

# Cardinal: Analytic tools for mass spectrometry imaging

*Kylie A. Bemis and April Harry*

May 8, 2019

## Contents

|       |  |    |
|-------|--|----|
| 1     | Introduction . . . . .                         | 2  |
| 2     | Input/Output . . . . .                         | 4  |
| 2.1   | Input . . . . .                                | 4  |
| 2.1.1 | Analyze 7.5 . . . . .                          | 4  |
| 2.1.2 | imzML . . . . .                                | 4  |
| 2.1.3 | readMSIData . . . . .                          | 5  |
| 2.2   | Output . . . . .                               | 5  |
| 2.2.1 | RData files . . . . .                          | 5  |
| 3     | Data exploration and visualization . . . . .   | 6  |
| 3.1   | An example dataset . . . . .                   | 6  |
| 3.2   | The <code>MSImageSet</code> object . . . . .   | 6  |
| 3.3   | Subsetting a <code>MSImageSet</code> . . . . . | 8  |
| 3.4   | Plotting mass spectra . . . . .                | 9  |
| 3.5   | Plotting ion images . . . . .                  | 9  |
| 4     | Pre-processing . . . . .                       | 12 |
| 4.1   | Normalization . . . . .                        | 12 |
| 4.2   | Smoothing . . . . .                            | 12 |
| 4.3   | Baseline reduction . . . . .                   | 12 |
| 4.4   | Peak picking . . . . .                         | 13 |
| 4.5   | Peak alignment . . . . .                       | 14 |
| 4.6   | Peak filtering . . . . .                       | 14 |
| 4.7   | Data reduction . . . . .                       | 15 |
| 4.8   | Batch processing . . . . .                     | 15 |
| 5     | Analysis . . . . .                             | 17 |
| 5.1   | Principal components analysis (PCA) . . . . .  | 17 |
| 5.2   | Partial least squares (PLS) . . . . .          | 17 |

## Cardinal: Analytic tools for mass spectrometry imaging

|       |  |    |
|-------|--|----|
| 5.3   | Orthogonal partial least squares (O-PLS) | 18 |
| 5.4   | Spatially-aware k-means clustering       | 19 |
| 5.5   | Spatial shrunken centroids               | 19 |
| 6     | Examples                                 | 21 |
| 7     | Advanced Topics                          | 23 |
| 7.1   | Apply                                    | 23 |
| 7.1.1 | pixelApply                               | 23 |
| 7.1.2 | featureApply                             | 25 |
| 7.2   | Simulation                               | 27 |
| 7.2.1 | Simulation of spectra                    | 27 |
| 7.2.2 | Simulation of images                     | 28 |
| 7.2.3 | Advanced simulation                      | 29 |
| 8     | Session info                             | 31 |

## 1 Introduction

---

The *R* package *Cardinal* has been created to fill the need for an efficient, open-source tool for the analysis of imaging data—specifically mass spectrometry imaging data. *Cardinal* is built upon data structures which follow Bioconductor (<http://www.bioconductor.org/>) standards for data classes in order to provide an additional level of convenience and familiarity to those who may be used to performing bioinformatic analyses in *R*.

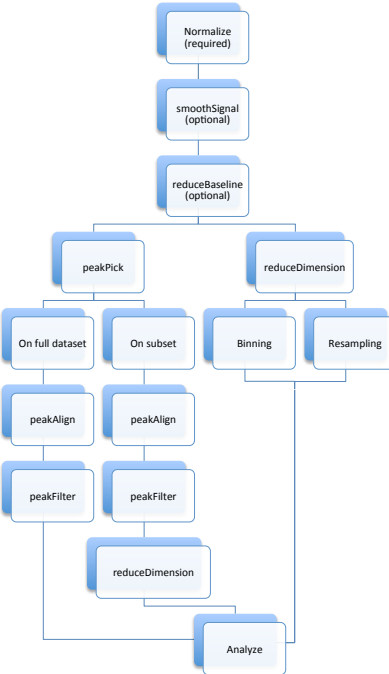
Analysis in imaging data includes many things, such as visualization, pre-processing, and multivariate statistical techniques. Both supervised and unsupervised statistical methods are supported in *Cardinal*, including image segmentation (clustering), principle components analysis, and classification techniques such as partial least Squares discriminant analysis.

Figure 1 charts out the workflow for mass spectrometry imaging data analysis.

This is a brief walkthrough of some of the basic functionality of *Cardinal*. For a more detailed view of the functionality of a given method, see the *R* help file.

Additional *R* packages useful for the analysis of mass spectrometry experiments are *MSnbase* [1] and *MALDIquant* [2], which are both designed for traditional proteomics analyses. *MALDIquant* also has limited support for mass spectrometry imaging data.

**Cardinal: Analytic tools for mass spectrometry imaging**



**Figure 1: Cardinal workflow for pre-processing and analysis**

## 2 Input/Output

---

### 2.1 Input

In order to be analyzed in *Cardinal*, input data must be in either Analyze 7.5 and imzML format. These are two of the most common data exchange formats in imaging mass spectrometry.

#### 2.1.1 Analyze 7.5

Originally designed for MRI data by the Mayo Clinic, Analyze 7.5 is a common format used for exchange of mass spectrometry imaging data.

The Analyze format uses a collection of three files with extensions '.hdr', '.img', and '.t2m' to store data. To read datasets stored in the Analyze format, use the `readAnalyze` function. All three files must be present in the same folder and have the same name (except for the file extension) for the data to be read properly.

```
> name <- "This is the common name of your .hdr, .img, and .t2m files"
> folder <- "/This/is/the/path/to/the/folder/containing/the/files"
> data <- readAnalyze(name, folder)
```

Large Analyze files can also be attached on-disk without fully loading them into memory by using the `attach.only=TRUE` option. Not all *Cardinal* features are supported for on-disk datasets.

For more information on reading Analyze files, type `?readAnalyze`.

#### 2.1.2 imzML

The open XML-based format imzML is a more recently developed format specifically designed for interchange of mass spectrometry imaging datasets [3]. Many other formats can be converted to imzML with the help of free applications available online. See <http://www.imzml.org> for more information and links to free converters.

The imzML format uses two files with extensions '.imzML' and '.ibd' to store data. To read datasets stored in the imzML format, use the `readImzML` function. Both files must be present in the same folder and have the same name (again, except for the file extension) for the data to be read properly.

```
> name <- "This is the common name of your .imzML and .ibd files"
> folder <- "/This/is/the/path/to/the/folder/containing/the/files"
> data <- readImzML(name, folder)
```

Large imzML files can also be attached on-disk without fully loading them into memory by using the `attach.only=TRUE` option. Not all *Cardinal* features are supported for on-disk datasets.

Both 'continuous' and 'processed' imzML format are supported, but currently only 'continuous' format can be attached using `attach.only=TRUE`.

For more information on reading imzML files, type `?readImzML`.

## **Cardinal: Analytic tools for mass spectrometry imaging**

### 2.1.3 readMSIData

*Cardinal* also provides the convenience function `readMSIData`, which can automatically recognize whether the data format is Analyze or imzML based on file extensions. The same rules for naming conventions apply as described above, but one need only provide the path to any of the data files. For example, to read an Analyze file, providing the path to any of the 'hdr', 'img', or 't2m' will work. Likewise, providing the path to either the 'imzML' or 'ibd' file will work for reading data stored in the imzML format.

```
> file <- "/This/is/the/path/to/an/imaging/data/file.extension"  
> data <- readMSIData(file)
```

## 2.2 Output

### 2.2.1 RData files

Any *R* object, including those created by *Cardinal*, can be saved as an **RData** file using the `save` and loaded using the `load` function.

```
> save(data, file="/Where/to/save/the/data.RData")  
> load("/Where/to/save/the/data.RData")
```

When an **RData** file is loaded, the saved object appears in the global environment for the *R* session and is available for access by name, just as it was in the session during which it was saved. This functionality is part of *R*; see `?save` and `?load` for more details.

## 3 Data exploration and visualization

Mass spectrometry imaging datasets in *Cardinal* are stored in `MSImageSet` objects. This allows *Cardinal* to keep track of the spectra, pixel coordinates,  $m/z$  values and more in one place for the dataset. The `MSImageSet` object is described in more detail below. There are many methods for both creating and manipulating `MSImageSet` objects in *Cardinal*. We now describe some of these methods.

