Characterization of Gas Engine Oils by IR and Chemometrics

Julia Felkel [1], Nicole Dörr [1], Tibor Leimeter [2]
Franz Novotny-Farkas [3], Kurt Varmuza [4]*

[1] Austrian Center of Competence for Tribology
AC²T research GmbH, www.ac2t.at
Viktor Kaplan-Straße 2, 2700 Wiener Neustadt, Austria

[2] Gubkin Russian State University of Oil and Gas, Production and Application of Lubricants, Moscow, Russia

[3] OMV Refining and Marketing GmbH, Schwechat, Austria

[4] Vienna University of Technology, Institute of Chemical Engineering Laboratory for Chemometrics, Vienna, Austria

* Presenting author
Laboratory for Chemometrics,
Institute of Chemical Engineering, Vienna University of Technology
Getreidemarkt 9/166, A-1060 Vienna, Austria
kvarmuza@email.tuwien.ac.at, www.lcm.tuwien.ac.at

Poster Presentation: CAC 2008
11th International Conference on Chemometrics in Analytical Chemistry
June 30 - July 4, 2008, Montpellier, France

Lubricating oil in internal combustion engines is exposed to various strains determined by the operating conditions, the fuel quality and the ambient conditions.

The **ageing of engine oils**, leading to deterioration of the base oil and depletion of performance additives, is a highly complex process.

For instance acidic combustion gases contaminate the engine oil, resulting in increased consumption of base reserve, reduction of oil lifetime and even to corrosion in the engine.

The **total acid number (*TAN*)** is an important indicator of oil quality, characterizing oil oxidation and the extent of acidic contamination.

- We investigated the applicability of **IR data** together with **chemometrics** for a qualitative and quantitative evaluation of used gas engine oils.
- PLS models have been developed for the prediction of ***TAN*** **from IR data**.
- Prediction performance has been evaluated by **repeated double cross validation (in R)**.

Gas engine oil samples

One monograde mineral oil of type SAE 40 was used in two engines running with different biogas. Oil samples were taken in intervals of 250 operating hours. When an oil has reached the end of its lifetime (typ. after 2,000 to 3,000 hours operating time), an oil change with fresh oil was carried out remaining approximately 10 % of used oil in the engine.

Engine 1: 3 new fillings, 35 samples;
Engine 2: 4 new fillings, 34 samples. Total n = 69 oil samples.

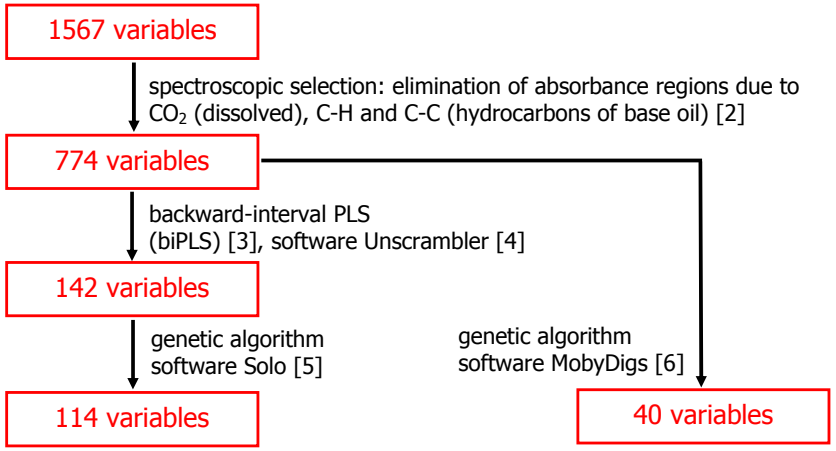
Total acid number (*TAN*)

Potentiometric titration with KOH (ASTM D 664, [1]); the oil sample was dissolved in a mixture of toluene and isopropanol.
Typical values: 1.5 - 4.2 mg KOH/g oil (y-variable).

IR spectra

Tensor 27 FTIR system (Bruker Optics, Germany), 0.1 mm ZnSe oil cell; 652 - 3672 cm^{-1}; m = 1567 data points (1567 x-variables).

Variable selection



Evaluation

The used methods for variable selection do not optimize the prediction performance for new cases (prediction of *TAN* from IR data). Therefore all variable sets have been thoroughly tested by

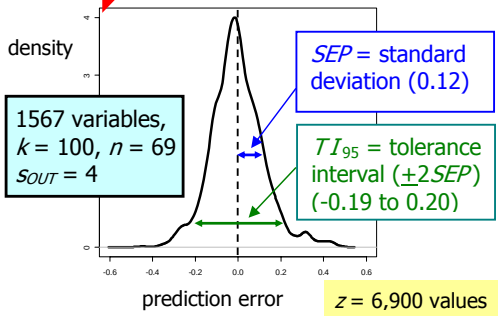
repeated double cross validation (rdCV)

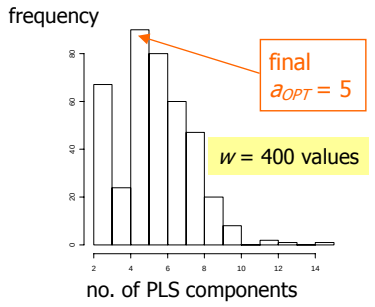
as implemented in the R-package [7] *chemometrics* [8].
The regression method was PLS as available in the R-package *pls* [9].

