Title

Local climate and social inequality drive spatio-temporal variation in malaria risk across a complex urban landscape

Short title:

Urban malaria across scales

Authors

Mauricio Santos-Vega[1,2], Rachel Lowe [3,4,5], Luc Anselin [6], Vikas Desai [7], Keshav G. Vaishnav [8], Ashish Naik [9], Mercedes Pascual[1*].

Affiliations:

Affiliations

1. Department of Ecology and Evolution, University of Chicago, Chicago, United States of America.
2. Departamento de Ingeniería Biomédica, Grupo de Investigación en Biología Matemática y Computacional BIOMAC, Universidad de los Andes, Bogotá Colombia.
3. Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, United Kingdom,
4. Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom,
5. Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain
6. Center for Spatial Data Science, University of Chicago, Chicago, United States of America.
7. Urban Health and Climate Resilience Center of Excellence (UHCRCE) Surat VBDC, Health Department, Surat Municipal Corporation.
8. Insecticide officer, Vector Borne Diseases Control Department, Surat Municipal Corporation
9. MOH, Health Department, Surat Municipal Corporation

*Corresponding author: mpsacual@uchicago.edu

Abstract

Rapid urbanization makes cities an increasingly important habitat for mosquito-borne infections. Although these diseases are both climate and poverty driven, the interaction of environmental and socio-economic factors at different spatial scales within cities remain poorly understood. We analyze 10 years of extensive surveillance dataset of falciparum malaria resolved at three different spatial resolutions for the city of Surat in Northwest India. The spatial pattern of malaria risk is found to be largely stationary in time. A Bayesian hierarchical mixed model that combines spatially explicit indicators of temperature, population density and poverty with interannual variability in humidity best explains and predicts these patterns. For the Indian subcontinent, which harbors a truly urban mosquito vector, malaria elimination should target disease

hotspots within cities. We show that urban malaria risk patterns are strongly driven by fixed spatial structures, highlighting the key role of social and environmental inequality and the need for targeted control efforts.

MAIN TEXT

Introduction

Urban areas have become the new dominant ecosystem around the world and are characterized by large spatial heterogeneity (Harpham 2009). Urbanization is a complex phenomenon that has been associated with rapid and pronounced environmental variation, such as flooding events caused by extreme rainfall or variation in the increase of sensible heat, which in turn can increase temperature by 2–10 °C in highly urbanized areas (Vittal et al. 2016, Shepperd et al. 2002). The remarkable expansion of cities has also increased heterogeneity in terms of population density, economy, and infrastructure, exacerbating inequalities in the provision of urban services (Ahern 2013, Pickett et al. 2017, Zhou et al. 2017). Pronounced socio-economic inequalities are evident in the unprecedented scale and growth of vast informal settlements (slums) in low and middle-income cities (Bolay 2006, Mitlin and Satterthwaite 2010). This heterogeneity is expected to have important consequences for the spatio-temporal population dynamics of vector borne diseases, such as malaria. However, there is limited understanding of the joint effects of climatic, demographic and socio-economic conditions on vector-borne disease risk across a range of spatial scales (Zhao et al. 2014, Mishra et al. 2015).

Our understanding of the complex relationship between urbanization and vector-borne disease transmission remains incomplete. Urbanization can alter important ecological and physiological parameters of mosquito vectors, which can promote the spread and emergence of mosquito-borne diseases. It can also improve infrastructure and environmental health, leading to better health care provision (Alirol et al 2011). Studies of urban malaria have focused on Africa where the disease remains a predominantly rural problem as the mosquito vectors have adapted to breed in rural environments (Donelly et al. 2005, Hay et al. 2005, Keiser et al 2004). Given the lack of suitable breeding sites for the vectors in African cities (Qi et al. 2012), the better access to health care services and an increased ratio of humans to mosquitoes (Hay et al. 2005), a common assumption is that urbanization reduces malaria transmission. Nevertheless, transmission continues to persist in cities, in some cases at even higher levels than in surrounding (rural) areas (De Silva and Marshall 2012).

In contrast to Africa, the Indian subcontinent harbors a truly urban vector, *Anopheles stephensi*, enabling the transmission of urban malaria for both parasites, *Plasmodium falciparum and Plasmodium vivax*. This mosquito breeds in various artificial containers within homes and construction sites (Murdock et al. 2014, Cator et al. 2013). Given the rapid urbanization rate in this region, there is growing interest in understanding

how the spatio-temporal structure of urban malaria reflects the socio-economic and environmental heterogeneities of large cities (Ahern et al. 2005, Santos-Vega et al. 2016). A previous study described within-city spatial patterns in malaria risk (Santos-Vega et al. 2016). However, the relative influence of several key drivers and how this varies across a range of spatial scales remain poorly understood, including the role population density humidity and the impact of high temperatures (Cator et al. 2013). The recent reports of *A. stephensi* in some locations in Africa (REFS) provides a warning of the possible future expansion of urban malaria beyond its current geographical distribution.

Climate variability and climate change are expected to impact urban areas in particular ways, and as such are major determinants of global health (Watts et al. 2015). Some authors have argued that at coarser spatial resolutions (e.g. >10 km), the effect of climate in urban areas could be negligible (Trusilova et al. 2013). Others have argued that even at coarse aggregated scales, the effects of climate change should be evident, not only in cities, but also in larger peri-urban areas (Georgescu et al. 2014). Finally, mosquito vectors could experience environmental variation at fine spatial scales (Potter et al. 2013, Pincebourde et al. 2016) and this variation would interact with local features such as housing density and material, vegetation cover, and distance to water (Afrane et al. 2004, Murdock et al. 2016). Since climatic variables at coarse scales are averages over large regions, land types and populations, they tend to hide extremes and can lead to spurious correlations with disease risk (Baker 2008, Crane and Daniere 1996). The level of aggregation at which infectious disease models should be formulated to best capture effects of spatial heterogeneity in climatic factors remains unclear.

Here, we investigate the spatial distribution of urban malaria risk and the influence of climatic, demographic, and socio-economic drivers in a large city in India. Based on an extensive surveillance data set, we document a spatial pattern of malaria risk that is largely stationary in time. Using a Bayesian hierarchical modeling framework, we compare thee different levels of aggregation within the city (zones, units and worker units corresponding to the different levels at which the reporting is performed, see Figure 1) and analyze the factors that explain patterns of spatial variation at these different spatial aggregations. By zooming in and out to higher and lower spatial resolution, we investigate if changes in spatial scale modify the association with climate and economic variables and the prediction accuracy of the model. Finally, we provide an evaluation of prediction accuracy for our best statistical model at these different resolutions over a 10-year time period. Implications of these findings for malaria control and elimination efforts in the Indian subcontinent are discussed.

[FIGURE 1]

Results

The pattern that results from ranking malaria risk based on incidence is largely stationary, as shown in Fig 1 and S1 Fig 1, with locations of high and low risk persisting over time, independent of the inter-annual

variation in total malaria cases. This temporal regularity of the spatial pattern suggests the existence of strong underlying determinants that are themselves largely stationary over the temporal scales of malaria variation considered here (Fig 2A and S1 Fig 1). This stationary pattern is supported by a significant spatial autocorrelation through time (Fig 2C). Results from the local indicators of the spatial association test (LISA) show that the units spatially associated with high malaria risk differ significantly between the center of the city and its periphery (Fig 2 B). Specifically, units in the central and northern part of the city exhibit (positive/positive) associations in malaria risk, whereas some units in the periphery show (negative/negative) associations. The clusters identified through spatial associations also show a striking regularity from season to season (Fig 1 B).

[FIGURE 2]

For the socioeconomic data from census and surveys in the city (S1 table 1), we first assessed if we could simplify the variation in a low dimensional space. The results from PCA showed that just the first three dimensions explained 80.9% of the variability among units. Dimensions 1, 2 and 3 respectively explained 45.1%, 27% and 3, 8.84% of the variance (Fig. 3 A). These three dimensions were retained for further analysis. Fig 3 A shows the distribution of units along the first three PCA axes (PC1 vs PC2 in the upper panel and PC2 vs PC3 in the lower panel), with greater dispersion along the horizontal than vertical axes. In the PCA bi-plot, the origin represents the average unit and the dispersion around it indicates how units differ in relation to this average. The contribution of each of the variables to the three components in the analysis are given in S1 Fig 2. These contributions show that PC1 largely represents economic level and is also correlated with the amount of water stored. PC2 is associated with labor and employment, likely representing the effect of movement and exposure in particular environments within the city, and PC3 exhibits a strong contribution from population density. Fig 3 B shows the spatial distribution for PC1 and PC3. The spatial pattern of the economic level and population density summarized by the PC closely matches that of disease risk (Fig 2 B and S1 Fig 1).

