

Claudia Colini, Lucas F. Voges, Stephan Seifert

Chapter 8: Analysis of Paper Components with FTIR (DRIFTS) Spectroscopy

Abstract: In this chapter, Fourier transform infrared spectroscopy (FTIR), independently of microscopy, will be applied to study the *tsakali* paper components, especially for the identification of fillers, sizing agents and degradation processes. Previously, this was also used for fibre identification, although the presence of sizing and of multiple types of fibres in the same object might compromise the results. This method is based on the excitation of chemical bonds in a molecule by IR light with characteristic frequencies, resulting in specific types of vibrations. These vibrations are detected and recorded in spectra: the position and intensity of the peaks provide information about the chemical composition and crystalline structure or geometry of the molecule. Comparing reference spectra to those of unknown substances proved extremely effective for their identification. Furthermore, chemometric approaches like principal component analysis (PCA) will be applied for fingerprinting. PCA generates new variables (principal components) that represent the main variances of the spectra, and can thus be used to analyse the similarities and differences between them.

1 Introduction

Based on the sieve used for the formation of the writing support, described in Chapters 1 and 5, the *tsakali* can be divided into two main groups: laid papers (occurring only in three folios: set A, cards 23 and 25, and set B, cards D) and woven papers (occurring in all other instances). The woven papers can be further subdivided into two groups on the basis of the technological process received by the fibres: in fact, thirty-nine folios in set A and four of set B show the presence of bundles of unbeaten fibres, thereby resulting in a higher fibre density, while fifteen folios in set A and three in set B appear to have been more evenly pulped.

We applied Fourier transform infrared spectroscopy (FTIR) to study the *tsakali*, especially for the identification of additives, fillers and sizing agents that may have been added to the paper. As this technique has also been used for fibre iden-

tification in textile and paper objects in the past,¹ we calculated the intensity ratio of characteristic peaks to compare the paper of the cards. The results have been evaluated in light of the identification performed via optical microscopy (in Chapter 6). Furthermore, chemometric approaches like principal component analysis (PCA) have been applied to compare all the spectra and thus carry out a fingerprint analysis of the material. PCA generates new variables (called principal components) that represent the main variances of the spectra, and can thus be used to analyse the similarities and differences between them.

2 FTIR analysis

Fourier transform infrared spectroscopy (FTIR) is based on the excitation of vibrations of chemical bonds in molecules by characteristic frequencies of IR light. The light used for the excitation is detected and displayed in spectra, and the position and intensity of the absorption peaks in the spectra provide information about the chemical composition and crystalline structure or geometry of the molecules involved. The identification of the unknown substances is normally obtained by comparing their spectra to those of reference materials. FTIR spectroscopy has been successfully utilised for fibre identification in previous studies, though mostly in attenuated total reflectance (ATR) mode.²

In this study, a portable 4100 Exoscan FTIR spectrometer (Agilent) was used in diffuse reflectance mode (DRIFTS). This is a surface non-destructive and non-invasive mode, although it requires stable contact with the analysed object. In fact, the equipment was placed perpendicularly to the object, touching it. The spectrometer is equipped with a ZnSe beam splitter, a Michelson interferometer and a thermoelectrically cooled dTGS detector. It has a spectral range of 4000 to 650 cm^{-1} and a spectral resolution of 4 cm^{-1} . A gold reference cap was used for background calibration and 256 scans were collected. Chemometric analysis was carried out in R 4.2.3³ with the packages *hyperSpec*,⁴ *EMSC*⁵ and *prospectr*.⁶

Measurements of the *tsakali* were done in triplicate on the unwritten paper area for fifteen cards. Nine cards from set A (3, 7, 8, 19, 23, 25, 49, 51, 55) and six

¹ See e.g. Garside and Wyeth 2003; Espejo et al. 2010; Peets et al. 2019.

² See e.g. Garside and Wyeth 2003; Espejo et al. 2010; Peets et al. 2019.

³ The R Core Team 2022.

⁴ Beleites et al. 2021.

⁵ Liland 2021.

⁶ Stevens and Ramirez-Lopez 2025.

cards from set B (B, C, D, E, F, G) were analysed. The raw data is available at the Research Data Repository of Universität Hamburg under CC BY 4.0.⁷

Spectra were imported in a hyperSpec R object, and the spectral range was subsequently reduced to 700 to 3800 cm^{-1} . Extended multiplicative scatter correction (EMSC) was applied to reduce scattering and apply normalisation. Smoothing (together with the first or second derivative) was applied with a window size of 11 spectral points (resulting in a total range of 20.5 cm^{-1}) and a polynomial order of two. Baseline correction was achieved with polynomial baselines (polynomial order 1).

