

Assessment of spatio-temporal direction of impervious surface area surface temperature in Pretoria, South Africa

Adeniyi Adeyemi^{a,e}, Abel Ramoelo^b, Moses Azong Cho^{c,d} and Jacobus Strydom^e

^aDepartment of Environmental Science, College of Agriculture and Environmental Science, University of South Africa, Pretoria, South Africa; ^bCentre for Environmental Studies, Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa; ^cPrecision Agriculture Research Group, Council for Scientific and Industrial Research, Pretoria, South Africa; ^dDepartment of Plant and Soil Sciences, College of Agriculture and Environmental Science, University of Pretoria, Pretoria, South Africa; ^eGeoTerralimage (Pty) Ltd., Agri Hub Business Park, Pretoria, South Africa

ABSTRACT

Over the years, rapid urban growth has led to the conversion of natural lands into large man-made landscapes due to enhanced political and economic growth. This study assessed the spatio-temporal change characteristics of impervious surface area (ISA) expansion using its surface temperature (LST) at selected administrative subplace units (i.e., local region scale). ISA was estimated for 1995, 2005 and 2015 from Landsat-5 Thematic Mapper (TM) and Landsat-8 OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor) images using a Random Forest (RF) algorithm. The spatio-temporal trends of ISA were assessed using an optimal analytical scale to aggregate ISA LST coupled with weighted standard deviational ellipse (SDE) method. The ISA was quantified with high predictive accuracy (i.e., AUROC = 0.8572 for 1995, AUROC = 0.8709 for 2005, AUROC = 0.8949 for 2015) using RF classifier. More than 70% of the selected administrative subplaces in Pretoria experienced an increase in growth rate (415.59%) between 1995 and 2015. LST computations from the Landsat TIRS bands yielded good results (RMSE = ~1.44OC, 1.40OC, ~0.86OC) for 1995, 2005 and 2015 respectively. Based on the hexagon polygon grid (90x90), the aggregated ISA surface temperature weighted SDE analysis results indicated ISA expansion in different directions at the selected administrative subplace units. Our findings can represent useful information for policymakers in evaluating urban development trends in Pretoria, City of Tshwane (COT).

ARTICLE HISTORY

Received 13 September 2021
Accepted 19 December 2021

KEYWORDS

Landsat; ISA; random forest; LST; standard deviational ellipse

1. Introduction

Development is sometimes reflected by the chaotic expansion of urban expansion and the spontaneous appearance of urban buildings in rural areas or on the peripheries of cities (Bouzekri et al. 2015). This rapid urban landscape change as indicated by Odindi et al. (2012) has been because of the exceptional increase in population concentration in cities.

CONTACT Adeniyi Adeyemi  adedayoadeyemi01@gmail.com

This article has been republished with minor changes. These changes do not impact the academic content of the article.

© 2021 Informa UK Limited, trading as Taylor & Francis Group

According to Demographia (2017) and Sahana et al. (2018) 53% of the global population lives in urban settlements and by 2030 this number is expected to have increased more than 60%. Developing countries are more likely to experience the highest rate of urbanization and industrialization in the near future i.e., conversion of vegetation areas to impervious surface materials (Sahana et al. 2018). Over the last decades, Pretoria, South Africa has been facing major land use or cover changes, such as loss of natural land, i.e., forest or plantations, agricultural lands and grasslands coupled with growing impervious surface areas (ISA) such as roads, sidewalks, parking lots, rooftops and bare lands due to a continuous increase in the population (Adeyemi et al. 2015).

In past decades, depending upon the research objectives, many methods have been proposed to extract impervious surfaces using satellite images. However, ISA mapping in major cities in African in the body of literature is still very few. An overview of the methods for ISA mapping can be grouped into four major categories: classification-based (i.e., pixel or object-based), mixture analysis (i.e., sub-pixel-based), spectral index-based and deep learning-based segmentation (Weng 2012; Yu et al. 2017; Tian et al. 2018; Wei and Blaschke 2018; Zhang and Huang 2018; Hua et al. 2020; Adeyemi et al. 2021). Most classification-based methods (i.e., supervised classifiers) require training samples e.g., maximum likelihood classifier (Masek et al. 2000), machine learning classifiers such as artificial neural networks (ANN) (Hu and Weng 2009; van de Voorde et al. 2009), decision tree (DT) (Xian and Crane 2006; Lu et al. 2011; Xu 2013), classification and regression tree (CART) (Xu and Wang 2016), random forest (RF) (Zhang et al. 2014; Adeyemi et al. 2015; Xu et al. 2018), support vector machine (Sun 2011; Okujeni et al. 2015; Shi et al. 2017; Xu et al. 2017) and regression modelling (Okujeni et al. 2018; Yu et al. 2018). Among the above-mentioned pixel-scale, classification-based methods on multispectral imagery, the non-parametric Random Forest (RF) algorithm has been reported to perform excellently in ISA estimation from multispectral imagery (Adeyemi et al. 2015). Nonetheless, the potential and effectiveness of random forest machine learning algorithms based on different training sample sizes in spatio-temporal analysis of urban impervious surfaces in major cities in Africa using remote sensing is still very little and needs to be explored.

Furthermore, the most significant environmental impact of the high degree of imperviousness as documented by many studies is land surface temperature (LST) and atmospheric temperature variability (Deng and Wu 2013; Artmann 2014; McGregor et al. 2015; Morabito et al. 2016; Ward et al. 2016; Tian et al. 2018). With the development of satellite thermal infrared remote sensing data, considerable LST measurements can be retrieved (Nie and Xu 2015). Consequently, there have been several algorithms and methods used for LST retrieval from remote sensing data. Qin et al. (2001) developed the split window and mono window algorithm and demonstrated their effectiveness of using Landsat data. Jiménez-Munoz and Sobrino (2003) and Jiménez-Muñoz and Sobrino (2010) also developed the single-channel algorithm for LST retrieval from Landsat and ASTER data respectively. In this study, the single-channel algorithm was employed due to its advantage of being used when the ground truth data is not available Alipour et al. (2011). Even though, increase in ISA results in surface temperature rise due to change in land surface component distinctive radiative, thermal, moisture and aerodynamic properties according to Owen et al. (1998), there is still limited explicit information using surface temperature as a complementary metric for spatio-temporal urban expansion trend analysis.

An analytical method such as standard deviational ellipse (SDE) (Lefever 1926) has been widely used in recent studies to evaluate the spatial distribution evolution and trends

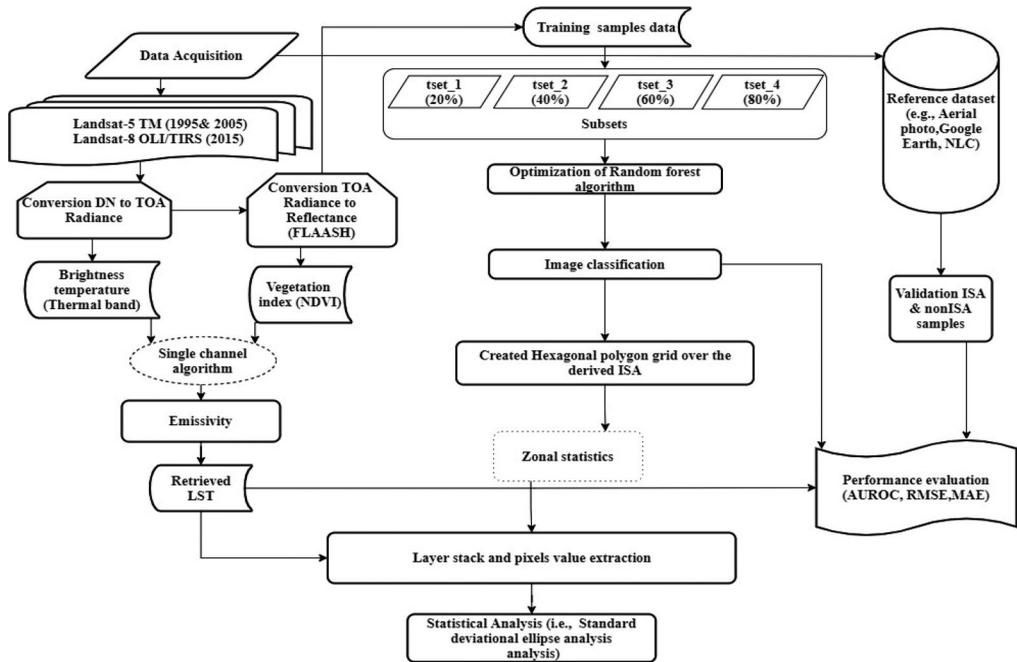


Figure 1. Flowchart depicting the methodological framework.

in various fields (Vanhulsel et al. 2011; Al-Kindi et al. 2017; Li et al. 2017; Xu et al. 2018), because it can reveal the spatial concentration of geographical phenomena and the change characteristics of the geospatial distribution. Recent studies in developed countries have used the SDE to examine the spatio-temporal dynamics of urban expansion over a long-time period by using the impervious surfaces estimated with remote-sensing data (Jian et al. 2016; Qiao et al. 2018; Xu et al. 2018; Man et al. 2019). Nevertheless, sufficient spatio-temporal details may still be required to understand the spatio-temporal urban expansion at different spatial scales coupled with the trends. Since none of this studies have been performed in major cities in Africa, we undertook a study using selected Pretoria administrative subplaces as a pilot area of comprehensive innovation reform. The aim of this study was to improve understanding of the spatio-temporal developing trend of ISA expansion at a local spatial scale based on surface temperature (i.e., a complementary metric) in Pretoria, South Africa during the past 30 years. The key research questions are:

1. Can the random forest algorithm based on different training sample subsets influence the accuracy of estimated ISA from optical Landsat imagery?
2. At local spatial scale, can the spatio-temporal changes of the extracted ISA be revealed within ten years interval time frame?
3. With an optimal analytical scale, is it possible to reveal the principle direction of urban expansion at local region level using the weighted standard deviational ellipse (SDE) method?

2. Materials and methods

The overall methodological workflow is summarized in the flowchart of Figure 1.

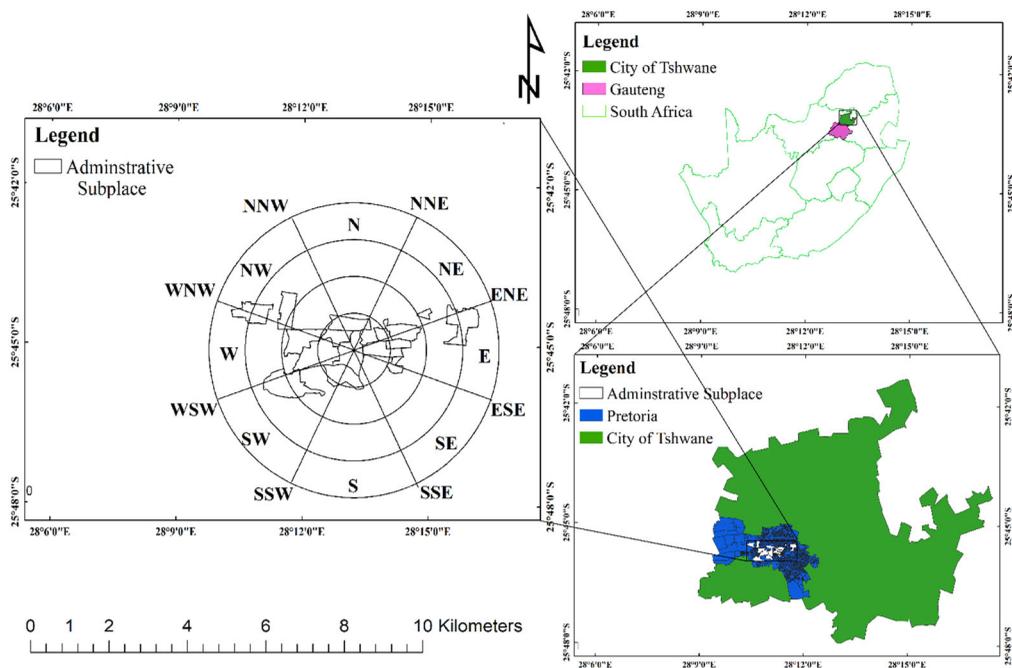


Figure 2. The location of the study area relative to Pretoria, South Africa.

