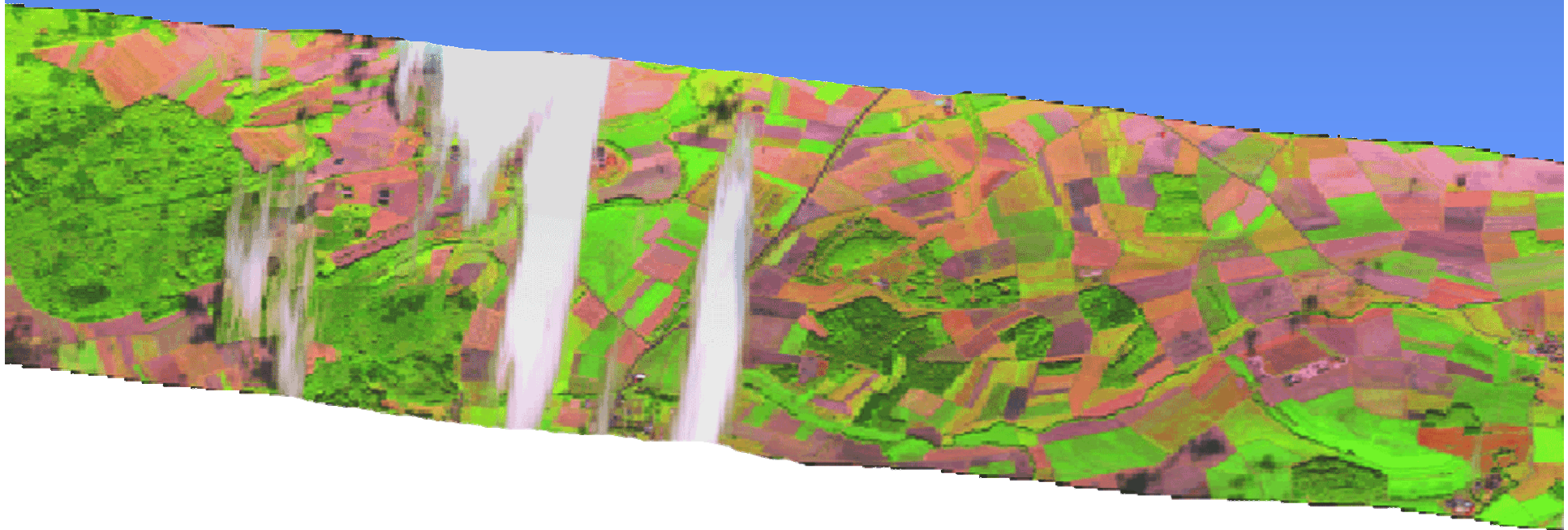


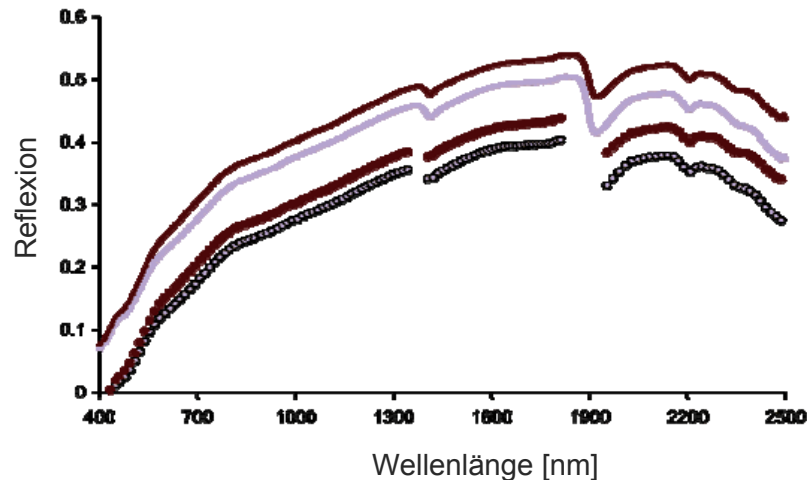
Charakterisierung von organischer Bodensubstanz mit (abbildender) Spektroskopie



Michael Vohland, Marie Eggemann, Christoph Emmerling, Sören Thiele-Bruhn



Reflexionssignal von Böden ist ein komplexes Mischsignal, das hervorgerufen wird aus einer heterogenen Mischung bestehend aus ...



- *Bodenmineralen*
- *sekundären Verwitterungsprodukten*
- *OBS*
- *bodenphysikalische Größen (Bodenart, Rauigkeit der Bodenoberfläche)*

→ Maskierungseffekte, Überlagerung der unterschiedlichen spektralen Fingerabdrücke,

Zur Schätzung bodenchemischer Größen kommen lediglich empirische Methoden in Frage, die den jeweiligen spektralen Vorhersagemechanismus nutzen; auf den Bodenchemismus abzielende und allgemeingültige Reflexionsmodelle existieren nicht.

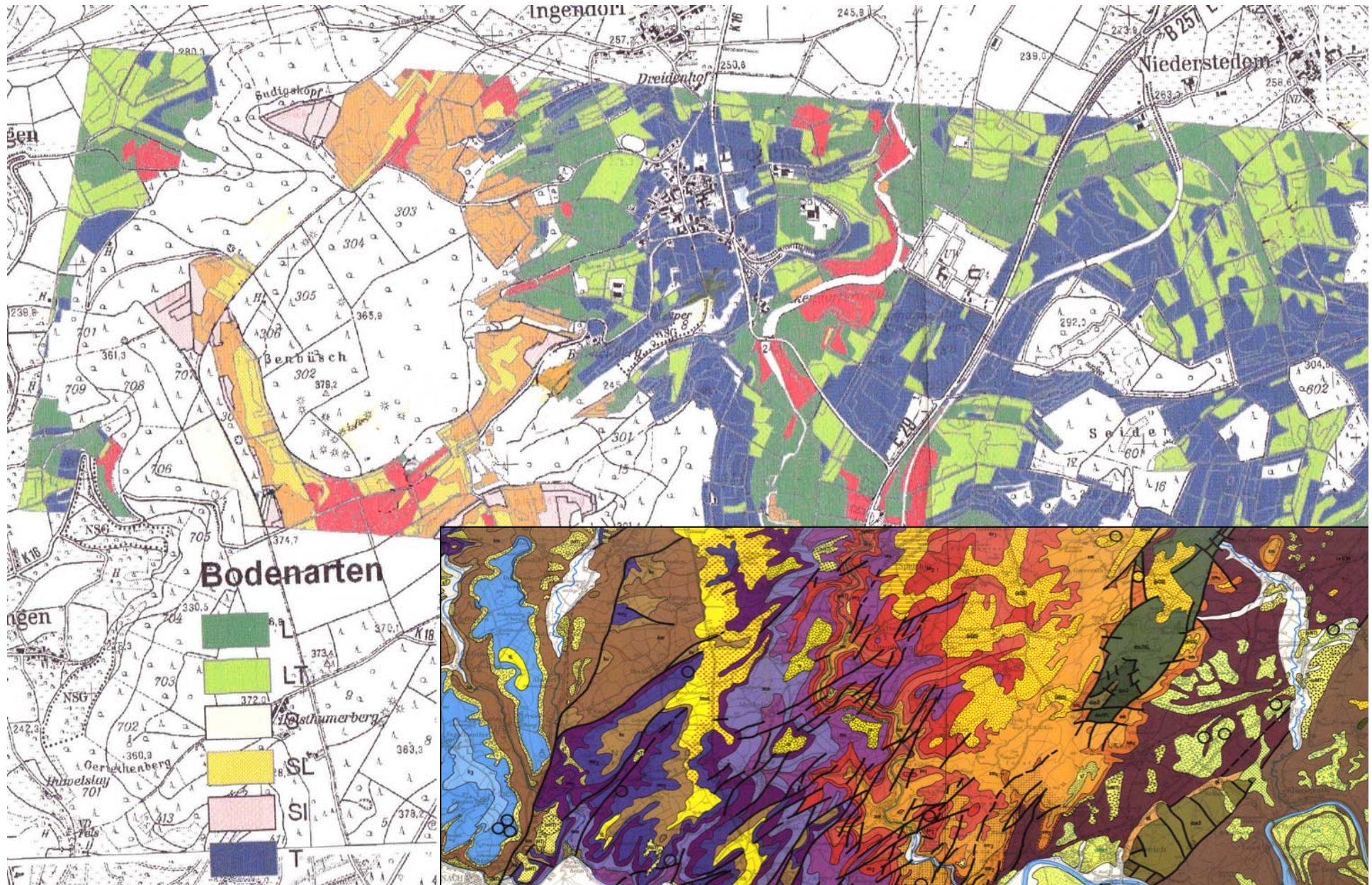


*Spectral predictive mechanisms may vary from one population to another depending on the decomposition stage of organic matter, the nature of existing compounds, and the influence of other relevant factors like texture or soil colour. This variability is the main reason that researchers often tend to develop only **local calibration models for each field they measure with NIR spectroscopy** (Ben-Dor & Banin 1995, Mouazen et al. 2007).*

