Big city, small world: Density, contact rates, and transmission of dengue across Pakistan

Kraemer, M.U.G.*1, Perkins, T. A.2,3, Cummings, D.A.T.4, Zakar, R.5, Hay, S.I.3,6,7, Smith, D.L.1,3,8, and Reiner, R.C.*3,9

1. Department of Zoology, University of Oxford, Oxford, OX1 3PS, United Kingdom;
2. Department of Biological Sciences and Eck Institute for Global Health, University of Notre Dame, Notre Dame, IN 46556, United States;
3. Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, United States;
4. Department of Epidemiology, Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD 21205, United States;
5. Department of Public Health, University of Punjab, Lahore, 54590, Pakistan;
6. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, United Kingdom;
7. Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA 98121, United States;
8. Sanaria Institute for Global Health and Tropical Medicine, Rockville, MD 20850, United States;
9. Department of Epidemiology and Biostatistics, Indiana University School of Public Health, Bloomington, IN 47405, United States;

* Author for correspondence: moritz.kraemer@zoo.ox.ac.uk or rcreiner@indiana.edu

Abstract:

Macroscopic descriptions of populations commonly assume that encounters between individuals are well mixed; i.e., each individual has an equal chance of coming into contact with any other individual. Relaxing this assumption can be challenging though, due to the difficulty of acquiring detailed knowledge about the non-random nature of encounters. Here, we fitted a mathematical model of dengue virus transmission to spatial time series data from Pakistan and compared maximum-likelihood estimates of “mixing parameters” when disaggregating data across an urban-rural gradient. We show that dynamics across this gradient are subject not only to differing transmission intensities but also to differing strengths of nonlinearity due to differences in mixing. We furthermore show that neglecting spatial variation in mixing can lead to substantial underestimates of the level of effort needed to control a pathogen with vaccines or other control efforts. We complement this analysis with relevant contemporary environmental drivers of dengue.

Keywords: dengue, epidemiology, heterogeneity, mixing, nonlinearity, spatial dynamics, vector-borne diseases, victim-exploiter dynamics
Introduction:

The transmission dynamics of all infectious diseases depend on a few basic but key determinants: the availability of susceptible and infectious hosts, contacts between them, and the potential for transmission upon contact. Susceptibility is shaped primarily by historical patterns of transmission, the natural history of the pathogen, the host’s immune response, and host demography (Grenfell et al. 2004). What constitutes an epidemiologically significant contact depends on the pathogen’s mode of transmission (Stoddard et al. 2009), and structure in contact patterns can be influenced by transportation networks and the spatial scale of transmission (Brockmann & Helbing 2013), by host heterogeneities such as age (Kilpatrick et al. 2006), and dynamically in response to the pathogen’s influence on host behavior (Fenichel et al. 2011). Whether transmission actually occurs during contact between susceptible and infectious hosts often depends heavily on environmental conditions (Shaman et al. 2010; Gilbert et al. 2014; Weiss et al. 2014b). Disentangling the relative roles of these factors in driving patterns of disease incidence and prevalence is a difficult but central pursuit in infectious disease epidemiology, and mathematical models that are specific about the biology of how these mechanisms interact represent an indispensable tool in this pursuit (Anderson & May 1991).

The time-series susceptible-infected-recovered (TSIR) model was developed by Finkenstädt & Grenfell (2000) to offer a more accurate and straightforward way to statistically connect mechanistic models of infectious disease transmission with time series data. Among other features, TSIR models readily account for inhomogeneous mixing in a phenomenological way by allowing for nonlinear dependence of rates of contact between susceptible and infectious hosts on their densities. Although this is a simple feature that can be incorporated into any model based on mass-action assumptions – indeed, earlier applications pertained to inhomogeneous mixing in predator-prey systems (Pascual & Levin 1999) – the “mixing parameters” that determine the extent of this nonlinearity have primarily been fit to empirical data in applications of the TSIR model to transmission of measles, cholera, rubella, and dengue (Bjørnstad et al. 2002; Koelle & Pascual 2004; Metcalf et al. 2011; Reich et al. 2013).

Applied to discrete-time models such as the TSIR, mixing parameters also have an interpretation as corrections for approximating a truly continuous-time process with a discrete-time model (Liu et al. 1987; Glass et al. 2003). In no application of the TSIR model to date has the potential for variation in these parameters been assessed, leaving the extent to which
inhomogeneity of mixing is different in space and time as an open question in the study of infectious disease dynamics.

There are a number of reasons why mixing might vary in time or space. Seasonal variation in mixing might arise because of travel associated with labor (Bharti et al. 2011), religious events (Lessler et al. 2014), or vacation (Deville et al. 2014). A major driver of flu is associated with the timing of school openings (Salathé et al. 2010), social networks (Cauchemez et al. 2011), and spatial variation in mixing could arise because of cultural differences and the geographic scales of interest (Brockmann & Helbing 2013; Vazquez-Prokopec et al. 2013; Read et al. 2014), because of variation in the density and quality of roads (Uchida & Nelson 2008), or because of variation in human densities and myriad factors associated with that (Bjørnstad et al. 2002; Simini et al. 2012). For vector-borne diseases, variation in mixing is amplified even further by variation in vector densities (Perkins et al. 2013), which effectively mediate contact between susceptible and infectious hosts.

Dengue is a mosquito-borne viral disease with a strong potential for spatial variation in mixing (Brady et al. 2012; Bhatt et al. 2013). The dominant vectors of dengue viruses (Aedes spp.) thrive in areas where they are able to associate with humans, as humans provide not only a preferred source of blood but also water containers that the mosquitoes use for egg laying and for larval and pupal development (Morrison et al. 2008). Two additional aspects of Aedes ecology – limited dispersal distance of the mosquito (Harrington et al. 2005) and daytime biting (Akram et al. 2009) – imply that human movement should be the primary means by which the viruses spread spatially (Stoddard et al. 2009). Indeed, analyses of dengue transmission dynamics at a variety of scales have strongly supported this hypothesis (Allicock et al. 2012; Stoddard et al. 2013; Bhoomiboonchoo et al. 2014; Reiner et al. 2014). To the extent that human movement in dense urban environments is more well-mixed than elsewhere, there are likely to be differences in the extent of inhomogeneous mixing in peri-urban and rural areas. This is also presumably the case for directly transmitted pathogens, but with a potentially even stronger discrepancy for dengue due to the urban-rural gradient in mosquito densities.

