Applied Geography 31 (2011) 533–544

Contents lists available at ScienceDirect

Applied Geography

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

Comparison of methods for land-use classification incorporating remote sensing and GIS inputs

Offer Rozenstein, Arnon Karnieli*

The Remote Sensing Laboratory, Jacob Blaustein Institute for Desert Research, Ben-Gurion University of the Negev, Sede Boker Campus 84990, Israel

Keywords:
Remote-sensing
Land-use
GIS
Negev

A B S T R A C T

Over the last few decades, dramatic land-use changes have occurred throughout Israel. Previously-grazed areas have been afforested, converted to irrigated or rain-fed agriculture, turned into natural reserves, often used as large military training sites, converted to rural and urban settlements, or left unused. Land-use maps provided by the Israeli governmental are more detailed for agricultural and urban land-use classes than for others. While rangelands still account for a substantial part of the northern Negev, their extent today is not well defined. In light of continuous land-use changes and lack of regard to rangelands in existing land-use maps, there is a need for creating a current land-use information database, to be utilized by planners, scientists, and decision makers. Remote-sensing (RS) data are a viable source of data from which land-use maps could be created and updated efficiently. The purpose of this work is to explore low-cost techniques for combining current satellite RS data together with data from the Israeli Geographic Information System (GIS) in order to create a relatively accurate and current land-use map for the northern Negev. Several established methods for land-use classification from RS data were compared. In addition, ancillary land-use data were used to update and improve the RS classification accuracy within a GIS framework. It was found that using a combination of supervised and unsupervised training classes produces a more accurate product than when using either of them separately. It was also found that updating this product using ancillary data and GIS techniques can improve the product accuracy by up to 10%. The final product's overall accuracy was 81%. It is suggested that applying the presented technique for more RS images taken at different times can facilitate the creation of a database for land-use changes.

Introduction

The observed biophysical cover of the earth's surface, termed land-cover, is composed of patterns that occur due to a variety of natural and human-derived processes. Land-use, on the other hand, is human activity on the land, influenced by economic, cultural, political, historical, and land-tenure factors. Remotely-sensed data (i.e., satellite or aerial imagery) can often be used to define land-use through observations of the land-cover (Brown, Pijanowski, & Duh, 2000; Karl & Maurer, 2010). Up-to-date land-use information is of critical importance to planners, scientists, resource managers, and decision makers.

One way to extract land-use information from remote-sensing data is through visual interpretation. However, visual interpretation is limited to a single band or a three-band (RGB) color composite. Manual digitization of land-use patches is extremely tedious as well as subjective (Bolstad, Gessler, & Lillesand, 1990). Therefore, automatic classification of remote sensing is more suitable for mapping land-use in a large area. While land-use and land-cover patterns may be obvious to an image interpreter, automatically mapping them could be difficult because automated classification techniques do not possess the superior pattern recognition capabilities of the human brain (Hudak & Brockett, 2004). When automatically classifying a complex landscape from remote-sensing imagery, it is challenging to achieve an accurate classification (Manandhar, Odeh, & Ancev, 2009). It has been claimed before that the eastern Mediterranean landscapes are considered the most heterogeneous of all (Alrababah & Alhamad, 2006). Therefore, classifying the landscape in this region is not a trivial task.

Nevertheless, previous studies show that Landsat Thematic Mapper (TM) images with the spatial resolution of 30 m are sufficient to accurately classify a large variety of landscapes from the homogeneous tropical landscapes to the heterogeneous Mediterranean landscapes (Alrababah & Alhamad, 2006; Koutsias & Karteris,

* Corresponding author.
E-mail address: karnieli@bgu.ac.il (A. Karnieli).

0143-6228/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apgeog.2010.11.006

2003; Manandhar et al., 2009; Sader, Ahl, & Liou, 1995; Schulz, Cayuela, Echeverria, Salas, & Rey Benayas, 2010). Landsat has been providing a nearly continuous record of global land surface change since 1972 (Cohen & Goward, 2004). Currently, two Landsat sensors in orbit are operational: TM on board Landsat-5 and Enhanced Thematic Mapper Plus (ETM+) on board Landsat-7. Both sensors acquire measurements in all major portions of the solar electromagnetic spectrum (visible, near-infrared, and shortwave-infrared), providing significant advantage over less capable sensor systems. However, Landsat-7, launched in 1999, developed a problem with the scan-line corrector in 2003, leading to reduced data quality for land-use mapping applications (Powell, Pflugmacher, Kirschbaum, Kim, & Cohen, 2007). Today, Landsat-5, launched in 1984, has far exceeded its 3-year life expectancy but continues to provide quality data products, although it was expected to run out of fuel by late 2010 (Wulder et al., 2008). Landsat data are widely applied for land-use classification on a regional scale due to their relatively lower cost, longer history, and higher frequency of archives in comparison to other remote-sensing data sources.

It has been previously determined that satellite image classification results did not improve over a period of 15 years in spite of vigorous and creative efforts to establish new classification algorithms during this period (Wilkinson, 2005). Therefore, it was concluded there is little value in continued research efforts to improve classification algorithms in remote sensing (Manandhar et al., 2009). Recently, the trend amongst researchers has been to let geographical data "have a stronger voice" rather than let statistically-derived parameters dictate the analysis. Integration of remotely-sensed data with other sources of georeferenced information, such as previous land-use data, spatial texture, and digital elevation models (along with their derivatives: slope, aspect, etc.), geology, soils, hydrology, transportation network, vegetation, and climate enable greater classification accuracy to be achieved (Lillesand & Kiefer, 2000; Manandhar et al., 2009; Stefanov, Ramsey, & Christensen, 2001; Tateishi & Shalaby, 2007). The particular sources of data used and how and when they are employed in a given application are normally determined through a set of decision rules formulated by the image analyst. The integration of several data sources in a Geographic Information System (GIS) allows the analyst to develop a series of post-classification decision rules utilizing all the data sources in combination (Lillesand & Kiefer, 2000). The integration of remote-sensing data, GIS and "expert system" techniques to form Decision Support Systems (DSS) can provide better classification accuracies than any of the individual data sources used alone.

The purpose of this work was to explore low-cost techniques for land-use mapping. Landsat TM imagery was classified by two widely used and established classification approaches, and these two methods were combined and compared. Next, the hypothesis that integrating current satellite remote-sensing data together with data from the Israeli GIS will improve the land-use mapping significantly was tested. The land-use classification technique presented in this work can be used to produce information pertaining changes in land-uses, such as monitoring of land-use conversion and land degradation. The information could be further used to study the relations between land-use changes and other phenomena such as carbon fixation, biodiversity, climate change, and sustainable management of natural resources.

Specific objectives of the current study are:

1. To compare between supervised and unsupervised land-use classification techniques;
2. To examine whether combining signatures from both supervised and unsupervised training data (hybrid classification) provides significantly more accurate results then each approach separately;
3. To examine whether using a decision support system for updating the map based on expert knowledge and ancillary GIS data improves the classification accuracy significantly.

Study area

Located in the northern Negev, on the desert fringe, the study area (Fig. 1) is about 4000 km^2 in size. The study area's borders are delimited by Ramat-Hovav in the south, Yatir forest in the east, Kiryat-Gat and Ashkelon in the North, and the Mediterranean Sea, Gaza and Sinai in the west. This area is particularly diverse since it lies on the transition zone between arid, semi-arid, and Mediterranean climate zones. Average annual precipitation decreases along two climate gradients from north to south and from west to east; from more than 450 mm/year in the north-eastern part to less than 150 mm/year in the arid parts of the Negev (southern part of the study area). Examples for several distinct geomorphologic structures can be found in this area, including flood and alluvial plains, calcareous crust, crescentic dunes, and sand fields, with diverse parent rocks. As a result, there are many soil types including skeletal soils on unconsolidated materials (regosols), coarse desert alluvium, sand dunes, loess with hard pan (loessial sierozem), rocky desert soils (lithosols), rendzinas and terrarosa. This diversity of the environment results in diverse communities of flora and fauna.

