

Article

Assessment of the Added-Value of Sentinel-2 for Detecting Built-up Areas

Martino Pesaresi [†], Christina Corbane ^{*,†}, Andreea Julea [†], Aneta J. Florczyk [†], Vasileios Syrris [†] and Pierre Soille [†]

European Commission, Joint Research Center, Institute for Protection and Security of the Citizen, Global Security and Crisis Management Unit, Via Enrico Fermi 2749, I-21027 Ispra, Italy; martino.pesaresi@jrc.ec.europa.eu (M.P.); andreea.julea@jrc.ec.europa.eu (A.J.); anejadflo@gmail.com (A.J.F.); vasileios.syrris@jrc.ec.europa.eu (V.S.); pierre.soille@jrc.ec.europa.eu (P.S.)

* Correspondence: christina.corbane-clairotte@jrc.ec.europa.eu; Tel.: +39-0332-78-3545; Fax: +39-0332-78-5154

† These authors contributed equally to this work.

Academic Editors: Clement Atzberger and Prasad S. Thenkabail

Received: 29 December 2015; Accepted: 25 March 2016; Published: 1 April 2016

Abstract: Monitoring of the human-induced changes and the availability of reliable and methodologically consistent urban area maps are essential to support sustainable urban development on a global scale. The Global Human Settlement Layer (GHSL) is a project funded by the European Commission, Joint Research Centre, which aims at providing scientific methods and systems for reliable and automatic mapping of built-up areas from remote sensing data. In the frame of the GHSL, the opportunities offered by the recent availability of Sentinel-2 data are being explored using a novel image classification method, called Symbolic Machine Learning (SML), for detailed urban land cover mapping. In this paper, a preliminary test was implemented with the purpose of: (i) assessing the applicability of the SML classifier on Sentinel-2 imagery; (ii) evaluating the complementarity of Sentinel-1 and Sentinel-2; and (iii) understanding the added-value of Sentinel-2 with respect to Landsat for improving global high-resolution human settlement mapping. The overall objective is to explore areas of improvement, including the possibility of synergistic use of the different sensors. The results showed that noticeable improvement of the quality of the classification could be gained from the increased spatial detail and from the thematic contents of Sentinel-2 compared to the Landsat derived product as well as from the complementarity between Sentinel-1 and Sentinel-2 images.

Keywords: Sentinel-2; Landsat; Sentinel-1; global human settlement layer; symbolic machine learning; land cover mapping; automatic classification

1. Introduction

There is a strong demand for evidence-based global policies in the economic, social, environmental and disaster risk reduction spheres. The need for accurate, reliable, and timely estimates of land cover of the world at sufficiently detailed spatial and thematic resolutions is underscored by national and international programs and the global change research community [1]. In particular, the understanding of the global human settlements and urban expansion are critical for a large number of issues including housing and urban development, poverty reduction, sustainable development, climate change, biodiversity conservation, ecosystem services provision and disaster management [2].

Several indicators, such as the access to infrastructure, pressure on biodiversity, urban planning and management [3,4], exposure and vulnerability to natural hazards, can be derived from the information on human settlements. Data from satellite imagery and new technologies for processing earth observation data facilitate the study of population distribution on a global scale. Earth observation is a unique source of information for deriving globally-consistent and evidence-based

indicators for improved accountability and coherence of international frameworks. Free and open earth observation data (e.g., Landsat and Sentinel) offer a great potential for large area mapping of human settlements.

The availability of such data has been a driver for different global urban mapping efforts. Gamba and Herold [5] and Schneider *et al.* [6] provide a summary of the data, methods and the estimates of global urban area obtained from these past efforts. The most recent global urban maps exploit essentially low and moderate resolution optical imagery including Système Pour l'Observation de la Terre (SPOT)-Vegetation data [7], Medium Resolution Imaging Spectrometer (MERIS) data [8] and Moderate Resolution Imaging Spectroradiometer (MODIS) data [9]. The geometric resolution of these products is limited to 300 m–1 km, resulting in lack of spatial detail, omission of small cities or towns and poor accuracies for precise characterization of urban settlements and for supporting decision makers and urban planners. More recently, the availability of a global coverage of Synthetic Aperture Radar (SAR) data obtained from TanDEM-X mission (TDM) has been exploited in order to generate a worldwide inventory of human settlements—the Global Urban Footprint (GUF) layer that will be provided publically at a spatial resolution of ~3.0 arcsec (*i.e.*, ~50–75 m) [10].

An ambitious project for Global Land Cover mapping from 30 m resolution Landsat data was launched in China. Two versions of global land cover maps were produced: the Finer Resolution Observation and Monitoring of Global Land Cover (FROM-GLC) and GlobeLand30. Accuracy assessment of the FROM-GLC product showed that several land cover classes, including impervious areas, were poorly classified—as reported in [1,11]. Recognizing the difficulties in making use of fully automated classification techniques for mapping land cover classes, Chen *et al.* [12] developed a pixel-object-based classification (POK-based) approach for generating the GlobeLand30 product with Landsat TM/ETM+ images of 2000 and 2010, as the primary data sources (supplemented by imagery from the Chinese Environmental and Disaster satellite (HJ-1) for the year 2010). The POK-based approach involves knowledge-based interactive verification procedure for the improvement of the quality of the data products. The need for manual editing by domain experts makes it difficult to automate the POK-based approach for continual updating, for developing long-time series of global lands cover datasets and for monitoring of human settlements.

Despite the tremendous progress made in the last decade regarding the purpose of mapping and monitoring human settlements at the global level, there are still numerous challenges to overcome including: (i) the availability of new and finer resolution data with a global consistent coverage; (ii) the technical capacities to handle large volume of satellite data; (iii) high performance computing for processing the data and mosaicking the final product; and (iv) more robust classifiers that minimize human intervention.

The Sentinel-2 mission, launched in June 2015, aims at providing a full and systematic coverage of the land surface at a global level. With a span of 13 spectral bands at a spatial resolution as high as 10 m, Sentinel-2 offers a great potential for addressing the satellite data requirements for fine scale mapping of human settlements. Together with Sentinel-1 SAR data, the prospect of compiling a cloud-free global image database at a fine spatial resolution for mapping human settlements is being gradually achieved.

Recently, a new method for supervised classification of remotely sensed data based on Symbolic Learning [13–15] and Association Analysis [16] was developed by [17]. This novel approach proved to be capable of solving the issue of global processing of multi-sensor and multi-temporal large scale data for the purpose of mapping the global built-up environment over a 40 year period [18]. It was successfully applied on Landsat imagery corresponding to the epochs 1975, 1990, 2000 and 2014. Preliminary validation tests confirmed the higher quality of the information on built-up areas delivered by the Landsat-made Global Human Settlement Layer (Landsat-GHSL) compared to other available global information layers. In GHSL, the built-up area class is defined as the union of all the spatial units collected by the specific sensor and containing a building or part of it as stated in [8]. Human settlements are made by population and physical infrastructures. The building is one essential part of the settlement infrastructure and it is the basic sign of the human presence that can be physically

observable by remote sensing technologies. The concept of Human Settlement adopted in this study relies on the classical notion inherited from the settlement geography, defined as “the description and analysis of the distribution of buildings by which people attach themselves to the land” by [19]. Besides, the GHSL classification schema does not assume any embedded urban/rural dichotomy (built-up structures are mapped independently if they belong to the “rural” or “urban” area definitions).

The latest developments presented in [17] in terms of the robust and computationally efficient SML classifier, combined with the recent release of more than 2000 Sentinel-2 products for a first assessment by the scientific community, represent a unique opportunity to analyze the suitability of Sentinel-2 for fine resolution mapping of human settlements.

In this paper, we investigated the added-value of Sentinel-2 data compared to Landsat in the frame of the GHSL scope. The purpose is to assess the possibility of increased spatial and thematic detail along with improved quality of the geospatial information layer describing human settlements by exploiting the capabilities offered by Sentinel-2. Areas of improvement, including the possibility of combined use of different sensors (*i.e.*, Sentinel-1 and Sentinel-2, Landsat and Sentinel-2), are also explored. The new SML classifier is applied to the different datasets in its simplest form for the purpose of the comparative assessment of Sentinel-2, Sentinel-1 and Landsat outputs.

