Statistische Verfahren:
Projekt 4 - Nahinfrarotspektroskopie I

Tobias Giesemann
Ferdinand Rewicki
Moritz Preuß

March 21, 2019

Abstract

We describe and assess ...

Contents

Statistische Verfahren:
Projekt 4 - Nahinfrarotspektroskopie I

Tobias Giesemann
Mail@me.com

Ferdinand Rewicki
...@gmail.com

Moritz Preuß
moleo.preuss@gmail.com

Abstract

We describe and assess ...

1 Einleitung

Obtaining ...

2 Background

2.1 Soil Parameters

Let A be any substance in a given soil sample dissolved in a solution of volume $V \in (0, \infty)$ and let $n_A \in (0, \infty)$ be the amount of A in the sample. Then the molar concentration c_A of A is given by

$$c_A := \frac{n_A}{V}$$

Now let c_0 be the molar concentration of this whole sample and n_0 the amount of the whole sample....

2.2 Near Infrared Spectroscopy

NIRS uses electromagnetic waves [**?** , 246],

The reflectance

$$\varrho\colon (0, \infty) \to (0, \infty), \qquad \varrho(\lambda) := \frac{P_\mathrm{r}(\lambda)}{P_0}$$

of a

3 Methodology

3.1 Measured Data

As a ... The reflectance $\varrho(\lambda)$ of a sample at a wavelength $\lambda \in \Lambda$ is recorded as

$$-\lg \varrho(\lambda) = -\frac{\ln \varrho(\lambda)}{\ln 10}$$

Figure 1 shows six randomly chosen soil spectra in a diagram.

3.2 Statistical Model

3.3 Modellwahl im klassischen linearen Modell

Im klassischen linearen Modell wird meist ... (s. Skript, S. XY) Das Problem mit der Wahl von α ... Beim Hinzufügen neuer Einflussgrößen in das Modell wird die Nullhypothese H_0 ...

3.4 Mallow's C_p

At this point, the model is specified using $k + 1 = 320$ predictors for each response variable, using the whole domain of measured spectrum for each soil sample.

3.5 Simulated Annealing

We can

3.6 Model Validation

3.7 Assessment by Simulation

4 Implementation

4.1 Choosing a Neighbour

We stated in section 3.5 that we want to select a "good" model for the prediction.

4.2 Additional Functions

All other functions were defined following a standard scheme. It follows from 3.4 that

$$\mathrm{cost}(M) := C_\mathrm{p}^{(M)}$$

4.3 Preprocessing

Implementing the algorithm described in 3.5 takes a sizeable toll on computing power. The most expensive calculations are performed in the computation of the residual sum of squares

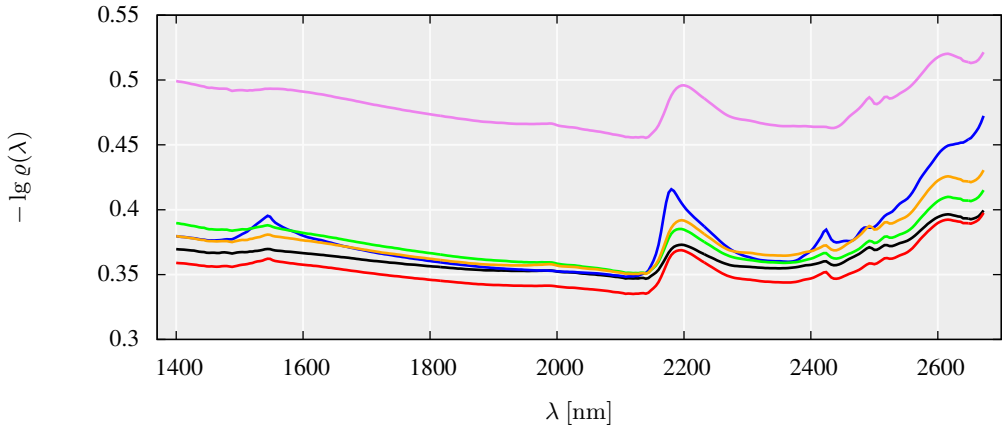
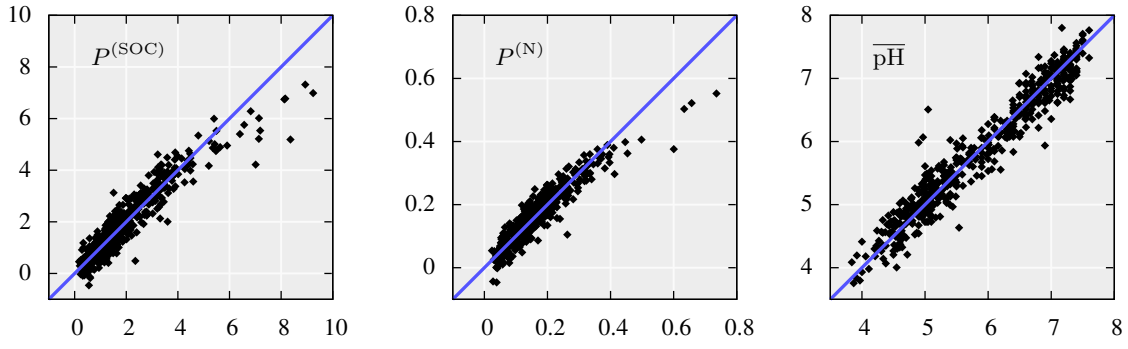


