This article was downloaded by: On: 17 January 2011 Access details: Access Details: Free Access Publisher Taylor & Francis Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37- 41 Mortimer Street, London W1T 3JH, UK

To cite this Article Geyer, Wolfgang , Brüggemann, Lutz and Hanschmann, Günter(1998) 'Prediction of Properties of Soil Humic Substances from FTIR Spectra using Partial Least Squares Regression', International Journal of Environmental Analytical Chemistry, 71: 2, 181 — 193

To link to this Article: DOI: 10.1080/03067319808032626 URL: <http://dx.doi.org/10.1080/03067319808032626>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use:<http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Intern. 1. Envimn. AMI. Chem., **Vol. 71(2). pp. 181-193 Reprints available directly from** *the* **publisher** Photocopying permitted by license only

PREDICTION OF PROPERTIES OF SOIL HUMIC SUBSTANCES FROM FTIR SPECTRA USING PARTIAL LEAST SQUARES REGRESSION

WOLFGANG GEYER*, **LUTZ** BRUGGEMANN and GUNTER HANSCHMANN

UFZ Centre for Environmental Research Leipzig-Halle Ltd. Dept. of Analytical Chemistry, Permoserstr: 15, 0-04318 Leipzig, Germany

(Received 20 May 1997; In final form 10 October 1997)

Humic substances (HS) play a key role in aquatic and terrestrial ecosystems. The understanding of the ecological functionalities of HS is based on the analysis of their properties, which is normally a very time-consuming procedure. Therefore we tested the possibility to apply the partial least squares regression **(PLSR)** method in connection with the mid infrared Fourier transform **(ITIR)** spectra of a series of soil humic substances for the prediction of different HS properties. The results with humic acid (HA) and fulvic acid (FA) fractions of soil HS from different environments show the possibility to predict several properties of an unknown soil HA with satisfying reliability above all the elemental composition.

Keywords: Soil humic substances; mid-infrared; **PLS** regression

INTRODUCTION

The vital role of humic substances (HS) in aquatic and terrestrial ecosystems has stimulated manifold efforts to uncover the mechanisms by which these refractory organic substances realize their environmental functionalities. Especially in studies concerning ecosystem dynamics, **a** number of analyses are needed **to** describe the usually complex and mostly heterogeneous environmental compartment under investigation. The measurements of the required parameters should be rapid but precise enough to allow reliable conclusions. Therefore, spectroscopic methods have been often used for these purposes. However, in the case of HS additional difficulties arise from the extreme molecular heterogeneity and complexity of these substances^[1,2], that are frequently available only in very small

^{*} Corresponding author. Fax No.:+49-341-2352625. E-mail: geyer@suny.rz.ufz.de.

quantities, e. **g.** from waters. So it will be not ever possible to perform all the measurements needed for the characterization of the HS. *An* expedient from this situation could be to use an analytical method which needs only a small quantity of substance for the measurement, but contains so much information, that other properties than the measured one can be predicted by calculation. Such an approach is already known from early work in applied infrared spectroscopy. Some of the topics in this line are the characterization of $\text{coal}^{[3,4]}$, peat^[5] and soil^[6] and the prediction of environmental fate parameters for chemicals^[7] using methods of multivariate **data** analysis.

The aim of our study was to evaluate the possibility to predict various properties of soil HS from infrared spectra. We started our investigations with soil HS, because in **this** case we could extract amounts of HS enough **to** carry out analyses with as many methods **as** possible. Additionally we included results from measurements of aqueous soil extracts for comparison with an ecologically meaningful parameter.

EXPERIMENTAL

Material

Soil samples of the A horizon were taken from 19 different environments in urban and industrial regions surrounding Leipzig, Halle, and Bitterfeld (Germany). Nine samples originated from floodplain areas (series A and series K), six samples from a municipal landfill (series T) and four samples from fallow land (series **V).** Sample sites differed in pollution, inundation, topology, and tillage. Air-dried samples were sieved through **a** 2 mm mesh.

Preparation of fulvic (FA) and humic (HA) acids

HS were extracted by use of **0.1M** NaOH and separated into FA and HA according to the recommendations of the **IHSS** (International Humic Substances Society), following the procedure described by Kuwatsuka et al.^[8]

Analyses

The FA and HA were characterized by elemental analyses **(LECO CHN932),** UVNIS spectra (Varian Cary 3, 0.05 M NaHCO₃ solution) and ¹H-NMR spectra (Varian Unity 400, 0.5 M NaOD/D₂O solution).

