Statistical Analysis of Advanced Very High Resolution Radiometer Data (AVHRR) Soil Relationship

[1] E. DOBOS, [2] E. MICHÉLI and [3] M. F. BAUMGARDNER

[1] University of Miskolc, [2] Gödöllő University of Agricultural Sciences (Hungary) and [3] Purdue University, West Lafayette (USA)

Introduction

In 1986 the International Society of Soil Science initiated an international project to create a World Soil and Terrain Digital Database, called SOTER (ISRIC, 1993). The original idea of doing so is based on existing data from all over the world. Even if the data are available for the entire Earth surface, much research about how to generalize, translate, relate and compile data into a uniform database at a scale of 1:1M should be done. A new methodology was designed, and scientists, who understand the data format and methodology of the original (national) and the SOTER database have faced the challenge of transforming the data and the knowledge into the SOTER. This process is currently going on and sooner or later – depending on the financial sources – the SOTER database will be complete. Unfortunately, data distribution and quality are not homogeneous over the world. There are regions where the soil and terrain data necessary to complete the basic level of SOTER are limited in quality or density, or even may be missing. Hence, a field data collection and analysis has to be carried out or secondary data sources, that are appropriate to extract soil and terrain information from them, have to be utilized in order to reduce financial input and speed up the completion process of the database.

After a literature review, the coarse spatial resolution satellite imageries have been found to be a promising data source for characterizing soils in small scale. The only available coarse spatial resolution satellite data is the AVHRR (Advanced Very High Resolution Radiometer), however a more advanced, state-of-the-art satellite with much better spectral resolution, the MODIS (Moderate Resolution Imaging Spectroradiometer), will be launched soon. Therefore, this study is not only the evaluation of the AVHRR data but also a preliminary study of the possible use of such satellite data sources for the future as the MODIS.

Within the frame of this study we aimed to quantify the level of similarity between the information held by the different channels. The *„similarity of the* *information content"* – meaning the similarity of the value distribution and the spatial variation of the pixel values – had to be identified in order to understand the possibilities of lowering the dimensionality. This goal has been reached by generating the correlation cross matrix of channels. The effect of image dates and their spectral information on the characterization of soil variability had to be analyzed as a second step. In order to do so, a multivariate analysis of variance was performed for the 30 channels covering the area of Hungary. The Coefficient of Multiple Determination (R^2) was calculated for all channels to describe the variation that accounts for the soil variability. This value can give some general information about the *AVHRR –soil relationship.*

Material and Methods

The AVHRR data

The primary data used in this project are from the Advanced Very High Resolution Radiometer (AVHRR) on the National Oceanic and Atmospheric Administration (NOAA) polar orbiting weather satellites. The 10-day composite images were downloaded from the 1-km AVHRR Global Land Data Set of the U.S. Geological Survey EROS Data Center at Sioux Falls, South Dakota (TOWNSHEND et al., 1994). All five spectral bands were used. The characteristics of NOAA/AVHRR system are shown in Table 1. The decadal composites are made up of picture elements (pixels) that have the maximum value out of the 10-day period for a given pixel (Maximum Value Composite-MVC). The basis of data selection was: the cloudfreeness, the lack of snow cover, and the representation of different stages of the vegetative periods. The data that have been chosen are from the periods of 1-10 May, 11-20 August and 11-20 September from 1992, and 1-10 June and 21-30 September from 1993. These data were preprocessed, the raw images were first radiometricly calibrated and georeferenced with the use of ground control points and DEM, and hence the spatial accuracy of the recommended 1000 meter could be reached. The original projection was the Interrupted Goode Homolosine. After the computation of NDVI, the final step was the atmospheric correction for Ozone and Rayleigh Scattering (EIDENSHINK & FAUNDEEN, 1994). The AVHRR data has an original resolution of 1.1 km, which was resampled into 1 km spatial resolution. The downloaded data were then stacked into 6 layer images by date – containing the five channels and the NDVI –, and a transformation into the Hungarian Unified Projection System (EOV) was performed to provide compatibility with the reference maps and digital elevation data. This transformation was done in a way suggested by EIDENSHINK & FAUNDEEN (1994), using hydrologic features from vector datasets in EOV projection. Characteristic points and features were selected from both the image and the vector coverage and within the ARC/INFO register command „links" between the point pairs were established. As

