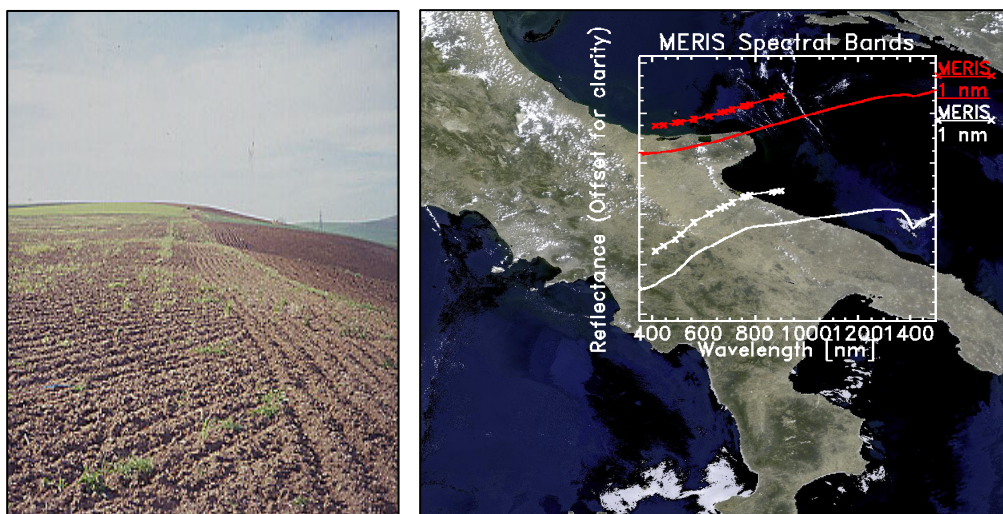


Chemometric Modelling and Remote Sensing of Arable Land Soil Organic Carbon as Mediterranean Land Degradation Indicator

A Case Study in Southern Italy

Kristin Boettcher, Thomas Kemper, Miriam Machwitz, Wolfgang Mehl, Stefan Sommer



EUR 23513 EN - 2008

The mission of the Institute for Environment and Sustainability is to provide scientific-technical support to the European Union's Policies for the protection and sustainable development of the European and global environment.

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information

Address: Stefan Sommer
Via E. Fermi 2749, TP 280
21027 Ispra (VA), Italy
E-mail: Stefan.Sommer@jrc.it
Tel.: +39 0332 78 9631
Fax: +39 0332 78 6394

<http://ies.jrc.ec.europa.eu>
<http://www.jrc.ec.europa.eu>

Legal Notice

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

A great deal of additional information on the European Union is available on the Internet.

It can be accessed through the Europa server
<http://europa.eu/>

JRC 47615

EUR 23513 EN
ISBN 978-92-79-09811-6
ISSN 1018-5593
DOI: 10.2788/98099

Luxembourg: Office for Official Publications of the European Communities

© European Communities, 2008

Reproduction is authorised provided the source is acknowledged

Printed in Italy

*Cover: Photograph of soil from Apulia courtesy of Dr. F. Castellani;
MERIS satellite image and soil spectral measurements*

Table of Contents

I	EXECUTIVE SUMMARY	5
II	PART 1: CHEMOMETRIC MODELLING OF SOIL ORGANIC MATTER FROM LABORATORY DATA	8
1	Introduction and objectives.....	9
2	Material and methods.....	10
2.1	Data set.....	10
2.1.1	Physical-chemical analysis of soil properties	11
2.1.2	Spectral measurements.....	14
2.2	Chemometric modelling.....	14
2.2.1	Multivariate calibration.....	14
2.2.2	Modelling results	20
2.2.3	Application to simulated satellite spectra	27
3	Discussion.....	29
III	PART 2: ESTIMATION OF SOIL ORGANIC CARBON FROM MERIS SATELLITE DATA	31
1	Introduction and objectives.....	32
2	Material and methods.....	32
2.1	Study area and field data.....	32
2.2	Reflectance measurements and pre-processing	33
2.3	Adaptation of laboratory SOC modelling to MERIS satellite data....	35
2.3.1	Re-grouping of laboratory spectra according to landscape units 35	
2.3.2	Sensitivity of models to vegetation influence.....	41
2.4	MERIS satellite data and pre-processing.....	42
3	Model application to MERIS minimum vegetation composites.....	47
3.1	Organic carbon estimates for Apulia and Abruzzo test areas	48
3.2	Comparison with independent data sets.....	50
3.2.1	Comparison for the Apulia test area	50
3.2.2	Comparison for the Abruzzo test area	54
4	Conclusions.....	56
IV	REFERENCES	57
	Acknowledgements	65

I Executive summary

This report gives an overview on the results of a research study exploring the potential of laboratory reflectance spectroscopy and of remote sensing spectral reflectance data for the estimation of top soil organic carbon (SOC), respectively soil organic matter (SOM) content of Mediterranean agricultural soils at regional level as required for addressing soil quality/state aspects in the context of land degradation and desertification assessment and monitoring.

Part of the in-situ data analysis had started under the JRC FP6 exploratory research programme 2003/2004 and as contribution to the DesertLinks project (<http://www.kcl.ac.uk/projects/desertlinks/>) resulting in a web based desertification indicator system for Mediterranean Europe (Brandt et al., 2003).

But most of the overall work, in particular the satellite application, was conducted under the FP6 Integrated Project DeSurvey (A Surveillance System for Assessing and Monitoring of Desertification) overall aiming at complementing assessment of desertification status with early warning of risks and vulnerability evaluation of the involved land use systems, operating primarily on the entire Mediterranean basin and selected non-Euro-Mediterranean test areas in Africa, Asia and South America (<http://www.desurvey.net/>; <http://www.nateko.lu.se/desurvey10/DeSurvey.htm>).

SOM is a key component in soil functioning, as it is an essential part of several processes maintaining soil fertility and stability as well as an important carbon pool, which is increasingly considered a highly relevant component for the mitigation of climate change (Stolbovoy et al., 2008).

It is a complex association of organic material (plant root remains, leaves and excrements), living organisms (bacteria, fungi, earthworms and other soil fauna) and humus, the stable end-product of the decomposition of organic material in the soil by the slow action of soil organisms.

A decrease of soil organic matter content is an indicator for lowered soil quality as well as the degradation of ecosystem functioning. Following the expansion and intensification of agriculture during the 20th century, there is clear evidence of a decline in the SOM contents in many soils as a consequence (Jones et al., 2005). In Southern Europe the decline in soil organic matter content of many soils has been identified as a major element of land degradation processes (Zdruli et al., 2004).

Accordingly, the EC Communication 'Towards a Thematic Strategy for Soil Protection' (Commission of the European Communities [CEC], 2002) and the 'Thematic Strategy for Soil Protection' (Commission of the European Communities [CEC], 2006) consider declining soil organic matter as one of the most serious of eight main threats to soil, which implies the necessity to assessing and monitoring the abundance and geographical distribution of SOM across Europe.

In the context of Mediterranean land degradation and desertification the DesertLinks project had identified SOM as a physical and ecological indicator of desertification risk and included it in the web-based DIS4ME indicator system (<http://www.kcl.ac.uk/projects/desertlinks/accessdis4me.htm>).

Recently, with reference to the work of DesertLinks, the FP 6 project **ENV**ironmental **AS**essment of **SO**il for **MO**nitoring (ENVASSO), has emphasized SOM decline being a key issue of soil degradation in dry sub-humid, semi-arid and arid regions in Europe and consequently proposed topsoil SOC content as one of six headline indicators out of a larger set of so-called soil-linked desertification indicators (SLDIs) which should be applicable at S-European and Mediterranean levels (Kibblewhite et al., 2007).

In most cases organic matter in soil is measured as organic carbon, and the values are converted to SOM using a standard conversion ratio OC:OM of 1:72, being considered satisfactory for providing broad scale environmental assessments and monitoring in support to the policy making process, as SOC is more clearly definable and measurable with less uncertainty (Jones et al., 2004; Jones et al., 2005). Recently the distribution of topsoil SOC was mapped throughout Europe by Jones et al. (2005)

and it was found that the largest areas of soils with SOC contents $\leq 1\%$ are found within the arid, semi-arid and dry sub-humid regions of Mediterranean Europe typically with lowest levels on arable lands. Given the need to avoid further decline in SOM, ENVASSO proposed that current SOC contents should be considered as thresholds for these regions (Kibblewhite et al., 2007).

Currently, measurement of SOC is included in most of the existing field-sampling based soil monitoring networks but due to gaps in the geographical coverage of sampling sites, especially in Southern Europe, ENVASSO estimates a requirement for measurements at about 800 additional field measurement sites in Mediterranean Europe alone (Kibblewhite et al., 2007). That appears at its best an ambitious long term objective for the next 10 years in European conditions while comparable measurement network density may be beyond feasibility e.g. in the non-European southern and eastern Mediterranean regions not speaking of even more affected regions e.g. in the South-Saharan Africa.

Hence there is a vital interest to investigate alternative methods possibly allowing to complement existing networks of monitoring sites and to contribute to the filling of significant gaps in SOC monitoring especially in areas susceptible to land degradation and desertification.

As a consequence, the DeSurvey-IP project (2005-2010) included as part of its monitoring component the investigation of advanced remote sensing techniques to derive indicators related to soil surface parameters such as organic and inorganic carbon or iron content as an exploratory element, to which this study contributed, in particular with its part on the satellite data application.

Against this background our work focused on the derivation of soil organic carbon (SOC) as proxy for soil organic matter (SOM), by investigating:

- ✓ Firstly, in a laboratory study the potential of chemometric modelling using hyperspectral reflectance measurements of available Italian national sets of field samples for the estimation of soil organic matter and carbonate content
- ✓ Secondly, based on results of the laboratory study, the potential of ENVISAT MERIS data for the regional derivation of SOC of arable lands

The strong potential of using hyperspectral in-situ and airborne data for quantitative derivation of soil properties has been discussed and successfully demonstrated in numerous studies world wide. Up-to-date overview on the published literature, current progress and the state-of-the art of the applied methodologies can be found in recent publications such as Ben-Dor et al. (2008) or Stevens (2008).

The application of chemometric models for the quantitative estimation of soil organic matter (SOM) from laboratory reflectance data from samples taken on the regional/national level from Italian sites is explored in Part 1 of this report. In addition, the possibility to transfer the developed models from the spectral resolution of lab/field instrumentation to the one of operational satellite systems has been evaluated, by using the laboratory spectra to simulate the respective soil reflectance signatures of Landsat-TM, MODIS and MERIS.

Soil physical and chemical laboratory analyses results were provided by the JRC-IES SOIL action (formerly JRC FP6 MOSES action). The 376 soil samples, used in this study, were collected for previous projects of the IES SOIL action and its partners within a wide range of environmental settings in Italy. Reflectance measurements were obtained on disturbed soil samples using an ASD Field Spec Pro spectro-radiometer. Data transformation methods (standardisation, vector-normalisation and first and second order derivatives) have been applied on the spectral data. The transformed spectral data have been used for the prediction of SOM and carbonate content using the partial least squares regression (PLSR). The results (R^2 between 0.57 and 0.8) demonstrate the successful application of reflectance spectroscopy combined with chemometric modelling for the estimation of SOM and carbonate content. The calibration models demonstrated a tolerable stability over a variety of different soil types, which is a positive factor for opening the opportunity to use this methodology for

monitoring larger areas. Furthermore it could be shown, that the spectral resolution of the MERIS sensor is sufficient for approximation of the SOC/SOM content from pure soil spectra.

Consequently, the second part of the study focused on the use of MERIS satellite data for the estimation of soil organic carbon content of bare soils at regional scale. The study concentrated on the Apulia region, where we had high density of available field sampling sites, and on parts of the coastal areas of the Abruzzi region South of Pescara, which are known to be amongst the more critical areas in Italy suffering from land degradation problems and desertification risk.

For specific morphological-lithological units simple spectral models, based on soil colour and spectral shape attributes, were built to derive soil organic carbon content.

In order to apply these models to MERIS satellite data, a time series of images covering the years 2003 and 2004 were acquired for Southern Italy. Pre-processing of image data aimed at extracting those pixels with negligible vegetation abundance at least at one date of observation per year, i.e. practically showing pure bare soil signatures only, and consisted of:

- Geometrical co-registration and superposition of images from different acquisition dates
- Derivation of minimum vegetation composites for each year applying simple minimum value criteria for MERIS vegetation indices
- Determination of soil and vegetation abundance at sub-pixel level based on spectral mixture modelling.
- Removal of residual vegetation influence from image spectra

Soil colour attributes (soil lightness, R coordinate of R-G-B model) and coefficients of a second order polynomial fitted through the pixel reflectance signatures were derived from the minimum vegetation composites of both years. The spatial distribution of soil organic carbon was estimated for each year within specific morphological-lithological units in the Apulia region. In addition models could be applied to other regions in Southern Italy. Estimation results showed good agreement with independent field data and the pedo-transfer rules based estimations of Jones et. al. (2004; 2005).

It is worth noting that the field samples used for the development, calibration and validation of the chemometric and spectral models were sampled independently from our study, mainly in the framework of regional to national studies of field based characterisation of soil mapping units in view of their vulnerability to soil degradation (European Soil Bureau and Regione Puglia, 2001). While the results of most other studies were obtained primarily at local levels addressed by highly site specific sampling (e.g. Kooistra et al., 2003), the samples used here were collected country-wide in a wide range of environmental conditions. Nevertheless the performance of our models was close to the results of the more detailed site specific studies, which indicates the robustness of the field/laboratory spectrometry based chemometric models, which is an important pre-requisite to establish a limited number of calibration models that are stable over a variety of different soil types thus opening the possibility to use the methodology quasi operationally in support of SOM field monitoring campaigns of larger areas.

The MERIS based remote sensing study demonstrated that the simplified spectral models resulting from the MERIS data simulation can in principle be successfully applied to real satellite data, however, with some constraints presumably primarily related to the uncertainties in the removal of residual green and senescent vegetation from image spectra. Currently the most reliable way to minimise these effects is the application of a very strict FAPAR minimum value threshold, which however considerably reduces the number of pixels available for the SOM estimates, which in addition may vary from year to year. Nevertheless, the reasonable agreement of the MERIS based SOM estimates with the modelled values of the OCTOP data base indicates the good potential of the method to obtain spatially coherent information on SOM in arable top-soils at landscape level, provided that the method of removing residual vegetation influence from the MERIS spectra can be further improved.

II Part 1: Chemometric modelling of soil organic matter from laboratory data

1 Introduction and objectives

Soil reflectance spectra are very rich in information, which allows inferring chemical, physical or biological properties of soils from a single spectrum. In comparison to conventional analysis, reflectance spectroscopy is faster, more cost-efficient and non-destructive. In this respect it could be an alternative to costly field survey campaigns and has therefore raised interest of the soil science community.

Soil organic matter (SOM) is a soil property that significantly influences the reflectance characteristics of a soil. Stoner and Baumgardner (1981) studied the spectral reflectance of 239 soils mainly from the United States. They identified five characteristic curve shapes on the basis of visual inspection (Figure 1). Each shape is associated with certain soil characteristics. The classes are: organic dominated (organic carbon > 2%, iron oxide < 1%, fine texture), minimally altered (organic carbon < 2%, iron oxide < 1%), iron affected (organic carbon > 2%, iron oxide 1–4%), organic affected (organic carbon > 2%, iron oxide < 1%), iron dominated (iron oxides > 4%).

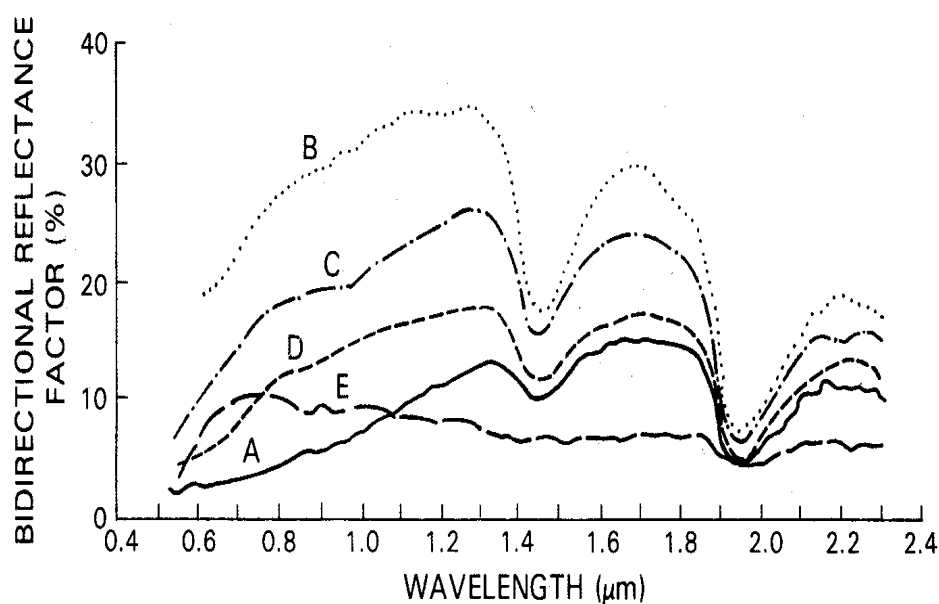


Figure 1. Reflectance spectra of mineral soils representing different soil characteristics: organic dominated (A), minimally altered (B), iron affected (C), organic affected (D) and iron dominated (E) (from Stoner & Baumgardner 1981)

From this and other studies [e.g. Condit (1970), Baumgardner et al. (1985)] it is understood that SOM, on the one hand, influences the general brightness (albedo) of soil and, on the other hand, affects the shape of the spectral continuum in the visible to near infrared part of the spectrum (0.35 – 1.4 μm). In case of soils with low SOM content the reflectance curve tends to be convex and the overall albedo relatively high (Figure 3, curve B). In case of soils with a high content of SOM the overall albedo decreases, but mainly the shape of the soil spectrum becomes more concave from the visible (VIS) to near infrared (NIR) (Figure 3, curve A). However, soil reflectance is also influenced by other parameters like iron content, soil moisture, texture or mineralogy, which make a simple quantification of SOM from soil reflectance difficult. Diffuse reflectance spectroscopy and multivariate calibration techniques have been successfully applied to estimate soil properties from reflectance spectra in different wavelength ranges of the electromagnetic spectrum [VIS, NIR, mid-infrared (MIR)].

Most of the early studies focussed on the VIS part of the solar spectrum (Al-Abbas & Baumgardner, 1972; Krishnan et al., 1980; Leger et al., 1979; Page, 1974; Shields et al., 1968). Krishnan et al. (1980) extended the range to the NIR part and tried to include also variables describing more closely the

shape of the spectral profile. Ben-Dor & Banin (1990; 1994) were the first to apply multivariate statistics on the full spectral range for the prediction of different soil constituents.

Viscarra Rossel et al (2006b) demonstrated the potential of diffuse reflectance spectroscopy using VIS, NIR, MIR regions for the prediction of several soil properties.

Research focussed mainly on spectral calibration and prediction of soil properties using multivariate calibration using methods like multiple linear regression (MLR) (Ben-Dor et al., 1991; Kemper & Sommer, 2002), principal component regression (PCR) (Chang et al., 2001), partial least square regression (PLS) (Cohen et al., 2005; Cozzolino & Morón, 2006; Kooistra et al., 2003; Masserschmidt et al., 1999; McCarty et al., 2002; Reeves et al., 2002) and artificial neural networks (Fidêncio et al., 2002; Udelhoven & Schütt, 2000). Several studies targeted specifically at the determination of soil organic carbon/ organic matter content (Cozzolino & Morón, 2006; Fidêncio et al., 2002; Masserschmidt et al., 1999; Reeves et al., 2002; Stevens, 2008).

Recently, McBratney et al. (2006) presented an interference system for the determination of soil properties whereby latent predictor variables are derived from spectra in contrast to the classical pedo-transfer rule approach for the determination of functional soil properties. In the first step, soil MIR spectra and PLS regression were used to estimate soil pH, clay, silt, sand and organic carbon and cation exchange capacity (CEC). In the second step, predictions and uncertainties of predictions were used in the interference system to obtain soil water content and soil pH buffering from defined pedo-transfer rules (PTF).

Apart from a few studies, which are valid for a large geographical region (Chang et al., 2001; McCarty et al., 2002), chemometric models are often limited to the local scale or even single fields (McBratney et al., 2006; Udelhoven et al., 2001). Furthermore the potential to determine soil properties from recent satellite sensor systems (MERIS, ASTER, MODIS) has not been assessed in this context.

Our objectives were therefore:

- to build a robust chemometric model for the estimation of SOM from laboratory data for soils in different environmental settings in Italy
- to evaluate possibilities to transfer the developed models to operational satellite systems (Landsat TM, MODIS, MERIS)

The model should be applicable on a regional/countrywide level in order to allow large area surveys in contrast to previous local studies. For a better understanding of the model principles a thorough statistical analysis was carried out to overcome the often criticised black-box character of statistically based calibration models.

2 Material and methods

2.1 Data set

Soil samples used in this study consist of two subsets: INTERREG 'Italia II-Albania' and NATURA 2000 (Figure 2). Data had been acquired for other projects and hence, sampling strategies and physical and chemical analysis methods were not harmonised and adapted to the purpose of this study. This may have implications on the quality of the modelling, because the concentrations obtained with different analytical methods may not be directly comparable. Furthermore, not all parameters were analysed for both data sets. Therefore, only those soil parameters were included in the analysis, which were available for both subsets [texture (sand, silt, clay), SOM, cation exchange capacity, pH, total carbonate].

The two data sets will be described in the following section.



Figure 2. Sampling locations from the INTERREG data set (green) and from the NATURA 2000 data set (blue).

2.1.1 Physical-chemical analysis of soil properties

The first subset was prepared within the INTERREG project 'Italia II-Albania' (INTERREG). The Soil & Waste Unit developed in collaboration with the region of Apulia a soil information system. For the information system 250 soil profiles were analysed, which include in total 652 samples from different profiles.

The analysed parameters are:

- Texture (sand, silt, clay) [%]
- pH (H₂O & KCl)
- Total carbonates [%] (tot. carb.)
- Active carbonates [%]
- Organic carbon [%]
- Total sulphur
- Cation exchange capacity [meq/100g]
- Ca & Mg exchange capacity [meq/100g]
- Base saturation, conductivity [mS/cm]

Methods for the laboratory analysis of soil properties followed Italian standard methods¹.

276 samples were selected from the samples provided by the INTERREG project. Selected samples come mainly from the upper horizons, because the main objective of the study is the determination of the soil organic matter content, which is usually limited to the A-horizons. Moreover, the final goal is

¹ Methods are described in: Gazzetta Ufficiale del 25 maggio 1992, decreto ministeriale 11/5/92 "Metodi ufficiali di analisi chimica del suolo" and Gazzetta Ufficiale del 2 settembre 1997, decreto ministeriale 1/8/97 "Metodi ufficiali di analisi fisica del suolo" e "Metodi di analisi fisica del suolo" MIPA coord. M.Pagliai, Franco Angeli 1997

the application of the models to spectra derived by field-spectrometry or remote sensing data, which are also limited to the surface horizons. The descriptive statistics for this data set are listed in Table 1. The soil organic matter concentration was derived from the measured organic carbon content by multiplying it with 1,724, which is derived from the assumed average organic carbon concentration of humus of 58 % (Schlichting et al., 1994).

