Package: RamanMP (via r-universe)

November 6, 2024

Type Package

Title Analysis and Identification of Raman Spectra of Microplastics

Version 1.0

Author Veronica Nava [aut, cre], Maria Luce Frezzotti [ctb], Barbara Leoni [ctb]

Maintainer Veronica Nava <veronicanava245@gmail.com>

Description Pre-processing and polymer identification of Raman spectra of plastics. Pre-processing includes normalisation functions, peak identification based on local maxima, smoothing process and removal of spectral region of no interest. Polymer identification can be performed using Pearson correlation coefficient or Euclidean distance (Renner et al. (2019), <doi:10.1016/j.trac.2018.12.004>), and the comparison can be done with a user-defined database or with the database already implemented in the package, which currently includes 356 spectra, with several spectra of plastic colorants.

License GPL (>= 2)

Encoding UTF-8

LazyData true

Imports ggplot2, dplyr, ggrepel, imputeTS

Depends R (>= 3.5.0)

Config/pak/sysreqs libicu-dev libjpeg-dev libpng-dev libxml2-dev libssl-dev

Repository https://veronicanava.r-universe.dev

RemoteUrl https://github.com/veronicanava/ramanmp

RemoteRef HEAD

RemoteSha 312260153de86bfd82089c8d4aec5426b0c97512

Contents

matrix_unknown	*Matrix with 4 unknown Raman spectra of plastic polymers*

Description

Database with frequency data as a first column ("freq"), and intensity values of 4 different unknown plastic polymers (purely by way of example).

Usage

```
data("matrix_unknown")
```

Examples

```
data("matrix_unknown")
str(matrix_unknown)
summary(matrix_unknown)
```

MPdatabase	*Database with Raman spectra of plastic polymers and pigments*

Description

Database with frequency data as a first column ("freq"), and intensity values of different plastic polymers and plastic additives.

Usage

```
data("MPdatabase")
```

Examples

```
data("MPdatabase")
str(MPdatabase)
summary(MPdatabase)
```

---

norm.min.max   *Min-max normalisation*

Description

The function performs a min-max normalisation on one or multiple spectra. Normalisation is performed subtracting at each peak intensity the minimum intensity value of the spectra and then dividing for the difference between the maximum and the minimum peak values of the spectra.

Usage

```
norm.min.max(spectra)
```

Arguments

spectra   A dataframe/matrix with frequency values as first column and at least one column with intensity values.

Value

Return the normalised spectra: the first column represent the frequency data, the second the intensity values normalised

Author(s)

Veronica Nava

Examples

```
data("MPdatabase")
norm.database<-norm.min.max(MPdatabase)
norm.spectra<-norm.min.max(MPdatabase[,c(1,2)])
```

---

norm.SNV   *Z-score normalisation*

Description

The function performs a Standard normal variate (SNV) transformation of a spectra. Normalisation is performed subtracting at each peak intensity the mean intensity value of the spectra and then dividing for the standard deviation of the spectra intensities.

Usage

```
norm.SNV(spectra)
```

Arguments

spectra   A dataframe/matrix with frequency values as first column and at least one column with intensity values.

Value

Return the normalised spectra: the first column represent the frequency data, the second the intensity values normalised by Z-score

Author(s)

Veronica Nava

Examples

```
data("MPdatabase")
norm.database<-norm.SNV(MPdatabase)
norm.spectra<-norm.SNV(MPdatabase[,c(1,2)])
```

---

peak.finder   *Peaks identification*

Description

The function identifies peaks based on local maxima. The function returns a list of the peaks and a plot with the peaks labeled. Missing values (NA) are removed.

Usage

```
peak.finder(spectrum, threshold=0, m=5, max.peak=0)
```

Arguments

spectrum	A dataframe/matrix with only two columns: the first column must report the frequency values; the second column must report the intensity values.
threshold	Numeric. It indicates the value on y-axis that the peak intensity must exceed to be considered a peak. This can be helpful in case of noisy Raman spectrum. The default value is 0.
m	Numeric. It indicates the interval on x-axis for the determination of the interval for the calculation of the peak. Default value is 5.
max.peak	Numberic. It indicates the number of peaks that should be displayed. The default is 0, which indicates that all peaks are showed.

Value

Return the normalised spectra: the first column represent the frequency data, the second the intensity values normalised by Z-score

Examples

```
data("MPdatabase")
peak.data<-peak.finder(MPdatabase[,c(1,7)], threshold = 500, m=7)
```

---

region.remove	*Removal of spectral region*

Description

The function removes a spectral region of no interest for further analysis. The user must specify range values for the region that has to be removed.

Usage

```
region.remove(spectra, min.region, max.region)
```

Arguments

spectra	A dataframe/matrix with frequency values as first column and at least one column with intensity values.
min.region	Numeric. Minimum frequency value of the region that should be removed.
max.region	Numeric. Maximum frequency value of the region that should be removed.

Value

Return the spectra with the removed region. The rows corresponding to the range specified are removed.

