

Article

# Towards a universal calibration model to predict biogenic silica and organic carbon percentages in lacustrine sediment core samples

Rana Gahwagy<sup>1,2,†,\*</sup> , Lauren Meyer<sup>2,†,‡</sup>, Grace Hartley<sup>2,†</sup>

<sup>1</sup> Smith College Department of Statistical and Data Sciences Northampton, MA 01063; [rgahwagy@smith.edu](mailto:rgahwagy@smith.edu), [lmeyer@smith.edu](mailto:lmeyer@smith.edu), [ghartley@smith.edu](mailto:ghartley@smith.edu)

<sup>2</sup> Your department Street, City, Country; [mail@mail.com](mailto:mail@mail.com)

\* Correspondence: [leutnant@fh-muenster.de](mailto:leutnant@fh-muenster.de); Tel.: +XX-000-00-0000.

† Current address: Updated affiliation

‡ These authors contributed equally to this work.

Version March 8, 2022 submitted to Water



**Simple Summary:** A Simple summary goes here.

**Abstract:** In many settings, biogenic silica (BSi) and total organic carbon (TOC) are widely used as proxies for temperature and/or environmental variations that are helpful in paleoclimate and paleoenvironmental reconstructions. Often, the methodology for analyzing these parameters in sediments can be expensive and time consuming (particularly for BSi). However, Fourier Transform Infrared (FTIR) Spectroscopy offers an efficient alternative where many samples can be run with minimal amount of sediment and time. This technique is advantageous in that it requires small volumes of sediment (~0.01g), minimal sample preparation (mixing sample with potassium bromide powder), and instrumental analysis times are relatively rapid (a few minutes per sample). FTIR Spectroscopy quantifies BSi and TOC using infrared radiation (IR) absorbance units—as opposed to percentages of BSi or TOC—which are difficult to compare across different studies and localities. Therefore, there is a need for a systematic way to convert the results from the FTIR Spectrometer into percentages. In this research project, we address this need by building a universal calibration model using partial least squares (PLS) regression that converts BSi absorbance to percentages. We developed this model using a PLS package in R and based our model on samples from Arctic lakes in Greenland and Alaska. Our preliminary model uses a k-fold cross-validation method and utilizes three components. Ongoing work intends to improve on the model's prediction accuracy, expand our calibration model to include TOC percentages, and incorporate more BSi samples from other locations. We aim for the model to be universal and integrated into a Shiny app, where paleoclimatologists can use it on samples from various localities and compare their results. This model will prove a valuable tool in paleoclimate reconstruction by facilitating FTIR Spectroscopy on lake sediments.

**Keywords:** keyword 1; keyword 2; keyword 3 (list three to ten pertinent keywords specific to the article, yet reasonably common within the subject discipline.).

## 1. Version

This Rmd-skeleton uses the mdpi Latex template published 2019/02. However, the official template gets more frequently updated than the 'rticles' package. Therefore, please make sure prior to paper submission, that you're using the most recent .cls, .tex and .bst files (available [here](#)).

## 2. Introduction

Studying the content of biogenic silica and other organic compounds present in lake sediment cores can be a powerful tool in gaining insight into our reconstruction of past climates. The wet chemistry processes used to determine these proportions can be time consuming and costly. As a result, data on the amount of biogenic silica present in different samples is limited in quantity and resolution. Fourier-transform infrared (FTIR) spectroscopy is a promising technique to reduce the time and money needed to determine the proportion of these compounds in lake sediments. FTIR spectroscopy involves measuring the absorbance of the infrared radiation at different wavelengths. These absorbance values are arbitrary and unitless, so interpretation is needed to make them relevant between researchers. To interpret these absorbance values, we seek to use a Partial Least Squares Regression (PLSR) model to predict the percentage of the biogenic silica in each sample. This technique has been pioneered successfully by Vogel *et al.* [1], but is not accessible to those who might wish to use FTIR spectroscopy for their samples and lack the statistical background to implement a PLSR model.