To assess the potential for spatial variation in the inhomogeneity of mixing as it pertains dengue transmission, we performed an analysis of district-level time series of dengue transmission in the Punjab province of Pakistan using a TSIR model with separate mixing parameters for urban and rural districts. We likewise made estimates of the relationships
between density-independent transmission potential and putative drivers thereof, such as
temperature, to allow for the relative roles of extrinsic and intrinsic factors to be teased apart.

Finally, we performed mathematical analyses of the fitted model to assess the significance of
spatial variation in mixing inhomogeneity for how time series data are interpreted and used to
guide control efforts.

Material and Methods:

We obtained daily dengue case data aggregated at hospital level from Punjab province
provided by the Health Department Punjab, Pakistan between 2011 and 2014. In total, 41,300
suspected and confirmed dengue cases were reported in 109 hospitals. All hospitals were
subsequently geo-located using ‘Google maps’ (http://www.maps.google.com). Hospitals that
could not be identified were removed from the database. 21,182 cases alone were reported
from the year 2011, which affected the almost the entire province. Many more cases occurred
in Lahore (35,348) compared to all other districts (5,952) (Table 1). A breakdown per year and
each province is provided in Table S1 and additional information about collection can be found
in the supplementary appendix.

Covariate selection and processing:

Environmental conditions are instrumental in defining the risk of transmission of dengue (Bhatt
et al. 2013). Transmission is limited by the availability of a competent disease vector. Due to a
lack of resources and political instability no comprehensive nation wide entomological surveys
have been performed in Pakistan. Therefore we use a probabilistic model to infer the
probability of occurrence of *Ae. aegypti* and *Ae. albopictus* in Pakistan derived from a globally
comprehensive dataset containing more than 20,000 records for each species (Figure 1a and
b). Such model outputs have proven useful in identifying areas of risk of transmission of
dengue as well as malaria (Gething et al. 2011; Sinka et al. 2012; Bhatt et al. 2013). Other
important environmental conditions defining the risk of transmission of dengue are
temperature, water availability, and vegetation cover (Messina et al. 2015). To account for such
variation, raster layers of daytime land surface temperature (LST) were processed from the
MOD11A2 satellite, gap-filled to remove missing values, and then averaged to a monthly
temporal resolution for all four years (Weiss et al. 2015). The density of vegetation coverage
has been shown to be associated with vector abundance (Eisen & Lozano-Fuentes 2009).
Moreover, vegetation indices are useful proxies for precipitation and may be used to derive the
presence of standing water buckets that are habitats for the *Aedes* mosquitoes (Maciel-de-
Freitas *et al.* 2007). The same method was again applied to derive the Enhanced Vegetation
Index (EVI) from the MOD11A2 satellite to produce 16 day and monthly pixel based estimates
for 2011-2014 (Figure 1g) (Weiss *et al.* 2014a). Due to the inherent delay between rainfall and
daily temperature influencing mosquito population dynamics and those mosquitoes contributing
to an increase in DENV transmission, we consider both, the influence of the current
temperature, vegetation indices and precipitation, data on current transmission as well as the
values of those covariates the time step before (Figure 1f).

We used population count estimates on a 100m resolution that were subsequently aggregated
up to match all other raster layers to a 5 km x 5 km resolution for the year 2015
(http://www.worldpop.org) (Figure 1e). In a follow-up analysis to our primary investigation into
the climatological drivers of dengue transmission we included several density-based
covariates. We derived a weighted accessibility metric that includes both, population density
and urban accessibility, a metric commonly used to derive relative movement patterns (Uchida
& Nelson 2008; Tatem *et al.* 2012). This map shows a friction surface, i.e. the time needed to
travel through a specific pixel (Figure 1d). We also used an urban, peri-urban and rural
classification scheme to quantify patterns of urbanicity based on a globally available grid
(Center for International Earth Science Information Network (CIESIN) 2010) (Figure 1c). All
covariates and case data were aggregated and averaged (where appropriate) to a district level.

**Model:**

Following Finkenstädt & Grenfell (2000), we assume a general transmission model of

$$ I_{t,i} = \beta_{t,i} \cdot \frac{I_{t-1,i}}{N_i} \cdot S_{t-1,i} \cdot \epsilon_{t,i}, \quad (1) $$

where $I_{t,i}$ is the number of infected and infectious individuals and $S_{t,i}$ the number of susceptible
individuals, at time $t$ in district $i$, $N_i$ is the population of district $i$, and $\beta_{t,i}$ is the covariate driven
contact rate. The mixing parameter for the $i$th district is given by $\alpha_i$; when $\alpha_i$ is equal to 1, the
population has homogenous mixing where values less than one can either indicate
inhomogeneous mixing or a need to correct for the discretization of the continuous-time
process. Finally, the error terms $\epsilon_{t,i}$ are assumed to be independent and identically log-
normally distributed random variables. For more details please see SI.

**Model selection:**
The term $\beta_{i,t}$ is fit using generalized additive model regression (GAM) (Hastie & Tibshirani 1990; Dominici et al. 2002; Wood 2011). The time-varying climatological covariates are all fit as a smooth spline, while all other covariates enter $\beta_{i,t}$ linearly. Additionally, unexplained seasonal variation is accounted for using a 12-month periodic smooth spline.

Model selection was performed using backwards selection. Two base models were investigated. First, a climate only model was created using only the climatological and climate suitability covariates. Second, a “full” model using the density-dependent covariates as well as the climatological covariates were combined into a single model which was then subjected to backwards selection. For both models the mixing coefficient was initially set equal for each district and once a final model was arrived upon, the mixing coefficient for Lahore was allowed to vary separately from the other coefficients. All model fitting was conducted using R (R Core Team 2014) and the “mgcv” package (Wood 2011). Models are fit by maximizing the restricted maximum likelihood (Patterson & Thompson 1971) to reduce bias and over-fitting of the smooth splines. The model source code and processing of covariates will be made available in line with previous projects (Pigott & Kraemer 2014).

