

# RAPID ESTIMATION OF BRILLIANT BLUE CONCENTRATIONS IN SOIL BY DIFFUSE REFLECTANCE SPECTROSCOPY

Christina Bogner<sup>1</sup>, Iris Schmiedinger<sup>2</sup> and Bernd Huwe<sup>2</sup>.

<sup>1</sup>Ecological Modelling, BayCEER, University of Bayreuth, Dr.-Hans-Frisch-Straße 1-3, 95448 Bayreuth, Germany

<sup>2</sup>Soil Physics Group, BayCEER, University of Bayreuth, 95440 Bayreuth, Germany

**This is a poster paper presented at the “Art Science and Application of Reflectance Spectroscopy Symposium” in Boulder, CO USA, February 23-25, 2010.**

## Abstract

Brilliant Blue is often used to trace water movement in soils. Its concentration in soil samples is usually determined by extraction – a laborious procedure with varying accuracy. We show that Brilliant Blue can be estimated directly by visible diffuse reflectance spectroscopy. We build a PLSR (partial least squares regression) model for the concentration range of 0.1 to 15 mg Brilliant Blue per g soil with an RMSE (root mean square error of prediction) of 1 mg g<sup>-1</sup> and an R<sup>2</sup><sub>adj</sub> of 0.9. However, as the method is based on visible spectra, prediction accuracy can be altered by variations in soil colour between calibration and prediction data sets.

## Introduction

Water flow and solute transport in soils are complex phenomena. Usually, water chooses certain preferred pathways through the soil and bypasses a large portion of the profile (e.g. Flury *et al.*, 1994). This type of flow is called preferential flow and is hardly predictable. Often visualising preferential flow in tracer studies is the best way to study it. Brilliant Blue FCF (C<sub>37</sub>H<sub>34</sub>N<sub>2</sub>Na<sub>2</sub>O<sub>9</sub>S<sub>3</sub>, molar mass 792.9 g mol<sup>-1</sup>) is an artificial organic dye frequently used in soil hydrology to trace water movement in soils and to identify preferential flow paths. In a tracer study, an aqueous solution of the dye is evenly applied at the soil surface by an automated sprinkler. Then vertical soil profiles are excavated to uncover infiltration patterns of the stained water. These patterns are photographed and used to determine concentration maps of Brilliant Blue to understand water and solute transport in soils (e.g. Forrer *et al.*, 1999).

Preferential flow of water can affect soil chemistry and biology. Several authors identified preferential flow paths in Brilliant Blue tracer studies and analysed the content of total and organic C and N (Hagedorn *et al.*, 1999; Bundt *et al.*, 2001a; Bundt *et al.*, 2001b). They showed that soil matrix and preferential flow paths differed in their physical, chemical and biological properties. The sorption of the dye on soil particles during infiltration of the tracer solution changes the C and N contents of the soil. Therefore, the concentration of Brilliant Blue in the soil has to be determined and the additional amounts of C and N due to the dye have to be subtracted.

To sum up, the estimation of Brilliant Blue concentrations in soil is a crucial step. Usually, the quantity of the dye is determined by extraction with a water acetone solution or a 0.5 M K<sub>2</sub>SO<sub>4</sub> (e.g. Bundt *et al.*, 2001a). This is a laborious procedure with changing accuracy due to varying mass recovery (Forrer *et al.*, 2000). We propose to estimate the content of Brilliant Blue by visible diffuse reflectance spectroscopy (VIS-DRS) directly on soil samples without extraction.