In addition the contents of dissolved organic carbon (DOC) and of **HS** were estimated for the aqueous soil extracts to determine the **HS/DOC** ratio, which is an ecologically important quantity (compare with the "humus coefficient", mentioned by Prakash et al.^[9]).

The samples were prepared as KBr pellets (about 1 mg of the HA or FA was mixed with about 300 mg of coarse disperse KBr by milling for 2 min and resulting powder pressed under vacuum for **4** min to a thin pellet with a diameter of 13 mm). Absorbance spectra were obtained with a **FTIR** spectrometer (System **2000** FTIR, Perkin Elmer) in the frequency region from 3800 to 450 cm^{-1} using resolution of 2 cm⁻¹ and data interval of 0.2 cm^{-1} . 128 scans were accumulated. The spectra were scaled to 1 mg sample/cm and normalized to the carbon content of the compound.

Data analysis

In the last years the partial least squares regression (PLSR) has been established **as** a statistical standard approach, that we used for the prediction of chemical properties of the prepared **HS** from the IR spectra.

If multivariate statistical modelling tools, based on the maximum likelihood principle, were applied to real data with many collinear variables and a modest number of observations, grave identification and convergence problems were often encountered. In the basic papers of this field $[10-15]$ it has been shown that PLSR solves the multivariate prediction problem for collinear data with satisfactory predictive ability.

PLSR is a non least squares biased linear regression method that relates a set of predictor variables, **X,** to a set of response variables, **Y. A** least squares regression is performed on a set of uncorrelated variables, **T,** that are standardized linear combinations of the original predictor variables, **X.** The variables **T** are called the **X** latent variables. PLS models the responses as

$$
\mathbf{Y} = \mathbf{T} \mathbf{B} \mathbf{Q} + \mathbf{E}
$$

where **B** is a diagonal matrix containing the least squares coefficients from the regression on the latent variables, *Q* contains the weights of the responses, and E is the residual matrix. The latent variables are calculated on at a time, maximizing their covariance.

We worked with the chemometrics programs PLSplus for GRAMS/386 (PLS-1 approach), Galactic Industries Corporations and SCAN 1 *.O* from the Minitab Inc. The SCAN program works with X-weights, X-latent variables (X-scores), Y-weights and Y-latent variables (Y-scores). The X-weights resemble to the usual PLS-loadings (spectral loadings) and the Y-weights refer to chemical loadings. The optimal number of latent variables was determined by means of validation statistics. A measure of the goodness of the prediction is calculated from the cross-validated residuals (labelled as standard error of prediction):

$$
SECV = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Ya_i - Yp_i)^2}
$$

where Ya_i are the actual and Yp_i the predicted values.

Moreover, the **SCAN** program facilitates the calculation of regression coefficients for the original predictors and in **this** way the analysis of the importances of these regression coefficients. For each response, the importance of a predictor is measured by the standardized regression coefficient (equals to the corresponding regression coefficient multiplied by the standard deviation of the response). We used the predictor importances for the detection of the spectral ranges contributed mostly to the model. These importances are suitable for our purpose because the predictor importances reflect the whole regression model and not only a contribution to single factor loadings.

In order to handle the **SCAN** program, the **IR** spectra of our **19** samples, each of them consisting of **16751** points, were transformed into variable pattern with **838** points by means of an average moving procedure of PLSplus.

RESULTS

In the Tables I and II the data concerning elemental analysis, the N/C, O/C, and **WC** ratios and the ash content at **750" C (4** hours) of the HA and FA, respectively, are reported. Some data, commonly used for the characterization of **HS,** were calculated from the UV/VIS and ¹H-NMR spectra of the HA solutions and shown in Table **III** together with the **HS/DOC** ratios of aqueous soil extracts.

The infrared spectra of **HS** are generally poorly resolved and show a considerable degree of similarity. Differences between the IR spectra of the samples (Figure la and b) happened mainly in the frequency region between **2950** and **2850** cm^{-1} (aliphatic CH₃ and CH₂) and between 1750 and 1500 cm^{-1} (C=O of CO₂H, **C=O** of ketonic carbonyl, hydrogen-bonded **C=O, COO-,** NH-deformation of amide **I1** type), which therefore are commonly used for the interpretation of IR-spectra of **HS.**

However, we used the full spectra to test the predictive ability of the IR spectra for various properties of **HS,** because we assume that the structural informations of these polyfunctional macromolecules are more or less spread out over the whole IR spectral region.