Model analysis:

To explore the potential significance of spatial variation in mixing parameters, we conducted an analysis to probe the inherent mathematical tradeoff between the mixing parameter $\alpha$ and the density-independent transmission coefficient $\beta$. Specifically, to answer the question, what difference in local transmission would be necessary to account for a difference in mixing while achieving identical transmission dynamics. To explore this, we used eqn. (1) to establish:

\[ \beta_1 \frac{i_1r^\alpha_1}{N_1} = \beta_2 \frac{i_2r^\alpha_2}{N_2} \] (2),

from which we obtained

\[ \frac{\beta_1}{\beta_2} = \frac{i}{i_2} \frac{r^\alpha_2}{r^\alpha_1} \] (3).

We then examined how variation in $i$ and $\alpha_2 - \alpha_1$ affected the left hand side of eqn. (3) and likewise the critical proportion of the population to control in order to effect pathogen elimination, which, under our model, is $p_c = 1 - \frac{1}{\beta}$.

Results:

Description of case distribution:
The majority of cases are clustered in Lahore, the capital of Punjab province. Ongoing transmission seems to be focal in three (Vehari, Rawlpindi and Lahore) districts and spread through infective sparks to smaller more rural provinces. In early investigations, all hospitals that reported dengue cases are located in areas with high EVI values indicating a clear environmental signal (Figure 1b).

Model selection:

To disentangle the different aspects of dengue dynamics and their drivers we used a model containing only the climatological covariates and performing backwards model selection until each covariate in the model was significant at a 5% level resulted in a model that explained 76.9% of the deviance and whose adjusted R-squared was 0.746. Amongst the yearly-averaged covariates, EVI and precipitation remained in the model as well as the derived Ae. albopictus range map (p-values of $8.21 \times 10^{-4}$, 0.01, and $3.9 \times 10^{-5}$ respectively). Interestingly, if the derived Ae. aegypti map is substituted for the Ae. albopictus map, the deviance explained increases slightly to 76.8%. For climatological covariates that were fit as smooth splines, temperature, lagged temperature and EVI remained in the model (Figure 2, p-values of 0.010, 0.030 and 0.030 with effective degrees of freedom 7.55, 5.47, and 1.83, respectively). There was a significant amount of periodic variation unexplained by the climatological covariates alone and the ‘seasonality’ covariate remained (Figure 2, p-value = 0.0034). The estimated median values for $R_0$ per district are clustered around two (mean = 2.1), their geographical distribution indicates a clear trend towards districts with higher population (Figure 3). Finally, the mixing coefficient was significantly lower than 1 ($\alpha = 0.69$, 95% CI = (0.614, 0.771), p-value = $1.6 \times 10^{-14}$).

To understand these differences the final model was then compared to a nested model where the coefficient for Lahore was allowed to vary independently of all other districts. Deviance explained increased to 77% and adjusted R-squared increased to 0.753. Further, the mixing coefficient for Lahore ($\alpha = 0.74$) was significantly larger than the mixing coefficient for the other districts ($\alpha = 0.59$, p-value=0.0068) (Figure S1). The median $R_0$ for Lahore was estimated at 3.28, the highest for all districts.

In assess how far the variation of mixing coefficients can be explained by other covariates we consider the possibility that movement accounts for the differences in the mixing coefficient between Lahore and the other districts. The density-dependent covariates (described earlier)
were then added to the full model and backwards selection was repeated. The resulting model explained 78.6% of the variance, had an adjusted R-squared of 0.763 and is superior to the final climatological model based on AIC (699.23 versus 714.83). Yearly averaged EVI, NDVI and precipitation were all significant (p-values of $8.7 \times 10^{-5}$, 0.00024, and 0.00028, respectively). Again, the derived *Ae. albopictus* map was significant (p-value = 0.008163). For climatological covariates fit as smooth splines, only temperature and lagged temperature were found significant (Figure 2b, p-value of $4.0 \times 10^{-7}$ and 0.0013, and effective degrees of freedom 7.61 and 4.81, respectively), and there was still a significant ‘seasonality’ (Figure 2b, p-value of $4.0 \times 10^{-7}$, effective degrees of freedom 4.48). The mixing coefficient was fit at $\alpha = 0.58$, barely lower than the mixing coefficient for non-Lahore districts in the climatological model. The estimated median $R_0$ again clustered around 2 (mean = 1.8), and again the $R_0$ for Lahore was largest, but in this model it was considerably larger than in the climatological model (Lahore $R_0 = 7.82$, Figure 2b). Full details of the model parameters are shown in Table S2-S5.

Two of the density-dependent covariates remained in the model: the urban map (p-value = 0.01) and the weighted access map (p-value= $3 \times 10^{-5}$). When the nested model that allows Lahore’s mixing coefficient to vary was fit, there was no significant difference in the two mixing coefficients (p-value≈1).

Model analysis:

Given a difference in estimates of the mixing parameters between Lahore and elsewhere of 0.15, we analyzed eqn. (3) to assess the bias in estimates of the transmission coefficient that would result from ignoring this extent of variation in the inhomogeneity of mixing displayed between two areas. For the purpose of *ceteris paribus* comparisons, we assumed equal force of infection but varied it across several orders of magnitude. Depending on the order of magnitude, estimates of transmission coefficients made if overestimating the mixing parameter by 0.15 could easily result in a two- to three-fold underestimate in the transmission coefficient (Figure 4). For realistic ranges of the transmission coefficient for dengue, and more generally the basic reproductive number $R_0$, this extent of underestimation of $R_0$ could lead to underestimating the critical proportion of the population to which vaccines or other control measures must be applied by 20-30% (Figure 5).

