Available online at www.sciencedirect.com

SCIENCE DIRECT

Remote Sensing of Environment 93 (2004) 42–52

Remote Sensing of Environment

www.elsevier.com/locate/rse

Defining approaches to settlement mapping for public health management in Kenya using medium spatial resolution satellite imagery

Andrew J. Tatem[a,*], Abdisalan M. Noor[a,b], Simon I. Hay[a,b]

[a]TALA Research Group in the Department of Zoology, Tinbergen Building, University of Oxford, South Parks Road, Oxford OX1 3PS, UK
[b]Public Health Group, KEMRI/Wellcome Trust Research Laboratories, P.O. Box 43640, 00100 GPO, Nairobi, Kenya

Received 17 March 2004; received in revised form 31 May 2004; accepted 3 June 2004

Abstract

This paper presents an appraisal of satellite imagery types and texture measures for identifying and delineating settlements in four Districts of Kenya chosen to represent the variation in human ecology across the country. Landsat Thematic Mapper (TM) and Japanese Earth Resources Satellite-1 (JERS-1) synthetic aperture radar (SAR) imagery of the four districts were obtained and supervised per-pixel classifications of image combinations tested for their efficacy at settlement delineation. Additional data layers including human population census data, land cover, and locations of medical facilities, villages, schools and market centres were used for training site identification and validation. For each district, the most accurate approach was determined through the best correspondence with known settlement and non-settlement pixels. The resulting settlement maps will be used in combination with census data to produce medium spatial resolution population maps for improved public health planning in Kenya.

Keywords: Landsat TM; JERS-1 SAR; Neural network; Texture; Settlement mapping; Kenya; Public health; Population

1. Introduction

Ninety percent of projected global urbanization will be concentrated in low-income countries (United Nations, 2002). In Africa 38% of the 784 million inhabitants were urban dwellers in 2000. This is estimated to increase to 55% by 2030 as virtually all of this population doubling will be concentrated in urban areas (United Nations, 2002). The profound development and epidemiological impacts of these changes are increasingly being realised (Harpham, 1997; McMichael, 2000; Prothero, 2001). Urban dwellers face a very different set of health risks compared to their rural counterparts (Tatem & Hay, 2004). In Sub-Saharan Africa (SSA) for example, urban residents are more at risk from directly transmitted diseases such as tuberculosis (Banerjee et al., 1999; Floyd et al., 2002) and human immunodeficiency virus (HIV) (Abebe et al., 2003; Lagarde et al., 2003; Mhalu & Lyamuya, 1996) as well as certain vector-borne diseases such as dengue fever (Lines et al., 1994), but contrastingly are around 10 times less likely to receive a malaria-infected mosquito bite (Hay et al., 2000; Robert et al., 2003) and have significantly better access to health care facilities (Noor et al., 2003). Planning for the health consequences of urbanization in the countries of SSA relies on the provision of information and maps of settlement location, size and distribution. Whilst high-income countries often have mapping resources at their disposal to define and delineate settlements for such public health planning and management, this is often not the case for low-income regions of the world.

Satellite remote sensing offers a cheap and effective solution to mapping settlements and monitoring urbanization at a range of spatial scales. From continental scale

* Corresponding author. Tel.: +44 1865 271262.
E-mail address: andy.tatem@zoo.ox.ac.uk (A.J. Tatem).

0034-4257/$ - see front matter © 2004 Published by Elsevier Inc.
doi:10.1016/j.rse.2004.06.014

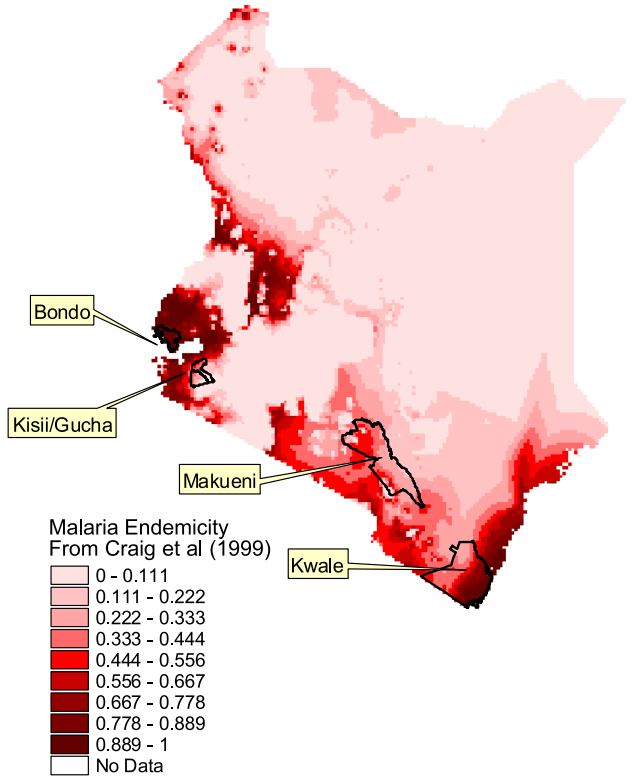

Fig. 1. Map showing location of the four districts on a malaria endemicity map of Kenya (Craig et al., 1999).

urban land use mapping using Advanced Very High Resolution Radiometer (AVHRR), Moderate Resolution Imaging Spectrometer (MODIS) and night-time imagery (Civco et al., 2002; Imhoff et al., 1997; Schneider et al., 2003; Vogelmann et al., 1998), to medium-scale regional settlement mapping and classification using Landsat Thematic Mapper (TM), Systeme Pour L'Observation de la Terre (SPOT) High Resolution Visible (HRV) and Earth Resources Satellite (ERS) 1/2 (Baraldi & Parmiggiani, 2000; Dell'Acqua & Gamba, 2003; Forster, 1983), down to the recent wave of settlement-scale studies using fine spatial resolution (Giada et al., 2003; Quartulli & Tupin, 2003; Roth, 2003; Tatem et al., 2001a,b), there now exist well-documented approaches for settlement delineation, classification and validation at all scales, reviewed in (Tatem & Hay, 2004). Few attempts have been made to use satellite sensor imagery in combination with census data to produce global and continental population maps (Landscan: Dobson et al., 2003, 2000; UNEP GRID: Deichmann, 1996; GPW 2.0: Deichmann et al., 2001). However, the administrative level of input census data for SSA is so coarse that population estimates at an administrative level fine enough to facilitate effective public health management can be inaccurate. The scarcity of reliable data for map validation and the difficulty in obtaining other data such as census statistics are cited as the main obstacles to settlement mapping in the region.