During the last few decades, considerable land-use changes have occurred in Israel (Orenstein & Hamburg, 2009). Historically most of the northern Negev was Bedouin grazing territory. The geographical distribution of accessible rangeland in the Negev changed due to afforestation programs of the Jewish National Fund (JNF) and concentration of population in townships, along roads, and along water lines. As a result, the available rangeland areas and stocking rates in the Negev have fluctuated. Previously-grazed areas have been afforested, converted to irrigated or rain-fed agriculture, turned into natural reserves, often used as large military reserves, converted to rural and urban settlements, or left unused. During this period, the Negev pastoralists, nearly all of them Bedouin (of several tribes), have been affected by social, economic, policy and political factors that have brought about changes in demography, lifestyle, livelihoods, and dependence on livestock.

While rangelands still account for a substantial part of the northern Negev, their extent today is not well defined. Land-use maps provided by the Israeli government are more detailed for agricultural and urban land-use classes, than for others. Currently, there are only two national land-use maps available from the state of Israel: one from the Israeli Central Bureau of Statistics (CBS) and another from the Survey of Israel. Both maps lack a definition of rangelands. In addition, the Ministry of Agriculture does not possess any rangeland maps of the northern Negev, where land devoted for pasture is not defined by fences (Shmuel Friedman, Director of Open Spaces, Ministry of Agriculture, personal communication). In light of continuous land-use changes and lack of regard to rangelands in existing land-use maps, there is a need for creating a current land-use information database, to be utilized by planners, scientists, and decision makers.

Methodology

Initially, a Landat-5 TM image of the northern Negev was pre-processed and then classified in several ways using ERDAS IMAGINE 2010. Post-classification, a decision support system based on expert knowledge was used to update the classification products according to existing land-use databases using ArcGIS 9.3. The accuracy of each of the derived classification products was assessed in several ways, after which different product accuracies were

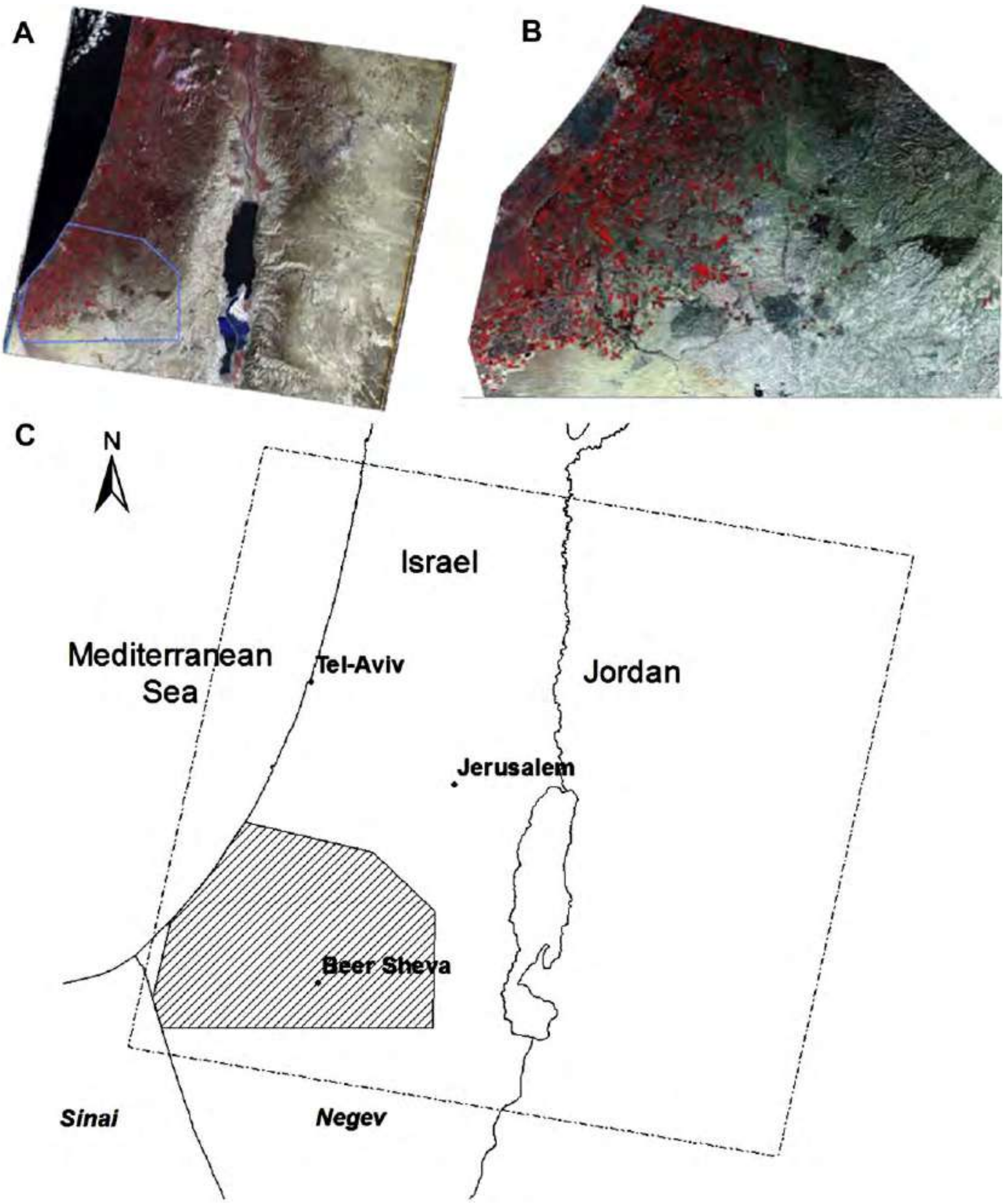

Fig. 1. (A) The research area over the Israeli Negev desert fringe (imposed over a false-color composite of Landsat-5 TM (RGB = 4,3,2)). Note the sharp contrast across the climatic transition zone between the arid and semi-arid zones; (B) Blowup of the study area; (C) Schematic map of the research area, Israel borders, and the Landsat image footprint.

compared using statistical means with STATISTICA 9.0. Fig. 2 presents a flowchart of the work.

Image pre-processing

The primary source for land-use classification is a Landsat-5 TM image (Path 174, Row 38) acquired on 30-Jan-2009. The selected area appears cloud free. Only the reflective bands (1–5 and 7) of the sensor were used in this study. Pre-processing of the image included one-step radiometric and atmospheric corrections using the dark-object subtraction method (Chavez, 1996; Song, Woodcock, Seto, Lenney, & Macomber, 2001) and the latest radiometric calibration coefficients published (Chander, Markham, & Helder, 2009).

To facilitate incorporation of ancillary data, the radiometrically and atmospherically corrected image was then geo-registered to an up-to-date orthophoto of Israel (Survey of Israel, 2009) using ERDAS AUTOSYNC feature; the Automatic Point Measurement (APM) software was used to generate 1095 Ground Control Points (GCPs) automatically and 40 manually generated GCPs were added in order to assure GCP distribution throughout the entire research area. Afterwards, the satellite image was resampled and projected to the Israeli Transverse-Mercator coordinate system with pixel size of 30 m using nearest-neighbor resampling and second order polynomial transformation equations. The total Root-Mean-Square Error (RMSE) achieved was 0.35 pixels, which is well under the conventional requirements of less than 1 pixel (Coppin, Jonckheere, Nackaerts, Muys, & Lambin, 2004; Huang et al., 2009; Mas, 1999) and even less than strict requirements of 0.5 pixels (Elvidge & Yuan,

