A Scalable Geospatial Web Service for Near Real-Time, High-Resolution Land Cover Mapping

Konstantinos Karantzalos, *Senior Member, IEEE*, Dimitris Bliziotis, and Athanasios Karmas, *Member, IEEE*

***Abstract*—A land cover classification service is introduced toward addressing current challenges on the handling and online processing of big remote sensing data. The geospatial web service has been designed, developed, and evaluated toward the efficient and automated classification of satellite imagery and the production of high-resolution land cover maps. The core of our platform consists of the *Rasdaman* array database management system for raster data storage and the open geospatial consortium web coverage processing service for data querying. Currently, the system is fully covering Greece with Landsat 8 multispectral imagery, from the beginning of its operational orbit. Datasets are stored and preprocessed automatically. A two-stage automated classification procedure was developed which is based on a statistical learning model and a multiclass support vector machine classifier, integrating advanced remote sensing and computer vision tools like Orfeo Toolbox and OpenCV. The framework has been trained to classify pansharpened images at 15-m ground resolution toward the initial detection of 31 spectral classes. The final product of our system is delivering, after a postclassification and merging procedure, multitemporal land cover maps with 10 land cover classes. The performed intensive quantitative evaluation has indicated an overall classification accuracy above 80%. The system in its current alpha release, once receiving a request from the client, can process and deliver land cover maps, for a 500-km^2 region, in about 20 s, allowing near real-time applications.**

***Index Terms*—Automation, classification, earth observation (EO), land cover, land use, machine learning, multispectral.**

I. Introduction

THE PROCESSING of more than 20 000 LANDSAT and HJ-1 satellite images and the development of the first global 30-m resolution land cover dataset for 2 baseline years (i.e., GlobeLand30) took almost 4 years [1]. The initial assessment of GlobeLand30's water layer in Northern Europe indicated high accuracy rates; however, a comprehensive validation has yet to be completed.

Although this can be correctly considered as a milestone in the history of obtaining global geospatial information from satellite imagery, it indicates the urgent need for efficient and automated classification tools able to process and deliver rapidly accurate land cover maps [2], [3].

Indeed, the current generation of space-borne sensors are generating nearly continuous streams of massive earth observation (EO) datasets. Shortly, high-resolution multispectral images will be available almost once a week and in some regions twice per week. These huge imaging streams, which are received through satellite downlink channels at gigabit rates, increase with tremendous velocity, reaching currently several petabytes in many satellite archives [4]–[7]. However, it is estimated that most of the datasets in existing archives have never been accessed and processed [8]. Harvesting valuable knowledge and information from big EO data is not a trivial task.

Big EO data processing turns out to be extremely challenging [7], [9]–[11], while the increasing data volumes are not the only consideration. As the wealth of data increases, the challenge of indexing, searching, and transferring increases exponentially as well. Open issues include the efficient data storage, handling, management, and delivery, the processing of multimodal and high-dimensional datasets as well as the increasing demands for real-, or near real-time processing for many critical geospatial applications [2], [4], [12]. Among them, land cover information and timely observations over the (bio)physical cover of the earth's surface are of significant importance.

Moreover, the development of novel geospatial web services [11], [13]–[15] for on-demand remote sensing analysis is a key issue. Geospatial web services enable users to leverage distributed geospatial data, products, and computing resources over the network and from third-party geospatial applications and to automate geospatial data integration and analysis procedures. These services should be interoperable and allow for collaborative processing of geospatial data for information and knowledge discovery. The aforementioned features can be accomplished through the utilization of the service computing and workflow technologies [16], [17].

To this end, a scalable geospatial web service has been designed, developed, and evaluated toward the efficient and automated processing of high-resolution satellite data for the production of land cover maps. The core of our platform consists of the *Rasdaman* array database management system (DBMS) [18], [19] for big raster data storage and the web coverage processing service (WCPS) interface standard which is maintained by the open geospatial consortium (OGC) for data querying [20]. The WebGIS client is based on the OpenLayers and GeoExt javascript libraries. Currently, the system is fully covering Greece with LANDSAT 8 multispectral data, from the beginning of the satellite's mission. Datasets are stored and preprocessed automatically. The automated classification framework is based on a comprehensive statistical training model and on a multiclass support vector machine (SVM) classifier, integrating advanced remote sensing and computer vision libraries like GDAL, Orfeo Toolbox, and OpenCV. The framework has been trained to classify pansharpened images at 15-m

Manuscript received November 02, 2014; revised May 13, 2015; accepted July 10, 2015.

The authors are with the Remote Sensing Laboratory, National Technical University of Athens, Athens 15780, Greece (e-mail: karank@central.ntua.gr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2015.2461556

ground resolution toward the detection of 31 spectral classes which can address at this particular scale the diversity of the terrain objects that can be found in Greece. The final product delivers, after a postclassification and merging procedure, multitemporal land cover maps with 10 land cover classes. The performed quantitative evaluation has indicated a classification accuracy above 80%. Toward the operational use of the service, a comprehensive validation on numerous scenes and dates along with the migration to a cloud environment has been scheduled.

A. Related Work

We have been inspired by the PlanetServer system [5] which is a service component of the EU-funded EarthServer project aimed at serving and analyzing planetary data online. EarthServer project[1] is creating an on-demand online open access and ad hoc analytics infrastructure for massive (100+ TB) Earth Science data based on cutting-edge array database platform and OGC WCPS standard. The WCPS has already proven its scalable and efficient capabilities on quickly process large amounts of data and deliver finished products to the end-user at an extremely low cost by integrating a full hyperspectral unmixing chain as part of the NASA Sensor Web suite of web services [6] and by integrating standard processing and vegetation analysis procedures for agricultural applications [21], [22].

Such frameworks, cloud-based platforms [23]–[25], and high performance computing [26], [27] form the single way forward for applying, on the fly, over the web, on the server-side, image analysis and data analytics tasks for the production of valuable geo-information. To this end, the Global Forest Watch Service [28], which is a milestone online platform powered by Google Earth Engine, forms a dynamic online forest monitoring and alert system which exploits satellite technology, open data, and crowdsourcing to guarantee access to timely and reliable information about forest status globally.

Moreover, along with the importance of employing efficient tools and platforms for big EO data storing, handling, and delivery, advanced algorithms able to learn, retrieve, and classify information from large datasets are required [29]–[32]. Retrieval from satellite image datasets has been proposed based on a semi-supervised method for the annotation of images [33] and on the enrichment of metadata, of the semantic annotations, and the image content [34]. A multistage active-learning procedure has been also employed for pattern retrieval in large image databases [35], while GPU-based implementations have been employed for addressing the computational cost of extracting spectral information from hyperspectral data [36]. Last but not least, a web-based system for the unsupervised classification of satellite and airborne images has been developed [31], which allows the user to perform classification tasks at different zoom levels allowing control and supervision over the final result.