2. Study aims

The research detailed in this paper forms part of a larger project aimed at extending the application of GIS and remote sensing technology to the quantification of human population distribution to allow more accurate malaria risk mapping and disease burden estimation across Africa. The findings of this study will therefore form an important input to this research and be used to scale up settlement maps and subsequently population maps to Kenya level. This study hence has two main aims:

- Assess the utility of both multispectral and radar imagery, along with derived texture layers, in producing settlement maps across four contrasting Kenyan districts at a spatial scale fine enough to facilitate application in public health management.
- Use the findings to inform on which methods to use in genarating Kenya-wide settlement maps.

Investigations into appropriate spatial scales for human population mapping in terms of public health management in SSA are ongoing (Hay et al., 2004; Noor et al., 2004; Noor et al., 2003), therefore this study is directed at producing settlement maps at as fine a spatial scale as possible computationally, to allow for flexibility in future population map production.

2.1. Study areas and data

2.1.1. The districts

Four study districts were chosen to represent the full range of land cover and use types, topography, climate, settlement distribution and public health requirements found in Kenya. Fig. 1 shows the location of the four districts on a malaria endemicity map of Kenya (Craig et al., 1999), and Table 1 lists the features of each.

2.1.2. Data

Orthorectified Landsat 5 TM 30 m spatial resolution imagery in six spectral bands were acquired for each district (the coarse-spatial resolution thermal infrared band

Table 1
Features of each of the four study districts

District	Area (km^2)	Population	No. of Government of Kenya health facilities	Malaria ecology
Bondo	960	238,780	21	Perennial, intense
Kisii/Gucha	1310	952,725	44	Highland, acutely seasonal
Kwale	8295	496,133	50	Seasonal, intense
Makueni	8266	771,545	59	Semi-arid, acutely seasonal

Table 2
Satellite imagery specifications and features

District	Number of Landsat TM scenes required	Date of each Landsat TM scene	Number of JERS-1 SAR scenes required	Date of each JERS-1 SAR scene
Bondo	1	Mar 1995	5	Jan 1996
Kisii/ Gucha	4	Jan–Mar 1995	4	Jan 1996
Kwale	4	Jan–Feb 1995	15	Jan–Feb 1996
Makueni	4	Jan–Feb 1995	16	Jun 1996

was excluded). Japanese Earth Resources Satellite 1 (JERS-1) Synthetic Aperture Radar (SAR) 12.5 m spatial resolution imagery (processed to level 2.1) of wavelength 0.24 m was also acquired. Table 2 shows the number of individual scenes required to cover each district and the date each were taken on. For Kenya (1996) JERS-1 SAR coverage was the most recently available, and therefore Landsat TM of a similar time period was obtained to match this. The Landsat TM imagery was then atmospherically corrected (Richter, 1990), but no high-resolution DEMs of the districts were available for topographic correction of the imagery sets. For each district, the component Landsat TM and JERS-1 SAR images were mosaiced together using Erdas Imagine™ version 8.6 (ERDAS, 2002) to create single TM and SAR images for each district. Each image was then georegistered using Erdas Imagine™ to vector polygon enumeration area images of their respective district (Noor et al., 2003). Rather than risk valuable information loss and complications in texture measure derivation by resampling the SAR imagery to the spatial resolution of the Landsat TM imagery, nearest neighbour resampling was used to upscale the Landsat imagery to the spatial resolution of the JERS-1 SAR data. Within the Landsat TM images, approximately 8% of the Makueni image, and approximately 5% of the Kwale image contained cloud cover. Fig. 2(i) and (ii) show the Landsat TM and JERS-1 SAR images of Bondo District, respectively.

Data for map validation were a set of vector points and ancillary data for each district collected through ground survey in 2001 and 2002 using a GPS (Noor et al., 2003). These included the locations of health facilities, market centres, schools, villages and selected households. In addition, 1999 census data at the enumeration area level were secured for each district (Hay et al., 2004). Vector layers of roads, rivers, dams and swamps were also available and are described elsewhere (Noor et al., 2003, 2004). For classifier training on those parts of the imagery not covered by settlement land use, three different land cover maps of Kenya were obtained, so that specific land covers could be identified. These were the global 1 km land cover classification (Hansen et al., 2000), the FAO Africover map for Kenya (FAO, 2003) and an Africover-derived spatially aggregated land cover database (FAO, 2003).

2.2. Methodology

2.2.1. Classifier

Although the principal aim of the study was to assess the utility of various combinations of satellite sensor imagery in producing accurate settlement maps, an initial testing of parametric, non-parametric, per-pixel and super-resolution classifiers was carried out to ensure an effective classifier would be used in the major part of the study. The classifiers tested included minimum distance (ERDAS, 2002), Mahalanobis distance (ERDAS, 2002) and maximum likelihood (Chan et al., 2001; Stefanov et al., 2001). Additionally, a feed-forward artificial neural network with two hidden layers using the back-propagation learning rule (Kavzoglu & Mather, 2003; Paola & Schowengerdt, 1995a) was tested, along with a Hopfield neural network-based superresolution algorithm (Tatem et al., 2001a,b, 2003) run on the output of an area estimation neural network (Lewis et al., 1998).

The classifier judged to be most effective was used to produce settlement maps for each district. For the district of Bondo, which has areas of distinctly differing land covers,

(i) (ii) (iii)

Fig. 2. (i) Landsat TM RGB image of whole of Bondo with enumeration area administrative unit boundaries overlain. (ii) SAR image of whole of Bondo with enumeration area administrative unit boundaries overlain. (iii) Segmentation map for per-parcel classification.

Table 3
Band combinations assessed for settlement mapping

Band combination
SAR
LSAT
TEXTURE
SAR+TEXTURE
LSAT+TEXTURE
SAR+LSAT
SAR+LSAT+TEXTURE

SAR=JERS-1 SAR; LSAT=Landsat TM; TEXTURE=JERS-1 SAR-derived texture measures.

consequently resulting in differing signature mixing at the edge of settlements, a pre-classification segmentation was also evaluated. This involved firstly the application of an image segmentation algorithm (Ruefenacht et al., 2003) to all image layers. The algorithm was used to segment the district of Bondo into five spectrally unique and spatially contiguous zones. A map of these zones is shown in Fig. 2(iii). Separate training and classification was then carried out within each zone. To test the utility of each image type, seven combinations were tested for each district (see Table 3 for details).