[FIGURE 3]

We then tested if the spatially stationary pattern in malaria risk was associated with variation in environmental factors (Relative humidity (RH) and temperature) and/or in socioeconomic and demographic indicators, including population density, as summarized in the three main components of the PCA axes (Fig 3). S1 Fig 3 A illustrates the temporal associations between ranked malaria risk and mean temperature and humidity, respectively. A statistically significant linear correlation was found only between malaria risk and humidity ($p < 0.01$). S1 Fig 3 B shows spatial correlations between the three main principal components and the mean ranked cases. The results also detect a significant positive correlation ($p<0.05$) of income level/water management and population density.

Goodness-of-fit metrics for models of increasing complexity are shown in Table 1. Comparisons between different models that account for, or neglect, the effect of climate, economy, demography, and neighborhood structure show that the model that best accounts for spatiotemporal variation in malaria risk/incidence based on WAIC (Watanabe-Akaike information criterion) is the one including the combined effects of temperature, humidity, population density, economy, and spatial random variations. Specifically, the effects of population density and economic level/water storage practices (represented by PC1) are significant when considered together with structured and unstructured random effects, explaining 56% of the variance based on the R^2 likelihood ratio test (Table 1). Of the 56% variance explained overall, 28% can be explained by temperature and a monthly autoregressive term of order one. An additional 10% is accounted for by including relative humidity. To test if the climate variables better explain malaria risk at the regional or local level, we fitted models with temperature and humidity that were either spatially resolved or averaged over the spatial units. The best fitting model incorporated aggregated humidity and spatially explicit temperature. These findings suggest that interannual variations in humidity explains temporal variation in malaria, whereas variations in temperature across the city strongly influences spatial malaria variability. Importantly, accounting for spatial fixed and random effects accounted for 17% of the total variance explained by the model.

[TABLE 1]

Table 2 summarizes the posterior mean parameter estimates for the best model. All parameters are significantly different from zero, with posterior distributions from the two chains well mixed and converged based on the Gelman-Rubin diagnostic (Table 2). Note that the overdispersion parameter of the negative binomial (i.e. the reciprocal of the scale parameter) has a posterior mean value of 2.519 with a 95% credible interval (CI) of [1.456, 3.243]. Thus, the estimated overdispersion parameter (θ) is significantly different from zero (the value expected for the Poisson special case of the negative binomial). Additionally, population density exhibits a positive and statistically significant association with malaria incidence. The effect of temperature is negative, consistent with the non-monotonic effect of temperature on the reproductive number R_0 for malaria at the high end of the temperature spectrum (Mordecai et al. 2013, Pharman et al. 2010). Increases in temperature above a certain threshold negatively affect mosquito and parasite physiological and demographic parameters that influence transmission intensity (Pharman et al. 2010). Finally, relative humidity shows a significant and important enhancing effect on malaria risk across the city.

[TABLE 2]

A comparison of modeled and observed malaria cases is illustrated in Fig 4. In general, the predicted patterns reflect the observations for individual units (Fig 4 C-F and S1 Fig S4) and for averages over the units (Fig 4 A-B and S1 Fig 6), although the best model tends to over predict the number of cases. The maps show

comparisons for observed and predicted cases at the unit level for 2008 (a high incidence year) (Fig 4 C and D), and for 2011 (a low incidence year) (C-F). We observe correct quantile predictions for 73% and 66% of the units respectively. The GLMM tends to under predict malaria in certain areas and is better able to capture instances of very high malaria risk across the southwest part of the city.

[FIGURE 4]

To compare the best model coefficients at different spatial levels (zones, units, workers), we refitted the unit model for 2008 to 2014 (the period of time for which we also have high resolution malaria data), and then fitted the model at both the aggregated zone level (7 zones) and the worker level (478 worker units). Table 3 shows changes in the values of the model coefficients and the level of significance of the coefficients for the best model (see Table 2) at the different spatial scales (parameter estimates are considered to be statistically significant if their 95% credible interval do not contain zero). Temperature and humidity show variation in their contribution across the spatial levels: when we disaggregate the system, the effect of temperature strengthens whereas the effect of humidity weakens. Also, the effect of the economy and population density intensifies at the highest resolution.

[TABLE 3]

Fig 5 shows the added value of modeling the system at different levels analyzed (zone, unit and workers unit). The model fitted at the unit level (32 units) appears to better capture the spatio-temporal dynamics of urban malaria. A model fitted at the zone level (7 zones) only performs better or adds information in 22% of the locations (Fig 5 A), and the model fitted at the workers level (486 areas) provides added value to 33% of the units. When the system is modeled at the zone level, malaria risk falls in the same observed quantile for only 29% of the locations. By contrast, when the system is disaggregated further, this proportion increases to 57% and 49% for the worker and unit levels respectively.

[FIGURE 5]

Discussion

Urban environments exhibit pronounced heterogeneity from rapid and unplanned urbanization (Santos-Vega et al. 2016, Reiner et al. 2012, Perkins et al. 2013). Our results underscore the importance of considering this spatial heterogeneity for urban malaria in the Indian subcontinent. This study extends previous findings of Santos et al. (2016) for a different, inland, city of Northwest India, by providing an explicit evaluation of the role of socioeconomic and demographic factors on the stationarity of the spatial distribution of the disease for a hierarchy of spatial scales. The models presented here build upon the results of Santos et al. (2016) and explicitly incorporate the effect of population density and economic level in the space-time patterns of urban

malaria. Importantly, we found that two climate covariates contribute differentially to explain the variation in the data: humidity helps in capturing inter-annual variability and peak timing of outbreaks; temperature explains some of the spatial variation synergistically with economic and demographic covariates. These effects are confirmed when we fit the models at different spatial scales and find the variation explained by temperature is greatest at the highest spatial resolution, while the variation explain by humidity is greater when aggregated across the whole city.

Our analyses show the presence of three distinct explanatory components in the social and economic covariates. The first component of the PCA is related to economic level and water management, including income, the quality and size of housing, and the amount of water stored, which can all influence the recruitment of mosquitoes (Sharma, 1996). Access to water is an important determinant for malaria in India, given that water is supplied irregularly, leading to water storage within houses, which in turn creates multiple breeding sites for the mosquito in overhead tanks, cisterns and cement tanks (Salje et al. 2016). The second component corresponds to labor and employment, and as such is likely related to human mobility (Romeo-Aznar et al. 2018). Vector movement is expected to act at very local scales since experiments have demonstrated that mosquitoes do not travel very far (Quraishi et al. 1965) and are likely to stay within the same residence for days (Pharman 2010). The third component relates to human density.

Importantly, population density was shown here to positively influence malaria risk. This finding opposes the typical expectation from mathematical models for the transmission of the disease in which the rate of individual infection (the force of infection) decreases with population. For urban malaria, higher population density could result in higher water storage concentrations in close proximity to people. Our result is consistent with Romeo-Aznar et al. (2018) who recently showed that linking vector abundance through mosquito recruitment to human population density in a transmission model can generate an increasing trend in the force of infection. This pattern arises when vector recruitment grows at a greater rate than human density, . In our statistical model, the effect of population density is accounted for explicitly and does not simply map onto that of other socio-economic variables.