The average values for SOM are with 1.76 % relatively low. Moreover, the high coefficient of variation (C.V.) and the strong positive skewness and kurtosis values indicate an asymmetrical distribution with few high values. However, the Kolmogorov-Smirnov test for normal distribution showed that the data still can be considered normally distributed with a high significance ($\alpha > 0.01$).

Table 1. Descriptive statistics of selected constituents of 276 surface samples from Apulia.

276 samples	Min	Max	Mean	Std.Dev.	C.V.	Skew.	Kurtosis
pH	5.96	9.67	7.81	0.41	5.28	-0.74	3.36
Tot. carb. [%]	0.00	86.60	11.67	16.26	139.34	2.07	4.92
Clay [%]	0.00	68.90	28.47	13.41	47.12	0.34	-0.35
CEC	4.40	95.00	25.40	9.95	39.17	1.31	8.19
Sand [%]	1.00	88.70	34.24	20.93	61.12	0.79	-0.16
Silt [%]	4.80	86.90	37.29	16.17	43.36	0.02	-0.26
SOM [%]	0.03	12.42	1.76	1.40	79.55	3.42	18.65

The correlation matrix of elements (Table 2) does not highlight any close relation between selected constituents. There are some significant negative correlations mainly between the texture classes. Another negative correlation is found between cation exchange capacity (CEC) and sand. SOM only shows some positive correlations with CEC.

Table 2. Correlation matrix of selected soil constituents for surface samples from Apulia.

276 samples	pH	Total carbonate.	Clay	CEC	Sand	Silt	SOM
pH	1.000						
Tot. carb.	0.194	1.000					
Clay	0.147	-0.139	1.000				
CEC	0.243	0.225	0.300	1.000			
Sand	-0.342	0.020	-0.635	-0.525	1.000		
Silt	0.321	0.091	-0.008	0.427	-0.768	1.000	
SOM	-0.012	0.092	-0.225	0.477	-0.034	0.265	1.000

The second subset used in this study was collected for the NATURA 2000 sites in Italy (NATURA 2000). It consists of 100 samples, which were collected from the different regions in Italy and analysed in a laboratory of ARPA Emilia-Romagna in Podenzano (S.I.L.P.A., 2001) using Italian standard methods².

The following soil properties were analysed:

- Soil texture (sand, silt, clay (g/100g))
- pH
- Carbonate content. (g/100g)
- Organic Carbon (g/100g)
- Nitrogen (g/kg) C/N-Relation
- Phosphorus (mg/kg)

² Methods are described in: "Supplemento ordinario alla Gazzetta Ufficiale, n. 248 del 21 ottobre 1999".

- Cation exchange capacity (meq/100g)

Table 3. Descriptive statistics of selected constituents of 100 surface samples from the NATURA 2000 sites in Italy.

100 samples	Min	Max	Mean	Std.Dev.	C.V. [%]	Skew	Kurtosis
pH	4.75	8.65	6.89	1.05	15.22	-0.22	-1.32
Tot. carb.	0.00	95.00	8.79	18.76	213.38	2.72	7.68
Clay	0.00	70.00	16.28	14.88	91.43	1.29	1.36
CEC	2.27	36.92	20.08	9.89	49.28	0.07	-1.18
Sand	5.25	94.00	51.27	24.88	48.53	-0.33	-1.05
Silt	2.75	67.25	32.46	14.41	44.40	0.06	-0.35
SOM	0.16	17.67	3.25	3.70	114.08	1.96	3.79

Table 4. Correlation matrix of selected soil constituents for surface samples from NATURA 2000 sites in Italy.

100 samples	pH	Total carbonate	Clay	CEC	Sand	Silt	SOM
pH	1						
Tot. carb.	0.542	1					
Clay	0.564	0.217	1				
CEC	0.110	-0.116	0.391	1			
Sand	-0.446	-0.229	-0.855	-0.611	1		
Silt	0.188	0.170	0.443	0.651	-0.844	1	
SOM	-0.171	-0.163	-0.210	0.634	0.023	0.177	1

The comparison of the descriptive statistics highlights some differences between the two data sets. The average values of total carbonates, clay, CEC and silt are considerably higher in the INTERREG data set, while average sand and organic matter concentrations are higher in the NATURA 2000 data set (Figure 3). These differences are statistically significant as confirmed by a t-test ($\alpha = 0.05$).

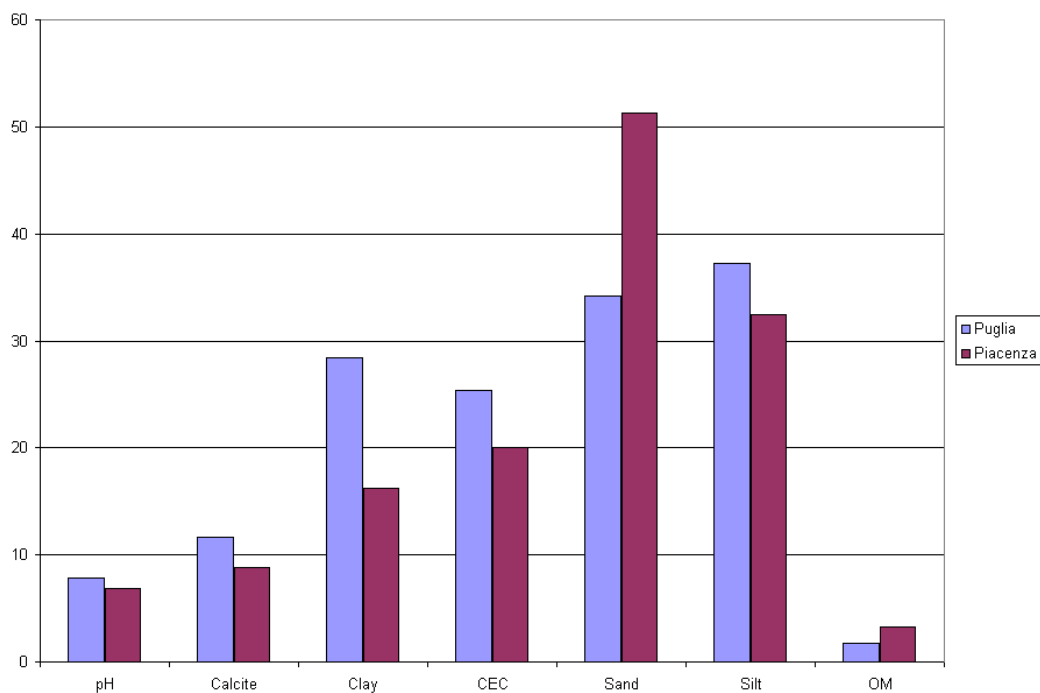


Figure 3. Average concentrations of the two data sets used in this study. Concentrations of total carbonates, clay, sand, silt and organic matter (OM) are weight percentages, pH between zero and seven, cation exchange capacity (CEC) in meq/100g.

The differences are due to the different areas of soil sampling. While the INTERREG data set is restricted to samples from the relatively homogeneous region of Apulia, the NATURA 2000 data set consists of samples from geologically very heterogeneous areas of Italy. Furthermore, the land use for the samples from Puglia was restricted to agricultural land, whereas the NATURA 2000 set includes different land cover classes.

This has also consequences for the modelling, since it can be expected that calibrating a model with one data set may not allow the estimation of soil constituents in the other, because the data ranges are too different.

2.1.2 Spectral measurements

Spectral measurements were obtained with the ASD FieldSpec FR[®] (Analytical Spectral Devices) spectroradiometer, which is a portable spectroradiometer designed for field use with a sampling interval of 2 nm and a spectral resolution of 10 nm. It covers a spectral range from 350 to 2500 nm, interpolated to 1 nm full width half maximum (FWHM). The spectrometer has a line array of Si-photodiodes in the VIS and NIR (350 to 1000 nm); the shortwave-infrared (SWIR) part is measured with two spectrometers with indium gallium arsenide (InGaAs) detectors. Reflectance was measured relative to a calibrated Spectralon[®] and converted to absolute reflectance. As illumination source the ASD High Intensity Reflectance Probe[®] was used. It is designed for measuring small surfaces of not more than 60 mm with a built in light source.

The samples were measured four times with a different field-of-view; each measurement is an average of 50 scans with the ASD FieldSpec FR.

2.2 Chemometric modelling

2.2.1 Multivariate calibration

In this study, the relationship between soil constituents and soil reflectance has been analysed and multivariate calibration is used to predict different soil constituents using the spectra measured in the laboratory.

Using chemometric modelling we can formalize the process of correlating properties to spectra. In chemometrics this modelling is often referred to as multivariate calibration. Multivariate calibration is the collective term used for the development of a quantitative method for the reliable prediction of properties of interest (y_1, y_2, \dots, y_q) from a number of predictor variables (x_1, x_2, \dots, x_p). The goal of the calibration is to replace a measurement of the property of interest by one that is cheaper, faster, or better accessible, yet sufficiently accurate (Figure 4). Developing the calibration model includes stating of the objective of the study, designing the experiment (sample measurement, data treatment), choosing the type of model, estimating its parameters and finally assessing the precision of the predictions (Massart et al., 1998).

A major application of multivariate calibration is in analytical chemistry, specifically the development and application of quantitative predictive calibration models, e.g. for simultaneous determination of the concentrations of various analytes in a multi-component mixture, where one may choose from different spectroscopic methods [e.g. ultraviolet (UV), infrared (IR), NIR, X-ray fluorescence (XRF)]. The application of near infrared analysis (NIRA) to analyse samples, needing little or no pre-treatment, has found widespread use in chemical and food industries (Hildrum et al., 1992; Williams & Norris, 1987). The use of multivariate calibration in the field of environmental science particularly in soil sciences is still limited. Relatively few researchers have extended the methods for the extraction of soil/sediment information. The ultimate goal of multivariate calibration is the indirect determination of a property of interest by measuring the predictor variables only. Therefore, it is not sufficient to describe the calibration data as adequate, but the model must be able to generalise unknown observations. The assessment of the optimum extent to which this is possible has to be done carefully.

When the calibration model chosen is too simple (underfitting) systematic errors are introduced, when it is too complex (overfitting) large random errors may result.

In order to understand chemometric models, a deeper understanding of the underlying chemistry and physics is as important as a set of statistical parameters describing the models. This means for spectral data that a basic understanding of spectra and their errors is useful and that spectral representation should be included in the utility of the data treatment (Geladi & Martens, 1996).

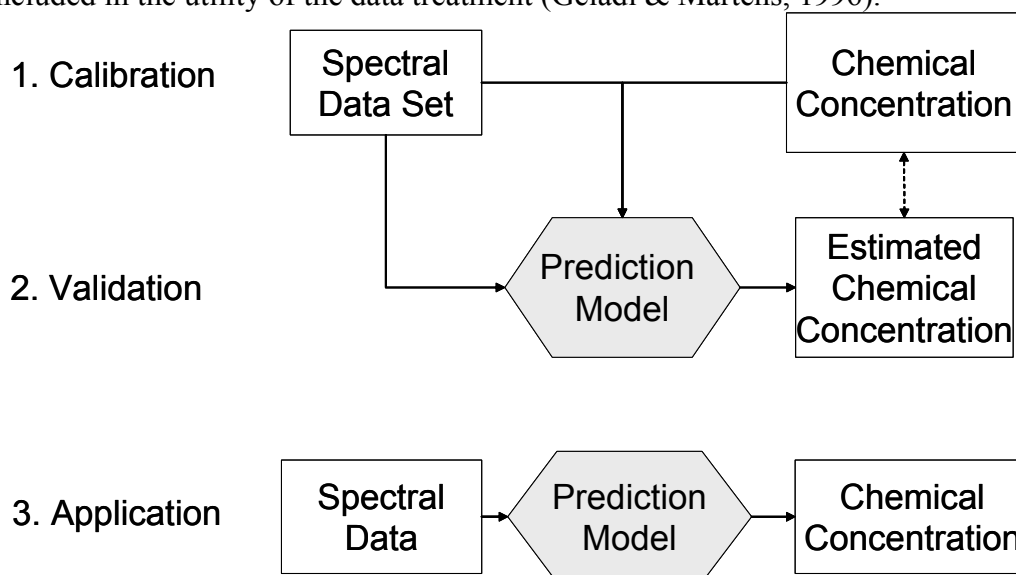


Figure 4. Principles of chemometric modelling.

Data pre-treatment

Spectral data obtained with the spectroradiometer have a very high level of detail, which for the application of multivariate statistics may cause problems due to noise and a high level of collinearity. These can be reduced by resampling the spectra to wider wavelength intervals and to transform the spectra, which enhances subtle but important spectral features.

The spectral resampling smoothes the spectra and reduces the number of wavelengths, which will be used later as independent variables in multivariate calibration, speeding up calculations and reducing the problems of over-fitting (Otto, 1999). The resampling, also termed ‘spectral channel degradation technique’, has been found effective for prediction of different soil properties (Ben-Dor & Banin, 1994). Kemper (2003) found in a study for estimation of heavy metals that the best results were obtained with bandwidths between 10 and 20 nm. Furthermore, the final goal is the application of multivariate calibration to airborne and eventually satellite imaging spectrometers, which usually have a bandwidth between 10 and 50 nm. Accordingly, the spectra were resampled to a bandwidth of 10 nm, which reduces the number of spectral bands from 2151 to 215.

The spectral data were transformed applying three different methods widely used in chemometrics modelling: standardisation, vector-normalisation and first and second order derivatives.

The standardisation transforms the data to a set of data with a mean of zero and a standard deviation of one:

$$x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{s_k} \quad \text{Eq. II-1}$$

where i = row index,

k = column index,

\bar{x}_k = column mean and

$$s_k = \sqrt{\frac{\sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}{n-1}} \quad \text{Eq. II-2}$$

where n = number of objects

The length of a vector $x = (x_1, x_2, \dots, x_n)$ is defined as the square root of the sum of the squared elements in a vector. The normalisation of a data vector to length one is achieved by division of the vector x with the length of the vector (Otto 1999):

$$x_{ik}^* = \frac{x_{ik}}{\|x_k\|} \quad \text{Eq. II-3}$$

where $\|x_k\| = \sqrt{x_{1k}^2 + x_{2k}^2 + \dots + x_{nk}^2}$

Derivatives are a well-established signal processing approach in analytical chemistry. Derivatives of VIS/NIR spectra carry predominantly chemical information of the sample, because albedo effects are reduced and overlapping spectral features are resolved. For example, the OH-absorption of a soil at 1400 nm is shown in the second derivative of this spectrum as a distinct peak. The FWHM is smaller for the peak in the second derivative compared with the original spectrum. Consequently, the smaller FWHM of the derivatives allows overlapping absorption features that are recognised as a single broad absorption in the reflectance spectrum to be distinguished. Baseline shifts are removed because the curvature of a straight line is zero and hence, the derivative of a spectrum plus a constant offset is the same as the derivative of the pure spectrum. The first order derivative has also these effects, but to a lesser extent. It can be interpreted as the slope of a spectrum at each wavelength. Higher order derivatives have the same two basic effects and will resolve overlapping absorptions even better than lower order derivatives. However, they are more sensitive to noise and generate more artefacts than lower order derivatives. Furthermore, they do not have any easily visualised geometric interpretation and hence, have not been widely used in VIS/NIR spectroscopy (Hruschka, 1987).

There are three common methods of calculating derivatives: the finite-differences, Fourier transform and Savitzky-Golay methods; each has its advantages and disadvantages. The Fourier transform is most useful, if the Fourier transform of a spectrum is already available for other purposes. The finite-difference is easiest to calculate but is more sensitive to noise than the Savitzky-Golay method, which includes a smoothing based on a polynomial function. Therefore the latter has been used for this study. The Savitzky-Golay method fits the spectrum in a wavelength interval with a polynomial and then takes the derivative of that polynomial. Savitzky-Golay (1964) used simplified least-square-fit convolution for smoothing and computing of derivatives of a spectrum. The general equation is represented as follows:

$$Y_j^* = \frac{\sum_{i=-m}^m C_i Y_{j+i}}{N} \quad \text{Eq. II-4}$$

where Y is the original spectrum, Y^* is the resultant spectrum, C_i is the coefficient of the i -th spectral value of the filter window, and N is the number of convoluting integers. The index j is the running index for the wavelength of the original spectrum data table. The filter window consists of $2m+1$ points, where m is the half width of the filter window. Savitzky & Golay (1964) provided several tables of coefficient values. Corrections to these tables were provided by Steinier et al. (1972). The use of these tables limits the application of the method. The maximum filter window size of the tables provided by Savitzky & Golay (1964) is 25 ($m=12$). Furthermore, for application of several

combinations of polynomials and derivatives, a general analytical form for calculation would be preferable. According to Madden (1978), equation Eq. II-4 can be rewritten as:

$$Y_j^* = \frac{\sum_{i=-m}^m P_i^{(0)} Y_{j+i}}{N} \quad \text{Eq. II-5}$$

where $P_i^{(0)}$ is the coefficient of the i -th point of the filter in the zeroth order of derivative computation. Accordingly, the smoothed q -th order derivative point of the midpoint is represented as:

$$\frac{d^q \bar{Y}_j}{d x^q} = \sum_{i=-m}^m P_i^{(0)} Y_{j+i} \quad \text{Eq. II-6}$$

Madden's equations allow calculation of the coefficients of least-square-fit convolution from zeroth order to the sixth order of smoothed derivatives.

It has to be mentioned that the derivative spectra obtained with the procedures described above are shorter than the original spectrum by the width of the filter because the procedures cannot be applied correctly at the ends of the spectrum.

The data set of elements, which was chemically analysed using conventional methods, represents quite different properties of the soils, so that the metric differs substantially. This implies absolute values as well as variances. Both types of distortion may affect the statistically based multivariate methods. Scaling, e.g. by standardisation (see Eq. II-1) or range scaling, the data to similar ranges can eliminate these differences (Otto, 1999).

Partial least squares regression

The spectral data described above will be used for the prediction of the chemical soil components. If it is possible to model known data, the models could then be transferred to unknown spectral measurements. However, the first step in the modelling process is calibration. Calibration plays an important role in analytical chemistry. All analytical instrumentation is dependent on a calibration that uses some regression model for a set of calibration samples. This has led to the development of a number of different calibration methods, which cannot all be treated here (more detailed discussions can be found in the standard literature for chemometrics such as Martens and Naes (1989), Massart et al. (1997; 1998) and Otto (1999)).

A valuable classification is given by Martens and Naes (1989). The most basic distinction is between univariate and multivariate calibration. Due to the huge number of variables that are produced with today's spectrometers, we will concentrate on multivariate data.

Another basic distinction is between linear and non-linear calibration, e.g. methods that yield linear or non-linear functions for the X-variables. Most of the models focus on linear models, because theory may indicate a linear relationship, e.g. the Lambert-Beer law of the linear relationship between concentration and absorption. Even when the linear relationship does not hold, it can be a sufficiently good local approximation (Massart et al., 1998).

An important question in multivariate calibration is whether all available X-variables should be used or only a few of them. Many methods can employ only a limited number of predicting variables for purely mathematical reasons. In contrast, the full spectrum methods use all available wavelengths of relevance.

The distinction between direct and indirect calibration is related to the amount of a priori information available. All model parameters are known for direct calibration and can be used to construct a predictor. This may be possible in some simple situations (e.g. linear mixtures), but in most analytical situations, such causal calibration is not possible because not all parameters are known due to unknown interferences.

The last distinction concerns two ways of thinking about the calibration modelling: causal vs. predictive modelling. The latter is referred to as forward or inverse modelling, the causal modelling as reverse or classical; this method was the original and basic approach to calibration and is therefore sometimes referred to as classical calibration. In controlled calibration experiments, the X-variables are designed. They can be set exactly at the values prescribed by the experimental design, e.g. calibration standards of known composition can be prepared. In the classical way of thinking, a spectrum can be expressed as a function of its composition. It may work well for one-constituent problems and simple mixtures, but it does not work with spectra, which are affected by unidentified constituents or whose constituents interact.

In this study partial least squares regression (PLSR) was selected for modelling. It is a multivariate, indirect models, which perform forward calibration from X to Y. It is a linear full-spectrum method. PLSR is a so-called bilinear modelling method such as principal components regression (PCR) or PLSR. It is a powerful, flexible approach and yields informative and reliable predictors $\hat{Y} = f(X)$ by projecting the many variables $X = (x_1, x_2, \dots, x_k)$ onto the few variables $\hat{T} = (\hat{t}_1, \hat{t}_2, \dots, \hat{t}_a)$. The compressed variables are then used as regressors for Y. In this way, the common structures in the X-variables are compressed into a stabilised, more easily interpretable model, leaving out much of the noise and not relevant information (Martens & Naes, 1989).

The data compression applied in PCR gives substantial improvement over ordinary MLR in the modelling of collinear data. The important problem is the choice of the right number of eigenvectors. Nevertheless, sometimes improvements in PCR can be obtained by leaving out some major eigenvectors, since they correspond to phenomena in X of no relevance for modelling Y.

PLSR is an extension of classical PCR. The concept of partial least squares (PLS) was developed by H. Wold (1982) in the field of economy and social sciences and propagated to the field of chemometrics by his son S. Wold (1983). PLSR differs from PCR by using the Y-variables actively during the bilinear decomposition of X. By balancing the X- and Y-information, the method reduces the impact of large, but irrelevant variations in the calibration modelling (Martens & Naes, 1989). PLSR is now dominating the practice in multivariate calibration because of the quality of calibration models and because it is implemented in many commercial software packages (Brown et al., 1996) and Lavine (1998) and references therein). For more detailed discussions of the algorithm, see e.g. Haaland & Thomas (1988a; 1988b) or Massart et al. (1997); (1998). The description used here is based on the review of Wold et al. (2001).

The PLSR model is developed from a training set of N observations (soil samples in this study) with K X-variables (spectral bands) and M Y-variables (soil constituents). These training data form the matrices X and Y with dimensions (N*K) and (N*M), respectively.