## Materials and methods

### Soils

We took soil material on an agricultural field at the experimental site of INRA Avignon, France (43°54'58.76"N, 4°52'54.81"E), and in a Norway spruce forest in the Fichtelgebirge in south-eastern Germany (50° 08'N, 11°52'E). The soil at the first site is a highly structured Calcisol (IUSS Working Group WRB, WRB, 2007) with a homogeneously greyish yellow colour (Oyama & Takehara, 1999) throughout the profile. Due to its high clay content it develops deep fissures during desiccation. These cracks can reach one meter and more and traverse the plough pan situated in about 25-30 cm depth. The forest soil is a Podzol (IUSS Working Group WRB, WRB, 2007) with five different horizons. The soil colour varies considerably between horizons from greyish yellow brown to light yellow. Table 1 summarizes the distribution of the soil fine fraction (soil particles < 2 mm) and the soil colour.

Table 1: Soil characteristics.

Site	Horizon or depth (cm)	Sand (%)	Silt (%)	Clay (%)	Colour <sup>†</sup>	Colour name <sup>†</sup>
I <sup>‡</sup>	0-30	5.18	48.35	46.47	2.5Y 6/2	greyish yellow
	30-60	4.94	48.14	46.93	2.5Y 6/2	greyish yellow
	60-100	15.53	47.63	36.85	2.5Y 6/2	greyish yellow
II <sup>§</sup>	Ea	43.16	48.47	8.37	10YR 5/2	greyish yellow brown
	Bsh	38.63	51.48	9.90	10YR 4/2	greyish yellow brown
	Bs	33.94	50.38	15.69	10YR 4/4	brown
	Bw	31.61	54.53	13.87	2.5Y 7/4	light yellow
	Bw/C	36.62	49.71	13.67	2.5Y 7/4	light yellow

<sup>†</sup>determined on dry samples according to Oyama & Tekahara (1999); <sup>‡</sup>agricultural site; <sup>§</sup>forest site

### Sample preparation and measurements

We prepared 40 calibration samples from site one (calibration set I) and 65 from site two (calibration set II) by mixing five g of sieved soil (< 2 mm) with 5 ml of Brilliant Blue solution with different concentrations. The final concentration of the dye for every depth or horizon ranged from about 0.1 to about 15 mg g<sup>-1</sup> soil for calibration set I and from 0.1 to about 10 mg g<sup>-1</sup> soil for calibration set II. After mixing the soil with the Brilliant Blue solution the samples were dried at 40°C and ground. Two independent data sets of 20 samples each from site I (validation set I) and from site II (validation set II) were prepared in the same way.

Visible diffuse reflectance spectra of the soil samples were collected using a Cricket accessory (Harrick Scientific Products) installed in a Cary 100 Conc UV-VIS spectrometer (Varian). An aliquot of a ground sample was scanned between 400 to 700 nm in 1 nm steps averaging 10 measurements per step. Then, a baseline correction procedure was applied:

$$\text{Corrected spectrum} = \frac{\text{Raw spectrum} - 0\% \text{Transmission}}{100\% \text{Transmission} - 0\% \text{Transmission}}$$

where 0%Transmission is the zero reference baseline collected with the sample beam covered and the 100%Transmission is the spectrum collected on spectralon (100% reflectance reference). Spectra were recorded in reflectance units (R) and transformed to absorbance units, i.e. log (1/R).

## Partial least squares regression

In order to estimate the content of Brilliant Blue in soil samples, we need a model relating soil spectra to the concentration of the dye. Partial least squares regression is a multivariate linear regression technique particularly suitable when predicting a dependent variable from a large number of independent (possibly collinear) variables. We used the software ParLes (Viscarra Rossel, 2008) that allows multivariate calibration and prediction based on partial least squares regression combined with bootstrap aggregation (bagging-PLSR). For an introduction to PLSR the reader is referred to Geladi & Kowalski (1986). During bagging, the calibration data set is repeatedly sampled with replacement and a PLSR model is calculated for each of these subsamples. Then, a mean predictor with confidence intervals can be derived. A tutorial on bootstrapping for chemical application can be found in Wehrens *et al.* (2000). Viscarra Rossel (2007) showed that bagging improved the robustness of PLSR, was less susceptible to overfitting and led to more accurate models while providing a measure for model uncertainty.