Table 1

NOAA/AVHRR System Characteristics

(from KIDWELL, 1990; IGBP, 1992 and EHRLICH et al., 1994)

Characteristics	
Sensor Characteristics	
Spectral Bandwidth	1. 580-680 nm 2. 735-1100 nm 3. 3550-3930 nm 4. 10300-11300 nm 5. 11500-12500 nm
Radiometric Resolution	10 bits (1024 level)
IFOV (nadir)	1.1 km
View Angle	55.4 degrees (IFOV = 6 km at swath edge) (i)
Swath Width	2700 km
Platform Characteristics	
Orbit	Near-polar, Sun-synchronous
Altitude	833-870 km
Inclination	98.7
Period	102 min
Equat. crossing time (ii)	07.30 and 19.30 (even numbered satellites) 14.00 and 02.00 (odd numbered satellites)
Repeat cycle	12 hours
Global Frequency Coverage	1-2 days

(i) the most usable within the swath of 2700 km is the area within ± 15 degrees. At 15 degrees, the area covered by a pixel is approximately 1.5 km and the repeated coverage for this reduced swath width is about 6 days. (ii) Greenwich standard time (14.30 ascending and 2.30 descending (local time))

an average of 13 links were selected and then an affine transformation was applied to calculate the amount of scaling and rotating required to align the image to map coordinates. Following the transformation the 30 layer stack was formed using the nearest neighbor resampling method.

Meteorological conditions of Hungary at the used image dates

The data used for the characterization of the climatic conditions in Hungary are from the Monthly Weather Report („Időjárási Havijelentés') published monthly by the National Meteorological Service of Hungary.

May, 1992. – The image was taken in the first decade of May, the beginning of the growing season in Hungary. The vegetation in general is about to come into leaf, the leafy crown of the natural forests is not closed yet, pastures and meadows are somewhat greenish, but the fresh vegetative parts do not cover more than 60-70% of the field. Among the most important agricultural crops, corn and sunflower are about to be planted – thus the soil is still bare –, while wheat and barley have an average coverage of 70%.

1992 was an extremely dry and hot year as compared to the average. The May precipitation – as the percentage of the normal value – ranged between 10 and 90% with an average of 37.86%, while the temperature was 0.4 °C milder than the average. This month followed a relatively dry winter and spring period, too, however – due to the early stage of the growing season – the soil water content was still high enough to satisfy the needs of the vegetation.

August, 1992. – Due to the low precipitation and high temperature in July and the record heat and extremely low precipitation in August, this was the month of drought. The average temperature anomaly was +4.7 °C, while the total hours of sunshine reached 120-130% of the average. The average precipitation was 12.7% of the long-term average in August.

In this part of the year, many of the agricultural crops have already been harvested, so there are huge areas with bare or crop residue covered soils. The pastures and meadows – except the ones with high groundwater table – are dry and have large amounts of dead plant tissue on the surface.

September, 1992. – The weather extremity of August had gone in September, and the temperature and precipitation values were close to the long-term averages. Precipitation was 93.1% of the average, while the temperature was 0.3 °C warmer than the average. The main part of the precipitation fell in early September, so it could have had a great effect on the 11-20 September image. A country-wide raining occurred during the period, too.

In this part of the year the proportion of bare soils increases because many sunflower and part of the corn fields have already been or are being harvested, while the autumn crops are just about to sprout or be sown. The natural grass-like vegetation is green again due to the autumn aspect, and the natural forest starts to get brownish

June, 1993. – 1993 was still a warmer and less rainy year than the long-term average, however it was far less extreme than the previous one. Spring and summer were still droughty, however, from September on the weather became rainy and cooler.