2.2.2. Training area selection

For each of the four districts, the point data on market centres, health facilities and schools were utilised in the identification of settlement training sites, and census data were used as a check to ensure a settlement had indeed been identified. As a further test, the spectral signature, radar return and texture statistics of each settlement training site were then compared to those of other known settlements to ensure representative statistics had been extracted. The various land cover maps obtained were then used in combination with the census data and GIS layers to identify representative training samples for all non-settlement areas of the four districts. Training samples were only selected where all three land cover maps were in agreement with each other and the census and GIS data, to minimise error and its propagation through the methodology. This process was undertaken three times in order to provide alternative sets of training data to test the stability of classifiers.

2.2.3. Texture measures

For each JERS-1 SAR district image, eight texture feature images calculated from the grey level cooccurrence matrix (GLCM) method (Haralick, 1979; Haralick et al., 1973) were produced; mean, variance, contrast, homogeneity, dissimilarity, correlation, entropy and angular second moment (Baraldi & Parmiggiani, 1995; Haralick et al., 1973). Mean and variance are measured in terms of the GLCM, contrast measures local spatial frequency and homogeneity results in large values when the elements of the GLCM are concentrated on the main diagonal (i.e., spatially similar). Dissimilarity measures local contrasts, correlation examines GLCM similarities, while entropy measures image disorder and angular second moment indicates local uniformity. After initial exploratory analysis on average settlement sizes and following work on urban areas using similar spatial resolution imagery and texture, a 7×7 moving window size was used (Zhang et al., 2003).

2.2.4. Accuracy assessment

The ground-collected set of points was used as the principal source of settlement mapping assessment. These points were taken to represent locations of settlements to validate the accuracy with which each classification identified settlement pixels. In addition to these, the same points were used in combination with the land cover maps, census data and visible Landsat TM bands to identify an equal set of non-settlement points for each district. Half of these points were located on the edge of known settlements to assess the accuracy by which the settlement maps had delineated the extent of settlements. Buffer areas around known settlements were identified and points were randomly selected from within these. The other half were located in areas clearly at a distance from settlements to assess whether the predicted settlement maps had identified false areas of settlement. Non-settlement areas were identified and again points were randomly selected from within these.

Given the variation in populations and number of settlements in each district, the numbers of validation points for each district, excluding those falling under cloud cover in the Landsat TM imagery, were as follows: Bondo: 606 points; KisiiGucha: 222 points; Makueni: 144 points; Kwale: 94 points. The satellite sensor image-derived

Table 4
Accuracy statistics for each classifier type

	OA(%)	K	Settlement		Non-settlement	
			C(%)	O(%)	C(%)	O(%)
Minimum distance	81.3	0.626	21.3	24.8	24.8	21.3
Mahalanobis distance	90.3	0.811	12.6	15.2	15.2	12.6
Maximum likelihood	94.5	0.892	5.0	4.3	4.1	5.0
Artificial neural network	97.6	0.957	2.2	1.9	1.9	2.2
Hopfield network superresolution	97.8	0.961	0.8	2.0	2.0	0.9

OA=overall percentage correct, K=kappa, C=error of commission, O=error of omission.

Table 5
Accuracy statistics of each band combination and classification type

Band combination	Bondo		Kisii/Gucha		Makueni		Kwale	
	OA (%)	K	OA (%)	K	OA (%)	K	OA (%)	K
Full image classification								
SAR	44.77	−0.111	26.68	−0.476	31.53	−0.362	66.6	0.433
LSAT	90.38	0.806	89.53	0.792	73.79	0.415	83.54	0.69
TEXTURE	47.61	0.005	64.59	0.31	55.6	0.111	73.32	0.442
SAR+TEXTURE	58.1	0.164	38.29	−0.234	38.22	−0.184	75.16	0.523
LSAT+TEXTURE	92.22	0.821	82.15	0.662	74.51	0.477	87.33	0.8
LSAT+SAR	92.78	0.851	74.85	0.498	75.55	0.502	95.2	0.906
SAR+LSAT+TEXTURE	97.6	0.957	79.94	0.599	79.37	0.576	95.78	0.921
Pre-segmented classification								
SAR	52.34	0.044	n/a	n/a	n/a	n/a	n/a	n/a
LSAT	91.3	0.842	n/a	n/a	n/a	n/a	n/a	n/a
TEXTURE	59.12	0.17	n/a	n/a	n/a	n/a	n/a	n/a
SAR+TEXTURE	62.13	0.34	n/a	n/a	n/a	n/a	n/a	n/a
LSAT+TEXTURE	92.96	0.857	n/a	n/a	n/a	n/a	n/a	n/a
LSAT+SAR	94.85	0.893	n/a	n/a	n/a	n/a	n/a	n/a
SAR+LSAT+TEXTURE	98.52	0.974	n/a	n/a	n/a	n/a	n/a	n/a

OA=overall accuracy, K=kappa.

settlement maps were binary encoded into settlement/non-settlement pixels and the vector-point validation datasets were used to extract pixels for each district and measures of accuracy were calculated. Health facilities were found to be on average 400 m from settlement centres, so any pixels predicted as containing a settlement within this distance of a health facility validation point were counted as correct. Accuracy assessment measures included errors of omission and commission, overall percentage correct (Campbell, 1996) and Kappa (Ma & Redmond, 1995). Finally, a visual comparison between 1999 enumeration area census counts and settlement maps was made to check for any obvious inaccuracies. To test prediction stability, each classifier was tested on three sets of training data and accuracy assessed using the data described above. The three sets of accuracy statistics were then averaged to provide a set of overall accuracy statistics.

3. Results

Mahalanobis distance, minimum distance, maximum likelihood, artificial neural network and superresolution classification for Bondo were undertaken on all image types combined three times, using different training data each time. The results are shown in Table 4, and from these the artificial neural network approach was identified for use in the remaining part of the study.

