputation. Selecting – “Optimization” will provide estimate of the best imputation method for the dataset. Optimization can be performed with either full search or fast, partial search.

Download data and check out the effect of imputation with histograms, kurtosis and skewness analysis as well as PCA representation of results in the Visualization tab above



ImpLiMet v1.0

- Getting started
- Download sample data
- Analyze
- Troubleshoot
- Authors and citing
- Return to CompLiMet

Imputation

Visualization

User Manual

Step 1

This box must be selected prior to data upload if input includes information about multiple feature measurement groups (see download sample data for information about the required input format).

Upload a file for imputation (*.csv).

Browse... Input_data_without_group_info.csv
Upload complete

Total number of sample(s): 45
 Total number of feature(s): 40
 Total number of measurement groups: 1
 Total number of missing values in the dataset: 21

Step 2

Remove samples with the specified % of missing values

Don't remove any samples

Remove features with the specified % of missing values

Don't remove any features

The sample(s) left : 45
 and the feature(s) left: 40

Cleaned Data

Step 3

Select imputation method (Note: a minimum of 6 samples or 3 features without missing values is required for the full optimization option). Final results are graphically represented in the Visualization tab above.

Select imputation method:

Optimization

full parameter search (slow for large dataset)

Run IMPLIMET

Imputed Data

If user selected full or partial "Optimization" ImpLiMet presents mean absolute percentage error (MAPE) for each imputation method for simulated missigness of three types. Selected for imputation is the method with the lowest overall MAPE value across all three missingness types and all methods (indicated in orange). ImpLiMet also shows the lowest average MAPE value which user can chose by simply rerunning ImpLiMet with this imputation type selected.

Orange label indicates the minimal MAPE value across all tests and the method used for the imputation following this optimization. Blue label indicates the minimal average of MAPE values for the three missingness types. If this imputation method is preferred please select it and run ImpLiMet.

	missing_type	mean	median	maximum	minimum	one_fifth_minimum	KNN	RF	MICE
1	MCAR	0.584	0.584	4.708	0.567	0.907	0.585 (k:10)	0.486 (trees:500)	0.696 (iteration index:2)
2	MNAR	1.183	1.183	5.674	0.655	0.901	1.125 (k:10)	1.061 (trees:500)	1.492 (iteration index:2)
3	MAR	0.864	0.864	5.511	0.655	0.916	0.784 (k:10)	0.669 (trees:500)	0.912 (iteration index:2)
4	Average	0.877	0.877	5.298	0.626	0.908	0.831	0.739	1.033

Selecting the “**Visualization**” tab will open visual representation of the dataset properties for the set with removed rows and columns with missing values (in blue) as well as dataset imputed with selected method (in orange). Histogram as well as skewness and kurtosis show whether there is any skewness in the data distribution possibly indicating that missingness in some of the features comes from MNAR and sensitivity problem in the data quantification. User can download these images as .svg files.