The paper is structured as follows: Section 2 describes the methodological context with the principles of the SML classifier and its relevance for mapping human settlements. Section 3 presents the experimental set-up, including the testing of the SML classifier on Sentinel-2 data and the comparison with Landsat and Sentinel-1 outputs. Considerations on the potential of scaling up the new classification methodology to deal with the spectral, spatial and temporal characteristics of Sentinel-2 data are discussed in Section 4.

2. Principles of the SML Classifier

The new classification method developed is based on the concept of the empirical multidimensional histogram. In short, associations among class labels and tuples of features are defined. The feature values are derived from the grouping/binning of the multidimensional histogram that describes the data under consideration. Then, a normalized measure scores these tuples according to the number of their occurrences in each class. The analysis of the cumulative distribution of the estimated scores results at the end in their reclassification. The method is designed for working in a complex and information-abundant environment, where it is important to assess the relationships between different data layers in model free and computationally efficient ways. This is particularly relevant for the global processing of multi-sensor and multi-temporal large-scale remote sensing data where the main challenges are related to the large variety of landscapes that introduce spectral heterogeneity within a single land cover class. Traditional parametric classifiers (*e.g.*, maximum likelihood and Mahalanobis distance classifiers) can easily fail in this complex scenario as they rely on high stability both in the input data and in the reflectance characteristics of the target to be detected.

The singularity of the proposed non-parametric approach resides in the fact that the learning is driven by the data, in particular by the frequency of “similar feature sequences” in each class. In the context of this method, “symbolic” refers to the representation of the input data as categorical, as opposed to continuous or quantitative.

The main steps of the method are data reduction and association analysis. The flowchart diagram with the processing steps implied by the SML method is shown in Figure 1. For our example, the attributes (features) of the input data set are spectral and textural descriptors of a high resolution satellite image. The pre-processing stage for data reduction consists of three main phases. The first one implies the generation of a taxonomy (*i.e.*, quantization of numerical values in our case), the second phase constructs the data sequences while during the last phase, the unique sequences are identified and their frequencies of occurrence are calculated. The results of the input data reduction stage, the unique data sequences ($\hat{X}_{q_i,F}$), are used in the following association analysis step. The coarse resolution reference data set originating from available high abstraction semantic layers (Y) is also needed in the

association analysis step where the calculation of the interestingness measure and the classification are performed.

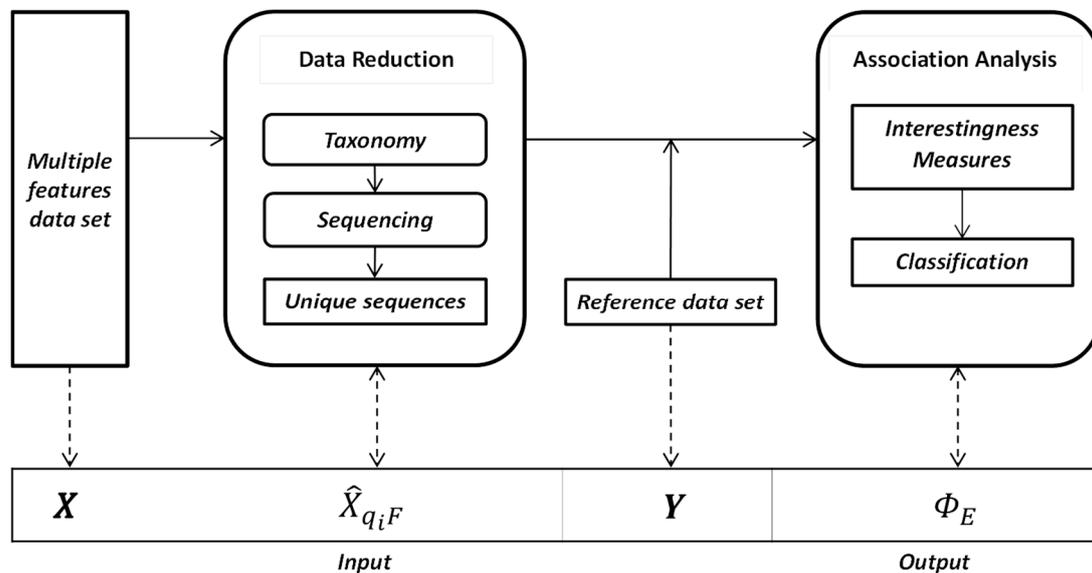


Figure 1. The Symbolic Machine Learning approach for remote sensing data classification.

2.1. Data Reduction

The specific clustering approach for data (re)quantization and data sequencing uses two main criteria. The first criterion comes from the requirement of an objective method—as much as possible independent from assumptions about specific statistical distributions. The second criterion is linked to the necessity of adopting a sparse representation of the input data features and discriminant functions. Such coding allows the compression of the feature space, increasing the computational efficiency for global processing of remote sensing data.

Let $D_{m \times n \times F}$ be a satellite scene with $m \times n$ pixels and F features. The latter can be spectral features—as bands—or derived image features as textural and morphological measurements.

Let $X_{mn \times F}$ be a 2-dimensional data matrix with mn = number of rows and F columns, where $X = [x_1, x_2, \dots, x_i, \dots, x_F]$ and $x_i \in \mathbb{Z}_+^{mn}$. X has the cardinality $|X|$ expressing the number of input data instances.

Let \hat{X} be the set of all the unique data instances of X with having as cardinality $|\hat{X}| \leq mn$. This magnitude depends on the specific number of symbols s_i used to encode the x_i values and on the number of features F .

The average support of \hat{X} can be estimated as $supp_\mu = |X|/|\hat{X}|$. The $supp_\mu$ influences the generalization and computational performances of the SML classifier and thus it must be controlled. Assuming as given the number of features F , the control of $supp_\mu$ is achieved by controlling the number of symbols used to encode X by quantization. Several quantization or re-discretization approaches including uniform, adaptive and statistical approaches may be found in literature as reviewed by [20]. In the method presented in this manuscript, a uniform quantization approach was implemented, with the quantized data instances defined as:

$$X_{q_i F} = \lceil [x + 0.5] / q_i : x \in x_i \rceil_{mn \times F} \quad (1)$$

with $q_i = \max(x_i) / s_i$ and $i = 1, \dots, F$.

Accordingly, the respective quantized unique data sequences are denoted by $\hat{X}_{q_i F}$. The quantization parameter q_i is empirically determined by choosing a value so that $supp_\mu$ is in the range of 10^2 – 10^3 . These $supp_\mu$ orders of magnitude were tested as satisfactory for classification

exercises incorporating noisy reference sets, with noise characteristics similar to the ones experimented in [17].

2.2. Association Analysis

Association analysis is used to search for relevant, systematic relationships between image data instances and spatial information encoded in a selected reference set containing class memberships.

Let $A = \{A_1, A_2, \dots, A_j, A_{j+1}\}$ be a set of $j + 1$ distinct attributes (in Association Analysis context, the term “attribute” has been established in place of the term “feature”; in our framework these two concepts are used interchangeably. For any two disjoint attribute-value subsets $\{X\}$ and $\{Y\}$ of A , the patterns of the form $X \rightarrow Y$ are called association rules. The attribute-value sets X and Y are named antecedent and consequent of the association rule, respectively. In the associative classification framework started by [21], the class association rules are the association rules with a class label attribute as the only consequent ($X = \{A \setminus C\}$ and $Y = A_{j+1} = C$, where C is the attribute set containing the class labels).

The SML method consists in using association analysis to quantify the association $X \rightarrow Y$ between multi-attribute unique data sequences \hat{X}_{q_iF} and available high-abstraction semantic reference layers Y .

The method quantifies the associations using an index inspired by the family of probability-based objective interestingness measures [22,23], namely the Evidence-based Normalized Differential Index (ENDI). This measure scores the data sequences in \hat{X}_{q_iF} according to the number of their occurrences in each reference class in Y . Finally, using discriminant thresholds determined through the analysis of the cumulative distribution of these scores, the classification results are obtained.