Discussion:
Our results point to considerable spatial heterogeneity in the extent of mixing and strength of an associated nonlinearity in transmission along an urban-rural gradient. This regional variability in mixing has direct implications for estimates of the basic reproductive number of dengue in our study region and elsewhere. Although the potential for such bias in estimates of the basic reproductive number has been shown in a theoretical context (Hu et al. 2013; Perkins et al. 2013), we provide quantitative estimates of the extent of this problem in interfacing models with a rich spatio-temporal data set. Such analyses have implications for estimates of population-level parameters not only for dengue but also for other infectious diseases (Bartlett 1960; Keeling & Grenfell 1997; Bjørnstad et al. 2002; Smith et al. 2002; Keeling & Eames 2005; Kiss et al. 2008) and possibly even more broadly in ecology.

We revealed significant differences in mixing components between urban and rural settings and found that a population-weighted urban accessibility metric was able to absorb differences in mixing between these settings indicating that this specific covariate accounts for aspects influencing mixing. Mixing is presumably influenced directly by human behavior and has been shown to be highly unpredictable, largely dependent on the local context and the spatial and temporal scale (Yang et al. 2014). In this study however we could show that the density-dependent covariate selected was able to capture the influence of these key encounters on a district level. Once differences in mixing were accounted for, estimated $R_0$ values indicated considerably larger differences between transmission potential in Lahore versus all other districts. Synchronizing more accurate geo-referenced data would allow to assess the importance of spatial scale on the relationship between “mixing parameters” and urban accessibility (Perkins et al. 2013; Mills & Riley 2014). In the case of dengue this again has been limited by the availability of high resolution data (Ruberto et al. 2015). Complementing this analysis with measurement of direct social contact patterns could be important to explore this relationship in even more detail (Bauch & Galvani 2013; Vazquez-Prokopec et al. 2013; Heesterbeek et al. 2015) and could be informed by mathematical models that explored this relationship previously for other diseases (Bjørnstad et al. 2002; Reiner et al. 2012). Another encouraging finding is that large-scale mosquito suitability surfaces help capture the environmental determinants of dengue transmission (Bhatt et al. 2013).

Intervention strategies are contingent on both understanding key environmental drivers of transmission and the dynamics of ongoing human-to-human transmission, particularly in outbreaks situations (Perkins et al. 2015). Environmental drivers such as seasonal fluctuations...
in rainfall, temperature, vegetation coverage or mosquito abundance will help guide surveillance and control efforts targeted mostly towards the ecological aspects of mosquito dispersal (Johansson 2015). Once infection occurs much debate has been focused around optimizing intervention strategies to reduce disease incidence, which is largely determined by $R_0$. The presented framework shows that the interaction between mixing parameters and force of infection has potentially large implications for optimizing targeted intervention, particularly in countries where resources are scarce (Cesare et al. 2015). This in fact is even more important in areas of low infection where transmission seems to be more focal (Salje et al. 2012). Again however importance needs to be focused on the spatial and temporal resolution of appropriate intervention strategies and the respective effect of the selected covariates and model parameters (Mills & Riley 2014). Empirical understanding, however, on which spatial resolution is most appropriate to carry out large-scale vector-borne disease interventions remains unknown.

Once transmission has occurred in one place, understanding not only spatial heterogeneities of transmission dynamics but their subsequent spread in mechanistic stochastic models as shown for measles would help to empirically determine the propagation of the disease (Grenfell et al. 2001). The spatial spread dynamics have been given considerable interest globally as the risk of importation of dengue into yet endemic areas continentally and internationally is increasing with travel and trade (Schaffner & Mathis 2014). Exploration of the case data in Pakistan suggests that spread happens along major transport routes from Lahore to Karachi and north to Rawalpindi. Using results presented here on mixing components and environmental drivers will help pinpoint areas of major risk of importation more accurately especially in the case of recurring epidemics. Using the fitted relationships of the environmental drivers of transmission and $R_0$ will enable future analyses and comparisons between diseases and geographic regions. In this context it will be instrumental to integrate a variety of movement and social network models with the evidence presented here to infer more accurately how the geographical spread of dengue is determined.

To allow for comparison of these results in a broader context and across diseases, possibly even in outbreak situations and in real time, it is essential to make data widely available by open access (Heesterbeek et al. 2015). Moreover, the complexity of infectious disease dynamics is not fully understood and we are limited by computational capacities to fully account for stochasticity and nonlinearity.
Author contributions:
Conceived and designed the experiment: MUGK, RCR, TAP, DLS. Performed the experiments: RCR, TAP, MUGK. Analysed the data: RCR, MUGK. Edited the manuscript: DLS, SIH, DATC, ZR. Wrote the paper: MUGK, RCR, TAP.

Acknowledgements and funding:
The authors would like to thank the WHO Punjab office, Health Department Punjab, and PID for providing the epidemiological data, and all participants that helped collect the data. MUGK acknowledges funding from the German Academic Exchange Service (DAAD). TAP, SIH, RCR acknowledge the support from the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health. SIH is funded by a Senior Research Fellowship from the Wellcome Trust (#095066) and a grant from the Bill & Melinda Gates Foundation (#OPP1093011). Funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References:
1. Akram, W., Hafeez, F., Ullah, U.N., Kim, Y.K., Hussain, A. & Lee, J.J. (2009). Seasonal distribution and species composition of daytime biting mosquitoes. Entomol. Res., 39, 107–13.
2. Allicock, O.M., Lemey, P., Tatem, A.J., Pybus, O.G., Bennett, S.N., Mueller, B. a., et al. (2012). Phylogeography and population dynamics of dengue viruses in the Americas. Mol. Biol. Evol., 29, 1533–43.
3. Anderson, R.M. & May, R.M. (1991). Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford, U.K.
4. Bartlett, M.S. (1960). The critical community size for measles in the U.S. J. R. Stat. Soc. A, 123, 37–44.
5. Bauch, C.T. & Galvani, A.P. (2013). Epidemiology. Social factors in epidemiology. Science, 342, 47–9.
6. Bharti, N., Tatem, a J., Ferrari, M.J., Grais, R.F., Djibo, A. & Grenfell, B.T. (2011). Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. Science, 334, 1424–7.
7. Bhatt, S., Gething, P.W., Brady, O.J., Messina, J.P., Farlow, A.W., Moyes, C.L., et al. (2013). The global distribution and burden of dengue. Nature, 496, 504–7.
8. Bhoomboonchoo, P., Gibbons, R. V., Huang, A., Yoon, I.-K., Buddhari, D., Nisalak, A., et al. (2014). The spatial dynamics of Dengue virus in Kamphaeng Phet, Thailand. PLoS Negl. Trop. Dis., 8, e3138.
9. Bjørnstad, O.N., Finkenstädt, B.F. & Grenfell, B.T. (2002). Dynamics of measles epidemics: Estimating scaling of transmission rates using a Time Series SIR model. Ecol. Monogr., 72, 169–84.
10. Brady, O.J., Gething, P.W., Bhatt, S., Messina, J.P., Brownstein, J.S., Hoen, A.G., et al. (2012). Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Negl. Trop. Dis.*, 6, e1760.

