CERES

CHEMICAL ENGINEERING AND RENEWABLE RESOURCES FOR SUSTAINABILITY

1290 UNIVERSIDADE D COIMBRA

Predicting the coagulation potential of waste lubricant oil using multiblock machine learning of NIR and MIR spectroscopy

Rúben Gariso[a], Tiago J. Rato[a], Margarida J. Quina[a], Licínio Ferreira[a], Marco S. Reis[a*]

[a]CERES, Department of Chemical Engineering, University of Coimbra, Rua Sílvio Lima, 3030-790, Coimbra, Portugal, marco@eq.uc.pt

Introduction

- Effective management of waste lubricant oil (WLO) is imperative to mitigate environmental hazards.
- In the waste management hierarchy of the European Union, WLO regeneration is the preferable treatment.
- However, WLO may coagulate during the regeneration process, leading to premature shutdowns of the process for cleaning and WLO disposal.
- The coagulation potential of WLO is assessed using an alkaline treatment with KOH. This test is time-consuming, labor-intensive, subjective and poses safety risks.

Goal

Develop a model for rapid and reliable prediction of the WLO coagulation potential leveraging Process Analytical Technology (PAT) and multiblock machine learning.

Data Collected:

- 41 WLO samples that coagulate.
- 66 WLO samples that do-not-coagulate.
- NIR spectra (2530 wavenumbers, 7000 to 3950 cm^{-1}) collected in triplicate.
- MIR spectra (1814 wavenumbers, 4000 to 500 cm^{-1}) collected in triplicate.

Methodology:

- The best prediction model was determined by search over multiple combinations of preprocessing techniques, block scaling, and modeling methodologies.
- A total of 1701 multiblock models and 54 single-block models were evaluated.
- The models' performance was evaluated and compared using the SS-DAC framework.

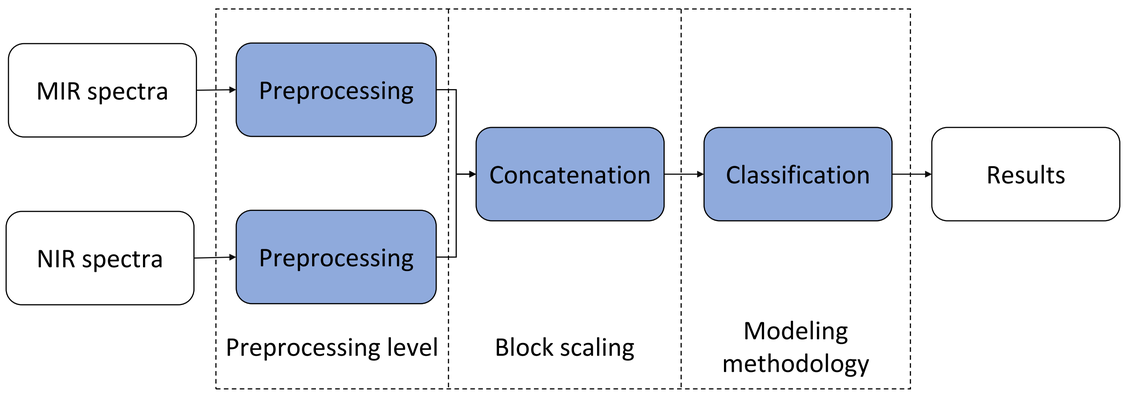

Figure 1 Diagram of the proposed methodology and its three levels to be optimized (spectra preprocessing, block scaling, and modeling methodology).

Spectral Preprocessing:

- Focus on three prominent categories:
  - Standard Normal Variate (Geladi et al., 1985);
  - Multiplicative Scatter Correction (Barker and Rayens, 2003);
  - Savitzky-Golay differentiation (SGD) (Savitzky and Golay, 1964).
- Various combinations and parameterizations of these techniques were explored.

Block Scaling:

- Block scaling methods are divided into two categories (Campos et al., 2020):
  - Block Scaling (BS): considering the number of variables within the block.
  - Block Variance Scaling (BVS): considering the standard deviation of each block.

Modeling Methodology:

- Partial least squares (PLS) for discriminant analysis was the primary modeling technique:
  - PLS (Wold et al., 2001);
  - Forward interval PLS (FiPLS) (Xiaobo et al., 2007);
  - Backward interval PLS (BiPLS) (Xiaobo et al., 2007).

Acknowledgements

The authors gratefully acknowledge the financial support of SOGILUB – Sociedade de Gestão Integrada de Óleos Lubrificantes Usados, Lda. The authors also acknowledge the support from the Chemical Engineering and Renewable Resources for Sustainability (CERES), which is financed by national funds from FCT/MCTES (reference UIDB/EQU/00102/2020).

