

RESEARCH

Open Access



# Mapping intra-urban malaria risk using high resolution satellite imagery: a case study of Dar es Salaam

Caroline W. Kabaria<sup>1\*</sup>, Fabrizio Molteni<sup>2,3</sup>, Renata Mandike<sup>2</sup>, Frank Chacky<sup>2</sup>, Abdisalan M. Noor<sup>1,4</sup>, Robert W. Snow<sup>1,4</sup> and Catherine Linard<sup>5,6</sup>

## Abstract

**Background:** With more than half of Africa's population expected to live in urban settlements by 2030, the burden of malaria among urban populations in Africa continues to rise with an increasing number of people at risk of infection. However, malaria intervention across Africa remains focused on rural, highly endemic communities with far fewer strategic policy directions for the control of malaria in rapidly growing African urban settlements. The complex and heterogeneous nature of urban malaria requires a better understanding of the spatial and temporal patterns of urban malaria risk in order to design effective urban malaria control programs. In this study, we use remotely sensed variables and other environmental covariates to examine the predictability of intra-urban variations of malaria infection risk across the rapidly growing city of Dar es Salaam, Tanzania between 2006 and 2014.

**Methods:** High resolution SPOT satellite imagery was used to identify urban environmental factors associated malaria prevalence in Dar es Salaam. Supervised classification with a random forest classifier was used to develop high resolution land cover classes that were combined with malaria parasite prevalence data to identify environmental factors that influence localized heterogeneity of malaria transmission and develop a high resolution predictive malaria risk map of Dar es Salaam.

**Results:** Results indicate that the risk of malaria infection varied across the city. The risk of infection increased away from the city centre with lower parasite prevalence predicted in administrative units in the city centre compared to administrative units in the peri-urban suburbs. The variation in malaria risk within Dar es Salaam was shown to be influenced by varying environmental factors. Higher malaria risks were associated with proximity to dense vegetation, inland water and wet/swampy areas while lower risk of infection was predicted in densely built-up areas.

**Conclusions:** The predictive maps produced can serve as valuable resources for municipal councils aiming to shrink the extents of malaria across cities, target resources for vector control or intensify mosquito and disease surveillance. The semi-automated modelling process developed can be replicated in other urban areas to identify factors that influence heterogeneity in malaria risk patterns and detect vulnerable zones. There is a definite need to expand research into the unique epidemiology of malaria transmission in urban areas for focal elimination and sustained control agendas.

**Keywords:** Urban malaria, Remote sensing, Boosted Regression Trees, Dar es Salaam

\*Correspondence: ckabaria@kemri-wellcome.org

<sup>1</sup> Spatial Health Metrics Group, Kenya Medical Research Institute/  
Wellcome Trust Research Programme, Nairobi, Kenya  
Full list of author information is available at the end of the article

## Background

The rapid rate of urban growth in sub-Saharan Africa will mean that the majority of the population on the continent will be classified as urban by 2030 [1]. The process of urbanization is associated with changes in the demographic, environmental and socioeconomic landscapes which in turn impact on the health of urban residents [2–4], including their risks of vector-borne diseases [4–6].

Malaria in Africa has long been regarded as a rural disease with the process of urbanization reducing suitable breeding environments for the dominant vector species complexes of *Anopheles gambiae* s.l. and *A. funestus* [6–9]. However, the risk of malaria infection does persist within densely populated, urban settings of Africa. In particular, *A. gambiae* s.l. is more likely to be bred in urban aquatic habitats [10–12] than other vector species and has been found in domestic containers and highly organically polluted habitats in urban areas [13, 14]. The focal transmission risk in urban areas is associated with proximity to breeding sites due to within urban water bodies, urban agriculture and proximity to peri-urban peripheries more likely to support vector breeding [6, 8, 15–19]. This heterogeneity of intra-urban risk is not captured in continental malaria risk mapping initiatives [20–23] and are not considered as part of current national control strategies that focus protecting less densely populated rural communities where risk of infection is typically higher compared to neighbouring urban areas.

Remote Sensing and Geographical Information Systems (GIS) provide cost-effective tools to identify environmental risk factors for high risk areas of vector-borne diseases. Previous studies have showed that satellite-derived or ground defined mapped extents of water bodies, swampy areas and agricultural land use, can distinguish with some precision areas of higher malaria risk within the urban settings of Dar es Salaam, Tanzania [14, 16] Ouagadougou, Burkina Faso [24] and Dakar, Senegal [8, 25]. The studies in Dar es Salaam during the early 2000s used aerial photography and hand-drawn maps to digitize potential mosquito breeding sites which were compared to empirical measures of mosquito larval densities and limited data on the prevalence of malaria infection among school children [14, 16]. The authors were able to demonstrate spatial declines in school children's infection prevalence from the periphery to the centre of Dar es Salaam [16] and higher larval densities were associated with closer proximity to urban agriculture [14]. High resolution satellite imagery are useful for accurate mapping of malaria risk factors in urban area and can be used to detect heterogeneity in land cover classes over small distances and improve the ability to identify urban vector breeding sites are often small, partially or completely covered by vegetation. High resolution imagery

was shown to be most accurate in identifying *A. gambiae* larval habitats compared to lower resolution satellite imagery [26]. The main aim of this study was to identify at a high resolution environmental factors that influence localized heterogeneity of malaria transmission in the city of Dar es Salaam, Tanzania. In this study, we combine data derived from high resolution SPOT satellite image and a wider suite of remotely sensed variables to estimate their impact of intra-urban variations of malaria infection risk in Dar es Salaam between 2006 and 2014.

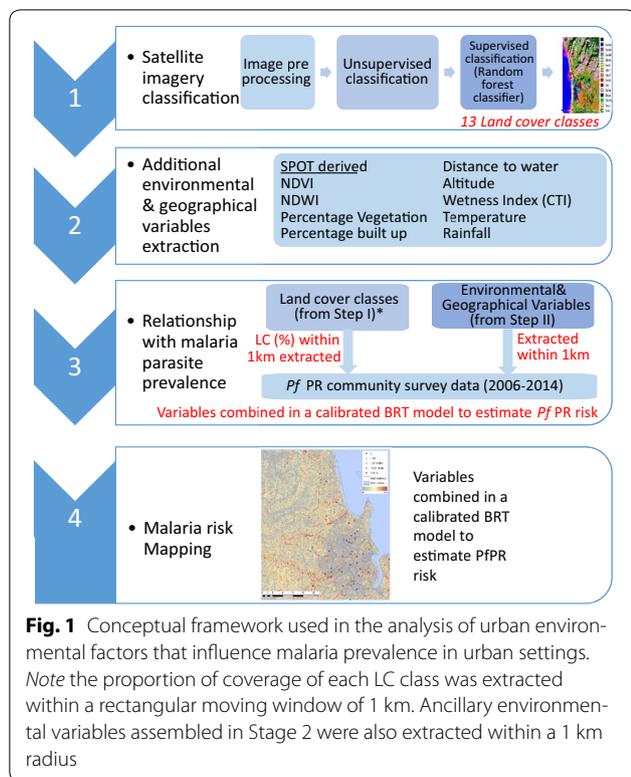
## Methods

### Study area

Dar es Salaam is Tanzania's largest city with a population of 4.6 million people [27]. With an annual population growth rate of 5.6 %, Dar es Salaam is among the fastest growing cities in Africa and the metropolitan population of Dar es Salaam is projected to reach over 5 million by 2020. The city is located on the East African coast and has been endemic for *Plasmodium falciparum* transmission since the turn of the last century when it was occupied by German Colonial authorities [28]. *P. falciparum* accounts for over 90 % of cases treated within the city [16, 29] with over a million malaria cases reported annually by the health facilities in Dar es Salaam [29] although some of these cases could be imported cases. The commercial and administrative significance of the port city of Dar es Salaam, meant that it enjoyed a long history of aggressive control through mass drug administration under German control [16, 28], environmental management under British Colonial rule [30, 31] and since the 1970s periods of integrated vector management as part of municipality control efforts [16, 32] culminating in the current programme referred to as the Urban Malaria Control Project (UMCP) [15, 16, 29, 33, 34]. Largely community-based, the UMCP mainly focuses on integrated malaria vector control based on ground-based mapping and surveillance of potential mosquito breeding sites. Routine mosquito surveillance and larviciding is conducted by community-based resource persons (CORPs), recruited from local communities via the elected local government [29, 33, 35, 36]. However, ground-based mapping and surveillance has been reported as labour-intensive and expensive [15, 37]. The application of remote sensing as a faster and less labour intensive alternative for targeted and effective control application is explored in this study. In addition, we explore the use of parasite prevalence surveys in estimating urban malaria risk.