Different variations of the ENDI measure, noted Φ_E , exist and several discriminant thresholds can be applied on each of these variations leading to different classification results. Five automatic discriminant thresholds were tested in [17] (no data values are excluded from the calculations):

$$cl = \begin{cases} 1, \forall x_{kj} > T_l : x_{kj} \in \Phi_E \\ 0, otherwise \end{cases} \quad (2)$$

with threshold values given by

$$T_{l=} = [0, \operatorname{argmin}_i |y_i - \Sigma R|, m_1, m_0, (m_1 - m_0)/2] \quad (3)$$

where y_i are the values of the cumulative sum over the quantized values of Φ_E , $m_0 = \operatorname{mean}(x_{lp}) : (x_{lp} \in \Phi_E) \wedge (r_{lp} \in R) \wedge (r_{lp} = 0)$ and $m_1 = \operatorname{mean}(x_{lp}) : (x_{lp} \in \Phi_E) \wedge (r_{lp} \in R) \wedge (r_{lp} = 1)$, for $l = 0, \dots, 4$. Particularly, for the case of $l = 0$, the “>” operator becomes “ \geq ”.

In [17,24], an extensive experiment involving a combination of three variations of the ENDI measure and the five threshold values (15 combinations) was proposed to rank the classification outputs in terms of error rates (omission and commission errors). The authors found that a good measure, providing balanced error rates, is the one defined as follows:

$$\Phi_E^{ab} = \frac{\Phi_E^a + \Phi_E^b}{2} \quad (4)$$

with:

$$\Phi_E^a(X, Y^+, Y^-) = \frac{f_{pos} - f_{neg}}{f_{pos} + f_{neg}} \quad (5)$$

$$\Phi_E^b(X, Y^+, Y^-) = \frac{p_{pos} - p_{neg}}{p_{pos} + p_{neg}} \quad (6)$$

where f_{pos} and f_{neg} are the frequencies of the joint occurrences among X data instances and the positive and negative reference instances respectively (Y^+ and Y^-); p_{pos} and p_{neg} represent the empirical

probabilities calculated as $p_{pos} = \frac{f_{pos}}{N_{pos}}$; $p_{neg} = \frac{f_{neg}}{N_{neg}}$; N_{pos} and N_{neg} denote the total of the positive and negative elements of the reference set, respectively.

The Φ_E^{ab} has values ranging between $[-1, +1]$. The best threshold according to the tests performed in [18], is an unsupervised threshold at zero level, which results in a more balanced behavior of commission and omission errors.

An extensive benchmarking exercise to assess both the computational complexity of the SML classifier and its robustness against different types and levels of noise ingested in the reference set was implemented in [17,24]. These issues often appear when learning is guided by low resolution or low quality reference sets. Seven supervised classifiers used in the statistical learning and data mining community were considered for comparison: Maximum Likelihood, Logistic regression, Linear Discriminant Analysis, Naive Bayes, Decision Tree, Random Forest and Support Vector Machines. The analysis showed that the SML classifier outperforms other parametric and non-parametric classifiers both in terms of computational efficiency and accuracy, especially if applied in the following operative conditions: (i) ill-defined and spectrally heterogeneous target classes (as built-up areas); (ii) large input data with heterogeneous characteristics (including lack of seasonal standardization and/or radiometric calibration); and (iii) noisy reference set (scale generalization noise, displacement and spatial misalignment error noise, and thematic random noise).

3. Application of the SML Classifier to Sentinel-2

The new classification method briefly described above has been successfully applied to solve the first global processing of Landsat data records and monitoring the evolution of human settlements over the past 40 years [18]. The prospect of deploying the system to the classification of Sentinel-2 imagery is a pressing technological issue that requires a series of tests to analyze the suitability of this new sensor data for global mapping of human settlements.

In this section, a preliminary test has been implemented with the purpose of: (i) assessing the applicability of the SML classifier to Sentinel-2 imagery; (ii) evaluating the complementarity of Sentinel-1 and Sentinel-2; and (iii) understanding the added-value of Sentinel-2 with respect to Landsat for the improvement of the global high-resolution human settlement mapping. For this experiment, a Sentinel-2 scene covering the city of Porto Viro in the area of the Po river delta, northeast of Italy was used. The study area combines both highly populated settlements and large agricultural plains with many small scattered built-up areas. The image acquired on 8 July 2015 was provided as part of the first pre-qualified expert user data released by the European Space Agency (ESA). The subset shown in Figure 2 covers an area of 110×110 km².

This test was designed to exploit the key feature of Sentinel-2 imagery, which is the availability of four 10 m spatial resolution bands: B2 (490 nm), B3 (560 nm), B4 (665 nm) and B8 (842 nm). The data were taken as delivered without trying to compensate for any problems that may still exist in this very pre-operational dataset (e.g., geometrical precision, coarse atmospheric corrections, and sensor geometry).

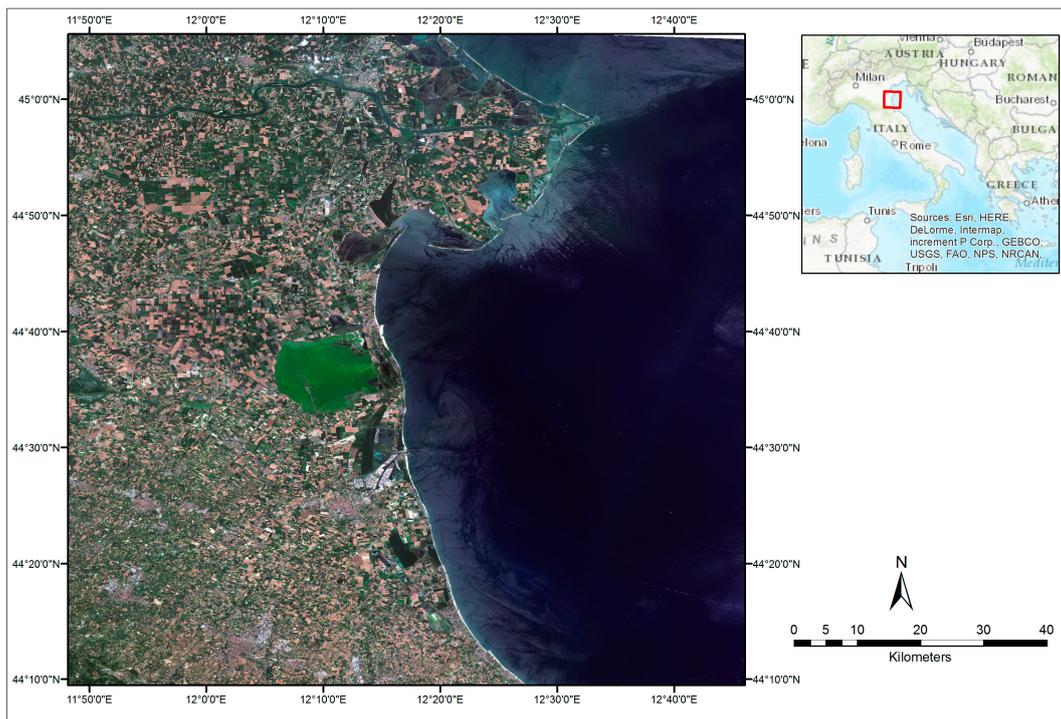


Figure 2. True color composite of Sentinel-2 image (Red: Band 3, Green: Band 2, Blue: Band 1) acquired on 8 July 2015 and covering the Po Delta on the Adriatic Sea, northeast (central coordinates: 12° 13'E, 44° 35'N).

3.1. Input Features

The standard set of image features used in the GHSL production workflow [18] includes radiometric, textural and multi-scale morphological descriptors. In the example reported here, only the first two were used as input to the SML classifier because they represent a good compromise between computational cost and accuracy. Besides, the morphological descriptors are mostly used with finer scale input sensors typically in the range of 0.5–2.5 m and were considered less relevant with the 10-m-resolution Sentinel-2 data. The following features were finally used for the classification of the Sentinel-2 image with the SML approach.

