CLASSIFICATION USING OBJECTS FROM HIGH AND LOW RESOLUTION IMAGES. AN EXAMPLE OF BURROW SYSTEM AND RODENT HABITAT MAPPING IN KAZAKHSTAN

L.I. Wilschut [a]*, E.A. Addink[a], S.M. de Jong[a], S.A. Davis[b], A. Laudisoit[c], L.A. Burdelov[d], V. Dubyanskiy[d], V.S. Ageyev[d], V.G. Meka-Mechenko[d], M. Begon[c]

[a]Dept. of Physical Geography, Utrecht University, PO Box 80115, 3508 TC Utrecht, The Netherlands - (l.i.wilschut, e.a.addink, s.m.dejong)@uu.nl
[b]School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, Victoria 3000, Australia
[c]Institute of Integrative Biology, The University of Liverpool, Crown Street, Liverpool L69 7ZB, UK
[d] Kazakh Scientific Centre for Quarantine and Zoonotic Diseases (KSCQZD), Almaty, Kazakhstan

KEY WORDS: landscape units, Great Gerbil, bubonic plague, SPOT5XS, Landsat ETM+, PreBalkash region

ABSTRACT:

Bubonic plague is a zoonotic disease and has a number of hot spots in the world among which southern Kazakhstan. In this study we investigated the use of earth observation to improve our understanding of the spatial and temporal patterns of this disease. The bubonic plague bacteria, *Yersinia pestis*, are spread by fleas that feed on rodents. In southeastern Kazakhstan, a semi-desert area, these rodents are great gerbils that live in burrow systems. The burrows can be seen as stepping stones for the spreading of plague; when a burrow system is occupied, it can be infected by the plague bacteria and then serve as an infection source for gerbil families in neighboring burrow systems. The density of the burrow systems is thus an indicator for plague risk; lower densities indicate lower infection and transmission risks. We used high resolution earth observation images to map individual burrow systems and low-resolution images to assess environmental factors like food availability, topography and soil type. We created strata based on the environmental factors and assessed the quality of the classification when applying stratified and unstratified classification. We applied an object-based approach during the stratification and the classification. The segmentation for the stratification resulted in objects that were grouped according to their mean values on the environmental factors producing eight different strata. Each stratum was classified with a Random Forest trained with the specific characteristics as well as with an overall Random Forest based on the overall characteristics of the area. We used 2.5m-resolution SPOT5XS images to classify the burrow systems and 30m-resolution Landsat ETM+ images combined with 60m-SRTM data to define the strata. Results show that the stratification improves the classification accuracy for most strata. We conclude that as the environment of the burrow systems is different within different landscape units, their classification clearly benefits from a stratified approach taking into account the environmental factors.

1. INTRODUCTION

Bubonic plague is a zoonotic disease and has several foci in the world among which southern Kazakhstan. In this study we investigated the use of earth observation to improve our understanding of the spatial and temporal patterns of this disease.
The bubonic plague bacterium, *Yersinia pestis*, is spread by fleas that feed on rodents (Gage & Kosoy, 2005). In southeastern Kazakhstan, a semi-desert area, these rodents are great gerbils which are social animals that live in burrow systems (Naumov & Lobachev, 1975). These burrows measure around 25m in diameter and have an entirely bare surface in the ecological centre and are heavily grazed directly around it. Each burrow system hosts a gerbil family with one male, several females and their off-spring. The burrow systems can be seen as stepping stones for the spreading of plague (Davis et al., 2008); when a burrow system is occupied, it can be infected by the plague bacteria and then serve as an infection source for other burrow systems. Adult gerbils may visit colonies up to 400 m away while dispersal movements seem limited to approximately 5 km (Kausrud et al., 2007). The density of the burrow systems is thus likely to play an important role in the infection and transmission risks of bubonic plague. The aim of this study was to investigate whether the classification of burrow systems improves when a stratification into landscape units is applied. In pixel-based studies stratification has been used several times to enhance classification studies, but there are few examples of using stratification in object-based studies.
We used objects derived from low-resolution images to assess environmental factors like food availability, topography and soil type and to create strata. Next we used high spatial resolution earth observation images to map individual burrow systems within the strata.