Supervised neural network classification was carried out on the band combinations detailed in Table 3, both for the entirety of all four districts and within the segmented zones of Bondo shown in Fig. 2(iii). This was undertaken three times for each band combination, using a different set of training data each time to test classifier stability. The averaged results of accuracy assessment of each band combination's predicted settlement map are shown in Table 5, with the errors of commission and omission figures for the most accurate classification of each district in Table 6. No significant difference was found between accuracy statistics output from stability testing. Fig. 3 shows an area of Bondo from JERS-1 SAR imagery and Landsat TM red, green and blue bands, compared to the validation points, census data and predicted most accurate settlement map. For Bondo, the Landsat TM, JERS-1 SAR and texture combination of layers using the pre-segmented classification was found to most closely correspond to the validation points, with a Kappa value of 0.964. For Kisii/Gucha, the results differed, with the classification of just the Landsat TM imagery proving most accurate, classifying 86.5% of the verification points correctly. The verification points of both Makueni and Kwale were most accurately mapped using all available image sources, with kappa values of 0.569 and 0.915, respectively.

Table 6
Errors of commission (C) and omission (O) for the four settlement maps where the validation points were most accurately classified

	Settlement		Non-settlement	
	C(%)	O(%)	C(%)	O(%)
Bondo	1.7	1.8	1.8	1.7
Kisii/Gucha	11.0	14.3	12.8	12.2
Makueni	18.1	16.6	16.6	18.1
Kwale	6.8	0.0	0.0	7.9

4. Discussion

The principal aim of the study was to assess the utility of various combinations of satellite sensor imagery in produc-

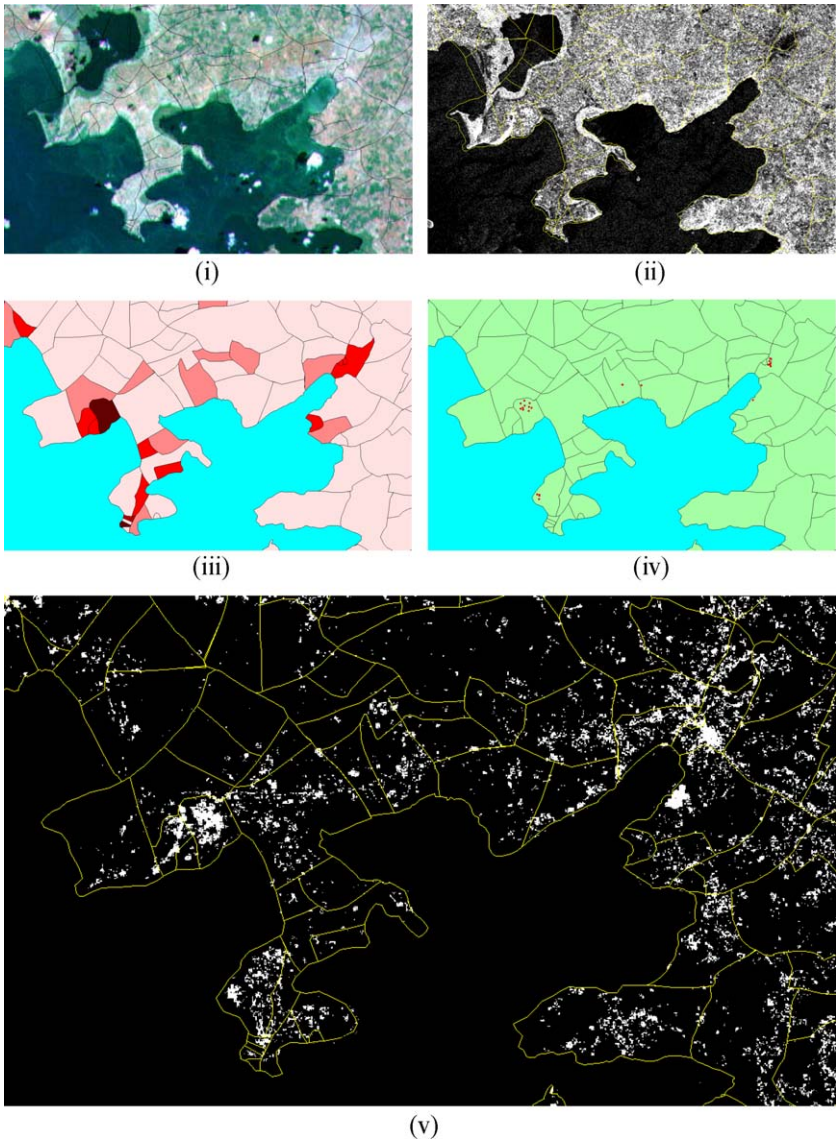

Fig. 3. Close-up view of an area of Bondo with enumeration area administrative unit boundaries overlain with: (i) Landsat TM RGB image, (ii) JERS-1 SAR image, (iii) Population density derived from 1999 census map of the area, (iv) Hosp/Mkt points for same area, (v) Predicted settlement map of the area.

ing accurate settlement maps, so testing of classifiers was restricted to one image combination for Bondo (the district with most land cover and topographical variation) and discussion here is kept brief. Supervised classification is not typically the process of choice in large area land cover or land use classifications due to the requirement for significant amounts of training data (Bauer et al., 1994), but the comprehensive set of ground-collected data available removed this constraint.

Table 4 shows that the superresolution algorithm produced the most accurate results, but the large increase in imagery size it produces at a spatial scale likely to be overly fine for public health purposes means such an approach is likely to be impractical when extending this study Kenya-wide. However, as computing power continues to grow it may find future application. The neural network proved the most accurate of the remaining classifiers, possibly due to the heterogeneous nature of settlements leading to multimodal responses, something that neural networks are better able to cope with than parametric classifiers that assume unimodal normally distributed class response. Therefore, the sole use of a feed-forward neural network classifier, accurately applied in other settlement mapping studies (Kanellopoulos et al., 1992; Paola & Schowengerdt, 1995b) ensured that the maps produced using each image combination could be compared quantitatively, thus not detracting from the overall focus of the paper. Preliminary testing found a perceptron architecture with two hidden layers to be most accurate for settlement mapping, although as other studies have found (e.g., Ardo et al., 1997) little difference in results were exhibited with other architectures.

Many studies have shown the benefit of applying texture algorithms to SAR imagery in order to extract further land

cover information (Arzandeh & Wang, 2002; Dekker, 2003; Gong et al., 1992; Karathanassi et al., 2000; Marceau et al., 1990), an approach currently underutilised in image classification (Franklin & Wulder, 2002). The methodology adopted for this study aimed to assess whether the addition of JERS-1 SAR derived textural information could increase settlement mapping accuracy. Several methods for image texture quantification exist (Carr & de Miranda, 1998; Dekker, 2003; Laine & Fan, 1993), but those based on the grey-level co-occurrence matrix (GLCM) have often been found to be most effective for satellite sensor imagery (Baraldi & Parmiggiani, 1995; Haralick et al., 1973; Zhang et al., 2003). Recent texture studies have also shown that the highest classification accuracies are achieved with the inclusion of the maximum number of texture layers available (Dekker, 2003; Zhang et al., 2003). This, along with combinations of Landsat TM bands was explored briefly for Bondo, but findings showed that the loss of even a single texture measure or TM band proved detrimental to mapping accuracy.