CENTRO 2020 PORTUGAL 2020 UNIÃO EUROPEIA Fundos Europeus Estruturais e de Investimento

Results

Single-Block Models:

- The best single-block model using MIR spectra (F-A-9-0-0) ranked in position #21, with an accuracy of 0.88.
- The best single-block model using NIR spectra (B-B-0-4-0) ranked in position #1708 with an accuracy of 0.56.
- MIR-based models significantly outperformed the NIR-based models.

Multiblock Models:

- 478 out of 567 FiPLS based multiblock models solely utilized MIR data, becoming equivalent to single-block models.
- The best multiblock model (F-C11-6-3-III) ranked in position #1, with an accuracy of 0.94, representing a 6.82% improvement over the best MIR single-block model.

Table 1 Accuracy and ranking of the best sing-block and multiblock models.

		Best model	Accuracy	Ranked position
Single-Block	MIR	F-A-9-0-0	0.88	#21
	NIR	B-B-0-4-0	0.56	#1708
Multiblock	MIR&NIR	F-C11-6-3-III	0.94	#1

Feature Selection:

- Multiblock models using MIR and NIR spectra showed higher performances.
- From the MIR spectra the relevant intervals were [734.4 - 966] and [1668 - 1900] cm^{-1}.
- From the NIR spectra the [4951 - 5150] cm^{-1} was frequently selected.
- The selected MIR and NIR intervals are linked to the presence of esters in WLO.

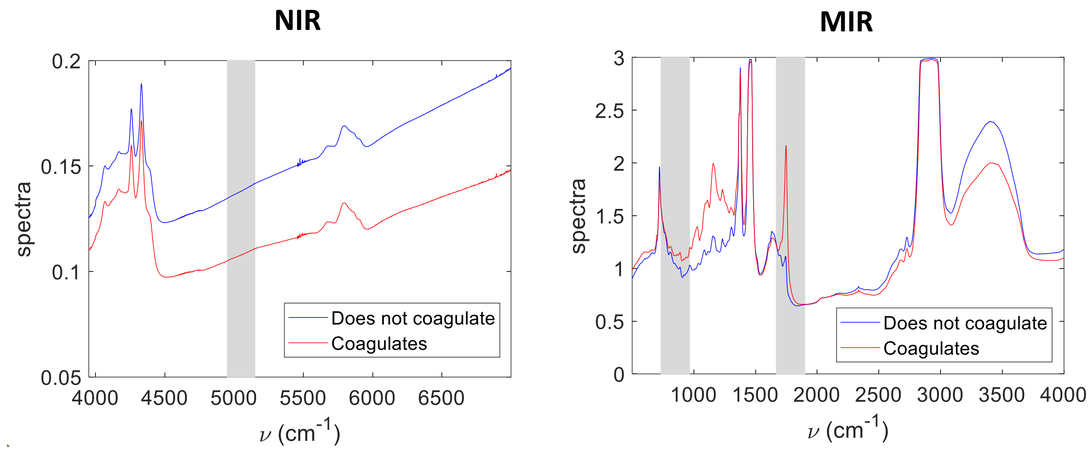

Figure 2 Average spectra of the WLO classes. The selected intervals are highlighted.

Conclusions

- Single-block models using NIR spectra had the lowest performance, with even the best model achieving only 0.56 accuracy.
- MIR single-block models were more informative, reaching an accuracy of 0.88.
- Multiblock models combining FiPLS with SGD showed the best performance, with the top model achieving 0.94 accuracy, a 6.82% improvement over single-block models.
- Block scaling had a limited impact on performance, as FiPLS mainly selected intervals solely from the MIR spectra.
- Incorporating both MIR and NIR spectral information significantly enhanced model prediction capabilities.
- The presence of esters emerged as a critical factor for WLO coagulation, providing valuable insights into the coagulation phenomenon.

Bibliography

1. P. Geladi, D. MacDougall, H. Martens, 1985, Applied Spectroscopy, 39, 491–500.
2. M. Barker, W. Rayens, 2003, Journal of Chemometrics, 17, 166–173.
3. A. Savitzky, M.J.E. Golay, 1964, Analytical Chemistry, 36, 1627–1639.
4. M. Campos, M. Reis, 2020, Chemometrics and Intelligent Laboratory Systems, 199, 103959.
5. T.J. Rato, M.S. Reis, 2019, Computers & Chemical Engineering, 128, 437–449.
6. S. Wold, M. Sjöström, L. Eriksson, 2001,Chemometrics and Intelligent Laboratory Systems, 58, 109–130.
7. Z. Xiaobo, Z. Jiewen, H. Xingyi, L. Yanxiao, 2007, Chemometrics and Intelligent Laboratory Systems, 87, 43–51