Geostatistical modelling of the malaria risk in Mozambique: effect of the spatial resolution when using remotely-sensed imagery

Federica Giardina,[1,2] Jonas Franke,[3] Penelope Vounatsou[1,2]

[1]Swiss Tropical and Public Health Institute, Basel; [2]University of Basel, Basel, Switzerland; [3]Remote Sensing Solutions GmbH, Baierbrunn, Germany

Abstract

The study of malaria spatial epidemiology has benefited from recent advances in geographic information system and geostatistical modelling. Significant progress in earth observation technologies has led to the development of moderate, high and very high resolution imagery. Extensive literature exists on the relationship between malaria and environmental/climatic factors in different geographical areas, but few studies have linked human malaria parasitemia survey data with remote sensing-derived land cover/land use variables and very few have used Earth Observation products. Comparison among the different resolution products to model parasitemia has not yet been investigated. In this study, we probe a proximity measure to incorporate different land cover classes and assess the effect of the spatial resolution of remotely sensed land cover and elevation on malaria risk estimation in Mozambique after adjusting for other environmental factors at a fixed spatial resolution. We used data from the Demographic and Health survey carried out in 2011, which collected malaria parasitemia data on children from 0 to 5 years old, analysing them with a Bayesian geostatistical model. We compared the risk predicted using land cover and elevation at moderate resolution with the risk obtained employing the same variables at high resolution. We used elevation data at moderate and high resolution and the land cover layer from the Moderate Resolution Imaging Spectroradiometer as well as the one produced by MALAREO, a project covering part of Mozambique during 2010-2012 that was funded by the European Union's 7th Framework Program. Moreover, the number of infected children was predicted at different spatial resolutions using AFRIPOP population data and the *enhanced* population data generated by the MALAREO project for comparison of estimates. The Bayesian geostatistical model showed that the main determinants of malaria presence are precipitation and day temperature. However, the presence of wetlands and bare soil are also very important factors. The model validation performed on a subset of locations revealed that the use of high-resolution covariates (MALAREO land cover and elevation data) improved prediction performance.

Correspondence: Federica Giardina, Swiss Tropical and Public Health Institute, Socinstrasse 57, 4002 Basel, Switzerland.
Tel: +41.61.2848109 - Fax: +41.61.2848101.
E-mail: federica.giardina@unibas.ch

Key words: Spatial epidemiology; Geostatistical models; Malaria; Remote sensing.

Acknowledgements: we would like to acknowledge Measure Demographic and Health Survey for providing the survey data (Mozambique DHS 2011) and NASA for providing the MODIS data.

Note: the present paper includes results from MALAREO (2011-2013; http://cordis.europa.eu/project/rcn/97892_en.html), a project in the Seventh Framework Programme of the EU, aiming to stimulate and facilitate the use of earth observation in malaria control and management in South-Africa, Swaziland and Mozambique.

Received for publication: 24 February 2015.
Revision received: 1 July 2015.
Accepted for publication: 17 August 2015.

Geospatial Health 2015; 10:333
doi:10.4081/gh.2015.333

Introduction

Malaria, a leading cause of morbidity and mortality in the developing world, especially in sub-Saharan Africa, where it constitutes also a major impediment to economic development (WHO, 2014), remains one of the most important human parasitic diseases. Recent research in the spatial epidemiology of malaria has benefited from the significant progress in the development of Geographic Information Systems (GIS) (*e.g.* the MARA/ARMA project, Craig *et al*., 1999; the Malaria Atlas Project, Hay and Snow, 2006), computerised systems capable of collecting, storing, handling, analysing and displaying geographically referenced information. Further gains have been achieved due to advances in Earth Observation (EO) systems, where gathering of information about Earth via remote sensing (RS) technologies, have led to the development of moderate (MR), high (HR) and very high (VHR) spatial resolution products. The growing availability of RS data, some of them accessible free of charge via the Internet, has played a crucial role in determining the environmental predictors of malaria transmission (Ceccato *et al*., 2005).

The readily available up-to-date information on environmental variables pertinent to malaria transmission over large regions makes RS a useful source of information for identification of pockets of transmission and the development of epidemic early warning systems (EWS).

Data emanating from RS can further assist malaria control and elimination programs through spatial decision support systems enabling accurate and timely resource allocation (Clements *et al*., 2013), while spatial statistics based on RS facilitate mapping malariometric indices, such as the presence and persistence of vector (mosquitoes of the species *Anopheles*) breeding sites, larval densities, the entomological inoculation rate as well as prevalence, morbidity and mortality in the human host (Machault *et al*., 2011). Further developments in Bayesian geostatistical modelling (Diggle *et al*., 1998) have recently boosted research in this area (Gosoniu *et al*., 2009; Giardina *et al*., 2014).