The goal of this project is to create an interface where a user can input their FTIR spectroscopy data into a PLSR model to calculate the approximate biogenic silica and total organic carbon percentages. Our work primarily includes improving the accuracy of the existing model (<https://github.com/people-r-strange/PLSmodel>) and preparing for universal input. This includes selection of the most applicable diagnostic plots to determine the accuracy of the model. Because this work will be accessible to the public, our model and corresponding interface satisfy an existing need for efficient FTIR spectroscopy interpretation. Further, greater accessibility and use of this technology may eliminate the need for expensive wet chemical processes.

## 3. Materials and Methods

Materials and Methods should be described with sufficient details to allow others to replicate and build on published results. Please note that publication of your manuscript implicates that you must make all materials, data, computer code, and protocols associated with the publication available to readers. Please disclose at the submission stage any restrictions on the availability of materials or information. New methods and protocols should be described in detail while well-established methods can be briefly described and appropriately cited.

Research manuscripts reporting large datasets that are deposited in a publicly available database should specify where the data have been deposited and provide the relevant accession numbers. If the accession numbers have not yet been obtained at the time of submission, please state that they will be provided during review. They must be provided prior to publication.

Interventionary studies involving animals or humans, and other studies require ethical approval must list the authority that provided approval and the corresponding ethical approval code.

## 4. Results

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation as well as the experimental conclusions that can be drawn.

### 4.1. Subsection Heading Here

Subsection text here.

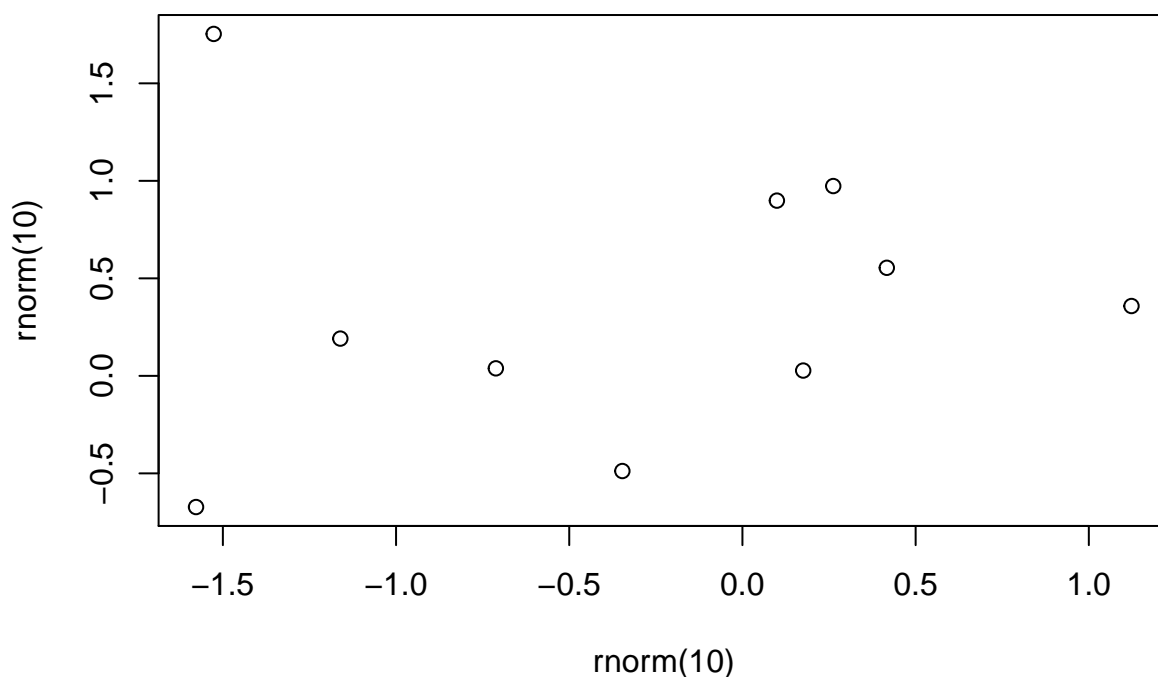
#### 4.1.1. Subsubsection Heading Here

Bulleted lists look like this:

- First bullet
- Second bullet
- Third bullet



**Figure 2.** This is a figure, Schemes follow the same formatting. If there are multiple panels, they should be listed as: (a) Description of what is contained in the first panel. (b) Description of what is contained in the second panel. Figures should be placed in the main text near to the first time they are cited. A caption on a single line should be centered.



