



**HAL**  
open science

## A Semantic-Based Approach for Landscape Identification

Anne Toulet, Emmanuel Roux, Anne-Elisabeth Laques, Eric Delaître, Laurent Demagistri, Isabelle Mougenot

► **To cite this version:**

Anne Toulet, Emmanuel Roux, Anne-Elisabeth Laques, Eric Delaître, Laurent Demagistri, et al.. A Semantic-Based Approach for Landscape Identification. Bruno Pinaud; Fabrice Guillet; Fabien Gandon; Christine LARGERON. Advances in Knowledge Discovery and Management, 834, Springer, pp.119-136, 2019, Studies in Computational Intelligence, 978-3-030-18129-1. 10.1007/978-3-030-18129-1\_6 . hal-02164143

**HAL Id: hal-02164143**

**<https://hal.science/hal-02164143>**

Submitted on 24 Jun 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A semantic-based approach for landscape identification

Anne Toulet and Emmanuel Roux and Anne-Élisabeth Laques and Éric Delaître  
and Laurent Demagistri and Isabelle Mougenot

**Abstract** Here we present an original method for the automation of landscape identification in a satellite image. There are two major challenges in this process. The first lies in the ability to take all expert knowledge into account for the full time it takes to analyze the image. The second is successfully structuring and persisting this knowledge so that it becomes interoperable and usable in the Semantic Web context. In this paper, we explain how the combination of several strategies associating image processing, the calculation of specific characteristics and inductive logic programming (ILP) can feed into the automation process, and how the integration of knowledge via the construction of dedicated ontologies can meet these challenges.

## 1 Introduction

Particular emphasis is placed upon successful landscape identification in satellite images, since the study of landscapes and their evolution over time is one approach

---

Anne Toulet

A. Toulet, LIRMM, University of Montpellier, France, e-mail: [anne.toulet@lirmm.fr](mailto:anne.toulet@lirmm.fr)

Emmanuel Roux

E. Roux, IRD, Montpellier, France, e-mail: [emmanuel.roux@ird.fr](mailto:emmanuel.roux@ird.fr)

Anne-Élisabeth Laques

A.-E. Laques, IRD, Montpellier, France, e-mail: [anne-elisabeth.laques@ird.fr](mailto:anne-elisabeth.laques@ird.fr)

Éric Delaître

E. Delaître, IRD, Montpellier, France, e-mail: [eric.delaitre@ird.fr](mailto:eric.delaitre@ird.fr)

Laurent Demagistri

L. Demagistri, IRD, Montpellier, France, e-mail: [laurent.demagistri@ird.fr](mailto:laurent.demagistri@ird.fr)

Isabelle Mougenot

I. Mougenot, University of Montpellier, France

e-mail: [isabelle.mougenot@umontpellier.fr](mailto:isabelle.mougenot@umontpellier.fr)

of addressing major social, economic and environmental challenges [11]. To identify landscapes, the geographer can use a variety of techniques: field missions, aerial photographs, mapping tools and satellite imagery. A geography expert will use all of these approaches to characterize different types of landscape and draw up a map. It is, however, a very time-consuming task. Satellite imagery is currently used as a faster and more focused way of obtaining much of this information. Access to images is facilitated by the exponential growth in their production and much wider availability. It is actually very difficult to extract knowledge automatically from satellite images. Indeed, satellite image analysis techniques remain limited and this work cannot be done without human input. One of our goals is therefore to facilitate the geographers task by automating landscape extraction from satellite images. This goal is a very big challenge: it is very complex to reproduce the different stages carried out by the geographer. To achieve this, it is necessary to implement several strategies that use knowledge from the subject area (landscape geography), satellite imaging and inductive logic programming. In this article, we present our approach and implement it to automatically extract landscapes on an area of the Brazilian Amazon. The second goal is to structure and persist all this knowledge via dedicated ontological modules so that it is interoperable and reusable in the context of the Semantic Web. Thus, the use of ontologies makes it possible to initiate an approach towards the production of FAIR data: Findable, Accessible, Interoperable and Reusable [33]. Satellite images have been subject to different studies involving ontologies e.g. [6, 8, 13]. These ontologies actually play a wide variety of roles. The high volume of data and the exclusively digital nature of the data extracted from the image means that ontologies have to be constructed and regularly used to supplement other approaches, for example with supervised classification or clustering. This point will be detailed in the section 2. In our study, ILP is used instead of the conventional classification techniques. By opting for ILP, we are able to apply an approach that combines statistical learning and logic programming and thus formulate hypotheses that involve several variables simultaneously to describe objects in our image. Another reason for our choice is that we can work - at ILP level and with the ontologies - with formalisms that are all underpinned by predicate logic, even though the calculating processes differ. The paper proceeds as follows. In section 2, we first present related works and explain the similarities and the differences with our work. In section 3, we present the approach we applied to achieve our objectives: we first set out the geographical context and the preliminary processing for satellite images; then we explain how landscapes are extracted using ILP; and finally, we describe how these different stages feed knowledge into the integration process via specific ontologies: one dedicated to geography (landscape ontology) and the other to data taken from the image (image ontology). In section 4, we describe a practical application to present our approach. For a given satellite image, we explain the preliminary processing and calculations required before describing how classification rules are obtained to categorize the landscapes in the image. We will also show how we use this processing to populate the image ontology. In section 5, we will go over the results obtained for landscape prediction and for the role played by the ontologies. In our conclusion, we discuss the perspectives for this type of work.

