ISTAT DATA UTILIZATION TO ENHANCE LANDSAT 8 IMAGES CLASSIFICATION PROCESS

by Stefano Mugnoli and Raffaella Chiocchini

THIS PAPER SHOWS AN EXPERIMENTAL STUDY BASED ON A LANDSAT 8 IMAGE THAT COVER COMPLETELY 5 PROVINCES IN THE NORTH OF ITALY, WHERE IT'S SHOWN THAT ISTAT DATA, DEM AND COMBINE OF NDVI AND NDBI INDICES CAN IMPROVE THE RESULTS OF THE SATELLITE IMAGE CLASSIFICATION PROCESS.

Fig. 1 - LANDSAT 8 image clipped on ISTAT administrative boundaries (R,G,B 4,3,2).

Since the end of the Second World War Italian landscape has suffered the consequences derived from many occurrences that have played an important role in improving the demand of housing estates, factories, roads and highways, to name but a few.
The increase in the use of cars has influenced urban sprawl in terms of amplification of residence areas option that are now further away from working and commercial ones (*European Environment Agency, Urban Sprawl in Europe - the ignored challenge, EEA report n. 10/2006* [1]).
So, not only the residential role of the town has been walked away from urban centers but productive and commercial too. This fact has created a gradual expulsion process from centers to suburbs, leaving urban unoccupied places and making difficult a really rational urban planning.
Thus, it's very important assess how much the 'worn out soil' in Italy is, and this need comes from many factors, some of these are described below:

- landscape and ecosystem fragmentation with negative impact on vegetation, wildlife and that can produce critical hydrogeological factors;
- damage in a socio-cultural sense, since landscape is a sort of human perceptual and cultural identity;
- services and urbanization cost increase. In fact the costs of a new 'not planned' urban areas are highest in relation to planned ones.

As far as 'urban sprawl' is concerned Italian National Institute of Statistics (ISTAT) has an important institutional role overall in statistics assessment of the phenomenon.
As ISTAT is also considered in 'Welfare Assessment Scientific Commission' (BES Project) for the themes 'Landscape and Cultural heritage' and 'Environment', the most urgent task is to have further geographic data to assess 'urban sprawl' has arisen.
This paper describes an experimental study where Statistical data and satellite ones are integrated to produce a geographic dataset that can help to estimate urban sprawl at medium scale.

USED DATA AND PRELIMINARY ELABORATIONS

The experimental study has been carried out on an area in the North of Italy. This area covers 5 Provinces in the Po valley. In order to cover the entire zone by a satellite image, a full LANDSAT 8 (LANDSAT Images can be downloaded from http://earthexplorer.usgs.gov) scene was chosen (scene LC81920292013183LGN00). The file name better explain that the scene was taken in the July 2nd 2013.
Beginning February 3, 2014, all Landsat 8 data held in the USGS archives will be reprocessed. All Landsat 8 scenes will be removed from the online cache at this time and the data will then be reprocessed starting with the most recent acquisitions and proceeding back to the beginning of the mission. Data will then become available for download. Scenes waiting to be reprocessed will also be available for on-demand product orders. Reprocessing is expected to take approximately 50 days. Many corrections will be made to the data, affecting both the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). These corrections include all calibration parameter file updates since launch; improved OLI reflectance conversion coefficients for the cirrus band; improved OLI radiance conversion coefficients for all bands; refined OLI detector linearization to decrease striping; a radiometric offset correction for both TIRS bands; and a slight improvement to the geolocation of the TIRS data (details about these changes are available on http://landsat.usgs.gov/calibration_notices.php).
However, the aim of the project is to extend the study to the whole Italian territory in order to evaluate, at medium scale, anthropic land cover and use overall in extra-urban areas.
Satellite image was clipped using ISTAT administrative boundaries and so was extracted just the area that covers completely five provinces: Modena, Bologna, Ferrara, Ravenna, Forlì-Cesena. In fig.1 satellite image in combination R,G,B, 4,3,2 of the area in exam.
Other data used have been ISTAT enumeration areas, Digital Elevation Model (20 m) elaborated by ISPRA (Superior Institute for environment Protection and Research) and ISTAT census 2010 population data.
First of all LANDSAT image was classified by an unsupervised classification ISODATA algorithm; this in order to understand how many land cover classes could have been extracted from the image. The ISODATA clustering method uses the minimum spectral distan-