B. Contribution

Building upon similar efforts [5], [6], [21], we introduce a framework that integrates cutting edge geospatial tools for raster data handling through DBMSs, OGC services, and advanced data processing algorithms. The developed geospatial web service is scalable and has been designed to execute on the fly, over the web, on the server-side, image classification tasks for the production of land cover maps. To our knowledge, this is the first web-based system that is addressing these challenges offering at a national scale (currently, fully covering Greece) geospatial products.

It should be noted that the main contribution is not the classification framework. The system has been designed to operate with any classifier or training procedure. However, we have experimented with various advanced training models and classifiers, with different (numbers of) land cover classes and hierarchies. We propose an automated two-stage classification framework which can deliver in near real-time 10 land cover classes at 15-m spatial resolution based on LANDSAT 8 multispectral, multitemporal data.

II. Land Cover Geospatial Service

The main objective of this work was to design and implement a framework for the online analysis of multispectral, multitemporal satellite imagery for the production of land cover maps. Various components and processing steps are involved in setting, running, and utilizing the developed service.

The overall system architecture is shown in Fig. 1. The core functionality of the developed framework, consists of the *Rasdaman* Array DBMS for storage of remote sensing data and OGC WCPS interface standard for querying them. *Rasdaman* was selected as the core system of the implementation due to its proven robustness, novelty, and efficiency in handling big imaging data[2] [5], [11], [21], [22], [37].

A. Data Acquisition and Preprocessing

Currently, the open satellite imaging data that systematically are stored in our database are derived from the US LANDSAT Data Continuity Mission (LDCM).[3] The LANDSAT 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) instruments acquire multitemporal, multispectral data with a spatial resolution of 30 m. LANDSAT 8 raw data are downloaded, stored, and preprocessed automatically through our system.

As far as the download, storage, and preprocessing stages are concerned, a number of Python scripts were developed which control, facilitate, and automate the entire operation. First, any newly acquired LANDSAT 8 dataset over the Greek territory has been detected and downloaded in an automated manner. Then, another script archives the data, i.e., uncompresses them and performs all necessary image preprocessing radiometric correction steps that are required. It should be mentioned that a cloud-screening process does not let any scene with a cloud coverage above 70% to be stored in our system.

Regarding the radiometric correction steps, the Unites States Geological Survey's (USGS's) instructions were followed in

[1][Online]. Available: www.earthserver.eu

[2][Online]. Available: http://www.copernicus-masters.com/index.php?kat=winners.html&anzeige=winner_t-systems2014.html

[3][Online]. Available: http://LANDSAT.usgs.gov/

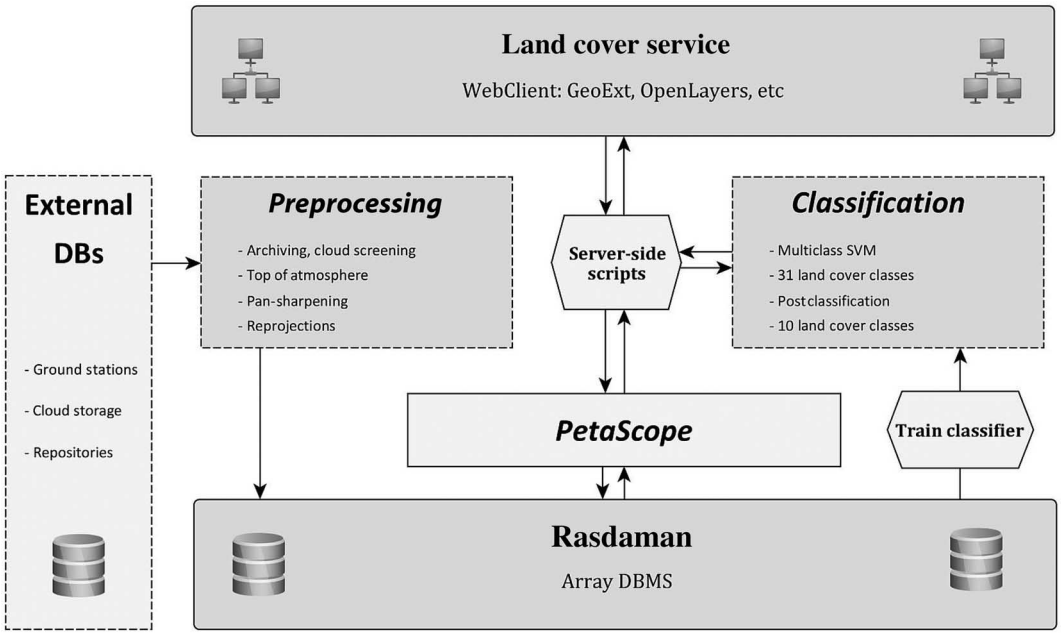

Fig. 1. System architecture of the developed land cover geospatial service.

using the LANDSAT 8 product.[4] Digital Numbers are converted to top of atmosphere (TOA) reflectance using radiometric rescaling coefficients provided in the product metadata file (MTL file). The MTL file also contains the thermal constants needed to convert TIRS data to the at-satellite brightness temperature.

After the radiometric corrections, the multispectral bands (MBs) are fused with the panchromatic (PAN) one, in order to increase their spatial resolution to 15-m ground resolution. The pansharpening is performed based on a standard process $(HR_{MB}/PAN_{sm}) * PAN$ through the Orfeo ToolBox.[5] HR_{MB} is the multispectral image interpolated into the same size as the PAN, while PAN_{sm} is an isotropically smoothed version of the raw panchromatic band. The resulting pansharpened multispectral images have a 15-m ground resolution and are the data which are stored and processed by the service. The pansharpened images are then reprojected through GDAL[6] to WGS 84 (EPSG:4326).

After the preprocessing stages, data are ready to be inserted into the *Rasdaman* database. Thereupon, a final script reads the delivered metadata and inserts the datasets and their metadata into the *Rasdaman* database.

B. Rasdaman and WCPS

Multidimensional arrays of large size are not supported by traditional DBMSs. As a consequence, these data are served through custom-made ad hoc servers which support arrays, but, on the other hand, lack database features such as query languages, query optimization and parallelization, and access-efficient storage architectures. Array DBMSs, however, support multidimensional arrays with unlimited size of dimensions while offering all the classical databases' advantages.

Rasdaman supports multidimensional arrays of very large sizes and thus can handle inherently big remote sensing data. *Rasdaman*'s architecture is based on a transparent array partitioning, called tiling. Conceptually, there is no size limitation for *Rasdaman* as a central DBMS of raster datasets. Additionally, *Rasdaman* features parallel server architecture that offers a scalable, distributed environment to efficiently process very large numbers of concurrent client requests and serve distributed datasets across the web.

The *Rasdaman* database of the developed system currently contains multitemporal imagery from the LANDSAT 8 satellite that covers the entire Greek territory (approximately 45 paths and rows) from the beginning of the mission. Prior to the insertion of data in *Rasdaman*, an appropriate data type for the LANDSAT 8 data needs to be defined. This type definition initially defines the "pixel type" by setting the amount of bands and the value type for each band. Then, a raster type (multidimensional array) is created which is specified as being two-dimensional (2-D), with completely open bounds in all directions; thus, the *Rasdaman* server will allow for images at any coordinate and with a dynamically growing extent. This feature allows to serve all LANDSAT 8 images through the definition of only one data type regardless of their spatial extent.