The linear PLSR model finds new variables, which are estimates of the latent variables or their rotations. They are few in number and orthogonal. These new variables T are called X-scores and are linear combinations of the original variables X and with the weights (or coefficients) W:

$$T = XW \quad \text{Eq. II-7}$$

The X-scores T, multiplied by the loadings P, good summaries of X, so that the residuals E are small:

$$X = TP' + E \quad \text{Eq. II-8}$$

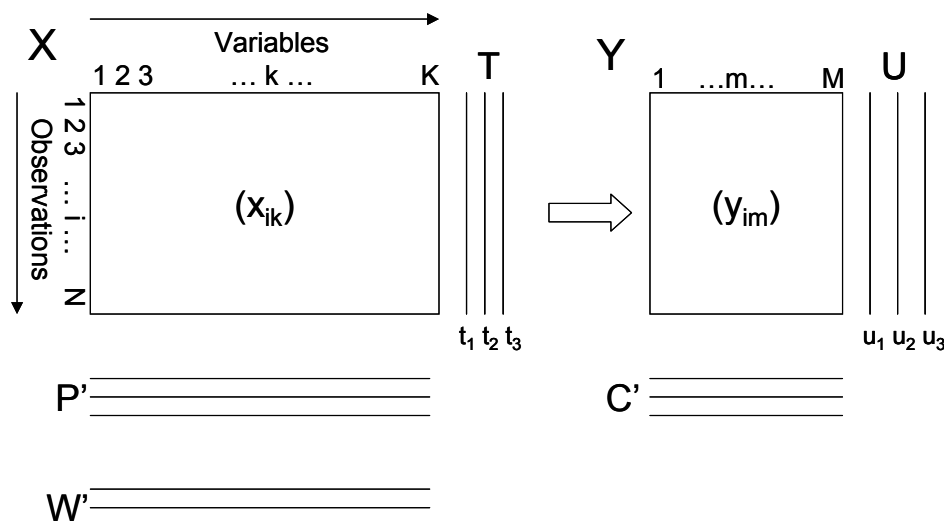


Figure 5: Data arrangement for PLSR from Wold et al. (2001). The variables are described in the text.

With multivariate Y (when $M > 1$), the corresponding Y -scores U , multiplied by the loadings Q are good summaries of Y , so that the residuals F are small:

$$Y = UC' + G \quad \text{Eq. II-9}$$

The X -scores T are also good predictors of Y :

$$Y = TC' + F \quad \text{Eq. II-10}$$

The Y residuals G express the deviations between observed and modelled responses and comprise the elements of the Y residual matrix G .

Because of Eq. II-7, Eq. II-10 can be rewritten to resemble a multiple regression model:

$$Y = XW \cdot C' + F = XB + F \quad \text{Eq. II-11}$$

The PLS-regression coefficients can be written as:

$$B = W \cdot C' \quad \text{Eq. II-12}$$

The score matrices T and U contain information about the objects and their similarities/dissimilarities with respect to the given model and problem for the interpretation of the PLSR model. The weights W and C give information about how the variables combine to form a quantitative relationship between X and Y , thus providing an interpretation of the weights T and U . Therefore, the weights are important for an understanding of which X -variables are important (e.g. having large numerical weight values) and which provide the same information (e.g. having similar profiles of weights) (Wold et al., 2001). All parameters T , U , W , P and C are calculated by a PLSR algorithm. There are several variants developed for different types of data. The algorithms work either with (scaled and centred) original data or with variance-covariance matrices. The most common algorithm is the non-linear partial least squares (NIPALS) algorithm, which is numerically and statistically very stable (Wold et al., 1983). It is also used in this study. The calibration with PLSR was achieved using the N-way toolbox for

Matlab, which is a freely available collection of functions and algorithms for modelling multiway data sets by a range of multilinear models (Andersson & Bro, 2000).

2.2.2 Modelling results

Correlograms

The influence of the different transformations becomes obvious when comparing the correlations between soil constituents, reflectance values and transformed values, respectively (Figure 6). The correlations for the reflectance values are generally relatively low with a maximum correlation coefficient for total carbonates of about $r = 0.55$. They show for all soil constituents negative correlations (that is, for an increase in their amounts there is a decrease in reflectance), except for total carbonate and sand, whose reflectance increases with increasing concentrations, due to the bright colour of calcite and quartzite, respectively. The similarity in the shape of the majority of the correlation curves is associated with the strong positive covariance relationships between some soil constituents. After the standardisation the soil spectra are much more congruent; the albedo differences are removed, retaining the shape differences. Hence, the correlation curves show strong enhanced features. With respect to SOM the most prominent feature is a negative correlation at about $0.6 \mu\text{m}$. This feature is very interesting; because it is clearly linked to the concavity/convexity of the reflectance curve of SOM affected soils. Two other positive correlations at about $1.4 \mu\text{m}$ and $1.9 \mu\text{m}$ are related to water absorption bands. The smaller positive feature at $2.2 \mu\text{m}$ is related to the hydroxyl groups, often related to clay mineralogy. The vector normalised data show a similar correlation structure. The first derivative describes the slope or gradient of the spectrum. Hence, the maxima of the derivative are located at the wavelengths with the steepest slope and absorption maxima or minima have a value of zero. Consequently, there is a negative correlation at about $0.5 \mu\text{m}$ and a positive feature at about $0.8 \mu\text{m}$; the first is caused by the strong increase of reflection of not SOM affected soil, the latter is caused by the convexity of SOM affected soils. Another strong feature is the hydroxyl absorption at about $2.2 \mu\text{m}$, which is expressed by a strong positive feature before and a negative feature after the actual absorption feature.

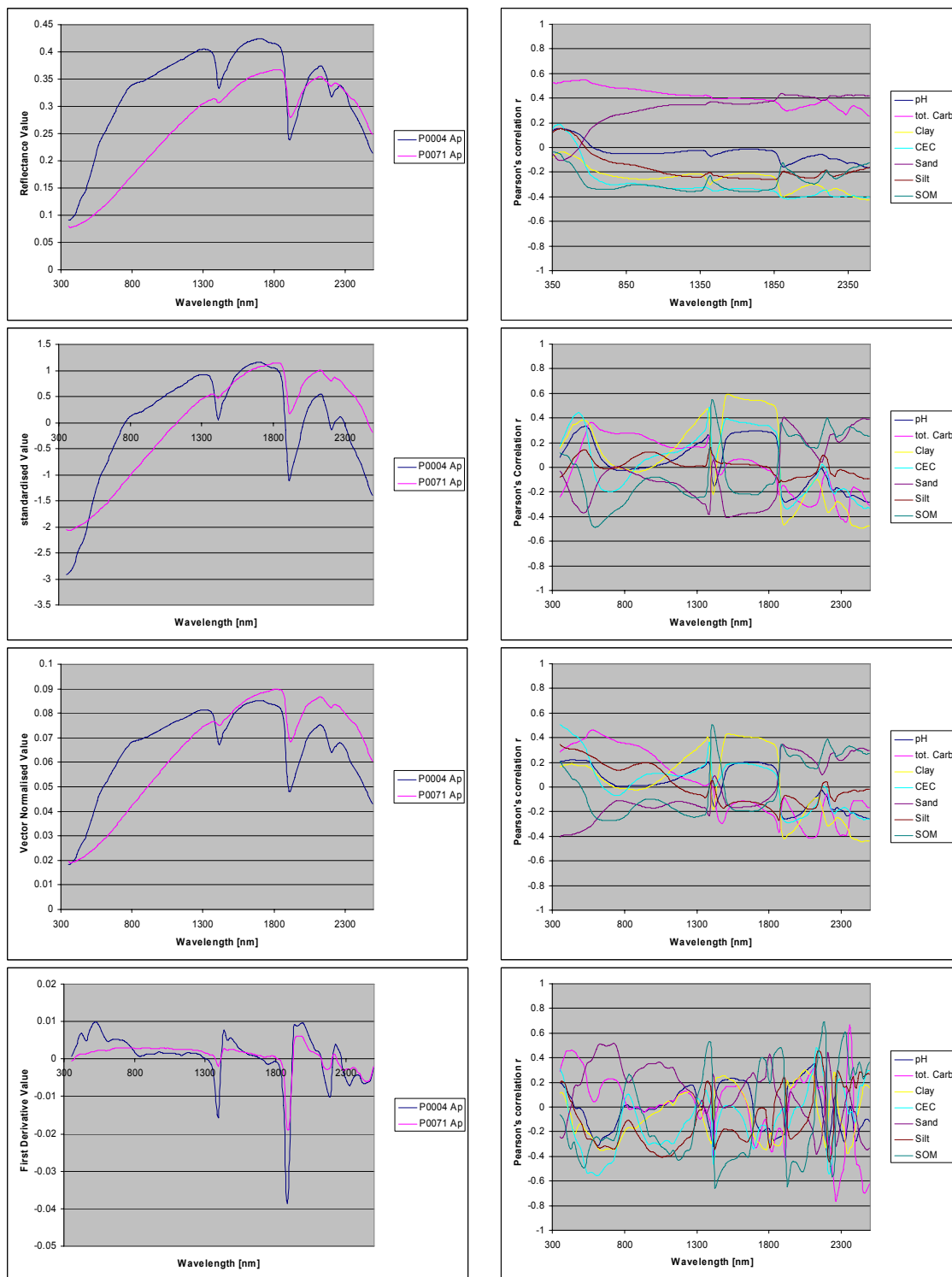


Figure 6. Reflectance and transformed values of selected spectra (left) and correlations between all soil constituents and reflectance for different transformations (right): reflectance, standardised, vector normalised and first derivative (from top to bottom). Sample P0004 with low SOM (0.25 %), sample P0071 with high SOM (5.23 %).

Validation of modelling set-up

Validation of the models is of fundamental importance for multivariate calibration. To achieve this, the data set was split into a calibration set and a validation set and a test data set. The data were sorted from lowest to highest SOM concentration. The samples were divided into two groups by taking odd-

numbered spectra for the calibration group and even-numbered samples for the validation group. In such a way, both sets represent approximately the full range of concentrations.

The prediction quality of the models and the selection of the best model are based on the validation set. This approach gives the best estimate of the model's performance since none of the samples in the validation set were used to build the model. In general, the prediction error is calculated as predictive residuals sum of squares (PRESS) from:

$$PRESS = \sum (y_i - \hat{y}_i)^2 \quad \text{Eq. II-13}$$

The root mean square value of the prediction error is derived from PRESS as:

$$RMSPE = \sqrt{\frac{PRESS}{n}} \quad \text{Eq. II-14}$$

Correlations of measured versus predicted concentration are expressed as coefficients of determination (R^2). Due to the skewed distribution of concentration values, the coefficient of determination would overestimate the quality of the results. Therefore, additional error measures were calculated. The standard error of prediction (SEP) was calculated as standard deviation of differences between reference values and predicted values. From the SEP, two other measures were derived for the evaluation of the model quality. The relative percent difference (RPD) is the ratio of the standard deviation of the reference chemistry to the SEP. It should be as high as possible ($RPD > 3$ for prediction purposes). The relative error ratio (RER) is the ratio of the range in the prediction set to the SEP. It should be greater than 10 (Malley & Williams, 1997).

Single data set

The 276 INTERREG samples were sorted from lowest to highest SOM content and then split into two parts as described earlier. The first part with 138 samples was used for the model calibration. The second part was split again into a validation and a test data set with 69 samples each.

One of the most difficult tasks in using PLSR is determining the correct number of loading vectors to use to model the data. The first vectors in the model are most likely to be the ones related to the constituents of interest, while vectors of higher order generally have less information that is useful for predicting concentration. In fact, if these vectors are included in the model, the predictions can actually be worse than if ignored altogether. Thus, decomposing spectra with PLSR and selecting the correct number of loading vectors is a very effective way of filtering out noise. However, if too few vectors are used to construct the model, the prediction accuracy for unknown samples will suffer since not enough terms are used to model all the spectral variations that make up the constituents of interest. Therefore, it is very important to define a model that contains enough vectors to properly model the components of interest without adding too much contribution from noise.

To avoid building a model that is either an overfit or an underfit, the number of factors where the PRESS plot reaches a minimum would be the obvious choice for the best model. While the minimum of the PRESS may be the best choice for predicting the particular set of samples, it is however, unlikely to be the optimum choice for all unknown samples.

A solution to this problem has been suggested by Haaland & Thomas (1988a) in which the PRESS values for all previous factors are compared to the PRESS value at the minimum. The ratio between these values can be calculated and assigned a statistical significance based on the number of samples used in the calibration set:

$$Fratio_i = \frac{PRESS_i}{PRESS_{\min}} \quad \text{Eq. II-15}$$

This ratio is an indicator of the relative significance of each model to the model with the number of factors at the minimum of the PRESS. The number of factors where the F-ratio falls below a predefined significance level determines the optimum number of factors for a model used for predicting unknowns:

$F_{ratio_i} < F_{\alpha, m, m}$, where $F_{\alpha, m, m}$ is the $(1-\alpha)$ percentile of Snedecor's F distribution with m and m degrees of freedom (m is the number of calibration samples). Haaland & Thomas (Haaland & Thomas, 1988a) found that $\alpha = 0.25$ is a good compromise in practice. However, in this study better results were obtained with a more restrictive significance of $\alpha = 0.05$.

The presence of outliers may have a detrimental effect on the quality of the calibration model. Therefore, the identification of outliers is an important part of the modelling process. Outliers usually arise from some incorrect measurement, whether it is in the concentration data, or in the spectral data. The inclusion of outlier samples in the training set will introduce a bias to the final model. In effect, outlier samples will tend to 'pull' the model in their direction, causing the predicted concentrations of valid samples to be less accurate (or even erroneous) than if the sample was completely eliminated from the training set. A discussion of standard diagnostic tools for the detection of calibration and prediction outliers is given by Martens & Naes (1989). There are two approaches for outlier detection. The first approach is to fit the data with least squares, construct regression diagnostics and remove the outliers. The second approach is to construct estimators that fit the majority of data and examine the residuals from this fit to detect the outliers (Pell, 2000).

A tool for outlier detection is the cross-validation procedure, which is often used as a validation method. The most common approach is the Leave-Out-One (LOO) procedure, where n calibration steps are calculated leaving out one sample at a time. The samples not included are then used to assess the PRESS of the cross validation. The model attempts to account for all the variations in the training data when the calibration calculations are performed, thus rendering approximately the same prediction error for most of the samples should be approximately the same. Samples that have significantly larger concentration residuals than the rest of the training set are known as concentration outliers. The F-test method for determining the optimum number of factors from PRESS analysis is also useful for determining the statistical significance of a sample's concentration residual with respect to the rest of the training set. In this case, the F-ratio value is calculated by:

$$F_{\alpha, 1, n-1}(c_i) = \frac{(n-1)(e_{c_i})^2}{\left(\sum_{i \neq j} e_{c_j}^2\right)} \quad \text{Eq. II-16}$$

where i is the number of the sample being tested, n is the number of samples in the training set and e_{c_i} is the difference between measured and predicted concentration for the i -th sample left out during cross validation. Generally, samples that exhibit probabilities of 0.99 ($\alpha=0.01$) are considered outliers and should be removed from the training set before calculating the final calibration model (Haaland & Thomas, 1988a).

The first models were built to test the general ability to predict the different soil constituents. However, as shown in Table 5 only for total carbonate and SOM significant results could be obtained. The same could be observed for the other transformation methods. Thus, in the following we will concentrate on these two constituents. Nevertheless, the models were built for all seven soil parameters.

Table 5. Prediction results for ‘unknown’ test data with a PLSR model. Calibration with standardised spectra.

	SOM	Total carbonates.	pH	CEC	Clay	Silt	Sand
R ²	0.56	0.72	0.09	0.35	0.27	0.26	0.05
SEP	0.54	0.93	0.47	0.60	0.78	0.68	0.51
RPD	2.45	1.88	0.76	1.53	1.74	1.46	0.73
RER	14.34	10.53	3.72	8.06	7.49	6.68	4.82

The modelling with the different transformations produced very similar results for all transformations (Table 6 and Table 7). The two results (test 1, test 2) refer to the exchange of test and validation set. In the models using standardised and vector normalised spectra there are some consistent differences in quality between the two results, which may be caused by outliers in the data set. The models based on the first derivative seem to handle such differences better, but with a minor overall model performance. Compared to the SOM model the total carbonate estimation perform slightly better.

Table 6. Prediction results for SOM contents of unknown test data with PLSR models for different transformation methods.

	Standardised		Vector normalised		1 st derivative	
	Test 1	Test 2	Test 1	Test 2	Test 1	Test 2
R ²	0.56	0.67	0.56	0.67	0.62	0.63
SEP	0.54	0.73	0.54	0.73	0.52	0.58
RPD	2.45	2.55	2.45	2.55	2.01	1.57
RER	14.34	16.08	14.33	16.05	10.12	8.93

Table 7. Prediction results for total carbonates contents of unknown test data with PLSR models for different transformation methods.

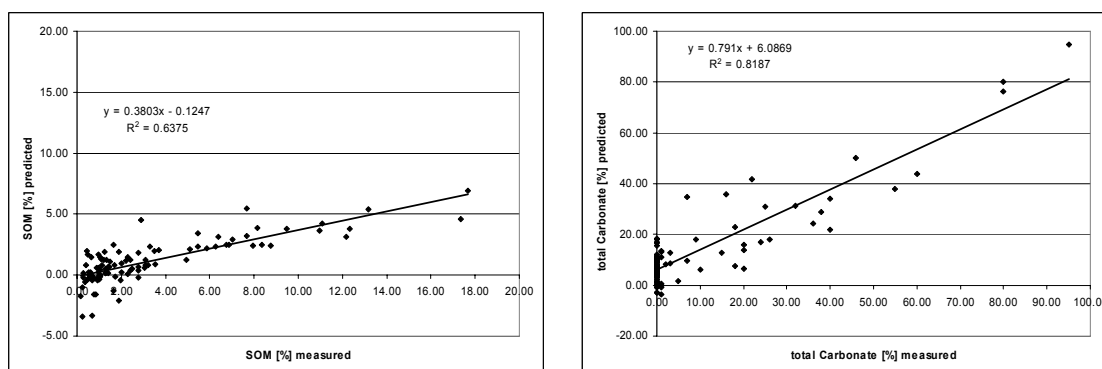
	Standardised		Vector normalised		1 st derivative	
	Test 1	Test 2	Test 1	Test 2	Test 1	Test 2
R ²	0.72	0.71	0.72	0.71	0.76	0.71
SEP	0.93	0.64	0.93	0.64	0.88	0.60
RPD	1.88	2.12	1.88	2.12	2.07	1.93
RER	10.53	8.31	10.53	8.32	11.59	7.89

Combined data set

In the next step the models developed for the first data set (INTERREG) were directly applied to the second data set (NATURA 2000). The model performance on this new data set was only for the first derivative satisfactory (Table 8). This confirms the observation that the derivative-model is less sensitive to influencing factors. However, when analysing also the other error measures and the scatterplots it becomes obvious that the model underestimated the SOM concentrations strongly. This highlights the problem of extrapolation with chemometric models: The model was calibrated for a small data range, but the new data set has a significantly wider range and higher average. In contrast to the SOM prediction, the estimation of total carbonates is possible, because the range of the new data match better the range in the calibration model.

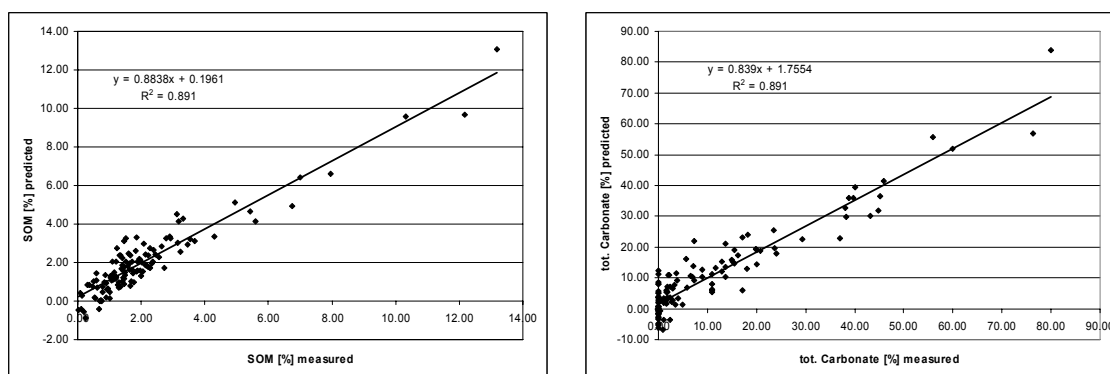
Table 8. Prediction result for the second data set (NATURA 2000) with a PLSR model calibrated and validated with the first data set (INTERREG).

	Standardised		Vector normalised		1 st derivative	
	SOM	total carbonates	SOM	total carbonates	SOM	total carbonates
R ²	0.464	0.379	0.465	0.379	0.638	0.819
SEP	0.963	0.882	0.964	0.883	0.752	0.964
RPD	5.849	8.408	5.865	8.434	0.744	3.139
RER	1.065	1.491	1.067	1.493	4.348	18.831

**Figure 7. Scatterplots of measured versus predicted contents of SOM (left) and total carbonates (right) using first derivatives and PLSR.**

As a consequence, it was necessary to combine both data sets into the calibration model in order to improve the estimation of SOM. The combined data set of 376 samples was split into three data sets with an equivalent data range and size. For the modelling all combinations of calibration, validation and test data sets were applied.

The following gives an example for the results obtained in one calibration run. In the first step the model is calibrated. The results for the calibration data set after the removal of outliers reach an R² of almost 0.9 (Figure 7). During the calibration the validation set was used to estimate the number of factors to be used in the calibration model and to monitor the generalisation capability of the model. The validation results are after the removal of outliers comparable to the results of the unknown test data sets (with R² of 0.79 for SOM and 0.85/0.89 for total carbonates (Figure 8, Figure 9). This proves that the generalization capability is still preserved. The scatterplots show some problems for the prediction in the low concentration range, especially for total carbonates.

**Figure 8. Scatterplots of measured versus predicted contents of SOM (left) and total carbonates (right) for the calibration set using first derivatives and PLSR.**

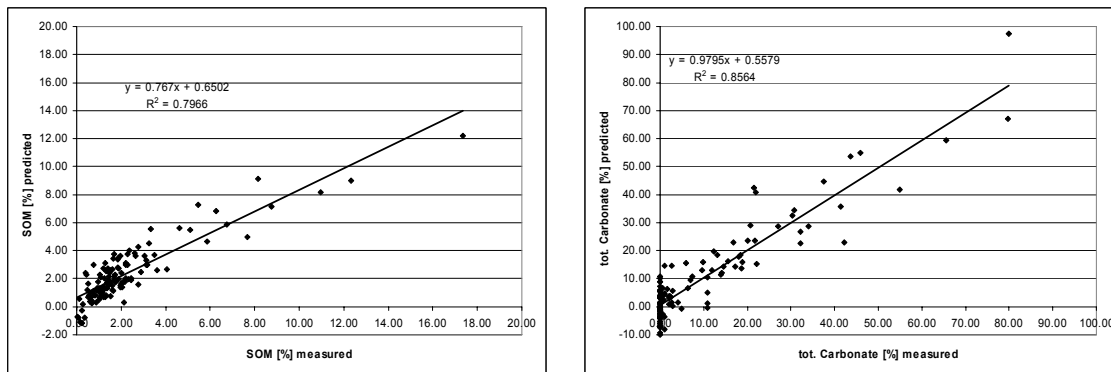


Figure 9. Scatterplots of measured versus predicted contents of SOM (left) and total carbonates (right) for the validation set using first derivatives and PLSR.

