Import and Export of Spectra Files

Vignette for the R package hyperSpec

Claudia Beleites < chemometrie@beleites.de> CENMAT and DI3, University of Trieste Spectroscopy · Imaging, IPHT Jena e.V.

January 11, 2013

Supported File Formats

hyperSpec supports a number of file formats relevant for different types of spectroscopy. This is naturally only a subset of the file formats produced by different spectroscopic equipment. If you use hyperSpec with data formats not mentioned in this document, please send an email to Claudia Beleites <chemometrie@beleites.de>, so that this document can be updated. The information should include

- The type of spectroscopy
- Spectrometer model, manufacturer, and software
- The "native" file format (including a sample file)
- Description of relevant procedures to convert the file
- R code to import the data together with an example file that can actually be read by R.
- Documentation, particularly the description of the data format

If you need help finding out how to import your data, *hyperSpec* has a mailing list hyperspec-help@lists.r-forge.r-project.org, subscription and archives are available at http://r-forge.r-project.org/mail/?group_id=366.

Reproducing the Examples in this Vignette

The source code of this vignette including the spectra files are available as .zip file at *hyperSpec*'s home page: http://hyperspec.r-forge.r-project.org/fileio.zip
Note that some definitions are in file vignettes.defs.

Contents

1. Introduction 2

2.	Creating a hyperSpec object with new	3
3.	Reading Multiple files into one hyperSpec object	3
4.	ASCII files 4.1. ASCII files with samples in columns 4.2. JCAMP-DX 4.3. ASCII Export	4 5 6
5.	Binary file formats 5.1. Matlab Files	66 66 67 77
6.	Manufacturer-Specific Discussion of File Import 6.1. Manufacturer Specific Import Functions 6.2. Bruker FT-IR Imaging 6.3. Nicolet FT-IR Imaging 6.4. Varian/Agilent FT-IR Imaging 6.5. Kaiser Optical Systems Raman 6.5.1. Kaiser Optical Systems ASCII Files 6.5.2. Kaiser Optical Systems Raman Maps 6.6. Renishaw Raman 6.6.1. Renishaw ASCII data 6.7. Horiba Jobin Yvon 6.8. Witec	99 99 100 100 100 111 111 113 133
7.	Writing your own Import Function 7.1. A new ASCII Import Function: scan.txt.PerkinElmer 7.2. Deriving a More Specific Function: read.ENVI.Nicolet 7.3. Deriving import filters for spc files	14 14 15 17
Α.	File Import Functions by Format	18
В.	File Import Functions by Manufacturer	19
C.	File Import Functions by Spectroscopy	20

1. Introduction

This document describes how spectra can be imported into hyperSpec objects. Some possibilities to export hyperSpec objects as files are mentioned, too.

The most basic funtion to create *hyperSpec* objects is **new** ("hyperSpec") (section 2). It makes a *hyperSpec* object from data already in R's workspace. Thus, once the spectra are imported into R, conversion to *hyperSpec* objects is straightforward.

In addition, hyperSpec comes with predifined import functions for different data formats. This document divides the discussion into dealing with ASCII files (section 4, p. 4) and binary file formats (section 5, p. 6). If data export for the respective format is possible, it is discussed in the same sections. As sometimes the actual data written by the spectrometer software exhibits peculiarities,

hyperSpec offers several specialized import functions. These are in general named after the data format followed by the manufacturer (e. g. read.ENVI.Nicolet).

Overview lists of the directly supported file formats are in the appendix: sorted by file format (appendix A, p. 18), manufacturer (appendix B, p. 19), and by spectroscopy (appendix C, p. 20).

2. Creating a hyperSpec object with new

```
To create a hyperSpec object from data in R's workspace, use:

spc <- new ("hyperSpec", spc, wavelength, data, labels)}

With the arguments:

spc the spectra matrix (may also be given as matrix inside column $spc of data)

wavelength the wavelength axis vector

data the extra data (possibly already including the spectra matrix in column spc)
```

a list with the proper labels. Do not forget the wavelength axis label in \$.wavelength and the spectral intensity axis label in \$spc.

Thus, once your data is in R's workspace, creating a hyperSpec object is easy. I suggest wrapping the code to import your data and the line joining it into a hyperSpec object by your own import function. You are more than welcome to contribute such import code to hyperSpec. Secion 7, (p. 14) discusses examples of custom import functions.

3. Reading Multiple files into one hyperSpec object

Many of the function described below will work on one file, even though derived functions such as read.spc.KaiserMap (see section 6.5.2, p. 11) may take care of measurements consisting of multiple files

Usually, the most convenient way to import multiple files into one *hyperSpec* object is reading all files into a list of *hyperSpec* objects, and then collapseing this list into a single *hyperSpec* object:

```
> files <- Sys.glob ("spc.Kaisermap/*.spc")</pre>
> files <- files [seq (1, length (files), by = 2)] # import low wavenumber region only
> spc <- lapply (files, read.spc)</pre>
> length (spc)
[1] 54
> spc [[1]]
hyperSpec object
   1 spectra
   6 data columns
   1340 data points / spectrum
wavelength: x/"a. u." [numeric] 1 2 ... 1340
data: (1 rows x 6 columns)
   1. z: x/"a. u." [numeric] 1
   2. z.end: x/"a. u." [numeric] 1
   3. fexper: [factor] General
   4. fres: [factor]
   5. fsource: [factor]
   6. spc: Counts [matrix1340] 2782.7 2229.8 ... 932.02
```

```
> spc <- collapse (spc)
> spc

hyperSpec object
   54 spectra
   6 data columns
   1340 data points / spectrum
wavelength: x/"a. u." [numeric] 1 2 ... 1340
data: (54 rows x 6 columns)
   1. z: x/"a. u." [numeric] 1 1 ... 1
   2. z.end: x/"a. u." [numeric] 1 1 ... 1
   3. fexper: [factor] General General ... General
   4. fres: [factor] ...
   5. fsource: [factor] ...
   6. spc: Counts [matrix1340] 2782.7 2678.6 ... 789.49
```

Note that in this particular case, the spectra are more efficiently read by read.spc.KaiserMap (see section 6.5.2, p. 11).

