Grain size and heavy metals estimation from airborne hyperspectral data

E. CASTILLO, V. BAYARRI, R. FERRER
Departamento de Ingeniería Geográfica y Técnicas de Expresión Gráfica
Universidad de Cantabria
Avda. Los Castros s/n, Santander 39005, Cantabria
SPAIN

Abstract: - The Bay of Santander is a very important environmental element, located in the north, of Spain (Cantabria). Human spills to convert sea into land started about 1850 and gradually change the tide prism. This induced the Bay to start filling up and consequently, problems for safety navigation. Whence the navigational channel must be periodically measured and dredged.

Marsh, dunes or estuaries use to be characterized by their high natural values what usually cause them to be subjected to high protection levels that make the activities sought to develop inside. The current work compiles and developed process to estimate the grain size (sand, silt, clay, silty loam, sandy clay loam,...) from hyperspectral data and procedures of calibration in laboratory of the CEDEX with the ASD-FR (Analytical Spectral Devices-“Full Resolution”) spectrometer.

The data were obtained in June 2004 by the Catalonia Cartographic Institute (ICC). The ICC owns an airborne sensor (Compact Airborne Spectrographic Imager-CASI) which let the user setup temporal, radiometric and spatial resolution according to the necessities (36 bands and 4 m. spatial resolution). This makes possible to collect data in the optimal tidal and brightness conditions and so they can be considered as an interesting tool to monitor depths, grain size and ripples.

Results show that accuracy is highly dependent of the radiometric and atmospheric corrections, but it can be used as a practical tool for deriving size grain and inorganic sediments in coastal environments more economical that the classical chemical methods.

Key-Words: - Grain size, sand, silt, clay, airborne hyperspectral data, radiometric corrections, atmospheric corrections, CASI, spectrometer, CEDEX, ICC.

1 Introduction

The most important aspects of the project that appears are that it uses the airborne hyperspectral data as a tool to analyze the hydrodynamics, the quality of the water and the spills in Santander Bay. The results will be compared with the habitually used methods and that are based on chemical and oceanographic procedures.

Historically the rivers and estuaries have been suitable zones for the location of urban and industrial centers, had to a great extent to their potential like routes of communication and transport. Therefore, it is no wonder some of the contaminated regions more of the world are indeed estuaries. Nevertheless, although the estuaries have been considered traditionally like traps of the polluting agents spilled to aquatic means, its capacity for the retention of these substances varies to a great extent according to its characteristic specific hydrodynamics, being necessary to analyze the environmental situation in each case.

During the last decade it has increased considerably to the interest by the study and characterization of the contamination of aquatic sediments as an obligatory step in the correct environmental management of you laugh and estuaries. And therefore every time they are plus the projects of investigation dedicated to this subject.

Thanks to the I+D project "Characterization of the territory by means of numerical cartography, bathymetry and remote sensing [REN 2002-04202-C02-01]" financed by the Ministry of Science and Technology, with a duration of 3 years from November of 2002 to November of 2005, have obtained hyperspectral images for the teledeteción applied in three zones of Cantabria: the Bay of Santander, the Salt Marshes of Santoña and Ria of San Martin (Suances).
Traditionally satelital and airborne remote sensing have been used to determine parameters on the quality of the water (contained organic total, turbidity...). Despite, the appearance of the multispectral and hyperspectral images it opened the possibility of exploring new aspects in coastal zones. In the present work one is going away to use the images obtained in each one of the zones from the airborne sensor (Compact Airborne Spectrographic Imager) pertaining to the Cartographic Institute of Catalonia (ICC).

2 Problem Formulation

The Bay of Santander is located in a privileged place within Cantabrian community (Spain). It is a depression whose ecosystem is very rich both from biological and from the socio-economical point of view. More than 250,000 inhabitants are concentrated in this area, which means more than 50% of the community population. The bay has been conceiving from the Mesozoic and it is very important for a population which develops several activities, all of them creating residually spills.