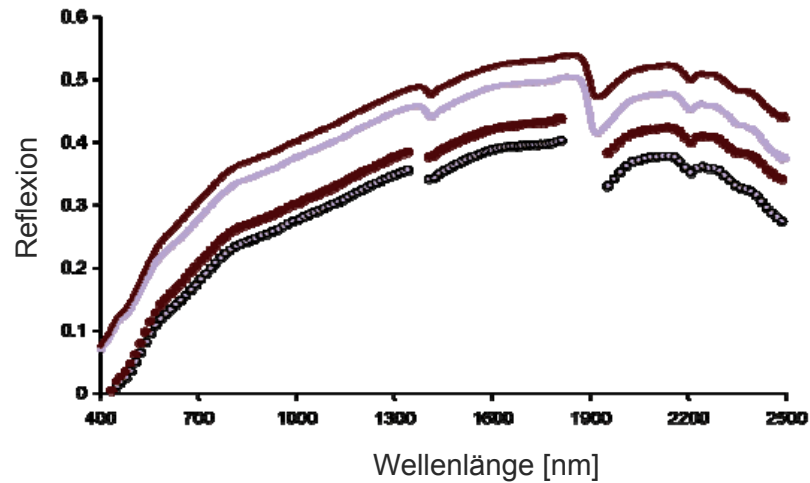
Auf VISNIR-Spektroskopie basierende Schätzmodelle auf der Landschaftsskala können nicht die Güte von Schätzmodellen auf Feldskala erreichen. Optimierungsmöglichkeiten bestehen auf Seiten der

- Datenqualität und Samplingstrategie
- Vorprozessierung der Daten (Originaldaten, scatter corrections, Ableitungsspektren)
- multivariaten Kalibrations-Werkzeuge (PCA, PLS, SVM-R)
- Techniken der Selektion von Spektralvariablen (s. Multikollinearität, Rauschen) - UVE, GA, stepwise forward/backward selection

HyMap Kampagne 27.08.2009 – Trier-Bitburger Mulde (Bitburger Gutland)



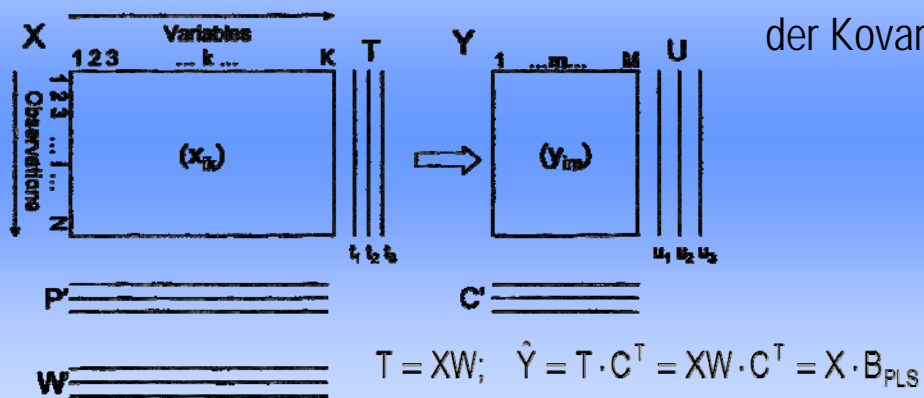
Laborspektroskopie (ASD FieldSpec-II Pro FR)



- n = 60, davon 42 sampling points auch in HyMap Daten auswertbar
- Messkonfiguration: 30° Beleuchtungszenit, Nadirsicht, Öffnungswinkel des Sensors 23°
- Zielgrößen: Corg, Cmic, Chwl, N, Nmic
- Probenset Kalibrierung: n = 42; Validierung: n= 18

Das Standardwerkzeug...

Partial Least Squares (PLS) Regression



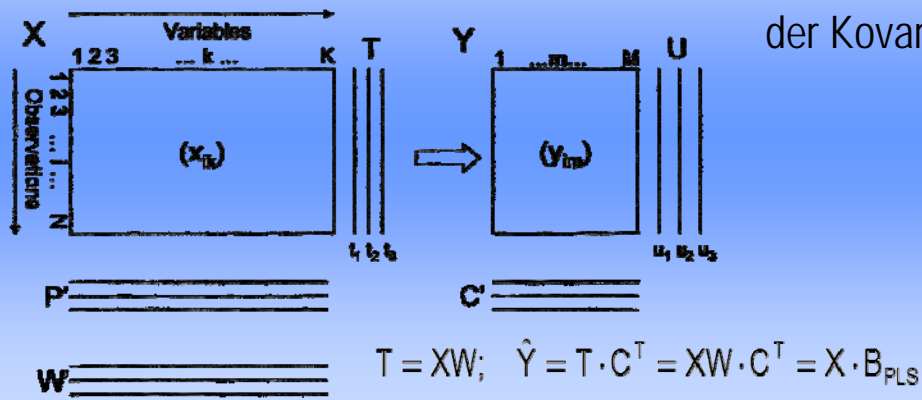
neue latente Variablen werden gesucht, die möglichst viel der Kovarianz zwischen X und Y erklären

„ability to analyze data with many, noisy, and collinear variables“ (Wold et al. 2001)

„leads to stable, correct, and highly predictive models even for correlated descriptors“ (Eriksson et al. 2001)

Partial Least Squares (PLS) Regression

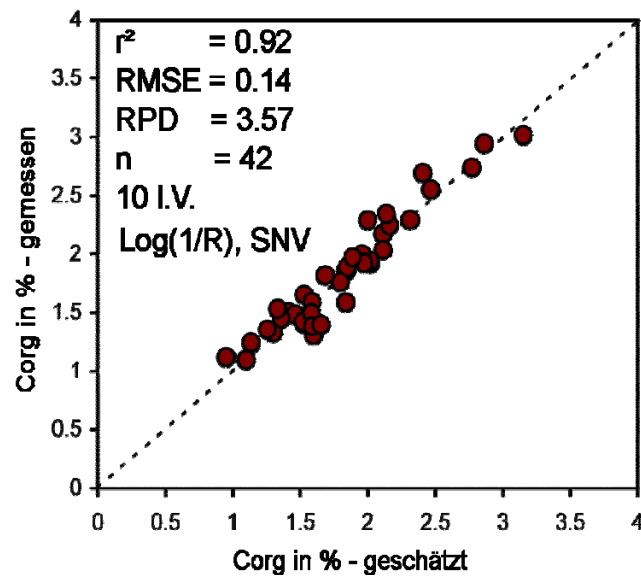
neue latente Variablen werden gesucht, die möglichst viel der Kovarianz zwischen X und Y erklären



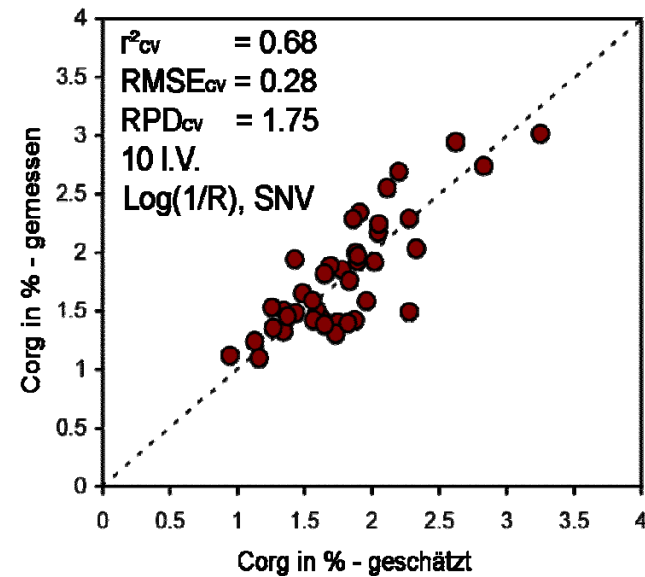
„ability to analyze data with many, noisy, and collinear variables“ (Wold et al. 2001)

„leads to stable, correct, and highly predictive models even for correlated descriptors“ (Eriksson et al. 2001)

Kalibrierung



Kreuzvalidierung



Kalibrierung SVM-R

INPUT

PROZESS

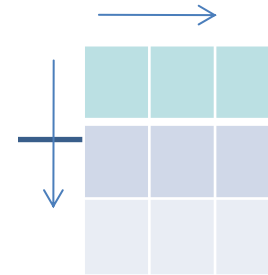
OUTPUT

Train
Test



Datenskalierung

Parameterselektion
(grid search)



Train



SVM-R CV

MSE (Modell_subopt.)