If you regularly import huge maps or images, writing a customized import function is highly encouraged. You may gain speed and memory by using the internal workhorse functions for the file import. In that case, please contact the package maintainer (Claudia Beleites <chemometrie@beleites.de>) for advise (contributions to hyperSpec are welcome and all authors are listed appropriately in the function help page's author section).

4. ASCII files

Currently, hyperSpec provides two functions for general ASCII data import:

```
read.txt.long imports long format ASCII files, i.e. one intensity value per row
```

read.txt.wide imports wide format ASCII files, i.e. one spectrum per row

The import functions immediately return a hyperSpec object.

Internally, they use read.table, a very powerful ASCII import function. R supplies another ASCII import function, scan. scan imports numeric data matrices and is faster than read.table, but cannot import column names. If your data does not contain a header or it is not important and can safely be skipped, you may want to import your data using scan.

Note that R allows to use a variety of compressed file formats directly as ASCII files (for example, see section 6.6.1 on p. 11). Also, both read.txt.long and read.txt.wide accept connections instead of file names.

4.1. ASCII files with samples in columns

Richard Pena asked about importing another ASCII file type:

Triazine5_31.txt file corresponds to X ray powder diffraction data (Bruker AXS). The native files data ".raw" are read with EVA software then they are converted into .uxd file with the File Exchange software (Bruker AXS). The .uxd file are opened with Excel software and saved as .txt file, csv file (ChemoSpec) or xls.

The first and following columns corresponds to the angle diffraction and the intensity values of samples respectively.

This file thus differs from the ASCII formats discussed above in that the samples are actually in columns whereas *hyperSpec* expects them to be in rows. The header line gives the name of the sample. Import is straightforward, just the spectra matrix needs to be transposed to make a *hyperSpec* object:

```
> file <- read.table ("txt.t/Triazine 5_31.txt", header = TRUE, dec = ",", sep = "\t")
> triazine <- new ("hyperSpec", wavelength = file [,1], spc = t (file [, -1]),
                    data = data.frame (sample = colnames (file [, -1])),
                    labels = list (.wavelength = expression (2 * theta / degree),
                                   spc = "I / a.u."))
> triazine
hyperSpec object
   25 spectra
   2 data columns
   1759 data points / spectrum
wavelength: 2 * theta/degree [numeric] 5.0025 5.0173 ... 31.004
data: (25 rows x 2 columns)
   1. sample: [factor] DIV1208200 DIV1208300 ... VCA0106703
   2. spc: I / a.u. [matrix1759] 92 96 ... 163
> plot (triazine [1])
    0009
   4000
   2000
              10 12 14 16 18 20 22 24 26 28 30 32
            8
         6
                          2θ/°
```

4.2. JCAMP-DX

1 data columns

1216 data points / spectrum

data: (1 rows x 1 columns)

wavelength: tilde(nu)/cm^-1 [numeric] 3030 3028 ... 600

1. spc: A [matrix1216] -3.1244e-05 0.0000e+00 ... 0

Limited import of JCAMP-DX files v. 4.24 [1] is available in function read.jdx. These files can contain multiple spectra, supported data formats are tabular XY..XY and X++(Y..Y).

> read.jdx ("jcamp-dx/shimadzu.jdx", encoding = "latin1", keys.hdr2data=TRUE)

```
hyperSpec object
   1 spectra
   9 data columns
   295 data points / spectrum
wavelength: [numeric] 60 61 ... 550.3
data: (1 rows x 9 columns)
   1. spc: [matrix295] 111 276 ... 99
   2. title: [numeric] 44
   3. mw: [numeric] 382
   4. molform: [factor] C25 H50 O2
   5. casregistryno: [factor] 2442 - 49 - 1
   6. datatype: [factor] Mass Spectrum
   7. sampledescription: [factor] ...\11-12-12\pAS+OS Lauf1.qgd"\n20000 Da/s, ET 30ms, m/z 60-550\nPeak-Apex-Spektrum, H
   8. casname: [factor] Methyltetracosanoat (LRI C24)
   9. $retentionindex: [numeric] 2400
> read.jdx ("jcamp-dx/virgilio.jdx")
hyperSpec object
   1 spectra
```

Note

read.jdx.Shimadzu is deprecated.

4.3. ASCII Export

ASCII export can be done in wide and long format using write.txt.long and write.txt.wide. If you need a specific header or footer, use R's functions for writing files: write.table, write, cat and so on offer fine-grained control of writing ASCII files.

5. Binary file formats

5.1. Matlab Files

Matlab files can be read and written using the package R.matlab[2], which is available at CRAN and can be installed by install.packages ("R.matlab").

```
spc.mat <- readMat ("spectra.mat")</pre>
```

If the .mat file was saved with compression, the additional package Rcompression is needed. It can be installed from omegahat:

```
install.packages("Rcompression", repos = "http://www.omegahat.org/R")
```

See the documentation of R.matlab for more details and possibly needed further packages.

readMat imports the .mat file's contents as a list. The variables in the .mat file are properly named elements of the list. The *hyperSpec* object can be created using new, see 2 (p. 3).

Again, you probably want to wrap the import of your matlab files into a function.

5.1.1. Matlab Export

R.matlab's function writeMat can be used to write R objects into .mat files. To save an hyperSpec object x for use in Matlab, you most likely want to save:

- the wavelength axis as obtained by wl (x),
- the spectra matrix as obtained by x [[]], and
- possibly also the extra data as obtained by x\$...
- as well as the axis labels labels (x).
- Alternatively, x\$. yields the extra data together with the spectra matrix.

However, it may be convenient to transform the saved data according to how it is needed in Matlab. The functions as.long.df and as.wide.df may prove useful for reshaping the data.

5.2. ENVI Files

ENVI files are binary data accompanied by an ASCII header file. *hyperSpec*'s function read.ENVI can be used to import them. Usually, the header file name is the same as the binary data file name with the suffix replaced by .hdr. Otherwise, the header file name can be given via parameter *headerfile*.

As we experienced missing header files (Bruker's Opus software frequently produced header files without any content), the data that would usually be read from the header file can also be handed to read. ENVI as a list in parameter *header*. Arguments given in *header* replace corresponding entries of the header file. The help page gives details on what elements the list should contain, see also the discussion of ENVI files written by Bruker's OPUS software (section 6.2, p. 9).