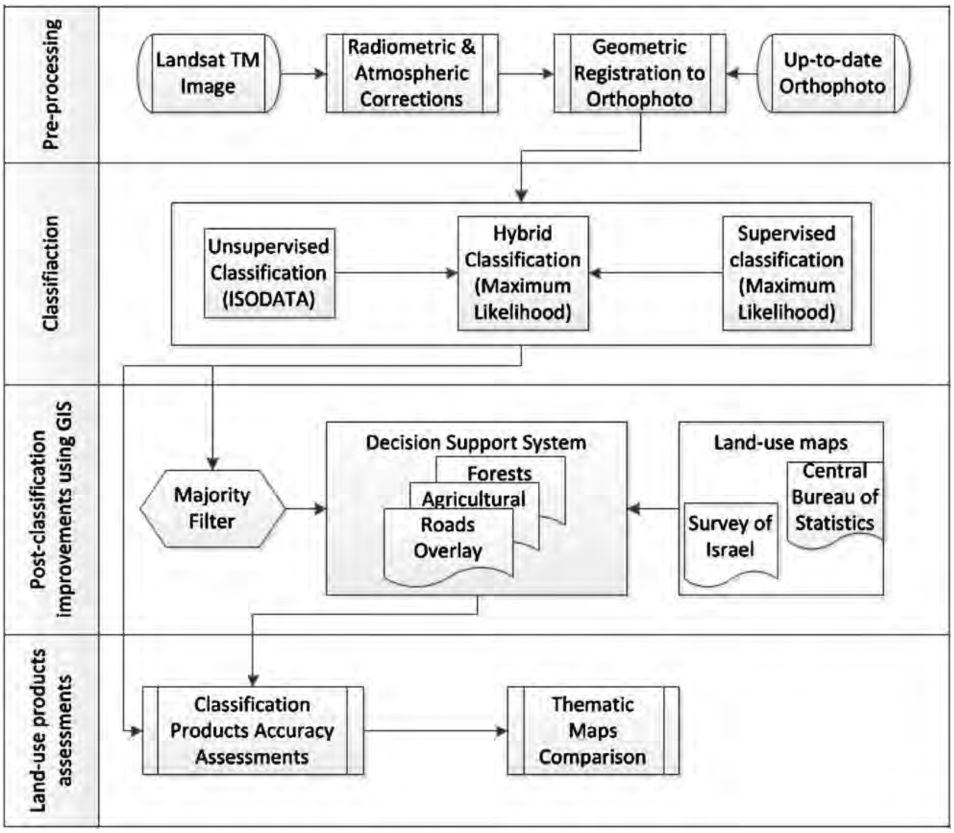

Fig. 2. Research flowchart.

1998; Kennedy, Cohen, & Schroeder, 2007). The successful geo-registration allows comparison to the national orthophoto and to additional maps.

Classification

For this study, Level 1 of the Anderson classification system was used (Anderson, Hardy, Roach, & Witmer, 1976). This classification system is designed to mainly rely on remote sensing; therefore only land-use and land-cover types identifiable by remote sensing are used as the basis for organizing this classification. Level 1 of the Anderson classification system is recommended for use with Landsat resolution data. Although this classification scheme is coarse, it eliminates misclassification errors and makes delineation of categories more substantial (Mallinis, Emmanoloudis, Giannakopoulos, Maris, & Koutsias, 2011; Zomeni, Tzanopoulos, & Pantis, 2008). The different land-uses and land-covers included in the six classes used by this study are detailed in Table 1.

ISODATA unsupervised classification

The pre-processed reflective bands image was classified into 80 classes using ISODATA classification technique. Following classification, each of the 80 classes was assigned into one of the six land-use classes by masking each class and projecting it on the up-to-date orthophoto of Israel for visual interpretation. Finally, the image was recoded according to the six land-use classes.

Maximum likelihood supervised classification

The image was classified using signatures from training sites that include all the land-cover types detailed in Table 1. A total of 120 signatures were collected from all land-use and land-cover classes. The signatures were collected by digitizing polygons on the up-to-date, high-resolution orthophoto of Israel, and then projecting them onto the image to collect the training samples. This allows for greater accuracy then simply digitizing from the Landsat TM image itself. When collection of training sites was done, the Euclidean distance between their spectral signatures served as a measure of separability for the signatures collected for each land-use class; spectrally similar signatures of the same class were united. The maximum likelihood classification (MLC) was run with a feature-space non-parametric decision rule. Classes of the resulting image were recoded into the six land-use classes.

Hybrid classification

An iterative classification approach was used, whereby spectral signatures for specific land-use and land-cover classes were created using unsupervised training followed by supervised training (Bakr, Weindorf, Bahnassy, Marei, & El-Badawi, 2010). After

Table 1
Land-use classification system for use with Landsat data (After Anderson et al., 1976).

Land-use class	Land-uses and land-covers included in class
1 Urban or built-up land	Structures of all types: residential, industrial, agricultural commercial and services. Transportation and utilities. Mixed urban or built-up land.
2 Agricultural fields	Cropland, orchards, vineyards, and nurseries.
3 Rangeland	Herbaceous, shrub and brush, and mixed rangeland.
4 Forest	Deciduous, evergreen, and mixed forests.
5 Water bodies	Reservoirs, coastal water.
6 Barren land	Bare exposed rock, quarries and disturbed ground at building sites, and dirt roads.

evaluating the classification product accuracies, signatures that contributed to the most accurate class assignments from both supervised and unsupervised training were appended together. MLC was applied again with the improved signature set.

Post-classification processing

While governmental land-use maps have their flaws, some relevant information could be extracted from them in order to improve the remote-sensing based classification. Created in 2004, the CBS map has been based on data from 2002. Although delineation of built-up terrain (housing and agricultural buildings) in the Bedouin diaspora of the Negev was added in 2007, this map is not updated for recent changes in land-use. While the CBS map is very detailed for urban and built-up land-uses and moderately detailed for agricultural land-uses, it does not account for areas used for pasture. These rangelands are categorized under “other open grounds” together with everything that does not fit into one of the other land-use classes of the CBS map. The Survey of Israel map is much more expensive and thus, only parts were available for use. Moreover, those parts were updated for 2004. Morphological cover features and orchards are mapped to a great detail but areas used for pasture are not defined. Also, some, but not all urban land-uses are defined. Most built-up areas are included in the “area without known characteristics” class together with other land-cover classes that do not fit in any class. Since both maps are relatively up-to-date and contain some useful information for land-use classification, they were combined to enhance the remote-sensing based classification efforts.

Land-use polygon layers were clipped according to the research area boundaries and converted to ERDAS raster format using ARCGIS. The data was resampled to 30 m resolution to match the Landsat TM data. Each of the land-use and land-cover classes of the original maps was recoded into the most fitting of the six land-use classes (Fig. 3). No classes were recoded into the “Water” and “Barren land” classes.

A Decision Support System (DSS) was designed (Fig. 4) based on a set of logical land-use trends, and the “convergence of evidence” approach (Sader et al., 1995) whereby a pixel’s value is updated only if an indicator exists in all data layers. It was decided not to update water and barren land pixels since the land-use maps do not account for them. To clear up some of the “salt & pepper” noise

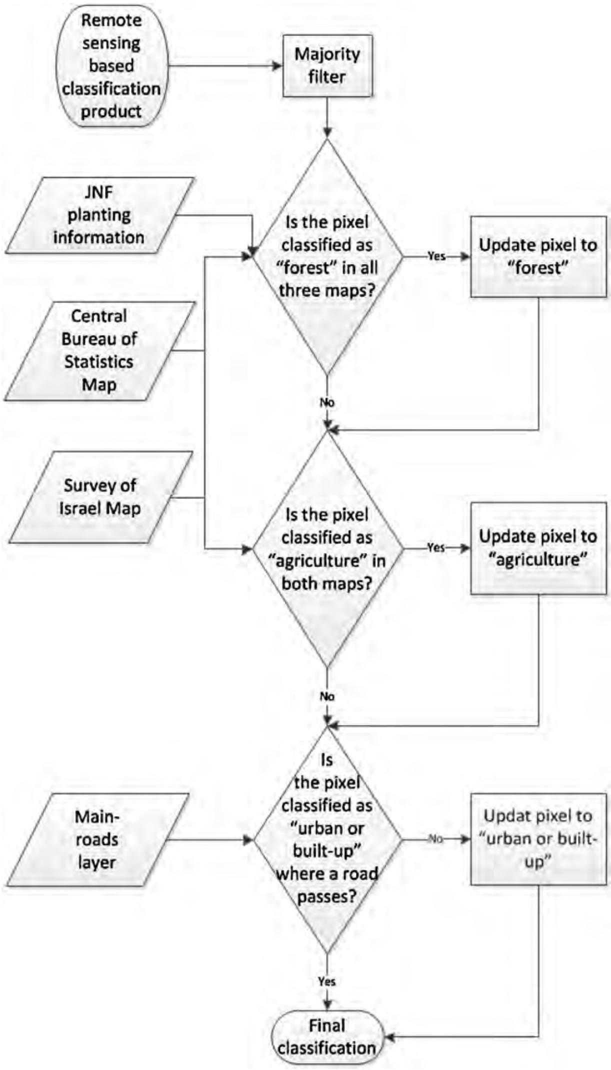

Fig. 4. Flowchart for the post-classification accuracy improvements by noise filtering and incorporation of ancillary land-use data using a Decision Support System (DSS).

