Expert Systems With Applications 187 (2022) 115838

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Semantic modelling of Earth Observation remote sensing

José F. Aldana-Martín, José García-Nieto *, María del Mar Roldán-García, José F. Aldana-Montes

Dept. de Lenguajes y Ciencias de la Computación, ITIS Software, University of Málaga, ETSI Informática, Campus de Teatinos, Málaga 29071, Spain

ARTICLE INFO

Keywords:
Remote sensing
Earth Observation
Semantic web
Ontology
Linked data
Reasoning

ABSTRACT

Earth Observation (EO) based on Remote Sensing (RS) is gaining importance nowadays, since it offers a well-grounded technological framework for the development of advanced applications in multiple domains, such as climate change, precision agriculture, smart urbanism, safety, and many others. This promotes the continuous generation of data-driven software facilities oriented to advanced processing, analysis and visualization, which often offer enhanced computing capabilities. Nevertheless, the development of knowledge-driven approaches is still an open challenge in remote sensing, besides they provide human experts with domain knowledge representation, support for data standardization and semantic integration of sources, which indeed enhance the construction of advanced on-top applications. To this end, the use of ontologies and web semantic technologies have shown high success in knowledge representation in many fields, in which the Earth Observation is not an exception. However, as argued by the research community, there is large room for improvement in the specific case of remote sensing, where ontologies that consider the special nature and structure of different satellital and airborne data products are demanded. This article addresses, in first instance, part of this need by proposing a semantic model for the consolidation, integration, reasoning and linking of data (and meta-data), in the context of satellital remote sensing products for EO. With this objective, an OWL ontology has been developed and an RDF repository has been generated to allow advanced SPARQL querying. Although the proposal has been designed to consider remote sensing data products in general, the current study is mainly focused on the Sentinel 2 satellite mission from the Copernicus Programme of the European Space Agency (ESA). Four different use cases are showcased to check potentials of the proposed semantic model in terms of ontology integration, federated querying, data analysis and reasoning.

1. Introduction

The development of new sensors and the growing ease of access to data generated with remote sensing techniques are pushing data-driven research and the development of new innovative algorithms for its analysis. In this context, Earth Observation's satellite systems are continuously generating a great quantity of data, which are nowadays essential for applications in diverse areas, such as: climate change monitoring (Plummer, Lecomte, & Doherty, 2017), precision agriculture (Weiss, Jacob, & Duveiller, 2020), smart urban design (Reba & Seto, 2020), and many others. This promotes the continuous generation of data-driven software facilities oriented to advanced processing, analysis and visualization, which often offer enhanced computing capabilities. However, while promising, such new-generation applications still require the symbolic representation of scenes and objects in images, as well as the setting of threshold values of computed vegetation indexes for the generation of knowledge rule systems (Belgiu, Drăguţ, & Strobl, 2014). In this regard, the development of knowledge-driven deductive methods represents an important research line in remote sensing (Arvor, Belgiu, Falomir, Mougenot, & Durieux, 2019; Chen, et al., 2016), as they complement inductive data-driven techniques to make them more actionable. A significant example in this direction is GEOBIA (Geographic Object-Based Image Analysis) (Blaschke, et al., 2014) that allows to set objects in satellite images (based on grouping pixels according to common features) to classify them. GEOBIA enables the representation of complex spatial topological and non-topological relationships. However, as argued in the literature (Arvor et al., 2019) and (Belgiu et al., 2014), GEOBIA rules are usually highly biased to specific scenarios and hence, they are rarely suited to be generalized.

In this sense, the development of general knowledge-driven approaches constitutes an open challenge in remote sensing, besides they provide human experts with domain knowledge representation, support for data standardization and semantic integration of multiple sources, such as multi-spectral (and hyper-spectral) data from various satellites

* Corresponding author.
E-mail addresses: jfaldanam@gmail.com (J.F. Aldana-Martín), jnieto@lcc.uma.es (J. García-Nieto), mmar@lcc.uma.es (M. del Mar Roldán-García), jfam@lcc.uma.es (J.F. Aldana-Montes).

https://doi.org/10.1016/j.eswa.2021.115838
Received 26 May 2020; Received in revised form 19 May 2021; Accepted 30 August 2021
Available online 11 September 2021

and linked open data (meteorological, plant phenotype, etc.). Therefore, there is a clear need of studying integration aspects of existing ontologies in the context of remote sensing (Arvor et al., 2019), as well as to show the potentials of such integration for feeding advanced analysis. To this end, the use of ontologies and web semantic technologies has shown high success in many fields. In the specific domain of remote sensing, there are several distinct attempts of ontologies, although they still constitute local prototypes that illustrate the potential in their use (Andrés, Arvor, Mougenot, Libourel, & Durieux, 2017; Arvor et al., 2019).

On the basis of this necessity, this article proposes a semantic model for the consolidation, integration, reasoning and linking of data (and meta-data), in the context of satellital remote sensing products for EO. With this objective, an OWL (Web Ontology Language) (Group, 2012) ontology has been developed and an RDF repository has been generated to allow advanced SPARQL querying. The proposed ontology, called RESEO (REmote SEnsing Ontology), has been designed to consider remote sensing data and meta-data products in general, including satellite constellations, unmanned aerial vehicles (UAVs), airborne, etc. For the sake of better understanding, the current study is mainly focused on the Sentinel 2 satellite mission of the Copernicus Programme of the European Space Agency (ESA), due to the growing popularity it is exhibiting among the scientific community since its launch in 2015 (Pahlevan, Sarkar, Franz, Balasubramanian, & He, 2017). Several use cases are showcased to check potentials of the proposed semantic model in terms of ontology integration, federated querying, data analysis and reasoning.

The main contributions of this article are summarized as follows:

- A new ontology called RESEO[1] is proposed for the semantic modelling of remote sensing data and meta-data, produced in the scope of Earth observation. This ontology is developed for the first time (to the best of our knowledge) to cover multiple kinds of data products of multi/hyper-spectral images and meta-data collected from well-known satellite EO programs, UAVs, etc. RESEO is indeed linked with other existing ontologies in the field of Earth observation, as well as with ontologies devoted to meteorological open data, so an enriched knowledge framework is obtained as a result.
- In terms of materialization, the proposed ontology is developed in OWL 2 and it has been linked with related external ontologies according to the same standard. Then, a series of mapping functions have been developed for data consolidation in RDF (Resource Description Framework) standard, including automatic storage in common RDF repository and Endpoint service. From this, a series of advanced SPARQL queries are set in form of API service to promote the use from the research community.
- For validation purposes, a series of use cases have been worked that comprise: time series analysis, multiple satellital data product consolidation (Sentinel 2 and Landsat 8), data integration for analysis enrichment, and semantic reasoning for land-cover classification. RESEO is then shown to be useful to provide a knowledge framework for the data integration and enriched analysis, in the scope of remote sensing.

This article is organized as follows. In Section 2 background concepts and literature overview are given. Section 3 describes the semantic approach, focusing on the OWL Ontology. The procedure to validate this approach is described in Section 4. Section 5 provides discussions. Finally, Section 6 presents concluding remarks and future works.

[1] Available at URL https://github.com/KhaosResearch/RESEO.

2. Background and related work

This section describes required background concepts about semantic web and remote sensing. A review of related articles in the specialized literature is also provided to clarify the contributions or our proposal with regards to the current state of the art.

2.1. Background concepts

- *Ontology.* An ontology defines a simplified representation of the world, so that it can be represented for some purpose (Gruber, 1993). Ontology languages define a set of representational primitives which are used to model a body of knowledge. The main elements of an ontology are classes (or concepts), properties (or attributes), instances (or class members) and relationships. The Web Ontology Language (OWL) is a semantic markup language used to define and publish ontologies. OWL its build on top of RDF and it is a standard by the W3C. To formalize the proposed ontology in this work, a description logic syntax OWL-DL is used as summarized in Table 1.

- *RDF.* Resource Description Framework is also a W3C standard for the representation of information in the Web (Schreiber & Raimond, 2014). W3C encourages the use of RDF in applications where the data are going to be processed by other applications instead of only being shown to users. RDF offers a common framework where information can be shared between applications without losing their meaning, identifying each resource by an URI (Uniform Resource Identifier). To define resources, RDF uses statements in the form of triples, which contain a subject, a predicate and an object.

- *SPARQL.* SPARQL is a query language for RDF graphs (Harris & Seaborne, 2013), allowing the execution of queries between several graphs in different repositories (federated queries). SPARQL queries use RDF patterns to retrieve the set of triples in the RDF repository that match.

- *SWRL.* This language incorporates mechanisms to identify semantic relationships between individuals (Horrocks, Patel-Schneider, Bechhofer, & Tsarkov, 2005), hence providing OWL-based ontologies with extra inference capabilities. SWRL is based on rule expressions in form of "Antecedent ⇒ Consequent" to represent semantic relationships. Antecedent, as well as consequent, can be formulated as conjunctions of elements, which are associated to one or more attributes defined by a question mark and a variable (e.g., $?x$) in the rule. SWRL is being used to perform reasoning tasks for object classification in remote sensing imagery (Andrés et al., 2017; Gu, et al., 2017).