Here is how to use read. ENVI:

```
> spc <- read.ENVI ("ENVI/example2.img")
.read.ENVI.header: Guessing header file name (ENVI/example2.hdr)
> spc
hyperSpec object
    420 spectra
    3 data columns
    1738 data points / spectrum
wavelength: [numeric] 649.90 651.83 ... 3999.7
data: (420 rows x 3 columns)
    1. x: [integer] 0 0 ... 13
    2. y: [integer] 0 1 ... 29
    3. spc: [matrix1738] 0 0 ... 0
```

Please see also the manufacturer specific notes in section 6.1, p. 9.

5.2.1. ENVI Export

Use package caTools or rgdal with GDAL for writing ENVI files.

5.3. spc Files

Thermo Galactic's .spc file format[3] can be imported by read.spc.

A variety of sub-formats exists. *hyperSpec*'s importread.spc function does *not* support the old file format that was used before 1996. In addition, no test data with *w planes* was available — thus the import of such files could not be tested. If you come across such files, please contact the package maintainer (Claudia Beleites <chemometrie@beleites.de>).

Here are some tests using Thermo Galactic's example files:

```
spc/BARBITUATES.SPC: list of 286 spectra, 4 - 101 data pts / spc
spc/barbsvd.spc: list of 286 spectra, 4 - 101 data pts / spc
\ensuremath{\mathtt{spc}}/\ensuremath{\mathtt{BENZENE}}.\ensuremath{\mathtt{SPC}}\colon \ensuremath{\mathtt{1}} \ensuremath{\mathtt{spectrum}}(a) \,, \ensuremath{\mathtt{1842}} \ensuremath{\mathtt{data}} \ensuremath{\mathtt{pts}} \ensuremath{/} \ensuremath{\mathtt{spc}}.
spc/DRUG SAMPLE_PEAKS.SPC: list of 6 spectra, 80 - 253 data pts / spc
spc/DRUG SAMPLE.SPC: list of 400 spectra, 2 - 254 data pts / spc
spc/FID.SPC: 1 spectrum(a), 8192 data pts / spc.
spc/HCL.SPC: 1 spectrum(a), 8361 data pts / spc.
spc/HOLMIUM.SPC: 1 spectrum(a), 901 data pts / spc.
spc/IG_BKGND.SPC: 1 spectrum(a), 4096 data pts / spc.
spc/IG_MULTI.SPC: 10 spectrum(a), 4096 data pts / spc.
spc/IG_SAMP.SPC: 1 spectrum(a), 4645 data pts / spc.
spc/KKSAM.SPC: 1 spectrum(a), 751 data pts / spc.
spc/POLYR.SPC: 1 spectrum(a), 1844 data pts / spc.
spc/POLYS.SPC: 1 spectrum(a), 1844 data pts / spc.
spc/SINGLE POLYMER FILM.SPC: 1 spectrum(a), 1844 data pts / spc.
spc/SPECTRUM WITH BAD BASELINE.SPC: 1 spectrum(a), 1400 data pts / spc.
spc/TOLUENE.SPC: 1 spectrum(a), 801 data pts / spc.
spc/TriVista-linear.spc: 1 spectrum(a), 1149 data pts / spc.
spc/TriVista-normal.spc: 1 spectrum(a), 1024 data pts / spc.
spc/TUMIX.SPC: 1 spectrum(a), 1775 data pts / spc.
spc/TWO POLYMER FILMS.SPC: 1 spectrum(a), 1844 data pts / spc.
spc/XYTRACE.SPC: 1 spectrum(a), 3469 data pts / spc.
```

The header and subheader blocks of spc files store additional information of pre-defined types (see the file format specification[3]). Further information can be stored in the so-called log block at the end of the file, and should be in a key-value format (although even the official example files do not always). This information is often useful (Kaiser's Hologram software e.g. stores the stage position in the log block).

read.spc has four arguments that allow fine-grained control of storing such information in the hyperSpec object:

keys.hdr2data parameters from the spc file and subfile headers that should become extra data columns

keys.hdr2log parameters from the spc file and subfile headers that should be stored as list entries in the long.description of the log entry

keys.log2data parameters from the spc file log block that should become extra data columns

The value of these arguments can either be logical (amounting to either use all or none of the information in the file) or a character vector giving the names of the parameters that should be used. Note that the header file field names are always lowercase, while the log entries are treated case sensitive.

.spc files may contain multiple spectra that do *not* share a common wavelength axis. In this case, read.spc returns a list of *hyperSpec* objects with one spectrum each. collapse may be used to combine this list into one *hyperSpec* object:

```
> barbiturates <- read.spc ("spc/BARBITUATES.SPC")
> save (barbiturates, file = "barbiturates.rda")
> class (barbiturates)
[1] "list"
> length (barbiturates)
[1] 286
> barbiturates <- do.call (collapse, barbiturates)
> barbiturates <- orderwl (barbiturates)
> barbiturates
```

```
hyperSpec object
   286 spectra
   6 data columns
   375 data points / spectrum
wavelength: frac(m, z)/frac(u, e) [numeric] 25.95 26.05 ... 244.05
data: (286 rows x 6 columns)
   1. z: t/min [numeric] 4.0272 4.0341 ... 5.9978
   2. z.end: t/min [numeric] 4.0272 4.0341 ... 5.9978
   3. fexper: [factor] General General ... General
   4. fres: [factor]
   5. fsource: [factor] MS_5971 MS_5971 ... MS_5971
   6. spc: I/"a. u." [matrix375] NA NA ... NA + NA
> barbiturates [[1:10, , 25 ~ 30]]
      25.95 26.05 26.15 26.95 27.05 27.15 28.05 28.15 29.05 29.15 29.95
 [1,]
                                              NA 11511 6146
         NΑ
               NΑ
                      NΑ
                            NΑ
                                 562
                                        NΑ
                                                                        NΑ
 [2,]
         NA
               NA
                      NA
                            NA
                                  NA
                                        618 10151
                                                     NA
                                                         5040
 [3,]
         NA
               NA
                      NA
                            NA
                                 638
                                        NA
                                              NA 10722
                                                         5253
                                                                  NA
                                                                        NA
 [4,]
         NA
               NA
                      NΑ
                            NA
                                  NA
                                        NA 10548
                                                    NΑ
                                                         5865
                                                                 NA
                                                                        NΑ
 [5,]
         NA
               NA
                      NA
                            NA
                                              NA 10519
                                                         4664
 [6,]
                      NA
                           796
                                        NA 10519
                                                         5110
         NA
               NA
                                  NA
                                                     NA
                                                                  NA
                                                                        NΑ
 [7,]
         NA
               NA
                      NA
                            NA
                                  NA
                                        NA 10096
                                                     NA
                                                         4769
                                                                  NA
                                                                       907
 [8,]
         NA
               NA
                      NA
                            NA
                                  NA
                                        NA
                                              NA 10929
                                                         5400
                                                                  NA
                                                                        NA
                                        NA 10235
         NΑ
               NΑ
                      NΑ
                            NΑ
                                  NΑ
                                                     NΑ
                                                         4930
                                                                  NΑ
                                                                        NΑ
 [9,]
[10,]
         NA
               NA
                      NA
                            NA
                                  NA
                                        NA
                                               NA 10663
                                                         4690
                                                                  NA
                                                                       799
```