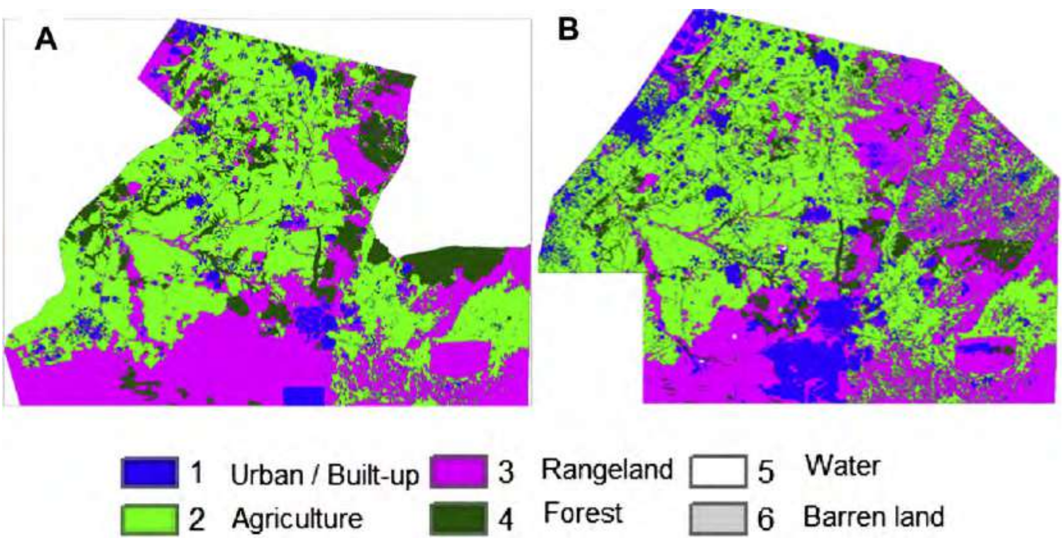

Fig. 3. Land-use maps assigned to Level 1 of the Anderson classification system (Anderson et al., 1976): (A) Central Bureau of Statistics (2004); (B) Survey of Israel (2004). The legend applies to all maps in this work.

apparent in the remote-sensing based classification products, a majority filter was employed for all of them. Following the majority filter, the three steps outlined below were executed in the order described for all the products:

- Since there is mainly afforestation and little or no deforestation in Israel, it was assumed that a pixel classified as forest in both land-use maps, should be classified as forest in the final classification product. However, since the forest areas in the maps represent JNF development areas, which are actually not all forests, additional information about forest planting was integrated into the decision making process. Thus, information layers containing older planted forests and trees planted on stream banks were used to confirm the afforestation in the JNF development areas.
- In a similar fashion, agricultural land could theoretically be transformed into built-up land, but rarely would it be abandoned to become a forest or rangeland. Most classification accuracy assessments showed only few pixels misclassified as built-up land instead of agricultural fields. Also, it was assumed that not a great deal of land has undergone the transition from agricultural to built-up in the five-year gap between the productions of the land-use maps and the acquisition of the satellite image. Due to all these assumptions, it was decided a pixel classified as agricultural land in both maps should be classified as so in the final classification product.
- As for built-up land, since the Survey of Israel map is not very accurate for that class it was decided not to update this class from the two land-use maps. However, after being paved, main roads are usually not converted to any other land-uses. Therefore, they were extracted from a GIS layer containing all the roads in Israel and overlaid on top of the final classification product.

Classification accuracy assessment

Throughout the classification process, the accuracy of classification maps was assessed by a set of 600 points sampled using the stratified random sampling; 100 points were randomly selected for each of the classes in the first generated classification map of this research (the ISODATA classification). These validation points were projected on to the up-to-date orthophoto of Israel, visually interpreted and each point was assigned to one of the land-use classes defined in Table 1. When using a coarse classification scheme such as the Anderson Level 1 classification, the analyst's interpretation based on a detailed orthophoto (1m spatial resolution) is not only as accurate as collecting ground truth data, but also faster and more efficient. The same validation set was further used for all the generated classification products to help ensure that the differences in accuracy could be attributed to the nature of class allocation and not the selected validation set (Foody, 2004). For each map, a confusion matrix was created and accuracy measures were calculated. The use of measures such as overall accuracy, Kappa statistics, producer's accuracy, user's accuracy and the conditional Kappa, are quite common and explained in detail in numerous publications (e.g. Campbell, 1996; Congalton, 1991; Foody, 2002; Lillesand & Kiefer, 2000; Rosenfield & Fitzpatrick-Lins, 1986).

Thematic maps accuracy comparison

When comparing classification methods, to decide which one is better, a researcher can simply choose the classifier yielding a better accuracy as measured by the overall accuracy, or the Kappa statistic (Cingolani, Renison, Zak, & Cabido, 2004). However, not every difference is significant and therefore, statistical significance tests are required. Comparison of Kappa coefficients using a *Z*-test is perhaps the method most advocated for thematic maps accuracy comparison (Congalton, 1991; Congalton, Oderwald, & Mead, 1983; Elmahboub, Scarpace, & Smith, 2009; Foody, 2002, 2004; Rosenfield & Fitzpatrick-Lins, 1986; Sader et al., 1995). Having used the same validation set for all the classification products, one cannot assume the samples used to derive each Kappa coefficient are independent and therefore the parametric test for comparing Kappa coefficients is inappropriate (Foody, 2004). If a *Z*-test was performed, it may result in overly large variance estimates and too conservative inference about the difference in accuracy between the two methods (De Leeuw et al., 2006).

An alternative approach for comparison of related samples has emerged in recent years; instead of comparing Kappa coefficients, the statistical significance of the difference between two proportions may be evaluated using McNemar's test (De Leeuw et al., 2006; Foody, 2004). This is a non-parametric test that is based on a binary distinction between correct and incorrect class allocations (Table 2). The McNemar test is based upon the standardized normal test statistic in equation (1).

$$z = \frac{b-c}{\sqrt{b+c}} \tag{1}$$

Since the square of *z* follows a Chi-squared distribution with one degree of freedom (Foody, 2004), the test equation could be expressed as equation (2):

$$\chi^2 = \frac{(b-c)^2}{b+c} \tag{2}$$

While χ^2 distribution is continuous, the distribution of sample frequencies in tests based on *z* is discrete (Dietterich, 1998). Therefore, a continuity correction is recommended. It is particularly important if the sample size used is small, but its impact diminishes for large sample sizes (Foody, 2004).

Equation (3) incorporates such a continuity correction:

$$\chi^2 = \frac{(|b-c|-1)^2}{b+c} \tag{3}$$

Results

Reflective band classification

The accuracies of both the supervised and unsupervised classification of the reflective bands were assessed. Confusion matrixes and accuracy measures can be found in Tables 3 and 4. Judging by the overall accuracy and overall Kappa statistics, it is apparent that the unsupervised classification is superior to the supervised classification (overall accuracy of 70.67% vs. 60.83%, respectively, Kappa statistic of 0.65 vs. 0.53, respectively). A McNemer's test confirmed that the unsupervised classification was significantly better in comparison to the supervised classification ($\chi^2 = 19.67$, $p < 0.0001$). However, when looking at specific class accuracy measures, such as the conditional Kappa, a different reality unfolds; for most classes, the supervised classification accuracies are better or similar to the

Table 2
Cross tabulation of number of correct and wrongly classified pixels for two alternative classifiers; the definition of matrix elements used in equations (1)–(3).

