



OPEN

DATA DESCRIPTOR

Annual Impervious Surface Data from 2001–2020 for West African Countries: Ghana, Togo, Benin and Nigeria

Andrews Korah  & Michael C. Wimberly 

Impervious surface data are increasingly important for research and planning. Despite the availability of global and local urban land cover maps, regional data are lacking in Africa. We generated annual 30 m impervious cover data from 2001–2020 for Ghana, Togo, Benin, and Nigeria using the Landsat archive. We used random forest to predict impervious cover using 11 spectral indices and applied pixel-level temporal segmentation with the LandTrendr algorithm. Processing with LandTrendr improved the accuracy of the random forest predictions, with higher predicted-observed r^2 (0.81), and lower mean error (-0.03), mean absolute error (5.73%), and root mean squared error (9.93%). We classified pixels $>20\%$ impervious as developed and $<=20\%$ impervious as undeveloped. This classification had 93% overall accuracy and similar producer's (79%) and user's (80%) accuracies for developed area. Our maps had higher accuracy and captured more developed areas than comparable global datasets. This is the first regionally calibrated 30 m resolution impervious dataset in West Africa, which can support research on drivers and impacts of urban expansion and planning for future growth.

Background & Summary

Globally, the number of people living in cities is increasing rapidly, with most of this growth occurring in urbanizing Asia and Africa^{1–3}. Future projections are that Africa alone will add 930 million urban residents by 2050⁴. The growth of urban populations results in the expansion of impervious surfaces, including building roofs, glass, concrete, asphalt, and paved roads. These increases in developed areas have altered many of Earth's surface processes, including energy balance, hydrological cycle, atmospheric circulation, and phenology⁵. These changes manifest through increased heat stress, urban fires, and floods combined with losses of cropland, natural habitats, and biodiversity^{6–10}. With increasing urbanization in low-income countries, urban impervious surface data is needed to support governments, planners, and policymakers in managing the resulting impacts of impervious surface expansion. However, there is currently a lack of data on urban land cover change with 30 m or higher spatial resolution, broad-scale regional coverage, and consistent thematic content for Africa¹¹.

Satellite Earth observations are increasingly used to develop data products for monitoring and evaluating urban expansion at global, regional, and local scales over a range of temporal and spatial resolutions. Commonly used data sources include daytime optical imagery, nighttime lights, and synthetic aperture radar from missions such as MODIS, VIIRS, Landsat, and Sentinel 1 and 2^{12–17}. Several global data products are available, including Global Impervious Surface Area¹⁸, Global Human Settlement Layers^{19,20}, Global Man-Made Impervious Surface²¹, Global Human Built-up and Settlement Extent²², Multi-temporal Global Impervious Surfaces²³, Global 30-m Impervious Surface Map^{24,25}, and the Global Artificial Impervious Area²⁶. Although these global urban change datasets are often used for assessments at continental, national, and landscape scales, their local accuracies can be highly variable and are usually unknown⁹. Most of them are only available for a few time points separated by periods of 5–10 years, and none fully encompasses the numerous small cities that are critical components of the urban system. There are also a variety of national and local level urban land cover datasets throughout Africa, but differences in classification schemes and inconsistent processing methods hinder their usage across countries and cities^{11,27}. Many of the impacts of global change manifest at the regional level;

Department of Geography and Environmental Sustainability, University of Oklahoma, Norman, OK, 73019, USA.
 ✉e-mail: andrews.korah@ou.edu

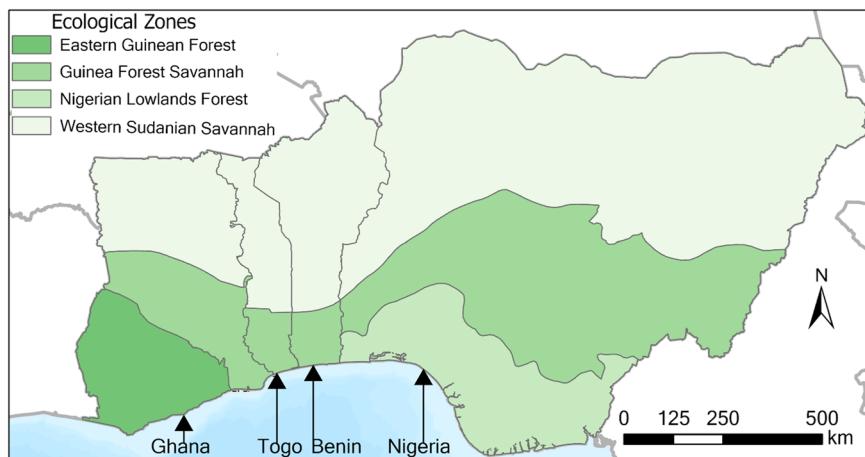


Fig. 1 Study area countries and ecological zones. These ecological zones were generated by merging smaller terrestrial ecoregions into larger ones based on proximity and similar biophysical characteristics.

therefore, planning and policy interventions need to respond at this scale⁹. Thus, regionally optimized and consistent urban cover data are needed to support large-scale research, regional planning, future projections, and assessment of urban expansion impacts across multiple study areas.

Advances in machine learning and change detection algorithms create opportunities for improved regional land use and land cover monitoring compared with traditional classifiers such as maximum likelihood, minimum distance, and k-means^{28–30}. Newer machine learning techniques such as support vector machines, neural networks, and random forests are more robust and computationally efficient, resulting in higher accuracy when used for land cover mapping^{31–33}. Additionally, time series algorithms applied at the individual pixel level, including breaks for additive season and trend (BFAST)³⁴, continuous change detection and classification (CCDC)³⁰, and Landsat-based detection of Trends and Recovery (LandTrendr)^{35–37}, can improve land cover change detection. BFAST and CCDC detect land cover change as deviations from long-term and seasonal trends using harmonic regression models, whereas LandTrendr estimates change trajectories by fitting segmented linear regression models to annual data.

Open access to the Landsat archive combined with machine learning and time series algorithms through the cloud-based Google Earth Engine (GEE) platform have revolutionized urban impervious surface cover mapping over large areas^{23,35,38,39}. Studies in North America and Asia have leveraged the GEE cloud computing platform to generate regional impervious cover datasets^{30,37,40,41}. However, there are currently no regionally calibrated and publicly available historical impervious surface data for West Africa, which is one of the fastest urbanizing regions in the world¹¹. A major reason is limited internet resources, including unstable and slow bandwidth that limits the processing of thousands of images over large areas⁴². Impervious surface data that reflect neighborhood patterns and capture annual growth are particularly needed to support future projections and regional assessments of urban change impacts on people and the environment.

We addressed this gap by developing an accurate and consistent 20-year time series of continuous impervious surface data from 2001–2020 at a 30 m grid cell size using all available Landsat images. These data covered four heavily urbanized countries in West Africa: Ghana, Togo, Benin and Nigeria. In addition, we classified developed areas based on an impervious surface threshold and generated a separate product of their change over time. We carried out a comprehensive validation of both datasets to quantify their accuracies and compare them with currently available global impervious surface datasets. Our data products have been made available as TIFF images that can be used in standard software for geospatial data processing, and the code is available for updating the datasets and refining and extending the approach to new areas.

Materials and Methods

Study area. The coverage is 1.3 million square kilometers across four West African countries, consisting of Ghana, Togo, Benin, and Nigeria. These countries contain 73% of the 165 million urban population in West Africa and are rapidly expanding their developed area, with an annual urban expansion rate of 4.70% in Nigeria, 4.46% in Ghana, 4.19% in Benin, and 3.89% in Togo from 2001–2020^{43–45}. Most (70%) of the cities in West Africa are in these four countries, with 1231 in Nigeria, 209 in Ghana, 110 in Benin, and 53 in Togo. Within these countries, natural population increase, rural–urban migrations, reclassification of rural areas into urban centers, uneven distribution of resources, and ethnic and political conditions increase demand for housing and other urban infrastructure and drive high rates of urban expansion^{46,47}.

Each country has several ecological zones with different climates, soil moisture, vegetation, and land cover patterns (Fig. 1). The Western Sudanian Savannah (WSS) mainly consists of grasslands with short, dispersed trees with 600 mm to 1000 mm precipitation annually, resulting in semi-arid conditions during the dry season. The Guinea Forest Savannah (GFS) is a transitional zone with a mixture of dense tree cover and open grassland. This zone mostly receives 1600 to 2000 mm per year. The Eastern Guinean Forest (EGF) and Nigerian Lowlands Forest (NLF) receive more precipitation, ranging from 1500 mm to 2500 mm per year, and support dense tropical forests.

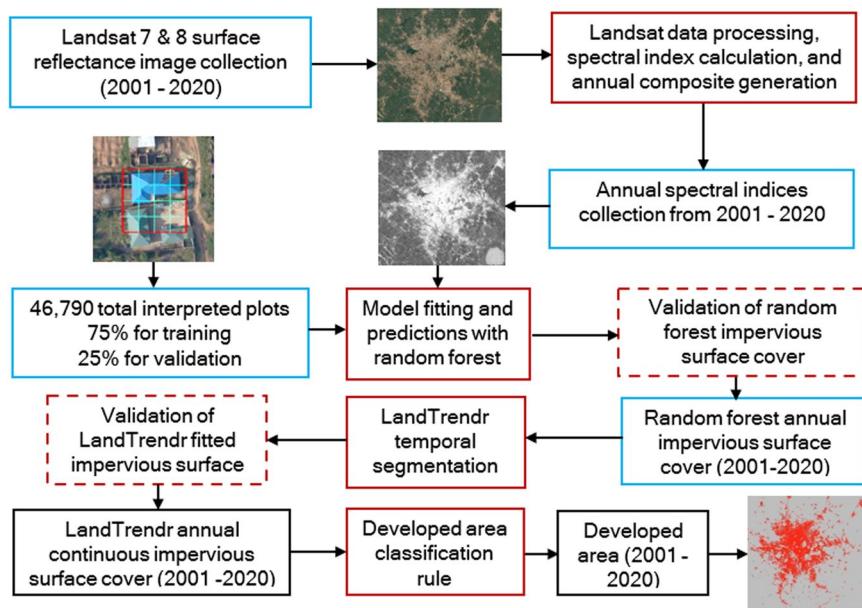


Fig. 2 Technical workflow for generating the impervious cover dataset. The data at each stage is bounded in blue, processing methods in red, validation in dashed red, and final outputs in black.