## 2 Review of related works

### 2.1 *Satellite imagery domain*

Making sense of data that take the form of raster images requires well-structured thinking that may be aided by logic-based approaches. Because of key goals to discover new associations between data and to learn reasoning patterns, some existing works [5] adequately exploit ILP techniques due to the natural representation of relation and the use of background theories. Some other works [1, 2, 7] take benefit of ontologies and associated reasoning capabilities to extract meaning from the content of images. Acquiring new knowledge is very demanding especially within the context of satellite images and can be partially automated by applying a combination of machine learning techniques with ontology to improve the performances of these techniques. We aim to use an inductive and deductive approach that will lead to a mixed knowledge-based strategy to automate the classification of different landscapes from satellite imagery. We try to investigate whether a knowledge base that is incomplete by definition can be automatically enriched using ILP. Additionally, a major challenge is to extract knowledge from satellite images that would meet the experts requirements as experts views have become a matter of real concern. A few works [14, 26] combine machine learning and deductive reasoning to boost creating, adapting and generalizing knowledge within the context of remote sensing. To our knowledge, there is no consistent approach using both ontologies and ILPs as in the work presented here.

### 2.2 *Life sciences domain*

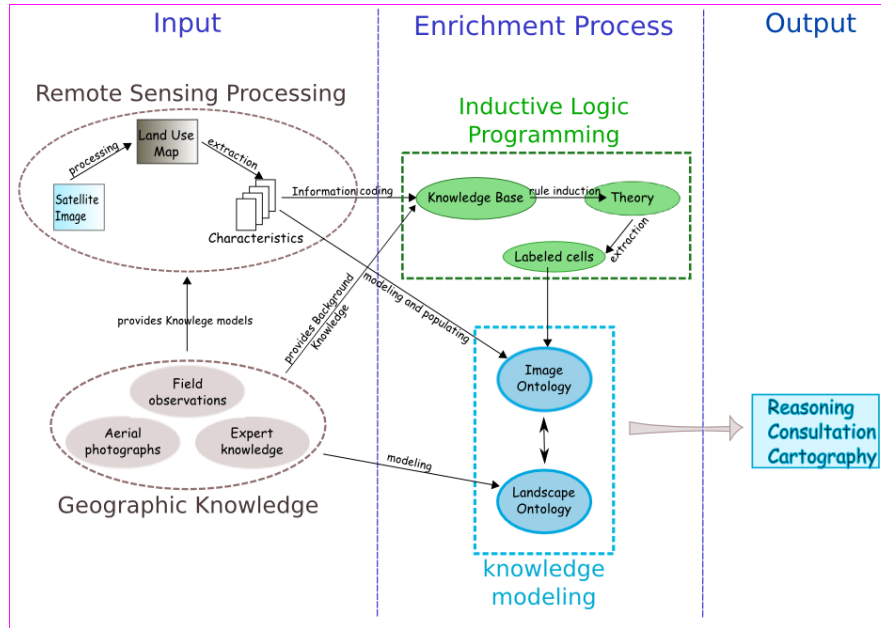
At the opposite, in life sciences, different approaches combine in a very effective manner ILP and ontology-based methods to acquire new knowledge and to validate some biological hypothesis [18, 25, 29, 31]. Existing controlled vocabularies in life sciences as Gene Ontology are of importance and constitute a real help to build such approaches. Although the approaches are similar, it is difficult to compare results in areas as different as satellite imagery and biomedicine.

This state of the art shows that our approach to extract knowledge in satellite image combining ontologies and ILP is quite new and original.

## 3 Methodological approach

The different steps of our work are summarized in Fig. 1. In this diagram, we explain the global approach that we implemented to move from a raw satellite image to services exploitable by the geographer. To achieve clarity and accuracy, we divide

our approach in three parts: the input (data & knowledge), the enrichment process and the output.



**Fig. 1** Overall approach

The first part (“Input”) concerns all data and knowledge used to feed the process as well as the preliminary processing of satellite images. By preliminary processing, we mean classical remote sensing treatments. The expertise provided by the geographer guides these treatments and the calculations are built to replicate landscape metrics. The second step (“Enrichment process”) describes how to extract knowledge from the pre-processed satellite image, using two complementary methods: Inductive Logic Programming (ILP) and knowledge modeling. The geographic knowledge is modelled in the landscape ontology. The results of the calculations made in the preliminary processing of the satellite image are coded so that they can be exploited by ILP to produce classification rules that can then be applied to identify landscapes in the image. What is more, the characteristics obtained by the calculations are used to populate the image ontology. What is original about this method is that it combines two symbolic approaches: an inductive approach (ILP) and a deductive approach (ontology-based reasoning). The geography expert intervenes at two levels: he is involved in the construction of the ILP Knowledge Base and provides the necessary elements for ontological modeling. The last step (“Output”) shows what you get at the end of this process. At this point, it is possible to offer services adapted to the geographer’s needs: reasoning, consultation and

cartography for example. Each of these steps will be described in the following sections: first, from a generic point of view, and then illustrated in an example.

### ***3.1 Input data & knowledge***

We begin by setting out the context of landscape geography as envisaged in our study and the preliminary processing applied to the image to fulfill our requirements.