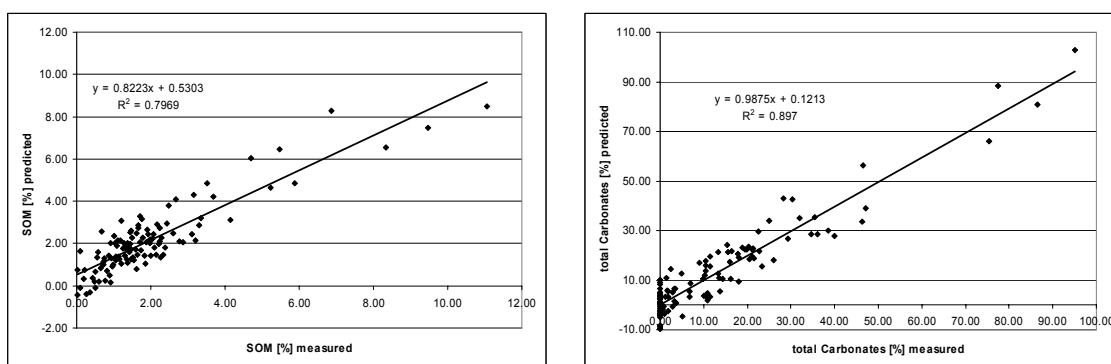


Figure 10. Scatterplots of measured versus predicted contents of SOM (left) and total carbonates (right) for the test set using first derivatives and PLSR.

The prediction results without outlier removal in the predicted data set are listed in Figure 9 and Figure 10. The differences between the prediction results show that there is still some variation in the data sets despite the large amount of data.

Table 9. Prediction results for SOM contents of unknown test data with PLSR models for the combined data set.

SOM	test 1	test 2	test 3
R ²	0.63	0.57	0.67
SEP	0.62	0.82	0.98
RER	8.07	12.60	13.74
RPD	1.23	1.89	2.05

Table 10. Prediction results for total carbonate contents of unknown test data with PLSR models for the combined data set.

Total carbonates	test 1	test 2	test 3
R ²	0.80	0.58	0.81
SEP	0.84	0.84	1.10
RER	17.52	10.85	20.25
RPD	2.86	1.61	3.05

For a better understanding of the calibration models, the factor loadings of the PLSR were analysed (Figure 11), which should be compared to the first derivative spectra and the correlograms in Figure 6. There are four important spectral regions: The first one between 500 and 600 nm can be attributed to the slope of soils in the visible, which is determined strongly by the SOM content. The areas 1350-1450 nm and 1850-1950 nm are clearly related to the major water absorptions and the area between 2130 and 2250 nm is mainly related to the hydroxyl absorption at 2200 nm, which is strongly influenced by the content of clay minerals. There are also some smaller regions, amongst which the most prominent one is the 2350 nm feature. It can be attributed to the calcite absorption and is thus very important for the estimation of total carbonates.

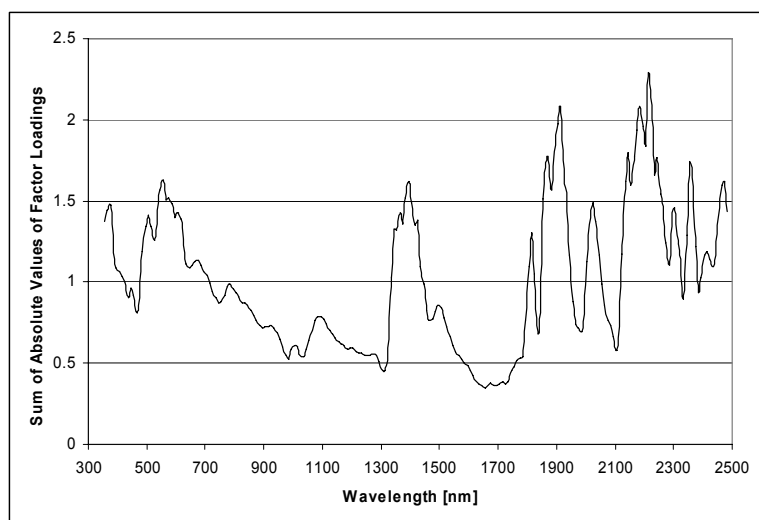


Figure 11. Sum of absolute values of factor loadings for the PLSR calibration with 20 factors.

2.2.3 Application to simulated satellite spectra

Since the regions with high factor loadings are relatively wide it might also be possible to estimate the concentrations of SOM and total carbonates also with lower spectral resolution. Therefore, the original spectra were resampled to the spectral resolution of the multispectral sensors Landsat TM, MODIS and MERIS, which are sensors that are operational and would allow a large scale estimation of SOM. The Landsat mission is part of NASA's Earth Science Enterprise (ESE) and gathers remotely sensed images of the land surface and surrounding coastal regions for global change research, regional environmental change studies and other civil and commercial purposes. Landsat provides repetitive, synoptic coverage of continental surfaces with spectral bands in the visible, near-infrared, short-wave, and thermal infrared regions of the electromagnetic spectrum at a spatial resolution of 30 meters and absolute radiometric calibration. The first Landsat TM (Thematic Mapper) was launched in July 1982, which offers the possibility to use historic timeseries.

MODIS and MERIS are both 'next-generation' satellites, whose scope lies beyond simple mapping but playing a vital role in the development of validated, global, interactive Earth system models able to predict global change accurately enough to assist policy makers in making sound decisions concerning the protection of our environment. With its sweeping 2330-km-wide viewing swath, MODIS sees every point on our world every 1-2 days in 36 discrete spectral bands (only seven used for land applications) at a ground resolution of 250-1000 m. MODIS is a key instrument aboard the Terra (EOS AM) and Aqua (EOS PM) satellites. Terra's orbit around the Earth is timed so that it passes from north to south across the equator in the morning, while Aqua passes south to north over the equator in the afternoon.

MERIS (MEdium Resolution Imaging Spectrometer), aboard ESA's ENVISAT satellite, is an imaging spectrometer that measures the solar radiation reflected by the Earth at a ground spatial resolution of

300 m and allows global coverage of the Earth in 3 days. The primary mission of MERIS is the measurement of sea colour in the oceans and in coastal areas. However, the MERIS instrument is capable of retrieving a variety of geophysical information also interesting for land applications. The band settings of Landsat TM and MODIS spectra only allow a standardisation and vector normalisation due to the limited and irregular spaced number of bands (Table 11 & Figure 12). For MERIS also the first derivative was calculated.

Table 11. Overview of satellites spectral band settings. Wavelength (WL) and Full Width Half Maximum (FWHM) in nm.

<i>Landsat TM</i>		<i>MODIS</i>		<i>MERIS</i>	
WL	FWHM	WL	FWHM	WL	FWHM
486.57	70.00	469.00	20.00	412.55	9.93
570.93	80.00	555.00	20.00	442.40	9.95
660.97	90.00	645.00	50.00	489.74	9.97
836.88	140.00	858.50	35.00	509.70	9.97
1676.65	200.00	1240.00	20.00	559.63	9.98
2215.48	270.00	1640.00	12.00	619.62	9.99
		2120.00	50.00	664.64	9.99
				680.90	7.49
				708.43	10.00
				753.47	7.49
				760.35	3.74
				778.50	15.00
				864.83	20.00
				884.85	9.99
				899.86	9.99

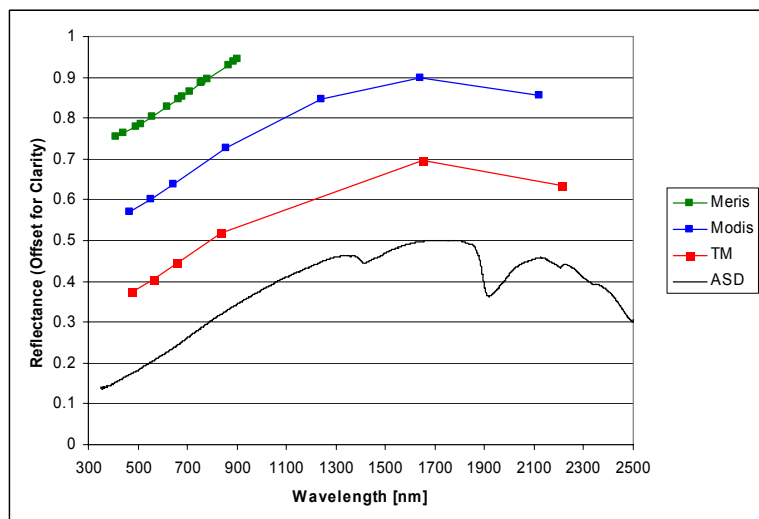


Figure 12. Spectral reflectance of sample P100_4 for the high resolution spectrum (1 nm) and the different satellites. Note the offset for clarity.

For Landsat TM and MODIS no significant results could be obtained, because the number of bands is too low and it is not possible to model well the spectrum. Moreover, the PLSR algorithm can compute only as many factors as spectral bands are available. As could be seen with the calibration of the 10 nm data, as much as 20 factors were necessary to model well the data. In contrast, with MERIS simulated spectra the prediction of SOM was possible when using the first derivative (Figure 13). The estimation of total carbonates was not possible, because the MERIS band setting is limited to the VIS and NIR,

and thus the calcite absorption is not covered. But it is this spectral region that is very important for the mapping of SOM.

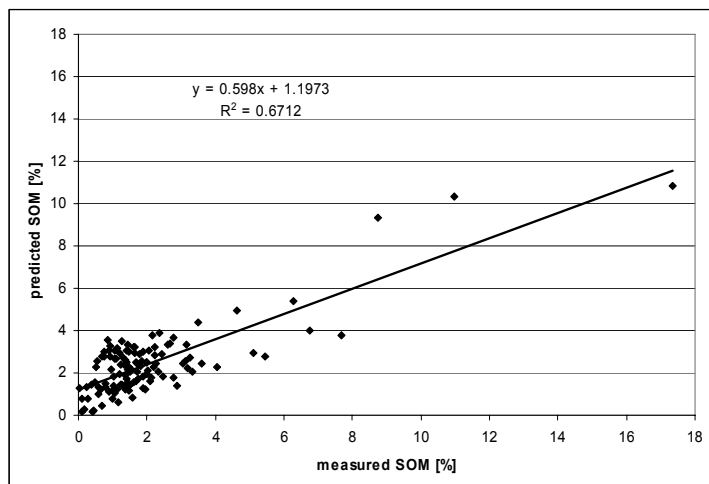


Figure 13. Scatterplots of measured versus predicted contents of SOM for the validation set using first derivatives and PLSR (13 factors).

3 Discussion

The soil properties discussed in this paper, SOM and total carbonates, have been previously predicted by laboratory reflectance spectroscopy in combination with PLSR or artificial neural networks (Kooistra et al., 2003; Reeves et al., 2002; Salgó et al., 1998). These studies found R^2 values for SOM between 0.51 and 0.9. These results are comparable with the best results obtained in the present study ($R^2 = 0.8$). Results could probably be better, if problems due to inconsistent chemical analysis of SOM could be harmonised and the time difference between chemical measurement and spectral measurement is minimised, which was not the case in this study.

Despite the fact that SOM does not exhibit distinct absorption features like for example calcite, the transformations enhanced more subtle features like the concavity in the VIS-NIR part of the spectrum. Particularly the first derivative, which is a measure of slope, worked well in that respect. Nevertheless, the high number of factors used in the PLSR model (up to 20 factors) indicates some non-linearities in the relationship. The spectrally important areas identified in this study, correspond well with the results of other studies (Kooistra et al., 2003; Salgó et al., 1998).

The results of most other studies were obtained for small study areas, whereas the samples used here were collected country-wide in a wide range of environmental conditions. Having a calibration model that is stable over a variety of different soil types opens the possibility to alleviate monitoring of larger areas. Today, large area surveys, like ICP Forest survey (Vanmechelen et al., 1997) or the geochemical baseline programme FOREGS (Association of the Geological Surveys of the European Union (EuroGeoSurveys)/ the Geological Survey of Finland, 2005; Salminen & Tarvainen T., 2006), rely on an enormous amount of samples that are analysed in laboratory, which is very time and cost intensive. Using a chemometric model in combination with a spectroradiometer, such large area surveys could be made faster and cheaper even with a denser sampling grid. Samples would have to be collected in the field and the reflectance measured in the laboratory, which would provide almost direct the SOM content. For a quality assessment, a part of the samples should still be analysed conventionally and could be integrated afterwards into the calibration model for model stabilisation. On a European scale (probably also on a regional scale) such an approach would imply a stratification based on a pedo-climatologic zoning.

Nevertheless, these measurements are point samples, which would have to be interpolated in one way or the other for deriving a synoptic view. Reflectance spectroscopy can also be used at macroscopic scales and can be carried out remotely without direct contact to the sample. As demonstrated in this study, the spectral resolution of the MERIS sensor onboard of the ENVISAT satellite is sufficient to estimate roughly the SOM content. Other operational sensors (Landsat TM, MODIS) do not have a sufficient spectral resolution. However, the MERIS spectra used in this study were simulated laboratory spectra. Real image spectra suffer quality losses from a number of influencing factors. Besides clouds and the general influence of the atmosphere on the signal, the mixing of different surface components has a strong impact on the reflectance signal, particularly for a sensor with a spatial resolution of 500 by 500 m². The problems of mixtures occur already with higher resolution sensors and even when trying to estimate soil constituents using spectra measured directly in the field (Kemper, 2003; Kooistra et al., 2003).

This study demonstrated successfully the application of reflectance spectroscopy combined with chemometric modelling for the estimation of SOM and other soil constituents. However, several aspects should to be analysed in order to enhance the models and to apply it eventually to satellite imagery. Future work should focus on the following aspects:

- Robustness of models for field use: influence of particle size, soil humidity, vegetation components in a mixed sample
- Pedo-climatologic zoning: influence of geo-environmental setting on the model performance
- Application to large scale satellite imagery: influence of scale on the reflectance signal of the pixel, mixed pixels

In order to enhance the model there is still a large amount of data available, which could be integrated into the modelling for model validation and further model improvement: For the eco-pedological map many more samples were analysed. And another independent data set sampled in the Fortore Beneventano area, central Italy is available (Leone & Sommer, 2000).

III Part 2: Estimation of soil organic carbon from MERIS satellite data

1 Introduction and objectives

Soil organic carbon (OC) is a key component in ecosystem functioning, as it is an essential part of several processes and an important carbon pool, which may be relevant for the regulation of the greenhouse effect. Various important soil characteristics concerning the nutrient budget are determined through soil organic matter, such as cation exchange capacity, soil acidity, the capacity of binding organic chemicals. Soil organic matter also influences the physical characteristics of the soil with regard to soil erosion processes, like soil structure and aggregate stability. A decrease of soil organic matter content is an indicator for lowered soil quality as well as the degradation of ecosystem functioning. The decline in soil organic matter content of many soils in Southern Europe is becoming a major process of degradation according to Zdruli et al. (2004). To assess the status and to monitor changes of soil organic carbon content, reliable estimates at regional and global scale are needed.

Remote sensing imagery could be an alternative to conventional methods to assess and monitor OC content of arable lands at regional scale. Condit (1970) and Baumgardner et al. (1985), among others, described that OC significantly influences the general brightness and shape of the spectral continuum in the visible to near infrared part of the spectrum. Soil colour has been recognised as a good proxy indicator for soil organic carbon/ soil organic matter content which is widely used in soil surveys. Many studies found significant correlations between soil colour attributes and soil organic carbon content (Blume & Helsper, 1987; Jarmer, 2005; Spielvogel et al., 2004; Viscarra Rossel et al., 2006a). The relevant VIS-NIR spectral region is well recorded by the Medium Resolution Imaging Spectrometer (MERIS) and therefore soil colour attributes and parameters to describe spectral shape could be derived from MERIS satellite data.

Accordingly, our objectives were the following:

- Development of models to derive soil organic carbon (OC) content from MERIS satellite images based on laboratory reflectance spectra resampled to MERIS spectral resolution
- Definition of pre-processing strategy to minimise vegetation related effects in the satellite imagery
- Mapping of soil organic carbon content in Southern Italy
- Comparison of results with independent modelling results

A set of laboratory reflectance measurements, limited to soils of the Apulia region (Southern Italy), were used for model calibration and validation. To derive stable models and to ensure the applicability to areas outside the calibration site, stratification into morphological-lithological units (regional scale) was implemented. As the retrieval of soil properties from remote sensing data presumes the absence of vegetation in a pixel a multi-temporal approach is needed, combining different acquisition dates in the assessment procedure. Pre-processing focussed on the derivation of a minimum vegetation composite and a subsequent decomposition of image spectra into soil and vegetation fractions through spectral unmixing.

2 Material and methods

2.1 Study area and field data

The study area is located in the South of Italy covering the administrative regions Apulia, Abruzzo, Marche and Basilicata (geographic coordinates: upper left corner 12°30'E/ 43°30' N; lower right 18°45'E/ 39°6'N). Sampling was limited to the Apulia region, situated in the south-eastern part of the study area (Figure 14). The Apulian region is a mostly flat to slightly sloping lowland except the Gargano Massiv, which is situated in the North-East. Agricultural lands cover about 72 % of the region, mainly cultivated with cereals (27.4 %), olive trees (23.9 %), vineyards and vegetable crops (Todorovic & Steduto, 2003).

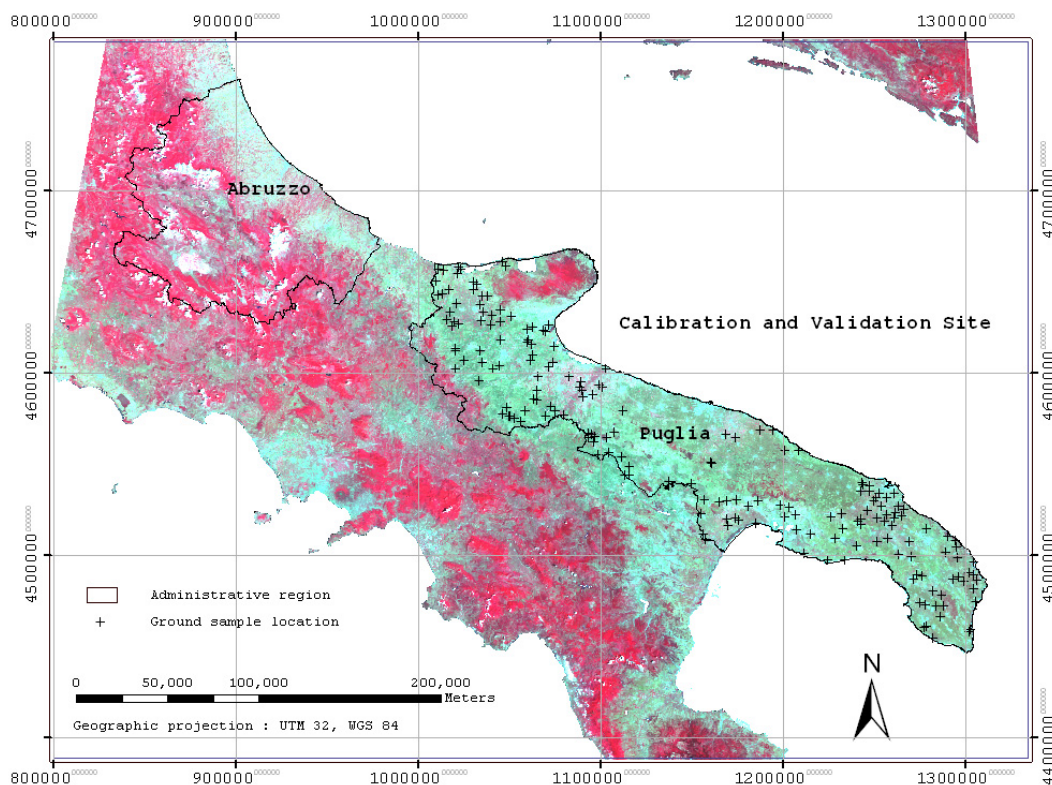


Figure 14. Study areas Apulia and Abruzzo region (Southern Italy). MERIS satellite image from 11/06/2003 (band combination 13-8-5) superimposed by administrative boundaries of study regions.

Soil samples for the study were taken from the INTERREG project 'Italia II –Albania' and are described in Part I of this report. Only samples from the top horizon (N=185) were included in the analysis. The average value for organic carbon is with 1.17 % relatively low. Moreover, a strong positive skewness (3.36) and kurtosis values (8.4) indicate an asymmetrical distribution with few high values.

2.2 Reflectance measurements and pre-processing

Reflectance measurements were obtained on sieved (fine fraction < 2 mm) soil samples using an ASD Field Spec Pro (Analytical Spectral Devices) spectroradiometer. Reflectance was measured relative to a calibrated Spectralon[®] and was converted afterwards to absolute reflectance. As illumination source the ASD High Intensity Reflectance Probe[®] was used. Samples were measured four times with a different field-of-view; each measurement is an average of 50 scans. Reflectance data were degraded to MERIS spectral characteristics (Table 14).

The close relation between soil colour and organic matter is widely applied in soil surveys to assess the status of soil organic matter in the field (AG Boden, 1996). Many studies found significant correlations between soil colour attributes and soil organic carbon content (Blume & Helsper, 1987; Jarmer, 2005; Spielvogel et al., 2004; Viscarra Rossel et al., 2006a). Viscarra Rossel et al. (2006a) compared correlations between soil colour parameters from different colour space models (MUNSELL HVC, R-G-B, CIE XYZ, CIELUV and CIELAB) and organic carbon content for a wide range of soil samples from France, Canada and Australia. Organic carbon was correlated with the lightness parameters of different colour systems and to a lesser extent to their chromaticity. The strongest single parameters

relationship was obtained with the R coordinate of the R-G-B model. Good correlation was also obtained for the L and v^* parameter of the CIELUV model. Jarmer (2005) found significant correlations between soil organic carbon and CIE chromaticity coordinates for soils developed on Cretaceous limestone in Israel.

Chromaticity coordinates (CIE xyz) of the 1931 colorimetric system of the Commission Internationale de l'Éclairage (CIE) (Wyszecki & Stiles, 1982) were included in our analysis. Furthermore soil colour parameters of the R-G-B, CIELAB and CIELUV colour models (Commission Internationale de l'Éclairage, 1978) were calculated from laboratory reflectance spectra. The CIE XYZ tristimulus values, obtained from the reduced resolution of the MERIS spectra, were corrected to full spectra chromaticity coordinates using empirical equations obtained from the laboratory dataset. Colour attributes for the CIELAB and CIELUV colour space were calculated from corrected CIE XYZ tristimulus values.

Besides soil brightness, soil organic matter influences the shape of the spectral continuum. Hill & Schütt (2000) used a parameterisation of the spectral shape through the coefficients of a polynomial for the estimation of soil organic carbon. Following this approach we described the shape of the reflectance curve by coefficients of a second order polynomial fitted through the vector-normalised reflectance spectra (413 nm – 890 nm):

$$p(\lambda) = b_3 + b_1 \lambda + b_2 \lambda^2 \quad \text{Eq. III-1}$$

where p = fitted polynomial

b_1, b_2, b_3 = coefficients of the polynomial

λ = wavelength.