rdCV applies cross validation in three nested loops (1) (2) (3)

- k repetitions (typ. 100) (1)
  - Outer loop (2)
    The available n objects are randomly split into a calibration set and a test set (s_{OUT} segments)
    - Inner loop (3)
      From the calibration set an optimum number of PLS components (a_{OPT}) is estimated by CV (s_{IN} segments)
    - A model with a_{OPT} components is calculated from the calibration set, and applied to the current test set, yielding test-set-predicted $\hat{y}$.
  - We have: n test-set predicted $\hat{y}$ and s_{OUT} values for a_{OPT}

FINALLY $z = k.n$ test-set-predicted $\hat{y}$ | $w = k.\ s_{OUT}$ values for a_{OPT}



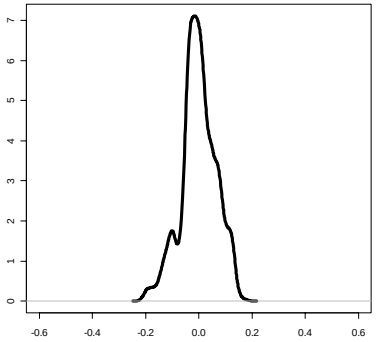


Results

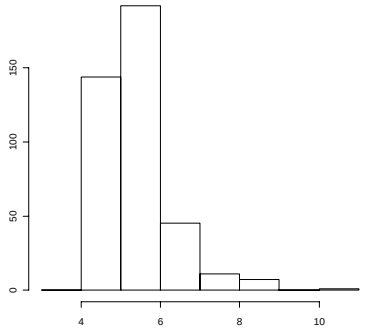
Number of variables	SEP_{TEST}	R^2	TI_{95}	a_{OPT}
1567	**0.124**	**0.9784**	**0.398**	**5**
774	**0.0896**	**0.9889**	**0.231**	**6**
142	**0.0689**	**0.9934**	**0.236**	**6**
114	**0.0663**	**0.9939**	**0.231**	**6**
40	**0.0801**	**0.9911**	**0.273**	**5**

best variable subset

Computing time for rdCV is between 20 and 200 s (40 - 1567 variables)



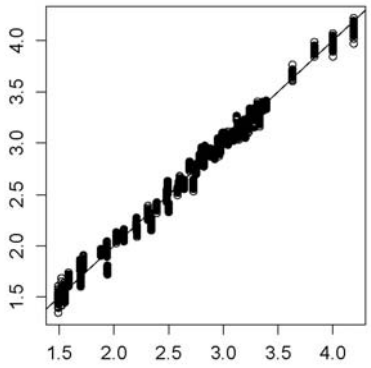
Probability density of 6,900 test-set-prediction errors (*TAN*)
SEP = 0.066 mg KOH/g oil



Histogram of 400 values obtained for optimum number of PLS components; most frequent are 6 components

test-set-predicted *TAN* (from IR data),

PLS model with 114 variables and 6 components



R^2 = 0.994
100 repetitions for 69 objects

experimental *TAN* (potentiometric titration)

Conclusions / References

- The total acid number (*TAN*) of used lubricating oil can be accurately determined from IR data by a linear regression model.
- Variable selection by a combination of a spectroscopic approach, biPLS and a Genetic Algorithm reduced the standard error of prediction (*SEP*) to 50% of the value for a model with all 1567 variables.
- Standard error of prediction (*SEP*) of the best model (114 variables) is 0.066 mg KOH/g oil, corresponding to 1.6 - 4.4% of the relevant *TAN* values.
- Repeated double cross validation (rdCV) as implemented in the R-package *chemometrics* allows a fast and comprehensive optimization of the model complexity (number of PLS components) and a realistic estimation of prediction errors for new cases.

References

[1] ASTM D 664: Standard test method for acid number of petroleum products by potentiometric titration.
[2] L. Youngbok, C. Hoeil: Spectral range optimization for the near infrared quantitative analysis of petrochemical and petroleum products: Naphtha and gasoline. *Appl. Spectrosc.* **60**, 892-897, 2006.
[3] R. Leardi, L. Norgaard: Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. *J. Chemom.* **18**, 486-497, 2004.
[4] Software "The Unscrambler", Camo Process AS, www.camo.no, Oslo, Norway.
[5] Software "Solo", Eigenvector Research Inc., www.eigenvector.com, Wenatchee, WA, USA.
[6] Software "MobyDigs", Talete srl, www.talete.mi.it, Milan, Italy.
[7] Software "R", R Development Core Team, www.r-project.org.
[8] K. Varmuza, P. Filzmoser: *Introduction to Multivariate Statistical Analysis in Chemometrics*. CRC-Press, Boca Raton, FL, USA. To be published end of 2008.
[9] B. H. Mevik, R. Wehrens: The pls package: Principal component and partial least squares regression in R. *Journal of Statistical Software*, **18**(2), 2007.

Acknowledgments. This work was funded from the "Austrian Kplus-Program" (governmental funding program for pre-competitive research) via the Austrian Research Promotion Agency (FFG) and the TecNet Capital GmbH (Province of Niederösterreich) and has been carried out within the "Austrian Center of Competence for Tribology" (AC²T research GmbH). We thank P. Filzmoser (Vienna University of Technology, Institute of Statistics and Probability Theory) for the development of the repeated double cross validation and implementation in R.