Parameter_opt. (MSE_min)

Train

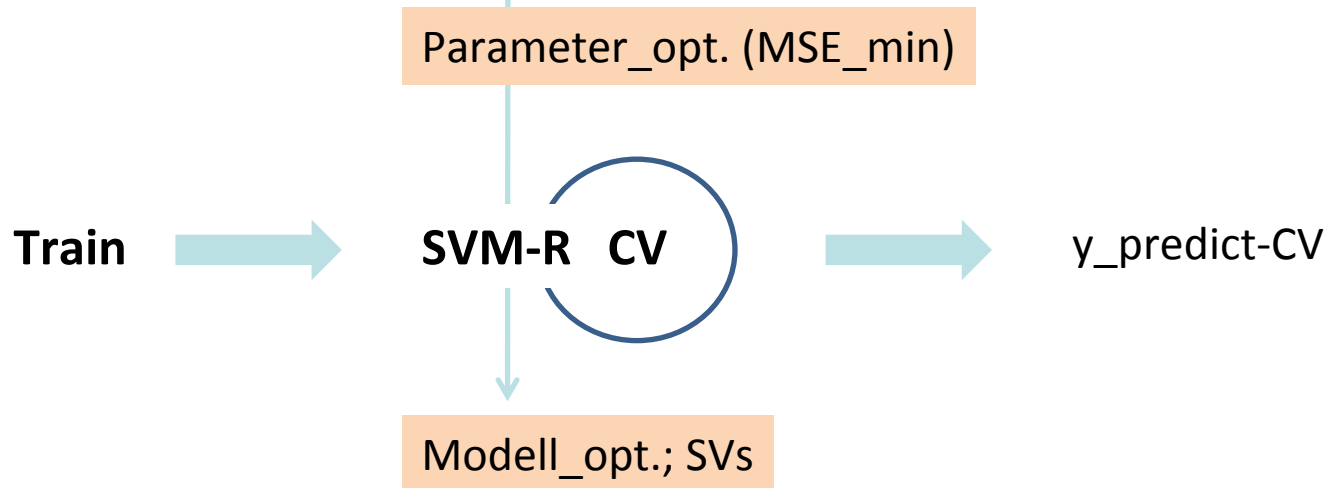
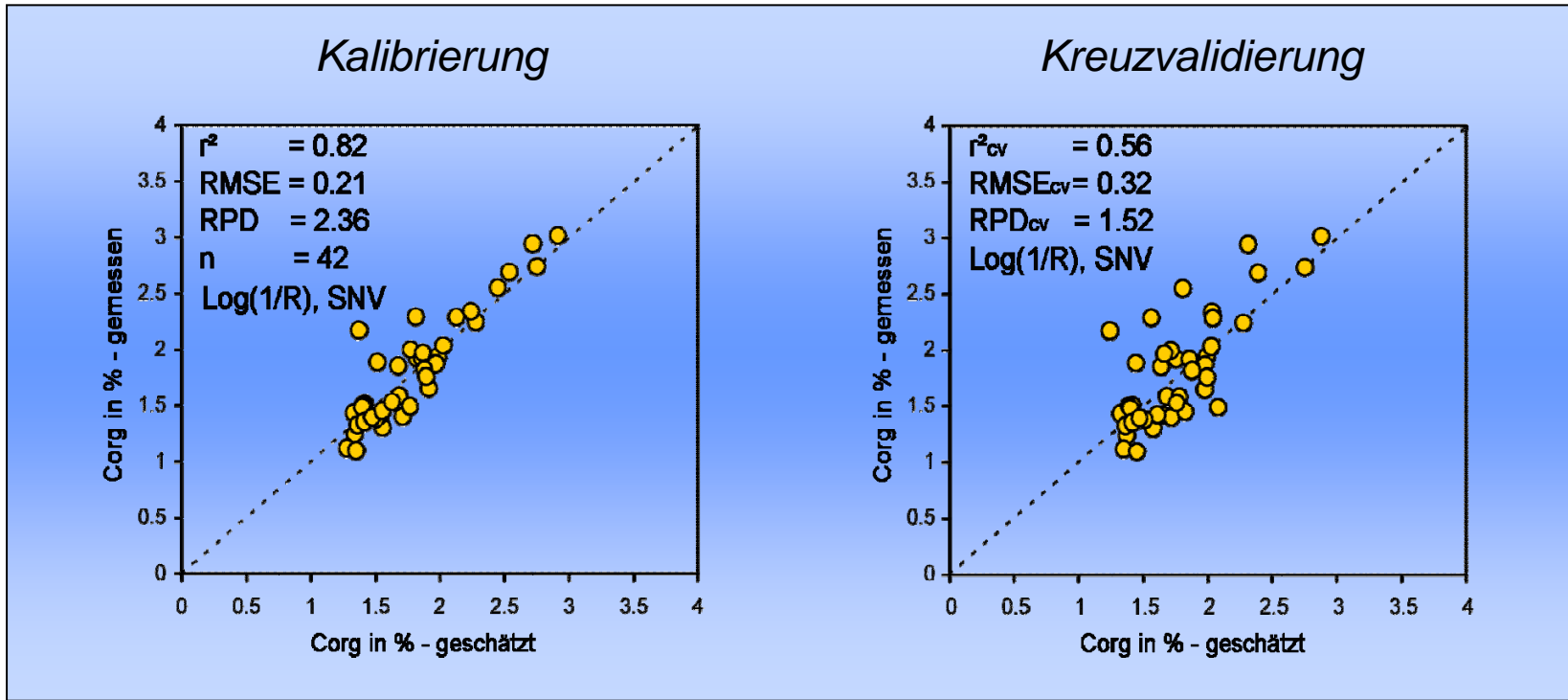


SVM-R CV

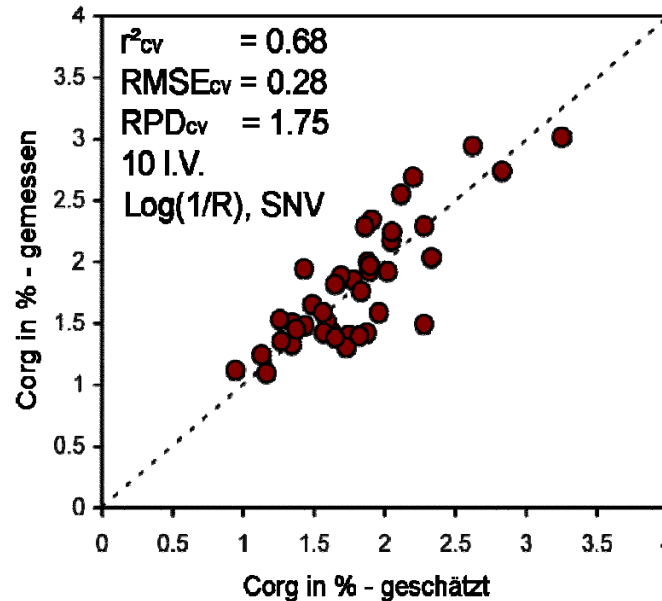
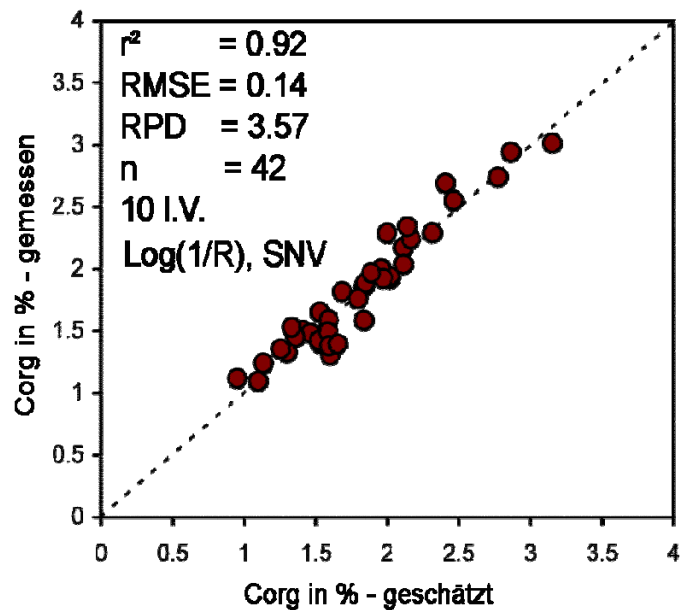
y_predict-CV

Modell_opt.; SVs

Kalibrierung SVM-R



Partial Least Squares (PLS) Regression



PLS ist rauschempfindlich!

