

Co-kriging of soil properties with Vis-NIR hyperspectral covariates in the Cap Bon region (Tunisia)

R. Ciampalini & P. Lagacherie

INRA, Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH), IRD-INRA-SupAgro, Montpellier, France

P. Monestiez & E. Walker

INRA, Unité de Biostatistique et Processus Spatiaux (BioSP), Avignon, France

C. Gomez

IRD, Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH), IRD-INRA-SupAgro, Montpellier, France

ABSTRACT: Visible and Near Infrared (Vis-NIR) hyperspectral airborne spectroscopy can be used for predicting soil surface properties but its use is constrained to bare soil surfaces. To extend its use to larger areas, a co-kriging procedure was applied across a 339 km² area located in the Cap Bon Region (northern Tunisia). The study used 262 soil surface analysed samples and, as covariates, the bare soil estimates of four topsoil properties (Clay, Sand, Iron contents and Cation Exchange Capacity) derived from a 30 meter resolution Vis-NIR AISA-Dual hyperspectral image. The resulting co-regionalisation models and co-kringed maps allowed to capture the soil pattern of the study area which indicates a promising opportunity for using Vis-NIR hyperspectral covariates for Digital Soil Mapping. However, the co-kriging performances were limited by the short-scale soil variations of this lithology-driven Mediterranean study area.

1 INTRODUCTION

In many parts of the world, the available legacy soil data are too sparse for the Digital Soil Mapping (DSM) of soil properties with acceptable precision (e.g. Ciampalini et al., 2012). It is therefore necessary to collect new input soil data while maintaining acceptable costs.

It has been showed recently (Gomez et al., submitted, Gomez et al., 2012) that Vis-NIR hyperspectral airborne spectroscopy could provide acceptable estimates of some key topsoil properties. However, prediction of the soil properties is limited to bare soil surfaces. In recognition of this, Lagacherie et al. (2012) used co-kriging and block-cokriging for extending the hyperspectral estimates to an entire study area.

From the encouraging results of Lagacherie et al. (2012), Vis-NIR hyperspectral imagery is expected to be largely used for DSM in the future, especially in Mediterranean and semi-arid areas where bare soil surfaces are common and where dry periods can allow to avoid soil moisture

perturbations of the spectrum (Lagacherie et al., 2008). Further Vis-NIR hyperspectral sensors are planned to be launched on board satellites within the next two years, PRISMA (Giampaolo et al., 2008) and EnMap (Stuffer et al., 2007).

In this paper, a cokriging procedure is applied to map four topsoil properties (Clay, Sand, iron contents and CEC) in the Cap Bon region (Northern Tunisia) using as covariates the accurate but incomplete estimates derived from an Vis-NIR hyperspectral image (Gomez et al., 2012).

2 THE CASE STUDY

2.1 Study area

The study area is located in the Cap Bon region in northern Tunisia (36°24'N to 36°53'N; 10°20'E to 10°58'E), 60 km east of Tunis, Tunisia (Figure 1a). This 300 km² area includes the Lebna catchment (Figure 1b), which is mainly rural (>90%) and devoted to cereals in addition to legumes,

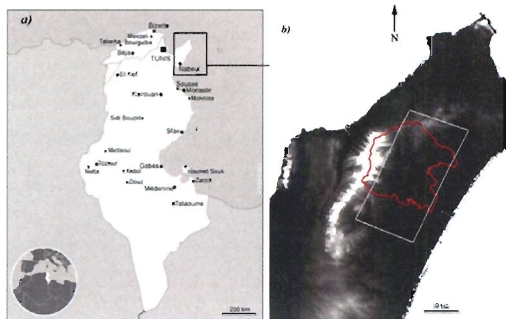


Figure 1. a) Location of the Cap Bon region in the Northern Tunisia, b) limits of the Lebna Catchment (in red) and the AISA hyperspectral image (in grey) plotted over the STRM DEM of the Cap Bon.

olive trees, natural vegetation for breeding and vineyards. It is characterized by relief areas, with an altitude between 0 and 226 m. The main soil types are Regosols, Eutric Regosols (9.6%) preferentially associated with sandstone outcrops, Calcic Cambisols, and Vertisols preferentially formed on marl outcrops and lowlands. The southeastern region of the study area has a flatter landscape with sandy Pliocene deposits yielding Calcisols and Rendzina.