2. DATA AND METHODS

2.1 Study area

Our study area is located in eastern Kazakhstan south of Lake Balkash and measures 200x300 km^2. The landscape is a steppe area shaped by aeolian and fluvial processes. The former delta of the river Ili can still be recognized through the presence of inactive river channels while large dune fields cover the area further to the east. Climate is harsh with annual minimum/maximum values of -40ºC/+40ºC and a mean precipitation of less than 200mm/yr (Naumov &

* Corresponding author.

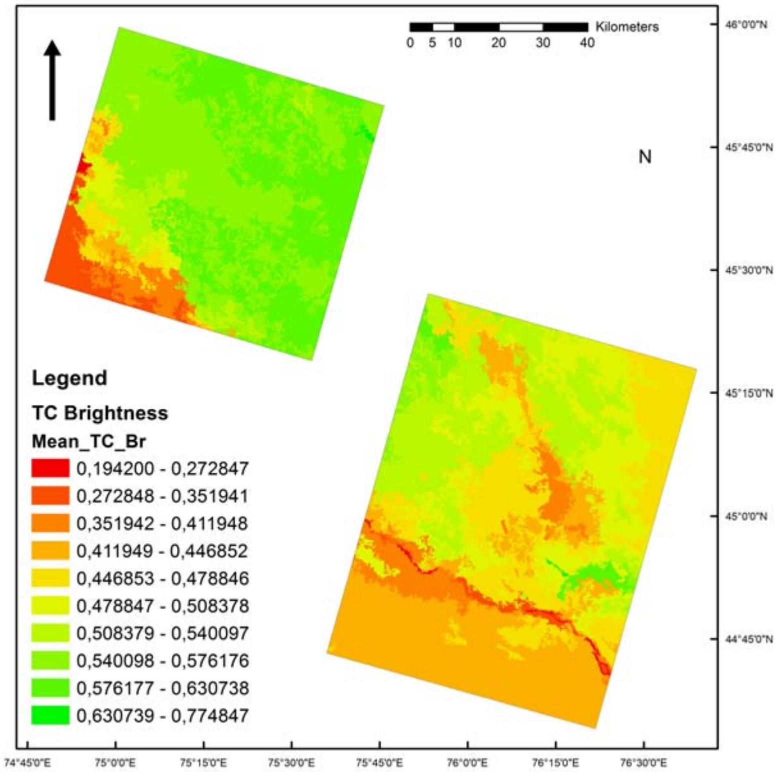

Figure 1. Values of the Tasselled Cap Brightness for the two areas covered by the SPOT5-XS images

