Spatial non-stationarity analysis to estimate dwelling units in Riyadh, Saudi Arabia

Mohammed Alahmadi

National Center for Remote Sensing Technology, Space and Aeronautics Research Institute, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia

ABSTRACT
Dwelling unit (population) data at finer zones are important for different applications. The census data are not available every year and published at coarse zones. While the remotely sensed data have addressed this limitation, analysis of spatial non-stationarity in the relationship between dwelling and detailed urban remote-sensing covariates for dwelling estimates is almost novel. Riyadh, Saudi Arabia, is chosen as a case study. The remote-sensing variables have been derived from QuickBird data using an object-based image analysis. To analyse the spatial non-stationarity, ordinary least squares (OLS) and geographically weighted regression (GWR) approaches are applied. The OLS models suggest that utilizing three-residential classes provides more accurate results than built area and residential built area. The GWR models are more accurate than the OLS model. This research proves that the GWR approach cannot completely handle the spatial non-stationarity problem without efficient explanatory variables. For example, the GWR model that uses three-residential classes is the best model to estimate dwelling compared with the GWR model that uses residential built area. This article confirms that the application of the OLS approach with efficient explanatory variables can successfully account the spatial non-stationarity and the results are relatively comparable with the GWR model.

Introduction
The size and distribution of small-area population are necessary for private business and governmental plans (Deng, Wu, & Wang, 2010). For instance, small-area population is powerful in transport planning and traffic engineering, land-use planning, site location analysis, urban sprawl analysis and national resource management (Martin & Williams, 1992; Pattnaik, Mohan, & Tom, 1998; Rees, Norman, & Brown, 2004). Despite of its importance, census population data have limitation in terms of (1) spatial resolution and (2) temporal resolution (Bhaduri, Bright, Coleman, & Urban, 2007). Regarding the former limitation, census population data are released at a larger geographical area to protect the confidentiality of respondents. In terms of the latter limitation, the census is commonly done every 10 years in most countries and they may be either outdated or not available at all in some developing countries (Rees et al., 2004). Moreover, the census and field survey are time consuming and labour intensive. These limitations have fostered the search for alternative methods to estimate population in a timely and effective way (Liu & Herold, 2007). The developed models for mapping population using a geographic information system (GIS) and remote sensing were classified into two groups: areal interpolation approaches and statistical modelling approaches (Wu, Qiu, & Wang, 2005). This paper is focused on the statistical modelling approaches. However, other studies used social media and phone calls for mapping population (Deville et al., 2014; Patel et al., 2017; Tsou, 2015).

The potential use of remotely sensed data for mapping population has been utilized since the 1950s by counting the dwelling units from aerial photographs (Anderson & Anderson, 1973; Hsu, 1971; Porter, 1956). As a result of the development in the sensor technologies and image classification algorithms, different approaches of population estimation using remotely sensed satellite data have been applied (Hostert, 2007). Spatial distribution of population has been estimated from coarse-spatial-resolution data (Bagan & Yamagata, 2015; Bhaduri et al., 2007; Dobson, Bright, Coleman, & Bhaduri, 2003; Lo & Welch, 1977; Tobler, 1969), medium-spatial-resolution data (Alahmadi, Atkinson, & Martin, 2013, 2016; Dong, Ramesh, & Nepali, 2010; Joseph, Wang, & Wang, 2012; Li & Lu, 2016; Lo, 2008; Patel et al., 2015), fine-spatial-resolution data (Alahmadi, Atkinson, & Martin, 2015; Lu, Im, Quackenbush, & Halligan, 2010; Silvan-Cardenas et al., 2010; Ural,
Hussain, & Shan, 2011) and Light Detection And Ranging (Dong et al., 2010; Silvan-Cardenas et al., 2010; Tomás, Fonseca, Almeida, Leonardi, & Pereira, 2016).

Inhabitants are commonly found in places classified as impervious surfaces, built area or residential built area. This information can be successfully extracted from the satellite data (Tatem, Noor, & Hay, 2005). Bivariate and multivariate regression models are commonly applied to demonstrate the correlation between population and urban remote-sensing variables (Langford, 2006; Lu et al., 2010; Lu, Weng, & Li, 2006; Silvan-Cardenas et al., 2010). A bivariate regression model was applied by Lu et al. (2006) to estimate population at the block group level in Marion, Indiana, USA, using Landsat data. Impervious surface, residential area and residential impervious surface were regressed with the square root of population density. Lu et al. (2006) concluded that the residential impervious surface model was the best model to estimate the population density. Joseph et al. (2012) applied a multivariate regression model using three urban remote-sensing variables: mean of impervious surface, mean of vegetation and standard deviation of vegetation.

A potential problem with the most commonly used statistical modelling approaches (bivariate and multivariate regression models) is that the regression coefficients of the dependent and explanatory variables are often assumed to be stationary over space (Langford, 2006). There are some reasons to assume that the relationship between the response and explanatory variables might be varied across the study area. One reason is that it is difficult to get a land-use/cover map with 100% accuracy. It is possible to have some omission and commission errors between the land-use/cover classes and this would be a source of spatial data variability (Lo, 2008). For example, in some parts of the study area, a residential built area could be classified as bare land, and any inhabitants who have lived in those areas would have been missed. Another reason is that the relationship between the variables examined could be fundamentally different over space as spatial variation can be recognized in human behaviour, administrative boundaries or environmental factors that lead to different processes over space (Fotheringham, Brunsdon, & Charlton, 2002; Su, Foody, & Cheng, 2012). For example, the modifiable areal unit problem indicates that the results of spatial data analysis could be affected by modifying the areal boundaries of data aggregation. For this reason, it is not clear to what extent the results obtained (for example, dwelling density) are an artefact of the particular areal boundaries used and whether different results would have been obtained using different boundaries superimposed on the same data.