Datasets to be processed are extracted from *Rasdaman* through the execution of retrieval queries written in a query language defined by OGC's WCPS standard. This language allows for retrieval, filtering, processing, and fast subsetting of multidimensional raster coverages, such as sensor, simulation, image, and statistics data. For WCPS query language, *Rasdaman* is the reference implementation.

WCPS queries are submitted to the *Rasdaman* database server through the PetaScope component [38]. PetaScope is a java servlet package which implements OGC standard

[4][Online]. Available: http://LANDSAT.usgs.gov/LANDSAT8_Using_Product.php

[5][Online]. Available: www.orfeo-toolbox.org/otb/

[6][Online]. Available: www.gdal.org

TABLE I
TEN LAND COVER CLASSES, THE NUMBER OF SUBCLASSES, AND THE TRAINING DATA (NUMBER OF POLYGONS, THEIR AREA, AND PERCENTAGE)

Class	# of subclasses	# of polygons	# of pixels (×1000)	Area (km^2)	%
1: Water bodies	3	22	13.68	3.1	28.1
2: Wetland	2	11	0.67	0.1	1.4
3: Artificial surfaces	4	36	1.66	0.4	3.4
4: Forest	4	36	7.39	1.7	15.2
5: Shrubland	1	16	6.09	1.4	12.5
6: Grassland	2	13	2.18	0.5	4.5
7: Bareland	3	25	2.77	0.6	5.7
8: Cultivated land	8	146	13.64	3.1	28.0
9: Cloud/shadows	2	15	0.62	0.1	1.3
10: Snow/ice	2	0	0	0.0	0.0
Total	31	320	48.71	11.0	100

interfaces, thus allowing on-demand submission of queries that search, retrieve, subset, and process multidimensional arrays of large size. Moreover, it adds geographic and temporal coordinate system support. The result of the execution of the WCPS query produced by the service on the *Rasdaman* server is the multispectral image to be classified by the service.

C. Land Cover Map Production

In this section, the core process of the land cover mapping service and, in particular, the "server-side scripts" component is described (Fig. 1).

A main server-side script (*task manager*) written in *php* programming language orchestrates the individual scripts, system components, and procedures that are involved in the production of land cover maps. It acts as the intermediate connection point between the service's web client (Section II-D), the *Rasdaman* server, and the various programming scripts that power the land cover service. When invoked from the Web Client, the *task manager* parses the user input (i.e., dataset metadata and task to fulfill) and retrieves through WCPS queries (data subsetting and extraction) the necessary dataset from *Rasdaman*, which is stored as a temporary file in the server. Once the dataset has become available, the image classification functions are executed. The resulting classified image is forwarded to a C program that merges the detected classes and creates the final land cover product. The product is then georeferenced and stored temporarily on the server.

D. Client

The Web Client of the system is heavily based on GeoExt and OpenLayers javascript libraries. It utilizes them for managing service's graphical interface and the interaction with the user, as well as, for placing the produced land cover maps as layers inside a map. In order to achieve this, metadata need to be determined. These metadata are the coordinates of the vertices of the bounding box that defines an area of interest (AOI) that the user has specified and the name of the collection in the *Rasdaman* database that hosts the image which contains

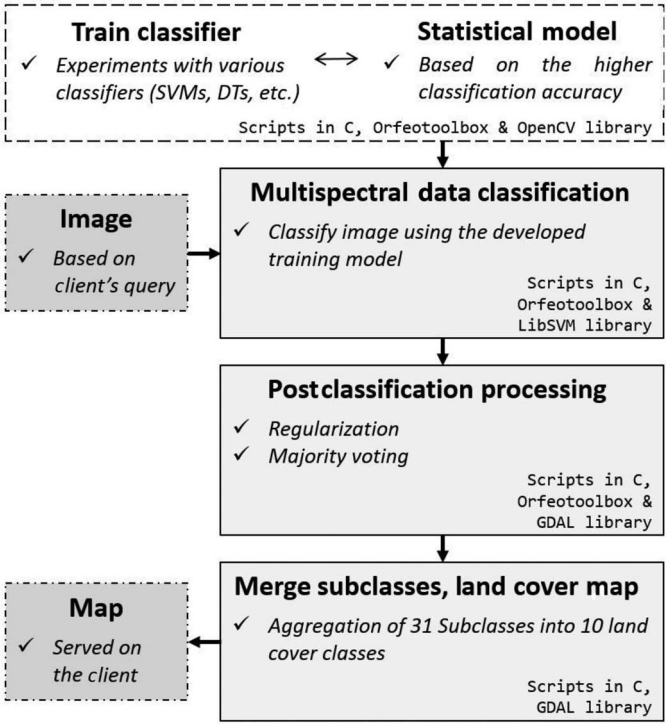

Fig. 2. Flowchart with the developed land cover classification procedure.

the given AOI. It should be noted that the AOI can take any arbitrary shape and there is not any limitation regarding the number of polygon nodes.

The coordinates of the bounding box that define an AOI are recorded on completion of the definition of the AOI by the user. Then, a server-side script utilizes them so as to find the collection that contains the AOI. The returned metadata are then employed to form the WCPS query that will be sent to the *task manager*. Once all necessary information has been determined, the web client sends it asynchronously to the *task manager* and waits for the response. In order to handle request timeout on the client side, a request timeout rule was also implemented. The implementation takes into account the areal extent of the AOI defined by the user and sets the timeout parameter accordingly.

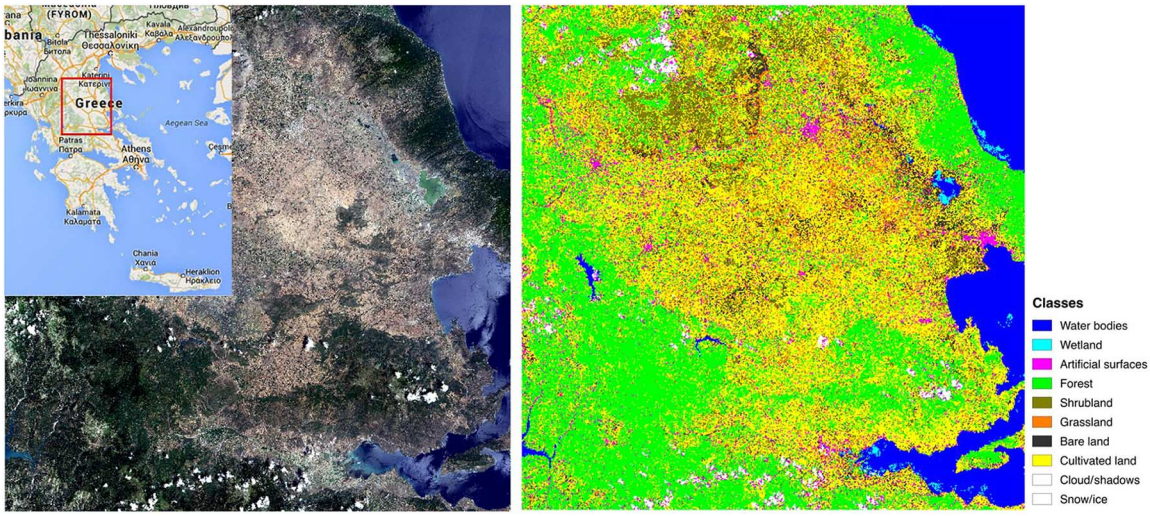

Fig. 3. Land cover classification results from the developed geospatial service. The large multispectral mosaic, from central Greece, with 15-m ground resolution of 8684×8676 pixels, covering 16 952 km^2, was classified, after approximately 10 min. Images were acquired on June 25, 2013. The quantitative evaluation indicated an OA rate of 82.8% with a Kappa coefficient of 0.801 (Table II).