#### **3.1.1 Geographical context.**

Technically speaking, landscape identification uses a number of criteria, including composition and configuration, measured using landscape metrics applied to landscape components [12]. At this level of observation, Landscape Components (LC) refer to a specific land use or what is known as the “patch” in ecology. It is the relationships between the various components that determine the classification of a Landscape Type (LT). The landscape metrics used in this study are classical in this context [10], such as composition, diversity or fragmentation. As an example, a forest ecosystem landscape may be made up of various landscape components in varying proportions and with specific configuration (terra ferme forest 90%, gallery forest 6%, secondary vegetation (wild land and fallow land) 2%, cassava field 2%). One of the difficulties lies in the fact that, in satellite images, there are no objects “naturally” identifiable as landscape components to which landscape metrics could be applied. Here we will explain the preliminary processing required to achieve this.

#### **3.1.2 Preliminary processing of satellite images.**

The aim of the processing work is to obtain objects that can be likened to landscape components. The first step is therefore to produce a land use map that will be used to establish this kind of object in the image. Land use labels are used to categorize objects that can be likened to landscape components, such as “Water”, “Vegetation” and “Forest” [28]. However, the semantic level attained by this initial output is not high enough to take landscape structure into consideration. Most importantly, we are unable to apply landscape metrics directly to the objects obtained. Two questions are raised: how do you define the areas to which metrics should be applied, and what calculations should be defined in relation to landscape metrics? To respond to the first question, we place a grid on the image. The grid comprises a set of regular, square-shaped cells, the size of which is defined according to need. This procedure lets us identify well-defined arbitrary zones in the image, to which the metrics can be applied. When it comes to metrics, multiple tools are already included in remote

sensing image processing software, such as ENVI<sup>1</sup> or GRASS<sup>2</sup>, and they can be used to make calculations for landscape metrics. We therefore replicate the geographers approach by applying these metrics to “landscape components” contained in each cell.

## 3.2 *Enrichment process*

### 3.2.1 Landscape extraction using ILP

Inductive logic programming is used to obtain landscape classification rules. In this section, we explain how these rules are produced.

Developing rules through automatic learning.

Inductive Logic Programming (ILP) was selected to induce rules capable of ‘predicting’ the classification of a cell as a landscape type, based on the landscape characteristics found in that cell. ILP was introduced by Muggleton [24] and is an automatic learning method using examples (or a supervised learning method) and the formalism of first-order logic. As such, it is especially well suited to the discovery of relational knowledge within data. The value of this kind of method has been demonstrated for the identification of classification rules in the field of geographic and remote sensing data interpretation and processing [21, 32].

Inductive Logic Programming (ILP): basic principles.

Inductive Logic Programming [20] works with

- a knowledge base  $B$  expressed in first-order logic and describing a body of knowledge and a set of constraints;
- a set of examples  $E$  split into two subsets  $E^+$  and  $E^-$  respectively, corresponding to positive and negative examples;
- a description language  $L$ .

ILP generates a “theory” (*i.e.* a set of rules),  $H$ , using the description language,  $L$ , and i) must cover positive examples  $E^+$  and ii) must not cover negative examples  $E^-$ . Among the existing inductive systems, the *Aleph* system [30] has been selected.

---

<sup>1</sup> Environment for Visualizing Images, the software commercialized by EXELIS

<sup>2</sup> GRASS GIS, GIS freeware from the GRASS Development Team

### Information coding within the knowledge base.

Each landscape metric described above corresponds to a predicate in the knowledge base, i.e. a function that returns a response (true or false, or a constant) when adequate information is input. For example, the metric *proportion of forest* corresponds to the binary predicate `forest_proportion/2`, applicable to a cell and returning the value of the proportion of forest in that cell. However, ILP can only reason with symbolic information. However, landscape metrics provide numerical data which therefore have to be recoded. The recoding method selected here is that put forward by [17]: for a given numerical data item, each value observed is recoded using inequalities that call on the characteristic values of the variable considered. In our application, these characteristics are the minimum (min), the 10th, 20th, ..., 90th percentiles and the maximum (max) of the variable considered. Hence, each landscape metric corresponds to a predicate that can be used to test 20 inequalities. In the Prolog language used by the Aleph system, an inequality can be written as follows:

```
forest_proportion_qual(M, ≤ max):- forest_proportion(M,X), X ≤ max
```

This coding is a good compromise between the loss of inherent information that occurs when converting numerical data to symbolic information, and the capacity for generalization during the learning process. Furthermore, it highlights significant value ranges [5].

### Example base and rule induction.

Given the goals set (*i.e.* the predicted classification of a given cell as a landscape type), ILP requires knowledge of the landscape characteristics of a significant number of cells, and the landscape type that they belong to. A cell that is associated with landscape metric values, and whose landscape type is known, therefore provides an example of a training set. It should be noted that a priori knowledge of landscape type classification can come from various sources; here it comes from a geographer's expertise. The learning process is launched for each individual landscape type. For a given landscape type, the cells belonging to that type are positive examples, and the remaining cells provide negative examples. The minimum accuracy of the rules is set at 0.7; it is established during the learning process using the ratio  $p/(p+n)$ , where  $p$  and  $n$  correspond to the number of examples, respectively positive and negative, covered by the rule. This value is seen as a good compromise between the conflicting requirements of accuracy and of generalization of the rules induced. Finally, the maximum length of the premises of the rules is set at five literals because, in practice, this kind of value is considered to be the limit enabling correct reading of the results [23].