2.1. Study area

The study covers Pretoria (Figure 2) which is a city in the north-northeast of Johannesburg in the northeast of South Africa. It is located between $S25^{\circ}35'29.68''$ $E28^{\circ}0'51.34''$ (top left) and $S25^{\circ}56'50.55''$ $E28^{\circ}28'37.59''$ (bottom right) of the central part of the City of Tshwane Metropolitan Municipality. According to the South African National Census (2011), Pretoria is situated in the Gauteng province with an estimated population of 12.2 million people which is 25% of South Africa total population and indicating it is the fastest-growing province that has witnessed a population grows of over 33% between 1996 and 2011. Pretoria has a topography 1330 m above sea level has an average temperature ranging from 29°C max to 18°C min in January; 19°C max to 5°C in June and precipitation of 674 mm South African Weather Service (2011). Similarly, South African Weather Services (2013) reported that during a nationwide heatwave in November 2011, Pretoria experienced temperatures that reached 39°C which was unusual for that time of the year. The all-time high recorded in Pretoria was 42°C (108°F) on 25 January 2013. The year 2014 saw one of the wettest years on record for the city with precipitation of 914 mm experienced to the end of December.

2.2. Data collection and pre-processing

In this study, three cloud-free springtime images (Table 1) recorded by Landsat 5-TM (Thematic Mapper) on 25th September 1995 at 07:03 h local time, 20th September 2005 at 07:50 h local time and Landsat-8 OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor) on 16th September 2015 at 08:02 h local time were obtained from the United States Geological Survey Global Visualization Viewer and reprojected to the UTM

Table 1. Summary of data used in research.

Data Types	Datasets		
	Date	Image type	Path/Row
Remote Sensing data	1995-09-25	Landsat 5-TM	170/78
	2005-09-20	Landsat 5-TM	170/78
	2015-09-16	Landsat-8 OLI/TIRS	170/78
Boundary file	2011	Pretoria Administrative boundary shape dataset based on Enumeration Area level (i.e., the smallest geographical unit, with typically 100 to 250 households, into which the country is divided for census or survey purposes) and subplace level (i.e., the second (lowest) level of the place name category, namely a suburb, section or zone of an (apartheid) township, smallholdings, village, sub-village, ward or informal settlement.	
Ancillary data	2012-2015	1:10,000 aerial photography from 2012-2015 with spatial resolution of 0.5 m and	
	2015	Google Earth Images	
	2015	NLC classification dataset of 2013-2014	

Zone 35S projection system with WGS84 datum. The Landsat images were pre-processed using two important steps: (1) converting digital numbers (DNs) to top-of-atmosphere (TOA) radiance and then to TOA reflectance; and (2) conversion of the TOA reflectance to surface reflectance using the Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) a first principle atmospheric correction tool that incorporates the standard MODTRAN model in ENVI software (Felde et al. 2003; Krause 2005; Environment for Visualizing Images ENVI 2015 2014). Also, in this study, we used the Pretoria administrative shapefile boundary at Enumeration Area (EA) and subplace (SP) level source from Statistics South Africa (STATSSA).

The dominant land use per EA was extracted by analysing the building-based land use dataset and then rolled up to subplace level. As pilot areas, the following subplaces within Pretoria were selected based on the dominant land use (EA type) i.e., Formal residential (Arcadia, Capital Park, Claremont, Eastwood 2, Philip Nel Park, Riviera, Rietondale), Commercial (Loftus Stadium, Pretoria Central, Pretoria West), Industrial (Kirkney, Koedoespoort, Koedoespoort Industrial, Pretoria Industrial) and Collective living quarters (Salvokop). Other data used in this study are listed in Table 1.

3. Methodology

3.1. Land surface temperature (LST) retrieval

The single-channel algorithm (SCA) developed by Jiménez-Munoz and Sobrino (2003) was employed to retrieve LST from the geometrically corrected TIR band 6 from Landsat 5-TM (10.44 – 12.42 μm) and mean of band 10 and 11 from Landsat-8 TIRS (10.6 – 11.19 μm) and (11.50 – 12.51 μm) respectively.

3.1.1. Conversion to at-satellite brightness temperature (TB)

The thermal bands digital numbers were first calibrated to minimize the noise caused by aerosols, water vapour etc. before being converted to top-of-the-atmosphere (TOA) radiance. The following formulas are used to perform this process (Equation 1):

For Landsat 5TM:

$$R = \left(\frac{L_{\max} - L_{\min}}{Q_{cal\max} - Q_{cal\min}} \right) (Band\ 6_{DN} - Q_{cal\min}) + L_{\min} \quad (1)$$

Where: R is TOA radiance (watts/(meter squared * ster * μm), $Q_{cal\max} = 255$, $Q_{cal\min} = 0$ while L_{\max} and L_{\min} can be obtained from the header file of the Landsat 5TM imagery (Markham and Barker 1985) in Equation 2:

For Landsat 8:

$$R = M_L^* Band10_{DN} + A_L \quad (2)$$

Where, R is TOA radiance (watts/(meter squared * ster * μm), M_L and A_L were also obtained from the header file of the Landsat 8 image (United States Geological Survey 2013). Thereafter, the radiance (R) images of the two Landsat sensors were converted to at-satellite brightness temperature, T_b i.e., blackbody temperature under the assumption of a uniform emissivity in Equation 3:

$$T_b = \frac{K_2}{\ln(K_1/R) + 1} \quad (3)$$

Where, T_b is at-satellite brightness temperature or black body temperature, R is radiance while K_1 (Watts/Meter Squared * Ster * mm) and K_2 (Kelvin) are constants which are 774.89 and 1321.08 respectively. The K_1 and K_2 constant for Landsat sensors are provided in the image header file. As indicated by many authors, T_b is not the true surface temperature due to atmospheric interference and variations in land cover (Weng and Lu 2008; Hu and Jia 2010). In this study, the (Sobrino et al. 2004) single channel atmospheric correction method was used to remove the noise.

3.1.2. Determination of surface emissivity (ϵ)

Firstly we derived the surface emissivity (ϵ) which is commonly carried out by differentiation of NDVI which has an advantage when the researcher has no detailed information on derived land cover in the study area (Carlson and Ripley 1997). Surface emissivity (ϵ) varies with land covers on ground surfaces (Sun et al. 2017). In urban environments, vegetated surfaces have stronger thermal holding capacity and higher cooling effects than non-vegetated areas. The Normalized Difference Vegetation Index (NDVI) was now used to estimate the Proportion of vegetation (P_v) which is to assess the role of vegetation in each pixel of the satellite images (Gutman and Ignatov 1998). The formula below was designed for calculating the NDVI and vegetation proportion (Equation 4 and 5):

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (4)$$

$$P_v = \left(\frac{NDVI - NDVI_{\min}}{NDVI_{\max} - NDVI_{\min}} \right) \quad (5)$$

Where the $NDVI_{\min}$ and $NDVI_{\max}$ were the maximum and minimum values obtained from the derived vegetation index image. (Sobrino et al. 2008) measured the relationships between ϵ and proportion of vegetation (P_v) on a variety of ground surfaces based on the Landsat-extracted NDVI, at each 30 m pixel with the formula established according to (Sobrino et al. 2004) in Equation 6:

Table 2. Summary of the ISA and NonISA classes assigned as training and validation dataset.

Samples	Training (ISA/NonISA)	Validation (ISA/NonISA)	Total (Σ)
1995	183(112/71)	315(163/152)	498
2005	164(88/76)	441(166/275)	605
2015	154(93/61)	273(144/129)	427
			1530

$$\varepsilon = \left\{ \begin{array}{ll} 0.979 - 0.035\text{Red} & NDVI < NDVI_{min} \\ 0.986 + 0.004Pv & NDVI_{min} \leq NDVI \leq NDVI_{max} \\ 0.99 & NDVI > NDVI_{max} \end{array} \right\} \quad (6)$$

Where: ε is the surface emissivity image and Red is the surface reflectance of the red band

3.1.3. Conversion of at-satellite brightness temperature to LST

Finally, the calculated land surface emissivity for each Landsat image was used to convert the brightness temperature image to Land Surface Temperature (LST) using the Planks equation described in Equation 7 (Weng et al. 2004);

$$LST_{(KELVIN)} = \frac{T_b}{1 + (\lambda + T_b/\rho) * \ln \varepsilon} \quad (7)$$

To convert the LST image to Celsius image using the Equation 8:

$$LST_{(CELSIUS)} = LST_{(KELVIN)} - 273.15 \quad (8)$$

Where: λ is the wavelength of radiation emitted in Landsat 5 TM (11.5 μm) (Markham and Barker 1985) and Landsat 8 LDCM (10.8 μm) (United States Geological Survey 2013). $\rho = h * c/\sigma$, σ = Stefan Boltzmann's constant, h = Plank's constant, C = velocity of light, ε = surface emissivity image, LST = surface temperature image. The rescaled to 30 m spatial resolution daytime time retrieved surface temperature (LST) images were also normalized.

3.2. Collection of training and validation samples

Training datasets are significant to understanding the features in real-world and to map a mental picture of the land use or cover type while the validation samples were used for independent validation of the obtained land cover maps (Bhaskaran et al. 2010; Aguilar et al. 2014; Ishimwe et al. 2014). Although the selected Pretoria subplaces comprises of the following dominant land use types formal residential, commercial, industrial and collective living quarters, unfortunately, we were unable to visually inspect with a handheld GPS receiver all ISA samples. As a result, the ISA used for validation samples were manually digitized from the reference data such as 1:10,000 aerial photography (0.5 m), Google Earth Engine (DigitalGlobe). The ISA polygons were uploaded using the ESRI ArcGIS software (Table 2). These exposed ISA created on the multispectral images were used as training and validation samples (obtained from reference datasets) for classification and accuracy assessment. Furthermore, Ramezan et al. (2019) recently detailed that the size and quality of training sample data coupled with sample selection method used can affect the classification and accuracy assessments. Therefore, in this study, we similarly assess the effect of the training sample sizes and the machine learning algorithm performance based on the classification accuracies. We randomly divided the training sample dataset into 4 different imbalanced datasets (i.e., tset_1, tset_2, tset_3, tset_4) with corresponding

sizes 20%, 40%, 60% and 80% of the total training data (Figure 1). The create Data Partition function in the caret package in the R statistical software environment (RStudio, Inc., Boston, MA, USA, Version 1.1.463) software was used to ensure that the number of pixels chosen in each class for every sub-dataset to keep the most consistent size in the imbalanced training sample size.