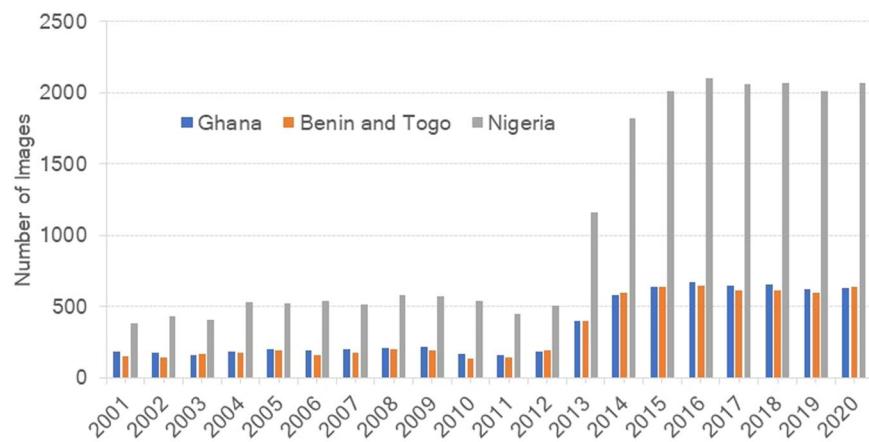


Fig. 3 Availability of Landsat 7 and 8 images from 2001–2020. Each bar shows the number of images in each year for Ghana in blue, Benin and Togo in orange and Nigeria in gray.

Technical framework. The workflow for generating the impervious cover dataset is shown in Fig. 2. Step one involves processing Landsat imagery to remove clouds and cloud shadows and generating annual composites of spectral indices from 2001–2020. Step two is training random forest models with impervious surface observations from very high-resolution images and using them to predict impervious cover with the Landsat imagery. Step three is temporally segmenting the impervious cover time series for each pixel with LandTrendr. Step four involves classifying the LandTrendr processed impervious cover into developed and undeveloped areas. The following sections describe in detail the major workflow steps.

Landsat data. We used Landsat 7 and 8 Collection 2 level-2 surface reflectance science products from 2001–2020. There was a total of 35065 images, including 7051 in Ghana, 6755 in Benin and Togo, and 21259 in Nigeria (Fig. 3). There was a mean of 850 Landsat 7 images per year from 2001–2012, and the addition of Landsat 8 increased the mean to 3110 images per year since 2013 (Fig. 3). We masked all clouds and cloud shadows using the pixel-quality layer with the C-function of the mask algorithm (CFMASK)^{48,49}.

Spectral indices and annual composites. Due to varying phenological and biophysical conditions across ecological regions, we used a combination of built-up, vegetation, burned area, and soil-based indices. Because the spectral properties of land surfaces vary geographically, we used diverse indices to minimize background noise and reduce errors. Previous comparisons of LandTrendr time series models found that using more than eight explanatory variables in land cover predictions resulted in lower errors than models based on fewer

| Spectral Index | Abbreviation | Equation | Source |
|---|--------------|---|---------------|
| Built-up area extraction index | BAEI | $baei = \frac{Red + 0.3}{Green + Swirl}$ | ⁵¹ |
| Biophysical composition index | BCI | $bci = \frac{TCB + TCW \div 2 - TCG}{TCB + TCW \div 2 + TCG}$ | ⁵² |
| Band ratio for built-up area index | BRBA | $brba = \frac{Red}{Swirl}$ | ⁵³ |
| Bare soil index | BSI | $bsi = \frac{(Red + Swirl) - (Nir + Blue)}{(Red + Swirl) + (Nir + Blue)}$ | ⁵⁹ |
| Combinational biophysical composition index | CBCI | $cbsi = 1.5 * mbsi - osavi + 0.5$ | ⁵⁴ |
| Normalized built-up area index | NBAI | $nbai = \frac{(Swirl2 - Swirl1) / Green}{(Swirl2 + Swirl1) / Green}$ | ⁵³ |
| Normalized burned ratio index | NBRI | $nbri = \frac{Nir - Swirl}{Nir + Swirl}$ | ⁶⁰ |
| Normalized difference built-up index | NDBI | $ndbi = \frac{Swirl - Nir}{Swirl + Nir}$ | ⁵⁵ |
| Normalized difference vegetation index | NDVI | $ndvi = \frac{Nir - Red}{Nir + Red}$ | ⁵⁸ |
| Urban index | UI | $uid = \frac{Swirl2 - Nir}{Swirl2 + Nir}$ | ⁵⁶ |
| Visible red-based built-up index | VRBI | $vrbi = \frac{Red - Nir}{Red + Nir}$ | ⁵⁷ |

Table 1. Spectral indices used to model impervious cover across the four countries. Tasseled cap indices, including brightness index (TCB), tasseled cap wetness index (TCW), tasseled cap greenness index (TCG), were used to generate biophysical composition index (BCI). Modified bare soil index (MBSI) and optimized soil adjusted vegetation index (OSAVI) were used to generate CBCI.

| Spectral Index | Eastern Guinea Forest/ Nigeria Lowlands Forest | Guinea Forest Savannah | Western Sudan Savannah |
|----------------|---|---------------------------|---------------------------|
| BAEI | 15 | 10 | 10 |
| BCI | 10 | 10 | 10 |
| BRBA | 5 | 5 | 5 |
| BSI | 25 | 10 | 10 |
| CBCI | 15 | 5 | 5 |
| NBAI | 5 | 5 | 5 |
| NBRI | 15 | 40 | 15 |
| NDBI | 5 | 5 | 5 |
| NDVI | 55 | 85 | 75 |
| UI | 10 | 5 | 5 |
| VRBI | 40 | 40 | 40 |

Table 2. Percentiles used to generate annual compositions by major ecological zones. Spectral index codes are described in Table 1.

explanatory variables⁵⁰. We computed 11 spectral indices (Table 1). They included 8 impervious surface indices: built-up area extraction index (BAEI)⁵¹, biophysical composition index (BCI)⁵², band ratio for built-up area index (BRBA)⁵³, combinational biophysical composition index (CBCI)⁵⁴, CBCI, normalized built-up area index (NBAI)⁵³, normalized difference built-up index (NDBI)⁵⁵, urban index (UI)⁵⁶ and visible red based built-up index (VRBI)⁵⁷. We also included the normalized difference vegetation index (NDVI)⁵⁸, the bare soil index (BSI)⁵⁹, and the normalized burn ratio index (NBRI)⁶⁰.

These indices were computed for all cloud-free pixels in all available Landsat images. Then, annual temporal metrics were generated for each index by summarizing across all cloud-free observations in a year^{61,62}. These metrics were based on percentiles ranging from 5 to 95 at 5% intervals. We manually selected the best percentile for each spectral index by systematically varying the thresholds and visually comparing the resulting metrics with high-resolution data for cities with known changes (Table 2). The spectral signature of impervious surfaces is relatively stable, whereas the spectral characteristics of vegetation and soils vary seasonally with climate. Therefore, annual composites of impervious surface indices were generated using low percentiles (mostly 5–15%) to identify pixels where these indices remained high throughout the entire year. Conversely, higher percentiles were used for NDVI to distinguish areas where greenness remained low for most of the year. The percentiles were also varied across ecoregions to account for geographic variation in the spectral characteristics of urban and non-urban areas.

Training and validation data. To ensure we had sufficient sample points in both urban and non-urban areas, we created random sample points within two strata: inside the Africapolis city boundaries⁴⁴ and outside of Africapolis city boundaries. We also used three country strata consisting of Ghana, Togo and Benin combined, and Nigeria, and four ecological zone strata. We randomly selected 300 points for each combination of the three strata, resulting in 600 points for each ecological zone within each country. In Ghana, we selected a total of 1800