The monthly mean temperature was a little bit warmer than the average, and the precipitation was only slightly higher than half of the average. At this time of the year the vegetation cover is at its maximum. All of the agricultural crops are still standing. Signs of drought are not characteristic.

September, 1993. - The September of 1993 was much rainier than the 1992 one, and followed a much less dramatic summer period. The image was taken one decade later, too. The average monthly precipitation was 132.7% of the long-term average, with a maximum of 221% and minimum of 68%. The temporal distribution of precipitation shows three peaks, one at the beginning of the month, one wider peak between 10 and 17 September and one at the end of the month. This steady distribution and relative high amount of rain presents a continuous water supply for the vegetation and soil. The monthly mean temperature was 0.4 °C lower than the average.

"TIM" Database (Soil Monitoring and Information System)

TIM is part of the Hungarian Environmental Monitoring System and was set up in 1995 (VÁRALLYAY et al., 1995). This point-vector database consists of 1236 soil profile descriptions. However, only 1117 data points were in our version, and 1073 were complete and hence used in the statistical analysis. The locations of these points were selected as representative points of the natural landscape units of Hungary, so the database can be considered as a realistic characterization of soil resources of the country. In addition to the detailed soil description data, it contains numerous soil physical and chemical measurement data for monitoring the soil changes in time occurring due to anthropogenic and natural processes. Unfortunately, we had access only to the soil classification categories but no other descriptive data. The classification categories are based on the Hungarian Soil Classification System (HSC). The subtype („Altípus') level was too detailed for this analysis, because on the one hand, it was based on only small differences between the subtypes, and on the other hand, the number of observations in a given class was lower than the number of layers or the dimension of the dataset, which caused the failure of matrix operations for the classification procedure. For these reasons the number of soil classes was reduced to 13. These 13 classes are mainly the major soil types complemented with soil types that have a high spatial extent or importance. The classes are as follows: Stony soils, Blown sand, Humous sandy soils, Rendzinas and Erubases, Brown forest soils, Ramann-type brown forest soils, Sandy forest soils, Chernozem-Brown forest soils, Chernozems, Chernozem-Meadow soils, Salt affected soils, Meadow and Peaty soils, and Alluvial soils. The types and subtypes belonging to the given classes are presented in Table 2. The grouping was based on the natural hierarchy of the HSC. In some cases, when the spatial extent of a certain Soil Type showed significant spatial extent, and the distinction of this Type was relevant, an additional classification category was added. A further subdivision was done in the case of the transitional soil types, and categories such as Chernozem-Meadow, Chernozem-Brown forest soils were added to represent the transitional zones between the Major Soil Types. Meadow soils and Peaty soils were grouped together too, because of their hydromorphic na-

Table 2

Grouping of soil subtypes into the classification classes used in the study

Class name	Soil Subtypes
Stony soils	Solid and unconsolidated rock at or near the surface, except the ones with sand parent material.
Blown sand	All Subtypes within the Blown sand Soil Type group
Humous sandy soils	All Subtypes within the Humous sandy Soil Type group
Rendzinas and Erubases	All Subtypes within the Rendzina and the Erubase Soil Type groups
Brown forest soils	All Subtypes within the Brown forest Soil Major Type group, except the Ramann and the Sandy forest type Brown forest soils, complemented with the Swampy forest soil
Ramann-type brown forest soils	All Subtypes within the Ramann-type brown forest Soil Type group
Sandy forest soils	All Subtypes within the Sandy forest Soil Type group
Chernozem-Brown forest soils	All Subtypes within the Chernozem-Brown forest Soil Type and within the Leached chernozem type
Chernozems	All Subtypes within the Chernozem Major Soil Type group except the Meadow-Chernozem and the ones under the Leached chernozem Soil Type
Chernozem-Meadow soils	All Subtypes within the Chernozem-Meadow and Meadow-Chernozem Soil Types
Salt-affected soils	All Solonchaks and Solonetzes
Meadow and Peaty soils	All Subtypes within the Peaty and Meadow Soil Type group except the Chernozem- Meadow Soil Type
Alluvial soils	All Subtypes within the Alluvial Major Soil Type

