

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

Modelling and mapping the intra-urban spatial distribution of *Plasmodium falciparum* parasite rate using very-high-resolution satellite derived indicators

Georganos, Stefanos; Brousse, Oscar; Dujardin, Sébastien; Linard, Catherine; Casey, Daniel; Milliones, Marco; Parmentier, Benoit; van Lipzig, Nicole P.M.; Demuzere, Matthias; Grippa, Tais; Vanhuysse, Sabine; Mboga, Nicholas; Andreo, Verónica; Snow, Robert W.; Lennert, Moritz

Published in:

International Journal of Health Geographics

DOI:

[10.1186/s12942-020-00232-2](https://doi.org/10.1186/s12942-020-00232-2)

Publication date:

2020

Document Version

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (HARVARD):

Georganos, S, Brousse, O, Dujardin, S, Linard, C, Casey, D, Milliones, M, Parmentier, B, van Lipzig, NPM, Demuzere, M, Grippa, T, Vanhuysse, S, Mboga, N, Andreo, V, Snow, RW & Lennert, M 2020, 'Modelling and mapping the intra-urban spatial distribution of *Plasmodium falciparum* parasite rate using very-high-resolution satellite derived indicators', *International Journal of Health Geographics*, vol. 19, no. 1, 38, pp. 38. <https://doi.org/10.1186/s12942-020-00232-2>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy


If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

RESEARCH

Open Access



Modelling and mapping the intra-urban spatial distribution of *Plasmodium falciparum* parasite rate using very-high-resolution satellite derived indicators

Stefanos Georganos^{1*} , Oscar Brousse², Sébastien Dujardin^{3,4}, Catherine Linard^{3,4}, Daniel Casey⁵, Marco Milliones⁶, Benoit Parmentier^{5,6}, Nicole P. M. van Lipzig², Matthias Demuzere⁷, Tais Grippa¹, Sabine Vanhuyse¹, Nicholus Mboga¹, Verónica Andreo^{8,9}, Robert W. Snow^{10,11} and Moritz Lennert¹

Abstract

Background: The rapid and often uncontrolled rural–urban migration in Sub-Saharan Africa is transforming urban landscapes expected to provide shelter for more than 50% of Africa's population by 2030. Consequently, the burden of malaria is increasingly affecting the urban population, while socio-economic inequalities within the urban settings are intensified. Few studies, relying mostly on moderate to high resolution datasets and standard predictive variables such as building and vegetation density, have tackled the topic of modeling intra-urban malaria at the city extent. In this research, we investigate the contribution of very-high-resolution satellite-derived land-use, land-cover and population information for modeling the spatial distribution of urban malaria prevalence across large spatial extents. As case studies, we apply our methods to two Sub-Saharan African cities, Kampala and Dar es Salaam.

Methods: Openly accessible land-cover, land-use, population and OpenStreetMap data were employed to spatially model *Plasmodium falciparum* parasite rate standardized to the age group 2–10 years (PfPR_{2–10}) in the two cities through the use of a Random Forest (RF) regressor. The RF models integrated physical and socio-economic information to predict PfPR_{2–10} across the urban landscape. Intra-urban population distribution maps were used to adjust the estimates according to the underlying population.

Results: The results suggest that the spatial distribution of PfPR_{2–10} in both cities is diverse and highly variable across the urban fabric. Dense informal settlements exhibit a positive relationship with PfPR_{2–10} and hotspots of malaria prevalence were found near suitable vector breeding sites such as wetlands, marshes and riparian vegetation. In both cities, there is a clear separation of higher risk in informal settlements and lower risk in the more affluent neighborhoods. Additionally, areas associated with urban agriculture exhibit higher malaria prevalence values.

Conclusions: The outcome of this research highlights that populations living in informal settlements show higher malaria prevalence compared to those in planned residential neighborhoods. This is due to (i) increased human exposure to vectors, (ii) increased vector density and (iii) a reduced capacity to cope with malaria burden. Since informal settlements are rapidly expanding every year and often house large parts of the urban population, this emphasizes

*Correspondence: sgeorgan@ulb.ac.be

¹ Department of Geoscience, Environment & Society, Université Libre de Bruxelles, 1050 Brussels, Belgium

Full list of author information is available at the end of the article



© The Author(s) 2020. 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/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

the need for systematic and consistent malaria surveys in such areas. Finally, this study demonstrates the importance of remote sensing as an epidemiological tool for mapping urban malaria variations at large spatial extents, and for promoting evidence-based policy making and control efforts.

Keywords: Urban malaria, Random forest, Kampala, Dar es Salaam, Remote sensing, Population

Introduction

Unprecedented rates of rural–urban migration and natural population increase in sub-Saharan Africa (SSA) have dramatically affected urban environments [1]. Low income housing has not kept up with population growth which has contributed to widely varying physical and socio-economic landscapes within cities where formal and informal settlements coexist [2]. Informal settlements are often characterized by residential areas where land tenure is not recognized by authorities, housing quality is sub-standard and access to several basic services is lacking [3–5]. These rapidly transforming environments, have an impact upon urban health, such as the risk of infection with vector-borne diseases [6, 7].

While malaria has widely been known as a rural disease, uncontrolled urbanisation has altered urban landscapes in ways that may increasingly support vector breeding, making the disease to be persistent in urban settings [7–9]. One reason for this is the increased likelihood of breeding sites for mosquitoes of the genus *Anopheles*, the vectors of the *Plasmodium falciparum* parasite [10]. Previous work has highlighted the focal nature of urban malaria and its link with human activities. For example, the development of urban agricultural areas, irrigation schemes, market gardens, open water storage, or even open excavation during the construction of building sites and roads have led to rain-fed breeding sites associated with increased malaria prevalence [7–9, 11–16]. Furthermore, the functional organization of cities can influence the heterogeneity of urban malaria risk. Areas with peripheral housing settlements and a central business district may exhibit different malaria patterns compared to those with business districts located on the periphery and housing located centrally.

The social vulnerability of a population, which can vary spatially, is also expected to affect the ability of a population to cope with the burden of malaria [17]. Previous work has shown that malaria prevalence can be significantly higher in informal settlements than in other urban landscapes due to poor housing infrastructure, lack of bed nets, and inadequate financial resources to buy anti-malarial drugs, among others [18, 19]. Hence, given the same levels of vector density, two communities with significantly different levels of income and education might have significantly

different prevalence levels of malaria. These variations, often dominated by differences between planned and unplanned settlements, might explain the clustered nature of urban malaria in SSA cities [20]. Indeed, as noted by Taubenböck [21], the physical urban surface reflects the underlying social processes that developed it. It would be reasonable to assume that spatial malaria models capturing a combination of the physical surface and the population's socio-economic levels are more informative than those only relying on a purely physical representation of the land cover.

Satellite images can contribute a vast amount of information for modeling and mapping malaria prevalence at the city scale. The intra-urban component of malaria dynamics, however, has neither been part of continental malaria risk mapping initiatives nor considered part of most national control strategies [22–25]. Satellite imagery can provide valuable input for epidemiological models such as detailed land-cover (LC), land-use (LU) maps, as well as socioeconomic indicators and population distribution maps. Recent research has demonstrated that several moderate or high-resolution geospatial and/or satellite-derived features could distinguish areas of higher malaria risk within the urban settings of Dar es Salaam, Tanzania [6, 11], Ouagadougou, Burkina Faso [26] and Dakar, Senegal [9], often due to the differences in building density and vegetation type coverage.

In this paper, we model and map the spatial distribution of malaria prevalence across two SSA cities—Dar es Salaam and Kampala, using very-high-resolution (VHR) satellite indicators and machine learning techniques. We investigate the use of VHR LC and LU classes as a composite of both physical and socio-economic information to predict malaria prevalence. We then examine the use of human population products to adjust our estimates according to the underlying population. Our research objectives can be summarized as follows:

- 1) Assessing the potential of VHR satellite-derived LC, LU and population information for modeling urban malaria prevalence.
- 2) Exploring their utility as tools to map and highlight intra-urban variations of malaria prevalence across large extents.

Methods

Case studies

Dar es Salaam, Tanzania

Dar Es Salaam is the former capital of Tanzania with an estimated population exceeding five million, and one of the fastest growing cities in the world [27]. According to recent estimates, 75% of the residential population lives in informal settlements where only a small part of the urban fabric is planned [28–30]. Malaria is endemic in the city, with over one million cases reported annually [6, 11, 31]. The Urban Malaria Control Project (UMCP) has been responsible for a large part of the efforts to control malaria transmission in the city through ground-based sampling and monitoring [6, 14, 19, 32, 33]. While entomological inoculation rates in urban areas can in general be considered lower than rural regions, this might not hold true in urban slums, as these are usually built around environments that favor mosquito vector breeding [19]. The dominant vector species in Dar es Salaam, come from the *Anopheles gambiae* complex with a smaller contribution coming from *Anopheles funestus* [19, 34]. *Anopheles gambiae* are usually found in small bodies of freshwater while *Anopheles funestus* is frequently encountered in permanent water bodies such as wetlands and marshes [19]. Additionally, *Anopheles arabiensis* of the *gambiae* complex are increasingly feeding outdoors, an adaptation to the high levels of bed nets usage and house protection [32, 35]. Analyzing samples from health care facilities in and around Dar es Salaam, Wang et al. [36] found that the differences of malaria prevalence between the urban–rural spectrum in the city were low. In addition, Kabaria et al. [6] identified increased risk across riparian vegetation and wetlands in the city.

Kampala, Uganda

Kampala is the capital and main economic center of Uganda, with a population of over 1.5 million [37]. Similar to Dar es Salaam, Kampala is growing rapidly every year at a rate of about 5% [38]. Informal settlements house roughly 60% of the urban residents [39, 40]. Malaria is endemic in the region, and previous research has noted significant spatial variations within the city associated with the residential characteristics of sampled locations such as the water sources utilized by a household [41]. The majority of the malaria vectors in Kampala are of the *Anopheles gambiae* complex, with a smaller population of *Anopheles funestus* vectors [42]. Informal settlements have been built in vicinity of marshes, streams and swamps because of their high viability for urban agriculture [43]. This influences the intra-urban spatial heterogeneity and is consistently implicated in increased malaria prevalence in these areas [44, 45]. In

addition, the sanitation conditions of the city's slums tend to deteriorate during the peak of the wet season in which malaria transmission intensifies [46]. In a study by Mukasa [18], it was shown that about 45% of the interviewed mothers from the Bwaese slum in Kampala, were not in possession of a bed net indicating high inability to cope with the burden of malaria.

Satellite derived indicators

Land-cover (LC)

The LC maps (50 cm resolution) used in this study were produced through a combination of Computer Assisted Photo Interpretation, Geographic Object Based Image Analysis GEOBIA and machine learning algorithms through open-access software ([47–49]; Additional file 1), and are openly accessible through the Zenodo scientific repository [50, 51]. The classifications were based on Pleiades satellite imagery of Kampala (collected in February 2013) and Dar es Salaam (stereo-images collected in March and January 2016 and July 2018). The LC of Kampala exhibited an overall accuracy of 86% (7 classes), while that of Dar es Salaam, an overall accuracy of 90% (9 classes). The accuracy metrics were the result of an assessment through out of bag error of a random forest (RF) classifier trained and validated at the date of acquisition. The building class in Dar es Salaam LC was further subdivided into three height subclasses due to the availability of photogrammetrically generated height elevation data [52]. The complete LC legend is shown in Figs. 1 and 2.

