CLASSIFICATION OF MULTISPECTRAL ASTER IMAGERY IN ARCHAEOLOGICAL SETTLEMENT SURVEY IN THE NEAR EAST

Bjoern H. Menze[1] and Jason A. Ur[2]

[1]Interdisciplinary Center for Scientific Computing (IWR), University of Heidelberg, Germany
[2]Department of Anthropology, Harvard University, USA
`bjoern.menze@iwr.uni-heidelberg.de, jasonur@fas.harvard.edu`

ABSTRACT:

In alluvial areas of the Near East, the former locations of settlements are often represented by "tells", small artificial mounds resulting from millennia of human settlement activity, especially the continual construction and decay of mud brick architecture. To identify such tells and other, smaller settlement sites in the modern landscape, we develop a classifier which screens wide areas for tell-specific soil-changes, based a characteristic spectral signature in ASTER imagery. Using data from sites identified from CORONA imagery and field survey on a north Syrian plain, a Random Forest classifier was trained, using the raw reflectances, vegetational features, correlation with prototype-spectra of the JPL ASTER SpecLib, and time flags as input features. A spatio-temporal sampling strategy allowed us to classify and fuse results from any ASTER images available for a certain region. The classifier was tested in an independent test area, centered around Tell Hamoukar, with close ground control from an archaeological survey. In this test area it was possible to identify 32 out of the 49 site bigger than $2ha$. Overall we found that multi-spectral ASTER imagery can be used to provide highly specific information on character and composition of the ground, a tool which can be used in survey planning or the screening of wide regions for conservational issues or studies in landscape archaeology.

1 INTRODUCTION

Archaeologists recognize the scale and spatial distribution of settlement as critical variables in the study of the origins of urbanism and social complexity, making their location and measurement an important component of archaeological research. Tells, the arabic name for the settlement mounds of the Near East, represent places which were occupied over millennia, often since the beginning of farming in the early Neolithic. Due to a predominantly mud brick-based building technique many of them grew to considerable heights (Rosen, 1986) and some even achieved urban status (Wilkinson, 1994) during the Bronze Age. Today, thousands of these settlement sites still can be found in Near Eastern landscapes. Some of them are mounds of sizes which even allow to spot them in global elevation models (Sherratt, 2004, Menze et al., 2006), other sites are only visible from characteristic changes of the soil, changed by the debris of millennia of settlement activity (Wilkinson et al., 2006). In the present work, the use of ASTER imagery for providing means of prospecting for these remains of the earliest human settlement system will be evaluated (Altaweel, 2005). So far, most archaeological applications use (high resolution) satellite images as an replacement for standard aerial photography. In these applications, the spectral information only provides qualitative information, e.g. enabling the interpretation of a scene in a false-colour coding (Lasaponara and Masini, 2007, Masini and Lasaponara, 2006, for example), but is not used explicitly to identify spectral signatures which are of archaeological interest. In the following, we will illustrate how the spectral information of multi-spectral satellite imagery can be used for such an approach. While this machine-based search for a specific ground cover is a well established tool in agricultural or geological remote sensing, such an automated classification of spectral imagery has not been pursued in archaeological remote sensing, so far.

Operating since the year 2000, the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), provides spectral imagery for any region worldwide and at many points in time. Few applications, however, use the full information of whole multi-temporal data sets. Often, the classification of surface features is limited to the classification of one single image and to data which were acquired at the same date and under similar environmental conditions (Apan et al., 2002, for example). Among the approaches which use images and spectral signals of different points in time, most search for a characteristic change pattern along the temporal dimension, e.g. in a change detection between different scenes (Bruzzone et al., 2004, Im et al., 2007), or in the classification of vegetational features according to their trajectory along the seasonal cycle (Chattopadhyay and Dutta, 2006, Hayes and Cohen, 2007, Xiao et al., 2006). Characteristic patterns found along the temporal dimension are used to identify the different types of ground cover and then serve as input features in a subsequent classification. Unfortunately, the spectral pattern of soils does not show such a characteristic seasonal dependence. Variations of the spectral pattern are dominated by short time changes due to precipitation and other meteorological and environmental factors. Thus, approaches using the temporal dimension as in (Chattopadhyay and Dutta, 2006, Hayes and Cohen, 2007, Xiao et al., 2006) are not easily applicable to identify tell-specific soil changes. Instead of searching for characteristic trajectories along the temporal axis in an ordered set of images, we will propose a classifier in the following which can be applied individually to any single image, irrespectively of its time of acquisition. Together with an appropriate fusion strategy, this will allow all information of all ASTER scenes available for a certain region to be used. To this end, we will employ a spatio-temporal sampling strategy (sections 2 & 3), allowing to obtain a classifier which can be trained on the maximal number of different environmental situations in different ASTER scenes. We will evaluate the use of different fusion strategies available from sensor fusion when applying this classifier to a multi-temporal data set (Benediktsson and Kanellopoulos, 1999, Jeon and Landgrebe, 1999, Bruzzone et al., 1999, and references therein) and finally test the optimized classifier in a region with close archaeological ground control (sections 4 & 5).