Note: The Class name does not refer directly to the Soil Type, it is more like a general description of the soils belonging to the given classification class

ture. The areas covered by Peaty soils are significant in extent, however only a small portion of this area is untouched and due to the intensive melioration – canalization and drainage of the area – the vegetation and the overall image of the area is getting closer to the meadow ecosystem, although some major differences still exist. That is why this unification of the two Major Types – the Meadow and Peaty Major Soil Types – was done.

The statistical analysis

The creation of the dataset that could serve as the input for the statistical analysis was done within the ARC/INFO. The AVHRR stack with the 30 layers in it was overlaid by the TIM point database. From each location where the TIM had valid data a column vector with x and y coordinates, the soil type identifier, and the pixel values of the 30 channels in it was formed, and hence the 1117 observation vectors formulated a 1117 row by 33 column matrix. The observations with missing values were deleted and finally 1073 observations remained in the dataset.

The first step that had to be done in this study was to earn some understanding about the statistical nature of the data, and its capability in discriminating among soils as natural bodies, utilizing its complexity instead of just a certain set of its features. For this reason, a statistical analysis of the data was performed with the use of UNIX-based SAS software (SAS, 1989).

First, a Correlation Coefficient cross-matrix was plotted to test against similarity and dependency between the 30 channels and to gain some idea about the possibilities of the reduction of dimensionality in the data set. For the same reason, the scattergrams – values of a certain channel plotted against the values of an other channel for the given pixel – of all the channel-pairs were generated for visual examination.

The best and most obvious way of measuring the usefulness of AVHRR in soil characterization is to perform a variance analysis on the data. Because of the unbalanced design of the classes, the General Linear Model (GLM) Procedure of the SAS was used instead of the ANOVA Procedure, hence twelve indicator variables for the thirteen classes were introduced and a linear model was set up. The goal was to compare the means of the response variable for various combinations of the classification variables, and to measure and compare the variations in response variables due to classification variables and variation due to random error. The ratio of the two kinds of variations describes the percentage of the classification variable accounting for the variability of the dependent variable. To quantify this ratio, the Coefficient of Multiple Determination (R^2) was calculated as followed:

$$R^2 = SSR/SSTO$$

where: SSR stands for „Regression Sum of Squares" and $SSTO$ is the „Total Sum of Squares".

Results and Discussion

Characterization of the AVHRR bands

The first step within this part of the study was the covariance analysis of the 30 channel data base. Due to the difficult interpretation of the covariance matrix, a correlation (r) cross matrix of the channels was generated. In general, the correlation between two data sets is considered high when the absolute value of r is higher than 0.7-0.8, however the decision boundary is always case dependent. In this study we considered the correlation between two channels significant if the r value was higher than 0.6 and considered it severe when the absolute value was above 0.8. According to this categorization the following were found:

The most severe correlation was found between the two thermal bands 4 and 5. The *r* values were always higher than 0.97 (0.97; 0.98; 0.98; 0.98; 0.99 in ascending time order), regardless of the date of the image. The scattergram for these two bands showed an almost one-to-one relationship, too.

A very high collinearity was also detected between bands 3 and 4 (thermal-infrared and the thermal band). In this case, however, the correlation was a little bit lower, but still around 0.9. The June and August images had relatively smaller values, 0.88 and 0.87 respectively, while the *r* value for the one May and the two September images were still around 0.92.

