

Modeling Approaches to Investigating Distributions, Abundances, and Connectivity of Mosquito Vector Species

By
Lindsay P. Campbell

Submitted to the graduate degree program in Ecology and Evolutionary Biology and the
Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for
the degree of Doctor of Philosophy.

Chair: A. Townsend Peterson

Jorge Soberón

Kirsten Jensen

Andrew Short

Victor Agadjanian

Date Defended: 28 November 2016

The dissertation committee for Lindsay P. Campbell certifies that this is the approved version of the following dissertation:

**Modeling Approaches to Investigating Distributions, Abundances,
and Connectivity of Mosquito Vector Species**

Chair: A. Townsend Peterson

Date Approved: 28 November 2016

ABSTRACT

Vector-borne and zoonotic diseases comprise a serious public health concern globally. Over the past 30 years, an increase in newly-emerging vector-borne pathogens, coupled with the broader dispersal of known pathogens, has resulted in substantial challenges for public health intervention and prevention programs. This burden highlights the need for continued improvement of modeling approaches and prediction methods to help identify areas vulnerable to infection, thereby contributing toward more efficient distributions of limited public health resources.

The field of disease ecology emphasizes interactions between disease system components and the natural environment, recognizing that humans are not always the catalyst for pathogen dispersal and distributions. While incorporating environmental factors in assessing potential pathogen risk is a logical first step, complexities in this approach exist because pathogens are nested within the broader community ecology of host, vector, and reservoir species, and often, not all of these elements are known. Although this element poses challenges to understanding limiting factors of specific environmental pathogens, the multitude of components within individual disease systems offer several avenues from which to study patterns, providing insight into risk. Mosquito vectors are one such component. This knowledge, coupled with advances in geospatial technologies, provides excellent opportunities to model environmental factors contributing to potential pathogen distributions and to help predict disease risk in humans.

Here, I present three ecological modeling approaches to quantify and predict suitable environments, abundances, and connectivity for three mosquito vector species important to human and domestic livestock health. The first chapter delivers a global model of suitable environments for *Aedes aegypti* and *Ae. albopictus* under present and future climate change

calibrated on presence-only data. The second chapter outlines a new approach to predicting *Ae. mcintoshi* abundances in Kenya and western Somalia at an 8-day temporal resolution during October to January from 2002 – 2015. The third chapter demonstrates the potential to investigate *Ae. mcintoshi* population genetic structure and associations between environmental variables across eastern Kenya using gene sequence data. Each of these chapters address individual research questions using a disease ecology approach, while contributing more broadly to knowledge of mosquito vector ecologies and the potential for human disease risk.

ACKNOWLEDEMENTS

I would like to acknowledge the hard work and dedication of my advisor, A. Townsend Peterson, to my academic and professional development. His guidance and willingness to support my interest in exploring a wide variety of research approaches led to my success in the doctoral program. Additionally, I would like to acknowledge Daniel Reuman for his expertise and guidance on the second chapter of this dissertation, and Alana Alexander for her collaboration on the third chapter. I would also like to thank Jorge Soberón for his openness to discussing research topics and helping me on a variety of projects during my time at KU. I met several wonderful students and friends through the KU Ecological Niche Modeling group, including in no particular order Lynnette Dornak, Naryani Barve, Hannah Owens, Erin Saupe, Corinne Meyers, Kate Ingenloff, Andrés Lira, Vijay Barve, Karen Olson, and many more! I would like to dedicate this dissertation work to my parents and sister, who have always believed in me and encouraged me to follow my interests, while inspiring me to dedicate my efforts toward a better future for everyone in our global community. Finally, I would like to express my profound gratitude to my husband David, who was willing to move to Kansas with me, so that I could pursue my academic and professional goals.

Table of Contents

ABSTRACT.....	iii
ACKNOWLEDGEMENTS	v
INTRODUCTION.....	1
CLIMATE CHANGE INFLUENCES ON GLOBAL GEOGRAPHIC	4
DISTRIBUTIONS OF DENGUE AND CHIKUNGUNYA VIRUS VECTORS	
Methods.....	5
Results	9
Discussion.....	16
Acknowledgements	22
References	22
A NEW APPROACH FOR PREDICTING ABUNDANCES.....	30
OF <i>Aedes mcintoshi</i>, A PRIMARY RIFT VALLEY FEVER VIRUS MOSQUITO	
VECTOR	
Methods.....	33
Results and Discussion.....	38
Acknowledgements	48
References	48
LANDSCAPE GENETICS OF <i>Aedes mcintoshi</i>, AN IMPORTANT	64
VECTOR OF RIFT VALLEY FEVER VIRUS IN EASTERN KENYA	
Methods.....	66

Results	69
Discussion.....	76
Acknowledgements	80
References	80
CONCLUSION	91
References	92

INTRODUCTION

Vector-borne and zoonotic diseases comprise a serious public health concern globally (Reisen 2010, Kilpatrick and Randolph 2012). Vector-borne diseases are infections transmitted to humans, usually by arthropods, such as insects or ticks (Meade and Earickson 2000). Over the past 30 years, an increase in newly-emerging vector-borne pathogens, coupled with the broader dispersal of known pathogens, has resulted in substantial challenges for public health intervention and prevention programs (Kilpatrick and Randolph 2012). This burden highlights the need for continued improvement of modeling approaches and prediction methods to help identify areas vulnerable to infection, thereby contributing toward more efficient distributions of limited public health resources.

The field of epidemiology centers largely on characterizing disease risk as a function of detected occurrence in humans (Last et al. 1995), but often, humans do not play a central role in pathogen maintenance in the environment (Pavlovsky 1966). This knowledge, although not recognized widely in traditional epidemiology (Peterson 2014), is not a new concept. In the 1930s, E.N. Pavlovsky stated formally that zoonotic disease systems exist in the natural environment independent of disease incidence in humans, noting that risk in humans occurred only when encountering these environments (Pavlovsky 1966). The field of disease ecology stems from this paradigm, emphasizing interactions between disease system components and the natural environment, recognizing that humans are not always the catalyst for pathogen dispersal and distributions (Ostfeld et al. 2008, Peterson 2014). Pavlovsky stated further that identifying environmental factors and geographic barriers that limit the spatial distribution of a disease provides the information necessary to map and to predict future disease risk in a given area (Pavlovsky 1966).

While incorporating environmental factors in assessing potential pathogen risk is a logical first step, complexities in this approach exist because pathogens are nested within the broader community ecology of host, vector, and reservoir species, and often, not all of these elements are known (Plowright et al. 2008). Although this element poses challenges to understanding limiting factors of specific environmental pathogens, the multitude of components within individual disease systems offer several avenues from which to study patterns, providing insight into risk (Plowright et al. 2008). Disease vectors, and more specifically, mosquito vectors are one such component. This knowledge, coupled with advances in geospatial technologies, provides excellent opportunities to model environmental factors contributing to potential pathogen distributions and to help predict disease risk in humans (Kalluri et al. 2007).

Here, I present three ecological modeling approaches to quantify and predict suitable environments, abundances, and connectivity for three mosquito vector species important to human and domestic livestock health. The first chapter delivers a global model of suitable environments for *Aedes aegypti* and *Ae. albopictus* under present and future climate change calibrated on presence-only data. *Aedes aegypti* and *Ae. albopictus* are important vectors of dengue, chikungunya, and Zika viruses, as well as being implicated as vectors of yellow fever, Japanese encephalitis, and a suite of additional pathogens (Rogers et al. 2006, Hayes 2009, Pages et al. 2009, Lambrechts et al. 2010, Jentes et al. 2011). The second chapter outlines a new approach to predicting *Ae. mcintoshi* abundances in Kenya and western Somalia at an 8-day temporal resolution during October to January from 2002 – 2015. *Aedes mcintoshi* is a primary mosquito vector for Rift Valley fever virus in East Africa (Linthicum et al. 1985, Pepin et al. 2010, Rosmoser et al. 2011, Tchouassi et al. 2014), an important human and livestock disease that results in devastating economic losses in affected regions (Murithi et al. 2011, Chengula et

al. 2013). The third chapter demonstrates the potential to investigate *Ae. mcintoshi* population genetic structure and associations between environmental variables across eastern Kenya using gene sequence data. Understanding connectivity between vector populations across a landscape provides a new toolset from which to predict the potential for pathogen dispersal across different geographic regions. Each of these chapters address individual research questions using a disease ecology approach, while contributing more broadly to knowledge of mosquito vector ecologies and the potential for human disease risk.

CHAPTER 1

Climate Change Influences on Global Distributions of Dengue and Chikungunya Virus Vectors

The global distributional potential of mosquito-borne viruses has seen considerable research attention in recent years, and particularly as regards viruses transmitted by *Aedes* mosquitoes (Brady et al. 2012; Charrel et al. 2007; Jentes et al. 2011; Rogers et al. 2006; Van Kleef et al. 2010). These diseases—e.g., dengue, yellow fever, and chikungunya—present serious public health concerns, particularly in light of recent spread events, in which each of the diseases has emerged either in new regions (Depoortere et al. 2008) or in new environments (Massad et al. 2003; Vasconcelos et al. 2001). As a consequence, *Aedes*-transmitted viral diseases have been of considerable concern in recent years across much of the Tropics and Subtropics worldwide.

The situation for this suite of diseases is complicated still more by the global spread of two vector species, *Aedes aegypti* and *Ae. albopictus* (Benedict et al. 2007; Caminade et al. 2012; Erickson et al. 2012; Simard et al. 2005). These two species have spread essentially worldwide at lower and middle latitudes in recent decades, providing new or newly reinfested, particularly efficient vectors for transmission of ‘forest diseases’ to humans, particularly in periurban settings (Massad et al. 2003). Of particular interest are a series of intriguing results regarding interactions between the two species: *Ae. albopictus* appears to be the superior larval competitor (Braks et al. 2004; Juliano et al. 2004; Lounibos et al. 2002), yet *Ae. aegypti* appears to be the vector species responsible for most or all massive outbreaks of dengue (Lambrechts et al. 2010). As a consequence, quite a bit is to be gained from a deep understanding of the present geographic distribution of these species, which has been the focus of several recent studies

(Benedict et al. 2007; Bhatt et al. 2013; Caminade et al. 2012; Fischer et al. 2014; Lambrechts et al. 2010); however, in light of already-occurring changes in climate and the possible distributional opportunities that these open for the mosquitoes, as well as the ongoing evolution of the actual distributions of these two globally invasive species (ALA 2014; Derraik 2006; Holder et al. 2010), a predictive view of their distributional potential into coming decades as global climates change is also useful.

In this paper, then, we use correlative ecological niche modeling approaches (Peterson et al. 2011) to evaluate and assess the distributional potential of *Aedes aegypti* and *Ae. albopictus* at present and, most importantly, into the future. We use diverse models and scenarios for future climate conditions, and constrain our results and interpretations carefully to avoid overinterpretation of pattern in the data. The chief result—that, with the exception of smaller regional shifts, distributional potential of these two species will likely be relatively stable in coming decades—can be understood as the consequence of broad climate tolerances by these two species, such that changes in global climate over coming decades may not translate into major distributional shifts in these vector species. Model results also, however, point to the potential for reorganization of the distributions of the two species, each in response to its own particular ecological niche profile, in several areas, which may have implications for disease transmission.

Methods

We collected primary occurrence data (i.e., data documenting occurrences of individual animals at points in time and space) for the two focal species (*Aedes aegypti* and *Ae. albopictus*), and indeed records for the entire genus *Aedes*, accumulating digital accessible records from

VectorMap (<http://vectormap.org/>), Atlas of Living Australia (<http://www.ala.org.au/>), speciesLink (<http://www.splink.org.br/>), and GBIF (<http://www.gbif.org/>; searches made 09 April 2013; individual totals of records not possible, as these data sources frequently overlap in their coverage of specific data records). The data for the two focal species were used to calibrate models, as is described in detail below; the data for the broader genus were used to characterize sampling effort across the Earth, which is quite uneven, focusing on the many species in the genus, such that collecting intensity and sampling techniques would be broadly representative of sampling that would yield records of the focal species; this step was necessitated by the almost-universal lack of reporting of negative data in biodiversity data sets (i.e., “searched for species X but did not find it”). Data were inspected to remove data records not referring to this genus of mosquitoes. Finally, to provide a qualitative independent test of model predictions, one of us (DML) extracted an independent data set from the broader scientific literature, gray literature, and personal collections; these data were not used in model calibration, for lack of information on underlying sampling, but provided a useful check on how comprehensive our model predictions really are.