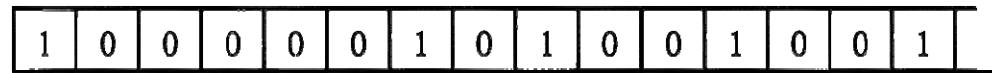
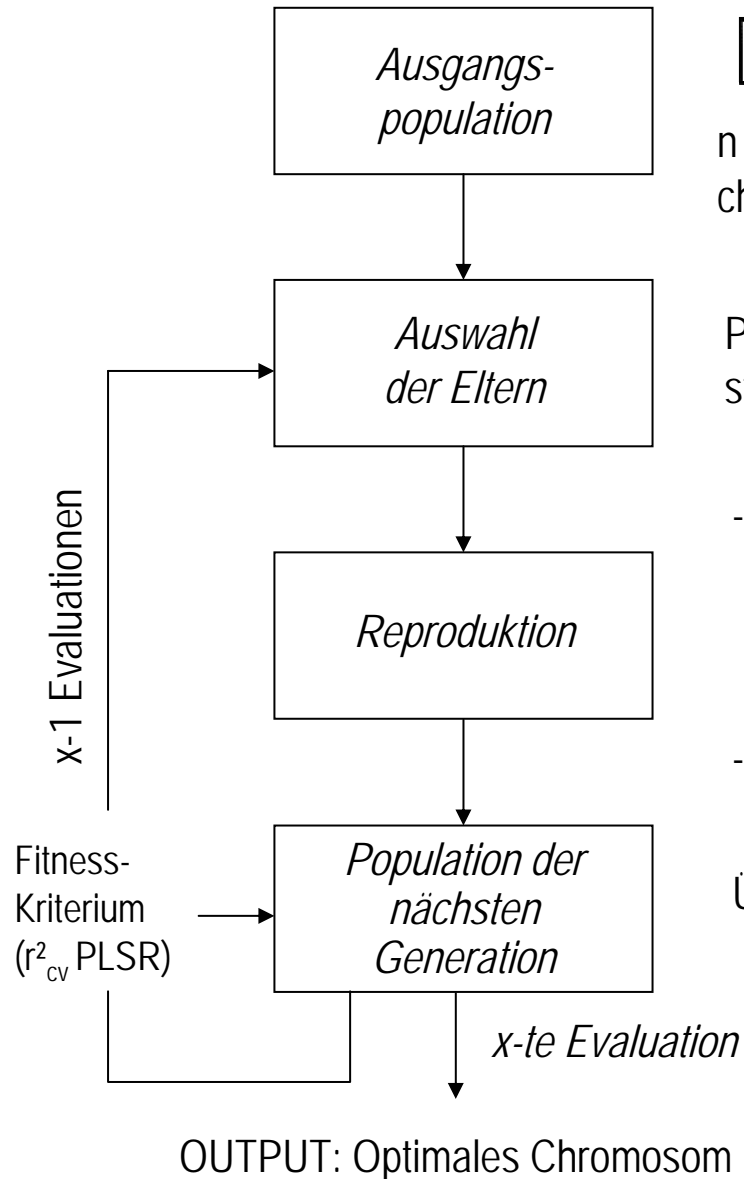
Lässt sich über eine Auswahl der geeignetsten Variablen („feature selection“) die Robustheit des Modells erhöhen?

Fehlende a priori Kenntnisse über die Zusammensetzung von C_{org} und das Zusammenspiel mit anderen Komponenten /Faktoren im Hinblick auf spezifische Reflexionsmuster führen zur Wahl eines **stochastischen Suchverfahrens (Genetischer Algorithmus)**.

Ähnlich der biologischen Evolution wird dabei eine Menge von Lösungskandidaten (Population) zufällig erzeugt und im Folgenden, über eine Vielzahl von Generationen, werden letztendlich diejenigen selektiert, die ein gewähltes Gütekriterium (r^2_{cv}) am besten erfüllen.

Genetische Algorithmen

Inspiriert von der Evolutionstheorie

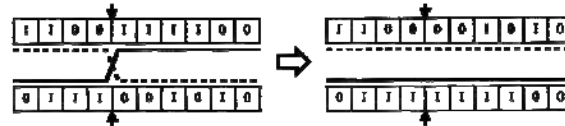


n Chromosomen mit 421 bzw. 126 Genen, Codierung durch binäre Zeichenketten („on“, „off“ → Variable „wird benutzt“, „wird nicht benutzt“)

Probabilistische Elternselektion
stochastisch universelle vs. fitnessproportionale Selektion

- Rekombination, „crossover“ als primärer Operator

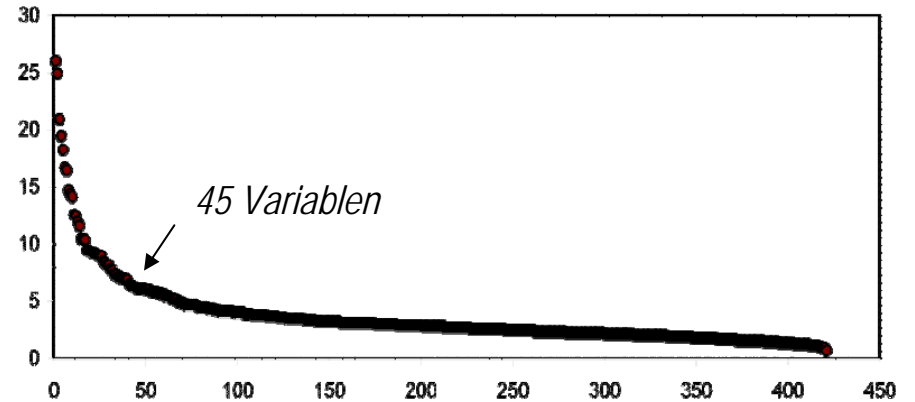
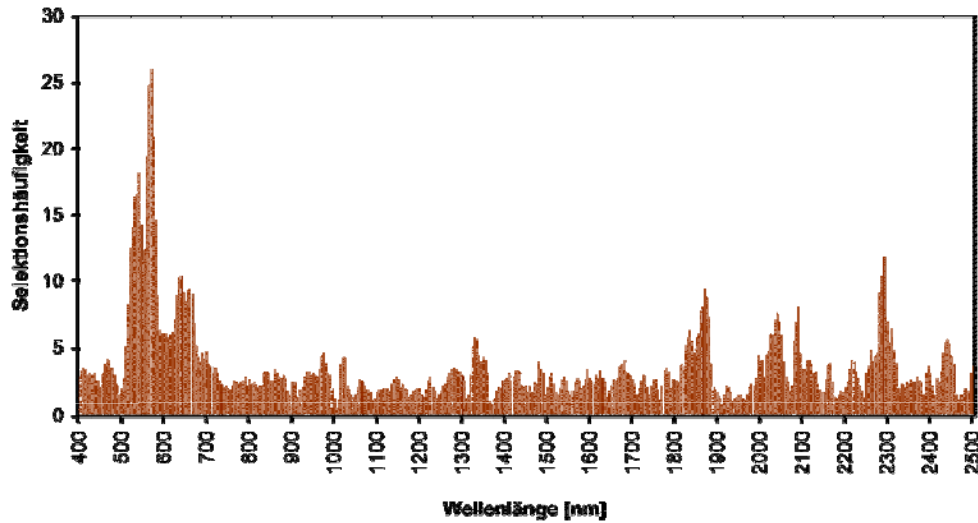
„single crossover“ vs. „uniform crossover“



- Mutation: Hintergrundoperator, geringe Mutationswahrscheinlichkeit (garantiert die Erreichbarkeit aller Punkte im Suchraum)

Überlappung der Eltern- und Kindergeneration erlaubt?