11. Brockmann, D. & Helbing, D. (2013). The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342, 1337–42.

12. Cauchemez, S., Bhattarai, A., Marchbanks, T.L., Fagan, R.P., Ostroff, S., Ferguson, N.M., et al. (2011). Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proc. Natl. Acad. Sci.*, 108, 2825–30.

13. Center for International Earth Science Information Network (CIESIN). (2010). *Gridded Population of the World, version 3 (GPWv3)* and the Global Rural-Urban Mapping Project (GRUMP).

14. Cesare, M. Di, Bhatti, Z., Soofi, S.B., Fortunato, L., Ezzati, M. & Bhutta, Z. a. (2015). Geographical and socioeconomic inequalities in women and children’s nutritional status in Pakistan in 2011: an analysis of data from a nationally representative survey. *Lancet Glob. Heal.*, 3, e229–39.

15. Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F.R., Gaughan, a. E., et al. (2014). Dynamic population mapping using mobile phone data. *Proc. Natl. Acad. Sci.*, 111, 15888–93.

16. Dominici, F., McDermott, A., Zeger, S.L. & Samet, J.M. (2002). On the use of generalized additive models in time-series studies of air pollution and health. *Am. J. Epidemiol.*, 156, 193–203.

17. Eisen, L. & Lozano-Fuentes, S. (2009). Use of mapping and spatial and space-time modeling approaches in operational control of Aedes aegypti and dengue. *PLoS Negl. Trop. Dis.*, 3, e411.

18. Fenichel, E.P., Castillo-Chavez, C., Ceddia, M.G., Chowell, G., Parra, P. a G., Hickling, G.J., et al. (2011). Adaptive human behavior in epidemiological models. *Proc. Natl. Acad. Sci.*, 108, 6306–11.

19. Finkenstädt, B.F. & Grenfell, B.T. (2000). Time series modelling of childhood diseases: a dynamical systems approach. *Appl. Stat.*, 49, 187–205.

20. Getling, P.W., Patil, A.P., Smith, D.L., Guerra, C. a, Elyazar, I.R.F., Johnston, G.L., et al. (2011). A new world malaria map: Plasmodium falciparum endemicity in 2010. *Malar. J.*, 10, 378.

21. Gilbert, M., Golding, N., Zhou, H., Wint, G.R.W., Robinson, T.P., Tatem, A.J., et al. (2014). Predicting the risk of avian influenza A H7N9 infection in live-poultry markets across Asia. *Nat. Commun.*, 5, 4116.

22. Glass, K., Xia, Y. & Grenfell, B.T. (2003). Interpreting time-series analyses for continuous-time biological models - Measles as a case study. *J. Theor. Biol.*, 223, 19–25.
Grenfell, B.T., Bjørnstad, O.N. & Kappey, J. (2001). Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414, 716–23.

Grenfell, B.T., Pybus, O.G., Gog, J.R., Wood, J.L.N., Daly, J.M., Mumford, J. a, *et al.* (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, 303, 327–32.

Harrington, L.C., Scott, T.W., Lerdthusnee, K., Coleman, R.C., Costero, A., Clark, G.G., *et al.* (2005). Dispersal of the dengue vector *Aedes aegypti* within and between rural communities. *Am. J. Trop. Med. Hyg.*, 72, 209–20.

Hastie, T.J. & Tibshirani, R.J. (1990). *Generalized additive models*. Vol. 43. CRC Press.

Heesterbeek, H., Anderson, R.M., Andreasen, V., Bansal, S., De Angelis, D., Dye, C., *et al.* (2015). Modeling infectious disease dynamics in the complex landscape of global health. *Science*, 347, aaa4339.

Hu, H., Nigmatulina, K. & Eckhoff, P. (2013). The scaling of contact rates with population density for the infectious disease models. *Math. Biosci.*, 244, 125–34.

Johansson, M.A. (2015). Chikungunya on the move. *Trends Parasitol.*, 31, 43–45.

Keeling, M.J. & Eames, K.T.D. (2005). Networks and epidemic models. *J. R. Soc. Interface*, 2, 295–307.

Keeling, M.J. & Grenfell, B.T. (1997). Disease extinction and community size: modeling the persistence of measles. *Science*, 275, 65–7.

Kilpatrick, A.M., Daszak, P., Jones, M.J., Marra, P.P. & Kramer, L.D. (2006). Host heterogeneity dominates West Nile virus transmission. *Proc. Biol. Sci.*, 273, 2327–33.

Kiss, I.Z., Green, D.M. & Kao, R.R. (2008). The effect of network mixing patterns on epidemic dynamics and the efficacy of disease contact tracing. *J. R. Soc. Interface*, 5, 791–9.

Koelle, K. & Pascual, M. (2004). Disentangling extrinsic from intrinsic factors in disease dynamics: A nonlinear time series approach with an application to Cholera. *Am. Nat.*, 163, 901–13.

Lessler, J., Rodriguez, I., Derek, B., Tini, A.T.C., Kerkhove, M., Van, Mills, H., *et al.* (2014). Estimating potential incidence of MERS CoV associated with Hajj pilgrims to Saudi Arabia, 2014. *PLOS Curr. Outbreaks*, November 2.