To summarize sampling effort globally, we created a ‘fishnet’ at a spatial resolution of 10’ that was coincident in position and orientation with the environmental data that would be used in model calibration (see below). Via this polygon shapefile, we counted individual records out of the 118,134 overall records of the genus falling in each of the square polygons—these sampling intensities varied from 0 (covering approximately 99.5% of the terrestrial portion of the Earth’s surface) to as many as 454 data records in a single 10’ pixel. Using these counts as values, we created a raster data set that could be used to represent sampling intensity for this genus.

We characterized present-day climates (1950-2000) via data layers provided as part of the WorldClim climate data archive (Hijmans et al. 2005). Specifically, we used the 19 ‘bioclimatic’ variables (10’ resolution) that are derived from monthly average maximum temperature, monthly average minimum temperature, and monthly precipitation, all across the half-century time period. We obtained parallel data layers for general circulation model (GCM) projections to 2050 from the CMIP4 future-climate model projections (CI 2014); these future-climate projections summarized 4-8 different GCMs (i.e., distinct implementations of simulation models for global climate dynamics) for three scenarios of future-climate emissions (A2, B1, A1B).

Ecological niche models were calibrated in Maxent, version 3.3.3 (Phillips et al. 2006). Initial exploratory runs served to detect a number of apparent artifacts in some of the climate datasets: bio 8-9, bio 18-19 were particularly notable in the present-day coverages, apparently as a consequence of their linking between temperature and precipitation variables, so we removed them from analysis at the outset; a few other variables presented odd artifacts upon visual inspection, and also were removed. To reduce dimensionality, and to reduce collinearity among layers, we applied principal components analysis to the remaining bioclimatic layers, and used the component loadings in the present to transform the future-climate model projections in parallel. We used initial, exploratory Maxent runs with its jackknife functionality to explore the relative contributions of each of the principal components, and eliminated layers that showed consistently low contribution to model predictions. As a consequence, we ran final models based on sets of 6 and 8 principal components for each species.

To provide a general evaluation of the predictive ability of our models, and also to allow decisions about whether the 6- or 8-component models were preferable for interpretation, we used partial receiver operating characteristic (ROC) analysis (Peterson et al. 2008). Given

concerns about the appropriateness of traditional ROC applications (Lobo et al. 2008), which plots omission and commission errors across a set of model thresholds, and compares the area under that curve to the area under a line of null expectations, the partial ROC approach rescales one axis of the ROC curve to reflect proportional area identified as suitable (instead of commission error), and focuses only on predictions that have acceptable levels of omission error (termed E ; in this case, we used $E = 5\%$). We subset spatially unique occurrence points randomly into equal portions for model calibration and model evaluation. Partial ROC calculations were carried out via programs developed by N. Barve (<http://hdl.handle.net/1808/10059>). Probability values were determined by direct count of AUC ratios ≤ 1.0 , among 1000 replicate 50% bootstrap subsamplings.

In general, final models were calibrated in Maxent with its bootstrapping/replicated run functionality. In view of the apparently massive invasive potential of both mosquito species involved (Benedict et al. 2007; Jansen & Beebe 2010), we used a very broad hypothesis of the accessible area (M) for them, considering much of the world, and eliminating only the highest southern latitudes (i.e., $< 60^\circ S$). We set aside 50% of input occurrence points as a ‘test percentage,’ and conducted 10 replicate analyses to consider effects of particular calibration data sets on model outcomes. The sampling bias layer described above was included in the Maxent “bias layer” function as a guide to sampling towards a characterization of the broader ‘background’ for model fitting, akin to pseudoabsence sampling in other niche modeling algorithms (Phillips et al. 2009); this layer is illustrated in the Supplementary Materials, and is available as a GeoTIFF dataset at <http://hdl.handle.net/1808/15275>. Models were calibrated for present-day conditions, and then transferred onto each of the future-climate scenarios and models that were available to us (see Table 1 for a summary).

Table 1. Summary of future-climate scenarios (B1, A1B, and A2) and climate models assessed.

Model	Host country	A1B and A2	B1
BCCR-BCM 2.0	Norway	X	X
CSIRO-MK 3.0	Australia	X	
CSIRO-MK 3.5	Australia	X	X
INM-CM 3.0	Russia	X	X
MIROC medium resolution	Japan	X	X
NCAR-CCSM 3.0	United States	X	X

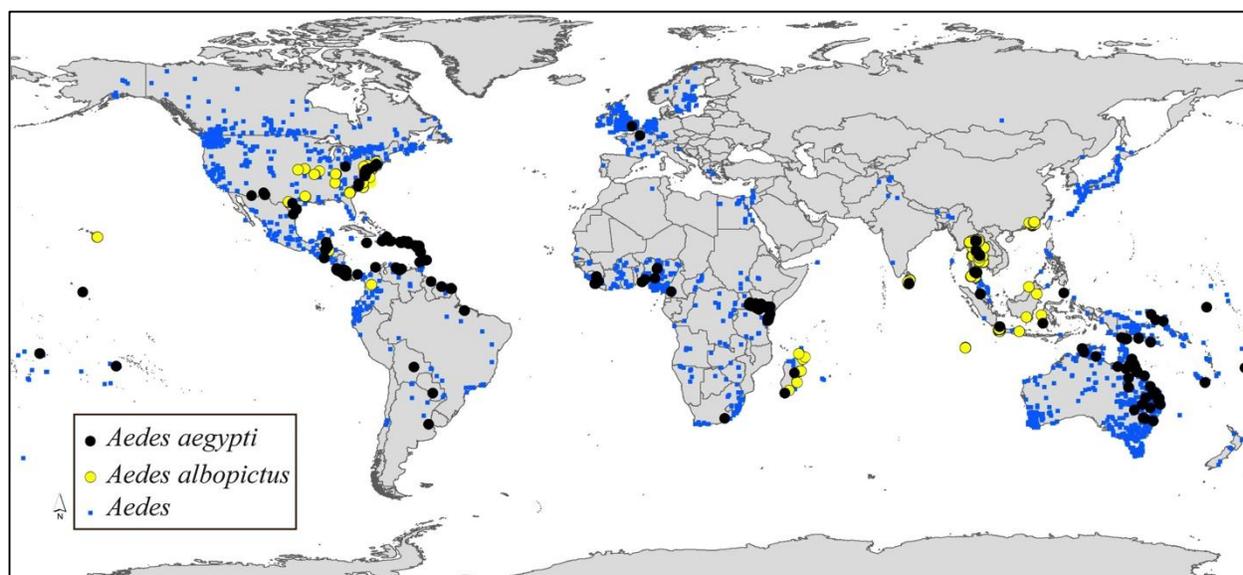
Once all models were calibrated and all future transfers developed, we used the median value for all replicate analyses for each combination of model, scenario, and number of principal components. We then calculated the median of medians for a given scenario and number of principal components, seeking a central tendency across all of the replicate models developed from different random resamplings from available occurrence data. We calculated differences between future and present in these logistic suitability scores; we also thresholded the data using an $E = 5\%$ training presence data criterion, in which we sought the highest suitability score that included 95% of the data used to calibrate the models, to convert raw suitability scores into binary estimates of potential presence and absence across regions, with the estimate of E derived from assessment of approximate error rates in input data (Peterson et al. 2008). Based on these binary outputs, we explored anticipated range expansions and contractions.

Results

In all, we assembled 2108 and 8040 occurrence records for *Aedes aegypti* and *Ae. albopictus*, respectively, which distilled down to 338 and 350 spatially unique records at the

resolution of our analysis. The overall picture of sampling of *Aedes* mosquitoes, however, indicated sampling of this genus as rather more concentrated in the temperate zone than are the populations of these two species (Figure 1). In preliminary explorations, inclusion of this surface as a ‘bias layer’ in model calibration made significant differences in model outputs, such that we emphasize the importance of considering sampling effort in any modeling exercises that are cast on an uneven landscape of sampling, even within the accessible area for a species.

Figure 1. Summary of primary occurrence data available globally for *Aedes* mosquitoes in general (blue), and *Ae. aegypti* (black) and *Ae. albopictus* (yellow) in particular.



Calibrating models across the entire accessible region for the species (i.e., most of the world) based on a subset of available occurrences resulted in model predictions that performed uniformly better than random expectations. That is, using partial ROC analysis on random 50% subsets of data set aside from model calibration, all comparisons yielded AUC ratios that were significantly above null expectations ($P < 0.001$). Comparing among different numbers of

principal components as descriptors of environmental spaces, AUC ratios for 8-component models were higher than those for 6-component models by 1.72-2.57%; hence, we focused all subsequent explorations on models based on 8 principal components (6-component results are presented in the Supplementary Materials, and do not differ markedly from those for 8-component models).

Model predictions for the present day reflected well the known global distributions of each of the two species (Figure 2). Indeed, some of the more interesting features include an indication of markedly broader distributional potential for *Aedes aegypti* than for *Ae. albopictus* in Australia, and indeed perhaps overall broader distributional potential in tropical and subtropical regions for *Ae. aegypti*, but then, curiously, a broader overall potential distribution of *Ae. albopictus* in North America. These explorations also confirm that these two successful invasive species have indeed fulfilled much or all of their invasive potential globally—broad regions that appear to be suitable climatically, but from which no occurrence data were available to us included eastern China, Vietnam, Philippines, Borneo, and northeastern Brazil for *Ae. aegypti*, and South America, Africa, India, Japan, and New Zealand for *Ae. albopictus*. Nonetheless, searches via Google Scholar quickly revealed known presences of these species in almost all of these areas (Dias et al. 1997; Honório et al. 2003; Huber et al. 2002; Kalra et al. 1997; Kobayashi et al. 2002; Macdonald & Rajapaksa 1972; Pagès et al. 2009; Savage et al. 1992; Schultz 1993; Tsuda et al. 2001), so the gap is one of availability of occurrence data, rather than absence of the species from these potential areas. The sole significant exception is that of New Zealand, which has been reached by *Ae. albopictus*, but apparently does not yet have established populations (Derraik 2006; Holder et al. 2010; Laird et al. 1994).

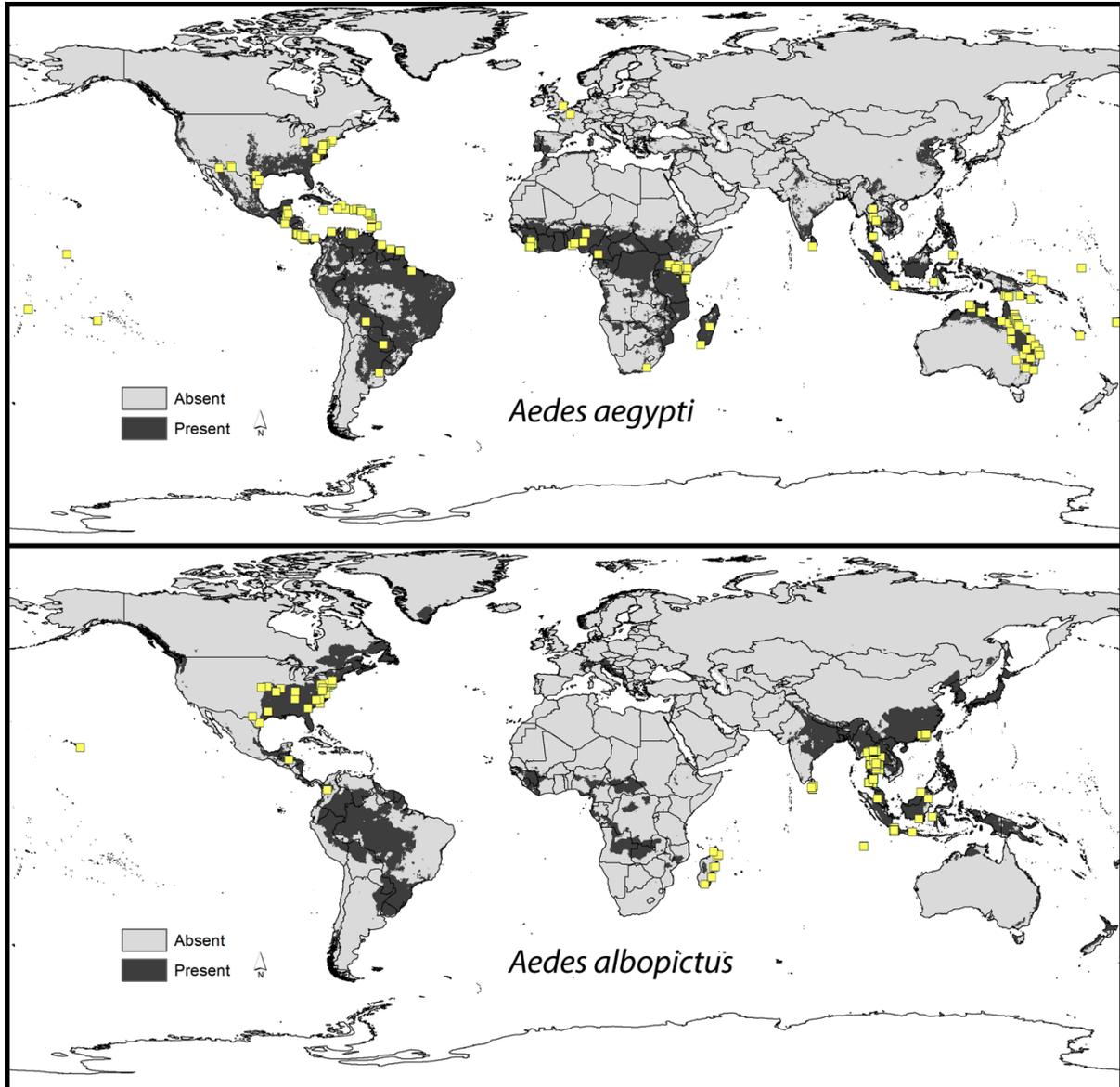


Figure 2. Summary of modeled potential distributional patterns of *Aedes aegypti* and *Ae. albopictus* under present-day conditions based on analysis of 8 principal components.