### Overview of analysis strategy

Figure 1 gives an overview of the framework for analysis used in this study and described in more detail in subsequent sections.



**Fig. 1** Conceptual framework used in the analysis of urban environmental factors that influence malaria prevalence in urban settings. Note the proportion of coverage of each LC class was extracted within a rectangular moving window of 1 km. Ancillary environmental variables assembled in Stage 2 were also extracted within a 1 km radius

**Satellite imagery classification**

**Image acquisition and pre-processing**

High resolution satellite image from SPOT 6 (Système Pour l’Observation de la Terre) at a resolution of 1.5 m, acquired during the short rains on 14th December 2012 was obtained for the city of Dar es Salaam. The satellite image includes four spectral bands blue (0.455–0.525 μm), green (0.530–0.590 μm), red (0.625–0.695 μm), and near-infrared (NIR) (0.760–0.890 μm). The image was geo-referenced and projected to UTM zones on the WGS84 datum. Atmospheric correction was then applied using a Dark Object Subtraction (DOS) model in image analysis software, ENVI version 5.0 (Exelis VIS, USA). Radiometric correction was conducted on the satellite image by first converting digital numbers to spectral radiance then calculating exoatmospheric reflectance (reflectance above the atmosphere) using published post-launch gain and offset values [38]. A coastline mask was digitized and applied to mask out pixels of the ocean from the subsequent analysis.

**Unsupervised classification**

The SPOT image was then classified to extract land cover (LC) classes. In the initial step of image classification, exploratory unsupervised classification was run to identify a manageable number of land cover classes for image training. Unsupervised classification using ISODATA

algorithm repeated over 20 iterations was used to classify the satellite image into 20 LC classes. Class validation was conducted by checking accuracy of the generated classification for a random set of points in Google Earth. Several LC classes were merged in order to reduce the number of classes to 13, which were subsequently used to identify training sites that would be used for supervised classification in the second step of image classification. Similar methods of hybrid classification combining manual digitizing and semi-automated techniques to generate training sites for supervised classification have been used in previous studies [39–43].

**Supervised classification**

The output of the unsupervised classification was used to identify training classes for the supervised classification. For each of the 13 LC classes, a training dataset was selected on the satellite image by manually digitizing multiple training polygons for each class. One hundred training sites were obtained for each of the LC class.

A supervised classification algorithm based on random forest (RF) modelling was used to classify the satellite image [44–46]. Unlike the statistical algorithm used for the unsupervised classification that is based on the assumption that each cluster comes from a spherical normal distribution which is often not true for remote sensing images; the RF algorithm does not start with a predetermined model but instead learns the relationship from the data [45]. To build the RF model, spectral values were extracted from the multi-band SPOT image for each pixel within the training polygons. Optimal values for the number of trees (*N*) and number of observations per node (*m*) that maximize the classification accuracy while minimizing the computational time were selected by testing different combinations. Error rate estimates and confusion matrices were used to assess classification accuracy. All analyses were conducted using the statistical environment R (version 2.15.3). The RF model was developed using the random forest package version 4.6–7 [46] and additional functions provided in [47].

**Additional environmental and geographical variables extraction**

Using the LC classes developed above, a vegetation predictor was determined by combining dense and riverine vegetation LC classes. Using focal statistics techniques, percentage vegetation was then calculated within a 1 km radius. Similarly, built-up classes were combined and percentage built-up pixel calculated within a 1 km radius. Euclidean distance tool in ArcGIS was used to calculate distance to water bodies for each pixel represented by the parasite prevalence survey location. The distance from inland water variable was calculated using water channels

identified in the image combined with data from the Global Lakes and Wetlands Database (GLWD) to account for seasonal water channels in the study area that could have been present over the study period but not identified in 2012 satellite imagery.

In addition, several additional variables including humidity, vegetation and soil indicators were calculated from the SPOT image. The Normalized Difference Vegetation Index (NDVI) was calculated using the NIR and Red spectral bands as  $(\text{NIR}-\text{Blue})/(\text{NIR} + \text{Blue})$  while the Normalized Difference Water Index (NDWI) was calculated by normalising the difference between the green and NIR bands calculated as  $(\text{NIR}-\text{Green})/(\text{NIR} + \text{Green})$  [48]. The NDWI is useful in potentially delineating open water features while eliminating the presence of soil and terrestrial vegetation features. Ancillary environmental and geographical datasets shown to be associated with malaria transmission in the literature were acquired from secondary sources. Altitude was obtained from ASTER Digital Elevation Model (DEM) available at 30 m spatial resolution [49]. The Aster DEM was also used to calculate a Compound Topographic Index (CTI), a wetness index that is a function of topographic slope and the upstream contributing area orthogonal to the flow direction [50]. Mean monthly temperature for each month in the period 2006 to 2014 was calculated from land surface temperatures (LST) dataset extracted from daily Moderate Resolution Imaging Spectro-radiometer (MODIS-Terra) images. These would then be matched to the month of malaria prevalence survey for each community site. MODIS LST is freely available at 1 km spatial resolution [51]. Annual precipitation estimates in 2012 were calculated from daily rainfall estimates obtained from African Rainfall Estimates version 2 (RFE 2.0) dataset developed as a collaborative programme between NOAA's Climate Prediction centre (CPC) and USAID/Famine Early Systems Network (FEWS) [52]. A summary of all covariates used in the model, their sources as well as their spatial resolution is given in the Table 1.

#### Relationship with malaria parasite prevalence: BRT modelling

In the third stage, community level parasite prevalence survey data was combined with LC classes (Stage 1) and other environmental factors (Stage 2) to identify environmental factors that influence malaria risk within the urban area.

#### *Plasmodium falciparum* parasite prevalence data

As part of continued support to the National Malaria Control Programme (NMCP) in Tanzania, the Information for Malaria Project (INFORM) has assembled from published and unpublished sources all available

**Table 1 Summary of variables used in BRT model to estimate  $PfPR_{2-10}$  risk**

Variable	Source	Data source spatial resolution
Land cover classes (C1–C13)	SPOT satellite image <sup>a</sup>	1.5 m
Percentage dense/riverine vegetation	SPOT satellite image <sup>a</sup>	1.5 m
Percentage built-up	SPOT satellite image <sup>a</sup>	1.5 m
Distance to inland water (m)	SPOT satellite image <sup>a</sup> and GLWD <sup>b</sup>	1.5 m –
NDVI	SPOT satellite image <sup>a</sup>	1.5 m
NDWI	SPOT satellite image <sup>a</sup>	1.5 m
Altitude	ASTER GDEM <sup>c</sup>	30 m
Wetness Index (CTI)	ASTER GDEM <sup>c</sup>	30 m
Temperature	MODIS LST <sup>d</sup>	1 km
Rainfall	RFE 2.0 <sup>e</sup>	1 km