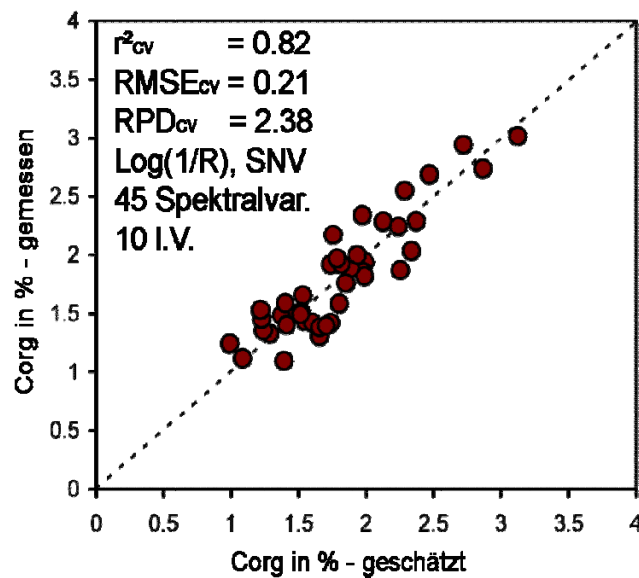
Ergebnisse GA-PLSR - C_{org}



Kalibration

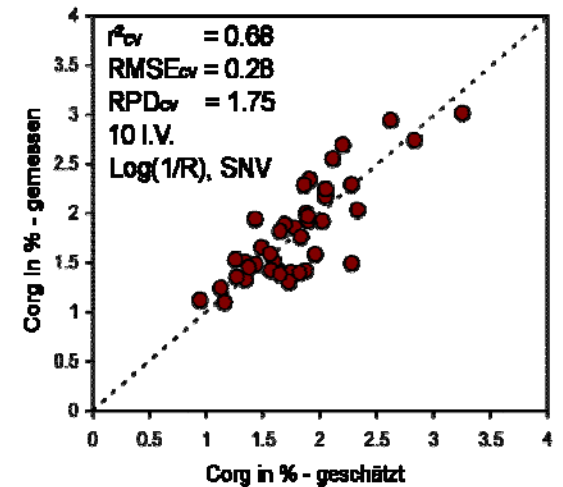
r^2 = 0.93
 RMSE = 0.13
 RPD = 3.89

Kreuzvalidierung

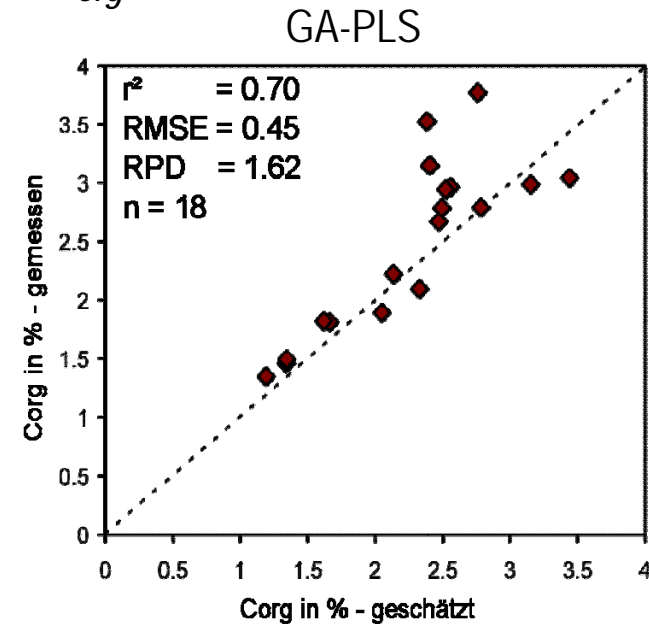
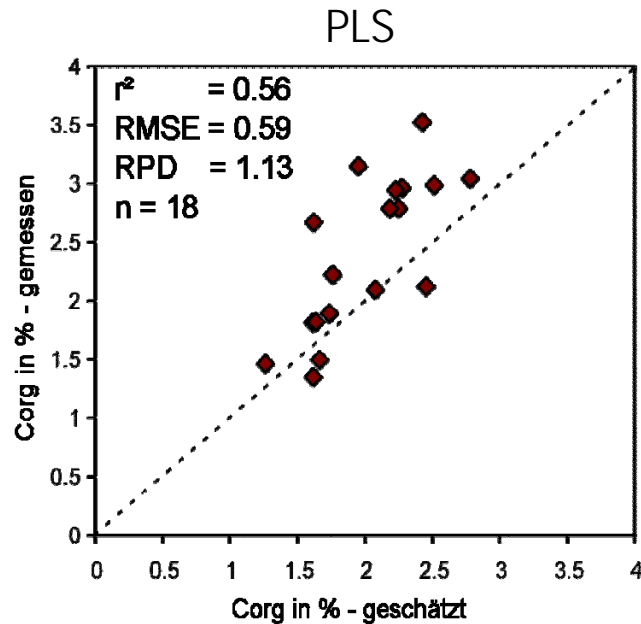


PLS (ohne GA), Kreuzvalidierung

>



Validierung - C_{org}



C_{hwl} (heißwasserlösliche C-Fraktion)

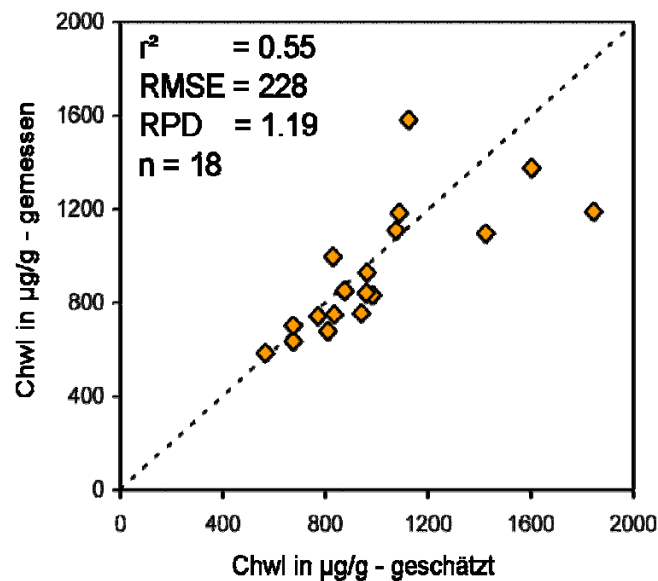
- Potenzial an sehr leicht umsetzbarer OBS mit schnellen Turnover Raten, reflektiert stärker als C_{org} die Bewirtschaftungsmaßnahmen (Düngung)
- enge Beziehungen zur *mikrobiellen Biomasse* im Boden und deren Aktivitäten
- Indikator des OBS-Versorgungszustandes landwirtschaftlich genutzter Böden

Quelle: Schulz, E. (2005), Hoyer, U. et al. (2007)

C_{hwI}

	PLS		GA-PLS	
	cv	val	cv	val
C_{org}				
r^2	0.68	0.56	0.82	0.70
pRMSE	0.15	0.25	0.11	0.18
RPD	1.75	1.13	2.38	1.62
C_{hwI}				
r^2	0.62	0.56	0.80	0.55
pRMSE	0.17	0.25	0.12	0.24
RPD	1.63	1.18	2.27	1.19