Model transfers to future conditions were generally similar to present-day patterns, at least in the broadest terms (A1B scenario shown in Figure 3; other scenarios provided in Supplementary Materials). For *Aedes aegypti*, model predictions indicated some potential for northward expansion in eastern North America, South Asia, and East Asia, and southward in Africa and Australia; broadening distributional potential was indicated in interior South America and Central Africa. For *Ae. albopictus*, model predictions gave clearer indications of expanding distributional potential in eastern North America and East Asia, plus expanding potential across Africa and in eastern and southern South America; distributional potential in Australia was anticipated to expand rather markedly for this species.

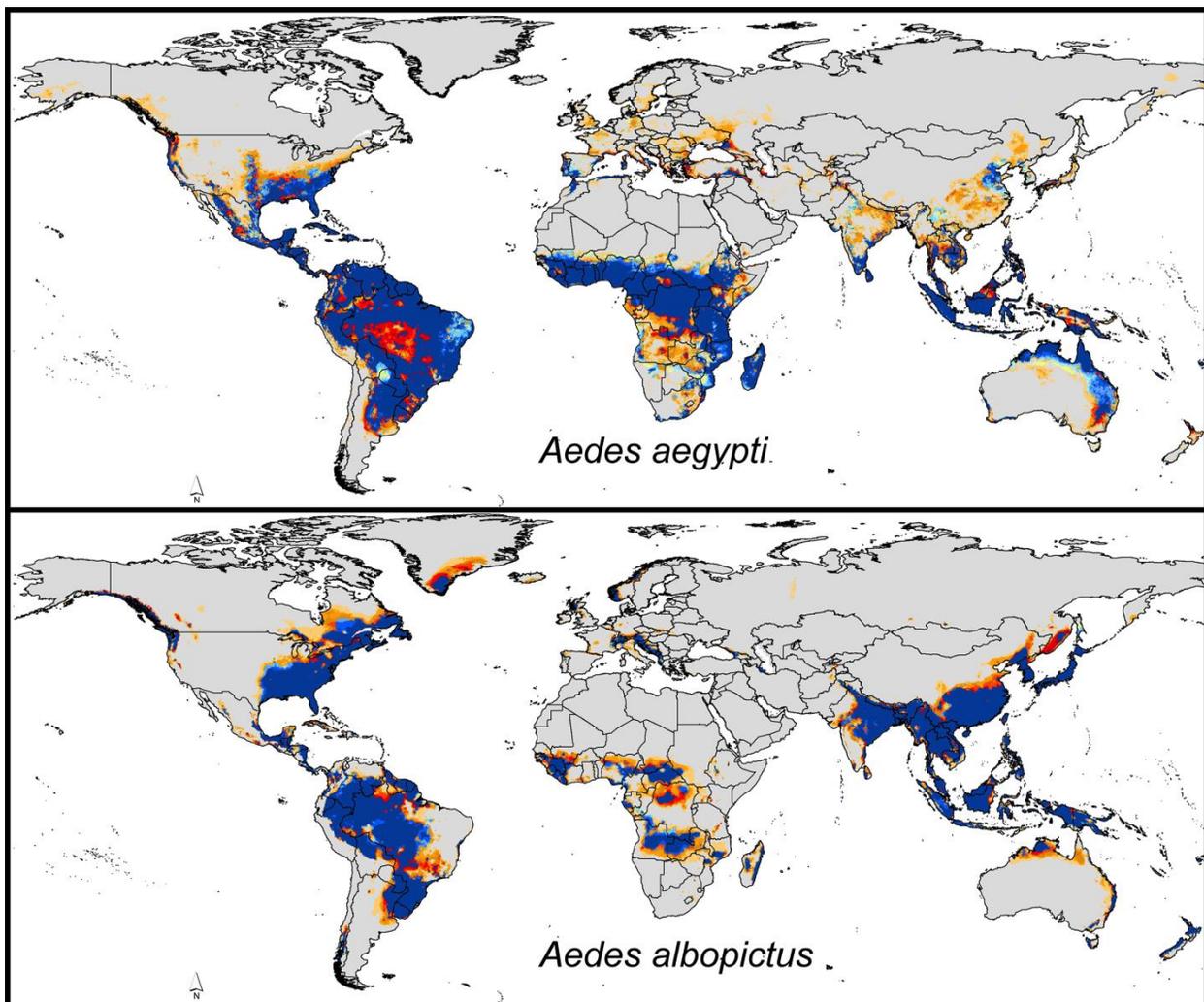


Figure 3. Summary of modeled potential distributional patterns of *Aedes aegypti* and *Ae. albopictus* under future conditions (A1B) based on analysis of 8 principal components.

Confidence in present-day and future distributional potential is based on agreement among 6 climate models (Table 1): present-day-only distributional areas are shown in blue, with model agreement regarding stability of present-day distributional areas shown via the intensity of blue shading (light blue = low, dark blue = high model agreement); future distributional potential is shown as shades of orange (light orange = low, dark orange = high model agreement in predicting future suitability).

Assessing the distributional potential of the two species in tandem suggests a complex mosaic of the distributions of the two species around the world (Figure 4): while *Aedes albopictus* has the broadest potential distribution in North America, *Ae. aegypti* appears to have a broader potential in Africa and Australia. Assessing how the distributional patterns of the two species will change, continent by continent, we see complex rearrangements (Figure 5): *Ae. aegypti* appears to gain potential distributional area in South America, lose ground in Australia, and rearrange its distributional area in Europe, while *Ae. albopictus* may be able to expand its distributional area more than *Ae. aegypti* in North America; Asia shows particularly complicated shifts in dominance of the two species geographically.

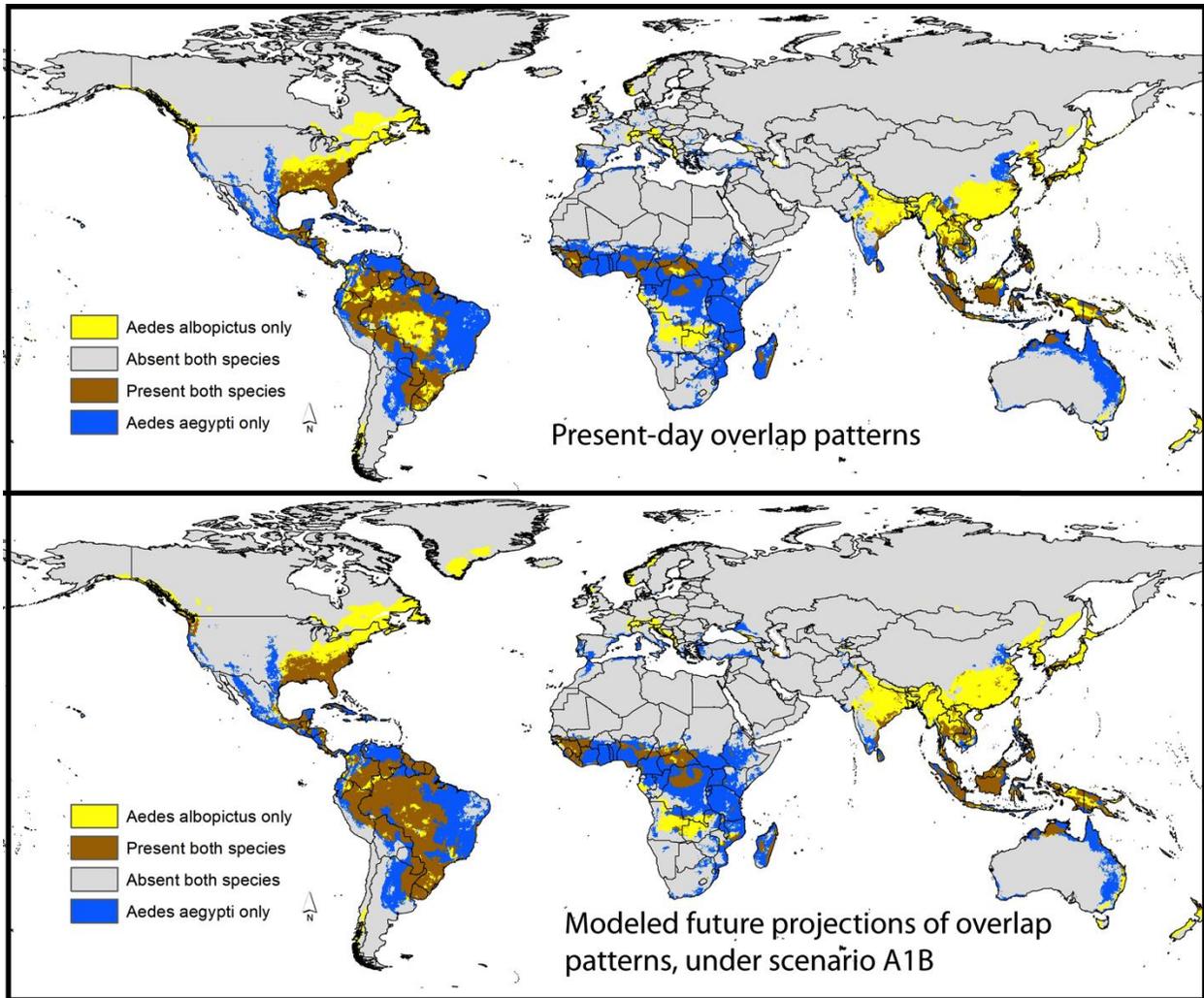


Figure 4. Summary of patterns of potential distributional overlap derived from ecological niche models of *Aedes aegypti* and *Ae. albopictus* worldwide, both under current conditions and under modeled future conditions (A1B scenario), based on 8 principal components.

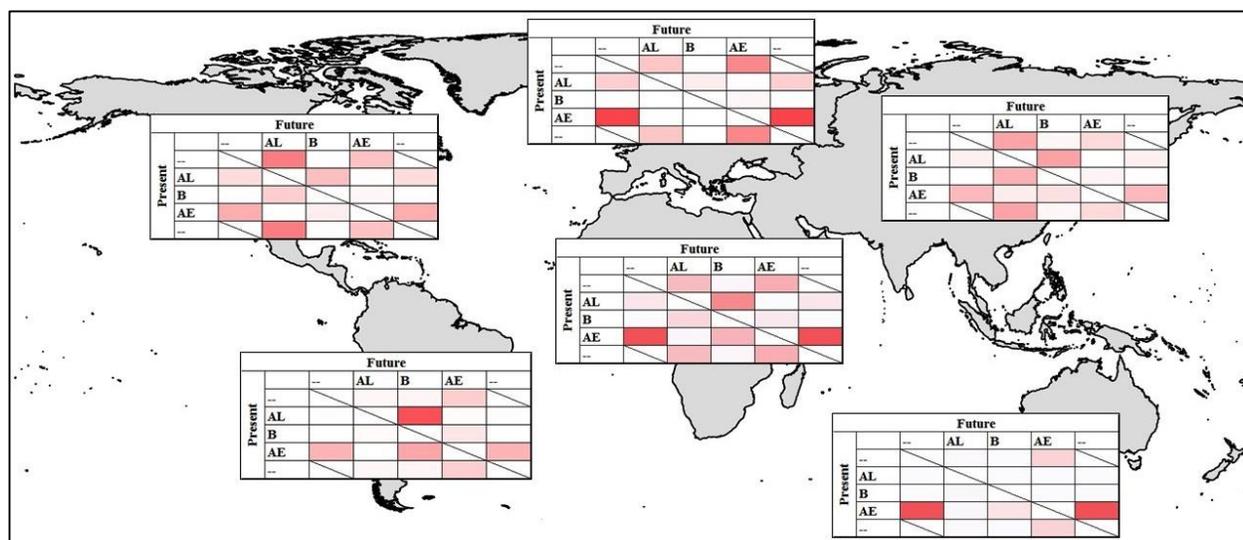


Figure 5. Summary of patterns of change in potential presence and distributional overlap of *Aedes aegypti* and *Ae. albopictus* worldwide under the A1B scenario, continent by continent. Patterns of shift under climate change scenario A1B are shown via transition matrices for each continent: rows indicate present-day situation, and columns indicate model projections of future potential. AL = *Ae. albopictus*, AE = *Ae. aegypti*, B = both species, -- = neither species.