Land-use (LU)

The LU maps utilized here were the outcome of a processing chain involving the computation of LC-based spatial metrics derived from the maps mentioned above and information derived from OpenStreetMaps (OSM) [53]. Linear elements extracted from OSM such as the street network and various parcel types were imported into a PostGIS database and processed to create street-block polygons. Subsequently, a machine learning classifier assigned a LU value to each street-block through supervised training. The complete processing chain was a reproduction of the work by Grippa et al. ([54]; Additional file 1). The LU classification allowed for the classification of the urban surface according to its different urban functions (i.e., residential/non-residential). The residential classes were categorized into either formal or informal settlements. The complete LU legend is shown in Figs. 1 and 2.

Population density

We used high-resolution population maps (100-m resolution) that were constructed through population

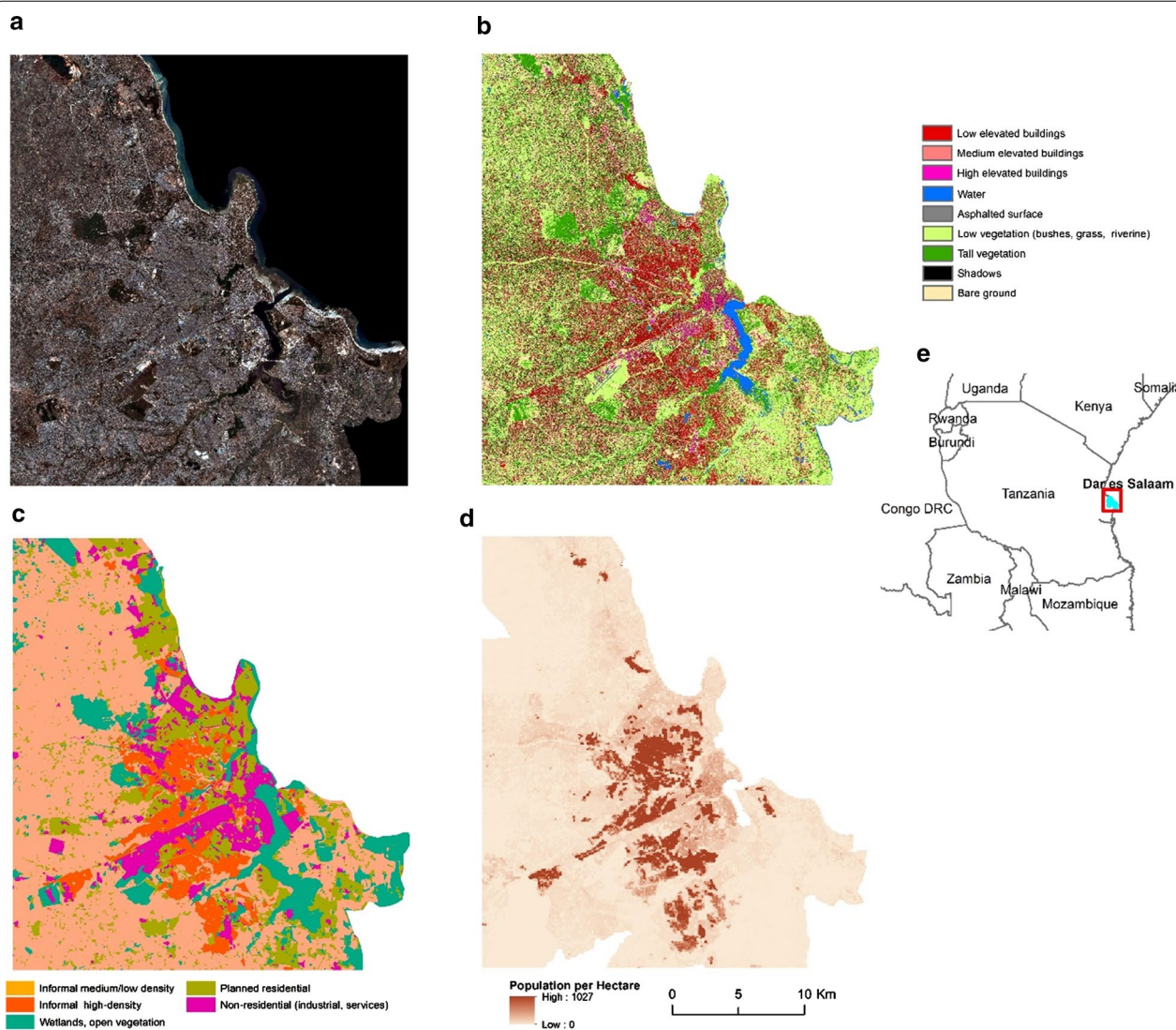


Fig. 1 **a** Pleiades satellite imagery of Dar es Salaam—RGB natural color composite, **b** land cover (0.5 m resolution), **c** land use at the street block level, **d** population counts per hectare and **e** location map of Dar es Salaam within Tanzania

disaggregation algorithms trained on census data, and using the LC and LU data as input [55, 56] and Additional file 1). The validation of the population models demonstrated R^2 values of 0.63 and 0.77 for Dar es Salaam and Kampala, respectively, which is in line with state-of-the-art results of similar studies [55–57].

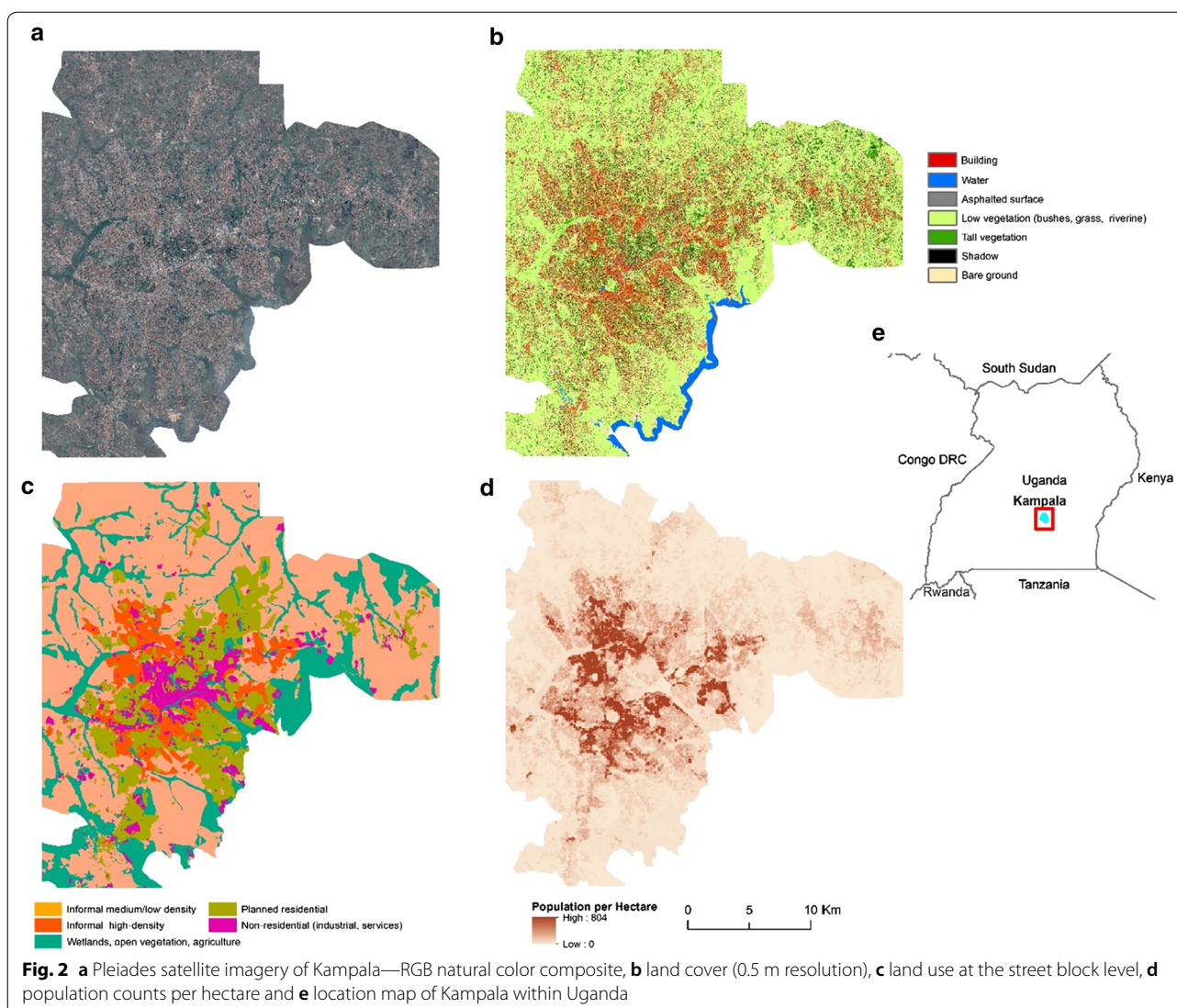
Ancillary data

To complement the previous datasets, we extracted the Normalized Difference Vegetation Index (NDVI) from the raw VHR satellite images, and terrain height information (30 m resolution) from the NASA/NGA Shuttle Radar Topography Mission (SRTM) [58]. In the case of Dar es Salaam, OSM vector features such as wetlands,

streams and rivers were used due to the high level of detailed information available for the city, in a large part thanks to community mapping projects. For instance, ‘Ramani Huria’, one of the largest community projects in Dar es Salaam aiming to mitigate hazard and flooding risk, has mapped detailed urban infrastructure such as the drainage network and buildings for more than 4 million people in the city since 2018 [29]. Table 1 summarizes the complete set of variables examined.

Plasmodium falciparum prevalence data

Data of community surveys were extracted from an open access online database [59] that accompanied the publication of changing malaria prevalence across



sub-Saharan Africa since 1900 [60]. From the available pool of surveys, we included those that have high degrees of spatial accuracy of the survey location (GPS coordinates or Google Earth validation) and consistent metadata information. Because different surveys often cover different age ranges, each parasite rate was standardized to the age group 2–10 (PfPR_{2-10}) [6, 61, 62]. As mentioned by Smith et al. [63], PfPR_{2-10} combines reliable epidemiological and statistical properties beneficial for multi-survey comparison and analysis. The temporal range of selected samples was 2005–2014, resulting in 39 surveys (at 38 unique locations) for Kampala and 90 surveys (at 57 unique locations) for Dar es Salaam (Fig. 3). In Dar es Salaam, 27 (30%) school surveys were included, undertaken in 2014. In Kampala, 21 school surveys (54%) were included, with 20 of them undertaken in 2014.

The average survey sample size in Kampala and Dar es Salaam, is 82 and 175, among individuals aged 0–16 years old, respectively. All surveys were random selections of communities or schools. Further information regarding the key characteristics of the malaria dataset can be found in Additional file 2. The mean PfPR_{2-10} values were 6.76% and 7.76% for Kampala and Dar es Salaam, respectively. 17.6% of the data points reported zero PfPR_{2-10} . Finally, we used 1-kilometer buffers around each geolocalized survey to extract aggregated values for each predictor mentioned previously in Table 1 (i.e. proportions for categorical features, mean values for continuous ones and the mean distance to the “Wetland”, “River” and “Stream” classes), similar to research employing survey data and geographical variables [6, 64]. Even though there was a temporal mismatch between the satellite imagery and the

Table 1 Predictive variables investigated in each city

	Kampala	Dar es Salaam	Type
Land cover			
Low vegetation (humid, riparian, grasses, bushes)	X	X	Proportion
Tall vegetation	X	X	Proportion
Bare ground (or dried out vegetation)	X	X	Proportion
Water	X	X	Proportion
Building	X		Proportion
Low elevated building		X	Proportion
Medium elevated building		X	Proportion
High elevated building		X	Proportion
Land use			
Planned residential	X	X	Proportion
Informal residential – high density	X	X	Proportion
Informal residential – medium/low density	X	X	Proportion
Wetlands, streams, marshes, rivers (mixed class)	X		Distance
OSM			
Wetlands		X	Distance
Rivers		X	Distance
Streams		X	Distance
Ancillary features			
Population per hectare	X	X	Average
Normalized difference vegetation index	X	X	Average
Terrain elevation	X	X	Average

malaria data, we presume a degree of stationarity across the main urban extent as most of the LU changes in SSA cities are characterized mostly by expansion rather than transition, and that the malaria data are likely to be representative of land use/ecology in the period of the satellite imagery, as done in similar studies [6]. Moreover, the collapse of the temporal dimension was inevitable due to the limited sample size and was the only means to increase spatial coverage.