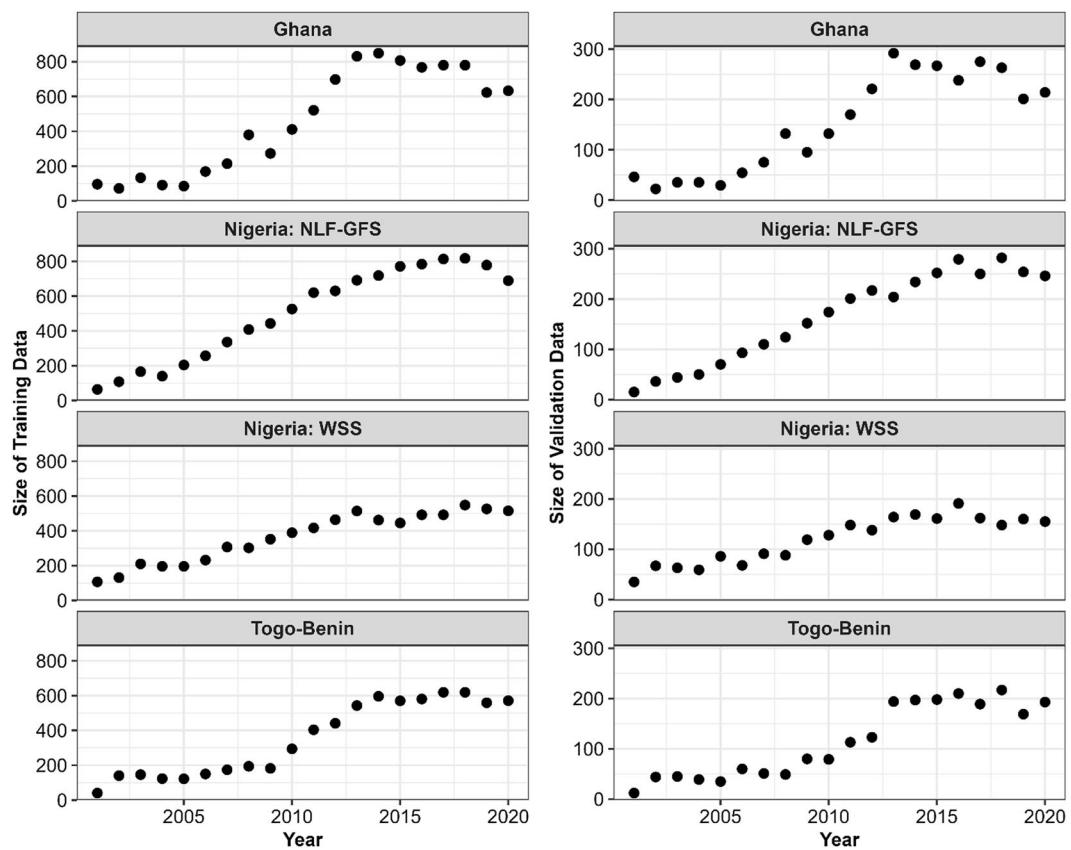


Fig. 4 Temporal distribution of training and validation plots. The black dots are the total number of plots in each year. The Nigeria Lowlands Forest and Guinea Forest Savannah is NLF-GFS, and the Western Sudanian Savannah is WSS.

| LandTrendr Parameters | Default Parameters | Study Parameters |
|------------------------|--------------------|------------------|
| maxSegments | 6 | 6 |
| spikeThreshold | 0.9 | 0.75 |
| vertexCountOvershoot | 3 | 2 |
| PreventOneYearRecovery | TRUE | TRUE |
| recoveryThreshold | 0.25 | 0.01 |
| pvalThreshold | 0.05 | 0.05 |
| bestModelProportion | 0.75 | 0.5 |
| minObservationsNeeded | 6 | 6 |

Table 3. LandTrendr segmentation parameters to process random forest impervious surface cover.

points, consisting of 900 within cities and 900 outside cities across three ecological zones. Likewise, in Benin and Togo, we selected a total of 1200 points, consisting of 600 within cities and 600 outside cities across two ecological zones. In Nigeria, we also selected 1200 total points consisting of 600 within cities and 600 outside cities across the Nigeria Lowlands Forest and Guinea Forest Savannah ecological zones. Because there were typically fewer years of historical images in the Western Sudanian Savannah ecological zone in Nigeria, we selected 350 points in each city strata, resulting in 700 points.

We generated a square of size 30 m for each sample location and overlaid a 5×5 grid of points. Each point on the grid represented a land cover of 4%. We overlaid these plots on all available very high-resolution (VHR) images in Google Earth Pro. There are geographical and temporal variations in the availability and quality of VHR images in Google Earth Pro. We collected training and validation for all years and locations where data were available and not obscured by clouds. We visually estimated the impervious cover as the number of points in each gridded polygon that covered impervious surfaces such as buildings, glass, concrete, and asphalt. The number of estimated points were multiplied by four to calculate the percent impervious cover for each plot. The total number of interpreted plots was 46,790, including 25,153 in Nigeria, 12,277 in Ghana, and 9,360 in Benin and Togo. We fitted the impervious surface models with 75% of the interpreted data and tested the accuracy

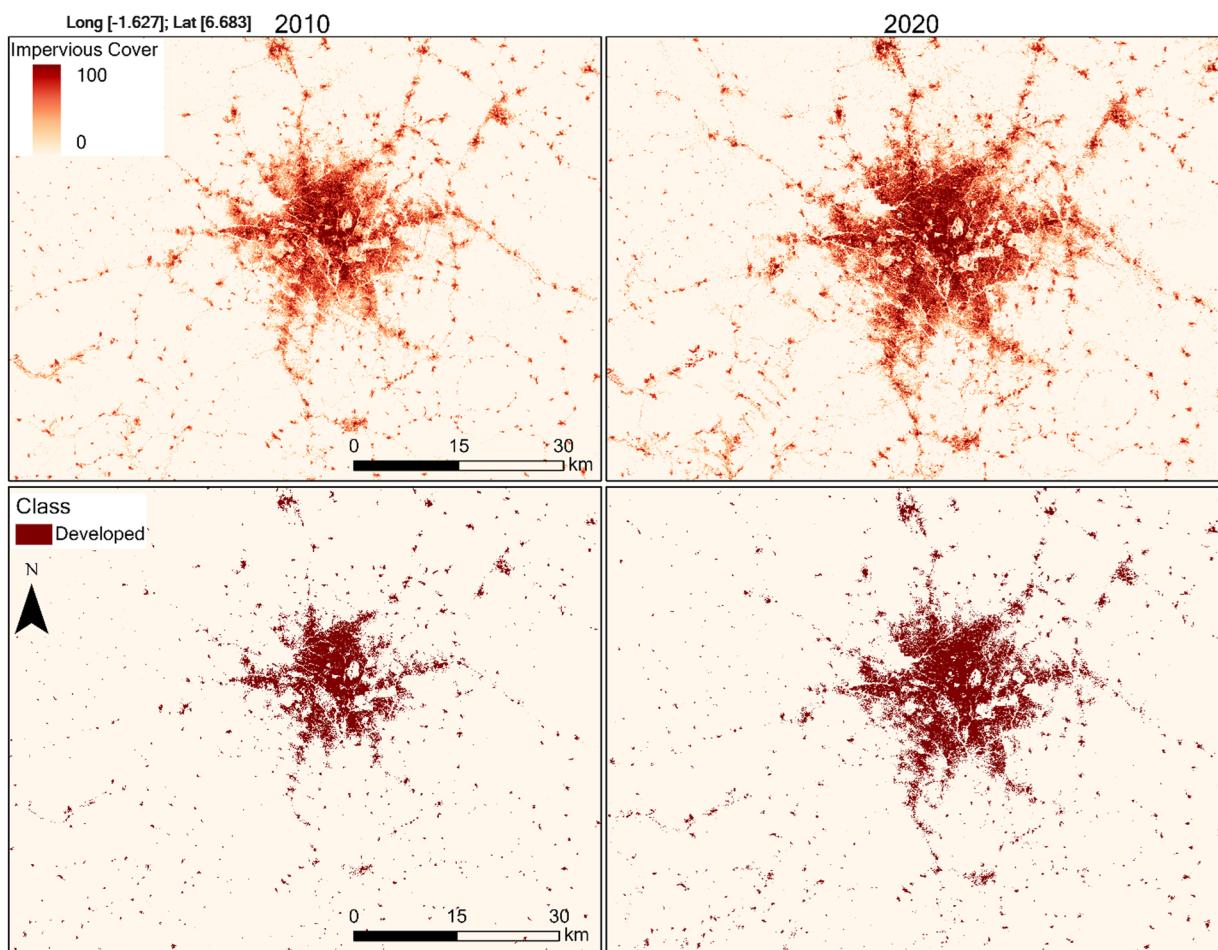


Fig. 5 Continuous impervious surface cover and developed area classification for Kumasi and surrounding smaller cities in the southern portion of Ghana. The top row shows the continuous data, and the bottom row shows the classified data.

| Data | Spatial Extent | Temporal Resolution | Spatial Resolution | Coordinate System | Format | Data Type | Pixel values |
|------------------|-----------------------------|---------------------|--------------------------------------|-----------------------|--------|------------------------|--------------------------------|
| Impervious Cover | Ghana, Togo, Benin, Nigeria | Annual | $0.00027^\circ \times 0.00027^\circ$ | WGS 1984 (ESPG: 4326) | TIFF | 8 Bit unsigned integer | 0–100 |
| Developed | Ghana, Togo, Benin, Nigeria | Annual | $0.00027^\circ \times 0.00027^\circ$ | WGS 1984 (ESPG: 4326) | TIFF | 8 Bit unsigned integer | 0: Undeveloped 1: Developed |

Table 4. Details of impervious cover and developed area data.

using the remaining 25% as an independent validation dataset. The training and validation plots were distributed across all four countries from 2001–2020 (Fig. 4).

Random forest regression. We used random forest regression to generate impervious cover using annual composites of spectral indices as predictors and percent impervious cover estimates from VHR as training data. Random forest is a machine learning algorithm that uses bootstrap sampling of the training data with random subsets of predictors to generate tree-based models and aggregates the results to generate ensemble predictions^{32,63}. This approach decreases correlations among the trees and improves accuracy of predictions. Using the 35,231 training plots, we extracted the spectral index values corresponding to each year and sampled location. To reduce computational time, we trained four separate random forest models, including one in Ghana, one in Togo and Benin together, and two in Nigeria (one for the Nigeria Lowlands Forest and Guinea Forest Savannah ecological zones and one for the Western Sudanian Savannah ecological zone). We used the GEE `smileRandom()` function to fit each model with 1000 trees, 0.65 bag fraction, 6 predictors per split, 7 minimum leaf population, and output mode set to regression. These parameters were used to predict annual continuous impervious surface cover from 2001 to 2020.