Discussion

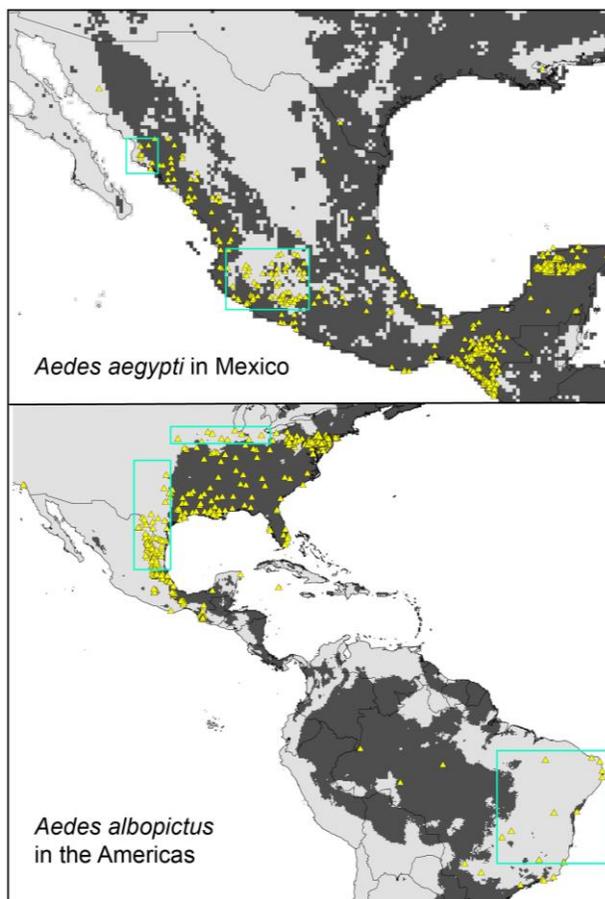
Mosquito distributions are highly dynamic in space and time, as their life cycles are short and heavily influenced by environmental variation (Crans 2004). Therefore, their distributions respond fluidly to new opportunities, e.g., in terms of (1) overcoming dispersal barriers to colonize new areas (e.g., the global invasion of the two species under analysis herein) (Benedict et al. 2007), (2) ready expansion into new areas as conditions change and improve for their population biology (Roiz et al. 2011), and (3) the dynamic interactions and potential competitive exclusion between the two species (Braks et al. 2004; Juliano et al. 2004; Lounibos et al. 2002). As such, for example, the difference between the two species in terms of Australian distribution is intriguing—while *Ae. aegypti* is present and established, *Ae. albopictus* is not established

there, in spite of having been introduced to at least four states (ALA 2014): based on our results, we view this lack of establishment as reflecting non-ideal conditions for the species in Australia, although Fischer et al. (2011) presented a more optimistic and suitable picture for the species under present-day conditions in Australia, and certainly the Australian Quarantine Service has expended considerable effort in assuring its non-establishment. Similarly demonstrated colonization pressure by this species without successful establishment has occurred in New Zealand, though probably thanks to rapid and effective mitigation efforts (Holder et al. 2010). More generally, in view of the impressive dispersal and colonization abilities of these two species, our models of potential distributional areas replicate quite well their actual present-day distributional areas.

Our models, while largely coincident with known present-day distributions of the two species, were developed to explore climate change implications for these two species. Still, these models did fail in replicating some aspects of the species' known, present-day distributions, as can be appreciated from Figure 6, with the overlay of independent occurrence data. For example, model predictions failed to identify suitable habitat for *Ae. albopictus* in northeastern Mexico and northeastern Brazil. This omission error was likely a consequence of incomplete sampling in our training data set, resulting in under-representation or no representation of environmental combinations manifested in these regions.. Although incomplete input data may account for some model failures, anthropogenic factors could also impact model results locally: for example, *Aedes albopictus* populations introduced into and established in Los Angeles, in southern California (Zhong et al. 2013) may be able to survive there thanks to urban subsidy of moisture, but the details are not clear. In addition to omissions, model predictions appeared to extrapolate into several high-latitude regions (e.g., Greenland), reflecting model extrapolation beyond the set

of environmental conditions associated with the calibration region. These inconsistencies may derive from peripherality of environmental characteristics of occurrences within the calibration region, leading to wild and inappropriate extrapolation (Owens et al. 2013).

Figure 6. Overlay of independent occurrence data set (yellow triangles) on model predictions for present-day potential distributions of *Aedes aegypti* and *Ae. albopictus*. Green boxes delimit areas in which model predictions failed to anticipate known occurrences.



Our future-climate projections, however, indicate considerable potential for shifting establishment of these two vector species in several directions, as follows: (1) more broadly in

eastern North America, (2) farther south in southern South America (particularly in *Ae. aegypti*); (3) locally northward into southern Europe, (4) more broadly in Central Africa, (5) more broadly in East Asia, and (6) across northern and eastern Australia (*Ae. albopictus*). While these potential range shifts in response to climate are relatively subtle, they are the implications of environmental shifts that would transform areas from presenting conditions outside of the modeled ecological niches of the vector species (e.g., no establishment in Australia despite repeated introductions) to presenting conditions that do permit maintenance of populations. Clearly, even within the present distributional areas of these species, changing conditions may cause changes in their abundance and dominance, although those shifts are not included in our present modeling efforts (Martínez-Meyer et al. 2012).

Regional-scale modeling efforts have shed considerable light on the distributional potential of *Ae. albopictus*, thanks to a series of detailed analyses (Fischer et al. 2014; Fischer et al. 2011). These studies have provided a level of rigor that is unusual in the present literature on climate change effects on species' distributions, and offers considerable detail on likely distributional shifts of this species across Europe as climates change. Both the Fischer et al. (2011) models and our own results indicate a curious middle-latitude focus of suitability for *Ae. albopictus* in Europe, which contrasts with the southern and southeastern suitability profiles for this species in North America.

The relative roles and importance of these two species in dengue transmission appear to be unequal. That is, a recent analysis indicated that *Ae. albopictus* does not appear to drive major dengue outbreaks (Lambrechts et al. 2010), such that *Ae. aegypti* emerges as the chief driver of large-scale dengue transmission. Both the pattern of known occurrences and the maps of potential distributions for the two species, in large part, suggest that *Ae. aegypti* is more

extensive distributionally than *Ae. albopictus*, with the exceptions of eastern North America and East Asia, neither of which has a long history of dengue transmission; both of these regions, however, appear now to have active transmission (Radke et al. 2012; Ramos et al. 2008; Wu et al. 2010). These imbalances, however, may not hold for other *Aedes*-transmitted diseases, such as chikungunya, which is readily transmitted by *Ae. albopictus*, at least in some cases (Tsetsarkin et al. 2011), and which has shown recent major distributional expansion (Fischer et al. 2013; Van Bortel et al. 2014), with importation events presenting potential for further colonization (Centers for Disease Control and Prevention 2006).

The global distribution of *Aedes*-borne viruses, particularly dengue, has seen a lot of attention in the scientific literature in recent years (e.g., Bhatt et al. 2013; Brady et al. 2012). However, these efforts have not included deep contemplation and detailed mapping of distributions and distributional shifts in vector species: rather, they have focused in largest part on human infections. While clearly vector populations can exist without virus presence, such populations set the stage for easy disease introduction, as is evidenced by recent dengue emergence in the southern USA (Radke et al. 2012; Ramos et al. 2008), particularly as both vector populations and human populations rearrange spatially in response to climate change (in the case of the vectors) and economic opportunity (humans). In this sense, we see the results of the present study as offering a quantitative input that can be incorporated into future summaries of human disease distributions—to this end, we have placed GeoTIFF versions of our raster model outputs in a data repository at the University of Kansas (<http://hdl.handle.net/1808/15275>).

The models presented herein are far from definitive, however. While this study is novel in its global scope with consideration of multiple models and climate change scenarios, several

‘next steps’ emerge as highly desirable. First, the availability of large quantities of occurrence data for mosquitoes globally is simultaneously an opportunity and a constraint: the opportunity is that as much data as we found were readily available for analysis; however, these data are still sparse and even lacking for some key regions (e.g., most of Asia, much of South America, parts of the United States). Clearly, more data exist, but are cloistered in national, institutional, and even personal research databases, and are not available to the broader community.

Second, distributional complexities need also to be incorporated into these models. The competitive interactions between the two species when they co-occur present a first dimension of complexity. Another major trend in dengue transmission has been its urbanization in many parts of the world—this trend, and the unique opportunities offered to the mosquitoes by urban environments, is not reflected in these models, and yet is crucial to understanding present-day dengue transmission (Jansen & Beebe 2010). Multiscalar, nested models may represent a useful way forward for dealing with such effects.

Finally, a major factor in transmission of mosquito-borne diseases is the effect of temporal and spatiotemporal dynamics of conditions. For example, a recent analysis indicated the potential for chikungunya transmission across the eastern United States, but in very limited seasons at the northernmost site analyzed (New York City), such that transmission would not be at all efficient (Ruiz-Moreno et al. 2012). A previous study by our research group (Peterson et al. 2005) indicated clear predictivity of spatiotemporal dynamics of dengue vector mosquito populations across Mexico, but further exploration of this potential has been stymied by lack of adequate and sufficiently dense (in time and space) occurrence data for the species. Overall, though, the indication is of a system that is highly predictable, and one in which climate change

is likely to produce distributional shifts that, while not massive, will have significant public health implications worldwide.

Acknowledgments

We thank our colleagues Rick Wilkerson and Des Foley at WRAIR, for their efforts to enable broad and open sharing of data regarding mosquito distributions, and we thank the organizers of this collection of papers for their kind invitation to contribute.

References

- ALA. 2014 *Atlas of Living Australia*; <http://www.ala.org.au/>, accessed 2 July 2014. Canberra: Atlas of Living Australia.
- Benedict, M. Q., Levine, R. S., Hawley, W. A. & Lounibos, L. P. 2007 Spread of the tiger: Global risk of invasion by the mosquito *Aedes albopictus*. *Vector-Borne and Zoonotic Diseases* **7**, 76-85.
- Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G. & Sankoh, O. 2013 The global distribution and burden of dengue. *Nature* **496**, 504-507.
- Brady, O. J., Gething, P. W., Bhatt, S., Messina, J. P., Brownstein, J. S., Hoen, A. G., Moyes, C. L., Farlow, A. W., Scott, T. W. & Hay, S. I. 2012 Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Neglected Tropical Diseases* **6**, e1760.
- Braks, M. A. H., Honório, N., Lounibos, L., Lourenço-de-Oliveira, R. & Juliano, S. 2004 Interspecific competition between two invasive species of container mosquitoes, *Aedes*

- aegypti* and *Aedes albopictus* (Diptera: Culicidae), in Brazil. *Annals of the Entomological Society of America* **97**, 130-139.
- Caminade, C., Medlock, J. M., Ducheyne, E., McIntyre, K. M., Leach, S., Baylis, M. & Morse, A. P. 2012 Suitability of European climate for the Asian tiger mosquito *Aedes albopictus*: recent trends and future scenarios. *Journal of the Royal Society Interface* **9**, 2708-2717.
- Centers for Disease Control and Prevention. 2006 Chikungunya fever diagnosed among international travelers--United States, 2005-2006. *Morbidity and Mortality Weekly Report* **55**, 1040-1042.
- Charrel, R. N., de Lamballerie, X. & Raoult, D. 2007 Chikungunya outbreaks--The globalization of vectorborne diseases. *New England Journal of Medicine* **356**, 769.
- CI. 2014 *Globally downscaled climate projections for assessing the conservation impacts of climate change*; <http://futureclimates.conservation.org/>, accessed 2 July 2014. Washington, D.C.: Conservation International.
- Crans, W. J. 2004 A classification system for mosquito life cycles: Life cycle types for mosquitoes of the northeastern United States. *Journal of Vector Ecology* **29**, 1-10.
- Depoortere, E., Salmaso, S., Pompa, M., Guglielmetti, P. & Coulombier, D. 2008 Chikungunya in Europe. *Lancet* **371**, 723.
- Derraik, J. G. 2006 A scenario for invasion and dispersal of *Aedes albopictus* (Diptera: Culicidae) in New Zealand. *Journal of Medical Entomology* **43**, 1-8.
- Dias, J., Pedral-Sampaio, D. B. & Jones, T. C. 1997 *Aedes aegypti* surveillance and correlation with the occurrence of dengue fever in Bahia, Brazil. *Brazilian Journal of Infectious Diseases* **1**, 36-41.