- Spectral features: Only the four input spectral bands at 10 m resolution, namely B2, B3, B4 and B8, were tested in this experiment for their potential in discriminating different land cover classes.
- Textural features: Several studies have demonstrated the utility of the textural measures for land cover mapping and in particular for built-up area extraction [25,26]. “The inclusion of one-band texture may not take full advantage of the available texture information contained in the multispectral image. By contrast, multi-band texture measures, derived from all available spectral bands of a multispectral image [27], allow identifying the spatial variation of multiple bands, thus avoiding the need to select a single band as required in the traditional one-band texture method” [25]. Therefore, the textural analysis was made on the four input bands by applying anisotropic rotation-invariant contrast textural measures extracted using the grey-level co-occurrence matrix (GLCM) and the Pantex methodology of Pesaresi *et al.* [28]. The applied operative parameters were a window size of 5 × 5 pixels (50 × 50 m) and a displacement vector distance of 1 pixel with 8 directions.

All eight features were quantized in $s_i = 16$ levels, $i = 1, \dots, 8$ based on the quantization-sequencing approach described in Section 2 using the quantization parameter $q = 4096$ (Equation (1)). The number of resulting unique sequences ($\hat{X}_{q_i F}$) in the image under test was 38,762.

3.2. Learning Set

The basic information components of the SML automatic image information extraction workflow are clouds, water, and built-up areas [18]. In the test reported here, only the water/land and the built-up/non-built-up semantics are considered. The reference set used for the example was extracted from Corine Land Cover (CLC) product of year 2006 (Copernicus Land Monitoring Services, <http://land.copernicus.eu/pan-european/corine-land-cover/clc-2006/>) used in the gridded version at 100 meter resolution. The positive examples of the target class “water” were extracted from the CLC classes 5.1 and 5.2, corresponding to inland and marine waters, respectively. The positive examples for the “built-up” class were extracted from the CLC classes of continuous urban fabric (code 1.1.1), discontinuous urban fabric (code 1.1.2), and industrial or commercial units (code 1.2.1). The CLC reference set can be considered as affected by thematic and spatiotemporal noise as regarding the target classes in the Sentinel-2 image. A time gap of nine years (2006 for the CLC and 2015 for the Sentinel-2 data) and a scale/generalization gap (a spatial resolution of 100 m for CLC size *versus* 10 m for Sentinel-2) can be assumed between the image data and the CLC data. Moreover, semantic differences may exist between the two datasets: for example, code 1.1.2 of CLC corresponding to discontinuous urban fabric includes large non built-up surfaces like vegetated areas and bare soils that are captured by the spatial resolution of Sentinel-2. Similarly, some of the CLC classes may include surface water, even if they are not included in the positive examples of the “water” class (e.g., the wetlands class of CLC).

The reference set used for the example is shown in Figure 3a. The positive examples for the built-up and water target classes, respectively, as extracted from the CLC source, are represented in white and blue.

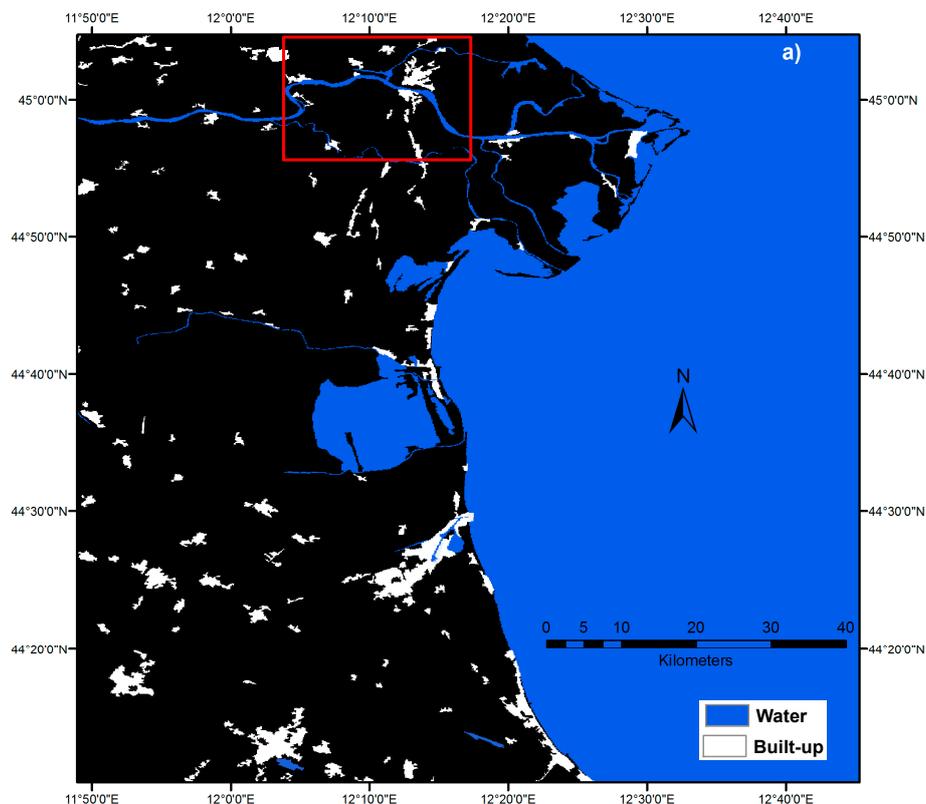


Figure 3. Cont.

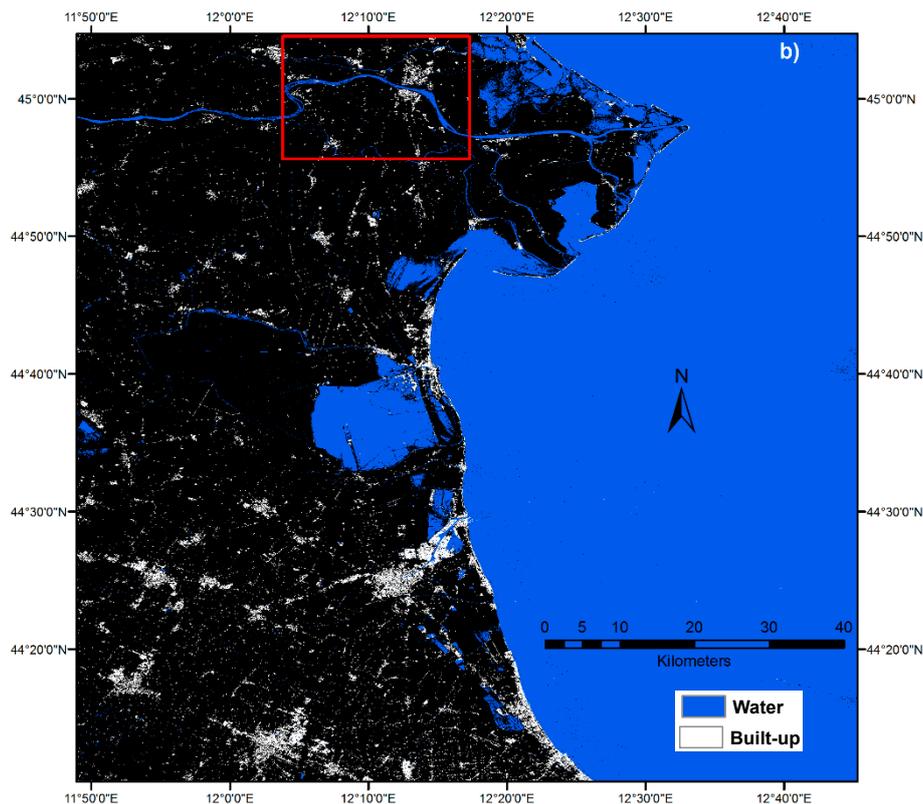


Figure 3. (a) Corine Land Cover (CLC) reference set; and (b) the classified Sentinel-2 imagery.

3.3. Results and Accuracy Assessment

Figure 3b shows the results of the SML classifier applied to Sentinel-2 data. The class corresponding to built-up areas is displayed in white and the water is shown in blue. The visual inspection of the results of the classification shows a noticeable increase of detail with respect to the thematic information of the CLC reference set Figure 3a. It is worth noting the SML outputs may significantly contradict the information provided in the reference set if the evidences collected from the image data sufficiently support the decision. In both the built-up and the water examples, important information contradicting the reference set was discovered: in the built-up case scattered settlement structures were discovered even if neglected by the reference, while, in the water case, a large water surface (red rectangle) was classified accordingly by the SML association analysis. This area (red rectangle) corresponds to salt marshes according to CLC data but was not included in the reference set for the “water” class. Moreover, thin water channels and small water bodies are classified as well. Analogously to the built-up case, these thin structures are the added-value provided by the higher spatial resolution of the image data with respect to the reference set.