3.3. Random Forest classification

Random Forests (RF) classifier, developed by Breiman (2001), is an ensemble algorithm developed in the field of machine learning that uses a similar but enhanced method of bagging (bootstrap aggregation) operation (Adelabu et al. 2013; Cracknell and Reading 2014; Adeyemi et al. 2015). According to Loosvelt et al. (2012), RF classifier operates by creating multiple classification trees, each trained on a bootstrapped sample of the original training data. In this technique, the diversity of trees is increased by making them develop from different training data subsets created through bagging (Breiman 1996). The majority of 'votes' from the assemblages of trees built by RF decide the class assignment of a given pixel i.e., The majority 'vote' is used to predict the final class for each unknown (Berhane et al. 2018; Maxwell et al. 2018; Guo et al. 2020). RF classifies the data that is not in the trees as out-of-bag (OOB) data, and the average OOB error rates from all trees give an error rate called the OOB classification error for each input variable i.e., an independent estimate of the overall accuracy of the RF classification (Breiman 2001). Furthermore, to implement the RF according to Breiman (2001), two parameters need to be set up which are the number of trees (ntree) and the number of features in each split (mtry). Regarding the mtry parameter, many studies use the default value mtry such as the number of predictor variables or bands according to (Belgiu and Drăgu 2016; Shrestha et al. 2021) while Feng et al. (2015) stated that with ntree \geq 200, RF could achieve accurate results. Although some studies stated that satisfactory results could be achieved with the default parameters while others indicated that large number of trees will provide a stable result of variable importance (Thanh Noi and Kappas 2017; Shrestha et al. 2021). In addition, RF classifier can determine the "best split" threshold of input values for given classes by implementing the Gini Index, which returns a measure of class heterogeneity within child nodes as compared to the parent node (Waske and Braun 2009). Compared to other non-parametric classifiers, RF has a number of advantages which are: (1) easy to implement as only two parameters (ntree and mtry) need to be optimized (Özçift 2011), (2) can be more reliable than other iterative techniques that do not always consider parameters as independent (Adelabu et al. 2013), (3) insensitive to noise (Watts and Lawrence 2008), does not suffer from over-fitting or a long training time (Loosvelt et al. 2012), faster computation and (4) ability to determine input variable importance by comparing the OOB error rate (Rodriguez-Galiano et al. 2012) and can handle imbalanced data sets (Maxwell et al. 2018).

3.4. Optimization for impervious surface area extraction

Thanh Noi and Kappas (2017) asserted that parameter tuning plays an important role in producing high accuracy results when using machine learning algorithms. Therefore, in this study to find the optimal RF classifier parameters that could accurately depict ISA, we tested a series of values for the tuning process. From the Landsat 5TM and Landsat 8OLI, we used 6 bands (VIS-SWIR) equalling 6 input predictor variables for the parameter tuning of the RF classifier. Four different sub-datasets with corresponding sizes 20%,

40%, 60% and 80% of the total training data were used to train the model and the rest to test the model (Figure 1). Finally, a range of values was used for the parameterization of both: $n_{tree} = 500:3000$ with a step size of 500.; $m_{try} = 1:6$ with a step size of 1. We implemented RF classification using Caret - RandomForest package in the R statistical software environment (RStudio, Inc., Boston, MA, USA, Version 1.1.463).

3.5. Model performance evaluation

In this study, the RF classifier performance evaluation was to assess the accuracy of the derived binary classification results for the three years based on stratified random evenly distributed samples obtained from reference data (aerial photo at 0.5 m spatial resolution and visual inspection on Google Earth). 10-fold cross-validation was used to rearrange the samples to ensure that each fold is a good representation of the whole datasets i.e., with a lower sample distribution variance compared to the hold-out cross-validation (Danjuma 2015). Finally, we implemented the performance evaluation metrics using the area under the receiver operating characteristic curve (AUROC) which is a graph that summarizes the performance of the indices (classifier) over all possible thresholds. The graph was generated by plotting the true positive rate (y-axis) against the false positive rate (x-axis) (Wieland and Pittore 2014). The performance evaluation metric was computed using InformationValue, plotROC and ggplot2 packages in the R statistical software environment RStudio, Inc., Boston, MA, USA, Version 1.1.463 (Prabhakaran2016; Sameen and Pradhan 2016).

3.6. Hexagon polygon grid to determine sampling scale

Grid analysis has been used to evaluate the composite effects due to its flexibility of analysis with scale variation, bounding of quantitative values and locations, and statistics of area proportions in these regular shapes (Xiao et al. 2018). Regular rectangular or square grid and hexagonal grid have also been compared in some studies and their relative merit was also examined. Aiazzi et al. (2002) earlier analyzed hexagonal sampling under general assumptions, compared with conventional rectangular sampling, and found out that hexagonal sampling was attractive for remote sensing applications where the acquisition process is crucial to preserve image quality without introducing data transmission overheads. He and Jia (2005) affirmed that hexagonal structure is considered to be preferable to the rectangular structure due to its higher sampling efficiency, consistent connectivity and higher angular resolution and is even proved to be superior to square structure in many applications. Birch et al. (2007) in earlier research also investigated the use of rectangular and hexagonal grids application in ecological observation, experiment and simulation such as the role of nearest neighbourhood in experimental design, the representation of connectivity in maps, and a new method for performing field surveys.

Since hexagonal grid is simpler and less ambiguous than a rectangular grid, we used QGIS (version 3.8) software to create a hexagonal polygon grid with matching centroids covering the RF extracted ISA extent for each year (i.e., 1995, 2005 and 2015) within the study area with the origin coordinate system. The decision to use the grid size of 90mx90m is similar to the study of Xiao et al. (2018) after empirically testing various grid sizes. At the optimal 90 m grid-scale, we observed that the hexagon grid size is much smaller than the impervious surface patches or thematic outputs, thereby preserving useful geometry for the interpretation of corresponding variables such as LST pixels i.e., ensuring that an adequate number of pixels is considered (Adeyemi et al. 2021). Afterward, the hexagon polygon grids were used to aggregate the LST raster layer pixels (i.e., ISA surface

temperature pixel values were averaged over the hexagon grid cells with the spatial analyst tool “zonal statistics module” of ArcGIS software) finally used to measure the geographical distribution (i.e., weighted standard deviational ellipse).

3.7. Spatial analytic method

The standard deviational ellipse (SDE) (Lefever 1926) methods were widely used to assess the spatial distribution evolution and distributional trends in many fields, because they can reveal the spatial concentration of geographical phenomena and the change characteristics of the geospatial distribution (Al-Kindi et al. 2017; Li et al. 2017; Qiao et al. 2018; Xu et al. 2018). To measure at local region scale the spatio-temporal developing trends of urban expansion in Pretoria, the weighted standard deviational ellipse (SDE) (Lefever 1926) method based on ISA surface temperature was used in this study. We also put into consideration the use of the Central Business District (CBD) as its reference point of expansion from the center to suburbs in the form of concentric circles, Qian and Wu (2019). Based on sampling hexagon polygon grid centroids representing the ISA surface temperature, the calculated parameters of the weighted SDE representing the dispersion and directional trends of the ISA at local region scale (sub-place units) were the long axis, short axis, and rotation angle. The rotation angle of the weighted SDE is calculated as follows:

$$a = \left(\sum_{i=1}^n w_i^2 \tilde{X}_i^2 - \sum_{i=1}^n w_i^2 \tilde{Y}_i^2 \right) \quad (9)$$

$$b = \sqrt{\left(\left(\sum_{i=1}^n w_i^2 \tilde{X}_i^2 - \sum_{i=1}^n w_i^2 \tilde{Y}_i^2 \right) - 4 \left(\sum_{i=1}^n w_i^2 \tilde{X}_i \tilde{Y}_i \right) \right)} \quad (10)$$

$$\tan \theta = \frac{a + b}{2 \left(\sum_{i=1}^n w_i^2 \tilde{X}_i \tilde{Y}_i \right)} \quad (11)$$

$$\begin{cases} \tilde{X}_i = X_i - \bar{X} \\ \tilde{Y}_i = Y_i - \bar{Y} \end{cases} \quad (12)$$

where θ is the rotation angle of the ellipse, indicating the angle measured clockwise from the North to the long axis of the ellipse (Equation 9, Equation 10 and Equation 11). The X and Y are the coordinates while \bar{X} and \bar{Y} are the mean X and Y coordinates (Equation 12). \tilde{X}_i and \tilde{Y}_i are the deviation between the i-th grid center in the X and Y direction respectively and w_i is the weight. In this study, the weight w_i indicates the ISA surface temperature of the i-th grid. The standard deviations σ_x and σ_y of the ellipse in the X and Y directions (Equation 13) are calculated as follows:

$$\begin{cases} \sigma_x = \sqrt{\frac{\sum_{i=1}^n (w_i \tilde{X}_i \cos \theta - w_i \tilde{Y}_i \sin \theta)^2}{\sum_{i=1}^n w_i}} \\ \sigma_y = \sqrt{\frac{\sum_{i=1}^n (w_i \tilde{X}_i \sin \theta - w_i \tilde{Y}_i \cos \theta)^2}{\sum_{i=1}^n w_i}} \end{cases} \quad (13)$$

The long axis, short axis and rotational angle represent the dispersion and directional trends of the urban ISA i.e., as the rotation angle changes in the SDEs show the spatio-temporal changes in the local impervious surfaces in a particular spatial direction. Subsequently, ellipse centre was drawn from the above SDE parameters, allowing us to

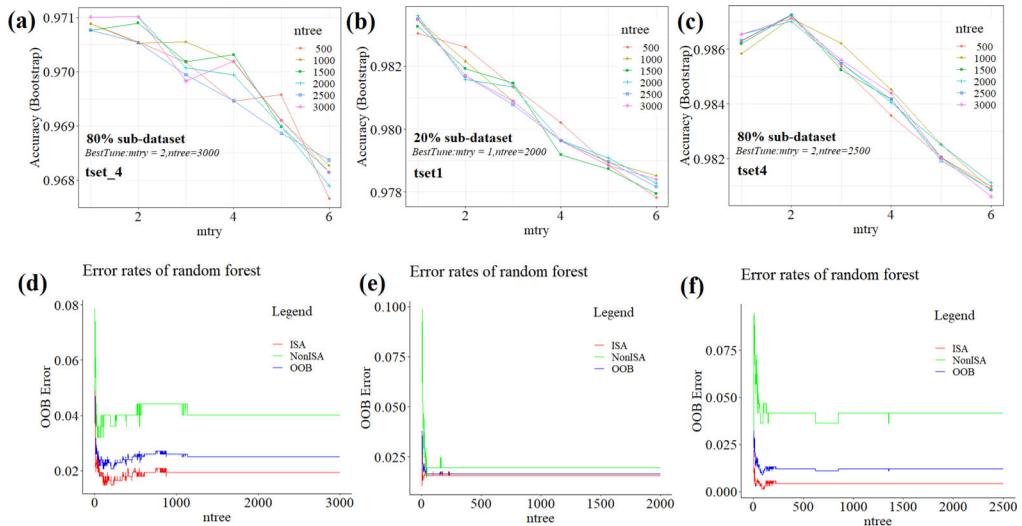


Figure 3. (a) – (c) Effect of the number of trees and the number of random split variables at each node (mtry) while (d) – (f) shows the relationship between OOB error (y-axis) and ntree parameter (x-axis) of the RF classifier based on the best sub-datasets of training sample data for 1995, 2005 and 2015.

see the elongation of the spatial distribution of impervious surface and its particular orientation. Lastly, we quantified the SDE using the spatial statistic “measuring geographical distribution” toolbox in ArcGIS.