2 MULTI-TEMPORAL CLASSIFICATION STRATEGY

Designing a classifier for a specific problem on a certain type of data typically follows a very standardized procedure. In a classification task on a single spectral image, for example, it is the collection of a training data set in a first step, to be followed by the definition of a feature representation and the classification algorithm. Parameters of feature selection and classification are then evaluated by an error measure which is appropriate to the learning task and are finally optimized accordingly.

To obtain a classifier in the given application which is robust against the variations of the data along the time-line, the normal sampling of the training data in the spatial dimensions is extended by an additional sampling in the temporal direction, in the combined space of coordinate-space and time. As the resulting classifier is supposed be applied to any available ASTER image of a certain region, the standard classifier design is also extended by a subsequent fusion step, to pool the results map of all individual images. In our search for a tell-like surface or soil pattern we used the following scheme, under particular consideration of the multi-temporal extension of the classification:

1. Sampling. The detection task is transformed to a binary classification. In addition to the verified tell sites, a number of "non-tell" locations are chosen randomly. For all locations of both groups, spectra are sampled from all imagery available at these locations. Relying on a non-parametric classification model in the following, the sampling is of crucial importance in the design of the classifier.

2. Features. In addition to the "raw" spectral reflectances, secondary features transport prior knowledge on: a) Invariants – Vegetation indices represent established normalization strategies on the reflectance of specific channels. b) Subclasses – Prototypes of expected spectral patterns, i.e. rock, soil-types, water, can be compared against the observed pattern, e.g. in a correlation with the signal. c) Temporal features – Indicate the date of acquisition.

3. Classifier. Due to the lack of information about the presence of subgroups (such as high-, low-mounded tells, tells under modern settlement in the tell class, but also like crop fields, bare soil, rocks, modern settlements in the non-tell class) and due to the inhomogeneity of the features (categorical, nominal), the non-linear and non-parametric "Random Forest" classifier (Breiman, 2001) has been chosen. It is a tree based ensemble learner with few, easily adjustable hyper-parameters.

4. Optimization. A cross-validation both over the spatial box-grid (Lahiri, 2003) and the temporal dimension, i.e. the single image covering the training area, is used to evaluate and adjust the parameters of steps 1.-3., i.e. the sampling strategy, the optimal choice of features, and the hyper-parameters of the classifier.

5. Fusion. As the classification procedure (steps 1.-4.) can be applied to any available image of the region under study, it allows for a fifth step: Pooling the result maps of all available images.

The resulting classifier differs from hierarchical approaches in multi-sensor fusion (Briem et al., 2002, Jeon and Landgrebe, 1999, Zhu and Tateishi, 2006) as it avoids the training of a new classifier for each sensor, or, in this application, for each image. It allows for a robust, independent classification of single images, regardless of their time of acquisition, without the need to manually select the data for specific time-points or specific environmental conditions in training or in application.
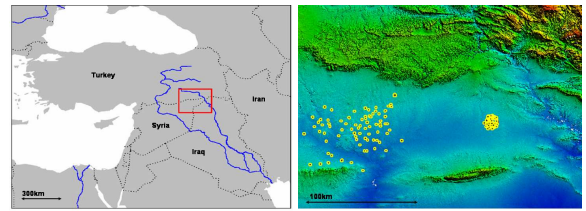
3 METHODS

Figure 1: The Khabur plain, region under study: location, with the Upper Khabur basin indicated (left) and digital elevation model (right). Indicated yellow are training sites (west, compare Fig. 2) and test area centered around Tell Hamoukar (east, Fig. 6).