Modelling and quality assessment methods

To train the $PfPR_{2-10}$ models, we used a random forest (RF) regressor which has shown to be resilient to overfitting, capturing non-linear relationships and appropriate for heavily contextual models [65]. RF is an ensemble of regression decision trees, trained on random data bootstraps (bagging). In a standard RF configuration, each computed tree is trained using a random sub-sample of about 70% of the initial data. The average prediction of all computed trees is used as the final output [65]. The hyperparameters that require fine-tuning in RF are (i) the number of considered features for each decision split in each tree (feature bagging) and (ii) the total number of decision trees built. In this study, the former was determined through cross validation (value of 1), while the latter was set at a computationally efficient number (1000) through

the R software's caret package [66]. To create the final predictive models, we employed feature selection methods, namely the Variable Selection with Random Forests (VSURF) algorithm [67]. VSURF is a well-documented and robust variable selection procedure that uses iterative and nested RF models to identify variables contributing to the task at hand eliminating useless, noisy and/or redundant features. By creating more parsimonious models, the model performance may increase as several studies have shown [68, 69].

For the evaluation metrics, we reported the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and the Predicted Coefficient of Determination (Predicted R^2). Given the relatively small sample sizes, and to reduce the prediction bias, we made use of a bootstrap approach [70]. The train and test data were split in an 80:20 ratio using stratified random sampling through 100 simulations and reported on the average RMSE, MAE and R^2 values. The sampling was stratified by the survey type (i.e., samples from a particular study) to make sure that the training and testing data distributions were similar. Finally, the variable importance and respective partial dependency plots for the most important variables in each model were extracted and visualized. In RF regression, the most common way to extract the variable importance is by the increase in Mean Squared Error

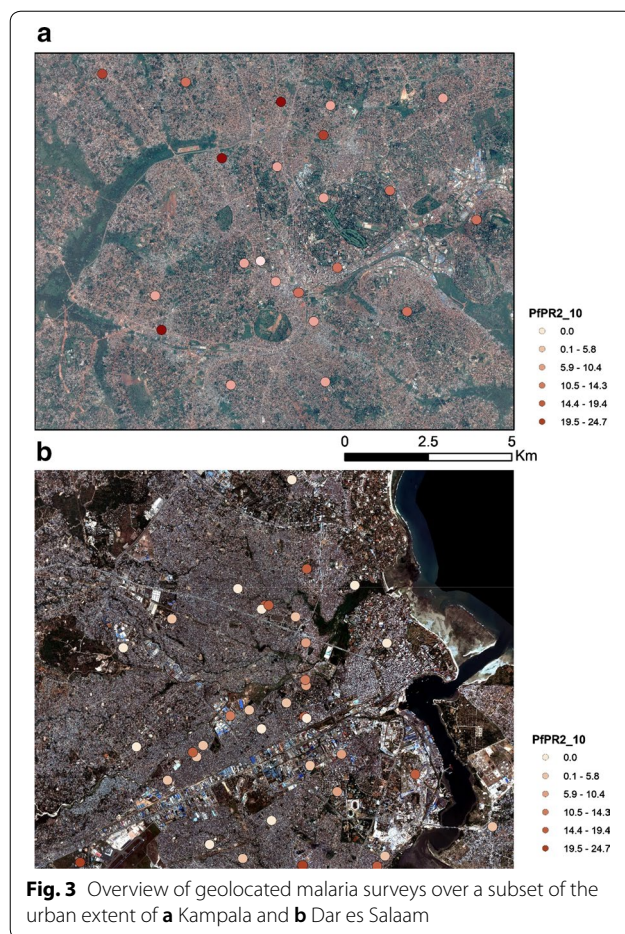


Fig. 3 Overview of geolocated malaria surveys over a subset of the urban extent of **a** Kampala and **b** Dar es Salaam

(iMSE). To compute the iMSE for a given feature, its values are randomly permuted and the internal RF performance metric, the Out of Bag (OOB) error is computed. Important variables are expected to significantly decrease model performance if permuted, reporting high values of iMSE [65, 71, 72]. For prediction, we used a 100-m grid resolution with variables aggregated at that level. Even though higher resolutions have been used such as 10 meters [6], 100 m was a reasonable scale for mapping intra-urban PfPR₂₋₁₀, capturing neighborhood variability and also matching with the spatial data used (i.e., population and land-use at the street block level).

To gain a deeper understanding of the urban malaria prevalence, we adjusted our predicted estimates according to the underlying population based on the population distribution maps. In each of the cities, we multiplied the predicted PfPR₂₋₁₀ with the population of each grid cell to obtain the predicted number of infected people. Afterwards, we summarized this information at the administrative level that the population model was trained at. Finally, we computed population adjusted PfPR₂₋₁₀ estimates, by dividing the total predicted number of infected

people aggregated of a census unit with its total population as in previous works [73–77]. The census delineation for Dar es Salaam was an aggregated version of the administrative level 5 of the 2002 census, and was computed through k-means cluster analysis [55]. This was done to find suitable training formations for the population distribution models resolution and as such, these units do not represent official administrative levels. For Kampala, we made use of the 2002 level 4 census delineation.

Hardware

The data processing and model training was performed with two Intel® Xeon® CPU E5-2690 (2 processors of 2.90 GHz, 16 cores and 32 processing threads) having 96 GB of RAM.

Results

Variable selection and importance

Using the results of the VSURF as an anchor, we filtered out variables that had minimal or zero influence for the task of predicting PfPR₂₋₁₀. Notably, in the case of Dar Es Salaam all initial features were kept, while in Kampala three features were dropped (SRTM, NDVI and the proportion of “water” class coming from the LC map). This could be explained both by the fact that the lower thematic detail of the Kampala LC information coupled with minimal coverage of inland water in the imagery. Figure 4 presents the variables used and their importance derived from the RF regressor for Dar es Salaam and Kampala. To account for uncertainty, we extracted the average importance over one hundred model runs, along with the corresponding standard deviation using all data points. The proportion of water was the most important variable in the prediction of PfPR₂₋₁₀ in Dar es Salaam, along with the proportion of tall vegetation, bare ground, distance to wetlands, medium density informal settlements and low elevated buildings. Four out of the six most important variables were derived from the LC map which indicates the importance of mapping the physical characteristics of the surface. On the other hand, in Kampala, the LU classes were dominant in terms of feature importance. The typology of street blocks (residential, informal) was the most discriminating predictor of PfPR₂₋₁₀, followed by the land cover classes of bare ground and tall vegetation. Other variables such as the population density, building density, the proportion of low vegetation and the distance to wetlands, while still important, contributed to lower degrees.

To further illustrate these outcomes, we present the partial dependency plots averaged over 100 model runs for the six most important variables in each city. The dependency plots illustrate the response of PfPR₂₋₁₀ with

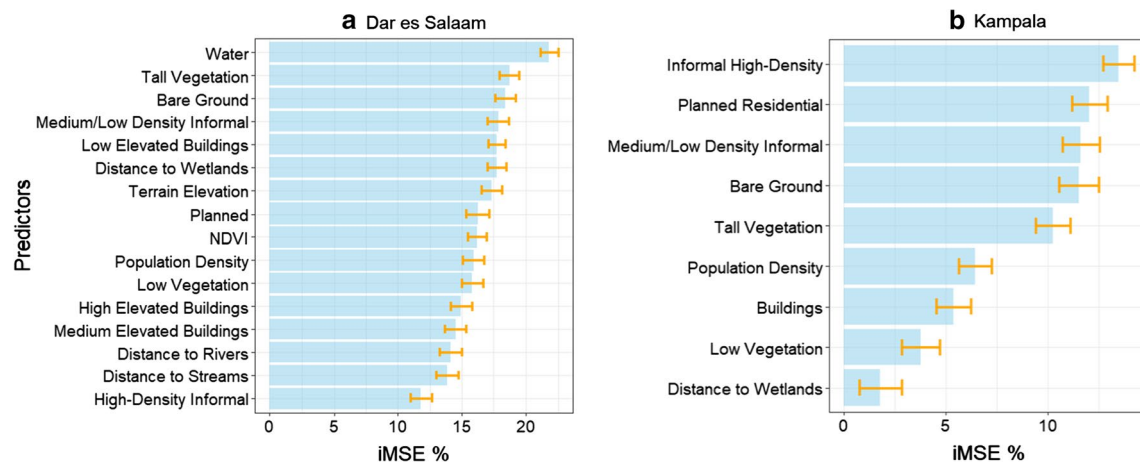


Fig. 4 Variable importance across 100 model runs using all data points. **a** Dar es Salaam and **b** Kampala

an increase in each explanatory variable, after adjusting for the effects of all other predictors. In Dar es Salaam (Fig. 5), the proportion of water was positively associated with $PfPR_{2-10}$. After the threshold of roughly 2% of water, the response of malaria prevalence spiked in a positive manner and then levels off. This could be an indicator of small patches of inland water such as ponds, or riparian

vegetation that is particularly humid, wetlands, or urban agriculture irrigation systems. Tall vegetation and bare ground are negatively associated. Moreover, there is a strong relationship between the distance from wetlands and a reduction in malaria prevalence. Low elevated buildings had a non-linear impact in $PfPR_{2-10}$, where a negative trend is exhibited up to 40% and then the