Liu, W.M., Hethcote, H.W. & Levin, S. a. (1987). Dynamical behavior of epidemiological models with nonlinear incidence rates. *J. Math. Biol.*, 25, 359–380.

Maciel-de-Freitas, R., Marques, W. a., Peres, R.C., Cunha, S.P. & De Oliveira, R.L. (2007). Variation in *Aedes aegypti* (Diptera: Culicidae) container productivity in a slum and a suburban district of Rio de Janeiro during dry and wet seasons. *Mem. Inst. Oswaldo Cruz*, 102, 489–96.

Messina, J.P., Brady, O.J., Pigott, D.M., Golding, N., Kraemer, M.U.G., Scott, T.W., *et al.* (2015). The many projected futures of dengue. *Nat. Rev. Microbiol.*, 13, 230–39.
Metcalf, C.J.E., Bjornstad, O.N., Ferrari, M.J., Klepac, P., Bharti, N., Lopez-Gatell, H., et al. (2011). The epidemiology of rubella in Mexico: seasonality, stochasticity and regional variation. Epidemiol. Infect., 139, 1029–38.

Mills, H.L. & Riley, S. (2014). The spatial resolution of epidemic peaks. PLoS Comput. Biol., 10, e1003561.

Morrison, A.C., Zielinski-Gutierrez, E., Scott, T.W. & Rosenberg, R. (2008). Defining challenges and proposing solutions for control of the virus vector Aedes aegypti. PLoS Med., 5, e68.

Pascual, M. & Levin, S. a. (1999). From individuals to population densities: searching for the scale of intermediate determinism. Ecology, 80, 2225–36.

Patterson, H.D. & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika, 58, 545–54.

Perkins, T.A., Metcalf, C.J.E., Grenfell, B.T. & Tatem, A.J. (2015). Estimating Drivers of Autochthonous Transmission of Chikungunya Virus in its Invasion of the Americas. PLoS Curr., 7, ecurrents.outbreaks.a4c7b6ac10e0420b1788c9767946d1.

Perkins, T.A., Scott, T.W., Le Menach, A. & Smith, D.L. (2013). Heterogeneity, mixing, and the spatial scales of mosquito-borne pathogen transmission. PLoS Comput. Biol., 9, e1003327.

Pigott, D.M. & Kraemer, M.U. (2014). Enhancing infectious disease mapping with open access resources. Eurosurveillance, 19, pii=20989.

R Core Team. (2014). R: A language and environment for computing. Vienna, Austria. R Found. Stat. Comput.

Read, J.M., Lessler, J., Riley, S., Wang, S., Tan, L.J., Kwok, K.O., et al. (2014). China social mixing patterns in rural and urban areas of southern China. Proc. R. Soc. B, 281, 20140268.

Reich, N.G., Shrestha, S., King, A. a, Rohani, P., Lessler, J., Kalayanarooj, S., et al. (2013). Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity. J. R. Soc. Interface, 10, 20130414.

Reiner, R.C., King, A. a, Emch, M., Yunus, M., Faruque, a S.G. & Pascual, M. (2012). Highly localized sensitivity to climate forcing drives endemic cholera in a megacity. Proc. Natl. Acad. Sci., 109, 2033–6.

Reiner, R.C., Stoddard, S.T. & Scott, T.W. (2014). Socially structured human movement shapes dengue transmission despite the diffusive effect of mosquito dispersal. Epidemics, 6, 30–6.

Ruberto, I., Marques, E., Burke, D.S. & Van Panhuis, W.G. (2015). The availability and consistency of dengue surveillance data provided online by the World Health Organization. PLoS Negl. Trop. Dis., 9, e0003511.
Salathé, M., Kazandjieva, M., Woo, J., Levis, P., Feldman, M.W. & Jones, J.H. (2010). A high-resolution human contact network for infectious disease transmission. *Proc. Natl. Acad. Sci.*, 107, 22020–5.

Saïje, H., Lessler, J., Endy, T.P., Curriero, F.C., Gibbons, R. V., Nisalak, A., et al. (2012). Revealing the microscale spatial signature of dengue transmission and immunity in an urban population. *Proc. Natl. Acad. Sci.*, 109, 9535–8.

Schaffner, F. & Mathis, A. (2014). Dengue and dengue vectors in the WHO European region: past, present, and scenarios for the future. *Lancet. Infect. Dis.*, 14, 1271–80.

Shaman, J., Pitzer, V.E., Viboud, C., Grenfell, B.T. & Lipsitch, M. (2010). Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol.*, 8.

Simini, F., González, M.C., Maritan, A. & Barabási, A.-L. (2012). A universal model for mobility and migration patterns. *Nature*, 484, 96–100.

Sinka, M.E., Bangs, M.J., Manguin, S., Rubio-Palis, Y., Charoenviriyaphap, T., Coetzee, M., et al. (2012). A global map of dominant malaria vectors. *Parasit. Vectors*, 5, 69.

Smith, D.L., Lucey, B., Waller, L. a, Childs, J.E. & Real, L. a. (2002). Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. *Proc. Natl. Acad. Sci.*, 99, 3668–72.

Stoddard, S.T., Forshey, B.M., Morrison, A.C., Paz-Soldan, V. a, Vazquez-Prokopec, G.M., Astete, H., et al. (2013). House-to-house human movement drives dengue virus transmission. *Proc. Natl. Acad. Sci.*, 110, 994–9.

Tatem, A.J., Hemelaar, J., Gray, R.R. & Salemi, M. (2012). Spatial accessibility and the spread of HIV-1 subtypes and recombinants. *AIDS*, 26, 2351–60.

Uchida, H. & Nelson, A. (2008). *Agglomeration index: towards a new measure of urban concentration*. Washington D.C.

Vazquez-Prokopec, G.M., Bisanzio, D., Stoddard, S.T., Paz-Soldan, V., Morrison, A.C., Elder, J.P., et al. (2013). Using GPS technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment. *PLoS One*, 8, e58802.

Weiss, D.J., Atkinson, P.M., Bhatt, S., Mappin, B., Hay, S.I. & Gething, P.W. (2014a). An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS J. Photogramm. Remote Sens.*, 98, 106–18.