The MALAREO project (www.malareo.eu; Gebreslasie and Bauwens, 2015), supported by the European Union's 7th Framework Program (FP7) for research aimed at building GIS, EO and spatial statistics capabilities and implement the their products to support the malaria control programme (MCP) in Southern Africa. The project focused on the area that corresponds to the geographic region targeted by the Lubombo Spatial Development Initiative (LSDI). Launched in 1999, the LSDI had the goal of accelerating development, particularly with regard to agriculture and tourism within an area of approximately 25,000 km^2 covering southern Mozambique, eastern Swaziland and north-eastern South Africa.

The main product created within the MALAREO project is a HR (5 m) land cover/land use (LULC) map based on RapidEye, a German geospatial information provider focused on assisting management decision-making through services based on their own EO imagery (Franke *et al*. 2013). Land cover and land use are often mapped together from RS images, because biophysical characteristics of the Earth surface (*e.g.* water, vegetation, bare soil, artificial structures), *i.e. land cover*, are strongly modified by human activities, such as agriculture, forestry and urban development (Machault *et al*., 2011). The LULC layer was classified into malaria-relevant classes including wetlands, permanent and flowing water bodies, large-scale agriculture, savannah and forests. A HR population density map was obtained by the combination of field data (detailed settlement extents and aggregated LULC classes) with census estimates from 2007 (Deleu *et al*., 2015). This approach has been used previously for the production of population layers in the AFRIPOP project (Tatem *et al*., 2007) as described by Linard *et al*. (2011). However, AFRIPOP used relatively large-scale data from the 300 m GlobCover and the 30 m Landsat Enhanced Thematic Mapper Plus (ETM+) (https://lta.cr.usgs.gov/LETMP) for settlement mapping.

Land cover/land use types have been associated with vector habitats based on simple classification techniques, as well as more sophisticated statistical models that link satellite-derived multi-temporal meteorological data and EOs with vector biology and abundance (Kalluri *et al*., 2007). Very few studies using LULC in mapping of malaria prevalence from survey data exist, but Stefani *et al*. (2013) have produced a review of studies characterising LULC features and their roles in malaria transmission. Omumbo *et al*. (2005) used LULC to map malaria risk in East Africa based on the Africover project (http://www.africover.org). The latter was produced by visual interpretation of Landsat digital ETM+ satellite imagery, and the authors defined two ecological zones using the classes *water body* and *urban/rural area type* representing the percentage area of each pixel occupied by each class. Craig *et al*. (2007) regrouped the thirteen United States Geological Survey (USGS) land cover classes from Anderson *et al*. (1976) into two categories, broadly corresponding to dry and moist land cover types in Botswana, while Gosoniu *et al*. (2009) grouped them into six categories, *i.e. urban area*, *cropland*, *grass/shrub land/savannah*, *water bodies*, *wetland* and *forest*. Both these studies used LULC as a categorical variable in their models. Riedel *et al*. (2010) assessed the role of LULC, from MR Imaging Spectroradiometer (MODIS), in the analysis of malaria indicator survey data (MIS) in Zambia. Five categories were defined: *wetlands*, *forests*, *urban areas*, *shrub lands* and *other*. At each cluster location, or group of households, the land cover covariate was summarised by the proportion of each land category within a radius of 3 km. Associations were found in particular with the urban class, where the odds of malaria were significantly lower. Overall, results generally varied from study to study.

The work presented here was undertaken as the effect of varying the spatial resolution on RS-derived environmental predictors on malaria has not yet been studied. We show the effect of the spatial resolution of RS-derived environmental covariates (LULC and elevation) and population density on the estimation of malaria risk and number of infected children, after adjusting for other environmental factors at fixed resolution. Furthermore, we probe a modelling strategy for the LULC covariate that allows direct estimation of the effect of each such class type, and we study associations with malaria risk in a geostatistical model. The malaria data used in the analysis were collected by the Demographic and Health Survey (DHS) conducted in 2011 in Mozambique testing children up to 5 years of age. Data have been analysed elsewhere without the inclusion of LULC classes (Giardina *et al*., 2014). Moderate resolution environmental variables were freely available on the Web. In the area of Mozambique belonging to the LSDI area (approximately 11,000 km^2 in the southern part of Maputo Province), LULC, elevation and population density layers were used for model validation. In particular, we produced spatially explicit estimates of malaria parasitemia risk and the number of infected children in the whole country and in the MALAREO area in Mozambique. We performed a predictive analysis using HR data and comparing the estimates in terms of their predictive ability with the lower resolution products.