Some researches (Langford, 2006; Lo, 2008; Yuan, Smith, & Limp, 1997) addressed this problem by subdividing the study area into smaller parts (based on administrative units) and then applying a regression model to each part. Langford (2006) indicated an important point about how the part size should be before applying the regression model. Other researches (Dmowska & Stepinski, 2014; Jia & Gaughan, 2016; Ural et al., 2011) applied dasymetric mapping to account the spatial non-stationarity.

Since the mid-2000s, a geographically weighted regression (GWR) has been applied in human geography. The GWR technique takes into account the spatially varying relationship between the investigated variables (Fotheringham et al., 2002). Mennis (2006) mapped the result of the GWR in Philadelphia, USA, at the census tract level using census data and concluded that the variation of the relationship between the variables was accurately explained by the GWR approach compared with the ordinary least squares (OLS) approach. A detailed study was undertaken by Lo (2008) who compared the use of one-class global, four-class global, four-class regional and four-class GWR regression models to estimate population at census tracts in Atlanta, USA, by utilizing Landsat TM data. Lo (2008) concluded that the four-class regional OLS model was more accurate than the one-class and the four-class global OLS models. Moreover, the four-class GWR model produced much more accurate results compared with all the other models (Lo, 2008). Landsat imagery and height data were used to estimate population in Riyadh by using regression through the origin (Alahmadi et al., 2013) and dasymetric mapping (Alahmadi et al., 2016). Alahmadi et al. (2013, 2016) concluded that the remotely sensed height data were useful to accurately estimate dwelling units.

Dong et al. (2010) extracted building counts, areas and volumes from Landsat TM imagery, remotely sensed height data and parcel boundaries to estimate population at the census block in the eastern part of Denton, USA. In that study, the results of the GWR models were more accurate than those of the OLS models. Moreover, the results indicated that areas with high-density population usually were underestimated (Dong et al., 2010). Joseph et al. (2012) derived houses and vegetation from Landsat ETM+ to estimate population at “Section d’énNumération” in Port-au-Prince, Haiti. In their study, the GWR models estimated population better than the OLS models (Joseph et al., 2012).

Mulatu, Van Der Veen, Becht, Van Oel, and Bekalo (2013) estimated population at the sub-location level in Lake Naivasha Basin, Kenya, using general land-cover classes obtained from Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data. In that attempt, the GWR model was more accurate than
the OLS model (Mulatu et al., 2013). Bagan and Yamagata (2015) estimated population density at a coarse spatial resolution (1 km) using general land-cover classes and concluded that the GWR model outperformed the OLS model.

From the above-mentioned studies, Landsat imagery was commonly used to estimate population. The spatial resolution of Landsat imagery is a source of limitation because it is difficult to demarcate urban objects empirically (e.g. villas) that are much smaller than the pixel size, while this would not be the case when the urban objects (e.g. built area) are relatively large compared with the pixel size of the Landsats imagery. However, objects such as built area may be classified inaccurately due to the presence of mixed pixels in the Landsat imagery. Most studies used urban remote-sensing variables such as impervious surface (Li & Lu, 2016; Lu et al., 2006; Patel et al., 2015), built area (Dong et al., 2010; Langford, 2006), residential built area (Lu et al., 2010, 2006; Silvan-Cardenas et al., 2010) and two-residential classes (Deng et al., 2010; Langford, 2006; Silvan-Cardenas et al., 2010). It is not realistic to deal with the city as homogenous. Transition can arise in building structure, dwelling density, family size, human attitude, landscape design and economic activity within the city blocks (Bhaduri et al., 2007; Cadenasso, Pickett, & Schwarz, 2007). Few studies (Lo, 2008; Mennis, 2003; Mulatu et al., 2013) gave attention to the city heterogeneity in terms of three-residential classes such as low-density, medium-density and high-density urban uses.

This paper examines the city heterogeneity in detail by (1) using various land-use/cover categories obtained from fine-spatial-resolution imagery and (2) exploring the application of different statistical modelling approaches for dwelling estimation. This research examines different models ranging from simple to complex models and demonstrating the amount of change in accuracy between them.

This research aims to analyse the spatial non-stationarity in the relationship between dwelling unit density and urban remote-sensing information for small-area population estimates. The study area and data are presented in the second section, while the methods are explained in the third section, covering object-based image analysis (OBIA), OLS, GWR and accuracy assessment. The results of image analysis, OLS models and GWR models are clarified in the fourth section. Finally, discussion and conclusions are provided in the fifth and sixth sections, respectively.

**Study area and data**

**Study area**

Riyadh is the capital and most populated city of Saudi Arabia. According to the census data, population of Riyadh was 2.6 million in 1992, 3.7 million in 2004, 5.2 million in 2010 and 6.5 million in 2016. This rapid increase needs timely and accurate approaches of population estimation to support urban and regional planning.

Riyadh city is heterogeneous in terms of urban structure, dwelling counts, population size, education services and municipal services. Unfortunately, it was not possible to obtain the dwelling unit data for a larger area or more than one ward due to the production cost and other issues. Um Alhamam ward locates in the west part of Riyadh (Figure 1). This ward has a population of 45,991 inhabitants, an area of 4 km², and contains 488 blocks and 2839 parcels. This ward is chosen because it represents various land covers (residential, services, vegetation and industrial) and includes different types of residential units (e.g. low-density, medium-density and high-density areas), thus making it a very good representative ward of Riyadh and a perfect ward to examine the spatial non-stationarity between dwelling unit density and urban remote-sensing variables.