The timeout parameter is set between 1 and 3 min depending on the exact areal extent. As far as very large AOIs are concerned, longer processing time is required. For such cases, the user receives an e-mail with a url address pointing to the result.

Upon receipt of the response (land cover map), the coordinates that define the areal extent of the AOI are again used through the image layer capability offered by OpenLayers, in order to place the land cover layer at its correct location in the Client's OpenLayers map. This allows for spectral data analysis in a web environment. The resulting land cover map can therefore be compared with overlapping visual imagery and geodata from various sources.

III. Classification Framework

The classification framework was developed to reflect the major land types in Greece with reference to major land cover products and mapping practices [1], [39], [40]. Focusing on delivering land cover classification maps with 15-m ground resolution, a two-stage classification procedure was employed, which discriminates initially 31 subclasses and then after a postprocessing and a grouping procedure delivers 10 aggregated classes. A flowchart with the developed land cover classification procedure is shown in Fig. 2.

The 31 subclasses, which were derived through a comprehensive, time-consuming manual selection of numerous training data all over Greece in various images from different dates, represent the land cover types, terrain objects, and surfaces that can be found in Greece at this particular spatial scale. It should be noted that experiments started with multiple unsupervised classification in various images, all over Greece, with 60 classes. After a demanding, thorough, trial and error procedure, these 31 subclasses were selected as representative. This procedure included study of the spectral properties, seasonal behaviors, and a cross validation from higher resolution imaging (e.g., google earth) and vector geospatial open data (e.g., geodata.gov.gr) [41].

A. Land Cover Classes and Hierarchy

These subclasses belong to the 10 major land cover types that are shown in Table I, i.e., *water bodies, wetland, artificial surfaces, forest, shrubland, grassland, bareland, cultivated land, cloud/shadows, and snow/ice*. For each class, the number of the corresponding subclasses and the number of the manually annotated training polygons are also shown in Table I. The *cultivated land* class includes the larger number of both subclasses and training samples due to the diverse spectral properties of the corresponding subclasses (e.g., row croplands, orchards, small grain croplands, fallow croplands, and harvested croplands). *Artificial surfaces, forest, water bodies,* and *bareland* also possess a number of subclasses and polygons.

In terms of the area that the training samples are covering (in square kilometer), the *cultivated land* and the *water bodies* are holding the larger share, i.e., 28% of all samples. *Forest* and *shrubland* follow with approximately 15% and then the rest of the classes with less than 5% each. In total, the training data covered approximately 11 km^2, i.e., approximately 49 000 pixels (Table I).

B. Training

All manually annotated samples, which were digitized in various images of different dates, were employed for the training of the SVM classifier. Our implementation was based on the Orfeo Toolbox[7] [42], OpenCV.[8] and LibSVM [43] software libraries, which provide an efficient framework for the integration of existing machine learning algorithms. Briefly, Orfeo Toolbox (OTB) is an open source library which is focusing on the analysis of large remote sensing images, OpenCV is a computer vision library which is focusing on computational efficiency and real-time applications and LibSVM is

[7][Online]. Available: http://www.orfeo-toolbox.org/SoftwareGuide/
[8][Online]. Available: http://opencv.org/documentation.html

TABLE II
CONFUSION MATRIX FOR THE QUANTITATIVE EVALUATION OF THE DELIVERED MULTITEMPORAL LAND COVER CLASSIFICATION MAP OF FIG. 3

	Reference data											
# of pixels	Water bodies	Wetland	Artificial surfaces	Forest	Shrub-land	Grass-land	Bareland	Culti-vated land	Cloud/ shadows	Snow/ ice	Total	**UA (%)**
Classification												
Water bodies	**3004**	0	23	0	0	0	0	0	0	0	3027	99.2
Wetland	14	**1296**	0	0	0	0	0	0	0	0	1310	98.9
Artificial surfaces	0	0	**1103**	0	0	21	95	15	125	0	1359	81.2
Forest	0	66	0	**3743**	9	0	0	16	0	0	3834	97.6
Shrubland	0	0	14	143	**3370**	6	20	643	0	0	4196	80.3
Grassland	0	0	32	0	1	**1069**	78	381	0	0	1561	68.5
Bareland	0	0	542	0	0	134	**942**	15	56	0	1689	55.8
Cultivated land	0	0	0	1015	58	83	247	**3107**	0	0	4510	68.9
Cloud/shadows	0	0	1	0	0	0	0	0	**2698**	0	2699	99.9
Snow/ice	0	0	0	0	0	0	0	0	0	**0**	0	0
Total	3018	1362	1715	4901	3438	1313	1382	4177	2879	0	**24 185**	
PA (%)	99.5	95.1	64.3	76.4	98	81.4	68.2	74.4	93.7	0		
Overall accuracy = 82.8%, Kappa coefficient = 0.801												

an integrated software tool for support vector classification, multiclass classification, regression, and distribution estimation (Fig. 2).

The training/prediction model was built based on image statistics including all multispectral bands, the normalized difference vegetation index (NDVI), and the normalized difference water index (NDWI). The Kappa coefficient, which was employed for validating the training model, indicated a Kappa coefficient above 80%.

C. Two-Stage Classification

The developed geospatial service was designed to deliver in a systematic and operational manner, validated land cover maps in near real-time. Toward addressing such challenges, a two-stage classification strategy was implemented in Fig. 2. First, experimenting on a national scale with LANDSAT 8 datasets from the beginning of its operational orbit (i.e., 4/2013) and validating land cover maps at 15-m resolution for 31 land cover subclasses was neither feasible nor within the perspectives of our work. Moreover, in all experiments, on all datasets, on all dates, and by employing different classifiers (e.g., SVMs and decision trees), the overall classification accuracy for the 31 subclasses was above 70% reaching approximately 80%.

Therefore, by employing a standard postprocessing (i.e., majority voting filter) technique and grouping the subclasses into 10 land cover classes (Table I), the framework managed to systematically deliver land cover maps with an overall accuracy (OA) of above 80% across different datasets and dates. This was acceptable for the current alpha system release.

