tidyspec: a gentle framework to handle with spectroscopy data using tidy data philosophy

Marcel Rodrigues Ferreira[1,2] Julia Ferreia Moraes[1]

Luisa Sutter[1] Willian Fernando Zambuzzi[3,*]

[1] Departament of Chemistry and Biochemistry, São Paulo State University (UNESP), Institute of Biosciences, Campus Botucatu
[2] Molecular Genetics and Bioinformatics Laboratory, Experimental Research Unit (Unipex), School of Medicine, São Paulo State University (UNESP)
[3] Departament of Chemistry and Biochemistry, São Paulo State University (UNESP), Institute of Biosciences, Campus Botucatu

[*] Correspondence: Willian Fernando Zambuzzi <w.zambuzzi@unesp.br>

```
<!-- convidar o grupo da USP para trabalhar nisso -->
```

Abstract

Introduction

Spectroscopy

Spectroscopy, the study of interactions between light and matter, has been a cornerstone of scientific exploration for centuries, enabling us to peer into the fundamental properties and intricate structures of materials across various domains. From the early days of visible light observations to the cutting-edge techniques of modern spectroscopy, researchers have continually pushed the boundaries of our understanding, unraveling the mysteries of atoms, molecules, and complex systems. This paper presents a comprehensive exploration of recent advancements in spectroscopic analyses, delving into the multifaceted applications, innovative methodologies, and emergent technologies that have revolutionized our ability to probe, characterize, and manipulate matter at its most fundamental levels. By harnessing the power of spectroscopy, scientists are embarking on a journey to unlock hidden dimensions of matter, uncovering new insights that have far-reaching implications in fields ranging from materials science and

chemistry to astronomy and biology. This paper not only reviews key developments in spectroscopic techniques but also highlights their profound impact on our comprehension of the physical world, underlining the ever-expanding role of spectroscopy as an indispensable tool for scientific discovery and technological innovation.

R language

The R programming language has emerged as a versatile and powerful tool in the realm of spectroscopy, offering researchers an extensive suite of specialized packages designed to facilitate data analysis, visualization, and interpretation. Leveraging R's rich statistical capabilities and its seamless integration with a diverse range of data formats, spectroscopists can efficiently process and manipulate complex spectral datasets. Notably, packages such as `{hyperSpec}` [1](Beleites and Sergo) and `{ChemoSpec}` (Hanson 2023) provide dedicated functions for handling hyperspectral data, allowing for preprocessing, spectral alignment, and exploratory data analysis. 'SpecMine' and 'SpecHelpers' further extend R's utility by enabling advanced spectral processing, peak detection, and quantification tasks. Visualization of spectroscopic data is greatly enhanced by packages like `{ggplot2}` (Wickham 2016) and `{plotly}` (Sievert 2020), enabling researchers to create insightful and interactive graphical representations. The extensibility of R also encourages the development of customized algorithms and methods tailored to specific spectroscopic challenges. As the field of spectroscopy continues to expand its horizons, R remains an invaluable asset, empowering scientists to unravel the intricacies of spectral information with precision and depth.

Tidydata and tidyverse

Data sets

Overview of `tidyspec` package

The `tidyspec` package was design to enable the data analysis of spectroscopy data (as IR, Raman, NMR) with the tidydata format. Similar to packages like `{stringr}` (Wickham 2022) and `{forcats}` (Wickham 2023), the tidyspec package contains a naming pattern at the beginning of its function name, in this case `spec_`. The functions were designed to solve 6 different problems in spectroscopy:

- **Transformation**: Convert data from absorbance to transmittance (`spec_abs2trans`) & from transmittance to absorbance (`spec_trans2abs`).

$$A = 2 - \log_{10} T(\%)$$

[1]In this paper, the R packages names will be formated as {package}, a format widely used by the R community.

- **Normalization**: Normalize the data to range 0-1 (`spec_norm_01`), normalize between a custom range (`spec_norm_minmax`), or normalize to have a standard deviation of one (`spec_norm_var`).
- **Baseline correction**: Correct the baseline using the the *rolling ball* algorithm (`spec_blc_rollingBall`). The function `spec_bl` return the baseline vectors (`spec_bl_rollingBall`).
- **Smooth correction**: Smooth the data using the average window (`spec_smooth_avg`) or using the Savitzky-Golay algorithm (`spec_smooth_sga`).
- **Derivative**: Create differential data from the spectra (`spec_diff`).
- **Preview**: Preview your data while applying changes statically (`spec_smartplot`) or interactively (`spec_smartplotly`).

Case studies

UV-VIS data

```
library(tidyspec)
library(tidyverse)
library(modeldatatoo)

UV <- data_chimiometrie_2019()
```

FTIR

Raman

{hyperSpec} chodro data.

```
library(hyperSpec)
Raman <- chondro
```

Conclusion and perspectives

Acknowledgments

References

Beleites, Claudia, and Valter Sergo. "hyperSpec: A Package to Handle Hyperspectral Data Sets in r." https://github.com/r-hyperspec/hyperSpec.

Hanson, Bryan A. 2023. "ChemoSpec: Exploratory Chemometrics for Spectroscopy." https://CRAN.R-project.org/package=ChemoSpec.

Sievert, Carson. 2020. "Interactive Web-Based Data Visualization with r, Plotly, and Shiny." https://plotly-r.com.

Wickham, Hadley. 2016. "Ggplot2: Elegant Graphics for Data Analysis." https://ggplot2.tidyverse.org.

———. 2022. "Stringr: Simple, Consistent Wrappers for Common String Operations." https://CRAN.R-project.org/package=stringr.

———. 2023. "Forcats: Tools for Working with Categorical Variables (Factors)." https://CRAN.R-project.org/package=forcats.