Our study further shows a negative relationship between malaria cases and high temperatures. This complements the more commonly documented positive relationship near or below optimal temperature conditions for physiological and epidemiological parameters of the mosquito, and for the malaria parasite within the mosquito (Mordecai et al. 2013). The negative relationship at high temperatures, beyond this optimum, emphasizes the need to better understand the high end of the temperature spectrum on vector-borne infections, the part of physiological curve least studied (Cator et al. 2013, Mordecai et al. 2013). Humidity and temperature show contrasting effects in our model. A model that incorporates aggregated humidity (averaged for the whole city) and spatially explicit temperature performs better than one with both spatially resolved temperature and humidity. This difference could be explained by the strong dependence

of humidity on the winds, which can alter evaporation by changing water vapor in the air. Winds tend to vary at a regional scale making humidity change over large regions (Singh et al. 2007). Conversely, temperature can exhibit large variation within a city at the local level, given the pronounced heterogeneity of impervious surfaces, with differing radiative, thermal, aerodynamic, and moisture properties (Siraj et al. 2015, Arnfield 2003).

Our model captures the seasonal pattern and the main trends in interannual variation in malaria cases, by including climatic and socioeconomic covariates, and allowing for spatial dependency between areas. Overall, a model including only autoregressive effects (to account for seasonality) and climate covariates accounted for 29% of the variance in urban malaria risk. Economic covariates and population density explained an additional 15%, and finally spatial dependencies added the biggest increase in the explained variance, underscoring the importance of considering local dependencies. Although our spatio-temporal statistical model is able to predict the peak of epidemics and also to capture the inter-annual variation in malaria cases, it could be improved in several directions. In particular, an extended formulation could include: (1) mobility fluxes derived with movement models from the spatial distribution of the population, to replace the near-neighbor effects on transition probabilities; (2) the explicit effect of population density on vector abundance; (3) further analysis of the local effect of other environmental heterogeneities such as river discharge and soil moisture. Temporal changes of the city structure itself would also be informative, including changes in the local speed of urbanization and the development of informal settlements, and their implications for mobility, population distribution, and economic level.

The city has experienced strong malaria interventions in the last three decades reflected in the negative trend in the number of reported cases from the 1980s and 1990s to the 2000s. From 2000 onwards, however, Surat has experienced seasonal outbreaks whose overall level has remained fairly constant. The stationary pattern in space described here, together with major drivers of this variation, indicate that targeted control could help reduce transmission even further, and that control measures could be implemented ahead of the season based on known spatial heterogeneity. In particular, ongoing efforts to provide better access to water may contribute to reduce urban malaria transmission, and possibly other vector-borne infections. Although we were unable to separate here the correlated effects of poverty and water access/storage, this is an important area for further study. Ultimately, at longer time scales, a reduction in poverty concomitant with better access to water is fundamental to reduce and eliminate malaria not only within cities but at a more regional level. A deeper understanding of humidity and temperature effects on malaria transmission is key for determining impacts of future climate variability and climate change on the space-time dynamics of vector-borne diseases. Humidity and temperature are expected to increase under future climate projections for the Indian subcontinent (Edmonson et al. 2016), specifically in the northwest part of India, which should experience a rise in humidity (Vittal et al. 2016, Dai et al. 2017). Under this scenario, a better understanding of how

humidity influences malaria transmission in urban environments can inform India's target of malaria elimination by 2030.

Materials and Methods

Study site and data description

Surat is located on the banks of the Tapi River in the western part of India in the state of Gujarat (Fig 1). It is one of the fastest growing Indian cities due to immigration from various parts of Gujarat and other Indian states. In this largely semi-arid state where malaria is seasonally epidemic, Surat reports more than 1300 *Plasmodium falciparum* cases every year, and an even larger number of *Plasmodium vivax* cases. We concentrate our analyses on the former since it is largely the target of control efforts. The current burden of malaria for both parasites has been greatly reduced in the city after the higher levels of the 1990s. Extensive and dedicated control efforts (indoor residual spraying, breeding sites detection or insecticide-impregnated bed nets to prevent transmission) by the SMC keep malaria outbreaks to relatively low levels, but do not eliminate the problem. Malaria exhibits seasonal outbreaks of varying size. The city presents ideal characteristics to investigate urban malaria transmission and their association with climatic and socioeconomic conditions at different spatial scales, given pronounced environmental and socioeconomic disparities, as well as an established surveillance program.

We obtained malaria data from 2004 to 2014 from the Surat Municipal Corporation disaggregated at different levels of resolution corresponding to 7 zones, 32 units, and 478 'worker' units. In this city, epidemiological data is recorded at the worker unit level and combines active and passive surveillance. The active case detection (ACD) is done through house to-house surveys by malaria workers. The workers collect data every 15 days. The main responsibility of the malaria workers is to find a fever case in the last 15 days in the family, collect slides if a fever case is found and provide radical treatment through follow-up visits. For passive case detection (PCD), each worker collects the data reported by the sentinel centers (hospitals, urban health centers, laboratories, clinics) within their area.

[FIGURE 5]

We collated a multi-sourced spatio-temporal (climatic and socioeconomic) dataset using the statistical computing software R. We reconciled data of different types and aggregation levels (i.e. economic, demographic) to the gridded climate data. The interpolation approaches and description of the generated datasets are explained below.

Climate and socioeconomic data

Socioeconomic data

A database of candidate drivers of urban malaria risk in Surat was generated for each reporting unit (zones, units, and workers) for 2010 (the year of the most recent census) using a universal kriging method to generate an estimated interpolation surface for each covariate (Supplementary text 2). For this, we used census data from the District Census Handbook at the district level for the year 2010 from the Directorate of Census Operations, Gujarat. In addition, variables such population density and slum density were calculated using annual population estimates from the Surat Municipal Corporation (SMC). Data on urbanization, housing, water provision, water storage, sanitation, and water scarcity were obtained from water management surveys conducted by Taru Leading Edge, a leading development advisory company in India (http://taru.co.in/). This household survey included 80 locations and 400 households across the city.

Climate data

We obtained gridded (1km x 1km latitude longitude grid) monthly mean surface air temperature data from MODIS (Moderate Resolution Imaging Spectroradiometer https://modis.gsfc.nasa.gov/data/). To estimate surface relative humidity (RH) within the city, we used MODIS data and meteorological parameters obtained from ground-based measurements extracted from GSOD (Global Summary of the Day from NOAA https://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD&countryabbv&georegionabbv) and 10 meteorological stations located throughout the city that have measured daily temperature, humidity and dew point in the city of Surat since 2014 (S1 text 3.5.2, (Peng et al. 2006)).

Data analyses

To examine spatial patterns of malaria risk independently from the inter-annual variation of incidence, we accumulated reported cases for a given year and normalized this sum by the total yearly cases for the whole city. We then conducted univariate statistical analyses to evaluate the spatial dependency of the malaria cases (Fig 1 B and C). First, a univariate Moran Index (Moran's I) was computed through time (Anselin 1988, Anselin 1996) (Fig 1 C). Moran's I identifies the global degree of spatial association (or how much the magnitude of an indicator in one location is influenced by the magnitude of the indicator in an area close to it) (Anselin 1996, Anselin 2001). Then the Local Moran statistic was obtained to identify local clusters and spatial outliers (Anselin 1995) (Fig 1 A). This indicator identifies units with spatial association in malaria incidence. Depending on the sign of the indicator (positive or negative), these local associations can express positive-positive, positive-negative, negative-positive or negative-negative associations. Where positive-positive and negative-negative associations represent spatial clusters and positive-negative, negative-positive are spatial outliers. Only positive-positive associations are considered and represent locations with significant positive local spatial autocorrelation are the core of a cluster (actual “cluster” includes neighbors as well as core).

In order to reduce the dimensionality of the socioeconomic variables (S1 Table 1) and to address the existence of a spatial pattern in these indicators, we used principal component analysis (PCA) to find the best low-dimensional representation of the variation in this multivariate data set. Graphical visualization reveals organization of the variables in a low-dimensional space as well as the spatial organization of the variables (Fig 2 A). We explored the spatial distribution of the components that account together for more than the 80% of the variance by plotting the spatial distribution of the factor loadings of each component. Then we evaluated the association between the loading values of each PC and the mean malaria rankings (Fig S1 3 B). We examined the environmental data for the existence of temporal associations between the interannual variation in relative humidity, temperature and malaria rankings (Fig S1 3 A). To explore spatial relationships between the malaria annual cases and the climate factors, we conducted a bivariate Moran analysis. This index shows the correlation between one variable (malaria incidence) at a location, and a different variable (temperature and humidity) at neighboring locations (S1 Table 2).