3.1 Satellite data and Region under study

The study is based on ground truth on a part of a north Mesopotamian plain (the Upper Khabur basin), in the province of Hassake, Syria, close to the Turkish and the Iraqi borders (Fig. 1, *left*). The training sites (128 sites) are situated in an area of approx. $90km * 60km$ in the western part of that plain (Fig. 1, *right*). These settlement sites were identified by analyzing declassified CORONA imagery in conjunction with several seasons of survey fieldwork (Wilkinson, 1997, Ur, 2004). The training sites include low- and high-mounded tells, as well as sites which are (partially) situated beneath modern settlements. The training area is partially covered by 16 cloud free ASTER swaths (Fig. 2), acquired during different times of the year between 2003-2006, most of them during the dry-season from May-October (Fig. 3). A second region approx. $100km$ east to this area was subject to an intensive ground survey (Ur, 2002b, Ur, 2002a) and used for the evaluation of the ASTER classification (Fig. 1, *right*). It is an circular area of approx. $125km^2$, and with a reported number of 75 archaeologically relevant sites, ranging from low mounded areas with moderate densities of pot sherds or minor soil changes to major mounds like Tell Hamoukar in its center. This test area is covered by 9 cloud-free ASTER swaths from 2003-2007, most of them acquired during the dry-season as well. For both training and test data, the 6 SWIR and 5 TIR channels were interpolated to the maximal resolution of the 3 VNIR channels ($15m * 15m$) using a nearest-neighbor approach.
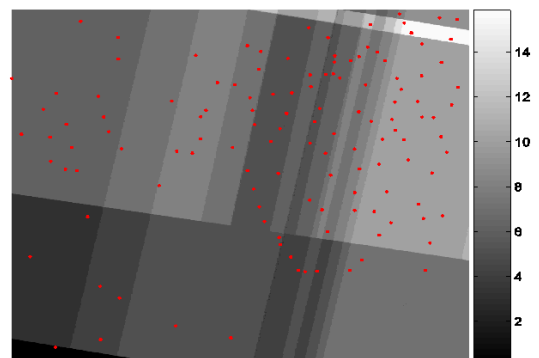

Figure 2: Number of ASTER images for a tell site (indicated red) in the training area. Tell sites are covered by 2 to 15 ASTER images (see scale bar).

3.2 Implementation of the classifier

The classifier was trained and optimized on the first training data set and then applied to the second test set. It was implemented and optimized on the training data according to the multi-temporal classification strategy (section 2):
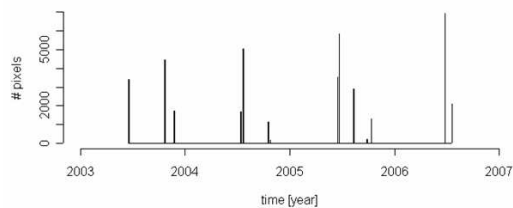
Figure 3: Pixels of tell sites in the training set, along time.

1. Sampling. The detection task was transformed to a binary classification, discriminating between the class of spectra from archaeologically relevant sites and a non-tell or "background" class. All spectra available from tell sites (ts) were included in the training data ($N_{site}^{ts} = 11215$ pixels, 0.24% of the 4 702 671 pixels in the test region; $N_{spec}^{ts} = 40306$). The locations of the background class (bg) were randomly sampled from the remaining regions ($N_{site}^{bg} = 9961$, 0.21%) and the spectra available from these pixels were subsampled to approximately match N_{spec}^{ts}, and included into the training set ($N_{spec}^{bg} = 50000$).

2. Features. In total 25 features were included into the training set. Beside the raw reflectance of the 14 spectral channels, two time flags were included (year, day of the year). To obtain invariance against seasonal influences, three vegetation indices were included, representing differently normalized spectral channels (NIR / R; NIR - R "VIDIFF"; (NIR-R) / (NIR+R) "NDVI"). For six classes of the JPL ASTER SpecLib[1] (manmade, minerals, rocks, soil, vegetation, water), averaged spectra were subsampled from full resolution to the 14 ASTER channels and used for correlation with the raw spectra. The resulting correlation coefficients were also used as input features.

3. Classifier. To classify the rather unstructured set of spectral features, Random Forest (Breiman, 2001) was chosen, a nonparametric and non-linear ensemble learner based on decision trees, a class of classifiers allowing for scale invariant learning. Prior comparisons of Random Forest, a RBF-kernel support vector machine (SVM) and a linear SVM on the raw spectral features of a subset of the training data (approx. 7000 tell class and 8000 background class samples) had indicated the need for a nonlinear classifier (linear SVM 21.4% classification error in a tenfold cross validation), and a gave slight advantage of a Random Forest (18.2%, using default parameters) against the non-linear SVM (20.4%, optimizing the kernel-width in an internal five-fold cross-validation). In the following, Random Forest was always applied with the default parameters of (Liaw and Wiener, 2002) (*ntree* = 300 trees, *mtry* = 4 variables).

4. Optimization. During training, feature subsets were evaluated in a random, unblocked cross-validation ("full Xval", table 1), in a blocked cross-validation over a spatial grid ("spatial Xval") and over both a spatial grid and a flag of the ASTER image, i.e. a temporal label ("spatial & temp Xval"). The spatial-box grid had a side-length of $6km$, an extension beyond the length of most structures on the ground. For comparison, classifiers solely trained on single images were also applied to the remaining data (training on one image, classifying all others, averaging prediction accuracies at the end). Using the posterior probability of the Random Forests in prediction, the performance was measured by the average image-wise area-under-the-curve (AUC) of the receiver-operator characteristic (ROC), in order to compare binary ground truth with the predicted probability of the presence of a tell-like soil change. Thresholding the posterior probability at 0.5, the classification accuracy was evaluated as well.