**Data**

Demographic information at the ward level for 2004 was taken from the General Authority of Statistics (GaStat). A ward in Saudi Arabia has an average total population of 27,500 and a mean area of 4 km². Dwelling unit data at the parcel and block levels for 2004 were obtained from Arriyadh Development Authority. A block includes 15–30 parcels that are encircled by the road network. A parcel is the smallest municipal unit at which statistical data are gathered and has an average area of 550 m². QuickBird data acquired on 27 February 2004 were taken from King Abdulaziz City for Science and Technology (KACST). The geometric and radiometric distortions had been corrected by the National Centre for Remote Sensing Technology (NCRST). The coordinate system of the vector and satellite data is Ain El Abd 1970 UTM Zone 38 N.

**Methods**

One aim of this study is to estimate population at the parcel level. However, the GaStat publishes the census population data at a coarser spatial level (ward level) and this prevents validation at the parcel level. Dwelling unit counts are available at the parcel level and logically there is a strong positive correlation among population and dwelling units. Thus, dwelling unit density is used as a dependent variable instead of population density.

In this research, all blocks are utilized in model calibrations and then regression coefficients are used to estimate dwelling unit density at the parcel level.
Count of dwelling units (model validation) is computed from the dwelling unit density. Then, based on the average number of people per dwelling (5.9 in the study area), the small-area population map is produced.

**Object-based image analysis**

A new collection of very-fine-spatial-resolution sensors has been launched since 1999 such as QuickBird (0.65 cm), Pleiades (0.50 cm), GeoEye (0.46 cm) and WorldView-4 (0.31 cm). The development of spaceborne sensor technologies motivates researchers to use the ≤1 m spatial resolution that can minimize the mixed-pixel issue (Lu & Weng, 2009) to map detailed urban structures and land-use information. However, the limitation due to the lower spectral resolution (only four bands) increases the difficulties of distinguishing urban information associated with similar spectral characteristics (Pinho, Fonseca, Korting, De Almeida, & Kux, 2012) when using a classical pixel-based approach. Fine-spatial-resolution satellite data contain urban information with heterogeneous spectral characteristics. Classical pixel-based approaches cannot produce accurate land-cover/use maps (Pu, Landry, & Yu, 2011) because they utilize only spectral characteristics and thus produce a salt-and-pepper effect (Guan et al., 2013; Hamedianfar, Shafri, Mansor, & Ahmad, 2014; Qin, Niu, Chen, Li, & Ban, 2013).

To overcome the above-mentioned problem, region- or object-based approach is used instead of the pixel-based approach (Blaschke, 2010; Pu et al., 2011). The object-based approach can provide accurate results with fine-spatial-resolution satellite data. For example, (1) image pixels can be grouped into objects in a similar way to human visual interpretation of the landscape; (2) land-use classes are obtained based on both spectral and spatial characteristics; and (3) different object levels (i.e. different scales) are incorporated to inherit classification between the different object levels and extract a certain land-use class (Kux & Araújo, 2008).

**OLS regression**

An OLS model has been utilized to explore the relationship between a dependent variable (dwelling unit density) and an independent variable (urban remote-sensing information). The OLS model estimates coefficients of the examined variables based on the best-fitting line of all observations (Equation (1)). The OLS coefficients are assumed to be stationary over space.
\[ Y_i = \beta_0 + \sum_{j=1}^{i-1} \beta_j X_{ij} + \epsilon_i, \quad (1) \]

where \( Y_i \) is the dependent variable for observation \( i \) \((i = 1, 2, \ldots, m)\), \( \beta_0 \) denotes the constant, \( \beta_j \) is the coefficient for variable \( j \), \( X_{ij} \) is the independent variable for the \( j \)th variable at \( i \)th observation, \( n \) denotes the number of independent variables and \( \epsilon_i \) is the residual term.

**Geographically weighted regression**

The GWR approach is built on the conventional regression model by allowing the relationship between variables to be local (spatial non-stationarity) rather than global. Thus, the model is expressed as:

\[ Y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{n} \beta_j(u_i, v_i) X_{ij} + \epsilon_i, \quad (2) \]

where \((u_i, v_i)\) represents the coordinate location for observation \( i \) \((i = 1, 2, \ldots, m)\). The idea of the GWR is that nearby observations have larger weights than distant observations. The GWR result is influenced by the bandwidth of the spatial weighting function chosen and is relatively unaffected by the choice of weighting function. Although the weighting function can be drawn in several forms, there are two common types: fixed and adaptive. The former is suitable in areas where observations are regularly distributed in the study region, whereas the latter is preferred when the observations are irregularly distributed and the sample points have a clustered pattern in the study region (Fotheringham et al., 2002). More detailed explanation of the GWR approach is found in Fotheringham et al. (2002).

**Accuracy assessment**

The block level is used to calibrate models, whereas the parcel level is used to validate models. Three different measurements are used for accuracy assessment, including overall relative error (ORE), mean absolute error (MAE) and root mean square error (RMSE).

\[ ORE = \frac{(EDW - DW)}{DW} \times 100. \quad (3) \]

\[ MAE = \frac{\sum_{i=1}^{TZ} |EDW_{iz} - DDW_{iz}|}{TZ}. \quad (4) \]

\[ RMSE = \sqrt{\frac{\sum_{i=1}^{TZ} (EDW_{iz} - DW_{iz})^2}{TZ}}. \quad (5) \]

where \( EDW \) denotes the overall estimation of dwelling unit, \( DW \) is the overall truth of dwelling unit, \( EDW_{iz} \) and \( DW_{iz} \) refer to the estimation and truth information of dwelling unit for target zones (i.e. parcels) \( tz \), respectively, and \( TZ \) is the overall number of target zones. Smaller values of the accuracy assessment measurements suggest a great model performance.

