

Norris Regression and Wavelength Selection with Applications to Molecular Spectroscopic Data

Abstract

This paper proposes new approaches to improve Norris Regression and Searching Combination Moving Window Partial Least Squares (SCMWPLS). Reducing calculation time and limiting user interaction during wavelength selection are issues that are under investigation. For Norris Regression, programming for automatic search and optimization is provided for the determination of the numerator and denominator positions, gap size of Norris Derivative and smoothing segment size. To modify SCMWPLS, several issues are discussed, such as determining moving window size by bandwidth, screening wavelength region candidates by RMSEC and PC number, and searching wavelength region combination by spectral intrinsic factors.

Introduction

Norris Regression and SCMWPLS try to build better regression models than those based on whole spectra.

Norris Regression [1]

The "Norris Regression" is a regression procedure to reduce the impacts of varying baseline, variable path lengths, and high stray lights due to scatter effects. The linear regression (LR) model is established between a Y and the ratio of two selected wavelengths. Moving average smoothing and Norris Gap Derivative at optimum settings by **searching** are usually applied onto the numerator and denominator positions.

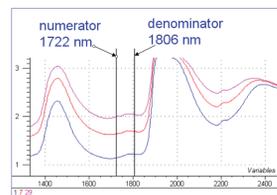


Fig 1. Typical MILK spectra (1 mm pathlength).

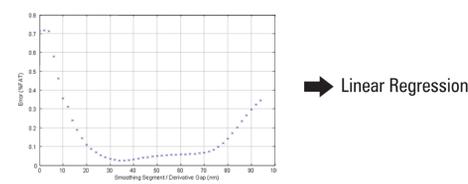


Fig 2. Optimizing smoothing and Derivative settings for numerator and denominator (4 plots total).

SCMWPLS [4]

Searching combination moving window PLS is a method for searching for either the optimized combination of informative regions or an optimized individual informative region by Moving Window PLS (MWPLS).

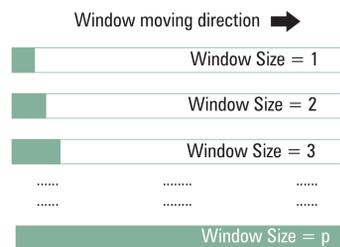


Fig 3. Scheme for explanation of MWPLS

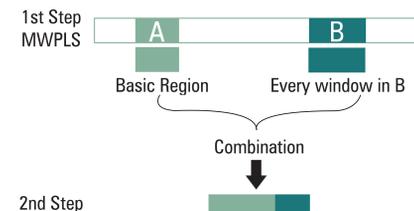


Fig 4. Scheme for explanation of SCMWPLS

Material and Methods

Applications are home-built programs running in matlab v6.5, and then confirmed in The Unscrambler v9.6. A two-stage approach is proposed for Norris Regression as shown in Fig 5: searching numerator and denominator, and optimizing derivative gap and smoothing segment. A total of 105 milk NIR diffuse transmittance measurements and corresponding %Fat references were kindly provided by Dr. Karl Norris.

A total of 60 beer NIR measurements and Y-references (unknown identity) were downloaded from KVL web site [3].

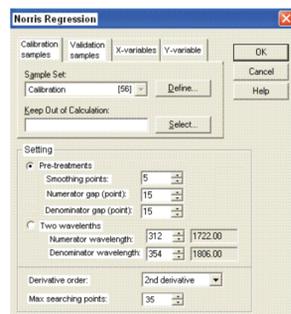


Fig 5. Norris Regression User Interface.

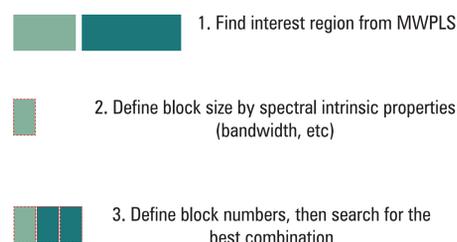


Fig 6. Proposed SCMWPLS approach

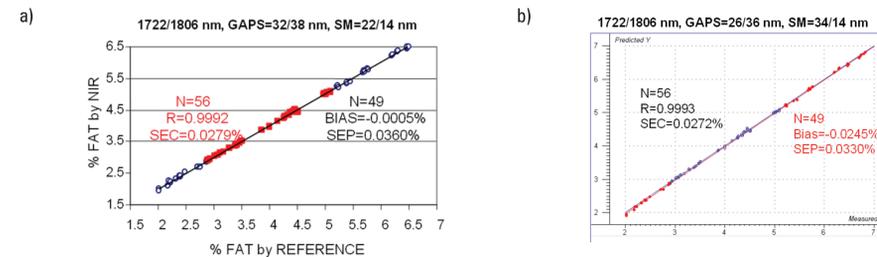


Fig 7. Comparison of Norris Regression results of MILK data: a) From Dr. Norris; b) 2-stage approach shown in Fig 5. Numerator and denominator positions are the same. There are small deviations in optimum Gap and Smoothing. Calculation takes about 25 minutes.

