

Machine Learning Used in Spectroscopic Analysis in Textile Recycling Processes

David Lilek, Sebnem Sara Yayla, Jeannie Egan, Melina Sophie Zirbs, Judith Rudolf-Scholik, Agnes Grünfelder, Birgit Herbinger, Christian B Schimper

Abstract:

The Josef Ressel Centre *ReSTex* – *Recovery Strategies for Textiles* – develops biotechnological and data-driven methods for recycling of cotton/polyester blends. An enzymatic process selectively separates cotton while preserving polyester, enabling sustainable material separation. Near-infrared spectroscopy is combined with machine learning to support process monitoring and optimization. A dataset of over 600 samples from enzymatic and alkaline treatments was analyzed using several preprocessing and regression models (PLS, KRR, RF, LS-SVM) for the prediction of cotton content. Unsupervised methods like t-SNE, k-mean clustering or PCA were used to find structural changes in cellulose during pretreatment. Suitable preprocessing methods and model selection significantly improve prediction accuracy and process understanding. The approach establishes a data-driven foundation for real-time decision support in textile recycling and will be extended to Raman spectroscopy.

Keywords: textile recycling, spectroscopy, machine learning, enzymatic separation, data-driven innovation

1. INTRODUCTION

Textile production has almost doubled within two decades, while the lifespan of garments has dropped by 40 %, resulting in more than 134 million tons of textile waste annually (Textile Exchange, 2025). Less than 1 % is recycled in a closed-loop process (Ellen MacArthur Foundation, 2017). A major bottleneck is the unavailability of recycling technology for blended fabrics, particularly cotton/polyester (CO/PET) textiles, which combine the two most used natural and synthetic fibers and are difficult to separate (Matsumura et al., 2024).

The Josef Ressel Centre *ReSTex* – *Recovery Strategies for Textiles* – addresses this challenge by a biotechnological and data-driven approach. The JRC develops enzymatic processes for selective separation of cotton in CO/PET blends, enabling recovery of both fractions under mild and environmentally friendly conditions. In parallel, spectroscopic techniques and machine learning are applied to analyze the fiber composition of intermediates, monitor reactions, and establish predictive models for process control.

Enzymatic methods for textile recycling offer a sustainable alternative to conventional chemical and energy intense chemical recycling processes that are being commercialized at the moment.

However, they require precise control of reaction conditions and continuous monitoring of material transformations.

One method that is applied is Near-infrared (NIR) spectroscopy. This method provides a fast and non-destructive tool for such monitoring, yet its use in textile recycling is challenged by overlapping absorption bands, color interference, and complex data structures (Cleve et al., 2000; Zhou et al., 2019). Classical chemometric approaches like PCA or PLS regression are limited when dealing with these nonlinearities. The project therefore explores machine-learning-based regression techniques as well as clustering techniques to improve predictive performance and extract deeper process insights.

2. METHODS

During the first funding period of the JRC ReSTex, a dataset of over 600 standardized textile samples was established, covering enzymatic treatments, alkaline (NaOH) pre-treatments, and controlled degradation resulting in varying cotton contents of the same precursor material (Egan et al., 2025; Yayla, 2025). Samples were measured using a Bruker MPA laboratory spectrometer with a resolution of 4 cm^{-1} and 16 scans using a diffuse reflectance unit.

For spectral preprocessing several options were tested: Standard Normal Variate (SNV), Detrending, MSC, FillPeaks, and Savitzky–Golay filtering. Regression models such as PLS, Kernel Ridge Regression (KRR), Random Forest (RF), and Least Squares SVM (LS-SVM) were compared regarding accuracy and robustness for the prediction of the cotton content. In addition, unsupervised learning (t-SNE, k-mean clustering or PCA) techniques were applied to explore the spectral effects of NaOH concentration and enzymatic activity.

Processing and model training was conducted using Python and R, with cross-validation and residual analysis to ensure statistical reliability.

3. RESULTS AND DISCUSSION

Bootstrap analysis revealed that five measurement spots per specimen are sufficient for statistically representative sampling (Yayla, 2025).

For the quantification of the cotton content, machine-learning models outperform classical chemometric approaches. Kernel Ridge Regression (KRR) and Random Forest (RF) achieve higher R^2 values and lower RMSE scores than Partial Least Squares (PLS) regression, confirming that nonlinear models are more suitable to capture spectral variability. PLS nevertheless proves to be robust and interpretable, with optimal results obtained using ten latent components without preprocessing. The analysis of PLS loadings revealed chemically meaningful features: negative loading regions around $7000\text{--}6200\text{ cm}^{-1}$ are associated with cotton-related O-H overtones, while positive peaks near 6000 cm^{-1} and $5250\text{--}4100\text{ cm}^{-1}$ correspond to polyester-specific C-H and C=O vibrations (Yayla, 2025).