**Results**

**Dependent and explanatory variables**

Square root of dwelling unit density is utilized as the dependent variable (computed as dwelling counts/area of zones). The proportions of urban remote-sensing variables in each zone are used as explanatory variables (computed as \( K \) extent of the urban remote-sensing variable/overall extent of the urban remote-sensing variables). For example, suppose a total area of a source zone is 100 m\(^2\) and it comprises of 40 m\(^2\) bare land and 60 m\(^2\) residential built area. Now, the residential built area proportion will be computed in that zone as \( 60/100 = 0.6 \).

**Land-use/cover information**

The eCognition software (v.8.8) is used to analyse the QuickBird imagery. The implementation of the OBIA requires three essential phases: (1) image segmentation, (2) feature selection and (3) classification algorithms. Image segmentation is defined as a function that groups image pixels into homogenous regions or objects (Ryherd & Woodcock, 1996). In terms of the first phase, multiple layers of image segmentation are created to allow hierarchical processing (Pinho et al., 2012). Layer-1 and Layer-2 image segmentations are dependent on the block and parcel (Erol & Akdeniz, 2005) boundaries, respectively, by adopting a chessboard algorithm (a top-down algorithm). The most important factor at Layer-1 and Layer-2 is defining the vector boundaries. Layer-3 image segmentation is dependent on the image channels (Lyons, Phinn, & Roelfsema, 2012) using a multi-resolution segmentation (a bottom-up algorithm). Shape and scale factors are essential to control the diversity of the image objects. The scale factor influences the size of the image object, and a value of 30 (Walker & Blaschke, 2008) is applied to obtain smaller image objects within Layer-2 (a segmentation based on the parcel boundaries). The shape factor is determined by giving a weight to the colour and shape parameters. In this article, the colour is emphasized (0.9) compared with the shape (0.1) to obtain meaningful image objects (Walker & Blaschke, 2008; Johnson, 2013). During the second phase, many attribute features (mean, saturation, hue, area, compactness and texture) are utilized to demarcate the detailed land-use/
cover classes at the different layers. The transferability is taken into account when writing the rules through two approaches. The first approach is that the rule must be transferable in theory. For example, a rule “distance to” is based on the first law of geography that near things are more strongly related than distant things (Tobler, 1970), and this is mostly more transferable than a rule that is based on the spectral values. The second approach is based on the influence (omission and commission errors) of the rule. For example, if the rule just corrects few urban areas, or it affects other urban areas, which are correctly classified, the rule is ignored. In the third phase, three classification algorithms are used: (1) knowledge-based expert system (Hayes-Roth, Waterman, & Lenat, 1983), (2) nearest neighbour (Hubert-Moy, Cotanec, Le Du, Chardin, & Perez, 2001) and (3) classification and regression tress (Breiman, Friedman, Olshen, & Stone, 1984). The general implementation of the OBIA can be summarized in four steps.

(1) At Layer-3, general land-cover classes are obtained based on the nearest neighbour algorithm and knowledge-based expert system.

(2) At Layer-1, detailed land-use classes are extracted based on the classification and regression tree (CART) algorithm and enhancement is done based on the knowledge-based expert system.

(3) At Layer-2, detailed land-use classes are inherited from Layer-1 and enhancement is done based on (i) the knowledge-based expert system at the same layer (Layer-2) and (ii) the hierarchical process using Layer-3.

(4) At Final Layer-3, the results from Layer-2 and Layer-3 are combined.

The land-use map at Layer-3 (Figure 2) is assessed by utilizing 50 random image objects for each class. Table 1 defines the land-use classes and that represents the situation of the urban structure in the Arab Gulf countries such as Riyadh. Table 2 shows the error matrix of the land-use classes. The overall accuracy and Kappa statistics are 93% and 0.92, respectively. The producer’s and user’s accuracies of the land-use classes indicate that most of the residential categories are accurately classified. However, there is confusion when discriminating low-density and medium-density villas due to the similarity in terms of the urban structures.

**Model-1-OLS: traditional model (built area)**

The different land-use categories (Figure 2) are processed by different routes to achieve the purpose of this article. The built area class is obtained by combining the residential and non-residential built areas and used as an independent variable in Model-1. The block level (488 blocks) is utilized to provide the data for the calibration, whereas estimation and assessment are performed at the parcel level (2839 parcels). The correlation coefficient between dwelling unit density and built area proportion is 0.6. The dependent variable is transformed (Joseph et al., 2012) using a square root transformation to maximize linearity of model calibrations. The correlation coefficient is increased when the dependent variable is transformed. Figure 3(a) shows a large positive correlation ($r = 0.73$) between the square root of dwelling unit density and built area proportion. Model-1 is estimated by:

$$\text{SQRT DUD} = 141.9B - 19.5, \quad R^2 = 0.53,$$

where SQRT DUD indicates the square root of dwelling unit density and $B$ denotes the built area.
where $B_{res}$ denotes the residential built area proportion. The correlation between the square root of dwelling unit density and the residential built area proportion (Figure 3(b)) is improved ($r = 0.79$) as well as the other accuracy measurements compared with Model-1.

**Model-3-OLS: three-residential classes**

Due to the complexity of urban structure, variation of dwelling densities (Liu & Herold, 2007) is existent within the residential built area. To address this issue, the six-residential classes are combined to produce (1) low-density, (2) medium-density and (3) high-density areas. Model-3 is estimated by:

$$\text{SQRT DUD} = 27.65 B_{Le} + 57.34 B_{Me} + 144.46 B_{He} + 4.22 \, \text{adj} R^2 = 0.85,$$

where $B_{Le}$, $B_{Me}$ and $B_{He}$ denote the proportions of the low-density, medium-density and high-density areas, respectively. All the accuracies are increased in Model-3 compared with Model-2.

