Ecological Informatics 26 (2015) 151–161

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Evaluation of urban sprawl from space using open source technologies

Gabriele Nolè [a,b], Beniamino Murgante [a,*], Giuseppe Calamita [b], Antonio Lanorte [b], Rosa Lasaponara [b]

[a] School of Engineering, University of Basilicata, Viale dell'Ateneo Lucano 10, Potenza 85100, Italy
[b] IMAA-CNR C.da Santa Loja, zona Industriale, Tito Scalo, Potenza 85050, Italy

ARTICLE INFO

Article history:
Received 3 January 2014
Received in revised form 20 April 2014
Accepted 6 May 2014
Available online 15 May 2014

Keywords:
Urban sprawl
Planning
Sustainability
Satellite data
SVM
Open source software

ABSTRACT

Up to nowadays, satellite data have become increasingly available, thus offering a low cost or even free of charge unique tool, with a great potential for quantitative assessment of urban expansion and urban sprawl, as well as for monitoring of land use changes and soil consumption. This growing observational capacity has also highlighted the need for research efforts aimed at exploring the potential offered by data processing methods and algorithms, in order to exploit as much as possible this invaluable space-based data source.

The work herein presented concerns an application study on the process of urban sprawl conducted with the use of satellite ASTER data. The selected test site is highly significant, being it a coastal zone (with the presence of sand and rocks) characterized by a fragmented ecosystem and small towns, with an increasing rate of urbanization and soil consumption. In order to produce synthetic maps of urban areas, ASTER images were classified using two automatic classifiers, Maximum Likelihood (MLC) and Support Vector Machines (SVMs) applied by changing setting parameters, with the aim to compare their respective performances in terms of robustness, speed and accuracy. All process steps have been developed integrating Geographical Information System and Remote Sensing, and adopting free and open source software.

Results pointed out that the SVM classifier with RBF kernel was generally the best choice (with accuracy higher than 90%) among all the configurations compared, and the use of multiple bands globally improves classification. One of the critical elements found in this case study is given by the presence of sand and sand mixed with rocks. The use of different configurations for the SVMs, i.e. different kernels and values of the setting parameters, allowed us to calibrate the classifier also to cope with a specific need, as in our case, to achieve a reliable discrimination of sand from urban area.

1. Introduction

In last centuries urbanization processes were strongly connected to industrial development, with related population growth; today, despite a profound revolution in labour market, this trend is still in place and can be more associated to the greater opportunities that cities offer compared to rural areas (Glaeser, 2011). Urban population will double within next twenty years and this phenomenon will lead the construction of new buildings and infrastructures. In many cases, urban growth is not due to a real need, but also to private interests, forms of rent, types of financing for local authorities and wrong lifestyles. The risk is that the application of urbanization policies without rules in the countryside, produces waterproofing of territory, wears out soils and distorts landscapes.

The confusing outward proliferation of buildings of a city to rural land is a phenomenon identified as “urban sprawl”. Outside the town centre, territory meshes become increasingly wider and less regular, entailing several inefficiencies, such as increase in travel time to move from home to work (and vice versa) and car-dependence, due to the growth of municipal extra mobility and, thus, increased road congestion and its related issues (Camagni et al., 2002). At the same time, costs of urbanization grow, making increasingly difficult to sustain investment for public transport sector, construction and maintenance of road infrastructure and public services (public lighting, garbage collection, etc.). Other criticisms linked to urban sprawl are aesthetic pollution, noise and air pollution, environmental impact, and soil consumption. Urban sprawl generates a highly fragmented agricultural and natural landscape (Murgante et al., 2011). This phenomenon produces natural islands, defined by the boundaries of new urbanized areas, which are too small to accommodate the life of certain animal species; consequently soil consumption has also a negative impact on biodiversity (Romano and Zullo, 2013). Soil is also linked to agricultural and zootechnical production; consequently, human nutrition depends on it. From a sustainability point of view, soil has a strong relationship with cycles of water and carbon. Urbanization process produces soil sealing. The soil loses its biological value, becoming unable to absorb and filter rainwater.

* Corresponding author. Tel.: +39 3204238518; fax: +39 0971205125.
E-mail address: beniamino.murgante@unibas.it (B. Murgante).

http://dx.doi.org/10.1016/j.ecoinf.2014.05.005
1574-9541/

This phenomenon is particularly intensive in Italy (Clementi et al, 1996; Indovina, 1990; Romano and Zullo, 2012), where the house has always been considered as a form of investment and homebuilders and concrete industry have always had a strong influence on governments.

In traditional approaches, studies on settlement system development are almost always dealt with, through the use of spatial data analysis methods (Cerreta and Poli, 2013; Modica et al, 2012; Murgante and Danese, 2011; Perchinunno et al, 2012).

Other approaches adopt different kinds of models, in order to predict urban sprawl phenomenon (Batty and Xie, 1994; Chaudhuri and Clarke, 2014; Clarke and Gaydos, 1998; Girvetz et al., 2008; Kok et al, 2001; Makse et al, 1998; Martellozzo, 2012; Martellozzo and Clarke, 2011; Olofsson et al, 2013).