The equation for the vector-normalisation of spectra is given in Eq. II-3 in part I of this report.

2.3 Adaptation of laboratory SOC modelling to MERIS satellite data

2.3.1 Re-grouping of laboratory spectra according to landscape units

Soils are complex dynamic systems, formed and developed as a result of the combined effect of climate and biotic activities, modified by topography, and acting on a parent material over periods of time (Brady & Weil, 1999). Chemical, physical and mineralogical properties of soil, which are reproduced in distinct spectral properties, are modified during soil development.

Many studies to predict soil organic carbon content from reflectance measurements were conducted successfully at local scale. Predictive ability may diminish for the calibration of spectral data over heterogeneous areas in terms of geology or soil type (Stevens, 2008; Udelhoven et al., 2003). Therefore, Baumgardner et al. (1985) suggested to study local geographic areas of similar parent material separately when trying to relate soil reflectance to other soil parameters.

Spatial variability of soil properties (excluding random variability) arises mainly from pedogenetic factors (climate, organisms, relief, parent material and time) and land use and management practices (Stevens, 2008). Spatial variability is scale dependent because soil properties are influenced by driving factors acting at different levels (Lin. et al., 2005).

A strategy to decrease spatial variability of soil properties could be a stratification based on “uniform” initial conditions at a specific scale. We propose stratification into units with a common landscape genesis, which reflects in characteristics of the actual mesorelief and lithology.

This approach seemed to be effective to improve soil organic matter predictions from ASTER satellite images as shown in a previous study conducted with the data set from the Apulia region (Machwitz, 2005). In addition, it would allow the transfer of models to areas with similar morpho-lithological characteristics beyond the calibration site within the same climatic region. Note that only arable land is considered in this study and therefore, variability due to different land use and management practice is negligible.

For the classification of soil samples into basic morphological-lithological units, classification into (nano)morphochores (*ambito / paesaggio di riferimento*), provided by the database of the soil information system (INTERREG II project) at a scale of 1:50,000 for each sample point was used. Data were successively aggregated to bigger units, similar in morphology and parent material. Three basic stratification units for modelling OC could be defined (see Figure 15):

- a) River terraces and floodplains characterised by alluvial deposits (unit 1),
- b) Morpho-structural depressions characterised by unconsolidated deposits (unit 2) and
- c) Structural highs characterised by limestone (unit 3).

For the transfer of models to the Abruzzo region, the correspondence of identified units with eco-pedological units from the eco-pedological map of Italy (European Soil Bureau, 2003.) has been assessed. The eco-pedological map of Italy delineates eco-pedological units at a spatial scale of 1: 250,000 (minimum mapping unit 1.5 km²). Stratification was also performed according to criteria: mesoform and lithology.

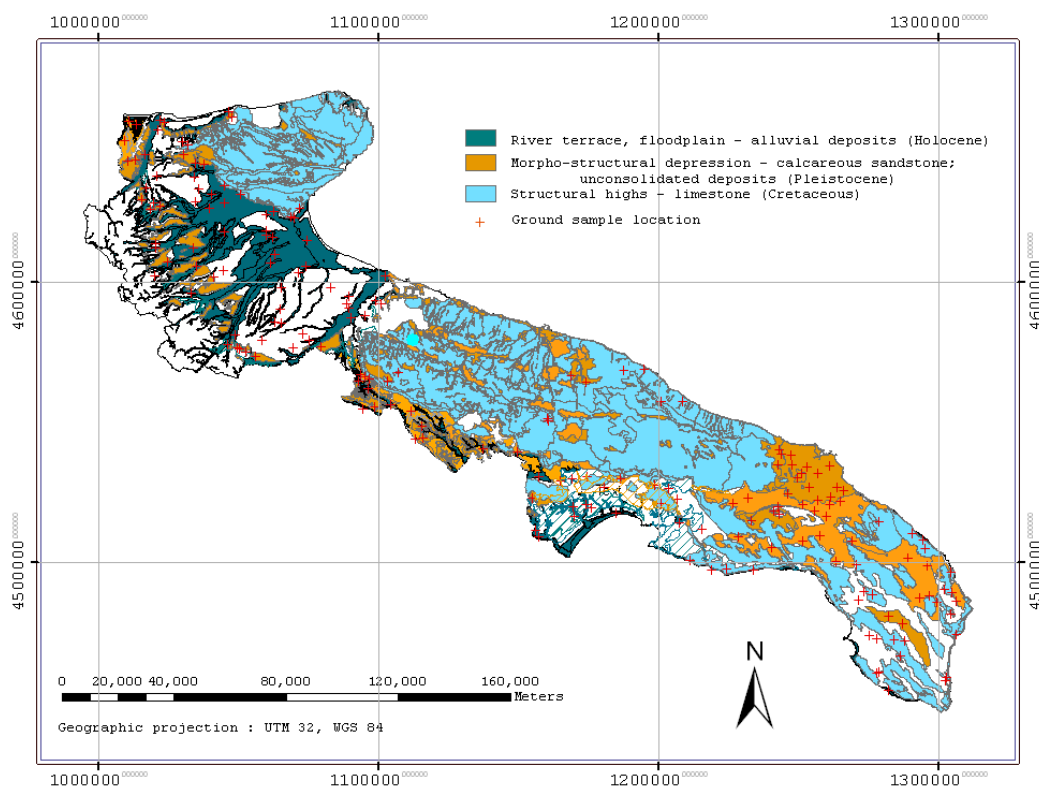


Figure 15. Stratification and sampling locations in the Apulia region (Southern Italy) for the estimation of soil organic carbon content from MERIS satellite data.

Despite stratification into morpho-lithological units, the ternary diagram of soil texture in Figure 16 reveals strong heterogeneity of texture within strata with a dominance of finer texture for soils developed on alluvial deposits. Heterogeneity of soil texture might hinder modelling of soil organic carbon content in function of soil reflectance since soil texture influence soil reflectance (Schulze et al., 1993).

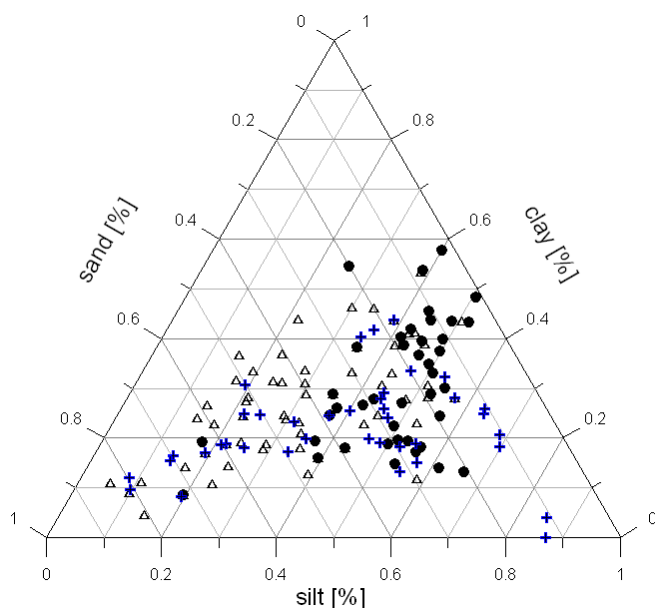


Figure 16. Ternary diagram of soil texture for stratification units: ● River terrace/ floodplain- alluvial deposits, Δ Morpho-structural depression – unconsolidated deposits and + Soils developed on limestone.

A correlation analysis was performed between spectral parameters (colour and shape) and soil organic carbon content to identify important parameters for modelling OC. The results for each morphological-lithological unit are listed in Table 12. Samples from areas characterised by alluvial deposits showed highest linear correlation of soil organic carbon content with coefficients of the second order polynomial fitted through the spectral reflectance curve. Significant correlations were also obtained for CIE chromaticity coordinates (CIE xyz), CIELAB a and b component, the CIELUV v component and the difference of the first derivative at 890 nm and 620 nm ($R'_{890\text{nm}} - R'_{620\text{nm}}$). For the second unit, correlations were highest (negative) with soil lightness L (CIELAB, $r = -0.76$, $\alpha = 0.01$), the X and Y (both $r = -0.76$, $\alpha = 0.01$) coordinate of the CIE colour space. Soils developed on limestone showed similar correlation patterns as soils on alluvial deposits with highest correlation between the coefficients of a second order polynomial and soil organic carbon.

Regression models for unit 1 and 3, areas characterised by alluvial deposits and limestone, were built using the coefficients of a second order polynomial (see Figure 17) and the parameter $R'_{890\text{nm}} - R'_{620\text{nm}}$, whereas soil lightness L, and the R component were chosen for the unit characterised by unconsolidated deposits. Due to the small sampling size, the leave-out-one procedure was applied for model validation. The procedure assures a nearly unbiased validation according to Otto (1999). The regression equations and modelling errors (cross-validation) are shown in Table 13.

Table 12. Correlations between soil organic carbon and spectral parameters for morphological-lithological units in the Apulia region.

Spectral parameter	River terrace, floodplain, alluvial deposits (N=39)	Depression, calcareous sandstone und Pleistocene deposits (N=56)	Structural highs; limestone (N=30)
	PEARSON correlation coefficient	PEARSON correlation coefficient	SPEARMAN correlation coefficient
CIE X	-0.40*	-0.76**	-0.28
CIE Y	-0.40*	-0.76**	-0.28
CIE Z	-0.24	-0.66**	0.1
CIE x	-0.6**	-0.23	-0.63**
CIE y	-0.7**	0.10	-0.63**
CIE z	0.65**	0.17	0.64**
CIELAB L	-0.4*	-0.76**	-0.28
CIELAB a	0.50**	0.08	-0.25
CIELAB b	-0.69**	-0.26	-0.68**
CIELUV u	-0.24	-0.11	-0.52**
CIELUV v	-0.69**	-0.32*	-0.65**
R-G-B R	-0.37*	-0.75**	-0.24
R' _{890nm} - R' _{620nm}	0.68**	-0.54**	0.77**
coefficient b ₁ 2 nd order polynomial	0.76**	0.29*	0.80**
coefficient b ₂ 2 nd order polynomial	-0.79**	-0.24	-0.78**
coefficient b ₃ 2 nd order polynomial	0.79**	0.17	0.74**

* Correlation is significant at the 0.05 level

** Correlation is significant at the 0.01 level

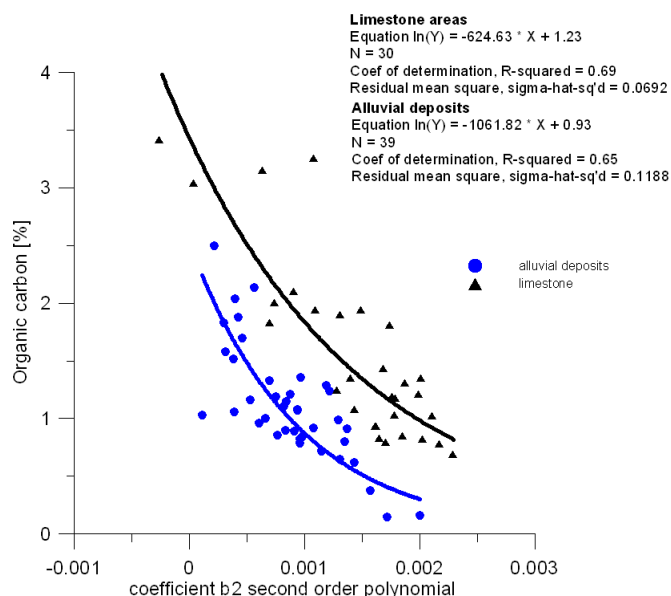
**Figure 17. Relationship between OC content [%] and the coefficient b₂ of a second order polynomial fitted through the vector-normalised reflectance curve for soils developed on alluvial deposits and limestone.**

Table 13. Regression equations for the estimation of soil organic carbon content based on laboratory spectra resampled to MERIS spectral resolution.

	Structural highs; limestone	River terrace, floodplain, alluvial deposits		Depression, calcareous sandstone und Pleistocene deposits	
	Coefficient b_2 2 nd order polynomial	Coefficient b_2 2 nd order polynomial	$R'_{890nm}-R'_{620nm}$	R (R-G-B)	L (CIELAB)
OC range [%]	0.69 – 3.4	0.15 – 2.5		0.16 -1.5	
N	30	39	39	56	56
Regression equation	$Y=3.43e^{-624.63x}$	$Y=2.53e^{-1061.8x}$	$Y=1.62e^{4318.47x}$	$Y=-2.24x-0.25$	$Y=-0.04x+2.67$
R^2_{adj}	0.69	0.68	0.67	0.56	0.56
$RMSE_{CV}$	0.40	0.31	0.40	0.20	0.19
R^2_{CV}	0.74	0.59	0.37	0.52	0.55

In contrast to results obtained by Viscarra Rossell et al. (2006a) and Spielvogel et al. (2004), the relationship between soil lightness and soil organic carbon content could be described by a linear model for unit 2, which is mainly due to the small range of OC values (0.16 – 1.5 %). For specific homogenous subunits regression equations could be improved: for calcarenites (*formazione di Gallipoli*, N=39) using soil lightness with $R^2=0.70$. A model, applicable to a wider range of soils, is preferred in this study.

The river terrace and floodplains, characterised by alluvial deposits, show a wider range of OC content and can be modelled by curve shape attributes, as the spectral shape is highly varying within this group. Again, a rather simple model is preferred to more accurate models for a sub regions [pure Holocene deposits using polynomial coefficient with $R^2=0.75$ (N=33)].

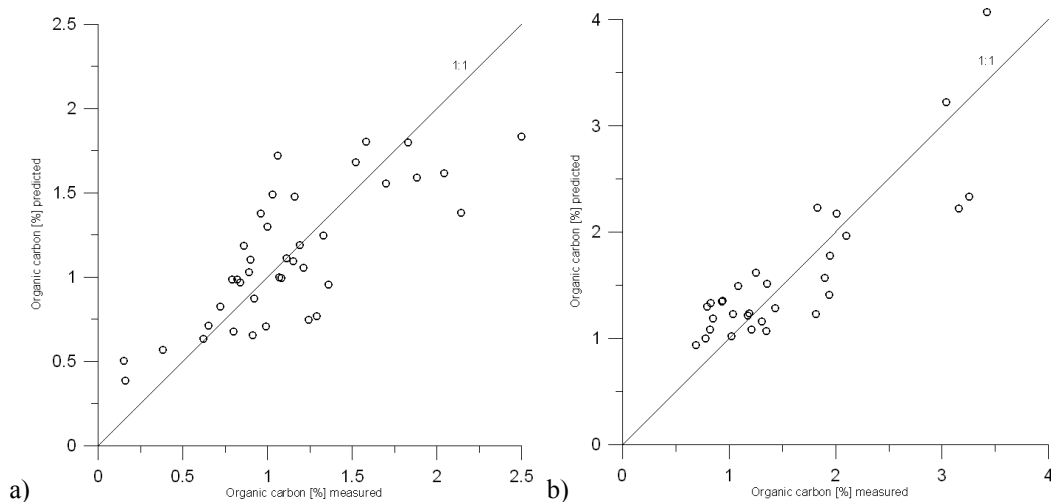


Figure 18. Cross-validation results for the estimation of OC content based on the coefficient b_2 of a polynomial fitted on the reflectance spectra for (a) areas characterised by alluvial deposits and (b) for soil developed on limestone.

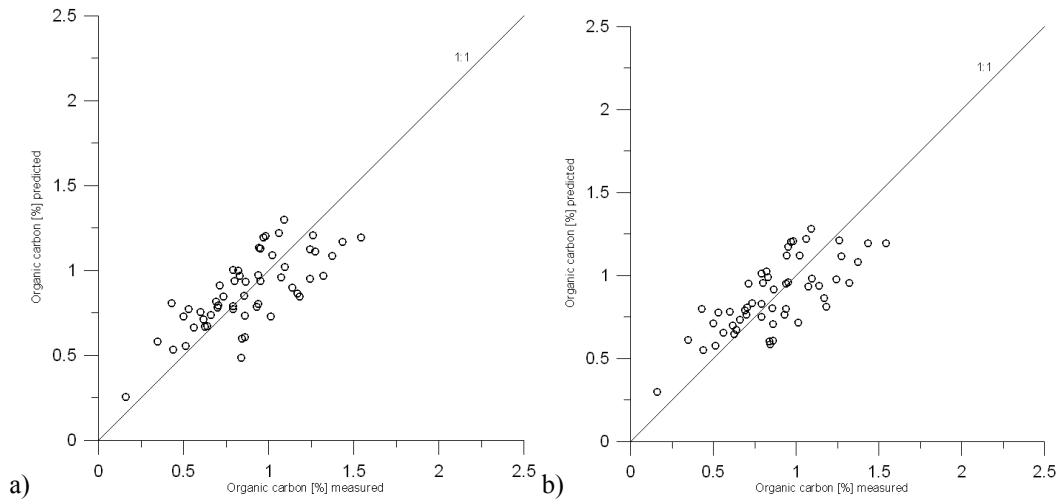


Figure 19. Cross-validation result for the estimation of OC content based on a) soil lightness L (CIELAB) and b) R (R-G-B) for areas characterised by calcareous sandstones and unconsolidated Pleistocene deposits.

Best cross-validation results were obtained using the coefficient b_2 of a second order polynomial ($R^2=0.59$ for unit 1 and $R^2=0.74$ unit 3) and soil lightness L ($R^2=0.54$ unit 2) as dependent variable in the regression equation. Regression results are with R^2 values between 0.5 – 0.7 in the same range as results by Viscarra Rossel et al. (2006a). Cross-validation results are shown in Figure 18 and Figure 19. Although model accuracy is rather low, regression equations are assumed to reflect still the general behaviour within a specific morpho-lithological unit and are therefore appropriate at regional scale.

2.3.2 Sensitivity of models to vegetation influence

The application of models, developed on laboratory reflectance spectra, to satellite imagery needs to address, among others, the problem of spectral mixture of different surface types in one pixel. Therefore the sensitivity of models to green and dry vegetation influences was assessed by synthetic linear mixtures of green and dry vegetation spectra and soil spectra for the specific morphological-lithological units. Green and dry vegetation spectra were both taken from the spectral library of the United States Geological Survey (USGS) (Clark et al., 1993). The percentage of green and dry vegetation in the spectral mixture was successively increased from 5 % to 30 %. OC estimates from pure spectra were compared to estimates from mixtures of soil and green vegetation and soil and dry vegetation, respectively.

Figure 20 and Figure 21 illustrate the results of the regression analysis between OC contents calculated from pure soil spectra with the ones calculated from mixtures of soil and green vegetation and soil and dry vegetation.

As with increasing vegetation amount the spectral shape changes strongly, the parameterisation of the spectral shape through a second order polynomial is highly sensitive to green vegetation influences, leading to an overestimation of OC content with increasing vegetation amount (see Figure 20). The influence of green vegetation is stronger for samples with higher content of soil organic carbon (maximal error 1.6 % for samples with OC 2.5 % and minimal 0.03 % for OC of 0.15 %).

Dry vegetation, present in the pixel, leads to an underestimation of OC content when applying the polynomial model (Figure 21). Assuming a fraction of 5% dry vegetation within a pixel, a minimum deviation of -0.1 % for OC content of 0.15 % and maximum deviation of -0.5 % at OC content of 2.5 % were obtained.

Models based on the parameter $R'_{890nm} - R'_{620nm}$ and on colour attributes are less affected by green and dry vegetation in comparison to the model based on polynomial coefficients (Figure 20 and Figure 21). Estimations of OC show an offset to higher values with increasing vegetation for the parameter $R'_{890nm} - R'_{620nm}$, leading to a small error between +0.23 % (for 0.15 % OC) and + 0.18% (for OC content of 2.5 %).

The influence of dry vegetation for the model based on parameter $R'_{890nm} - R'_{620nm}$ is tolerable up to 20% with a maximal error of 0.35 % for samples with 2.5 % OC content.

Soil lightness L (CIELAB) is less sensitive to vegetation influence for both green and dry vegetation mixtures than the R component of the R-G-B colour space model (Figure 20 and Figure 21). The maximal error for a mixture of 30 % green vegetation and soil is -0.16 %, whereas for R a maximal error value of +0.48 % is obtained. Green vegetation fraction up to 30 % can be tolerated for the model based on L.

The obtained maximal error is higher regarding dry vegetation influence; 30 % dry vegetation in the mixture leads to an underestimation of OC content of -0.67 % and -0.80 % for a model based on L and R, respectively. Again, the L model is more stable with a tolerable error of 0.29% for a dry vegetation contribution to the mixture of 10 %.

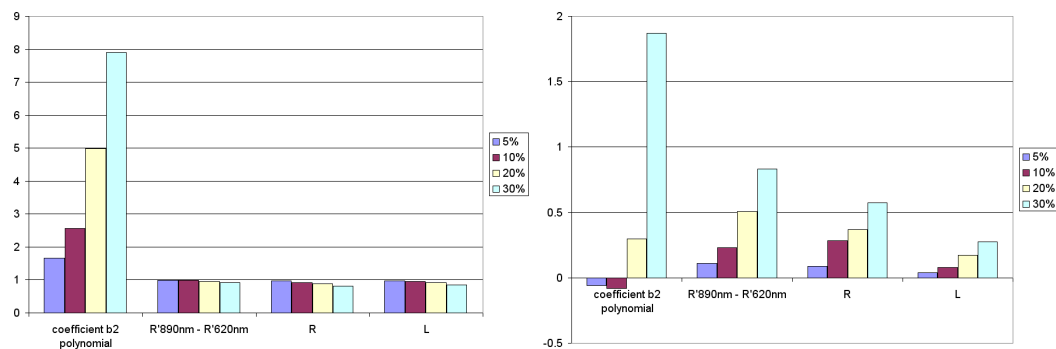


Figure 20. Regression analysis results between estimations of organic carbon from pure soil spectra and spectral mixtures of soil and green vegetation (5, 10, 20 and 30 %) showing the coefficient (left) and constant (right) of the regression equation.

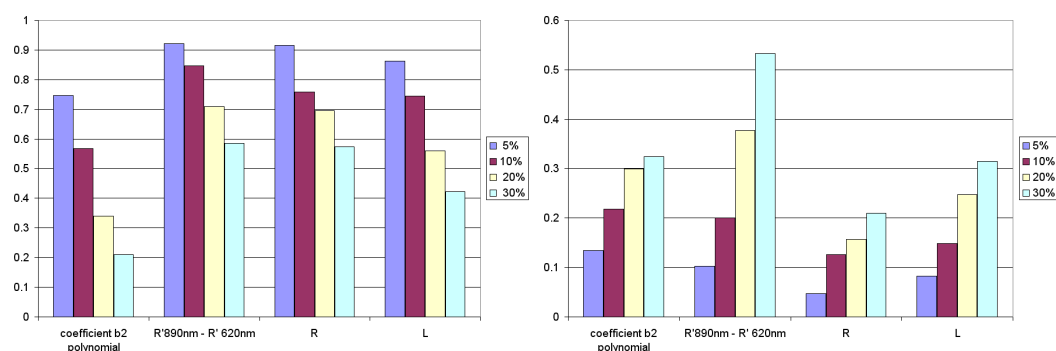


Figure 21. Regression analysis results between estimations of organic carbon from pure soil spectra and spectral mixtures of soil and dry vegetation (5, 10, 20, and 30 %) showing the coefficient (left) and constant (right) of the regression equation.