After the two above-mentioned statements, the high correlation between bands 3 and 5 was not a real surprise, however, the *r* value was lower for this relationship than for the previous ones, regardless of the date. The correlation ranged between 0.8 and 0.91 and was lowest in August 1992 and the highest in September 1993.

A significant correlation was also found between the NDVI and band 1 at all dates, however, these *r* values were not as high as those above, and did not necessarily mean a severe collinearity. The order of magnitude of the correlation was more dependent of the date when the image was made, than in the case of the correlation between the thermal bands. Higher *r* values were found for the August and September 1992 and June 1993 images (-0.8, -0.72 and -0.69, respectively) and relatively lower values for May 1992 and September 1993 (-0.52 and -0.47, respectively). As a first interpretation, this result suggests that the NDVI-band 1 correlation is higher for drier months than for those having higher water supply. However, the question seems to be more complex when we consider the R^2 values later in this study.

The correlation between the near-infrared channels (band 2) and NDVI was always much smaller, generally around 0.3 with the lowest (0.15) in June 1993 and the highest (0.44) in May 1992. The higher correlation between NDVI and the visible-red channels (band 1) suggests the higher dominance of the visible-red channels over the near-infrared channels in determining the NDVI values.

An interesting result was the high correlation between the NDVI and the thermal-infrared (band 3) channels. In August and September 1992 the *r* value was quite high (-0.7 and -0.8) while in the May 1992 and September 1993 images the correlation values were relatively low (-0.32, -0,57 and -0.17). This result is surprising, because none of the bands that were used for generating the NDVI – bands 1 and 2 – showed significant correlation with the thermal-infrared band, however, for the date with high *r* values between the NDVI and band 3, the band 1 and band 3 correlation shows a little bit higher *r* values than the ones with low NDVI-band 3 correlation.

The NDVI and band 4 and 5 correlation was found to be low, with the exception of the September 1992 image where the *r* values were -0.66 and -0.61.

We did not find high correlation between bands taken at different times of the year. The highest correlation value occurring between different date images was the NDVI August and NDVI September 1992 correlation. The *r* value was 0.58.

The first conclusion that can be drawn from these results is that a multitemporal database represents the independent variability of the scene. Even the same bands from different dates do not have significant correlation. The highest „interdate" *r* value was for the NDVI August and September 1992. These two months were extremely dry, and hence the vegetation could not recover from its sleeping stage. The September image was taken just after a short period of raining which caused a slight recover of the vegetation and a higher R^2 value. However, the vegetation changed from August to September, but the time period was still very short to eliminate vegetation spectral similarities between the two dates. As a conclusion, it can be said that significant correlation can only be found between the NDVI images taken close to each other in time, so the vegetation, that needs a longer period to react to the environmental changes, can reveal its characteristics.

The R^2 values

The Coefficient of Multiple Determination was calculated for all bands of the five dates, to quantify the correlation between the AVHRR pixel values (Digital Number, {DN}) and the soil types, and formulate general rules for the image-date and band selection for soil type characterization. The results are shown in Table 3.

From a general statistical point of view these R^2 values are very small and do not represent real correlation. A statistician would conclude no correlation based on these values. However, the results of previous studies carried out by AGBU et al. (1990), COLEMAN et al. (1993) and SKIDMORE et al. (1997) found the same magnitude in the R^2. In those works they studied specific soil charac-

Table 3

The Coefficient of Multiple Determination (R^2) for the soil-type-AVHRR correlation

	May 1992	August 1992	September 1992	June 1993	September 1993
Ch1	0.0282	0.1289	0.1649	0.1561	0.1802
Ch2	0.0470	0.0491	0.0248	0.0297	0.0246
Ch3	0.2919	0.1805	0.2633	0.2163	0.1374
Ch4	0.2646	0.0992	0.2787	0.1683	0.0754
Ch5	0.2386	0.0656	0.2682	0.1514	0.0717
NDVI	0.0543	0.1823	0.2402	0.1691	0.2592