Weiss, D.J., Bhatt, S., Mappin, B., Van Boeckel, T.P., Smith, D.L., Hay, S.I., et al. (2014b). Air temperature suitability for Plasmodium falciparum malaria transmission in Africa 2000-2012: a high-resolution spatiotemporal prediction. *Malar. J.*, 13, 171.
Weiss, D.J., Mappin, B., Dalrymple, U., Bhatt, S., Cameron, E., Hay, S.I., et al. (2015). Re-examining environmental correlates of Plasmodium falciparum malaria endemicity: a data-intensive variable selection approach. *Malar. J.*, 14, 1–18.

Wood, S.N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73, 3–36.

Yang, Y., Herrera, C., Eagle, N. & González, M.C. (2014). Limits of predictability in commuting flows in the absence of data for calibration. *Sci. Rep.*, 4, 5662.
Table 1: Reported cases by year in Lahore and all other districts.

|        | 2011  | 2012  | 2013  | 2014  | Total |
|--------|-------|-------|-------|-------|-------|
| Lahore | 18,020| 4,013 | 11,516| 1,799 | 35,348|
| Other  | 3,162 | 500   | 2,356 | 2,790 | 5,952 |

Figure 1: Covariates used in this study to derive environmental drivers of transmission.

**Aedes aegypti** probability of occurrence (a); **Ae. albopictus** probability of occurrence (b); Urbanicity (c); Weighted urban accessibility (d); Population density and study area (e); Precipitation (f); Enhanced Vegetation Index (EVI) mean (g).
Figure 2: Model outputs using a backwards model selection procedure in the model using climatological variables (a, A-D), and including the density dependent variables (b, A-D).
Figure 3: Average distribution of $R_0$ with green representing Lahore versus red all other districts (A), their geographical distribution (B), and over time in which the green line again is representing Lahore versus the red line representing all other districts (C).
Figure 4: Ratio of betas ($R_0$) assuming equal force of infection and a difference in $\alpha_2 - \alpha_1$ of 0.2, 0.15 (green), and 0.1, form top to bottom. The straight line indicates a ratio of 1.
Figure 5: Critical proportion of the population to control in population 2 as a function of $R_0$ in population 1, the order of magnitude of the infectious numbers in each population, and a difference in $\alpha_2 - \alpha_1$ of 0.1, 0.15 (green), and 0.2. The straight line indicates the critical proportion assuming the $\alpha$ in each population are equal.

Order of magnitude of infectious population size

$R_0 = 1.5$  $R_0 = 3$  $R_0 = 6$
Table S1: Reported case numbers per year for each district of the study region, Punjab Province, Pakistan.

| District          | 2011 | 2012 | 2013 | 2014 | Total |
|-------------------|------|------|------|------|-------|
| Attock            | 7    | 1    | 0    | 18   | 26    |
| Bagh              | 0    | 0    | 0    | 0    | 0     |
| Bahawalnagar      | 65   | 0    | 0    | 0    | 65    |
| Bahawalpur        | 65   | 1    | 40   | 3    | 109   |
| Bhakkar           | 33   | 0    | 0    | 0    | 33    |
| Bhimber           | 0    | 0    | 0    | 0    | 0     |
| Chakwal           | 0    | 0    | 0    | 0    | 0     |
| D. G. Khan        | 3    | 0    | 0    | 0    | 3     |
| Faisalabad        | 981  | 58   | 125  | 0    | 1164  |
| Gujranwala        | 61   | 0    | 0    | 0    | 61    |
| Gujrat            | 0    | 24   | 0    | 0    | 24    |
| Hafizabad         | 1    | 0    | 0    | 0    | 1     |
| Jhang             | 6    | 0    | 0    | 0    | 6     |
| Jhelum            | 75   | 0    | 0    | 33   | 108   |
| Kasur             | 77   | 0    | 0    | 0    | 77    |
| Khanewal          | 2    | 0    | 0    | 1    | 3     |
| Khushab           | 37   | 0    | 0    | 0    | 37    |
| Kotli             | 0    | 0    | 0    | 0    | 0     |
| Lahore            | 18020| 4013 | 11516| 1799 | 35348 |
| Leiah             | 27   | 1    | 0    | 0    | 28    |
| Lodhran           | 0    | 0    | 0    | 0    | 0     |
| Mandi Bahauddi    | 4    | 0    | 3    | 0    | 7     |
| Mianwali          | 6    | 0    | 0    | 0    | 6     |
| Mirpur            | 0    | 0    | 0    | 0    | 0     |
| Multan            | 371  | 1    | 64   | 4    | 440   |
| Muzaffarabad      | 0    | 0    | 0    | 0    | 0     |
| Muzaffargarh      | 0    | 0    | 0    | 0    | 0     |
| Narowal           | 26   | 2    | 0    | 0    | 28    |
| Neelum            | 0    | 0    | 0    | 0    | 0     |
| Okara             | 126  | 3    | 0    | 0    | 129   |
| Pakpattan         | 179  | 5    | 0    | 0    | 184   |
| Poonch            | 0    | 0    | 0    | 0    | 0     |
| Rahim Yar Khan    | 0    | 0    | 0    | 0    | 0     |
| Rajanpur          | 0    | 0    | 0    | 0    | 0     |
| Rawalpindi        | 736  | 132  | 1845 | 2532 | 5245  |
| Sahiwal           | 69   | 4    | 3    | 0    | 76    |
| Sargodha          | 14   | 0    | 0    | 0    | 14    |
| Sheikhupura       | 29   | 3    | 42   | 179  | 253   |
| Sialkot           | 49   | 0    | 0    | 0    | 49    |
| Sudhnoti          | 0    | 0    | 0    | 0    | 0     |
| Toba Tek Singh    | 7    | 1    | 6    | 2    | 16    |
| Vehari            | 106  | 264  | 228  | 18   | 616   |
Table S2: Environmental model: no variation in $\alpha$.