Prior to calibration, soil spectra were mean centred and smoothed with a wavelet filter. To choose the number of factors to use in the PLSR model, a leave-one-out cross validation was performed. The accuracy of the cross validation is measured by the root mean squared error of prediction (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

where  $N$  is the number of samples  $y_i$  the observed value,  $\hat{y}_i$  the predicted one. The optimal number of factors, i.e. the parsimonious model is chosen based on the corrected Akaike's Information Criterion ( $AIC_c$ ) (Sugiura, 1978):

$$AIC_c = N \ln(RMSE^2) + 2K + \frac{2K(K+1)}{N-K-1}$$

where  $N$  is the number of samples and  $K$  the number of factors. The model that represents an 'elbow' in the scree plot ( $AIC_c$  versus number of factors curve) is chosen. Some further measures to assess the goodness of the model are the bias (BIAS):

$$BIAS = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)$$

the standard deviation of the error distribution (SDE):

$$SDE = \sqrt{\frac{\sum_{i=1}^N ((\hat{y}_i - y_i) - BIAS)^2}{N-1}}$$

and RPD – the ratio of the standard deviation of the data to the RMSE.

## Results and discussion

### Calibration and validation

The diffuse spectra show a maximum of absorption at about 630 nm which corresponds to the Brilliant Blue absorption maximum (e.g. Forrer (1997)). The second smaller peak of the Brilliant Blue absorption spectrum is situated at about 430 nm (Figure 1).

The leave-one-out cross validation indicated a model with 4 factors for calibration set I and a model with 6 factors for calibration set II (Figure 2). The first 4 factors account for 95.61% of variation in the Brilliant Blue concentrations and 99.99% of variation in the spectra of calibration set I. In the calibration

set II, the model explained 95.17% of variation in the Brilliant Blue concentrations and 99.99% of variation in the spectra. The leave-one-out validation statistics are similar for the two calibration sets. The RMSE of about  $1 \text{ mg g}^{-1}$  and an  $R^2_{\text{adj}}$  of 0.94 indicate good models (Table 2). The fitted versus observed values plots did not show any non linearity (Figure 3). In the calibration set I, one point with a high residual was detected, but it did not deteriorate the RMSE of the validation.

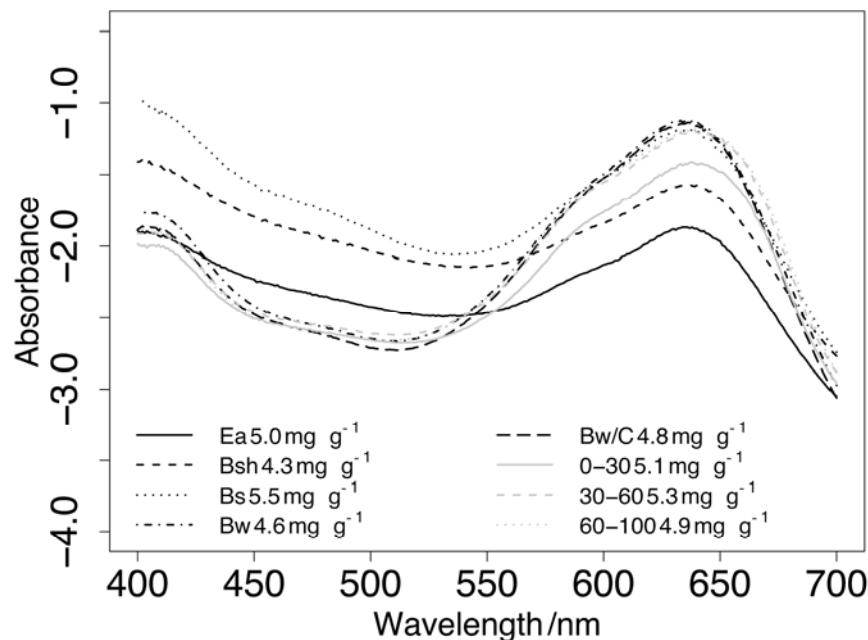


Figure 1: Soil visible diffuse spectra with about 5 mg Brilliant Blue per g soil.