Statistical models

A hierarchical model framework was applied to assess the relative contribution of climatic and socioeconomic factors to spatiotemporal urban malaria cases. Generalized linear mixed models (GLMM) were formulated, including random effects to account for unobserved confounding factors, such as additional spatial heterogeneity, quality of health care services, and local health interventions. A negative binomial model was used to allow for overdispersion found in the urban malaria count data. We also incorporated: (i) a spatially unstructured random effect to introduce an extra source of variability/overdispersion in space (a latent effect), and (ii) a spatially structured random effect to explicitly account for spatial autocorrelations and weight relative risk in a region according to the relative risks in neighboring regions. This is consistent with the effect of increased infectious disease risk from neighboring regions of high transmission introduced in both mathematical (Longini et al. 1988, Viboud et al. 2006, Bertuzzo et al. 2011) and statistical models (Lowe et al. 2013, Lowe et al. 2016).

The general form of the GLMM is as follows:

$$Y_{it} \sim NegBin(\mu_{it}, \theta) \tag{1}$$

$$Log(\mu_{it}) = \mathrm{Log}(\mathrm{p}_{si}) + \alpha + \sum_{j=1}^{2} \beta_j X_{jit} + \sum_{j=1}^{3} \gamma_j Y_{ji} + Z_{(t-1)} + v_i + \phi_i \tag{2}$$

Where μ_{it} are the malaria counts for each administrative unit (here we show an example of the model structure at the level of the 32 units, but this structure holds for the 7 zone and 486 worker units), i (i = 1,…, 32 and time t (t = 1,…, 132), where the population p_{it} is treated as an offset. The variables X_{it} represent the selected climate influences: temperature one month earlier (j=1) and humidity two months earlier (j=2). The variables Y_{ji} correspond to the components PC1 (j=1), PC2 (j=2) identified by the PCA analysis, described above, and to population density (j=3). As such, they vary only in space and not in time. They summarize the spatial heterogeneity in socio-economic and demographic conditions. Z_{t-1} represents a first order autoregressive term to account for temporal autocorrelation in the data. An independent diffuse Gaussian exchangeable prior was assumed for the unstructured random effect (ϕ_i), as well as v_i, a normal conditional autoregressive (CAR) prior distribution for the spatially structured random effects v_i:

$$v_i / v_i \sim N\left(\frac{\sum_j a_{ij} v_j}{\sum_j a_{ij}}, \frac{\sigma_i^2}{\sum_j a_{ij}}\right) \tag{3}$$

where $\sigma^2{}_i$ controls the strength of local spatial dependence, and a_{ij} are neighborhood weights for each unit as defined above, with simple binary values of 1 when unit i is a neighbor of unit j, and 0 otherwise. Since the CAR distribution is improper, we applied a 'sum to zero' constraint on each v_i. First, we fitted the Bayesian model at the intermediate level (32 units), via MCMC sampling implemented in R in conjunction with the OpenBUGS software (Sturtz et al. 2015). We generated two parallel MCMC chains, each of length 25,000 with a burn-in of 20,000 and a thinning of 10, to obtain 1000 samples from the joint posterior distribution. Convergence was assessed by inspecting plots of traces of simulations for individual parameters and monitoring the Gelman-Rubin diagnostic (Geman and Rubin 1992). We standardized the fixed explanatory variables for humidity, temperature, and population density to zero mean and unit variance, helping MCMC convergence. Model comparison and evaluation of goodness-of-fit of all the models was assessed at the intermediate level of aggregation (32 units) using different quantities, the deviance information criterion (DIC) (Spiegelhalter et al. 2002), the Watanabe-Akaike information criterion (Watanabe, 2010), and an R^2_{LR} statistic for mixed effects models, based on a likelihood ratio (LR) test between the candidate model and an intercept only (null) model (Kramer 2005; Magee 1990). Smaller values of DIC indicate a better-fitting model, whereas the likelihood ratio R^2_{LR} ranges from zero to 1, with 1 corresponding to a perfect fit for any reasonable model specification (Lowe et al. 2015).

Scale dependency

Based on the best model identified at the intermediate level of aggregation, we then tested the extent to which the association and significance of the coefficients in the model varied between different levels of spatial resolution. To this end, we first fitted the previously identified best model at both a higher spatial

resolution (478 worker units) and a lower resolution (7 zones) with the same approach outlined above. Then, for each of the resulting models we checked the convergence of the individual parameter estimates and calculated the potential scale reduction (see (Gelman et al. 2004) and note that values below 1.1 are considered to be acceptable in most cases). We also evaluated changes in the 95% credible interval (CI) to see if the interval contained zero. If the CI does not contain zero, the covariates contribute significantly to the model fit (Table 3).

Model comparisons

To assess the predictive ability of the best model, posterior predictive distributions of malaria incidence were obtained for each unit and month. New pseudo-observations were simulated by drawing random values from a negative binomial distribution with mean and scale parameter estimated using 10.000 samples from the posterior distribution of the parameters in the model and computing the median cases from the simulations. To summarize this information, the observed and posterior predictive mean malaria risk estimates were aggregated across space, and predictions for each unit in high and low incidence years were generated (Fig A-B). For the best model, we compared the temporal evolution of the fitted posterior median cases with the observed cases for Surat as a whole (S1 Fig 3.6).

We also compared the proportion of places in which the model accurately predicts malaria incidence. For this, we classified the cases into 5 categories, generated by considering all zero cases in one class and by subdividing all remaining non-zero cases into four equally sized intervals. The resulting categories correspond to no cases, very low, low, high and very high cases. We then mapped the categories of the spatial predictions and observed cases for 2008 (Fig 3 C-D) and 2011 (Fig 3 E-F), quantifying the number of times the values correctly matched. We also compared predictive accuracy across spatial scales (zones, units, workers) by calculating the root mean squared error (RMSE), a measure of the difference between modeled and observed values, over the 10-year time period and for each reporting unit (Lowe et al. 2013). Smaller values of RMSE indicate a better fitting model. Regions with positive values (RMSE null - RMS alternate > 0) indicate that modeling the system at the new scale locally improves the estimation of malaria relative risk.

Acknowledgments

We thank the commissioner of Surat Municipal Corporation for supplying the malaria data for the city. We also thank Dr Menno Bouma for insightful discussions in the early stages of this project. We thank Vimal Mishra from the Gandhinagar Institute of technology for the technical advice on the climatological data and patterns. The research was funded by the University of Chicago. RL was supported by a Royal Society

Dorothy Hodgkin Fellowship. We thank the Gardner High Performance Computing (HPC) cluster from the Center for Research Informatics of the University of Chicago for computational resources.

References

1. Harpham T. Urban health in developing countries: What do we know and where do we go? Heal Place. 2009;15(1):107–16.
2. Vittal H, Ghosh S, Karmakar S, Pathak A, Murtugudde R. Lack of Dependence of Indian Summer Monsoon Rainfall Extremes on Temperature: An Observational Evidence. Sci Rep. 2016;6. Shepperd et al . 2002
3. Shepherd JM, Pierce H, Negri AJ. Rainfall Modification by Major Urban Areas: Observations from Spaceborne Rain Radar on the TRMM Satellite. J Appl Meteorol [Internet]. 2002;41(7):689–701.
4. Ahern J. Urban landscape sustainability and resilience: The promise and challenges of integrating ecology with urban planning and design. Landsc Ecol. 2013;28(6):1203–12.
5. Pickett STA, Cadenasso ML, Rosi-Marshall EJ, Belt KT, Groffman PM, Grove JM, et al. Dynamic heterogeneity: a framework to promote ecological integration and hypothesis generation in urban systems. Urban Ecosyst. 2017;20(1).
6. Zhou W, Pickett STA, Cadenasso ML. Shifting concepts of urban spatial heterogeneity and their implications for sustainability. Landsc Ecol. 2017;32(1):15–30.
7. Bolay JC. Slums and urban development: Questions on society and globalisation. Eur J Dev Res. 2006;18(2):284–98.
8. Mitlin D, Satterthwaite D. (2013). Urban Poverty in the Global South. London: Routledge, https://doi.org/10.4324/9780203104316
9. Mishra V, Ganguly AR, Nijssen B, Lettenmaier DP. Changes in observed climate extremes in global urban areas. Environ Res Lett. 2015;10.
10. Zhao L, Lee X, Smith RB, Oleson K. Strong contributions of local background climate to urban heat islands. Nature. 2014;511(7508):216–9.
11. Alirol E, Getaz L, Stoll B, Chappuis F, Loutan L. Urbanisation and infectious diseases in a globalised world. Vol. 11, The Lancet Infectious Diseases. 2011. p. 131–41.
12. Donnelly MJ, McCall PJ, Lengeler C, Bates I, D'Alessandro U, Barnish G, et al. Malaria and urbanization in sub-Saharan Africa. Malar J. 2005;4(1):12
13. Hay SI, Guerra C a, Tatem AJ, Atkinson PM, Snow RW. Urbanization, malaria transmission and disease burden in Africa. Nat Rev Microbiol [Internet]. 2005;3(1):81–90.
14. Keiser J, Singer BH, Utzinger J. Reducing the burden of malaria in different eco-epidemiological settings with environmental management: A systematic review. Vol. 5, Lancet Infectious Diseases. 2005. p. 695–708.