### 3.1 An example dataset

To illustrate methods for the `MSImageSet` objects, we begin by creating a simple, simulated dataset using the *Cardinal* function `generateImage`. This dataset will be the running example for this section. For more details on simulating mass spectrometry images, see `?generateImage` or Section 7.2 *Simulation*.

```
> pattern <- factor(c(0, 0, 2, 2, 0, 0, 0, 0, 0, 0, 2, 2, 0,
+ 0, 0, 0, 0, 0, 0, 1, 2, 2, 0, 0, 0, 0, 0, 0, 2, 1, 1, 2,
+ 2, 0, 0, 0, 0, 0, 0, 1, 2, 2, 2, 2, 0, 0, 0, 0, 1, 2, 2,
+ 2, 2, 2, 0, 0, 0, 0, 2, 2, 2, 2, 2, 2, 2, 0, 0, 0, 2,
+ 2, 0, 0, 0, 0, 0, 0, 2, 2, 0, 0, 0, 0, 0, 0),
+ levels=c(0,1,2), labels=c("blue", "black", "red"))
> set.seed(1)
> msset <- generateImage(pattern, coord=expand.grid(x=1:9, y=1:9),
+ range=c(1000, 5000), centers=c(2000, 3000, 4000),
+ resolution=100, step=3.3, as="MSImageSet")
> summary(msset)

Class: MSImageSet
Features: m/z = 1000 ... m/z = 4999.6 (1213 total)
Pixels: x = 1, y = 1 ... x = 9, y = 9 (81 total)
x: 1 ... 9
y: 1 ... 9
Size in memory: 1 Mb
```

The above code creates a simulated MS imaging dataset called `msset`, which is  $9 \times 9$  pixels, with a mass range from  $m/z$  1000 to  $m/z$  5000. There are three peaks, occurring at  $m/z$  2000,  $m/z$  3000, and  $m/z$  4000. Each of these peaks corresponds to a distinct region of interest. These are saved in the `factor pattern`. A `factor` is the standard way of storing categorical variables in *R*. All pixels with `pattern` = 0 correspond to the region with peak at  $m/z$  2000, `pattern` = 1 corresponds to the peak at  $m/z$  3000, and `pattern` = 2 corresponds to  $m/z$  4000.

We'll label these regions of interest "blue" pixels, "black" pixels, and "red" pixels, respectively.

### 3.2 The `MSImageSet` object

Most important aspects of a mass spectrometry imaging dataset stored in an `MSImageSet` object can be accessed by simple methods.

## Cardinal: Analytic tools for mass spectrometry imaging

For example,  $m/z$ -values are accessed by the method `mz`, pixel coordinates are accessed by the method `coord`, and the mass spectra themselves are accessed by the method `spectra`. The mass spectra are stored as a matrix with each column corresponding to the mass spectrum at a single pixel.

```
> head(mz(msset), n=10) # first 10 m/z values
[1] 1000.0 1003.3 1006.6 1009.9 1013.2 1016.5 1019.8 1023.1 1026.4 1029.7

> head(coord(msset), n=10) # first 10 pixel coordinates
      x y
x = 1, y = 1 1 1
x = 2, y = 1 2 1
x = 3, y = 1 3 1
x = 4, y = 1 4 1
x = 5, y = 1 5 1
x = 6, y = 1 6 1
x = 7, y = 1 7 1
x = 8, y = 1 8 1
x = 9, y = 1 9 1
x = 1, y = 2 1 2

> head(spectra(msset)[,1], n=10) # first 10 intensities in the first mass spectrum
[1] 13.72308 11.93339 12.79512 12.50510 11.71788 13.92943 12.18719 12.24365 12.62183 12.30359
```

The methods `nrow` and `ncol` can be used to retrieve the number of features and number of pixels in an object, respectively. The method `dim` gives both number of features and number of pixels, while `dims` gives number of features as well as spatial dimensions of the image.

```
> nrow(msset)
Features
1213

> ncol(msset)
Pixels
81

> dim(msset)
Features  Pixels
1213      81

> dims(msset)
      iData
Features 1213
x         9
y         9
```

Two other helpful methods are `features` and `pixels`. These are useful for retrieving the feature number and pixel number (i.e., the row and column in the `spectra(msset)` matrix) corresponding to items of interest such as specific  $m/z$ -values or pixel coordinates.

## Cardinal: Analytic tools for mass spectrometry imaging

```
> features(msset, mz=3000) # returns the feature number most closely matching m/z 3000
m/z = 2999.8
      607

> mz(msset)[607]
[1] 2999.8

> pixels(msset, coord=list(x=5, y=5)) # returns the pixel number for x = 5, y = 5
x = 5, y = 5
      41

> pixels(msset, x=5, y=5) # also returns the pixel number for x = 5, y = 5
x = 5, y = 5
      41

> coord(msset)[41,]
      x y
x = 5, y = 5 5 5
```

See `?MSImageSet` for more details and additional methods.

*Technical note: `MSImageSet` is an `S4` class. It inherits from the more general `SImageSet` class, which itself inherits from the `iSet` virtual class. The `iSet` virtual class is designed around the same design principles as the `eSet` class provided by Biobase. See the “Cardinal development” vignette for more information.)*

### 3.3 Subsetting a `MSImageSet`

A `MSImageSet` can be subset by row and column like an ordinary *R* matrix or *data.frame*, where rows correspond to the features (*m/z*-values) and columns correspond to pixels (locations associated with mass spectra). Subsetting will return a new `MSImageSet`.

For example, we can subset by *m/z*-values so that we only keep the mass range from *m/z* 2500 to *m/z* 4500.

```
> tmp <- msset[2500 < mz(msset) & mz(msset) < 4500,]
> range(mz(msset))
[1] 1000.0 4999.6
> range(mz(tmp))
[1] 2501.5 4498.0
```

Alternatively, we can subset by pixel coordinates. To keep only pixels with *x*-coordinates greater than 5, we can do the following.

```
> tmp <- msset[, coord(msset)$x > 5]
> range(coord(msset)$x)
[1] 1 9
> range(coord(tmp)$x)
```

## Cardinal: Analytic tools for mass spectrometry imaging

```
[1] 6 9
```

We can also subset in both ways at once.

```
> tmp <- msset[2500 < mz(msset) & mz(msset) < 4500, coord(msset)$x > 5]
> range(mz(tmp))
[1] 2501.5 4498.0
> range(coord(tmp)$x)
[1] 6 9
```

It is also possible to manually select a region of interest and use it to subset the dataset. This is done using the `select` method, which will be introduced in Section 3.5 *Plotting ion images*.

### 3.4 Plotting mass spectra

Mass spectra from an `MSImageSet` can be displayed using the `plot` method. To plot the mass spectrum at the first pixel of our `MSImageSet`, we do the following:

```
> plot(msset, pixel=1)
```

The result of which is shown in Figure 2a.

Instead of pixel number, we can specify a set of coordinates corresponding to the mass spectrum we want to plot. The following produces Figure 2b, which is the mean spectrum for the pixel at spatial location (5,5), and all other spectra within a 2 pixel neighborhood of that location.

```
> plot(msset, coord=list(x=5, y=5), plusminus=2)
```

Finally, we can plot multiple spectra at once, as shown in Figure 2c. This is done below by specifying a vector for the `pixel` argument. The plots are displayed simultaneously by setting `superpose = TRUE` and `key = TRUE` generates a legend. The `pixel.groups` here indicates that the pixels should be grouped by their classifications as encoded by the pattern factor. By default, *Cardinal* averages over spectra in the same group.

```
> mycol <- c("blue", "black", "red")
> plot(msset, pixel=1:ncol(msset), pixel.groups=pattern, superpose=TRUE, key=TRUE, col=mycol)
```

### 3.5 Plotting ion images

Ion images from an `MSImageSet` can be plotted using the `image` method. To plot the ion image for the first feature, Figure 3a, we use:

```
> image(msset, feature=1)
```

The mean ion image for the neighborhood of  $m/z$  4000 with radius 10, i.e.  $m/z$  [3990,4010] is shown in Figure 3b.

## Cardinal: Analytic tools for mass spectrometry imaging

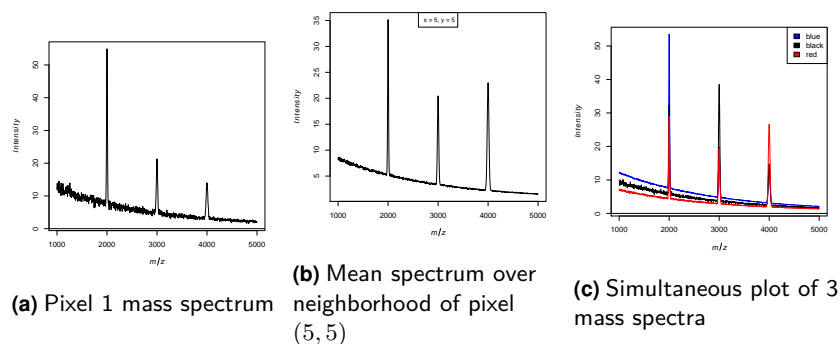


Figure 2: Plotting mass spectra

```
> image(msset, mz=4000, plusminus=10)
```

In Figure 3c the ion images for  $m/z$  2000,  $m/z$  3000, and  $m/z$  4000 are displayed simultaneously.

```
> mycol <- c("blue", "black", "red")
> image(msset, mz=c(2000, 3000, 4000), col=mycol, superpose=TRUE)
```

The ion image for  $m/z$  2000 is shown in Figure 3d, with a custom color scale from white to blue. The most intense "hotspots" are suppressed.