- Erickson, R. A., Hayhoe, K., Presley, S. M., Allen, L. J. S., Long, K. R. & Cox, S. B. 2012 Potential impacts of climate change on the ecology of dengue and its mosquito vector the Asian tiger mosquito (*Aedes albopictus*). *Environmental Research Letters* **7**, 034003.
- Fischer, D., Thomas, S., Neteler, M., Tjaden, N. & Beierkuhnlein, C. 2014 Climatic suitability of *Aedes albopictus* in Europe referring to climate change projections: Comparison of mechanistic and correlative niche modelling approaches. *Eurosurveillance* **19**, 20696.
- Fischer, D., Thomas, S. M., Niemitz, F., Reineking, B. & Beierkuhnlein, C. 2011 Projection of climatic suitability for *Aedes albopictus* Skuse (Culicidae) in Europe under climate change conditions. *Global and Planetary Change* **78**, 54-64.
- Fischer, D., Thomas, S. M., Suk, J. E., Sudre, B., Hess, A., Tjaden, N. B., Beierkuhnlein, C. & Semenza, J. C. 2013 Climate change effects on Chikungunya transmission in Europe: geospatial analysis of vector's climatic suitability and virus' temperature requirements. *Int J Health Geogr* **12**, 51.
- Hijmans, R., Cameron, S., Parra, J., Jones, P. & Jarvis, A. 2005 Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965-1978.
- Holder, P., George, S., Disbury, M., Singe, M., Kean, J. M. & McFadden, A. 2010 A biosecurity response to *Aedes albopictus* (Diptera: Culicidae) in Auckland, New Zealand. *Journal of Medical Entomology* **47**, 600-609.
- Honório, N. A., Silva, W. d. C., Leite, P. J., Gonçalves, J. M., Lounibos, L. P. & Lourenço-de-Oliveira, R. 2003 Dispersal of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) in an urban endemic dengue area in the state of Rio de Janeiro, Brazil. *Memórias do Instituto Oswaldo Cruz* **98**, 191-198.

- Huber, K., Le Loan, L., Hoang, T. H., Tien, T. K., Rodhain, F. & Failloux, A. 2002 Temporal genetic variation in *Aedes aegypti* populations in Ho chi Minh City (Vietnam). *Heredity* **89**, 7-14.
- Jansen, C. C. & Beebe, N. W. 2010 The dengue vector *Aedes aegypti*: What comes next. *Microbes and Infection* **12**, 272-279.
- Jentes, E. S., Pomeroy, G., Gershman, M. D., Hill, D. R., Lemarchand, J., Lewis, R. F., Staples, J. E., Tomori, O., Wilder-Smith, A. & Monath, T. P. 2011 The revised global yellow fever risk map and recommendations for vaccination, 2010: Consensus of the Informal WHO Working Group on Geographic Risk for Yellow Fever. *Lancet Infectious Diseases* **11**, 622-632.
- Juliano, S. A., Lounibos, L. P. & O'Meara, G. F. 2004 A field test for competitive effects of *Aedes albopictus* on *A. aegypti* in South Florida: Differences between sites of coexistence and exclusion? *Oecologia* **139**, 583-593.
- Kalra, N., Kaul, S. & Rastogi, R. 1997 Prevalence of *Aedes aegypti* and *Aedes albopictus*: Vectors of dengue and dengue haemorrhagic fever in north, north-east and central India. *Dengue Bulletin* **21**, 84-92.
- Kobayashi, M., Nihei, N. & Kurihara, T. 2002 Analysis of northern distribution of *Aedes albopictus* (Diptera: Culicidae) in Japan by geographical information system. *Journal of Medical Entomology* **39**, 4-11.
- Laird, M., Calder, L., Thornton, R., Syme, R., Holder, P. & Mogi, M. 1994 Japanese *Aedes albopictus* among four mosquito species reaching New Zealand in used tires. *Journal of the American Mosquito Control Association* **10**, 14-23.

- Lambrechts, L., Scott, T. W. & Gubler, D. J. 2010 Consequences of the expanding global distribution of *Aedes albopictus* for dengue virus transmission. *PLoS Neglected Tropical Diseases* **4**, e646.
- Lobo, J. M., Jiménez-Valverde, A. & Real, R. 2008 AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* **17**, 145-151.
- Lounibos, L., Suárez, S., Menéndez, Z., Nishimura, N., Escher, R., O'Connell, S. & Rey, J. 2002 Does temperature affect the outcome of larval competition between *Aedes aegypti* and *Aedes albopictus*? *Journal of Vector Ecology* **27**, 86-95.
- Macdonald, W. & Rajapaksa, N. 1972 A survey of the distribution and relative prevalence of *Aedes aegypti* in Sabah, Brunei, and Sarawak. *Bulletin of the World Health Organization* **46**, 203-209.
- Martínez-Meyer, E., Díaz-Porras, D., Peterson, A. T. & Yáñez-Arenas, C. 2012 Ecological niche structure determines rangewide abundance patterns of species. *Biology Letters* **9**, 20120637.
- Massad, E., Burattini, M. N., Coutinho, F. A. B. & Lopez, L. F. 2003 Dengue and the risk of urban yellow fever reintroduction in São Paulo state, Brazil. *Revista de Saúde Pública* **37**, 477-484.
- Owens, H. L., Campbell, L. P., Dornak, L. L., Saupe, E. E., Barve, N., Soberon, J., Ingenloff, K., Lira-Noriega, A., Hensz, C. M., Myers, C. E. & Peterson, A. T. 2013 Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling* **263**, 10-18.

- Pagès, F., Peyrefitte, C. N., Mve, M. T., Jarjaval, F., Brisse, S., Iteman, I., Gravier, P., Nkoghe, D. & Grandadam, M. 2009 *Aedes albopictus* mosquito: The main vector of the 2007 Chikungunya outbreak in Gabon. *PLoS One* **4**, e4691.
- Peterson, A. T., Martínez-Campos, C., Nakazawa, Y. & Martínez-Meyer, E. 2005 Time-specific ecological niche modeling predicts spatial dynamics of vector insects and human dengue cases. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **99**, 647-655.
- Peterson, A. T., Papeş, M. & Soberón, J. 2008 Rethinking receiver operating characteristic analysis applications in ecological niche modelling. *Ecological Modelling* **213**, 63-72.
- Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M. & Araújo, M. B. 2011 *Ecological Niches and Geographic Distributions*. Princeton: Princeton University Press.
- Phillips, S., Dudík, M., Elith, J., Graham, C., Lehman, A., Leathwick, J. & Ferrier, S. 2009 Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications* **19**, 181-197.
- Phillips, S. J., Anderson, R. P. & Schapire, R. E. 2006 Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231-259.
- Radke, E. G., Gregory, C. J., Kintziger, K. W., Sauber-Schatz, E. K., Hunsperger, E. A., Gallagher, G. R., Barber, J. M., Biggerstaff, B. J., Stanek, D. R. & Tomashek, K. M. 2012 Dengue outbreak in Key Eest, Florida, USA, 2009. *Emerging Infectious Diseases* **18**, 135-137.
- Ramos, M. M., Mohammed, H., Zielinski-Gutierrez, E., Hayden, M. H., Lopez, J. L. R., Fournier, M., Trujillo, A. R., Burton, R., Brunkard, J. M., Anaya-Lopez, L., Banicki, A. A., Morales, P. K., Smith, B., Munoz, J. L., Waterman, S. H. & The Dengue Serosurvey

- Working Group. 2008 Epidemic dengue and dengue hemorrhagic fever at the Texas-Mexico border: Results of a household-based seroepidemiologic survey, December 2005. *American Journal of Tropical Medicine and Hygiene* **78**, 364-369.
- Rogers, D., Wilson, A., Hay, S. & Graham, A. 2006 The global distribution of yellow fever and dengue. *Advances in Parasitology* **62**, 181-220.
- Roiz, D., Neteler, M., Castellani, C., Arnoldi, D. & Rizzoli, A. 2011 Climatic factors driving invasion of the tiger mosquito (*Aedes albopictus*) into new areas of Trentino, northern Italy. *PLoS ONE* **6**, e14800.
- Ruiz-Moreno, D., Vargas, I. S., Olson, K. E. & Harrington, L. C. 2012 Modeling dynamic introduction of chikungunya virus in the United States. *PLoS Neglected Tropical Diseases* **6**, e1918.
- Savage, H., Ezike, V., Nwankwo, A., Spiegel, R. & Miller, B. 1992 First record of breeding populations of *Aedes albopictus* in continental Africa: implications for arboviral transmission. *Journal of the American Mosquito Control Association* **8**, 101-103.
- Schultz, G. 1993 Seasonal abundance of dengue vectors in Manila, Republic of the Philippines. *Southeast Asian Journal of Tropical Medicine and Public Health* **24**, 369-375.
- Simard, F., Nchoutpouen, E., Toto, J. C. & Fontenille, D. 2005 Geographic distribution and breeding site preference of *Aedes albopictus* and *Aedes aegypti* (Diptera: Culicidae) in Cameroon, Central Africa. *Journal of Medical Entomology* **42**, 726-731.
- Tsetsarkin, K. A., Chen, R., Sherman, M. B. & Weaver, S. C. 2011 Chikungunya virus: Evolution and genetic determinants of emergence. *Current Opinion in Virology* **1**, 310-317.

- Tsuda, Y., Takagi, M., Wang, S., Wang, Z. & Tang, L. 2001 Movement of *Aedes aegypti* (Diptera: Culicidae) released in a small isolated village on Hainan Island, China. *Journal of Medical Entomology* **38**, 93-98.
- Van Bortel, W., Dorleans, F., Rosine, J., Bateau, A., Rousseau, D., Matheus, S., Leparc-Goffart, I., Flusin, O., Prat, C. & Césaire, R. 2014 Chikungunya outbreak in the Caribbean region, December 2013 to March 2014, and the significance for the European Union. *Euro Surveillance* **19**, 20759.
- Van Kleef, E., Bambrick, H. & Hales, S. 2010 The geographic distribution of dengue fever and the potential influence of global climate change. *TropIKA.net*.
- Vasconcelos, P. F., Rosa, A. P., Rodrigues, S. G., Rosa, E. S., Monteiro, H. A., Cruz, A. C., Barros, V. L., Souza, M. R. & Rosa, J. F. 2001 Yellow fever in Para state, Amazon region of Brazil, 1998-1999: Entomologic and epidemiologic findings. *Emerging Infectious Diseases* **7**, 565-569.
- Wu, J.-Y., Lun, Z.-R., James, A. A. & Chen, X.-G. 2010 Dengue fever in mainland China. *American Journal of Tropical Medicine and Hygiene* **83**, 664-671.
- Zhong, D., Lo, E., Hu, R., Metzger, M. E., Cummings, R., Bonizzoni, M., Fujioka, K. K., Sorvillo, T. E., Klueh, S. & Healy, S. P. 2013 Genetic analysis of invasive *Aedes albopictus* populations in Los Angeles County, California and its potential public health impact. *PLoS ONE* **8**, e68586.

CHAPTER 2

A New Approach for Predicting Abundances of *Aedes mcintoshi*, a Primary Rift Valley Fever Virus Mosquito Vector

Rift Valley fever virus (RVFV) is a mosquito-borne zoonotic disease of great economic, livestock, and human health importance in Africa and the Arabian Peninsula [1,2]. Transmission of RVFV to humans generally involves direct contact with infected tissues or body fluids of animals, or bites of infected mosquitoes [3]. Human illness from RVFV often goes undetected, or results in flu-like symptoms, but a more severe form of the virus may present, resulting in ocular disease, meningoencephalitis or hemorrhagic fever, the latter with a case-fatality rate of 50% [4]. Symptoms of RVFV in livestock are less likely to be asymptomatic. High numbers of simultaneous, spontaneous abortions among ruminants (so-called “abortion storms”) and high mortality rates among young animals accompany epizootics [5]. Effects of epizootics on domestic livestock herds are devastating, and result in tremendous economic losses and food insecurity for communities whose livelihoods depend on livestock [6]. RVFV has also been detected in a wide variety of wild ruminants, from African buffalo to giraffes, but without the pronounced symptoms displayed in livestock [7].