**Model-4-OLS: six-residential classes**

Model-3 provides more accurate results than Model-2. However, it is believed that the urban heterogeneity is still an issue. Thus, it is useful to consider whether there is a variation in dwelling unit densities within each class in Model-3. Model-4 is an attempt to utilize six-residential classes.

$$\text{SQRT DUD} = 25.4 B_1 + 33.86 B_2 + 49.53 B_3 + 61.26 B_4 + 145.62 B_5 + 144.78 B_6 + 3.42 \, \text{adj} R^2 = 0.85,$$

where $B_1$, $B_2$, $B_3$, $B_4$ and $B_5$ indicate the proportions of the extensive residence, large villa, classic villa, low-density villa, medium-density villa and residential blocks of flats, respectively. However, the use of six-residential classes does not improve the accuracy of the result. That means heterogeneity is not totally related to the urban areas, and it might be related to other socio-economic variables.

**GWR models**

All the previous models (Models 1, 2, 3 and 4) are replicated using the GWR approach. Thus, this allows a relevant comparison between and within each approach. The developed GWR models are:

- Model-5-GWR utilizing built area.
- Model-6-GWR utilizing residential built area.
- Model-7-GWR utilizing three-residential classes.
• Model-8-GWR utilizing six-residential classes.

The GWR tool in ArcGIS is used to run the models. The adaptive kernel is chosen because the observations are not regularly distributed in Um Alhamam ward and the Akaike information criterion (AICc) is used to determine the bandwidth. Table 4 shows the diagnostics of the GWR models. The $R^2$ values indicate that the square root of dwelling unit density is strongly correlated with the three-residential classes ($\text{adj}R^2 = 0.89$) in Model-7 than the built area ($R^2 = 0.79$) in Model-5 and the residential built area ($R^2 = 0.83$) in Model-6.

Figure 4 shows the GWR parameters for Model-5, Model-6, Model-7 and Model-8. The GWR coefficients suggest that the relationship between the square root of dwelling unit density and the urban remote-sensing variables in Model-5, Model-6, Model-7 and Model-8 vary across the study area. As expected, the results of Model-7 and Model-8 are more accurate than those of Model-5 and Model-6. For example, the RMSEs in Model-7 and Model-8 are 2.0 and 2.1, respectively, whereas they are 3.8 and 2.7 in Model-5 and Model-6, respectively. On the other hand, it is surprising to note that the 3-class GWR model (Model-7) is more accurate than the 4-class GWR model (Model-8).

The conventional model (Model-1), the best OLS model (Model-3) and the best GWR model (Model-7) are selected to map the residual dwelling estimates and population distribution (Figure 5). Most parcels are well estimated in Model-3 (Figure 5(b)) and Model-7 (Figure 5(c)), whereas overestimation and underestimation are obtained in the low-density and high-density parcels, respectively, in Model-1 (Figure 5(a)).

**Discussion**

This work reveals the efficiency of utilizing fine-spatial-resolution imagery to demarcate a detailed land-use/

---

**Figure 3.** Relationship between SQRT of dwelling unit density and (a) built area proportion and (b) residential built area proportion.

**Table 3.** Results of the OLS and GWR models at the parcel level.

| Models                  | DW   | EDW  | ORE (%) | MAE  | RMSE | P     | EP     |
|-------------------------|------|------|---------|------|------|-------|--------|
| Model-1-OLS             | 7771 | 9753 | 26      | 2.4  | 4.8  | 45,991| 57,512 |
| Built area              |      |      |         |      |      |       |        |
| Model-2-OLS             | 7771 | 9284 | 19      | 2.4  | 4.4  | 45,991| 54,709 |
| Residential built area  |      |      |         |      |      |       |        |
| Model-3-OLS             | 7771 | 8401 | 8       | 1.3  | 2.5  | 45,991| 49,541 |
| Three-residential classes| | | | | | | |
| Model-4-OLS             | 7771 | 8739 | 12      | 1.4  | 2.6  | 45,991| 51,521 |
| Six-residential classes |      |      |         |      |      |       |        |
| Model-5-GWR             | 7771 | 9835 | 27      | 1.7  | 3.8  | 45,991| 57,982 |
| Built area              |      |      |         |      |      |       |        |
| Model-6-GWR             | 7771 | 9428 | 21      | 1.5  | 2.7  | 45,991| 55,627 |
| Residential built area  |      |      |         |      |      |       |        |
| Model-7-GWR             | 7771 | 8566 | 10      | 1.2  | 2.0  | 45,991| 50,507 |
| Three-residential classes| | | | | | | |
| Model-8-GWR             | 7771 | 8753 | 13      | 1.3  | 2.1  | 45,991| 51,638 |
| Six-residential classes |      |      |         |      |      |       |        |

DW: true number of dwellings; EDW: estimated number of dwellings; ORE: overall relative error; MAE: mean absolute error; RMSE: root mean square error; P: truth number of population; EP: estimated number of population.

**Table 4.** Diagnostics of the GWR models.

| Models                  | NG   | EN   | Sigma | AICc  | adjR² |
|-------------------------|------|------|-------|-------|-------|
| Model-5-GWR: built area | 58   | 52.91| 30.46 | 4365  | 0.79  |
| Model-6-GWR: residential built area | 56 | 54.34| 27.40 | 4271  | 0.83  |
| Model-7-GWR: three-residential classes | 152 | 31.86| 20.13 | 3983  | 0.89  |
| Model-8-GWR: six-residential classes | 202 | 30.01| 20.28 | 3990  | 0.89  |

NG: neighbours; EN: effective number; AICc: Akaike information criterion.
cover map similarly to other works such as Silvancardenas et al. (2010) and Pu et al. (2011). However, this work is distinct because it demarcates six-residential classes and then manipulated them to produce a variety of urban remote-sensing variables. OBIA is used to analyse the QuickBird data. The accuracy measurements of the land-use map are greater than the common recommendation accuracies. However, it is challenging to discriminate classes that are similar in terms of the spectral and spatial properties such as low-density and medium-density villas. This challenge could be taken into account by using fine-spectral-resolution imagery or using ancillary urban information such as electricity and water consumption.