Deriving manufacturer specific import filters. Please note that future changes inside the read.spc function are likely to occur. However, if you just post-process the *hyperSpec* object returned by read.spc, you should be fine.

6. Manufacturer-Specific Discussion of File Import

6.1. Manufacturer Specific Import Functions

Many spectrometer manufacturers provide a function to export their spectra into ASCII files. The functions discussed above are written in a very general way, and are highly customizable. I recommend wrapping these calls with the appropriate settings for your spectra format in an import function. Please consider contributing such import filters to *hyperSpec*: send me the documented code (for details see the box at the beginning of this document). If you are able to import data of any format not mentioned in this document (even without the need of new converters), please let me know (details again in the box at the beginning of this document).

6.2. Bruker FT-IR Imaging

We use read.ENVI to import IR-Images collected with a Bruker Hyperion spectrometer with OPUS software. As mentioned above, the header files are frequently empty. We found the necessary information to be:

```
> header <- list (samples = 64 * no.images.in.row,
+ lines = 64 * no.images.in.column,
+ bands = no.data.points.per.spectrum,
+ 'data type' = 4,
+ interleave = "bip")</pre>
```

No spatial information is given in the ENVI header (if written). The lateral coordinates can be setup by specifying origin and pixel size for x and y directions. For details please see the help page.

The proprietary file format of the Opus software is not yet supported.

6.3. Nicolet FT-IR Imaging

Also Nicolet saves imaging data in ENVI files. These files use some non-standard keywords in the header file that should allow to reconstruct the lateral coordinates as well as the wavelength axes and units for wavelength and intensity axis. *hyperSpec* has a specialized function read.ENVI.Nicolet that uses these header entries.

It seems that the position of the first spectrum is recorded in µm, while the pixel size is in mm. Thus a flag *nicolet.correction* is provided that divides the pixel size by 1000. Alternatively, the correct offset and pixel size values may be given as function arguments.

```
> spc <- read.ENVI.Nicolet ("ENVI/example2.img", nicolet.correction = TRUE)
.read.ENVI.header: Guessing header file name (ENVI/example2.hdr)
> spc ## dummy sample with all intensities zero
hyperSpec object
    420 spectra
    3 data columns
    1738 data points / spectrum
wavelength: [numeric] 649.90 651.83 ... 3999.7
data: (420 rows x 3 columns)
    1. x: [integer] 0 0 ... 13
    2. y: [integer] 0 1 ... 29
    3. spc: [matrix1738] 0 0 ... 0
```

6.4. Varian/Agilent FT-IR Imaging

Agilent (Varian) uses a variant of ENVI (with binary header). A specialized form of read.ENVI will be coming soon.

6.5. Kaiser Optical Systems Raman

Spectra obtained using Kaiser's Hologram software can be saved either in their own .hol format and imported into Matlab (from where the data may be written to a .mat file readable by R.matlab's readMat. Hologram can also write ASCII files and .spc files. We found working with .spc files the best option.

The spectra are usually interpolated by Hologram to an evenly spaced wavelength (or $\Delta \tilde{\nu}$) axis unless the spectra are saved in a by-pixel manner. In this case, the full spectra consist of two files with consecutive file names: one for the low and one for the high wavenumber region. See the example for .spc import.

6.5.1. Kaiser Optical Systems ASCII Files

The ASCII files are long format that can be imported by read.txt.long (see section 4, p. 4).

We experienced two different problems with these files:

- 1. If the instrument computer's locale is set so that also the decimal separator is a comma, commas are used both as decimal and as column separator.
- 2. Values with a decimal fraction of 0 are written with decimal separator but no further digits (e.g. 2,). This may be a problem for certain conversion functions (read.table works fine, though).

Still the files may be imported, though care must be taken:

```
> ## 1. import as character
> tmp <- scan ("txt.Kaiser/test-lo-4.txt", what = rep ("character",4), sep = ",")
> tmp <- matrix (tmp, nrow = 4)
> ## 2. concatenate every two columns by a dot
> wl <- apply (tmp [1:2, ], 2, paste, collapse = '.')
> spc <- apply (tmp [3:4, ], 2, paste, collapse = '.')
> ## 3. convert to numeric and create hyperSpec object
> spc <- new ("hyperSpec", spc = as.numeric (spc), wavelength = as.numeric (wl))</pre>
```

6.5.2. Kaiser Optical Systems Raman Maps

hyperSpec provides the function read.spc.KaiserMap to easily import spatial collections of .spc files written by Kaiser's Hologram software. The filenames of all .spc files to be read into one hyperSpec object can be provided either as a character vector or as a wildcard expression (e.g. "path/to/files/*.spc").

The data for the following example was saved with wavelength axis being camera pixels rather than Raman shift. Thus two files for each spectrum were saved by Hologram. Thus, a file name pattern is difficult to give and a vector of file names is used instead:

```
> files <- Sys.glob ("spc.Kaisermap/*.spc")
> spc.low <- read.spc.KaiserMap (files [seq (1, length (files), by = 2)])
> spc.high <- read.spc.KaiserMap (files [seq (2, length (files), by = 2)])
> wl (spc.high) <- wl (spc.high) + 1340
> spc

hyperSpec object
    1 spectra
    1 data columns
    2110 data points / spectrum
wavelength: [numeric] 121.5 122.4 ... 2019.6
data: (1 rows x 1 columns)
    1. spc: [matrix2110] 1202.51 770.35 ... 141.01
```

6.6. Renishaw Raman

Renishaw's Wire software comes with an file format converter. This program can produce a long ASCII format, .spc, or .jdx files.