2.2 Field data

262 soil samples were collected between October 2008 and November 2010 (Figure 2). 129 of these soil samples were collected within the hyperspectral image perimeter in fields that were bare during the hyperspectral data acquisition in November 2010. The remaining 133 samples were located either outside this image perimeter (73 samples) or in vegetated areas (60 samples). The former 129 were used in Gomez et al. (2012) for calibrating the hyperspectral estimation functions. The samples were composed of five sub-samples collected to a depth of 5 cm at random locations within a 10 × 10 m square centered on the geographical position of the sampling plot, as recorded by a Garmin GPS instrument. Soil properties were measured using classical physico-chemical soil analysis. In this study only the four soil properties successfully predicted from hyperspectral data ($R^2 > 0.6$) were considered: clay, sand, free iron and cation exchange capacity (CEC).

2.3 Hyperspectral covariates

The covariates used in this study, called hereon "hyperspectral covariates", were the maps of topsoil properties derived from Vis-NIR AISA-Dual

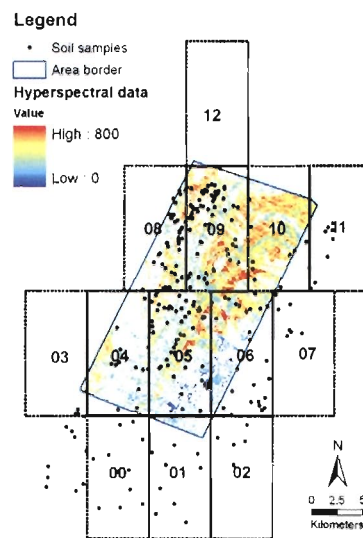


Figure 2. Locations of the samples and delimitation of computing zones for the co-kriging.

hyperspectral data which were acquired on 2nd November 2010 over the study area with a spatial resolution of 5 m and 359 non-contiguous bands covering the 400 to 2500 nm spectral domain. For computational efficiency, a pixel aggregation resampling method was used to provide Vis-NIR hyperspectral image at 30 m of spatial resolution. It must be noted that 30 meters is the resolution of the Vis-NIR hyperspectral sensors that are planned to be launched on board satellites (Giampaolo et al., 2008, Stuffer et al., 2007). After artifact filtering, it comprised 280 spectral bands covering the 450 to 2450 nm spectral domain. Vegetated surfaces, urban areas and water areas were masked to consider only bare soil surfaces for further soil predictions. The methodology for obtaining maps of predicted topsoil properties is fully detailed in (Gomez et al., 2012). Only a short summary is provided here.

The Partial Least Square Regression (PLSR) was used to establish relationships between the topsoil properties and the AISA-Dual spectra. PLSR-based prediction models were built using the 129 AISA-Dual spectra corresponding to the location of the soil samples collected over bare soils. Correct prediction models, with R^2 and RPD values greater than 0.6 and 1.4 respectively, were obtained for four soil properties: free iron, CEC, clay and sand content.

Figure 3 shows an example of the resulting 30 meter resolution images of soil properties that covered 41.6% of the studied area.

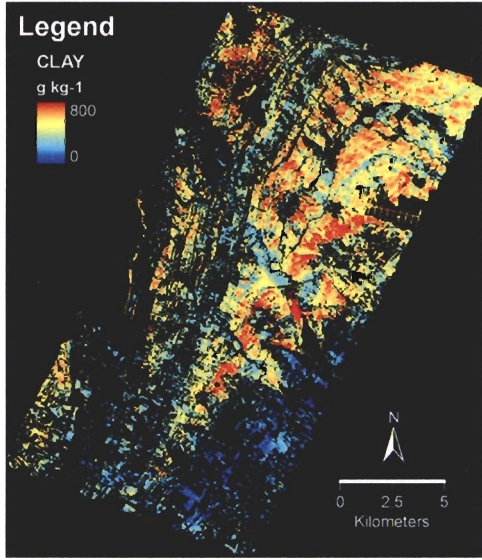


Figure 3. Clay content predicted over bare soils from AISA—Dual spectra (black areas correspond to non-predicted surfaces).

3 METHOD

The set of measured soil samples and the images of estimated soil properties were co-kriged to obtain new images that covered the whole study area. We applied a procedure that was recently developed for mapping topsoil properties from hyperspectral data in southern France (Lagacherie et al., 2012). In this study however; block-cokriging was replaced by punctual co-kriging.