5. Fusion. Using the optimized classifier, all cross-validated results were pooled using different fusion strategies: Averaging

---

[1]ASTER Spectral Library. `http://speclib.jpl.nasa.gov/`

thresholded probability maps using thresholds of 0.5, 0.7, and 0.9, averaging the posterior probability, and a Naive-Bayes approach on the probabilities was tested. Finally, classification accuracy and AUC of the ROC for the fused results were compared against the labels of the originally sampled training locations to find the best fusion strategy.

4 RESULTS

4.1 Training – *"What do we learn from the learning algorithm?"*

Evaluating the classification performance on the training set primarily aims at a better understanding of the learning problem and at an optimization of the design of the classification strategy. Three different cross-validation approaches measured or "simulated" different scenarios (table 1): The random cross-validation approximates an upper limit to the bayes error of the training data and is the most optimistic. The spatial cross-validation is an approximation to the optimal spatial generalization error, e.g. to results obtained when the same ASTER image is used in training and testing (or just imagery acquired on the same date). The spatio-temporal cross-validation is the most pessimistic of the three scenarios. It tests the performance on data which is both distant in space and in time, for example, on an ASTER image with a different date and from a different region than the training data. Assuming that training and test region have certain similarities, and that it might be possible to select images for their acquisition time –which will be given in most tests or applications– the real classification error will be somewhere in between the classification results of a spatial and a spatio-temporal resampling scheme.

Accuracy	all	w/o Spec	w/o Envir	w/o SpecLib	w/o Time
full Xval	82.	79.3	(82.0)	(83.2)	(81.8)
spatial Xval	78.1	75.6	77.1	78.2	77.2
spatial&temp Xval	69.9	(70.2)	67.7	69.5	70.4
single classifier	61.2	61.4	61.3	61.4	–
AUC ROC	all	w/o Spec	w/o Envir	w/o SpecLib	w/o Time
full Xval	90.5	87.0	(89.8)	(91.0)	(89.7)
spatial Xval	85.6	83.0	84.7	85.9	84.9
spatial&temp Xval	76.2	(76.4)	73.9	75.9	76.4
single classifier	66.7	66.6	66.3	67.4	–

Table 1: Accuracies and area under the curve (AUC) of the ROC on the differently cross-validated training data; on the full set of features (*all*) and with single feature groups removed (*w/o*). Nonsignificant differences to the results on the full set of features are in brackets.

We find that approx. 70-80% of the observation (table 1) are classified correctly when using the spatio-temporal sampling, as opposed to a classification accuracy of approx. 60% from classifiers trained on spectra sampled in coordinate-space only. Predicting full result maps of the training regions in a spatial cross-validation, the majority of the pixels in 124 out of the 128 tell sites were classified correctly. False negative pixels within the area of the 124 sites typically occurred at the borders of the tell sites. False positives pixels in the background region were primarily due to modern settlements. Studying false positive sites in high resolution imagery also revealed the presence of settlement mounds within the training region which were unrecorded in our training data (Fig. 7).

The contributions of the different feature classes (raw spectral channels, vegetation indices, correlation with prototype-spectra,

time-flags) were also tested in the different cross-validation scenarios. Each of the classes was removed from the feature set and the classification was repeated (table 1). These classification results were then compared against the results on the full set of features and tested for significant differences using a paired Cox-Wilcoxon test on the results of the single spatio-spectral blocks. We observe that raw reflectance spectra are relevant features in the random and the spatially blocked cross-validation, while secondary features like vegetation indices and prototype correlation are relevant in the test of spatio-temporal generalization. Time flags show an indifferent behavior. They improve the classification performance in random and a spatially blocked cross-validation, but decrease the spatio-temporal accuracy.
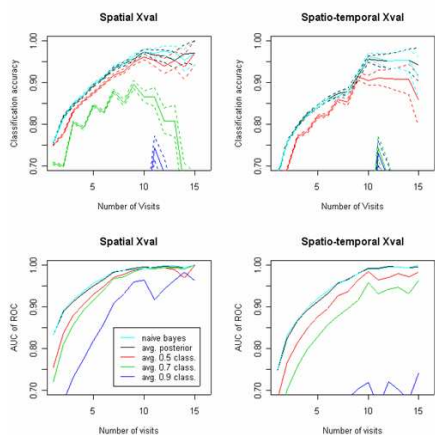

Figure 4: Fusion of single results at a location in a spatial and a spatio-temporal cross-validation. Details see text.