### 3.2.2 Constructing the ontologies

An ontology can be used to model a specific area of knowledge in a formal, structured manner [16]. This modelling is based on the definition of concepts and relations between those concepts. One of the stated objectives is to enable knowledge sharing via a representation of an area of knowledge that is as consensual as possible. To define our ontologies, we use description logics, which are a language family able to represent a knowledge base for a given area. In description logics, two components make up the knowledge base: the terminological box (or *TBox*), which models the concepts, and the assertional box (or *ABox*), which represents instances. The description logics draw on first-order logic and result in inferential reasoning. These mechanisms are used to enrich the knowledge base by specifying new facts deduced from the original items through reasoning. We then chose to operationalize these formal ontologies in OWL2 (Web Ontology Language) [9], the language recommended by the W3C.

Landscape ontology.

Landscape ontology is the ontology of an area of knowledge, reflecting the geographer's expertise. It models concepts that depend on classification (landscape type and landscape component) in the form of a tree structure expressing the specialization of the concepts (the subsumption relationship) and all the properties linking or defining those concepts (the meronymy relationship and metrics). The model is then formalized using description logics. Let us take the deliberately simplified example of a "forest ecosystem" landscape type. This kind of landscape must contain the "forest" landscape component, but it may also include other component types. The landscape is also characterized by its low diversity and limited heterogeneity. In this example, we can see the concepts of landscape type and landscape component, each specialized by a sub-concept (here, forest ecosystem and forest, respectively). Here, we also need the properties "diversity" and "heterogeneity", defined as functional properties in the mathematical sense. More generally speaking, the landscape metrics defined in the ontology are all functional properties. In description logics, we will write:

$$\begin{aligned}
 & \text{ForestEcosystem} \sqsubseteq \text{LandscapeType} \\
 & \text{Forest} \sqsubseteq \text{LandscapeComponent} \\
 & \left. \begin{array}{l} \top \sqsubseteq \leq 1 \text{ hasDiversity} \\ \top \sqsubseteq \leq 1 \text{ heterogeneity} \end{array} \right\} \text{ functional properties} \\
 & \text{ForestEcosystem} \equiv \text{LandscapeType} \sqcap \exists \text{hasPart.Forest} \\
 & \quad \sqcap \exists \text{hasDiversity}.\{\text{lowDiversity}\} \\
 & \quad \sqcap \exists \text{heterogeneity}.\{\text{limitedHeterogeneity}\}
 \end{aligned}$$

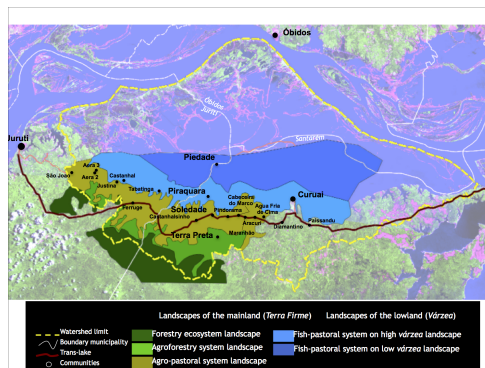
## Image ontology.

The image ontology is constructed to take into account various knowledge taken from the image, in connection with the landscape ontology. It is built as an application ontology, *i.e.* it is seen as a container for the cell descriptors obtained via the tasks run by the system. The concepts and relations in this ontology are obtained from preliminary processing and the calculation of the landscape characteristics. Most notably, it includes the key concept of 'cell'. The characteristics are assigned literal values. In section 4.3, we will see how this ontology is populated.

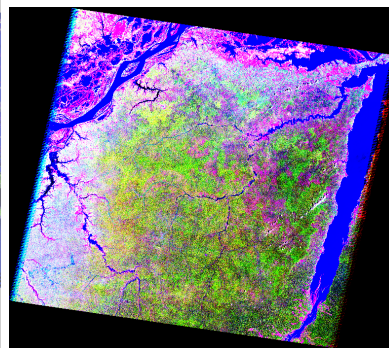
## 4 Application

### 4.1 Data

To illustrate the method, we have taken a Landsat 5 image acquired on 23 October 2009, covering part of Amazonia situated in the state of Pará, at the confluence of the Amazon and Río Tapajós. Five different types (LT1 to LT5) of landscape have been described and geolocated (Fig.2). They are input into the example database used to generate rules via ILP. The aim is to search for these landscapes within the image (Fig. 3), applying the strategy set out here above.



**Fig. 2** The 5 labelled landscapes.

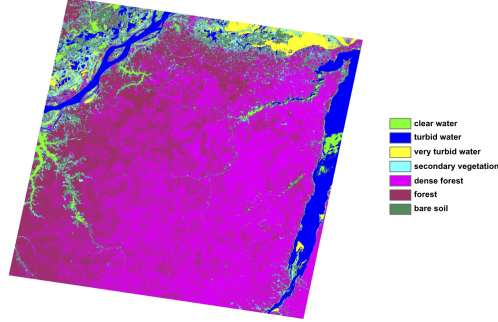


**Fig. 3** Lansat Image 2009.

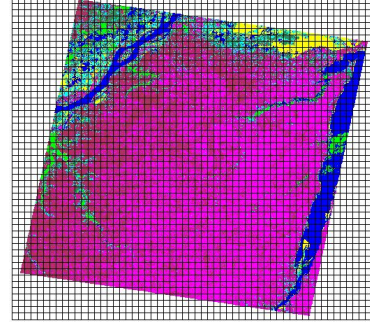
#### 4.1.1 Production of the land use map and grid.