We experienced that the conversion to .spc is *not* fully reliable: maps were saved as depth profile, loosing all spatial information. In addition, an evenly spaced wavelength axis was produced, although this was de-selected in the converter. We therefore recommend using the ASCII format. Otherwise the import using read.spc worked.

6.6.1. Renishaw ASCII data

An optimized import function for the ASCII files is available: scan.txt.Renishaw. The file may be compressed via gzip, bzip2, xz or lzma. zip compressed files are read via scan.zip.Renishaw. The ASCII files can easily become very large, particularly with linefocus- or streamline imaging. scan.txt.Renishaw provides two mechanisms to avoid running out of memory during data import. The file may be imported in chunks of a given number of lines (see the last example). scan.txt.Renishaw can calculate the correct number of wavelengths (i.e. data points per spectrum) if the system command wc is available on your computer.

In addition, the processing of the long ASCII format into the spectra matrix is done by reshaping the vector of intensities into a matrix. This process does not allow any missing values in the data. Therefore it is not possible to import multi-spectra files with individually "zapped" spectra using scan.txt.Renishaw.

The second argument to scan.txt.Renishaw decides what type of experiment is imported. Supported types are:

```
"xyspc"
                  maps, images, multiple spectra with x and y coordinates (default)
"spc"
                  single spectrum
"depth", "zspc" depth series
"ts"
                  time series
Instead of a file name, scan.txt.Renishaw accepts also a connection.
> paracetamol <- scan.txt.Renishaw ("txt.Renishaw/paracetamol.txt", "spc")
> paracetamol
hyperSpec object
   1 spectra
   1 data columns
   4064 data points / spectrum
wavelength: Delta * tilde(nu)/cm^-1 [numeric] 96.787 98.143 ... 3200.1
data: (1 rows x 1 columns)
   1. spc: I / a.u. [matrix4064] 2056.5 2224.8 ... 299.23
> save (paracetamol, file = "paracetamol.rda")
> scan.txt.Renishaw ("txt.Renishaw/laser.txt.gz", "ts")
hyperSpec object
   84 spectra
   2 data columns
   140 data points / spectrum
wavelength: Delta * tilde(nu)/cm^-1 [numeric] -199.08 -196.90 ... 99.934
data: (84 rows x 2 columns)
   1. t: t / s [numeric] 0 2 ... 5722
   2. spc: I / a.u. [matrix140] 29.801 32.093 ... 81.3
Very large files can be read in chunks to save memory:
> scan.txt.Renishaw ("txt.Renishaw/chondro.txt", nlines = 1e5, nspc = 875)
hyperSpec object
   875 spectra
   3 data columns
   1272 data points / spectrum
wavelength: Delta * tilde(nu)/cm^-1 [numeric] 601.62 602.66 ... 1802.2
data: (875 rows x 3 columns)
   1. y: y/(mu * m) [numeric] -4.77 -4.77 ... 19.23
   2. x: x/(mu * m) [numeric] -11.55 -10.55 ... 22.45
   3. spc: I / a.u. [matrix1272] 501.72 518.53 ... 151.92 + NA
R accepts a variety of compressed file formats for ASCII files:
> scan.txt.Renishaw ("txt.Renishaw/chondro.gz")
> scan.txt.Renishaw ("txt.Renishaw/chondro.xz")
> scan.txt.Renishaw ("txt.Renishaw/chondro.lzma")
> scan.txt.Renishaw ("txt.Renishaw/chondro.gz")
> scan.txt.Renishaw ("txt.Renishaw/chondro.bz2")
For .zip packed files, however, a connection must be used:
> scan.txt.Renishaw (unzip ("txt.Renishaw/chondro.zip"))
```

6.7. Horiba Jobin Yvon

Horiba Jobin Yvon's Labspec software saves spectra in a wide ASCII format which is read by read.txt.wide, e.g.:

```
> spc <- read.txt.wide ("txt.HoribaJobinYvon/test.txt",
                          cols = list (
                            x = expression ('/' (x, mu * m)),
                            y = expression ('/' (y, mu * m)),
                            spc = "I / a.u.",
                            .wavelength = expression (Delta * tilde (nu) / \text{cm}^-1)),
                          sep = '\t', header = TRUE)
> spc
hyperSpec object
   398 spectra
   3 data columns
   616 data points / spectrum
wavelength: Delta * tilde(nu)/cm^-1 [numeric] 50.00 55.61 ... 3500
data: (398 rows x 3 columns)
   1. x: x/(mu * m) [numeric] -60.612 -60.612 ... 43.234
   2. y: y/(mu * m) [numeric] -40.855 -35.600 ... 48.489
   3. spc: I / a.u. [matrix616] 0 0 ... 898.61
```

Note that this data set contains lots of spectra with zero intensity: Labspec saves a complete rectangular grid even if only part of it was measured. These spectra can be removed by

```
> ## a floating point precision tolerant version of ==
> `%~=%` <- function (x, y, tolerance = .Machine$double.eps^0.5)
+ abs (x - y) < tolerance
> spc <- spc [! apply (spc [[]] %~=% 0, 1, all)]
> spc

hyperSpec object
    141 spectra
    3 data columns
    616 data points / spectrum
wavelength: Delta * tilde(nu)/cm^-1 [numeric] 50.00 55.61 ... 3500
data: (141 rows x 3 columns)
    1. x: x/(mu * m) [numeric] -55.54 -55.54 ... 43.234
    2. y: y/(mu * m) [numeric] -35.600 -30.344 ... 48.489
    3. spc: I / a.u. [matrix616] 3126.3 2455.8 ... 898.61
```

6.8. Witec

In the Witec project software, export the spectra as ASCII X and Y files. These can be read by scan.txt.Witec:

```
> scan.txt.Witec ("txt.Witec/WitecExample-x.dat", points.per.line = 10, lines.per.image = 10)
hyperSpec object
  100 spectra
  3 data columns
  1024 data points / spectrum
wavelength: [numeric] 106.39 110.76 ... 3377.2
data: (100 rows x 3 columns)
  1. spc: [matrix1024] 1494 1486 ... 952
  2. x: [integer] 1 2 ... 10
  3. y: [integer] -1 -1 ... -10
```

7. Writing your own Import Function

This section gives examples how to write import functions. The first example implements an import filter for an ASCII file format basically from scratch. The second example shows how to implement more details for an already existing import filter.