Materials and Methods

Study area

The Republic of Mozambique is bordered by the Indian Ocean to the east, Tanzania to the north, Malawi and Zambia to the northwest, Zimbabwe to the west with Swaziland and South Africa to the southwest. Malaria, endemic throughout the country with regions ranging from mesoendemic to hyperendemic (Mabunda *et al*., 2008), remains a major cause of morbidity and mortality. The climate creates a favourable environment for the main malaria vectors: *Anopheles gambiae*, *A. arabiensi* and *A. funestus* species. *Plasmodium faliparum* is the most common species and it is responsible for approximately 90% of all malaria infections in the country. The peak of transmission occurs during and after the rainy season, between December and April, although malaria is transmitted year round. In the last decade the MCP has implemented large-scale indoor residual spraying (IRS) programmes in several areas of 42 districts (Ministerio da Saúde-Instituto Nacional de Estatística, 2011). Indoor residual spraying was also the major component of the LSDI. Distribution of insecticide treated nets (ITN) and long lasting insecticidal nets (LLIN) targeted all age groups since 2009 and coverage is estimated to have reached almost 40% (WHO, 2014).

Malaria data

The DHS 2011 in Mozambique was carried out between June and November 2011. It consisted of a stratified three-stage sampling design, where the primary sampling unit, referred to as cluster, was defined on the basis of the enumeration areas from the 2007 census frame. A total of 611 clusters were sampled with probability proportional to size, defined as the number of households. In the second sampling stage, 20 households were selected randomly in urban clusters and 25 households in rural clusters. A representative sample of around 13,000 households was selected and 4885 children, 0 to 5 years old, was tested for malaria parasitemia with rapid diagnostic test (RDT) and microscopy. Geo-reference and parasitemia measurements, freely accessible on the Measure DHS website, were available for 603 clusters in the survey.

Remote sensing

Land surface temperature (LST) data were obtained from MODIS at 1 km spatial resolution, while rainfall estimates (RFE 2.0) every 10 days were available at 8 km resolution via the Africa Data Dissemination Service (ADDS). RFE 2.0 were created by the Climate Prediction Center of United States' National Oceanic and Atmospheric Administration (NOAA). RFE 2.0 includes both warm cloud information and station precipitation data using an interpolation method which combines geostationary satellite infrared data from Meteosat (http://www.eumetsat.int/website/home/Satellites/CurrentSatellites/Meteosat/index.html) and Global Telecommunication System data (Xie and Arkin, 1997).

Elevation data were obtained from an interpolated global digital elevation model (GDEM) from the USGS at a spatial resolution of 1 km and the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) GDEM Version 2 at 30 m resolution (http://gdem.ersdac.jspacesystems.or.jp/). The climatic factors LST and rainfall were acquired for the 6-month period prior to the survey and the average were calculated and extracted for each data location. The environmental factors with available temporal resolution (LST and rainfall) were acquired for the 3-month period prior to the survey and the average were again calculated and extracted for each data location. In addition, AFRIPOP and MALAREO population density estimates resampled at 100 m spatial resolution were used. The proportion of children between 0 and 5 years living in Mozambique was obtained by the International Data Base of the U.S. Census Bureau, Population Division for the year 2011. All RS data at spatial resolutions between 30 and 1000 m are referred to as MR products and all those at spatial resolutions between 4 and 30 m are referred to as HR.

Land cover

The MODIS LULC categories were aligned with the MALAREO categories. The allocation was done on the basis of the available description of the layers as well as a graphical assessment. The final LULC categories are summarised in Table 1.

Statistical analysis

Land cover/land use proximity measures

While for the environmental factors we considered RS-derived values at locations only, we assumed that LULC classes might affect malaria parasitemia levels within larger areas surrounding the location. For this purpose, a measure of proximity was used to link LULC type with the DHS cluster spatial location. This was defined by the following equation:

$$LC_{ij} = \exp\left(-d^{*}_{i,LC_j^{cat}}\right), \forall j = 1,\ldots,k \qquad \text{eq. 1}$$

where $d^{*}_{i,LC_j^{cat}}$ indicates the minimum Euclidean distance between location i and the LULC category j.

Geostatistical model

Let Y_i and N_i be the number of malaria-infected and screened individuals at location $i(i=1,\ldots,n)$ and p_i the probability of infection. We assume that Y_i arises from a binomial distribution, $Y_i \sim \text{Bin}(p_i, N_i)$. The influence of environmental covariates X_i and location-specific spatial random effects ω_i are modelled on the logit scale, *i.e.*:

$$log\left(\frac{p_i}{1-p_i}\right) = X_i^T\beta + \omega_i \qquad \text{eq. 2}$$

where β is the vector of regression coefficients. Unobserved spatial variation is introduced on ω_i by assuming that $\omega = (\omega_1, \ldots, \omega_n)^T$ follows a latent stationary Gaussian process over the study region, $\omega \sim MVN(0,\Sigma)$. The matrix Σ has elements Σ_{ij} and represents the covariance between any pair in locations i and j. Assuming an isotropic exponential correlation function, the matrix elements Σ_{ij} are defined by

Table 1. Land cover/land use class alignment.