Classes	N. pixels	Interpretation
Class 1	1371212	Very wet vegetated areas; broad-leaved
Class 2	1169507	Coniferous
Class 3	1280340	Vegetation in plan areas
Class 4	803152	Rice fields; Coniferous
Class 5	371962	Coniferous
Class 7	302030	Water
Class 10	970482	Infrastructures; services
Class 11	1291884	Bare soil; not vegetated areas
Class 12	1515620	Urban areas; heterogeneous rural areas
Class 13	1294236	Dry bare soil
Class 14	369239	Dunes; beach; clouds

Tab. 1 - Unsupervised classification: summary report.

ce formula to form clusters. It begins with either arbitrary cluster means or means of an existing signature set, and each time the clustering repeats, the means of these clusters are shifted. The new cluster means are used for the next iteration. The ISODATA utility repeats the clustering of the image until either a maximum number of iteration has been performed, or a maximum percentage of unchanged pixels has been reached between two iterations.

After studying the clusters distance based on 'transformed divergence' method, eight land cover classes were identified. Table 1 shows the results of unsupervised classification.

The other classes are not significant. However, image classification is not the focal point of the study, because it will be corrected by integration with ISTAT and DEM data. Thus, supervised classification was realized considering just eight classes (Road networks, Coniferous, Broad-leaved, Water, Continuous urban fabric, Discontinuous urban fabric, Infrastructures and services, Bare soil). In order to rationalize further steps, it was decided to consider urban areas and extra-urban ones in two completely different ways. In fact, urban areas limits derived principally from ISTAT data, while their classification both ISTAT and satellite classification.

So, first of all, the classified image was clipped using ISTAT urban areas layer.

In figure 2 the classified image for Bologna municipality area.

NDVI AND NDBI INDICES [2]

In order to better specify land cover, overall in urban areas, some indices were calculated. So NDVI (Normalized Difference Vegetation Index) and NDBI (Normalized Difference Built-up Index) has been chosen because these can clarify green urban areas and urban fabric overall if these are utilized in combination.

In fact, NDBI, that was produced using the following equation (Zha *et al*. 2003 [3])

$$NDBI = \frac{MIR-NIR}{MIR-NIR}$$

can't explain by itself the built-up areas (and so the urban ones) because some studies found that many vegetated areas have positive NDBI values (i.e Cibula et al., 1992 [4]; Gao, 1996 [5]).

On the other hand NDVI, produced using canonical equation

$$NDVI = \frac{NIR-RED}{NIR+RED}$$

can generate positive values because some kind of artificial surfaces have a major reflectance in NIR respect in RED. So what it has been said above allows to understand that these two indices combination give us better results. The two indices has been calculated using ERDAS/IMAGINE sw [6].

So, after having reclassify NDBI and NDVI images, taking 0 value as break point for both, a combine image was produced using ARCGIS 10.1 combine algorithm [7]. The result of this elaboration is shown in figure 3.

For urban areas, just 311 pixel on 1166535 are not significant (0,026%).

Other indices can be calculated to extract vegetation or other land cover typologies {i.e. NDWI or MNDWI (McFeeters,1996 [8]) for water and SAVI for vegetation (Huete, 1988 [9])}, but in this study it was decided to calculate just NDVI and NDBI since ISTAT data can be better define the results derived from image processing.