7.1. A new ASCII Import Function: scan.txt.PerkinElmer

The raw spectra of the flu data set (see also the respective vignette) are in PerkinElmer's ASCII file format, one spectrum per file.

We need a function that automatically reads all files specified by a pattern, such as *.txt. In order to gain speed, the spectra matrix should be preallocated after the first file is read.

A short examination of the files (flu*.txt in directory txt.PerkinElmer) reveals that the actual spectrum starts at line 55, after a line containing #DATA. For now, no other information of the files is to be extracted. It is thus easier to skip the first 54 lines than searching for the line after #DATA.

A fully featured import function should support:

- Reading multiple files by giving a pattern
- hand further arguments to scan. This comes handy in case the function is used later to import other data types.
- Also skipping 54 lines would be a weird default, so we rather require it to be given explicitly.
- The same applies for the axis labels: they should default to reasonable settings for fluorescence spectra, but it should be possible to change them if needed.
- The usual log entry arguments should be supplied.
- A sanity check should be implemented: stop with an error if a file does not have the same wavelength axis as the others.
- Finally, if no file can be found, an empty *hyperSpec* object is a reasonable result: There is no need to stop with an error, but it is polite to issue an additional warning.

```
_ scan.txt.PerkinElmer.R
scan.txt.PerkinElmer <- function (files = "*.txt", ..., label = list (),</pre>
                                   short = "scan.txt.PerkinElmer", user = NULL, date = NULL) {
  ## set some defaults
 long <- list (files = files, ..., label = label)</pre>
  label <- modifyList (list (.wavelength = expression (lambda / nm),
                               spc = expression (I[f1] / "a.u.")),
                        label)
  ## find the files
  files <- Sys.glob (files)
  if (length (files) == 0){
    warning ("No files found.")
    return (new ("hyperSpec"))
  ## read the first file
 buffer <- matrix (scan (files [1], ...), ncol = 2, byrow = TRUE)</pre>
  ## first column gives the wavelength vector
  wavelength <- buffer [, 1]
```

```
## preallocate the spectra matrix:
  ## one row per file x as many columns as the first file has
  spc <- matrix (ncol = nrow (buffer), nrow = length (files))</pre>
  ## the first file's data goes into the first row
  spc [1, ] <- buffer [, 2]</pre>
  ## now read the remaining files
  for (f in seq (along = files)[-1]) {
    buffer <- matrix (scan (files [f], ...), ncol = 2, byrow = TRUE)
    ## check whether they have the same wavelength axis
    if (! all.equal (buffer [, 1], wavelength))
      stop (paste(files [f], "has different wavelength axis."))
    spc [f, ] <- buffer[, 2]</pre>
  ## make the hyperSpec object
  new ("hyperSpec", wavelength = wavelength, spc = spc,
       data = data.frame (file = files), label = label,
       log = list (short = short, long = long, user = user, date = date))
}
```

Note how the labels are set. The label with the special name .wavelength corresponds to the wavelength axis, all data columns should have a label with the same name. The spectra are always in a data column called spc.

Thus.

imports the spectra.

```
> source ("scan.txt.PerkinElmer.R")
> scan.txt.PerkinElmer ("txt.PerkinElmer/flu?.txt", skip = 54)
hyperSpec object
   6 spectra
   2 data columns
   181 data points / spectrum
wavelength: lambda/nm [numeric] 405.0 405.5 ... 495
data: (6 rows x 2 columns)
   1. file: [factor] txt.PerkinElmer/flu1.txt txt.PerkinElmer/flu2.txt ... txt.PerkinElmer/flu6.txt
   2. spc: I[f1]/"a.u." [matrix181] 27.150 66.801 ... 294.65
```

This function is not exported by *hyperSpec*. While it is already useful for importing files, it is not yet general enough to work immediately with new data: the file header is completely ignored. Thus information like the excitation wavelength is lost.

7.2. Deriving a More Specific Function: read.ENVI.Nicolet

The function read. ENVI. Nicolet is a good example for a more specific import filter derived from a general filter for the respective file type. Nicolet FT-IR Imaging software saves some non-standard keywords in the header file of the ENVI data. These information can be used to reconstruct the x and y axes of the images. The units of the spectra are saved as well.

 ${\tt read.ENVI.Nicolet}$ thus first adjusts the parameters for ${\tt read.ENVI.Then}$ read. ENVI does the main work of importing the file. The resulting ${\tt hyperSpec}$ object is post-processed according to the special header entries.