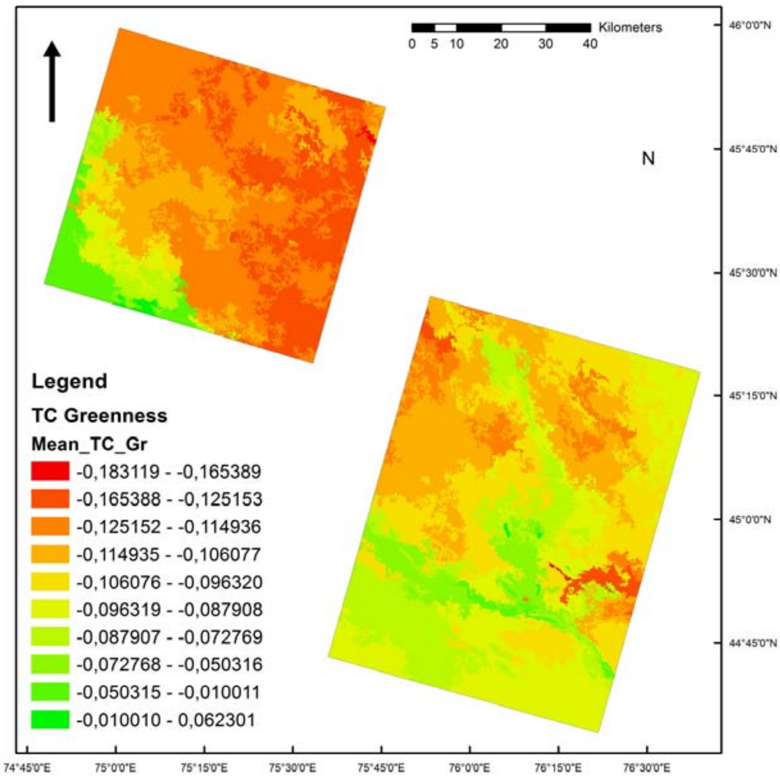

Figure 2. Values of the Tasselled Cap Greenness for the two areas covered by the SPOT5-XS images

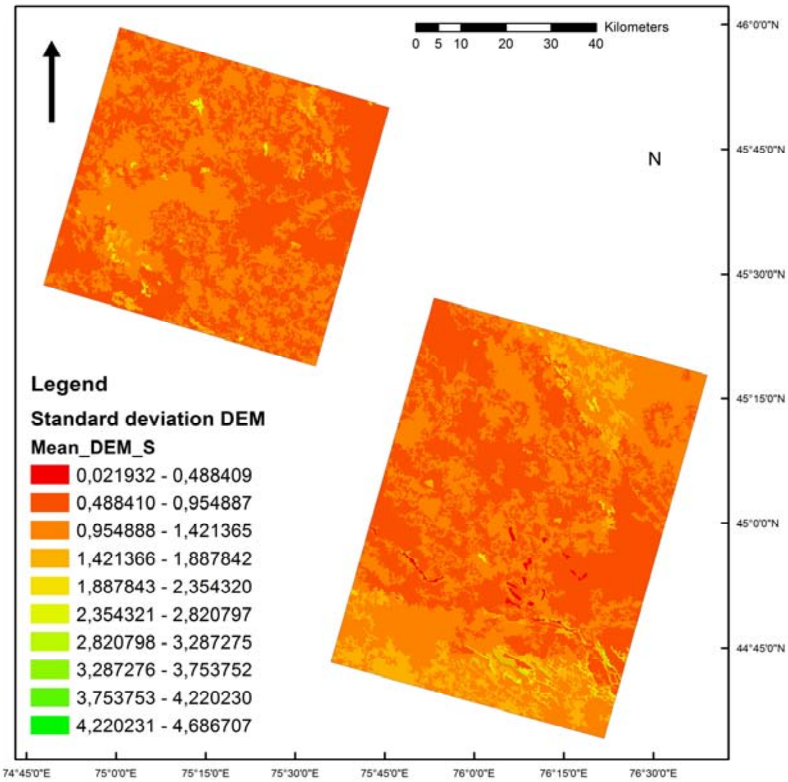

Figure 3. Values of the local variation of the SRTM DEM for the two areas covered by the SPOT5-XS images

Lobachev, 1975). The area is a plague focus, i.e. an area where host, vector and pathogen are present, and is intensely monitored by the Anti-Plague Institute in Almaty (Atshabar et al. 2010). It is divided in sectors of 10x10km^2 and each sector is sampled twice a year to test for the presence of plague in the great gerbils and the fleas, and to get an indication of occupancy density of the burrow systems.

2.2 Data

We had two field campaigns to collect data on the burrow systems in September 2010 and April 2011. We sampled seven sectors by inventorying all burrow systems within 200x200m^2 areas. In total, 54 squares were studied. For each encountered burrow system its size was measured and GPS coordinates were recorded. In total 904 burrow systems were included.