3500 30100 2500 2dOO Id00 Id00 **560 Absorbance** *I* **Wavenumber (cm-1)**

FIGURE 1 **PTIR-spectra of humic (a) and fulvic acids** (b) **from soils (one was chosen from each group)**

Sample	Elemental composition $(\%)^a$					Atomic ratio			$Ash(\%)$
	C	Н	N	Ω	S	N/C	O/C	H/C	
Al	51.38	4.40	3.59	39.18	1.45	0.0599	0.572	1.028	6.64
A2	51.05	4.12	4.05	39.75	1.03	0.0680	0.584	0.968	0.97
A ₃	51.35	3.97	3.43	39.97	1.28	0.0572	0.584	0.928	3.82
A ₄	50.12	5.22	4.51	38.78	1.37	0.0771	0.580	1.250	2.63
A ₅	51.77	4.97	4.16	38.25	0.85	0.0689	0.554	1.152	1.32
K1	52.43	5.41	3.90	35.47	2.79	0.0637	0.507	1.238	1.04

TABLE I Elemental composition and ash content of the humic acid fractions from soils

"ash free basis, 0 content was calculated from the weight difference to 100%.

TABLE Il Elemental composition and ash content of the fulvic acid fractions from soils

Sample	Elemental composition $(\%)^a$					Atomic ratio			Ash $\%$)
	\mathcal{C}	H	N	0	S	N/C	O/C	H/C	
A1	45.42	3.15	3.46	46.44	1.53	0.0653	0.767	0.833	2.45
A2	44.84	3.87	3.35	46.76	1.18	0.0640	0.782	1.052	3.10
A3	42.88	3.76	2.51	49.55	1.30	0.0502	0.867	1.052	2.45
A4	45.16	4,67	3.19	45.92	1.06	0.0657	0.763	1.241	2.88
A5	45.39	4.43	3.14	45.79	1.25	0.0593	0.757	1.171	2.78
Ki	45.92	4.47	2.27	44.70	2.64	0.0424	0.730	1.168	4.61
K ₂	44.49	4.27	2.87	45.75	2.62	0.0553	0.771	1.152	0.75
K3	44.37	5.11	3.77	45.19	1.56	0.0728	0.764	1.382	1.74
K4	44.50	4.85	3.62	45.91	1.12	0.0697	0.774	1.308	2.04
T1	44.93	2.69	2.41	48.59	1.38	0.0460	0.811	0.718	4.10
T ₂	44.83	3.70	2.35	47.70	1.42	0.0449	0.798	0.990	4.60
T3	45.42	3.58	2.64	46.70	1.66	0.0498	0.771	0.946	3.27
Т4	44.92	4.01	2.68	47.01	1.38	0.0511	0.785	1.071	1.36
T5	44.25	3.04	1.46	50.17	1.11	0.0283	0.850	0.824	1.34
T6	44.47	3.43	3.08	47.19	1.83	0.0594	0.796	0.926	2.88
V1	45.14	3.72	1.49	48.47	1.18	0.0283	0.805	0.989	4.15
V ₂	44.42	3.11	1.40	48.81	1.26	0.0270	0.824	0.840	3.10
V ₃	43,46	3.50	0.41	51.40	1.23	0.0296	0.887	0.966	2.64
V4	43.09	4.34	1.88	49.13	1.56	0.0374	0.855	1.209	2.41

%h free basis, 0 content was calculated from the weight difference to 100%.

Sample		UV/VIS Molar Absorbance ^a	¹ H-NMR Relative intensity (% of total area)				
	$E(280)^b$	$E\llap{/}E_6^c$	Aliphatic ^d protons $0.4 - 1.7$ ppm	Aromatic protons $6.5 - 8.1$ ppm	HCO ^e $3.3 - 4.6$ ppm	Arom./Aliph. protons	HS/ DOC
A1	51.38	3.59	32.0	13.3	28.2	0.416	0.572
A2	51.05	4.05	31.7	10.9	31.8	0.344	0.584
A3	51.35	3.43	31.2	12.6	28.2	0.404	0.584
A4	50.12	4.51	30.8	12.8	31.3	0.416	0.580
A5	51.77	4.16	33.0	13.0	27.4	0.394	0.554
K1	52.43	3.90	34.9	15.4	23.4	0.442	0.507
K ₂	51.46	4.82	35.3	13.0	25.5	0.369	0.521
K ₃	50.77	4.56	33.3	12.7	29.7	0.382	0.562
K4	49.96	4.41	30.4	12.0	30.9	0.395	0.585
T1	54.58	2.11	29.9	16.3	20.0	0.545	0.515
T ₂	52.61	2.12	29.1	16.8	19.4	0.577	0.560
T3	54.41	1.86	27.7	17.2	18.6	0.621	0.516
T4	55.15	3.11	29.3	16.4	22.1	0.560	0.477
T5	57.25	1.42	31.0	24.6	12.8	0.794	0.468
T ₆	53.65	2.92	30.3	16.6	20.7	0.548	0.505
V1	54.78	2.84	31.3	21.7	22.5	0.693	0.520
V ₂	53.58	2.69	26.6	22.1	23.3	0.831	0.547
V ₃	55.91	2.92	29.7	23.4	22.7	0.788	0.493
V ₄	56.50	3.07	28.8	21.9	23.5	0.760	0.477