Optimal fusion strategies were evaluated in a second step. For each location the cross-validated observations available from the training data set were fused to a single prediction (Fig. 4), which was then compared to the label of the location. If more observations than necessary were available for a single location, e.g. when assessing the performance of three combined observations while five were available for the location, observations were subsampled accordingly. We observe that the classification accuracy increases from approx. 70% to $>> 90\%$ with a nearly complete class separation both for the spatial and the spatio-temporal cross-validation (AUC ROC $>> .95$). The direct average of the class posterior probabilities performs best, together with the application of the product rule on the probabilities (Naive-Bayes). Averaging the classification error ranks third, while averages of classification maps thresholded at other probability than 0.5 – potentially allowing to trade sensitivity and specificity– perform the worst.

4.2 Testing – *"What do we learn from the learning problem?"*

While the evaluation of classification performances on the training data set primarily serves technical purposes, the application of the classifier to the test region allows us to assess the relevance of the classification algorithm in the application it has been designed for. Thus, all nine ASTER image covering the test area were classified using the the Random Forest classifier on the full set of features. In the fusion of the result maps, the two top performing approaches (average posterior, Naive-Bayes) were tested, but for reasons of comparison the classification performance at a 0.5 and a 0.7 threshold was determined as well. Results were compared qualitatively in a pixel-wise manner –using the precision-recall-curve– and in a quantitative approach determining the number of detected sites.
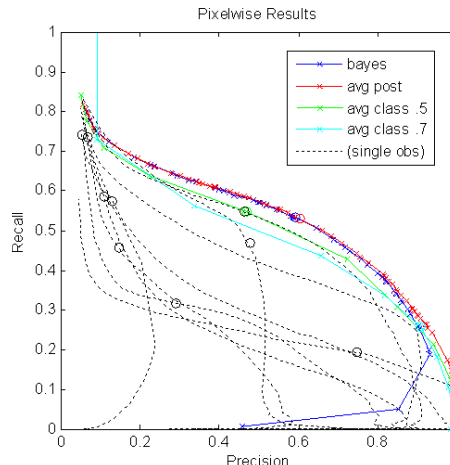

Figure 5: Pixel-wise precision and recall on the test data for the different fusion strategies and for single images. Points indicate classification results with a fixed threshold (e.g. 0.5 probability).

As the number of "background" pixels outnumber the pixels of tell sites in the given detection problem significantly, precision (true positive / (true positives + false positives)) and recall (true positives / (true positives + false negatives)) were used to concentrate the evaluation on the smaller, but more important class of tell-pixels (positives). We find that average posterior and Naive-Bayes perform best (Fig. 4), with nearly identical classification results when using the optimal threshold from the training (recall 0.54, precision 0.59). As the Naive-Bayes approach assigns high probabilities to pixels which do not belong to the tell class, its AUC is somewhat smaller than the one from the average posterior. The average classification strategy (threshold 0.5) shows acceptable results, still better than the best classification of a single image. The average 0.7 classification –representing a robust error measure on the posterior– has the highest precision among all approaches (0.91), but at a very low recall (0.26). Overall we observe again that the averaging improves the results significantly and that average posterior and Naive-Bayes perform best, although the latter has a somewhat smaller area under the convex curve of precision and recall.

To determine a "site detection rate", the true positive site "hit" and the false positive site "miss" has to be defined. Ground truth allowed to assign each *pixel* of the survey area either or the tell or to the background class (Fig. 6 c). As slight shifts occurred frequently between the ground-truthed tell sites and regions with a tell-like spectral signature (either due to registration or projection artifacts, or to real differences between these borders) areas around tells were extended by a distance of $90m$ (approx. TIR pixel-size, see dark blue region in Fig. 6 d). "Positive" pixels in these regions were considered to be "true positive". Remaining areas of the survey region (bright blue in Fig.6 d) were considered to belong to the background class. Here, "positive" pixels were "false positives". A *tell site* of the ground truth data set was "detected", if more than 15 adjacent "true positive" pixels (four-neighborhood) could be found in its vicinity and the center of the connected pixels was within the confirmed tell region. Areas of 15 or more "false positive" pixels in the background region were considered to be a false positive site. Fifteen pixel correspond to a site of approx. $0.3ha$, the smallest size of a tell site in the training data.