We have two sets of spaceborne imagery available to map the density of the burrow systems over the study area. We have two high-resolution multispectral SPOT5 images with 2.5m pixels recorded in October 2010. These images cover 60x60km^2 and 60x80 km^2. They only show part of the area but all field locations are included. Besides, we have four Landsat ETM+ images with 30m pixels recorded in 2000, 2001 and 2002 which do cover the full study area.

All four images were calibrated to Top-Of-Atmosphere (TOA) reflectance values (Chander et al., 2009). We applied the Tasselled Cap transformation (Crist and Cicone, 1984) to the Landsat ETM+ images to maximize the information about the landscape in the images. The Tasselled Cap transformation applies a linear rotation to the seven optical bands of Landsat ETM+ resulting in new dimensions such as Brightness and Greenness. The rotation is based on points with a known physical meaning like green vegetation, dark and bright soil. Brightness (figure 1) is a measure of overall reflectance differentiating soil types and Greenness (figure 2) displays the contrast between the sum of near-infrared spectral reflectance bands and the sum of visible reflectance and provides a measure for the presence and density of green vegetation. A Digital Elevation Model derived from SRTM data is available for the entire area with a spatial resolution of 60m. The DEM was used to calculate the topography standard deviation in a 3x3 window representing the local topography of floodplains, steppe and dunes (SRTMsd) (figure 3). The combination of Brightness, Greenness and local topography represents the soil characteristics relevant to building burrow systems, food availability and elevation differences relevant to potential flooding and shelter. It provides a detailed overview of the variation in the landscape and yields a good scientific basis for segmentation.

2.3 Methods

The aim of this study is to investigate whether classifying the landscape as separate units (based on Landsat ETM+) improves classification results of high resolution images (SPOT XS 2.5m) compared to classifying the entire area in one run without landscape differentiation. Therefore, we first created strata representing different landscape units. The first step was to apply a segmentation to the Brightness, Greenness and SRTMsd data. The heterogeneity was chosen such that the resulting objects matched the landscape units identified by eye. They had an average size of 7km^2. Next, the average Brightness, Greenness and SRTMsd values were determined for each object and the median values for all segments were determined. Each object was then labelled with one of eight classes, i.e. the low/high combinations of the three variables (figure 4).

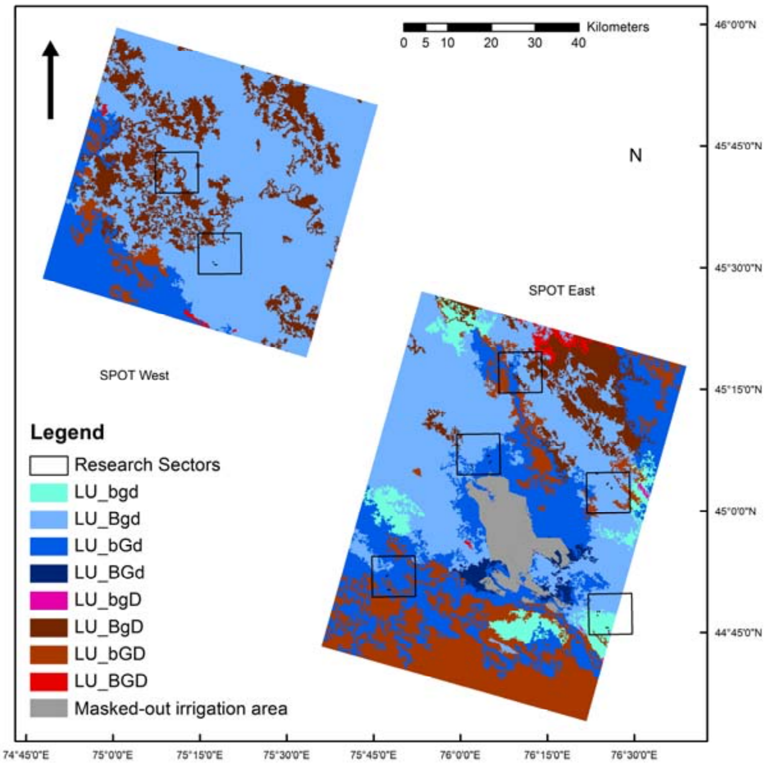

Figure 4. The landscape units used for stratification characterized by low (small letters) or high (capital letters) values for Brightness, Greenness and SRTMsd data.