Current RVFV forecasting models use persistence of above-average rainfall, positive Normalized Difference Vegetation Index anomalies, cloud coverage measurements, and El Niño Southern Oscillation information to guide early warning systems [8-13]. However, this approach treats mosquito population ecology as homogenous among vector species: predictions are made for the aggregate abundances of both primary and secondary mosquito vectors. Although this approach has seen some success in predicting RVFV risk to general locations and time periods

[8,9,14-17], model outputs do not identify the specific timing and locations of high abundances of primary vectors, and thus, potential virus emergence. Identifying drivers of high abundances of primary vector species will contribute to a better understanding of RVFV ecology, improve current prediction methods, and contribute to more efficient allocation of veterinary and public health resources.

Mosquito vectors in the genera *Aedes* and *Culex* are responsible for RVFV maintenance and amplification in natural environments [18]. Primary vector ecology is a key component in the RVFV disease system, warranting special consideration when inferring risk. Evidence suggests that adult *Aedes* mosquitoes in the subgenera *Neomelanicion* and *Aedimorphus* transmit the virus transovarially to their eggs, strongly implicating these species as primary disease vectors [3]. Adult *Aedes* mosquitoes can emerge already infected with the virus, before feeding on wild or domestic ungulates, thereby establishing low levels of virus activity within a geographic area. If suitable environmental conditions persist, *Culex* mosquitoes emerge, acting as secondary vectors that amplify the virus broadly across vulnerable populations [18]. In Kenya, where 11 epizootics occurred between 1951 and 2007 [19], the mosquito species *Ae. mcintoshi* (included within the subgenus *Neomelanicion*) has been implicated as an important RVFV primary vector [5,20], with evidence of transovarial transmission [21,22] and high RVFV prevalence in the species during the 2006-2007 epizootic in this region [3].

Early field studies provided fundamental information regarding *Ae. mcintoshi* population biology and ecology. *Ae. mcintoshi* prefer “dambo” habitats: shallow depressions in the landscape that become flooded following heavy rainfall [23,24]. Linthicum, Bailey, Davies and Kairo [25] flooded a dambo artificially for 18 continuous days and found that female emergence occurred at ~14 days, and blood feeding at ~18 days; mosquito life expectancy was <45 days,

female dispersal was low (~0.15 km), and direction of dispersal did not always correspond to surface winds. Logan, *et al.* conducted a sequence of artificial flooding events and found that ~90% of *Ae. mcintoshi* eggs hatched during the initial flooding of a dambo habitat, with fewer eggs hatching in subsequent flooding periods. Our study uses the biological knowledge gained from these initial field studies to predict *Ae. mcintoshi* adult abundances across broader landscapes and at different time periods, using georeferenced mosquito abundance data, remotely sensed environmental variables, and a predictive modelling approach.

Specifically, we investigated effects from land surface temperature, cumulative precipitation, compound topographic wetness index values, and percent clay in the soil on adult *Ae. mcintoshi* abundances, using zero-inflated negative binomial regression and a multimodel averaging approach. We hypothesized that wetness index values, land surface temperatures, and cumulative precipitation would all have positive effects on subsequent adult abundance. We hypothesized that because soils with more clay retain water better, higher percent clay would be associated with lower probabilities of so-called “structural” zeros in the zero-inflated model framework (see Methods). Having parameterized models with available *Ae. mcintoshi* survey data, we used the model to predict *Ae. mcintoshi* abundances retrospectively across Kenya and western Somalia over the months leading up to the 2006 – 2007 RVFV epizootic in the study region, and we compared those values to retrospective predictions for dates across a 14 year time period (2002 to 2016) to observe general differences between predicted values during epizootic and inter-epizootic time periods.

Methods

Rainfall varies greatly across Kenya, with more moisture in the West. Kenya experiences two rainy seasons referred to as the “short rains” (October to December) and the “long rains” (March through May), that coincide with movement of the intertropical convergence zone (ITCZ) [26]. More rainfall occurs from October to December during warm El Niño Southern Oscillation (ENSO) anomalies [27].

Mosquito abundance data for *Ae. mcintoshi* at 23 locations across Kenya were acquired by RS and JL as part of ongoing studies in the United States Army Medical Research Unit in Kenya (Figure 1). The data consisted of repeat sampling at each location for 3 to 4 days, usually twice yearly from 2007 to 2012, except for 2011. CO₂-baited CDC light traps were placed overnight at sample sites, and trapped mosquitoes were identified to species by Kenya Medical Research Institute entomology personnel.

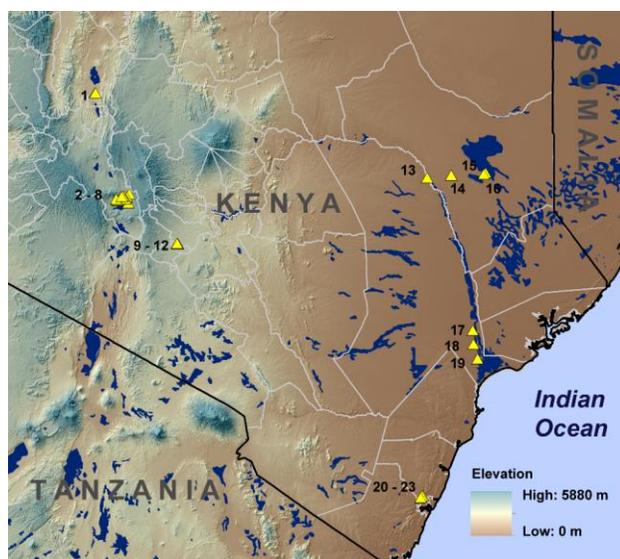


Figure 1. Sample site locations and elevation in study area.

A total of 158 sampling days were accumulated across all locations during the study period, with abundance values ranging from 0 to 4,426 individuals; dates on which sampling took place but the species was not recorded were assigned an abundance value of zero; such zero counts represented 41% of the total counts (Figure 2).

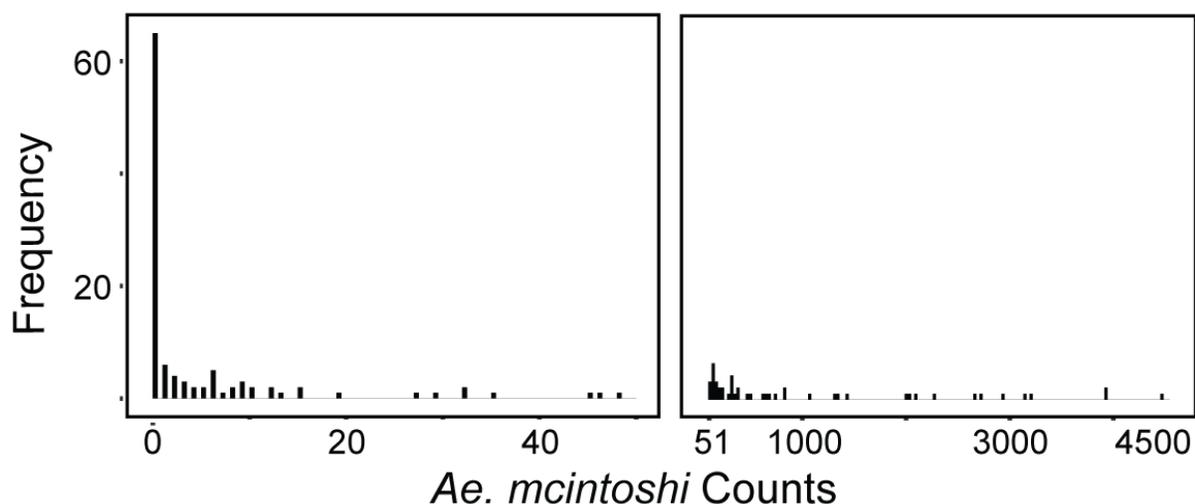


Figure 2. Count frequencies from 0 to 50 and 51 to 5,000.

We obtained Land Surface Temperature/Emissivity data from the Moderate Resolution Spectroradiometry Shuttle Mission (MODIS) Aqua sensor at an 8-day temporal resolution and 1 km spatial resolution through the Reverb ECHO NASA data portal (<http://reverb.echo.nasa.gov/reverb>). Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) data were obtained through the University of California at Santa Barbara data portal at a daily temporal resolution and a 5 km spatial resolution (<http://chg.ucsb.edu/index.html>). Compound topographic wetness index values [28] were derived from Shuttle Radar Topography Mission (SRTM) version 4.0 data at a 90 m spatial resolution accessed through the Consultative Group on International Agricultural Research Consortium for Spatial Information (<http://srtm.csi.cgiar.org/>). Minimum and maximum wetness index values within a 500 m radius

of each sampling location were calculated. Percentage of soil clay content was obtained at a 1 km spatial resolution from the Global Soil Dataset for Earth System Modeling Soils [29].

Environmental values were extracted from raster pixels at the geographic location of each sample site using the raster package in R [30]. Daily cumulative precipitation data corresponding to sampling dates were constructed in three time windows: sampling date to 14 days prior, 14 to 18 days before the sampling date, and 14 to 28 days prior. Land surface temperature data were acquired in 8-day mean composites. We identified the date of sampling and then subtracted 8, 16, and 24 days from the sampling date. The composite data with dates closest to this subtracted value were used in the analysis (Table 1).

Table 1. Environmental variables included in candidate models.

Variable	Spatial Resolution	Temporal Period
Minimum Wetness Index	90 m aggregated to 500 m	Static
Maximum Wetness Index	90 m aggregated to 500 m	Static
8 day Land Surface Temperature	1 km	8 day
16 day Land Surface Temperature	1 km	8 day
24 day Land Surface Temperature	1 km	8 day
Cumulative Precipitation 0 - 14	5 km	Daily
Cumulative Precipitation 14 - 18	5 km	Daily
Cumulative Precipitation 14 - 28	5 km	Daily
Percent Clay in the Soil	1 km	Static

Our count data included a greater number of zeros than may be expected under a Poisson or negative binomial distribution (Figure 2). Ignoring this phenomenon would have led to large biases in estimated parameters and their standard errors, and zeros can contribute to overdispersion [31,32], so we used a zero-inflated statistical modelling approach. Zeros may result from inevitable ecological factors or human error, including sampling error, observer error, or situations in which suitable habitat is present, but is not occupied due to essentially random

events [32]. Following standard zero-inflated modelling approaches, we refer to zeros recorded owing to inevitable circumstances as “structural” zeros, and to those recorded by chance due to sampling variation as “sampling” zeros [31]. The zero-inflated regression models that we used are mixture models that fit processes for both structural and sampling zeros [33]. Yau et al. [34] parameterized the zero-inflated negative binomial distribution as

$$P(Y = 0) = p + (1 - p)t^k$$

$$P(Y = y) = (1 - p) \binom{y + k - 1}{y} t^k (1 - t)^y, y = 1, 2, \dots$$

where $0 < p < 1$ is the probability of a structural zero, $t = \frac{k}{k + \mu}$, and μ is the mean of the negative binomial distribution used for the sampling part of the model. So the three parameters of the zero-inflated negative binomial used here are p , t and k (or, alternatively, p , k , and μ). Zero-inflated regression modelling is a standard technique used in ecology and entomology [35-37]. We evaluated the importance of relationships between mosquito abundances and our various environmental variables using an Akaike’s Information Criterion (AIC) and a Bayesian Information Criterion (BIC) approach [38-40]. Lower AIC or BIC scores indicate better-supported models. AIC model weights (denoted AIC_w) were also calculated; these values sum to 1 across all models, and indicate the weight of evidence supporting a model. Likewise, BIC weights (BIC_w) were computed. Best-performing models were considered the so-called 99% confidence set of models [39], i.e., for AIC, the models with highest AIC_w values and with a cumulative sum of these weights just exceeding 0.99. Importance of predictors was assessed using a standard sum-of-weights approach [39]: the sum of AIC_w (respectively, BIC_w) values for all models in the 99% confidence set that contained a given predictor was computed, this sum being an index of importance of that predictor. Model-averaged coefficients were also computed,

using a weighted average of the coefficients of the models in the 99% confidence set, with weights being the AIC_w (respectively, BIC_w) values. We fit zero-inflated negative binomial models to a random sample of 80% of our data set ($n = 127$), assigning *Ae. mcintoshi* abundance values as the response variable and the assembled environmental data as predictor variables (with different subsets of variables used in different models). All models were fitted using the `zeroinfl` function specifying a negative binomial distribution with a logit link in the `pscl` package in R 3.13 [41].