Table I. Comparison of SEC and SEP among 4 methods using the MILK and BEER datasets. SCMWPLS is sensitive to the size of wavelength blocks shown in Fig 6. Some of selected wavelengths show in 2 and more methods, but not always true.

Data Pre-treatment	(Full Range)	Norris Regression	MWPLS	SCMWPLS (3 Blocks)			SMLR
MILK Data (700 pts)	Parameter	numerator/denominator: 1722/1806 nm, Gap: 28/36 nm, SM: 34/14 nm	1768-1810 nm	Block width: 4 nm 1190-1194, 1276-1730, 1738-1742 nm	Block width: 8 nm 1376-1386, 1768-1776, 1788-1796 nm	Block width: 16 nm 1370-1386, 1764-1780, 1818-1834 nm	1732, 1190, 1598, 1898, 1638, 2084, 1188 nm
	SEC (0.0274)	0.0272	0.0758	0.0576	0.0548	0.0572	0.0722
	SEP (0.2918)	0.0330	0.1247	0.0751	0.0856	0.0975	0.0550
	Calculation Time (sec)	1490	10	95	15	2	10
BEER Data (926 pts)	Parameter	numerator/denominator: 1152/1244 nm, Gap: 2/16 nm, SM: 46/14 nm	1232-1272 nm	Block width: 8 nm 958-956, 1208-1216, 1228-1236 nm	Block width: 12 nm 922-934, 958-962, 1224-1236 nm	Block width: 16 nm 1160-1176, 1232-1248, 1268-1284 nm	1270, 1188, 980, 842, 930, 1430 nm
	SEC (0.2130)	0.1410	0.1956	0.1485	0.0998	0.1211	0.175
	SEP (2.1252)	0.0717	0.1256	0.0892	0.0668	0.1119	0.0697
	Calculation Time (sec)	2060	11	225	57	25	9

Conclusions

- Searching criteria: SEC vs. SEP vs. SEC+SEP vs. others? We prefer minimizing $1.5 * SEC + SEP$.
- Calculation time: Exhaustive search vs. stepwise? It depends. We prefer stepwise.
- User Interaction: As little as possible for commercial software.
- Norris regression: One of the best quantitative approaches for Diffuse Transmittance measurements in which low transmittance and unknown or unstable pathlength are the problems. Application rules from Dr. Norris.Spectrometer must have excellent wavelength stability. Must have very low errors in the reference data. Must have defined absorption band for numerator. Avoid denominators which approach zero. Avoid wavelengths with sharp peaks in correlation.
- SCMWPLS: Wavelength block width and number of blocks in search combination are important.
- Calibration samples should be representative; Validation samples should be true replicates.
- The Unscrambler for sample outlier detection.

Acknowledgements

Karl Norris for Norris Regression tutorial and MILK dataset
Lars Nørgaard for free iToolbox matlab program and BEER dataset
Fred McClure for SMLR communication
Dongsheng Bu & Valérie Lengard CAMO Software AS, Nedre Vollgate, 8, NO-0158 Oslo, Norway
Suresh Kumar B.V - CAMO Software India Pvt. Ltd., Domlur Layout, Bangalore, India

References

1. Norris K, NIR news Vol.9(4) (1998) 3, Vol.12(3) (2001) 6, and Vol.13(3) (2002) 8, 2. Hopkins DW, NIR news Vol.14(5) (2002) 10
3. Nørgaard L <http://www.models.kvl.dk/source/iToolbox/index.asp>, 4. Jiang JH, NIR news Vol. 16(5) (2005) 10



CAMO Software India Pvt. Ltd.
14 & 15, Krishna Reddy Colony
Domlur Layout, Bangalore - 560 071
INDIA

Tel: +91 (80) 4125 4242
Fax: +91 (80) 4125 4181



The Unscrambler®