The individual benefits of medium-spatial resolution multispectral imagery, SAR imagery and SAR-derived texture measures for settlement mapping have been widely extolled (e.g., Dell'Acqua & Gamba, 2003; Forster, 1980; Henderson & Xia, 1999; Iisaka & Hegedus, 1982; Yuan et al., 1997), but very few studies have attempted to quantify the potential accuracy gains from combining each (e.g., Lichtenegger et al., 1991; Toll, 1985), and no studies of note have examined their potential in SSA. The benefits to settlement mapping of combining these different layers are amply demonstrated in Fig. 4 where the strengths and weaknesses of each are highlighted. Small settlements of grass-roofed houses surrounded by grassland are impossible to identify solely by multispectral imagery, whereas the high radar return from such regularly shaped objects is detectable, a benefit of radar imagery also noted in other studies (Henderson & Xia, 1997; Lo, 1984, 1986). In contrast, settlements on steep slopes surrounded by forest prove difficult to isolate using SAR imagery alone, but produce enough of a spectral reflectance difference with their surroundings to be clearly delineated by a classifier using multispectral data. In addition, texture measures can add further discriminating layers of information.

For Bondo, Table 5 demonstrates that while solely Landsat TM imagery produces excellent results in classifying over 90% of the verification points correctly, the addition of extra layers of information in the form of SAR backscatter and SAR-derived texture measures enables the classifier to improve further on these results. Additional improvements are also witnessed when segmenting the imagery and classifying within each segment. The varying land cover zones of Bondo mean that each group of settlements have distinct reflectance and backscatter signatures where mixing with the surrounding land covers occur at the settlement edge. Segmentation by land cover zone therefore allows the classifier to more clearly identify settlement pixels in feature space. Kisii/Gucha by contrast does not exhibit the high levels of accuracy as Bondo does

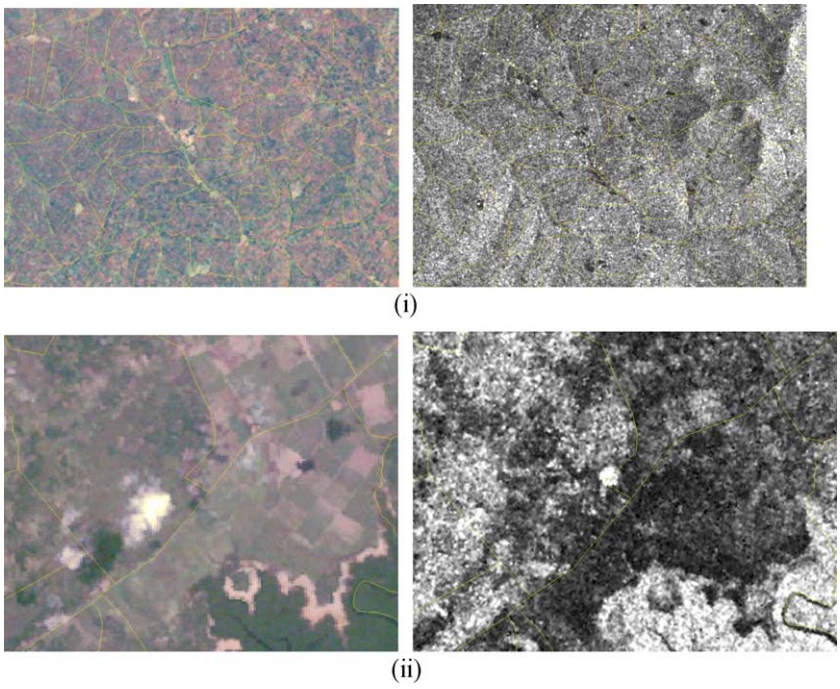

Fig. 4. (i) Close-up view of a small Kisii/Gucha settlement located in a mountainous region at the base of a valley, connected by a northwest to southeast road. It is visible in RGB Landsat TM bands, but invisible in SAR image next to it. (ii) Close-up view of a small Kwale settlement located in a flat agricultural region, surrounded by fields. It is invisible in RGB Landsat bands, but visible in SAR image next to it (enumeration area administrative unit boundaries overlain on each).

in correctly classifying the verification points. The use of just Landsat TM imagery produces the largest overall percentage correct and Kappa value, probably due to the mountainous nature of the district. The varying topography and lack of a high spatial resolution DEM to correct for the effects this has on the SAR imagery and derived texture layers results in these imagery layers being a hindrance in settlement mapping, rather than the help they were in relatively flat Bondo. Makueni represented the toughest challenge of the four districts, covering such a wide range of topographies, land covers and land uses, in addition to being sparsely populated. Whereas settlements make up a relatively large proportion of the land cover in Bondo, Kwale and Kisii/Gucha, many of them large enough to provide an excellent array of training pixels and most with consistently similar spectral reflectance characteristics, this was not the case in Makueni. The sparsely scattered settlements, each small enough to be made up almost entirely of pixels of mixed spectral reflectance from the various surrounding land covers, made identification of a unique 'settlement' reflectance signature very difficult, consequently leading to lower accuracy figures than the three other districts. Nevertheless, the use of all three types of imagery again proved to be the most accurate in classifying the verification points, with a kappa value of 0.569. Kwale produced a similar result, where combing Landsat TM, SAR and texture successfully mapped 95.74% of the verification points. The settlement error of omission of 0% showed also that all the settlement verification points were classified correctly. Finally, although the low number of validation points for Kwale may have inflated accuracy results, the flat topography of the district is reflected in the large percentage correct and kappa values when using solely SAR or texture layers, an indication of the potential benefits to the other districts of using topographically corrected SAR imagery.

In sum, the results clearly demonstrate that the combination of medium spatial resolution multispectral satellite imagery with similar scale SAR imagery and derived texture layers is effective in identifying and mapping settlements at medium-scale spatial resolution across the diverse landscapes of Kenya.