3 Results

All the peaks present in the spectra are compatible with the expected peaks associated with cellulose, hemicellulose and lignin. By comparison with reference spectra of substances used in papermaking in Asia and Europe, no fillers, additives or sizing could be identified.

The paper of the *tsakali* was analysed to see whether the two sets or the paper types had an influence on the resulting spectra. Principal component analysis (PCA) was used to visualise and detect whether differences in absorbance can be connected with the paper type or manuscript sets. The spectral range between 1900 and 2400 cm^{-1} was not included in the PCA, as signals from air carbon dioxide and monoxide are located here. After preprocessing, the spectra were comparable in intensities (see Fig. 1) and were used for PCA. The first and second derivatives were also used for PCA, as they are not influenced by the baseline.

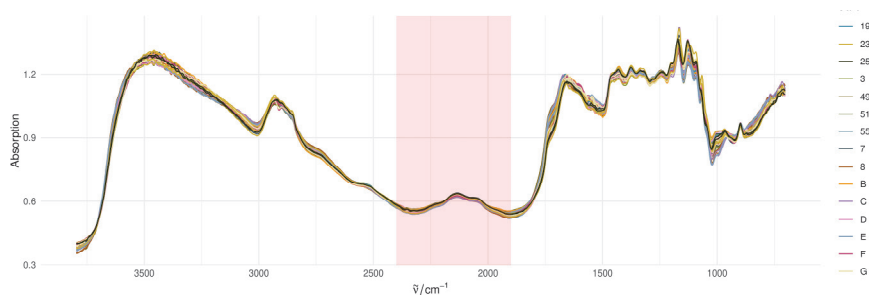


Fig. 1: Spectra after preprocessing (EMSC; smoothing and spectral range reduction to 700 to 3800 cm^{-1}). The area in light red was not used in the PCA.

⁷ Voges, Horn and Seifert 2024.

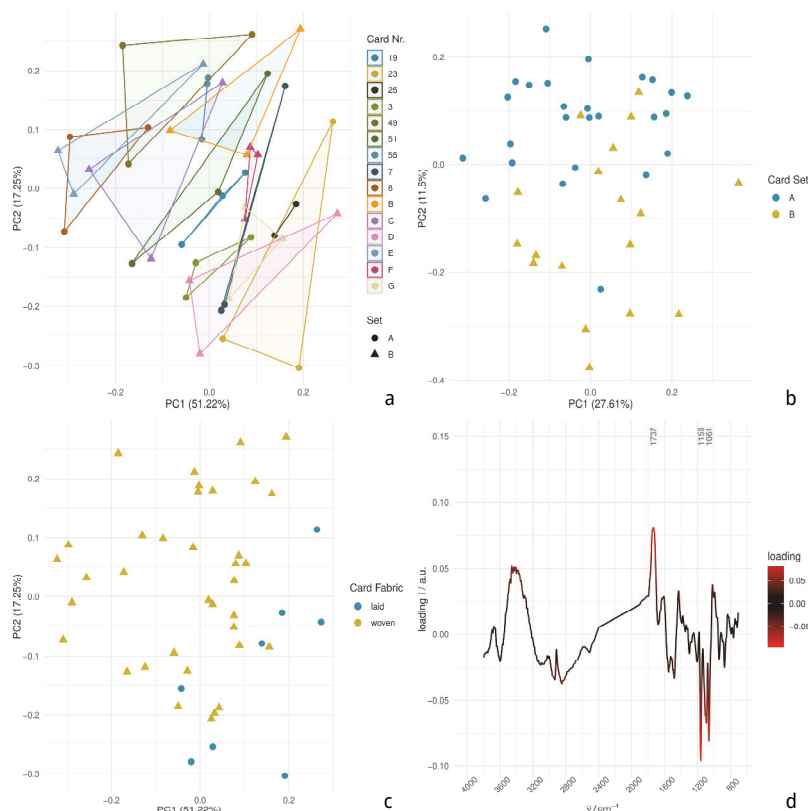
The results of the PCA are shown in Figs 2a–d. Around 50 per cent of the variance can be explained by the first component, and no differences separating the sets can be observed for the preprocessed spectra (Fig. 2a), while, in the second derivative (Fig. 2b), a trend seems to emerge. When considering the paper type, the scores of the principal components in the preprocessed spectra (Fig. 2c) seem to present some differences. The corresponding loadings for the first and second components (Fig. 2d) indicate that the peaks around 1737 cm^{-1} , 1000 to 1100 cm^{-1} and, in the higher wavenumbers, around 3400 to 3600 cm^{-1} are mainly responsible for these differences. The loadings in the first derivative present the same results (not shown here) with the addition of a strong contribution of the peak at 1160 cm^{-1} , which, therefore, might be the main reason for the paper type separation.