Note: All values are statistically significant at 0.01probability level teristics such as pH, available phosphorus, potassium and calcium, color and chroma value, percentage of sand, silt and clay etc. and its correlation with TM and SPOT bands. Even though the R^2 were around 0.2 – never exceeded the 0.3 value –, based on the classification results and the regression models, they concluded the remotely sensed data useful for predicting soil variables. They also concluded that these low values are partly because of the low spatial resolution of the data source. In the case of the AVHRR data the spatial resolution is 20-30 times lower than those of the SPOT and TM data, so we were expecting even lower correlation values. The results were surprising, because the order of magnitude was the same or even better than those mentioned in these previous studies.

The first surprise was the comparison between the NDVI and thermal images. In May and September 1992 and June 1993, the most widely used NDVI had a very low or at least a lower contribution to the explanation of soil variation than the thermal bands. Especially the thermal-infrared (band 3) resulted in a very high R^2 value. The term „high" should be interpreted in this context only and not in its general meaning! The thermal bands digital number (DN) values refer to the surface temperature, and are mainly determined by the emitted radiation. The reflected component – even in the case of band 3 – makes only 3% contribution to the overall signature. Among the thermal bands, band 3 always had the highest R^2 value (with the exception of September 1992, when all the thermal bands were about the same). KERBER & SCHUTT (1986) explained this phenomena with a higher sensitivity of band 3 over band 4 and 5 to temperature due to the exponential behavior of the Planck function in the 3.55-3.93 μm region and by the higher transmissivity within band 3 relative to the other thermal channels. In the case of band 4 for temperature higher than 3 °C, the emission sensed is on the downside of the black body curve, where the response is approaching linearity.

The same rule works in the case of band 4 and 5. These bands have always high correlation with each other, and hence probably refer to the same natural phenomena. However, band 4 always has a higher R^2 value than band 5. Band 5 is sensitive to the 11.5-12.5 μm region while band 4 has 10.3-11.3 μm wavelengths. As it was concluded above, the longer the sensitive wavelength, the lower the selectivity in the temperature range characteristic is to the Earth environment.

The reason for having high thermal band performance for many of the dates has to be described separately, because it was influenced by different natural and environmental conditions. In the May case the NDVI was necessarily low due to the early stage of vegetation development. The image was taken in late-mid spring, when the temperature was not hot yet and hence, despite of the low precipitation of the winter and spring period, the soil – depending on its physical characteristics – could retain a certain amount of water received earlier. This moisture condition is characteristic of the soil type itself and has a great influence on the surface temperature, that can be sensed with the thermal bands of the AVHRR.

September 1992 also has high thermal band performance, while the NDVI is much higher than the May image. The height of the R^2 for vegetation index is surprising, because the image was taken at the end of a very dry period. This month followed an extremely hot and dry month, when the NDVI had a 0.18 R^2 value. The precipitation in September was also very small, however just a few days before the image taking period, a country-wide rain occurred. The effect of this rain on the vegetation can be detected through the increase of the NDVI R^2 from August to September with an approx. 40 (relative) percent. The highest NDVI was found in the two September cases, probably due to the full maturity of the plants, combined with the effect of the starting process of leaf chromaticism, that is highly species specific. This relatively high NDVI-soil correlation in September occurred in 1993, too. However, the similarity between the two September images is not revealed in case of the thermal bands. The good performance of the thermal bands in September 1992 suggests a great spatial heterogeneity in the surface temperature, while contrary to our expectations the September 1993 image had low performances in bands 3-5. The possible reason of this difference may be found by examining the weather conditions of the two given periods. In 1992, the image was taken after a long drought period and the amount of precipitation was not high enough to saturate all soils for a long period, and only those with higher water retaining character could maintain some of its water income. This limited water supply can intensify the natural differences among soil types, and hence can help to distinguish among them with the use of thermal bands. The September 1993 case is somewhat different. This September followed a moderately dry August and had much more precipitation than the long-term average through the month. When the water supply is continuous and plentiful, and the temperature is lower, too, the soils maintain a wet condition and high water content. Even soils with lower field capacity would express wet conditions and hence have equalized temperature mainly due to the domination of the high thermal capacity of the water over the soil thermal capacity.