MODIS	MALAREO	LC-aligned category
Water	Standing water, flowing water	Water
Evergreen needleleaf forest, evergreen broadleaf forest, deciduous needleleaf forest, deciduous broadleaf forest, mixed forest, woody savannah	Forest/savannah	Forest
Grassland savannah	Grassland/savannah	Savannah
Barren or sparsely vegetated land	Bare soil/rock	Bare soil
Urban and built-up area	Roads urban/populated	Urban
Closed shrub-land, open shrub-land, cropland/natural vegetation mosaic	Bush/shrub-land	Bush
Permanent wetlands	Wetlands	Wetlands
Croplands	Large scale agriculture	Agriculture

MODIS, moderate resolution imaging spectroradiometer; LC, land cover.

$\Sigma_{ij} = \sigma^2 exp\,(-\rho d_{ij})$ with spatial variance σ^2, rate of correlation decay ρ with Euclidean distance between locations d_{ij}. The minimum distance for which the spatial correlation is less than 5% is referred to as range and can be calculated by $3/\rho$ in the exponential correlation function setting.

A Bayesian model formulation requires the specification of prior distributions of all model parameters. For the regression coefficients β, we assumed normal prior distributions with mean 0 and large variance. For the spatial parameters σ^2 and ρ, we chose non-informative inverse Gamma and Gamma distributions, respectively. The model was fitted using Markov chain Monte Carlo (MCMC) simulation implemented in the software 'Just Another Gibbs Sampler' (JAGS) (Plummer, 2003).

This model was initially used to obtain spatially explicit estimates of the malaria risk over the whole country by using a grid formed by pixels of 3 km resolution for computational reasons and for comparison with previously published work (Giardina *et al*., 2014). The same model was also used to obtain malaria risk estimates in the area of Mozambique belonging to the MALAREO project (Figure 1) at several higher spatial resolution by resampling (or aggregating) the environmental variables at the target spatial resolutions (*i.e.* 1 km, 500 m and 100 m).

The number of children between 0 to 5 years was calculated by resampling AFRIPOP data at the target spatial resolution and assuming the proportion of children between 0 to 5 years remained constant throughout the country (*i.e.* 21%, as reported by the International Data Base of the U.S. Census Bureau, Population Division for the year 2011). The number of infected children between 0 to 5 years was estimated sampling from the predictive distributions.

Assessing the effect of spatial resolution on model-based predictions

The model was validated using as *training set* all DHS data except the 35 locations belonging to the MALAREO area (Figure 1), which formed the *testing set*. The model used MR variables in the fitting part and MR as well as HR variables in the prediction part (Table 2). All environmental variables were resampled/aggregated at the different spatial resolutions that were assessed. Model performance was compared in terms of log-predictive density (Robert, 1996). Spatially explicit predictions (malaria risk and number of infected children) were obtained over grids covering this area with spatial resolutions of 1 km, 500 m and 100 m using both MR and HR variables.

Results

The effect of the environmental and climatic factors on malaria parasitemia risk estimated from the full DHS dataset is shown in Table 3. The main determinants of malaria presence were rainfall and LST_{day}. Among the LULC classes, the presence of large-scale agriculture and bare soil reduced the odds of parasitemia by 8% (95% BCI: 0-15%) and 44% (95% BCI: 26-60%), respectively, while the presence of bush, forest, savannah and wetlands increased the odds by 31% (95% BCI: 21-42%), 11% (95% BCI: 4-19%), 34% (95% BCI: 18-46%) and 37% (95% BCI: 55-75%). The estimates of the spatial parameters revealed a variance of 2.61 (95% BCI: 1.64-2.82) and a spatial range (the distance at which the correlation becomes negligible) of around 85.56 km (95%

Figure 1. The MALAREO project area. It is enclosed by the black line and includes the northern part of South Africa (KwaZulu-Natal Province), eastern Swaziland, and the southern part of Mozambique.

Table 2. Remotely sensed environmental variables.

Variable	Source/product for MR	Spatial resolution (m)	Source/product for HR	Spatial resolution (m)
LC	MODIS (MCD012Q1)	500	Rapid Eye	5
Elevation	MODIS	100	GDEM2	30
LST	MODIS (MOD13A2)	1000	-	-
Rainfall	MEFW (ADDS)	8000	-	-
Population	-	-	Afripop (Landsat) MALAREO (RapidEye)	100 100

MR, moderate resolution; HR, high resolution; LC, land cover; MODIS, moderate resolution imaging spectroradiometer; LST, land surface temperature.

pagepress

BCI: 56.22-127.32).