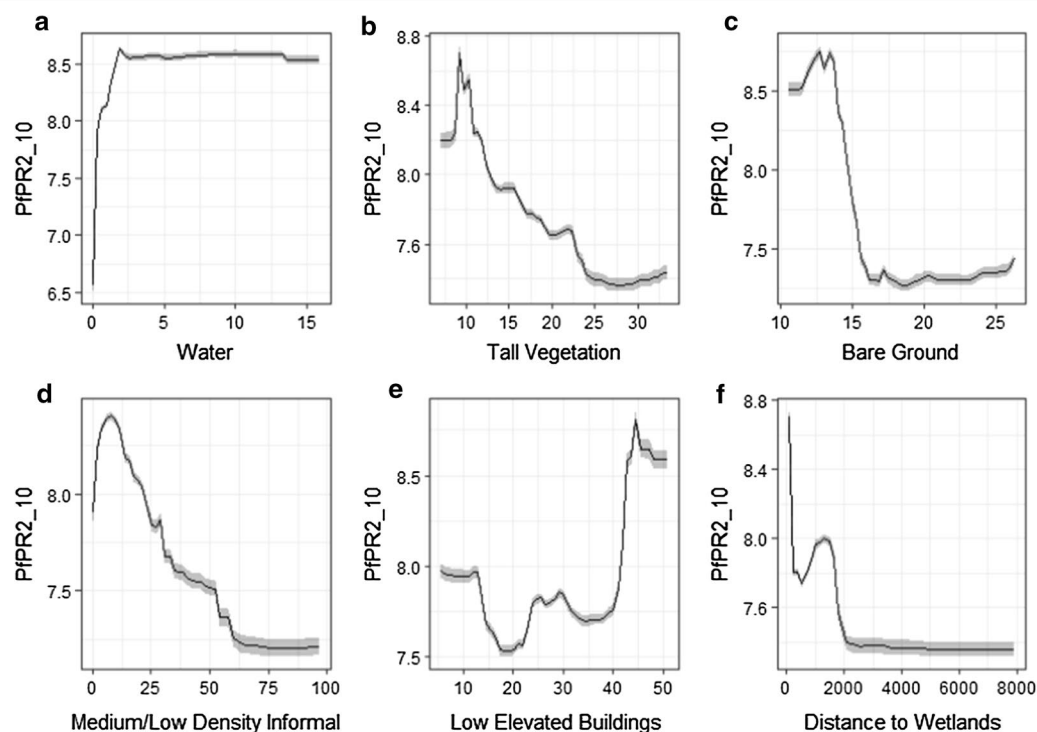


Fig. 5 Partial dependency plots for the six most important model predictors in Dar es Salaam. The shaded area represents the standard deviation over 100 simulations using all data points. The x-axis in **a** to **e** represents proportions while in panel **f**, the x-axis units represent meters. The y-axis refers to the *Plasmodium falciparum* parasite rate standardized in the 2–10 years age range

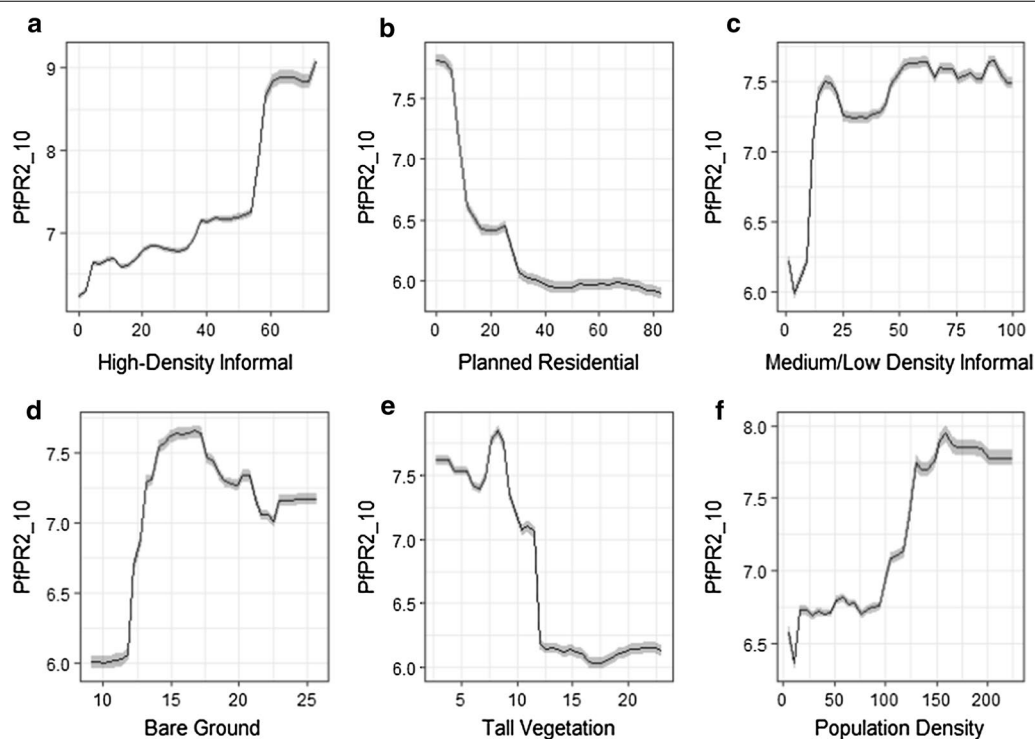


Fig. 6 Partial dependency plots for the six most important model predictors in Kampala. The shaded area represents the standard deviation over 100 simulations using all data points. The x-axis in **a** to **e** represents proportions while in panel **f**, the x-axis units represent population per hectare. The y-axis refers to the *Plasmodium falciparum* parasite rate standardized in the 2–10 years age range

relationship became positive. Finally, medium/low density informal settlements demonstrated a negative relationship with malaria prevalence. This can be explained as this residential class represents the average and most common building type in Dar es Salaam.

In Kampala, high-density informal settlements reported a strong positive relationship with malaria prevalence (Fig. 6). Different to Dar es Salaam, the proportion of medium/low informal settlements and bare ground exhibited a positive relationship with $PfPR_{2-10}$ in Kampala. Meanwhile, planned residential blocks revealed a negative association, highlighting the importance of residential typology for identifying malaria hotspots. As in Dar Es Salaam, tall vegetation showcased a negative association with $PfPR_{2-10}$.

Model performance

The model performance metrics for the two cities are presented in Tables 2 and 3. Both models performed satisfactorily, with a median R^2 of 0.39 and 0.43 in Kampala and Dar es Salaam, respectively. The model in Kampala indicated more dispersion across the bootstrap, and thus, increased uncertainty (Interquartile range of $R^2=0.45$) while a smaller dispersion was noted for the Dar es Salaam model (Interquartile range of $R^2=0.33$). With

Table 2 Descriptive statistics of the root mean squared error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2) model performance metrics for Kampala

Kampala			
	RMSE	MAE	R^2
1st quartile	4.64	3.91	0.12
Median	5.45	4.54	0.39
Mean	5.53	4.59	0.38
3rd quartile	6.59	5.19	0.57

The RMSE and MAE refer to $PfPR_{2-10}$ values

Table 3 Descriptive statistics of the root mean squared error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2) model performance metrics for Dar es Salaam

Dar es Salaam			
	RMSE	MAE	R^2
1st quartile	5.16	4.28	0.23
Median	6.02	4.81	0.43
Mean	6.05	4.88	0.39
3rd quartile	6.85	5.55	0.56

The RMSE and MAE refer to $PfPR_{2-10}$ values

respect to the RMSE, Kampala exhibited a median RMSE of 5.45 while Dar Es Salaam a median score of 6.02. The MAE distribution for both cities was less dispersed than the RMSE, as it is not influenced as much by large error deviations. The median MAE values were 4.54 and 4.81 for Kampala and Dar es Salaam, respectively. The RMSE and MAE values should not be compared across cities as they are dependent on the amount of surveys and distribution of $PfPR_{2-10}$ values in each city.

$PfPR_{2-10}$ predictions

Dar es Salaam

As exhibited in Fig. 7, the predicted distribution of malaria prevalence in Dar es Salaam was diverse and

did not follow a gradually increasing malaria risk as a function of the distance from the urban center. The spatial clustering of high $PfPR_{2-10}$ values appeared to be associated with the underlying physical and socio-economic environment and develops across riparian vegetation, urban agriculture and highly dense informal settlements. When adjusted for population, the aggregated census polygons that contain highly dense informal settlements displayed high $PfPR_{2-10}$ values, even if they were in the urban center. As Fig. 8 demonstrates, the predicted $PfPR_{2-10}$ values in Dar es Salaam were lower across the wealthier planned neighborhoods of the urban center, while were significantly higher for

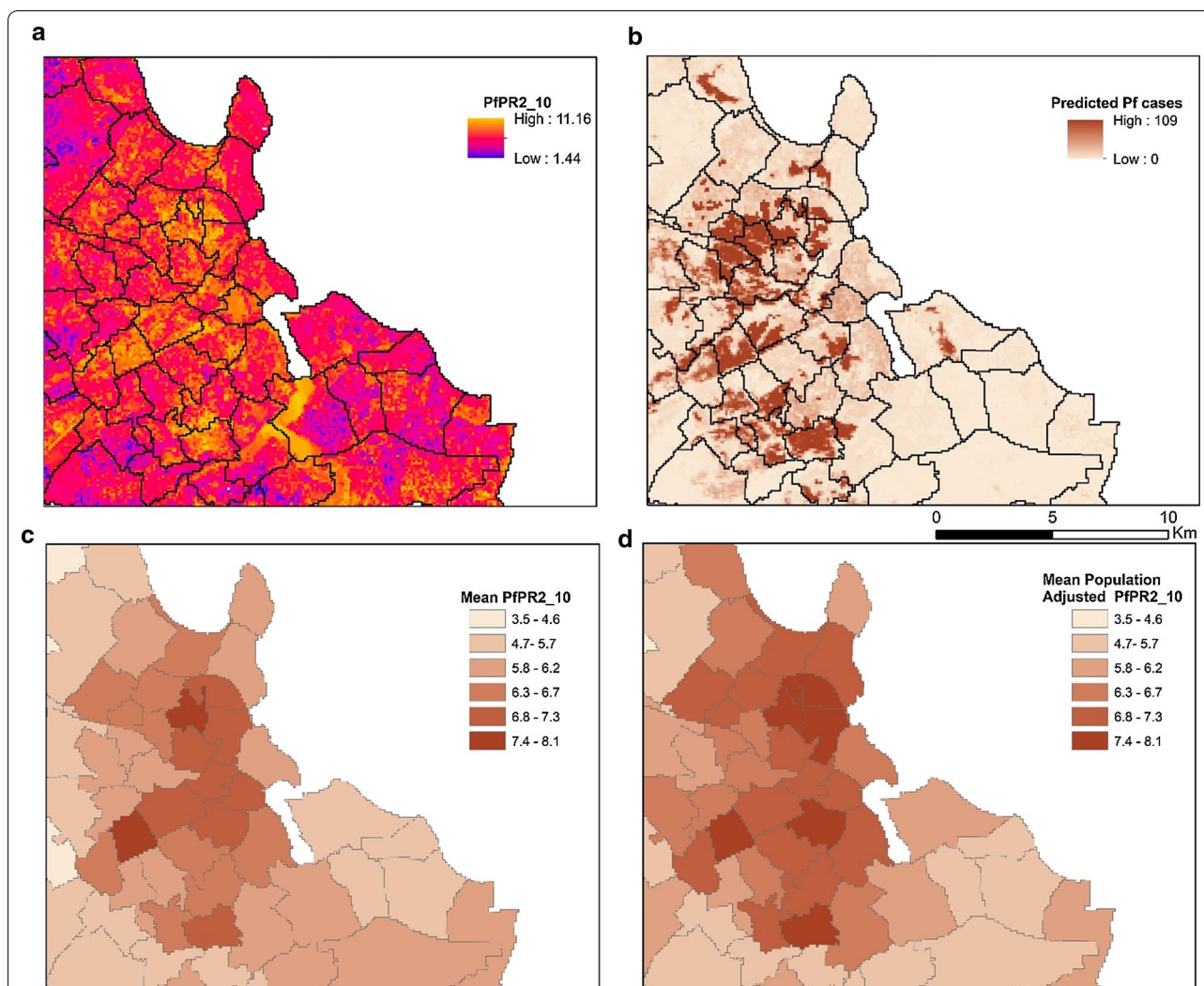
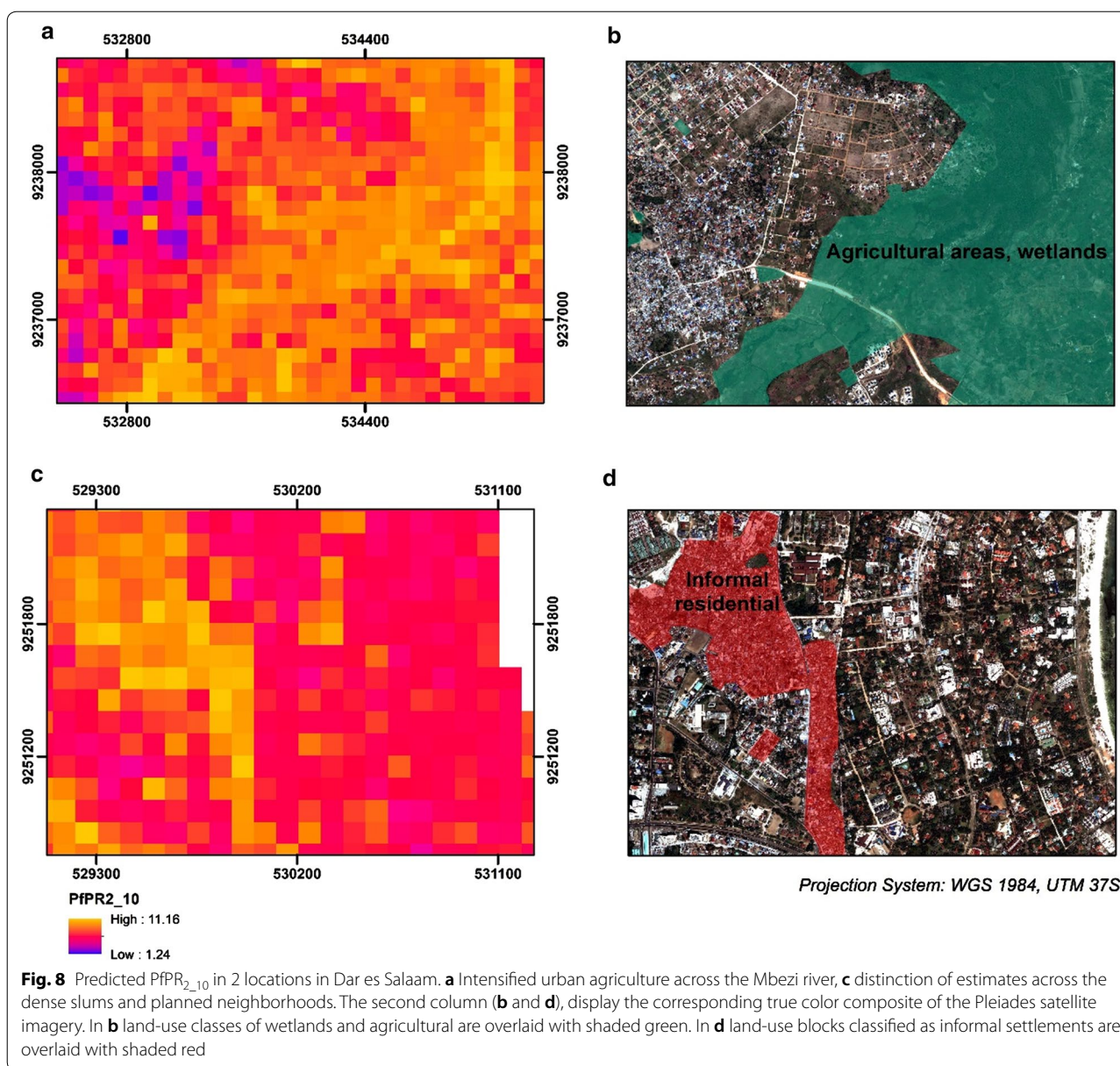