4.1. Future work

The natural continuation of this study is to utilise the findings and recommendations made here to extend to a settlement map of Kenya. Results indicate that accurate settlement maps can be produced using Landsat TM and JERS-1 SAR imagery, however, certain combinations appear to be more suited to certain topographies. This fact, and the success of the pre-segmented map of Bondo, suggests that any attempt to map settlements at Kenya level should be based on an initial division of the country into zones of similar topography, land cover, land use and settlement density. This could be achieved for example, by application of the spatial–spectral segmentation algorithm utilised in this paper on a countrywide DEM, coarse spatial resolution satellite imagery (e.g., MODIS) and road density layers. Results indicate that unless topographically corrected SAR imagery can be obtained, its application in Kenya-wide settlement mapping should be restricted to non-mountainous zones.

Extension to a Kenya level settlement map will require expansion from the 13 Landsat TM scenes and 40 JERS-1 SAR scenes used in this study to 35 Landsat and 360 SAR scenes. Although this will be technically challenging, it is well within modern processing and computing capabilities, as recent large-scale studies have shown (De Grandi et al., 2000; Franklin & Wulder, 2002; Fuller et al., 1994; Siqueira et al., 2003). Once a substantial imagery archive for Kenya is built up, future work may show that Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) imagery (Fujisada, 1994) represents a feasible finer spatial resolution substitute for Landsat TM imagery. Ideally, the application of a combination of very high spatial resolution imagery, such as that from IKONOS (Tanaka & Sugimura, 2001), Quickbird (Volpe & Rossi, 2003) or TerraSAR (Roth, 2003), to settlement mapping across Kenya would increase confidence with which settlements could be identified and delineated, but such an approach is both technically and financially prohibitive.

This study forms the starting point of a larger project aimed at quantifying human population distribution initially in Kenya, but ultimately across Africa, to increase the accuracy of malaria risk maps and disease burden estimates. This will be achieved by extending the findings of this work to produce settlement maps as input, along with other GIS and recent census information, to population mapping models (e.g., Deichmann, 1996; Deichmann et al., 2001; Eicher & Brewer, 2001; Langford, 2003; Tobler, 1979; Wright, 1936), and by investigating appropriate spatial scales for public health planning (Hay et al., 2004; Noor et al., 2003). The fine spatial resolution of these efforts will be more appropriate to the scale of human population and disease processes and an order of magnitude finer than previous attempts (Deichmann, 1996; Deichmann et al., 2001; Dobson et al., 2003, 2000; Sutton et al., 2001; Sutton, 2003). Furthermore, high spatial resolution human population distribution maps will help facilitate many of the wider aspirations of those involved in public health research across Africa with respect to commodity needs estimation, health service and intervention equity issues and some basic analyses on the efficacy of delivery mechanisms for control services.

5. Conclusions

Planning for the health consequences of urbanization in the countries of SSA relies on the provision of fine spatial resolution information and maps of settlement location, size and distribution. The scarcity of reliable map validation data

has meant that previous attempts at settlement and population maps in the region have been at too coarse a spatial resolution to facilitate effective public health management. Equally, this fact has meant that settlement mapping from satellite imagery has simply not been attempted, in an area of the world where its need is greatest.

Here, for the first time sufficient ground data has been obtained to assess approaches to settlement mapping across the unique landscapes of Kenya. This paper has presented an appraisal of the efficacy of two types of satellite imagery and derived texture layers in identifying and mapping settlements across four contrasting Kenyan districts. Results demonstrate that the combination of medium spatial resolution multispectral satellite imagery with similar scale SAR imagery and derived texture layers is effective in identifying and delineating settlements across Kenya, and that a neural network is an accurate approach to derive these maps. Information at such a scale has never before been produced, and will provide not only a valuable input to population mapping models designed for public health applications, but a valuable starting point for many other studies across SSA.

Acknowledgements

We thank Scott Goetz and Alastair Graham for comments on earlier drafts of the manuscript. SIH and AJT are funded by a Research Career Development Fellowship (to SIH) from the Wellcome Trust (#069045).

References

Abebe, Y., Ab, S., Mamo, G., Negussie, A., Darimo, B., Wolday, D., & Sanders, E. J. (2003). HIV prevalence in 72,000 urban and rural male army recruits, Ethiopia. *Aids*, *17*, 1835–1840.

Ardo, J., Pilesjo, P., & Skidmore, A. (1997). Neural networks, multitemporal Landsat Thematic Mapper data and topographic data to classify forest damage in the Czech Republic. *Canadian Journal of Remote Sensing*, *23*, 217–219.

Arzandeh, S., & Wang, J. (2002). Texture evaluation of Radarsat imagery for wetland mapping. *Canadian Journal of Remote Sensing*, *28*, 653–666.

Banerjee, A., Harries, A. D., & Salaniponi, F. M. L. (1999). Differences in tuberculosis incidence rates in township and in rural populations in Ntcheu District, Malawi. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *93*, 392–393.

Baraldi, A., & Parmiggiani, F. (1995). An investigation of the textural characteristics associated with the gray level cooccurrence matrix statistical parameters. *IEEE Transactions on Geoscience and Remote Sensing*, *33*, 653–666.

Baraldi, A., & Parmiggiani, F. (2000). Urban area classification by multispectral SPOT images. *IEEE Transactions on Geoscience and Remote Sensing*, *28*, 674–680.

Bauer, M. E., Burk, T. E., Ek, A. R., Coppin, P. R., Lime, S. D., Walsh, T. A., et al. (1994). Satellite inventory of Minnesota forest resources. *Photogrammetric Engineering and Remote Sensing*, *60*, 287–298.

Campbell, J. (1996). Introduction to remote sensing. London: The Guildford Press.

Carr, J. R., & de Miranda, F. P. (1998). The semivariogram in comparison to the co-occurrence matrix for classification of image texture. *IEEE Transactions on Geoscience and Remote Sensing*, *36*, 1945–1952.

Chan, J. C. -W., Chan, K. -P., & Yeh, A. G. -O. (2001). Detecting the nature of change in an urban environment: A comparison of machine learning algorithms. *Photogrammetric Engineering and Remote Sensing*, *67*, 213–225.

Civco, D. L., Hurd, J. D., Wilson, E. H., Arnold, C. L., & Prisloe, M. P. (2002). Quantifying and describing urbanizing landscapes in the Northeast United States. *Photogrammetric Engineering and Remote Sensing*, *68*, 1083–1090.

Craig, M. H., Snow, R. W., & le Sueur, D. (1999). A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitology Today*, *15*, 105–111.