In addition to the refined classification of the built-up areas and water surfaces, the analysis of Sentinel-2 with the SML classifiers reveals some disagreements with the CLC data. A close-up view of the upper right part (Figure 4) shows that the SML classifier correctly identifies the built-up industrial area (red rectangle) where the CLC dataset shows a much larger area as industrial including non-built-up features. However, misclassification errors were observed along the coastal areas where the sand dunes were detected in the built-up class because of similar spectral characteristics in the Sentinel-2 data.

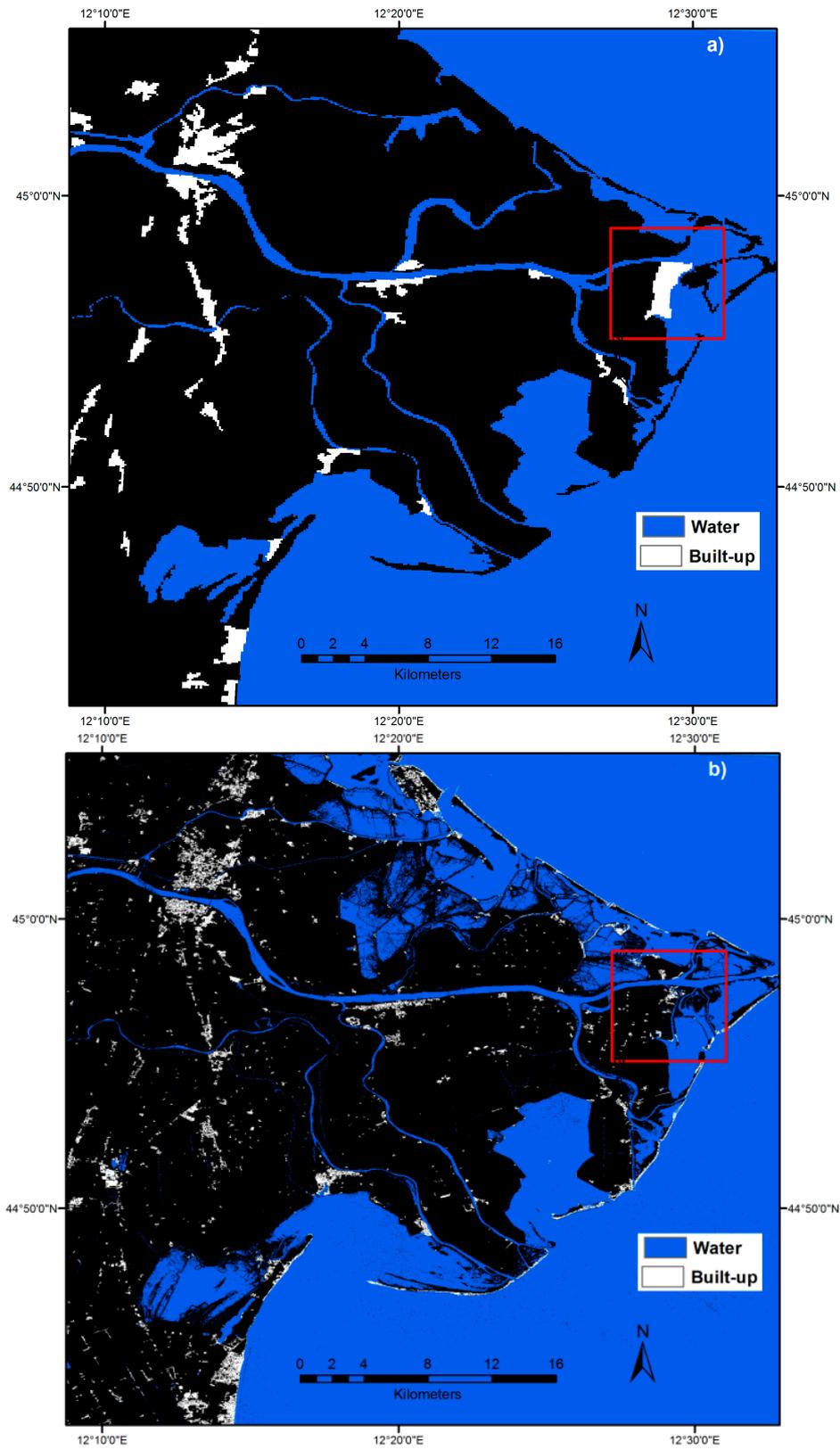


Figure 4. Cont.



Figure 4. (a) Close-up view of the CLC training; (b) classified Sentinel-2; and (c) the original Sentinel-2 image in natural colors.

The accuracy of the SML classification is evaluated for the built-up target class using visual interpretation of 446 random samples equally distributed in the built-up and non-built-up areas from the training set. The accuracy of the water class was not assessed in this test. Table 1 shows the main results of the accuracy assessment. Following the considerations of [18], the reporting of quantitative accuracy metrics is based on the error matrix.

Table 1. Bias adjusted accuracy measures using visual interpretation of 446 stratified random samples equally distributed in the built-up and non-built-up areas.

<i>Balanced accuracy</i>	0.80
<i>Informedness</i>	0.61
<i>Commission error</i>	0.80
<i>Omission error</i>	0.19

The overall accuracy represents the share of all correctly classified sample units of the entire error matrix. The Kappa coefficient, introduced by [29], presents a discrete multivariate technique commonly used in accuracy assessment to statistically test if classifications are significantly different from the reference data, being a chance corrected renormalized form of accuracy. Nevertheless, this very popular measure has one weakness when dealing with imbalanced datasets, as it proves to be the case of built-up classification where very few positive examples compared to the ones of non-built-up class exist. In such situations, bias and prevalence effects preclude right judgments by Kappa values. Consequently, we also use two quality measures that are skew-insensitive, thus better controlling inflated performance estimates on imbalanced datasets. The first one is the balanced accuracy (BA) from [30] and it corresponds to the Area Under the Curve from the Receiver-Operating Characteristic (ROC) analysis (Equation (7)). The second one is the Informedness [31] (Equation (8)), which can be

viewed as an unbiased renormalized form of recall (True Positive Rate—TPR). The measure can be directly expressed as the probability that an informed decision is made rather than a random guess (chance), as stated by [31].

$$\text{Balanced Accuracy} = \frac{\text{TPR} + \text{TNR}}{2} \quad (7)$$

$$\text{Informedness} = \text{TPR} - \text{FPR} \quad (8)$$

With the usual terms from the binary confusion matrix: TPR = True positive rate (TP/P); TNR = True negative rate (TN/N); FPR = False positive rate (FP/N); TP = True Positive; TN = True Negative; FP = False Positive; P = Positive instances; and N = Negative instances. The accuracy measures were all calculated after applying the bias-adjustment for stratified random sampling proposed by Card [32].

Table 1 summarizes the accuracy measures: the balanced accuracy reaches 81%, the Informedness value, which is more robust to skewed distributions, is also satisfactory and the omission error is low. However the classification resulted in high commission errors. These errors can be mainly observed in coastal areas (rocky shorelines or sand dunes) or artificial open spaces and bare ground. Omission errors are essentially related to scattered and isolated buildings that were difficult to detect on the Sentinel-2 data due to their average size, which is smaller than the typical side dimension ranging from 10 to 20 m [33].

4. Comparative Analysis

One of the advantages of the SML method is the possibility to apply it to other remotely sensed data, and the outcome may be complementary. Therefore, it is important to understand the differences in the detection of built-up areas produced from data from different sensors. In this section, we will provide the first results of such comparative studies.