Fig. 7 Model derivatives at a raster (**a, b**) and administrative (**c, d**) resolution for Dar es Salaam. **a** Predicted $PfPR_{2-10}$ at a 100 m resolution, **b** number of predicted positive malaria cases at 100 m resolution using the distributed population grid, **c** Mean predicted $PfPR_{2-10}$ at an aggregated version of the level 5 of the administrative level of the 2002 Tanzania Census and **d** Mean Population Adjusted $PfPR_{2-10}$ at an aggregated version of the level 5 administrative level of the 2002 Tanzania Census

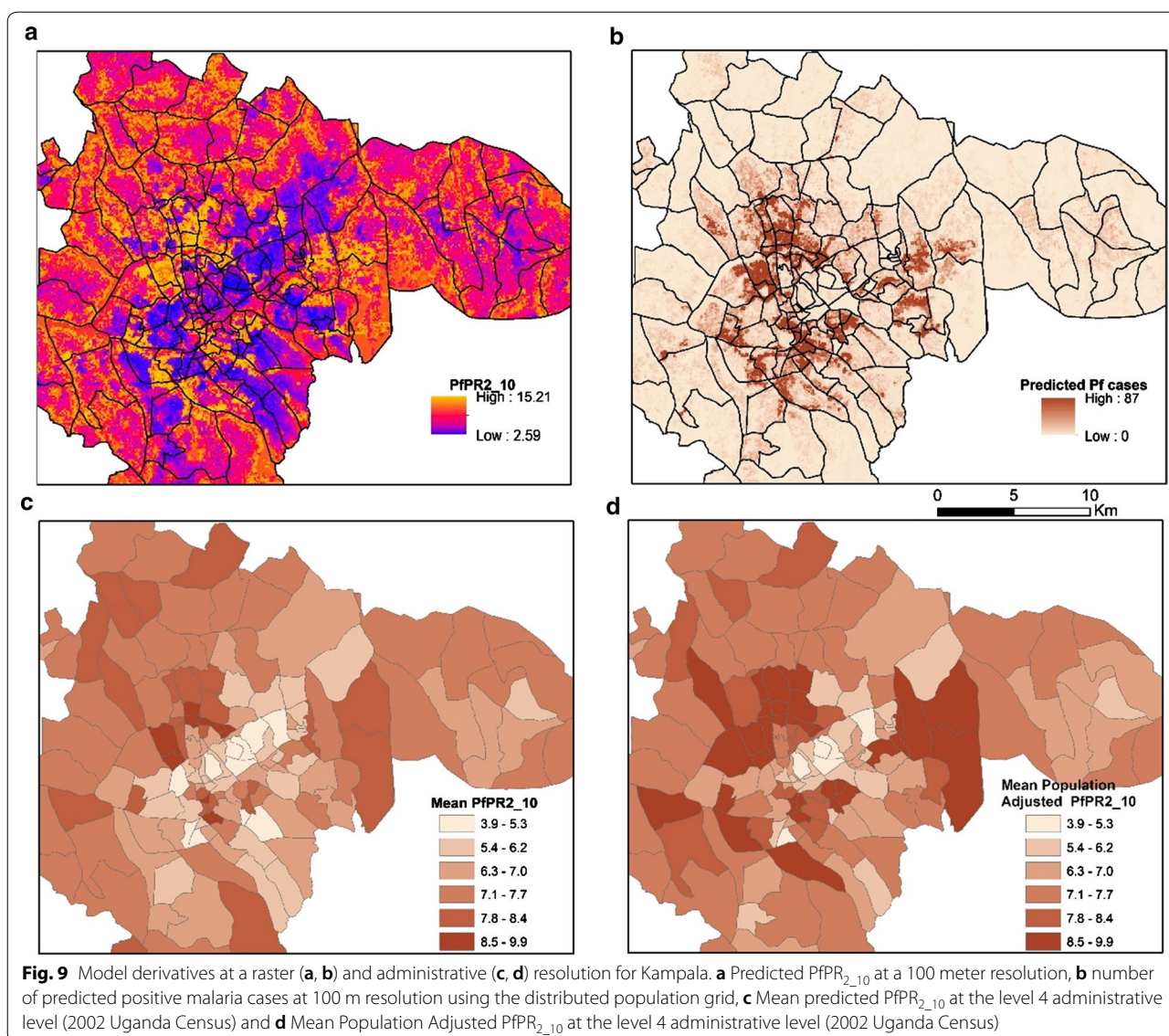


the dense informal settlements (Fig. 8c), and regions of urban agriculture and wetlands (Fig. 8a).

Kampala

In Kampala, the overall range of the parasite rate was higher than that of Dar es Salaam but with a different urban distribution. The highest values of predicted $PfPR_{2-10}$ are in regions combining a set of physical and socio-economic criteria such as highly dense slums bordering wetlands. The overall predicted $PfPR_{2-10}$ ranged from 2.6 to 15.2 at the grid level, while it decreased when summarized at an administrative level (3.9–9.9). Figure 9 illustrates these outputs across the main urban extent

of Kampala. The population adjusted estimates signify increased risk across administrative units that contain large extents of highly populated slums, developed across large bodies of water, wetlands and humid vegetation. The risk in the planned and commercial center was significantly lower than in peri-urban regions whether accounting for population or not. Figure 10 shows snapshots of the predicted $PfPR_{2-10}$ across different locations in Kampala. The model predicted increased values of $PfPR_{2-10}$ in informal settlements (Fig. 10c), regions of urban agriculture, wetlands and swamps (Fig. 10a) while the risk was decreased in the planned residential areas.



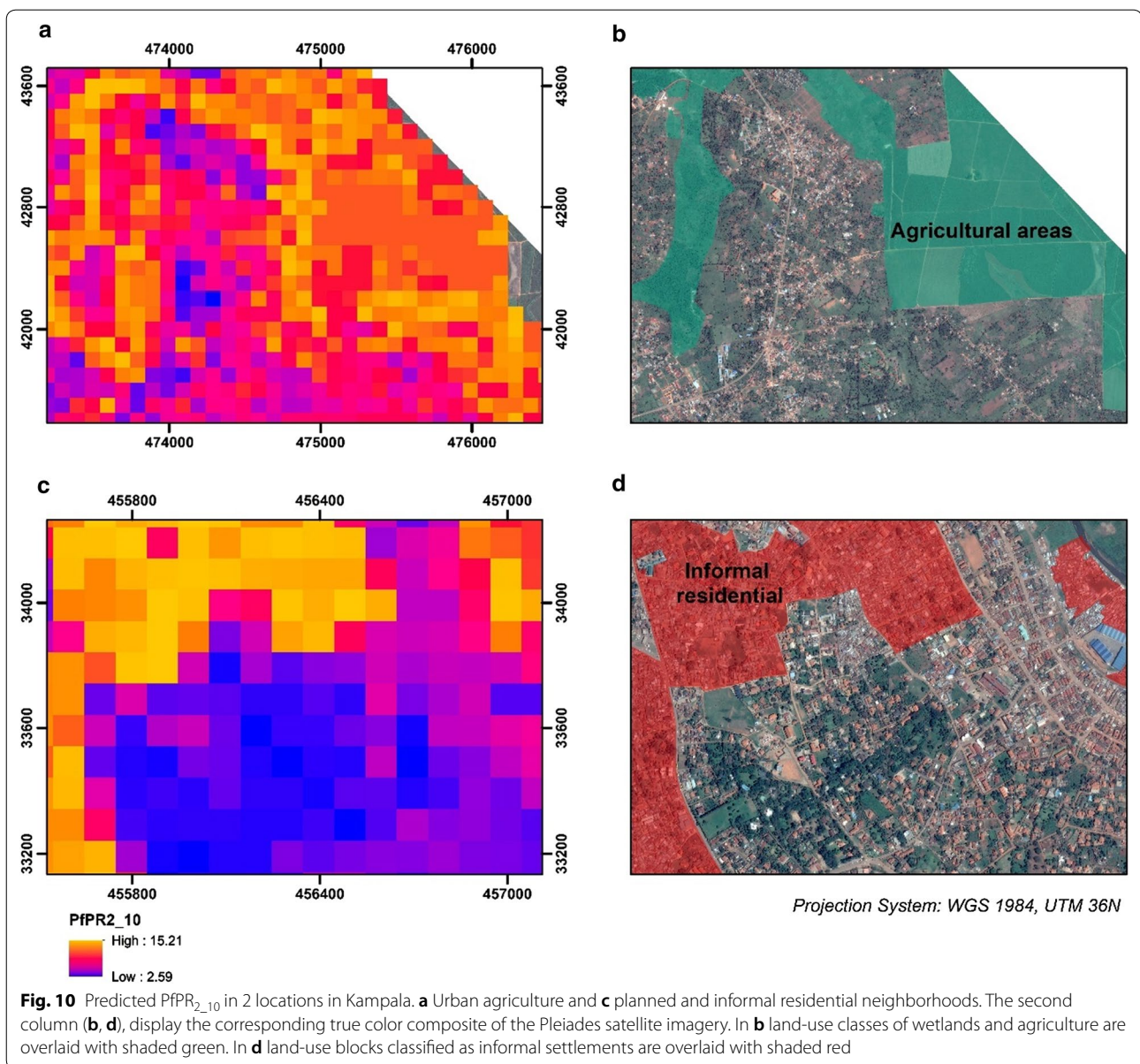
Spatial uncertainty

Figure 11 presents the spatial distribution of the coefficient of variation (CV), computed on the predictions of the 100 model runs in each city. The CV values were low in both cities, indicating low spatial prediction uncertainty. Nonetheless, in relative terms across the spatial domain, some differences emerge. In Kampala, the CV was higher in the urban center with decreased values across the peri-urban regions, while in Dar es Salaam higher values of CV were clustered mostly at the planned residential neighborhoods.