4. Results

4.1. Random Forest classifier

The success of the RF classifier depended on the optimization of key parameters i.e., ntree and mtry. The grid search method was used to optimize the RF classifier using 5-fold cross-validation. The concept behind the grid search technique is that different pairs of parameters were evaluated and the one yielding the highest level of accuracy is selected similar to the study of Kavzoglu and Colkesen (2009). To find the optimal parameters for the RF classifier, several values (mtry = 1:6; ntree = 500:3000) were tested for all 4 sub-datasets. A sub-datasets of 80%, 20% and 80% respectively had the highest results for the Landsat-derived ISA obtained with mtry equal to 2 or 3 for 1995, 1 for 2005 and 1 or 2 for 2015 (Figure 3a–c). Furthermore, Figure 3d–f show that out-of-bag (OOB) error decreased sharply when ntree increased. Increase in ntree to ≥ 2000 , ≥ 1000 and ≥ 1500 respectively based on different sub-datasets for each year had slightly different trends as indicated in Figure 3a–c, however, generally, the OOBs were slightly reduced at all sub-datasets (Figure 3d–f). Also, all OOBs of all sub-datasets remain stable when ntree increase from 2000 to 3000, 1000 to 2000 and 1500 to 2500 respectively for each year used in this study (Figure 3d–f). Hence, ntree = 3000, 2000 and 2500 coupled with the highest results of mtry were the best parameters used.

4.2. The performance evaluation of random Forest classifier on Sub-datasets

As shown in Figure 3a–c with the best sub-datasets for each year, the three highest accuracies were considered for the random forest model. The highest accuracy for the random

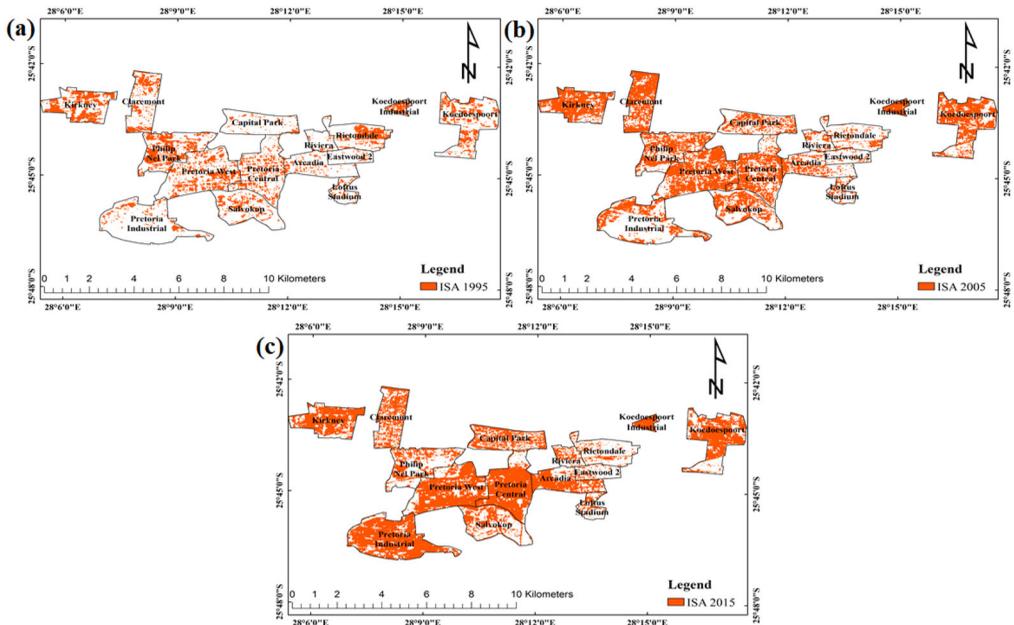


Figure 4. Random forest ISA classification maps of different periods (a)1995, (b) 2005 and (c) 2015.

forest model for 1995 and 2015 was observed when the training sample size was large enough (80%) i.e., $\sim 97\%$ and $\sim 98\%$ accuracies respectively. Whereas the highest accuracy for random forest model for 2005 occurred with 20% of the training samples i.e., $\sim 95\%$ (Figure 3b). In addition to visual examination of all the thematic images outputs shown in Figure 4a–c, the area under the receiver operating characteristic curve (AUROC) was used to evaluate the performance of the random forest classifier in extracting ISA or built-up areas across the study area over the 30 year period. As shown in Figure 5a–c, the computed AUROC values were: $AUROC = 0.8572$ for 1995, $AUROC = 0.8709$ for 2005, $AUROC = 0.8949$ for 2015 respectively. Therefore, based on the stratified 10-fold cross-validation, the results indicate that the random forest classifier effectively depicted ISA with relatively high precision.

4.3. Dynamic ISA change in the subplace units

Table 3 and Figure 6 reveal the spatial and temporal changes of the impervious surfaces area during the 30 study years. Based on the selected dominant land use of the administrative sub places, the results show an increase from 1995 to 2015 in ISA (hectares, ha) and growth rate(%) respectively. For instance, formal residential: Arcadia($\sim 41\text{ha}$ to $\sim 197\text{ha}$; 53%), Capital Park($\sim 31\text{ha}$ to $\sim 206\text{ha}$; 51%), Eastwood 2($\sim 13\text{ha}$ to $\sim 28\text{ha}$; 11%), Riviera(11 ha to $\sim 44\text{ha}$; 29%); Commercial: Loftus Stadium(21 ha to 37 ha; 14%), Pretoria Central($\sim 140\text{ha}$ to 419 ha; 55%), Pretoria West(197 ha to 454 ha; 42%); Industrial: Kirkney($\sim 178\text{ha}$ to $\sim 277\text{ha}$; $\sim 26\%$), Koedoespoort Industrial($\sim 38\text{ha}$ to 53 ha; 28%), Pretoria Industrial($\sim 59\text{ha}$ to 543 ha; $\sim 74\%$) and Collective living quarters: Salvokop(8 ha to $\sim 169\text{ha}$; 21%). Although the ISA relatively decreased in other places during the same period (e.g., Formal residential: Claremont($\sim 264\text{ha}$ to 204.03 ha), Philip Nel Park(243 ha to $\sim 168\text{ha}$), Rietondale($\sim 66\text{ha}$ to 59%) and Industrial: Koedoespoort(323 ha to 321 ha), it

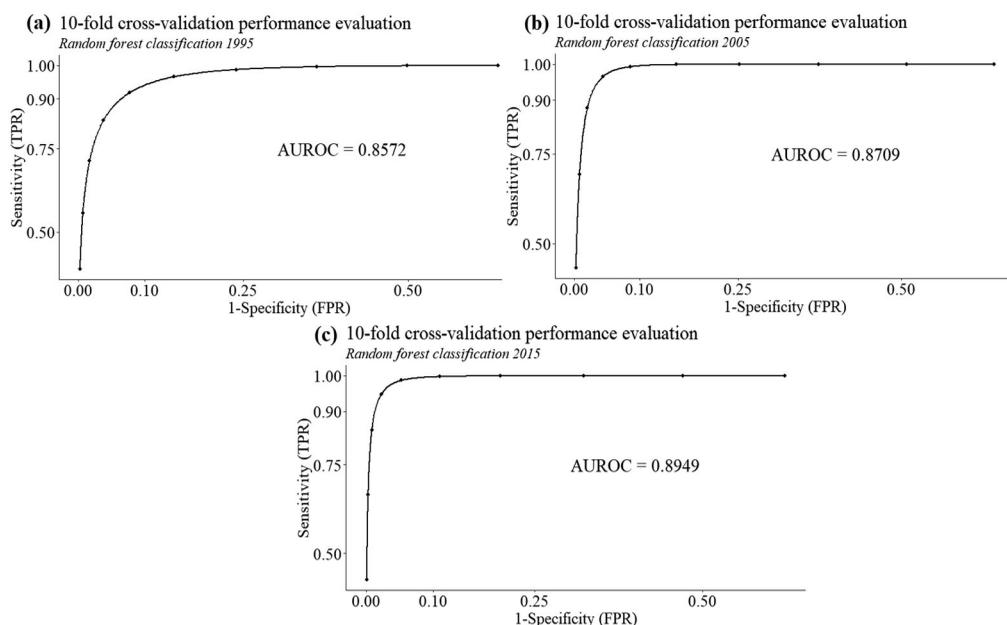


Figure 5. AUROC curve showing the performance evaluation of random forest ISA classification (a) 1995 (b) 2005 (c) 2015.

Table 3. Spatio-temporal analysis of the impervious surfaces area of selected sub-places in the study area.

Administrative Subplace	Dominant land use type	ISA 1995 (ha)	ISA 2005 (ha)	ISA 2015 (ha)
Arcadia	Formal residential	40.77	115.02	197.01
Capital Park	Formal residential	30.51	149.76	206.10
Claremont	Formal residential	102.24	263.61	204.03
Eastwood 2	Formal residential	12.87	21.96	27.54
Philip Nel-park	Formal residential	195.12	243.27	167.67
Rietondale	Formal residential	104.85	65.88	59.04
Riviera	Formal residential	11.25	33.12	43.56
Loftus stadium	Commercial	21.15	33.39	37.44
Pretoria central	Commercial	135.99	302.04	418.86
Pretoria West	Commercial	197.10	432	454.05
Kirkney	Industrial	177.75	285.66	276.48
Koedoessport Industrial	Industrial	37.80	52.74	53
Koedoessport Industrial	Industrial	150.93	323.37	321.93
Pretoria Industrial	Industrial	58.77	208.08	543.33
Salvokop	Collective living quarters	82.44	189.63	168.75

is still observed that more than 70% of the selected administrative subplace units in this study experienced dramatic growth in impervious surfaces.

$$Total\ Area = 5263ha; 1ha = pixel\ count \times 900m^2 \times 0.0001 \quad (14)$$

4.4. Land surface temperature retrieval (LST) for ISA

Figures 7–9 illustrate the spatial pattern of absolute normalized LST retrieved for the study. The computed LST map for the entire study area shows that for 1995, 2005 and

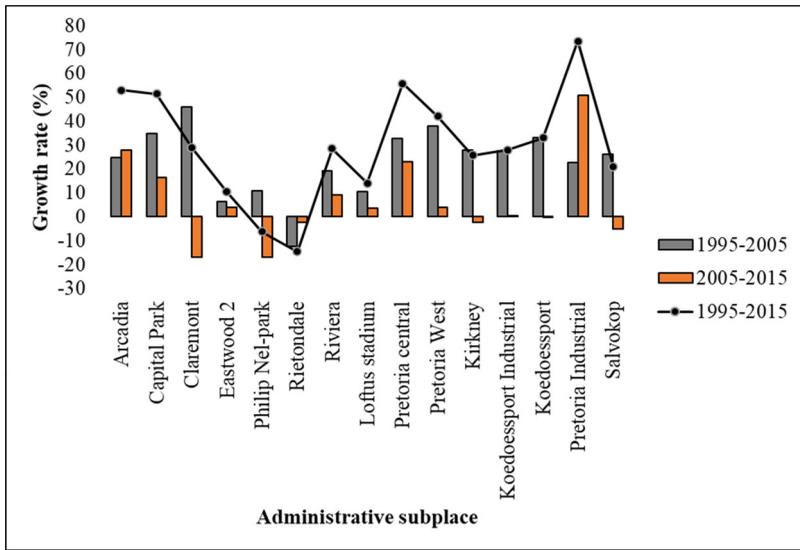


Figure 6. Variations in ISA growth rate for the selected administrative sub-places.