- *Data Product.* Compressed file that includes both, images from spectral bands and meta-data files needed to work with them. In Earth observation remote sensing, data products are collected from satellites, UAVs, or airborne devices, among others. Multi-spectral imagery data products generally refers to 3 to 13 bands regarding to channels, e.g., red, green, blue, near-infrared and short-wave infrared, whereas hyper-spectral imagery consists of much narrower bands (10–20 nm), with hundreds or thousands of them.

- *Sentinel 2 Data Product.* This kind of products are distributed at several levels of processing, in this work level 2 A with Bottom-of-atmosphere reflectance, correcting the atmosphere distortion effect on the image. Sentinel 2 products cover an area of $100 \times 100\ km^2$ and have a spatial resolution of 10, 20 or 60 metres per pixel depending on the band. This satellite takes 5 days to scan the globe, so it is the frequency to revisit a scene and to generate the corresponding data product. These data are distributed through the Copernicus Open Access Hub.[2] Fig. 1 illustrates the data product folder structure of Sentinel 2, including the ranges of spatial resolutions (% reflectance and wavelength) of the twelve bands.

[2] Copernicus Open Access Hub: https://SciHub.copernicus.eu/dhus/.

Table 1
Basic OWL-DL semantic syntax used to formally define the proposed ontology.

Descriptions	Abstract syntax	DL syntax
Operators	$intersection(C_1, C_2, \ldots, C_n)$ $union(C_1, C_2, \ldots, C_n)$	$C_1 \sqcap C_2 \sqcap \cdots C_n$ $C_1 \sqcup C_2 \sqcup \cdots C_n$
Restrictions	for at least 1 value V from C for all values V from C R is symmetric	$\exists V.C$ $\forall V.C$ $R \equiv R^{-}$
Class axioms	$A\ partial(C_1, C_2, \ldots, C_n)$ $A\ complete(C_1, C_2, \ldots, C_n)$	$A \sqsubseteq C_1 \sqcap C_2 \sqcap \cdots C_n$ $A \equiv C_1 \sqcap C_2 \sqcap \cdots C_n$

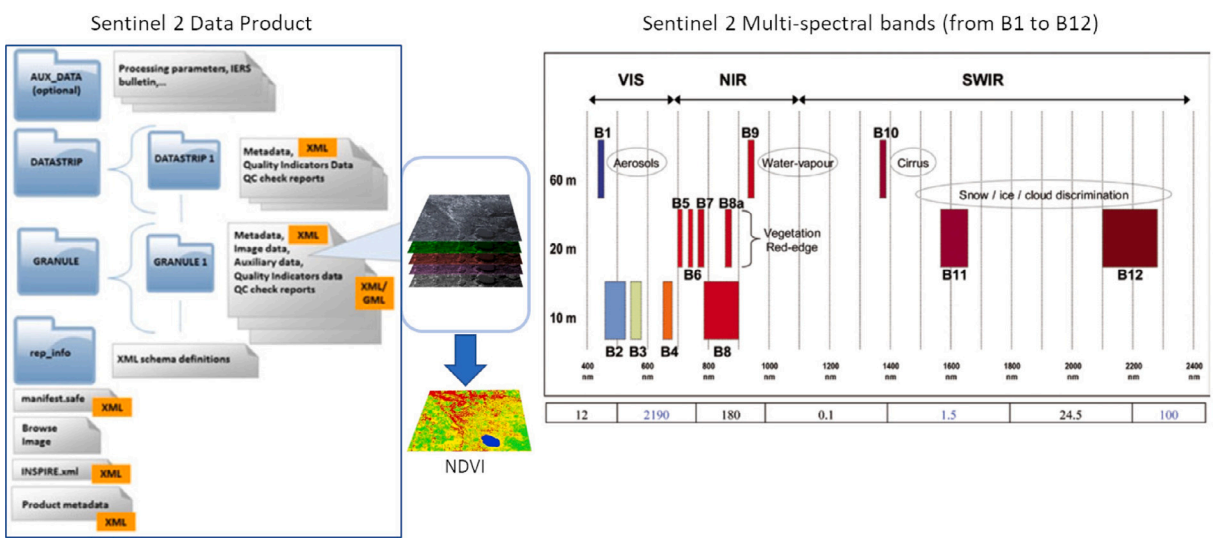

Fig. 1. Sentinel 2 Data Product structure (left). A scheme of spatial resolutions for bands (B1 to B12) with names and functionalities is shown at right. From some of these bands, vegetation indexes are computed, e.g., NDVI.

- *Vegetation indices.* A vegetation index (Abdou, Morin, Bonn, & Huete, 1996) is a value calculated from a set of channels (bands) from satellital sensors that quantifies the intensity of a complex phenomenon. Each channel (band) of a satellite image represents a different part of the electromagnetic spectrum, not limited to the visible light. An example of this type of indices is the Normalized Difference Vegetation Index (NDVI), which quantifies the liveliness of green vegetation in an area. It is calculated by Eq. (1), where NIR (B8 in Sentinel 2) is the near infrared channel of an image, and RED (B4 in Sentinel 2) is the red channel of an image.

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{1}$$

There exist many other indices that can be calculated from popular remote sensing satellite products (Sentinel 2, Landsat 8, MODIS, WordView, etc.), such as: Soil Adjusted Vegetation Index (SAVI), Enhanced Vegetation Index (EVI), Shadow Index (SI) or Normalized Difference Water Index (NDWI). All these can be obtained for the same sensed area and with different satellites, although showing different characteristics and resolutions (depending on the specific physical features of each sensor instrument). Therefore, data product harmonization is an important task in remote sensing, since it allows to complement information, hence enhancing the image analysis (Roy, et al., 2019). The RESEO ontology proposed in this work aims at covering this issue from a knowledge-driven perspective.

2.2. Related works

As commented before, a promising research line in Earth Observation consists in the development of knowledge-based solutions (Chen, et al., 2016). This will support domain experts to perform advanced analysis where context knowledge and the interpretation of remote sensing images are involved. In this sense, mechanisms like GEOBIA allows the classification of groups of pixels that share several properties in satellite images, by analysing them using knowledge from experts (Blaschke, 2010). However, GEOBIA is still limited to objects, so a complete contextual framework is required to capture the semantics of such observations, including their relationships in different levels, to better represent them in form of knowledge base.

A first attempt in this direction is the OBOE ontology (Madin, et al., 2007), which is oriented to represent ecological observations and to map real-world geographic entities with their corresponding objects in images. OBOE includes an extensive set of unit definitions and can facilitate automatic unit conversions, but is intended as a broadly applicable ontology, missing characteristics from specific research domain. OBOE can be aligned with the O&M ontology (Cox, 2013) by means of a property of equivalence between the classes *Measurement* (in the former) and *Observation* (in the latter). O&M is an OWL ontology that follows the ISO/OGC standard for Observations, as well as for other standard geographic information schemes.

From a different perspective, the Semantic Sensor Network (SSN) ontology (Compton, et al., 2012) comprises a contextual framework to represent sensor meta-data and observations, including remote sensing. In turn, SSN can be aligned with OBOE and O&M to compose a high level integration scheme, hence considering ecological and sensors knowledge domains. In this regard, the Semantic Web for Earth and Environmental Terminology (SWEET) (Raskin & Pan, 2005) is actually a collection of OWL ontologies considering such different domains (space, time, biological realms, physical quantities, etc.) and science knowledge concepts (phenomena, reactions, chemical processes, events, etc.). SWEET has been extended in some works to cover other domains, such as hydrogeology (Tripathi & Babaie, 2008) and Earth systems sciences in general (DiGiuseppe, Pouchard, & Noy, 2014).

From an orthogonal viewpoint, web semantic technologies can be also used to discover and integrate remote sensing services, which

Table 2
Summary ontologies' main features with regards to the proposed approach.

Domain/Ontology	OBOE	SNN	TIME-OWL	AEMET	GeoSPARQL	RESEO
Scientific observation and measurement	✓	✓				✓
Sensor's metadata		✓				✓
Temporal concepts			✓			✓
Meteorological data				✓		✓
Geospatial information					✓	✓
SWRL classification						✓

are disperse on the web, although devoted to similar and complementary functionalities. In Liu, Xue, Guang, and Liu (2015), an ontology-enabled framework is proposed for enabling collaboration among service providers and applications to semantically discover remote sensing services. This framework combines the use of ontologies and processing workflows.

Another interesting semantic model is the standard GeoSPARQL (Battle & Kolas, 2012). It supports the semantic representation and querying of geospatial data. GeoSPARQL defines a specific ontology for representing geospatial data in RDF, including an extension to the SPARQL query language for dealing with this kind of data. GeoSPARQL allows qualitative spatial reasoning and computations, so it is also suitable for being used in the context of remote sensing analysis (Viqueira, Villarroya, Mera, & Taboada, 2020).

All these semantic approaches can be extended to consider specific elements to the remote sensing domain, which lead to interpret EO images (e.g. spectral bands, indices, product's meta-data, etc.). Recent studies in this direction can be found in Andrés et al. (2017) and Gu, et al. (2017), although they are focused on the specific case of ontology-driven image classification.

The RESEO ontology aims at covering this gap, hence to improve the integration of remote sensing data and to enhance the generation of knowledge-driven approaches in this domain. Following the suggestions made in Arvor et al. (2019), RESEO can be used for modelling elements, such as: Sentinel 2, NDVI, NDVI Processing, MSI, etc., which indeed can be used as linking concepts for the alignment with other related ontologies: OBOE, SSN, and SWEET. As a summary, Table 2 shows the main features characterizing the reported ontologies, with regards to the proposed approach.