Following this definition, we were able to identify a major subgroup of 34 sites among the 75 confirmed sites of the test set, at a number of approx. 20 false positives. Among the false positive
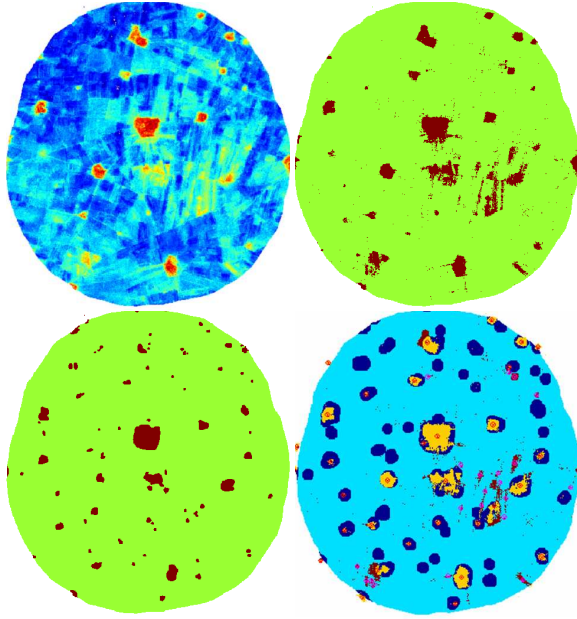
Figure 6: The test region. From top to bottom: a) average posterior (*left*), b) binary classification (*right*), c) ground truth (*left*), d) classification results (*right*). Tell Hamoukar is slightly north to the center.

"hits", most occurred in regions of known tells and coincide with the outlines of selected modern agricultural fields. Most false negative were minor sites; among the 49 sites with a size of more than $2ha$ (approx. two TIR pixels), 32 were detected successfully.

5 DISCUSSION

In most applications of archaeological remote sensing, satellite images are used primarily for visual interpretation and as a replacement for aerial imagery (Lasaponara and Masini, 2007, Masini and Lasaponara, 2006). Instead of using spectral information for a false-colouring of a scene and in a qualitative interpretation only, we explicitly used characteristic changes in the reflectance of the soil (Wilkinson et al., 2006) in a quantitative approach. Exploiting the spectral information of the satellite data directly, we were able to identify a spectral signature of archaeologically relevant soil marks and places of former settlement activity (Fig. 7). The ability to determine the locations of archaeological sites greatly enhances the efficiency of ground survey, the process by which the chronological range of occupation is determined. Ultimately, we will use these chronological patterns to build up a history of settlement in the region over some eight millennia which will allow us to investigate the spatial dimensions of the evolution of early urbanism and land use in one of the regions of the earliest cities in the Near East.

The spatio-temporal sampling strategy allowed us to obtain a classifier which is robust against temporal variations of the spectral signature of the ground. It can be applied to any ASTER image available for a region of interest without being restricted to certain periods of the year, and thus provides means to analyze and fuse all information which is available for that region. Fused result maps outperform any single observation both on training and on test data, reducing "noise" in the spectral signal due to environmental influences. Averaging the posterior probability performs better than a Naive-Bayes approach or the results from the ensemble-vote of a classification, indicating that the result maps of the ASTER images are indeed "noisy" realizations of an iid
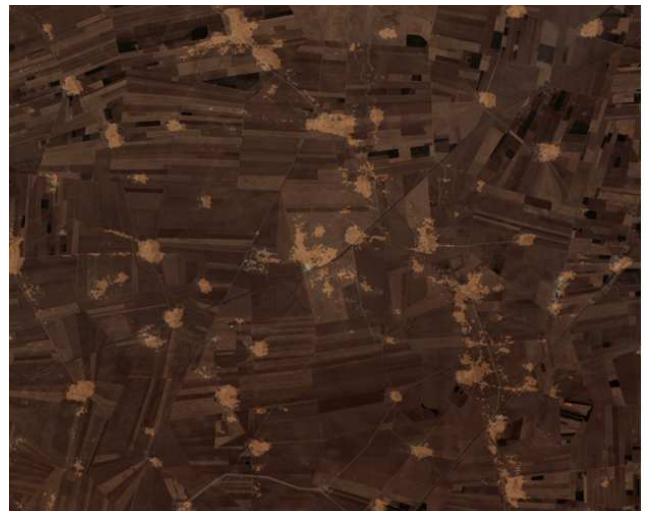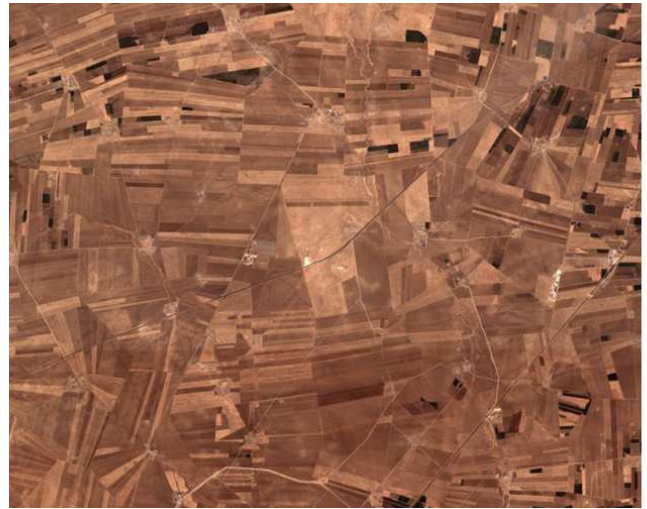

Figure 7: Survey of wide regions. Top: High resolution imagery of the Khabur plain. Bottom: Binary classification map provides adjunct information. Tell-like regions are indicated by brighter color. (Cross-validated and fused results for the western part of the training region.)