We calculated a Pearson's correlation matrix to assess the potential for multicollinearity between environmental variables, and found high correlation values within, but not between, sets of variables (S1 Table). Only one land surface temperature variable, precipitation variable, and wetness index variable were included in a candidate model at the same time. Percent clay could be included in the zero-inflated portion of a model, but not in the negative binomial portion.

We investigated model residuals from the lowest-AIC model for evidence of spatial autocorrelation using a spline correlogram in the `ncf` package in R [42,43], and found no evidence of it (S1 Fig). Although sampling dates were inconsistent across study site locations and time periods, we investigated the potential for residual temporal autocorrelation. Plots of model residuals against sampling dates did not identify obvious temporal patterns in the residuals (S2 Fig).

Estimated parameters from best performing models were used to predict mosquito abundances using a random sample of 20% of the abundance data ($n = 31$) withheld from the regression models, and a root mean square predictive error was calculated to evaluate accuracy. We used the `predict` function in the `pscl` package, with `type = "response"` to incorporate both the

structural zero portion and the sampling portion of the zero-inflated negative binomial model when predicting values [41,44].

For retrospective predictions, unsampled locations were generated across the study region at 10 km intervals, predictions were generated from each model in the 99% confidence set, and a weighted average of predictions was computed using the AIC_w values as weights. Predicted values were rasterized into a 10 x 10 km grid for visualization purposes. Values greater than 50,000 were considered extrapolative and excluded from visualizations, based on the fact that the greatest number of mosquitoes trapped in a single day across all species in our study locations was 47,694.

Results and Discussion

Model results indicated that relationships between environmental variables and *Ae. mcintoshi* abundances can be detected using remote sensing and that parameters derived from these models can be used to predict accurately the timing and location of primary vector emergence. Additionally, our results corroborated several factors known from local-scale studies to be important for *Ae. mcintoshi* development and population ecology. The 99% confidence set of models with respect to AIC consisted of 11 models (Table 2), and, with respect to BIC, of 19 models (S2 Table); these are the models best supported by data according to each information criterion, with the support for each model quantified by its AIC (or BIC) weight (AIC_w or BIC_w). The same variables tended to be included as predictors in the best-supported models according to both AIC and BIC, and model-averaged coefficients were similar (S3 Table), so the choice of information criterion did not substantially affect results. We henceforth use AIC because our objectives include prediction, for which AIC is considered more suitable [45,46].

Table 2. Model averaged coefficients and sums of AIC_w for each candidate variable using the 11 best models. AIC_w values close to one indicate strong support for the importance of a predictor.

	Negative Binomial Model Component									Zero-Inflation	
	Intercept	Wetness Index Max	Wetness Index Min	Precip 0 to 14 days	Precip 14 to 18 days	Precip 14 to 28 days	Land Surface Temp 8	Land Surface Temp 16	Land Surface Temp 24	Intercept	% Clay
Model avg. coefficient	-11.858	0.000	1.825	0.010	0.000	0.000	0.003	0.004	0.060	1.202	-0.051
Sum of AIC_w	0.994	0.000	0.994	0.870	0.004	0.042	0.061	0.072	0.819	0.994	0.994

The sum of AIC_w across the 11 best models including a given predictor indicated the importance of that predictor; these sums showed that minimum wetness index values within 500 m of a sampling site, cumulative precipitation from the sampling date to 14 days prior, land surface temperature values from 8-day mean composites 24 days prior to sampling, and percentage of clay in the soil (in the structural-zero portion of the model – see Methods) had the greatest importance for predicting abundance (Table 2).

The data strongly supported a positive relationship between minimum wetness index values within 500 m of a sampling site and mosquito abundance (Table 2). This result corroborated existing information regarding *Ae. mcintoshi* ecology: locations with high minimum wetness index values have low slopes and high flow accumulation, indicative of dambo habitats or landscapes likely to collect water during a precipitation event. We also found a positive effect on mosquito abundance of cumulative precipitation from the sampling date to 14 days prior, and no meaningful effect of cumulative precipitation 14 to 18 or 14 to 28 days prior to sampling. These results indicated that precipitation within a short time period prior to emergence has a greater impact on *Ae. mcintoshi* abundances, suggesting that forecasting may need to be performed at shorter time intervals. Model results indicated a positive effect on

mosquito abundances of 8-day mean land surface temperature composites at 8, 16, and 24 days prior to sampling, with 24 days prior having the greatest effect among these variables.

Temperature is known to play an important role in mosquito life stage development and may impact RVFV vector competency in some species [47,48], but the role of local temperatures in RVFV risk has been neglected in previous forecasting models [49]. Our results indicated that land surface temperature is indeed an important determinant of *Ae. mcintoshi* abundances in the study region. We found a negative relationship between percent clay in the soil and the probability of a structural zero within the zero-inflated portion of the model, indicating that areas with low clay content had a higher probability of a structural zero (Figure 3), presumably because standing-water pools are less likely to form over such soil types.

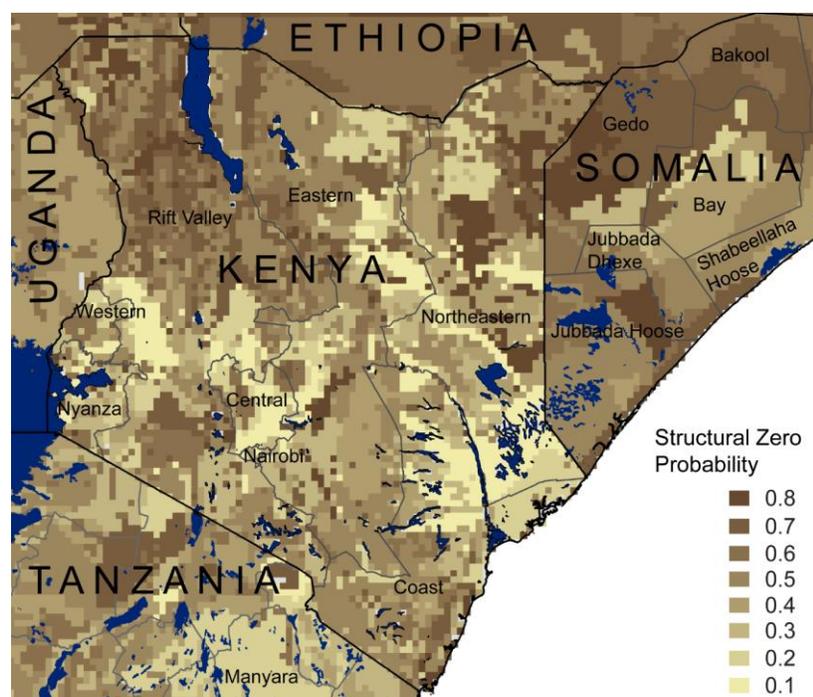


Figure 3. Probability of structural zeros.

Model-averaged out-of-sample predictions demonstrated the capacity of our models to predict elevated abundances (Figure 4). Our models' predictive accuracy was low for low

observed abundances: the relationship between observed and predicted values was weak on the left-most side of Figure 4. However, as the intended use of the models is to predict very high abundances, this point is of secondary importance. Prediction of high abundances was effective. Model performance statistics, including root mean square error values from out-of-sample predictions for the 11 best models are provided in more detail in S2 Table.

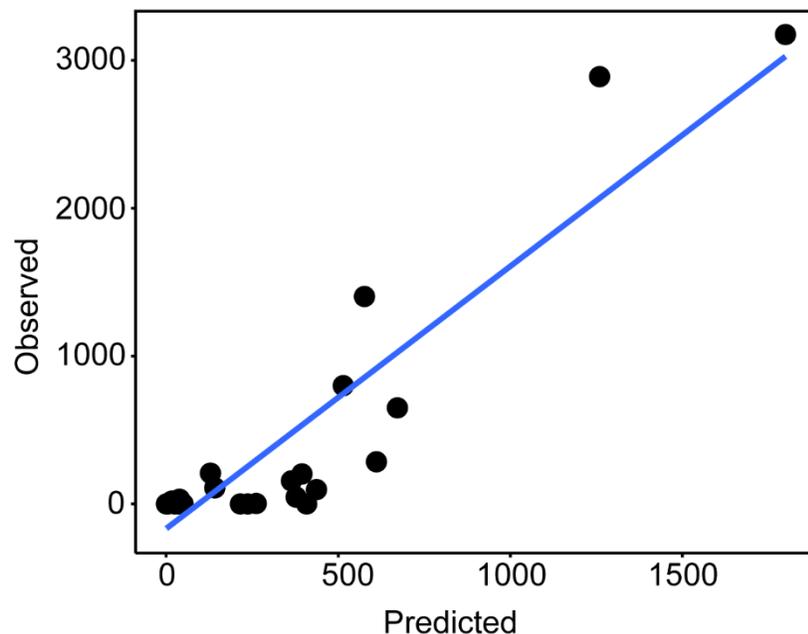


Fig 4. Observed vs. predicted values (n = 31). These are out-of-sample predictions (see Methods).

Retrospective model-averaged vector abundance predictions were produced at 8-day intervals from 30 September 2002 through 25 January 2016. Monitoring and human health agencies, such as the World Health Organization, reported [50-52] high abundances of RVFV cases at five times and locations during the 2006-2007 epizootic period (Figure 5, numbered circles). Retrospective vector abundance predictions were high pre-dating three of these (Figure

5, heat maps). Elevated abundances were predicted for 1 November 2006 in Bay, Bakool, Jubbuda Dhexe, and Gedo Provinces, Somalia (Figure 5, circle 1), ~7 weeks prior to reports of human cases [53]. Importantly, some ambiguity remains whether RVFV crossed into Somalia from Kenya, or whether foci developed independently within Somalia during the 2006-2007 epizootic. Our results supported the possibility of virus circulation within Somalia independent of Kenya, with high vector abundance predictions in isolated areas prior to elevated predicted abundances in Kenya (Figure 5). High abundances were predicted for 17 and 25 November in Garissa, Ijara, and Kitui Districts, Kenya; human cases of RVFV were first reported in mid-December in Garissa, with the potential index case presenting on 30 November (Figure 5, circle 2) [54]. Additionally, several of the Kenyan districts located within high predicted abundance areas for 25 November later reported RVFV activity for the first time: Kitui, Tharaka, Mwingi, Embu, Kirinyaga, Meru South, Meru Central, and Malindi District (Figure 5, circle 3) [51]. The World Health Organization recognized cases in these districts as a new RVFV focus during the outbreak, although little information is available regarding the exact date of onset of symptoms in humans or animals [52].