3. Semantic approach

One of the main objectives of RESEO is to provide an ontological framework for the semantic consolidation of the data captured by Earth Observation satellites, in a way that it can be easily extended for adding new data sources, such as different satellites, UAVs or linked open data. To this end, the proposed ontology has been defined in OWL 2, following the Ontology Development 101 (Noy & McGuinness, 2001) seven-step methodology as detailed next:

(i) *Determine the domain and scope of the ontology*. Although general enough to consider any kind of remote sensing product, for simplicity in this study, the scope of RESEO has been limited to the attributes of the Sentinel-2 and Landsat −8 meta-data. For example, the Sentinel 2 products include *platform name, orbit number, orbit direction, format, filename, data take identifier, processing level, etc.*

(ii) *Consider reusing existing ontologies*. Several existing ontologies have been used to make the proposal easier to align with others. Firstly, the OWL-Time.owl (Cox & Little, 2017) is used to describe temporal instants, and GeoSPARQL.owl (Perry & Herring, 2012) to describe geographical positions, both ontologies are W3C standards. To integrate meteorological data, RESEO is aligned to the AEMET.owl ontology (Poveda Villalón, 2011), that defines meteorological data and how it is captured. In the context of EO and sensors, the OBOE and SSN ontologies have been partially reused to consider classes related to satellital sensors and indexes. Table 2 shows the domains of all the linked ontologies. RESEO's goal is not to replace any of this standard ontologies, but to integrate all their fields to enhance the generation of knowledge-driven approach in the field of remote sensing.

(iii) *Enumerate important terms in the ontology*. The most important concepts for RESEO are the *Product*, its *DataSource*, the *Snapshot* (the analysis of a *Product*) and the *Scene* of interest for an analysis.

(iv) *Define classes and the class hierarchy*. The key concepts defined above have been modelled as the classes of the ontology. Fig. 2 shows the most important classes of RESEO. Some of these classes comprise a generalization of a set of more concrete classes. For example, *DataSource* is a general concept, which is specified in a hierarchy of subclasses that define more concrete data sources for a product, like *Satellite*, which later has other subclasses like *Sentinel 2* or *Landsat 8*. If a use case needs any extra data source, it can be integrated in the ontology as a subclass of *Data Source*.

(v) *Define the properties of classes and slots*. Object properties define relationships between classes. Some examples of them are: *Product* has data source *Data Source*, *Product* has scene *Scene*, *Satellite* has sensor *Sensor*, etc. Data properties define attributes that a member of a class can have. Some examples of them are: *Sentinel 2 product* has a *footprint*, a *format*, a *tile-id*, etc. General classes usually do not have any data property.

(vi) *Define the facets of the slots*. Each property in the ontology is constrained in the type and cardinality of its range and domain. For example, the range of the *hasDataSource* object property is a *DataSource*, and the domain of *hasDataSource* is *Product*; the range and domain of the *format* data property are *xsd:string* and *Sentinel2Product*, respectively. Value restrictions are used in our ontology to specify, for example, that if *Sentinel 2* is connected to *Sensor* through the *hasSensor* object property, at least one sensor has to be *MSI* (Multi-Spectral Imagery).

(vii) *Create instances*. Individuals are specific data that belong to a class. These instances are created by mapping the data from the data sources (Copernicus Data Hub,[3] AEMET Open Data,[4] etc.) to RDF. The mappings are done by following the model defined by the ontology. Apart from the mapping of data, some static instances are created for the satellites in study, namely: Sentinel 2 and Landsat 8. These instances are created so that all the new ones generated from the data have a place to link to. For example, if a new Sentinel 2 product is included, its data source must be an instance of *DataSource* and all the Sentinel 2 products will link to the same instance.

3.1. Ontology model

In its current version (1.0), RESEO includes 3319 axioms, 157 classes, 105 object properties, 118 data properties and 20 individuals. A number of classes have been integrated from other external ontologies,

[3] https://SciHub.copernicus.eu/.
[4] https://opendata.aemet.es/.

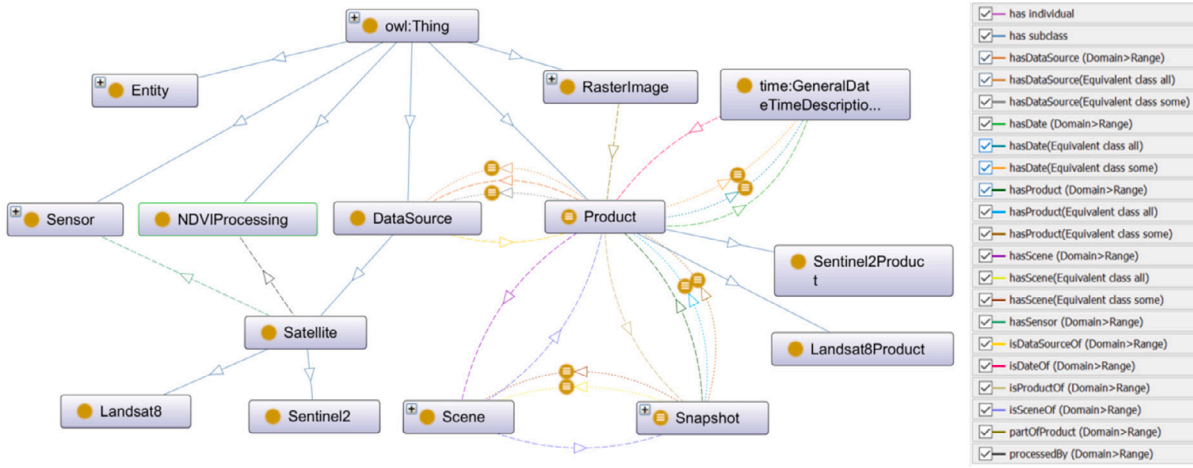

Fig. 2. Class diagram of the RESEO ontology. Continuous arrows refer to subclass of. Dotted arrows refer to specific properties as shown in the figure legend at right.

Table 3
Product: Object properties.

Object properties	Description logic
hasDataSource	$\equiv$ isDataSourceOf$^-$ $\exists$ hasDataSource Thing $\sqsubseteq$ Product $\top \sqsubseteq \forall$ hasDataSource DataSource
hasDate	$\equiv$ isDateOf$^-$ $\top \sqsubseteq \leq 1$ hasDate Thing $\exists$ hasDate Thing $\sqsubseteq$ Product $\top \sqsubseteq \forall$ hasDate GeneralDateTimeDescription
hasProduct	$\equiv$ isProductOf$^-$ $\exists$ hasProduct Thing $\sqsubseteq$ Snapshot $\top \sqsubseteq \forall$ hasProduct Product
hasScene	$\equiv$ isSceneOf$^-$ $\exists$ hasScene Thing $\sqsubseteq$ SpatialObject $\exists$ hasScene Thing $\sqsubseteq$ Product $\top \sqsubseteq \forall$ hasScene Scene

although a set of new generic ones have been designed to be inherited from. This allows the ontology to be easily expanded by adding new *Scene*, *DataSource* or *Product* types depending on the specific end user's case of study. For simplicity, a selection of the most important classes of RESEO are detailed as follows:

- ***Product.*** This class defines the data products received from the remote sensing devices, i.e., EO satellites and UAVs. Each *Product* has associated a timestamp given from the owl-TIME ontology of type *time:GeneralDateTimeDescription*.[5] The *Product* class is modelled in this version with two subclasses, *Sentinel2Product* and *Landsat8Product* which correspond to the data sources worked at the moment in the semantic model, although it can be easily extended with more of them. A set of main object properties of the *Product* class are: *hasDataSource*, *hasDate*, *hasProduct*, and *hasScene*, which are defined in description logic in Table 3. Regarding data properties, they are mostly defined for the subclasses of *Product*, since they cover specific aspects of the contextual use cases where they are used. For example, in the case of *Sentinel2Product*, data properties are referred to this particular data structure and attributes as defined in Table 4.

- ***Data Source.*** It represents a data provider for a *Product*. In its current state, this class has the subclass *Satellite*, which in turn has other two subclasses: *Sentinel2* and *Landsat8*. However, depending on the application, it could be extended with new hierarchies of subclasses to define different sources, such as: other EO satellites (MODIS, WorldView) and UAV products. *DataSource* includes two main object properties, *hasDataSource* and *isDataSourceOf*, which are defined in Table 5. It is worth noting that *DataSource* can also be used as linking element with other external ontologies, since it could cover sources of data in general, but focused on contextualizing with the remote sensing domain of knowledge. In this regard, the subclass *Satellite* has the object property *processedBy* with range *NDVIProcessing*, which is in turn linked with classes *Procedure* and *Observation* from the SSN ontology. *NDVIProcessing* has the data property *formula* to define how the NDVI is calculated for each satellite. In this regard, a subclass of *satellite* is *Sentinel2*, which is connected to *MSI*, being this last a subclass of *Sensor* (also taken from SSN). (See Table 4)

- ***Scene.*** A *Scene* defines a region of interest with a specific location in the Earth. A scene is contained inside one or more products, which could be captured from different remote sensing devices (Sentinel 2, Landsat 8, UAVs, etc.), although referring to the same specific location and preferably to similar (or close) time instants. *Scene* is modelled as a subclass of the *geosparql:Geometry* class of the GeoSPARQL ontology.[6] Object properties defined for *Scene* are: *hasNearestStation hasScene* and *isSceneOf*, which descriptions are detailed in Table 6. Data properties of this class are defined in *geosparql:Geometry*. The property *hasNearestStation* is used as linking element with the AEMET.owl ontology that incorporates meteorological data. In this way, the *Scene* allows to integrate different sensing data referring to a specific location, for a given time period and including the specific climatic conditions, e.g., the imagery products of Sentinel 2 and Landsat 8 capturing the area of the Strait of Gibraltar, and including maximum and minimum temperatures during the first week of August.