| Term                | Estimate | Std. Error | t-value | p-value |
|---------------------|----------|------------|---------|---------|
| Intercept           | 2.51     | 0.679      | 3.696   | 0.00028 |
| $\alpha$            | 0.690    | 0.040      | 17.288  | < 2e-16 |
| EVI (yearly average)| -8.78    | 2.499      | -3.392  | 0.00082 |
| Ae. albopictus      | 2.28     | 0.544      | 4.196   | 3.93e-5 |
| Precipitation (yearly average) | -0.021 | 0.0079 | -2.590 | 0.0102 |

| Term   | edf | Ref. df | F      | p-value |
|--------|-----|---------|--------|---------|
| “Seasonality” | 3.44 | 8.00    | 1.566  | 0.0034  |
| Temperature | 7.55 | 8.47    | 2.537  | 0.0102  |
| 2-week lagged temperature | 5.47 | 6.67    | 2.300  | 0.0304  |
| EVI    | 1.83 | 2.33    | 3.373  | 0.0299  |
Table S3: Environmental model: $\alpha$ in Lahore differs from other districts.

| Term               | Estimate | Std. Error | t-value | p-value |
|--------------------|----------|------------|---------|---------|
| Intercept          | 2.548    | 0.676      | 3.769   | 0.0002  |
| $\alpha$ (Lahore)  | 0.741    | 0.043      | 17.088  | $<2e$-16|
| $\alpha$ (Not Lahore) | 0.594 | 0.055 | 10.887  | $<2e$-16|
| EVI (yearly average) | -7.636 | 2.527 | -3.022  | 0.003   |
| Ae. albopictus     | 1.159    | 0.679      | 1.708   | 0.089   |
| Precipitation (yearly average) | -0.008 | 0.009 | -0.870  | 0.385   |

| Term        | edf | Ref. df | F     | p-value |
|-------------|-----|---------|-------|---------|
| “Seasonality” | 3.238 | 8.000 | 1.278 | 0.0095  |
| Temperature | 7.502 | 8.435 | 3.195 | 0.0016  |
| 2-week lagged temperature | 5.866 | 7.067 | 2.371 | 0.0231  |
| EVI         | 2.179 | 2.784 | 3.690 | 0.0154  |
Table S4: Full model: no variation in $\alpha$.

| Term                  | Estimate | Std. Error | t-value | p-value  |
|-----------------------|----------|------------|---------|----------|
| Intercept             | -34.83   | 9.803      | -3.553  | 0.00047  |
| $\alpha$              | 0.578    | 0.0451     | 12.824  | <2e-16   |
| EVI (yearly average)  | -60.12   | 15.04      | -3.998  | 8.71e-5  |
| NDVI (yearly average) | 0.0383   | 0.0103     | 3.737   | 0.00024  |
| Ae. albopictus        | 2.745    | 1.028      | 2.669   | 0.00826  |
| Weighted Access       | 2.742e-5 | 6.464e-6   | 4.242   | 3.26e-5  |
| Urbanicity            | -2.468   | 0.9494     | -2.600  | 0.00776  |
| Precipitation (yearly average) | -0.0684 | 0.0185 | -3.697 | 0.00028 |

| Term                  | edf      | Ref. df   | F        | p-value  |
|-----------------------|----------|-----------|----------|----------|
| “Seasonality”         | 4.480    | 8.000     | 4.173    | 3.97e-7  |
| Temperature           | 7.607    | 8.504     | 4.396    | 4.14e-5  |
| 2-week lagged temperature | 4.823  | 5.981     | 3.793    | 0.0013   |
Table S5: Full model: $\alpha$ in Lahore differs from other districts.

| Term               | Estimate | Std. Error | t-value | p-value   |
|--------------------|----------|------------|---------|-----------|
| Intercept          | -35.92   | 9.93       | -3.618  | 0.00037   |
| $\alpha$ (Lahore)  | 0.612    | 0.0643     | 8.510   | <2e-16    |
| $\alpha$ (Not Lahore) | 0.555    | 0.0566     | 9.802   | <2e-16    |
| EVI (yearly average)| -61.59   | 15.19      | -4.055  | 6.95e-5   |
| NDVI (yearly average) | 0.0395   | 0.010      | 3.8     | 0.00019   |
| Ae. albopictus     | 2.81     | 1.033      | 2.721   | 0.0070    |
| Weighted Access    | 2.42e-5  | 7.86e-6    | 3.072   | 0.0024    |
| Urbanicity         | -2.33    | 0.9697     | -2.403  | 0.0171    |
| Precipitation (yearly average) | -0.069   | 0.0186     | -3.733  | 0.00024   |

| Term               | edf  | Ref. df | F      | p-value   |
|--------------------|------|---------|--------|-----------|
| “Seasonality”      | 4.350| 8.000   | 3.817  | 1.31e-6   |
| Temperature        | 7.620| 8.511   | 4.335  | 4.97e-5   |
| 2-week lagged      | 4.764| 5.916   | 3.680  | 0.00178   |
| temperature        |      |         |        |           |
Figure S1: Predicted versus expected values for $\alpha$ in Lahore (green) and all other districts (red) for the environmental model.

Environmental Model:
Mixing in Lahore differs from other districts

Additional information about collection of epidemiological data.
Secondary data from hospital records were used from Punjab province in Pakistan. The data was initially collected by Punjab Health Department as part of the dengue prevention and eradication program. For ensuring accurate reporting from the health facilities, Punjab Health Department used the following three procedures: (i) Clinical case reporting, (ii) Lab case reporting, and (iii) Case management. All health facilities were liable to record the data and share them with Punjab Information Technology Board (PITB) within 24 hours. For clinical case reporting, as per Dengue Expert Advisory Group (DEAG), guidelines, dengue suspects, probable, and confirmed cases needed to be correctly entered on the PITB dashboard within 24 hours of admission. For lab case reporting, all private sector labs must send reports of
positive dengue cases in the line list format to the respective Executive District Officer Health (EDOH) for online entry on the dashboard, again within 24 hours. For case management, the healthcare facilities are liable to manage dengue cases regularly in specified Dengue care units, OPDs, emergency units, or wards.