All variables except distance to water were extracted within 1 km radius

Data sources: <sup>a</sup> SPOT imagery; complete section on SPOT data

<sup>b</sup> Global Lakes and Wetlands Database. [www.worldwildlife.org/GLWD](http://www.worldwildlife.org/GLWD)

<sup>c</sup> <http://www.jspacesystems.or.jp/ersdac/GDEM/E/index.html>

<sup>d</sup> [https://lpdaac.usgs.gov/dataset\\_discovery/modis/modis\\_products\\_table/mod11a1](https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod11a1)

<sup>e</sup> <http://www.cpc.noaa.gov/products/international/data.shtml>

community based survey data on malaria infection prevalence for the country, including survey data from Dar es Salaam [53]. In brief, data included the month and year of the survey, numbers of individuals examined, lower and upper ages of the population sampled, methods used to detect parasites and the longitude and latitude of the community surveyed. Due to diversity in the age ranges of sampled populations between studies, there was need for a standardized age range to make meaningful comparisons of *Plasmodium falciparum* parasite rates across surveys. We therefore standardised parasite prevalence into the 2–10 years age group ( $PfPR_{2-10}$ ) using algorithms based on catalytic conversion models first used in malaria by Pull and Grab [54] that uses the lower and upper range of the sample and the overall prevalence to transform into a predicted estimate in children aged 2–10 years as described in Smith et al. [55]. The working paper on data assembly under the INFORM project is available on the INFORM website (<http://www.inform-malaria.org/working-papers/>). Only data from 2006 to 2014 were selected for the analysis to coincide with periods of remotely sensed image used for urban risk classifications. The final dataset included 169 community surveys at 116 sample locations within Dar es Salaam. The majority of the data locations ( $N = 75$ , 65 %) were from surveys undertaken between 2011 and 2014, where 30 were undertaken as part of school based investigations of malaria risk undertaken by the NMCP in 2014. The majority (82 %) of the

surveys were tested using RDT. A summary of key characteristics of the PR survey data is given in the Additional file 1.

The geographic coordinates of each community survey, measured at the estimated geographic centre of the survey site, were used as a unique identifier to extract values of the LC classes at the survey location. The extractions were done in ArcGIS 10 (ESRI, USA) using spatial neighbourhood analysis technique to obtain the proportion of coverage of each LC class within a rectangular moving window of 1 km radius surrounding each grid cell. Ancillary environmental variables assembled in Stage 2 were also extracted within a 1 km radius.

### BRT modelling

Boosted Regression Tree (BRT) modelling was then used to examine the relationship between parasite prevalence ( $PfPR_{2-10}$ ), urban LC classes as well as other environmental variables sampled at each community survey site. BRT is a machine learning technique increasingly used for modelling event distribution in ecology and epidemiology [56–59], in remote sensing land cover classification [60] and land cover change modelling [61].

To build the BRT model, the optimal number of trees  $nt$  was determined using the *gbm* function provided by Elith et al. [62]. Several combinations of the learning rate (LR) (0.025, 0.05, 0.1) and tree complexity (tc) (1, 5, 9) parameters were tested. Cross-validation techniques were used to evaluate model predictive performance, by randomly separating the dataset into a modelling dataset that was used to fit the model and a testing dataset that was excluded from model fitting and was used for testing the model's predictive performance. The ratio model set was set at 75 % which defined the percentage of the data sampled at every run. This was further improved using bootstrapping techniques over 25 iterations. Root mean square error (RMSE) was used to select the optimal model using the smallest value. Important predictors of  $PfPR_{2-10}$  were identified using the relative contribution output of the BRT model while the relationship patterns between individual predictors and  $PfPR_{2-10}$  were examined through partial differential plots. Variables with zero influence or <1 % relative contribution were dropped from the model. BRT models were developed using the R package 'gbm' version 1.6–3.2 [63] and the additional functions provided in Elith et al. [62]. All analyses were conducted using R (version 2.15.3) [64].

### Malaria risk mapping

In the last stage, the final selected BRT model was used to predict malaria parasite prevalence on a 10 m grid level for the city of Dar es Salaam based on the identified associations between LC classes, environmental variables and

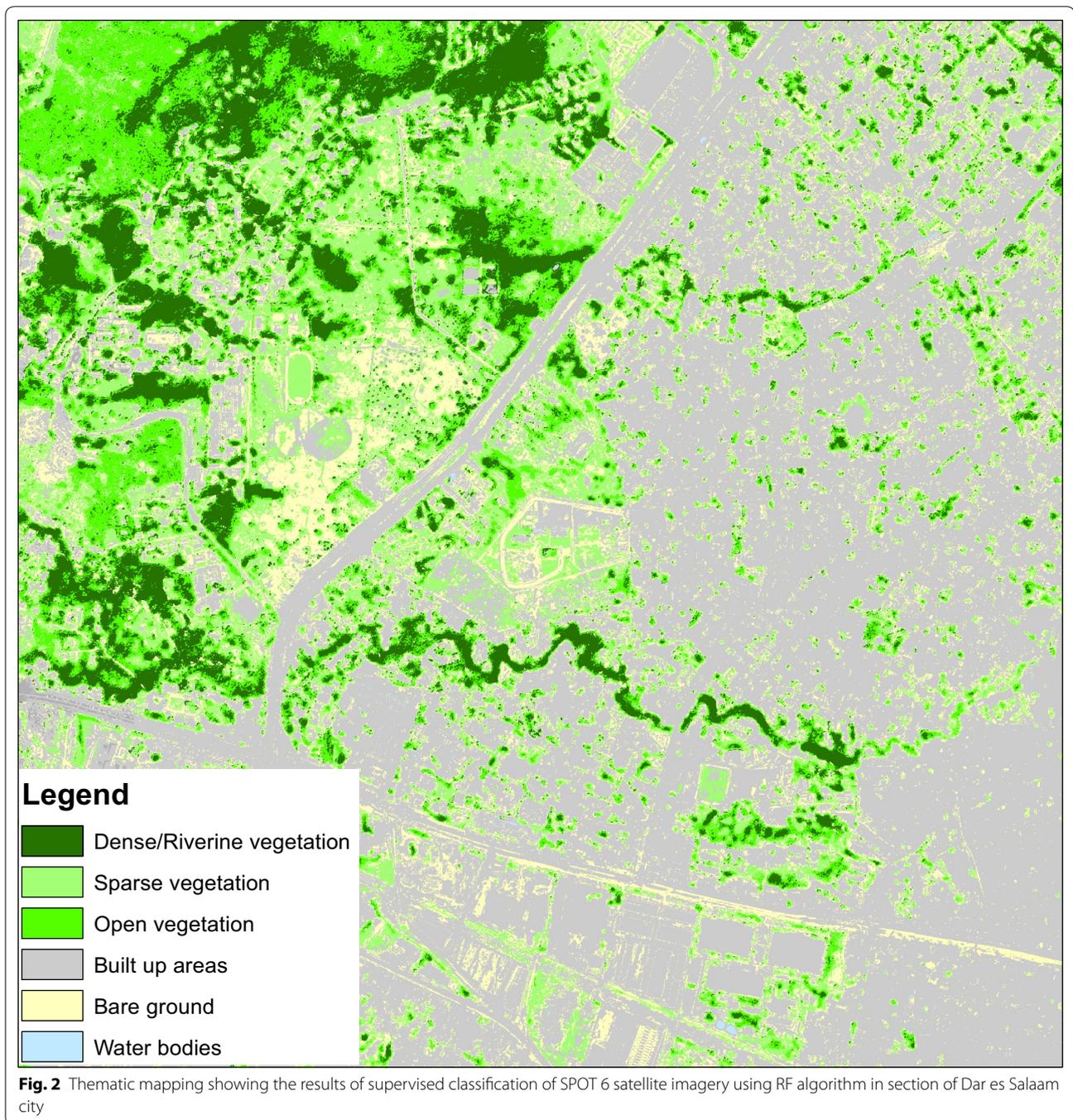
parasite prevalence. To obtain more reliable results, the BRT model runs were repeated in 25 iterations with the mean predicted value over the 25 iterations calculated as the final value for each grid cell. The mean parasite prevalence was then estimated by ward, the lowest administrative unit level used for municipal planning in Dar es Salaam.