For using the function, see section 6.3 (p. 10).

```
{\tt read.ENVI.Nicolet} {\tt <-} {\tt function} {\tt (..., \# goes to read.ENVI}
                # file headerfile, header
                x = NA, y = NA, # NA means: use the specifications from the header file if possible
                log = list (),
                keys.hdr2log = TRUE,
                nicolet.correction = FALSE) {
  ## set some defaults
  log <- modifyList (list (short = "read.ENVI.Nicolet",</pre>
                            long = list (call = match.call ())),
                     log)
  ## the additional keywords to interprete must be read
  if (! isTRUE (keys.hdr2log))
    keys.hdr2log <- unique (c ("description", "z plot titles", "pixel size", keys.hdr2log))
  ## most work is done by read.ENVI
  spc <- read.ENVI (..., keys.hdr2log = keys.hdr2log,</pre>
                    x = if (is.na(x)) 0 : 1 else x,
                     y = if (is.na (y)) 0 : 1 else y,
                    log = log)
  ## get the header for post-processing
  header <-spc@log$long.description [[1]]$header
### From here on processing the additional keywords in Nicolet's ENVI header **********************
  ## z plot titles ---
  ## default labels
  label <- list (x = expression ('/ (x, micro * m)),
                 y = expression ('/' (y, micro * m)),
                 spc = 'I / a.u.',
                  .wavelength = expression (tilde (nu) / cm^-1))
  ## get labels from header information
  if (!is.null (header$'z plot titles')){
    pattern <- "^[[:blank:]]*([[:print:]^,]+)[[:blank:]]*,.*$"</pre>
    tmp <- sub (pattern, "\\1", header$'z plot titles')</pre>
    if (grepl ("Wavenumbers (cm-1)", tmp, ignore.case = TRUE))
     label$.wavelength <- expression (tilde (nu) / cm^(-1))</pre>
    else
      label$.wavelength <- tmp
    pattern <- "^[[:blank:]]*[[:print:]^,]+,[[:blank:]]*([[:print:]^,]+).*$"</pre>
    tmp <- sub (pattern, "\\1", header$'z plot titles')</pre>
    if (grepl ("Unknown", tmp, ignore.case = TRUE))
      label$spc <- "I / a.u."
    else
      label$spc <- tmp
  ## modify the labels accordingly
  spc@label <- modifyList (label, spc@label)</pre>
  ## set up spatial coordinates -----
  ## look for x and y in the header only if x and y are NULL
  ## they are in `description` and `pixel size`
  ## set up regular expressions to extract the values
   \texttt{p.description} \leftarrow \texttt{paste} \ (\texttt{"`Spectrum position [[:digit:]]+ of [[:digit:]]+ positions,", } 
                           "X = ([[:digit:].-]+), Y = ([[:digit:].-]+)$")
  p.pixel.size <- "^[[:blank:]]*([[:digit:].-]+),[[:blank:]]*([[:digit:].-]+).*$"
  if (is.na (x) && is.na (y) &&
```

```
! is.null (header$description) && grepl (p.description, header$description) &&
      ! is.null (header$'pixel size') && grepl (p.pixel.size, header$'pixel size')) {
    x [1] <- as.numeric (sub (p.description, "\1", header$description))
    y [1] <- as.numeric (sub (p.description, "\\2", header$description))
    x [2] <- as.numeric (sub (p.pixel.size, "\1", header\'pixel size'))
    y [2] <- as.numeric (sub (p.pixel.size, "\\2", header$'pixel size'))
    ## it seems that the step size is given in mm while the offset is in micron
    if (nicolet.correction) {
      x [2] <- x [2] * 1000
      y [2] <- y [2] * 1000
    ## now calculate and set the x and y coordinates
    x \leftarrow x [2] * spc$x + x [1]
    if (! any (is.na (x)))
      spc@data$x <- x
    y \leftarrow y [2] * spc$y + y [1]
    if (! any (is.na (y)))
      spc@data$y <- y
  }
  spc
}
```

7.3. Deriving import filters for spc files

Please note that future changes inside the read.spc function are likely to occur. However, if you just post-process the *hyperSpec* object returned by read.spc, you should be fine.

References

- [1] Robert S. McDonald and Jr. Paul A. Wilks. Jcamp-dx: A standard form for the exchange of infrared spectra in computer readable form. *Applied Spectroscopy*, 42(1):151–162, 1988.
- [2] Henrik Bengtsson and Jason Riedy. R.matlab: Read and write of MAT files together with R-to-Matlab connectivity, 2012. URL http://CRAN.R-project.org/package=R.matlab. R package version 1.6.1.
- [3] Universal Data Format Specification. Galactic Industries Corp., 1997. URL https://ftirsearch.com/features/converters/gspc_udf.zip.

A. File Import Functions by Format

Format	Manufacturer	Function	section	Notes
ASCII file for	rmats			
ASCII long		read.txt.long	4, p. 4	
ASCII long	Renishaw (Raman)	scan.txt.Renishaw	6.6.1, p. 11	
ASCII long	Kaiser (Raman)	read.txt.long	6.5.1, p. 10	Not recommended, see discussion
ASCII long	Perkin Elmer (Fluorescence)	read.txt.PerkinElmer	7.1, p. 14	Reads multiple files, needs to be sourced.
ASCII wide		read.txt.wide	4, p. 4	
ASCII wide	Horiba Jobin Yvon	read.txt.wide	6.7, p. 13	
JCAMP-DX		-	4.2, p. 5	see read.jdx
JCAMP-DX	Renishaw (Raman)	-	4.2, p. 5	not available
JCAMP-DX	Shimadzu (GC,GC-MS)	jcamp-dx	4.2, p. 5	import for subset of the JCAMP-DX standard
JCAMP-DX	PerkinElmer (Infrared)	jcamp-dx	4.2, p. 5	import for subset of the JCAMP-DX standard
binary file for	rmats			
ENVI		read.ENVI	5.2, p. 6	
ENVI	Bruker (Infrared Imaging)	read.ENVI	6.2, p. 9	
ENVI	Nicolet (Infrared Imaging)	read.ENVI.Nicolet	6.3, p. 10	
hol	Kaiser (Raman)	-	6.5, p. 10	via Matlab
Matlab	Matlab	R.matlab::readMat	5.1, p. 6	
Opus	Bruker (Infrared Imaging)	-	6.2, p. 9	
spc		read.spc	5.3, p. 7	
spc	Kaiser (Raman Map)	read.spc.KaiserMap	6.5.2, p. 11	Reads multiple files
spc	Kaiser (Raman)	read.spc	5.3, p. 7	Reads multiple files
spc	Renishaw (Raman)	read.spc	6.6.1, p. 11	Not recommended, see discussion of ASCII files.

B. File Import Functions by Manufacturer

Manufacturer	Format	Function	section	Notes
Manufacturers				
Bruker (Infrared Imaging)	ENVI	read.ENVI	6.2, p. 9	
Bruker (Infrared Imaging)	Opus	-	6.2, p. 9	
Horiba Jobin Yvon	ASCII wide	read.txt.wide	6.7, p. 13	
Kaiser (Raman)	ASCII long	read.txt.long	6.5.1, p. 10	Not recommended, see discussion
Kaiser (Raman)	hol	-	6.5, p. 10	via Matlab
Kaiser (Raman Map)	spc	read.spc.KaiserMap	6.5.2, p. 11	Reads multiple files
Kaiser (Raman)	spc	read.spc	5.3, p. 7	Reads multiple files
Matlab	Matlab	R.matlab::readMat	5.1, p. 6	
Nicolet (Infrared Imaging)	ENVI	read.ENVI.Nicolet	6.3, p. 10	
PerkinElmer (Infrared)	JCAMP-DX	jcamp-dx	4.2, p. 5	import for subset of the JCAMP-DX standard
Perkin Elmer (Fluorescence)	ASCII long	read.txt.PerkinElmer	7.1, p. 14	Reads multiple files, needs to be sourced.
Renishaw (Raman)	ASCII long	scan.txt.Renishaw	6.6.1, p. 11	
Renishaw (Raman)	JCAMP-DX	-	4.2, p. 5	not available
Renishaw (Raman)	spc	read.spc	6.6.1, p. 11	Not recommended, see discussion of ASCII files.
Shimadzu (GC,GC-MS)	JCAMP-DX	jcamp-dx	4.2, p. 5	import for subset of the $JCAMP-DX$ standard
Witec (Raman)	Witec ASCII	scan.txt.Witec	6.8, p. 13	