LandTrendr temporal segmentation. We applied LandTrendr to the annual impervious cover predictions from random forest to process the time series at individual pixel levels. We used the LandTrendr time series

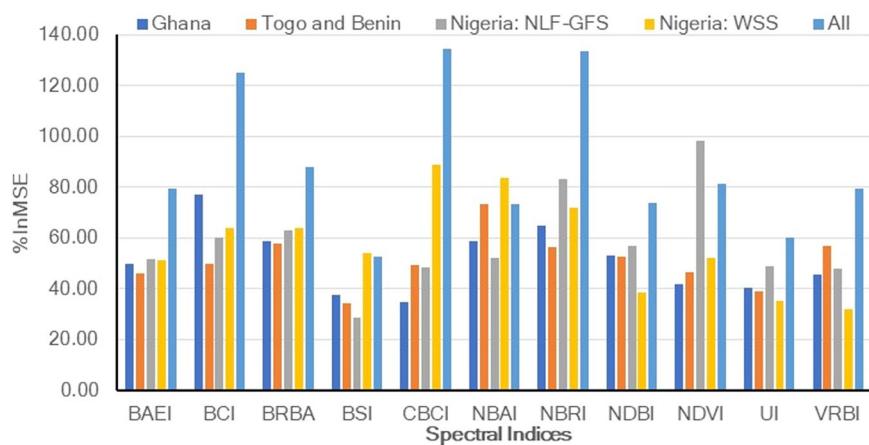


Fig. 6 Importance of spectral indices in random forest predictions using percent increase in mean squared error (%InMSE). The % increase in mean squared error represents the increase in prediction errors (out-of-bag) resulting from the permutation of a random predictor. Spectral index codes are described in Table 1.

| Country | Metric | Random Forest | Random Forest + LandTrendr |
|---|--------------------------|---------------|----------------------------|
| Ghana | Predicted-Observed r^2 | 0.67 | 0.79 |
| | Mean Error | 0.22 | -0.30 |
| | Mean Absolute Error | 8.19 | 6.73 |
| | Root Mean Squared Error | 13.40 | 11.03 |
| Togo and Benin | Predicted-Observed r^2 | 0.62 | 0.81 |
| | Mean Error | 0.38 | 0.44 |
| | Mean Absolute Error | 6.14 | 4.78 |
| | Root Mean Squared Error | 10.89 | 8.01 |
| Nigeria: Nigeria Lowlands Forest and Guinea Forest Savannah | Predicted-Observed r^2 | 0.71 | 0.81 |
| | Mean Error | 0.46 | 0.34 |
| | Mean Absolute Error | 6.59 | 4.36 |
| | Root Mean Squared Error | 12.17 | 8.05 |
| Nigeria: Western Sudan Savannah | Predicted-Observed r^2 | 0.70 | 0.82 |
| | Mean Error | 0.13 | -0.47 |
| | Mean Absolute Error | 8.62 | 6.88 |
| | Root Mean Squared Error | 14.53 | 11.65 |
| All | Predicted-Observed r^2 | 0.60 | 0.81 |
| | Mean Error | 0.28 | -0.03 |
| | Mean Absolute Error | 7.81 | 5.73 |
| | Root Mean Squared Error | 13.45 | 9.93 |

Table 5. Accuracy for Random Forest and LandTrendr fitted continuous impervious surface cover.

algorithm because it is less computationally demanding than CCDC and directly models annual change as a continuous variable, which was our desired output³⁵. Although CCDC and BFAST offer the potential for monitoring sub-annual timing of changes^{30,34}, previous studies of urban impervious surfaces show little to no variability across seasons^{64,65}. Thus, LandTrendr was better suited for our application because it uses annual composites, reducing computation time and filling data gaps.

LandTrendr uses a temporal segmentation approach that fits linear regression models to annual values for each pixel with breakpoints and segments representing change history^{35,37}. We used the fitted values from the segmented regression models generated by LandTrendr as our final estimated values of percent impervious cover. The reason for this step was to reduce noise in the random forest predictions of impervious cover and generate interpolated estimates for missing data points. Although the LandTrendr algorithm is designed to capture rapid declines in vegetation indices resulting from ecological disturbance, we adapted it to model increases in impervious cover resulting from urban expansion. We did this by reversing the numerical order of the impervious cover values so that zero represented 100% cover and hundred represented 0% cover. To accomplish this conversion, the random forest predictions of impervious cover were subtracted from 100, the LandTrendr algorithm was applied to these transformed data, and the resulting fitted values were subtracted from 100 to return them to their original order. We selected the best-fit parameters to run LandTrendr for our study area by conducting a sensitivity analysis, starting with the default parameters and systematically adjusting them by visually inspecting locations with known change histories (Table 3).

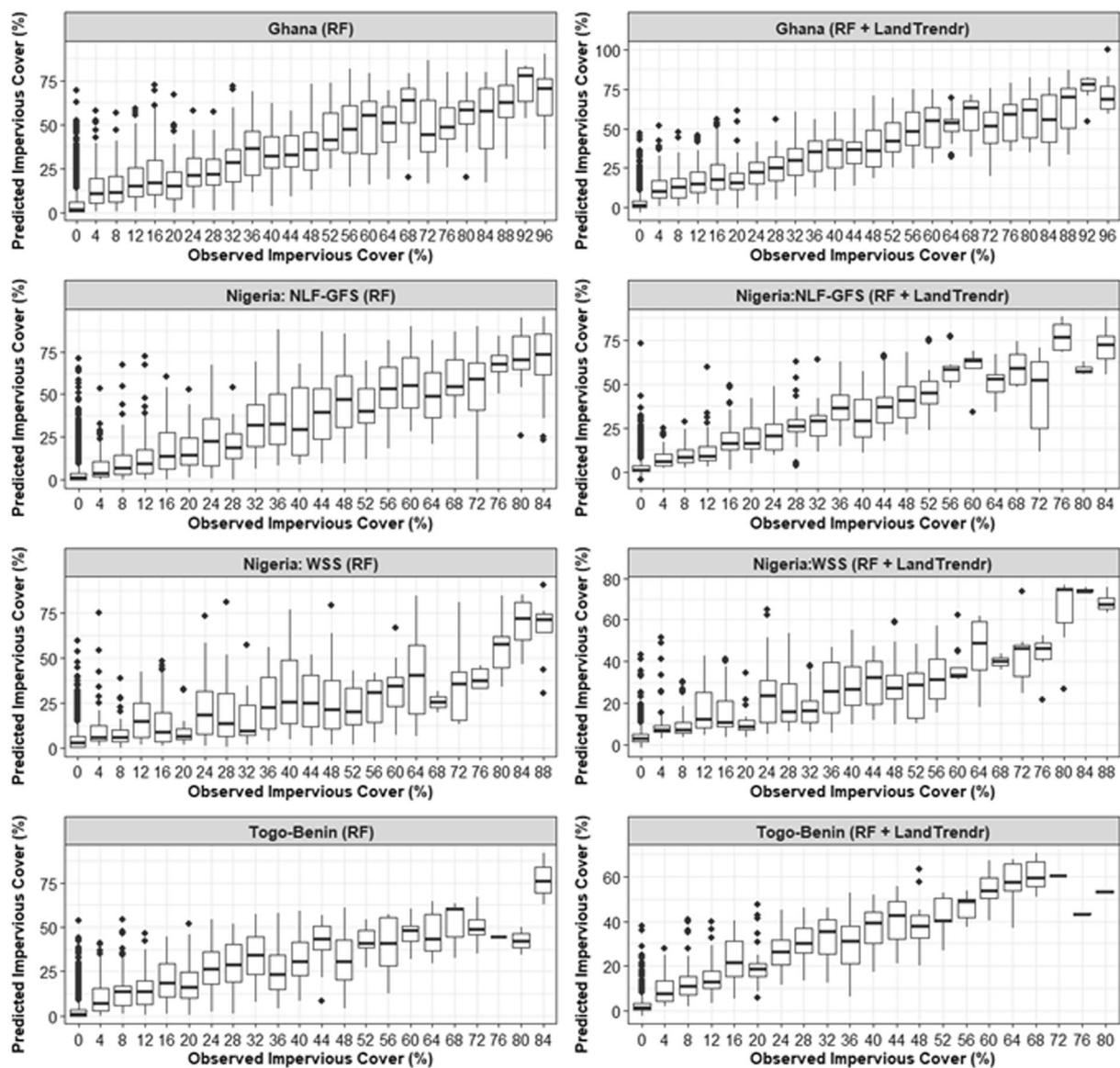


Fig. 7 Associations between observed and predicted impervious surface cover estimates from Random Forest (RF) and processed LandTrendr. The outliers are the individual points, the 25th percentile is the lower box edges, the 75th percentile is the upper box edges, and the 50% percentile is the line separating the lower and upper boxes.

Developed area classification. We classified LandTrendr fitted continuous impervious cover into developed and undeveloped. Impervious surface cover pixels values greater than 20% were classified as developed, and pixels less than or equal to 20% as undeveloped. We selected this threshold based on visual comparison of different thresholds with very high-resolution imagery. We applied a no-loss rule, which states that pixels remained developed even if the impervious surface cover for preceding years dropped below 20%. Similar rules are commonly used in urban land cover analysis to reduce false changes since impervious surfaces are mostly permanent^{40,41}. We have provided an example for the continuous and developed area classification for Kumasi, Ghana, one of the cities in the study area with rapid urban expansion (Fig. 5).

Model validation. We used 11,559 independent observations to evaluate impervious surface cover predictions from random forest and LandTrendr segmentation. We measured the association between observations and predictions using the predicted-observed r^2 . We measured prediction bias with mean error and quantified differences between observations and predictions using mean absolute error and root mean squared error. For the classification of developed areas, we computed a standard confusion matrix with overall accuracy, user's and producer's accuracies, and commission and omission errors.