TABLE **III** Data from UVNIS and **from** 'H-NMR spectra of the HA fractions **and** "Humic Coefficient" **(HS/DOC)** of aqueous soil extracts

^anormalized to the carbon content, ^brelated to the aromatic carbon content.

'E (465)/E(665), inversely proportional to molecular size, dupfield aliphatic protons only. ^eprotons on a carbon adjacent to a heteroatom.

The quality of the prediction can be viewed from the plots of the predicted component values versus the actual ones (see Figure 2a $-$ d with the properties H/C, E4/E6, HS/DOC and the ratio of aromatic to upfield aliphatic protons calculated from the 'H-NMR spectra).

The importances and the first and second factor loadings for the nitrogen content of HA are compared in Figure 3. With respect to the pattern of the curves of importances and to the positions and absolute values of maxima in **this** curves the results can be divided into two groups. One group shows characteristical patterns whose maxima can be assigned to absorption maxima of structural groups which are related to the corresponding property (Figure 4 a and b). The importances of the other group show small absolute values of maxima and a great similarity between the curves (see property HSDOC Figure4). Numerical results of the PLS calculations has been arranged in the Table IV.

FIGURE **2 Regression plot for PLS; predicted versus actual values for the** HA **properties: a** - *WC,* ^b- E4E6, **c** - **aromatichpfield aliphatic protons, d** - **HS/DOC**

The numerical values of properties vary little within the sample series but differ between the properties often in orders. For this reason the predictive ability of the IR spectra for each property was proved on the basis of a validation set of samples using a relative SECV-value^[16] (SECV_{rel} = SECV \times 100/c, where c is the mean of the actual values of each property). The smallest prediction errors were obtained for the properties H/C, N/C, and O/C, the nitrogen content and the proportion of upfield aliphatic protons and of the HCO groups determined from the ¹H-NMR spectra of HA and for O/C of FA.

Evidently a relation exists between the pattern of the importances and the prediction ability, which is distinct in the two groups mentioned above. Within the first group only slight errors emerge for the predicted values showing a small variance of the errors (Figure 5a for O/C), because of a strong correlation between the properties **and** the IR spectra. In the second group higher prediction errors appeared showing a large variance of the errors (Figure 5b for HS/DOC). Since characteristics such as HS/DOC or aromatic proton content from ¹H-NMR have only a little response to the IR spectra, we excluded these parameters from further considerations.

FIGURE 3 Importances, first **and** second factor loadings from PLSR for the nitrogen content of HA

Property ^a (HA)	Factors	R^2	SECV	Mean value	rel. SECV (%)
H/C	$\mathbf{2}$	0.851	0.070	1.017	6.9
N/C	9	0.932	0.0048	0.0537	8.9
O/C	8	0.700	0.024	0.533	4.5
N	9	0.920	0.286	3.294	8.7
S	10	0.836	0.262	1.506	17.4
E(280)	1	0.486	156.8	998.1	15.7
E_4/E_6	5	0.560	0.549	3.495	15.7
Protons:					
Aliphatic	15	0.728	1.212	30.85	3.9
Aromatic	3	0.606	2.845	16.46	17.3
HCO	2	0.711	2.704	24.31	8.4
Arom./Aliph.	3	0.711	0.0912	0.541	16.9
HS/DOC	ı	0.360	8.51	40.6	20.9
Property ^a (FA)	Factors	R^2	SECV	Mean value	rel. SECV (%)
H/C	1	0.307	0.150	1.044	14.4
N/C	5	0.737	0.0078	0.0498	15.7
O/C	13	0.604	0.029	0.758	3.8
N	6	0.754	0.401	2.57	15.6
S	8	0.824	0.195	1.488	13.1

TABLE IV Results of the PLS correlation using **full spectra**

^aabbreviations see Tables I, II, and III.