ISTAT [10] DATA INTEGRATION

ISTAT enumeration parcels data (ISTAT Enumeration parcels can be downloaded from http://www.istat.it/it/archivio/104317) have been updated during Census 2010 collecting data; in urban areas these parcels are very detailed and they have been drawn not only in relation to population and social parameters but land cover ones too. So these can be used, in combination to other elaborations to describe land cover/use entities at medium scale. ISTAT data was first rasterized because they are in vector format.

Thus, they have been combined to image processing results in order to enhance thematic resolution of the elaborations described above. So, combining ISTAT population data linked to enumeration parcels attribute it's possible to codify some areas according to their land cover/use 'vocation'.

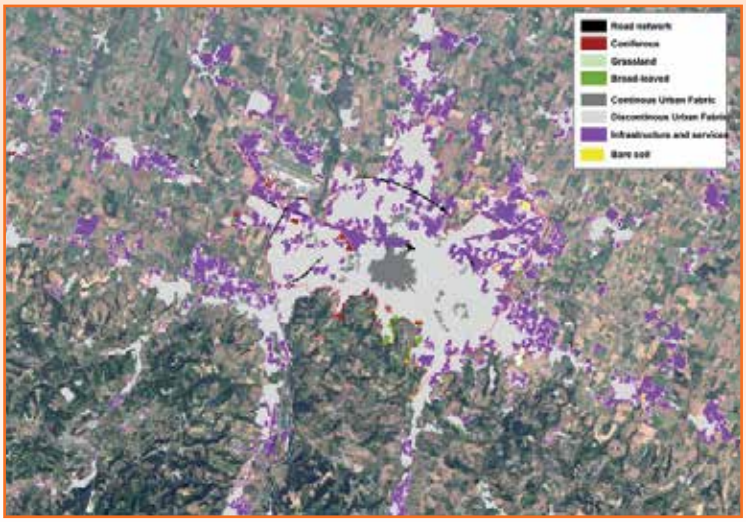

Fig. 2 - Supervised classification (Bologna municipality).

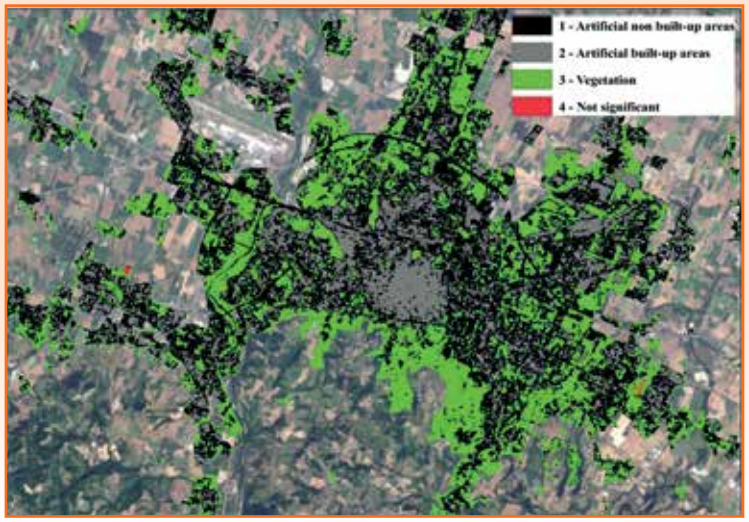

Fig. 3 - Combine NDBI & NDVI image (Bologna municipality).

Fig. 4 - Example of codified enumeration area (Cemetery).