There are different ways of producing land use maps [22]. We chose unsupervised classification (or clustering) using the k-means algorithm. The land use map ob-

tained is made up of seven labelled classes (Fig. 4) over which we lay a grid of square cells that are then used to calculate the landscape characteristics (Fig. 5).



**Fig. 4** Land use map.



**Fig. 5** Grid.

## 4.2 Production of the classification rules

To obtain the classification rules, we need to exploit the resulting characteristics, calculated on the basis of the cells used in the learning process. These cells are those contained in the five reference landscapes. The choice of metrics depends on the approach taken by the geographer. We establish in Table 1 the correspondences between the landscape metrics and the characteristics calculated in the satellite image:

**Table 1** Equivalence between landscape metrics and characteristics in the image.

Expert geographer	Characteristics calculated in the cells
Composition	Proportion of each class present
Diversity	Specific richness (no. of classes present)
Heterogeneity	Shannon index $H = -\sum_{k=1}^N p_k \cdot \ln(p_k)$
Fragmentation: patch surface area	Average surface area of objects
Shape	Perimeter/area: average measured

### 4.3 Populating the image ontology

As we have seen above, the image ontology models the cells and the properties that correspond to landscape characteristics. Each of these properties is used to associate a cell with the value obtained from the calculation of a given characteristic. We thus obtain information calculated cell by cell, which then instantiates the ontology by populating the *ABox*. We can take the example of an individual from the cell class - identified by m59 - which, as a proportion of the forest, has a value of 0.287. In logical form, this is written as follows:

```

Cell  $\sqsubseteq$  Top
Cell (m59)
 $\top \sqsubseteq \leq 1$  forest_prop (functional property)
forest_prop (m59, "0.287")

```

As such, all cells in the ontology are described by a set of attributes that correspond to the metric values calculated during preliminary processing. Each cell is georeferenced using latitude/longitude centroid coordinates, as well as its geometrical representation as a polygon. In GeoSPARQL<sup>3</sup>, the polygon has a literal value expressed as geo:gmlLiteral. In addition, the classification rules specify the cell's landscape type: this information is added to the cell description. In the image ontology, we therefore have a very comprehensive range of information. We can also note that the structure of the ontology means we can add as many additional attributes as we like, for example texture or vegetation index.

## 5 Results and discussion

### 5.1 Results

#### 5.1.1 Characteristics of the ruleset and performance.

Six grids were generated, made up of square cells with sides measuring from 32,640 m to 1,020 m (system of nested cells for a multi-resolution approach). The large grids (32,640 m, 16,320 m and 8,160 m) proved to be poorly suited to the surface area of the landscapes making up the learning base. Calculations were therefore made on the grids measuring 1,020, 2,040 and 4,080. The most useful results were obtained for the 1,020 and 2,040 grids, with little difference between them. Here we present the 2,040 grid. Respectively 1, 3, 1, 6 and 4 rules for landscape types LT1 to LT5 were induced for the full learning set made up of 300 cells with sides measuring 2,040 metres. Here is an example of the kind of rule obtained: *landscape\_type(A,tp5) :- turbidWater\_prop(A,eq0), forest\_prop(A,le0.009)*.

This rule means that a cell A belongs to LT5 if its proportion of turbid water is zero and its proportion of forest is below 0.009. While the rules induced are easy to read and intelligible, even for someone who is not expert in the learning method, the

<sup>3</sup> geospatial standard defined by OGC (Open Geospatial Consortium)

prediction results vary according to the landscape type considered. Table 2 shows sensitivity and specificity<sup>4</sup>, calculated using the full learning set and via crossed validation (average, minimum and maximum values). The accuracy of the predic-

**Table 2** Predication results for the 2,040-metre resolution grid.

		TP1	TP2	TP3	TP4	TP5
Full learning set	Sensitivity	0,98	0,73	0,97	0,93	0,96
	Specificity	0,98	1,00	0,97	0,99	0,99
Crossed validation average	Sensitivity	0,76	0,15	0,64	0,46	0,62
	[min; max]	[0,40; 1,00]	[0,00; 1,00]	[0,33; 0,83]	[0,30; 0,62]	[0,25; 0,80]
[min; max]	Sensitivity	0,92	0,92	0,82	0,75	0,83
	Specificity	[0,82; 1,00]	[0,80; 1,00]	[0,73; 0,89]	[0,60; 0,90]	[0,75; 0,94]