Towards 1750, the city of Santander and its port start experimenting an extraordinary growing. The small mariner village changed to become a city (this year is the 250 anniversary of this fact), when several circumstances were filled and then a revitalizing process started:

- The appreciation of the port of Santader; it is located in a big bay which offers a spacious and safe natural protection.
- The persistence of Cenón de Somadevilla, Marquis of La Ensenada, minister with a lot of power who tried to strengthen the port and upgrade the exportations as an alternative way to Bilbao’s which did not pay tariffs to the crown.
- The opening of Camino Real de Reinosa, an authentic ancient carriageway which linked the port to the Plateau (1748-1753). It was considered as the first road built in Spain with high capabilities.
- The eagerness of Francisco de Rábago (reverend Rábago), confessor of Fernando VI, king of Spain and personal counsellor, which was born in Tresabuela (Cantabria). He obtained the desired diocese of Santander and propitiated the rise and growing of the city.
- The infrastructures construction which needed big occupational surfaces such as the airport or the new commercial port of Raos.
- The construction of the railway Santader-Bilbao-Oviedo (1900) and the new Bilbao road, which fence still more the spaces of the primitive Bay.
- The urban pressure which demands soil, this means new spaces in the northern area along the coastal front of the Bay (1850-1950)
- The emergence of the railway Alar-Santander (1860) which needed to cross the marsh and acquire a large space (the area for the city entries) to being able to build the railway-station in the center of the city, gaining space to the sea.
- Iron minery (1875-1950) in the group of mountains Peña Cabarga converted the Bay into a deposit for the wastes produced by the 12 mineral laundries placed in the southern area of the Bay.
- The creation of the siderurgy industry in the Oleo Island (Nueva Montaña Society, 1898), which obliged to gain more surface to the Bay.
- The coats of the urbanization industry in the Oleo Island (Nueva Montaña Society, 1898), which obliged to gain more surface to the Bay.

Notwithstanding, nowadays the Bay is the great free space of the city. It is scenery that characterizes the Capital of Cantabria and gives it a great attractive. It is not possible to separate it since it constitutes a part of the urban frame.

Thereby and given that sediments of the Bay act as a trap for all those products that do not remain in the water column at first instance, these can contain a high degree of pollutants waiting to pass to the trophic chain, be returned to the water column or simply be located in more propitious medium for its mobilization.

It is essentially important to investigate the presence of polluting agents in sediments and their effects in the environment, trying to look for parameters that could inform about the quality of the sediment. It has been already proven that sediments...
are and important tool to measure the impact of men on the middle.

2.1 Materials

The imagery used in the study has been taken by CASI sensor, owned by Catalonian Cartographic Institute (ICC). CASI is a pushbroom imaging spectrograph with a two-dimensional CCD array of 512 spatial pixels and 288 spectral pixels scanning the scene in the VNIR (405-950nm). It lets the user to set up the number and width of the bands in which the sensor will record data.

The sensor was installed on board of the plane “Cessna Citation I” belonging to ICC (Fig.1). The flight was made the 4rd of June and to capture the whole bay, 10 tracks were needed. The imagery was atmospherically (6S model), radiometrically and geometrically corrected and then orthorectified. The spectral configuration of the images was configured by the ICC, with 36 channels and a spatial resolution of 4 x 4 m².

The images were recorded in enhanced spectral mode. This mode lets the users obtain a continuous channels configuration between a spectral range from 408 nm to 953 nm and minimum spectral width of up to 1.8 nm. The designed configuration has consisted in 15nm width per channels, and 36 spectral channels. The flight path of the three estuaries of interest are shown in Fig 2.

The Chemical Engineering group of E.T.S.I. Industriales y Telecomunicaciones of University of Cantabria where responsible of sampling the area, measuring the sediment concentrations between 0 to 20 cm. The analysed parameters were: humidity, total organic content (T.O.C.) grainsize distributions, concentration of plumbum, zinc, iron and Manganese oxide.

Accurate correction for the atmospheric, geometric and radiometric effects on the imagery[1]. Lack of corrections insert noise on data registered by CASI in each of the interest parameters. In the following points, the most used classical methods in multispectral remote sensing, as well as some advances that can be very useful when an analysis with the highest precision is made [2],[3].

The analysed methods include different complexity methods such as single bands, spectral rationing and classical and robust principal components analysis.

It is necessary taking into account that a double study has been done, that is to say on one hand the intertidal area, whose spectral response is similar to a soil with different degree of humidity, has been analysed and on the other hand the subtidal area, completely afloat of water. Imagery was taken one hour after low tide.

Fig.2.- Flight path.

The radiometric measures in field (soil and water) for the atmospheric and radiometric corrections have been made with spectrometer ASD-FR (Analytical Spectral Devices-"Full Resolution") property of the Hydrographic Training Center (CEDEX), in the interval between the 400 and the
1000 nm (Fig.3) and in laboratory in the interval between the 400 to 2500 nm with samples of grain size and different moisture percentage (Fig.4).