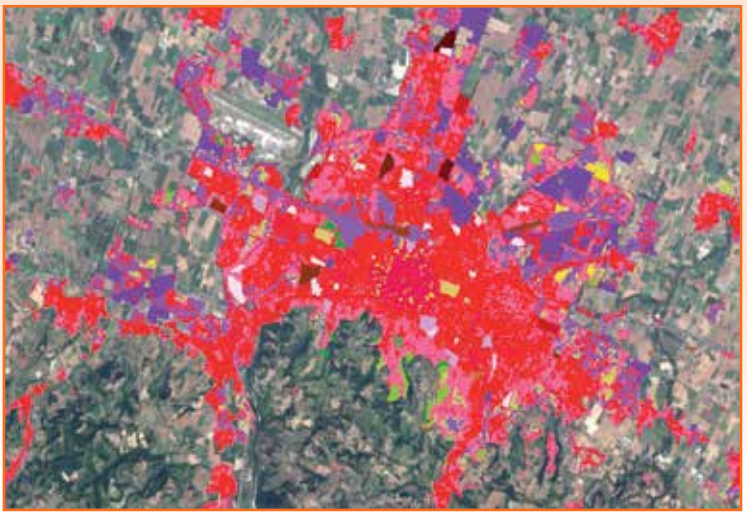

Fig. 5 - Synthesis of the data integration (Bologna municipality): See the text for details.

Not all enumeration parcels can be utilized for this porpoise, so it has been decided to extract just these parcels that have codes described in table 2 that have residential population less than 1 resident per hectare. In figure 4 an example of an enumeration area classified according to its land cover/use.

ISTAT Code	N. pixels	Legend
5	4694	Green urban areas
12	95969	Industrial or commercial units
9	1775	Hospitals
16	142	Sport and leisure facilities
2	203	Religious institutes
3	58	Monuments
6	1025	Port areas
7	824	Airports
11	824	Military installations
18	208	Research institutes
24	204	Prisons
50	235	Water treatment plants

Tab. 2 - ISTAT enumeration aparcels code chosen to combine step.

How we can see, many classes are represented by a little number of pixels and these are surely eliminated if we consider the spatial resolution of a cartography product derived from LANDSAT images (scale 1:50.000); but there is to say that all of these pixels come from single enumerations areas and so they are grouped in clusters absolutely recognizable on the classified image. But it's better to remember, that the product of this experimentation isn't it a cartography one.
So combining NDVI/NDBI image and LANDSAT classification with ISTAT enumeration parcels in raster format as describe above, we have now a new image that represent a sort of synthesis of all of these layer. In fig 5 the synthesis image (Bologna municipality). Some land cover/use indications:

- Red tones: residential areas;
- Purple: Infrastructures and services;
- Green: Green Urban areas;
- White: Sport and leisure facilities;
- Dark yellow: Hospitals

Fig. 6 - Water extraction and its land cover/use classification.

EXTRA-URBAN AREAS INTEGRATION

At this point of the study the problem is to relate and to define the extra-urban classification. ISTAT enumeration areas, in fact, aren't very suitable to describe land cover/use in extra-urban zones as they are for urban ones. Just few of them can be taken for this scope. For example, those that have a special area code (i.e. River, lakes, lagoons, etc.) since they are drawn exactly to the limit of these territorial features.
Another problem is linked to agricultural areas; heterogeneity of these areas in terms of reflectance make the extraction of agricultural areas quite impossible, just considering LANDSAT image classification results. So another approach is needed.
For this porpoise DEM and slope layer have been used. So, a sort of a 'step by step' extraction has been done.
First of all water has been isolated using NDVI, ISTAT layer and LANDSAT classification. In fact, water can be easily extracted from LANDAST image (especially from band 2 and infrared ones); furthermore water has ever negative NDVI values too. So this land cover class has been extracted using LANDSAT image and NDVI index. But, in order to better specify 'environmental' and use classification of the body of water, ISTAT attributes were used; in fact in ISTAT enumeration parcels attribute table there is a code that identify the most important basins and rivers by an unique national code; using this code, water land cover/use typology can be specify. An example of what is has just been said is shown on fig.6.

DEM AND SLOPE

The area in exam is not very nervous in morphological sense. In fact, considering Digital Elevation Model (DEM) realized by ISPRA (Superior Institute for environment Protection and Research) with a 20m*20m cell, it's evident that over 50% of the DEM cell have a value less than 50 m. Moreover, considering the slope value of a DEM single cell we can see that over the 55% of the cell are completely flat.