Examples

```
data("MPdatabase")
new.spectrum<-region.remove(MPdatabase[,c(1,6)], min.region=500, max.region=1200)
new.spectra<-region.remove(MPdatabase, min.region=500, max.region=1200)
```

---

savit.gol   *Savitzky–Golay smoothing*

Description

The function applies a Savitkzy-Golay smoothing filter on the spectra file based on settings defined by the user.

Usage

```
savit.gol(x, filt, filt_order = 4, der_order = 0)
```

Arguments

x	A vector with the intensity values that should be smoothed.
filt	Numeric.The length of the filter length, must be odd.
filt_order	Numeric. Filter order: 2 = quadratic filter, 4 = quartic. Default is 4.
der_order	Numeric. Derivative order: 0 = smoothing, 1 = first derivative, etc. Default is 0.

Value

Return the spectra with the removed region. The rows corresponding to the range specified are removed.

Examples

```
data("MPdatabase")
smooth.vect<-savit.gol(MPdatabase[,6], filt=11)
```

| single_unknown | *Matrix with 1 unknown Raman spectra of plastic polymer* |

Description

Database with frequency data as a first column ("freq"), and intensity values of 1 unknown plastic polymers (purely by way of example).

Usage

```
data("single_unknown")
```

Examples

```
data("single_unknown")
str(single_unknown)
summary(single_unknown)
```

| spectra.alignment | *Align spectra with different spectral resolution* |

Description

The function merges spectra with different spectral resolution using as a reference the spectra with highest resolution. The matching is done based on a span value defined by the user.

Usage

```
spectra.alignment(db1, db2, t)
```

Arguments

db1	Dataframe/matrix with frequency values as first column and at least one column with intensity values.
db2	Dataframe/matrix with frequency values as first column and at least one column with intensity values.
t	Numeric. It indicates the tolerance for the matching of the two spectra. For a given t-value, the intensity values that range in the frequency interval (f-t, f+t) are matched with the corresponding intensity values of the database with the highest spectral resolution.

Value

Return a matrix with frequency of the database with highest spectral resolution and intensity values of the two databases matched based on the 't' parameter.

---

spectra.corr   *Spectrum identification based on Pearson correlation coefficient*

Description

The function allows identification of Raman spectra of single unknown plastic polymer comparing the spectrum with a user-defined database or using the database included into the package using the Pearson correlation coefficient. The database is provided within the data of the package with the name 'MPdatabase' and includes different plastic polymers, pigments and additives.

Usage

```
spectra.corr(db1, db2, t, normal='no', plot=T)
```

Arguments

db1   Dataframe/matrix with frequency values as first column and at least one column
               with intensity values. This should be the database with the known spectra of
               plastics. This can be a user-defined database or the database implemented in the
               package ('MPdatabase').

db2   Dataframe/matrix with frequency values as first column and one column with
               intensity values of the unknown spectrum that should be identified.

t   Numeric. It indicates the tolerance for the matching of the two spectra. For a
               given t-value, the intensity values that range in the frequency interval (f-t, f+t)
               are matched with the corresponding intensity values of the database with the
               highest spectral resolution.

normal   This arguments indicates if the data of the database and the unknown spectra
               should be normalized and with which methods. Accepts the following inputs:
               'percentage' divides each peak for the peak of maximum intensity and then cal-
               culate the percentage; 'SNV' performs a Standard Normal Variate transforma-
               tion; 'min.max' applies a min-max normalisation; 'no' no normalisation proce-
               dure is applied. Default is 'no'.

plot   Logical. If TRUE, a plot of the unknown spectra and the spectrum of the
               database, for which the highest correlation value was found, are showed. This
               allows verification of the results obtained

Value

Return a matrix with Hit Quality Indexes (HQI) calculated using Pearson correlation coefficient of the unknown spectra vs spectra of the database, as reported in eq. 7 of Renner et al. (2019).The matrix reports only the top 10 polymers for which the correlation values are the highest, ordered from the largest to the smallest. If the database contains less than 10 spectra, all the correlation coefficients are reported.

References

Renner, G., Schmidt, T. C., Schram, J. (2019).Analytical methodologies for monitoring micro(nano)plastics: Which are fit for purpose?. Current Opinion in Environmental Science & Health, 1, 55-61, https://doi.org/10.1016/j.coesh.201

Examples

```
data("MPdatabase","single_unknown")
identif_spectra<-spectra.corr(MPdatabase, single_unknown, t=0.5, normal='min.max')
```

---

spectra.corr.mat	*Identification of multiple spectra identification based on Pearson correlation coefficient*

Description

The function allows identification of Raman spectra of multiple plastic polymers through the comparison with a user-defined database or using the database included into the package by means of Pearson correlation coefficient. The database is provided within the data of the package with the name 'MPdatabase' and includes different plastic polymers, pigments and additives.