2.4 MERIS satellite data and pre-processing

MERIS is a fine spectral and medium spatial resolution satellite sensor, which is part of the core instrument of Envisat, the European Space Agency's (ESA) environmental research satellite, launched in 2002.

MERIS is a 'push broom' instrument with five optical modules, each containing a two-dimensional charged-couple-device (CCD) array, scanning the Earth's surface in the across-track direction. Optical modules have a 14° field of view (FOV), providing an overlap of 10 pixels between modules and resulting in a FOV at nadir of 68.5°, covering a swath width of 1150 km from the platform altitude of 799 km (ESA online at <http://envisat.esa.int/instruments/meris>). Data is available in dual spatial resolution: reduced spatial resolution (RR) with 1200 m at nadir and full spatial resolution (FR) with 300 m at nadir.

The instrument was primarily designed for ocean and coastal remote sensing and its radiometric performance is therefore beyond that of a sensor required for land applications. Radiometric error is less than 2 % of the detected signal between 400 to 900 nm and less than 5 % of the detected signal between 900 to 1050 nm (Curran & Steele, 2005). MERIS has a 12 bit signal digitization (4096 intensity levels). Visible and near infrared radiation is recorded in 15 narrow bands (within the 390 – 1040 spectral region). Default band centres and band width are given in Table 14.

Table 14. MERIS full resolution (FR) level 2 spectral band settings. Wavelength (WL) and Full Width Half Maximum (FWHM) in nm.

MERIS		
Band	WL	FWHM
1	412.5	9.93
2	442.5	9.95
3	490	9.97
4	510	9.97
5	560	9.98
6	620	9.99
7	665	9.99
8	681.25	7.49
9	705	10.00
10	753.75	7.49
11	775	15.00
12	865	20.004
13	890	9.99

For land applications the MERIS level 2 data products, which contain measurements of geophysical variables, are provided by ESA. They include calibrated reflectance in all channels, the Normalised Difference Vegetation Index (NDVI) and the MERIS Global vegetation index (MGVI) (Gobron et al., 2004). The MERIS level 2 product consists of top of aerosol reflectance for 13 MERIS bands, which means reflectance corrected for gaseous absorption and Rayleigh scattering (Santer et al., 1999). The correction of surface reflectance for aerosol absorption and scattering is not included yet. Reflectance is only derived for the 13 bands which are not strongly contaminated by atmospheric gaseous absorption. The channels at 760 nm and at 900 nm are devoted to the retrieval of gaseous abundances (O_2 and water vapour respectively). Information on aerosol is given in the MERIS level 2 product in terms of aerosol optical thickness and angstrom coefficient over dark dense vegetation (DDV) (Université du Littoral Côte d'Opale, 2000). For the retrieval of aerosol optical thickness, angstrom coefficient over land surfaces and a subsequent aerosol correction, the Bremen Aerosol Retrieval (BAER) algorithm (von Hoyningen-Huene et al., 2006) was developed and implemented in the BEAM VISAT software (Brockmann Consult). However, the retrieval of aerosol optical thickness using the BAER algorithm has limitations for bright surfaces, which leads to voids in the data layer, in particular for bare soils. The improvement of the aerosol retrieval procedure for bright surfaces was outside the scope of this study. [A void filling through interpolation of retrieved aerosol optical thickness by incorporating a Digital Elevation Model can be proposed.]

MERIS FR level 2 data for land applications were acquired for different acquisition dates, covering the years 2003 and 2004. (in total 12 images, listed in Table 15). For the analysis MERIS level 2 data without any further atmospheric correction was used.

The presence of vegetation in image pixels hinders the derivation of soil properties of the soil surface. Agricultural soils are bare at different times of the year according to management. Therefore a multi-temporal assessment is necessary to obtain a maximum share of bare soil for the test site. The following pre-processing steps to minimize the effect of vegetation on image spectra were applied:

- Superposition of images from different acquisition dates
- Masking of clouds
- Derivation of a minimum vegetation composite
- Masking of vegetation
- Determination of residual vegetation at sub-pixel level through linear spectral unmixing

- Removal of the contribution of the vegetation fraction to surface reflectance
- Masking of dry vegetation
- Vector-normalisation of image spectra
- Fitting of second-order polynomials to image spectra
- Calculation of colour attributes

The pre-processing steps will be described in detail.

Geolocation accuracy of provided MERIS level 2 products was not sufficient for multi-temporal analysis. In order to improve geolocation accuracy between imagery, images were superposed on Landsat Image 2000 data (EUROPEAN COMMISSION, 2005). The original resolution of Landsat Image 2000 data of 25 m was reduced by averaging 6 x 6 pixels. The semi-automatic procedure FINDGCP (Mehl, 2004, Unpublished Work.) was applied to extract ground control points (GCP) between imagery. The geometric transformation was determined through linear interpolation within each triangle of a Delaunay triangulation at any point and described on a regular grid. The image was resampled through the grid file using cubic convolution as interpolation method. The root mean square error (RMSE) varied between 0.53 – 0.68 for the 12 corrected images. Nearest neighbour resampling from raw data in the distributed MERIS imagery limits the geolocation accuracy and could account for an error up to 0.5 pixels. Additional geolocation errors might be due to the five camera optics subassembly (COSA), which causes displacements in the overlapping area of two adjacent COSAs.

Clouds were masked using an improved MERIS level 2 cloud flag by applying the algorithm of von Hoyningen-Huene et al. (2006). Remaining clouds and cloud shadows, especially in mountain areas, can be neglected in this study.

Minimum compositing technique was applied using the lowest value of the fraction of absorbed photosynthetically active radiation (FAPAR equal to MGVI) throughout a year for each pixel as selection criterion. The MGVI product is recognised to have a high sensitivity to presence and changes of healthy green vegetation, while minimizing the sensitivity to atmospheric scattering and absorption effects, to soil colour and brightness effects and to temporal and spatial variations in the illumination and observation geometry (Gobron et al., 2000; Gobron et al., 2006). It has an uncertainty range of +/- 0.1. However, a validation exercise showed good agreement with ground based measurement also for sparse vegetation cover (Gobron et al., 2006).

Figure 22 demonstrates the composition of the minimum vegetation composite of the year 2003, showing the acquisition date for each pixel. The proportion of single images in the minimum composite of year 2003 and 2004 are given in Table 15. Highest proportions for both years stem from images in September, followed by July acquisitions.

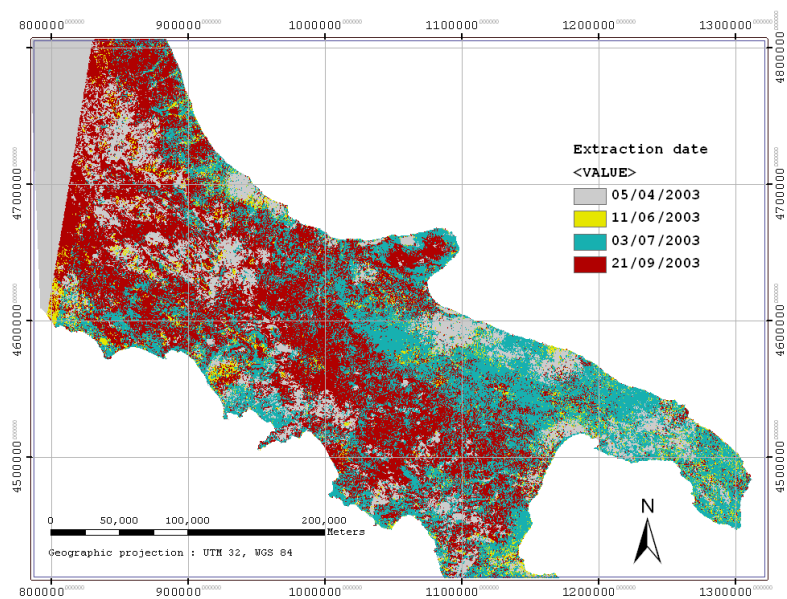


Figure 22. Acquisition dates for the MERIS minimum vegetation composite of the year 2003.

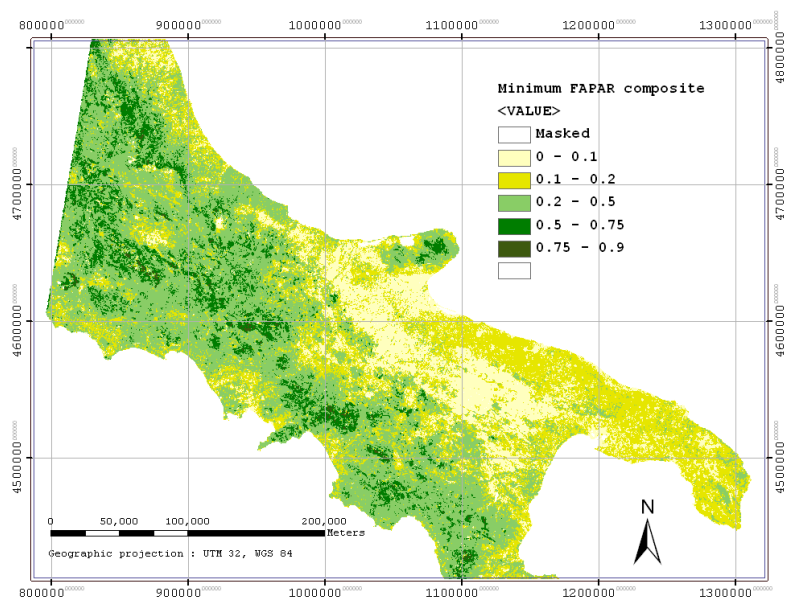


Figure 23. FAPAR value of MERIS minimum vegetation composite 2003.

Table 15. MERIS images used for the derivation of minimum vegetation composite and their proportion on minimum vegetation composite of the years 2003 and 2004.

Acquisition dates 2003	Proportion minimum composite [%]	Acquisition dates 2004	Proportion minimum composite [%]
05/04/2003	21.89	26/05/2004	6.72
11/06/2003	4.43	11/06/2004	1.93
03/07/2003	30.98	24/06/2004	0.77
21/09/2003	42.70	06/07/2004	11.21
		19/07/2004	9.45
		17/08/2004	3.42
		23/08/2004	11.33
		11/09/2004	55.17

The spatial distribution of FAPAR values of the minimum vegetation composite is shown in Figure 23. Arable lands in the Apulian plain demonstrate low values of FAPAR between 0 – 0.1.

To minimize further the influence of photosynthetic-active vegetation on image spectra, pixels with FAPAR value greater than 0.1 were masked. Furthermore water bodies and urban areas were excluded from further analysis using corresponding Corine Land Cover 2000 classes [CLC2000,EEA (2006)].

The proportions of CLC2000 classes for pixels retrieved using the FAPAR threshold (0.1) are given in Figure 24. About 70 % of the pixels correspond to arable land (irrigated and non-irrigated). Other agricultural classes (mainly complex cultivation patterns, olive groves, vineyards, fruit trees and pasture) account for 24 % and 22 % in year 2003 and 2004, respectively. Only about 7 % of the pixels in each year derive from non-agricultural land cover classes (e. g. natural grassland, sparsely vegetated areas).

Pixels not coinciding with CLC 2000 arable land were masked.

The minimum value compositing and subsequent masking of pixels with the FAPAR threshold allowed the monitoring of approximately 50 % of arable land as classified by CLC2000 for each year in Southern Italy.

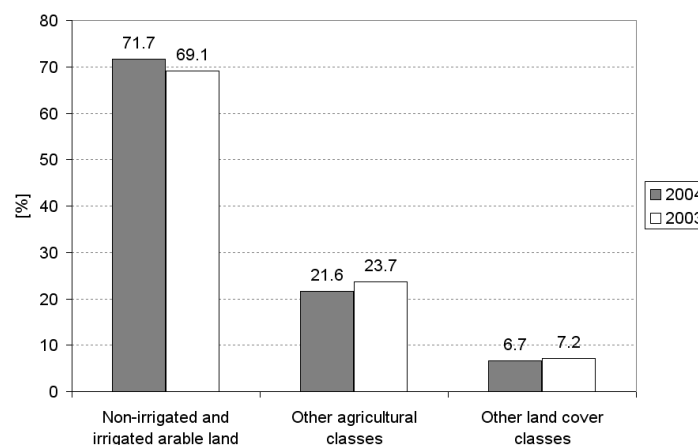


Figure 24. Proportions of land cover classes for pixels with FAPAR values lower or equal to 0.1 in the MERIS minimum vegetation composites of the years 2003 and 2004.

The residual amount of photosynthetic-active vegetation at subpixel level was determined through linear spectral mixture analysis using green vegetation, soil and shade as endmember (EM). The vegetation endmember was extracted from a single image (July 2003). The soil EM was deduced from the minimum vegetation composite by principle component analysis. The extracted soil endmember

spectrum was spectrally matched with a spectral library of soils from Apulia using spectral angle, standardised reflectance difference and reflectance difference as criteria (Garcia-Haro et al., 2005). The vegetation fraction determined by spectral unmixing ranges between 0 –15 % (mean value 2 %) for 2003 and between 0 and 20% (mean 9 %) for the year 2004. Spectral reflectance was corrected to “bare” soil reflectance by discarding the part of the reflectance related to residual vegetation. Accordingly, spectral reflectance of each pixel is determined using the normalised abundance of the soil endmember. Residuals between modelled and original reflectance in each channel are added to the vegetation-corrected reflectance, as they result partly from varying soil properties (Jarmer, 2005).

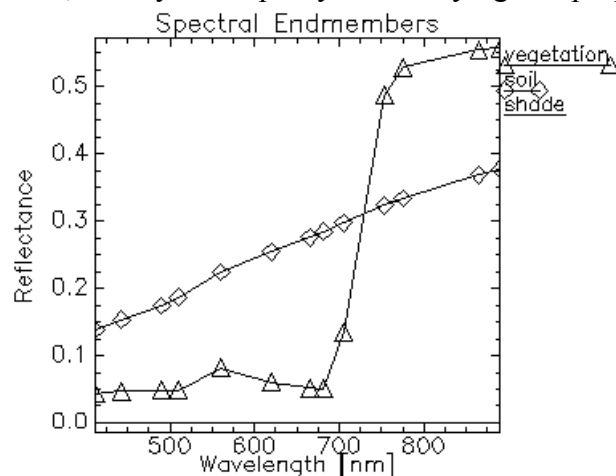


Figure 25. Spectral endmembers green vegetation, soil and shade.

In addition to photosynthetic-active vegetation, the influence of dry vegetation has to be considered in the pre-processing. Senescent vegetation and bare soil reflectance show distinct and separable features in the shortwave-infrared spectral region between 2100 and 2400 nm (Asner & Lobell, 2000), but are featureless in the spectral domain covered by the MERIS sensor. Information on ploughing date and management after harvesting of test areas was not available.

In order to assess the amount of dry vegetation, Spectral Angle Mapping (Kruse et al., 1993) has been applied. Spectral Angle Mapping (SAM) considers reflectance spectra of individual pixel as n -dimensional vector, where n is the number of image channels. The length of the vector represents brightness of the target whereas the direction represents spectral feature of the target. Variation in illumination mainly affects changes in the length of the vector. An endmember spectrum of straw was taken from field measurements by Sommer (personal communication). Pixels classified as straw (criteria: angle lower than 0.1 radians between pixel and endmember straw) have been excluded from further analysis. An analysis based on the laboratory spectra of field samples from Apulia (only top horizon) revealed that with this criterion also 15 % of the soil spectra would be classified as straw. However, pixels classified as straw in the minimum composites of both years stem mainly from July images (58 % in 2003 and 83 % in 2004). This finding seems to support the chosen approach as harvesting of crops occurs primarily in July.

In the next step, second order polynomials were fitted to vector-normalised image spectra and soil colour attributes (as described before) were calculated. Except from few outliers, obtained data ranges of derived parameters were similar to ranges of field samples.

3 Model application to MERIS minimum vegetation composites

Regression equations, as listed in Table 13, were applied to MERIS minimum vegetation composites of the years 2003 and 2004 to estimate OC content. Results have been analysed regarding data ranges, value distribution and the spatial coherency between two observations (2003 and 2004) in Apulia and in the Abruzzo region.

3.1 Organic carbon estimates for Apulia and Abruzzo test areas

Unit 1-River terraces and floodplains characterised by alluvial deposits

Figure 26 depicts the distribution of OC in unit 1, estimated from MERIS minimum composites. The histogram is positively skewed with few high values of OC. Mean values of OC content are with 0.97 % (2003) and 0.87 % (2004) in agreement with the average value of the calibration set (1 %). For 1 % of the data an OC content higher than the maximum value of laboratory measurements (2.5 %) was obtained. The minimum value (0.28 % OC) is in line with measurements (0.15 % OC). Results from two subsequent years revealed spatial instabilities which might be due to an overcorrection of vegetation abundance in image spectra. Spatial stability of OC estimations increased when considering only pixels with a vegetation fraction lower than 5 %. A graphical representation of the linear relationship between satellite observations from 2003 and 2004 is given in Figure 27. The coefficient of determination (R^2) for the linear regression is 0.85 (see Figure 27). Calculated organic carbon content in 2004 is lower than in 2003.

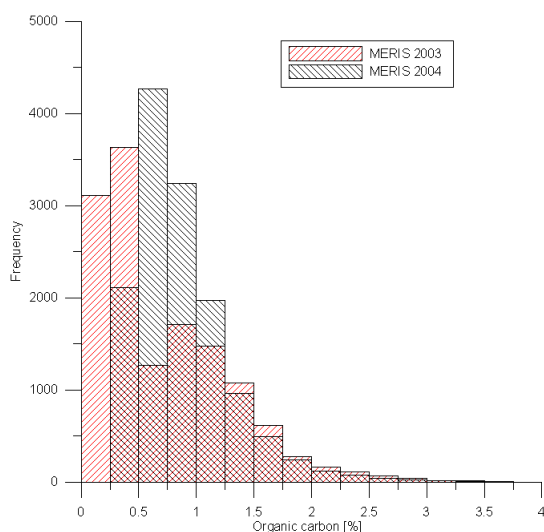


Figure 26. Histogram of OC [%] content from MERIS minimum composite for unit 1. Calculation of OC was based on the coefficient b_2 of a polynomial.

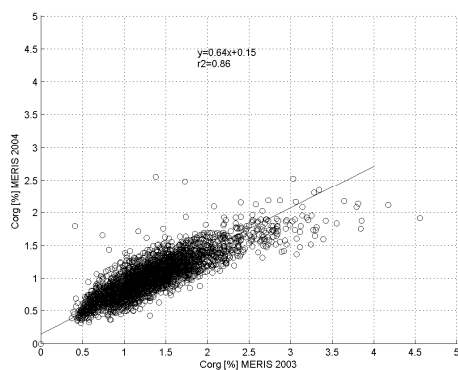


Figure 27. Correlation between OC estimations from MERIS data for the years 2003 and 2004 for unit 1. Estimations are limited to areas with a vegetation fraction lower than 5% in both years.

Unit 2 - Morpho-structural depressions characterised by unconsolidated deposits

MERIS OC estimates for unit 2, based on R, are in the same range as field samples of the calibration set (0.19 – 1.31 % OC; mean 0.87). The mean OC value obtained when using the model based on L is

slightly higher (mean OC 0.98 % in 2003). Although ranges and mean values agreed between two subsequent years (see also Figure 32 b), histograms revealed differences in the value distribution in 2003 and 2004 (Figure 28). This might be due to an overcorrection of the vegetation influence in 2004.

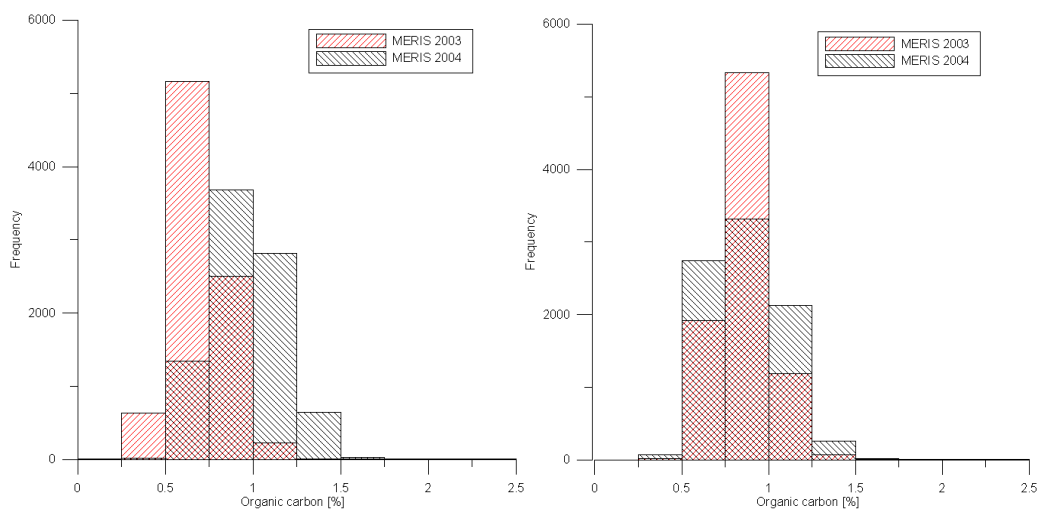


Figure 28. Histograms of OC [%] content from MERIS minimum composite for unit 2; using L (left) and R (right) as input parameters.

Unit 3 – Structural highs characterised by limestone

The distribution of OC content from MERIS data in unit 3 is in good agreement for both years. Values are slightly shifted to lower values in 2004 (Figure 29). Mean values are consistent between two observations with 1.87 % in 2003 and 1.79 % in 2004. The average OC content of field samples (calibration set) is with 1.59 % slightly lower.

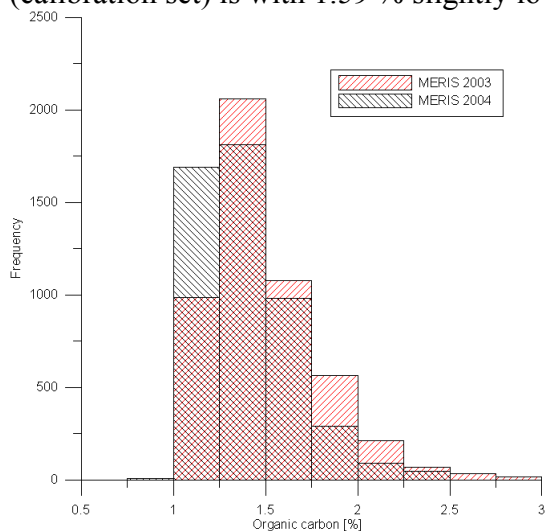


Figure 29. Histogram of OC content [%] from MERIS minimum composite for unit 3 using coefficient b_2 of a fitted polynomial as input parameter.