C. File Import Functions by Spectroscopy

Spectroscopy	Format	Manufacturer	Function	section	Notes
Fluorescence	ASCII long	Perkin Elmer	read.txt.PerkinElmer	7.1, p. 14	Reads multiple files, needs to be sourced.
GC,GC-MS	JCAMP-DX	Shimadzu	jcamp-dx	4.2, p. 5	import for subset of the JCAMP-DX standard
Infrared	JCAMP-DX	PerkinElmer	jcamp-dx	4.2, p. 5	import for subset of the JCAMP-DX standard
Infrared Imaging	ENVI	Bruker	read.ENVI	6.2, p. 9	
Infrared Imaging	ENVI	Nicolet	read.ENVI.Nicolet	6.3, p. 10	
Infrared Imaging	Opus	Bruker	-	6.2, p. 9	
Raman	ASCII long	Renishaw	scan.txt.Renishaw	6.6.1, p. 11	
Raman	ASCII long	Kaiser	read.txt.long	6.5.1, p. 10	Not recommended, see discussion
Raman	hol	Kaiser	-	6.5, p. 10	via Matlab
Raman	JCAMP-DX	Renishaw	-	4.2, p. 5	not available
Raman	spc	Kaiser	read.spc	5.3, p. 7	Reads multiple files
Raman	spc	Renishaw	read.spc	6.6.1, p. 11	Not recommended, see discussion of ASCII files.
Raman	Witec ASCII	Witec	scan.txt.Witec	6.8, p. 13	
Raman Map	spc	Kaiser	read.spc.KaiserMap	6.5.2, p. 11	Reads multiple files

Index

Agilent ENVI, 10	Shimadzu, 5 jdx, see JCAMP-DX
ASCII	
compressed, 11	Kaiser
JCAMP-DX, 5	ASCII long, 10
long, 4	hol, 10
Fluorescence, 14	Map, 11
Kaiser, 10	spc, 7, 10, 11
PerkinElmer, 14	M 44
Raman, 10, 11	Map, 11
Renishaw, 11	ENVI, 6, 9, 10
samples in columns, 4	Kaiser, 11
transposed, 4 wide, 4	Raman, 11
Horiba Jobin Yvon, 13	Matlab, 6
Witec, 13	now hyperSpec object 3
zip, 11	new hyperSpec object, 3 Nicolet
F,	ENVI, 10, 15
Bruker	Infrared, 15
AXS, 4	Map, 15
ENVI, 6, 9	map, 10
powder diffraction, 4	PerkinElmer
x-ray, 4	ASCII long, 14
	Fluorescence, 14
create hyperSpec object, 3	Infrared, 5
	JCAMP-DX, 5
ENVI	powder diffraction
Agilent, 10	Bruker, 4
Bruker, 6, 9	
Infrared, 6, 9, 10	Raman
Nicolet, 15	ASCII long, 10, 11
Map, 6, 9, 10, 15	hol, 10
Nicolet, 10, 15	Horiba Jobin Yvon
Varian, 6	ASCII wide, 13
T)	Kaiser, 7, 10, 11
Fluorescence	Map, 11
ASCII long, 14	Renishaw, 7
PerkinElmer	ASCII, 11
ASCII, 14	spc, 11
FT-IR, see Infrared	spc, 7, 10, 11
hol	Witec ASCII, 13
Kaiser, 10	Renishaw
Horiba Jobin Yvon	ASCII long, 11
ASCII	Raman, 11
wide, 13	spc, 7, 11
hyperSpec object	F 1) 1)
create, 3	Shimadzu
	GC-MS, 5
Image, see Map	JCAMP-DX, 5
Infrared	spc, 7
ENVI, 6, 9, 10, 15	Kaiser, 7, 10
JCAMP-DX, 5	
Map, 15	Raman, 7, 10, 11
wap, 15	Raman, 7, 10, 11 Renishaw, 7, 11
Nicolet, 15	Raman, 7, 10, 11
Nicolet, 15 PerkinElmer, 5	Raman, 7, 10, 11 Renishaw, 7, 11 TriVista, 7
Nicolet, 15	Raman, 7, 10, 11 Renishaw, 7, 11 TriVista, 7
Nicolet, 15 PerkinElmer, 5 initialize hyperSpec object, 3	Raman, 7, 10, 11 Renishaw, 7, 11 TriVista, 7
Nicolet, 15 PerkinElmer, 5 initialize hyperSpec object, 3 JCAMP-DX	Raman, 7, 10, 11 Renishaw, 7, 11 TriVista, 7 TriVista spc, 7
Nicolet, 15 PerkinElmer, 5 initialize hyperSpec object, 3 JCAMP-DX ASCII, 5	Raman, 7, 10, 11 Renishaw, 7, 11 TriVista, 7 TriVista spc, 7 Varian, see Agilent
Nicolet, 15 PerkinElmer, 5 initialize hyperSpec object, 3 JCAMP-DX	Raman, 7, 10, 11 Renishaw, 7, 11 TriVista, 7 TriVista spc, 7

Witec ASCII, 13

x-ray Bruker, 4

Session Info

R version 2.15.2 (2012-10-26)

Platform: x86_64-pc-linux-gnu (64-bit)

locale

[7] LC_PAPER=C LC_NAME=C LC_ADDRESS=C

[10] LC_TELEPHONE=C LC_MEASUREMENT=de_DE.UTF-8 LC_IDENTIFICATION=C

attached base packages:

[1] grid stats graphics grDevices utils datasets methods base

other attached packages:

[1] hyperSpec_0.98-20121216 mvtnorm_0.9-9994 lattice_0.20-10

loaded via a name space (and not attached):

[1] tools_2.15.2