GA-PLS, Validierung, 18 samples



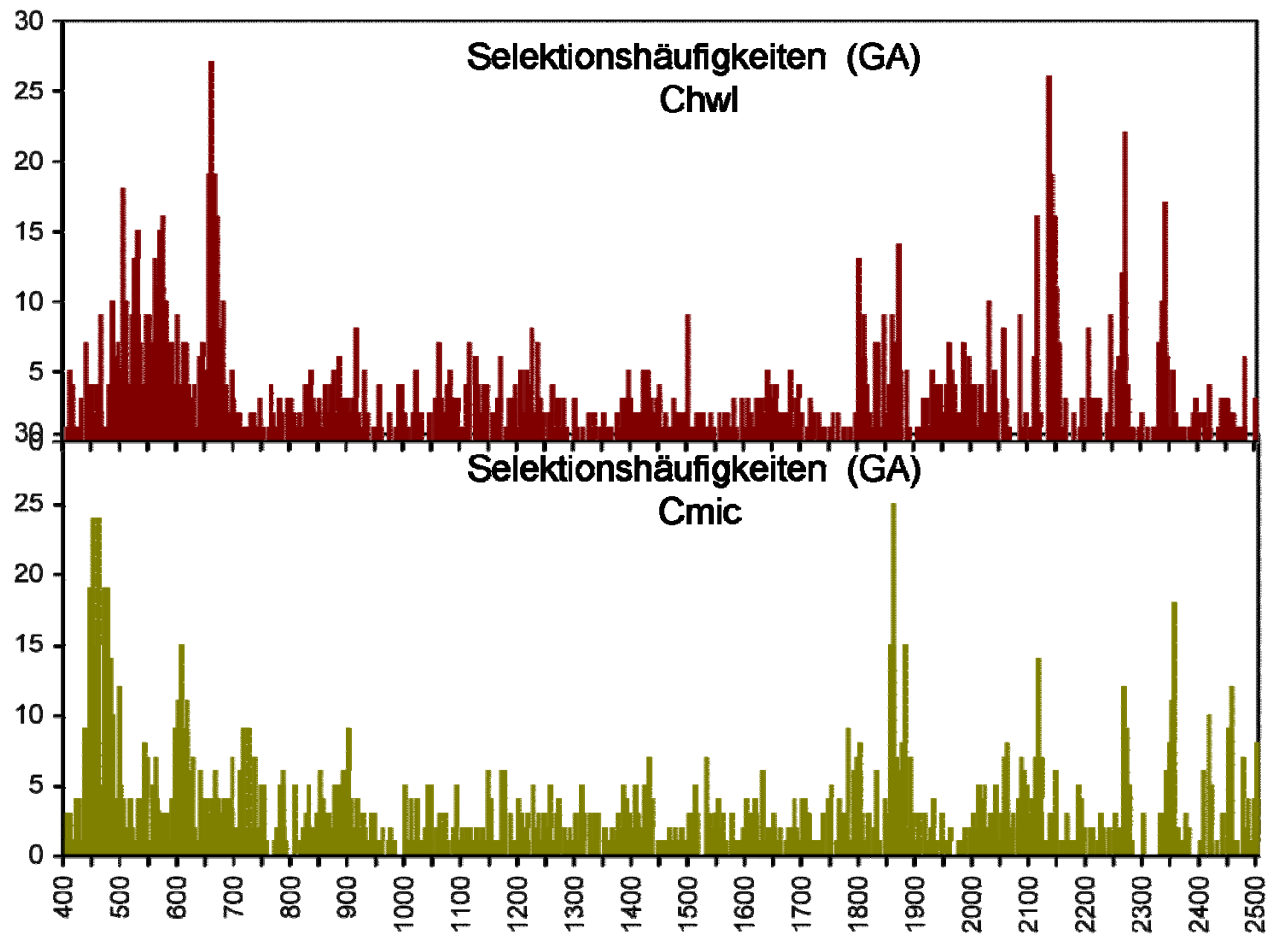
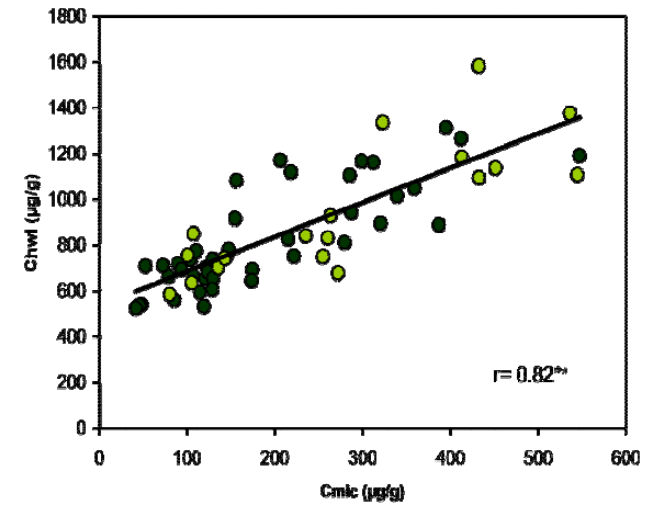
Vergleichsstudien:

$$C_{org} > C_{hwI}$$

Chodak, M. et al. (2003)
Vohland & Emmerling (2010)

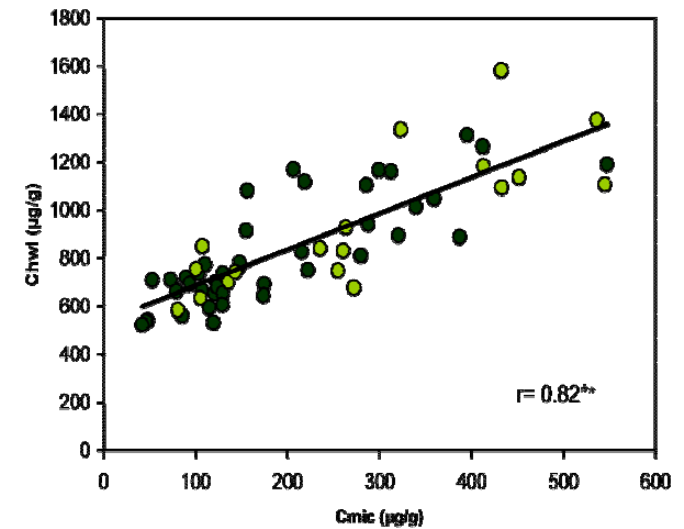
$$C_{mic}$$

C_{mic} ist ein Indikator für Veränderungen der OBS, bevor sie sich im Corg-Gehalt erkennen lassen. Der Quotient C_{mic}/C_{org} zeigt an, ob das Substrat das Wachstum der Mikroorganismen fördert und kann daher als Indikator für die Humusqualität gesehen werden.

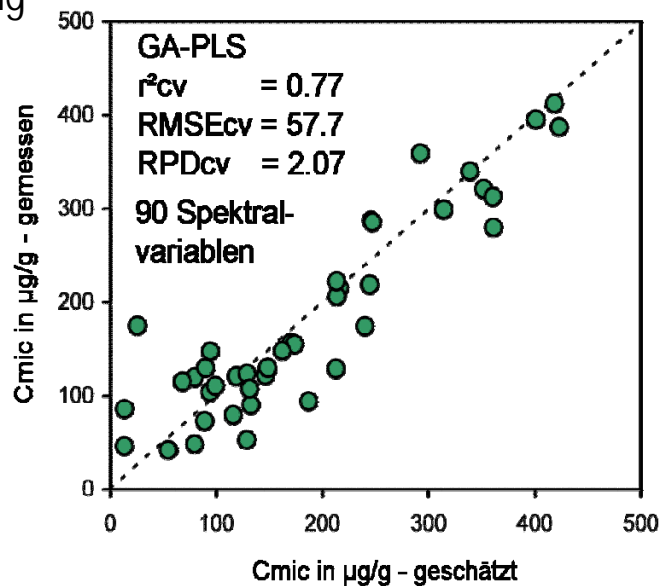


C_{mic}

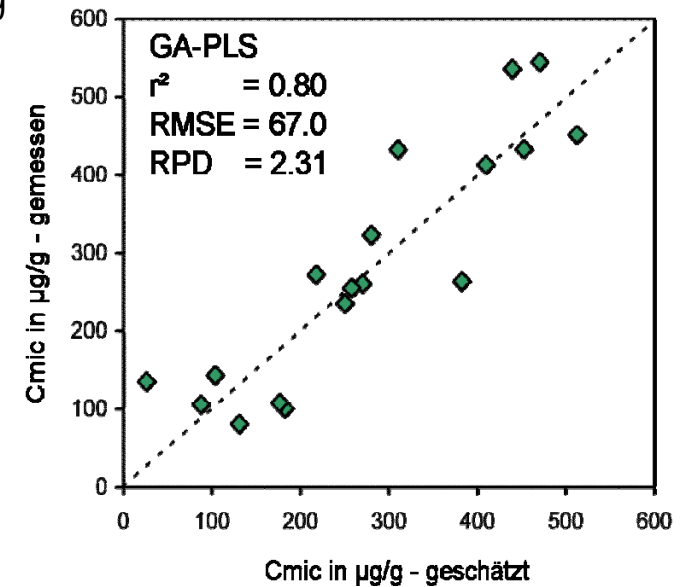
	PLS		GA-PLS	
	cv	val	cv	val
C_{hw1}				
r^2	0.62	0.56	0.80	0.55
pRMSE	0.17	0.25	0.12	0.24
RPD	1.63	1.18	2.27	1.19
C_{mic}				
r^2	0.67	0.82	0.77	0.80
pRMSE	0.37	0.23	0.30	0.24
RPD	1.69	2.33	2.07	2.31