Abruzzo test area

The application of MERIS satellite models to the Abruzzo test area was limited to small lowland areas, as only few pixels remained for the analysis after vegetation masking. Soil organic carbon was estimated for unit 2 (eco-pedological unit 9.01) using the model based on R (Table 13). MERIS OC estimates from 2004 showed a shift to higher values of OC and a higher standard deviation in comparison to the year 2003 (see Figure 33 and Figure 34).

3.2 Comparison with independent data sets

In order to verify our OC estimates from MERIS satellite data, results were compared with the map of OC content in the topsoil layer (OCTOP), which was calculated based on pedo-transfer rules (Jones et al., 2005).

The OCTOP dataset is a European wide assessment based on the European Soil Database (ESDB) with a grid size of 1 km x 1 km (Jones et al., 2005). OCTOP is given in the following classes:

- Very low: < 1 %
- Low: 1.1 – 2 %
- Medium: 2.1 – 6. %
- High: > 6 %.

Pedo-transfer rules (PTRs) for the determination of topsoil OC were constructed by Van Ranst et al. (1995) and modified by Jones et al. (2005). After assigning the OCTOP values on the basis of soil characteristics, these are modified on the basis of land use; for example, under grassland a soil is given a higher OCTOP content than the same soil type under cultivation. The OCTOP content obtained from applying the revised PTR was further modified to take account of changes in organic carbon content caused by variations in temperature. Main inputs, such as soil type and soil texture, can not be described more precisely, concerning their spatial location, than the initial Soil Mapping Unit in the ESDB. Modifications derive from land use, based on the CORINE LAND COVER at 1km grid resolution and the temperature gradient. A comparison of spatial distribution on a pixel by pixel basis is therefore not feasible.

Besides the OCTOP map, the soil profile data base of the eco-pedological map of Italy served as data set for comparison of SOC content of arable lands within specific eco-pedological units.

3.2.1 Comparison for the Apulia test area

The spatial distribution of OC content of arable land derived from MERIS 2003 and from the OCTOP map is demonstrated in Figure 30. Very low values of OC (0 -1 %) are dominating in both maps (50.8 % MERIS 2003, 53.3 % OCTOP), followed by low content (1 -2 %) of OC (45.4 % MERIS 2003 and 39.6 % OCTOP) and medium content (2 – 6 %) of OC with a proportion of only 3.7 % and 7.1 % for MERIS estimates and OCTOP, respectively. MERIS OC estimates demonstrate a higher heterogeneity of organic carbon content in areas characterized by alluvial deposits, showing few areas with medium organic carbon content.

In contrast to MERIS estimates, the OCTOP map shows medium OC values in the Higher Murge area, which could not be verified further.

In general, data comparison revealed a continuous histogram distribution for the satellite derived OC content and a n-modal distribution for the OCTOP dataset (Figure 31). The abrupt changes in the OCTOP dataset coincide with a change between two adjacent soil mapping units, which represent different soil attributes concerning soil type and soil texture and therefore different class of OC content. For satellite models (based on soil colour and coefficient of polynomial) good agreement could be observed with the OCTOP map concerning data ranges (Figure 32). Maximum values for all units revealed few outliers in MERIS estimates in contrast to field data (from Ecopedological map) and OCTOP map.

Unit 1- River terraces and floodplains characterised by alluvial deposits

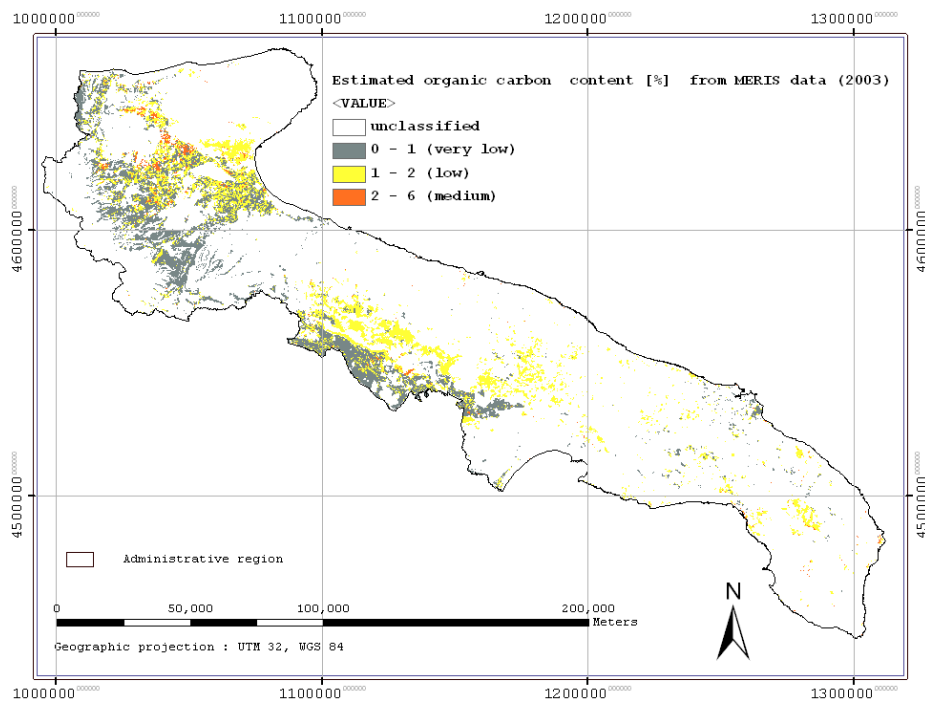
In accordance with MERIS OC estimates for unit 1, OC content modelled by pedo-transfer rules demonstrated a lower mean value (0.95 %) as field data (Figure 32 a). OC estimates from MERIS, limited to areas with a vegetation cover up to 5 %, suggested a higher mean value (MERIS 2003: 1.34 %; MERIS 2004: 1.04 %).

Unit 2- Morpho-structural depressions characterised by unconsolidated deposits

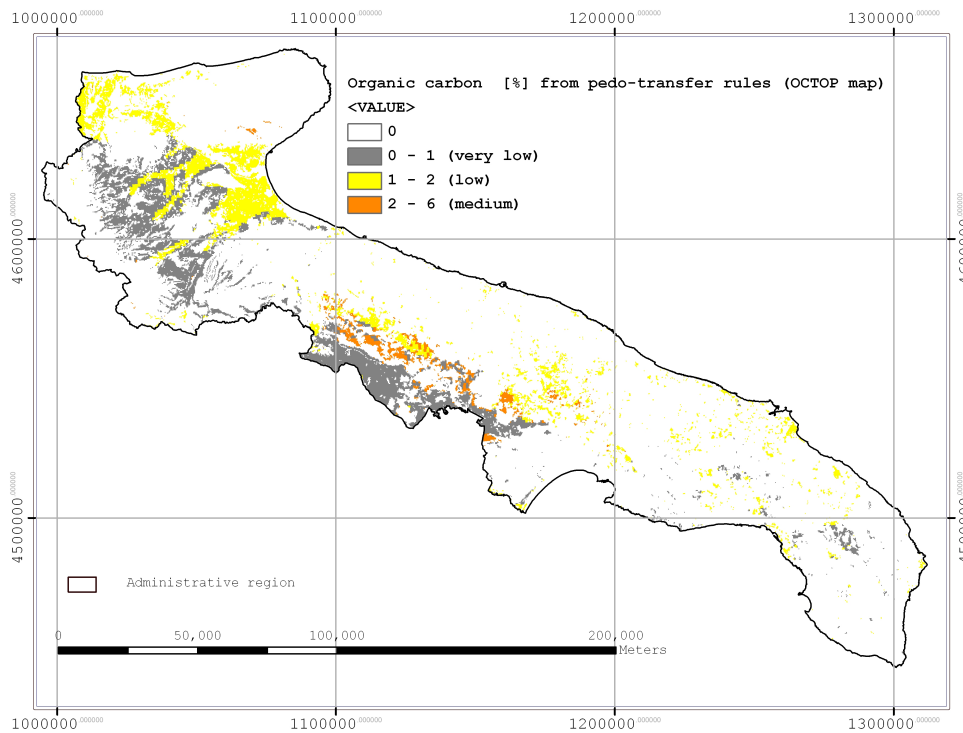
Mean values of MERIS OC content for 2003 and 2004 were in line with the OCTOP map for this unit. Field data from the eco-pedological map indicated a lower mean value and a higher standard deviation (Figure 32 b).

Unit 3- Structural highs characterised by limestone

Mean values derived for unit 3 from MERIS data were slightly higher than OCTOP estimates (Figure 32 c).



a)



b)

Figure 30. Estimations of OC content [%] for arable land in Apulia a) from MERIS data (2003) and b) modelled with pedo-transfer rules [OCTOP map, (JONES ET AL., 2005)]. Areas covered by vegetation in satellite data are masked in both maps.

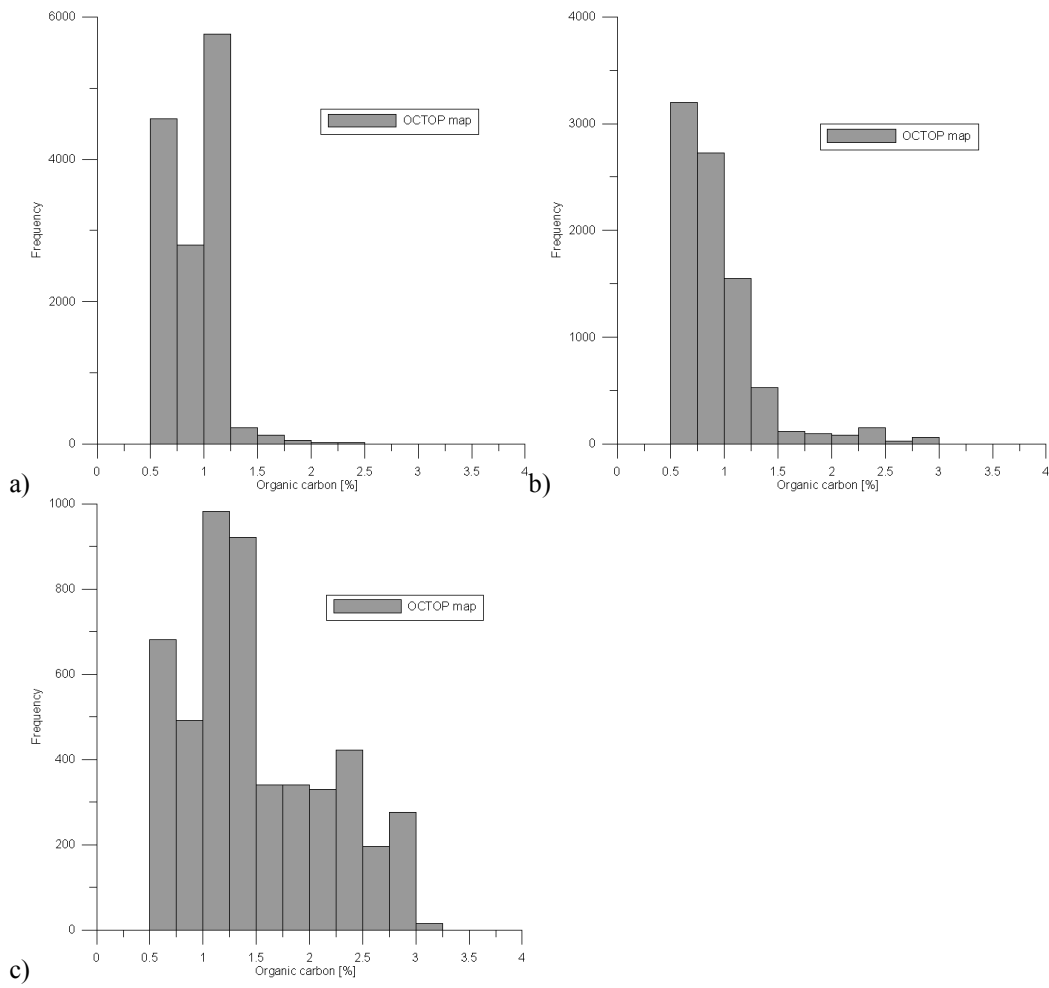


Figure 31. Distribution of OC content [%] derived from pedo-transfer rules [OCTOP map, (Jones et al., 2004)] for stratification unit 1 (a), unit 2 (b), unit 3 (c).

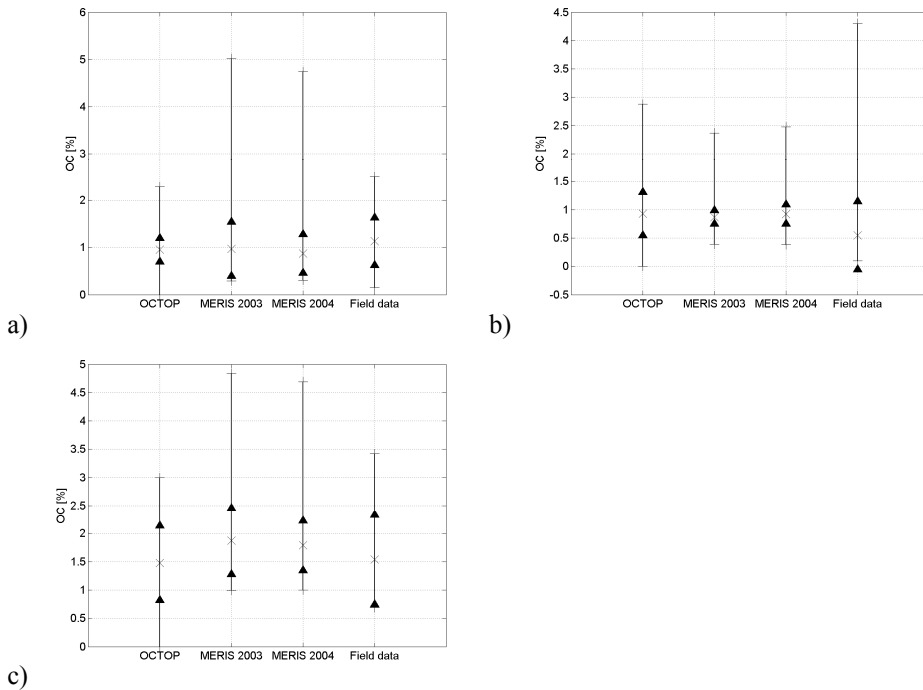


Figure 32. Basic statistics (mean x, standard deviation Δ , min and max -) of OC content [%] derived from pedo-transfer rules, MERIS satellite data and ground samples from the eco-pedological map for morpho-lithological units 1 (a), 2 (b), 3 (c) in Apulia.

3.2.2 Comparison for the Abruzzo test area

Results for unit 1 (alluvial floodplains and terraces) could not be validated as only few pixels remained after vegetation masking (~ 400 pixels). Furthermore these pixels were located at the edge of the unit and might not be representative. Therefore only results for unit 2 (unconsolidated deposits) were compared with the OCTOP map. Very low values of OC are dominating for all datasets. OC estimations from MERIS data agreed with modelling results from pedo-transfer rules (Figure 33 b) in mean values. In contrast to estimations from MERIS and OCTOP modelling results, the average OC content of field data (eco-pedological map) showed a lower mean value (0.55 % OC, $N = 46$, Figure 33). It has to be noticed that after vegetation masking only 6 % of the stratum (UE 9.01 1,203/21.138 pixels) remained for the derivation of OC from satellite data.

Despite stability in descriptive statistics for OC estimates from MERIS data for two observations (years 2003 and 2004) (Figure 33 b), local variations can still be observed as shown in Figure 34 for a subset of unit 2 in Abruzzo.

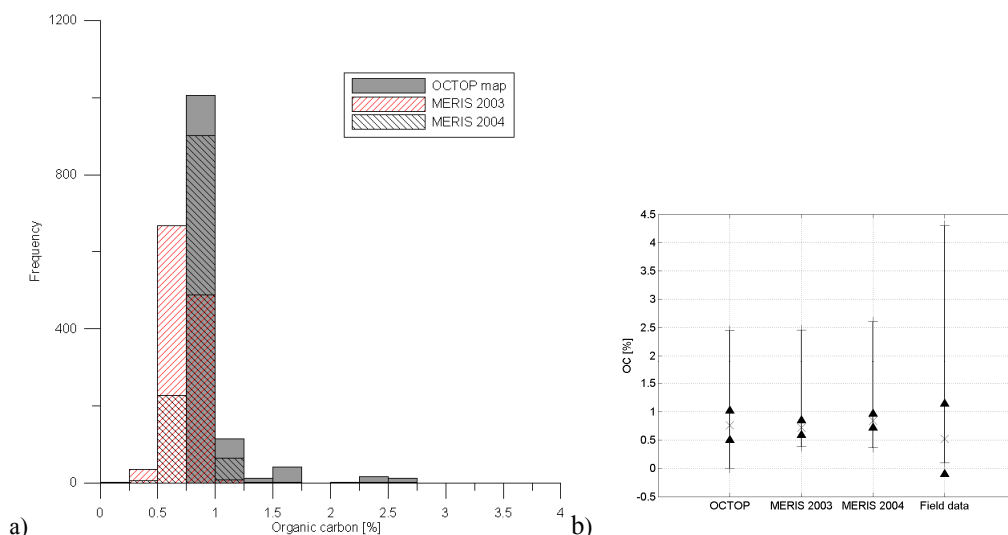


Figure 33. Histogram (a) and basic statistics (b) (mean \bar{x} , standard deviation Δ , min and max -) of OC content [%] derived from pedo-transfer rules, MERIS satellite data and ground samples (eco-pedological map, N = 46) units 9.01 in Abruzzo.

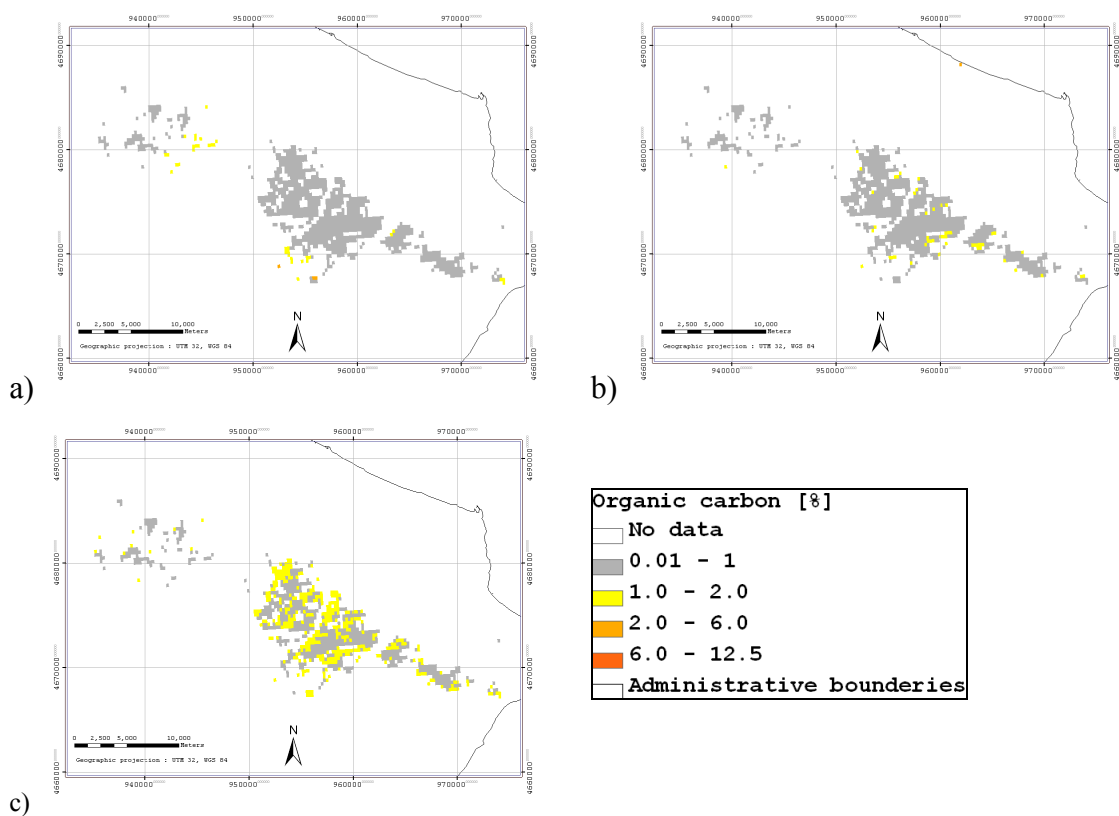


Figure 34. Distribution of organic carbon [%] within a subset of unit 2 (eco-pedological unit 9.01) in Abruzzo (a) from OCTOP map, (b) estimated from MERIS minimum composite of year 2003 and (c) estimated from MERIS minimum composite of the year 2004. Areas covered with vegetation in satellite data are masked for all maps.

4 *Conclusions*

The estimation of topsoil organic carbon content from MERIS reflectance data could be demonstrated for arable land at regional scale in Southern Italy. Simple models using soil colour attributes and spectral shape could be derived for specific morphological-lithological units in the Apulia region. Stratification into morphological-lithological units takes best into account the dependence of soil reflectance from soil parent material and soil development, which can not be well depicted by MERIS spectra due to the limited spectral range of the sensor. Furthermore, the INTEREG project sampling strategy being not optimised for our purpose might have been a limiting factor for the quality of models. Sampling was adapted to a delineation of soil types based on already known general spatial distribution of soil types focussing thus on a high variety of soils (European Soil Bureau & Regione Puglia, 2001).

The above-mentioned stratification approach instead allowed a transfer of models from the calibration site to another region in Southern Italy.

Vegetation related influences in image spectra could be minimised by a multi-temporal minimum vegetation criterion and a further decomposition of image spectra into its components soil and vegetation. The quantification of the contribution of senescent vegetation to image spectra remains to be improved.

Regarding models based on colour attributes a high stability of OC estimations could be observed for two years, since these models show a low sensitivity to vegetation. Stability of estimations with the model based on the coefficient of the polynomial fitted through the image spectra remains restricted to almost vegetation free pixels. Analysis of OC estimations indicates inaccuracies in the determination of vegetation and soil abundance in the pixel, which led to errors in the modelled reflectance for areas with overestimated vegetation abundance.

Hence, monitoring capability is given for models based on colour attribute but remains restricted to almost vegetation free areas for models based on polynomial coefficients.

Organic carbon estimations for the region Apulia and Abruzzo showed good agreement with independent modelling results (OCTOP map) in data ranges and mean values.