15. Qi Q, Guerra C a, Moyes CL, Elyazar IRF, Gething PW, Hay SI, et al. The effects of urbanization on global Plasmodium vivax malaria transmission. Malar J [Internet]. 2012;11(1):403.
16. De Silva PM, Marshall JM. Factors contributing to urban malaria transmission in sub-saharan Africa: A systematic review. Journal of Tropical Medicine. 2012.
17. Murdock CC, Blanford S, Luckhart S, Thomas MB. Ambient temperature and dietary supplementation interact to shape mosquito vector competence for malaria. J Insect Physiol. 2014;67:37–44.
18. Cator LJ, Thomas S, Paaijmans KP, Ravishankaran S, Justin J a, Mathai MT, et al. Characterizing microclimate in urban malaria transmission settings: a case study from Chennai, India. Malar J [Internet]. 2013;12(1):84.
19. Ahern M, Kovats RS, Wilkinson P, Few R, Matthies F. Global health impacts of floods: Epidemiologic evidence. Epidemiol Rev. 2005;27:36–46.
20. Santos-Vega M, Bouma MJ, Kohli V, Pascual M. Population Density, Climate Variables and Poverty Synergistically Structure Spatial Risk in Urban Malaria in India. PLoS Negl Trop Dis. 2016;10(12).
21. Rezza G, Nicoletti L, Angelini R, Romi R, Finarelli A, Panning M, et al. Infection with chikungunya virus in Italy: an outbreak in a temperate region. Lancet. 2007;370(9602):1840–6.
22. Smith DL, Dushoff J, McKenzie FE. The risk of a mosquito-borne infection in a heterogeneous environment. PLoS Biol [Internet]. 2004;2(11):e368.
23. Acevedo MA, Prosper O, Lopiano K, Ruktanonchai N, Caughlin TT, Martcheva M, et al. Spatial heterogeneity, host movement and mosquito-borne disease transmission. PLoS One [Internet]. 2015;10(6):e0127552.
24. Cosner C, Beier JC, Cantrell RS, Impoinvil D, Kapitanski L, Potts MD, et al. The effects of human movement on the persistence of vector-borne diseases. J Theor Biol. 2009;258(4):550–60.
25. Watts N, Adger WN, Agnolucci P, Blackstock J, Byass P, Cai W, et al. Health and climate change: Policy responses to protect public health. Vol. 386, The Lancet. 2015. p. 1861–914.
26. Trusilova K, Früh B, Brienen S, Walter A, Masson V, Pigeon G, et al. Implementation of an urban parameterization scheme into the regional climate model COSMO-CLM. J Appl Meteorol Climatol. 2013;52(10):2296–311.
27. Georgescu M, Morefield PE, Bierwagen BG, Weaver CP. Urban adaptation can roll back warming of emerging megapolitan regions. Proc Natl Acad Sci [Internet]. 2014;111(8):2909–14.
28. Potter KA, Arthur Woods H, Pincebourde S. Microclimatic challenges in global change biology. Glob Chang Biol. 2013;19(10):2932–9.
29. Pincebourde S, Murdock CC, Vickers M, Sears MW. Fine-Scale Microclimatic Variation Can Shape the Responses of Organisms to Global Change in Both Natural and Urban Environments. Integr Comp Biol [Internet]. 2016;56(1):45–61

30. Afrane YA, Klinkenberg E, Drechsel P, Owusu-Daaku K, Garms R, Kruppa T. Does irrigated urban agriculture influence the transmission of malaria in the city of Kumasi, Ghana? Acta Trop. 2004;89(2):125–34.
31. Murdock CC, Sternberg ED, Thomas MB. Malaria transmission potential could be reduced with current and future climate change. Sci Rep. 2016;6.
32. Baker JL. Urban Poverty : A Global View. World Bank Gr. 2008;v-27.
33. Crane R, Daniere A. Measuring access to basic services in global cities: Descriptive and behavioral approaches. J Am Plan Assoc. 1996;62(2):203–21.
34. Mordecai EA, Paaijmans KP, Johnson LR, Balzer C, Ben-Horin T, de Moor E, et al. Optimal temperature for malaria transmission is dramatically lower than previously predicted. Ecol Lett. 2013;16(1):22–30.
35. Parham PE, Michael E. Modeling the effects of weather and climate change on malaria transmission. Env Heal Perspect [Internet]. 2010;118(5):620–6.
36. Reiner Jr. RC, King AA, Emch M, Mohammad Y, Faruque ASG, Pascual M. Highly localized sensitivity to climate forcing drives endemic cholera in a megacity. Proc Natl Acad Sci U S A [Internet]. 2012;109(6):2033–6.
37. Perkins TA, Scott TW, Le Menach A, Smith DL. Heterogeneity, Mixing, and the Spatial Scales of Mosquito-Borne Pathogen Transmission. PLoS Comput Biol. 2013;9(12).
38. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math. 1991;43(1):1–20.
39. Sharma VP. Re-emergence of malaria in India. Vol. 103, Indian Journal of Medical Research. 1996. p. 26–45.
40. Salje H, Lessler J, Paul KK, Azman AS, Rahman MW, Rahman M, et al. How social structures, space, and behaviors shape the spread of infectious diseases using chikungunya as a case study. Proc Natl Acad Sci [Internet]. 2016;113(47):13420–5.
41. Quraishi MS, Mofidi MH, Motabar M, Ergul C, Ghadirian E, Rizvi SA, et al. Determination of flight range and length of gonotrophic cycles in nature in Anopheles stephensi. World Heal Organ. 1965;1:21.
42. Siraj AS, Bouma MJ, Santos-Vega M, Yeshiwondim AK, Rothman DS, Yadeta D, et al. Temperature and population density determine reservoir regions of seasonal persistence in highland malaria. Proc R Soc B Biol Sci [Internet]. 2015;282(1820):20151383.
43. Arnfield AJ. Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. Int J Climatol. 2003;23:1–26.
44. Romeo-Aznar V, Paul R, Telle O, Pascual M. Mosquito-borne transmission in urban landscapes: the missing link between vector abundance and human density," Proceedings of the Royal Society of London B: Biological Sciences. 2018. vol. 285, no. 1884.