4.1. Comparison to Landsat

The built-up areas obtained with Sentinel-2 data, as described in Section 3.3, were compared to the built-up classification outputs extracted from the Landsat data. The Landsat mage records used to produce the GHSL dataset, correspond to the four epochs 1975, 1990, 2000 and 2014 [18] (the most recent Landsat image over this area was acquired on 26 August 2014). This multi-temporal dimension was summarized through one class, *i.e.*, built-up area for the comparison with the outputs of Sentinel-2. Figure 5 provides a detailed look of the two classification outputs (Figure 5b for Landsat and Figure 5c for Sentinel-2). A visual inspection of the results shows a noticeable improvement of the quality of the classification of built-up areas with Sentinel-2 data. This is observable in agricultural fields that were misclassified as built-up with Landsat imagery. The textural features that were calculated from the Sentinel-2 bands with the Pantex methodology allow mitigating the problem of confusion between bare soils in agricultural fields and built-up areas. Another significant improvement when applying the SML classifier on Sentinel-2 data is related to the identification of thin water channels and other waterways, like the secondary branches of the Po River (e.g., Po di Volano).

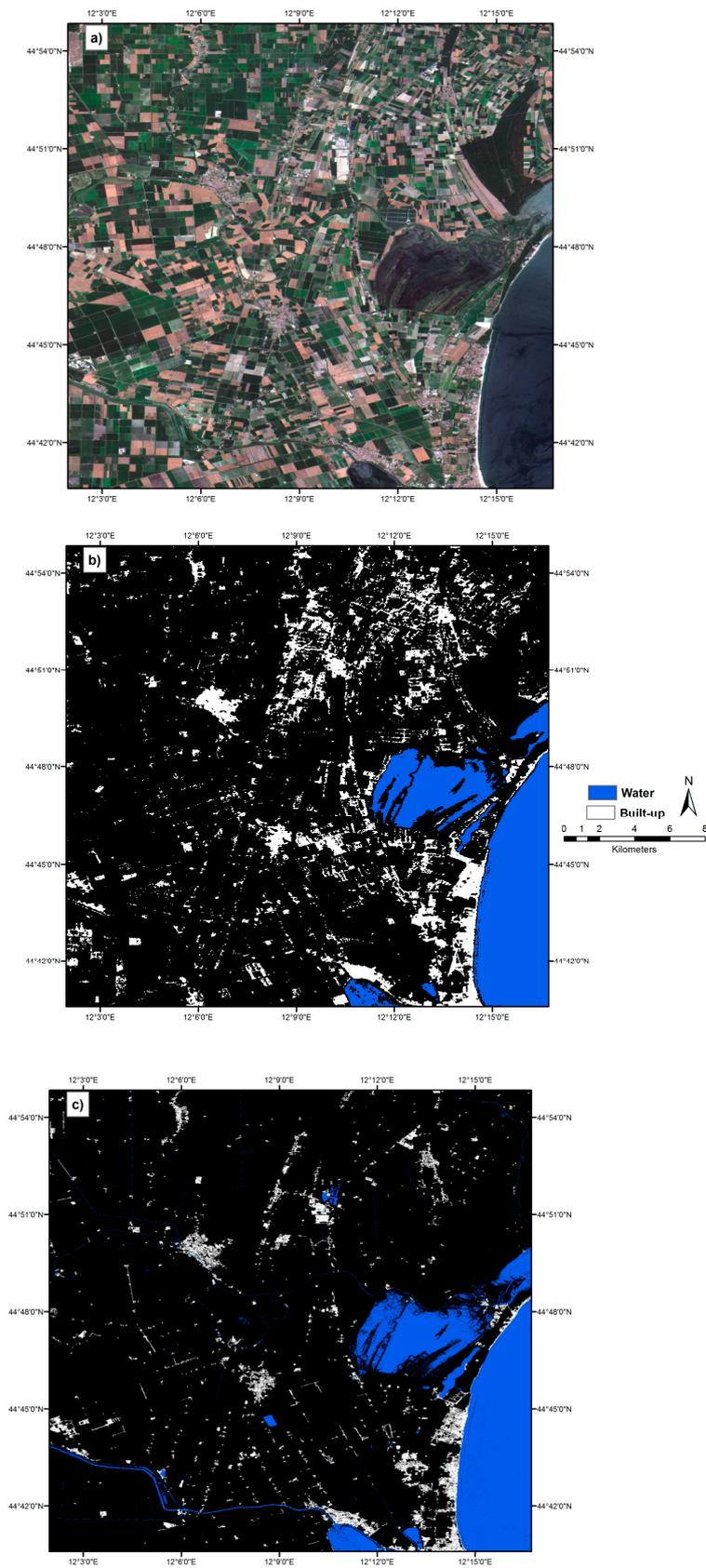


Figure 5. Detailed view of an area from the Sentinel-2 image in natural colors (a) and the two classification outputs (built-up and water classes) obtained with Landsat (b) and Sentinel-2 (c) (all three figures are at the same scale).

4.2. Comparison to Sentinel-1

Another area along the Po river centered on the town of Sannazzaro (Central coordinates: $45^{\circ}6'N$; $8^{\circ}54'E$) was selected for the next comparative analysis. This area is characterized by the presence of human settlements surrounded by agricultural fields, which are crossed by the Po River and a highway (Figure 6a). Both Sentinel-1 and Sentinel-2 were available over this area. The Sentinel-1 image is in the Stripmap mode (SM) (80 km swath, $5\text{ m} \times 5\text{ m}$ resolution, single-look) with a dual polarization (HH + HV). It was acquired on 26 June 2015. The SML classifier was applied on both datasets following the same approach described in Sections 3.1 and 3.2 using the CLC data as a reference set. In the case of Sentinel-1, the sequencing phase of the SML classifier included the following seven features: the HH and HV back-scattering amplitude values, two textural features computed on the HH and HV bands following the Pantex methodology described in [28] and three topographic features derived from the 90 m global digital elevation model of the Shuttle Radar Topography Mission (SRTM). The topographic features correspond to the slope, the aspect and ridgelines. They were integrated in the classification as contextual information to account for topographic distortions, typically the problems of foreshortening, layover and shadow inherent to the SAR imagery. Besides, ridgelines appear as bright linear features similar to building footprints and can be hence misclassified as built-up areas in the Sentinel-1 image.

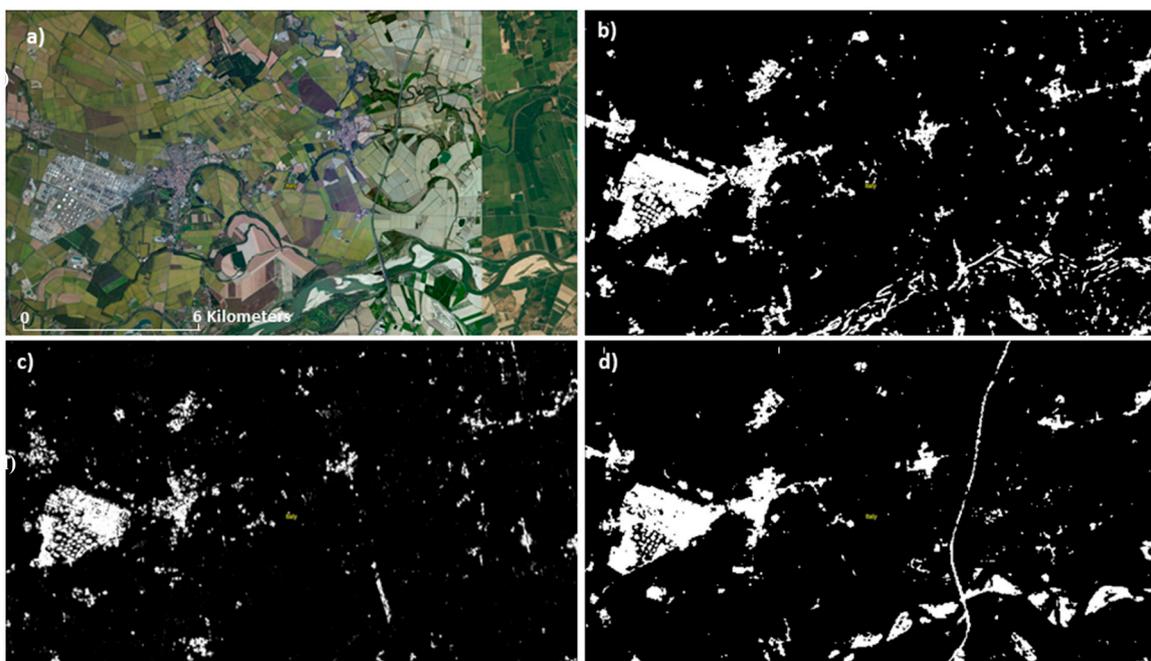


Figure 6. (a) The Google earth image of the area under study is displayed here with the: (b) results of the SML classifier applied on Sentinel-2 and (c) Sentinel-1 compared to (d) the GHSL derived from Landsat imagery.