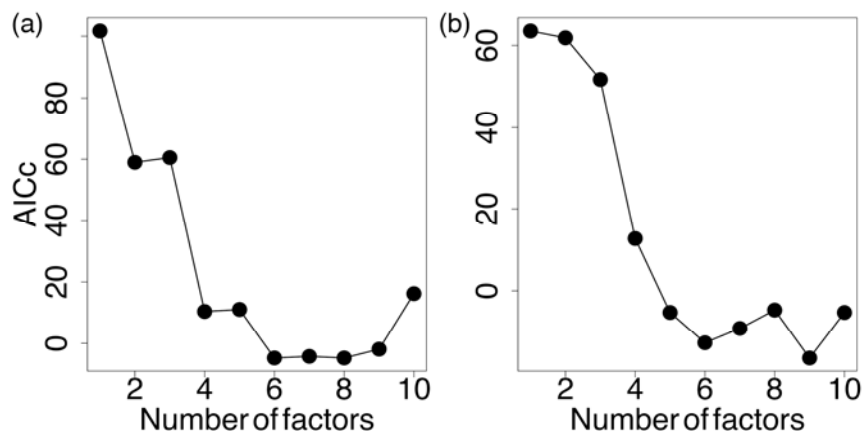


Figure 2: Evolution of the corrected Akaike's Information Criterion ( $AIC_c$ ) with increasing number of factors; (a) calibration set I and (b) calibration set II.

Table 2: Statistics of the calibration procedure based on leave-one-out cross validation.

Calibration set	RMSE <sup>†</sup> ( $\text{mg g}^{-1}$ )	$R^2_{\text{adj}}$	BIAS ( $\text{mg g}^{-1}$ )	SDE <sup>‡</sup> ( $\text{mg g}^{-1}$ )	RPD <sup>§</sup>
I	1.29	0.94	-0.01	1.31	4.04
II	0.82	0.94	-0.01	0.82	4.10

<sup>†</sup>Root mean square error of prediction; <sup>‡</sup>Standard deviation of the error distribution; <sup>§</sup>Ratio of the standard deviation of the data to the RMSE

The statistics of the leave-one-out cross validation and the independent validation are comparable (Fehler! Ungültiger Eigenverweis auf Textmarke.). Validation set I gave an RMSE of  $1.09 \text{ mg g}^{-1}$  and an  $R^2_{\text{adj}}$  of 0.95. For validation of set II we obtained an RMSE of  $0.98 \text{ mg g}^{-1}$  and an  $R^2_{\text{adj}}$  of 0.87. All depths in validation set I and horizons in validation set II were equally well predicted (Figure 4).

Table 3: Statistics of the validation procedure.

Validation set	RMSE <sup>†</sup> ( $\text{mg g}^{-1}$ )	$R^2_{\text{adj}}$	BIAS ( $\text{mg g}^{-1}$ )	SDE <sup>‡</sup> ( $\text{mg g}^{-1}$ )	RPD <sup>§</sup>
I	1.09 (0.83, 1.59) <sup>¶</sup>	0.95	-0.15	1.11	4.57
II	0.98 (0.74, 1.43) <sup>¶</sup>	0.87	-0.14	1.00	2.87

<sup>†</sup>Root mean square error of prediction; <sup>‡</sup>Standard deviation of the error distribution; <sup>§</sup>Ratio of the standard deviation of the data to the RMSE; <sup>¶</sup>95% confidence interval