Moreover, in all experiments, the SVM classifier [43] using the one-against-one strategy delivered the highest accuracy rates with an acceptable computational cost, allowing near real-time computations on the server side. In order to slightly refine the initial classification results and enhance its classification accuracy, a standard majority voting filtering (in a 3×3 window) was employed, which ameliorated the OA in all experiments by 1%–5%.

IV. EVALUATION

A. Experimental Results

The experimental validation was performed by applying the developed service on numerous regions of different sizes, on different dates, of various land cover types and complex terrain objects. A demo with the overall functionality and performance of the developed system can be found here.[9] In particular, the alpha version of our system is currently hosted on an eight-core machine with 32-GB RAM running Debian GNU/Linux (Release 7.5), Apache Tomcat 6, and the open source *Rasdaman* community version 8.5. Under this demonstration environment, the stored and preprocessed remote sensing data are fully covering the Greek territory providing every approximately 16-day satellite imagery from the beginning of the LANDSAT 8 operational orbit (April 2013). Obviously, the current hardware cannot support numerous concurrent users and a migration to a cloud environment has already been scheduled.

Along with the collection of the training samples for the 31 subclasses in Table I, the collection of the ground truth data was performed with another independent procedure. In particular, ground truth/ reference polygons were manually collected and annotated from various images and acquisition dates. About 50 000 pixels were used for the training process and about 50 000 for the quantitative evaluation, all referring to the 31 land cover classes. For the quantitative evaluation, the reference polygons were merged, respectively, to match the 10 land cover classes of the end product. These manually annotated polygons were verified by intensive comparisons from very-high resolution imaging data (e.g., orthophotos and google earth) and vector geospatial open data (e.g., geodata.gov.gr).

High-resolution imaging data and vector information were also employed for the qualitative evaluation of the produced land cover maps. Such systematic qualitative examinations of the delivered maps and comparisons (both qualitative and quantitative) with other maps and data sources indicated high

[9][Online]. Available: http://users.ntua.gr/karank/Demos/WebLC.html

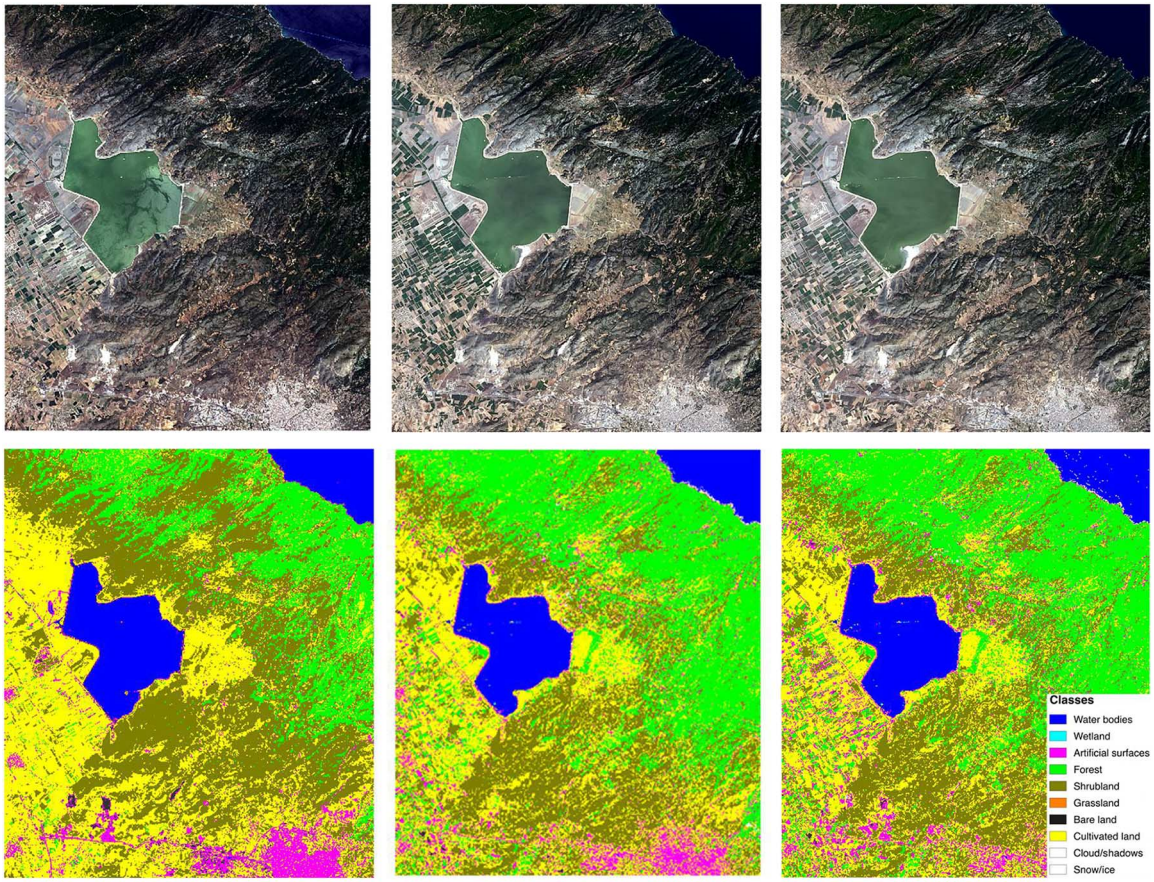

Fig. 4. Land cover classification results from the developed geospatial service over a multitemporal dataset. Each multispectral image of 1446×1682 pixels, covering 547 km^2 was processed and delivered to the client after approximately 22 s. The quantitative evaluation indicated an OA rate above 87% in all cases (Table III).

multitemporal and spatially distributed confidence values for the developed service. In all experiments, the qualitative assessment was generally in accordance with the quantitative assessment. During the evaluation of the first classification step (detection of 31 land cover classes), the performed assessment was mainly based on the quantitative results of various confusion matrices, the reported Kappa coefficient, and OA rates. In particular, the OA for this initial classification step was above 64%, including scenes and dates with cloud coverage above 15%. The lowest accuracy rates were found in certain subclasses, i.e., *artificial surfaces*, *cultivated land*, *bareland*, and *forest*. In particular, areas belonging to the dense urban fabric subclass of the *Artificial surfaces* class were systematically misclassified to the white sand or to the quarries subclasses of the *bareland* class.

In Fig. 3, the delivered land cover classification map from the developed geospatial service is shown over a region in Central Greece. A large multispectral mosaic of LANDSAT 8 data, with 15-m ground resolution of 8684×8676 pixels, covering almost 17 000 km^2 was processed by the system and the land cover map was delivered after approximately 10 min. The corresponding LANDSAT 8 images were acquired on June 25, 2013. The performed quantitative evaluation indicated an OA rate of 82.8% with a Kappa coefficient of 0.801 (Table II). It should be noted that the initial reported classification accuracy for the 31 subclasses was 72.1% with a Kappa coefficient of 0.713.