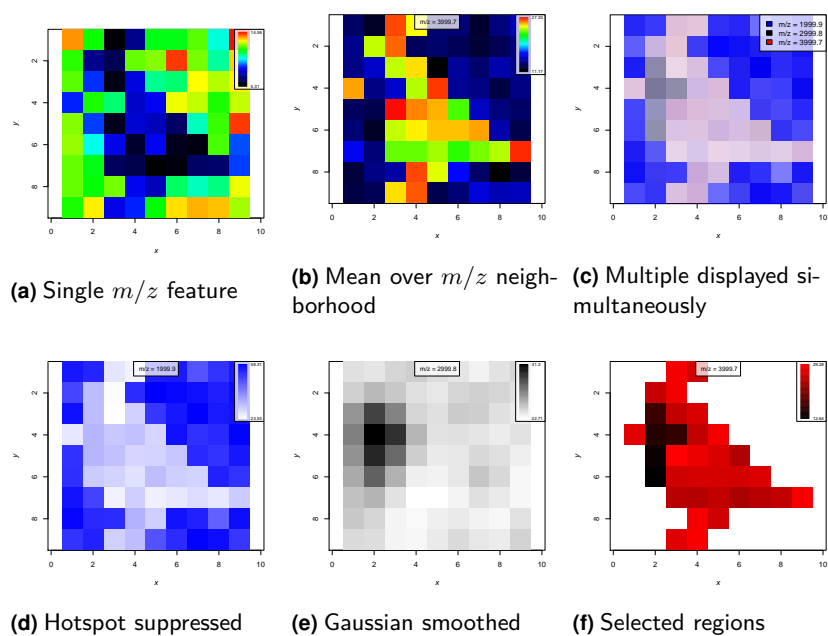
```
> mycol <- gradient.colors(100, start="white", end="blue")
> image(msset, mz=2000, col.regions=mycol, contrast.enhance="suppress")
```

In Figure 3e, a smoothed ion image for  $m/z$  3000 with a custom color scale from white to black is presented.

```
> mycol <- gradient.colors(100, start="white", end="black")
> image(msset, mz=3000, col.regions=mycol, smooth.image="gaussian")
```

Finally, in Figure 3f, for only those pixels defined as being from the "black" and "red" regions, we plot the ion image of  $m/z$  4000 with a custom color scale from black to red.

```
> msset2 <- msset[,pattern == "black" | pattern == "red"]
> mycol <- gradient.colors(100, start="black", end="red")
> image(msset2, mz=4000, col.regions=mycol)
```



**Figure 3: Plotting ion images**

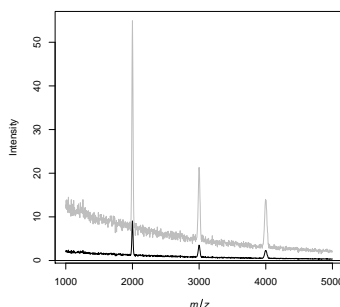
## 4 Pre-processing

### 4.1 Normalization

Normalization is perhaps the most important pre-processing step before any kind of analysis should be performed on biological datasets, and mass spectrometry imaging experiments are no different in this regard. *Cardinal* provides normalization to total ion current (TIC), commonly used in MSI analysis (see [4] for a discussion of this method). In the first command below, we only perform the normalization on the first pixel in order to show a plot of the processing results in Figure 4. In the second, we perform normalization on the whole dataset.

```
> normalize(msset, pixel=1, method="tic", plot=TRUE)
```

```
> msset2 <- normalize(msset, method="tic")
```



**Figure 4: Total ion current (TIC) normalization**

### 4.2 Smoothing

Smoothing the mass spectra is useful for reducing noise, which can improve detection of peaks. *Cardinal* provides several common methods for smoothing mass spectra, including Gaussian kernel smoothing (Figure 5a), Savitsky-Golay smoothing (Figure 5b), and a simple moving average filter [5].

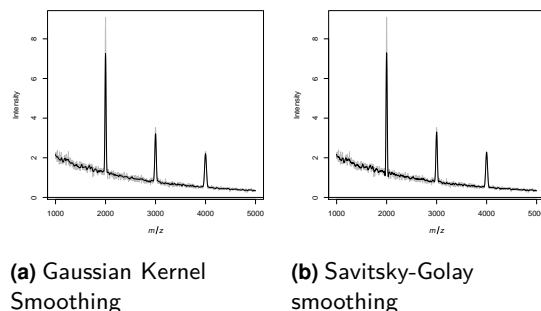
```
> smoothSignal(msset2, pixel=1, method="gaussian", window=9, plot=TRUE)
```

```
> smoothSignal(msset2, pixel=1, method="sgolay", window=15, plot=TRUE)
```

```
> msset3 <- smoothSignal(msset2, method="gaussian", window=9)
```

### 4.3 Baseline reduction

Baseline reduction is often necessary for many datasets, especially those obtained through matrix-assisted methods such as MALDI ([5]). *Cardinal* implements a simple version that interpolates a baseline from local medians or local minima, while attempting to preserve



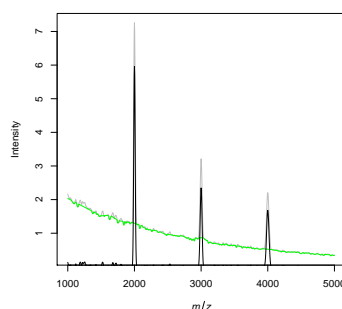
**Figure 5: Smoothing techniques**

the signal from mass spectral peaks. Figure 6 shows baseline reduction for a single pixel, where the green curve represents the estimated baseline and the baseline-reduced spectrum is plotted in black.

```
> reduceBaseline(msset3, pixel=1, method="median", blocks=50, plot=TRUE)
```

We can also reduce baseline across all pixels in the image.

```
> msset4 <- reduceBaseline(msset3, method="median", blocks=50)
```



**Figure 6: Baseline reduction using interpolation from medians**

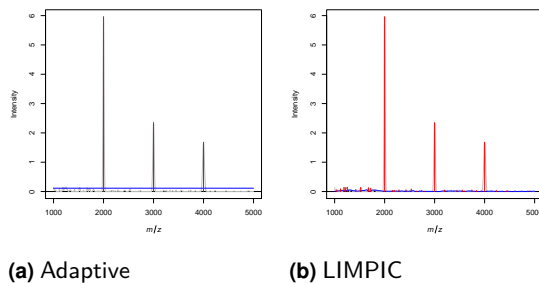
## 4.4 Peak picking

Peak picking is a common form of data reduction that reduces the signal to relevant data peaks. *Cardinal* implements three varieties based on a user-specified signal-to-noise ratio (SNR). The “simple” version interpolates a constant noise pattern, the “adaptive” version interpolates an adaptive noise pattern Figure 7a, and “limpic” implements the LIMPIC algorithm for peak detection Figure 7b.

```
> peakPick(msset4, pixel=1, method="adaptive", SNR=3, plot=TRUE)
```

```
> peakPick(msset4, pixel=1, method="limpic", SNR=3, plot=TRUE)
```

```
> msset5 <- peakPick(msset4, method="simple", SNR=3)
```



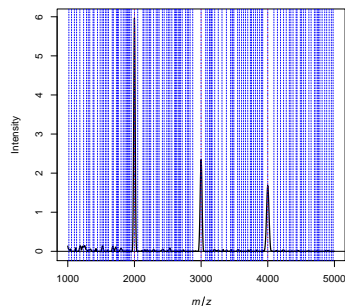
**Figure 7: Peak picking techniques**

## 4.5 Peak alignment

Peak alignment is necessary to account for possible inaccuracy in  $m/z$  measurements. Peaks can be aligned to a reference list of known  $m/z$  values, or to the local maxima in the mean spectrum. Figure 8 denotes the selected peaks by red vertical lines, and aligns the local maxima of the mean spectra to these peaks, as in [6].

```
> peakAlign(msset5, pixel=1, method="diff", plot=TRUE)
```

```
> msset6 <- peakAlign(msset5, method="diff")
```



**Figure 8: Peak alignment to the local maxima of the mean spectrum**

## 4.6 Peak filtering

Peak filtering removes peaks that occur infrequently, such as those which only occur in a small proportion of pixels. This is useful for removing extraneous peaks that are likely to be false positives.

```
> msset7 <- peakFilter(msset6, method="freq")  
> dim(msset6) # 89 peaks retained
```

```
Features  Pixels
    118      81

> dim(msset7) # 10 peaks retained

Features  Pixels
     3      81
```

## 4.7 Data reduction

Other common forms of data reduction include resampling and binning.

*Cardinal* can do binning for a fixed width, taken to be 25 in this example. The mean intensity of ions located in the same  $m/z$  bin is taken to be the response in the reduced version of the data. The results of binning on pixel 1 is plotted in Figure 9a. The original spectrum is plotted in black, with the binned version displayed simultaneously in red.