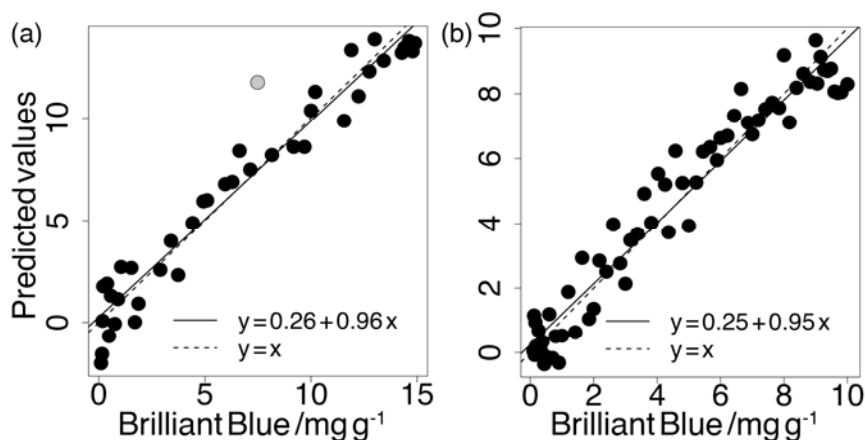


Figure 3: Results of the leave-one-out cross validation; (a) calibration set I and (b) calibration set II.

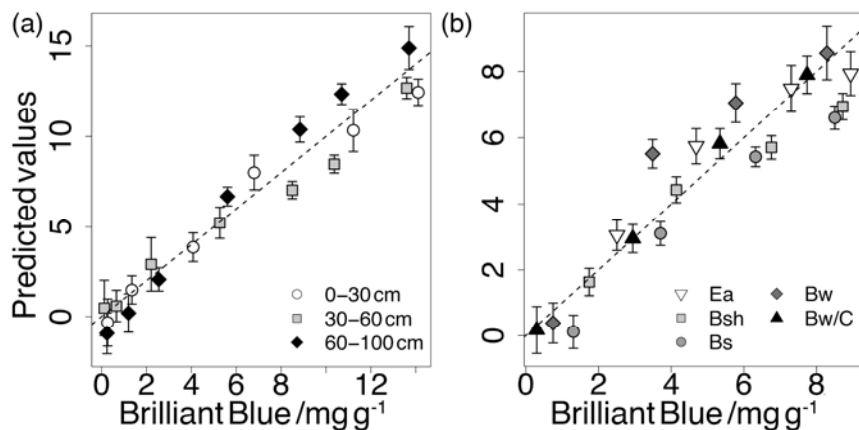


Figure 4: Results of the independent validation; (a) calibration set I and (b) calibration set II.

### Robustness

To infer the robustness of the models, we calculated the concentrations of Brilliant Blue in validation set I using the calibration set II and vice versa. Using the same PLSR models as described above yielded an  $R^2_{\text{adj}}$  of 0.22, an RMSE of  $13.51 \text{ mg g}^{-1}$  and a BIAS of  $-10.66 \text{ mg g}^{-1}$  for validation set II. Prediction for calibration set I based on calibration set II was better, leading an  $R^2_{\text{adj}}$  of 0.87, an RMSE of  $2.94 \text{ mg g}^{-1}$  and a BIAS of  $1.65 \text{ mg g}^{-1}$ . This deterioration of prediction ability is probably due to variations in soil colour, as the concentration range of Brilliant Blue is similar in both calibration sets. Soil colour must be considered as a nuisance parameter that changes the hue of samples with the same

concentration of Brilliant Blue, thus affecting their VIS diffuse spectra. To deal with this problem a regression model with both spectral and designed variables including additional information like horizon or soil colour could be applied (Jørgensen *et al.*, 2004). However, this is ongoing research and outside of the scope of this short paper.

To improve prediction we combined both calibration sets and built a new model. The pre-processing of the spectra had to be adjusted to account for a more heterogeneous data set. In addition to wavelet smoothing and mean-centring we applied a wavelet detrending and calculated the first derivative. Wavelet detrending is often used to correct baseline shifts or to remove curvilinearity in the spectra (Viscarra Rossel *et al.*, 2007). The prediction of both validation sets together was satisfactory and gave an  $R^2_{adj}$  of 0.88, an RMSE of 1.34 mg g<sup>-1</sup> and a BIAS of 0.11 mg g<sup>-1</sup>.