Additionally, we used validation data for 2001–2018 to compare the accuracy of WADISC with existing global datasets with similar temporal and spatial resolution, consisting of global artificial impervious area (GAIA)²⁶ and global impervious surface area (GISA)⁶⁶. GAIA and GISA have binary classes consisting of impervious and non-impervious at 30 m resolution. GISA is available from 1972–2019, and GAIA

| | Observation | | | | | |
|------------|--|-------------|-------------|-----------|--------------|--------------|
| | Country | Class | Undeveloped | Developed | User | Commission |
| Prediction | Ghana | Undeveloped | 2210 | 154 | 0.93.49 | 6.51 |
| | | Developed | 165 | 534 | 76.39 | 23.61 |
| | | Producer | 93.05 | 77.62 | 89.59 | |
| | | Omission | 6.95 | 22.38 | | 10.41 |
| | Togo and Benin | Undeveloped | 1868 | 63 | 96.74 | 3.26 |
| | | Developed | 84 | 282 | 77.05 | 22.95 |
| | | Producer | 95.70 | 81.74 | 93.60 | |
| | | Omission | 4.30 | 18.26 | | 6.40 |
| | Nigeria: Nigeria Lowlands & Guinea Forest Savannah | Undeveloped | 2910 | 74 | 97.52 | 2.48 |
| | | Developed | 79 | 396 | 83.37 | 16.63 |
| | | Producer | 97.36 | 84.26 | 95.58 | |
| | | Omission | 2.64 | 15.74 | | 4.42 |
| | Nigeria: Western Sudan Savannah | Undeveloped | 2175 | 134 | 94.20 | 5.80 |
| | | Developed | 73 | 398 | 84.50 | 15.50 |
| | | Producer | 96.75 | 74.81 | 92.55 | |
| | | Omission | 3.25 | 25.19 | | 7.45 |
| | All | Undeveloped | 9163 | 425 | 95.57 | 4.43 |
| | | Developed | 401 | 1610 | 80.06 | 19.94 |
| | | Producer | 95.81 | 79.12 | 92.88 | |
| | | Omission | 4.19 | 20.88 | | 7.12 |

Table 6. Confusion matrix of developed and undeveloped areas classifications from 2001–2020. The overall accuracy of the data is in bold, and the misclassified error is in bold italics.

| Observation | | | | | | | |
|-------------|---------|-------------|-------------|-----------|--------------|--------------|---------|
| | Dataset | Class | Undeveloped | Developed | User | Commission | F-score |
| Prediction | WADISC | Undeveloped | 7909 | 318 | 96.13 | 3.87 | |
| | | Developed | 361 | 1336 | 78.73 | 21.27 | 0.7974 |
| | | Producer | 95.63 | 80.77 | 93.16 | | |
| | | Omission | 4.37 | 19.23 | | 6.84 | |
| | GISA | Undeveloped | 7714 | 412 | 94.93 | 5.07 | |
| | | Developed | 556 | 1242 | 69.08 | 30.92 | 0.7196 |
| | | Producer | 93.28 | 75.09 | 90.25 | | |
| | | Omission | 6.72 | 24.91 | | 9.75 | |
| | GAIA | Undeveloped | 7728 | 522 | 93.67 | 6.33 | |
| | | Developed | 542 | 1132 | 67.62 | 32.38 | 0.6803 |
| | | Producer | 93.45 | 68.44 | 89.28 | | |
| | | Omission | 6.55 | 31.56 | | 10.72 | |

Table 7. Accuracy comparison with global impervious datasets from 2001–2018. The overall accuracy of the data is in bold, and the misclassified error is in bold italics.

has annual data from 1985–2018. We computed the average accuracies and errors for only these eighteen years because these are when data were available in all the impervious datasets we examined. We also compared the total developed area extents from these datasets across the four ecological zones: Eastern Guinea Forest (EGF), Guinea Forest Savannah (GFS), Nigeria Lowlands Forest (NLF) and Western Sudanian Savannah (WSS).

Data Records

The West Africa Dataset of Impervious Surface Change (WADISC) comprises 30-m resolution continuous impervious cover and classified developed area datasets and is publicly accessible through Figshare: <https://doi.org/10.6084/m9.figshare.24716481.v3>⁶⁷. The dataset consists of 40 Tiff files, including 20 continuous and 20 binary classifications, each containing developed area information across the four countries. The continuous impervious cover pixel values are percentages from 0 to 100, and the classified developed area values are 1 and 0, indicating presence or absence (Table 4).

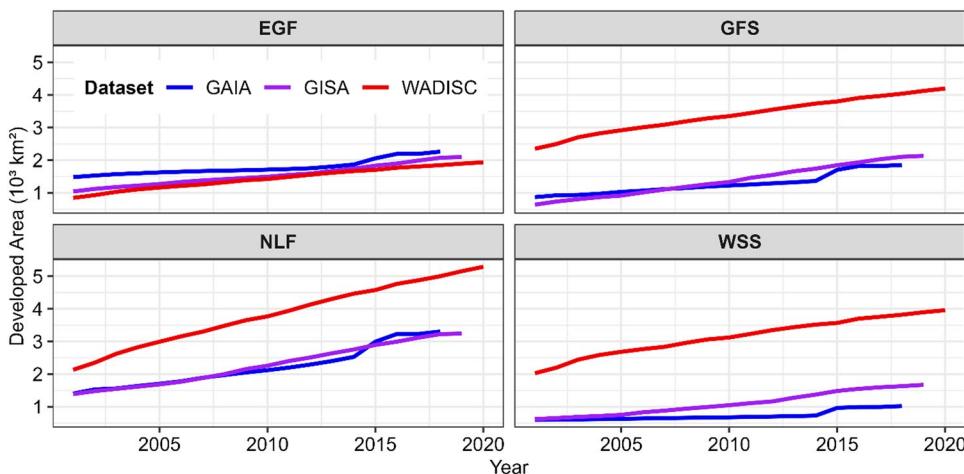


Fig. 8 Developed area of the GAIA and GISA datasets compared to WADISC across ecological zones in the study area. The red line shows the WADISC developed area from 2001–2020, the blue line shows the GAIA developed area from 2001–2018, and the purple line shows the GISA developed area from 2001–2019. The ecological zones are Eastern Guinea Forest (EGF), Guinea Forest Savannah (GFS), Nigeria Lowlands Forest (NLF), and Western Sudanian Savannah (WSS).

Technical Validation

We used all 11 spectral indices to generate random forest impervious cover estimates because models using all the predictors had higher accuracies than those based on subsets. This finding was consistent with previous assessments of land cover change models that found models with more predictors had lower errors⁵⁰. Across the four random forest models employed in this study, the relative importance of spectral indices varied, with BCI the most important predictor of impervious surface cover in Ghana, NBAI in Benin and Togo, NDVI in Nigeria Lowlands Forest and Guinea Forest Savannah in Nigeria, and CBCI in Western Sudanian Savannah in Nigeria (Fig. 6). BSI had relatively low importance, suggesting that the urban indices effectively discriminated impervious surfaces from soils.

Using LandTrendr for temporal segmentation of the random forest predictions of impervious cover resulted in higher accuracies than predictions based on random forests alone (Table 5). The high predicted-observed r^2 values indicated a strong linear association between the predictions and observations of impervious cover in the validation dataset. The strength of this association can also be seen in scatterplots of the predicted-observed relationship (Fig. 7). The magnitude of the mean error was close to zero when summarized across the entire study area, indicating that the predictions were unbiased. The overall mean absolute error was less than 6% cover and the root mean squared error was less than 10% cover, giving us confidence that the predictions can effectively distinguish areas with high versus low impervious cover.

The overall accuracy of the developed area classification was high, with a value of 93% across the entire study area, and 90% in Ghana, 94% in Togo-Benin, 96% in Nigeria Lowlands Forest and Guinea Forest Savannah, and 93% in Nigeria Western Sudan Savannah (Table 6). Class-level accuracy statistics were generally higher for non-developed areas than for developed areas. However, the overall producer's and user's accuracies for developed areas were still relatively high at ~80%. The errors of omission and commission had similar magnitudes for each class, indicating that there was not a strong bias toward overpredicting or underpredicting developed areas.

Comparison with global impervious data. The average overall accuracy from 2001 to 2018 was highest in WADISC with 93.16%, whereas GISA was 90.25%, and GAIA was 89.28% (Table 7). Also, the omission error, representing the probability of misclassifying developed areas in the reference data, was 19.23% in WADISC but increased to 24.91% in GISA and 31.56% in GAIA. The commission error, indicating the probability of misclassifying undeveloped areas, was 21.27% in WADISC but increased considerably to 30.92% in GISA and 32.38 in GAIA. In addition, the F-score representing the harmonic mean between the producer and user's accuracy was highest for WADISC followed by GISA and GAIA.

The developed area comparison across the various datasets shows that GISA and GAIA were similar, and WADISC was higher (Fig. 8). Examples of developed area in primary cities with over a million population and secondary cities with between ten thousand and one million population in different ecological zones are shown in Fig. 9. Developed area extent in all three datasets were similar in the Eastern Guinea Forest (EGF). The developed area for WADISC, however, was higher than GAIA and GISA in the Guinea Forest Savannah (GFS), Nigeria Lowlands Forest (NLF), and Western Sudanian Savannah (WSS). Our findings are consistent with previous studies that indicate most existing global datasets underestimate urban extent in the study area^{1,11}.