### Results

The final classification defined 13 LC classes, distributed as five urban classes (depending on the type of buildings and tarmacked roads), three vegetation classes (light, dense or riverine), two water classes (inland or sea) and three bare soil classes (sand, bare soil, or bare soil mixed with vegetation) (Fig. 2). The accuracy of the image classification result was evaluated using an error matrix, one of the most widely used post-classification accuracy assessment methods. Overall classification accuracy of the satellite image covering Dar es Salaam was 87.1 % with Kappa co-efficient of 0.854.

Empirical estimates of malaria infection risk ( $PfPR_{2-10}$ ) ranged from 0 to 38.8 %, with 18 % of surveys reporting zero infection among the 169 sample locations. The lowest RMSE value, indicative of the best fitting model, was used to determine the parameters of the final BRT model. Several combinations of the learning rate (LR) (0.025, 0.05, 0.1) and tree complexity (tc) (1, 5, 9) parameters were tested. The BRT model with the smallest RMSE value of 16.02 was selected as the most optimal model with the final tuning parameters set to: learning rate (LR) = 0.1, tree complexity = 1 and number of trees = 100. An evaluation of the model prediction accuracy measured using the Area under Curve (AUC) over 25 iterations showed AUC = 0.89.

The relative contribution of significant predictor variables on the outcome (malaria positivity) is summarized in Table 2. Among the LC variables, the percentage of dense/riverine vegetation, built-up areas and proximity to water were found to be important predictors of  $PfPR_{2-10}$ . The percentage of dense/riverine vegetation within a 1 km radius was found to be the most important predictor of parasite prevalence with a relative contribution of close to 30 % (Table 2). Partial differential plots showing the effect of the environmental variables on the  $PfPR$  are shown in Fig. 3. The risk of malaria infection was shown to increase as the percentage of dense vegetation increased. Urban land cover, measured using percentage built-up area, was ranked as the second most important predictor of parasite prevalence with a relative contribution of 27 % (Table 2). Partial responses indicated malaria infection risk decreased with increase in built-up land cover beyond 10 % of built-up areas. Proximity to inland water was found to be the third most important predictor



of malaria infection risk, with an overall relative contribution of 9 % (Table 2). Communities closer to water bodies were found to be at a higher risk of infection than those living further away.

Topographic variables were also found to influence parasite prevalence. Altitude derived from ASTER 30 m DEM was ranked 4th with 9 % relative contribution, with a trend toward malaria risk decreasing with increasing altitude. The topography Derived Wetness Index (CTI)

was also shown to be associated with malaria infection risk, ranked as the fifth most important predictor and a relative contribution of approximately 7 % (Table 2). The partial differential plot indicates that wetter areas were associated with higher values of  $P/PR_{2-10}$ . NDVI showed similar trends with an increase in NDVI (above 0.15) associated with an increase in parasite prevalence. However, NDWI, temperature and precipitation performed poorly as predictors of malaria infection risk with  $<0.01$

**Table 2 Summary of the average contributions of the significant predictor variables using a Boosted Regression Trees (BRT) model**

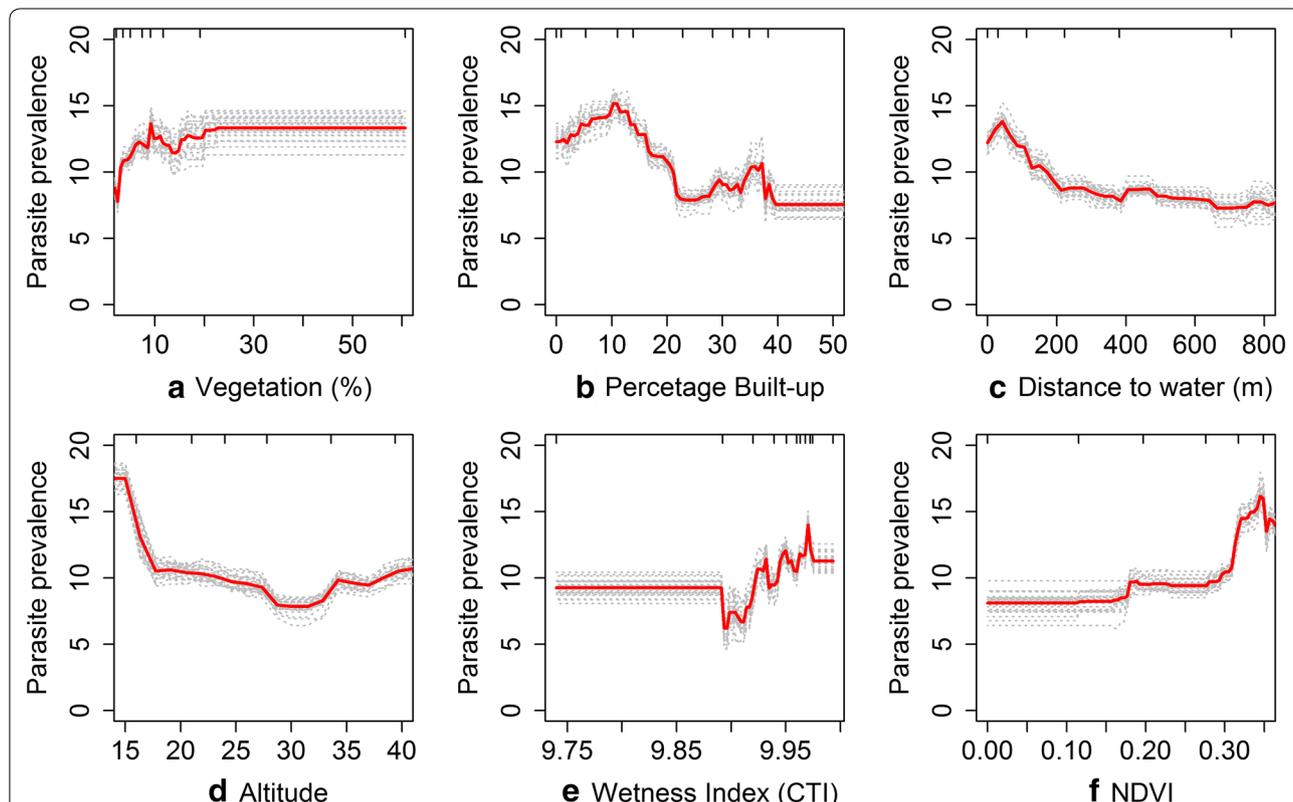
Variable	Average relative contribution (%)
Percentage dense/riverine vegetation	29.94
Percentage built-up	26.85
Distance to inland water (m)	8.98
Altitude	8.34
Wetness Index (CTI)	6.61
NDVI	5.44

BRT model developed with cross-validation over 25 bootstraps. Average relative contribution refers to the influence of each variable to the BRT model calculated as the proportion of times that a variable was selected for splitting, weighted by the squared improvement to the model as a result of each split [73]. This was then averaged over the 25 iterations of the BRT model run. This has added to the footnote of Table 2. Variables with zero influence or relative contribution <1 % were dropped from the analysis. Effect of other land cover classes was controlled for

relative contribution and not included in the subsequent model.