De Grandi, G., Mayaux, P., Rauste, Y., Rosenquist, A., Simard, M., & Saatchi, S. S. (2000). The global rain forest mapping project JERS-1 radar modaic of tropical Africa: Development and product characterization aspects. *IEEE Transactions on Geoscience and Remote Sensing*, *38*, 2218–2233.

Deichmann, U., 1996. A review of spatial population database design and modelling. Santa Barbara, California, USA, National Center for Geographic Information and Analysis (NCGIA), University of California, Santa Barbara (UCSB).

Deichmann, U., Balk, D. & Yetman, G. (2001). Transforming population data for interdisciplinary usages: From census to grid.

Dekker, R. J. (2003). Texture analysis and classification of ERS SAR images for map updating of urban areas in the Netherlands. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 1950–1958.

Dell'Acqua, F., & Gamba, P. (2003). Texture-based characterization of urban environments on satellite SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 153–159.

Dobson, J. E., Bright, E. A., Coleman, P. R., & Bhaduri, B. L. (2003). Landscan: A global population database for estimating populations at risk. In V. Mesev (Ed.), *Remotely sensed cities* (pp. 267–279). London: Taylor and Francis.

Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., & Worley, B. A. (2000). LandScan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing*, *66*, 849–857.

Eicher, C. L., & Brewer, C. A. (2001). Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Science*, *28*, 125–138.

ERDAS (2002). *ERDAS field guide*. (6th ed.). Atlanta, GA: ERDAS.

FAO (2003). Africover, FAO. 2003.

Floyd, K., Blanc, L., Raviglione, M., & Lee, J. W. (2002). Resources required for global tuberculosis control. *Science*, *295*, 2040–2041.

Forster, B. (1983). Some urban measurements from Landsat data. *Photogrammetric Engineering and Remote Sensing*, *14*, 1693–1707.

Forster, B. C. (1980). Urban residential ground cover using Landsat digital data. *Photogrammetric Engineering and Remote Sensing*, *46*, 547–558.

Franklin, S. E., & Wulder, M. A. (2002). Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. *Progress in Physical Geography*, *26*, 173–205.

Fujisada, H. (1994). Overview of ASTER instrument on EOS AM-1 platform. *Proc. SPIE*, *2268*, 14–36.

Fuller, R. M., Wyatt, B. K., & Barr, C. J. (1994). The landcover map of Great Britain: An automated classification of Landsat Thematic Mapper data. *Photogrammetric Engineering and Remote Sensing*, *60*, 553–562.

Giada, S., De Groeve, T., Ehrlich, D., & Soille, P. (2003). Information extraction from very high resolution satellite imagery over Lukole refugee camp, Tanzania. *International Journal of Remote Sensing*, *24*, 4251–4266.

Gong, P., Marceau, D. J., & Howarth, P. J. (1992). A comparison of spatial feature extraction algorithms for land-use classification with SPOT HRV data. *Remote Sensing of Environment*, *40*, 137–151.

Hansen, M., DeFries, R., Townshend, J. R. G., & Sohlberg, R. (2000). Global land cover classification at 1km resolution using a decision tree classifier. *International Journal of Remote Sensing*, *21*, 1331–1365.

Haralick, R. M. (1979). Statistical and structural approaches to texture. *Proceeding of the IEEE*, *67*, 786–804.

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Texture features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *3*, 610–621.

Harpham, T. (1997). Urbanisation and health in transition. *The Lancet*(in preparation), 349.

Hay, S. I., Noor, A. M., Nelson, A., & Tatem, A. J. (2004). Demography for epidemiology: The precision of large-area human population maps. *International Journal of Epidemiology*, Submitted.

Hay, S. I., Rogers, D. J., Toomer, J. F., & Snow, R. W. (2000). Annual *Plasmodium falciparum* entomological inoculation rates (EIR) across Africa: Literature survey, internet access and review. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *94*, 113–127.

Henderson, F. M., & Xia, Z. -G. (1997). SAR applications in human settlement detection, population estimation and urban land use pattern analysis: A status report. *IEEE Transactions on Geoscience and Remote Sensing*, *35*, 79–85.

Henderson, F. M., & Xia, Z. -G. (1999). SAR applications in human settlement detection, population estimation and urban land use pattern analysis: A status report. *IEEE Transactions on Geoscience and Remote Sensing*, *35*, 79–85.

Iisaka, J., & Hegedus, E. (1982). Population estimation from Landsat imagery. *Remote Sensing of Environment*, *12*, 259–272.

Imhoff, M. L., Lawrence, W. T., Stutzer, D. C., & Elvidge, C. D. (1997). A technique for using composite DMSP.OLS "city lights" satellite data to map urban area. *Remote Sensing of Environment*, *61*, 361–370.

Kanellopoulos, I., Varas, I., Wilkinson, G. G., & Megier, J. (1992). Land-cover discrimination in SPOT HRV imagery using an artificial neural network—a 20-class experiment. *International Journal of Remote Sensing*, *13*, 917–924.

Karathanassi, V., Iossifidis, C., & Rokos, D. (2000). A texture-based classification method for classifying built areas according to their density. *International Journal of Remote Sensing*, *21*, 1807–1823.

Kavzoglu, T., & Mather, P. M. (2003). The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing*, *24*, 4907–4938.

Lagarde, E., van der Loeff, M. S., Enel, C., Holmgren, B., Dray-Spira, R., Pison, G., et al. (2003). Mobility and the spread of human immunodeficiency virus into rural areas of West Africa. *International Journal of Epidemiology*, *32*, 744–752.

Laine, A., & Fan, J. (1993). Texture classification by wavelet package signatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *15*, 1186–1191.

Langford, M. (2003). Refining methods for dasymetric mapping using satellite remote sensing. In V. Mesev (Ed.), *Remotely sensed cities* (pp. 137–156). London: Taylor and Francis.

Lewis, H. G., Brown, M., Tatnall, A. R. L., Nixon, M. S., & Manslow, J. (1998). Data analysis and empirical classification in FLIERS. Southampton: University of Southampton.

Lichtenegger, J., Dallemand, P., Reichert, P., & Rebillard, P. (1991). Multisensor analysis for land use mapping in Tunisia. *Earth Observation Quarterly*, *33*, 1–6.

Lines, J., Harpham, T., Leake, C., & Schofield, C. (1994). Trends, priorities and policy directions in the control of vector-borne diseases in urban environments. *Health Policy and Planning*, *9*, 113–129.