The same phenomena occurred in August 1992 and June 1993. However, in these cases the homogeneity in surface temperature was due to the long dry period, and therefore almost all soils were dry and the effect of soil moisture condition – that refers to some of the physical and chemical characteristics of the soil – was eliminated.

Band 2 (Near-Infrared) always showed the lowest correlation with soil types, regardless of the season when the image was taken.

As it was concluded above – based on the *r* values –, the NDVI-Band 1 correlation seems to be higher for the dry months and lower for the wet ones. This idea suggests similarity between the May 92 and the September 93 images, however, after a further examination, it is unlikely to be true, or at least something else has to cause the result. In the May case, both band 1 and the NDVI have low R^2 values, while in the September case these two bands are the best information holders with much higher R^2 values. The correlation definitely exists in both cases, however the situation differs greatly. The May case is at the beginning of the growing season when the NDVI does not represent a „good" data source, meaning no real significant link between the soil type and the spec-tral condition of the vegetation. It is probably due to the lower vegetation cover percentages, and to the characteristics of the immature vegetative parts of the plants, such as the higher water content, softer tissues, and lighter green color, that comformizes the appearance of the plant. The September case represents a matured vegetation which shows much more species dependent spectral features.

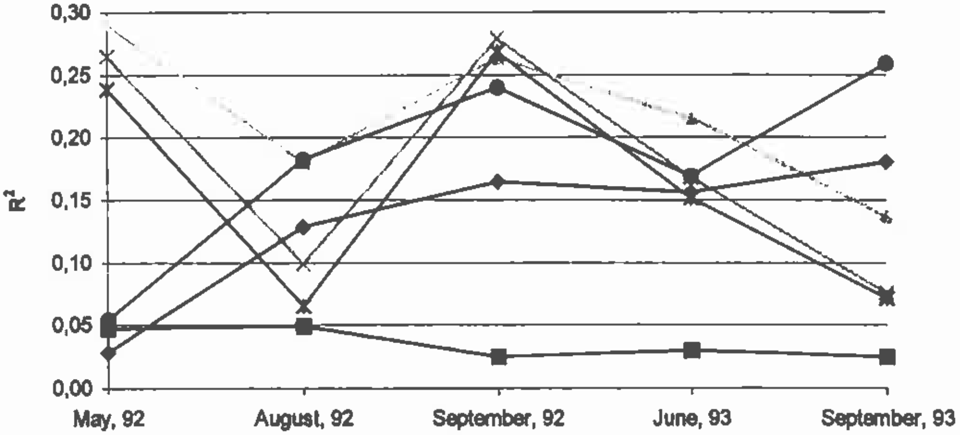

Figure 1

The R^2 values for the soil-types/AVHRR channels correlation

♦: Ch 1; ■: Ch2; Δ: Ch3; x: Ch4; *: Ch5; ●: NDVI

It is worth looking at the temporal changes of the R^2 values (Figure 1).

The figure clearly shows the temporal changes of the correlation. As we concluded before, the thermal bands show a parallel tendency of temporal changing. The R^2 values are high in May and September 1992, when a relatively high amount of water was in the soil, however, in general the water supply was not exceeded or did not even approach the available water holding capacity of the soils. When the weather was dry or too wet, the temperature differences were lower among the soils, and hence the performances of the thermal channels were lower, too. Channel 1 and NDVI show similar shapes too, because of the high correlation between them. These channels usually follow the long-term, seasonal trend of the development of the vegetation, and are not affected directly by short-term environmental effects, such as the temporary weather condition of the given area. It shows a continuous increase through the growing season, reaching its maximum in early fall. The channel 2 curve looks independent from the temporally changing environment and always has a very low R^2 value.