All the analyses in this study were performed using R (R Development Core Team, 2007).

3.1 Modelling multivariate spatial correlations

In this study, the measured soil properties and their hyperspectral covariates are denoted as Z_1 and Z_2 , respectively. Suppose that u is a location in two-dimensional space and $Z_1(u)$ and $Z_2(u)$ are spatial random functions. Assuming that the soil property (Z_1) is spatially cross-correlated with the hyperspectral covariate (Z_2), the spatial cross correlation between Z_1 and Z_2 can be quantified by a cross-variogram (Wackernagel, 1995). In univariate or bivariate frameworks, the covariance and variogram functions can be estimated as follows:

$$\hat{\gamma}_{z_i z_j}(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} (z_i(u_\alpha + h) - z_i(u_\alpha)) \times (z_j(u_\alpha + h) - z_j(u_\alpha)) \quad (1)$$

In Equation (1) i and j belong to $\{1, 2\}$. When $i = j$, Equations (1) denotes the variogram estimates. When $i \neq j$, Equations (1) denotes the cross variogram estimates. h is the separation vector between the data locations u_α and $u_\alpha + h$ (the translated of h from u_α). $Z_1(u_\alpha)$ and $Z_2(u_\alpha + h)$ are observations of the variable z_i and z_j at spatial locations u_α and $u_\alpha + h$, respectively, and $N(h)$ is the number of distinct pairs of observations at distance h .

To undertake the co-kriging, a variogram matrix in which the diagonal entries are variograms and the off-diagonal entries are cross variograms must be strictly conditionally negative definite. To ensure this condition, one can use intrinsic or linear co-regionalisation models. The formulation of the latter in the bivariate case with two nested spatial structures is (Wackernagel, 1995).

$$\Gamma(h) = B_1 g_1(h) + B_2 g_2(h) \quad (2)$$

where $g_1(h)$ and $g_2(h)$ are two normalized variograms, one for each spatial structure, and B_1 and B_2 are positive semi-definite 2×2 matrices.

3.2 Co-kriging

The co-kriging estimator is a best linear unbiased estimator (BLUE) and has minimum estimation error variance (Wackernagel, 1995). In the two variables case, the ordinary co-kriging estimator is a linear combination of weights w_α^1 and w_α^2 with data from the two variables Z_1 and Z_2 located at sample points in the neighborhood of a spatial location u_0 . Each variable is defined on a set of samples of possibly different sizes n_1 and n_2 , and the estimator is defined as:

$$\hat{Z}_1(u_0) = \sum_{\alpha=1}^{n_1} w_\alpha^1 Z_1(u_\alpha) + \sum_{\alpha=1}^{n_2} w_\alpha^2 Z_2(u_\alpha) \quad (3)$$

where the weights w_α^1 and w_α^2 are solutions of a co-kriging system and sum to 1 and 0, respectively.

The co-kriging variance of the estimation error of \hat{Z}_1 in the two variables case can be estimated from the variogram $\gamma_{z_1 z_1}$ and the crossvariogram $\gamma_{z_1 z_2}$ (Wackernagel, 1995) by using the following expression:

$$\sigma_E^2(u_0) = \sum_{\alpha=1}^{n_1} w_\alpha^1 \gamma_{z_1 z_1}(u_\alpha - u_0) + \sum_{\alpha=1}^{n_2} w_\alpha^2 \gamma_{z_1 z_2}(u_\alpha - u_0) + \mu_{z_1} \quad (4)$$

where μ_{z_1} is the lagrange multiplier of the co-kriging system and $u_\alpha - u_0$ denotes the distance between u_α and u_0 locations.

3.3 Co-kriging with large datasets

The large number of sites with hyperspectral covariates made it difficult to co-krige from the whole set of these sites at every prediction location. Subsets were therefore considered at the neighborhood of the prediction locations by randomly selecting at most 500 sites within a $300\text{ m} \times 300\text{ m}$ area centered on the prediction location and at most 1000 additional sites within a $1000\text{ m} \times 1000\text{ m}$ area centered on the prediction location.