Kreuzvalidierung



Validierung



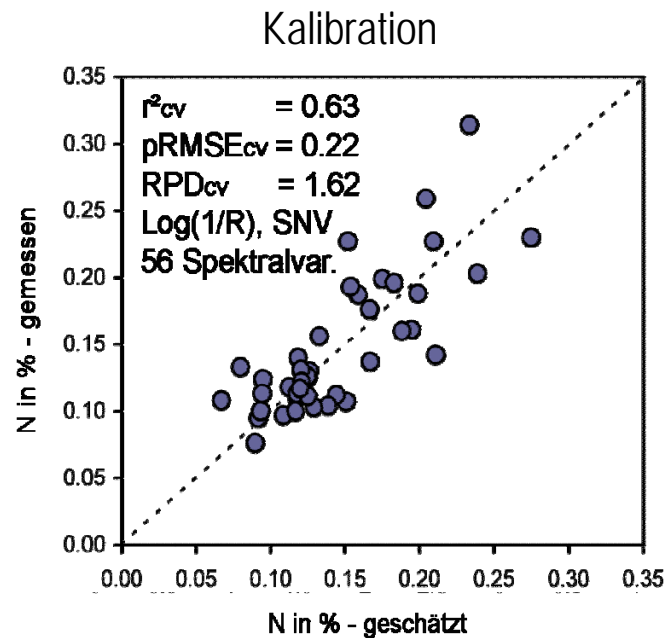
N und N_{mic}

C/N – Verhältnis: Bewertung der Bodenfruchtbarkeit, Stickstoffverfügbarkeit, Humusform bzw. Humifizierungsgrad

N_{mic} : Humifizierungsgrad, freigesetzter Stickstoff wird in die Biomasse eingebaut

GA-PLS

N:

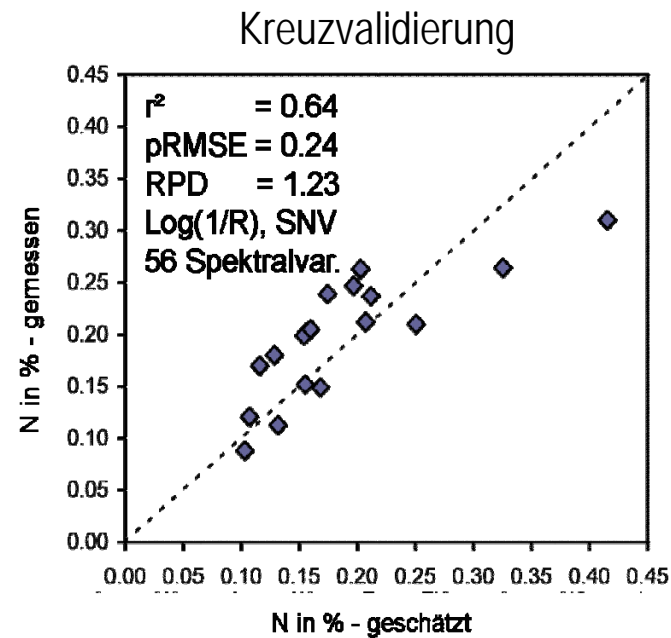


Vergleichswerte PLS:

$$r^2_{cv} = 0.50$$

$$pRMSE_{cv} = 0.26$$

$$RPD_{cv} = 1.34$$



Vergleichswerte PLS:

$$r^2 = 0.61$$

$$pRMSE = 0.24$$

$$RPD = 1.22$$

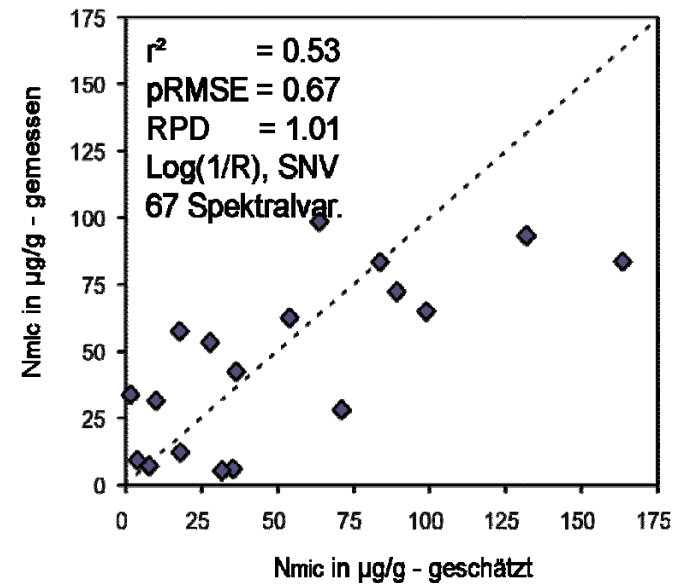
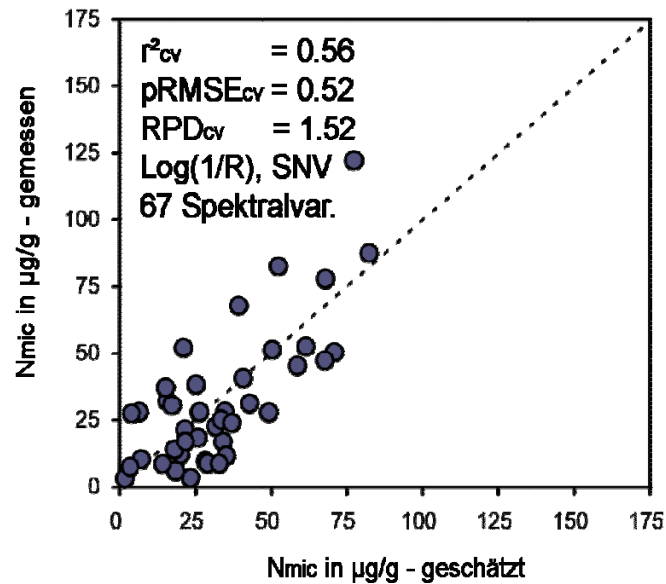
N und N_{mic}

N_{mic} :

GA-PLS

Kalibration

Kreuzvalidierung



Vergleichswerte PLS:

$$r^2_{cv} = 0.41$$
$$pRMSE_{cv} = 0.62$$
$$RPD_{cv} = 1.27$$

Vergleichswerte PLS:

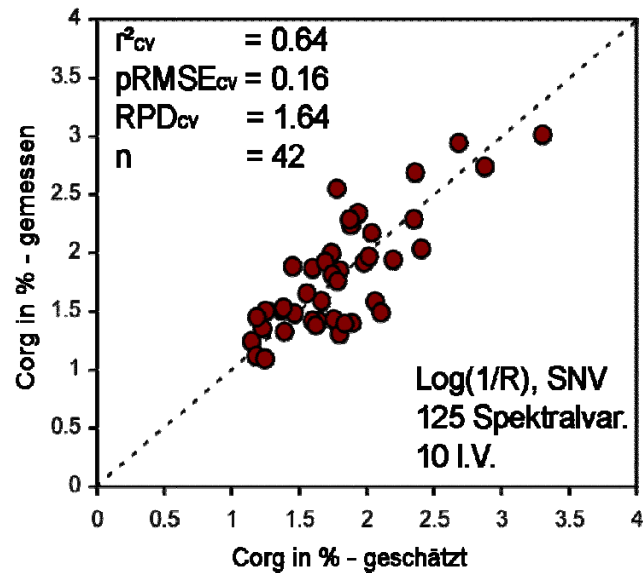
$$r^2 = 0.47$$
$$pRMSE = 0.55$$
$$RPD = 1.23$$

HyMap_{sim} – simuliert aus FieldSpec Daten

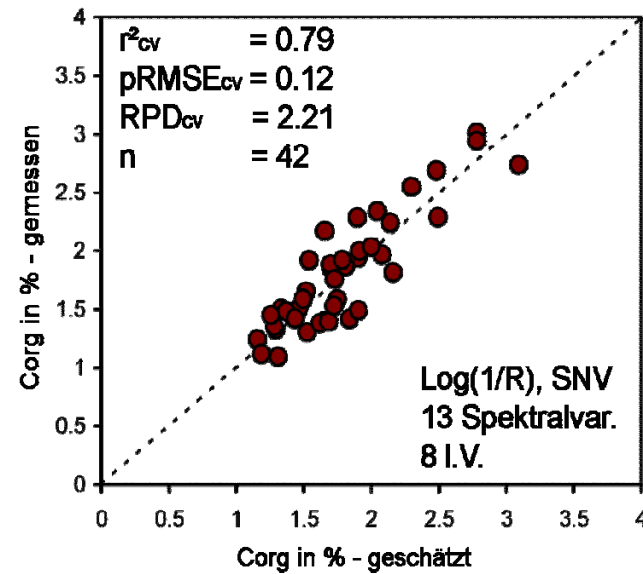
Schätzergebnisse Laborspektren

	PLS		GA-PLS	
	cv	val	cv	val
C_{org}				
r ²	0.68	0.56	0.82	0.70
pRMSE	0.15	0.25	0.11	0.18
RPD	1.75	1.13	2.38	1.62

Reicht die spektrale Auflösung (125 Kanäle)?