Summary and Conclusions

A statistical analysis and quantitative evaluation of the AVHRR data as a potential data source for small scale soil characterization were undertaken. First, a correlation matrix of the thirty AVHRR channels was created and analyzed. It was concluded, that a severe correlation exists between the thermal channels (band 3, 4 and 5). The band 4 and 5 correlation was always the highest, regardless of the date, while the band 3 and 4 and the band 3 and 5 correlations were somewhat smaller. The NDVI and band 1 correlations were also significant, however the *r* values were much smaller and thus the correlations were not as severe as the ones between the thermal channels. High correlation between bands taken in different times of the year was not found, and thus it can be concluded that a multitemporal database represents so much independent variability of the scene.

The statistical correlation between the AVHRR and Soil Types was also tested and the R^2 values for all the thirty channels were calculated. Despite of the coarse spatial resolution the order of magnitude of the R^2 was the same or even better than the ones that can be found in the literature for TM and SPOT data. The general conclusion was that the thermal bands and the NDVIs have the highest soil type correlation in general. The R^2 values for the NDVIs are more dependent on the date and period of the growing season when the image was made, and less dependent on weather and environmental conditions. It has the lowest value at the beginning of the growing season and reaches its maximum at the end of the growing season, just about the time when leaves start to change their color. The vegetation index can show a relatively high value even under dry conditions, and can be low just after a higher precipitation. The thermal bands performance, however, are much more dependent on the temporal environmental conditions and react much more strongly to a different environment. When the weather was dry or too wet, the temperature differences were lower among the soils, and hence the performance of the thermal channels was lower, too. The Near-infrared band (band 2) was found to have low R^2 and relatively independent of temporal changes.

References

AGBU, P. A., FEHRENBACHER, D. J. & JANSEN, I. J., 1990. Soil property relationship with SPOT satellite digital data in East Central Illinois. Soil Sci. Soc. Am. J. **54.** 807-812.

COLEMAN, T. L., AGBU, P. A. & MONTGOMERY, O. L., 1993. Spectral differentiation of surface soils and soil properties: Is it possible from space platforms? Soil Science. 155. 283-293.

EHRLICH, D., ESTES, J. E. & SINGH, A., 1994. Applications of NOAA-AVHRR 1 km data for environmental monitoring. Int. J. Remote Sensing. 15. 145-161.

EIDENSHINK, J. C. & FAUNDEEN, J. L., 1994. The 1-km AVHRR global land data set: First stages in implementation. Int. J. Remote Sensing. 15. 3443-3462.

IGBP, 1992. The International Geosphere-Biosphere Program: A Study of Global Change. Improved Global Data for Land Applications. Report No. 20. Stockholm, Sweden.

ISRIC, 1993. Global and National Soils and Terrain Digital Databases (SOTER). Procedures Manual. Wageningen, The Netherlands.

KERBER, A. G. & SCHUTT, J. B., 1986. Utility of AVHRR channels 3 and 4 in land-cover mapping. Photogrammetric Engineering and Remote Sensing. 52. 1877-1883.

KIDWELL, K. B., 1990. Global Vegetation Index User's Guide. NOAA/NESDIS. National Climatic Data Center. Washington, D. C.

SAS Institute, 1989. SAS Stat User's Guide. Version 6, 4th edition. SAS Institution, Gary, NC

SKIDMORE, A. K. et al., 1997. Remote sensing of soils in a eucalypt forest environment. Int. J. Remote Sensing. 18. 39-56.

TOWNSHEND, J. R. G. et al., 1994. The 1-km AVHRR data set: Needs of the International Geosphere and Biosphere Program. Int. J. Remote Sensing. 15. 3319-3332

VÁRALLYAY, G. et al., 1995. Soil Conservation Information and Monitoring System. Vol. 1. Methodology. (In Hungarian) Ministry of Agriculture. Budapest.