- ***Snapshot.*** A *Snapshot* represents the results of an analysis of a *Product*, over a concrete *Scene* of interest. Table 7 shows all the object properties of this class, namely: *hasProduct*, *isProductOf* and *isSceneOf*, while Table 8 contains a selection of representative data properties. Among these properties, it is worth mentioning those referring to vegetation indexes, such as EVI or NDVI, which are computed with different combination of spectral bands, depending on the remote sensing devices involved in the specific *Scene* (e.g., Sentinel 2, Landsat 8, etc.). The remaining indexes (SAVI, NSDI, etc.) are defined in this class by following similar schemes of data properties as done with EVI and NDVI. In this sense, NDVI is linked with the OBOE class *Characteristic*

[5] More information about the owl-TIME ontology: https://www.w3.org/TR/owl-time/.

[6] More information about the GeoSPARQL standard: https://www.ogc.org/standards/geosparql.

Table 4
Sentinel 2 Product: Data properties.

Data properties	Description logic
sentinel2ProductProperties	$\exists$ sentinel2ProductProperties Datatype rdfs:Literal $\sqsubseteq$ Sentinel2Product
baresoilpercentage	$\sqsubseteq$ sentinel2ProductProperties $\top \sqsubseteq \leq 1$ baresoilpercentage $\exists$ baresoilpercentage Datatype rdfs:Literal $\sqsubseteq$ Sentinel2Product $\top \sqsubseteq \forall$ baresoilpercentage Datatype xmls:decimal
beginposition	$\sqsubseteq$ sentinel2ProductProperties $\top \sqsubseteq \leq 1$ beginposition $\exists$ beginposition Datatype rdfs:Literal $\sqsubseteq$ Sentinel2Product $\top \sqsubseteq \forall$ beginposition Datatype xmls:dateTime
datatakesensingstart	$\sqsubseteq$ sentinel2ProductProperties $\top \sqsubseteq \leq 1$ datatakesensingstart $\exists$ datatakesensingstart Datatype rdfs:Literal $\sqsubseteq$ Sentinel2Product $\top \sqsubseteq \forall$ datatakesensingstart Datatype xmls:dateTime
endposition	$\sqsubseteq$ sentinel2ProductProperties $\top \sqsubseteq \leq 1$ endposition $\exists$ endposition Datatype rdfs:Literal $\sqsubseteq$ Sentinel2Product $\top \sqsubseteq \forall$ endposition Datatype xmls:dateTime
filename	$\sqsubseteq$ sentinel2ProductProperties $\top \sqsubseteq \leq 1$ filename $\exists$ filename Datatype rdfs:Literal $\sqsubseteq$ Sentinel2Product $\top \sqsubseteq \forall$ filename Datatype xmls:string
footprint	$\sqsubseteq$ landsat8ProductProperties $\sqsubseteq$ sentinel2ProductProperties $\top \sqsubseteq \leq 1$ footprint $\exists$ footprint Datatype rdfs:Literal $\sqsubseteq$ Sentinel2Product $\top \sqsubseteq \forall$ footprint Datatype xmls:string
gmlfootprint	$\sqsubseteq$ sentinel2ProductProperties $\top \sqsubseteq \leq 1$ gmlfootprint $\exists$ gmlfootprint Datatype rdfs:Literal $\sqsubseteq$ Sentinel2Product $\top \sqsubseteq \forall$ gmlfootprint Datatype xmls:string
highprobacloudspercentage	$\sqsubseteq$ sentinel2ProductProperties $\top \sqsubseteq \leq 1$ highprobacloudspercentage $\exists$ highprobacloudspercentage Datatype rdfs:Literal $\sqsubseteq$ Sentinel2Product $\top \sqsubseteq \forall$ highprobacloudspercentage Datatype xmls:decimal

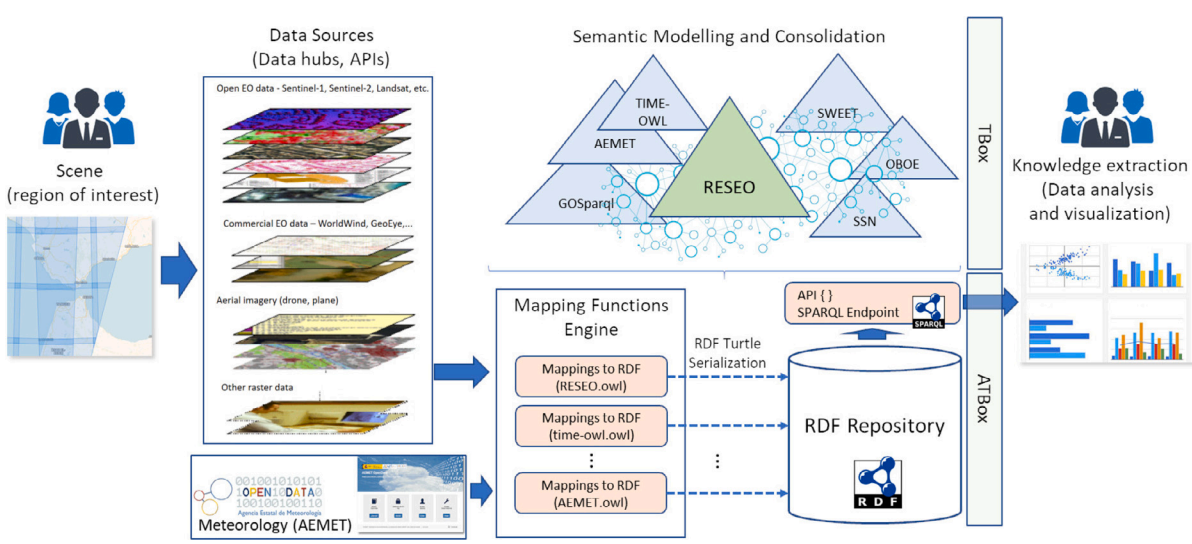

Fig. 3. Overall semantic model driven by the RESEO ontology as terminology component (TBox) in the knowledge-base of remote sensing data. The associated ABox is materialized with mapping functions to generate RDF linked data, the repository to consolidate them, and the SPARQL Endpoint to provide access by querying.

and with the SWEET ontology by means of the property *hasCharacteristic*. Another interesting property is the cloud cover percentage in sensed images, which is often used as a threshold to select an specific product, or discard it, for the analysis.

- ***Entity.*** This class has been reused from the OBOE ontology to model in RESEO those subclasses related to land-cover classification, in the current version: *BareSoil*, *Building*, *Vegetation* and *Water*. These classes are indeed subclasses of *FeatureOfInterest* from the SSN ontology. They are used as consequent elements in reasoning rules to perform ontology classification of remote sensing imagery (to be explained in Section 4.4).

Table 5
Data source: Object properties.

Object properties	Description logic
hasDataSource	$\equiv$ isDataSourceOf^{-} $\exists$ hasDataSource Thing $\sqsubseteq$ Product $\top \sqsubseteq \forall$ hasDataSource DataSource
isDataSourceOf	$\equiv$hasDataSource $^{-}$ $\exists$ isDataSourceOf Thing $\sqsubseteq$ DataSource $\top \sqsubseteq \forall$ isDataSourceOf Product

Table 6
Scene: Object properties.

Object properties	Description logic
hasNearestStation	$\exists$ hasNearestStation Thing $\sqsubseteq$ Scene $\top \sqsubseteq \forall$ hasNearestStation WeatherStation
hasScene	$\equiv$ isSceneOf^{-} $\exists$ hasScene Thing $\sqsubseteq$ SpatialObject $\exists$ hasScene Thing $\sqsubseteq$ Product $\top \sqsubseteq \forall$ hasScene Scene
isSceneOf	$\equiv$ hasScene^{-} $\exists$ isSceneOf Thing $\sqsubseteq$ Scene $\top \sqsubseteq \forall$ isSceneOf Snapshot $\top \sqsubseteq \forall$ isSceneOf Product

Table 7
Snapshot: Object properties.