The same model was used to predict malaria risk among children up to 5 years of age over a grid of 3 km resolution. Figure 2 shows that the two provinces with the highest malaria risk were in the northern part of the country (the Nampula and Zambezia Provinces). The southern parts of the country were characterised by lower risk compared to the rest of the country (<10%), especially Maputo (city and province) and Gaza Province. Estimates of the number of children between 0 and 5 years infected by malaria parasites were obtained using the predictive distribution of the risk and the population data at 100 m spatial resolution provided by AFRIPOP (Figure 3). In most of the country, the number of infected children per 9 km^2 ranges from 1 to 10. In some densely populated areas, *e.g.* Maputo and Matola (in the South), and in very high-risk areas, *e.g.* the Zambezia Province in the central coastal region, the number can reach up to 1800 children.

The model validation revealed that the use of HR covariates in the 35 testing locations improved prediction performance. In particular, the model that employed the MALAREO layer for LULC and GDEM values had a log-predictive density of -115.12 (95% BCI: -122.32,-104.21), whilst the model that used MR covariates estimated -132.22 (95% BCI: -143.11,-121.17).

Predictions in the same area were carried out at several spatial resolutions. Figure 4 depicts the predicted malaria risk among children aged 0-5 years at 1 km, 500 m and 100 m resolution using MR and HR data.

Table 4 shows how the estimated number of infected children is affected by the population layer (and indirectly by the spatial resolution of the environmental covariates). On average, the total number of infected children estimated by the models increased with increasing resolution of the predictive grid. The use of MR variables tended to result in an overestimation in the number of infections.

Table 3. Posterior estimates arising from the geostatistical model fitted on the full Demographic and Health Survey dataset with moderate resolution imaging spectroradiometer land cover/land use.

Covariate	Median (95% BCI)
Rainfall	0.14 (0.07,0.22)
LST_{Night}	-0.11 (-0.40,0.16)
LST_{Day}	0.31 (0.09,0.54)
Elevation	-0.03 (-0.14,0.07)
LC category	
Agriculture	-0.09 (-0.17,-0.01)
Bush	0.27 (0.19,0.35)
Forest	0.11 (0.04,0.18)
Savannah	0.30 (0.17,0.45)
Urban	0.05 (-0.16,0.41
Water	0.09 (-0.2,0.40)
Bare soil	-0.59 (-0.91,-0.30)
Wetlands	0.44 (0.32,0.56)
Spatial parameter	
σ^2	2.61 (1.64,2.82)
ρ	2.31 (1.51,3.43)

BCI, Bayesian confidence interval; LST, land surface temperature; LC, land cover. The LULC categories refer to the aligned variable.

Discussion

This study focuses on the use of MR and HR of mapped variables derived by RS to obtain spatially explicit malaria burden estimates in geostatistical models. In particular, the work shows the effect of different spatial resolutions of elevation data and LULC layers (and derived population estimates) on the estimation of risk and number of infected children below the age of 5 years. An alternative definition of the LULC covariate based on a proximity measure is proposed to study associa-

Table 4. Estimated total number of infected children in the MALAREO area based on moderate and high-resolution products.

	1 km	500 m	100 m
MR	43,555 (42,334-44,234)	45,171 (44,525-46,123)	45,605 (44,532- 46,892)
HR	37,901 (36,884-38,424)	37,919 (37,011-38,626)	38,111 (37,773-39,100)

MR, moderate resolution; HR, high resolution. Median and (95% Bayesian confidence interval).