The optimal BRT model (Stage 3) was used to estimate parasite prevalence rates for each grid cell across Dar es

Salaam with the predictions improved using bootstrapping techniques averaged over 25 iterations. The final ensemble prediction map identifies the spatial patterns of parasite prevalence across the city of Dar es Salaam (Fig. 4a). The spatial patterns of predicted parasite prevalence from the composite model suggest that the risk of malaria transmission increases away from the city centre. There is a higher risk of infection along water channels and close to dense vegetation and lower risks among the dense, built up areas of the city where vegetation is sparse (Fig. 4a). Transformation of these high resolution risk predictions to zonal estimates by administrative wards is shown in Fig. 4b. Predicted mean  $PfPR_{2-10}$  ranged between 1 and 5 % in wards within central Dar es Salaam, with the lowest estimates,  $PfPR_{2-10}$  less than 1 %, predicted in Makurumia and Tandale wards in the city centre (Fig. 4b). A slightly higher mean  $PfPR_{2-10}$  of 6 % was predicted in Upanga ward, which borders mangrove swamps at the mouth of Msimbazi River. There was an increase in malaria prevalence in peri-urban wards as a result of increasing vegetation cover and decreasing built-up areas. The predicted mean  $PfPR_{2-10}$  ranged



**Fig. 3** Partial dependence plots showing the effect of different environmental predictors on the  $PfPR$ . Note effect of each environmental predictors after accounting for the average effect of other explanatory variables Results are shown for **a** vegetation with 1 km (%), **b** percentage built-up **c** Proximity to inland water, **d** altitude, **e** Wetness Index (CTI), **f** NDVI. Results of each of the 25 bootstrap runs are shown in grey dashed lines while average/mean plot is shown in red line

between 5 and 10 % in the peri-urban areas. The highest mean  $PfPR_{2-10}$  of above 10 % was predicted in Pugu ward in the outskirts of Dar es Salaam and close to the Pugu forest reserve (Fig. 4b).

## Discussion

Dar es Salaam, on the Tanzania coast, is characteristic of many rapidly growing, densely populated cities in Africa. Malaria transmission is generally considered lower in urban areas of Africa compared to neighbouring rural communities with an average entomological inoculation rates (EIR) of 18.8 infective bites per year estimated in urban areas compared to 126.3 in rural areas in a review of 33 independent surveys [7]; a similar trend was observed using 286 urban–rural pairs of parasite prevalence data [65]. However, the results of this study show that malaria risks do exist within the urban extents and that malaria risk within urban areas is not homogenous. In this study, we predicted heterogeneity in malaria risk using high resolution SPOT satellite image and ancillary environmental data without recourse to highly labour intensive ground mapping of risks. The variation in malaria risk within Dar es Salaam was shown to be influenced by varying environmental factors in different parts of the city with higher malaria risk associated with proximity to dense vegetation, inland water and wet/swampy areas while lower malaria risks were predicted in densely built-up areas. These results correspond to findings from mosquito vector abundance studies in Dakar where proximity to dense vegetation and large marshland areas was associated with increased mosquito densities and increased risk of malaria infection [8, 25, 66]. In Accra, Ghana, an increase in malaria cases was reported for people living within 1 km of urban agricultural activity [67] and in Ouagadougou, Burkina Faso, built up areas were significantly associated with declining malaria infection risks [24].

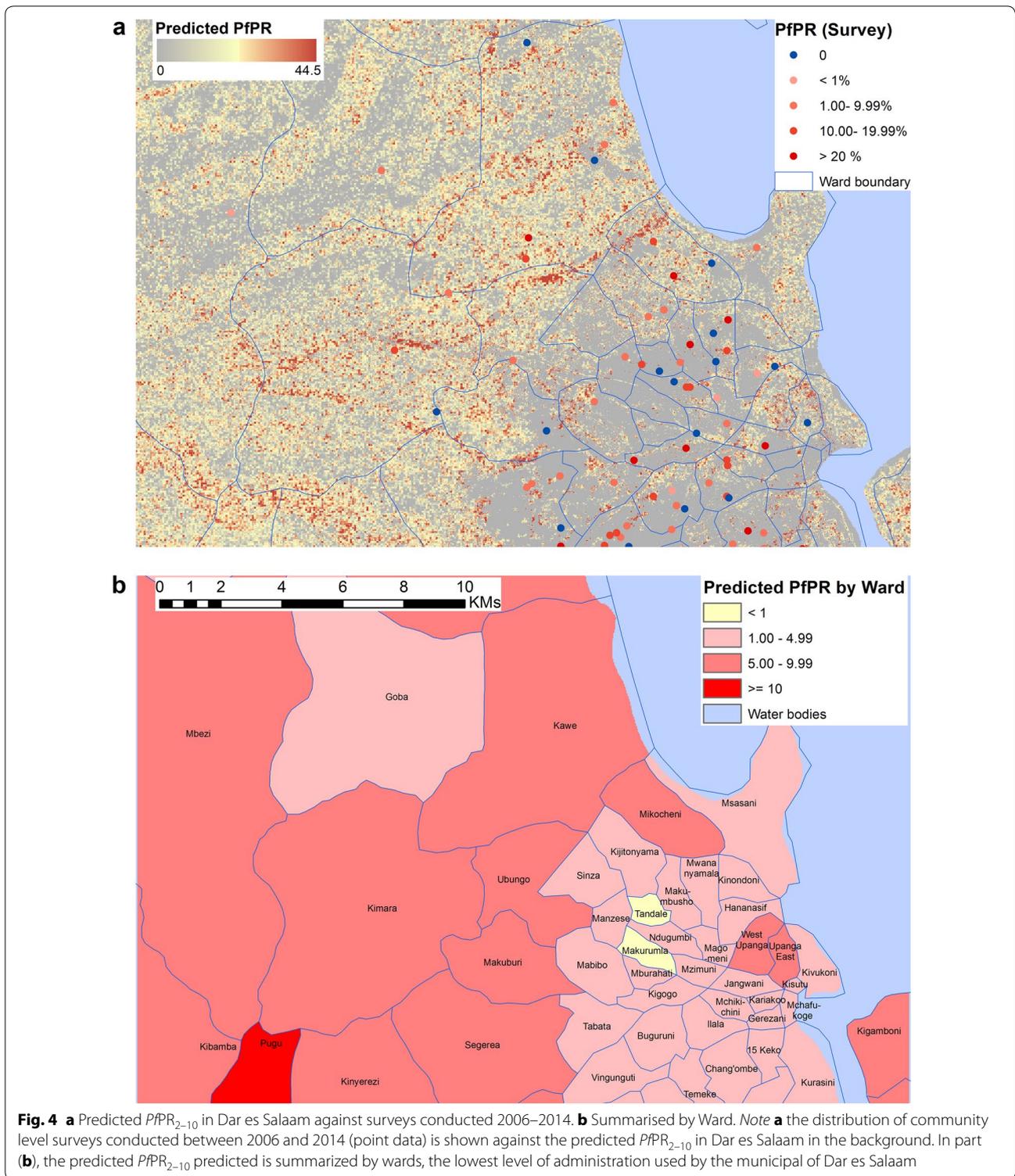
The findings of this study have implications for the recent efforts to model the intensity of malaria transmission across Africa through time [20–23]. These studies applied a single rule in all urban and peri-urban areas assuming malaria risk of infection was uniformly distributed within urban areas. However, the results of this study show that empirically measured risks across Dar es Salaam vary considerably and that this heterogeneity can be predicted based on the varying landscape within the city (Fig. 4a, b).