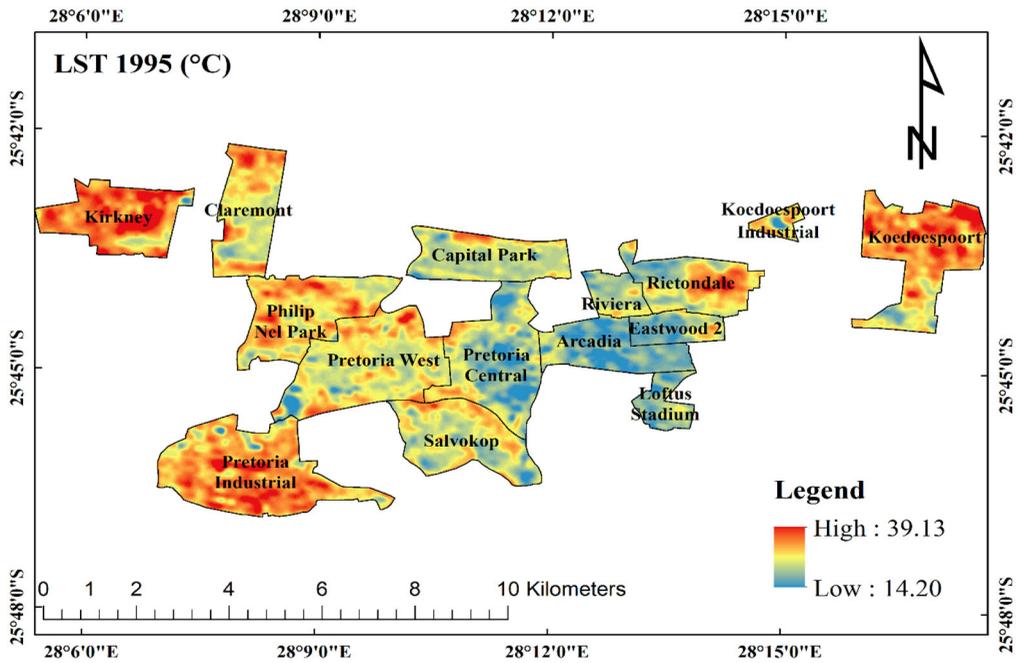


Figure 7. LST maps ($^{\circ}\text{C}$) of selected sub-place in Pretoria city in 1995.

2015, LST values range between $14.20^{\circ}\text{C} - 39.13^{\circ}\text{C}$, $16.72^{\circ}\text{C} - 44.23^{\circ}\text{C}$ and $18.15^{\circ}\text{C} - 48.25^{\circ}\text{C}$ respectively. This study revealed that the maximum LST for the whole area went up by $\sim 9^{\circ}\text{C}$ from 1995 to 2015, which were 42.12°C to 53.26°C ; the minimum temperature increased by 3°C from 14.20°C to 18.15°C , during the same season with the ten-year interval. This result indicates that the changes in land cover types thermal emittance have resulted in climate change as reported by the South African Weather Service (SAWS) in recent years. An earlier study by Adeyemi et al. (2015) revealed that ISA can

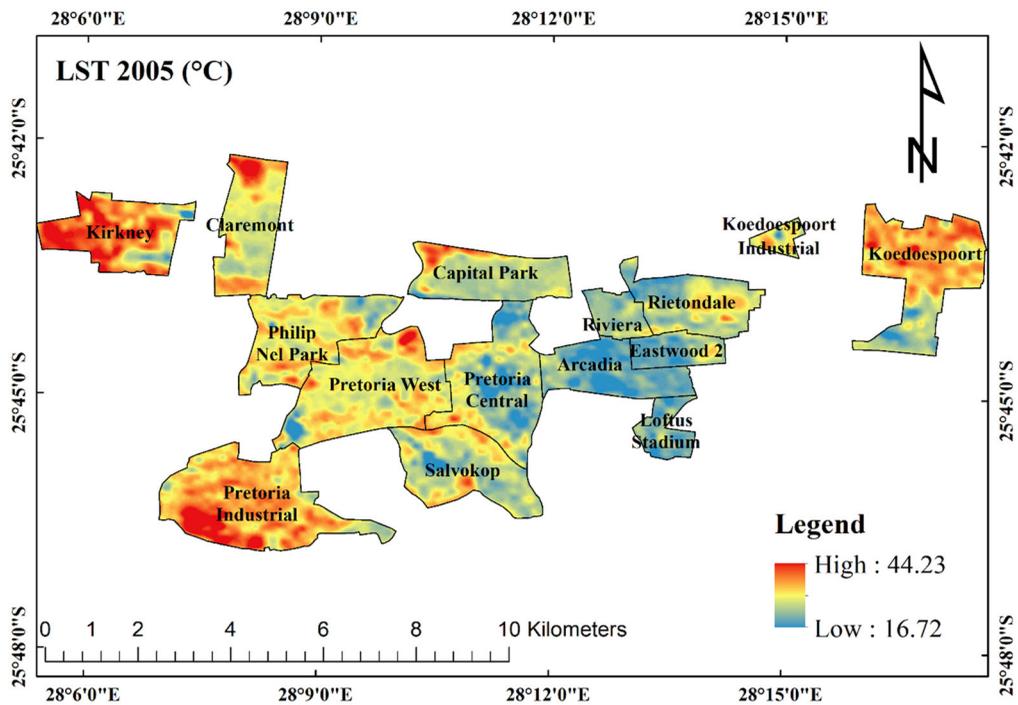


Figure 8. LST maps (°C) of selected sub-place in Pretoria city in 2005.

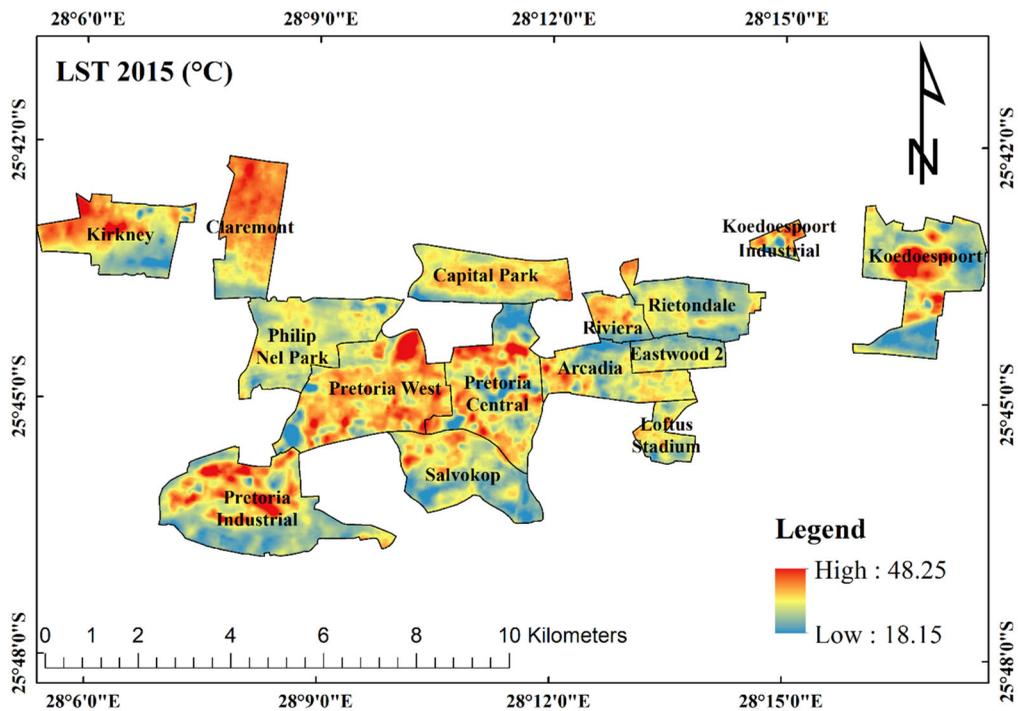


Figure 9. LST maps (°C) of selected sub-place in Pretoria city in 2015.

be used as a complementary metric for surface urban heat island studies, in this study we also examined the variation in ISA thermal emittance for 1995, 2005 and 2015 based on the above pixel samples. Based on the spatial distribution of land surface temperatures of ISA derived from the Landsat images for the selected administrative subplace units study years in Pretoria, the mean ISA surface temperature for 1995 was 22.51 °C (standard deviation of 1.38 °C), followed by 2005 with a mean of 27.01 °C (standard deviation of 1.62 °C) and the highest mean ISA LST of 29.48 °C (standard deviation of 2.21) in 2015. Also, we verified the overall metrics error between the predicted and actual LST to assess the accuracy of the retrieved normalized LST image for the study area. The result indicates that the overall retrieval error for the study area is quite low: ~1.44 °C for RMSE and ~1.05 °C for MAE (1995), ~1.40 °C for RMSE and ~1.08 °C for MAE (2005) and ~0.86 °C for RMSE and ~0.59 °C for MAE (2015) respectively. Thus, the thermal bands of Landsat 5TM and Landsat 8 TIRS data employed for this study provided good results and can be used for further temperature variability analysis.

4.5. Spatio-temporal developing trends of ISA expansion

In this study, the ISA surface temperature weighted standard deviation ellipse (SDE) was used to further reveal the spatio-temporal developing trends of ISA expansion. We used the rotation angle of SDE to analyse the spatial direction of impervious surface expansion (Table 4 and Figure 10a–o). It can be seen from Figure 10a–o that the SDEs of 15 administrative subplaces at the local region scale indicated significantly different ISA expansion directions. In Eastwood 2, with an approximated rotation angle of ~92° the spatial direction of ISA expansion was eastern in 1995 and 2005 (Figure 10d). After that the rotation angle decreased by 3.4° (Table 4), indicating a change in ISA distribution in an east-north-east in 2015. In Riviera, the rotation was maintained at an angle of ~114° - ~116° from 1995-2015 (Figure 10n). This implies that during the study period, the ISA mainly expanded towards the southeast. In Arcadia (Figure 10a), Pretoria Industrial (Figure 10k), Rietondale (Figure 10m) and Salvokop (Figure 10o) with rotation angles > 90°, the ISA mainly expanded to the east-south-east from 1995 and 2005. Also, a north-north-east ISA expansion trend was observed from 1995-2015 in Claremont (Figure 10c), Koedoespoort (Figure 10f), Loftus Stadium (Figure 10h) and Pretoria Central (Figure 10j) with rotation angles < 45°. In Capital Park (Figure 10b) and Kirkney (Figure 10e) with rotation angles slightly > 90°, the ISA expanded to the east. In Koedoespoort Industrial (Figure 10g), Philip Nel Park (Figure 10i) and Pretoria West (Figure 10l), with rotation angle maintained < 90°, the ISA significantly expanded towards the east-northeast in these subplaces during the study period.

5. Discussion

Over the years, the urban expansion experienced in the selected administrative subplaces for this study in Pretoria is not only seen as a sign of growth and prosperity but has continuously brought about expanded infrastructure which are impervious surfaces (Adeyemi et al. 2015). These increase in man-made features (i.e., ISA) and their sequential relationship with climatic variables such as surface temperature (e.g., LST) are crucial to understanding urban sprawl (Tian et al. 2018). In this study, the ISA data extracted with multispectral Landsat-5 TM and Landsat 8 with six bands (VIS-SWIR) images, were used to investigate the spatio-temporal dynamics and the expansion direction of urban sprawl at local administrative subplace units in Pretoria from 1995 to 2015. The first results in

Table 4. SDE parameters of impervious surface expansion from 1995 to 2015.