3.4 Validation

The validation of the co-kriging outputs was done on the 189 soil samples located in the image perimeter by comparing the true measurements with the predicted values obtained from a leave-one-out cross-validation. To evaluate the prediction quality, the classical indicators R^2 and RMSE were calculated. To match GlobalSoilMap.net specifications, we also calculated the observed fractions A of the true values falling into the estimated 95%-probability interval (PI_{95}) bounded by 0.025 and 0.975 quantiles. It was expected that the computed fractions A were close to 95% and that the PI_{95} were as narrow as possible.

4 RESULTS

4.1 Co-regionalisation models

Linear co-regionalisation models were built for the pairs “soil properties-hyperspectral covariates” from the set of 129 sites at which these two variables were available. The two direct variograms were first modeled as linear combinations of two selected basic structures. The same basic structures were then fitted to the cross-semi-variograms under the positive semi-definite constraints (Goovaerts, 1997).

Figures 4a-c show the fitted co-regionalisation models for clay content, CEC, and iron content as well as their analytical expressions. In general the models fitted adequately to the data. None of the models had a nugget effect. The ranges were very similar across the soil properties with a short range of 250 and large range between 1500 and 2000 m. Short range semi-variances were clearly larger than the large range ones.

4.2 Validation results

Table 1 shows the summary of the performance parameters of the cross validation procedure. RMSE and R^2 values showed moderate accuracies with clear decreases at the sites located in the gaps of the hyperspectral image (digits between

brackets in table 1). At such sites, the distances from any available data were often greater than the shortest ranges at which most of the variances were observed, especially for iron.

This result was fairly well anticipated by the model by estimating 95%-probability interval widths. These widths were however slightly underestimated as shown by the fractions of the true values falling in the PI_{95} that were less than the expected 0.95 values.

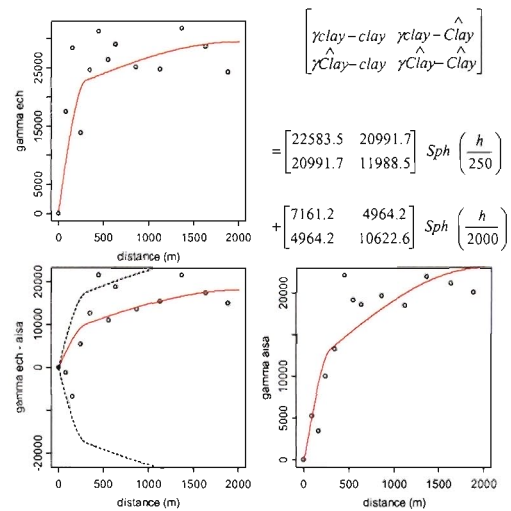


Figure 4a. The linear model of co-regionalization for clay content (clay) and its hyperspectral estimate (\hat{Clay}).

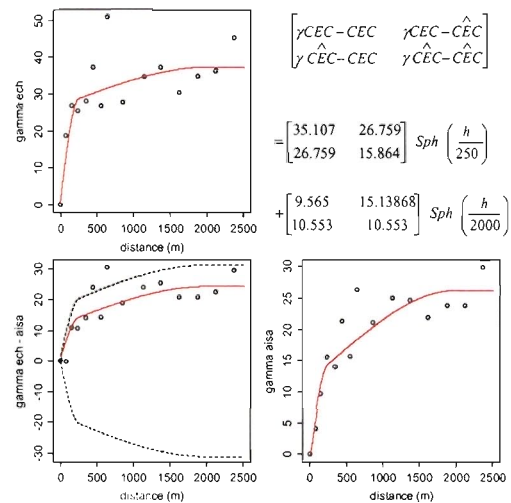


Figure 4b. The linear model of co-regionalization for Cation Exchange Capacity (CEC) and its hyperspectral estimate (\hat{CEC}).

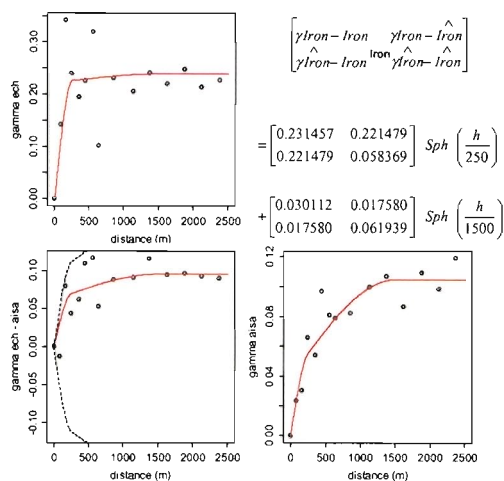


Figure 4c. The linear model of co-regionalization for iron content (Iron) and its hyperspectral estimate (\hat{Iron}).