Remote sensing of the urban environment as used in this study offers some valuable information when aiming to identify areas within urban settlements in Africa where malaria continues to pose a significant problem. The predicted malaria risk map of Dar es Salaam, when reformatted to administrative areas used by the

municipality (Fig. 4b), shows areas of high, moderate and low risk influenced by the distribution of predictor variables. These maps can serve as valuable resources for municipal councils aiming to shrink the extents of malaria across cities, target resources for vector control or intensify mosquito and disease surveillance. The semi-automated modelling process developed in this study can be updated with new data for monitoring and estimating trends in malaria risk over time. There is also potential to scale up malaria risk evaluation to other urban areas in Africa using the methods developed in this study which can easily be replicated to identify factors that influence heterogeneity in malaria risk patterns and detect vulnerable zones.

The current malaria control strategy in Dar es Salaam implemented by UMCP focuses on integrated malaria vector control based on ground-based mapping and surveillance of potential mosquito breeding sites. However, this has been reported as labour-intensive and expensive [15, 37] while the translation of entomological-based measures into disease outcomes, such as the prevalence of malaria infection, is not straightforward [15, 37]. Remote sensing of the urban environment as used in this study provides a faster and less labour intensive alternative for targeted and effective control application. We also explore the use of parasite prevalence which is simpler to measure in the field with standardized methods and has previously been used in urban settings [15, 24]. We used  $PfPR$  measured at community level which allowed directly linkage to local environment characteristics when evaluating malaria risk factors. This is an advantage over infection estimates collected at health facilities that rarely include information on the community of residence of patients making it difficult to directly estimate the impact of environmental variables on local malaria risk. Further, community parasite prevalence surveys are obtained through active detection by screening populations irrespective of the presence symptoms of malaria unlike health facility level data that relies on symptomatic patients presenting for diagnosis and treatment. Finally, with community level parasite rates, populations at risk of malaria infection can be accurately estimated.

There are however some limitations when interpreting the results of this study. First, the temporal/seasonal trend in malaria risk could not be estimated in this study due to the unavailability of frequently collected parasite prevalence survey data that is well distributed across Dar es Salaam. We therefore used PR estimates aggregated over the period of study to predict a single risk map which does not account for change in environmental variables over the study period. Secondly, there were no details from the surveys included on whether infections were acquired locally, or whether an individual had travelled



outside of his/her usual residence. The assumption made throughout this study has been that all infections were locally acquired from the point where someone was surveyed. However, studies suggest that some of malaria

cases reported in urban areas are imported from travel to areas with high levels of malaria transmission and in the presence of proficient malaria vectors, increases malaria risk in urban areas [17, 18, 68, 69]. The inclusion of travel

histories of children positive for malaria in future parasite prevalence surveys would therefore be important in overcoming this limitation and can improve the results of the study.

There are also some constraints in using geospatial datasets at varying scales as used in this study. Although high resolution data is necessary for accurate mapping of malaria risk, remote sensing datasets are rarely available at the high resolution needed and it is often necessary to combine datasets of varying resolution to estimate risk. All resampling methods introduce some level of spatial errors as they preserve the pattern recognized at coarser resolution without increasing the information content. Secondly, there are some constraints in identifying a limited number of land-cover classes when using very high resolution satellite imagery. To minimise this constraint, a two-step hybrid classification method was used to aggregate land cover classes with similar characteristics with numerous training samples distributed across the study area taken to account for spectral range within a land cover class.

In addition, by using parasite prevalence summarised at community level, variability in risk within the community is not accounted for and thus the estimated relationship with environmental variables is effective at the community level. There is potential to explore variability in predicted risk using higher resolution parasite prevalence data. A recent study in Mozambique showed that models matching national household-level malaria infection data to high resolution environmental datasets resulted in more precise prediction compared to models using lower resolution data of greater than 30 m [70]. There is need to explore the application of household level datasets with urban contexts in Africa. Lastly, in order to account for the true effect of time on environmental determinants of malaria, the environmental covariates used must be matched with the observed data on malaria transmission. However, the environmental datasets are rarely available at time points that correspond with the date of surveys as most are derived from long-term processed remotely sensed satellite imagery or modelled climatic data generated as synoptic estimates that do not represent a specific year [71, 72].

Finally, while the development of the malaria risk map does not require extensive ground survey work of environmental risk factors, there are a number of caveats to their wider applicability. First, the model development depends on high resolution satellite imagery that is not free to public health practitioners which limits their use in public health. Public health practitioners including NMCP could benefit from collaborations with donors that lower or subsidize the cost of high resolution satellite imagery. Second, the model requires that there are some survey data that enables a training of environmental

data. This in itself is not a caveat, models unencumbered by data often do not reflect the complexities of disease under controlled and real life conditions. However, the implications are that there is a need to test the externality of findings presented here across of a range of urban settings in Africa where data do exist and their wider applications across Africa's urban extents.

## Conclusion

With more than half of Africa's population expected to live in urban settlements by 2030, the burden of malaria among urban populations in Africa will continue to rise with the increasing number of people at risk of infection. However, malaria intervention across Africa remains focused on rural, highly endemic communities with far fewer strategic policy directions for the control of malaria in rapidly growing African urban settlements. The complex and heterogeneous nature of urban malaria requires a better understanding of the spatial and temporal patterns of urban malaria risk in order to design effective urban malaria control programs targeted to specific zones. The semi-automated image classification and modelling process developed in this study can easily be replicated in other urban areas to identify environmental factors that influence heterogeneity in malaria risk patterns and detect zones of vulnerability. There is a definite need to expand research into the unique epidemiology of malaria transmission in urban areas for focal elimination and sustained control agendas.

## Additional file

**Additional file 1.** Descriptive Statistics of community level parasite prevalence surveys assembled in Dar es Salaam between 2006 and 2014.

## Abbreviations

ASTER: Advanced Spaceborne Thermal Emission and Reflection Radiometer; BRT: Boosted Regression Tree; CTI: Compound Topographic Index; DEM: Digital Elevation Model; DOS: Dark Object Subtraction model; GIS: Geographical Information Systems; INFORM: Information for Malaria Project; LC: land cover; LST: land surface temperatures; MODIS: Moderate Resolution Imaging Spectroradiometer; NDVI: Normalized Difference Vegetation Index; NDWI: Normalized Difference Water Index; NMCP: National Malaria Control Programme; RF: random forest; RFE: rainfall estimates;  $PPR_{2-10}$ : age standardised malaria parasite prevalence; SPOT: Système Pour l'Observation de la Terre; UMCP: Urban Malaria Control Project.

## Authors' contributions

CWK, CL and RWS developed the study concept and study design. CWK was responsible for data assembly, model development, data analysis and wrote the first draft of the manuscript. CL and RWS guided study design, data interpretation and helped in drafting the manuscript. CWK, AMN and RWS were involved in data assembly, cleaning and archiving under the Information for Malaria Project (INFORM). FM, RM, FC provided additional data on parasite prevalence and contributed to the drafting of the paper. AMN and RWS provided the policy and epidemiological interpretation of the study results. All authors reviewed the manuscript and contributed to the final submission. All authors read and approved the final manuscript.

### Author details

<sup>1</sup> Spatial Health Metrics Group, Kenya Medical Research Institute/Wellcome Trust Research Programme, Nairobi, Kenya. <sup>2</sup> National Malaria Control Programme, Ministry of Health and Social Welfare, Dar es Salaam, Tanzania. <sup>3</sup> Swiss Tropical and Public Health Institute, Dar es Salaam, Tanzania. <sup>4</sup> Centre for Tropical Medicine and Global Health, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK. <sup>5</sup> Department of Geography, Université de Namur, Rue de Bruxelles 61, 5000 Namur, Belgium. <sup>6</sup> Biological Control and Spatial Ecology, Université Libre de Bruxelles CP160/12, Av. F.D. Roosevelt 50, 1050 Brussels, Belgium.