Robust models of population estimation require (1) appropriate explanatory variables and (2) accurate land-use/cover data. Model-1 is based on the OLS approach and uses built area proportion as the explanatory variable. This variable includes non-residential uses such as public and commercial buildings. The influence of the non-residential uses in Model-1 is double. First, it decreases the slope value ($\beta$) in model calibration. Second, it distributes inhabitants in places which are not residential in reality (model

Figure 4. GWR coefficients of Model-5, Model-6, Model-7 and Model-8 at the block level.
validation) and this will cause overestimations. Thus, the results of Model-1 are disappointing (ORE = 27%, MAE = 1.7 and RMSE = 3.8). The limitation of Model-1 (the non-residential uses) has been addressed in Model-2 by using residential built area as the explanatory variable. It is clear that excluding the non-residential uses does improve the model calibration in Model-2 and this is proved by the $R^2$ value (0.62) as well as the model accuracies (ORE = 21%, MAE = 1.5 and RMSE = 2.7) compared with Model-1.

Model-1 and Model-2 assume that the relationship between the independent variables (built area and residential built area, respectively) and the square root of dwelling unit density is stationary over space. This is not the case in the study area because the study area comprises different types of residential housing. Thus, Model-1 and Model-2 fail to address the spatial non-stationarity. The limitation of Model-1 and Model-2 has been taken into account in Model-3 by using three explanatory variables (low-density, medium-density and high-density areas). The results of Model-3 are more accurate than those of Model-1 and Model-2. For example, the ORE and RMSE results improve to 8% and 2.5 in Model-3, respectively, compared with the results of Model-1 and Model-2. The different coefficients of the independent variables in Model-3

Figure 5. Residual dwelling estimates for (a) conventional model (Model-1); (b) best OLS model (Model-3); (c) best GWR model (Model-7) and population transformations of (d) Model-1; (e) Model-3; (f) Model-7 at the parcel level.
Model-7 aims to reduce the limitation of Model-6 by utilizing three residential classes. The explanatory variables in Model-7 show different coefficients across the study area. For example, the local coefficients for the low-density area (Figure 4) (Model-7 low-density areas) show high positive values in the east part where low-density area is concentrated and low negative values in the middle where high-density area is distributed. The AICc (3983) and the $R^2$ (0.89) in Model-7 compared with Model-6 suggest that the calibration of Model-7 is better than Model-6. Moreover, the accuracies of Model-7 are more than Model-6. For example, the ORE, MAE and RMSE are decreased to 10%, 1.2 and 2.0, respectively, in Model-7 compared with Model-6.

As a result of complexity of urban systems, the six-residential class (Model-8) is examined. The results of Model-8 are more accurate than those of Model-5 and Model-6 but not than of Model-7. That means advanced models do not always give accurate results. It is important to mention that the results of Model-3 (the OLS approach) are similar to those of Model-7 (the GWR approach). This assumes that the spatial non-stationarity can be successfully addressed by using representative and appropriate urban covariates.

It should be noted that Um Alhamam ward is used to develop the OLS and GWR models. This indicates that all the regression coefficients are specifically relevant to the urban characteristics of that ward and they are not transferable to other sites. However, the modelling process (i.e. OBIA), findings and conclusions will be applicable to other sites. The population maps (Figure 5(d–f)) are a final conversion of the dwelling estimates. This means that the variation in terms of the family size between the different urban covariates is not addressed. This limitation could be overcome by additional information such as utility usage because a large family is likely to consume more electricity than a small family.

**Conclusion**

Small-area population distributions are essential for different applications such as national resource management and urban-sprawl analysis. This research analyses the spatial non-stationarity in the relationship between dwelling unit density and land-use/cover map to estimate population at the parcel level in Um Alhamam ward. General and detailed land-use/cover classes are obtained from QuickBird satellite data using the OBIA method. Most of the land-use/cover classes are correctly classified. Two approaches are applied to estimate dwelling density at the parcel level: OLS and GWR.

The OLS approach assumes constant regression coefficients across the study area, thus failing to deal with the
spatial non-stationarity. Model-1 uses built area which is not suitable to describe the variation of the square root of dwelling unit density, and thus, unsatisfactory results are expected. A refined explanatory variable such as residential built area (Model-2) does improve the accuracy, but the spatial non-stationarity is still an issue. The three-residential classes (Model-3) account the spatial non-stationarity, and thus, the results are improved. The complex model (Model-4) does not improve the accuracies compared with the representative model (Model-3).

Generally, most of the GWR models are more accurate than the OLS models because the GWR approach does account the spatial non-stationarity unlike the OLS approach. However, the GWR approach is not enough to completely address the spatial non-stationarity. The explanatory variables are also important likewise the model construction. Thus, the application of the GWR approach in Model-7 (three-residential classes) provides more accurate results compared with Model-6 (residential built area) and Model-5 (built area).

Acknowledgements
The author thanks King Abdulaziz City for Science and Technology (KACST) and High Commission for the Development of Arriyadh (HCDA) for provision of data.

Disclosure statement
No potential conflict of interest was reported by the author.

Funding
This research is supported by King Abdulaziz City for Science and Technology (KACST).