All objects belonging to one class formed a stratum in the next step where the burrow systems were to be classified. In classification studies, stratification in advance of classification, is often used when the spectral signatures of the studied features varies across the study area. In those cases, it may help to have separate rule sets per landscape. In this study, the spectral characteristics of the burrow systems are not unique (Addink et al. 2010) as they mainly consist of bare soil, a characteristic they share with waterlogged areas, roads and patches of bare sand. Moreover, the surroundings can differ significantly in the different landscape types, varying from hardly any vegetation cover, to a dense bush cover (dominated by *Saxaul specs*), to a complete lichen cover. The shape of the burrow systems is circular in most landscape types but changes to more elongated shapes in the dune areas where the burrows are built on the slopes parallel to the dune ridges.

To identify the burrow systems in the images we applied Random Forests (RF) (Breiman, 2001), a machine learning algorithm that uses multiple classification trees. A classification tree is a decision tree, where each split results in more homogenous sub-groups. So starting from the entire training set with an overall heterogeneity with burrow systems and non-burrow systems, each split will provide two subgroups with a higher homogeneity. At each split a decision is made based on the threshold of one variable. The attribute value for an object is either higher or lower than the threshold value and the object follows the path through the tree until it reaches the end and gets assigned to a class. The variables used at the splits are found by iterating through all variables and associated values in the training set and finding the combination that provides the most homogenous subgroups.

We selected 65 spectral, shape or neighbourhood variables as input to the Random Forests. The classification we applied in this study is the same as described by Addink et al. (2010) except that we replaced the classification trees by the random forests. So by the Random forests objects were identified that met the criteria of burrow systems. Next a rule was applied that prohibits neighbouring objects both being burrow systems and that selects the brightest neighbour as the final burrow system.

From the fieldwork we had 904 locations of burrow systems and 3098 locations of non-burrow systems. We created a training set containing 80% of the burrow systems from each 200m-square and made sure that the number of non-burrow systems equalled the number of burrow systems. The remaining 20% were used for validation.

2.4 Results

The stratification revealed eight different landscape types with high or low values for Brightness, Greenness and SRTMsd. For five of those we had collected field data, so we could create a Random Forests to classify the burrow systems. In total we had six RF's one for the overall unstratified area and five for the represented strata.

The Random Forests show that different variables are being selected for different landscape units. For all forests the neighbour variables were dominant, but the ratio between the spectral and the shape variables varies between 0% to 100% shape and vice versa.

The overall accuracy values improved for four out of five strata. The differences between individual strata for stratified classification varied from a decrease of accuracy of 6% to an increase of 8%. The difference between User's and Producer's accuracy values for individual strata was 15% for

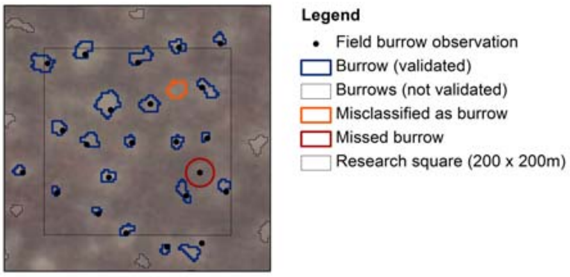

Figure 5. SPOT5 image with classification results. Validated/not validated refers to the presence/absence of field observations.

the unstratified classification and 10% for the stratified approach, indicating a more robust classification result with the stratified approach.
Figure 5 shows one of the field squares in the SPOT image and its classification. It clearly shows that the burrow systems can be recognized as lighter spots in the images. The field observations of the burrow systems were all recognized during the classification. The one exception is a burrow system without a lighter spot. Besides, one of the burrow systems in the classification does not match a field observation.