In brief, for each cell, the slope value is calculated as the maximum rate of change in value from that cell to its neighbors. Basically, the maximum change in elevation over the distance between the cell and its eight neighbors identifies the steepest downhill descent from the cell. The lower the slope value, the flatter the terrain; the higher the slope value, the steeper the terrain. Below a scheme that explain the concept.
In order to prevent critical hydrogeological factors regional technical services propose ordinary piece of legislation to plan agricultural activities. Overall, their attention is focused on areas that have a degree of slope as high as, or higher than, 30%. In this study this value has been used to discriminate agro-forestry areas.

Dem has been used to characterized flat areas and areas over 1700 m of altitude; in fact this value represent the maximum limit of woods for the area in exam. Thus using DEM and 'slope' layer it' possible to classify extra-urban areas in six classes: two agricultural classes (completely plain and hilly); an 'agro-forestry' class; a broad-leaved class and finally a 'sparsely vegetated areas'.

RESULTS

In figure 7 is shown a map that represent the final result of the activities briefly describe above. Furthermore, in table 3 a summary report with the extension of each classes is shown. The supervised algorithm was able to distinct coniferous woods in coastal zone, but unfortunately there are some kind of agricultural crops (especially very wet ones), that have a reflectance very similar to coniferous woods. So , in table 4 coniferous woods class is in red to highlight the fact that coniferous class data embody some pixels that belong to some specific crops.
In annotation column just some pointers to specify the way in which some classes have been extracted.

About results it might be said that they have quite good quality in urban areas, while in extra-urban one agro-forest class is overestimated at the expense of 'forest and semi-natural' one. This is an example of the fact that ISTAT data are fundamental to calculate 'urban sprawl' in rural areas.
In table 3 it' shown a comparison among 'ISTAT/LANDSAT' layer with CORINE Land cover 2006 (CLC 2006) and Land Use Map 2008 realized by Emilia-Romagna regional technical services (LC ER), that represent the reality. All the class codes have been related to first CORINE LC level. Area is expressed in hectares.
As it has been said before, the resulting map is not a really cartography product but it can be a good example of the fact that ISTAT data can be integrate with other base map that are often used to produce cartography. In this regard, ISTAT has instituted an activity in order to produce an integrated layer called Statistical Synthetic Map, with the aim to put into a cartography product not only 'statistical data' but administrative, cover and land use too; about that at the end of 2014 six Italian regions will be completely realized.

CONCLUSION AND FUTURE DEVELOPMENT

Nowadays, the activities of the study are in progress, and these are not utterly. It is necessary to individuate both some statistic strategies to meliorate classification especially in extra-urban areas, and a methodology to validate and to estimate the accuracy of this approach both thematically and globally. For the first point ISTAT residence population data can be used to better specify anthropic land use in extra-urban areas.
The best way to test this study is to individuate a procedure of accuracy assessment using the land cover and land use cartography like the Regional Land cover and land use map of Emilia Romagna Region. The reason for this is because this map is one of the best and complete that we have and because it use a legend quite similar at the CORINE Land Cover Legend. This map is produced at 1:10.00 scale and it is easily available in shape file format at the Technical Regional Office responsible.
The main activity, at the beginning, is to harmonize the legend of the two product, because they have been realized with completely different aims. All LANDSAT image derived products not only have a return scale of 1:50.000 but the legend linked to them derived especially from the layer used to catch spectral signature of the classes. In contrast, the Emilia Romagna CUS map has a return scale of 1:10.000 with a very good accuracy.
If the results of the further elaboration will be acceptable, it is possible to consider to enlarge the portion of territory to investigate using more different territorial characteristics that can better help us to individuate an approach that can be extended to whole Italian territory. At a later time it is possible to plan an integration with other geographic datasets in order to upgrade the legend details and to complete and to explain some shortcomings in the classification system adopted.