ORCID
Mohammed Alahmadi ∗ http://orcid.org/0000-0003-4739-4855

References
Alahmadi, M., Atkinson, P., & Martin, D. (2013). Estimating the spatial distribution of the population of Riyadh, Saudi Arabia using remotely sensed built land cover and height data. Computers, Environment and Urban Systems, 41, 167–176.

Alahmadi, M., Atkinson, P., & Martin, D. (2015). Fine spatial resolution residential land-use data for small-area population mapping: A case study in Riyadh, Saudi Arabia. International Journal of Remote Sensing, 36(17), 4315–4331.

Alahmadi, M., Atkinson, P.M., & Martin, D. (2016). A comparison of small-area population estimation techniques using built-area and height data, Riyadh, Saudi Arabia. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(5), 1959–1969.

Anderson, D.E., & Anderson, P.N. (1973). Population estimates by humans and machines. Photogrammetric Engineering, 39(2), 147–154.

Bagan, H., & Yamagata, Y. (2015). Analysis of urban growth and estimating population density using satellite images of nighttime lights and land-use and population data. GIScience & Remote Sensing, 52(6), 765–780.

Bhaduri, B., Bright, E., Coleman, P., & Urban, M.L. (2007). LandScan USA: A high-resolution geospatial and temporal modeling approach for population distribution and dynamics. Geoforum, 69(1–2), 103–117.

Blaschke, T. (2010). Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing, 65(1), 2–16.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and regression trees. New York: Wadsworth International Group.

Cadenasso, M.L., Pickett, S.T.A., & Schwarz, K. (2007). Spatial heterogeneity in urban ecosystems: Reconceptualizing land cover and a framework for classification. Frontiers in Ecology and the Environment, 5(2), 80–88.

Deng, C., Wu, C., & Wang, L. (2010). Improving the housing-unit method for small-area population estimation using remote-sensing and GIS information. International Journal of Remote Sensing, 31(21), 5673–5688.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F.R., Gaughan, A.E., . . . Tatem, A.J. (2014). Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences, 111(45), 15888–15893.

Dmowska, A., & Stepinski, T.F. (2014). High resolution dasymetric model of U.S demographics with application to spatial distribution of racial diversity. Applied Geography, 53(Supplement C), 417–426.

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

Dong, P., Ramesh, S., & Nepali, A. (2010). Evaluation of small-area population estimation using LiDAR, Landsat TM and parcel data. International Journal of Remote Sensing, 31(21), 5571–5586.

Erol, H., & Akdeniz, F. (2005). A per field classification method based on mixture distribution models and an application to Landsat thematic mapper data. International Journal of Remote Sensing, 26(6), 1229–1244.

Fotheringham, A., Brunsdon, C., & Charlton, M. (2002). Geographically weighted regression: The analysis of spatially varying relationships. Chichester: John Wiley & Sons Inc.

Guan, H., Li, J., Chapman, M., Deng, F., Ji, Z., & Yang, X. (2013). Integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests. International Journal of Remote Sensing, 34(14), 5166–5186.

Hamedianfar, A., Shafri, H.Z.M., Mansor, S., & Ahmad, N. (2014). Improving detailed rule-based feature extraction of urban areas from worldview-2 image and lidar data. International Journal of Remote Sensing, 35(5), 1876–1899.

Hayes-Roth, F., Waterman, D., & Lenat, D. (1983). Building expert systems (1st ed.). Reading, MA: Addison Wesley.

Hostert, P. (2007). Advances in urban remote sensing: Examples from Berlin (Germany). In M. Netzband, W. L. Stefanov, & C. Redman (Eds.), Applied remote sensing for urban planning, governance and sustainability (pp. 37–51). Heidelberg: Springer.
Hsu, S.-Y. (1971). Population estimation. *Photogrammetric Engineering, 37*(5), 449–454.

Hubert-Moy, L., Cotanenc, A., Le Du, L., Chardin, A., & Perez, P. (2001). A comparison of parametric classification procedures of remotely sensed data applied on different landscape units. *Remote Sensing of Environment, 75*(2), 174–187.

Jia, P., & Gaughan, A.E. (2016). Dasyometric modeling: A hybrid approach using land cover and tax parcel data for mapping population in Alachua County, Florida. *Applied Geography, 66*(Supplement C), 100–108.

Johnson, B.A. (2013). High-resolution urban land-cover classification using a competitive multi-scale object-based approach. *Remote Sensing Letters, 4*(2), 131–140.

Joseph, M., Wang, L., & Wang, F. (2012). Using Landsat imagery and census data for urban population density modeling in Port-au-Prince, Haiti. *GIScience & Remote Sensing, 49*(2), 228–250.

Kux, H.J.H., & Araújo, E.H.G. (2008). Object-based image analysis using QuickBird satellite images and GIS data, case study Belo Horizonte (Brazil). In T. Blaschke, S. Lang, & G.J. Hay (Eds.), *Spatial concepts for knowledge-driven remote sensing* (pp. 571–588). Berlin, Heidelberg: Springer Berlin Heidelberg.

Langford, M. (2006). Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. *Computers, Environment and Urban Systems, 30*(2), 161–180.

Li, L., & Lu, D. (2016). Mapping population density distribution at multiple scales in Zhejiang province using Landsat thematic mapper and census data. *International Journal of Remote Sensing, 37*(18), 4243–4260.

Liu, X., & Herold, M. (2007). Population estimation and interpolation using remote sensing. In Q. Weng & D. Quattrochi (Eds.), *Urban remote sensing* (pp. 269–290). Boca Raton: Taylor & Francis Group.

Lo, C. (2008). Population estimation using geographically weighted regression. *GIScience & Remote Sensing, 45*(2), 131–148.