2.5 Discussion

The stratification improves the accuracy of the classification. So far we created one set of stratification objects, which already resulted in an improvement of the classification. Fine-tuning the scale of the stratification will probably result in even more accurate classifications.

For one land unit the classification with stratification resulted in a less accurate classification result (i.e. overall accuracy). The validation set for this land unit was however very small i.e. only two field reference points, and we believe that this might give a bias in the classification results of the stratified approach versus the non-stratified approach.

In this study we used the Greenness and Brightness values of one moment for the stratification. Alternatives could be found by adding a temporal component to strata definition by using time series of NDVI or meteorological data.

2.6 Conclusions

In this study we investigated whether classification of burrow systems in a semi-desert in environment in Kazakhstan would benefit from a stratification into landscape units. For four out of five strata the results did indeed improve, which is a clear recommendation to apply a landscape stratification. The result of the classification is a burrow-density map which is essential to model the spreading of bubonic plague.
The stratification stemmed from a Tasselled Cap transformation of Landsat ETM+ images and local variance of SRTM DEM data. These variables were segmented into landscape units. The burrow systems within those units were then identified through an object-based approach combined with Random Forests to distinguish burrow systems from non-burrow systems. As both the landscape units and the burrow systems do not have a unique spectral signature, this study would not have been possible without object-based image analysis.

2.7 Acknowledgements

This project was part of the Wellcome Trust project (WT 046498) ‘Advances in the prediction of plague outbreaks in Central Asia’ and the NWO project (NWO Vici 918.56.620) ‘Population dynamics and control of infectious diseases: modelling interaction at different scales’.

The SPOT5 images were obtained through the ISIS programme of CNES and SPOTIMAGE in ISIS projects 395 and 476 ‘Plague Surveillance’, Copyright CNES.

2.8 References

Addink, E. A., De Jong, S. M., Davis, S. A., Dubyanskiy, V., Burdelov, L. A. & Leirs, H. (2010). The use of high-resolution remote sensing for plague surveillance in Kazakhstan. Remote Sensing of Environment, 114, 674-681.

Atshabar, B.B., Burdelov, L.A., Ageyev, V.S., Aukabirov, S.A., Dubyanskiy, V.M., Grazhdanov, A.K., et al. (2010). Atlas of Bacterial and Virus Zoonotic Infections Distribution in Kazakhstan. (pp. 122). Almaty, Kazakhstan: M.Aikimbayev´s KSCQZD

Breiman, L. (2001). Random Forests. Machine Learning, 45, 5-32

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

Crist E.P. and R.C. Cicone, 1984, A physically-based transformation of the Thematic Mapper data: the TM tasselled cap. IEEE Transactions on Geoscience and Remote SensingGE-22, 256-263.

Davis, S., Trapman, P., Leirs, H., Begon, M., & Heesterbeek, J.A.P. (2008). The abundance threshold for plague as a critical percolation phenomenon. Nature, 454, 634-637

Gage, K.L., & Kosoy, M.Y. (2005). NATURAL HISTORY OF PLAGUE: Perspectives from More than a Century of Research*. Annual Review of Entomology, 50, 505-528

Kausrud L.K., H. Viljugrein, A. Frigressi, M. Begon, S. Davis, H. Leirs, V. Dubyanskiy and N.C. Stenseth, 2007, Climatically driven synchrony of Gerbil populations allows large-scale plague outbreaks. Proceedings of the Royal Society 274, 1963-1969.

Naumov, N.P., & Lobachev, V.S. (1975). Ecology of desert rodents of the U.S.S.R. (Jerboas and Gerbils). In I. Prakash, & P.K. Ghosh (Eds.), Rodents in desert environments (pp. 549-598). The Hague, The Netherlands: Dr. W. Junk Publishers