Table 1. Validations results.

Soil property	RMSEcv*	R ² cv	PI95*** width*	A****
Sand	163 (207)	0.51 (0.32)	464	0.86
CEC	4.9 (6.2)	0.54 (0.32)	14.2	0.88
Iron	0.39 (0.47)	0.41 (0.31)	1.3	0.87

* in g/kg for clay and sand, in meq/100 g for CEC and g/kg for iron.

** between brackets: results for validation points located in the gaps of the hyperspectral image.

*** PI95 is the estimated 95%-probability interval (PI₉₅) bounded by 0.025 and 0.975 quantiles.

**** A is the fraction of the true values of soil property falling in 95% PI.

4.3 Mapping results

Figures 5 show the estimated maps of the four studied soil properties and their associated uncertainty maps that were obtained from co-kriging with the hyperspectral covariates.

The soil property maps exhibit short scaled and striped patterns that are similar across the set of studied properties. Areas with high clay and iron contents, low sand contents and high CEC contrasted with areas with moderate clay and iron contents, high sand contents and moderate CEC. These patterns strongly resemble those of the hyperspectral covariate (Figure 3) with the noticeable exception of the south-east area where the low extreme values of clay contents disappeared in the co-kringed maps (Figure 5a).

The uncertainty maps show variations that were influenced by the locations of the measured sites around which the predicted uncertainties were

minimal—and by the presence of hyperspectral covariates—as shown by the north-south increase of uncertainty with the decrease of hyperspectral covariate availability. The maximal uncertainty was predicted for a small area located at the east of the study area due to the distance from measured sites and hyperspectral covariate data (Figure 2).

5 DISCUSSION & CONCLUSION

5.1 Modelling a Mediterranean soil pattern

The motivation of this study was to deal with soil variations within a sample area of the North Africa region which is poor in available legacy soil data. However, through the use of Vis-Nir hyperspectral imagery that covers the study area (Gomez et al., 2012), the complex soil pattern of this region could be quantitatively modeled in great detail (30 meter resolution). The co-regionalization models (Figures 4) and the co-kriging maps (Figures 5) of four correlated soil properties (clay, sand, iron and CEC) respectively revealed a multi-scaled and anisotropic soil pattern that we interpret as mainly driven by successions of tectonised marl and sandstone outcrops with a decrease of the distances between these successive outcrops from East to West. This lithology driven soil pattern is expected to frequently occur within Mediterranean and semi-arid regions having a geology as complex as the one of the Cap Bon Region.

5.2 Handling incomplete spatial covariates

Spatially incomplete landscape covariates have been rarely handled in DSM studies. Yet it may be of great interest to mask the covariate layers in locations where the local conditions weaken the correlation with a soil property as e.g. vegetated areas for hyperspectral covariates. In such situations, we confirmed after Lagacherie et al. (2012) that co-kriging is a suitable DSM procedure provided that a co-regionalization model can be fitted to the data. We observed however that the co-kriging accuracy could rapidly decrease for sites located in the gaps of a covariate layer if short scale soil variations predominate. Co-kriging may also introduce a bias if the soil of the masked areas differ significantly from the not-masked ones. This may occur if soil properties are correlated with the local conditions that are considered for masking the landscape covariates.