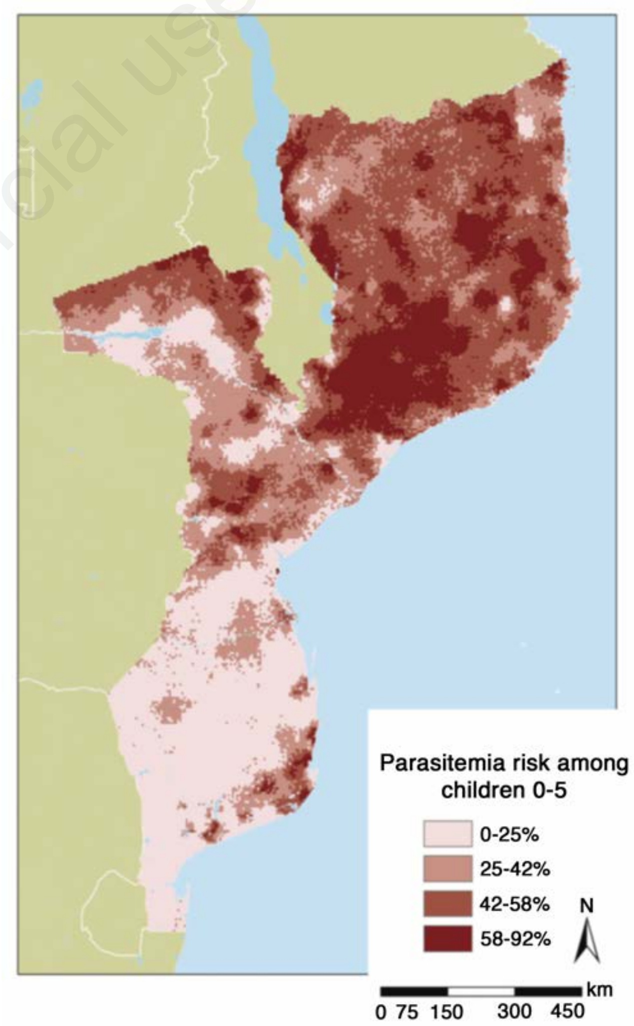

Figure 2. The predicted malaria risk among children up to 5 years of age. Median estimates are plotted at the 3-km resolution.

tions of 8 different LULC types with malaria risk and obtain explicit effect estimation.

The analysis was performed using data collected by the Mozambican DHS in 2011 with a geostatistical model utilising MR and HR environmental variables obtained by RS. The model was fitted with MR variables and a grid of 3 km resolution was chosen for the prediction in order to make direct comparison with previous work by the lead author (Giardina *et al*., 2014). Indeed, the coefficients' estimates of the common variables (Rainfall, LST_{Night} and LST_{Day}) as well as the total malaria burden measure (number of infected children) were in agreement with values reported in this paper with the spatial parameters estimates (variance and decay parameter) showing similar values.

A relatively small number of studies have included LULC classes in geostatistical models for malaria risk mapping despite their important role in determining the suitability for transmission of the disease. This may be due to difficulties in the definition of the variable to be used in the models. In some applications (*e.g.* Riedel *et al*., 2010), the LULC covariate has been considered as a categorical variable indicating the relative frequency of each LULC type within a buffer. This approach might conceptually be the best way of defining the variable. However, it has drawbacks, *e.g.*, parameter estimates have to be expressed relative to a baseline category, and certain arbitrariness in the choice of the reference category as well as with regard to the size of the buffer. We propose here a proximity measure that does not account for the area covered by a specific LULC class surrounding the locations, but which is instead based on the distance between location and each LULC class. This work shows that *wetlands* and *bare soil* are important factors with regard to risk and protection in malaria modelling. The effect of large-scale agriculture on malaria risk has always been controversial, as it has often been assumed that a high number of malaria vectors, resulting from irrigation schemes leads to increased malaria in local communities. However, recent studies in Africa reveal that, for many sites, there is instead less malaria in irrigated communities than in the surrounding areas. It has been suggested that communities near irrigation schemes would benefit from the greater wealth created and consequently use impregnated bed nets more commonly, have better access to improved healthcare and receive fewer infective bites compared with those residing outside such development schemes (Ijumba and Lindsay, 2001).

Within the MALAREO project a HR LULC map covering the study area at 5 m resolution was produced. A secondary outcome was an *enhanced* population map, obtained by the combination of the LULC layer with census data, aggregated at 100 m resolution. MODIS LULC categories (MR) were aligned with the MALAREO LULC categories and used for validation purposes in the prediction of the malaria parasitemia risk at locations in the MALAREO area.

The comparison showed that the model which used HR products (MALAREO LULC and DEM elevation) had a higher predictive ability than the one that used MR data. The data used for validation were the locations at which both MR and HR were available, *i.e*. the locations in Mozambique that were part of the MALAREO study area. Unfortunately, a randomly sampled validation set was not possible due to the scarcity of data. Spatially explicit estimates over the grids of 1 km, 500 m and 100 m showed large differences with respect to risk and its spatial pattern. This could be due to our results being sensitive to different allocation of MODIS categories to the final variable used for the model, and/or some local features might have been missed as the MODIS LULC layer is based on a global classification methodology. In particular, the *wetland* category showed the largest differences in the compar-