Up to nowadays, remotely sensed data with an acceptable spatial and spectral resolution, together with GIS software tools, have become increasingly available (Boyd and Foody, 2011; Mesev, 2007; Steiniger and Hay, 2009) and provide a unique tool to retrieve a number of key variables, relevant to the quantification of urban sprawl and soil consumption from a local up to a global scale. However, this growing observational capacity is also increasing the need for research efforts aimed at exploring the potential offered by data processing methods and algorithms, in order to exploit as much as possible this invaluable space-based data source (Lasaponara and Lanorte, 2012; Lasaponara et al, 2014; Nolè et al, 2013).

The work here presented concerns an application study on the process of urban sprawl, carried out by the use of remote sensed information, from ASTER data which already proved to be suitable for these topics. In order to produce synthetic maps of urban areas of the territory, ASTER images were classified using two automatic classifiers, which are Maximum Likelihood Classifier (MLC) and Support Vector Machines (SVMs), the latter based on machine learning theory (Zhu, and Blumberg, 2002). The aim was to compare performances in terms of robustness, speed and accuracy of the two classifiers, regarding urban pixels.

All process steps have been developed integrating Geographical Information System and Remote Sensing, and adopting free and open source software.

2. Materials and methods

2.1. Data and tools

Multi-temporal satellite data with medium spatial resolution is very suitable in urban sprawl phenomena evaluation. Data adopted in this case study is an image acquired by ASTER sensor on August 24, 2007 (source: http://glovis.usgs.gov); additional details can be found in Abrams, 2000, on EOS-TERRA platform (http://asterweb.jpl.nasa.gov/). ASTER captures data at 15 m spatial resolution in 14 bands, from the visible to the thermal infrared wavelengths (Table 1).

ASTER data have been processed and handled with GRASS GIS (Neteler and Mitasova, 2004), Quantum GIS and R (www.r-project.org). GRASS and Quantum GIS are two desktop GIS software, which are integrated by means of a specific plug in, avoiding to change environment in data manipulation. R, on the contrary, is a language and a software environment that can be interfaced with both GRASS and Quantum GIS. The database has been implemented in GRASS core and repository of the application, while Maximum Likelihood Classification and supervised classification algorithms are available in R.

2.2. Study area

The study area is located in Apulia region, more precisely in the area close to the southern part of Bari, the main town of the region (Fig. 1). The area has a dimension of approximately 253 km^2 and a coastline of nearly 17 km. The analysis is mainly focused in Polignano a Mare and Monopoli municipalities and concerns also a portion of Conversano Municipality. This region was characterized by a remarkable phenomenon of urban sprawl for many years (Romano and Zullo, 2014). According to recent data by the Italian National Institute of Statistics (ISTAT, 2011), the area is characterized by a population density included in a range between 140 and 319 inhabitants per km^2.

Apulia is one of the Italian regions with a high trend of urban sprawl (Romano and Zullo, 2014), that often has no correlation with population growth. Fig. 2 shows a substantially constant demographic trend.

2.3. Methodology

The semiautomatic methodology herein described has been implemented in order to carry out a quantitative analysis of the urban sprawl process using ASTER satellite imagery and open source software. The importance of our objective is connected with the fact that today satellite time series and ancillary data are currently available, even free of charge (from national and international spatial agencies) and offer a great potential for a quantitative assessment of urban expansion; nevertheless suitable data processing (free and open source easy to be applied and generalized) are needed to extract and map reliable information from space.

In this paper, in order to produce synthetic maps of urban areas for operative applications, ASTER images were classified using two automatic classifiers, MLC and SVMs, described in Sections 2.4, respectively, with the aim to compare their respective performances in terms of robustness, speed and accuracy. All process steps have been developed integrating Geographical Information System and Remote Sensing, and adopting free and open source software.

In order to assess capabilities and reliabilities of MLC and SVMs for the discrimination of urban from non urban areas, both of them were applied to the same spectral bands in two different ways: (i) firstly as binary classifications with only two classes (urban/not-urban) as output, and then extended to (ii) 8 classes (urban, agricultural soils, forests and green areas, coastal seawater, non-coastal seawater, bare soil, sand and rock, pure sand).

The process started with the identification of training areas (ROI — Region Of Interest) using Quantum GIS, with the creation of a spatial database in GRASS and subsequently the application of MLC and SVMs algorithms implemented in the *rasclass*[1] contributed package of R. A procedure of estimation accuracy has been applied to the results of supervised classification. Accuracy parameters are an important aid in evaluating the effectiveness of a classification algorithm. Accuracy means the quantitative assessment of the pixels correctly assigned to the corresponding classes on the ground. A criterion usually employed for the evaluation of accuracy is based on the comparison between the reference data to earth (test), identified independently from those used for classification (training), and classified data.

This method is based on the "confusion matrix" (Table 2) (Congalton, 1988; Congalton and Mead, 1983) where a number of classes identified by the classifier and from direct observation are listed on rows and columns, respectively. In practice, matrix elements are represented by the number of pixels that the procedure (classification or direct observation) assigns to a given class: on the main diagonal are pixels assigned simultaneously to the same class by both procedures and which, therefore, are correct; those outside of the diagonal are considered errors. The calculation of confusion matrix is done on a limited number of pixels and not on the whole image, otherwise this would paradoxically require the knowledge of ground truth of the whole examination area.

Then, it is possible to obtain from confusion matrix, the total accuracy (overall accuracy) which represents a parameter that indicates the number of pixels correctly classified (i.e. the sum of the cells in the

[1] Package rasclass contains functions to perform supervised and pixel based raster image classification.