From early studies^[17] it is known that one can use information from the importances to improve the prediction results by means of a suitable selection of the spectral windows. Exemplarily we included only such spectral regions in further calculations for which large values of importances were observed. So we achieved improvements of the forecast of **the** *WC* value from 6.9 % with full spectra calculation to **5.3 9%** using spectral selection, and for the HCO proportion determined from the 'H-NMR spectra from **8.4** % to 6.0 %, respectively.

FIGURE 4 Importances for N. O and H/C (a), E4/E6, upfield aliphatic protons and HS/DOC (b) of HA

FIGURE5 Relative prediction error of the single soils refemng to the soil series (A,K,T,V) for *OK* **(a) and HSDOC** (b) **of** HA

CONCLUSIONS

In this study we showed that several properties of soil **HS,** above all the elemental composition as well as elemental ratios, are readily predictable from the **FTIR** spectra using the **PLSR** method. From the elemental ratios ecologically meaningful conclusions can **be** drawn concerning the soil humus quality, e. **g.** the degree of humification.

The aliphatic and aromatic proton contents of HA, which are important for the interactions with hydrophobic xenobiotics in the environment, cannot predicted with satisfying reliability. This is also true for the absorbance at 280 nm (related

to the aromatic carbon), the E_4/E_6 ratio (inversely proportional to the molecular size), and the **HS/DOC** ratio, which is a measure of the stability of a soil ecosystem. The reason for these unsatisfying results could be that the corresponding properties of the **HS** are generally underestimated in the IR spectra.

The PLS regression worked better with **HA** than with **FA** fractions, probably because of the distinct preparation methods for these fractions, by which the natural differences of **HS** sometimes may be equalized. In the HA preparations more structurally different proportions of **HS** remain preserved.

It is possible that the predictive power of the applied procedure can be improved, for instance by inclusion of more samples with a greater variability of properties of HS and therefore greater differences in the IR spectra.

References

- [I] **F.** J. Stevenson, *Humus chemistry. Genesis, composition, reactions* (Wiley, New York, 1982).
- [2] M. H. B. Hayes, P. MacCarthy, R. L. Malcolm and R. *S.* Swift, *Humic Substances II. In Search of Structure* (Wiley, Chichester, 1989).
- [3] **P.** M. Fredericks, J. B. Lee, P. R. Osborn and D. A. J. Swinkels, *Appl. Specrmsc.* 39,303-316 (1985).
- 141 **S.** Tesch, K.-H. Rentrop and M. Otto, *Fresenius J. Anal. Chem.,* 344,206-208 (1992).
- [5] A. Holmgren and B. Norden. *Appl. Spectmsc.,* 42,255-262 (1988).
- [6] L. J. Janik, J. 0. Skjernstand and M. D. Raven, *Aust.* J. *Soil Res.* 33, 621-636 and 637-650 (1995).
- [7] T.W. Collette, TRAC 16, 24-36 (1997).
- [8] S. Kuwatsuka, A. Watanabe, K. Itoh and *S. Arai, Soil Sci. Plant Nutr.* 38, 23–30 (1992).
- [9] A. Prakash, D.J. MacGregor, in: *Aquatic and terrestrial humic materials,* (R. F. Christman and E.T. Gjessing, eds. Ann Arbor Sci. Ann Arbor, Michigan, 1983), pp. 481-493.
- [101 H. Wold, in: *Systems under indirect observation, causality-structure prediction* (K.G. Joreskog and H. Wold, eds. North Holland, Amsterdam 1981).
- [11] S. Wold, H. Martens and H. Wold, *Proc. Conf. Matrix pencils* (A. Ruhe and B. Kagström eds. 1983).
- [121 P. Geladi and B.R. Kowalski, *Anal. Chim. Acta* **185,** 1-17 (1986).
- [13] H. Martens, *NCC Note STAT/35/87* from the Norwegian Computing Center (1987).
- [141 H. Martens and T. N. Naes, *Multivariate Calibration,* (John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapure 1989).
- [15] D. Haaland and E.V. Thomas, Anal. Chem. **60**, 1193-1202 and 1202-1208 (1988).
- 1161 *S.* Seculik, M. B. Seasholtz, **Z.** Wang, B. R. Kowalski, *S.* E. Lee and B.R. Holt. *Anal.* Chem. 65,835A-845A (1993).
- [I71 V. Centner, D.-L. Massart. 0. E. de Nord, S. de Jong, B. M. Vandeginste and C. Sterna, *Anal.* Chem. 68,3851-3858 (1996).