### Availability of data and materials

The high resolution SPOT 6 satellite image used in this study can be obtained from Spot Image website ([www.spotimage.com](http://www.spotimage.com); [www.geo-airbusds.com/](http://www.geo-airbusds.com/)). Licencing terms of purchase restrict access to licenced users only. All environmental and geographical variables used in this study are freely available to the public [49, 51, 52]. All data assembled in Tanzania under the INFORM project will be available through the Ministry of Health, Tanzania who will be the custodians and have full responsibility for its distribution to national, regional and international scientists as well as control partners. All other data used in this study are presented in the publication.

### Acknowledgements

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

### Competing interests

The authors declare that they have no competing interests.

### Consent for publication

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

### Ethics approval

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.

### Funding

AM is supported by the Wellcome Trust, UK as an intermediate fellow (# 095127); RWS is supported by the Wellcome Trust as Principal Research Fellow (# 079080 and # 103602), that's also supported CWK. CL is supported by funding from the Belgian Science Policy (SR/00/304). CWK is also grateful to the KEMRI-Wellcome Trust Overseas Programme Strategic Award (# 084538) for additional support during her PhD. FM, RM, and FC are supported under the National Malaria Control Programme, Ministry of Health and Social Welfare, Tanzania.

Received: 11 February 2016 Accepted: 27 June 2016

Published online: 30 July 2016

### References

- United Nations. Department of Economic and Social Affairs, Population Division. World Urbanization Prospects: The 2014 Revision. Highlights (ST/ESA/SER.A/352). New York: United Nations; 2015.
- Alirol E, Getaz L, Stoll B, Chappuis F, Loutan L. Urbanisation and infectious diseases in a globalised world. *Lancet Infect Dis*. 2011;11(2):131–41.
- Dye C. Health and urban living. *Science*. 2008;319(5864):766–9.
- Utzinger J, Keiser J. Urbanization and tropical health—then and now. *Ann Trop Med Parasitol*. 2006;100(5–6):517–33.
- Keiser J, Utzinger J, Caldas de Castro M, Smith TA, Tanner M, Singer BH. Urbanization in sub-saharan Africa and implication for malaria control. *Am J Trop Med Hyg*. 2004;71(2 Suppl):118–27.
- Robert V, Macintyre K, Keating J, Trape JF, Duchemin JB, Warren M, Beier JC. 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. Urbanization, malaria transmission and disease burden in Africa. *Nat Rev Microbiol*. 2005;3(1):81–90.
- Machault V, Vignolles C, Pages F, Gadiaga L, Gaye A, Sokhna C, Trape JF, Lacaux JP, Rogier C. Spatial heterogeneity and temporal evolution of malaria transmission risk in Dakar, Senegal, according to remotely sensed environmental data. *Malar J*. 2010;9:252.
- Warren M, Billig P, Bendahmane D, Wijeyaratne P. Malaria in urban and peri-urban areas in sub-Saharan Africa. In: EHP Activity Report 71. 1999.
- Awolola TS, Oduola AO, Obansa JB, Chukwura NJ, Unyimadu JP: *Anopheles gambiae* s.s. breeding in polluted water bodies in urban Lagos, southwestern Nigeria. *J Vector Borne Dis*. 2007;44(4):241–44.
- Khaemba BM, Mutani A, Bett MK. Studies of anopheline mosquitoes transmitting malaria in a newly developed highland urban area: a case study of Moi University and its environs. *East Afr Med J*. 1994;71(3):159–64.
- Omlin FX, Carlson JC, Ogbunugafor CB, Hassanali A. *Anopheles gambiae* exploits the treehole ecosystem in western Kenya: A new urban malaria risk? *Am J Trop Med Hyg*. 2007;77(6 Suppl):264–9.
- Kasili S, Odemba N, Ngere FG, Kamanza JB, Muema AM, Kutima HL. Entomological assessment of the potential for malaria transmission in Kibera slum of Nairobi, Kenya. *J Vector Borne Dis*. 2009;46(4):273–9.
- Sattler MA, Mtsiwa D, Kiama M, Premji Z, Tanner M, Killeen GF, Lengeler C. Habitat characterization and spatial distribution of *Anopheles* sp. mosquito larvae in Dar es Salaam (Tanzania) during an extended dry period. *Malar J*. 2005;4:4.
- 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.
- Castro MC, Yamagata Y, Mtsiwa D, Tanner M, Utzinger J, Keiser J, Singer BH. Integrated urban malaria control: a case study in dar es salaam, Tanzania. *Am J Trop Med Hyg*. 2004;71(2 Suppl):103–17.
- Wang S-J, Lengeler C, Mtsiwa D, Mshana T, Manane L, Maro G, Tanner M. Rapid Urban Malaria Appraisal (RUMA) II: epidemiology of urban malaria in Dar es Salaam (Tanzania). *Malar J*. 2006;5:28.
- 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;5:45.
- Wang S-J, Lengeler C, Smith TA, Vounatsou P, Diadie DA, Pritroipa X, Convelbo N, Kientga M, Tanner M. Rapid urban malaria appraisal (RUMA) I: epidemiology of urban malaria in Ouagadougou. *Malar J*. 2005;4:43.
- Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, Battle KE, Moyes CL, Henry A, Eckhoff PA, et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature*. 2015;526(7572):207–11.
- Gething PW, Van Boeckel TP, Smith DL, Guerra CA, Patil AP, Snow RW, Hay SI. Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. *Parasit Vectors*. 2011;4:92.
- Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, Noor AM, Kabaria CW, Manh BH, Elyazar IR, Brooker S, et al. A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med*. 2009;6(3):e1000048.
- Noor AM, Kinyoki DK, Mundia CW, Kabaria CW, Mutua JW, Alegana VA, Fall IS, Snow RW. The changing risk of *Plasmodium falciparum* malaria infection in Africa: 2000–10: a spatial and temporal analysis of transmission intensity. *Lancet*. 2014;383(9930):1739–47.
- Baragatti M, Fournet F, Henry MC, Assi S, Ouedraogo H, Rogier C, Saleem G. Social and environmental malaria risk factors in urban areas of Ouagadougou, Burkina Faso. *Malar J*. 2009;8:13.
- Machault V, Vignolles C, Pages F, Gadiaga L, Tourre YM, Gaye A, Sokhna C, Trape JF, Lacaux JP, Rogier C. Risk mapping of *Anopheles gambiae* s.l. densities using remotely-sensed environmental and meteorological data in an urban area: Dakar, Senegal. *PLoS ONE*. 2012;7(11):e50674.
- Mushinzimana E, Munga S, Minakawa N, Li L, Feng CC, Bian L, Kitron U, Schmidt C, Beck L, Zhou G, et al. Landscape determinants and remote sensing of anopheline mosquito larval habitats in the western Kenya highlands. *Malar J*. 2006;5:13.
- National Bureau of Statistics (NBS), Office of Chief Government Statistician (OCGS) Zanzibar: 2012 Population and housing census: population distribution by administrative units; Key findings. In: Dar es Salaam, Tanzania: NBS and OCGS; 2013.
- Clyde DF. Malaria in Tanzania. Oxford: Oxford University Press; 1967.