The global datasets we examined have longer temporal extents, allowing for historical developed area analysis as far back as 1972 for GAIA and 1985 for GISA. Although WADISC covers only four countries, it encompasses the majority of cities in West Africa, and it is the first annual regional impervious cover dataset with higher accuracies and lower errors than existing global datasets. The accuracy of WADISC was also high even in the Western Sudanian Savannah ecological zone with semi-arid conditions where urban areas are often difficult to classify with remote sensing data¹. The WADISC calibration accounted for sub-regional ecological differences in

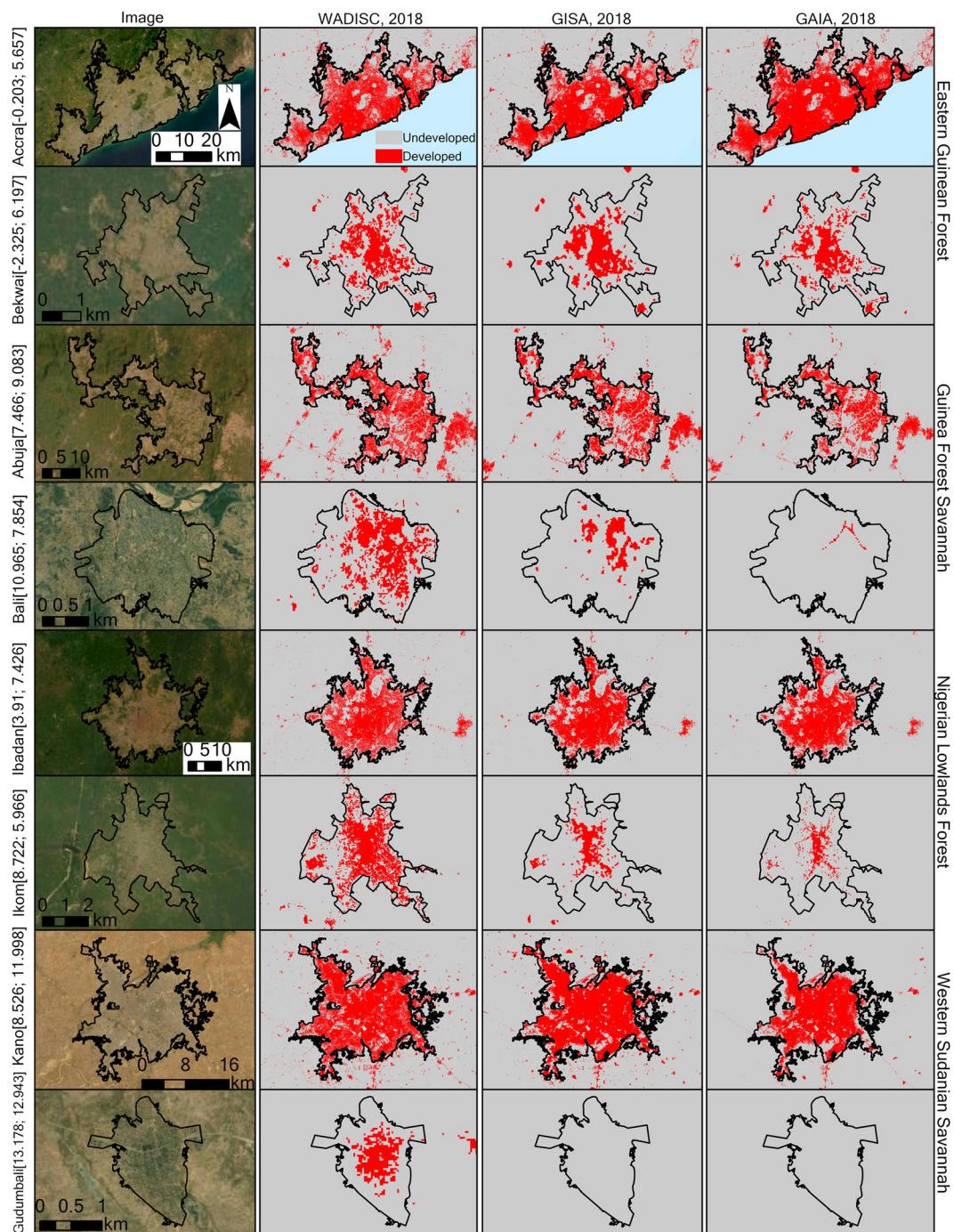


Fig. 9 Comparison across cities in different ecological zones. The primary cities with over one million urban populations are Accra, Abuja, Ibadan, and Kano, and the smaller cities with between ten thousand and one million population are Bekwai, Bali, Ikom, and Gudumbali. The numbers in parenthesis are the longitudes and latitudes.

urban morphology and background conditions, resulting in higher accuracy and better detection of developed areas than the global datasets (Fig. 9). Additionally, the WADISC data captured smaller settlements that are often excluded or misclassified in global datasets (Fig. 9).

Usage Notes

Change analysis. We demonstrated the potential uses of this data by quantifying and comparing the rates of developed area changes across countries and comparing the patterns of developed area changes for selected cities. The total developed area in each country more than doubled between 2001 and 2020, with a 2.2-fold increase in Benin, 2.3 in Ghana, 2.4 in Nigeria, and 2.1 in Togo (Table 8). Thus, Nigeria had the highest annual increase in developed areas, followed by Ghana, Benin, and Togo.

| Country | 2001 | 2020 | Change |
|---------|-----------------------------------|-----------------------------------|---------------------------|
| | Developed Area (km ²) | Developed Area (km ²) | Annual Expansion rate (%) |
| Benin | 217.52 | 474.33 | 4.19 |
| Ghana | 690.75 | 1583.14 | 4.46 |
| Nigeria | 2931.72 | 7012.58 | 4.70 |
| Togo | 160.75 | 331.77 | 3.89 |
| All | 4001.73 | 9401.81 | 4.60 |

Table 8. Total developed area and percentage annual expansion rate from 2001–2020.

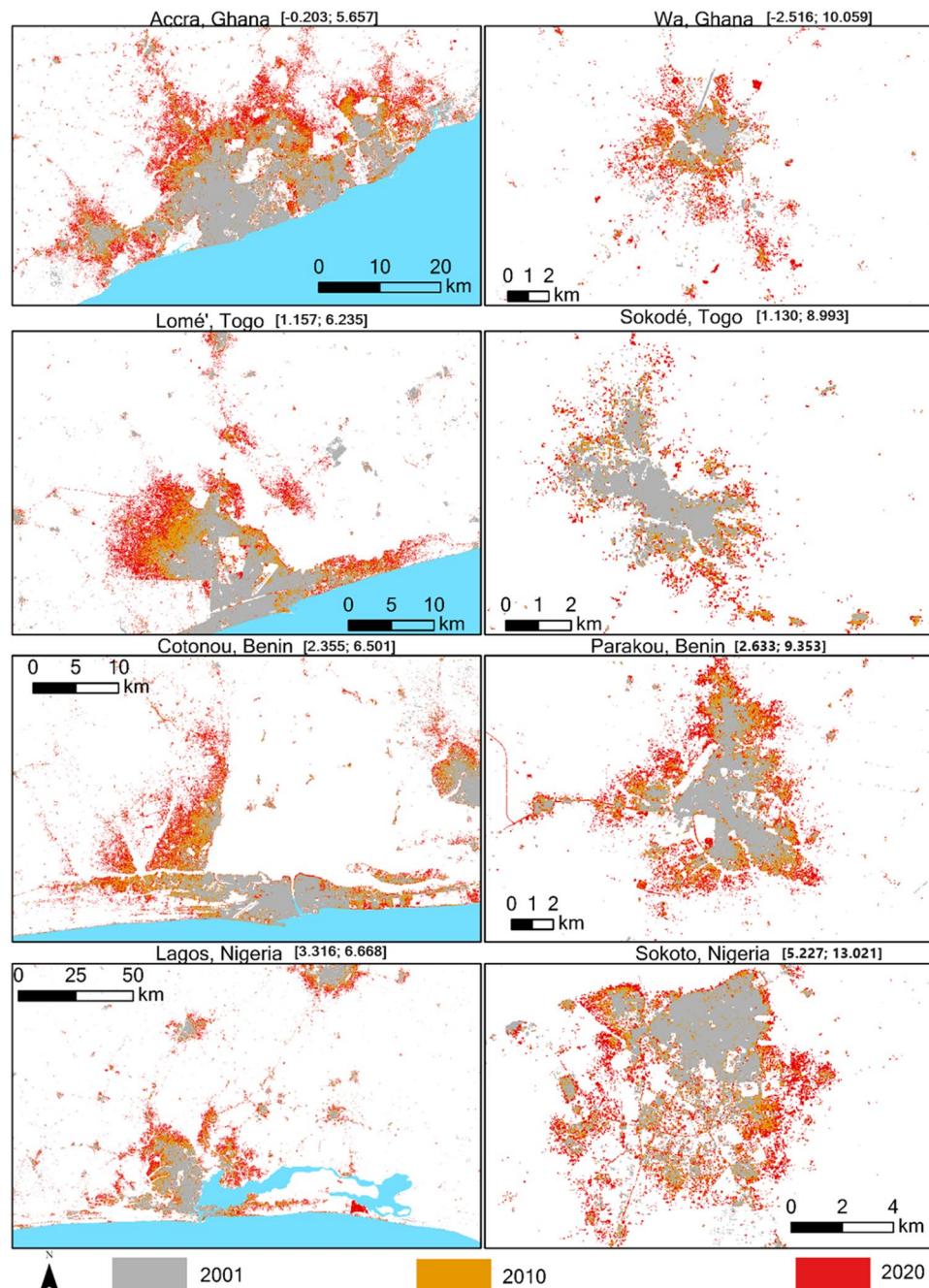


Fig. 10 Change in developed area for subset of cities in each country for three-time points. The gray shows the initially developed area in 2001, the orange shows the developed area in 2010, and the red shows the developed area in 2020. The numbers in parenthesis are the longitudes and latitudes of each city, respectively.