![Fig.4.- Reflectance in sand samples with different moisture percentage.](image)

Therefore it is had
- multispectral images of the CASI sensor;
- spectral libraries of some of some parameters of interest (measures of the CEDEX) (Fig.5);
- information of four weather stations of the CIMA (Government of Cantabria);
- calibration and validation points taken and analyzed by the group of Chemistry;
- data of sensor CASI of 2003 and 2004 flights.

![Fig.5.-Protocol of measurement of spectrometer ASD-FR in soil.](image)

2.2. Methods:

Single band analysis is the oldest method of the ones that are going to be analyzed and it is based in the existing relation between the analyzed parameter (percentage of sand, silt and clay, Pb, Zn, MnO concentration, etc.) and the obtained spectral response in each of the bands of the CASI sensor. Those bands whose correlations with data are the highest will provide the better final result and on this way the error between the observed and the estimated value is minimized.

The spectral rationing between two bands or combinations of them let decrease the radiometric differences produced when atmospherical corrections are performed. It is very difficult to apply an atmospherical model which took into account all the atmospherical elements that influence imagery. This type of methods increases, at the same time, the spectral differences due to the presence of a determinated parameter or sediment.

Therefore, the objective of this type of methodology is to determine a band or bands in which this parameter of interest has the maximum reflectance and, at the same time, those bands in which the signal received is minimum. Rationing between bands or band combinations evidences the presence of such element letting discriminate the parameter easily.

Classical Principal Components Analysis (Classical PCA) is a very common statistical process in multispectral remote sensing that was proposed by Pearson in 1901. Original data is transformed into a new set of uncorrelated bands (orthogonal principal components) sorted by the amount of variance that any of them explains [8].

Robust Principal components Analysis (Robust PCA) has the objective of “robusting” the classical data obtained in the multivariate analysis detecting the presence of anomalous data. To be able to do this It is necessary to robust the covariance matrix since the most of results are based on it.

3 Problem Solution

The first stage to know to what corresponds the observed spectral response on each of the evaluated methods, consists on studying the correlation between each one of the parameters (Tab.1).

It is observed that the degree of humidity of the sediment is directly related to the grain size (sand, silt and clays) obtaining a really high correlation index.

Isoline planes were generated for each one of the interest parameters aimed to assess in a general way the distribution the parameter in the Bay. The methods used to calculate the surface were 2nd degree polynomial, Inverse Distance Weighting and ordinary kriging.

The work has consisted in two phases; the first one deals with calibrating the different methods with the on-site samples and the second one has
consisted in validating the results.

<table>
<thead>
<tr>
<th></th>
<th>humedad</th>
<th>gravas</th>
<th>arenas</th>
<th>limos/arcillas</th>
</tr>
</thead>
<tbody>
<tr>
<td>humedad</td>
<td>1.00</td>
<td>-0.099</td>
<td>-0.887</td>
<td>0.887</td>
</tr>
<tr>
<td>gravas</td>
<td>-0.099</td>
<td>1.000</td>
<td>0.068</td>
<td>-0.094</td>
</tr>
<tr>
<td>arenas</td>
<td>-0.887</td>
<td>0.068</td>
<td>1.000</td>
<td>-1.000</td>
</tr>
<tr>
<td>limos/arcillas</td>
<td>0.887</td>
<td>-0.094</td>
<td>-1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>COT</td>
<td>0.587</td>
<td>0.065</td>
<td>-0.582</td>
<td>0.579</td>
</tr>
<tr>
<td>pH</td>
<td>-0.268</td>
<td>-0.124</td>
<td>0.405</td>
<td>-0.401</td>
</tr>
<tr>
<td>LOI</td>
<td>0.746</td>
<td>0.100</td>
<td>-0.732</td>
<td>0.728</td>
</tr>
</tbody>
</table>

Tab.1.-Correlation between parameters

The calibration phase (Tab.2) has been developed in the SPLUS software and different classical and robust algorithms methods (linear regression model, Least Median of Squares Robust Regression, Least Trimmed Squares Robust Regression, M-Estimates of Regression, Linear Least-Squares, Minimum Absolute Residual Regression, High Breakdown and High Efficiency Robust Regression) have been applied obtaining adjust functions as it can be appreciated.

Both proceedings have been applied to CASI imagery by using the Program “R: A Language and Environment for Statistical Computing” v. 2.0.0 and a robust statistics module implemented by Professor Alfonso García Pérez (UNED). All the routines needed to convert imagery into matrices and undo such process. This procedure has been implemented in IDL 6.0. The minimum volume ellipsoid estimator has been used as a covariance matrix. This estimator reaches a breaking point close to 0.5, since it has less efficacy than the one of least determinant of the covariance matrix but it has higher breaking point.