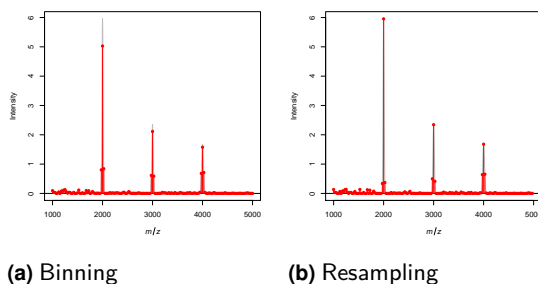
```
> reduceDimension(msset4, pixel=1, method="bin", width=25, units="mz", fun=mean, plot=TRUE)
```

There is also the option of doing resampling for a fixed step size. The results of resampling with step size 25  $m/z$  on pixel 1 is plotted in Figure 9b. The original spectrum is plotted in black, with the resampled version displayed simultaneously in red.

```
> reduceDimension(msset4, pixel=1, method="resample", step=25, plot=TRUE)
```

Data reduction can be done on the whole dataset at once.

```
> msset8 <- reduceDimension(msset4, method="resample", step=25)
```



**Figure 9: Data reduction via binning and resampling**

## 4.8 Batch processing

To simplify the above pre-processing steps, as well as save memory when processing on-disk data, *Cardinal* provides a batch processing method.

```
> msset9 <- batchProcess(msset, normalize=TRUE, smoothSignal=TRUE, reduceBaseline=TRUE)
> summary(msset9)
```

## Cardinal: Analytic tools for mass spectrometry imaging

```
Class: MSImageSet
Features: m/z = 1000 ... m/z = 4999.6 (1213 total)
Pixels: x = 1, y = 1 ... x = 9, y = 9 (81 total)
x: 1 ... 9
y: 1 ... 9
Size in memory: 1 Mb
```

```
> processingData(msset9)
```

```
Processing data
Cardinal version: 2.2.2
Files:
Normalization: tic
Smoothing: gaussian
Baseline reduction: median
Spectrum representation:
Peak picking:
```

Each step can be set to its default parameters by setting it to TRUE, or a list of options can be provided.

```
> msset10 <- batchProcess(msset, normalize=TRUE, smoothSignal=TRUE, reduceBaseline=list(blocks=200),
+   peakPick=list(SNR=12), peakAlign=TRUE)
> summary(msset10)
```

```
Class: MSImageSet
Features: m/z = 1046.2 ... m/z = 4953.4 (177 total)
Pixels: x = 1, y = 1 ... x = 9, y = 9 (81 total)
x: 1 ... 9
y: 1 ... 9
Size in memory: 0.4 Mb
```

```
> processingData(msset10)
```

```
Processing data
Cardinal version: 2.2.2
Files:
Normalization: tic
Smoothing: gaussian
Baseline reduction: median
Spectrum representation: centroid
Peak picking: simple
```

This method is particularly useful when processing larger-than-memory on-disk datasets to a smaller processed form, without loading the full data into memory.

See `?batchProcess` for more details and differences in behavior from the individual processing methods.

## 5 Analysis

For example workflows with analyses of real datasets, please see the vignettes in the companion data package *CardinalWorkflows*.

### 5.1 Principal components analysis (PCA)

Principal components analysis (PCA) is a multivariate statistical tool used for dimension reduction and exploratory data analysis. PCA can be useful when first exploring a dataset beyond plotting molecular ion images, but additional statistical analysis is usually necessary to extract meaningful results.

Below, we fit the first two principal components using *Cardinal*'s PCA method and plot their loadings and scores.

```
> pca <- PCA(msset4, ncomp=2)
```

```
> plot(pca)
```

```
> image(pca)
```

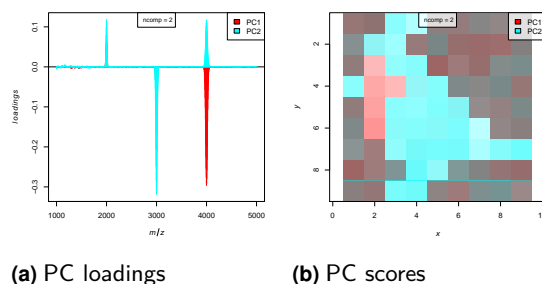


Figure 10: Principal components analysis

See ?PCA for more details.

### 5.2 Partial least squares (PLS)

Partial least squares (PLS), also called projection to latent structures, is a multivariate method from chemometrics that has been shown to be useful for classification of mass spectrometry images [7].

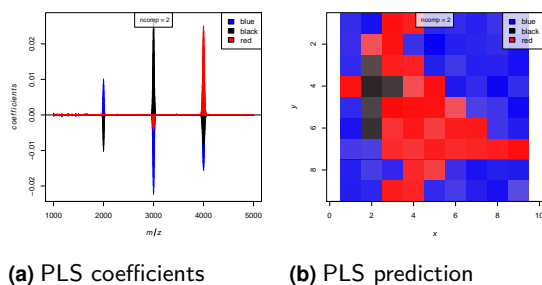
When used for classification, it is known as partial least squares discriminant analysis, or PLS-DA. PLS-DA works similarly to PCA, but it is a supervised method, so it requires data annotated with known labels.

Here, we train a PLS classifier using the `pattern` variable from earlier as our labels, and plot the results.

```
> pls <- PLS(msset4, y=pattern, ncomp=2)
```

```
> plot(pls, col=c("blue", "black", "red"))
```

```
> image(pls, col=c("blue", "black", "red"))
```



**Figure 11: Partial least squares**

When working with classification on real data, cross-validation should always be used, using the `cvApply` method, to avoid biased results. See `?cvApply` and `?PLS` for more details.

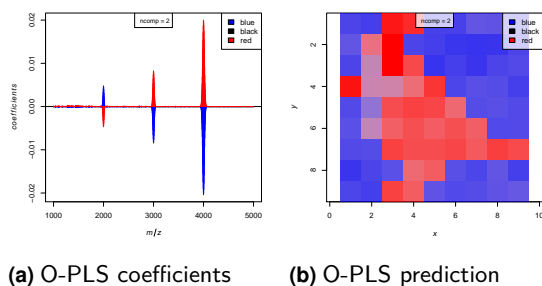
### 5.3 Orthogonal partial least squares (O-PLS)

Orthogonal partial least squares (O-PLS) is a variation on PLS. O-PLS can sometimes improve the interpretability of the PLS model coefficients, while producing similar accuracy. O-PLS-DA is also implemented in *Cardinal*

```
> opl <- OPLS(msset4, y=pattern, ncomp=2)
```

```
> plot(opl, col=c("blue", "black", "red"))
```

```
> image(opl, col=c("blue", "black", "red"))
```



**Figure 12: Orthogonal partial least squares**

O-PLS is primarily useful when many PLS components are required to fit an accurate model, since this often leads to unstable model coefficients. Try O-PLS when the PLS model coefficients are difficult to interpret, or the best PLS model uses a large number of components. See `?OPLS` for more details.

## 5.4 Spatially-aware k-means clustering

Spatially-aware clustering using k-means is available [6] through the `spatialKMeans` method. This method uses a spatial distance function to project the data to a kernel space before performing ordinary k-means clustering.

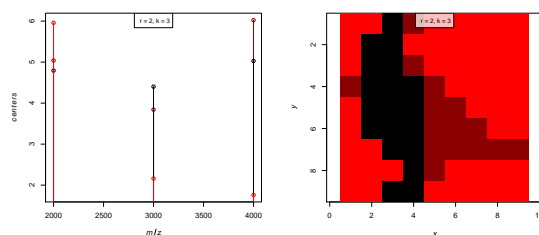
The parameters `r` and `k` are the neighborhood smoothing radius and the initial number of clusters.

Below, we create a spatial segmentation using spatially-aware clustering.

```
> set.seed(1)
> skm <- spatialKMeans(msset7, r=2, k=3, method="adaptive")

> plot(skm, col=c("darkred", "black", "red"), type=c('p','h'), key=FALSE)

> image(skm, col=c("darkred", "black", "red"), key=FALSE)
```



(a) Predicted centers (b) Predicted clusters

**Figure 13: Spatially-aware k-means clustering**

See `?spatialKMeans` for more details.

## 5.5 Spatial shrunk centroids

*Cardinal* offers a novel clustering and classification method based on the spatial smoothing [6] and nearest shrunk centroids [8]. This is the `spatialShrunkCentroids` method, which can be used both for clustering and for classification.

The parameters `r`, `k`, and `s` are the neighborhood smoothing radius, the initial number of clusters, and the sparsity parameter, respectively.

Below, we create a spatial segmentation using the spatial shrunk centroids method.

```
> set.seed(1)
> ssc <- spatialShrunkCentroids(msset7, r=1, k=5, s=3, method="adaptive")
```

## Cardinal: Analytic tools for mass spectrometry imaging

```
> plot(ssc, col=c("red", "black"), type=c('p','h'), key=FALSE)
```

```
> image(ssc, col=c("red", "black"), key=FALSE)
```

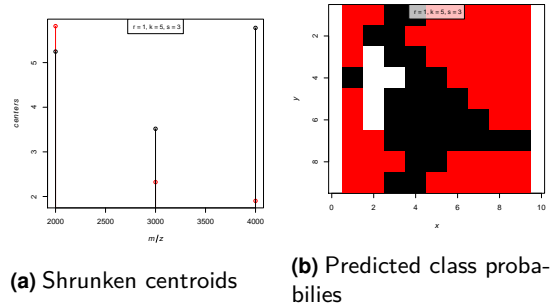


Figure 14: **Spatially-aware nearest shrunken centroids clustering**

A unique property of *Cardinal*'s spatial shrunken centroids method is that it allows for the automated selection of the number of clusters, driven in part by the sparsity parameter. Although we initialized the clustering above with 5 clusters, only 2 were used in the final segmentation.