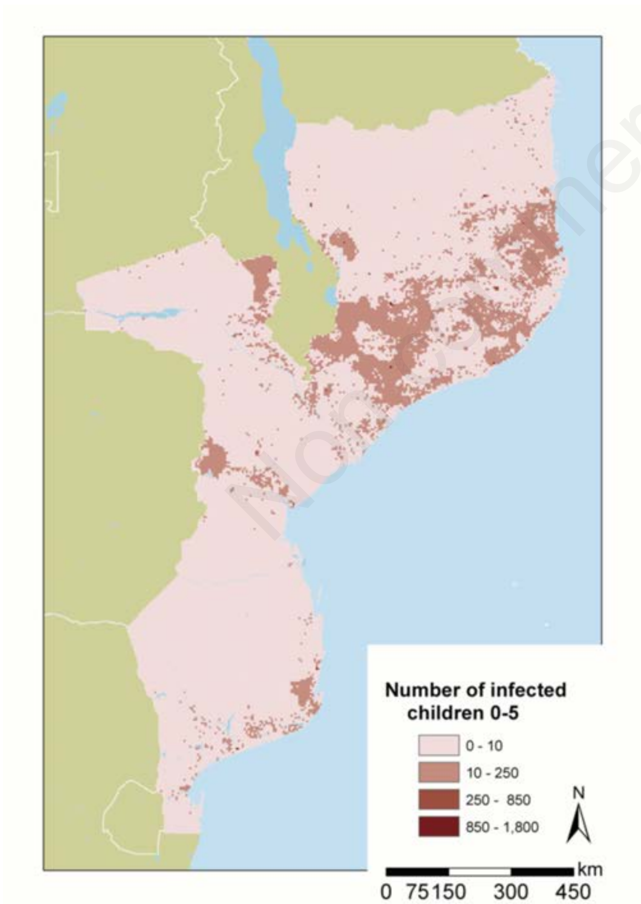

Figure 3. The predicted number of malaria-infected children up to 5 years of age. Median estimates are plotted at the 3-km resolution.

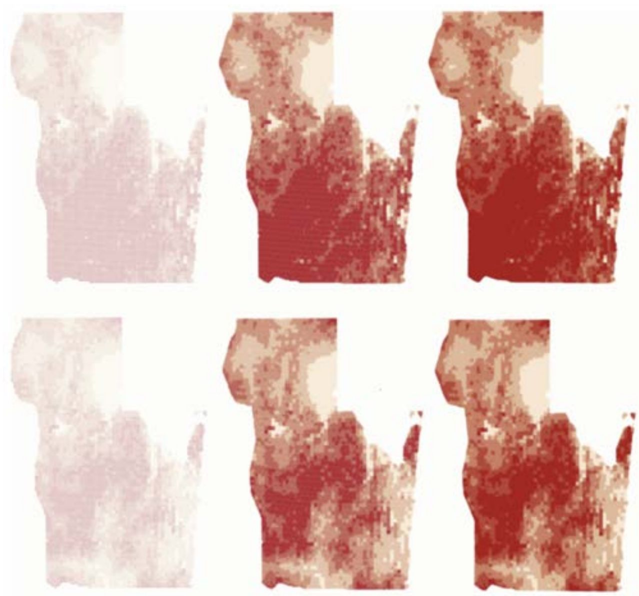

Figure 4. Predicted malaria risk (median) obtained by modelling with covariates. Moderate resolution covariates (first row); high-resolution covariates (second row) at spatial resolutions of 1 km (first column), 500 m (second column), and 100 m (third column).

ison with the MALAREO layer.

The MALAREO LULC layer is more accurate since the categories were assigned by a *supervised* algorithm (expert knowledge was incorporated) that allowed a more detailed description of the LULC. However, HR products like the MALAREO LULC are still expensive and may not be feasible over large areas, which will probably be overcome through future EO missions like the Sentinels and increased computational capabilities.

In this study, the estimated total number of infected children increased with increasing resolution of the predictive grid, which was independent from the spatial resolution of the covariates used for prediction. The use of MR variables tended to result in an overestimation of the number of infections. Observed differences between the 1 km resolution grid and the 500 m one using MR covariates were the result of aggregation of environmental covariates as well as population density over larger areas. However, the differences between the 500 m resolution grid and the 100 m one were only due to population density, as the MODIS LULC original resolution was 500 m.

Conclusions

Accurate estimation of malaria parasitemia risk has important implications on the planning of cost-effective control measures such as distribution of impregnated treated nets and IRS. The estimation of numbers of infected can further support National National Malaria Control MCPs in the determination of treatment needs.

References

Anderson JR, Hardy EE, Roach JT, Witmer RE, 1976. A land use and land cover classification system for use with remote sensor data. US Government Printing Office, Washington, DC, USA.

Ceccato P, Connor S, Jeanne I, Thomson M, 2005. Application of geographical information systems and remote sensing technologies for assessing and monitoring malaria risk. Parassitologia 47:81-96.

Clements AC, Reid HL, Kelly GC, Hay SI, 2013. Further shrinking the malaria map: how can geospatial science help to achieve malaria elimination? Lancet Infect Dis 13:709-18.