**Figure 3.** test figure

Numbered lists can be added as follows:

1. First item
2. Second item
3. Third item

The text continues here.

All figures and tables should be cited in the main text as Figure ??, Table 1, etc.



**Figure 1.** This is a figure, Schemes follow the same formatting. If there are multiple panels, they should be listed as: (a) Description of what is contained in the first panel. (b) Description of what is contained in the second panel. Figures should be placed in the main text near to the first time they are cited. A caption on a single line should be centered.

Figure ?? above Figure ?? below.

**Table 1.** caption here,

Title 1	Title 2	Title 3
entry 1	data	data
entry 2	data	data

table 1 above

Bibliography reference Gujer [2], [2]

**Table 2.** This is a table caption. Tables should be placed in the main text near to the first time they are cited.

Title 1	Title 2	Title 3
entry 1	data	data
entry 2	data	data

This is an example of an equation:

$$\S \quad (1)$$

Example of a theorem:

**Theorem 1.** *Example text of a theorem.*

The text continues here. Proofs must be formatted as follows:

Example of a proof:

**Proof of Theorem 1.** Text of the proof. Note that the phrase ‘of Theorem 1’ is optional if it is clear which theorem is being referred to.  $\square$

The text continues here.

## 5. Discussion

Authors should discuss the results and how they can be interpreted in perspective of previous studies and of the working hypotheses. The findings and their implications should be discussed in the broadest context possible. Future research directions may also be highlighted.

## 6. Conclusion

This section is not mandatory, but can be added to the manuscript if the discussion is unusually long or complex.

## 7. Patents

This section is not mandatory, but may be added if there are patents resulting from the work reported in this manuscript.

wordcountaddin::text\_stats()

**Acknowledgments:** All sources of funding of the study should be disclosed. Please clearly indicate grants that you have received in support of your research work. Clearly state if you received funds for covering the costs to publish in open access.

**Author Contributions:** For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used “X.X. and Y.Y. conceive and designed the experiments; X.X. performed the experiments; X.X. and Y.Y. analyzed the data; W.W. contributed

reagents/materials/analysis tools; Y.Y. wrote the paper.” Authorship must be limited to those who have contributed substantially to the work reported.

**Conflicts of Interest:** Declare conflicts of interest or state ‘The authors declare no conflict of interest.’ Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. Any role of the funding sponsors in the design of the study; in the collection, analyses or interpretation of data in the writing of the manuscript, or in the decision to publish the results must be declared in this section. If there is no role, please state ‘The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, an in the decision to publish the results’.

## Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	linear dichroism

## Appendix A

### Appendix A.1

The appendix is an optional section that can contain details and data supplemental to the main text. For example, explanations of experimental details that would disrupt the flow of the main text, but nonetheless remain crucial to understanding and reproducing the research shown; figures of replicates for experiments of which representative data is shown in the main text can be added here if brief, or as Supplementary data. Mathematical proofs of results not central to the paper can be added as an appendix.

## Appendix B

All appendix sections must be cited in the main text. In the appendixes, Figures, Tables, etc. should be labeled starting with ‘A’, e.g., Figure A1, Figure A2, etc.

## References

1. Vogel, H.; Rosén, P.; Wagner, B.; Melles, M.; Persson, P. Fourier transform infrared spectroscopy, a new cost-effective tool for quantitative analysis of biogeochemical properties in long sediment records. *Journal of Paleolimnology* **2008**, *40*, 689–702. doi:10.1007/s10933-008-9193-7.
2. Gujer, W. *Systems Analysis for Water Technology*; Springer-Verlag: Berlin, Heidelberg, Germany, 2008.

**Sample Availability:** Samples of the compounds . . . . . are available from the authors.

© 2022 by the authors. Submitted to *Water* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).