See `?spatialShrunkenCentroids` for more options and details.

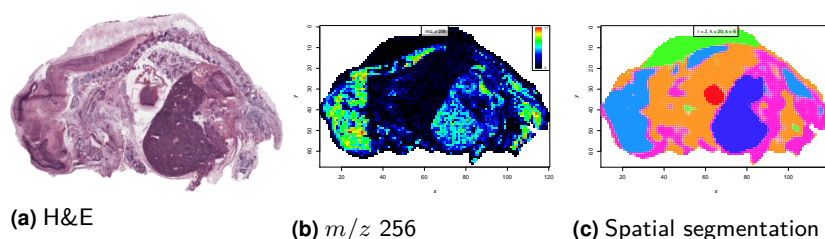
## 6 Examples

In-depth biological examples using real data can be found in the *CardinalWorkflows* package. Figure 15 shows an example using a cross-section of a whole pig fetus, and Figure 16 shows an example using a human renal cell carcinoma dataset. Both datasets and a thorough walkthrough of analyses are available in *CardinalWorkflows*.

To install *CardinalWorkflows*, run:

```
> BiocManager::install("CardinalWorkflows")
```

Please note that due to the size of the included datasets, downloading and installing *CardinalWorkflows* may take a long time.



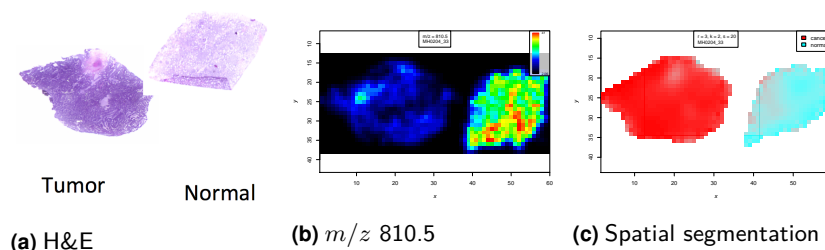
**Figure 15: Biological example of a pig fetus cross-section, showing the optical image, an ion image, and a segmentation created by Spatial Shrunken Centroids clustering**

Figure 15 uses a pig fetus cross-section as an example of unsupervised analysis of a mass spectrometry imaging experiment using *Cardinal*. To view the vignette associated with this dataset, install *CardinalWorkflows* and run:

```
> vignette("Workflows-clustering")
```

The dataset and its analyses can be loaded by running:

```
> data(pig206, pig206_analyses)
```



**Figure 16: Biological example of human renal cell carcinoma classification, showing an optical image, an ion image, and a segmentation created by Spatial Shrunken Centroids classification**

Figure 16 uses a human renal cell carcinoma dataset as an example of supervised analysis of a mass spectrometry imaging experiment using *Cardinal*. To view the vignette associated with this dataset, install *CardinalWorkflows* and run:

## **Cardinal: Analytic tools for mass spectrometry imaging**

```
> vignette("Workflows-classification")
```

The dataset and its analyses can be loaded by running:

```
> data(rcc, rcc_analyses)
```

See `?CardinalWorkflows` for more information.

## 7 Advanced Topics

### 7.1 Apply

The `apply` family of functions are a powerful feature of *R*. The `apply` function applies a function over margins of an array, while `sapply` applies a function over every element of a vector-like object. The function `tapply` applies a function over a “ragged” array, so that the function is applied over groups of values given by levels of another variable (usually a factor). In *Cardinal*, the methods `pixelApply` and `featureApply` allow `apply`-like functionality that combine traits of each of these, tailored for imaging datasets.

We need to mark which pixels are blue, black, and which are red, as in the `factor` `pattern` in Section 3.1.

```
> pData(msset)$pg <- pattern
```

Then we need to mark which features (which regions of the mass spectrum) belong to the peaks associated with “blue” ( $m/z$  2000), “black” ( $m/z$  3000), or “red” ( $m/z$  4000) pixels; the rest of the spectrum is marked as background noise (`bg`).

```
> fData(msset)$fg <- factor(rep("bg", nrow(fData(msset)))), levels=c("bg", "blue", "black", "red"))
> fData(msset)$fg[1950 < fData(msset)$mz & fData(msset)$mz < 2050] <- "blue"
> fData(msset)$fg[2950 < fData(msset)$mz & fData(msset)$mz < 3050] <- "black"
> fData(msset)$fg[3950 < fData(msset)$mz & fData(msset)$mz < 4050] <- "red"
```

Now we can experiment with different ways of plotting an imaging dataset.

#### 7.1.1 pixelApply

The method `pixelApply` allows functions to be applied over all pixels. The function is applied pixel-by-pixel to the feature vectors (mass spectra). Here, we use `pixelApply` to find the pixel-by-pixel mean intensity of different regions of the mass spectrum. We provide `fData(msset)$fg` as a grouping variable, since it indicates different regions of the mass spectrum we expect to be associated with either background noise, or blue, red, or black pixels. Since `pixelApply` knows to look in `msset` for the variable, we only need to provide `fg` to the argument `.feature.groups`.

```
> p1 <- pixelApply(msset, mean, .feature.groups=fg)
> p1[,1:30]
```

|       | x = 1, y = 1 | x = 2, y = 1 | x = 3, y = 1 | x = 4, y = 1 | x = 5, y = 1 | x = 6, y = 1 | x = 7, y = 1 |
|-------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| bg    | 5.591239     | 5.516053     | 3.395579     | 3.185078     | 5.615956     | 5.650231     | 4.957623     |
| blue  | 17.306620    | 17.046174    | 9.112457     | 8.557135     | 17.313809    | 17.481563    | 14.981038    |
| black | 9.974539     | 9.025602     | 8.412205     | 8.844019     | 9.631894     | 9.739096     | 8.005345     |
| red   | 7.628184     | 7.010724     | 12.982994    | 12.059920    | 7.883425     | 7.838796     | 7.077551     |
|       | x = 8, y = 1 | x = 9, y = 1 | x = 1, y = 2 | x = 2, y = 2 | x = 3, y = 2 | x = 4, y = 2 | x = 5, y = 2 |
| bg    | 5.285583     | 5.523104     | 5.173204     | 3.589339     | 3.361056     | 5.409369     | 5.962200     |
| blue  | 16.141836    | 16.873337    | 15.842321    | 11.037839    | 8.760469     | 16.625576    | 18.391835    |
| black | 9.232942     | 9.679532     | 8.708974     | 8.632293     | 8.432287     | 9.626762     | 9.028662     |
| red   | 7.870530     | 8.436895     | 7.819247     | 11.809760    | 12.806359    | 7.634099     | 7.807045     |
|       | x = 6, y = 2 | x = 7, y = 2 | x = 8, y = 2 | x = 9, y = 2 | x = 1, y = 3 | x = 2, y = 3 | x = 3, y = 3 |

## Cardinal: Analytic tools for mass spectrometry imaging

|  |           |           |           |           |           |           |           |
|--|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| bg   | 5.913981  | 5.710577  | 5.681514  | 5.860902  | 5.770548  | 4.274740  | 3.146736  |
| blue   | 18.392521 | 17.591820 | 17.528622 | 18.254246 | 17.893589 | 11.131247 | 8.295846  |
| black  | 10.234255 | 10.160928 | 10.370499 | 9.540406  | 10.013747 | 14.349932 | 7.718586  |
| red  | 7.588946  | 7.244261  | 7.563269  | 7.941176  | 8.005526  | 8.245206  | 11.762278 |
| x = 4, y = 3 x = 5, y = 3 x = 6, y = 3 x = 7, y = 3 x = 8, y = 3 x = 9, y = 3 x = 1, y = 4 |           |           |           |           |           |           |           |
| bg   | 3.316402  | 5.317308  | 5.688381  | 6.096776  | 5.727348  | 5.816680  | 3.290318  |
| blue   | 9.750688  | 16.232286 | 17.530305 | 18.911349 | 17.500409 | 17.829095 | 9.134986  |
| black  | 8.035290  | 9.444263  | 10.440994 | 10.710668 | 9.470964  | 9.166336  | 8.047249  |
| red  | 12.332335 | 6.764745  | 8.139496  | 7.897016  | 8.365797  | 8.130600  | 12.516607 |
| x = 2, y = 4 x = 3, y = 4  |           |           |           |           |           |           |           |
| bg   | 4.533675  | 4.376095  |           |           |           |           |           |
| blue   | 11.546565 | 11.061959 |           |           |           |           |           |
| black  | 15.589934 | 14.972439 |           |           |           |           |           |
| red  | 7.659493  | 8.236587  |           |           |           |           |           |