Craig M, Sharp B, Mabaso M, Kleinschmidt I, 2007. Developing a spatial-statistical model and map of historical malaria prevalence in Botswana using a staged variable selection procedure. Int J Health Geogr 6:44.

Craig M, Snow R, le Seur D, 1999. A climate-based distribution model of malaria transmission in Sub-Saharan Africa. Parasitol Today 15:105-11.

Diggle PJ, Tawn, J, Moyeed R, 1998. Model-based geostatistics. J Roy Stat Soc C-App 47:299-350.

Deleu J, Franke J, Gebreslasie M, Linard C, 2015. Improving AfriPop dataset with settlement extents extracted from RapidEye for the border region comprising South-Africa, Swaziland and Mozambique. Geospat Health 10:336.

Franke J, Bauwens I, Deleu J, de Montpellier C, Dlamini S, Gebreslasie M, Giardina F, Siegert, F, Vounatsou P, 2013. MALAREO MapAtlas. Exploring the spatial dimension of malaria and its explaining environmental factors in Southern Africa by Earth Observation. Available from: www.eurosense.com/documents/ pdf/news /malareo_mapatlas_web.pdf

Gebreslasie MT, Bauwens I, 2015. MALAREO: a user-driven project. Geospat Health 10:329.

Giardina F, Kasasa S, Sie A, Utzinger J, Tanner M, Vounatsou P, 2014. Effects of vector-control interventions on changes in risk of malaria parasitemia in sub-Saharan Africa: a spatial and temporal analysis. Lancet Glob Health 2:601-16.

Gosoniu L, Vounatsou P, Sogoba N, Maire N, Smith T, 2009. Mapping malaria risk in West Africa using a Bayesian nonparametric non-stationary model. Comput Stat Data An 53:3358-71.

Hay SI, Snow RW, 2006. The malaria atlas project: developing global maps of malaria risk. PLoS Med 3:e473.

Ijumba, J, Lindsay S, 2001. Impact of irrigation on malaria in Africa: paddies paradox. Med Vet Entomol 15:1-11.

Kalluri S, Gilruth P, Rogers D, Szczur M, 2007. Surveillance of arthropod vector-borne infectious diseases using remote sensing techniques: a review. PLoS Pathog 3:e116.

Linard C, Gilbert M, Tatem AJ, 2011. Assessing the use of global land cover data for guiding large area population distribution modelling. Geoj Lib 76:525-38.

Mabunda S, Casimiro S, Quinto L, Alonso P, 2008. A country-wide malaria survey in Mozambique. I. *Plasmodium falciparum* infection in children in different epidemiological settings. Malaria J 7:216.

Machault V, Vignolles C, Borchi F, Vounatsou P, Briolant S, Lacaux, JP, Rogier C, 2011. The use of remotely sensed environmental data in the study of malaria. Geospat Health 5:151-68.

Ministerio da Saúde-Instituto Nacional de Estatística, 2011. Moçambique. Inquérito demográfico e de saúde. Instituto Nacional de Estatística-Ministério da Saúde, Maputo, Moçambique. Available from: https://www.dhsprogram.com/pubs/pdf/FR266/FR266.pdf

Omumbo J, Hay S, Snow R, Tatem A, Rogers D, 2005. Modelling malaria risk in East Africa at high-spatial resolution. Trop Med Int Health 10:557-66.

Plummer M, 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd International Workshop on Distributed Statistical Computing, pp 20-22.

Riedel N, Vounatsou P, Miller JM, Gosoniu L, Chizema-Kawesha E, Mukonka V, Steketee RW, 2010. Geographical patterns and predictors of malaria risk in Zambia: Bayesian geostatistical modelling of the 2006 Zambia national malaria indicator survey (ZMIS). Malaria J 9:37.

Robert CP, 1996. Intrinsic losses. Theor Decis 40:191-214.

Stefani A, Dusfour I, Corrêa APS, Cruz MC, Dessay N, Galardo AK, Galardo CD, Girod R, Gomes MS, Gurgel H, Lima AC, Moreno ES, Musset L, Nacher M, Soares AC, Carme B, Roux E, 2013. Land cover, land use and malaria in the Amazon: a systematic literature review of studies using remotely sensed data. Malaria J 12:1-8.

Tatem AJ, Noor AM, von Hagen C, Di Gregorio A, Hay SI, 2007. High resolution population maps for low income nations: combining land cover and census in East Africa. PLoS One 2:e1298.

WHO, 2014. World malaria report: 2014. World Health Organization, Geneva, Switzerland. Available from: http://www.who.int/malaria /publications/world_malaria_report_2014/en/

Xie P, Arkin PA, 1997. A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. B Am Meteorol Soc 78:2539-58.