The peaks identified by the PCA analysis as contributing to the differentiation of paper types and card sets can be assigned to the relevant bonds present in the cellulose, hemicellulose and lignin, which are the main components of paper materials.⁸ The area around 3500 cm^{-1} can be assigned to OH valence vibration and may indicate different cellulose compositions of the paper types. The peak around 1170 cm^{-1} is sometimes shifted to lower wavenumbers (1166 – 1160), and from the first derivative, this peak has an impact depending on the paper type. This peak may be assigned to asymmetric C-O-C valence vibrations, and can be caused by cellulose and hemicellulose as well as by lignin. The area between 1000 and 1130 cm^{-1} can be assigned to C-O valence vibrations of the functional groups of primary and secondary alcohols found in cellulose, hemicellulose and lignin.

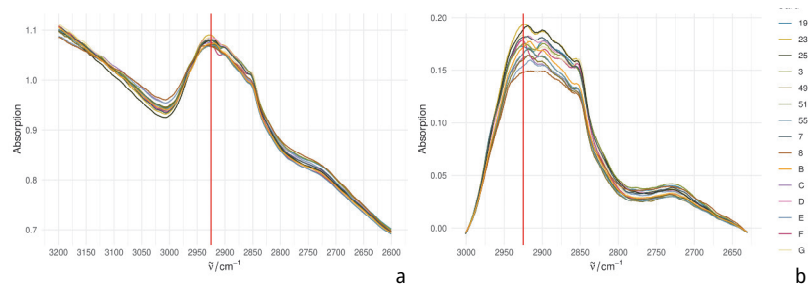
To analyse the paper types in detail, various bands from cellulose, hemicellulose and lignin have been compared as intensity ratios as described by Paul Garside and Paul Wyeth.⁹ Three ratios have been calculated from the spectra using peaks at 2925 cm^{-1} (general organic content, C-H vibration), 1595 cm^{-1} (lignin, C=C vibration) and 1110 cm^{-1} (glycosidic ether, mainly cellulose, C-O-C vibration) – in the same way as reported by Garside and Wyeth – with the addition of the peak at 1740 cm^{-1} (C=O vibration, corresponding to the carbonyl groups of oxycelluloses and lignin in the case of degraded materials or, though less likely in the case of paper, to pectin), as it was a promising feature identified by the PCA loadings. Intensities were measured at the mentioned wavenumbers on the raw, preprocessed and local baseline-corrected spectra. The baseline for the corresponding peaks was fitted in the spectral ranges 2600 – 3200 cm^{-1} , 1485 – 1780 cm^{-1} , 1025 – 1185 cm^{-1} and 1670 – 1800 cm^{-1} .

⁸ Garside and Wyeth 2003; Peets et al. 2019.

⁹ Garside and Wyeth 2003.



Figs 2a–d: (a) PCA of preprocessed spectra, with polygons highlighting the individual cards; (b) PCA of the second derivative of the spectra, with different highlighting of the cards, set and paper type; (c) PCA of preprocessed spectra, with fabric type indicated by colour; and (d) loading plot for the second component of the PCA (peaks with the highest absolute values are marked [cm^{-1}]).



Figs 3a–b: Mean spectra for each *tsakali* measured. (a) Preprocessed spectra. (b) Spectra with base-line fitted. The red vertical line shows the peak position at 2930 cm^{-1} used for the analysis.

An example of the differences in the baseline fitted and preprocessed spectra can be seen in Figs 3a–b for the 2925 cm⁻¹ peak. The mean values of the calculated ratios and their standard deviation are listed in Table 1.

Table 1: Mean values with standard deviation for the peak intensity ratios for the different paper types of spectra with different processing.

Spectra	Paper Type	R1 (1595/1110)	R2 (1595/2925)	R3 (1735/2925)
Raw	laid	0.86 ± 0.02	1.03 ± 0.01	0.80 ± 0.02
	woven	0.90 ± 0.03	1.05 ± 0.01	0.87 ± 0.04
Preprocessed	laid	0.87 ± 0.02	1.03 ± 0.01	0.81 ± 0.01
	woven	0.90 ± 0.02	1.05 ± 0.01	0.87 ± 0.03
Baseline	laid	1.09 ± 0.18	1.12 ± 0.08	0.31 ± 0.03
	woven	1.15 ± 0.15	1.28 ± 0.09	0.49 ± 0.14

Although the preprocessing and especially baseline fitting have a strong influence on the spectra, the ratios are quite robust regardless of the different processing. This is true except for R3, with baseline correction, which could be based on the fact that 1735 is a shoulder peak, hence the baseline correction is more strongly influenced by the neighbouring peaks.