Maps of the developed area extent for the largest primary city and secondary city each in Ghana, Togo, Benin, and Nigeria with rapid urban growth shows that new developed areas rapidly filled open spaces and expanded outwards from the initial developed area (Fig. 10). A recent study used WADISC and quantified the expansion patterns across cities of varying population sizes, and found cities mostly sprawled faster than infilled, especially in smaller secondary cities⁴³. Different growth patterns are influenced by differences in national urban development plans, rates of urbanization, land tenure systems, and foreign direct investments. The sprawling outward expansion and high expansion rates are challenging to manage due to limited resources. Although urban expansion occurs locally, the impacts are much broader, meaning regional-level analysis can better inform sustainable development decisions.

WADISC can support research and assessment of urban expansion impact on heat stress^{68,69}, urban malaria incidence^{70–72}, greenhouse gas emission^{73,74}, urban floods⁷⁵, cropland loss^{76,77}, biodiversity loss and habitat fragmentation^{2,8}, within and across countries. For example, Van Vliet (2019) found urban expansion resulted in 122 million tons decrease in cereal production per year from 1992–2015. Also, urban areas are susceptible to increased risk of heat stress in the Sydney region⁶⁸. These studies mostly found that urban expansion negatively impacts people and the environment. However, some studies found that compact urban expansion had a negative association with air pollution⁷⁸ and energy and resource usage⁷⁹ but was positively associated with economic productivity⁸⁰. Thus, WADISC can support diverse research and policy assessments across multiple scales in Ghana, Togo, Benin, and Nigeria.

Potential limitations. When classifying change in developed area, we applied a rule of no-developed area loss, which states that developed pixels are irreversible. This is a standard approach in land cover change analysis and had a relatively minor effect on the predictions of developed areas and their accuracy. There were only a few cities where this rule had a noticeable effect, correcting for false transitions of developed to non-developed areas that arose from noise in the spectral data. However, there are other areas, particularly in the northern portions of the study area, where climate fluctuations or conflict may result in abandonment of developed areas with a resulting transition to non-developed.

Additionally, all Landsat 7 data collected after May 31, 2003 have data gaps resulting from the scan line corrector (SLC) off. The scan line corrector removes the zigzag motion produced by across and along track movement of the sensor. Despite the SLC failure, these data are still geometrically and radiometrically correct and are measured at the same precision as prior to the anomaly. Although most SLC gaps are filled in the process of generating the annual temporal metrics, the process sometimes results in different values in the SLC gaps, particularly between 2003 and 2013 when only Landsat 7 images are available. These spatial anomalies in the temporal metrics influence the patterns of developed pixels in some locations. However, the overall accuracy of the continuous impervious surface estimates and developed area classification is still high.

We have provided access to the underlying continuous impervious surface data as well as the classified developed area data⁶⁷. These data can be used to generate developed area classifications using different impervious surface thresholds and to analyze change using different types of transition rules. All land cover datasets contain some error, but our rigorous quantitative validation combined with visual assessment of the predicted changes indicate that the predictions of impervious surfaces and developed areas accurately capture their patterns and magnitudes of change. We have provided access to the underlying code for generating the data products in Google Earth Engine and GitHub, and users are thus able to conduct more detailed evaluations of the underlying techniques and potentially update and extend the methods to improve their accuracy.

Code availability

The scripts used to generate the impervious surface cover data are available and accessible in figshare⁶⁷. The scripts are available in Google Earth Engine: https://code.earthengine.google.com/?accept_repo=users/korahandrews/WADISC, and on GitHub through this link: <https://github.com/Kora0003/WADISC>.

Received: 15 December 2023; Accepted: 5 July 2024;

Published online: 18 July 2024

References

1. Adhikari, P. & de Beurs, K. M. Growth in urban extent and allometric analysis of West African cities. *J. Land Use Sci.* **12**, 105–124 (2017).
2. Güneralp, B., Lwasa, S., Masundire, H., Parnell, S. & Seto, K. C. Urbanization in Africa: challenges and opportunities for conservation. *Environ. Res. Lett.* **13**, 015002 (2017).
3. Meng, L., Sun, Y. & Zhao, S. Comparing the spatial and temporal dynamics of urban expansion in Guangzhou and Shenzhen from 1975 to 2015: A case study of pioneer cities in China's rapid urbanization. *Land Use Policy* **97**, 104753 (2020).
4. OECD, United Nations Economic Commission for Africa, & African Development Bank. *Africa's Urbanisation Dynamics 2022: The Economic Power of Africa's Cities*. <https://doi.org/10.1787/3834ed5b-en> (OECD, 2022).
5. Seto, K., Fragkias, M., Güneralp, B. & Reilly, M. A Meta-Analysis of Global Urban Land Expansion. *PLOS ONE* **6**, (2011).
6. He, C., Zhang, D., Huang, Q. & Zhao, Y. Assessing the potential impacts of urban expansion on regional carbon storage by linking the LUSD-urban and InVEST models. *Environ. Model. Softw.* **75**, 44–58 (2016).
7. Liu, X. *et al.* Impacts of Urban Expansion on Terrestrial Carbon Storage in China. *Environ. Sci. Technol.* **53**, 6834–6844 (2019).
8. Seto, K. C., Güneralp, B. & Hutyra, L. R. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proc. Natl. Acad. Sci.* **109**, 16083–16088 (2012).
9. Tulbure, M. G., Hostert, P., Kuemmerle, T. & Broich, M. Regional matters: On the usefulness of regional land-cover datasets in times of global change. *Remote Sens. Ecol. Conserv.* **8**, 272–283 (2022).
10. Yao, R., Cao, J., Wang, L., Zhang, W. & Wu, X. Urbanization effects on vegetation cover in major African cities during 2001–2017. *Int. J. Appl. Earth Obs. Geoinformation* **75**, 44–53 (2019).
11. Korah, A., Koch, J. A. M. & Wimberly, M. C. Understanding urban growth modeling in Africa: Dynamics, drivers, and challenges. *Cities* **146**, 104734 (2024).