By comparing side-by-side with the ground truth (which we have stored in the variable `pData(msset)$pg`), we see the result is as we expected. For “blue” pixels, the mean intensity of features belonging to the “blue”-associated peak ( $m/z$  2000) is higher. For “black” pixels, the mean intensity of features belonging to the “black”-associated peak ( $m/z$  3000) is higher. Finally, for “red” pixels, the mean intensity of features belonging to the “red”-associated peak ( $m/z$  4000) is higher.

```
> cbind(pData(msset), t(p1))[1:30,c("pg", "blue", "black", "red")]
```

|              | pg    | blue      | black     | red       |
|--------------|-------|-----------|-----------|-----------|
| x = 1, y = 1 | blue  | 17.306620 | 9.974539  | 7.628184  |
| x = 2, y = 1 | blue  | 17.046174 | 9.025602  | 7.010724  |
| x = 3, y = 1 | red   | 9.112457  | 8.412205  | 12.982994 |
| x = 4, y = 1 | red   | 8.557135  | 8.844019  | 12.059920 |
| x = 5, y = 1 | blue  | 17.313809 | 9.631894  | 7.883425  |
| x = 6, y = 1 | blue  | 17.481563 | 9.739096  | 7.838796  |
| x = 7, y = 1 | blue  | 14.981038 | 8.005345  | 7.077551  |
| x = 8, y = 1 | blue  | 16.141836 | 9.232942  | 7.870530  |
| x = 9, y = 1 | blue  | 16.873337 | 9.679532  | 8.436895  |
| x = 1, y = 2 | blue  | 15.842321 | 8.708974  | 7.819247  |
| x = 2, y = 2 | red   | 11.037839 | 8.632293  | 11.809760 |
| x = 3, y = 2 | red   | 8.760469  | 8.432287  | 12.806359 |
| x = 4, y = 2 | blue  | 16.625576 | 9.626762  | 7.634099  |
| x = 5, y = 2 | blue  | 18.391835 | 9.028662  | 7.807045  |
| x = 6, y = 2 | blue  | 18.392521 | 10.234255 | 7.588946  |
| x = 7, y = 2 | blue  | 17.591820 | 10.160928 | 7.244261  |
| x = 8, y = 2 | blue  | 17.528622 | 10.370499 | 7.563269  |
| x = 9, y = 2 | blue  | 18.254246 | 9.540406  | 7.941176  |
| x = 1, y = 3 | blue  | 17.893589 | 10.013747 | 8.005526  |
| x = 2, y = 3 | black | 11.131247 | 14.349932 | 8.245206  |
| x = 3, y = 3 | red   | 8.295846  | 7.718586  | 11.762278 |
| x = 4, y = 3 | red   | 9.750688  | 8.035290  | 12.332335 |
| x = 5, y = 3 | blue  | 16.232286 | 9.444263  | 6.764745  |
| x = 6, y = 3 | blue  | 17.530305 | 10.440994 | 8.139496  |
| x = 7, y = 3 | blue  | 18.911349 | 10.710668 | 7.897016  |
| x = 8, y = 3 | blue  | 17.500409 | 9.470964  | 8.365797  |
| x = 9, y = 3 | blue  | 17.829095 | 9.166336  | 8.130600  |

## Cardinal: Analytic tools for mass spectrometry imaging

```
x = 1, y = 4 red 9.134986 8.047249 12.516607
x = 2, y = 4 black 11.546565 15.589934 7.659493
x = 3, y = 4 black 11.061959 14.972439 8.236587
```

We can manually construct the images corresponding to the mean intensity of the three peaks centered at  $m/z$  2000,  $m/z$  3000, and  $m/z$  4000 and plot their images. This is shown in Figure 17

```
> tmp1 <- MSImageSet(spectra=t(as.vector(p1["blue",])), coord=coord(msset), mz=2000)
> image(tmp1, feature=1, col.regions=alpha.colors("blue", 100), sub="m/z = 2000")
```

```
> tmp1 <- MSImageSet(spectra=t(as.vector(p1["black",])), coord=coord(msset), mz=3000)
> image(tmp1, feature=1, col.regions=alpha.colors("black", 100), sub="m/z = 3000")
```

```
> tmp2 <- MSImageSet(spectra=t(as.vector(p1["red",])), coord=coord(msset), mz=4000)
> image(tmp2, feature=1, col.regions=alpha.colors("red", 100), sub="m/z = 4000")
```

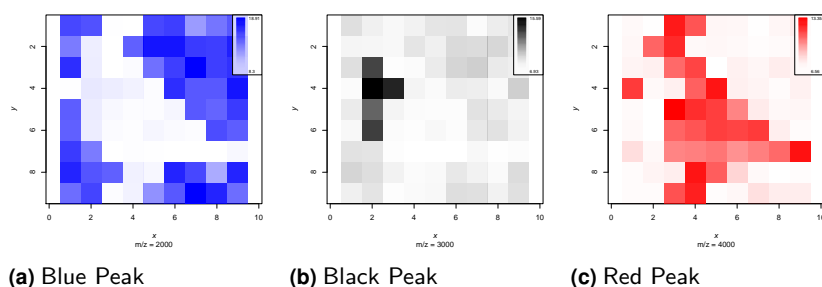


Figure 17: Mean intensities of the three peaks centered at  $m/z$  2000,  $m/z$  3000 and  $m/z$  4000

If only the plots are desired rather than the actual data, then `image` can be used to perform these steps automatically while producing the plot. See *Cardinal plotting* for how to do this.

### 7.1.2 featureApply

The method `featureApply` allows functions to be applied over all features. The function is applied to the flattened false-image vectors. These vectors are the pixel-by-pixel intensities of a single-feature image, not including missing pixels. Here, we use `featureApply` to find the mean spectrum for different groups of pixels. We provide `pData(msset)$pg` as a grouping variable, since it indicates the kind of pixel. We desire mean spectra for the black pixels, the red pixels, and the blue pixels. As before, since `featureApply` knows to look in `msset`, we only need to provide `pg` to the argument `.pixel.groups`.

```
> f1 <- featureApply(msset, mean, .pixel.groups=pg)
> f1[,1:30]
```

|       | m/z = 1000 | m/z = 1003.3 | m/z = 1006.6 | m/z = 1009.9 | m/z = 1013.2 | m/z = 1016.5 | m/z = 1019.8 |
|-------|------------|--------------|--------------|--------------|--------------|--------------|--------------|
| blue  | 12.113878  | 12.077172    | 12.271484    | 12.011245    | 12.106858    | 12.291783    | 12.011774    |
| black | 10.005194  | 9.491585     | 8.920648     | 8.395593     | 9.349371     | 8.972390     | 8.498503     |
| red   | 7.019996   | 7.095126     | 7.206785     | 7.046938     | 6.770562     | 7.135863     | 6.915438     |

## Cardinal: Analytic tools for mass spectrometry imaging

|       |              |              |              |              |              |              |              |
|-------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|       | m/z = 1023.1 | m/z = 1026.4 | m/z = 1029.7 | m/z = 1033   | m/z = 1036.3 | m/z = 1039.6 | m/z = 1042.9 |
| blue  | 11.936433    | 12.210184    | 11.989588    | 11.848585    | 11.790103    | 12.031041    | 11.732924    |
| black | 9.054659     | 9.949408     | 8.628287     | 8.868370     | 8.991401     | 9.277913     | 8.660284     |
| red   | 7.033159     | 7.091420     | 6.707759     | 7.258157     | 6.922927     | 7.025668     | 7.069312     |
|       | m/z = 1046.2 | m/z = 1049.5 | m/z = 1052.8 | m/z = 1056.1 | m/z = 1059.4 | m/z = 1062.7 | m/z = 1066   |
| blue  | 11.867339    | 11.796285    | 11.848332    | 11.996489    | 11.853608    | 11.766950    | 12.006615    |
| black | 8.677262     | 9.554605     | 9.319258     | 8.995928     | 8.754828     | 9.730654     | 9.159634     |
| red   | 7.120718     | 6.970966     | 6.852143     | 6.636296     | 6.992209     | 6.835762     | 6.493719     |
|       | m/z = 1069.3 | m/z = 1072.6 | m/z = 1075.9 | m/z = 1079.2 | m/z = 1082.5 | m/z = 1085.8 | m/z = 1089.1 |
| blue  | 11.757288    | 11.696140    | 11.544314    | 11.550182    | 11.684293    | 11.747299    | 11.787957    |
| black | 8.890601     | 8.980763     | 10.102702    | 8.745331     | 9.359917     | 8.196362     | 9.204949     |
| red   | 6.675064     | 6.787080     | 6.970741     | 6.662426     | 6.664785     | 6.829065     | 6.849054     |
|       | m/z = 1092.4 | m/z = 1095.7 |              |              |              |              |              |
| blue  | 11.472797    | 11.710456    |              |              |              |              |              |
| black | 9.279710     | 9.049170     |              |              |              |              |              |
| red   | 7.000153     | 6.704578     |              |              |              |              |              |

Again, we can check the results by plotting them in Figure 18.

```
> plot(mz(msset), f1["blue",], type="l", xlab="m/z", ylab="Intensity", col="blue")
> plot(mz(msset), f1["black",], type="l", xlab="m/z", ylab="Intensity", col="black")
> plot(mz(msset), f1["red",], type="l", xlab="m/z", ylab="Intensity", col="red")
```

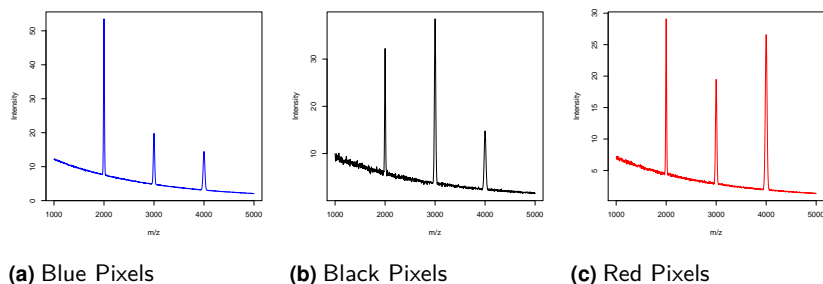


Figure 18: Mean spectra of blue, black, and red regions

As expected, we see the mean spectrum of the blue pixels has a higher peak at  $m/z$  2000, we see the mean spectrum of the black pixels has a higher peak at  $m/z$  3000, while the mean spectrum of the red pixels has a higher peak at  $m/z$  4000. As before, if only the plots are desired rather than the actual data, then `plot` can be used to perform these steps automatically. See *Cardinal plotting* for how to do this.