Lo, C. P. (1984). Chinese settlement pattern analysis using Shuttle Imaging Radar-A data. *International Journal of Remote Sensing*, *5*, 959–967.

Lo, C. P. (1986). Settlement, population and land use analyses of the North China Plain using Shuttle Imaging Radar-A data. *Professional Geographer*, *38*, 141–149.

Ma, Z. K., & Redmond, R. L. (1995). Tau-coefficients for accuracy assessment of classification of remote: Sensing data. *Photogrammetric Engineering and Remote Sensing*, *61*, 435–439.

Marceau, D. J., Howarth, P. J., Dubois, J. M., & Gratton, D. J. (1990). Evaluation of the grey-level co-occurrence matrix method for land-cover classification using SPOT imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *28*, 513–519.

McMichael, A. J. (2000). The urban environment and health in a world of increasing globalization: Issues for developing countries. *Bulletin of the World Health Organization*, *78*, 1117–1126.

Mhalu, F. S., & Lyamuya, E. (1996). Human immunodeficiency virus infection and AIDS in East Africa: Challenges and possibilities for prevention and control. *East African Medical Journal*, *73*, 13–19.

Noor, A. M., Gikandi, P. W., Hay, S. I., Muga, R. O., & Snow, R. W. (2004). Creating spatially defined databases for equitable health service planning in low-income countries: The example of Kenya. *Acta Tropica*, *91*, 239–251.

Noor, A. M., Zurovac, D., Hay, S. I., Ochola, S. A., & Snow, R. W. (2003). Defining equity in physical access to clinical services using geographical information systems as part of malaria planning and monitoring in Kenya. *Tropical Medicine & International Health*, *8*, 917–926.

Paola, J. D., & Schowengerdt, R. A. (1995a). A review and analysis of backpropagation neural networks for classification of remotely-sensed multispectral imagery. *International Journal of Remote Sensing*, *16*, 3033–3058.

Paola, J. D., & Schowengerdt, R. A. (1995b). A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification. *IEEE Transactions on Geoscience and Remote Sensing*, *33*, 981–996.

Prothero, R. M. (2001). Migration and malaria risk. *Health, Risk & Society*, *3*, 19–38.

Quartulli, M., & Tupin, F. (2003). *Information extraction from high resolution SAR data for urban scene understanding*. *Urban 2003*. Berlin: IEEE.

Richter, R. (1990). A fast atmospheric correction algorithm applied to Landsat TM images. *International Journal of Remote Sensing*, *11*, 159–166.

Robert, V., MacIntyre, K., Keating, J., Trape, J. F., Duchemin, J. B., Warren, M., et al. (2003). Malaria transmission in urban sub-Saharan Africa. *American Journal of Tropical Medicine and Hygiene*, *68*, 169–176.

Roth, A. (2003). *TerraSAR-X: A new perspective for scientific use of high resolution spaceborne SAR data*. *Urban 2003*. Berlin: IEEE.

Ruefenacht, B., Vanderzanden, D., & Morrison, M. (2003). New technique for segmenting images. Utah: USDA Forest Service Remote Sensing Applications Center.

Schneider, A., Friedl, M. A., McIver, D. K., & Woodcock, C. E. (2003). Mapping urban areas by fusing multiple sources of coarse resolution remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, *69*, 1377–1386.

Siqueira, P., Chapman, B., & McGarragh, G. (2003). The coregistration, calibration, and interpretation of multiseason JERS-1 SAR data over South America. *Remote Sensing of Environment*, *87*, 389–403.

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

Sutton, P. (2003). Estimation of human population parameters using nighttime satellite imagery. In V. Mesev (Ed.), *Remotely sensed cities* (pp. 301–333). London: Taylor and Francis.

Sutton, P., Roberts, D., Elvidge, C., & Baugh, K. (2001). Census from Heaven: An estimate of the global human population using nighttime satellite imagery. *International Journal of Remote Sensing*, *22*, 3061–3076.

Tanaka, S., & Sugimura, T. (2001). A new frontier of remote sensing from IKONOS images. *International Journal of Remote Sensing*, *22*, 1–5.

Tatem, A. J., & Hay, S. I. (2004). Measuring urbanization pattern and extent for malaria research: A review of remote sensing approaches. *Journal of Urban Health*, *81*, 363–376.

Tatem, A. J., Lewis, H. G., Atkinson, P. M., & Nixon, M. S. (2001). Super-resolution mapping of urban scenes from IKONOS imagery using a Hopfield neural network. International Geoscience and Remote Sensing Symposium, Sydney, Australia, IEEE, .

Tatem, A. J., Lewis, H. G., Atkinson, P. M., & Nixon, M. S. (2001). Super-resolution target identification from remotely sensed images using a Hopfield neural network. *IEEE Transactions on Geoscience and Remote Sensing*, *39*, 761–796.

Tatem, A. J., Lewis, H. G., Atkinson, P. M., & Nixon, M. S. (2003). Increasing the spatial resolution of agricultural land cover maps using a Hopfield neural network. *International Journal of Geographic Information Science*, *17*, 647–672.

Tobler, W. R. (1979). Smooth pycnophylactic interpolation of geographical regions. *Journal of the American Statistical Association*, *74*, 519–530.

Toll, D. (1985). Analysis of digital LANDSAT MSS and SEASAT SAR data for use in discriminating land cover at the urban fringe of Denver, CO. *International Journal of Remote Sensing*, *6*, 1209–1229.

United Nations (2002). World urbanisation prospects, 2002 revision, United Nations.

Vogelmann, J. E., Sohl, T., & Howard, S. M. (1998). Regional characterization of land cover using multiple sources of data. *Photogrammetric Engineering and Remote Sensing*, *64*, 45–57.

Volpe, F., & Rossi, L. (2003). Quickbird high resolution satellite data for urban applications. *Urban 2003*. Berlin: IEEE.

Wright, J. K. (1936). A method of mapping densities of population: With Cape Cod as an example. *Geographical Review*, *26*, 103–110.

Yuan, Y., Smith, R., & Limp, W. F. (1997). Remodeling census population with spatial information from landsat TM imagery. *Computers, Environment and Urban Systems*, *21*, 245–258.

Zhang, Q., Wang, J., Gong, P., & Shi, P. (2003). Study of urban spatial patterns from SPOT panchromatic imagery using textural analysis. *International Journal of Remote Sensing*, *24*, 4137–4160.