Code I liv.	ISTAT/LANDSAT	CLC 2006	LC ER
1	117534,7	67969,4	126460,8
2	1059996,8	957112,8	828397,3
3	127981,8	270308,4	313306,2
4	9746,1	6854,8	24889,5
5	11993,1	22825,0	29763,0
TOT	1327252,5	1325070,4	1322816,8

Tab. 3 - Comparison among ISTAT/LANDSAT map, CORINE Land Cover 2006 and Regional Land Cover Map. Area is in hectares.

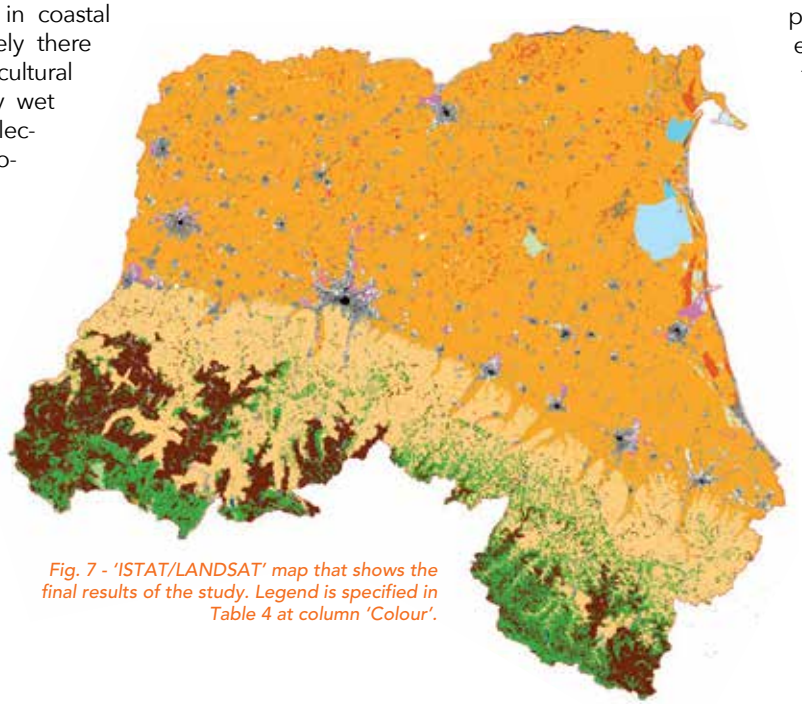

Fig. 7 - 'ISTAT/LANDSAT' map that shows the final results of the study. Legend is specified in Table 4 at column 'Colour'.

This experimental study represent a proposal to integrate geographic datasets that are usually used in other application fields and for very different purpose. One of the target of this project is to suggest how is a possible way to best integrate ISTAT geographic dataset that rarely are used as land/cover data.
There are a lot of benefits derived from the use of LANDSAT images; first of all they are free and then they cover, for the areas of overlapping scenes, large territorial portions every eight days. To use ISTAT data gives a great advantage overall in evaluation of 'urban sprawl' phenomenon; in fact from ISTAT enumeration parcels layer can be extracted little urban agglomerations that are cancelled in medium scale cartography, and this fact generate an error in quantification of land cover/use classes.