Subplace	Dominant land use type	Year	Rotation angle (°)	ISA Expansion Direction
Arcadia	Formal Residential	1995	97.433	ESE-ESE-ESE
		2005	97.532	
		2015	97.825	
Capital Park	Formal Residential	1995	95.234	E-E-E
		2005	94.998	
		2015	95.261	
Claremont	Formal Residential	1995	11.915	NNE-NNE-NNE
		2005	11.169	
		2015	10.763	
Philip Nel Park	Formal Residential	1995	66.276	ENE-ENE-ENE
		2005	67.229	
		2015	67.625	
Rietondale	Formal Residential	1995	103.429	ESE-ESE-ESE
		2005	98.664	
		2015	101.424	
Riviera	Formal Residential	1995	114.033	SE-SE-SE
		2005	116.313	
		2015	116.343	
Eastwood 2	Formal Residential	1995	91.457	E-E-ENE
		2005	91.588	
		2015	88.604	
Kirkney	Industrial	1995	93.274	E-E-E
		2005	96.762	
		2015	90.893	
Koedoespoort	Industrial	1995	28.309	NNE-NNE-NNE
		2005	24.712	
		2015	21.535	
Koedoespoort Industrial	Industrial	1995	73.335	ENE-ENE-ENE
		2005	73.995	
		2015	74.045	
Pretoria Industrial	Industrial	1995	101.356	ESE-ESE-ESE
		2005	99.848	
		2015	100.258	
Loftus Stadium	Commercial	1995	18.671	NNE-NNE-NNE
		2005	16.692	
		2015	11.769	
Pretoria Central	Commercial	1995	8.319	NNE-NNE-NNE
		2005	8.005	
		2015	6.561	
Pretoria West	Commercial	1995	68.786	ENE-ENE-ENE
		2005	68.396	
		2015	68.495	
Salvokop	Collective living quarters	1995	107.829	ESE-ESE-ESE
		2005	110.535	
		2015	110.813	

Figure 3 highlight the potential use of random forest classifier with different sample sizes to estimate ISA from Landsat image for the entire study years. Based on the four sub-datasets with corresponding sizes 20%, 40%, 60% and 80% of the total training data for each year, two different trends were clear: when the training sample size was good enough (80%), the highest accuracy for the RF model was observed for 1995 and 2015 i.e., ~97% and ~98% respectively (Figure 3a and c). Whereas highest accuracy of ~95% for rf model for 2005 occurred with 20% of the training samples (Figure 3b). The RF model accuracies can be attributed to the varying samples sizes of the imbalanced training data. Though this might be contradictory to many past studies on different satellite images such as Jin et al. (2014), Colditz (2015) and Mellor et al. (2015) to mention a few that asserted that the bigger the land cover class area is, the more training samples that are required to produce the best classification accuracy. The use of RF modelling of ISA in our study reveals

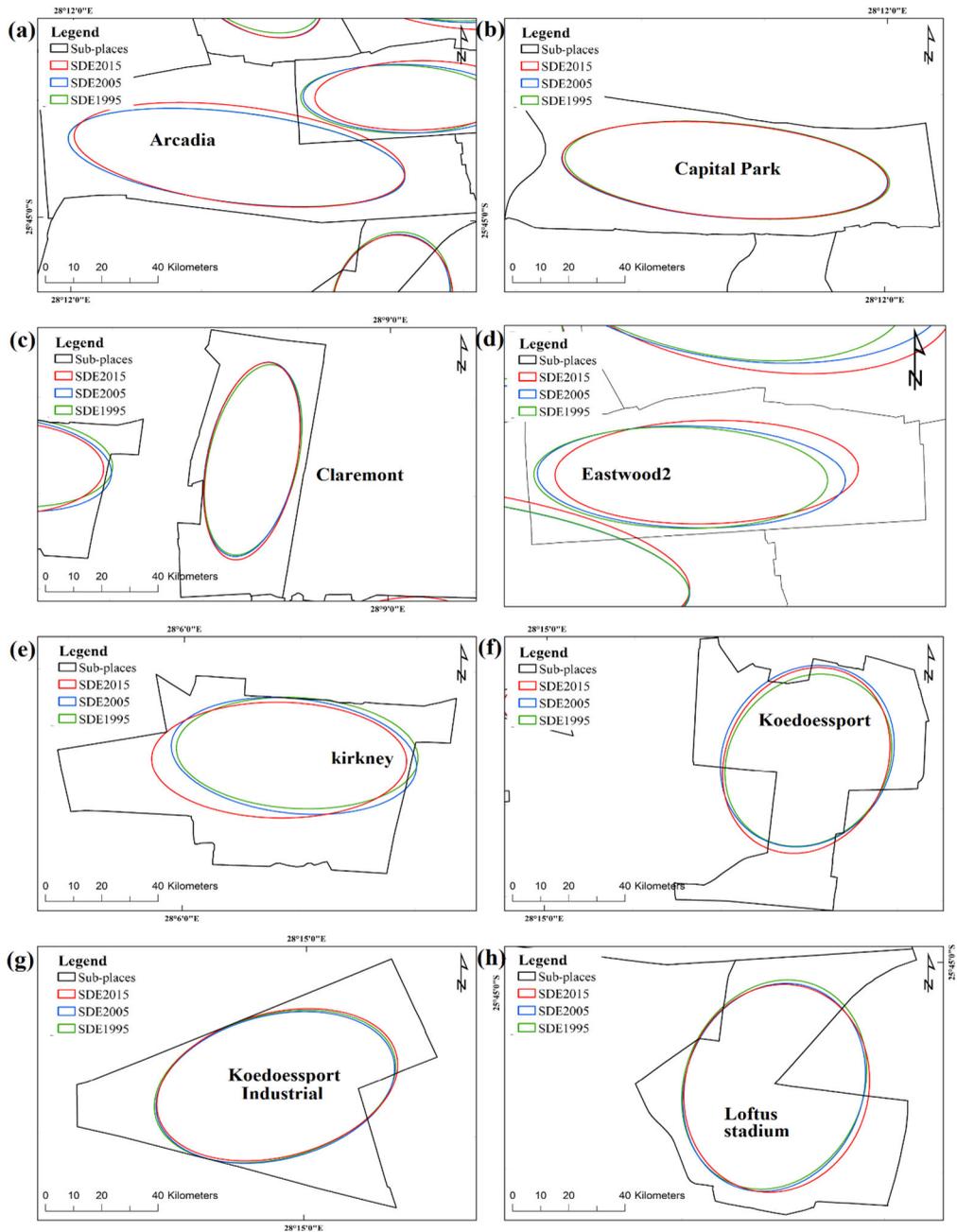


Figure 10. SDEs of ISA surface temperature for the selected administrative sub-places (i.e., local region scale) for different periods (continued overleaf).

similarity to the studies of Zhang et al. (2021), Liu et al. (2021), Shrestha et al. (2021), Guo et al. (2020) and Emanuel Gombe et al. (2017). Also similar to the study of Thanh Noi and Kappas (2017) the RF classifier employed in this study showed less sensitivity to the imbalanced training data as long as the training sample size is representative enough i.e., either large or small. After the visual examination of the random forest classifier thematic ISA outputs for the study years presented in Figure 4, their quantitative assessment

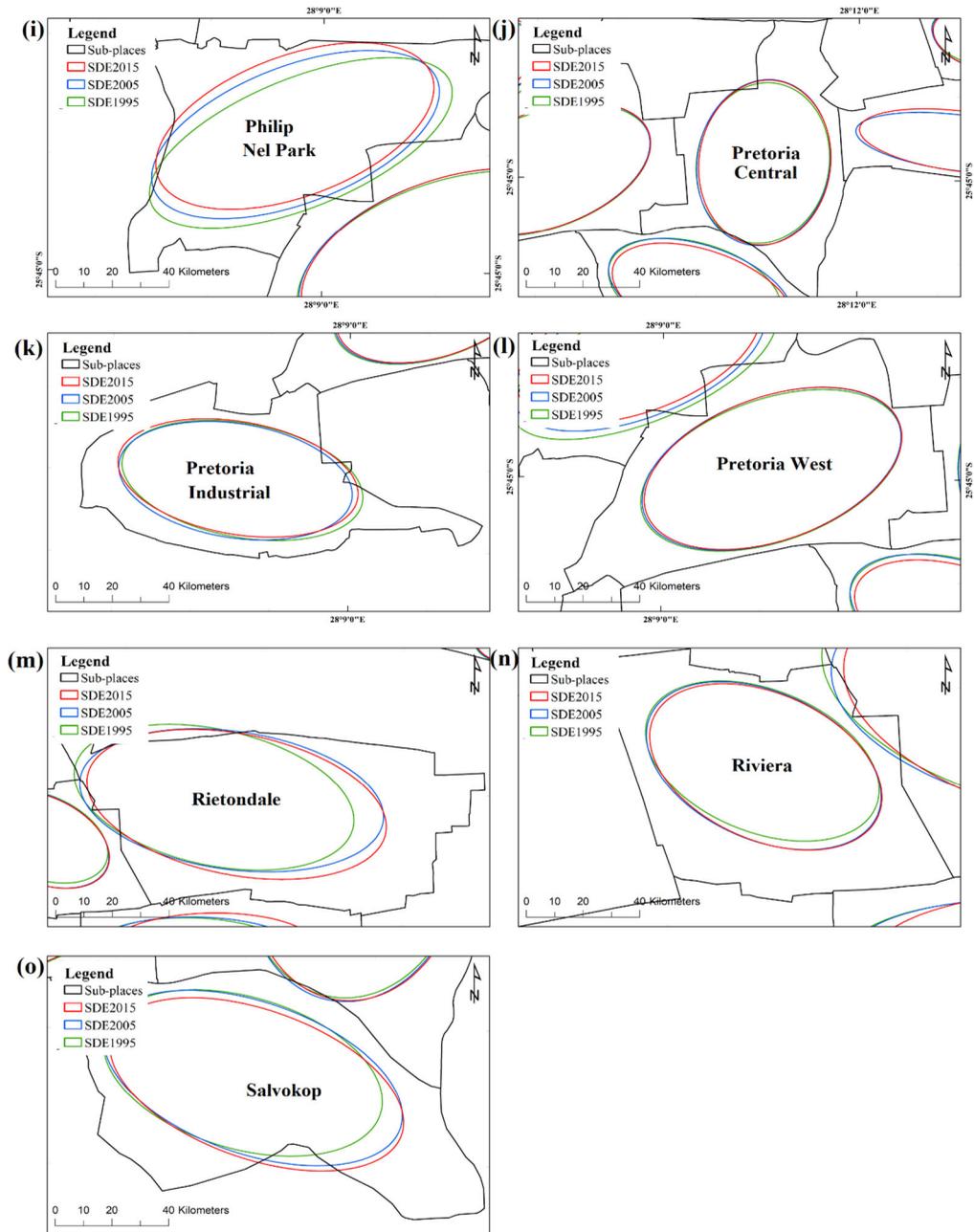


Figure 10. Continued.

based on 10-fold cross-validation, the AUROC was used to assess the unbiased predictive accuracy (Figure 5). Although the random forest classifier overall predictive accuracy was fairly high (i.e., AUROC = 0.8572 for 1995, AUROC = 0.8709 for 2005, AUROC = 0.8949 for 2015) because of the selection of representative training samples or pixels (Maxwell et al. 2018), there were still errors observed in the final thematic outputs due to mixed pixels i.e., ISA and vegetation (Xu et al. 2018) associated with the use of medium resolution multispectral satellite imagery. Secondly, we examined the ISA spatio-temporal

dynamics within ten years interval time frame (i.e., 1995 – 2015) at local region level. Results in [Table 3](#) and [Figure 6](#) above reveals while more than 70% of the selected administrative subplaces (i.e., Arcadia, Capital Park, Eastwood 2, Loftus Stadium, Koedoespoort Industrial, Pretoria Central, Pretoria Industrial, Pretoria West, Riviera) in this study experienced dramatically increase in ISA growth rate. Generally, the ISA spatio-temporal dynamics in the study area could be attributed to the incessant urban sprawl resulting in many places across Pretoria. Since Pretoria is one of the three capital cities in South Africa, the remarkable ISA growth over the years could also be due to political and socio-economic factors. Finally in our study, guided by the previous study of [Xiao et al. \(2018\)](#) on an optimal analytical scale, we used the hexagon polygon grid covering and aggregating the depicted ISA surface temperature pixels to examine the spatio-temporal characteristics or developing trends of ISA expansion with the aid of weighted standard deviational ellipse (SDE) method. Similar to the recent studies of [Xu et al. 2018](#); [Man et al. 2019](#) and [Hua et al. 2020](#)), our results indicated that the ISA exhibited an expansion trend generally in the east-south-east, east, north-north-east, east-north-east and south-east directions. This can be attributed to the change ISA growth rate coupled with population and various land use activities at the local administrative subplace units. In this study, it can therefore be asserted that the spatio-temporal pattern of ISA surface temperature is an important metric in understanding the principle direction of ISA expansion.