Object properties	Description logic
hasProduct	$\equiv$ isProductOf^{-} $\exists$ hasProduct Thing $\sqsubseteq$ Snapshot $\top \sqsubseteq \forall$ hasProduct Product
isProductOf	$\equiv$ hasProduct^{-} $\exists$ isProductOf Thing $\sqsubseteq$ Product $\top \sqsubseteq \forall$ isProductOf Snapshot
isSceneOf	$\equiv$ hasScene^{-} $\exists$ isSceneOf Thing $\sqsubseteq$ Scene $\top \sqsubseteq \forall$ isSceneOf Snapshot $\top \sqsubseteq \forall$ isSceneOf Product

3.2. Data consolidation

At this point, the ontological framework of RESEO is then defined, including its linking mechanisms with other existing ontologies (OBOE, AEMET, GEOSParql, etc.). This constitutes the terminological component (TBox) of the proposed semantic approach. For model materialization through the associated ABox, a series of mapping functions have been defined to convert all processed data into RDF, according to the RESEO ontological scheme. All these RDF data are then stored and consolidated in a common RDF repository, which enables an SPARQL Endpoint for data querying.

An overall representation of the proposed semantic model is illustrated in Fig. 3. In the current version, two main data sources have been used for feeding the model consisting in: (1) Sentinel 2 data products collected from the Copernicus Data Hub and (2) meteorological data from the AEMET Open Data Portal.[7] This last source is obtained from the corresponding API in JSON format, so specific mappings have been also developed to transform them into RDF according to the *AEMET.owl* ontology. These data are collected for a given scene that is geo-localized in the area of interest (see Section 4), for a specific time period.

[7] AEMET Open Data Portal: http://www.aemet.es/es/datos_abiertos/AEMET_OpenData.

As commented before, this semantic model could be easily extended to consider other kinds of remote sensing data products (Landsat, MODIS, UAVs, etc.), as well as observation and meteorological data. In this regard, the linking properties defined in RESEO leads the RDF repository to be connected with other external knowledge graphs, so the use of federated SPARQL queries will allow semi-transparent data fusion for feeding advanced applications.

All the services and material generated for the semantic model are available in a CKAN organization devoted to the Green Senti Project (University of Malaga).[8] This Linked Open Data repository contains the OWL ontologies (RESEO and AEMET), use cases datasets, images, and SPARQL Endpoint.[9]

4. Validation

To validate the proposed semantic model, four cases of study have been developed that represent featured functionalities to be provided by knowledge-based approaches (Arvor et al., 2019). These functionalities are mainly focused on: (1) data processing from multiple satellite products to generate time series; (2) for a given scene, querying to merge data from different data products; (3) querying to fuse different kind of data, e.g. vegetation indexes and meteorology; and (4) land-cover semantic classification of remote sensing imagery based on reasoning rules.

Most of these cases have been conducted on a common scene located in the Teatinos Campus of the University of Malaga, which is a semi-urbanized area on the outskirts of the city of Malaga (Spain). Fig. 4 shows the selected area of this scene, which comprises 185 hectares in the west side of the metropolitan area of Malaga, containing: green zones, buddings, parkings, sports area, roads, and lakes.

4.1. Use case 1: Time series

One of the main tasks in remote sensing analysis is the generation of time series of a set of attributes, where vegetation indexes are often arranged with the observation dates, for monitoring the evolution of a certain factor. In these time series, additional information such as, climatic conditions or topological attributes, are usually incorporated, which are indeed useful for time series forecasting.

Listing 1 SPARQL Query: Q1

```
PREFIX reseo:
<http://khaos.uma.es/green-senti/reseo#>

SELECT ?uuid ?date ?link
WHERE {
   ?product reseo:hasScene ?scene .
   ?product reseo:hasDate ?date .
   ?product rdf:type reseo:Sentinel2Product .
   ?product reseo:link ?link .
   ?product reseo:uuid ?uuid .
   FILTER(?scene = reseo:teatinos) .
}
```

[8] CKAN Organization Green Senti https://opendata.khaos.uma.es/organization/green-senti.

[9] Green Senti SPARQL Endpoint for RESEO https://khaos.uma.es/opendata/sparql/.

Table 8
Snapshot: Data properties.

Data properties	Description logic
SnapshotProperties	∃ SnapshotProperties Datatype rdfs:Literal ⊑ Snapshot
cloudcoverpercentage	⊑ SnapshotProperties ⊤ ⊑ ≤ 1 cloudcoverpercentage ∃ cloudcoverpercentage Datatype rdfs:Literal ⊑ Snapshot ⊤ ⊑ ∀ cloudcoverpercentage Datatype xmls:decimal
evi	⊑ SnapshotProperties ⊤ ⊑ ≤ 1 evi ∃ evi Datatype rdfs:Literal ⊑ Snapshot ⊤ ⊑ ∀ evi Datatype xmls:decimal
evi_image	⊑ SnapshotProperties ∃ evi_image Datatype rdfs:Literal ⊑ Snapshot ⊤ ⊑ ∀ evi_image Datatype xmls:anyURI
ndvi	⊑ SnapshotProperties ⊤ ⊑ ≤ 1 ndvi ∃ ndvi Datatype rdfs:Literal ⊑ Snapshot ⊤ ⊑ ∀ ndvi Datatype xmls:decimal
ndvi_image	⊑ SnapshotProperties ∃ ndvi_image Datatype rdfs:Literal ⊑ Snapshot ⊤ ⊑ ∀ ndvi_image Datatype xmls:anyURI
true_color	⊑ SnapshotProperties ⊤ ⊑ ≤ 1 true_color ∃ true_color Datatype rdfs:Literal ⊑ Snapshot ⊤ ⊑ ∀ true_color Datatype xmls:decimal
true_color_image	⊑ SnapshotProperties ∃ true_color_image Datatype rdfs:Literal ⊑ Snapshot ⊤ ⊑ ∀ true_color_image Datatype xmls:anyURI

Fig. 4. Selected area (Scene) located in the University campus of Teatinos, Malaga. The complete area surface is 198 has, with green areas and buildings, centred at coordinates Lat: 36.71618, Lng: -4.48431.

Listing 2 SPARQL Query: Q2

```
PREFIX reseo:
<http://khaos.uma.es/green-senti/reseo#>
PREFIX aemet:
<http://aemet.linkeddata.es/ontology/>

SELECT distinct ?date ?prop ?val
WHERE {
  ?scene reseo:hasNearestStation
    ?station .
  ?obs aemet:isCapturedBy ?station .
  ?obs aemet:valueOfObservedData ?val .
  ?obs aemet:observedProperty ?prop .
  ?obs aemet:observedInInterval ?date .
  FILTER(?scene = reseo:teatinos) .
}
```

In this use case, the goal is to monitor the evolution of green zones in the university campus. To this end, a procedure has been developed to calculate the surface of green zones, as the number of pixels with $NDVI > 0.5$ for each product and observation date. In this way, the number of green hectares have been registered from 2016 to the date. It is worth noting that, as the first Sentinel 2 product that was distributed at a 2 A level was released in march 2018, the previous period was feed with products of the 1C level, including the application of the atmosphere correction.

All these data have been mapped into RDF (following RESEO) and stored in the repository, which indeed includes meteorological data (from AEMET) related to the same time period. The integrated data can be then queried by means of the SPARQL queries on Listings 1 and 2. The resulting time series can be plotted as shown in Fig. 5, where each point identifies the number of green hectares computed for each product and for each date of observation. The regression line fitting the time series reflects the seasonality induced by the vegetative stage of plants, as well as the amount of surface identifies as green zone.

The increasing tendency can be explained not only by the growth of plants, but also for the planting of new ones.

4.2. Use case 2: Merging remote sensing data from different products

Besides Sentinel 2, there are many other EO satellites in orbit and aerial imagery vehicles, all of them supplying images that can provide information about the same region of interest (scene). Therefore, an interesting option is to use not only one remote sensing source, but several of them, getting the most amount of data for a given analysis. However due to the differences of the inboard sensors, trajectories, positions, etc., these remote sensing devices generate different data products of multi-spectral bands, which also involve differences in the calculation of indices. For example, there is a smooth different between the NDVI values calculated by Sentinel 2 and Landsat 8. In addition, there are different indices that could be calculated with a given remote sensor, but not with others. This entails the need of integrating information captured by multiple sensors, for a common scene.

Listing 3 SPARQL Query: Q3

```
PREFIX reseo:
<http://khaos.uma.es/green-senti/reseo#>

SELECT ?date
?productS2 ?ndviS2 ?productL8 ?ndviL8
WHERE {
  ?productS2 reseo:hasDate ?date .
  ?productS2 rdf:type
    reseo:Sentinel2Product.
  ?snapshotS2 reseo:hasProduct ?productS2 .
  ?snapshotS2 reseo:hasScene
    reseo:teatinos .
  ?snapshotS2 rdf:type reseo:Snapshot .
  ?snapshotS2 reseo:NDVI ?ndviS2 .
  ?productL8 reseo:hasDate ?date .
  ?productL8 rdf:type
    reseo:Landsat8Product .
  ?snapshotL8
    reseo:hasProduct ?productL8 .
  ?snapshotL8 reseo:hasScene
    reseo:teatinos .
  ?snapshotL8 rdf:type reseo:Snapshot .
  ?snapshotL8 reseo:NDVI ?ndviL8 .
}
```

For instance, in order to integrate data captured by Sentinel 2 and Landsat 8, it is required to obtain data products from both satellites captured at the same date. To perform this task manually, the human expert has to visit each satellite data portal to get a list of the products and check which dates have a product available in common. This manual step could be automated by including Landsat 8 in RESEO semantic model, so a SPARQL query would return integrated data from the two satellites in study.