COD	Legend	N. pixel	Area (Ha)	Annotation	Colour
1	Continuous urban fabric	33369	1334,7		Black
2	Discontinuous urban fabric	1212909	48517,1		Grey 50%
3	Discontinuous urban fabric (large green areas)	713194	28528,2		Grey 30%
4	Road networks	13817	552,6		Grey 10%
5	Infrastructures and services	553776	22151,4		White
6	Industrial or commercial units	236562	9462,6		Fuchia
7	Hospitals	33369	1334,8		Lemon
8	Sports facilities	8628	345,1		Cipria
9	Monuments	123	4,9		Beige
10	Port areas	3180	127,2		Lilac
11	Airports	14324	572,9		Coral
12	Barracks	2850	114		Dark violet
13	Railways stations	8628	345,1		Amethyst
14	Cemeteries	3220	128,8		Cabernet
15	Universities and research institutes	470	18,8		Lichen green
16	Prisons	463	18,5		Red
17	Green urban areas (Parks)	87501	3500,1		Dark green
18	Green urban areas (Tree)	11111	444,4		Olive green
19	Agricoltural 1	16613135	664536,2	$0 < Alt < 100m;\ \vartheta=0°$	Orange
20	Agricoltural 2	6971422	278861,4	$101m < Alt < 600m;\ 0° < \vartheta = 1°$	Mango
21	Lagoon	274808	10992,5		Very light blue
22	Farm fish	63434	2537,4		Light blue
23	Ponds	33367	1334,7		Light green
24	Very shallow waters	146848	5874,0		Blue white
25	Rivers	12269	490,7		Blue
26	Lakes	12746	509,8		Dark blue
27	Agro-forestry areas	2914930	116599,1	Over 600m; $1° < \vartheta < 15°$	Brown
28	Broad leaved forest	2656377	106256,8	Over 600m; $\vartheta > 15°$	Green
29	Coniferous forest	503307	20132,61	NDVI >0,3 AND NDVI <0,5	Flame red
30	Sparsely vegetated areas	39810	1592,43	Alt >1700 m	Sage green
31	Water treatment plant	826	33,0		Light pink

Tab. 4 - Summary report with the extension of each class.

BIBLIOGRAFIA

1. European Environment Agency (EEA): Urban Sprawl in Europe - the ignored challenge, EEA report n. 10/2006. ISBN 92-9167-887-2 (2006)
2. Hanqiu Xu, 2007. Extraction of built-up land features from LANDSAT imagery using a thematic oriented index combination technique. Photogrammetric Engineering & Remote Sensing, Vol. 73, No. 12, December 2007, pp. 1381–1391
3. Zha, Y., J. Gao, and S. Ni, 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery, International Journal of Remote Sensing, 24(3):583–594
4. Cibula, W.G., E.F. Zetka, and D.L. Rickman, 1992. Response of thematic bands to plant water stress, International Journal of Remote Sensing, 13(10):1869–1880
5. Gao, B.C., 1996. NDWI–A normalized difference water index for remote sensing of vegetation liquid water from space, Remote Sensing of Environment, 58(3):257–266
6. Intergraph ERDAS Imagine® 2013 - Version 13.00.0002.00598 ©1991-2013 Intergraph Corporation
7. ESRI® ArcMap TM 10.1 – ArcGIS 10.1 SP1 for Desktop, License type: Advanced
8. McFeeters, S.K., 1996. The use of normalized difference water index (NDWI) in the delineation of open water features, International Journal of Remote Sensing, 17(7):1425–1432
9. Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI), Remote Sensing of Environment, 25(3):295–309
10. ISTAT (Italian National Institute of Statistics) – DIQR – DCIQ – INT/C

ABSTRACT

ISTAT geographic data, updated to realize census 2010 project in October 2011, represents an useful resource to improve the results derived from Land cover/use cartography or satellite image processing. In fact, both ISTAT vector data and other cartography data (i.e. satellite image classification) can be integrated to realize a product that can help to better understand land cover data especially in urban environment (i. e. urban sprawl), although it can't be considered a cartography product in a strict sense. This paper summarizes an experimental study based on a LANDSAT 8 image that cover completely 5 provinces in the north of Italy, where it's shown that ISTAT data, DEM and combine of NDVI and NDBI indices can improve the results of the satellite image classification process, especially in urban areas. Used SW: ARCGIS 10.1 for desktop (ArcInfo license) and ERDAS Imagine.

PAROLE CHIAVE

LANDSAT 8; ISTAT; NDVI; NDBI; ARCGIS 10.1; ERDAS Imagine; Supervised classification; Urban sprawl

AUTORE

Stefano Mugnoli
mugnoli@istat.it

Raffaella Chiocchini
rachiocc@istat.it

National Institute of Statistics (ISTAT),
Rome, Italy