The results of the classifications are shown in Figure 6b,c for Sentinel-2 and Sentinel-1, respectively. The detected built-up areas are shown in white. An extract over the same area as characterized by the Landsat based GHSL is also shown here for comparison purposes (Figure 6d). Significant improvement associated with the exclusion of the highway from the detected built-up areas can be observed when comparing the results of Sentinel-1 and Sentinel-2 to those obtained with Landsat data. This may be attributed to the textural features that were computed on the 5 meter and 10 meter resolutions of Sentinel-1 and Sentinel-2 data, respectively, and that allow discriminating built-up structures from other image features (edges and lines) as demonstrated in [28].

It is interesting to note how the confusion between the built-up areas and the fluvial gravel observed on the results of Sentinel-2 and Landsat is overcome with Sentinel-1 SAR imagery. Given the high sensitivity of the SAR imagery to surface roughness compared to optical images, the distinction between built-up areas and fluvial gravel, similar in terms of radiometric characteristics but having different surface roughness, becomes possible. This characteristic of SAR imagery would also allow discriminating sand dunes, which exhibit a smooth texture on Sentinel-1 imagery and hence reduce the problem of the confusion between built-up areas and sand dunes obtained with Sentinel-2 data. Conversely, the Sentinel-1 data seem to underestimate the built-up surfaces in high density areas where shadows hamper the automated classification. These results outline the complementarity between the Sentinel-1 (SM mode) and Sentinel-2 sensors for increased accuracy in the detection of the built-up surfaces. Further analysis to assess the utility of combining/fusing the two images and testing the SML classifier on the resulting dataset for improving the discrimination of built-up areas is needed.

5. Discussion

The recent availability of Sentinel-2 data is expected to bring land cover mapping and monitoring to an unprecedented level. Among the main advantages of Sentinel-2, which make the satellite highly suitable for mapping and monitoring human settlements at a global level, it is the combination of the wide swath and the frequent revisiting time (currently 10 days with one satellite, and every five days once Sentinel-2B is in orbit) at high spatial resolutions. The Sentinel-2 mission establishes a key European source of data to the Copernicus environmental monitoring program. The sensor will not only enhance the earth observation data archive but it will also deliver new capabilities for mapping built-up areas at a global level.

The experiments presented in this paper demonstrate the added-value of Sentinel-2 for mapping built-up areas at a higher spatial resolution than the Landsat-derived GHSL product. The spectral confusions with other land cover classes observed in the Landsat-derived GHSL product were partially alleviated using Sentinel-2 data, especially for asphalt concrete roads and barren lands in agricultural fields. Despite this improvement, omission errors were observed in the results obtained with Sentinel-2 data (omission rate = 38%). These omissions are strongly inversely correlated to the density of the settlement patterns. Built-up surface associated with rural and scattered low-density settlement patterns are expected to be relatively under-represented. This was also observed in the GHSL product using the Landsat data where the omission rates, assessed with detailed building footprints, were in the order of 50% to 60% [18]. The capacity to discriminate such spatially scattered settlements is strongly linked to the spatial resolution of the input sensor. With Sentinel-2 data, omission rates tend to decrease, nevertheless single buildings and small settlement patterns surrounded by vegetation may still be difficult to identify, requiring spatial resolutions higher than 5 m.

As shown in this paper, deeper insight into the mapping of human settlements can be gained from the exploitation of the complementarity between Sentinel-1 and Sentinel-2 data. In this paper, Sentinel-1 in Stripmap mode was tested with a spatial resolution of 5 m. This type of imagery is used in exceptional cases only, to support emergency management actions. Additional tests are currently being conducted on Sentinel-1 Interferometric Wide Swath mode, which is the main acquisition mode over land satisfying the requirements for global mapping. These tests would help in the assessment of the added-value of combining Sentinel-1 and Sentinel-2 data for improving the discriminatory ability over that of either sensor alone.

Further experiments are ongoing to consider the possibility of exploiting the Sentinel-2 bands which were not included in the current classifications. The use of the 20 meter bands especially in the Near Infrared and Shortwave Infrared may help in alleviating the problem of confusion between built-up areas and other types of land cover such as artificial open spaces, river gravel and sand dunes. As suggested by [34], the physical and chemical characteristics of different urban surfaces are represented in all parts of the VIS-NIR-SWIR-TIR with important spectral signatures in the SWIR

region. It is therefore necessary to examine the benefit of exploiting both the 10 and 20 meter bands of Sentinel-2 without compromising the ability to detect fine scale human settlements.

6. Conclusions

The application of the SML classifier to Sentinel-2 data is encouraging because of improvements in terms of spatial detail and thematic contents with respect to the Landsat derived product. It also demonstrates the capacity of the SML methodology to handle different sets of input features such as radiometric, textural and morphological descriptors derived from Sentinel-2 data. Among the important characteristics of the SML methodology, there is the possibility to extract other types of land cover other than built-up areas. Due to the symbolic nature of the SML classifier, the methodology can be seamlessly expanded to multi-class classification problems.

New insights into the global mapping of human settlements can be expected by taking advantage of the singularities of the SML classifier and of the rich information content and operational capacities of Sentinel-2 satellites. Several application areas and a wide user community can benefit from the deployment of the SML classifier to the classification of Sentinel-2 imagery as has been demonstrated by the different GHSL applications: population spatial modeling [35], global exposure mapping [36], identification of informal human settlements [37], mapping and monitoring of refugees and internally displaced people [38,39], disaster risk reduction [2] and as a support to damage and needs assessment [40].

The first results presented in this paper show that opportunities exist to confront the global human settlement challenges thanks to the possibility of mapping built-up areas, their heterogeneities, dynamics and interactions with the environment.

Acknowledgments: The work was supported by the European Commission's Joint Research Centre institutional work program relatively to the Global Human Settlement Layer (GHSL) project. The authors would like to thank the whole GHSL team for the substantial help provided during all the phases of the experimental activities reported.

Author Contributions: Martino Pesaresi performed the experiments; Christina Corbane drafted the manuscript; Andreea Julea assessed the SML method and the results; Aneta Florczyk prepared the Landsat dataset and performed the quality assessment; Vasileios Syrris engineered the SML prototype and assessed the method from the perspective of the computational complexity and memory load; and Pierre Soille prepared and pre-processed Sentinel-1 and Sentinel-2 datasets. All the authors contributed equally in reviewing and finalizing the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Ban, Y.; Gong, P.; Giri, C. Global land cover mapping using Earth observation satellite data: Recent progresses and challenges. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 1–6. [[CrossRef](#)]
- Pesaresi, M.; Ehrlich, D.; Ferri, S.; Florczyk, A.; Freire, S.; Haag, F.; Halkia, M.; Julea, A.M.; Kemper, T.; Soille, P. Global human settlement analysis for disaster risk reduction. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *XL-7/W3*, 837–843. [[CrossRef](#)]
- Triantakoustantis, D.; Chrysoulakis, N.; Sazonova, A.; Esch, T.; Feigenwinter, C.; Düzgün, S.; Parlow, E.; Marconcini, M.; Tal, A. On-line evaluation of earth observation derived indicators for urban planning and management. *Urban Plan. Des. Res.* **2015**, *3*, 17–33. [[CrossRef](#)]
- Chrysoulakis, N.; Feigenwinter, C.; Triantakoustantis, D.; Penyevskiy, I.; Tal, A.; Parlow, E.; Fleishman, G.; Düzgün, S.; Esch, T.; Marconcini, M. A conceptual list of indicators for urban planning and management based on earth observation. *ISPRS Int. J. Geo-Inf.* **2014**, *3*, 980–1002. [[CrossRef](#)]
- Gamba, P.; Herold, M. *Global Mapping of Human Settlement: Experiences, Datasets, and Prospects*; CRC Press: Boca Raton, FL, USA, 2009.
- Schneider, A.; Friedl, M.A.; Potere, D. Mapping global urban areas using MODIS 500-m data: New methods and datasets based on “urban ecoregions”. *Remote Sens. Environ.* **2010**, *114*, 1733–1746. [[CrossRef](#)]
- Bartholome, E.; Belward, A.S. GLC2000: A new approach to global land cover mapping from Earth observation data. *Int. J. Remote Sens.* **2005**, *26*, 1959–1977. [[CrossRef](#)]

