List of Papers

This thesis is based on the papers listed below. They are referred to in the text by the corresponding Roman numeral (I-IV).

I.) PLS Regression on Wavelet Compressed NIR Spectra
J. Trygg and S. Wold
Chemometrics Intell. Lab. Syst. **42** (1998) 209-220

II.) 2D wavelet analysis and compression of on-line industrial process data
J. Trygg, N. Kettaneh, L. Wallbäcks
J. Chemometr. **15** (2001) 299-319

III.) Orthogonal projections to latent structures, O-PLS
J. Trygg, S. Wold
J. Chemometr. **16** 3 (2002)119-128

IV.) O2-PLS for qualitative and quantitative analysis in multivariate calibration
J. Trygg
J. Chemometr. (2002), accepted

Abbreviations

CLS	Classical least squares
DModW	Distance to model in wavelet space
DModX	Distance to model in X space
DOSC	Direct orthogonal signal correction
FIR	Finite impulse response
FT	Fourier transform
GT	Gabor transform
IWT	Inverse wavelet transform
MLR	Multiple linear regression
MVA	Multivariate analysis
NIPALS	Non-linear iterative partial least squares
NIR	Near infrared reflectance
O2-PLS	Modified O-PLS, hybrid OSC+CLS&PLS method
OLS	Ordinary least squares
O-PLS	Orthogonal-PLS, hybrid OSC+PLS method
OSC	Orthogonal signal correction
PC	Principal component
PCA	Principal component analysis
PCR	Principal component regression
PLS	Partial least squares projections to latent structures
PLS-DA	Partial least squares discriminant analysis
PRESS	Prediction error sum of squares
RMSEP	Root mean squared error of prediction
SIMCA	Soft independent modelling of class analogy
STFT	Short time Fourier transform
SURE	Stein's unbiased risk estimate
UV	Ultra Violet
WPT	Wavelet packet transform
WT	Wavelet transform

Notations

The following notation has been used throughout the thesis.
Vectors are denoted by bold, lower-case letters, e.g. **t**, and matrices by bold capital letters, e.g. **X**.
A matrix inverse is denoted $\mathbf{X}^{-1}$. Vectors are assumed to be column vectors unless indicated by a transposition, e.g. $\mathbf{t}^{T}$.

X	Matrix of predictor variables, [N×K]
Y	Matrix of response variables, [N×M]
E	Residual matrix of predictor variables, [N×K]
F	Residual matrix of response variables, [N×M]
B	Matrix of regression coefficients for **X**, [K×M]
W	Matrix of weight vectors for **X**, [K×A]
C	Matrix of weight vectors for **Y**, [M×A]
T	Matrix of score vectors for **X**, [N×A]
P	Matrix of loading vectors for **X**, [K×A]
U	Matrix of score vectors for **Y**, [N×A]
p	Loading vector for **X**, [K×1]
w	Weight vector for **X**, [K×1]
b	Regression vector for X, [Nx1]
t	Score vector for **X**, [Nx1]
c	Weight vector for **Y**, [M×1]
u	Score vector for **Y** [N×1]
A	Number of components in model (not incl. OSC comp.)
D	Number of Y-orthogonal comp. in **X**
Dy	Number of X-orthogonal comp. in **Y**
K	Number of columns in **X**
M	Number of columns in **Y**
N	Number of rows in **X** and **Y**
k	Column index for **X** (k=1,…K)
m	Column index for **Y** (m=1,…M)
n	Row index for **X** and **Y** (n=1,…N)
a	Model dimension index (a=1,…A)
z	Threshold value in wavelet denoising

Summary of Papers I-IV

Paper I

This paper describes how the wavelet transform can be used as an effective compression tool for near-infrared reflectance (NIR) spectra used in multivariate calibration. NIR spectra naturally contain a great deal of redundant data, and are, therefore, suitable for compression. NIR-VIS spectra in the wavelength region 400-2500 nm were collected from 227 different sheets of cellulose-derivatives. Viscosity of the sheets was measured using a standard reference method. Measuring viscosity is both expensive and time-consuming, and NIR spectroscopy has proven useful in providing a fast, cheap and non-invasive assessment of viscosity. The wavelet compressed PLS model was almost identical to the uncompressed PLS model. The reconstructed loadings of the wavelet compressed PLS model were very similar to the uncompressed PLS loadings. The compression ratio of 30:1 was impressive.

Paper II

Two-dimensional wavelet analysis and compression of NIR spectra for the on-line monitoring of wood chips are reviewed. This paper was part of a 1997 project investigating the use of on-line NIR spectroscopy for monitoring wood chips at ASSI Domän plan in Piteå, Sweden. A new parameter for outlier detection is introduced, distance to model in wavelet space (DModW). In addition, the wavelet power spectrum (WPS), the wavelet analogue of the power spectrum is described. The WPS provides an overview of the time-frequency content of a signal. In the example given, wavelets improved the detection of spectral shift and compressed data about 1000 times without adversely affecting the quality of the 2-D wavelet compressed PCA model.

Paper III

Data acquired from the measurement of complicated samples and complicated processes contain contributions from many sources, as well as several types of noise. Pre-processing methods can be applied in such situations to enhance the relevant information while decreasing or eliminating any structured noise in the data. OSC filters remove structured variation from **X** (e.g. spectral data) that is not related (Y-orthogonal) to **Y** (e.g. yield, cost or toxicity). Partial least squares projections to latent structures (PLS) is, frequently, applied to the OSC-filtered **X** matrix. In Paper III, we describe our development of an OSC filter specifically designed for PLS. The two methods were combined in a new hybrid regression method called Orthogonal-PLS (O-PLS). The O-PLS model removes the structured

Y-orthogonal variation from **X** and thereby increases the quality of the predictive model and makes the Y-orthogonal variation easier to interpret. Predictions are similar to those produced using the standard PLS model.

Paper IV

In Paper IV, we describe our development of an OSC filter specifically designed for classical least squares (CLS) combined with PLS. The new hybrid method is called O2-PLS. Compared with O-PLS, O2-PLS further increased ease of interpretation by providing estimates of: the pure constituent profiles in **X**; the Y-orthogonal variation in **X**; the X-orthogonal variation in **Y**; and the joint X-Y co-variation. It is also predictive in both directions, **X**✍ ✍**Y**. In real examples and in simulations, the O2-PLS method improved the model's interpretability and provided a good estimate of the pure constituent profiles in **X**. In addition, the predictive ability was similar to that of the standard PLS model.