IV References

- AG Boden, 1996. Bodenkundliche Kartieranleitung. 4. Aufl., E. Schweizerbart'sche Verlagsbuchhandlung, Bundesanstalt für Geowissenschaften und Rohstoffe und die Geologischen Landesämter in der Bundesrepublik Deutschland, Hannover.
- Al-Abbas, A. H. and Baumgardner, M. F., 1972. Relating organic matter and clay content to the multispectral radiance of soils. *Soil Science*, 477-485.
- Andersson, C. A. and Bro, R., 2000. The N-Way Toolbox for Matlab. *Chemometrics and Intelligent Laboratory Systems*, 52, 1-4.
- Asner, G. P. and Lobell, D. B., 2000. A biogeophysical approach for automated SWIR unmixing of soils and vegetation. *Remote Sensing of Environment*, 74, 1, 99-112.
- Association of the Geological Surveys of the European Union (EuroGeoSurveys)/ the Geological Survey of Finland, 2005. *Geochemical Atlas of Europe*.
<http://www.gtk.fi/publ/foregsatlas/index.php> [9-21-2008].
- Baumgardner, M. F., Silva, L. F., Biehl, L. L., and Stoner, E. R., 1985. Reflectance Properties of Soils. *Advances in Agronomy*, 38, 1-44.
- Ben-Dor, E. and Banin A., 1990. Near-infrared reflectance analysis of carbonate concentration in soils. *Applied Spectroscopy*, 44, 6, 1064-1069.
- Ben-Dor, E., Banin A., and Singer A., 1991. Simultaneous determination of six soil properties from the soil diffuse reflectance spectrum in the near infrared region (1-2,5 μm), 5th International Colloquium - Physical Measurements and Signatures in Remote Sensing, 1-14-1991, Courchevel, France, pp. 159-163 .
- Ben-Dor, E. and Banin, A., 1994. Visible and near-infrared (0.4-1.1 μm) analysis of arid and semiarid soils. *Remote Sensing of Environment*, 48, 3, 261-274.
- Ben-Dor, E., Taylor, R. G., Hill, J., Demattê, J. A. M., and Whiting, M. L. Chabrilat S. Sommer S., 2008. Imaging spectrometry for soil applications. *Advances in Agronomy*, 97, 321-391.
- Blume, H. P. and Helsper, M., 1987. Schätzung des Humusgehaltes nach der Munsell- Farbhelligkeit. *Zeitschrift für Pflanzenernährung und Bodenkunde*, 150, 354-356.
- Brady, N. C. and Weil, R. R., 1999. *The nature and properties of soil*. Prentice Hall, Upper Saddle River.
- Brandt, J., Geeson, N., and Imeson, A., 2003. A desertification indicator system for Mediterranean Europe. [Online] European Commission, DG-Research:Contract No. EVK2-CT-2001-00109.
<http://www.kcl.ac.uk/projects/desertlinks/downloads.htm> [10-15-2008].
- Brown, S. D., Sum, S. T., Despagne, F., and Lavine, B. K., 1996. Chemometrics. *Analytical Chemistry*, 68, 21R-61R.

- Chang, C.-W., Laird, D. A., Mausbach, M. J., and Hurburgh Jr., C. R., 2001. Near-infrared reflectance spectroscopy - principle components regression analyses of soil properties. *Soil Science Society of America Journals*, 65, 480-490.
- Clark, R. N., Swayze, G. A., Gallagher, A. J., King, T. V. V., and Calvin, W. M., 1993. The U.S. Geological Survey, Digital Spectral Library: Version 1: 0.2 to 3.0 microns. Open file report, U.S. Geological Survey.
- Cohen, M. J., Prenger, J. P., and DeBusk, W. F., 2005. Visible-near infrared reflectance spectroscopy for rapid, nondestructive assessment of wetland soil quality. *Journal of Environmental Quality*, 34, 4, 1422-1434.
- Commission Internationale de l'Eclairage, 1978. Proceedings of the 13th session. Cambridge, England, Bureau Central de la CIE, Paris.
- Commission of the European Communities [CEC], 2002. Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions - Towards a Thematic Strategy for Soil Protection - COM 179.
- Commission of the European Communities [CEC], 2006. Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions - Thematic Strategy for Soil Protection - COM 231.
- Condit, H. R., 1970. The spectral reflectance of American soils. *Photogrammetric Engineering & Remote Sensing*, 36, 955-966.
- Cozzolino, D. and Morón, A., 2006. Potential of near-infrared reflectance spectroscopy and chemometrics to predict soil organic carbon fractions. *Soil and Tillage Research*, 85, 1-2, 78-85.
- Curran, P. J. and Steele, C. M., 2005. MERIS: The re-branding of an ocean sensor. *International Journal of Remote Sensing*, 26, 9, 1781-1798.
- EEA, European Environment Agency, 3-2-2006. Corine land cover 2000 (CLC2000) 100 m - version 8/2005. <http://dataservice.eea.europa.eu/dataservice/metadetails.asp?id=822> [1-2-2007].
- European Commission, 2005. Image 2000 and CLC 2000 Products and Methods. EUR 21757 EN, European Commission, Joint Research Centre, Ispra, Italy.
- European Soil Bureau, 2003. Carta Ecopedologica d'Italia, 1:250,000 Map and Database. Relazione Nr. 3, CD, Ministero Ambiente Italia and European Commission - DG Joint Research Centre, Ispra, Italy.
- European Soil Bureau and Regione Puglia, 2001. INTERREG II Italia - Albania - Sviluppo di una base dati georiferita relativa al suolo dell'intera regione Puglia (Italia).
- Fidêncio, P. H., Poppi, R. J., and de Andrade, J. C., 2002. Determination of organic matter in soils using radial basis function networks and near infrared spectroscopy. *Analytica Chimica Acta*, 453, 125-134.

- Garcia-Haro, F. J., Sommer, S., and Kemper, T., 2005. A new tool for variable multiple endmember spectral mixture analysis (VMESMA). *International Journal of Remote Sensing*, 26, 10, 2135-2162.
- Geladi, P. and Martens, H., 1996. A calibration tutorial for spectral data. Part 1. Data pretreatment and principal component regression using Matlab. *Journal of Near Infrared Spectroscopy*, 4, 1-4, 225-242.
- Gobron, N., Ausetat, O., Lavergne, T., Taberner, M., Pinty, B., Brito, F., Faber, O., Brockmann, C., Lin, F., Robustelli, M., and Snoeij, P., 2006. Validation of the operational MERIS FAPAR products. European Space Agency, (Special Publication) ESA SP.
- Gobron, N., Ausetat, O., Pinty, B., Taberner, M., and Verstraete, M., 2004. Medium Resolution Imaging Spectrometer (MERIS) Level 2 Land Surface Products Algorithm Theoretical Basis Document. JRC Technical Report, EUR 21387 EN, European Commission, Joint Research Centre, Ispra, Italy.
- Gobron, N., Pinty, B., Verstraete, M. M., and Widlowsky, J.-L., 2000. Advanced vegetation indices optimized for up-coming sensors: Design, performance and applications. *IEEE Transactions on Geosciences and Remote Sensing*, 38, 2489-2505.
- Haaland, D. M. and Thomas, E. V., 1988a. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Analytical Chemistry*, 60, 11, 1193-1202.
- Haaland, D. M. and Thomas, E. V., 1988b. Partial least-squares methods for spectral analyses. 2. Application to simulated and glass spectral data. *Analytical Chemistry*, 60, 11, 1202-1208.
- Hildrum, K. I., Isaksson, T., Naes, T., and Tandberg, A., 1992. Near infrared spectroscopy, bridging the gap between data analysis and NIR applications. Ellis Horwood, New York.
- Hill, J. and Schütt, B., 2000. Mapping Complex Patterns of Erosion and Stability in Dry Mediterranean Ecosystems. *Remote Sensing of Environment*, 74, 557-569.
- Hruschka, R. W., 1987. Data analysis: wavelength selection methods. In: Williams, P. C. and Norris, K., Ed(s), *Near-infrared technology in the agriculture and food industry*. American Association of Cereal Chemists, St. Paul, Minnesota.
- Jarmer, T., 2005. Der Einsatz von Reflexionsspektrometrie und Satellitenbildern zur Erfassung padochemischer Eigenschaften in semi-ariden und ariden Gebieten Israels. *Trierer Geographische Studien*, Heft 29, Geographische Gesellschaft Trier.
- Jones, R. J. A., Hiederer, R., Rusco, E., Loveland, P. J., and Montanarella, L., 2004. The map of organic carbon in topsoils in Europe, Version 1.2, September 2003: Explanation of Special Publication Ispra 2004 No. 72 (S.P.i.04.72). European Soil Bureau Research Report No. 17, EUR 21209 EN, Office for Official Publications of the European Communities, Luxembourg.
- Jones, R. J. A., Hiederer, R., Rusco, E., and Montanarella, L., 2005. Estimating organic carbon in the soils of Europe for policy support. *European Journal of Soil Science*, 56, 655-671.

- Kemper, T., 2003. Reflectance spectroscopy for mapping and monitoring of metal mining related contamination. A case study of the Aznalcóllar mining accident, Spain. PhD Universität Trier, Germany.
- Kemper, T. and Sommer, S., 2002. Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. *Environmental Science and Technology*, 36, 12, 2742-2747.
- Kibblewhite, M. G., Rubio, J. L., Kosmas, C., Jones, R. J. A., and Arrouays, D., 2007. Environmental assessment of soil for monitoring desertification in Europe. Eighth session of the Conference of the Parties (COP 8) to the United Nations Convention to Combat Desertification (UNCCD), Madrid, 3 - 14 September 2007, ISBN 1-871315-97-2, Cranfield UK.
- Kooistra, L., Wanders, J., Epema, G. F., Leuven, R. S. E. W., Wehrens, R., and Buydens, L. M. C., 2003. The potential of field spectroscopy for the assessment of sediment properties in river floodplains. *Analytica Chimica Acta*, 484, 189-200.
- Krishnan, P., Alexander, D. J., Butler, B. J., and Hummerl, J. W., 1980. Reflectance technique for predicting soil organic matter. *Soil Science Society of American Journals*, 44, 1282-1285.
- Kruse, F. A. H., Lefkoff, A. B., Boardman, J. B., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., and Goetz, A. F. H., 1993. The spectral image processing system (SIPS) - Interactive visualisation and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 44, 145-163.
- Lavine, B. K., 1998. Chemometrics. *Analytical Chemistry*, 70, 12, 209R-228R.
- Leger, R. G., Millette, G. J. F., and Chomchan, S., 1979. The effects of organic matter, iron oxides and moisture on the soil colour of two agricultural soils in Quebec. *Canadian Journal of Soil Science*, 59, 191-202.
- Leone, A. P. and Sommer, S., 2000. Multivariate analysis of laboratory spectra for the assessment of soil development and soil degradation in the Southern Apennines (Italy). *Remote Sensing of Environment*, 72, 346-359.
- Lin, H., Wheeler, D., Bell, J., and Wilding, L., 2005. Assessment of soil spatial variability at multiple scales. *Ecological Modelling*, 182, 271-290.
- Machwitz, M., 2005. Methoden zur Bestimmung der organischen Substanz auf landwirtschaftlichen Flächen in Süditalien mit Hilfe von Spektrometernmessungen und Anwendung des Aufnahmesystems ASTER. Diplomarbeit Universität Trier, Trier.
- Madden, H., 1978. Comments on the Savitzky-Golay convolution method for least square fit smoothing and differentiation of digital data. *Analytical Chemistry*, 50, 9, 1383-1386.
- Malley, D. F. and Williams, P. C., 1997. Use of near-infrared reflectance spectroscopy in prediction of heavy metals in freshwater sediment by their association with organic matter. *Environmental Science and Technology*, 31, 12, 3461-3467.
- Martens, H. and Naes, T., 1989. *Multivariate Calibration*. Wiley & Sons, Chichester.

- Massart, D. L., Vandeginste, B. G. M., Deming, S. N., Buydens, L. M. C., De Jong, S., Lewi, P. J., and Smeyers-Verbeke, J., 1997. Handbook of chemometrics and qualimetrics, part A. Elsevier, Amsterdam.
- Massart, D. L., Vandeginste, B. G. M., Deming, S. N., Buydens, L. M. C., De Jong, S., Lewi, P. J., and Smeyers-Verbeke, J., 1998. Handbook of chemometrics and qualimetrics, part B. Elsevier, Amsterdam.
- Masserschmidt, I., Cuelbas, C. J., Poppi, R. J., de Andrade, J. C., de Abreu, C. A., and Davanzo, C. U., 1999. Determination of organic matter in soils - FTIR/Diffuse reflectance and multivariate calibration. *Journal of Chemometrics*, 13, 265-273.
- McBratney, A. B., Minasny, B., and Viscarra Rossel, R., 2006. Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis. *Geoderma*, 136, 1-2, 272-278.
- McCarty, G. W., Reeves III, J. B., Reeves, V. B., Follett, R. F., and Kimble, J. M., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Science Society of American Journals*, 66, 640-646.
- Mehl, W., 2004. Operational generation of time-series of high resolution satellite imagery. [Unpublished]
- Otto, M., 1999. *Chemometrics - statistics and computer application in analytical chemistry*. Wiley-VCH, Weinheim.
- Page, N. R., 1974. Estimation of organic matter in Atlantic coastal plain soils with a colour difference meter. *Agronomy Journal*, 652-653.
- Pell, R. J., 2000. Multiple outlier detection for multivariate calibration using robust statistical techniques. *Chemometrics and intelligent Laboratory Systems*, 52, 1, 87-104.
- Reeves, J., McCarty, G., and Mimmo, T., 2002. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. *Environmental Pollution*, 116, 277-284.
- S.I.L.P.A., 2001. Studio delle correlazioni fra metodi ISO e metodi italiani per le analisi chimiche e chimico-fisiche dei campioni di terreno rappresentativi prelevati in Italia nel programma siti NATURA 2000, European Commission: Contract No. 17957-2001-05F1ED ISP IT.
- Salgó, A., Nagy, J., and Tarnóy, J., 1998. Characterization of soils by the near infrared technique. *Journal of Near Infrared Spectroscopy*, 6, 199-203.
- Salminen, R. and Tarvainen T., 2006. EuroGeoSurveys-FOREGS Geochemical Baseline Mapping Programme. Geological Survey of Finland, Espoo. <http://en.gtk.fi/Research/CR/foregs.html> [9-21-2008].
- Santer, R., Carrere, V., Dubuisson, P., and Roger, J. C., 1999. Atmospheric correction over land for MERIS. *International Journal of Remote Sensing*, 20, 9, 1819-1840.
- Savitzky, A. and Golay, M. J. E., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36, 8, 1627-1639.
- Schlichting, E., Blume, H. P., and Stahr, K., 1994. *Bodenkundliches Praktikum*. Blackwell, Berlin.

- Schulze, D. G., Nagel, J. L., Van Scoyoc, G. E., Henderson, T. L., Baumgardner, M. F., and Stott, D. E., 1993. Significance of organic matter in determining soil colors. SSSA Special Publication, Soil Color, 31, Soil Science Society of America, San Antonio, Texas, USA, pp. 71-90 .
- Shields, J. A., Paul, E. A., Arnaud, S., and Head, W. K., 1968. Spectrophotometric measurement of soil colour and its relationship to moisture and organic matter. Canadian Journal of Soil Science, 45, 271-280.
- Spielvogel, S., Knicker, H., and Kögel-Knabner, I., 2004. Soil organic matter composition and soil lightness. Journal of Plant Nutrition and Soil Science, 167, 545-555.
- Steinier, J., Termonia, Y., and Deltour, J., 1972. Comments on smoothing and differentiation of data by simplified least square procedure [2]. Analytical Chemistry, 44, 11, 1906-1909.
- Stevens, A., 2008. Changes in soil organic carbon at regional scale: Strategies to cope with spatial variability. PhD Catholic University of Louvain, Belgium.
- Stolbovoy, V., Marechal, B., Jones, A. Rusco E., and Montanarella, L., 2008. Climate change - soil can make a difference! Slide presentation at the conference: Climate change - can soil make the difference? Brussels, 12th June 2008, European Commission, Joint Research Centre, pp. 1-14. Contact: Luca Montanarella, <http://eussoils.jrc.ec.europa.eu>
- Stoner, E. R. and Baumgardner, M. F., 1981. Characteristic variations in reflectance of surface soils. Soil Science Society of America Journal, 45, 6, 1161-1165.
- Todorovic, M. and Steduto, P., 2003. A GIS for irrigation management. Physics and Chemistry of the Earth, 28, 163-174.
- Udelhoven, T., Emmerling, C., and Jarmer, T., 2003. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: a feasibility study. Plant and Soil, 251, 319-329.
- Udelhoven, T., Jarmer, T., and Hill, J., 2001. Die Nutzung von Hyperspektraldaten zur Erfassung von Bodeneigenschaften - Grundlagen und Anwendungsperspektiven, 21. Wissenschaftlich-Technische Jahrestagung der DGPF, Konstanz, pp. 369-377.
- Udelhoven, T. and Schütt, B., 2000. Capability of feed-forward neural networks for a chemical evaluation of sediments with diffuse reflectance spectroscopy. Chemometrics and intelligent Laboratory Systems, 51, 9-22.
- Université du Littoral Côte d'Opale, 2000. Atmospheric products over land. http://envisat.esa.int/instruments/meris/pdf/atbd_2_15.pdf [9-21-2008].
- Van Ranst, E., Thomasson, A. J., Daroussin, J., Hollis, J. M., Jones, R. J. A., Jamagne, M., King, D., and Vanmechelen, L., 1995. Elaboration of an extended knowledge database to interpret the 1:1,000,000 EU Soil Map for environmental purposes. In: King, D., Jones, R. J. A., and Thomasson, A. J., Ed(s), European Land Information Systems for Agro-environmental Monitoring. Office for Official Publications of the European Communities, Luxembourg, pp. 71-84.

- Vanmechelen, L., Groenemans, R., and Van Ranst, E., 1997. Forest soil condition in Europe. Results of the Large-Scale Soil Survey. Brussels Geneva.
- Viscarra Rossel, R. A., Minasny, B., Roudier, P., and McBratney, A. B., 2006a. Colour space models for soil science. *Geoderma*, 133, 3-4, 320-337.
- Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., and Skjemstad, J. O., 2006b. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, 131, 1-2, 59-75.
- Von Hoyningen-Huene, W., Berthelot, B., Kokhanovsky, A., and Burrows, J. P., 2006. An ATBD for the ENVISAT radiometer: Determination of aerosol optical thickness over land surfaces, using Bremen Aerosol Retrieval (BAER) and its applications to atmospheric correction over lands. [Online] European Space Agency. http://www.brockmann-consult.de/beam/software/plugins/baer-1.0.0/NOV-3341-NT-3352_v1.0.pdf [9-21-2008].
- Williams, P. C. and Norris, K., 1987. Near-infrared analysis in agriculture and food industries. American Association of Food Chemists Inc., Williams, P. C. and Norris, K., St. Paul, Minnesota.
- Wold, H., 1982. Soft modelling. The basic design and some extensions. In: Jöreskog, K.-G. and Wold, H., Ed(s), *System under indirect observation*. Amsterdam.
- Wold, S., Martens, H., and Wold, H., 1983. The multivariate calibration problem in chemistry solved by the PLS method. *Proceedings of the Conference of Matrix Pencils*, Ruhe, A. and Kagström, B., Springer, Heidelberg, pp. 286-289.
- Wold, S., Sjöström M., and Eriksson L., 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58, 109-130.
- Wyszecki, G. and Stiles, W. S., 1982. *Color science: concepts and methods, quantitative data and formulas*. 2nd, John Wiley and Sons, New York.
- Zdruli, P., Jones, R. J. A., and Montanarella, L., 2004. Organic matter in the soils of Southern Europe. European Soil Bureau Technical Report, EUR 21083, Office for Official Publications of the European Communities, Luxembourg.

Acknowledgements

Part II of this work was performed under the IP DESURVEY project (IP Contract FP6 GCR Programme contract No 003950) funded by the EC DG RTD 6th Framework Programme. MERIS FR level 2 data were provided by the European Space Agency under the ESA AO project 631.

In particular we wish to thank our colleagues of the JRC-IES SOIL action (formerly JRC FP6 MOSES action), especially Luca Montanarella, Bob Jones, Ezio Rusco, Vladimir Stolbovoy and Nicola Filippi, for many constructive discussions and last but not least for sharing with us the field samples including the analytical data and the OCTOP data base.

European Commission

EUR 23513 EN – Joint Research Centre – Institute for Environment and Sustainability

Title: Chemometric Modelling and Remote Sensing of Arable Land Soil Organic Carbon as Mediterranean Land Degradation Indicator

Author(s): Kristin Böttcher, Thomas Kemper, Miriam Machwitz, Wolfgang Mehl, Stefan Sommer

Luxembourg: Office for Official Publications of the European Communities

2008 – 65 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1018-5593

ISBN 978-92-79-09811-6

Abstract

The application of chemometric models for the quantitative estimation of soil organic matter (SOM) from laboratory reflectance data on a regional/country-wide level is explored in the first part of this report. In addition, the possibility to transfer the developed models to operational satellite systems has been evaluated. 376 soil samples were collected for previous projects within a wide range of environmental settings in Italy. Soil physical and chemical laboratory analyses results were provided by activities of JRC-IES under the auspices of the European Soil Bureau. Reflectance measurements were obtained on disturbed soil samples using an ASD Field Spec Pro spectroradiometer. Data transformation methods (standardisation, vector-normalisation and first and second order derivatives) have been applied on the spectral data. The transformed spectral data have been used for the prediction of SOM and carbonate content using the partial least squares regression (PLSR). The results (R^2 between 0.57 and 0.8) demonstrate the successful application of reflectance spectroscopy combined with chemometric modelling for the estimation of SOM and carbonate content. The calibration models demonstrated a tolerable stability over a variety of different soil types, which gives the opportunity for monitoring larger areas. Furthermore it could be shown, that the spectral resolution of the MERIS sensor is sufficient to estimate roughly the SOM content.

A second study was conducted on the use of MERIS satellite data for the estimation of soil organic carbon content, which is described in the second part of this report. This study concentrated on the Apulia region, as sampling in other regions in Italy was not sufficiently dense to assure sound model calibration and validation. For specific morphological-lithological units simple spectral models, based on soil colour and shape attributes, were applied to derive soil organic carbon content.

In order to apply these models to MERIS satellite data, a time series of images covering the years 2003 and 2004 was acquired for Southern Italy. Pre-processing of image data aimed at the reduction of vegetation related effects on image spectra including the derivation of minimum vegetation composites and the unmixing of soil and vegetation. The spatial distribution of soil organic carbon was estimated for each year within the specific morphological-lithological units in the Apulia region. In addition the models could be applied to other regions in Southern Italy, especially the Southern part of the Abruzzo region. The results showed good agreement with independent field data and with the Europe wide estimations of the OCTOP data base described by Jones et al. (2005).

How to obtain EU publications

Our priced publications are available from EU Bookshop (<http://bookshop.europa.eu>), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