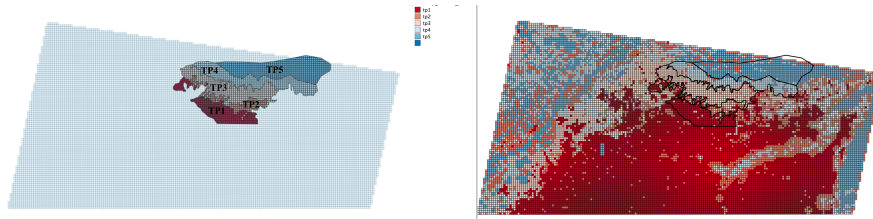
tion calculated using the full learning set comes to 94.7%. Stratified 10-fold cross-validation resulted in an estimated 58.6% precision accuracy with a maximum (and a minimum) of 71.0% for the ten subsets (respectively 45.7%). The results obtained for sensitivity are particularly good, reaching 76.3% for landscape type 1 and exceeding 60% for types 3 and 5 (respectively 63.8 and 61.9%). Landscape type 2 is especially difficult to predict with a success rate of just 15%, which is mainly explained by the limited number of related examples (15), but also by less clear-cut identification and characterization by the expert (in his own view). Other tests were run to produce rules using ILP. The spatial adjacency relationship, where two cells are only adjacent if they share a side, was thus coded in the learning process knowledge base, along with the multi-resolution character (inclusion relationships from one cell to another). This final learning strategy, exploiting contiguity information and the multi-resolution character of the spatial information, appears particularly relevant: it corresponds to the way in which the expert geographer comprehends the territory, using different mapping products with varying geographic reach and resolution. However, the inclusion of spatial relations and multi-resolution to produce classification rules did not demonstrate any real improvement in results.

Figure 6 shows the landscapes that feed the learning base for ILP, and Figure 7 illustrates the prediction results for the complete study area. The prediction results, most of which fall outside the reference landscape zone, lead to consistent characterization of the territory, the relevance of which is confirmed a posteriori by the expert.

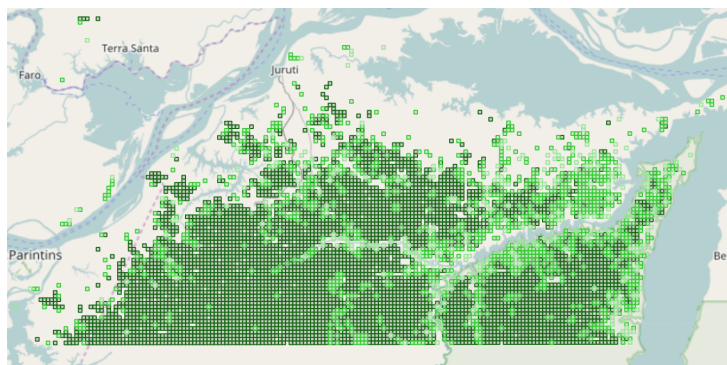
### 5.1.2 Ontology modules.

Two ontology modules were built. The landscape ontology that "captures" the geographer's work and thus represents the conceptual modelling of an area of knowledge. And the image ontology that contains all the information obtained from the satel-

<sup>4</sup> **Sensitivity**: capacity of the classifier to predict belonging to an LT class, given that the objects in question belong to that LT / **Specificity**: capacity of the classifier to predict non-belonging to an LT class, given that the objects in question do not belong to that LT.

**Fig. 6** Learning landscapes.**Fig. 7** Prediction results.

lite image and the ILP activity: georeferenced cells, landscape characteristic values and prediction results. The image ontology also extends the core ontology found in GeoSPARQL [4], facilitating coverage of the cell geometry. Some links connecting the two ontological modules have already been explored. Initial work involved developing synergies between the geographer’s expertise concerning a landscape type, as described in the landscape ontology, and the information on the cells classified by ILP as belonging to that landscape type. As such, we define adequate CONSTRUCT queries in SPARQL language [19], which produce new representations of the knowledge contained in the image ontology, making it more consistent with the content of the landscape ontology. Depending on the data searched, the information can be exploited in several ways. We can then look into what identifies the transition from one landscape to another by focusing on the boundary cells. The screen shot shown in Figure 8 shows the result of a CONSTRUCT query (LT1 type cells predominantly containing “forest” or “sparse forest” type landscape components (or patches) that we translate as cells in the forest ecosystem. These cells are shown in green in the study zone<sup>5</sup>. In the landscape ontology, the forest ecosystem is described as a landscape that is predominantly covered by forest, with little diversity. The different shades of green illustrate the diversity of the patches found in the cells.

**Fig. 8** Forest ecosystem cells on an OSM background map

<sup>5</sup> Visualization makes use of the Javascript OpenLayers library

The forest component is omnipresent in the dark green cells. The light green cells show greater diversity with intermediate patches of water or vegetation. The cells with a richer range of landscape components are frequently found at the boundaries of the forest ecosystem landscape and diversity is therefore a good indicator of a change of landscape, at least where the forest ecosystem is concerned. The content of the two ontological modules can also be linked to open external data sources, also represented using Semantic Web standards. Hence, in Figure 8, the cells are viewed on an *Openstreetmap* background map [27]. The geographical entities defined in OSM may be exploited via their RDF<sup>6</sup> representation [3] to better clarify the contours of the landscapes. Likewise, the ecological data can also contribute to better comprehension of landscapes through their ecosystem aspects. As such, data that comes from the observation of occurrences of living species can also be integrated into landscape characterization. We provide the example of data on the study zone found on the *GBIF* portal [15].