The influence of preprocessing was found to be substantial. Standard Normal Variate (SNV) and Multiplicative Scatter Correction (MSC) effectively reduce scattering effects, while Detrend provides the best baseline alignment. The combination of preprocessing and model choice proved to be critical for achieving generalizable and interpretable models. Cross-validation shall be performed at the sample level to ensure that measurements from the same specimen are not included in both training and test sets, thereby avoiding bias.

Compared to typical industrial systems, which achieve accuracies of around 10 % depending on material variability and sample preparation (Stipanovic et al., 2025), we demonstrated that with the optimized workflow used in this study, errors below 1 % could be achieved, enabling highly precise and reliable detection of cotton content in blended textiles.

For the alkaline pre-treated dataset, the unsupervised learning algorithms used in this study, outperform simple dimensional-reduction methods such as PCA. Preliminary findings suggest that data-driven methods can sensitively detect chemical and structural modifications, offering new possibilities for real-time process monitoring and automated decision support in textile recycling.

4. CONCLUSIONS AND OUTLOOK

The study aimed to detect and push the limits of accuracy in the quantification of cotton content in blended textiles (cotton/PET) by combining infrared spectroscopy with data-driven analysis. An optimized workflow was established to evaluate the influence of preprocessing and regression strategies on model performance. The results show that machine-learning approaches clearly outperform classical chemometric methods, achieving errors below 1%. Overall, the study demonstrates that proper preprocessing combined with machine-learning models significantly enhances quantitative fiber prediction and provides a deeper understanding of enzymatic and chemical treatment processes. The next phase will focus on Raman spectroscopy to complement NIR and overcome current limitations with dark or wet textiles (Yan & Siesler, 2018). The combination of NIR, MIR, and Raman data will increase robustness and support industrial applications.

5. LITERATURE

- Cleve, E., Bach, E., & Schollmeyer, E. (2000). Using chemometric methods and NIR spectrophotometry in the textile industry. *Analytica Chimica Acta*, 420(2), 163–167. [https://doi.org/10.1016/S0003-2670\(00\)00888-6](https://doi.org/10.1016/S0003-2670(00)00888-6)
- Egan, J., Barta, M., Pointner, P., Herbinger, B., Rudolf-Scholik, J., Gruenfelder, A., Lilek, D., Rosenau, T., Guebitz, G. M., & Schimper, C. B. (2025). Diving into commercial cellulase formulations for circular polyester/cotton separation through targeted depolymerization of cotton. *Frontiers in Bioengineering and Biotechnology*, Volume 13-2025. <https://doi.org/10.3389/fbioe.2025.1632772>
- Ellen MacArthur Foundation. (2017). *A New Textiles Economy: Redesigning Fashion's Future*. Ellen MacArthur Foundation. <https://ellenmacarthurfoundation.org/publications>
- Matsumura, M., Inagaki, J., Yamada, R., Tashiro, N., Ito, K., & Sasaki, M. (2024). Material Separation from Polyester/Cotton Blended Fabrics Using Hydrothermal Treatment. *ACS Omega*, 9(11), 13125–13133. <https://doi.org/10.1021/acsomega.3c09350>
- Stipanovic, H., Koinig, G., Fink, T., Schimper, C. B., Lilek, D., Egan, J., & Tischberger-Aldrian, A. (2025). Quantifying Cotton Content in Post-Consumer Polyester/Cotton Blend Textiles via NIR Spectroscopy: Current Attainable Outcomes and Challenges in Practice. *Recycling*, 10(4). <https://doi.org/10.3390/recycling10040152>
- Textile Exchange. (2025). *Materials Market Report*. Textile Exchange. <https://textileexchange.org/knowledge-center/reports/materials-market-report-2025/>
- Yan, H., & Siesler, H. W. (2018). Identification of textiles by handheld near infrared spectroscopy: Protecting customers against product counterfeiting. *Journal of Near Infrared Spectroscopy*, 26(5), 311–321. <https://doi.org/10.1177/0967033518796669>

Yayla, S. S. (2025). *Supervised Machine Learning for the Quantification of Cotton/Polyester Textiles Using IR Spectroscopy* [Master's Thesis]. University of Vienna.

Zhou, J., Yu, L., Ding, Q., & Wang, R. (2019). Textile Fiber Identification Using Near-Infrared Spectroscopy and Pattern Recognition. *Autex Research Journal*, 19(2), 201–209.
<https://doi.org/10.1515/aut-2018-0055>