6. Conclusion

Satellite imagery that measures spatio-temporal dynamics of impervious surface areas (ISA) in the context of rapid development, is key to understanding the process of urban expansion. The information obtained this way can serve as valuable input when dealing with challenges related to the environment, climate (for example shifts in land surface temperature (LST)), population health, natural resources etc.). Using a combination of quantitative remote sensing images such as Landsat 5 Thematic Mapper (TM), Landsat 8 Thermal Infrared Sensor (TIRS) and Operational Land Imager (OLI), and spatial statistical methods, the study investigated the spatio-temporal direction of ISA expansion at a local spatial scale, based on its surface temperature and within a time frame interval of ten years. The study displayed two main strengths. Firstly, the use of random forest algorithm (RF) based on different training sample subsets, enabled the researchers to accurately estimate and reveal the spatio-temporal dynamics of ISA in selected administrative sub place levels in Pretoria. Secondly, the researchers were able to identify the principal direction of urban expansion at a local spatial scale in Pretoria by combining zonal statistics with weighted SDE spatial statistical method. The findings of this study could be used by policymakers and urban planners as a key measure to detect places where urbanization is rapid, and prioritize areas of immediate attention and development of smart growth strategies. Future studies should focus on spatio-temporal urban expansion at different spatial scales (e.g., local and regional), depending on the coverage and commercial availability of fine resolution multispectral satellite imagery. Population, gross domestic product (GDP), topography, hydrology, socio-economic settings etc. may also be considered as drivers when modelling spatio-temporal urban dynamics.

Acknowledgments

We acknowledge the United States Geological Survey Global Visualization Viewer (<https://glovis.usgs.gov>) due to their freely accessible Landsat data. Much appreciation also goes to Tema Matilda Yebga - Adeyemi and Minette Teessen (Professional Editors Guild) for their support.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The author would like to thank the University of South Africa Student Funding Directorate (UNISA, DSF) and GeoTerraImage (Pty) Ltd for providing the research bursary to conduct this study.

References

- Adelabu S, Mutanga O, Adam E, Cho MA. 2013. Exploiting machine learning algorithms for tree species classification in a semiarid woodland using RapidEye image. *J Appl Remote Sens.* 7(1):073480.
- Adeyemi A, Botai J, Ramoelo A, Van der Merwe F, Tsela P. 2015. Effect of impervious surface area and vegetation changes on mean surface temperature over Tshwane metropolis, Gauteng Province, South Africa. *SA J Geomatics.* 4(4):351. <https://www.ajol.info/index.php/sajg/article/view/125779>.
- Adeyemi A, Ramoelo A, Cho M, Masemola C. 2021. Spectral index to improve the extraction of built-up area from WorldView-2 imagery. *J Appl Rem Sens.* 15(02):1–19 .
- Adeyemi A, Ramoelo A, Cho MA, Strydom J. 2021. Spatio-temporal analysis of built-up impervious surface area and interplay with land surface temperature in Pretoria, South Africa. *Geocarto Int.* DOI:10.1080/10106049.2021.1980617
- Aguilar MA, Bianconi F, Aguilar FJ, Fernández I. 2014. Object-based greenhouse classification from GeoEye-1 and WorldView-2 stereo imagery. *Remote Sensing.* 6(5):3554–3582.
- Aiazzi B, Baronti S, Capanni A, Santurri L, Vitulli R. 2002 Mar. Advantages of hexagonal sampling grids and hexagonal shape detector elements in remote sensing imagers. In: *European Signal Processing Conference.*
- Al-Kindi KM, Kwan P, Andrew NR, Welch M. 2017. Modelling spatiotemporal patterns of dubas bug infestations on date palms in northern Oman: a geographical information system case study. *Crop Prot.* 93:113–121.
- Alipour T, Sarajian M, Esmacely A. 2011. Land surface temperature estimation from thermal band of landsat sensor, case study: Alashtar City. *Int Arch Photogramm Remote Sens Spat Inf Sci.* XXXVIII(2003):1–6.
- Artmann M. 2014. Assessment of soil sealing management responses, strategies, and targets toward ecologically sustainable urban land use management. *Ambio.* 43(4):530–541.
- Belgiu M, Drăgu L. 2016. Random forest in remote sensing: a review of applications and future directions. *ISPRS J Photogramm Remote Sens.* 114:24–31.
- Berhane TM, Lane CR, Wu Q, Autrey BC, Anenkhonov OA, Chepinoga VV, Liu H. 2018. Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory. *Remote Sens (Basel).* 10(4):580.
- Bhaskaran S, Paramananda S, Ramnarayan M. 2010. Per-pixel and object-oriented classification methods for mapping urban features using Ikonos satellite data. *Appl Geogr.* 30(4):650–665.
- Birch CPD, Oom SP, Beecham JA. 2007. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecol Modell.* 206(3–4):347–359.
- Bouzekri S, Lasbet AA, Lachehab A. 2015. A new spectral index for extraction of built-up area using Landsat-8 data. *J Indian Soc Remote Sens.* 43(4):867–873.
- Breiman L. 1996. Bagging predictors. *Mach Learn.* 24(2):123–140.
- Breiman L. 2001. Random Forests. *Mach Learn.* 45(1):5–32.
- Carlson TN, Ripley DA. 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sens Environ.* 62(3):241–252.
- Colditz RR. 2015. An evaluation of different training sample allocation schemes for discrete and continuous land cover classification using decision tree-based algorithms. *Remote Sensing.* 7(8):9655–9681.
- Cracknell MJ, Reading AM. 2014. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers and Geosciences.* 63:22–33.
- Danjuma KJ. 2015. Performance evaluation of machine learning algorithms in post-operative life expectancy in the lung cancer patients. *Inter J Computer Sci.* 12(2):1–11.
- Demographia. 2017. Demographia world urban areas. In: 13th Annual Edition. <http://www.demographia.com/db-worldua.pdf>.
- Deng C, Wu C. 2013. Examining the impacts of urban biophysical compositions on surface urban heat island: A spectral unmixing and thermal mixing approach. *Remote Sens Environ.* 131:262–274.

- Emanuel Gombe K, Asanuma I, Park J-G. 2017. Quantification of annual urban growth of Dar es Salaam Tanzania from Landsat time Series data. *Adv Remote Sens.* 06(03):175–191.
- Environment for Visualizing Images ENVI 2015. 2014. Environment for visualizing images: Release 5.2; Exelis Visual Information Solutions. Boulder, CO: ITT Industries.
- Felde GW, Anderson GP, Cooley TW, Matthew MW, Adler-Golden SM, Berk A, Lee J. 2003. Analysis of Hyperion data with the FLAASH atmospheric correction algorithm. In: *International Geoscience and Remote Sensing Symposium (IGARSS)*. Vol. 1; p. 90–92.
- Feng Q, Gong J, Liu J, Li Y. 2015. Flood mapping based on multiple endmember spectral mixture analysis and random forest classifier-the case of Yuyao, China. *Remote Sens.* 7(9):12539–12562.
- Guo X, Zhang C, Luo W, Yang J, Yang M. 2020. Urban impervious surface extraction based on multi-features and random forest. *IEEE Access.* 8: 226609–226623.
- Gutman G, Ignatov A. 1998. The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *Int J Remote Sens.* 19(8):1533–1543.
- He X, Jia W. 2005. Hexagonal structure for intelligent vision. In: *Proceedings of 1st International Conference on Information and Communication Technology, ICICT 2005*; p. 52–64.
- Hu Y, Jia G. 2010. Influence of land use change on urban heat island derived from multi-sensor data. *Int J Climatol.* 30(9):1382–1395.
- Hua L, Zhang X, Nie Q, Sun F, Tang L. 2020. The impacts of the expansion of urban impervious surfaces on urban heat islands in a coastal city in China. *Sustainability (Switzerland)*. 12(2):475.
- Hu X, Weng Q. 2009. Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks. *Remote Sens Environ.* 113(10): 2089–2102.
- Ishimwe R, Abutaleb K, Ahmed F, Ngie A. 2014. Comparison of pixel-based and object-based classification approaches using Landsat 8 OLI and TIR spectral bands. In: *Proceedings of the 10th International Conference of AARSE*; p. 276–284.
- Jian P, Yanxu L, Hong S, Pan X, Xiaoxu H, Yanglin W. 2016. Using impervious surfaces to detect urban expansion in Beijing of China in 2000s. *Chin Geogr Sci.* 26(2):229–243.
- Jiménez-Munoz JC, Sobrino JA. 2003. A generalized single-channel method for retrieving land surface temperature from remote sensing data. *J Geophys Res Atmos.* 108(22):1–9.
- Jiménez-Munoz JC, Sobrino JA. 2010. A single-channel algorithm for land-surface temperature retrieval from ASTER data. *IEEE Geosci Remote Sens Lett.* 7(1):176–179.
- Jin H, Stehman SV, Mountrakis G. 2014. Assessing the impact of training sample selection on accuracy of an urban classification: a case study in Denver, Colorado. *Int J Remote Sens.* 35(6):2067–2081.
- Kavzoglu T, Colkesen I. 2009. A kernel functions analysis for support vector machines for land cover classification. *Int J Appl Earth Obs Geoinf.* 11(5):352–359.
- Krause K. 2005. Radiometric use of quickbird imagery. Colorado. <http://www.mendeley.com/research/radiometric-quickbird-imagery/>.
- Lefever DW. 1926. Measuring geographic concentration by means of the standard deviational ellipse. *Am J Sociol.* 32(1):88–94.
- Li Y, Gong H, Zhu L, Li X. 2017. Measuring spatiotemporal features of land subsidence, groundwater drawdown, and compressible layer thickness in Beijing Plain, China. *Water (Switzerland)*. 9(1):64.
- Liu Y, Zuo R, Dong Y. 2021. Analysis of temporal and spatial characteristics of urban expansion in Xiaonan District from 1990 to 2020 using time series Landsat imagery. *Remote Sens.* 13(21):4299.
- Loosvelt L, Peters J, Skriver H, Lievens H, Van Coillie FMB, De Baets B, Verhoest NEC. 2012. Random Forests as a tool for estimating uncertainty at pixel-level in SAR image classification. *Int J Appl Earth Obs Geoinf.* 19(1):173–184.
- Lu D, Moran E, Hetrick S. 2011. Detection of impervious surface change with multitemporal Landsat images in an urban-rural frontier. *ISPRS J Photogramm Remote Sens.* 66(3):298–306.
- Man W, Nie Q, Hua L, Wu X, Li H. 2019. Spatio-temporal variations in impervious surface patterns during urban expansion in a coastal city: Xiamen, China. *Sustainability (Switzerland)*. 11(8):2404.
- Markham BL, Barker JL. 1985. Spectral characterization of the landsat thematic mapper sensors. *Int J Remote Sens.* 6(5):697–716.
- Masek JG, Lindsay FE, Goward SN. 2000. Dynamics of urban growth in the Washington DC metropolitan area, 1973–1996, from Landsat observations. *Int J Remote Sens.* 21(18):3473–3486.
- Maxwell AE, Warner TA, Fang F. 2018. Implementation of machine-learning classification in remote sensing: An applied review. *Int J Remote Sens.* 39(9):2784–2817.
- McGregor GRM, Bessemoulin P, Ebi K, Menne B. 2015. Heatwaves and health: guidance on warning-system development. Geneva, Switzerland: World Meteorological Organization. <https://public.wmo.int/en/resources/library/heatwaves-and-health-guidance-warning-system-development>.