Figure 1: Six near infrared soil spectra of randomly chosen soil samples obtained from the data set, where λ is the wavelength and $\rho(\lambda)$ the corresponding reflectance and each colour refers to one sample



(a) $\hat{p}^{(\mathrm{SOC})} \sim p^{(\mathrm{SOC})}$
$(R^2)^{(\mathrm{SOC})} = 0.871$

(b) $\hat{p}^{(\mathrm{N})} \sim p^{(\mathrm{N})}$
$(R^2)^{(\mathrm{N})} = 0.861$

(c) $\widehat{\overline{\mathrm{pH}}} \sim \mathrm{pH}$
$(R^2)^{(\mathrm{pH})} = 0.940$

Figure 2: Correlation diagrams plotting $\hat{y}$ on y and the blue line representing the id

5 Calibration

5.1 Model Selection

SANN returned as minimum values for the respective models

$$C_{\mathrm{p}}^{(\mathrm{SOC})} = -17.46$$
$$C_{\mathrm{p}}^{(\mathrm{N})} = -28.51$$
$$C_{\mathrm{p}}^{(\mathrm{pH})} = -57.76$$

Figures 3a, 3b and 3c in appendix A show

5.2 Goodness of Fit

6 Simulation - Schätzgenauigkeit von Mallow's Cp

Die hier vorgestellte Untersuchung soll zeigen, ob und inwiefern sich Mallow‘s Cp als Schätzwert für den SPSE an diesen annähert, wenn sich der Umfang der Stichprobe verändert. Genauer gesagt ziehen wir ein Sample von oben erstelltem Modell, zu dem die exakte Berechnung des SPSE für alle von uns gewählten Stichprobengrößen (z.B. n = [5, 10, 25, 50, 100, 200, 500]) möglichist. Der ermittelte SPSE Wert dient uns im späteren Vergleich als Ground Truth. Für jeden Stichprobenumfang ermitteln wir zudem den Wert von Mallow‘s Cp, der ein Schätzwert des ursprünglichen SPSE darstellt. Ermittelt werden soll dabei, inwiefern sich der Umfang der Stichprobe auf die Abweichung des Schätzers von der Ground Truth (SPSE) auswirkt.

7 Conclusion

Using Mallow's C_{p} criterion, we calibrated three predictive models for the soil parameters $p^{(\mathrm{SOC})}$, $p^{(\mathrm{N})}$ and pH. To construct ...

References

A Prediction Parameters and Models

Table 1: Estimated model parameters of $P^{(\mathrm{SOC})}$ on selected model

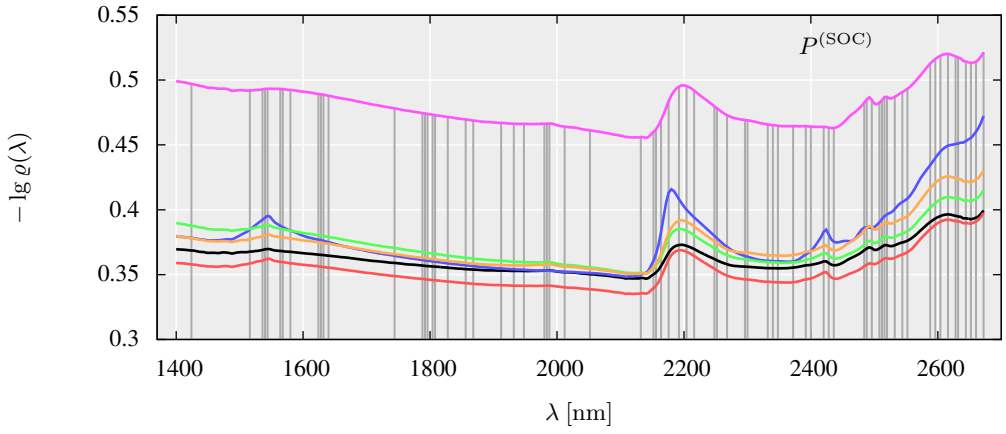
λ_i [nm]	$\beta_i^{(\mathrm{SOC})}$	λ_i [nm]	$\beta_i^{(\mathrm{SOC})}$	λ_i [nm]	$\beta_i^{(\mathrm{SOC})}$	λ_i [nm]	$\beta_i^{(\mathrm{SOC})}$
—	-1.47103	1808	1991.63	2204	-2319.71	2496	1956.13
1424	-811.326	1828	1568.91	2216	1075.21	2508	-5057.56

Table 2: Estimated model parameters of $\overline{\mathrm{pH}}$ on selected model

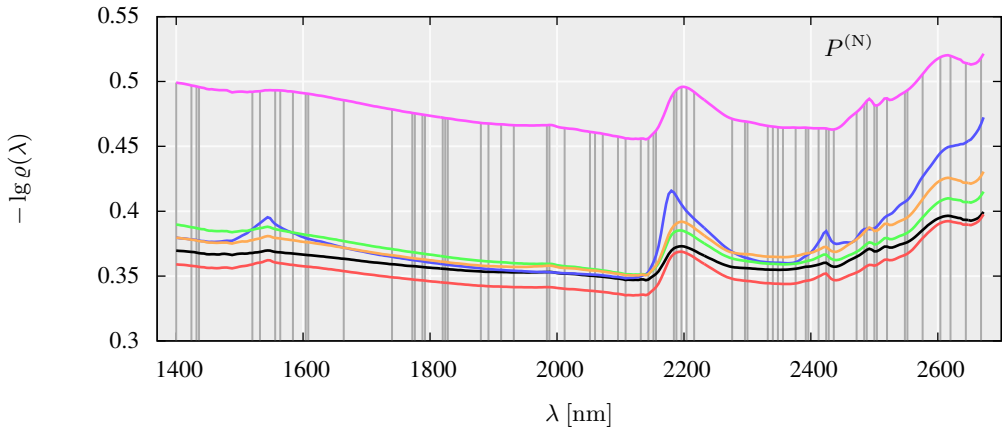
λ_i [nm]	$\beta_i^{(\mathrm{pH})}$	λ_i [nm]	$\beta_i^{(\mathrm{pH})}$	λ_i [nm]	$\beta_i^{(\mathrm{pH})}$	λ_i [nm]	$\beta_i^{(\mathrm{pH})}$
—	5.57628	1864	-508.298	2220	699.185	2460	545.634
1436	135.244	1896	-623.655	2224	-659.264	2464	-519.611

Table 3: Estimated model parameters of $P^{(\mathrm{N})}$ on selected model

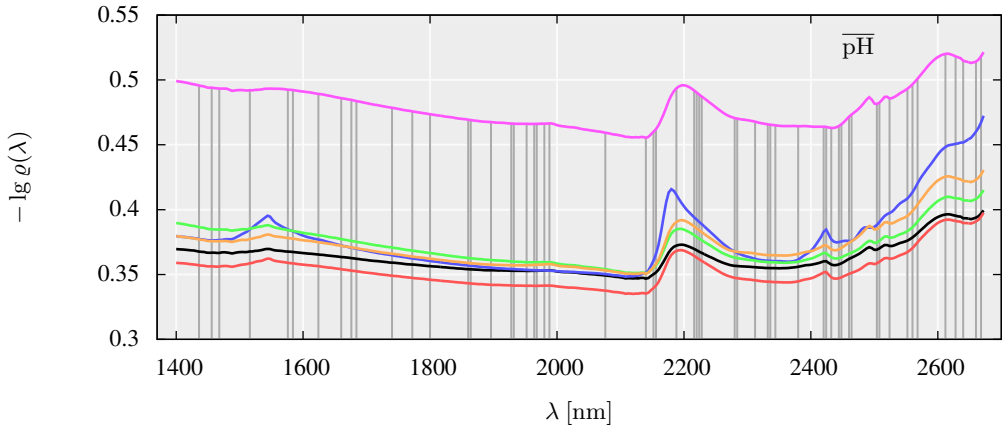
λ_i [nm]	$\beta_i^{(\mathrm{N})}$	λ_i [nm]	$\beta_i^{(\mathrm{N})}$	λ_i [nm]	$\beta_i^{(\mathrm{N})}$	λ_i [nm]	$\beta_i^{(\mathrm{N})}$
—	-0.0287506	1820	169.949	2156	95.2657	2428	116.231
1400	48.2214	1824	-272.304	2184	-99.54	2436	-60.6976



(a) $p^{(\mathrm{SOC})}$ with $\mathrm{p}^{(\mathrm{SOC})} = 72$



(b) $p^{(\mathrm{N})}$ with $\mathrm{p}^{(\mathrm{N})} = 69$



(c) pH with $\mathrm{p}^{(\mathrm{pH})} = 58$

Figure 3: Displaying the spectra from figure 1 with wavelength included in the selected models for each response highlighted by vertical grey lines

B R Source Code

Statutory Declaration

We herewith declare that ...

Jena, 25th of August 2016

Kazimir Menzel Markus Pawellek