45. Singh P, Kumar V, Thomas T, Arora M. Changes in rainfall and relative humidity in river basins in northwest and central India. Hydrol Process. 2008;22(16):2982–92.
46. Peng G, Li J, Chen Y, Norizan AP, Tay L. High-resolution surface relative humidity computation using MODIS image in Peninsular Malaysia. Chinese Geogr Sci. 2006;16(3):260–4.
47. Anselin L. Spatial Econometrics: Methods and Models [Internet]. Operational Regional Science Series. 1988. 289 p.
48. Anselin L. The Moran Scatterplot as an ESDA tool to assess local instability in spatial association. In: Spatial analytical perspectives on GIS. 1996. p. 111–25.
49. Anselin L. Spatial econometrics. A companion to Theor Econom [Internet]. 2001;310–30. Available from: http://edutexts.org/tw_files2/urls_33/60/d-59764/7z-docs/1.pdf
50. Longini IM, Koopman JS, Haber M, Cotsonis GA. Statistical inference for infectious diseases. Risk-specific household and community transmission parameters. Am J Epidemiol [Internet]. 1988;128(4):845–59.
51. Viboud C, Bjørnstad ON, Smith DL, Simonsen L, Miller MA, Grenfell BT. Synchrony, waves, and spatial hierarchies in the spread of influenza. Science (80-). 2006;312(5772):447–51.
52. Bertuzzo E, Mari L, Righetto L, Gatto M, Casagrandi R, Blokesch M, et al. Prediction of the spatial evolution and effects of control measures for the unfolding Haiti cholera outbreak. Geophys Res Lett. 2011;38(6).
53. Lowe R, Bailey TC, Stephenson DB, Jupp TE, Graham RJ, Barcellos C, et al. The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. Stat Med. 2013;32(5):864–83.
54. Lowe R, Cazelles B, Paul R, Rodó X. Quantifying the added value of climate information in a spatio-temporal dengue model. Stoch Environ Res Risk Assess. 2016;30(8):2067–78.
55. Sturtz S, Ligges U, Gelman A. R2WinBUGS: a package for running WinBUGS from R. J Stat Softw. 2005;12(3):1–16.
56. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. Stat Sci [Internet]. 1992;7(4):457–72.
57. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. J R Stat Soc Ser B Stat Methodol. 2002;64(4):583–616.
58. Kramer M, Service BC, Beltsville a RS. R2 Statistics for Mixed Models. Proc Conf Appl Stat Agric [Internet]. 2005;148–60.
59. Magee L. R2measures based on wald and likelihood ratio joint significance tests. Am Stat. 1990;44(3):250–3.
60. Edmondson JL, Stott I, Davies ZG, Gaston KJ, Leake JR. Soil surface temperatures reveal moderation of the urban heat island effect by trees and shrubs. Sci Rep. 2016;6.
61. Dai A. Precipitation characteristics in eighteen coupled climate models. J Clim. 2006;19(18):4605–30.

Figures and Tables

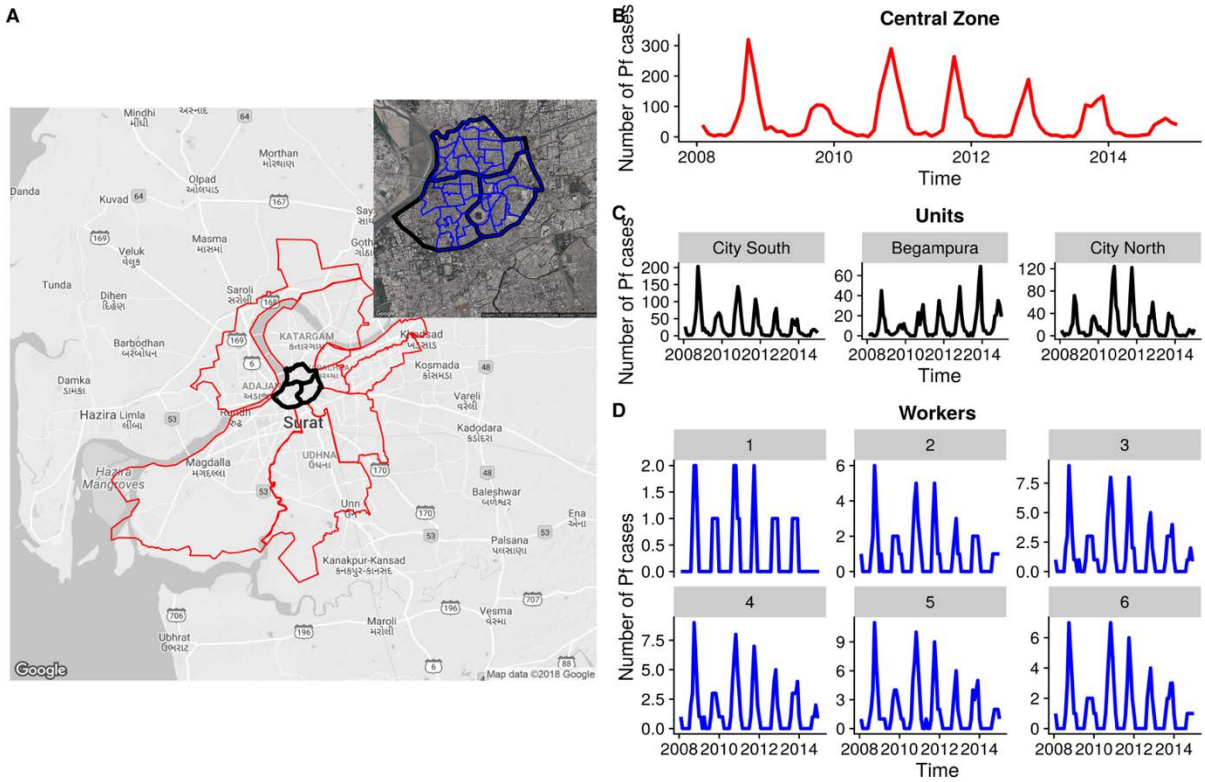

Figure 1. Location of the study area (A). The 7 administrative zones of the city are depicted in red, and for comparison some of the units (of a total of 32) are also shown in black for the central zone. Embedded within these, the subplot in the upper right illustrates the smallest reporting scale, that of the worker units. Time series of *P.falciparum* cases for (B) the central zone, (C) the three units and (D) the six worker units within the central zone.