Usage

```
spectra.corr.mat(db1, db2, t, normal='no')
```

Arguments

db1	Dataframe/matrix with frequency values as first column and at least one column with intensity values. This should be the database with the known spectra of plastics. This can be a user-defined database or the database implemented in the package ('MPdatabase').
db2	Dataframe/matrix with frequency values as first column and columns with intensity values of the unknown spectra that should be identified.
t	Numeric. It indicates the tolerance for the matching of the two spectra. For a given t-value, the intensity values that range in the frequency interval (f-t, f+t) are matched with the corresponding intensity values of the database with the highest spectral resolution.
normal	This arguments indicates if the data of the database and the unknown spectra should be normalized and with which methods. Accepts the following inputs: 'percentage' divides each peak for the peak of maximum intensity and then calculate the percentage; 'SNV' performs a Standard Normal Variate transformation; 'min.max' applies a min-max normalisation; 'no' no normalisation procedure is applied. Default is 'no'.

Value

Return a list of two elements. The first is "Score", which reports all the Hit Quality Index (HQI) calculated using the Pearson correlation coefficients as reported in eq. 6 of Renner et al. (2019). The second element of the list is "Maximum score" which reports for each unknown spectra (reported in col names) the name of the polymer for which the maximum value of the HQI was identified.

References

Renner, G., Schmidt, T. C., Schram, J. (2019).Analytical methodologies for monitoring micro(nano)plastics: Which are fit for purpose?. Current Opinion in Environmental Science & Health, 1, 55-61, https://doi.org/10.1016/j.coesh.201

Examples

```
data("MPdatabase","matrix_unknown")
identif_spectra<-spectra.corr.mat(MPdatabase, matrix_unknown, t=0.5, normal="min.max")
score<-identif_spectra[1]
maximum_match<-identif_spectra[2]
```

---

spectra.dist   *Spectrum identification based on Euclidean distance*

Description

The function allows identification of Raman spectra of single unknown plastic polymer comparing the spectrum with a user-defined database or using the database included into the package using the Euclidean distance. The database is provided within the data of the package with the name 'MPdatabase' and includes different plastic polymers, pigments and additives.

Usage

```
spectra.dist(db1, db2, t, plot=T)
```

Arguments

db1	Dataframe/matrix with frequency values as first column and at least one column with intensity values. This should be the database with the known spectra of plastics. This can be a user-defined database or the database implemented in the package ('MPdatabase').
db2	Dataframe/matrix with frequency values as first column and one column with intensity values of the unknown spectrum that should be identified.
t	Numeric. It indicates the tolerance for the matching of the two spectra. For a given t-value, the intensity values that range in the frequency interval (f-t, f+t) are matched with the corresponding intensity values of the database with the highest spectral resolution.
plot	Logical. If TRUE, a plot of the unknown spectra and the spectrum of the database, for which the highest correlation value was found, are showed. This allows verification of the results obtained

Value

Return a matrix with Hit Quality Indexes (HQI) calculated using the Euclidean distance for the unknown spectra from the database spectra following the equation 6 reported in Renner et al. (2019).The matrix reports only the top 10 polymers for which the HQI are the highest, ordered from the largest to the smallest. If the database contains less than 10 spectra, all the HQI are reported.

References

Renner, G., Schmidt, T. C., Schram, J. (2019).Analytical methodologies for monitoring micro(nano)plastics: Which are fit for purpose?. Current Opinion in Environmental Science & Health, 1, 55-61, https://doi.org/10.1016/j.coesh.201

Examples

```
data("MPdatabase","single_unknown")
identif_spectra<-spectra.dist(MPdatabase, single_unknown, t=0.5)
```

---

spectra.dist.mat	*Identification of multiple spectra identification based on Euclidean distance*

Description

The function allows identification of Raman spectra of multiple plastic polymers through the comparison with a user-defined database or using the database included into the package by means of Euclidean distance. The database is provided within the data of the package with the name 'MPdatabase' and includes different plastic polymers, pigments and additives.

Usage

```
spectra.dist.mat(db1, db2, t)
```

Arguments

db1	Dataframe/matrix with frequency values as first column and at least one column with intensity values. This should be the database with the known spectra of plastics. This can be a user-defined database or the database implemented in the package ('MPdatabase').
db2	Dataframe/matrix with frequency values as first column and columns with intensity values of the unknown spectra that should be identified.
t	Numeric. It indicates the tolerance for the matching of the two spectra. For a given t-value, the intensity values that range in the frequency interval (f-t, f+t) are matched with the corresponding intensity values of the database with the highest spectral resolution.

Value

Return a list of two elements. The first is "Score", which reports all the Hit Quality Indexes (HQI) calculated using the Euclidean distance for the unknown spectra from the database spectra following the equation 6 reported in Renner et al. (2019). The second element of the list is "Maximum score" which reports for each unkown spectra (reported in col names) the name of the polymer for which the maximum HQI (based on Euclidean distance) was identified.

References

Renner, G., Schmidt, T. C., Schram, J. (2019).Analytical methodologies for monitoring micro(nano)plastics: Which are fit for purpose?. Current Opinion in Environmental Science & Health, 1, 55-61, https://doi.org/10.1016/j.coesh.201

Examples

```
data("MPdatabase","matrix_unknown")
identif_spectra<-spectra.dist.mat(MPdatabase, matrix_unknown, t=0.5)
score<-identif_spectra[1]
maximum_match<-identif_spectra[2]
```

Index