12. Bilintoh, T. M., Korah, A., Opuni, A. & Akansobe, A. Comparing the Trajectory of Urban Impervious Surface in Two Cities: The Case of Accra and Kumasi. *Ghana. Land* **12**, 927 (2023).
13. He, T. *et al.* Global 30 meters spatiotemporal 3D urban expansion dataset from 1990 to 2010. *Sci. Data* **10**, 321 (2023).
14. Ju, Y., Dronova, I. & Delclòs-Alió, X. A 10 m resolution urban green space map for major Latin American cities from Sentinel-2 remote sensing images and OpenStreetMap. *Sci. Data* **9**, 586 (2022).
15. Song, C., Jiang, X., Fan, C. & Li, L. High-resolution circa-2020 map of urban lakes in China. *Sci. Data* **9**, 747 (2022).
16. Sun, G. *et al.* A 10-m resolution impervious surface area map for the greater Mekong subregion from remote sensing images. *Sci. Data* **10**, 607 (2023).
17. Zhao, C., Cao, X., Chen, X. & Cui, X. A consistent and corrected nighttime light dataset (CCNL 1992–2013) from DMSP-OLS data. *Sci. Data* **9**, 424 (2022).
18. Elvidge, C. *et al.* Global Distribution and Density of Constructed Impervious Surfaces. *Sensors* **7**, 1962–1979 (2007).
19. Pesaresi, M. *et al.* GHS-BUILT R2015B - GHS built-up grid, derived from Landsat, multitemporal (1975, 1990, 2000, 2014). (2015).
20. Pesaresi, M., Melchorri, M., Siragusa, A. & Kemper, T. Atlas of the Human Planet - Mapping Human Presence on Earth with the Global Human Settlement Layer. *JRC Publications Repository* <https://publications.jrc.ec.europa.eu/repository/handle/JRC103150>, <https://doi.org/10.2788/582834> (2016).
21. Brown de Colstoun, E. C. *et al.* Global Man-made Impervious Surface (GMIS) Dataset From Landsat. (2017).
22. Wang, P., Huang, C., Brown de Colstoun, E. C., Tilton, J. C. & Tan, B. Global Human Built-up And Settlement Extent (HBASE) Dataset From Landsat. (2017).
23. Liu, X. *et al.* High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform. *Remote Sens. Environ.* **209**, 227–239 (2018).
24. Chen, J. *et al.* Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* **103**, 7–27 (2015).
25. Xiao, Z. & Liangyun, L. Development of a global 30-m impervious surface map using multi-source and multi-temporal remote sensing datasets with the Google Earth Engine platform. *Zenodo* <https://doi.org/10.5281/zenodo.3505079> (2019).
26. Gong, P. *et al.* Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. *Remote Sens. Environ.* **236**, 111510 (2020).
27. Güneralp, B., Reba, M., Hales, B. U., Wentz, E. A. & Seto, K. C. Trends in urban land expansion, density, and land transitions from 1970 to 2010: a global synthesis. *Environ. Res. Lett.* **15**, 044015 (2020).
28. Yu, L. *et al.* Meta-discoveries from a synthesis of satellite-based land-cover mapping research. *Int. J. Remote Sens.* **35**, 4573–4588 (2014).
29. Yu, Z. *et al.* Land Use and Land Cover Classification for Bangladesh 2005 on Google Earth Engine. in *2018 7th International Conference on Agro-geoinformatics (Agro-geoinformatics)* 1–5, <https://doi.org/10.1109/Agro-Geoinformatics.2018.8475976> (IEEE, Hangzhou, 2018).
30. Zhu, Z. & Woodcock, C. E. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.* **144**, 152–171 (2014).
31. Amini, S., Saber, M., Rabiei-Dastjerdi, H. & Homayouni, S. Urban Land Use and Land Cover Change Analysis Using Random Forest Classification of Landsat Time Series. *Remote Sens.* **14**, 2654 (2022).
32. Belgui, M. & Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **114**, 24–31 (2016).
33. Maxwell, A. E., Warner, T. A. & Fang, F. Implementation of machine-learning classification in remote sensing: an applied review. *Int. J. Remote Sens.* **39**, 2784–2817 (2018).
34. Verbesselt, J., Hyndman, R., Newnham, G. & Culvenor, D. Detecting trend and seasonal changes in satellite image time series. *Remote Sens. Environ.* **114**, 106–115 (2010).
35. Kennedy, R. *et al.* Implementation of the LandTrendr Algorithm on Google Earth Engine. *Remote Sens.* **10**, 691 (2018).
36. Kennedy, R. E., Yang, Z. & Cohen, W. B. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation algorithms. *Remote Sens. Environ.* **114**, 2897–2910 (2010).
37. Kennedy, R. E. *et al.* Spatial and temporal patterns of forest disturbance and regrowth within the area of the Northwest Forest Plan. *Remote Sens. Environ.* **122**, 117–133 (2012).
38. Arévalo, P., Bullock, E. L., Woodcock, C. E. & Olofsson, P. A Suite of Tools for Continuous Land Change Monitoring in Google Earth Engine. *Front. Clim.* **2**, 576740 (2020).
39. Mugiraneza, T., Nascetti, A. & Ban, Y. Continuous Monitoring of Urban Land Cover Change Trajectories with Landsat Time Series and LandTrendr-Google Earth Engine Cloud Computing. *Remote Sens.* **12**, 2883 (2020).
40. Ding, Q., Shao, Z., Huang, X., Altan, O. & Hu, B. Time-series land cover mapping and urban expansion analysis using OpenStreetMap data and remote sensing big data: A case study of Guangdong-Hong Kong-Macao Greater Bay Area, China. *Int. J. Appl. Earth Obs. Geoinformation* **113**, 103001 (2022).
41. Lin, Y., Zhang, H., Lin, H., Gamba, P. E. & Liu, X. Incorporating synthetic aperture radar and optical images to investigate the annual dynamics of anthropogenic impervious surface at large scale. *Remote Sens. Environ.* **242**, 111757 (2020).
42. Wimberly, M. C. *et al.* Historical trends of degradation, loss, and recovery in the tropical forest reserves of Ghana. *Int. J. Digit. Earth* **15**, 30–51 (2022).
43. Korah, A. & Wimberly, M. C. Smaller cities have large impacts on West Africa's expanding urban system. *Sustain. Cities Soc.* **106**, 105381 (2024).
44. OECD/SWAC. *Africa's Urbanization Dynamics 2020: Africapolis, Mapping a New Urban Geography*. West African Studies. (OECD Publishing, Paris, 2020).
45. Walther, O. J. *Urbanisation and Demography in North and West Africa, 1950-2020*. vol. 33, https://www.oecd-ilibrary.org/development/urbanisation-and-demography-in-north-and-west-africa-1950-2020_4fa52e9c-en (2021).
46. Mlambo, V. H. An overview of rural-urban migration in South Africa: its causes and implications. *Arch. Bus. Res.* **6**, (2018).
47. Zimmer, A. *et al.* Dynamics of population growth in secondary cities across southern Africa. *Landsc. Ecol.* **35**, 2501–2516 (2020).
48. Foga, S. *et al.* Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sens. Environ.* **194**, 379–390 (2017).
49. Stillinger, T., Roberts, D. A., Collar, N. M. & Dozier, J. Cloud Masking for Landsat 8 and MODIS Terra Over Snow-Covered Terrain: Error Analysis and Spectral Similarity Between Snow and Cloud. *Water Resour. Res.* **55**, 6169–6184 (2019).
50. Cohen, W. B., Yang, Z., Healey, S. P., Kennedy, R. E. & Gorelick, N. A LandTrendr multispectral ensemble for forest disturbance detection. *Remote Sens. Environ.* **205**, 131–140 (2018).
51. Bouzekri, S., Lasbet, A. A. & Lachehab, A. A New Spectral Index for Extraction of Built-Up Area Using Landsat-8 Data. *J. Indian Soc. Remote Sens.* **43**, 867–873 (2015).
52. Deng, C. & Wu, C. BCI: A biophysical composition index for remote sensing of urban environments. *Remote Sens. Environ.* **127**, 247–259 (2012).
53. Waqar, M. M., Mirza, J. F., Mumtaz, R. & Hussain, E. Development of New Indices for Extraction of Built-Up Area & Bare Soil from Landsat Data. **1**, 6 (2012).

54. Zhang, S., Yang, K., Li, M., Ma, Y. & Sun, M. Combinational Biophysical Composition Index (CBCI) for Effective Mapping Biophysical Composition in Urban Areas. *IEEE Access* **6**, 41224–41237 (2018).
55. Zha, Y., Gao, J. & Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* **24**, 583–594 (2003).
56. Kawamura, M., Jayamanna, S. & Tsujiko, Y. Quantitative evaluation of urbanization in developing countries using satellite data. *Doboku Gakkai Ronbunshu* **1997**, 45–54 (1997).
57. Estoque, R. C. & Murayama, Y. Classification and change detection of built-up lands from Landsat-7 ETM+ and Landsat-8 OLI/TIRS imageries: A comparative assessment of various spectral indices. *Ecol. Indic.* **56**, 205–217 (2015).
58. Tucker, C. J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **8**, 127–150 (1979).
59. Rikimaru, A., Roy, P. S. & Miyatake, S. Tropical forest cover density mapping. *Tropical Ecology* **43**, 39–49 (2002).
60. García, M. J. L. & Caselles, V. Mapping burns and natural reforestation using thematic Mapper data. *Geocarto Int.* **6**, 31–37 (1991).
61. Hansen, M. C. *et al.* Monitoring conterminous United States (CONUS) land cover change with Web-Enabled Landsat Data (WELD). *Remote Sens. Environ.* **140**, 466–484 (2014).
62. Zhu, Z. Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. *ISPRS J. Photogramm. Remote Sens.* **130**, 370–384 (2017).
63. Breiman, L. Machine Learning, Volume 45, Number 1 - SpringerLink. *Mach. Learn.* **45**, 5–32 (2001).
64. Schneider, A. Monitoring land cover change in urban and peri-urban areas using dense time stacks of Landsat satellite data and a data mining approach. *Remote Sens. Environ.* **124**, 689–704 (2012).
65. Zhang, L., Zhang, M. & Yao, Y. Mapping seasonal impervious surface dynamics in Wuhan urban agglomeration, China from 2000 to 2016. *Int. J. Appl. Earth Obs. Geoinformation* **70**, 51–61 (2018).
66. Huang, X. *et al.* 30 m global impervious surface area dynamics and urban expansion pattern observed by Landsat satellites: from 1972 to 2019. *Zenodo* <https://doi.org/10.5281/zenodo.5136330> (2021).
67. Korah, A. & Wimberly, M. WADISC: Annual Impervious Surface Data for Ghana, Togo, Benin, and Nigeria from 2001–2020. *figshare* <https://doi.org/10.6084/m9.figshare.24716481.v3> (2024).
68. Argüeso, D., Evans, J. P., Pitman, A. J. & Di Luca, A. Effects of City Expansion on Heat Stress under Climate Change Conditions. *PLOS ONE* **10**, e0117066 (2015).
69. Cao, Q., Yu, D., Georgescu, M., Wu, J. & Wang, W. Impacts of future urban expansion on summer climate and heat-related human health in eastern China. *Environ. Int.* **112**, 134–146 (2018).
70. Neiderud, C.-J. How urbanization affects the epidemiology of emerging infectious diseases. *Infect. Ecol. Epidemiol.* **5**, 27060 (2015).
71. Tatem, A. J., Gething, P. W., Smith, D. L. & Hay, S. I. Urbanization and the global malaria recession. *Malar. J.* **12**, 133 (2013).
72. Tekla, H. *et al.* Trend analysis of malaria in urban settings in Ethiopia from 2014 to 2019. *Malar. J.* **22**, 235 (2023).
73. Luqman, M., Rayner, P. J. & Gurney, K. R. On the impact of urbanisation on CO₂ emissions. *Npj Urban Sustain.* **3**, 6 (2023).
74. Mitchell, L. E. *et al.* A multi-city urban atmospheric greenhouse gas measurement data synthesis. *Sci. Data* **9**, 361 (2022).
75. Feng, B., Zhang, Y. & Bourke, R. Urbanization impacts on flood risks based on urban growth data and coupled flood models. *Nat. Hazards* **106**, 613–627 (2021).
76. Tu, Y. *et al.* How does urban expansion interact with cropland loss? A comparison of 14 Chinese cities from 1980 to 2015. *Landscape Ecol.* **36**, 243–263 (2021).
77. Van Vliet, J. Direct and indirect loss of natural area from urban expansion. *Nat. Sustain.* **2**, 755–763 (2019).
78. Frank, L. D. & Engelke, P. Multiple Impacts of the Built Environment on Public Health: Walkable Places and the Exposure to Air Pollution. *Int. Reg. Sci. Rev.* **28**, 193–216 (2005).
79. Creutzig, F. *et al.* Urban infrastructure choices structure climate solutions. *Nat. Clim. Change* **6**, 1054–1056 (2016).
80. Vernon Henderson, J. Understanding knowledge spillovers. *Reg. Sci. Urban Econ.* **37**, 497–508 (2007).

Acknowledgements

We thank Ecological and Geospatial Research Applications in Planetary Health (EcoGRAPH) lab members for their feedback on the original version. Michael Glessner contributed to the manual interpretation of the training and validation data. We also thank Dr. Dawn Nekorchuk for contributing to the development of the Google Earth Engine code. Financial support for publication was provided by the OU Libraries' Open Access Fund.

Author contributions

Andrews Korah created and interpreted the validation and training data, implemented and modified the code, and drafted and revised the manuscript. Michael C. Wimberly developed the original code, provided funding, and reviewed and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.