## 7.2 Simulation

*Cardinal* provides functions for the simulation of mass spectra and mass spectrometry imaging datasets. This is of interest to developers for testing newly developed methodology for analyzing mass spectrometry imaging experiments.

### 7.2.1 Simulation of spectra

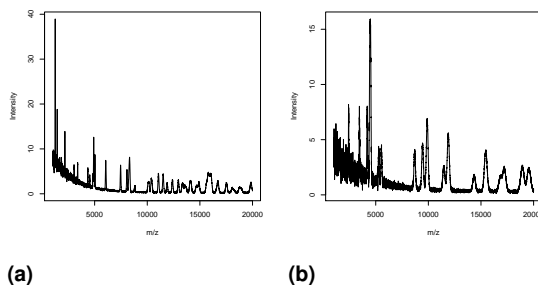
The `generateSpectrum` function can be used to simulate mass spectra. Its parameters can be tuned to simulate different kinds of mass spectra from different kinds of machines, and different protein and peptide patterns.

One spectrum with  $m/z$  range from 1001 to 20000, 50 randomly selected peaks, baseline 3000, and  $m/z$  resolution 100 is generated below and plotted in Figure 19a.

```
> set.seed(1)
> s1 <- generateSpectrum(1, range=c(1001, 20000), centers=runif(50, 1001, 20000), baseline=2000, resolution=100)
> plot(x ~ t, data=s1, type="l", xlab="m/z", ylab="Intensity")
```

An example with fewer peaks, larger baseline, and lower resolution (Figure 19b):

```
> set.seed(2)
> s2 <- generateSpectrum(1, range=c(1001, 20000), centers=runif(20, 1001, 20000), baseline=3000, resolution=50)
> plot(x ~ t, data=s2, type="l", xlab="m/z", ylab="Intensity")
```

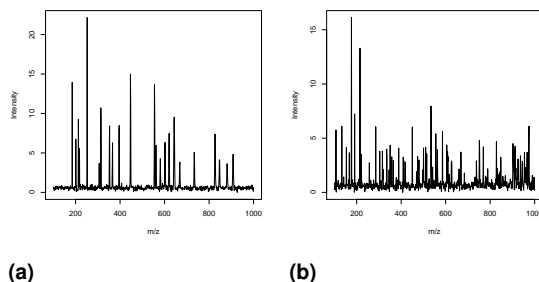


**Figure 19: MALDI-like simulated spectra**

Above we simulated MALDI-like spectra. We can also simulate DESI-like spectra, shown in Figure 20.

```
> set.seed(3)
> s3 <- generateSpectrum(1, range=c(101, 1000), centers=runif(25, 101, 1000), baseline=0, resolution=250, r)
> plot(x ~ t, data=s3, type="l", xlab="m/z", ylab="Intensity")
```

```
> set.seed(4)
> s4 <- generateSpectrum(1, range=c(101, 1000), centers=runif(100, 101, 1000), baseline=0, resolution=500, r)
> plot(x ~ t, data=s4, type="l", xlab="m/z", ylab="Intensity")
```



**Figure 20: DESI-like simulated spectra**

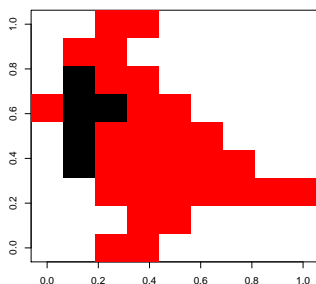
## 7.2.2 Simulation of images

The `generateImage` function can be used to simulate mass spectral images. This is a simple wrapper for `generateSpectra` that will generate unique spectral patterns based on a spatial pattern. The generated mass spectra will have a unique peak associated with each region. The pattern must have discrete regions, most easily given in the form of an integer matrix. We use a matrix in the pattern of a cardinal.

```
> data <- matrix(c(NA, NA, 1, 1, NA, NA, NA, NA, NA, NA, 1, 1, NA, NA,
+ NA, NA, NA, NA, NA, 0, 1, 1, NA, NA, NA, NA, NA, 1, 0, 0, 1,
+ 1, NA, NA, NA, NA, NA, 0, 1, 1, 1, 1, NA, NA, NA, NA, 0, 1, 1,
+ 1, 1, 1, NA, NA, NA, NA, 1, 1, 1, 1, 1, 1, 1, NA, NA, NA, 1,
+ 1, NA, NA, NA, NA, NA, NA, 1, 1, NA, NA, NA, NA, NA), nrow=9, ncol=9)
```

As seen in Figure ??, we can plot the ground truth image directly.

```
> image(data[,ncol(data):1], col=c("black", "red"))
```



**Figure 21: Ground truth image used to generate the simulated dataset**

Now we generate the dataset. To make it easy to visualize, we set up the `range` and `step` size so that the feature indices correspond directly to their values. We create two peaks at  $m/z$  100 and  $m/z$  200, one of which is associated with each region in the image.

```
> set.seed(1)
> img1 <- generateImage(data, range=c(1,1000), centers=c(100,200), step=1, as="MSImageSet")
```

## Cardinal: Analytic tools for mass spectrometry imaging

Now to confirm the reasonability of our simulated dataset, we plot images corresponding to the two peaks associated with each region in Figure 22b. (Note that rows in the original matrix correspond to the x-axis in the image and the columns correspond to the y-axis.)

```
> image(img1, mz=100, col.regions=alpha.colors("black", 100))
```

```
> image(img1, mz=200, col.regions=alpha.colors("red", 100))
```

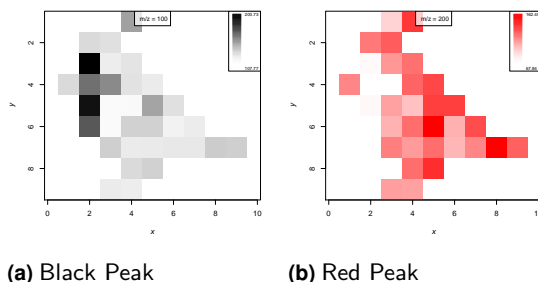


Figure 22: Generated image from an integer matrix

We can generate the same kind of dataset using a `factor` and a `data.frame` of coordinates, as is done in the running example for earlier sections of this vignette.

```
> pattern <- factor(c(0, 0, 2, 2, 0, 0, 0, 0, 0, 0, 0, 2, 2, 0,
+ 0, 0, 0, 0, 0, 0, 1, 2, 2, 0, 0, 0, 0, 0, 2, 1, 1, 2,
+ 2, 0, 0, 0, 0, 0, 1, 2, 2, 2, 2, 0, 0, 0, 0, 1, 2, 2,
+ 2, 2, 2, 0, 0, 0, 0, 2, 2, 2, 2, 2, 2, 2, 0, 0, 0, 2,
+ 2, 0, 0, 0, 0, 0, 0, 2, 2, 0, 0, 0, 0, 0),
+ levels=c(0,1,2), labels=c("blue", "black", "red"))
> coord <- expand.grid(x=1:9, y=1:9)
> set.seed(2)
> msset <- generateImage(pattern, coord=coord, range=c(1000, 5000), centers=c(2000, 3000, 4000), resolution=100)
```

Again, we can plot the images to see that the simulated dataset is the same pattern as before (though the exact intensities will differ, because we have used a different seed for the random number generator), Figure 23.