	Classification 1		
Classification 2		Correct	Incorrect
	Correct	a	b
	Incorrect	c	d

Table 3
Confusion matrix and accuracy measures for ISODATA unsupervised classification.

	Class	Reference data						Sum
		1	2	3	4	5	6	
Classified data	1	**61**	20	19	0	0	0	100
	2	7	**55**	35	3	0	0	100
	3	2	15	**83**	0	0	0	100
	4	1	25	36	**38**	0	0	100
	5	0	4	4	2	**90**	0	100
	6	0	0	3	0	0	**97**	100
	Sum	71	119	180	43	90	97	**600**

Producer's accuracy	User's accuracy	Conditional Kappa
85.92%	61.00%	0.56
46.22%	55.00%	0.44
46.11%	83.00%	0.76
88.37%	38.00%	0.33
100.00%	90.00%	0.88
100.00%	97.00%	0.96

The bold parts are used for the computation of the overall accuracy measure.
Overall Classification Accuracy = 70.67%.
Overall Kappa Statistics = 0.65.

unsupervised accuracies. When closely observing the first class (Urban or Built-up Land), it is apparent that it exhibits a very low user's accuracy and conditional Kappa for the supervised classification, and much higher values for the unsupervised classification.

Following this notion, the hybrid classification used the signatures that were generated by the ISODATA classification as the training set for the first class instead of the training signatures obtained in a supervised manner from digitizing areas of interest around settlements. The error matrix and accuracy measures for this hybrid classification are presented in Table 5. While the hybrid MLC results seem synergetic in a sense that both the overall accuracy and the overall Kappa statistics were improved, the McNemer's test confirmed that the accuracy improvement by the hybrid classification was statistically significant in comparison to the supervised classification ($\chi^2 = 42.15$, $p < 0.0001$), but not significant in comparison to the unsupervised classification ($\chi^2 = 1.43$, not significant (NS)).

GIS decision support system (DSS)

Each of the filtered remote-sensing based classification products has undergone improvement using a DSS. "Before" and "After" products are presented in Fig. 5. Confusion matrices and additional accuracy measures are presented in Tables 6–8. It was found that the DSS improved all the classification products by up to 10%. This improvement was found to be statistically significant ($\alpha = 0.01$) in every case (Table 9). The most accurate classification product was produced from reflective bands using the hybrid classification and the DSS improvement (81% overall accuracy, Kappa = 0.7681). When comparing its accuracy to those of the rest of the classification products, it was found to be significantly different from all of them (Table 10).

Fig. 6 summarizes the accuracies obtained for all the products, including the separate contribution of each stage of the DSS. Out of all the components, updating for agricultural land contributed most for the accuracy increase by the DSS.

Table 4
Confusion matrix and accuracy measures for supervised MLC.

	Class	Reference data						Sum
		1	2	3	4	5	6	
Classified data	1	**70**	81	81	11	13	3	259
	2	0	**27**	7	3	0	0	37
	3	1	5	**86**	13	0	2	107
	4	0	6	2	**13**	0	0	21
	5	0	0	1	3	**77**	0	81
	6	0	0	3	0	0	**92**	95
	Sum	71	119	180	43	90	97	**600**

Producer's accuracy	User's accuracy	Conditional Kappa
98.59%	27.03%	0.17
22.69%	72.97%	0.66
47.78%	80.37%	0.72
30.23%	61.90%	0.59
85.56%	95.06%	0.94
94.85%	96.84%	0.96

The bold parts are used for the computation of the overall accuracy measure.
Overall Classification Accuracy = 60.83%.
Overall Kappa Statistics = 0.53.

Table 5
Confusion matrix and accuracy measures for hybrid MLC.

	Class	Reference data						Sum
		1	2	3	4	5	6	
Classified data	1	**67**	49	28	4	5	0	153
	2	1	**35**	7	5	0	0	48
	3	3	23	**129**	3	1	0	159
	4	0	12	13	**29**	0	0	54
	5	0	0	0	2	**84**	0	86
	6	0	0	3	0	0	**97**	100
	Sum	71	119	180	43	90	97	**600**

Producer's accuracy	User's accuracy	Conditional Kappa
94.37%	43.79%	0.36
29.41%	72.92%	0.66
71.67%	81.13%	0.73
67.44%	53.70%	0.50
93.33%	97.67%	0.97
100.00%	97.00%	0.96

The bold parts are used for the computation of the overall accuracy measure.
Overall Classification Accuracy = 73.50%.
Overall Kappa Statistics = 0.68.

Discussion

A Landsat TM image was pre-processed and classified using three methods: supervised, unsupervised, and hybrid classification methods. Following this, the classification products' accuracy was assessed. A comparison of the products' accuracy was conducted to find out if the accuracy differences are statistically significant. It was found that unsupervised training produces more accurate results than supervised training. A hybrid supervised–unsupervised classification also produced more accurate classifications than the supervised classification; however, it did not improve the accuracy significantly in comparison to the unsupervised classification. All of the classification products were improved using GIS. First, a majority filter was used to smooth "salt & pepper" noise. Then, ancillary data from land-use maps and road maps was applied through a DSS to update the classification products. As a result, all products were improved significantly by up to 10%.

While others found supervised classification worked better than unsupervised classification (Alrababah & Alhamad, 2006), the opposite was found in this work, suggesting the training did not account for all the complex spectral variations of land-cover in the area. It is therefore concluded that one must be intimately familiarized with the research area to be able to train the maximum likelihood classifier properly. In a very heterogeneous area, collection of representative signatures is challenging. It was shown that statistically based clustering using ISODATA can produce superior results in such a situation.