(independent and identically distributed) process, rather than independent sources of information (independent distributed), often assumed in sensor-fusion (Benediktsson and Kanellopoulos, 1999, Jeon and Landgrebe, 1999, for example). From a practical perspective, this relieves the operator from the deliberate choice of selecting the "best" available ASTER image, when relying on the standard approach of only using one single image for the classification. However, restricting the fusion process to a selection of optimal result maps might enhance the performance of the fusion step even further (either determined in a visual inspection of the results or in an automated procedure (Petrakos et al., 2001)). The fusion of all information itself has proven to be of major relevance in the detection of a characteristic soil signature of former settlement sites.

Comparing classification results in the Tell Hamoukar region with ground truth from field survey (Ur, 2002a, Ur, 2002b), we observe a high correspondence for major sites. Minor sites, a group not represented overly well in the training data, could not be identified with the same reliability. Obviously, the spatial resolution of the ASTER sensor sets lower limits to the identification of small sites from satellite imagery, but it might be argued as well that small sites do not show the same soil composition, erosion be-

havior, or response to precipitation as major ones do. Thus, using a classifier which is explicitly trained to identify subgroups like minor, low mounded sites, but also like high mounded tells or mounds in vicinity or beneath modern settlement, might alleviate this problem. In addition to reduce false negatives, such a classifier might also allow to gain more insights in the character of false positive sites. It remains open, for example, whether in regions of recent agricultural earthworks, the leveling of former sites might have been the major source of these classification errors. False positives in the area of modern settlements might result from covariates in the (current) training data, or might resemble the spectral characteristic of ancient debris indeed, eventually indicating the presence of (unrecorded) former sites at the same place. Studying false positives should be a major objective in a more "prospective" ground study.

So far the classifier generalized well over the distance of around $100km$, from our training area in the southwestern region of the Upper Khabur basin to our testing area in the eastern end (Fig. 1). Regions with with different soil or precipitation pattern, however, might require a "local" retraining or recalibration of the classifier. Here, a combined approach with other modalities, e.g. with results from the SRTM model (Menze et al., 2006), might be preferable. In addition to the "extrapolation" of knowledge from one region to another, as evaluated in this study, an "interpolation" of information in regions already surveyed (Fig. 7) will always be possible, providing a tool to classify the spectral signature of site already known or even to identify sites which remained undetected on the ground.

Soil changes provide relevant adjunct information in the identification of former settlement places. In the ground survey of the test region, for example, a visible soil change was one of three criteria out of which two had to be fulfilled to mark a former settlement site, the others being topographic moundedness and presence of surface artifacts above a certain density threshold (Ur, 2002a, p. 61). Being able to identify soil changes not only on the ground, but also remotely from satellite data, a combined evaluation of CORONA data –or other high resolution imagery– together with the proposed ASTER classification, might now allow to pursue a relevant part of such a survey in advance and "offsite". Potential sites as identified via the ASTER classification could be used to guide ground control in a more efficient manner, which would allow broader geographic coverage during the limited time available for field survey.

6 CONCLUSIONS

In the identification of a spectral signature for ancient settlement places in ASTER satellite imagery, we developed a novel spatio-temporal sampling strategy. It yields a robust classifier which can be applied to any data available for a certain region, enables a fusion of all available images, and allows the identification of the spectral signature of ancient settlement sites in ASTER imagery. As an adjunct source of information in addition to highly resolved monochrome imagery (CORONA, QuickBird) the main potential of this approach is in a fast screening of wide regions and the mapping of whole settlement systems (Fig. 7), for survey planning and risk assessment in conservational issues. In regions which are not accessible to ground survey at present –in the Middle East, for example, wide areas of Iran, Iraq, Afghanistan (Petrie, 2007)– spectral imagery will provide the only means to access information on character and composition of the ground at all.

REFERENCES

Altaweel, M., 2005. The use of aster satellite imagery in archaeological contexts. Archaeological Prospection 12, pp. 151–166.