<table>
<thead>
<tr>
<th></th>
<th>PCA1</th>
<th>PCA2</th>
<th>PCA3</th>
<th>PCA4</th>
<th>PCA5</th>
</tr>
</thead>
<tbody>
<tr>
<td>arenas</td>
<td>0.79084</td>
<td>0.32038</td>
<td>0.41939</td>
<td>0.21818</td>
<td>0.13606</td>
</tr>
<tr>
<td>limos</td>
<td>-0.74182</td>
<td>-0.46587</td>
<td>-0.36996</td>
<td>0.09187</td>
<td>-0.09963</td>
</tr>
<tr>
<td>arcillas</td>
<td>-0.00286</td>
<td>0.23461</td>
<td>-0.03514</td>
<td>-0.42364</td>
<td>-0.04029</td>
</tr>
</tbody>
</table>

Tab.2.-Correlation between grain size of the calibration samples and the classical principal component.

Last the validation phase (Tab.3) for the results has been made by using a new validation data set so new chemical analysis of each of the parameters are needed.

<table>
<thead>
<tr>
<th></th>
<th>PCAr1</th>
<th>PCAr2</th>
<th>PCAr3</th>
<th>PCAr4</th>
<th>PCAr5</th>
</tr>
</thead>
<tbody>
<tr>
<td>arenas</td>
<td>0.89542</td>
<td>0.43279</td>
<td>0.44936</td>
<td>0.20325</td>
<td>0.14536</td>
</tr>
<tr>
<td>limos</td>
<td>-0.85643</td>
<td>-0.54632</td>
<td>-0.39284</td>
<td>0.09187</td>
<td>-0.12356</td>
</tr>
<tr>
<td>arcillas</td>
<td>-0.03445</td>
<td>0.29248</td>
<td>-0.08322</td>
<td>-0.51274</td>
<td>-0.05236</td>
</tr>
</tbody>
</table>

Tab.3.-Correlation between grain size of the calibration samples and the classical principal component.

The spectral responses on each one of the previous cases in the way that different maximums are observed. In later analysis phases it has been checked that they correspond to the signal corresponding to each one of the interest parameters.

Both proceedings have been applied to CASI imagery by using the Program “R: A Language and Environment for Statistical Computing” v. 2.0.0 and a robust statistics module implemented by Professor Alfonso García Pérez (UNED). All the routines needed to convert imagery into matrices and undo such process. This procedure has been implemented in IDL 6.0. The minimum volume ellipsoid estimator has been used as a covariance matrix. This estimator reaches a breaking point close to 0.5, since it has less efficacy than the one of least determinant of the covariance matrix but it has higher breaking point.

2.0.0 and a robust statistics module implemented by Professor Alfonso García Pérez (UNED). All the routines needed to convert imagery into matrices and undo such process. This procedure has been implemented in IDL 6.0. The minimum volume ellipsoid estimator has been used as a covariance matrix. This estimator reaches a breaking point close to 0.5, since it has less efficacy than the one of least determinant of the covariance matrix but it has higher breaking point.

Proceedings of the 5th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Madrid, Spain, February 15-17, 2006 (pp447-452)
4 Conclusion
Robust techniques have performed better results both in the calibration and validation phases than the classical methods.
To develop this types of studies, a coordinated work of all the implied parts is needed, since field spectroradiometer and the chemical group must be sampling the area while ICC is flying to obtain imagery.
Obtained results conclude that spectral analysis of imagery ease by this sensor is an effective and viable alternative that cheapen the sampling, analysis and interpolation processes; being able to extend the results to the whole study area providing at the same time more precise results.

5 Future work
In order to carry out a more complete analysis of grain size and heavy metals in the three estuaries of interest mentioned with new contributions of data provided by the sensor AHS-160 [ Airbone Hyperspectral Scanner ] property of the INTA.
The images of airborne sensor AHS-160, with inclusion in the thermal infrared, that also will be programmed, allow to work with 80 bands. East sensor was not used from the beginning of this work because it was acquired by the INTA in April of 2003 and was necessary to come to his calibrated during a period of about 6 months to be able to offer to the user a quality product. It was not possible to wait for all that time to have the work material because the project of the Ministry was within dates and was much work to make.
New methods based on the spectral derived ones will be use with the purpose of obtaining more precise results for the heavy metal detection.

References: