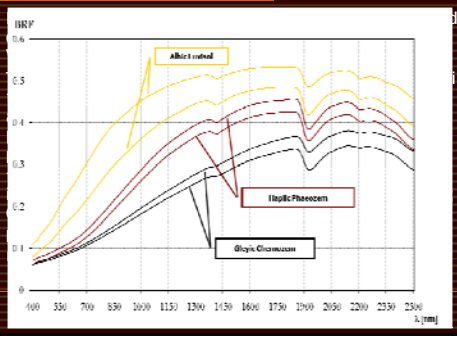
**K. Kuśnierek**
*Department of Soil Science and Remote Sensing of Soils, Adam Mickiewicz University, Poznań, Poland*
*(kus@amu.edu.pl)*

# *Pre-processing of soil visible and near infrared spectra taken in laboratory and field conditions to improve the within-field soil organic carbon multivariate calibration*

*2nd GWPSS, Montreal 2011*

## Introduction
### Soil reflectance

- d
- in



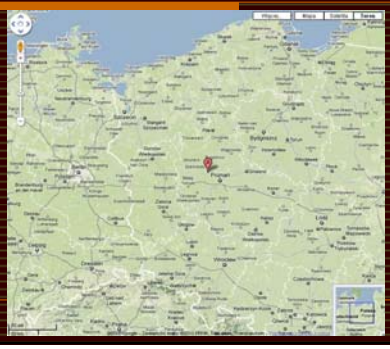

## Introduction
### Calibration methods

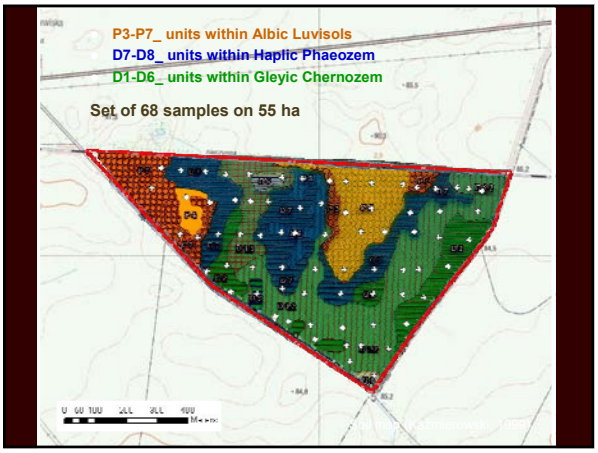
- The relationship between SOC and reflectance is used to constuct statistical regression models.
- Multivariate regression methods: PLSR, PCR, RT, ANN...
- Most common method: PLSR (Wold, 1966)
- Most effective method for total carbon estimation: PLSR (Vasques et al., 2008).
- Aadvanced machine learning techniques yielded better results than PLS regression (Viscarra Rossel and Behrens, 2010). Methods for large datasets.

## Introduction
### Pre-processing

- Pre-processing – transformations – pretreatment
- Pre-processing of data is an important step in multivariate calibration and its application improve the acuracy of prediction models (Boysworth and Booksh, 1988)
- Chang et al. (2001) concluded that the main challenge limiting application of NIRS for the evaluation of soil properties, is finding suitable data pre-treatment and calibration strategies.
- Preprocessing of spectral data for SOC estimation: Sorensen and Daalsgard (2005), Brunet et al. (2007), Stevens et al. (2008), Vasques et al. (2008).

## Study area
### Chlewiska study site in Poznań Lakeland



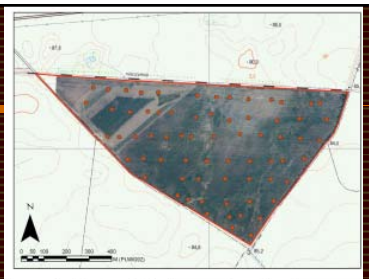


## Measurement methods
Laboratory

- Spectral measurement

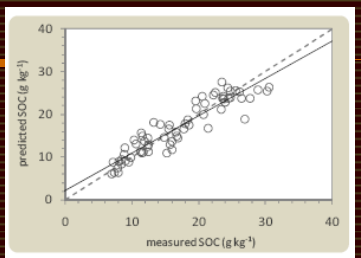
- Walkley-Black method

## Measurement methods
Field conditions



| | Texture (%) | | | SOC | pH | | CaCO3 |
|---|---|---|---|---|---|---|---|
| | sand | silt | clay | (g kg-1) | H2O | KCl | (g kg-1) |
| n.ob.[1] | 68 | 68 | 68 | 68 | 68 | 68 | 42 |
| min | 51 | 14 | 2 | 6.94 | 5.01 | 4.14 | 5.7 |
| mean | 71.3 | 21.4 | 7.4 | 17.10 | 6.86 | 6.35 | 22.9 |
| median | 72 | 20 | 7 | 16.34 | 7.28 | 6.84 | 18.5 |
| max | 83 | 37 | 15 | 30.54 | 7.84 | 7.43 | 59.6 |
| std[2] | 6.0 | 4.2 | 3.0 | 6.70 | 0.79 | 0.92 | 15.4 |

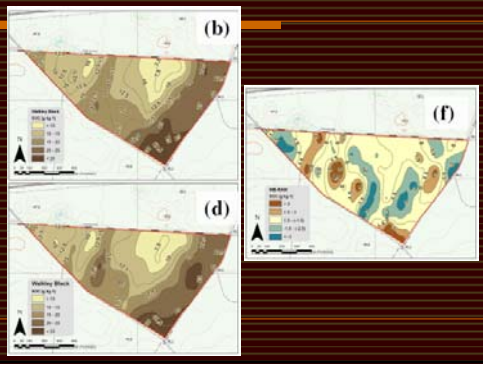
[1] number of observations, [2] standard deviation

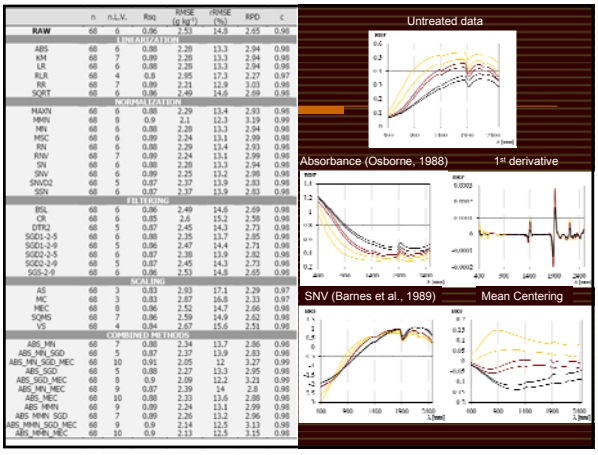



| n[1] | no.LV[2] | Rsq | RMSE | rRMSE | RPD | c |
|---|---|---|---|---|---|---|
| 68 | 6 | 0.86 | 2.53 | 14.8 | 2.65 | 0.98 |

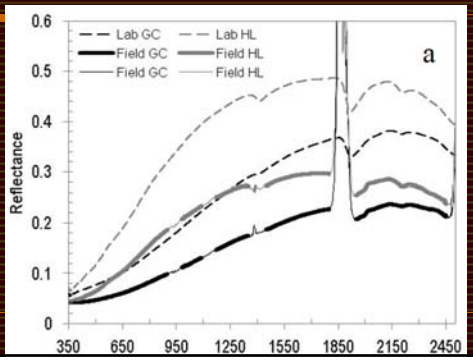
[1] number of observations, [2] number of PLS latent variables

The spectrally modeled SOC content map the Walkley Black reference SOC content map and the difference



The sample spectra of Gleyic Chernozems (GC, dark lines) and Haplic Luvisols (HL, bright lines) taken in the laboratory (dashed) and field (solid) conditions

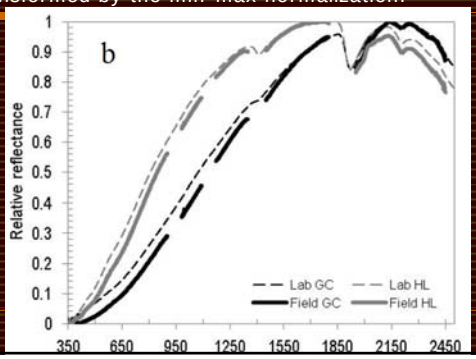



The accuracy of the SOC calibration using raw and pre-processed spectra from laboratory and field spectral measurements.

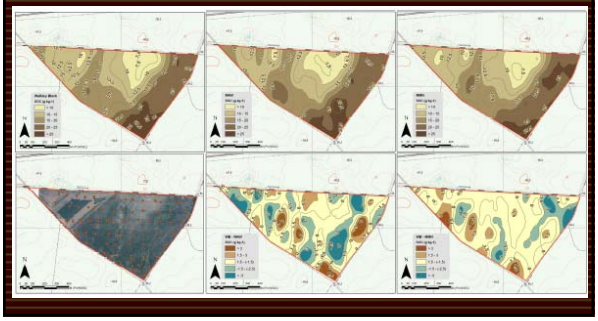
| | Laboratory | | | | | Field | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | n | n.LV[1] | $R^2$ | RMSE (g kg$^{-1}$) | RPD | n | n.LV[1] | $R^2$ | RMSE (g kg$^{-1}$) | RPD |
| **raw reflectance** | **68** | **6** | **0.86** | **2.53** | **2.65** | **68** | **6** | **0.78** | **3.12** | **2.15** |
| LINEARIZATION | | | | | | | | | | |
| absorbance (ABS) | 68 | 6 | 0.88 | 2.28 | 2.94 | 68 | 2 | 0.88 | 2.27 | 2.95 |
| Kubelka Munk | 68 | 7 | 0.89 | 2.28 | 2.94 | 68 | 3 | 0.87 | 2.47 | 2.71 |
| NORMALIZATION | | | | | | | | | | |
| **min-max (MMN)** | **68** | **8** | **0.9** | **2.1** | **3.19** | **68** | **7** | **0.89** | **2.19** | **3.06** |
| MSC | 68 | 6 | 0.89 | 2.24 | 2.99 | - | - | - | - | - |
| FILTERING | | | | | | | | | | |
| continuum removal | 68 | 6 | 0.85 | 2.6 | 2.58 | - | - | - | - | - |
| Sav.Gol. derivative(SGD) | 68 | 5 | 0.86 | 2.47 | 2.71 | - | - | - | - | - |
| SCALING | | | | | | | | | | |
| mean centering | 68 | 3 | 0.83 | 2.87 | 2.33 | 68 | 3 | 0.84 | 2.75 | 2.44 |
| median cent. (MEC) | 68 | 8 | 0.86 | 2.52 | 2.66 | 68 | 8 | 0.82 | 2.89 | 2.32 |
| COMBINED METHODS | | | | | | | | | | |
| ABS_MEC | 68 | 10 | 0.88 | 2.33 | 2.88 | 68 | 10 | 0.86 | 2.5 | 2.68 |
| ABS_MMN | 68 | 9 | 0.89 | 2.24 | 2.99 | 68 | 7 | 0.88 | 2.34 | 2.86 |
| ABS_MMN_SGD_MEC | 68 | 9 | 0.9 | 2.14 | 3.13 | - | - | - | - | - |
| ABS_MMN_MEC | 68 | 10 | 0.9 | 2.13 | 3.15 | 68 | 10 | 0.81 | 3.06 | 2.19 |

[1] number of latent variables

The sample spectra of Gleyic Chernozems (GC, dark lines) and Haplic Luvisols (HL, bright lines) taken in laboratory (solid) and field (dashed) conditions transformed by the min-max normalization.




SOC content mapped using the reference Walkley-Black method (a), and the multivariate calibration on raw reflectance spectra (b) and min-max pre-processed spectra and the respective difference maps



## Conclussions

- The use of the pre-processing methods allows reducing the uninformative variance from the soil spectra taken in laboratory and field conditions.
- The application of the pre-processed field spectra allows producing the soil maps of the accuracy, corresponding to the maps derived from the laboratory spectroscopy.
- The comparison between the results of this study and the previously published results indicates that the selection of the best performing pre-processing method is dataset dependant.
- Therefore, the future research on automated and unsupervised procedures using a large number of pre-processing methods in multivariate modeling is desired.

## References (1)


- Barnes, R.J., Dhanoa, M.S. and Lister, S.J., 1989. Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* 43: 772–777.
- Bowers, S.A. and Hanks, A.J., 1965. Reflection of radiant energy from soil. *Soil Science*, 100: 130-138.
- Boysworth, M.K., Booksh, K.S., 2000. Aspects of multivariate calibration applied to near-infrared spectroscopy, in Burns, D.A. and Ciurczak, E.W. (Eds.) *Handbook of Near-Infrared Analysis*, Marcel Dekker: New York[k1] .
- Brunet, D., Barthes, B.G., Chotte, J., and Feller, Ch., 2007. Determination of carbon and nitrogen contents in Alfisols , Oxisols and Ultisols from Africa and Brazil using NIRS analysis: Effects of sample grinding and set heterogeneity. *Geoderma*, 139: 106-117.
- Chang, Ch., Laird, D.A., Mausbach, M.J., and Hurburgh, Ch.R., 2001. Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties. *Soil Sci. Soc. Am. J.*, 65: 480-490.
- Hunt, G.R., 1980. Spectroscopic properties of rock and minerals in: Steward C.R. (Ed.), *Handbook of Phisical Properties of Rocks*, CRC Press, Boca Raton, 295 pp.

## References (2)


- Osborne, B.G., 1988. Comparative study of methods of linearization and scatter correction in near infrared reflectance spectroscopy, *Analyst*, 113: 263-267. doi: 10.1039/AN9881300263.
- Condit, H.R., 1970. The Spectral Reflectance of American Soils. *Photogrammetric Engineering*, 36: 955-966.
- Shepherd, K.D. and Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal* 66: 988–998.
- Sørensen, L.K. and Dalsgaard, S., 2005. Determination of clay and other soil properties by near infrared spectroscopy, *Soil Science Society of America Journal*, 69: 159–167.
- Vasques, G., Grunwald, S. and Sickman, J., 2008, Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma*. 146:14-25.
- Viscarra Rossel R.A. and Behrens T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 158: 46-54.
- Wold, H., 1966. Estimation of principal components and related models by iterative least squares, in P.R. Krishnaiaah (Ed.). Multivariate Analysis, New York: Academic Press, 391-420.