Discussion

Relationship between satellite indicators and malaria prevalence

Previous research using remotely sensed datasets was able to distinguish malaria risk across the urban fabric in SSA cities, albeit using coarser resolution information [6, 9]. Given the nature of datasets and differences in objectives, only basic distinctions could be made (land cover classes such as building density and vegetation). Building density was in general negatively associated with malaria prevalence while vegetation exhibited positive associations. Although very informative, these models often neglect the importance of the underlying socio-economic relationship of different urban settlements with malaria risk. Here, our results support the notion that malaria



prevalence is combination of physical factors such as urban land cover that favors the emergence of vector breeding sites, but also the type of surrounding communities (informal, formal), which provides an indication on their capability to cope with the burden of malaria. This aligns with previous work suggesting that malaria prevalence can be significantly higher in informal settlements in comparison to other urban landscapes [18, 19]. Additionally, we show that malaria prevalence can be linked to neighborhood location, where settlements located close to wetlands or agricultural fields were more affected. Recent evidence of increased insecticide resistance of malaria vectors in SSA cities is also fortifying the link

between persistent malaria prevalence and urban agriculture [78]. The strength of VHR remotely sensed products resides in their ability to discriminate, with relative ease various types of urban communities based on their built-up characteristics (orientation, size, density, elevation). Moreover, with the latest advents in computer vision, analysis of very large areas (city extent) can be feasible with standard computers and open source software [49, 54]. Nonetheless, VHR imagery can be particularly costly for institutions in the Global South to acquire. Soon, we expect more VHR satellite data to be publicly distributed, as is the case already in some areas.

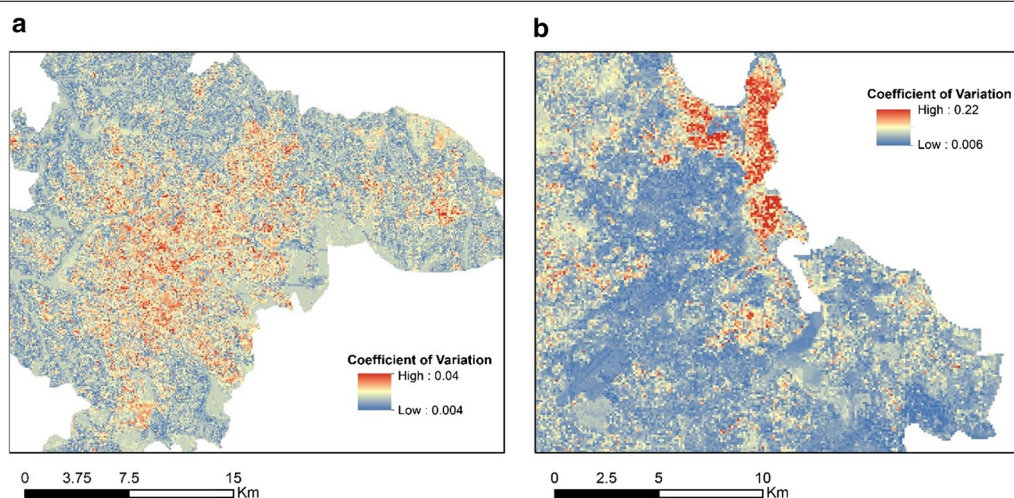


Fig. 11 Coefficient of variation in **a** Kampala and **b** Dar Es Salaam. The coefficient of variation is computed on the predictions from 100 model runs in each city

Malaria data limitations and model assessment

The $PfPR_{2-10}$ RF models are temporal composites of surveys ranging almost a decade, and consequently the temporal dimension was assumed stationary. We assumed that the extracted signal is mostly invariant as we focused on urban regions that have not undergone major transitional changes but might have expanded (i.e., large informal settlements, the planned residential center). Moreover, it was a necessary sacrifice in order to assemble a dataset large enough to capture the fine-scale spatial variability. Nonetheless, improvements in the modeling process can be expected if temporal effects are to be integrated. Furthermore, significant variations exist within the malaria datasets used in this study, with respect to sample sizes, survey locations and years of survey, which are likely to bias the results. Although, we attempted to mitigate these effects through stratified sampling and intense bootstrapping, a rigorous sensitivity analysis should be investigated when facing situations of multi-survey information as input to spatial models. Information regarding potential anti-malarial interventions was not incorporated as the number of surveys and information in both cities was limited. In future work, indicators pertaining to intervention campaigns should be investigated as some are already available at the national or regional levels [79]. As informative as they are, the model results should be used with caution and as complementary material with other malaria sources and expert knowledge.

Our models explained about half of the variance, which is in range with predictive studies of malaria prevalence across various spatiotemporal scales [6, 80–84]. The results are expected to improve when information

regarding human decisions and behavior is integrated, such as the use of insecticidal bed nets and type of infection (imported or acquired locally from rural–urban migration). Furthermore, we must acknowledge that the predictors used in this study cannot be considered absent of error. As with any LC and LU classification, there is a certain degree of misclassification error which can propagate in any subsequent analysis. This can be investigated further, using the maps of the coefficient of variation in the predictions. Areas that highlight hotspots of variation might indicate that a local refinement in the predictive variables of the LC and LU maps is needed and the classification process subsequently revisited. Alternatively, it might indicate uncertainties pertaining to the influence of the variables in the task of predicting $PfPR_{2-10}$. Nonetheless, all the LC and LU products were produced with recent, state-of-the-art analysis and high degrees of accuracy, and validated by the high level of model performance and minimal spatial uncertainty. It should be noted that the models developed here are applicable only in an urban context and lose generalization ability in the rural or dominantly rural peri-urban regions.

The importance of urban geography when addressing urban malaria

With respect to the predicted $PfPR_{2-10}$ distributions in Dar es Salaam and Kampala we conclude there is not a straightforward urban–rural trend in malaria prevalence. As mentioned by previous urban malaria reviews [85], the underlying physical and socio-economic geography may dictate part of the malaria distribution in the city. This would explain why some cities exhibit hotspots of malaria prevalence in densely urbanized areas

or intermediate zones rather than in surrounding, more rural regions [86, 87]. Aligning with these findings, our analysis demonstrated that central hotspots can be found when certain criteria are met, e.g., proximity to water bodies and humid, low, marsh-like vegetation, slums and agriculture. In Dar es Salaam, while low-elevated building density was mostly negatively associated with prevalence, a spike in PfPR₂₋₁₀ was observed once its density exceeded a certain threshold—which can be described as a highly dense, informal settlement. On the contrary, in Kampala, the land-use predictors were the most dominant in terms of importance for predicting malaria prevalence as a clear dichotomy between slums and planned neighborhoods was exhibited, which was not obvious in Dar es Salaam. These variations could be explained by two main factors. First, the land cover product of Kampala is generally less detailed than that of Dar es Salaam since it contains only a single building density class. Given that stereoscopic images were used for the LC classification in Dar es Salaam, the building elevation was extracted, offering more discrimination capabilities. Second, there exists intrinsic historical differences with respect to the way each city has been built and developed. Kampala exhibits a clearer pattern of clustered, wealthier areas built in elevated topography, with slums developed around them. In Dar es Salaam, most of the city can be considered to have a more informal nature, and as such, the discriminatory power of the LU map might be more limited—at least to the level that was used in this study. Notably, the data extracted from OSM in Dar es Salaam were highly predictive and should be considered as an additional source of information when they have enough degree of completeness for a given study area.

Intra-urban human population distribution maps as an additional tool to address urban malaria

To our knowledge this was the first study making use of fine-scale population data distributed through VHR information in order to adjust the PfPR₂₋₁₀ estimates across two cities. As population density varies greatly across the urban fabric, efforts should be made to not only present the malaria prevalence as an abstract variable, but according to the underlying population at risk. It should also be emphasized that the population information used in both cities comes from a census carried out in 2002 and a temporal mismatch between the imagery and population counts can exist, even though relative patterns are expected to be similar. Nonetheless, population projection techniques can be applied to simulate both population and malaria cases in future dates.

Future prospects

This work also serves as a call for the intensification of geolocated urban malaria surveys and their dissemination, while not neglecting privacy issues. With almost half of the SSA population predicted to be residing in cities by 2030 [88], understanding and mapping malaria prevalence across the various urban environments is of utmost importance for building more resilient cities. Our study suggests that more attention should be paid to informal settlements. The second point of note relates to secondary urban areas (SUA's) in SSA. Most of the malaria research is focused on rural communities or main urban centers of economic growth. Nonetheless, SUA's currently absorb about 75% of rural–urban migration and their growth rates can be considerably higher compared to the already large urban centers [89]. Zimmer et al. [90], analyzed SUA's in 8 southern African countries and concluded that they account for about half the urban population. These secondary cities are undergoing severe transformations. However, not much is known about them either from an epidemiological or a geographical perspective, with a large part of available information coming through studies employing satellite remote sensing [90]. In order to reduce and eliminate urban malaria in the coming years and to encourage sustainable urbanization, these cities should become a focus of interest for the research and policy-making community to prevent situations that might lead to high degrees of sustained and persistent intra-urban malaria prevalence. Along with the development of a systematic ground survey network in these forthcoming urban centers, the use of remote sensing should be heavily exploited. Due to the high transferability of malaria models based on earth observation datasets even with a very limited number of ground data available, hotspots of malaria prevalence could be detected, facilitating evidence-based allocation of resources and enhancing evidence-based policy making in these cities.

Conclusions

This research provides a framework to predict intra-urban malaria at a fine scale spatial resolution, coupling machine learning algorithms, very-high-resolution satellite derived indicators, and geospatial and survey data. Focusing on Dar es Salaam and Kampala as case studies, we conclude that the predictive dataset appears to be robust for modelling the intra-urban spatial distribution of malaria prevalence across large scales. Within both cities, urban malaria prevalence is not evenly distributed and varies intrinsically across the urban fabric. Informal settlements, urban agriculture

and locations near wetlands and riparian vegetation are highlighted as potential hotspots. Population adjusted estimates indicate higher prevalence values in highly populated administrative units. Finally, the outcome of this work further encourages the use of satellite data to understand and investigate urban malaria enhancing evidence-based policy making and control efforts in SSA cities.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12942-020-00232-2>.