In this use case, after querying all the available Sentinel 2 and Landsat 8 products for the region of and date of interest, a number of 4 pairs of products matched the requirements. This is mainly due by the difference in periods for both satellites. To reach a compromise between the amount of data and the difference in time between products, the time window for accepting products was set to 1 day. This means that if a Sentinel 2 product is released a given day, any Landsat 8 products released the day after or before, will be considered too. This increases the number of products available for this study to 16 pairs, from a starting pool of 140 Landsat 8 products and 96 Sentinel 2 products. From these selected products, an RDF subgraph including pixel per pixel values for all bands from both satellites and some indices (NDVI and EVI) was integrated. An example of SPARQL query to get the NDVI values calculated from Sentinel 2 and Landsat 8 products is shown below in Q3 (Listing 3).

Listing 4 SPARQL Query: Q4

```
PREFIX aemet:
<http://aemet.linkeddata.es/ontology/>
PREFIX reseo:
<http://khaos.uma.es/green-senti/reseo#>

SELECT ?date ?ndvi ?temp
WHERE {
  ?obs aemet:observedInInterval ?date .
  reseo:teatinos
    reseo:hasNearestStation ?station .
  ?obs aemet:isCapturedBy ?station .
  ?obs aemet:observedProperty
    aemet:TemperatureAverage .
  ?obs aemet:valueOfObservedData ?temp .
  ?product reseo:hasDate ?date .
  ?snapshot reseo:hasProduct ?product .
  ?snapshot reseo:hasScene reseo:teatinos .
  ?snapshot rdf:type reseo:Snapshot .
  ?snapshot reseo:NDVI ?ndvi .
}
```

Listing 5 SPARQL Query: Q5

```
PREFIX aemet:
<http://aemet.linkeddata.es/ontology/>
PREFIX reseo:
<http://khaos.uma.es/green-senti/reseo#>

SELECT ?date ?moisture ?temp
WHERE {
  ?obs aemet:observedInInterval ?date .
  reseo:teatinos
    reseo:hasNearestStation ?station .
  ?obs aemet:isCapturedBy ?station .
  ?obs aemet:observedProperty
    aemet:TemperatureAverage .
  ?obs aemet:valueOfObservedData ?temp .
  ?product reseo:hasDate ?date .
  ?snapshot reseo:hasProduct ?product .
  ?snapshot reseo:hasScene reseo:teatinos.
  ?snapshot rdf:type reseo:Snapshot .
  ?snapshot reseo:MOISTURE ?moisture .
}
```

4.3. Use case 3: Querying for merging heterogeneous data

Semantic integration and consolidation of heterogeneous data from multiple sources is a key functionality of ontological models, since it

enables to enrich the initial knowledge bases with additional variables, hence allowing advanced analysis (Barba-González, et al., 2019).

This use case is oriented to exploit such a functionality in the context of RESEO, by merging information about the vegetation stage of the region of interest (University Campus of Teatinos) with the weather conditions registered in this region and for a given time period (Aug. 2017 to Feb. 2019).

To this end, SPARQL query Q4 is formulated to obtain the NDVI calculated from Sentinel 2 products together with the registered temperature, while SPARQL query Q5 selects the Moisture index (also from Sentinel 2 products) with the temperature.

With the resulting data, it is now possible to calculate the correlation between these two couples of variables. In this way, Figs. 6 and 7 show the Pearson correlation of the NDVI with regards to the average temperature (after normalizing) and the same correlation of the Moisture with regards the temperature, respectively. In both cases, there is a negative correlation between these variables, since the NDVI and the Moisture indices decrease whereas the temperature is higher. This is a typical observation in southern Spain, where high temperatures are usually accompanied by a dry environment.

Listing 6 SWRL Rule: R1

```
Pixel(?pixel)
^ PixelValue(?ndviPixelValue)
^ NDVI(?ndviImage)
^ hasValue(?pixel, ?ndviPixelValue)
^ value(?ndviPixelValue, ?ndviValue)
^ partOfRasterImage(?ndviPixelValue, ?ndviImage)
^ PixelValue(?band11PixelValue)
^ Band11(?band11Image)
^ hasValue(?pixel, ?band11PixelValue)
^ value(?band11PixelValue, ?band11Value)
^ partOfRasterImage(?band11PixelValue, ?band11Image)
^ PixelValue(?band4PixelValue)
^ Band04(?band4Image)
^ hasValue(?pixel, ?band4PixelValue)
^ value(?band4PixelValue, ?band4Value)
^ partOfRasterImage(?band4PixelValue, ?band4Image)

^ swrlb:lessThanOrEqual(?ndviValue, 0.246)
^ swrlb:lessThanOrEqual(?band11Value, 0.120)
^ swrlb:lessThanOrEqual(?band4Value, 0.114)

   -> Water(?pixel)
```

4.4. Use case 4: Remote sensing pixel classification with semantic reasoning

Land-cover classification of high resolution imagery is one of the main functionalities demanded in remote sensing, since it provides a framework for the identification, monitoring and traceability of important elements appearing in such images. It is used in important applications, such as: crop-land classification, urban monitoring and water reservoirs evolution. This problem has been successfully approached with two main strategies (Belgiu & Csillik, 2018; Weih & Riggan, 2010) by the remote sensing community, namely: object-based and pixel-based classification. Object-based classification has been recently tackled with semantic reasoning in some works (Andrés et al., 2017; Gu, et al., 2017) with success, although pixel-based semantic classification still remains an alternative to be checked (to the best of our knowledge).

In this regard, this use case is focused on performing pixel classification in Sentinel 2 products by means of a series of semantic reasoning tasks with SWRL rules under the knowledge-base of RESEO. Therefore, a set of rules have been constructed from a previous labelling process, where a series of thresholds were identified (on bands and NDVI) by training a decision tree on several products of the same scenario, spread across a year. These thresholds have been obtained with the ArcGIS[10] tool for discriminating the values of NDVI and Bands spectra, hence to separate the ground information of the study area into different land covers. In addition, colour mapping and class labelling were done to complete the classification process.

Listing 7 SWRL Rule: R5

```
Pixel(?pixel)
^ PixelValue(?ndviPixelValue)
^ NDVI(?ndviImage)
^ hasValue(?pixel, ?ndviPixelValue)
^ value(?ndviPixelValue, ?ndviValue)
^ partOfRasterImage(?ndviPixelValue, ?ndviImage)
^ PixelValue(?band11PixelValue)
^ Band11(?band11Image)
^ hasValue(?pixel, ?band11PixelValue)
^ value(?band11PixelValue, ?band11Value)
^ partOfRasterImage(?band11PixelValue, ?band11Image)
^ PixelValue(?band4PixelValue)
^ Band04(?band4Image)
^ hasValue(?pixel, ?band4PixelValue)
^ value(?band4PixelValue, ?band4Value)
^ partOfRasterImage(?band4PixelValue, ?band4Image)
^ PixelValue(?band2PixelValue)
^ Band02(?band2Image)
^ hasValue(?pixel, ?band2PixelValue)
^ value(?band2PixelValue, ?band2Value)
^ partOfRasterImage(?band2PixelValue, ?band2Image)

^ swrlb:lessThanOrEqual(?ndviValue, 0.469)
^ swrlb:greaterThan(?band11Value, 0.120)
^ swrlb:lessThanOrEqual(?band4Value, 0.094)
^ swrlb:lessThanOrEqual(?band2Value, 0.056)

   -> BareSoil(?pixel)
```

Listing 8 SWRL Rule: R11

```
Pixel(?pixel)
^ PixelValue(?ndviPixelValue)
^ NDVI(?ndviImage)
^ hasValue(?pixel, ?ndviPixelValue)
^ value(?ndviPixelValue, ?ndviValue)
^ partOfRasterImage(?ndviPixelValue, ?ndviImage)

^ swrlb:greaterThan(?ndviValue, 0.520)

   -> Vegetation(?pixel)
```

As a result, Table 9 contains the identified thresholds on pixel observation attributes (Bands and NDVI), together with the labelled classes, namely: Water, Bare Soil, Vegetation and Building. Column at right contains an identifier (from R1 to R11) to define the corresponding SWRL rule associated with each decision path in the classification procedure. Examples of three of these representative rules are R1, R5 and R11, which are defined in Listings 6, 7 and 8, respectively. These SWRL definitions show a common structure with a first block in the

[10] Available in URL https://desktop.arcgis.com/.