We compared the values of the ratios obtained in this study with (1) the intensity ratios obtained by Garside and Wyeth¹⁰ for flax, jute, hemp, cotton, ramie and sisal (both native and processed fibres); and (2) the theoretical spectral regions corresponding to flax, jute, hemp and cotton, calculated by Teresa Espejo, Adrian Duran, Ana Lopez-Montes and Rosario Blanc¹¹ from the materials' theoretical percentages of cellulose, hemicellulose, pectin and lignin. Such comparison did not lead to any identification of fibres, as our values fell in different spectral regions. As demonstrated by Pilleriin Peets, Karl Kaupmees, Signe Vahur and Ivo Leito,¹² however, the position, shape and intensity of peaks change from ATR to DRIFTS, which might be the reason why no identification could be obtained. Additionally, the fibres were identified as *Stellera* with the possible admixture of *Daphne* fibres, according to optical microscopy (see Chapter 6), species that were not considered in the previous studies, and for which we do not have intensity ratios. In the

¹⁰ Garside and Wyeth 2003.

¹¹ Espejo et al. 2010.

¹² Peets et al. 2019.

future, we intend to build our own reference library in reflectance, including additional fibres used in Asian papermaking, in order to be able to compare the sets of data.

The scatter plot of the three ratios based on the preprocessed spectra is shown in Fig. 4; it is worth mentioning that for all the different types of processing, the plots show comparable results. As previously mentioned, the same mixture of fibres was found in all the *tsakali* paper analysed; however, all three ratios, but particularly R2 (1595/2925) and R3 (1735/2925), show a strong separation between high- and low-fibre density paper. This is to be expected, as the presence of fibre bundles is naturally associated with a higher content of lignin, found in the macrofibre walls, as well as with a shorter time of cooking. Furthermore, lignin oxidises faster than cellulose, resulting in the formation of additional carbonyl groups¹³ and thus in higher values of R3. R3 also shows a separation between paper types, with lower values connected to laid paper, perhaps suggesting that different sheet formations might play a role in paper degradation. However, the separation between paper types is ambiguous at the area of overlap between the two types, which is still reasonable, as the material is inhomogeneous and there are many factors influencing the measurement. Hence, for measurements of this kind, it is advised to take the mean of multiple measurements from the sample, as ratios from the same card can show fluctuations.

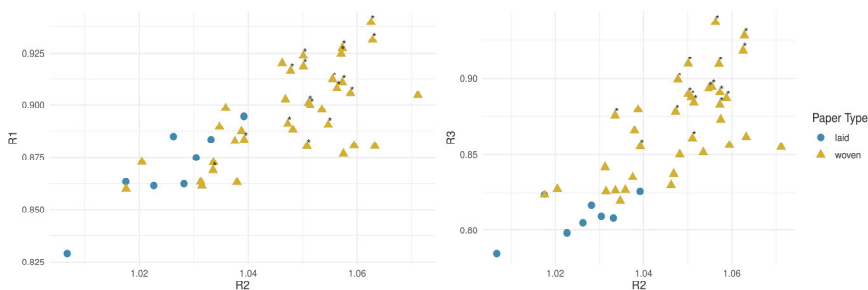


Fig. 4: Scatter plots for ratios R1/R2 and R3/R2 woven paper type with high fibre density are marked with asterisks. The peak intensities are from the preprocessed spectra.

4 Conclusions

The analysis with ratios proves to be a very robust method to measure differences in the paper types. The peaks identified in ATR by Garside and Wyeth¹⁴ for the calculation of ratios R1 and R2 do not correspond precisely to those in DRIFTS, as the latter seems to be broader and shifted to higher wavenumbers.¹⁵ Further investigation of other relevant peaks determined by PCA (such as the one at 1735 cm⁻¹, which we used to calculate R3) could advance the identification of paper types and the understanding of their composition. This workflow shows a promising application to written artefacts.

As for the identification of plant fibres through the calculation of R1 and R2, it must be remarked that paper is the result of a complex set of procedures and materials, and that several variables (i.e. the part of the plant used, the extent of pulping, the addition of sizing agents like starch glue and the use of a mixture of fibres) can contribute to the alteration of these ratios. This very case study is an example of such inherent data contamination, as papers with the same plant fibres, but with high- and low-fibre density due to different pulping processes, resulted in different ratio values.

¹⁴ Garside and Wyeth 2003.

¹⁵ Peets et al. 2019.