Additional file 1 Methodological information regarding the creation of the land-cover, land-use and population products that were used as input to the malaria models.

Additional file 2 Descriptive Statistics of parasite prevalence surveys assembled in Dar es Salaam and Kampala, respectively.

Abbreviations

LU: Land use; LC: Land cover; NDVI: Normalized vegetation difference index; UMCP: Urban Malaria Control Program; PfPR₂₋₁₀: Age standardized malaria parasite prevalence; OSM: OpenStreetMaps; CV: Coefficient of variation; RF: Random forest; VSURF: Variable selection using random forests; SSA: Sub-Saharan Africa; VHR: Very-high-resolution.

Acknowledgements

All persons involved in the analysis and interpretation of data as well as drafting and revision of the manuscript are included in the authorship of the paper.

Authors' contributions

SG, ML and CL developed the study concept. SG developed the first draft. RWS provided the assembled malaria survey data. SG, OB, TG, NM, SV and SD helped with the technical analysis. DC, VA and RWS helped with the epidemiological results interpretation. MD, MM and BP provided input and guidance for the geospatial analysis, variable selection and interpretation. All authors have contributed to consequent improvements of the first draft. All authors read and approved the final manuscript.

Funding

This research was funded by BELSPO (Belgian Federal Science Policy Office) in the frame of the STEREO III program, as part of the REACT (SR/00/337) project. RWS is supported as a Wellcome Trust Principal Fellow (# 103602 and # 212176) and is grateful to the support of the Wellcome Trust to the Kenya Major Overseas Programme (# 203077).

Availability of data and materials

The datasets that support the conclusions of this manuscript are deposited in the open access Zenodo scientific repository (<https://zenodo.org/record/3871497#XtTWPPZuKDA>). The raw satellite imagery is property of AIRBUS DS, France, all rights reserved. The OpenStreetMap data used in this are publicly accessible and copyrighted under the mention "©OpenStreetMap contributors, CC BY-SA". The Shuttle Radar Topography Mission information were extracted from <https://www.usgs.gov/>.

Ethics approval and consent to participate

Not applicable. In this study we have used secondary sources of parasite prevalence rates assembled from published and unpublished literature summarised at cluster level and does not contain any individual person's data.

Consent for publication

Not applicable. The manuscript does not contain any individual person's data.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Department of Geoscience, Environment & Society, Université Libre de Bruxelles, 1050 Brussels, Belgium. ² Department of Earth and Environmental Sciences, KU Leuven, Celestijnenlaan 200E, 3001 Louvain, Belgium. ³ Institute of Life, Earth and Environment, University of Namur, 5000 Namur, Belgium. ⁴ Department of Geography, University of Namur, Rue de Bruxelles 61, 5000 Namur, Belgium. ⁵ Senator George J. Mitchell Center for Sustainability Solutions, University of Maine, 5710 Norman Smith Hall, Orono, ME 04469-5710, USA. ⁶ Department of Geography, University of Mary Washington, 1301 College Avenue, Fredericksburg, VA 22401, USA. ⁷ Department of Geography, Ruhr-University Bochum, Bochum, Germany. ⁸ Instituto de Altos Estudios Espaciales "Mario Gulich". Comisión Nacional de Actividades Espaciales (CONAE), Universidad Nacional de Córdoba (UNC), Córdoba, Argentina. ⁹ Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina. ¹⁰ Population and Health Unit, Kenya Medical Research Institute/Wellcome Trust Research Programme, Nairobi, Kenya. ¹¹ Centre for Tropical Medicine and Global Health, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK.

Received: 3 June 2020 Accepted: 8 September 2020

Published online: 21 September 2020

References

- Wolff E, Grippa T, Forget Y, Georganos S, Vanhuyse S, Shimoni M, et al. Diversity of urban growth patterns in Sub-Saharan Africa in the 1960–2010 period. *African Geogr Rev*. 2020;39(1):45–57.
- Goncalves L, Santos Z, Amado M, Alves D, Simoes R, Delgado AP, et al. Urban planning and health inequities: looking in a small-scale in a city of Cape Verde. *PLoS ONE*. 2015;10(11):1–28.
- United Nations Human Settlements Programme (UN-Habitat). Slums: Some Definitions. *State of the World's Cities 2006/7*. 2007. p. 2.
- Kironde JML. The regulatory framework, unplanned development and urban poverty: findings from Dar es Salaam, Tanzania. *Land Use Policy*. 2006;23(4):460–72.
- Mboga N, Persello C, Bergado JR, Stein A. Detection of informal settlements from VHR images using convolutional neural networks. *Remote Sens*. 2017;9(11):1106.
- Kabaria CW, Molteni F, Mandike R, Chacky F, Noor AM, Snow RW, et al. Mapping intra-urban malaria risk using high resolution satellite imagery: a case study of Dar es Salaam. *Int J Health Geogr*. 2016. <https://doi.org/10.1186/s12942-016-0051-y>.
- Robert V, Macintyre K, Keating J, Trape JF, Duchemin JB, Warren M, et al. Malaria transmission in urban sub-Saharan Africa. *Am J Trop Med Hyg*. 2003;68(2):169–76.
- Hay SI, Guerra CA, Tatem AJ, Atkinson PM, Snow RW. Opinion—tropical infectious diseases: urbanization, malaria transmission and disease burden in Africa. *Nat Rev Microbiol*. 2005;3(1):81–90. <https://doi.org/10.1038/nrmicro1069>.
- Machault V, Vignolles C, Pages F, Gadiaga L, Gaye A, Sokhna C, et al. Spatial heterogeneity and temporal evolution of malaria transmission risk in Dakar, Senegal, according to remotely sensed environmental data. *Malar J*. 2010;9:252.
- Wilson ML, Krogstad DJ, Arinaitwe E, Arevalo-Herrera M, Chery L, Ferreira MU, et al. Urban malaria: understanding its epidemiology, ecology, and transmission across seven diverse ICEMR network sites. *Am J Trop Med Hyg*. 2015;93(3 Suppl):110–23.
- Dongus S, Nyika D, Kannady K, Mtasiwa D, Mshinda H, Gosoniu L, et al. Urban agriculture and Anopheles habitats in Dar es Salaam, Tanzania. *Geospat Health*. 2009;3(2):189–210.
- Chinery WA. Effects of ecological changes on the malaria vectors *Anopheles funestus* and the *Anopheles gambiae* complex of mosquitoes in Accra, Ghana. *J Trop Med Hyg*. 1984;87(2):75–81.
- Klinkenberg E, McCall PJ, Hastings IM, Wilson MD, Amerasinghe FP, Donnelly MJ. Malaria and irrigated crops, Accra, Ghana. *Emerg Infect Dis*. 2005;11(8):1290.
- Castro MC, Tsuruta A, Kanamori S, Kannady K, Mkude S. Community-based environmental management for malaria control: evidence from a small-scale intervention in Dar es Salaam, Tanzania. *Malar J*. 2009;8:57.