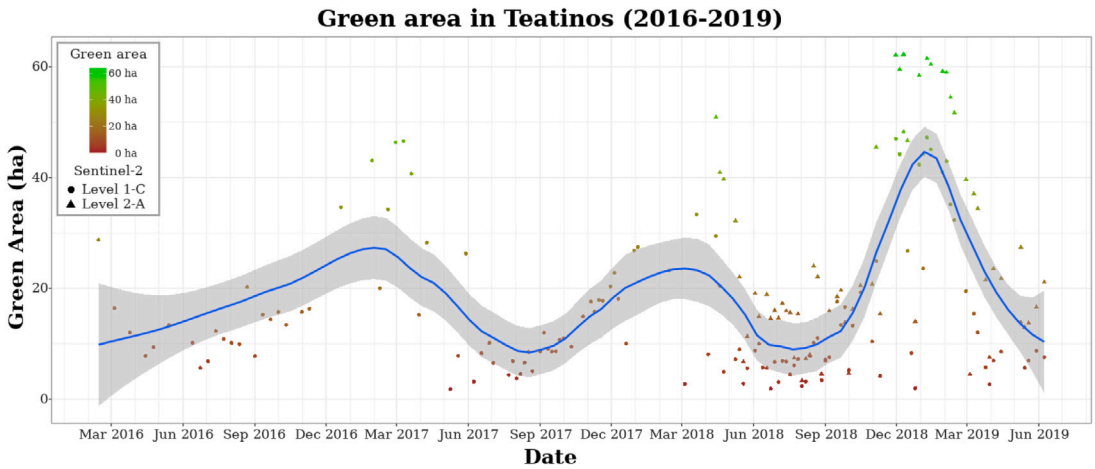

Fig. 5. Time series reflecting the green zone evolution in hectares of the university campus of Teatinos, from March 2016 to the date. Each point corresponds to a Sentinel 2 product for which, the observation area is used to compute the NDVI and to extract the number of green hectares.

Table 9
NDVI and Band thresholds calculated by decision tree.

Simplified rules				Class	Rule
B11 <= 0.120		NDVI <= 0.246	B04 <= 0.114	Water	R1
			B04 >0.114	Bare soil	R2
NDVI<=0.469	B11 <= 0.120	NDVI >0.246	B01 <= 0.040	Bare soil	R3
			B01 >0.040	Vegetation	R4
	B11 >0.120	B04 <= 0.094	B02 <= 0.056	Bare soil	R5
			B02 >0.056	Vegetation	R6
		B04 >0.094	B02 <= 0.232	Bare soil	R7
			B02 >0.232	Building	R8
NDVI >0.469	NDVI <= 0.520	B12 <= 0.144		Vegetation	R9
		B12 >0.144		Bare soil	R10
NDVI >0.520				Vegetation	R11

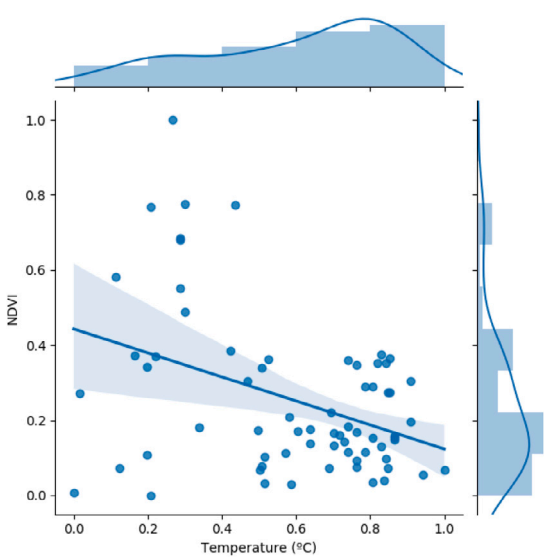

Fig. 6. Pearson's correlation between the NDVI vegetation index and the temperature in the university campus of Teatinos, Málaga.

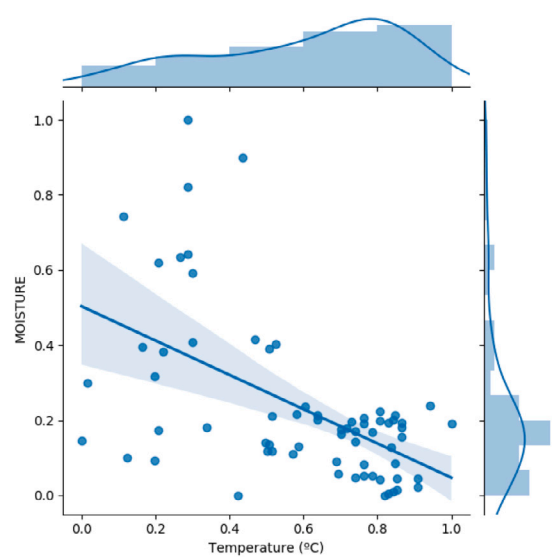

Fig. 7. Pearson's correlation between the Moisture index and the temperature in the university campus of Teatinos, Málaga.

antecedent of semantic element declarations, a second block (also in antecedent) of conditional definitions (with numeric thresholds defined in Table 9), and the labelled class in the consequent. For simplicity, the

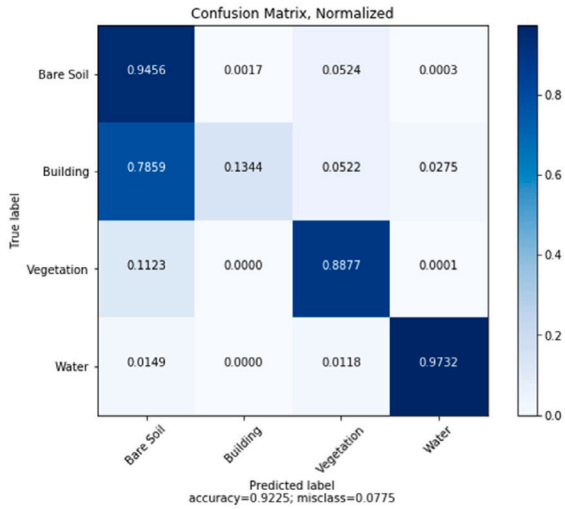

Fig. 8. Confusion matrix resulting from pixel classification in remote sensing imagery driven by RESEO ontology.

Table 10
Accuracy scores of the classifier.

	Precision	Recall	F1-score	Support
Bare soil	0.95	0.95	0.95	2623939
Building	0.60	0.13	0.22	51854
Vegetation	0.82	0.89	0.85	710049
Water	0.98	0.97	0.97	94936
Accuracy			0.92	3480778
Macro avg	0.84	0.74	0.75	3480778
Weighted avg	0.92	0.92	0.92	3480778

remaining of rules are provided in supplementary material CKAN site,[11] although they can be easily extracted from the examples.

These SWRL rules are then incorporated to RESEO and used to classify a test set on the selected image products, by means of reasoning tasks with Pellet reasoner (Sirin, Parsia, Grau, Kalyanpur, & Katz, 2007).[12] The confusion matrix obtained from this semantic classification is shown in Fig. 8, resulting a global prediction accuracy of 92%. More in detail, Table 10 contains, for each class and in general, the precision, recall, f1-score and support registered for this classification. Taking into account the high number of classified pixels (close to 3.5e+6), the resulting precision and recall are in general successful, as only for class "building" there is a low recall, probably due to class imbalance in the whole training dataset for this label.

5. Discussions

The development of knowledge-driven approaches represents an active research line in the remote sensing community (Arvor et al., 2019), since it offers potentials enough to provide human experts with domain knowledge representation, support for data standardization and semantic integration of sources. These functionalities are highly valuable for the creation of advanced on-top applications.

In this sense, the capacity of ontologies to offer formal framework for symbolic representation entails a new component in the associated artificial intelligence processes, since they allow to make the expert knowledge explicit (in such processes), hence contributing to the interpretation of results. This is of especial importance in the field of remote sensing imagery for Earth observation (Arvor, Durieux, Andrés, & Laporte, 2013), where research and industry communities are expending considerable efforts towards the knowledge extraction from sensed data and the interpretation of image products. RESEO aims at collaborating in this direction by offering an integration semantic model, as well as by incorporating mechanisms for the symbolic representation and standard access of data, in the specific field of remote sensing.

In terms of integration, a key aspect in RESEO design is the possibility of alignment with other interdisciplinary domains, such as: ecology, agriculture, urbanism, geology, etc, which indeed will contribute to the generation of new extended versions of this ontology.

With regards to symbolic representation and standard access, the proposed semantic model allows the generation of well-defined APIs and services, with the capacity of federated querying and data enrichment with foreign variables, such as: meteorologic measures, plant phenotype features or soil geo-morphology tastings. These features incorporate added value in current analysis, as worked in the context of the research project *Green-Senti 2019 PP Smart Campus* of the University of Malaga. This project initiated the design and development of RESEO, including the use cases described in Section 4 for demonstration purposes.

Finally, concerning societal and economic implications, the European Commission identified data standardization and harmonization in Earth observation as one of the main challenges to be tackled in this field (BDVA, 2017), since they constitute central aspects in the data value chain, from data acquisition and storage, to data usage for supporting in business decision-making. In fact, the main players in this field are currently large companies with capacity to access major infrastructures, although new opportunities are appearing for SMEs in respond to the emerging demand of remote sensing services. Initiatives like the Green-Senti project and the RESEO ontology are focused on contributing in this direction.

6. Conclusions

In this work, the RESEO ontology is proposed for the semantic modelling of remote sensing data and meta-data, in the scope of Earth observation. This ontology is conceived to cover multiple kinds of data products of remote sensing imagery and their associated meta-data. RESEO is indeed linked with other existing ontologies in the field of Earth observation, as well as with ontologies devoted to meteorological open data, so an enriched knowledge framework is obtained as a result.

The proposed ontology is developed in OWL 2 and it has been linked with related external ontologies according to the same standard (OBOE, SSN, TIME-OWL, AEMET, and GeoSPARQL). A series of mapping functions have been developed for data consolidation in RDF, including automatic storage in common repository and Endpoint service. From this, a series of advanced SPARQL queries are set in form of API service to promote the use from the research community.