## 5.2 Discussion

The first objective was to automate landscape identification in a satellite image as far as possible. The second objective was to express and persist all that knowledge via dedicated ontological modules so that it is interoperable and reusable in the context of the Semantic Web. These objectives were attained by combining different approaches, most notably associating ILP and ontologies. Furthermore, we have obtained data that has been greatly enriched compared to the image's initial content and restructured in Semantic Web format. As we have demonstrated, this restructuring opens up possibilities for the exploitation of the web of data. All these results thus validate our approach. The method now needs perfecting and we have a number of proposals in this respect, for example:

- refining landscape description (the geographer's work) to improve the learning base used by ILP to produce the rules;
- in the process of obtaining classification rules, making the most of the topological relations and cell multi-resolution;
- in ontological modelling, using adjacency relations between cells by making use of the GeoSPARQL model;
- using other open data sources, for example GeoNames<sup>7</sup>.

We have demonstrated all the potential of an approach that takes symbolic induction mechanisms to partially populate an applicative ontology. The usual data mining methods applied when processing an image, for example supervised classification, treat the input attributes as dimensions that are independent of one another. In this respect, ILP, which establishes dependencies between attributes, is better equipped

---

<sup>6</sup> OSM data was previously collected in RDF format in the LinkedGeoData project

<sup>7</sup> GeoNames : <http://www.geonames.org/>

to take advantage of the spatial organization of radiometric values within the cells seen as a set of connected pixels. Spatial analysis requires the processing of spatial entities according to their descriptors but also according to the descriptors of the adjacent entities and the relationships with those adjacent entities. We were thus able to extract the predicates involving several descriptors (for example, the proportion of each landscape component, diversity or fragmentation) to analyze landscape types as closely as possible. However, it now remains for us to enrich the ILP activities by including topological relationships, either within the same grid of cells or within nested cells. Likewise, we will continue to develop the cell and landscape ontologies. Most notably, the spatial relationships between cells will be described within the cell ontology and we also expect to work on new descriptors (such as texture). Finally, the links used to exploit the application ontology alongside the landscape ontology will be reinforced. The aim is to continue to develop conversion rules using SPARQL so that we are able to take a regularity on the cell ontology graph and extract a regularity from the landscape ontology graph, and vice versa. This bidirectional approach will enable us, on the one hand, to reconcile image data with the geographer's expertise and, on the other hand, potentially further the expert's knowledge according to the image's content.

## 6 Conclusion

We have presented a hybrid approach that draws on both expert knowledge formalized in ontological modules and new knowledge developed through symbolic learning. The goal is to work as closely as possible to the geography expert's reality while developing knowledge acquired via ILP. We have set out some encouraging initial results on the ability to automate the extraction of interesting patterns in a satellite image. We intend to enrich the ontological representation, especially in terms of metric description to develop the potential of reasoning mechanisms. The approach described in this paper can be improved upon as concerns the precision of the results, but it has helped us fulfil our objectives: achieving a shift towards automated landscape extraction from a satellite image while structuring, perpetuating, enriching and establishing interoperability of the knowledge and data exploited throughout the process, clearly tailoring it for the Semantic Web.

**Acknowledgements** This work was achieved within the COCLICO project funded by the French National Research Agency (grant ANR-12-MONU-0001).

## References

1. Andrés, S., Arvor, D., Mougénot, I., Libourel, T., Durieux, L.: Ontology-based classification of remote sensing images using spectral rules. *Computers & Geosciences* **102**, 158 – 166 (2017).