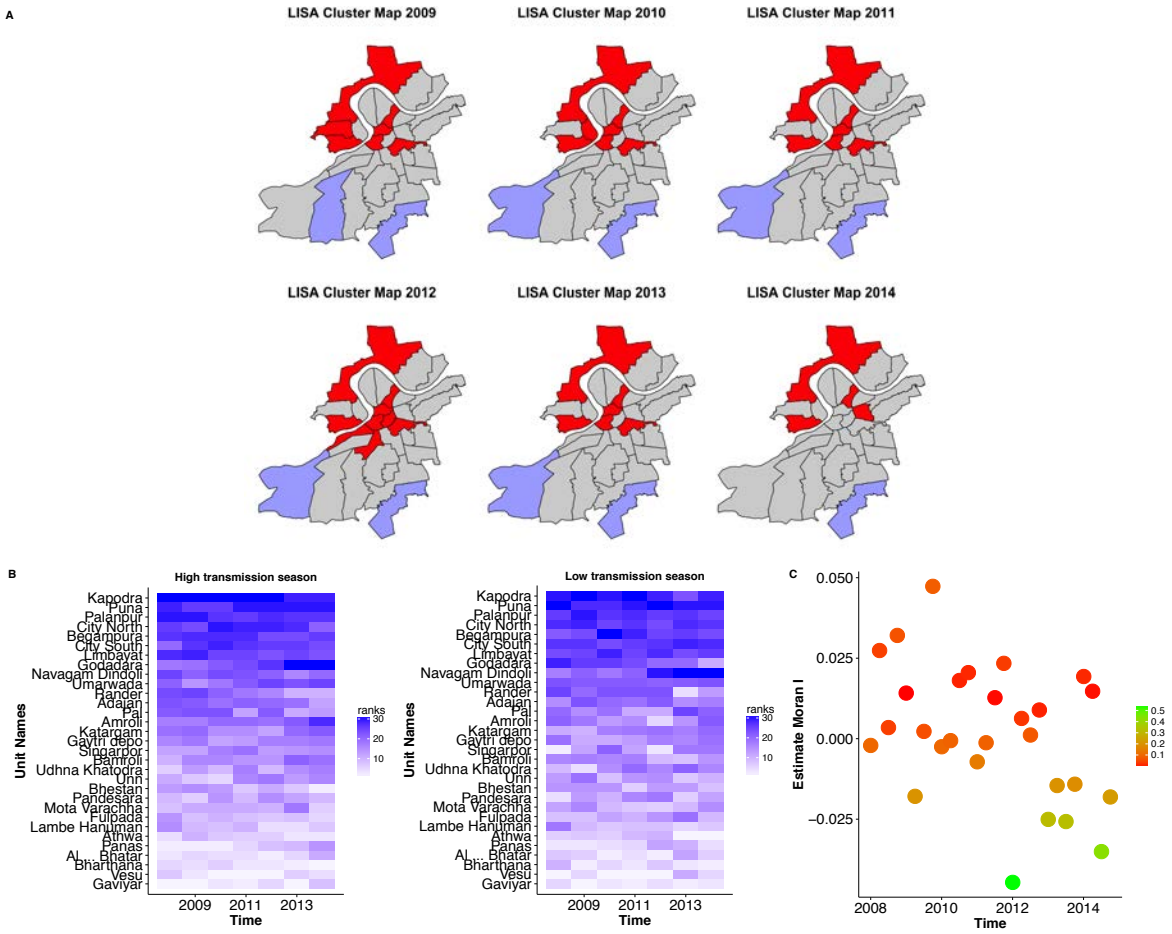

Figure 2. Spatial patterns. (A) Clusters in malaria incidence for the period 2008-14, identified using the local indicators of spatial associations (LISA) test/analysis. The resulting LISA cluster maps depict locations of significant Local Moran's I statistics, classified by type of spatial association, with red corresponding to positive/positive associations (locations where high incidence is surrounded by high incidence) and light blue, to negative/negative associations (locations where low incidence is surrounded by low incidence). Clusters are significant at $p = 0.05$ (based on 9999 permutations). (B) Distribution of cases normalized by population, with color intensity (from low [white] to high [blue]) corresponding to the ranking of incidence at the levels of the units. (C) The Moran's I autocorrelation index is calculated bimonthly from 2008-2014. Red dots represent significant values ($p < 0.05$) of spatial autocorrelation.

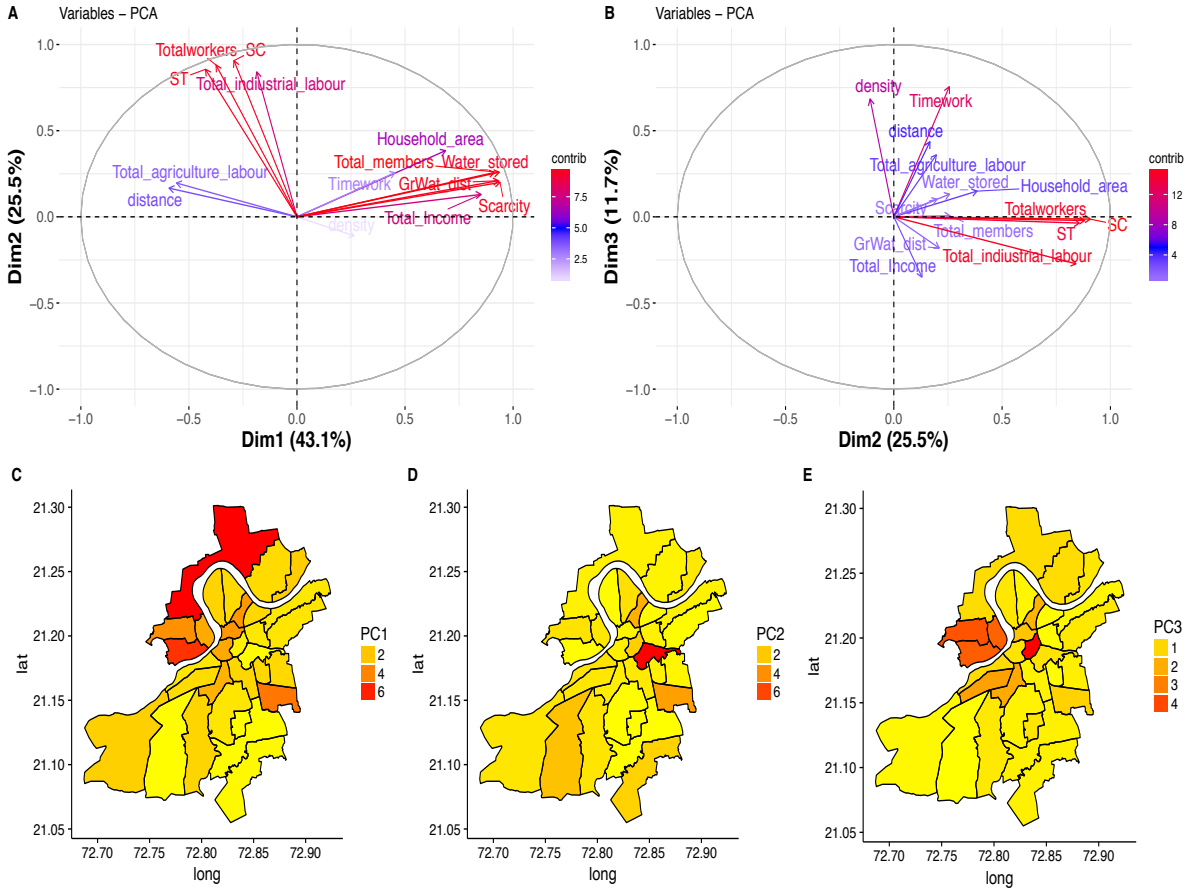

Figure 3. PCA Analysis and PCA spatial distribution. A-B) Distribution of the units in the first-second and second-third dimensions of the PCA. The colors from light blue to red show the respective contributions of the variables to the principal components; (C-E) The maps show the spatial distribution of PC1-PC3 loadings, with color intensity (from low [yellow] to high [red]) corresponding to the values of the PCAs.

Table 1. Comparison of goodness of fit. Comparison of goodness of fit for the different models tested based on different model selection criteria (DIC, WAIC and R^2 based on a likelihood ratio test). Global humidity or temperature correspond to averages over all the units; whereas spatial humidity and temperature indicated their local value for the 32 units (Smaller DIC and WAIC values signify better fit.)

Model	**DIC**	**R^2LR**	**WAIC**
Fixed effects only climate factors			
(A) Autoregressive + Spatial temperature	4937	0.28	821.63
(B) Autoregressive + Global temperature + Global Humidity	4934	0.32	818.32
(C) Autoregressive + Spatial temperature + Global humidity	4927	0.38	815.85

(D) Autoregressive +Spatial temperature+ spatial humidity	4932	0.34	816.43
(E) Autoregressive +Polynomial Temperature+ spatial humidity	4935	0.31	819.83
Fixed effects, climate factors and socioeconomic factors			
(F) Autoregressive +Climate factors+ PC1	4919	0.41	811.22
(G) Autoregressive +Climate factors+ PC1+PC2	4920	0.40	812.41
(H) Autoregressive +Climate factors+ PC1+ population density	4917	0.43	808.52
Climate factors, socioeconomic factors and random effects			
(I) Autoregressive + Climate +PC1 + Population density	4896	0.56	802.76

Table 2. Parameter estimates for the coefficients associated to variables included in the selected model (i.e. explaining more variation in urban malaria cases). The CI correspond to the 5% and 95% quantiles of the marginal posterior distribution. Each value of $\hat{R}$ correspond to the Gelman-Rubin diagnostic for chains convergence.

Covariate	**Mean**	**95% CI**	$\hat{R}$
Temperature	-0.2001	[-0.121, -0.267]	1.005
Relative humidity	0.6299	[0.445, 0.823]	1.002
Population density	0.1055	[0.034, 0.226]	1.001
Socioeconomic (PC1)	0.546	[0.369, 0.723]	1.002
Spatially unstructured hyperparameter	0.3257	[0.134, 0.456]	1.001
Spatially structured hyperparameter	0.026	[0.012, 0.16]	1.001
Lagged Malaria (lagged term)	1.4270	[0.9674, 2.465]	1.001
Overdispersion parameter	2.5190	[1.4567, 3.243]	1.001