On top of this semantic model, a series of pilot applications have been generated in form of use cases on a selected area in the campus university of Malaga (Spain). These use cases consist of: time series analysis for environmental monitoring, multiple satellital data product consolidation (Sentinel 2 versus Landsat 8), data integration for analysis enrichment, and semantic reasoning for land-cover classification. RESEO has been shown to be useful to provide a knowledge framework for the data integration and enriched analysis, in the scope of remote sensing.

As future work, more data sources will be integrated in the ontology, including more Earth observation satellites like MODIS or Proba-V, and others types of data products involving hyper-spectral imagery from UAVs. In addition, the integration and linkage of other ontologies is a future task, since it will allow scaling the use cases to more complex scenarios, such as global climate change, bio-habitats conservation, forest monitoring, etc.

[11] Available in URL https://opendata.khaos.uma.es/organization/green-senti.

[12] Available at URLhttps://www.w3.org/2001/sw/wiki/Pellet.

CRediT authorship contribution statement

José F. Aldana-Martín: Conceptualization, Investigation, Methodology, Software, Data curation, Writing – original draft. **José García-Nieto:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **María del Mar Roldán-García:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Validation. **José F. Aldana-Montes:** Supervision, Conceptualization, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has been partially funded by FEDER Grants TIN2017-86049-R and PID2020-112540RB-C41 (AEI/FEDER, UE) (Spanish Ministry of Education and Science), Andalusian PAIDI program with grant P18-RT-2799, and Green-Senti 2019 & 2021 PP Smart Campus UMA. It is also granted by the LifeWatch-ERIC initiative Smartfood Lifewatch (FEDER & Junta de Andalucia).

References

Abdou, B., Morin, D., Bonn, F., & Huete, A. (1996). A review of vegetation indices. *Remote Sensing Reviews*, *13*, 95–120. http://dx.doi.org/10.1080/02757259509532298.

Andrés, S., Arvor, D., Mougenot, I., Libourel, T., & Durieux, L. (2017). Ontology-based classification of remote sensing images using spectral rules. *Computers & Geosciences*, *102*(C), 158–166. http://dx.doi.org/10.1016/j.cageo.2017.02.018.

Arvor, D., Belgiu, M., Falomir, Z., Mougenot, I., & Durieux, L. (2019). Ontologies to interpret remote sensing images: why do we need them? *GIScience & Remote Sensing*, 1–29. http://dx.doi.org/10.1080/15481603.2019.1587890.

Arvor, D., Durieux, L., Andrés, S., & Laporte, M.-A. (2013). 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. http://dx.doi.org/10.1016/j.isprsjprs.2013.05.003, URL http://www.sciencedirect.com/science/article/pii/S092427161300124X.

Barba-González, C., García-Nieto, J., del Mar Roldán-García, M., Navas-Delgado, I., Nebro, A. J., & Aldana-Montes, J. F. (2019). BIGOWL: Knowledge centered big data analytics. *Expert Systems with Applications*, *115*, 543–556.

Battle, R., & Kolas, D. (2012). Enabling the geospatial semantic web with parliament and GeoSPARQL. *Semantic Web*, *3*, 355–370.

BDVA (2017). Whitepaper big data in earth observation. URL http://bdva.eu/sites/default/files/TF7%20SG5%20Working%20Group%20-%20White%20Paper%20EO_final_Nov%202017.pdf.

Belgiu, M., & Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment*, *204*, 509–523.

Belgiu, M., Drăguţ, L., & Strobl, J. (2014). Quantitative evaluation of variations in rule-based classifications of land cover in urban neighbourhoods using WorldView-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *87*, 205–215. http://dx.doi.org/10.1016/j.isprsjprs.2013.11.007, URL http://www.sciencedirect.com/science/article/pii/S0924271613002670.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(1), 2–16. http://dx.doi.org/10.1016/j.isprsjprs.2009.06.004, URL http://www.sciencedirect.com/science/article/pii/S0924271609000884.

Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., et al. (2014). Geographic object-based image analysis – Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, *87*, 180–191. http://dx.doi.org/10.1016/j.isprsjprs.2013.09.014, URL http://www.sciencedirect.com/science/article/pii/S0924271613002220.

Chen, J., Dowman, I., Li, S., Li, Z., Madden, M., Mills, J., et al. (2016). Information from imagery: ISPRS scientific vision and research agenda. *ISPRS Journal of Photogrammetry and Remote Sensing*, *115*, 3–21. http://dx.doi.org/10.1016/j.isprsjprs.2015.09.008, URL http://www.sciencedirect.com/science/article/pii/S092427161500218X (Theme issue 'State-of-the-art in photogrammetry, remote sensing and spatial information science').

Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S., et al. (2012). The SSN ontology of the W3C semantic sensor network incubator group. *Journal of Web Semantics*, *17*, 25–32. http://dx.doi.org/10.1016/j.websem.2012.05.003, URL http://www.sciencedirect.com/science/article/pii/S1570826812000571.

Cox, S. J. D. (2013). An explicit OWL representation of ISO/OGC observations and measurements. In *Proceedings of the 6th international conference on semantic sensor networks (vol. 1063)* (pp. 1–18). Aachen, DEU: CEUR-WS.org.

Cox, S., & Little, C. (2017). *Time ontology in OWL*: *Tech. rep.*, URL https://www.w3.org/TR/owl-time/.

DiGiuseppe, N., Pouchard, L. C., & Noy, N. F. (2014). SWEET ontology coverage for earth system sciences. *Earth Science Informatics*, *7*(4), 249–264.

Group, W. O. W. (2012). *OWL 2 web ontology language document overview (second edition)*: *Tech. rep.*, URL https://www.w3.org/TR/owl2-overview/.

Gruber, T. R. (1993). A translation approach to portable ontologies. *Knowledge Acquisition*, *5*(2), 199–220.

Gu, H., Li, H., Yan, L., Liu, Z., Blaschke, T., & Soergel, U. (2017). An object-based semantic classification method for high resolution remote sensing imagery using ontology. *Remote Sensing*, *9*, 329. http://dx.doi.org/10.3390/rs9040329.

Harris, S., & Seaborne, A. (2013). *SPARQL 1.1 query language*: *Tech. rep.*, URL https://www.w3.org/TR/sparql11-query/.

Horrocks, I., Patel-Schneider, P. F., Bechhofer, S., & Tsarkov, D. (2005). OWL rules: A proposal and prototype implementation. *Web Semantics: Science, Services and Agents on the World Wide Web*, *3*(1), 23–40.

Liu, L., Xue, Y., Guang, J., & Liu, J. (2015). Description of an ontology-based remote sensing model service with an integrated framework environment for remote sensing applications. *Remote Sensing Letters*, *6*(10), 804–813.

Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., & Villa, F. (2007). An ontology for describing and synthesizing ecological observation data. *Ecological Informatics*, *2*(3), 279–296.

Noy, N. F., & McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology*: *Stanford knowledge systems laboratory technical report KSL-01-05 and stanford medical informatics technical report SMI-2001-0880*.

Pahlevan, N., Sarkar, S., Franz, B., Balasubramanian, S., & He, J. (2017). Sentinel-2 MultiSpectral instrument (MSI) data processing for aquatic science applications: Demonstrations and validations. *Remote Sensing of Environment*, *201*, 47–56.

Perry, M., & Herring, J. (2012). *OGC GeoSPARQL - A geographic query language for RDF data*: *Tech. rep.*, URL https://www.ogc.org/standards/geosparql.

Plummer, S., Lecomte, P., & Doherty, M. (2017). The ESA Climate Change Initiative (CCI): A European contribution to the generation of the Global Climate Observing System. *Remote Sensing of Environment*, *203*, 2–8.

Poveda Villalón, M. (2011). The aemet network of ontologies. URL http://aemet.linkeddata.es/ontology/.

Raskin, R. G., & Pan, M. J. (2005). Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). *Computers & Geosciences*, *31*(9), 1119–1125.

Reba, M., & Seto, K. C. (2020). A systematic review and assessment of algorithms to detect, characterize, and monitor urban land change. *Remote Sensing of Environment*, *242*, Article 111739.

Roy, D. P., Huang, H., Boschetti, L., Giglio, L., Yan, L., Zhang, H. H., et al. (2019). Landsat-8 and sentinel-2 burned area mapping - A combined sensor multi-temporal change detection approach. *Remote Sensing of Environment*, *231*, Article 111254.

Schreiber, G., & Raimond, Y. (2014). *RDF 1.1 primer*: *Tech. rep.*, URL https://www.w3.org/TR/rdf11-primer/.

Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., & Katz, Y. (2007). Pellet: A practical OWL-DL reasoner. *Journal of Web Semantics*, *5*(2), 51–53, Software Engineering and the Semantic Web.

Tripathi, A., & Babaie, H. A. (2008). Developing a modular hydrogeology ontology by extending the SWEET upper-level ontologies. *Computers & Geosciences*, *34*(9), 1022–1033.

Viqueira, J. R. R., Villarroya, S., Mera, D., & Taboada, J. A. (2020). Smart environmental data infrastructures: Bridging the gap between earth sciences and citizens. *Applied Sciences*, *10*(3).

Weih, R., & Riggan, N. (2010). Object-based classification vs. Pixel-based classification: Comparitive importance of multi-resolution imagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *38*.

Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, *236*, Article 111402.