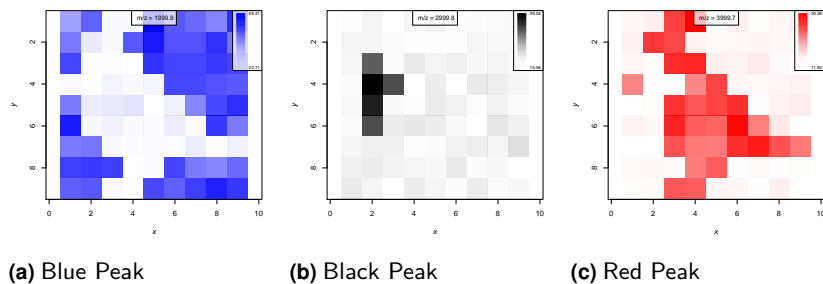
```
> image(msset, mz=2000, col.regions=alpha.colors("blue", 100))
```

```
> image(msset, mz=3000, col.regions=alpha.colors("black", 100))
```

```
> image(msset, mz=4000, col.regions=alpha.colors("red", 100))
```

### 7.2.3 Advanced simulation

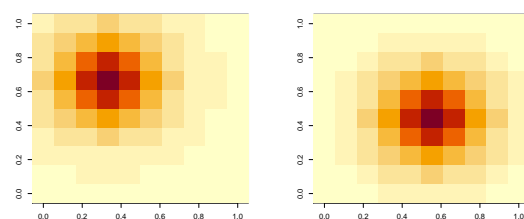
The `generateImage` function provides a straightforward method for rapid simulation of many kinds of images to test classification and clustering models, but suppose we wish to simulate a more complex dataset with spatial correlations. Below we simulate a dataset with two overlapping regions. In each of these regions, the intensity degrades with distance from the center of the region, implying spatial correlation, Figure 24.



**Figure 23:** Generated images from factor and coordinates

```
> x1 <- apply(expand.grid(x=1:10, y=1:10), 1,
+             function(z) 1/(1 + ((4-z[[1]])/2)^2 + ((4-z[[2]])/2)^2))
> dim(x1) <- c(10,10)
> image(x1[,ncol(x1):1])
```

```
> x2 <- apply(expand.grid(x=1:10, y=1:10), 1,
+             function(z) 1/(1 + ((6-z[[1]])/2)^2 + ((6-z[[2]])/2)^2))
> dim(x2) <- c(10,10)
> image(x2[,ncol(x2):1])
```



**(a) Region 1** **(b) Region 2**

**Figure 24:** Ground truth images of a dataset with overlapping regions

We generate the image by using `generateSpectrum` with the calculated mean intensities. We use two peaks for the two regions with nearly overlapping peaks at  $m/z$  500 and  $m/z$  510.

```
> set.seed(1)
> x3 <- mapply(function(z1, z2) generateSpectrum(1, centers=c(500,510), intensities=c(z1, z2), range=c(1,1000)),
+             expand.grid(x=1:10, y=1:10), 2)
> img3 <- MSImageSet(x3, coord=expand.grid(x=1:10, y=1:10), mz=1:1000)
```

Now we can plot the ion images for each of the two peaks in 25.

```
> image(img3, mz=500, col=intensity.colors(100))
```

```
> image(img3, mz=510, col=intensity.colors(100))
```

Finally, we plot the mass spectrum for a pixel from each region in Figure 26

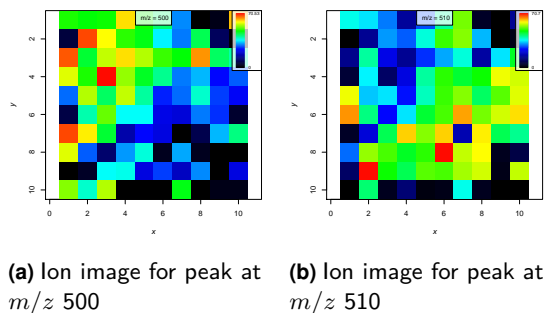


Figure 25: Simulated mass spectral images at the two peaks

```
> plot(img3, coord=list(x=4, y=4), type="l", xlim=c(200, 800))
```

```
> plot(img3, coord=list(x=6, y=6), type="l", xlim=c(200, 800))
```

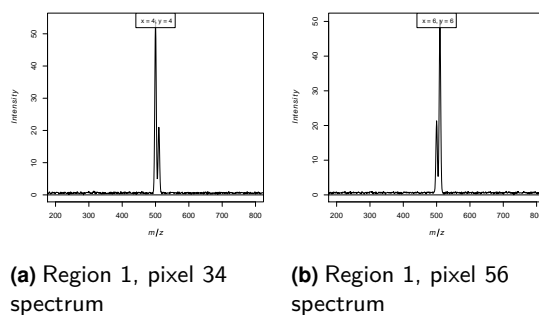


Figure 26: Simulated mass spectra from each of the two regions

By creating spatial correlation patterns and combining them with the `intensities`, `sd`, and `noise` arguments in `generateSpectrum`, it is possible to simulate more complex mass spectrometry imaging datasets.

## 8 Session info

- R version 3.6.0 (2019-04-26), x86\_64-w64-mingw32
- Locale: LC\_COLLATE=C, LC\_CTYPE=English\_United States.1252, LC\_MONETARY=English\_United States.1252, LC\_NUMERIC=C, LC\_TIME=English\_United States.1252
- Running under: Windows Server 2012 R2 x64 (build 9600)
- Matrix products: default
- Base packages: base, datasets, grDevices, graphics, methods, parallel, stats, stats4, utils

## Cardinal: Analytic tools for mass spectrometry imaging

- Other packages: BiocGenerics 0.30.0, BiocParallel 1.18.0, BiocStyle 2.12.0, Cardinal 2.2.2, EBIImage 4.26.0, ProtGenerics 1.16.0, S4Vectors 0.22.0
- Loaded via a namespace (and not attached): Biobase 2.44.0, BiocManager 1.30.4, DBI 1.0.0, MASS 7.3-51.4, Matrix 1.2-17, R6 2.4.0, RCurl 1.95-4.12, Rcpp 1.0.1, abind 1.4-5, assertthat 0.2.1, biglm 0.9-1, bitops 1.0-6, bookdown 0.9, compiler 3.6.0, crayon 1.3.4, digest 0.6.18, dplyr 0.8.0.1, evaluate 0.13, fftwtools 0.9-8, glue 1.3.1, grid 3.6.0, htmltools 0.3.6, htmlwidgets 1.3, irlba 2.3.3, jpeg 0.1-8, knitr 1.22, lattice 0.20-38, locfit 1.5-9.1, magrittr 1.5, matter 1.10.0, mclust 5.4.3, nlme 3.1-139, pillar 1.3.1, pkgconfig 2.0.2, png 0.1-7, purrr 0.3.2, rlang 0.3.4, rmarkdown 1.12, signal 0.7-6, sp 1.3-1, stringi 1.4.3, stringr 1.4.0, tibble 2.1.1, tidyselect 0.2.5, tiff 0.1-5, tools 3.6.0, viridisLite 0.3.0, xfun 0.6, yaml 2.2.0

## References

- [1] L. Gatto and K. S. Lilley. MSnbase—an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics*, 28(2):288–289, 2012.
- [2] S. Gibb and K. Strimmer. MALDIquant: a versatile R package for the analysis of mass spectrometry data. *Bioinformatics*, 28(17):2270–2271, 2012.
- [3] Thorsten Schramm, Alfons Hester, Ivo Klinkert, Jean-Pierre Both, Ron M. A. Heeren, Alain Brunelle, Olivier Laprévote, Nicolas Desbenoit, Marie-France Robbe, Markus Stoeckli, Bernhard Spengler, and Andreas Römpp. imzml – a common data format for the flexible exchange and processing of mass spectrometry imaging data. *Journal of Proteomics*, 75(16):5106 – 5110, 2012. Special Issue: Imaging Mass Spectrometry: A User’s Guide to a New Technique for Biological and Biomedical Research. URL: <http://www.sciencedirect.com/science/article/pii/S1874391912005568>, doi:<http://dx.doi.org/10.1016/j.jprot.2012.07.026>.
- [4] Sören-Oliver Deininger, Dale S. Cornett, Rainer Paape, Michael Becker, Charles Pineau, Sandra Rauser, Axel Walch, and Eryk Wolski. Normalization in MALDI-TOF imaging datasets of proteins: practical considerations. *Analytical and Bioanalytical Chemistry*, 401(1):167–181, 2011. URL: <http://dx.doi.org/10.1007/s00216-011-4929-z>, doi:10.1007/s00216-011-4929-z.
- [5] C. Yang, Z. He, and W. Yu. Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis. *BMC Bioinformatics*, 10, 2009.
- [6] Theodore Alexandrov and Jan Hendrik Kobarg. Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering. *Bioinformatics (Oxford, England)*, 27(13):i230–8, July 2011. URL: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3117346&tool=pmcentrez&rendertype=abstract>, doi:10.1093/bioinformatics/btr246.
- [7] A. L. Dill, L. S. Eberlin, C. Zheng, A. B. Costa, D. R. Ifa, L. Cheng, T. A. Masterson, M. O. Koch, O. Vitek, and R. G. Cooks. Multivariate statistical differentiation of renal cell carcinomas based on lipidomic analysis by ambient ionization imaging mass spectrometry. *Analytical and Bioanalytical Chemistry*, 398:2969, 2010.
- [8] R. Tibshirani, T. Hastie, B. Narasimhan, and G. Chu. Class prediction by nearest shrunken with applications to DNA microarrays. *Statistical Science*, 18:104, 2003.