8. Bontemps, S.; Defourny, P.; Bogaert, E.V.; Arino, O.; Kalogirou, V.; Perez, J.R. *GLOBCOVER 2009—Products Description and Validation Report*; ESA Technical Report; European Space Agency (ESA): Paris, France; Université catholique de Louvain (UCL): Louvain-la-Neuve, Belgium, 2011.
9. Schneider, A.; Friedl, M.A.; Potere, D. A new map of global urban extent from MODIS satellite data. *Environ. Res. Lett.* **2009**, *4*. [[CrossRef](#)]
10. Marconcini, M.; Esch, T.; Felbier, A.; Heldens, W. Unsupervised high-resolution global monitoring of urban settlements. In Proceedings of the 2013 IEEE International Geoscience and Remote Sensing Symposium, Melbourne, Australia, 21–26 July 2013; pp. 4241–4244.
11. Gong, P.; Wang, J.; Yu, L.; Zhao, Y.; Zhao, Y.; Liang, L.; Niu, Z.; Huang, X.; Fu, H.; Liu, S.; *et al.* Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens.* **2013**, *34*, 2607–2654. [[CrossRef](#)]
12. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M.; *et al.* Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 7–27. [[CrossRef](#)]
13. Hoffman, A. *Paradigms of Artificial Intelligence: A Methodological and Computational Analysis*; Springer-Verlag Singapore Pte. Limited: Singapore, 1998.
14. Ganascia, J.-G.; Pitrat, J.; Laurière, J.-L.; Brazdil, P.; Delobel, C.; Diday, E.; Gaudel, M.-C.; Kodratoff, Y. *AGAPE et CHARADE: Deux Techniques D'apprentissage Symbolique Appliquées à la Construction de Bases de Connaissances*; Faculté des Sciences d'Orsay, Université de Paris-Sud: Paris, France, 1987.
15. Ganascia, J.-G. Induction Symbolique et Numérique à partir de données. In *Charade: Apprentissage de Bases de Connaissances*; Diday, E., Kodratoff, Y., Eds.; Cépaduès Editions: Toulouse, France, 1991; pp. 309–326.
16. Tan, P.N.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*, 1st ed.; Addison-Wesley Longman Publishing Co., Inc.: Boston, MA, USA, 2005.
17. Pesaresi, M.; Syrris, V.; Julea, A. *Benchmarking of the Symbolic Machine Learning Classifier with State of the Art Image Classification Methods—Application to Remote Sensing Imagery*; JRC Technical Report; European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen: Ispra, Italy, 2016.
18. Pesaresi, M.; Ehrlich, D.; Ferri, S.; Florczyk, A.; Freire, S.; Halkia, M.; Julea, A.; Kemper, T.; Soille, P.; Syrris, V. *Operating Procedure for the Production of the Global Human Settlement Layer from Landsat Data of the Epochs 1975, 1990, 2000, and 2014*; JRC Technical Report; European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen: Ispra, Italy, 2016.
19. Stone, K.H. The Development of a focus for the geography of settlement. *Econ. Geogr.* **1965**, *41*, 346–355. [[CrossRef](#)]
20. Liu, H.; Hussain, F.; Tan, C.L.; Dash, M. Discretization: An enabling technique. *Data Min. Knowl. Discov.* **2002**, *6*, 393–423. [[CrossRef](#)]
21. Liu, B.; Hsu, W.; Ma, Y. Integrating classification and association rule mining. In Proceedings of the International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 27–31 August 1998; pp. 80–86.
22. Tew, C.; Giraud-Carrier, C.; Tanner, K.; Burton, S. Behavior-based clustering and analysis of interestingness measures for association rule mining. *Data Min. Knowl. Discov.* **2014**, *28*, 1004–1045. [[CrossRef](#)]
23. Geng, L.; Hamilton, H.J. Interestingness measures for data mining: A survey. *ACM Comput. Surv.* **2006**, *38*. [[CrossRef](#)]
24. Pesaresi, M.; Syrris, V.; Julea, A. Analyzing big remote sensing data via symbolic machine learning. In Proceedings of the 2016 conference on Big Data from Space, Santa Cruz de Tenerife, Spain, 15–17 March 2016.
25. Zhang, J.; Li, P.; Wang, J. Urban built-up area extraction from Landsat TM/ETM+ images using spectral information and multivariate texture. *Remote Sens.* **2014**, *6*, 7339–7359. [[CrossRef](#)]
26. Guindon, B.; Zhang, Y.; Dillabaugh, C. Landsat urban mapping based on a combined spectral-spatial methodology. *Remote Sens. Environ.* **2004**, *92*, 218–232. [[CrossRef](#)]
27. Li, P.; Cheng, T.; Guo, J. Multivariate image texture by multivariate variogram for multispectral image classification. *Photogramm. Eng. Remote Sens.* **2009**, *75*, 147–157. [[CrossRef](#)]
28. Pesaresi, M.; Gerhardinger, A.; Kayitakire, F. A robust built-up area presence index by anisotropic rotation-invariant textural measure. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2008**, *1*, 180–192. [[CrossRef](#)]
29. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [[CrossRef](#)]

30. Brodersen, K.H.; Ong, C.S.; Stephan, K.E.; Buhmann, J.M. The balanced accuracy and its posterior distribution. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 3121–3124.
31. Powers, D.M.W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Int. J. Mach. Learn. Technol.* **2011**, *2*, 37–63.
32. Card, D.H. Using known map category marginal frequencies to improve estimates of thematic map accuracy. *Photogramm. Eng. Remote Sens.* **1982**, *48*, 431–439.
33. Pesaresi, M.; Ehrlich, D.; Caravaggi, I.; Kauffmann, M.; Louvrier, C. Towards global automatic built-up area recognition using optical VHR imagery. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2011**, *4*, 923–934. [[CrossRef](#)]
34. Ben-Dor, E.; Levin, N.; Saaroni, H. A spectral based recognition of the urban environment using the visible and near-infrared spectral region (0.4–1.1 μm). A case study over Tel-Aviv, Israel. *Int. J. Remote Sens.* **2001**, *22*, 2193–2218. [[CrossRef](#)]
35. Freire, S.; Kemper, T.; Pesaresi, M.; Floczyk, A.; Syrris, V. Combining GHSL and GPW to improve global population mapping. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Milan, Italy, 26–31 July 2015; pp. 2541–2543.
36. Deichmann, U.; Ehrlich, D.; Small, C.; Zeug, G. *Using High Resolution Satellite Data for the Identification of Urban Natural Disaster Risk*; World Bank: Washington, DC, USA, 2011.
37. Kemper, T.; Mudau, N.; Mangara, P.; Pesaresi, M. Towards an automated monitoring of human settlements in South Africa using high resolution SPOT satellite imagery. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *XL-7/W3*, 1389–1394. [[CrossRef](#)]
38. Jenerowicz, M.; Kemper, T.; Soille, P. An automated procedure for detection of IDP's dwellings using VHR satellite imagery. *Proc. SPIE* **2011**. [[CrossRef](#)]
39. Kemper, T.; Jenerowicz, M.; Gueguen, L.; Poli, D.; Soille, P. Monitoring changes in the Menik Farm IDP camps in Sri Lanka using multi-temporal very high-resolution satellite data. *Int. J. Digit. Earth* **2011**, *4*, 91–106. [[CrossRef](#)]
40. Gueguen, L.; Pesaresi, M.; Gerhardinger, A.; Soille, P. Characterizing and Counting Roofless Buildings in Very High Resolution Optical Images. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 114–118. [[CrossRef](#)]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).