The performance of the land cover service was demonstrated over this particular extended scene because it was representative of all land cover classes, subclasses, and terrain objects in Greece, covering thousands of square kilometer. The results from its quantitative evaluation were also representative of the numerous experimental results performed over different multitemporal scenes. In particular, the quantitative assessment (Table II) indicated that the *water bodies* class has been detected with the highest rates (above 99%) based both on the producer's accuracy (PA), which indicates the probability of a reference pixel being correctly classified, and the user's accuracy (UA), which indicates the probability of a pixel in the map actually representing this category on the ground. This is in accordance with the literature [1], [44]. High accuracy rates (above 93%) have also been obtained for the *cloud/shadows* class for both PA and UA. The *shrubland* class has also been systematically detected with very high PA rates (98% in this case) and both the *wetland* and *forest* classes with very high UA rates (above 97% in this case).

The lowest PA has been reported for the *artificial surfaces* and *bareland* classes (with 64.3% and 68.2%, respectively).

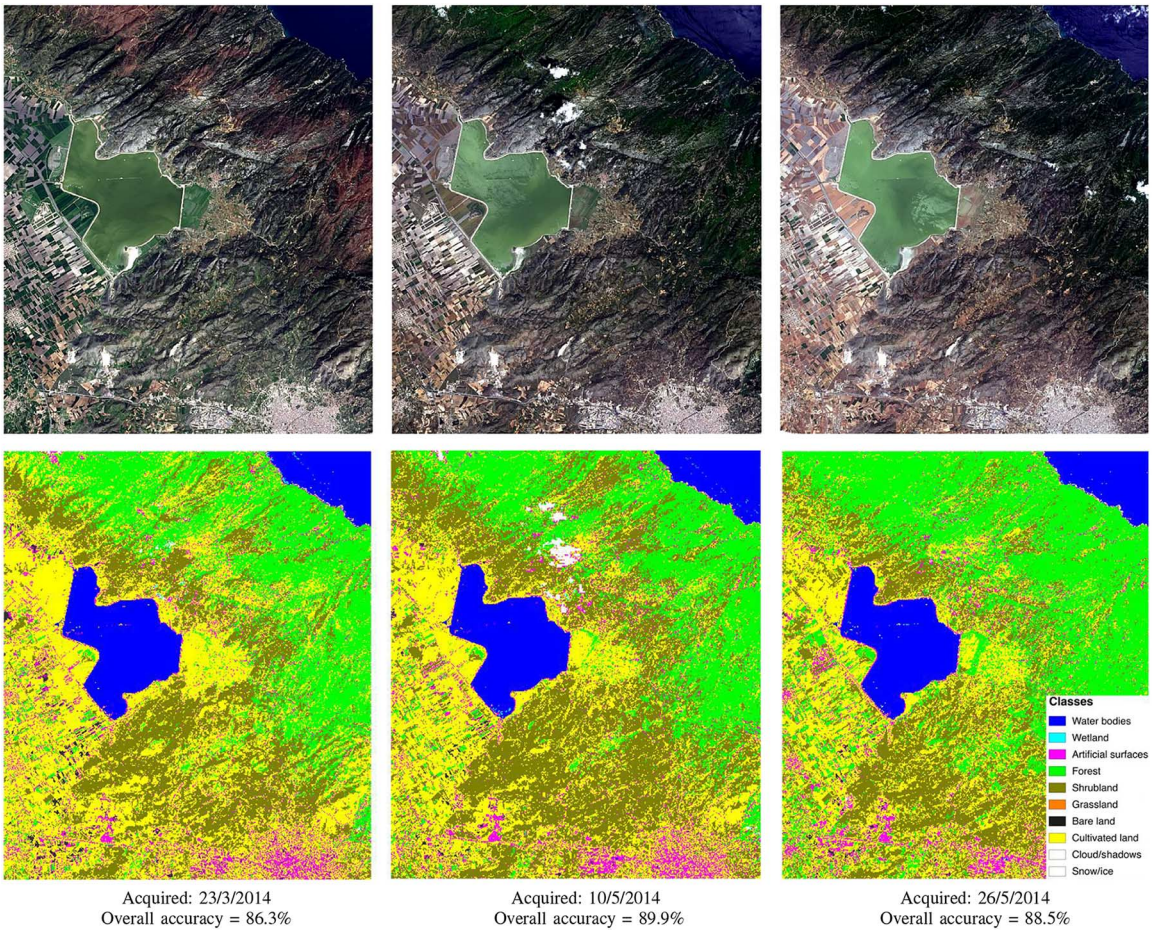

Fig. 5. Land cover classification results from the developed geospatial service over a multitemporal dataset. Each multispectral image of 1446×1682 pixels, covering 547 km^2 was processed and delivered to the client after approximately 22 s. The quantitative evaluation indicated an OA rate above 86% in all cases (Table III).

The *bareland* class was also calculated with the lowest UA at 55.8% (Table II). In particular, in all conducted experiments, pixels from the *artificial surfaces* class have been systematically misclassified as *bareland*. This is mainly inherited by the initial classification step with the 31 subclasses. *Bareland* pixels were also misclassified as *cultivated land*, while *cultivated land* pixels were mixed with the shrubland ones.

In addition, in order to demonstrate the performance of the land cover service for multitemporal data, experimental results from six different acquisition dates are shown in Figs. 4 and 5. In the top row of both figures, a natural color composite is shown and the corresponding land cover map, which was delivered by the system, is shown in the bottom. Land cover classes' colors are the same ones as in Fig. 3. This particular region, covering approximately 550 km^2, contains almost all land cover classes, mountainous, rural and urban regions. The size of the images was 1446×1682 pixels and in all cases, the system delivered the land cover product to the client after approximately 22 s. The time that the developed service required for the entire process (from the client request to the delivered map on the client) was, in all cases, measured using the *Firebug 2.0.4* application. In Fig. 6, an example of a WCPS request script is shown which retrieves an image from the *Rasdaman* database.

Qualitative comparison between the experimental results and different color composites and other higher resolution images showed that the delivered multitemporal land cover maps agree spatially with the actual land cover types in the ground. This observation is verified by the quantitative assessment that indicated OA rates ranging from 86% to 89% (Table III). In particular, there is a rather small difference between the calculated accuracy rates at all dates and this can indicate the stable performance of the developed classification procedure. A small difference in the OA rates during the initial classification with the 31 subclasses is also reported in all cases, i.e., from 75% to 81%. It should be mentioned that the lowest accuracy rates during the initial step do not necessarily result into the lowest accuracy rates at the end-product. Such a correlation has not been observed in any of our experiments. For example, the initial 31 subclasses have been detected with a relative low OA (75%) for May 10, 2014, while after the postprocessing and merging procedure, the evaluation indicated an OA of 89%.

In general, the corresponding confusion matrices from all dates of Figs. 4 and 5 agree with the systematic observations regarding the performance of the developed classification procedure as previously reported. The worth-mentioning differences are the misclassified pixels from the *artificial surfaces*

```
for c in (image_collection) return encode (
   trim(
      struct {
         band1 : (unsigned short) c.band1;
         band2 : (unsigned short) c.band2;
         .
         .
         band_i : (unsigned short) c.band_i
      },
      {x(low:high), y(low:high)}
   ),
   "tiff"
)
```

Fig. 6. Example of a WCPS request script which retrieves from the *Rasdaman* database an image named "c" as a multiband geotiff file, in order to form the input for the classification procedure.