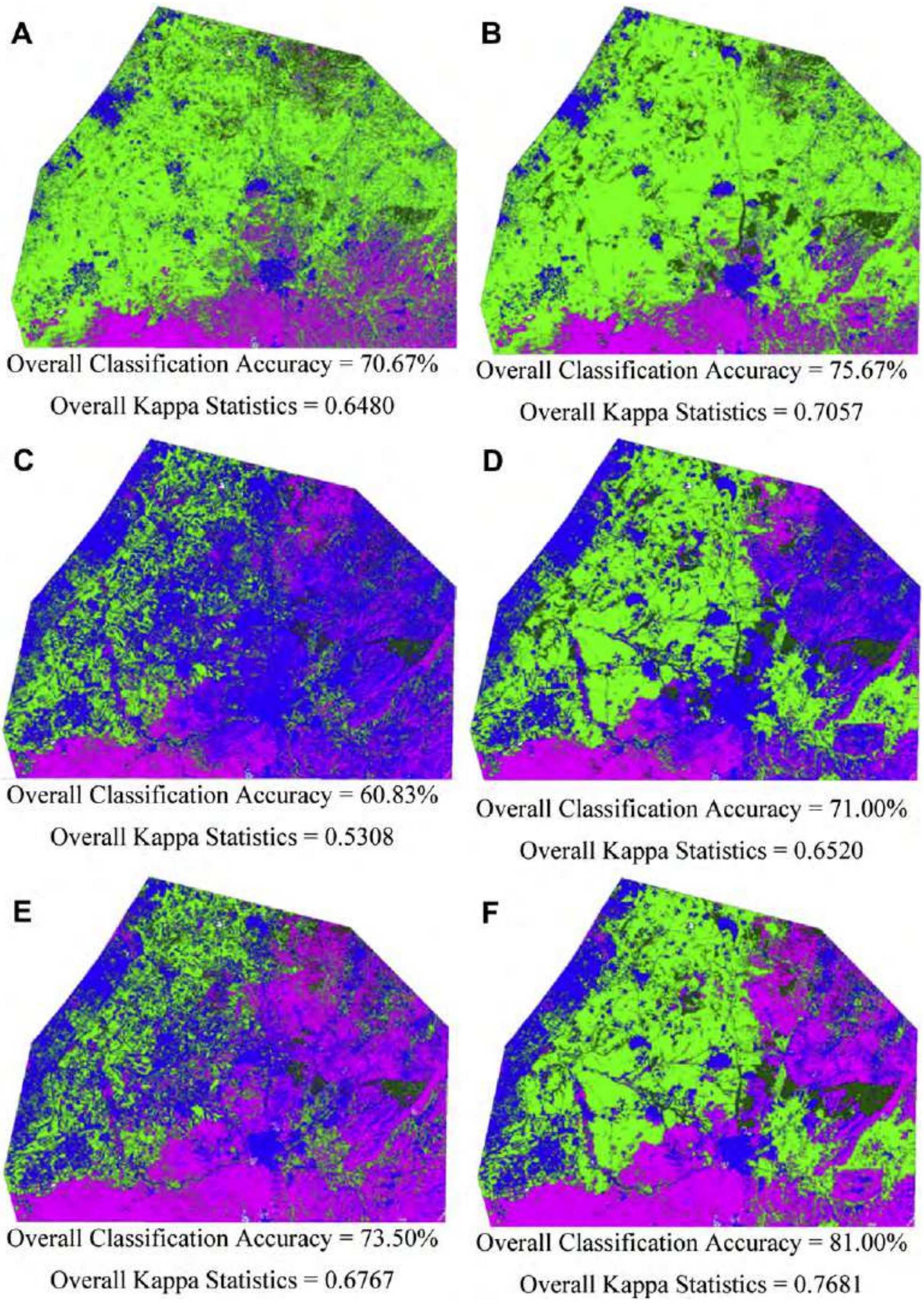

Fig. 5. Classification products: (A) ISODATA unsupervised classification; (B) ISODATA + DSS; (C) MLC supervised classification; (D) MLC + DSS; (E) Hybrid classification; (F) Hybrid classification + DSS.

The best overall accuracy achieved in this work is 81%, which is below the 85% level set as satisfactory for planning and management purposes (Anderson et al., 1976). However, in reality the accuracy of the classifications reported in many publications was also generally below the commonly recommended 85% target (Foody, 2002; Wilkinson, 2005) and its usefulness as a standard is unclear (Foody, 2008; Wulder, Franklin, White, Linke, & Magnussen, 2006). While some studies of land-use and land-cover achieved better accuracies, it is not always clear from the publication how the accuracy assessment was conducted (Koutsias & Karteris, 2003). Others conduct assessments limited by the amount of available data points (Cingolani et al., 2004; Elmahboub et al., 2009; Fuller, Groom, & Jones, 1994). Several land-cover mapping projects for the US using Landsat data present accuracies such as 65–82% (Homer, Huang, Yang, & Wylie, 2002), 70–98% with an overall average accuracy across all mapping zones of 83.9% (Homer, Huang, Yang, Wylie, & Coan, 2004; Homer et al., 2007), and 78.3–88.5% (Xian, Homer, & Fry, 2009). Therefore, while aspiring to uphold the accepted standards, the accuracy obtained in this study is quite satisfactory, as it resembles that of analogous studies.

It has been claimed that "the remote-sensing community appears to have a somewhat masochistic tendency in accuracy

Table 6
Confusion matrix and accuracy measures for ISODATA unsupervised classification with post-classification improvements.

	Class	Reference data						Sum
		1	2	3	4	5	6	
Classified data	1	**59**	8	15	0	0	1	83
	2	12	**100**	46	2	0	0	160
	3	0	4	**79**	0	0	9	92
	4	0	3	34	**39**	0	0	76
	5	0	4	4	2	**90**	0	100
	6	0	0	2	0	0	**87**	89
	Sum	71	119	180	43	90	97	**600**

Producer's accuracy	User's accuracy	Conditional Kappa
83.10%	71.08%	0.67
84.03%	62.50%	0.53
43.89%	85.87%	0.80
90.70%	51.32%	0.48
100.00%	90.00%	0.88
89.69%	97.75%	0.97

The bold parts are used for the computation of the overall accuracy measure.
Overall Classification Accuracy = 75.67%.
Overall Kappa Statistics = 0.71.

assessment, subjecting its thematic maps to an overly harsh and critical appraisal using pessimistically biased techniques yet accepting other maps with little question to their accuracy" (Foody, 2008). In this work, the use of ancillary maps for post-classification processing was done while the authors were aware of this critique. Assessing the ancillary map accuracies using the same set of validation data as for the maps derived from remote-sensing classified data is difficult, since they have different thematic classes. The interoperability problem associated with differences in map legends is often the greatest problem encountered in the comparison of thematic maps (Foody, 2007). Translating between legends is not necessarily straightforward, and can be a major source of error. It is also implicitly assumed that the maps are perfectly co-registered, an assumption which has not been verified. Since the accuracy of the ancillary maps is unknown, the authors hoped to avoid incorporation of errors from these maps into the products of this work through the combined use of several data sources. Still, some error might have been introduced to the final thematic maps by the use of these maps. Assessing the accuracy improvement for each stage of the DSS separately, verifies that this error is not greater than the contribution to the accuracy improvements. Such verification has not been reported by others who incorporated several sets of ancillary data in their post-classification accuracy improvements (e.g. Alrababah & Alhamad, 2006; Sader et al., 1995).

Table 7
Confusion matrix and accuracy measures for supervised MLC with post-classification improvements.

	Class	Reference data						Sum
		1	2	3	4	5	6	
Classified data	1	**67**	32	66	4	9	9	187
	2	4	**81**	20	2	0	0	107
	3	0	4	**79**	3	0	2	88
	4	0	2	11	**32**	0	0	45
	5	0	0	1	2	**81**	0	84
	6	0	0	3	0	0	**86**	89
	Sum	71	119	180	43	90	97	**600**

Producer's accuracy	User's accuracy	Conditional Kappa
94.37%	35.83%	0.27
68.07%	75.70%	0.70
43.89%	89.77%	0.85
74.42%	71.11%	0.69
90.00%	96.43%	0.96
88.66%	96.63%	0.96

The bold parts are used for the computation of the overall accuracy measure.
Overall Classification Accuracy = 71%.
Overall Kappa Statistics = 0.65.

Table 8
Confusion matrix and accuracy measures for hybrid classification of the reflective bands with post-classification improvements.

	Class	Reference data						Sum
		1	2	3	4	5	6	
Classified data	1	**66**	26	18	2	5	0	117
	2	5	**81**	16	3	0	0	105
	3	0	7	**125**	2	1	1	136
	4	0	5	18	**34**	0	0	57
	5	0	0	0	2	**84**	0	86
	6	0	0	3	0	0	**96**	99
	Sum	71	119	180	43	90	97	**600**

Producer's accuracy	User's accuracy	Conditional Kappa
92.96%	56.41%	0.51
68.07%	77.14%	0.71
69.44%	91.91%	0.88
79.07%	59.65%	0.57
93.33%	97.67%	0.97
98.97%	96.97%	0.96

The bold parts are used for the computation of the overall accuracy measure.
Overall Classification Accuracy = 81%.
Overall Kappa Statistics = 0.76.

It is advised for future studies to adopt this approach of examining the added value of each component in the process to the final accuracy. It is not trivial that any addition of data will result in improvement as this depends on the area, the quality of the initial classification and of the ancillary data. Each stage in the process should be evaluated by itself, and also within the context of the work flow; an addition of one data source might damage the accuracy, but in combination with another processing stage, or additional data sources it could prove to be of added value, and vice versa. Therefore, while it is generally recommended to incorporate additional spatial data sources with the remote-sensing data, there is no guarantee for improved accuracy as more spatial data is added and a trial and error process to eliminate redundant or damaging data is mandatory.

It was observed that out of all the components of the DSS, updating of agricultural areas contributed the most for improving the final products' accuracies. In spite of including training samples of crops in varied phonological stages in the MLC, there was still confusion with rangelands, urban, and forest classes. Since there are very few fields that are not cultivated for long periods of time, this problem could be solved by classifying multiple images of the same year, to capture vegetation peaks in all fields. This sort of technique was previously used for mapping of traditionally-managed rice fields (Turner & Congalton, 1998) and forest type classification (Wolter, Mladenoff, Host, & Crow, 1995). Unfortunately, for the current project, only one Landsat TM 5 image is available in 2009 over the study from USGS archives, not considering the Landsat ETM+ 7 with the scan-line corrector off. Therefore, multi-date classification approach was not attempted.