Lo, C., & Welch, R. (1977). Chinese urban population estimates. *Annals of the Association of American Geographers, 67*(2), 246–253.

Lu, D., & Weng, Q. (2009). Extraction of urban impervious surfaces from an IKONOS image. *International Journal of Remote Sensing, 30*(5), 1297–1311.

Lu, D., Weng, Q., & Li, G. (2007). Residential population estimation using a remote sensing derived impervious surface approach. *International Journal of Remote Sensing, 27*(16), 3553–3570.

Lu, Z., Im, J., Quackenbush, L., & Halligan, K. (2010). Population estimation based on multi-sensor data fusion. *International Journal of Remote Sensing, 31*(21), 5587–5604.

Lyons, M.B., Phinn, S.R., & Roelfsema, C.M. (2012). Long term land cover and seagrass mapping using Landsat and object-based image analysis from 1972 to 2010 in the coastal environment of South East Queensland, Australia. *ISPRS Journal of Photogrammetry and Remote Sensing, 71*, 34–46.

Martin, D., & Williams, H. (1992). Market-area analysis and accessibility to primary health-care centres. *Environment and Planning A, 24*(7), 1009–1019.

Mennis, J. (2003). Generating surface models of population using dasymetric mapping. *The Professional Geographer, 55*(1), 31–42.

Mennis, J. (2006). Mapping the results of geographically weighted regression. *The Cartographic Journal, 43*(2), 171–179.

Mulatu, D.W., van der Veen, A., Becht, R., van Oel, P.R., & Bekalo, D.J. (2013). Accounting for spatial non-stationarity to estimate population distribution using land use/cover. Case study: The lake Naivasha basin, Kenya. *Journal of Settlements and Spatial Planning, 4*(1), 33–44.

Patel, N.N., Angiulli, E., Gamba, P., Gaughan, A., Lisini, G., Stevens, F.R., . . . Trianni, G. (2015). Multitemporal settlement and population mapping from Landsat using Google Earth Engine. *International Journal of Applied Earth Observation and Geoinformation, 35*, 199–208.

Patel, N.N., Stevens, F.R., Huang, Z., Gaughan, A.E., Elyazar, I., & Tatem, A.J. (2017). Improving large area population mapping using geotweet densities. *Transactions in GIS, 21*(2), 317–331.

Pattnaik, S., Mohan, S., & Tom, V. (1998). Urban bus transit route network design using genetic algorithm. *Journal of Transportation Engineering, 124*(4), 368–375.

Pinho, C.M.D., Fonseca, L.M.G., Korting, T.S., De Almeida, C.M., & Kux, H.J.H. (2012). Land-cover classification of an intra-urban environment using high-resolution images and object-based image analysis. *International Journal of Remote Sensing, 33*(19), 5973–5995.

Porter, P.W. (1956). Population distribution and land use in Liberia (PhD). London School of Economics and Political Science, London, UK.

Pu, R., Landry, S., & Yu, Q. (2011). Object-based urban detailed land cover classification with high spatial resolution IKONOS imagery. *International Journal of Remote Sensing, 32*(12), 3285–3308.

Qin, Y., Niu, Z., Chen, F., Li, B., & Ban, Y. (2013). Object-based land cover change detection for cross-sensor images. *International Journal of Remote Sensing, 34*(19), 6723–6737.

Rees, P., Norman, P., & Brown, D. (2004). A framework for progressively improving small area population estimates. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 167*(1), 5–36.

Ryherd, S., & Woodcock, C. (1996). Combining spectral and texture data in the segmentation of remotely sensed images. *Photogrammetric Engineering & Remote Sensing, 62*(2), 181–194.

Silvan-Cardenas, J.L., Wang, L., Rogerson, P., Wu, C., Feng, T., & Kamphaus, B.D. (2010). Assessing fine-spatial-resolution remote sensing for small-area population estimation. *International Journal of Remote Sensing, 31*(21), 5605–5634.

Su, Y.-F., Foody, G.M., & Cheng, K.-S. (2012). Spatial non-stationarity in the relationships between land cover and surface temperature in an urban heat island and its impacts on thermally sensitive populations. *Landscape and Urban Planning, 107*(2), 172–180.

Tatem, A.J., Noor, A.M., & Hay, S.I. (2005). Assessing the accuracy of satellite derived global and national urban maps in Kenya. *Remote Sensing of Environment, 96*(1), 87–97.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*, 234–240.
Tobler, W.R. (1969). Satellite confirmation of settlement size coefficients. Area, 1(3), 30–34.
Tomás, L., Fonseca, L., Almeida, C., Leonardi, F., & Pereira, M. (2016). Urban population estimation based on residential buildings volume using IKONOS-2 images and lidar data. International Journal of Remote Sensing, 37(1), 1–28.
Tsou, M.-H. (2015). Research challenges and opportunities in mapping social media and big data. Cartography and Geographic Information Science, 42 (sup1), 70–74.
Ural, S., Hussain, E., & Shan, J. (2011). Building population mapping with aerial imagery and GIS data. International Journal of Applied Earth Observation and Geoinformation, 13(6), 841–852.
Walker, J., & Blaschke, T. (2008). Object based land cover classification for the Phoenix metropolitan area: Optimization vs. transportability. International Journal of Remote Sensing, 29(7), 2021–2040.
Wu, S.-S., Qiu, X., & Wang, L. (2005). Population estimation methods in GIS and remote sensing: A review. GIScience & Remote Sensing, 42(1), 80–96.
Yuan, Y., Smith, R., & Limp, W. (1997). Remodeling census population with spatial information from Landsat TM imagery. Computers, Environment and Urban Systems, 21(3–4), 245–258.