TABLE III
QUANTITATIVE EVALUATION OF THE LAND COVER CLASSIFICATION MAPS OF FIGS. 4 AND 5

Acquisition date	Initial classification 31 subclasses		Delivered product 10 land cover classes	
	OA (%)	Kc	OA (%)	Kc
25/06/2013	80.1	0.764	87.7	0.829
27/08/2013	81.2	0.781	87.9	0.839
12/09/2013	78.3	0.788	88.6	0.847
23/03/2014	79.3	0.767	86.3	0.815
10/05/2014	75.2	0.791	89.9	0.851
26/05/2014	80.3	0.796	88.5	0.849

class to the *cultivated land* and vice versa, apart from June 25, 2013 case. In a smaller extent, *shrubland* pixels have also been misclassifed as *grassland* ones. These differences are mostly due to the particular spectral variation among the different land cover classes in this specific region and available ground truth polygons (reference pixels employed for the evaluation were approximately 2100 in total).

V. CONCLUSION AND PERSPECTIVES

In this paper, a land cover classification web service has been designed, developed, and evaluated toward the efficient and automated processing of high-resolution satellite data. The core of the platform consists of the *Rasdaman* Array DBMS for big raster data storage and the OGC WCPS interface standard for data querying. The WebGIS client is based on the OpenLayers and GeoExt javascript libraries. Currently, the system is fully covering Greece with LANDSAT 8 multispectral data, from the beginning of its operational orbit. Datasets are stored and preprocessed automatically in our hardware. The automated classification framework is based on a comprehensive statistical training model and on a multiclass SVM classifier, integrating advanced remote sensing and computer vision libraries like GDAL, OTB, and OpenCV. The framework has been trained to classify pansharpened images at 15-m ground resolution toward the detection of 31 different land cover classes which can address at this particular scale the diverse land cover types that can be found in Greece. The final product of our system, in its current alpha release, delivers, after a postclassification and merging procedure, multitemporal land cover maps with 10 land cover classes. The performed quantitative evaluation has indicated a land cover classification accuracy above 80%.

Toward the operational use of the service and the forthcoming beta release, comprehensive validation on numerous scenes, which will also include ground truth data from field surveys, will be carried out. Furthermore, the migration of the platform to a cloud environment is already scheduled. Moreover, since the system has been designed to operate with any processing and classification algorithm, a benchmark of state-of-the-art approaches on training, learning, and classification on the continuously increasing dataset (and ground truth) is to be performed. Last but not least, toward improving significantly current computation times, parallel computing formulations and GPU implementations will be also employed.

REFERENCES

[1] J. Chen *et al.*, "Global land cover mapping at 30m resolution: A POK-based operational approach," *Int. J. Photogramm. Remote Sens.*, vol. 103, pp. 7–27, 2014.
[2] E. Zell, A. Huff, A. Carpenter, and L. Friedl, "A user-driven approach to determining critical earth observation priorities for societal benefit," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 6, pp. 1594–1602, Dec. 2012.
[3] C. Giri, B. Pengra, J. Long, and T. Loveland, "Next generation of global land cover characterization, mapping, and monitoring," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 25, pp. 30–37, 2013.
[4] A. J. Plaza and C. I. Chang, *High Performance Computing in Remote Sensing*. Boca Raton, FL, USA: Chapman & Hall/CRC Press, 2007.
[5] J. Oosthoek *et al.*, "Planetserver: Innovative approaches for the online analysis of hyperspectral satellite data from Mars," *Adv. Space Res.*, vol. 53, pp. 219–244, 2013.
[6] P. Cappelaere, S. Sanchez, S. Bernabe, A. Scuri, D. Mandl, and A. Plaza, "Cloud implementation of a full hyperspectral unmixing chain within the NASA web coverage processing service for EO-1," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 408–418, Apr. 2013.
[7] Y. Ma, L. Wang, P. Liu, and R. Ranjan, "Towards building a data-intensive index for big data computing—A case study of remote sensing data processing," *Inf. Sci.*, vol. 319, pp. 171–188, 2014.
[8] C. Nikolaou *et al.*, "Big, linked and open data: Applications in the german aerospace center," in *The Semantic Web: ESWC 2014 Satellite Events*. New York, NY, USA: Springer, 2014, pp. 444–449.
[9] Y. Ma, L. Wang, A. Zomaya, D. Chen, and R. Ranjan, "Task-tree based large-scale mosaicking for massive remote sensed imageries with dynamic DAG scheduling," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 8, pp. 2126–2137, Aug. 2014.
[10] B. C. Pijanowski, A. Tayyebi, J. Doucette, B. K. Pekin, D. Braun, and J. Plourde, "A big data urban growth simulation at a national scale: Configuring the GIS and neural network based land transformation model to run in a high performance computing (HPC) environment," *Environ. Modell. Softw.*, vol. 51, pp. 250–268, 2014.
[11] A. Karmas, A. Tzotsos, and K. Karantzalos, "Scalable geospatial web services through efficient, online and near real-time processing of earth observation data," in *Proc. IEEE Int. Conf. Big Data Comput. Serv. Appl.*, 2015, pp. 194–201.
[12] A. J. Plaza, "Special issue on architectures and techniques for real-time processing of remotely sensed images," *J. Real-Time Image Process.*, vol. 4, no. 3, pp. 191–193, 2009.
[13] P. Zhao, T. Foerster, and P. Yue, "The geoprocessing web," *Comput. Geosci.*, vol. 47, pp. 3–12, 2012.
[14] P. Yue, L. Di, Y. Wei, and W. Han, "Intelligent services for discovery of complex geospatial features from remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 83, pp. 151–164, 2013.
[15] A. Karmas, A. Tzotsos, and K. Karantzalos, "Big geospatial data for environmental and agricultural applications," in *Big Data Concepts, Theories and Applications*, S. Yu and S. Guo, Eds. New York, NY, USA: Springer, 2016.

[16] P. Yue *et al.*, "GeoPW: Laying blocks for the geospatial processing web," *Trans. GIS*, vol. 14, no. 6, pp. 755–772, 2010.

[17] W. Han, Z. Yang, L. Di, and P. Yue, "A geospatial web service approach for creating on-demand cropland data layer thematic maps," *Trans. Amer. Soc. Agric. Biol. Eng.*, vol. 57, no. 1, pp. 239–247, 2014.

[18] P. Baumann, "Management of multidimensional discrete data," *Int. J. Very Large Data Bases*, vol. 4, no. 3, pp. 401–444, 1994.

[19] P. Baumann, "Array databases and raster data management," in *Encyclopedia of Database Systems*, T. Ozsu and L. Liu, Eds. New York, NY, USA: Springer, 2009.

[20] P. Baumann, "The OGC web coverage processing service (WCPS) standard," *GeoInformatica*, vol. 14, no. 4, pp. 447–479, 2010.