One of the main conclusions of this work is that the incorporation of the governmental land-use maps was very effective in improving the remote-sensing based classification. Alas, the

Table 9
Comparison of remote-sensing based classification products vs. those products after DSS improvements.

	χ^2	p-value	Overall classification accuracy improvement
ISODATA	11.68	0.0006	5%
MLC	35.64	<0.0001	10.17%
HYBRID	26.94	<0.0001	8%

Table 10
Comparison of all classification products to the hybrid classification with DSS improvement (the most accurate of all based on overall accuracy and Kappa statistic).

	χ^2	*p*-value
ISODATA	24.81	<0.0001
MLC	90.57	<0.0001
HYBRID	26.94	<0.0001
ISODATA + DSS	8.90	0.0029
MLC + DSS	40.48	<0.0001

overlap between the maps and the research area was not complete; the CBS map had only contributed information within the “Green line” boundaries (the 1949 armistice lines established between Israel and its neighbors), and the Survey of Israel map was missing a section in the south-western corner of the research area. In addition, military installations were masked in the governmental maps. Therefore, the DSS contributed to updating only parts of the classification map. It is the authors' belief that the accuracy improvement potential through the use of ancillary data has not been exhausted; had more complete maps coverage been available, the DSS improvements would yield better accuracy.

This is the first time (to our knowledge) that rangelands were defined in a land-use map of this area. This is a very heterogeneous class, which includes many land-cover types, consisting of different kinds of soil and diverse vegetation cover types. While the rangeland class consists of the natural, undeveloped environment of the northern Negev, in which grazing traditionally took place, agriculture land and forests both still support grazing for parts of the year. Forest managers allow grazing between February and May to prevent accumulation of burning material, remove vegetation and encourage water flow into sink areas called limans (Karnieli, Ben-Asher, Dodi, Issar, & Oron, 1988) for storage, and increasing the biological diversity through geophyte encouragement (Isaac Moshe, JNF southern region manager, personal communication).

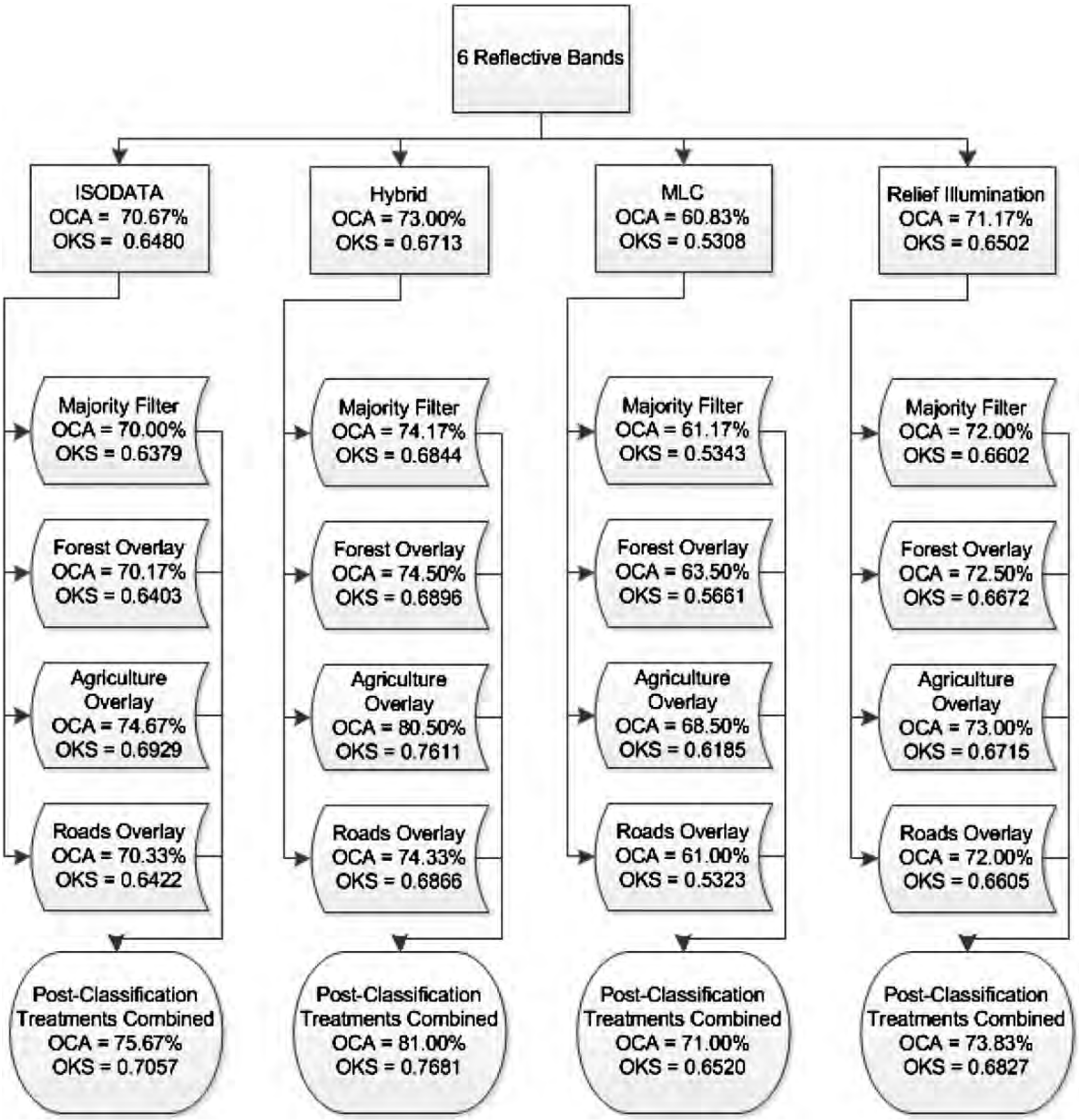

Fig. 6. Accuracies obtained for reflective bands products, including the separate contribution of each stage of the DSS by itself.

During the summer, Bedouin herds graze on stubble in cultivated crop fields, but not in orchards, both incorporated in the agricultural land class. This is an example of the limitations of mapping land-use using remote sensing; even extensive activities covering large tracts of land are not always amenable to interpretation from remote sensor data (Anderson et al., 1976). For this reason, the products of this work should be used carefully, together with ancillary knowledge in order to draw conclusions and conduct estimates of grazing territories.

Conclusions

- When lacking intimate familiarization with a large, complex, and heterogeneous area, unsupervised classification has a potential to produce more accurate results than supervised classification.
- Hybrid supervised–unsupervised classification produced more accurate classifications than the supervised classification; however, it did not improve the accuracy significantly in comparison to the unsupervised classification.
- Using a decision support system for updating the map based on expert knowledge and ancillary GIS data improved the classification accuracy significantly in all cases by 5–10%. There is great potential in this technique, but it depends on the availability of quality ancillary data. In the case of the northern Negev this potential is not fully realized yet due to partial coverage of the available land-use maps.

Acknowledgements

This Research program was sponsored by the Mid-East Program, International Programs, United States Forest Service.

The authors wish to thank the helpful comments and advise from both anonymous reviewers.

References

Alrababah, M. A., & Alhamad, M. N. (2006). Land use/cover classification of arid and semi-arid Mediterranean landscapes using Landsat ETM. *International Journal of Remote Sensing, 27*, 2703.

Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). *A land use and land cover classification system for use with remote sensor data*. U.S. Geological Survey Professional Paper, 964, 28.

Bakr, N., Weindorf, D. C., Bahnassy, M. H., Marei, S. M., & El-Badawi, M. M. (2010). Monitoring land cover changes in a newly reclaimed area of Egypt using multi-temporal Landsat data. *Applied Geography, 30*, 592–605.

Bolstad, P. V., Gessler, P., & Lillesand, T. M. (1990). Positional uncertainty in manually digitized map data. *International Journal of Geographical Information Systems, 4*, 399.