PLS

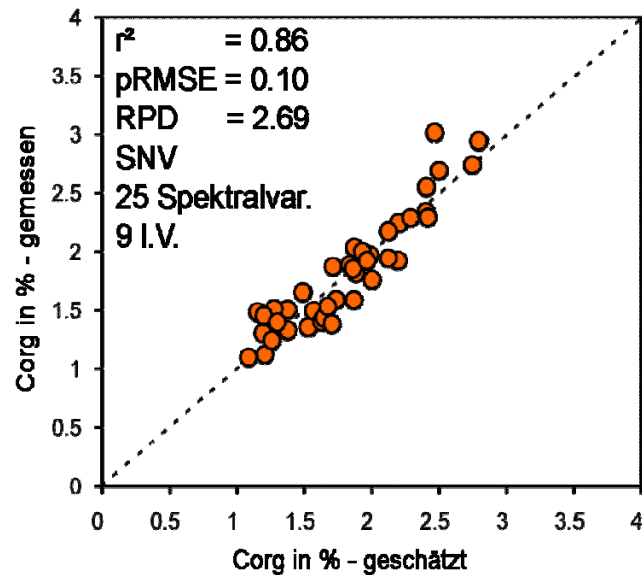


GA-PLS

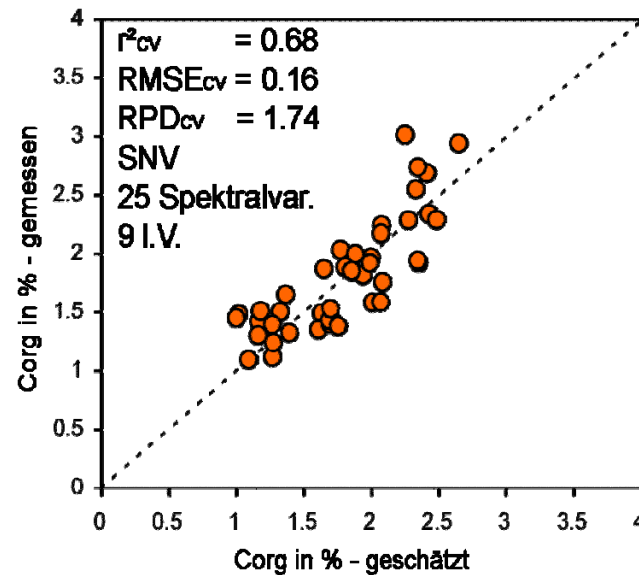
HyMap - C_{org}

	PLS		GA-PLS	
	cv	val	cv	val
C_{org} (Labor)				
r^2	0.68	0.56	0.82	0.70
pRMSE	0.15	0.25	0.11	0.18
RPD	1.75	1.13	2.38	1.62
C_{org} (HyMap_{elm})				
r^2	0.64		0.79	
pRMSE	0.16		0.12	
RPD	1.64		2.21	

GA-PLS



Kalibration



Kreuzvalidierung

Vergleichswerte PLS

$$r^2_{cv} = 0.48$$

$$pRMSE_{cv} = 0.20$$

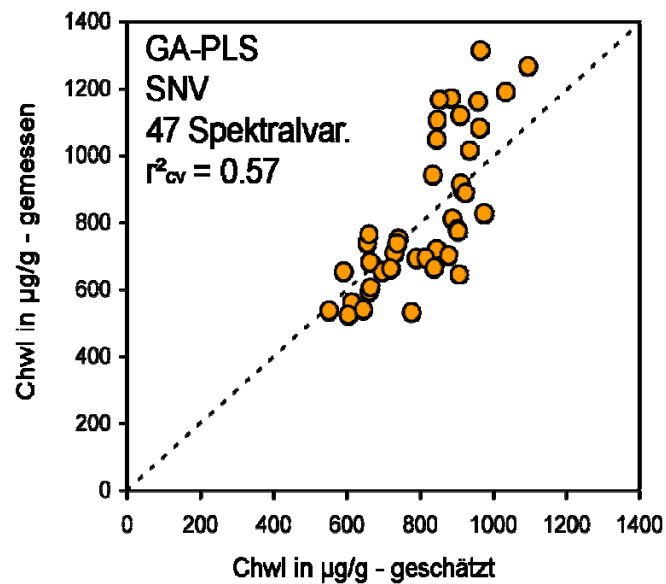
$$RPD_{cv} = 1.36$$

$$9 \text{ I.V.}$$

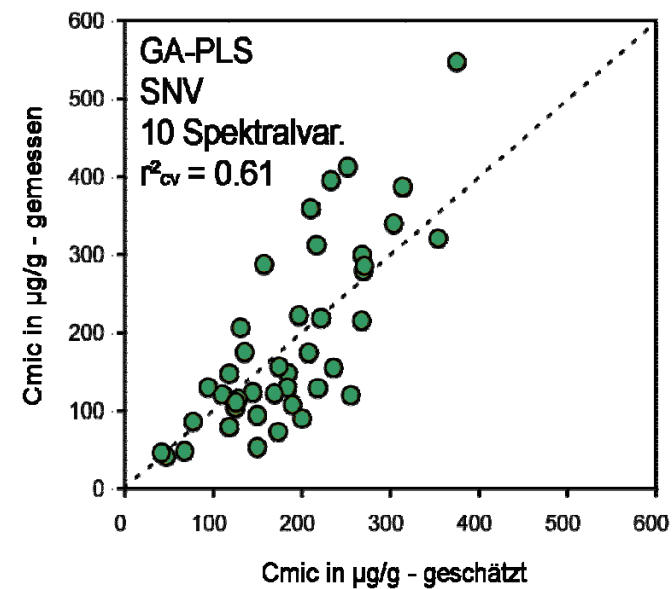
HyMap – C_{hwl} und C_{mic}

	PLS _{cv}	GA-PLS _{cv}	PLS _{cv}	GA-PLS _{cv}
C_{hwl} (Labor)			C_{mic} (Labor)	
r^2	0.62	0.80	0.67	0.77
pRMSE	0.17	0.12	0.37	0.30
RPD	1.63	2.27	1.69	2.07
C_{hwl} (HyMap)			C_{mic} (HyMap)	
r^2	0.26	0.57	0.40	0.61
pRMSE	0.24	0.18	0.49	0.40
RPD	1.13	1.49	1.29	1.48

C_{hwl} - GA-PLS



C_{mic} - GA-PLS



Labor:

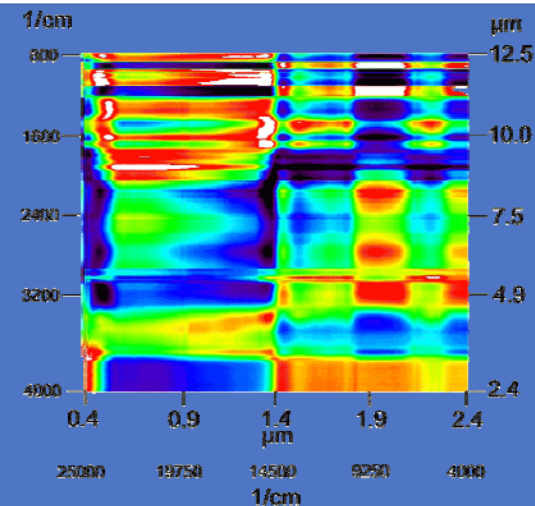
Validierungsergebnisse zeigen die Rangfolge $C_{mic}, C_{org} > N > C_{hwl}, N_{mic}$
Variablenselektion (GA) hilfreich, um für heterogene Probenkollektive zufriedenstellende Schätzgüten zu erreichen

Genetischer Algorithmus:

- effizientes stochastisches Werkzeug zur Variablenselektion
- Aufdeckung und Interpretation von Abhängigkeiten zwischen Zielgröße und Spektralinformation

Handelt es sich hierbei um direkte oder indirekte Korrelationen zwischen Bodenkonstituent und Spektralsignatur?

2 D-Korrelations-
Spektroskopie
MIR-VISNIR



Assessing Soil Organic Matter on a Landscape Scale by Combining Non-Invasive (Spectroscopic) and Invasive Methods (DFG)