29. Dongus S, Nyika D, Kannady K, Mtsiwa D, Mshinda H, Gosoni L, Drescher AW, Fillinger U, Tanner M, Killeen GF, et al. Urban agriculture and Anopheles habitats in Dar es Salaam, Tanzania. *Geospatial Health*. 2009;3(2):189–210.
30. Colonial Development Fund. Malarial Research Scheme. Report on Work Done at Dar es Salaam during the period January 1932–1934. New York: Garden City Press; 1935. p. 1–76.
31. Colony of Tanganyika Territory: Annual medical report for the year 1921–1926. In: *Dar es Salaam*; 1927.
32. Bang YH, Sabuni IB, Tonn RJ. Integrated control of urban mosquitoes in Dar es Salaam using community sanitation supplemented by larviciding. *East Afr Med J*. 1975;52(10):578–88.
33. Fillinger U, Kannady K, William G, Vanek MJ, Dongus S, Nyika D, Geissbuhler Y, Chaki PP, Govella NJ, Mathenge EM, 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;7:20.
34. 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.
35. Chaki PP, Dongus S, Fillinger U, Kelly A, Killeen GF. Community-owned resource persons for malaria vector control: enabling factors and challenges in an operational programme in Dar es Salaam, United Republic of Tanzania. *Hum Resour Health*. 2011;9:21.
36. Castro MC, Kanamori S, Kannady K, Mkude S, Killeen GF, Fillinger U. The importance of drains for the larval development of lymphatic filariasis and malaria vectors in Dar es Salaam, United Republic of Tanzania. *PLoS Negl Trop Dis*. 2010;4(5):e693.
37. Geissbuhler Y, Kannady K, Chaki PP, Emidi B, Govella NJ, Mayagaya V, Kiama M, Mtsiwa D, Mshinda H, Lindsay SW, et al. Microbial larvicide application by a large-scale, community-based program reduces malaria infection prevalence in urban Dar es Salaam, Tanzania. *PLoS ONE*. 2009;4(3):e5107.
38. Chander G, Markham BL, Helder DL. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens Environ*. 2009;113(5):893–903.
39. Bwagoy J-RB, Hansen MC, Roy DP, Grandi GD, Justice CO. Wetland mapping in the Congo Basin using optical and radar remotely sensed data and derived topographical indices. *Remote Sens Environ*. 2010;114(1):73–86.
40. Dong J, Xiao X, Sheldon S, Biradar C, Duong ND, Hazarika M. A comparison of forest cover maps in Mainland Southeast Asia from multiple sources: PALSAR, MERIS, MODIS and FRA. *Remote Sens Environ*. 2012;127:60–73.
41. Hansen MC, Defries RS, Townshend JRG, Sohlberg R, Dimiceli C, Carroll M. Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sens Environ*. 2002;83(1–2):303–19.
42. Hansen MC, Roy DP, Lindquist E, Adusei B, Justice CO, Altstatt A. A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. *Remote Sens Environ*. 2008;112(5):2495–513.
43. Midekisa A, Senay GB, Wimberly MC. Multisensor earth observations to characterize wetlands and malaria epidemiology in Ethiopia. *Water Resour Res*. 2014;50(11):8791–806.
44. Archer KJ, Kimes RV. Empirical characterization of random forest variable importance measures. *Comput Stat Data Anal*. 2008;52(4):2249–60.
45. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
46. Liaw A, Wiener M. Classification and regression by randomForest. *R News*. 2002;2:18–22.
47. Horning N. New open source tools for segment-based classification. <http://blog.remote-sensing-conservation.org/author/ned-horning/>. Accessed Aug 2014.
48. McFeeters SK. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int J Remote Sens*. 1996;17(7):1425–32.
49. Ministry of Economy Trade and Industry of Japan (METI), National Aeronautics and Space Administration (NASA): ASTER Global Digital Elevation Model (ASTER GDEM). <http://www.jspacesystems.or.jp/ersdac/GDEM/E/index.html>. Accessed Sept 2014.
50. Evans JS, Oakleaf J, Cushman SA: An ArcGIS toolbox for surface gradient and geomorphometric modeling, version 1.01. <http://evansmurphy.wix.com/evansspatial>. Accessed Sept 2014.
51. Land Processes Distributed Active Archive Center (LP DAAC), ASA Earth Science Data and Information System (ESDIS): MODIS Terra LST/E Daily L3 Global 1 km Grid product (MOD11A1). [https://lpdaac.usgs.gov/dataset\\_discovery/modis/modis\\_products\\_table/mod11a1](https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod11a1). Accessed Sept 2014.
52. NOAA Climate Prediction centre (CPC): USAID/Famine Early Systems Network (FEWS): RFE 2.0 rainfall estimates. <http://www.cpc.noaa.gov/products/international/data.shtml>. Accessed Sept 2014.
53. Snow R, W., Amratia P, Mundia CW, Alegana VA, Kirui VC, Kabaria CW, Noor AM. Assembling a geo-coded repository of malaria infection prevalence survey data in Africa 1900–2014. In: *INFORM working paper*; 2015.
54. 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):507–16.
55. 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.
56. Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, et al. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*. 2006;29(2):129–51.
57. Leathwick JR, Francis MP, Hastie T, Taylor P. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Mar Ecol Prog Ser*. 2006;321:267–81.
58. Martin V, Pfeiffer DU, Zhou X, Xiao X, Prosser DJ, Guo F, Gilbert M. Spatial distribution and risk factors of highly pathogenic avian influenza (HPAI) H5N1 in China. *PLoS Pathog*. 2011;7(3):e1001308.
59. Van Boeckel TP, Thanapongtharm W, Robinson T, Biradar CM, Xiao X, Gilbert M. Improving risk models for avian influenza: the role of intensive poultry farming and flooded land during the 2004 Thailand epidemic. *PLoS ONE*. 2012;7(11):e49528.
60. Schneider A, Friedl MA, Potere D. Mapping global urban areas using MODIS 500-m data: new methods and datasets based on 'urban ecoregions'. *Remote Sens Environ*. 2010;114(8):1733–46.
61. Linard C, Tatem AJ, Gilbert M. Modelling spatial patterns of urban growth in Africa. *Appl Geogr*. 2013;44:23–32.
62. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *J Anim Ecol*. 2008;77(4):802–13.
63. Ridgeway G. Package 'gbm': Generalized Boosted Regression Models. R package Version 2.1.1; 2015.
64. R Development Core Team: The R Project for Statistical Computing. <https://www.r-project.org/>.
65. Tatem AJ, Guerra CA, Kabaria CW, Noor AM, Hay SI. Human population, urban settlement patterns and their impact on *Plasmodium falciparum* malaria endemicity. *Malar J*. 2008;7:218.
66. Adlaoui E, Faraj C, El Bouhmi M, El Aboudi A, Ouahabi S, Tran A, Fontenille D, El Aouad R. Mapping malaria transmission risk in northern morocco using entomological and environmental data. *Malar Res Treat*. 2011;2011:391463.
67. Stoler J, Weeks JR, Getis A, Hill AG. Distance threshold for the effect of urban agriculture on elevated self-reported malaria prevalence in Accra, Ghana. *Am J Trop Med Hyg*. 2009;80(4):547–54.
68. Siri JG, Wilson ML, Murray S, Rosen DH, Vulule JM, Slutsker L, Lindblade KA. Significance of travel to rural areas as a risk factor for malarial anemia in an urban setting. *Am J Trop Med Hyg*. 2010;82(3):391–7.
69. Rabarijaona LP, Arie F, Matra R, Cot S, Raharimalala AL, Ranaivo LH, Le Bras J, Robert V, Randrianarivelojosia M. Low autochthonous urban malaria in Antananarivo (Madagascar). *Malar J*. 2006;5:27.
70. Giardina F, Franke J, Vounatsou P. Geostatistical modelling of the malaria risk in Mozambique: effect of the spatial resolution when using remotely-sensed imagery. *Geospatial Health*. 2015;10(2):333.
71. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol*. 2005;25(15):1965–78.
72. Scharlemann JP, Benz D, Hay SI, Purse BV, Tatem AJ, Wint GR, Rogers DJ. Global data for ecology and epidemiology: a novel algorithm for temporal Fourier processing MODIS data. *PLoS ONE*. 2008;3(1):e1408.
73. Friedman JH, Meulman JJ. Multiple additive regression trees with application in epidemiology. *Stat Med*. 2003;22(9):1365–81.