- Mellor A, Boukir S, Haywood A, Jones S. 2015. Exploring issues of training data imbalance and mislabeling on random forest performance for large area land cover classification using the ensemble margin. *ISPRS J Photogramm Remote Sens.* 105:155–168.
- Morabito M, Crisci A, Messeri A, Orlandini S, Raschi A, Maracchi G, Munafò M. 2016. The impact of built-up surfaces on land surface temperatures in Italian urban areas. *Sci Total Environ.* 551-552: 317–326.
- Nie Q, Xu J. 2015. Understanding the effects of the impervious surfaces pattern on land surface temperature in an urban area. *Front Earth Sci.* 9(2):276–285.
- Odindi J, Mhangara P, Kakembo V. 2012. Remote sensing land-cover change in Port Elizabeth during South Africa's democratic transition. *S Afr J Sci.* 108(5/6):5–6. (
- Okujeni A, Canters F, Cooper SD, Degerickx J, Heiden U, Hostert P, Priem F, Roberts DA, Somers B, van der Linden S. 2018. Generalizing machine learning regression models using multi-site spectral libraries for mapping vegetation-impervious-soil fractions across multiple cities. *Remote Sens Environ.* 216:482–496.
- Okujeni A, van der Linden S, Hostert P. 2015. Extending the vegetation-impervious-soil model using simulated EnMAP data and machine learning. *Remote Sens Environ.* 158:69–80.
- Owen TW, Carlson TN, Gillies RR. 1998. An assessment of satellite remotely-sensed land cover parameters in quantitatively describing the climatic effect of urbanization. *Int J Remote Sens.* 19(9):1663–1681.
- Özçift A. 2011. Random forests ensemble classifier trained with data resampling strategy to improve cardiac arrhythmia diagnosis. *Comput Biol Med.* 41(5):265–271.
- Prabhakaran S. 2016. Performance analysis and companion functions for binary classification models. [Internet]. <https://www.rdocumentation.org/packages/InformationValue/versions/1.2.3>.
- Qian Y, Wu Z. 2019. Study on urban expansion using the spatial and temporal dynamic changes in the impervious surface in Nanjing. *Sustainability (Switzerland).* 11(3):933.
- Qiao K, Zhu W, Hu D, Hao M, Chen S, Cao S. 2018. Examining the distribution and dynamics of impervious surface in different function zones in Beijing. *J Geogr Sci.* 28(5):669–684.
- Qin Z, Karnieli A, Berliner P. 2001. A mono-window algorithm for retrieving land surface temperature from Landsat TM data and its application to the Israel-Egypt border region. *Int J Remote Sens.* 22(18): 3719–3746.
- Ramezan CA, Warner TA, Maxwell AE. 2019. Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. *Remote Sens.* 11(2):185.
- Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J Photogramm Remote Sens.* 67(1):93–104.
- Sahana M, Hong H, Sajjad H. 2018. Analyzing urban spatial patterns and trend of urban growth using urban sprawl matrix: a study on Kolkata urban agglomeration, India. *Sci Total Environ.* 628–629: 1557–1566.
- Sameen MI, Pradhan B. 2016. A novel built-up spectral index developed by using multiobjective particle-swarm-optimization technique. *IOP Conf Ser: Earth Environ Sci.* 37(1):012006.
- Shi L, Ling F, Ge Y, Foody GM, Li X, Wang L, Zhang Y, Du Y. 2017. Impervious surface change mapping with an uncertainty-based spatial-temporal consistency model: a case study in Wuhan City using landsat time-series datasets from 1987 to 2016. *Remote Sens.* 9(11):1148.
- Shrestha B, Ahmad S, Stephen H. 2021. Fusion of Sentinel-1 and Sentinel-2 data in mapping the impervious surfaces at city scale. *Environ Monit Assess.* 193(9):556.
- Shrestha B, Stephen H, Ahmad S. 2021. Impervious surfaces mapping at city scale by fusion of radar and optical data through a random forest classifier. *Remote Sens.* 13(15):3040.
- Sobrino JA, Jiménez-Muñoz JC, Paolini L. 2004. Land surface temperature retrieval from LANDSAT TM 5. *Remote Sens Environ.* 90(4):434–440.
- Sobrino JA, Jiménez-Muñoz JC, Sòria G, Romaguera M, Guanter L, Moreno J, Plaza A, Martínez P. 2008. Land surface emissivity retrieval from different VNIR and TIR sensors. *IEEE Trans Geosci Remote Sens.* 46(2):316–327.
- South African Weather Service. 2011. Climate data for Pretoria [Internet]. Pretoria. <http://www.weathersa.co.za/Climat/Climstats/BloemfonteinStats.jsp>.
- South African Weather Services. 2013. Climate data for Pretoria. Pretoria.
- Sun Z. 2011. Estimating urban impervious surfaces from Landsat-5 TM imagery using multilayer perceptron neural network and support vector machine. *J Appl Remote Sens.* 5(1):053501.
- Sun Z, Wang C, Guo H, Shang R. 2017. A modified normalized difference impervious surface index (MNDISI) for automatic urban mapping from landsat imagery. *Remote Sens.* 9(9):942.

- Thanh Noi P, Kappas M. 2017. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors (Basel, Switzerland)*. 18(2):18.
- Tian Y, Chen H, Song Q, Zheng K. 2018. A novel index for impervious surface area mapping: development and validation. *Remote Sens*. 10(10):1521.
- United States Geological Survey. 2013. Landsat 8 Landsat data continuity mission. Fact sheet 2013–3060. [Internet]. Reston, VA. <https://www.usgs.gov/>.
- van de Voorde T, de Roeck T, Canters F. 2009. A comparison of two spectral mixture modelling approaches for impervious surface mapping in urban areas. *Int J Remote Sens*. 30(18):4785–4806.
- Vanhulsel M, Beckx C, Janssens D, Vanhoof K, Wets G. 2011. Measuring dissimilarity of geographically dispersed space-time paths. *Transportation*. 38(1):65–79.
- Ward K, Lauf S, Kleinschmit B, Endlicher W. 2016. Heat waves and urban heat islands in Europe: A review of relevant drivers. *Sci Total Environ*. 569–570:527–539.
- Waske B, Braun M. 2009. Classifier ensembles for land cover mapping using multitemporal SAR imagery. *ISPRS J Photogramm Remote Sens*. 64(5):450–457.
- Watts JD, Lawrence LR. 2008. Merging random forest classification with an object-oriented approach for analysis of agricultural lands. *Int Archiv Photogramm Remote Sens Spatial Inf Sci*. 37(2018):579–582.
- Wei C, Blaschke T. 2018. Pixel-wise vs. object-based impervious surface analysis from remote sensing: correlations with land surface temperature and population density. *Urban Sci*. 2(1):2.
- Weng Q. 2012. Remote sensing of impervious surfaces in the urban areas: requirements, methods, and trends. *Remote Sens Environ*. 117:34–49.
- Weng Q, Dengsheng L, Jacquelyn S. 2004. Estimation of land surface temperature–vegetation abundance relationship for urban heat island studies. *Remote Sens Environ*. 89(4):467–483.
- Weng Q, Lu D. 2008. A sub-pixel analysis of urbanization effect on land surface temperature and its interplay with impervious surface and vegetation coverage in Indianapolis, United States. *Int J Appl Earth Obs Geoinf*. 10(1):68–83.
- Wieland M, Pittore M. 2014. Performance evaluation of machine learning algorithms for urban pattern recognition from multi-spectral satellite images. *Remote Sens*. 6(4):2912–2939.
- Xian G, Crane M. 2006. An analysis of urban thermal characteristics and associated land cover in Tampa Bay and Las Vegas using Landsat satellite data. *Remote Sens Environ*. 104(2):147–156.
- Xiao H, Kopecká M, Guo S, Guan Y, Cai D, Zhang C, Zhang X, Yao W. 2018. Responses of urban land surface temperature on land cover: a comparative study of Vienna and Madrid. *Sustainability (Switzerland)*. 10(2):260.
- Xu H. 2013. Rule-based impervious surface mapping using high spatial resolution imagery. *Int J Remote Sens*. 34(1):27–44.
- Xu Z, Chen J, Xia J, Du P, Zheng H, Gan L. 2018. Multisource earth observation data for land-cover classification using random forest. *IEEE Geosci Remote Sens Lett*. 15(5):789–793.
- Xu R, Liu J, Xu J. 2018. Extraction of high-precision urban impervious surfaces from sentinel-2 multispectral imagery via modified linear spectral mixture analysis. *Sensors*. 18(9):2873.
- Xu Z, Mountrakis G, Quackenbush LJ. 2017. Impervious surface extraction in imbalanced datasets: integrating partial results and multi-temporal information in an iterative one-class classifier. *Int J Remote Sens*. 38(1):43–63.
- Xu H, Wang M. 2016. Remote sensing-based retrieval of ground impervious surfaces. *Yaogan Xuebao/J Remote Sens*. 20(5):1270–1289.
- Xu J, Zhao Y, Zhong K, Zhang F, Liu X, Sun C. 2018. Measuring spatio-temporal dynamics of impervious surface in Guangzhou, China, from 1988 to 2015, using time-series Landsat imagery. *Sci Total Environ*. 627:264–281.
- Yu Y, Li J, Zhu C, Plaza A. 2018. Urban impervious surface estimation from remote sensing and social data. *Photogramm Eng Remote Sens*. 84(12):771–780.
- Yu S, Sun Z, Guo H, Zhao X, Sun L, Wu M. 2017. Monitoring and analyzing the spatial dynamics and patterns of megacities along the Maritime Silk Road. *Yaogan Xuebao/J Remote Sens*. 21(2):169–181.
- Zhang T, Huang X. 2018. Monitoring of urban impervious surfaces using time series of high-resolution remote sensing images in rapidly urbanized areas: a case study of Shenzhen. *IEEE J Sel Top Appl Earth Observ Remote Sens*. 11(8):2692–2708.
- Zhang X, Liu L, Chen X, Gao Y, Jiang M. 2021. Automatically monitoring impervious surfaces using spectral generalization and time series landsat imagery from 1985 to 2020 in the Yangtze River Delta. *J Remote Sens*. 2021:1–16.
- Zhang Y, Zhang H, Lin H. 2014. Improving the impervious surface estimation with combined use of optical and SAR remote sensing images. *Remote Sens Environ*. 141:155–167.