Apan, A., Kelly, R., Jensen, T., Butler, D., Strong, W. and Basnet, B., 2002. Spectral discrimination and separability analysis of agricultural crops and soil attributes using aster imagery. In: Proc 11th ARSPC (Australasian Remote Sensing and Photogrammetry Conference), Brisbane/Australia, pp. 396–411.

Benediktsson, J. and Kanellopoulos, I., 1999. Classification of multi-source and hyperspectral data based on decision fusion. IEEE Transactions on Geoscience and Remote Sensing 37, pp. 1367–1377.

Breiman, L., 2001. Random forests. Machine Learning 45, pp. 5–32.

Briem, G., Benediktsson, J. and Sveinsson, J., 2002. Multiple classifiers applied to multisource remote sensing data. IEEE Transactions on Geoscience and Remote Sensing 40, pp. 2002–2291.

Bruzzone, L., Cossu, R. and Vernazza, G., 2004. Detection of land-cover transitions by combining multidate classifiers. Pattern Recognition Letters 25, pp. 1491–1500.

Bruzzone, L., Prieto, D. and Serpico, S., 1999. A neural-statistical approach to multitemporal and multisource remote-sensing image classification. IEEE Transactions on Geoscience and Remote Sensing 37, pp. 1350–1359.

Chattopadhyay, A. and Dutta, S., 2006. Mapping monsoonal soil wetness regions from multi-temporal vegetation dataset. International Journal of Remote Sensing 27, pp. 4693–4700.

Hayes, D. and Cohen, W., 2007. Spatial, spectral and temporal patterns of tropical forest cover change as observed with multiple scales of optical satellite data. Remote Sensing of Environment 106, pp. 1–6.

Im, J., Rhee, J., Jensen, J. and Hodgson, M., 2007. An automated binary change detection model using a calibration approach. Remote Sensing of Environment 106, pp. 89–105.

Jeon, B. and Landgrebe, D., 1999. Decision fusion approaches for multitemporal classification. IEEE Transactions on Geoscience and Remote Sensing 37, pp. 1227–1233.

Lahiri, S., 2003. Resampling Methods for Dependent Data. Springer-Verlag, New York and Heidelberg.

Lasaponara, R. and Masini, N., 2007. Detection of archaeological crop marks by using satellite QuickBird multispectral imagery. Journal of Archaeological Science 34, pp. 214–221.

Liaw, A. and Wiener, M., 2002. Classification and Regression by randomForest. R News 2, pp. 18–22.

Masini, N. and Lasaponara, R., 2006. Satellite-based recognition of landscape archaeological features related to ancient human transformation. Journal of Geophysical Engineering 3, pp. 230–235.

Menze, B., Ur, J. and Sherratt, A., 2006. Detection of ancient settlement mounds – Archaeological survey based on the SRTM terrain model. Photogrammetric Engineering and Remote Sensing 72, pp. 321–327.

Petrakos, M., Benediktsson, J. and Kanellopoulos, I., 2001. The effect of classifier agreement on the accuracy of the combined classifier in decision level fusion. IEEE Transactions on Geoscience and Remote Sensing 39, pp. 2539–2546.

Petrie, C., 2007. Remote sensing inaccessible Routes: Traceing Old Routes from India to Afghanistan. online. http://www.archatlas.org/.

Rosen, A., 1986. Cities of Clay: The Geoarchaeology of Tells. University of Chicago Press, Chicago and London.

Sherratt, A., 2004. Spotting tells from space. Antiquity 78, online. http://antiquity.ac.uk/ProjGall/sherratt/.

Ur, J., 2002a. Settlement and Landscape in Northern Mesopotamia: The Tell Hamoukar Survey. Akkadica 123, pp. 57–88.

Ur, J., 2002b. Surface Collection and Offsite Studies at Tell Hamoukar, 1999. Iraq 64, pp. 15–43.

Ur, J., 2004. Urbanism and Society in the third Millennium upper Khabur Basin. PhD thesis, The Department of Near Eastern Languages and Civilizations, The University of Chicago.

Wilkinson, K., Beck, A. and Philip, G., 2006. Satellite Imagery as a Resource in the Prospection for Archaeological Sites in Central Syria. Geoarchaeology 21, pp. 735–750.

Wilkinson, T., 1994. The Structure and Dynamics of Dry-Farming States in Upper Mesopotamia. Current Anthropology 35, pp. 483–520.

Wilkinson, T., 1997. Archaeological Survey of the Tell Beydar Region, Syria. Prepols, Turnhout/Belgium.

Xiao, X., Boles, S., Frolking, S., Li, C., Babu, J., Salas, W. and Moore, B., 2006. Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. Remote Sensing of Environment 100, pp. 95–113.

Zhu, L. and Tateishi, R., 2006. Fusion of multisensor multitemporal satellite data for land cover mapping. International Journal of Remote Sensing 27, pp. 903–918.