Brown, D. G., Pijanowski, B. C., & Duh, J. D. (2000). Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. *Journal of Environmental Management, 59*, 247–263.

Campbell, J. B. (1996). *Introduction to remote sensing*. New York: The Guilford Press. (pp. 622).

Chander, G., Markham, B. L., & Helder, D. L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment, 113*, 893–903.

Chavez, P. S. (1996). Image-based atmospheric corrections - revisited and improved. *Photogrammetric Engineering and Remote Sensing, 62*, 1025–1036.

Cingolani, A. M., Renison, D., Zak, M. R., & Cabido, M. R. (2004). Mapping vegetation in a heterogeneous mountain rangeland using Landsat data: an alternative method to define and classify land-cover units. *Remote Sensing of Environment, 92*, 84–97.

Cohen, W. B., & Goward, S. N. (2004). Landsat's role in ecological applications of remote sensing. *Bioscience, 54*, 535–545.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment, 37*, 35–46.

Congalton, R. G., Oderwald, R. G., & Mead, R. A. (1983). Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogrammetric Engineering and Remote Sensing, 49*, 1671–1678.

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E. (2004). Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing, 25*, 1565–1596.

De Leeuw, J., Jia, H., Yang, L., Liu, X., Schmidt, K., & Skidmore, A. K. (2006). Comparing accuracy assessments to infer superiority of image classification methods. *International Journal of Remote Sensing, 27*, 223–232.

Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation, 10*, 1895–1923.

Elmahboub, W., Scarpace, F., & Smith, B. (2009). A highly accurate classification of TM data through correction of atmospheric effects. *Remote Sensing, 1*, 278–299.

Elvidge, C., & Yuan, D. (1998). NALC land cover change detection Pilot study: Washington D.C. area experiments. *Remote Sensing of Environment, 66*, 166–178.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment, 80*, 185–201.

Foody, G. M. (2004). Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing, 70*, 627–634.

Foody, G. M. (2007). Map comparison in GIS. *Progress in Physical Geography, 31*, 439.

Foody, G. M. (2008). Harshness in image classification accuracy assessment. *International Journal of Remote Sensing, 29*, 3137–3158.

Fuller, R. M., Groom, G. B., & Jones, A. R. (1994). The land-cover map of Great Britain: an automated classification of Landsat thematic mapper data. *Photogrammetric Engineering and Remote Sensing, 60*, 553–562.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., et al. (2007). Completion of the 2001 national land cover database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing, 73*, 337–341.

Homer, C. G., Huang, C., Yang, L., & Wylie, B. (2002). *Development of a circa 2000 landcover database for the United States*.

Homer, C., Huang, C., Yang, L., Wylie, B., & Coan, M. (2004). Development of a 2001 national landcover database for the United States. *Photogrammetric Engineering and Remote Sensing, 70*, 829–840.

Huang, C., Goward, S. N., Schleeweis, K., Thomas, N., Masek, J. G., & Zhu, Z. (2009). Dynamics of national forests assessed using the Landsat record: case studies in eastern United States. *Remote Sensing of Environment, 113*, 1430–1442.

Hudak, A. T., & Brockett, B. H. (2004). Mapping fire scars in a southern African savannah using Landsat imagery. *International Journal of Remote Sensing, 25*, 3231.

Karl, J. W., & Maurer, B. A. (2010). Multivariate correlations between imagery and field measurements across scales: comparing pixel aggregation and image segmentation. *Landscape Ecology, 25*, 591–605.

Karnieli, A., Ben-Asher, J., Dodi, A., Issar, A., & Oron, G. (1988). An empirical approach for predicting runoff yield under desert conditions. *Agricultural Water Management, 14*, 243–252.

Kennedy, R. E., Cohen, W. B., & Schroeder, T. A. (2007). Trajectory-based change detection for automated characterization of forest disturbance dynamics. *Remote Sensing of Environment, 110*, 370–386.

Koutsias, N., & Karteris, M. (2003). Classification analyses of vegetation for delineating forest fire fuel complexes in a Mediterranean test site using satellite remote sensing and GIS. *International Journal of Remote Sensing, 24*, 3093.

Lillesand, T. M., & Kiefer, R. W. (2000). *Remote sensing and image interpretation*. New York: John Wiley & Sons. (pp. 724).

Mallinis, G., Emmanoloudis, D., Giannakopoulos, V., Maris, F., & Koutsias, N. (2011). Mapping and interpreting historical land cover/land use changes in a natura 2000 site using earth observational data: the case of Nestos delta, Greece. *Applied Geography, 31*, 312–320.

Manandhar, R., Odeh, I. O. A., & Ancev, T. (2009). Improving the accuracy of land use and land cover classification of Landsat data using post-classification Enhancement. *Remote Sensing, 1*, 330–344.

Mas, F. J. (1999). Monitoring land-cover changes: a comparison of change detection techniques. *International Journal of Remote Sensing, 20*, 139–152.

Orenstein, D. E., & Hamburg, S. P. (2009). To populate or preserve? Evolving political-demographic and environmental paradigms in Israeli land-use policy. *Land Use Policy, 26*, 984–1000.

Powell, S. L., Pflugmacher, D., Kirschbaum, A. A., Kim, Y., & Cohen, W. B. (2007). Moderate resolution remote sensing alternatives: a review of Landsat-like sensors and their applications. *Journal of Applied Remote Sensing, 1*. 012506.

Rosenfield, G. H., & Fitzpatrick-Lins, K. (1986). A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing, 52*, 223–227.

Sader, S. A., Ahl, D., & Liou, W. (1995). Accuracy of Landsat-TM and GIS rule-based methods for forest wetland classification in Maine. *Remote Sensing of Environment, 53*, 133–144.

Schulz, J. J., Cayuela, L., Echeverria, C., Salas, J., & Rey Benayas, J. M. (2010). Monitoring land cover change of the dryland forest landscape of central Chile (1975–2008). *Applied Geography, 30*, 436–447.

Song, C., Woodcock, C., Seto, K. C., Lenney, M. P., & Macomber, S. A. (2001). Classification and change detection using Landsat TM Data - when and how to correct atmospheric effects? *Remote Sensing of Environment, 75*, 230–244.

Stefanov, W. L., Ramsey, M. S., & Christensen, P. R. (2001). Monitoring urban land cover change: an expert system approach to land cover classification of semi-arid to arid urban centers. *Remote Sensing of Environment, 77*, 173–185.

Tateishi, R., & Shalaby, A. (2007). Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. *Applied Geography, 27*, 28–41.

Turner, M. D., & Congalton, R. G. (1998). Classification of multi-temporal SPOT-XS satellite data for mapping rice fields on a West African floodplain. *International Journal of Remote Sensing, 19*, 21.

Wilkinson, G. G. (2005). Results and implications of a study of fifteen years of satellite image classification experiments. *IEEE Transactions on Geoscience and Remote Sensing, 43*, 433–440.

Wolter, P. T., Mladenoff, D. J., Host, G. E., & Crow, T. R. (1995). Improved forest classification in the northern Lake states using multi-temporal Landsat imagery. *Photogrammetric Engineering and Remote Sensing, 61*, 1129–1144.

Wulder, M. A., Franklin, S. E., White, J. C., Linke, J., & Magnussen, S. (2006). An accuracy assessment framework for large-area land cover classification products derived from medium-resolution satellite data. *International Journal of Remote Sensing, 27*, 663–683.

Wulder, M. A., White, J. C., Goward, S. N., Masek, J. G., Irons, J. R., Herold, M., et al. (2008). Landsat continuity: issues and opportunities for land cover monitoring. *Remote Sensing of Environment, 112*, 955–969.

Xian, G., Homer, C., & Fry, J. (2009). Updating the 2001 National land cover database land cover classification to 2006 by using Landsat imagery change detection methods. *Remote Sensing of Environment, 113*, 1133–1147.

Zomeni, M., Tzanopoulos, J., & Pantis, J. D. (2008). Historical analysis of landscape change using remote sensing techniques: an explanatory tool for agricultural transformation in Greek rural areas. *Landscape and Urban Planning, 86*, 38–46.