ACKNOWLEDGMENTS

The authors are indebted to UMR LISAH (IRD, France) and to CNCT (Centre National

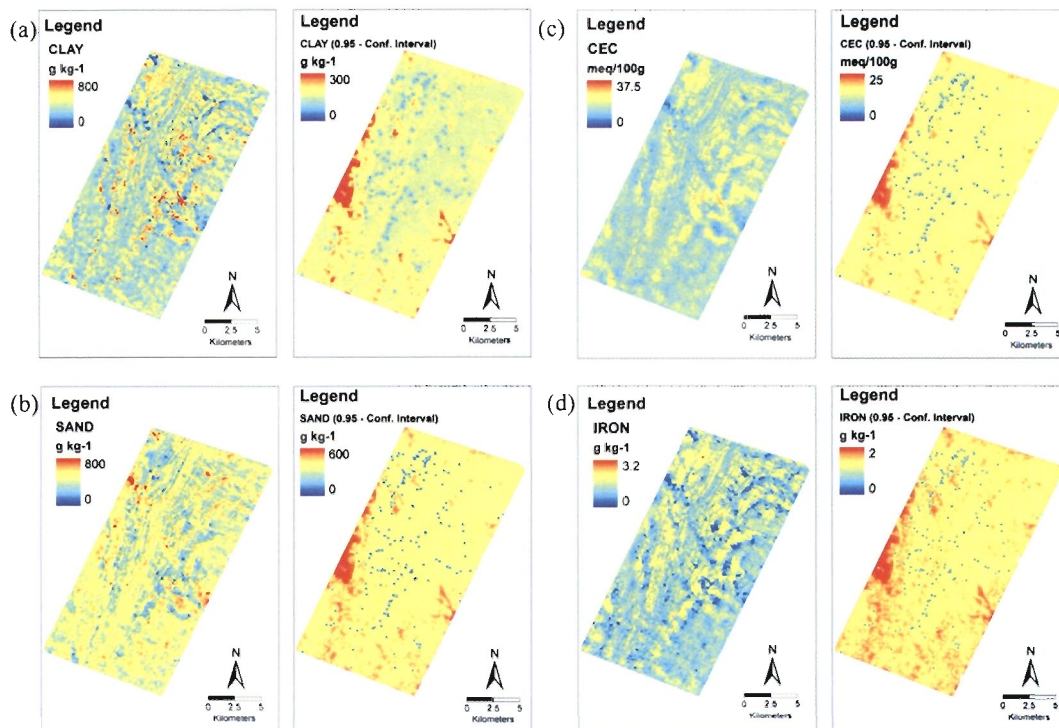


Figure 5. Maps of predicted soil properties and of their associated uncertainty (width of 95%-probability interval).

de Cartographie et de Télédétection, Tunisia), for providing the AISA-Dual images for this study. This hyperspectral data acquisition was granted by IRD, INRA and the French National Research Agency (ANR) (ANR-O8-BLAN-C284-01) ». We are also indebted to Yves Blanca (IRD-UMR LISAH Montpellier), Zakia Jenhaoui (IRD-UMR LISAH Tunis) for the soil sampling in 2009 and 2010 over the Lebna catchment and to Hedi Hamrouni (DG/ACTA Sol, Tunis) for his significant support to this study.

REFERENCES

- Ciampalini, R., Lagacherie, P. & Hamrouni, H. 2012. Documenting GlobalSoilMap.net grid cells from legacy measured soil profile and global available covariates in Northern Tunisia in *DSM2012*, Minasny et al. (eds). Sydney.
- Preti, G., Cisbani, A., De Cosmo, V., Galeazzi, C., Labate, D. & Melozzi, M. (2008). Hyperspectral Instruments for Earth Observation. International conference on Space Optics, October 14–17 Toulouse, France.
- Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press.
- Gomez, C., Lagacherie, P. & Bacha, S. 2012. Using Vis-NIR hyperspectral data to map topsoil properties over bare soils in the Cap Bon region, Tunisia in *DSM2012*, Minasny et al. (eds). Sydney.
- Gomez, C., Lagacherie, P. & Coulouma, G. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis-NIR data. Under Review in *Geoderma*.
- Lagacherie, P., Bailly, J.S., Monestiez, P. & Gomez, C. 2012. Using scattered hyperspectral imagery data to map the soil properties of a region. *Eur. J. Soil Science*, 63:110–119.
- Ouerghemmi, W., Gomez, C., Nacer, S. & Lagacherie, P. (2011). Applying Blind Source Separation on hyperspectral data for clay content estimation over partially vegetated surfaces. *Geoderma*, 163(3–4), 227–237.
- Stuffer, T., Kaufmann, H., Hofer, S., Förster, K.-P., Schreier, G., Müller, A., Eckardt, A., Bach, H., Penne, B., Benz, U. & Haydn, R. 2007. The EnMAP hyperspectral imager—An advanced optical payload for future applications in Earth observation programmes, *Acta Astronautica*, 61, 1–6, 115–120.
- Wackernagel, H. 1995. *Multivariate geostatistics*. Springer—Verlag. p. 255.