## Conclusions

Rapid and accurate estimation of Brilliant Blue by diffuse reflectance spectroscopy is possible. However, the accuracy of prediction can be altered when soil colour varies between calibration and prediction data sets.

## Acknowledgements

The first author was financially supported by the Deutsche Forschungsgemeinschaft (DFG FOR 562) and the Deutsch-Französische Hochschule (Grant Nr. CT-31-06).

## References

- Bundt, M., Jaggi, M., Blaser, P., Siegwolf, R. & Hagedorn, F. (2001a). Carbon and nitrogen dynamics in preferential flow paths and matrix of a forest soil. *Soil Science Society of America Journal*, 65, 1529-1538.
- Bundt, M., Widmer, F., Pesaro, M., Zeyer, J. & Blaser, P. (2001b). Preferential flow paths: biological 'hot spots' in soils. *Soil Biology & Biochemistry*, 33, 729-738.
- Flury, M., Flühler, H., Jury, W.A. & Leuenberger, J. (1994). Susceptibility of soils to preferential flow of water: a field study. *Water Resources Research*, 30, 1945-1954.
- Forrer, I.E. (1997). *Solute transport in an unsaturated field soil: visualisation and quantification of flow patterns using image analysis*. Ph.D. thesis, Swiss Federal Institute of Technology, Zurich.
- Forrer, I.E., Kasteel, R., Flury, M. & Flühler, H. (1999). Longitudinal and lateral dispersion in an unsaturated field soil. *Water Resources Research*, 35, 3049-3060.
- Forrer, I.E., Papritz, A., Kasteel, R., Flühler, H. & Luca, D. (2000). Quantifying dye tracers in soil profiles by image processing. *European Journal of Soil Science*, 51, 313-322.
- Geladi, P. & Kowalski, B.R. (1986). Partial least-squares regression - a tutorial. *Analytica Chimica Acta*, 185, 1-17.
- Hagedorn, F., Mohn, J., Schleppei, P. & Flühler, H. (1999). The role of rapid flow paths for nitrogen transformation in a forest soil: A field study with micro suction cups. *Soil Science Society of America Journal*, 63, 1915-1923.
- IUSS Working Group WRB. (2007). World reference base for soil resources 2006, first update 2007. World Soil Resources Reports No. 103.
- Jørgensen, K., Segtnan, V., Thyholt, K. & Næs, T. (2004). A comparison of methods for analysing regression models with both spectral and designed variables. *Journal of Chemometrics*, 18, 451-464.
- Oyama, M. & Takehara, H. (1999). *Revised standard soil color charts*. Eijkelpamp Agrisearch Equipment, Soil colour book 08.11, Netherlands.
- Sugiura, N. (1978). Further analysts of the data by akaike's information criterion and the finite corrections. *Communications in Statistics-Theory and Methods*, 7, 13-26.
- Viscarra Rossel, R.A. (2007). Robust modelling of soil diffuse reflectance spectra by "bagging-partial least squares regression". *Journal of near Infrared Spectroscopy*, 15, 39-47.
- Viscarra Rossel, R.A. (2008). ParLeS: Software for chemometric analysis of spectroscopic data. *Chemometrics and Intelligent Laboratory Systems*, 90, 72-83.
- Viscarra Rossel, R.A., Taylor, H.J. & McBratney, A.B. (2007). Multivariate calibration of hyperspectral  $\gamma$ -ray energy spectra for proximal soil sensing. *European Journal of Soil Science*, 58, 343-353.
- Wehrens, R., Putter, H. & Buydens, L.M.C. (2000). The bootstrap: a tutorial. *Chemometrics and Intelligent Laboratory Systems*, 54, 35-52.