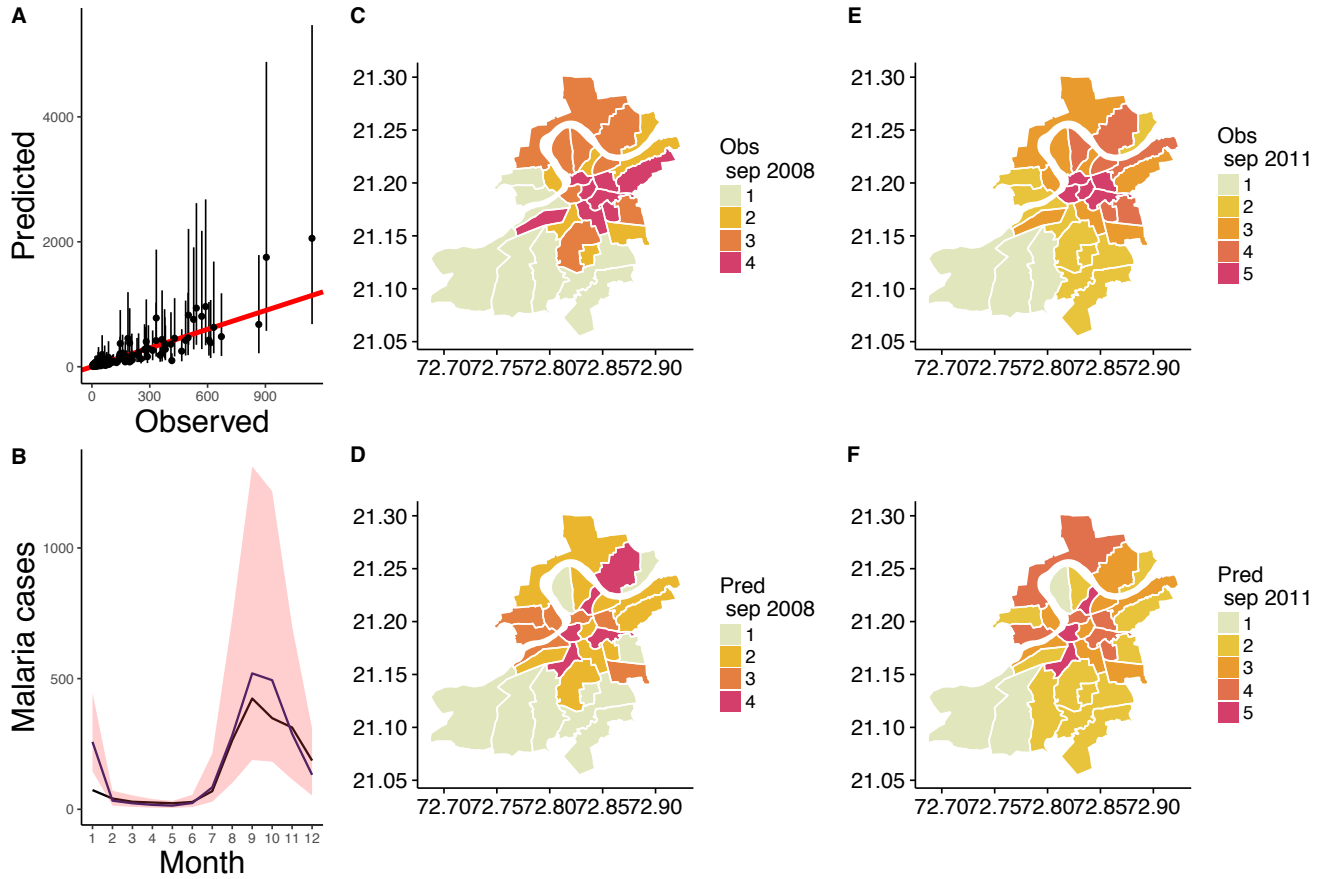

Figure 4. Observed versus predicted *Plasmodium falciparum* (PF) cases. A) Total malaria cases. The identity line (in red) is shown for comparison. B) Seasonal pattern for the observed cases (red) and the median of 1000 model simulations (blue). The 5% and 95% percentiles of the simulated data are shaded in light red. Comparison of observed and predicted cases for 2008 (C-D) and 2011 (E-F). The colors in the maps progress from yellow to red based on quantiles generated by considering all zero cases in one class, and by subdividing all remaining non-zero cases into four intervals. The resulting 5 categories correspond to no cases, very low, low, high and very high cases. Comparisons of the quantiles in the maps reveal that 73% and 66% of their values correctly match for years 2008 and 2011, respectively.

Table 3. Model comparisons. Comparison between the beta coefficients fitted at three levels of aggregation (zone, unit and workers unit) (significant values are depicted in bold). The colors represent whether the credible interval includes zero (red) or not (blue), and the intensity of the blue (from light to dark) represents significance.

Level of aggregation	*Temperature*	*Humidity*	*Population density*	*Socioeconomic (PC1)*	*Spatial unstructured parameter*	*Spatial structured parameter*	*Lagged malaria relative risk*

Zone	[0.091, -0.147]	**[0.634, 0.911]**	[-0.121, 0.078]	[-0.069, 0.362]	[-0.067, 0.237]	[-0.110, 0.124]	**[0.890, 2.232]**
Unit	**[-0.121, -0.267]**	**[0.445, 0.823]**	**[0.034, 0.226]**	**[0.369, 0.723]**	**[0.134, 0.456]**	**[0.012, 0.16]**	**[0.9674, 2.465]**
Worker unit	**[-0.234, -0.323]**	[0.225, -0.018]	**[0.126, 0.467]**	**[0.624,1.002]**	**[0.323, 0.728]**	**[0.144, 0.328]**	**[0.922, 2.639]**

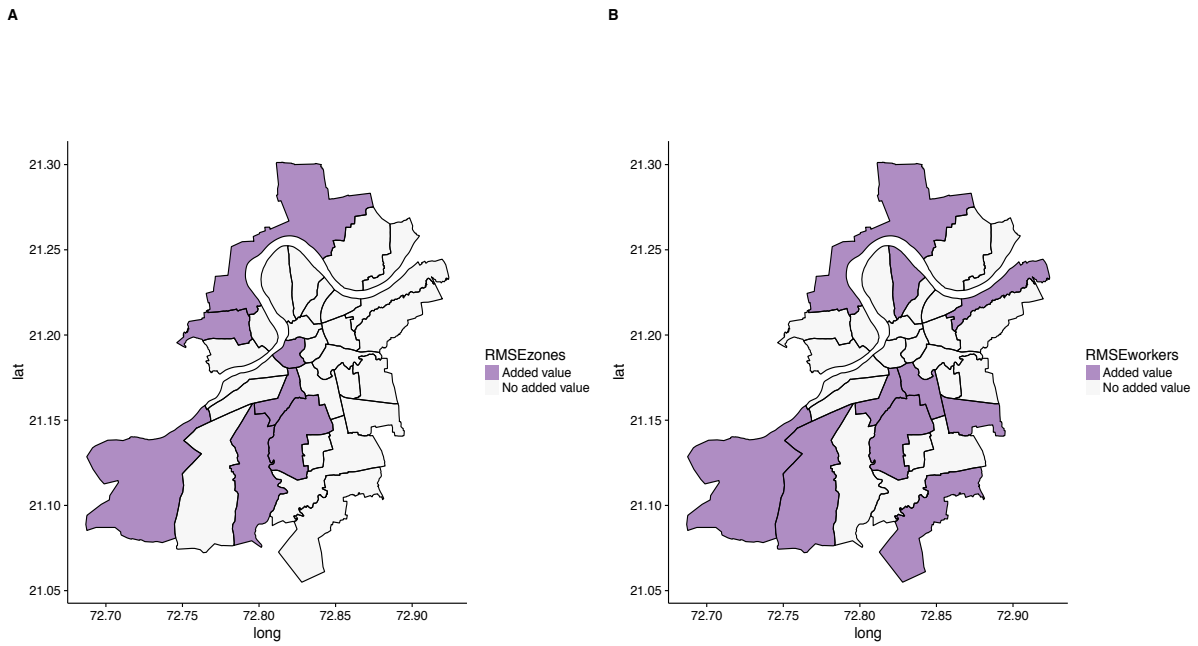

Figure 5. Model predictions scale comparisons. The lower the RMSE (root-mean-square error), the better the model fit. Differences between RMSE for the models fitted at (A) the workers level and (B) zone level vs. the unit level. Districts with positive values of (RMSE null - RMSE alternate) (in purple) suggest that disaggregating/aggregating the data improves the model in these areas. Districts with negative values of this quantity (in white) suggest that the new scale does not add information to the model in that area. We repeated the values for the units within a zone to be able to subtract the alternative RMSE (the same process was performed for the workers areas within each unit).