15. Wang S-J, Lengeler C, Smith TA, Vounatsou P, Diadie DA, Pritroipa X, et al. Rapid urban malaria appraisal (RUMA) I Epidemiology of urban malaria in Ouagadougou. *Malar J*. 2005;16:1–16.
16. Byrne N. Urban malaria risk in sub-Saharan Africa: where is the evidence? *Travel Med Infect Dis*. 2007;5(2):135–7.
17. Kienberger S, Hagenlocher M. Spatial-explicit modeling of social vulnerability to malaria in East Africa. *Int J Health Geogr*. 2014;13(1):29.
18. Mukasa DM. Malaria control and prevention among the under five children in slums: a case of Bwaise. 2014.
19. De Castro MC, Yamagata Y, Mtasiwa D, Tanner M, Utzinger J, Keiser J, et al. Integrated urban malaria control: a case study in Dar es Salaam, Tanzania. *Am J Trop Med Hyg*. 2004;71(2_suppl):103–17.
20. Kuffer M, Pfeffer K, Sliuzas R. Slums from space—15 years of slum mapping using remote sensing. 2016.
21. Taubenbock H, Wurm M, Setiadi N, Gebert N, Roth A, Strunz G, et al. Integrating remote sensing and social science. 2009 Jt Urban Remote Sens Event. 2009;1–7. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5137506>.
22. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature*. 2015;526(7572):207–11.
23. Gething PW, Van Boeckel TP, Smith DL, Guerra CA, Patil AP, Snow RW, et al. Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. *Parasit Vectors*. 2011;4(1):92.
24. Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, Noor AM, et al. A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med*. 2009;6(3):0286–302.
25. Noor AM, Kinyoki DK, Munda CW, Kabaria CW, Mutua JW, Alegana VA, et al. The changing risk of *Plasmodium falciparum* malaria infection in Africa: 2000–2010: a spatial and temporal analysis of transmission intensity. *Lancet*. 2014;383(9930):1739–47.
26. Baragatti M, Fournet F, Henry MC, Assi S, Ouedraogo H, Rogier C, et al. Social and environmental malaria risk factors in urban areas of Ouagadougou, Burkina Faso. *Malar J*. 2009;8:13.
27. Tanzania UR of. 2012 population and housing census: Population distribution by administrative areas. Dar es Salaam Natl Bur Stat Off Chief Gov Stat. 2013.
28. Kuffer M. Monitoring the dynamics of informal settlements in Dar es Salaam by remote sensing: exploring the use of SPOT, ERS and small format aerial photography. In: Schrenk M (Ed) Proc CORP 2003. 2003;473–83.
29. Msilanga M. Community mapping for flood resilience—the case of Dar es Salaam, Tanzania. In: Proceedings from association of geographic information laboratories in Europe conference. 2018. p. 12–5.
30. Rasmussen MI. The power of informal settlements. The case of Dar es Salaam, Tanzania. *Planum—the J Urban*. 2013;1:26.
31. Clyde DF et al. Malaria in Tanzania. *Malar Tanzania*. 1967.
32. Fillinger U, Kannady K, William G, Vanek MJ, Dongus S, Nyika D, et al. A tool box for operational mosquito larval control: preliminary results and early lessons from the Urban Malaria Control Programme in Dar es Salaam, Tanzania. *Malar J*. 2008;7(1):20.
33. Maheu-Giroux M, Castro MC. Impact of community-based larviciding on the prevalence of malaria infection in Dar es Salaam, Tanzania. *PLoS ONE*. 2013;8(8):e71638.
34. Chaki PP, Mlacha Y, Msellemu D, Muhili A, Mtema ZJ, Kiware SS, et al. An affordable, quality-assured community-based system for high-resolution entomological surveillance of vector mosquitoes that reflects human malaria infection risk patterns. *Malar J*. 2012;11(1):172.
35. Geissbühler Y, Chaki P, Emidi B, Govella NJ, Shirima R, Mayagaya V, et al. Interdependence of domestic malaria prevention measures and mosquito-human interactions in urban Dar es Salaam, Tanzania. *Malar J*. 2007;6(1):126.
36. Wang S-J, Lengeler C, Mtasiwa D, Mshana T, Manane L, Maro G, et al. Rapid urban malaria appraisal (RUMA) II: epidemiology of urban malaria in Dar es Salaam (Tanzania). *Malar J*. 2006;5(1):28.
37. Uganda Bureau of Statistics. Statistical abstract. Kampala Uganda Bur Stat. 2013.
38. Nyakaana JB, Sengendo H, Lwasa S. Population, urban development and the environment in Uganda: the case of Kampala city and its environs. Kampala: Fac Arts, Makerere Univ; 2007.
39. Richmond A, Myers I, Namuli H. Urban informality and vulnerability: a case study in Kampala, Uganda. *Urban Sci*. 2018;2(1):22.
40. Habitat UN. Situation analysis of informal settlements in Kampala. 2017.
41. Njama D, Dorsey G, Guwatudde D, Kigonya K, Greenhouse B, Musisi S, et al. Urban malaria: primary caregivers' knowledge, attitudes, practices and predictors of malaria incidence in a cohort of Ugandan children. *Trop Med Int Heal*. 2003;8(8):685–92.
42. Lindsay S, Egwang T, Kebba A, Oyena D, Matwale G. Activity report 122. 2003.
43. Vermeiren K, Van Rompaey A, Loopmans M, Serwajja E, Mukwaya P. Urban growth of Kampala, Uganda: pattern analysis and scenario development. *Landsc Urban Plan*. 2012;106(2):199–206.
44. Clark TD, Greenhouse B, Njama-Meya D, Nzarubara B, Maiteki-Sebuguzi C, Staedke SG, et al. Factors determining the heterogeneity of malaria incidence in children in Kampala, Uganda. *J Infect Dis*. 2008;198(3):393–400.
45. Staedke SG, Nottingham EW, Cox J, Kanya MR, Rosenthal PJ, Dorsey G. Proximity to mosquito breeding sites as a risk factor for clinical malaria episodes in an urban cohort of Ugandan children. *Am J Trop Med Hyg*. 2003;69(3):244–6.
46. Kwirigira J, Atekyereza P, Niwagaba C, Kabumbuli R, Rwabukwali C, Kulabako R, et al. Seasonal variations and shared latrine cleaning practices in the slums of Kampala city, Uganda. *BMC Public Health*. 2016;16(1):361.
47. Blaschke T. Object based image analysis for remote sensing. *ISPRS J Photogramm Remote Sens*. 2010;65(1):2–16. <https://doi.org/10.1016/j.isprsjprs.2009.06.004>.
48. Georganos S, Grippa T, Lennert M, Vanhuyse S, Johnson B, Wolff E, et al. Scale matters: spatially partitioned unsupervised segmentation parameter optimization for large and heterogeneous satellite images. *Remote Sens*. 2018;10(9):1440. <http://www.mdpi.com/2072-4292/10/9/1440>.
49. Grippa T, Lennert M, Beaumont B, Vanhuyse S, Stephenne N, Wolff E. An open-source semi-automated processing chain for urban object-based classification. *Remote Sens*. 2017;9(4):358. <http://www.mdpi.com/2072-4292/9/4/358>.
50. Georganos S, Grippa T. Kampala Very-High-Resolution Land Cover Map. Zenodo; 2020. <https://doi.org/10.5281/zenodo.3711905>.
51. Georganos S, Grippa T. Dar Es Salaam Very-High-Resolution Land Cover Map. Zenodo; 2020. <https://doi.org/10.5281/zenodo.3711903>.
52. Vanhuyse S, Grippa T, Lennert M, Wolff E, Idrissa M. Contribution of nDSM derived from VHR stereo imagery to urban land-cover mapping in Sub-Saharan Africa. In: 2017 Joint Urban Remote Sensing Event (JURSE). 2017. p. 1–4.
53. OpenStreetMap contributors. Planet dump retrieved from <https://planet.osm.org>. 2017.
54. Grippa T, Georganos S, Zarougui S, Bognounou P, Diboulo E, Forget Y, et al. Mapping urban land use at street block level using openstreetmap, remote sensing data, and spatial metrics. *ISPRS Int J Geo-Inf*. 2018;7(7):246.
55. Grippa T, Linard C, Lennert M, Georganos S, Mboga N, Vanhuyse S, et al. Improving urban population distribution models with very-high resolution satellite information. *Data*. 2019;4(1). <http://www.mdpi.com/2306-5729/4/1/13>.
56. Stevens FR, Gaughan AE, Linard C, Tatem AJ. Disaggregating census data for population mapping using Random forests with remotely-sensed and ancillary data. *PLoS ONE*. 2015;10(2):1–22.
57. Linard C, Kabaria CW, Gilbert M, Tatem AJ, Gaughan AE, Stevens FR, et al. Modelling changing population distributions: an example of the Kenyan Coast, 1979–2009. *Int J Digit Earth*. 2017. <https://doi.org/10.1080/17538947.2016.1275829>.
58. Rodriguez E, Morris CS, Belz JE. A global assessment of the SRTM performance. *Photogramm Eng Remote Sens*. 2006;72(3):249–60.
59. Snow RW. The prevalence of *Plasmodium falciparum* in sub Saharan Africa since 1900. *Nature*. 2017. <https://doi.org/10.7910/DVN/Z29FR0>.
60. Snow RW, Sartorius B, Kyalo D, Maina J, Amratia P, Munda CW, et al. The prevalence of *Plasmodium falciparum* in sub-Saharan Africa since 1900. *Nature*. 2017;550(7677):515–8.
61. Pull JH, Grab B. A simple epidemiological model for evaluating the malaria inoculation rate and the risk of infection in infants. *Bull World Health Organ*. 1974;51(5):507.
62. Smith DL, McKenzie FE, Snow RW, Hay SI. Revisiting the basic reproductive number for malaria and its implications for malaria control. *PLoS Biol*. 2007;5(3):e42.
63. Smith DL, Guerra CA, Snow RW, Hay SI. Standardizing estimates of the *Plasmodium falciparum* parasite rate. *Malar J*. 2007;6(1):1–10.

64. Georganos S, Gadiaga AN, Linard C, Grippa T, Vanhuysse S, Mboga N, et al. Modelling the wealth index of demographic and health surveys within cities using very high-resolution remotely sensed information. *Remote Sens.* 2019;11(21):2543.
65. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5–32.
66. Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, et al. caret: classification and regression training. R package version 6.0–21. CRAN Wien, Austria. 2014.
67. Genuer R, Poggi J-M, Tuleau-Malot C. VSURF: an R package for variable selection using random forests. *R J.* 2015;7(2):19–33.
68. Georganos S, Grippa T, Vanhuysse S, Lennert M, Shimoni M, Kalogirou S, et al. Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing ... Less is more: optimizing classification performance through feature selection in a very-. *GIScience Remote Sens.* 2017. <https://doi.org/10.1080/15481603.2017.1408892>.
69. Ma L, Fu T, Blaschke T, Li M, Tiede D, Zhou Z, et al. Evaluation of feature selection methods for object-based land cover mapping of unmanned aerial vehicle imagery using random forest and support vector machine classifiers. *ISPRS Int J Geo-Inf.* 2017;6(2):51. <http://www.mdpi.com/2220-9964/6/2/51>.
70. Borra S, Di Ciaccio A. Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. *Comput Stat Data Anal.* 2010;54(12):2976–89.
71. Georganos S, Grippa T, Gadiaga AN, Linard C, Lennert M, Vanhuysse S, et al. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto Int.* 2019:1–12 (**just-accepted**).
72. Liaw A, Wiener M. Classification and regression by random forest. *R news.* 2002;2:18–22.
73. Bennett A, Kazembe L, Mathanga DP, Kinyoki D, Ali D, Snow RW, et al. Mapping malaria transmission intensity in Malawi, 2000–2010. *Am J Trop Med Hyg.* 2013;89(5):840–9.
74. Noor AM, Kibuchi E, Mitto B, Coulibaly D, Doumbo OK, Snow RW. Sub-national targeting of seasonal malaria chemoprevention in the Sahelian countries of the Nouakchott Initiative. *PLoS ONE.* 2015;10(8):e0136919.
75. Noor AM, Uusiku P, Kamwi RN, Katokele S, Ntomwa B, Alegana VA, et al. The receptive versus current risks of *Plasmodium falciparum* transmission in Northern Namibia: implications for elimination. *BMC Infect Dis.* 2013;13(1):184.
76. Macharia PM, Giorgi E, Noor AM, Waqo E, Kiptui R, Okiro EA, et al. Spatio-temporal analysis of *Plasmodium falciparum* prevalence to understand the past and chart the future of malaria control in Kenya. *Malar J.* 2018;17(1):340.
77. Giorgi E, Osman AA, Hassan AH, Ali AA, Ibrahim F, Amran JGH, et al. Using non-exceedance probabilities of policy-relevant malaria prevalence thresholds to identify areas of low transmission in Somalia. *Malar J.* 2018;17(1):88.
78. Dia AK, Guèye OK, Niang EA, Diédhiou SM, Sy MD, Konaté A, et al. Insecticide resistance in *Anopheles arabiensis* populations from Dakar and its suburbs: role of target site and metabolic resistance mechanisms. *Malar J.* 2018;17(1):1–9.
79. Tangena J-AA, Hendriks CMJ, Devine M, Tammamo M, Trett AE, Williams I, et al. Indoor residual spraying for malaria control in sub-Saharan Africa 1997 to 2017: an adjusted retrospective analysis. *Malar J.* 2020;19:1–15.
80. de Oliveira Padilha MA, de Oliveira Melo J, Romano G, de Lima MVM, Alonso WJ, Sallum MAM, et al. Comparison of malaria incidence rates and socioeconomic-environmental factors between the states of Acre and Rondônia: a spatio-temporal modelling study. *Malar J.* 2019;18(1):1–13.
81. Loha E, Lindtjørn B. Model variations in predicting incidence of *Plasmodium falciparum* malaria using 1998–2007 morbidity and meteorological data from south Ethiopia. *Malar J.* 2010;9(1):166.
82. Homan T, Maire N, Hiscox A, Di Pasquale A, Kiche I, Onoka K, et al. Spatially variable risk factors for malaria in a geographically heterogeneous landscape, western Kenya: an explorative study. *Malar J.* 2016;15(1):1.
83. Xu J-W, Liu H. The relationship of malaria between Chinese side and Myanmar's five special regions along China–Myanmar border: a linear regression analysis. *Malar J.* 2016;15(1):368.
84. Weiss DJ, Mappin B, Dalrymple U, Bhatt S, Cameron E, Hay SI, et al. Re-examining environmental correlates of *Plasmodium falciparum* malaria endemicity: a data-intensive variable selection approach. *Malar J.* 2015;14(1):68.
85. De Silva PM, Marshall JM. Factors contributing to urban malaria transmission in sub-Saharan Africa: a systematic review. *J Trop Med.* 2012. <https://doi.org/10.1155/2012/819563>.
86. Wang S-J, Lengeler C, Smith TA, Vounatsou P, Akogbeto M, Tanner M. Rapid urban malaria appraisal (RUMA) IV: epidemiology of urban malaria in Cotonou (Benin). *Malar J.* 2006;5(1):45.
87. Mourou J-R, Coffinet T, Jarjavall F, Cotteaux C, Pradines E, Godefroy L, et al. Malaria transmission in Libreville: results of a one year survey. *Malar J.* 2012;11(1):40.
88. Desa UN. World urbanization prospects, the 2018 revision. *Popul Div Dep Econ Soc Aff United Nations Secr.* 2018.
89. Andreassen MH, Agergaard J, Kiunsi RB, Namangaya AH. Urban transformations, migration and residential mobility patterns in African secondary cities. *Geogr Tidsskr J Geogr.* 2017;117(2):93–104.
90. Zimmer A, Guido Z, Tuholske C, Pakalnisks A, Lopus S, Caylor K, et al. Dynamics of population growth in secondary cities across southern Africa. *Landsc Ecol.* 2020. <https://doi.org/10.1007/s10980-020-01086-6>.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