[21] A. Karmas, K. Karantzalos, and S. Athanasiou, "Online analysis of remote sensing data for agricultural applications," in *Proc. OSGeo's Eur. Conf. Free Open Source Softw. Geospat.*, Jul. 2014, pp. 1–9.

[22] K. Karantzalos, A. Karmas, and A. Tzotsos, "RemoteAgri: Processing online big earth observation data for precision agriculture," in *Proc. Eur. Conf. Precis. Agric.*, 2015, pp. 421–428.

[23] K. Evangelidis, K. Ntouros, S. Makridis, and C. Papatheodorou, "Geospatial services in the cloud," *Comput. Geosci.*, vol. 63, pp. 116–122, 2014.

[24] D. Fustes, D. Cantorna, C. Dafonte, B. Arcay, A. Iglesias, and M. Manteiga, "A cloud-integrated web platform for marine monitoring using GIS and remote sensing. Application to oil spill detection through SAR images," *Future Gener. Comput. Syst.*, vol. 34, pp. 155–160, 2014.

[25] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The rise of 'big data' on cloud computing: Review and open research issues," *Inf. Syst.*, vol. 47, pp. 98–115, 2015.

[26] C. Lee, S. Gasster, A. Plaza, C.-I. Chang, and B. Huang, "Recent developments in high performance computing for remote sensing: A review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 3, pp. 508–527, Sep. 2011.

[27] X. Su, J. Wu, B. Huang, and Z. Wu, "GPU-accelerated computation for electromagnetic scattering of a double-layer vegetation model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 4, pp. 1799–1806, Aug. 2013.

[28] M. C. Hansen *et al.*, "High-resolution global maps of 21st-century forest cover change," *Science*, vol. 342, no. 6160, pp. 850–853, 2013.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Neural Inf. Process. Syst. (NIPS'12)*, 2012, pp. 1097–1105.

[30] B. Liu, E. Blasch, Y. Chen, D. Shen, and G. Chen, "Scalable sentiment classification for big data analysis using Naive Bayes Classifier," in *Proc. IEEE Int. Conf. Big Data*, Oct. 2013, pp. 99–104.

[31] A. Ferran, S. Bernabe, P. Rodriguez, and A. Plaza, "A web-based system for classification of remote sensing data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 4, pp. 1934–1948, Aug. 2013.

[32] M. Muja and D. Lowe, "Scalable nearest neighbor algorithms for high dimensional data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2227–2240, Nov. 2014.

[33] P. Blanchart and M. Datcu, "A semi-supervised algorithm for auto-annotation and unknown structures discovery in satellite image databases," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 3, no. 4, pp. 698–717, Dec. 2010.

[34] D. Espinoza-Molina and M. Datcu, "Earth-observation image retrieval based on content, semantics, and metadata," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 11, pp. 5145–5159, Nov. 2013.

[35] P. Blanchart, M. Ferecatu, S. Cui, and M. Datcu, "Pattern retrieval in large image databases using multiscale coarse-to-fine cascaded active learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1127–1141, Apr. 2014.

[36] J. Sevilla, S. Bernabe, and A. Plaza, "Unmixing-based content retrieval system for remotely sensed hyperspectral imagery on GPUS," *J. Supercomput.*, vol. 70, pp. 1–12, 2014.

[37] P. Baumann, "Rasdaman: Array databases boost spatio-temporal analytics," in *Proc. 5th Int. Conf. Comput. Geospat. Res. Appl. (COM. Geo)*, Aug. 2014, pp. 54–54.

[38] A. Aiordachioaie and P. Baumann, "Petascope: An open-source implementation of the OGC WCS geo service standards suite," in *Scientific and Statistical Database Management*, vol. 6187, M. Gertz and B. Ludascher, Eds. New York, NY, USA: Springer, 2010, pp. 160–168.

[39] B. Waske, M. Chi, J. Atli Benediktsson, S. van der Linden, and B. Koetz, "Algorithms and applications for land cover classification—A review," in *Geospatial Technology for Earth Observation*, D. Li, J. Shan, and J. Gong, Eds. New York, NY, USA: Springer, 2009, pp. 203–233.

[40] M. Hansen and T. Loveland, "A review of large area monitoring of land cover change using Landsat data," *Remote Sens. Environ.*, vol. 122, pp. 66–74, 2012.

[41] G. Foody, A. Mathur, C. Sanchez-Hernandez, and D. Boyd, "Training set size requirements for the classification of a specific class," *Remote Sens. Environ.*, vol. 104, p. 1–14, 2006.

[42] E. Christophe and J. Inglada, "Open source remote sensing: Increasing the usability of cutting-edge algorithms," *IEEE Geosci. Remote Sens. Newslett.*, pp. 9–15, 2009.

[43] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 27:1–27:27, 2011.

[44] Z. Zhu and C. E. Woodcock, "Continuous change detection and classification of land cover using all available Landsat data," *Remote Sens. Environ.*, vol. 144, pp. 152–171, 2014.

Konstantinos Karantzalos (SM'05) received the Engineering Diploma degree in survey engineer from the National Technical University of Athens (NTUA), Athens, Greece, and the Ph.D. degree in remote sensing from NTUA, in collaboration with Ecole Nationale de Ponts et Chaussees (CERTIS, ENPC), Marne-la-Vallée, France, in 2007.

In 2007, he joined the Department of Applied Mathematics, Ecole Centrale de Paris, Paris, France, as a Postdoctoral Researcher working with Prof. Nikos Paragios (Head of the Center for Visual Computing). He is currently an Assistant Professor with the Remote Sensing Laboratory, National Technical University of Athens, joining the Remote Sensing Laboratory (Head: Prof. Demetre Argialas). He is also with i-SENSE Group, Institute of Communication and Computer Systems (ICCS), NTUA. He has several publications in top-rank international journals and conferences and a number of awards and honors for his research contributions. He has more than 15 years of research experience, involved with more than 14 EU and national excellence/competitive research projects. His research interests include geoscience and remote sensing, big data, hyperspectral image analysis, computer vision and pattern recognition, environmental monitoring, and precision agriculture. http:// users.ntua.gr/karank

Dimitris Bliziotis received the B.Sc. degree in rural and surveying engineering from the National Technical University of Athens (NTUA), Athens, Greece, in 2014. He is currently pursuing the M.Sc. degree in geoinformatics at the same university.

Since 2013, he has been a Researcher with the Remote Sensing Laboratory, NTUA. He is a Skilled Programmer and Free and Supporter of Open Source software. His research interests include remote sensing, geoinformatics, artificial intelligence, and machine learning.

Athanasios Karmas (M'15) received the Diploma degree in electrical and computer engineering from the National Technical University of Athens (NTUA), Athens, Greece, in 2014. He is currently pursuing the M.Sc. degree in geoinformatics at the same university.

Since 2013, he has been a Researcher with the Remote Sensing Laboratory, NTUA. He is a passionate Supporter of Free and Open Source software. His research interests include array DBMS, big data, cloud computing, geoinformatics, as well as earth observation data management and geospatial applications. http://users.ntua.gr/karank