- DOI <https://doi.org/10.1016/j.cageo.2017.02.018>. URL <http://www.sciencedirect.com/science/article/pii/S0098300417302078>
2. Arvor, D., Durieux, L., Andrés, S., Laporte, M.A.: Advances in Geographic Object-Based Image Analysis With Ontologies: a review of main contributions and limitations from a remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing* **82**, 125–137 (2013). URL <https://hal.archives-ouvertes.fr/hal-01102197>
  3. Auer, S., Lehmann, J., Hellmann, S.: Linkedgeodata: Adding a spatial dimension to the web of data. In: *Proceedings of the 8th International Semantic Web Conference, ISWC '09*, pp. 731–746. Springer-Verlag, Berlin, Heidelberg (2009). DOI 10.1007/978-3-642-04930-9\_46. URL [http://dx.doi.org/10.1007/978-3-642-04930-9\\_46](http://dx.doi.org/10.1007/978-3-642-04930-9_46)
  4. Battle, R., Kolas, D.: Enabling the geospatial semantic web with parliament and geosparql. *Semant. web* **3**(4), 355–370 (2012). URL <http://dl.acm.org/citation.cfm?id=2590208.2590211>
  5. Bayouhd, M., Roux, E., Richard, G., Nock, R.: Structural knowledge learning from maps for supervised land cover/use classification: Application to the monitoring of land cover/use maps in french guiana. *Computers and Geosciences* **76**, 31–40 (2015). DOI 10.1016/j.cageo.2014.08.013. URL <https://hal.archives-ouvertes.fr/hal-01369822>
  6. Belgiu, M., Thomas, J.: Ontology based interpretation of very high resolution imageries—grounding ontologies on visual interpretation keys. *AGILE 2013Leuven* **2013**(1), 14–17 (2013)
  7. Belgiu, M., Tomljenovic, I., Lampoltshammer, T.J., Blaschke, T., Hfle, B.: Ontology-based classification of building types detected from airborne laser scanning data. *Remote Sensing* **6**(2), 1347–1366 (2014). URL <http://www.mdpi.com/2072-4292/6/2/1347>
  8. Chahdi, H., Grozavu, N., Mougnot, I., Bennani, Y., Berti-Equille, L.: Towards ontology reasoning for topological cluster labeling. In: *International Conference on Neural Information Processing*, pp. 156–164. Springer International Publishing (2016)
  9. Dean, M., Schreiber, G.: OWL Web Ontology Language - Reference. W3C recommendation, W3C (2004). URL <http://www.w3.org/TR/2004/REC-owl-ref-20040210/>
  10. Enserink, M.: Biological invaders sweep in. *Science* **285**(5435), 1834–1836 (1999). DOI 10.1126/science.285.5435.1834. URL <http://science.sciencemag.org/content/285/5435/1834>
  11. Fahrig, L.: Effects of habitat fragmentation on biodiversity. *Annual review of ecology, evolution, and systematics* **34**(1), 487–515 (2003)
  12. Farina, A.: *Landscape Ecology in Action*. Springer Netherlands (2000)
  13. Forestier, G., Puissant, A., Wemmert, C., Ganarski, P.: Knowledge-based region labeling for remote sensing image interpretation. *Computers, Environment and Urban Systems* **36**(5), 470 – 480 (2012). DOI <http://dx.doi.org/10.1016/j.compenvurbsys.2012.01.003>. URL <http://www.sciencedirect.com/science/article/pii/S019897151200004X>
  14. Forestier, G., Wemmert, C., Puissant, A.: Coastal image interpretation using background knowledge and semantics. *Computers & Geosciences* **54**, 88 – 96 (2013)
  15. GBIF: Global biodiversity informatics facility (2016). URL <http://gbif.org/>. [Online: <http://gbif.org/>]
  16. Gruber, T.: A Translation Approach to Portable Ontology Specification. *Knowledge Acquisition* **5**(2), 199–220 (1993). DOI 10.1006/knac.1993.1008
  17. Kavurucu, Y., Senkul, P., Toroslu, I.H.: A comparative study on ilp-based concept discovery systems. *Expert Systems with Applications* **38**(9), 11,598 – 11,607 (2011)
  18. King, R.D.: Applying inductive logic programming to predicting gene function. *AI Mag.* **25**(1), 57–68 (2004)
  19. Kostylev, E.V., Reutter, J.L., Ugarte, M.: Construct queries in sparql. In: *ICDT* (2015)
  20. Lavrac, N., Dzeroski, S.: *Inductive Logic Programming: Techniques and Applications*. Ellis Horwood (1994)
  21. Malerba, D., Esposito, F., Lanza, A., Lisi, F., Appice, A.: Empowering a gis with inductive learning capabilities: the case of ingens. *Computers, Environment and Urban Systems* **27**(3), 265 – 281 (2003). DOI 10.1016/S0198-9715(02)00024-8

22. Masse, A.: Dveloppement et automatisaion de mthodes de classification partir de sries temporelles d'images de tldtection - application aux changements d'occupation des sols et l'estimation du bilan carbone. Ph.D. thesis, Universit P. Sabatier, Toulouse (2013)
23. Michalski, R.S.: Machine learning: An artificial Intelligence Approach, chap. a theory and methodology of inductive learning, pp. 110–161. TIOGA Publishing Co. (1983)
24. Muggleton, S.: Inductive logic programming. *New Generation Computing* **8**, 295–318 (1991)
25. Nguyen, H., Luu, T.D., Poch, O., Thompson, J.D.: Knowledge discovery in variant databases using inductive logic programming. In: *Bioinformatics and biology insights* (2013)
26. Nieland, S., Moran, N., Kleinschmit, B., Förster, M.: An ontological system for interoperable spatial generalisation in biodiversity monitoring. *Computers & Geosciences* **84**, 86–95 (2015). DOI 10.1016/j.cageo.2015.08.006. URL <https://doi.org/10.1016/j.cageo.2015.08.006>
27. OSM: Openstreetmap (2016). URL [\url{https://www.openstreetmap.org}](https://www.openstreetmap.org). [Online: <https://www.openstreetmap.org>]
28. Richards, J.A.: Remote Sensing Digital Image Analysis. An Introduction. Springer-Verlag Berlin Heidelberg (2013)
29. Ryeng, E., Alsberg, B.: Microarray data classification using inductive logic programming and gene ontology background information. *Journal of Chemometrics* **24**, 231–240 (2010)
30. Srinivasan, A.: The aleph manual (2007)
31. Sternberg, M.J., Tamaddoni-Nezhad, A., Lesk, V.I., Kay, E., Hitchen, P.G., Cootes, A., van Alphen, L.B., Lamoureux, M.P., Jarrell, H.C., Rawlings, C.J., Soo, E.C., Szymanski, C.M., Dell, A., Wren, B.W., Muggleton, S.H.: Gene function hypotheses for the campylobacter jejuni glycome generated by a logic-based approach. *Journal of Molecular Biology* **425**(1), 186 – 197 (2013)
32. Vaz, D., Ferreira, M., Lopes, R.: Spatial-yap: a logic-based geographic information system. In: *Proceedings of the 23rd international conference on Logic programming, ICLP'07*, pp. 195–208. Springer-Verlag, Berlin, Heidelberg (2007)
33. Wilkinson, M., et al: The fair guiding principles for scientific data management and stewardship. *Nature Scientific Data* **3**(160018) (2016). URL <http://www.nature.com/articles/sdata201618>