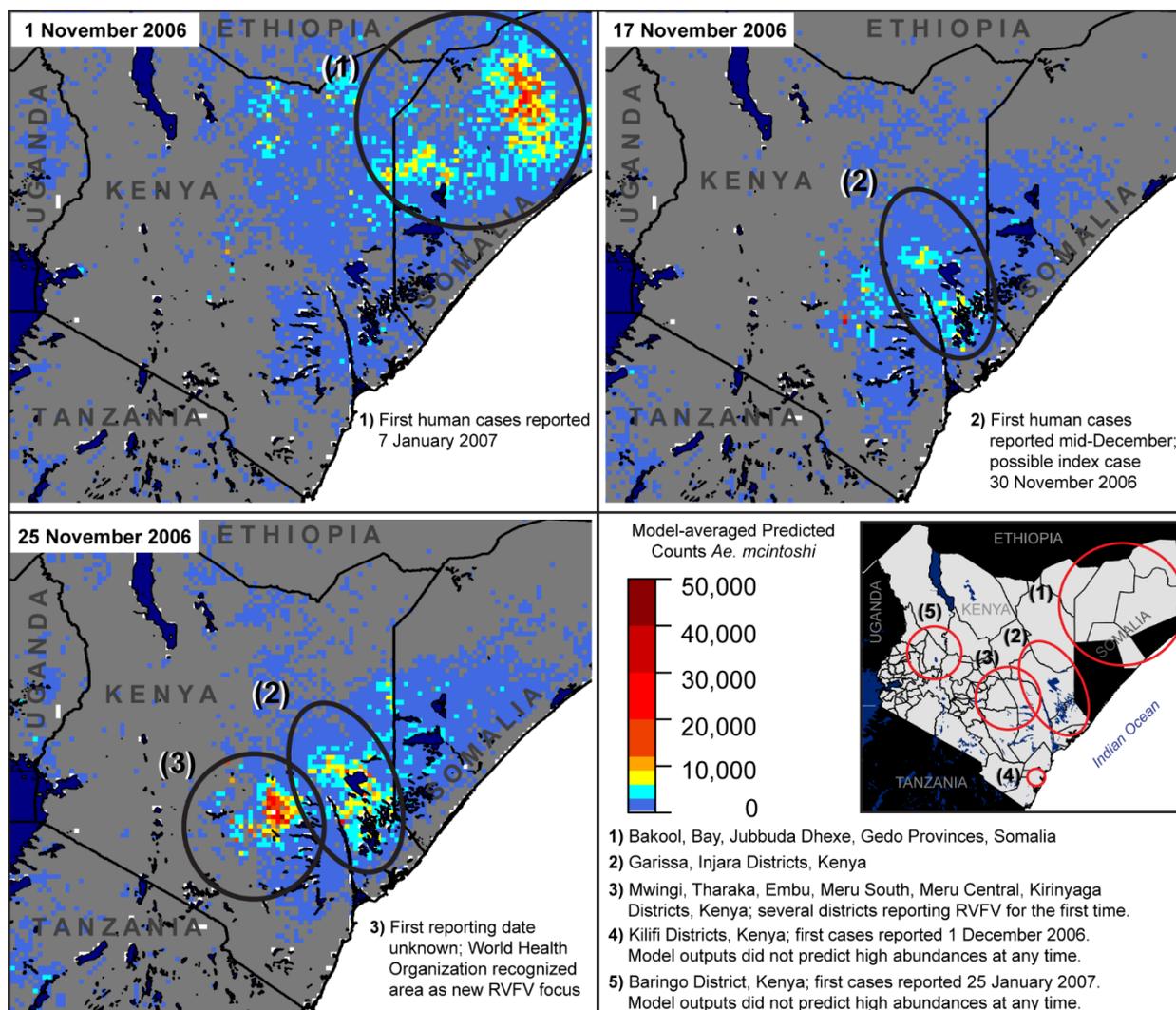


Fig 5. Predicted elevated abundances prior to and during the 2006-2007 epizootic (heat maps) and locations of independently documented RSVFV foci (numbered circles).

Human RSVFV cases were first reported retrospectively on 6 December 2006 in Kalifi District, near the eastern coast of Kenya (Figure 5, circle 4) [54]. Our models did not predict high *Ae. mcintoshi* abundances at any time in this area. Nguku et al. [54] found that illness in Kilifi District coincided with heavy rainfall, rather than emerging approximately one month after heavy rainfall, as was the case in other regions, and that movement of infected livestock from the outbreak area in Northeastern Province, Kenya into Kilifi District may have been the catalyst for

the outbreak. These observations suggest that primary vector emergence in Kilifi District may not have been responsible for virus circulation in the area, but further investigation is needed.

Illness in Baringo District was estimated to begin in late December, 2006, with the first human case reported on 25 January 2007 [51,54], but our model results did not predict elevated *Ae. mcintoshi* abundances in Baringo District at any time during the study period (Figure 5, circle 5). Although it is possible that our model predictions did not reflect the potential for *Ae. mcintoshi* abundance in this region, more likely the model was accurate and suitable habitat for this species was deficient in the area. Low numbers of *Ae. mcintoshi* were collected in Baringo District during the 2006-2007 epizootic, and follow-up sampling also indicated low densities of this species [3,24]. Our data set had only two records for *Ae. mcintoshi* in Baringo District with 2 and 6 females recorded, respectively. Alternative vectors or movement of infected mosquitoes or livestock into this region may have contributed to the outbreak.

Comparative predictions between 2002 and 2016 indicated lower predicted abundances across the study period (S1 Movie File), with the exception of one date in 2002 (Figure 6, circle 1), and a few small- and medium-scale (Figure 7) predictions of elevated abundances which apparently did not result in reported RVFV cases. High predicted abundances from 9 November 2002 were located to the northwest of the 2006-2007 focus in Isiolo and Laikipia Districts, Kenya (Figure 6, circle 1); virus activity was not reported there during this time period, nor was it reported in the regions shown in Figure 7. Even though conditions appeared to be suitable for high primary vector emergence, the possibility exists that ova deposited in these areas were not infected with RVFV transovarially and did not emerge with the capacity to transmit the disease; conditions were not suitable for large numbers of secondary vectors to amplify the virus; or susceptible livestock were not present in the area. In fact, some evidence indicates possible

livestock movement restrictions in this area and during this time period because of rinderpest virus detected in cattle on 23 October 2002 [55], but further investigation is needed before determining whether this factor could have impacted potential RVFV circulation.

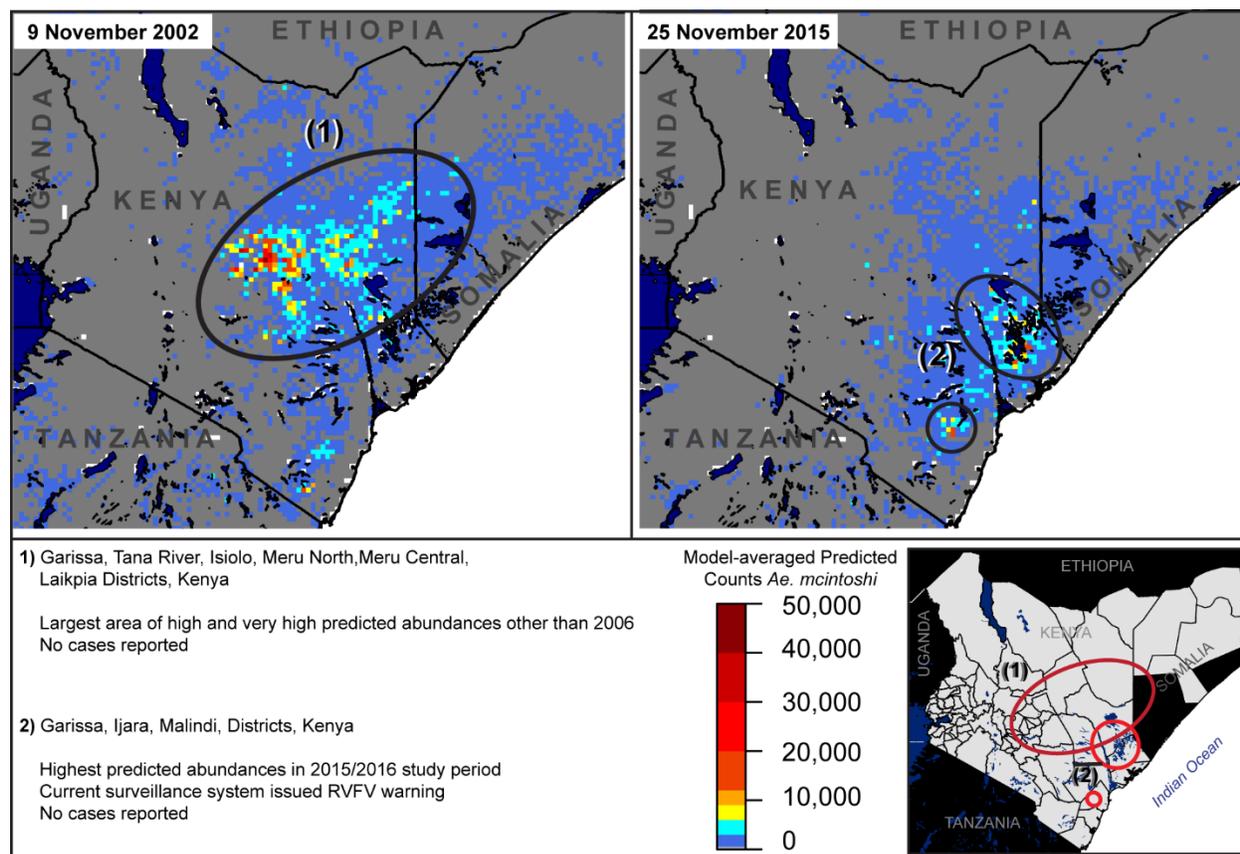


Figure 6. Predicted elevated abundances during interepizootic year 2002 (left panel), and lower predicted abundances during the 2015/2016 time period, when current RVFV surveillance systems issued warnings, but no virus was detected (right panel).

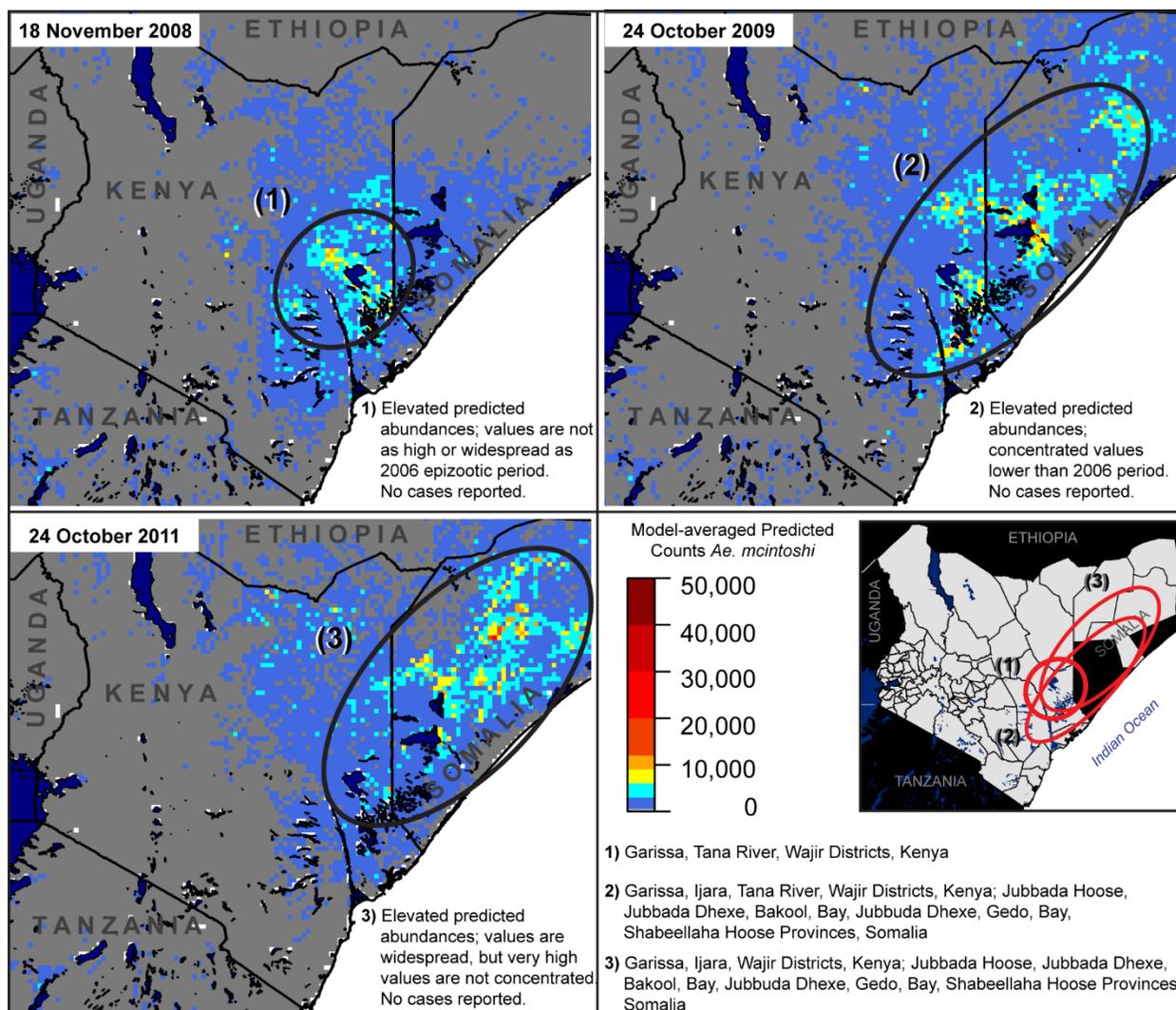


Figure 7. Medium-scale predictions of elevated abundances which apparently did not result in any reported RVFV cases, including 18 November 2008 (upper left), one week after animal health authorities issued an RVFV warning in the far, northeastern region of Kenya.

Importantly, abundances predicted by our models were low during two time periods in which RVFV warnings were issued in Kenya, but no virus was detected. Animal health authorities issued a warning on 13 November 2008 in the far northeastern region, near the borders of Ethiopia and Somalia because of persistent rainfall. Our model did not predict high vector abundances in this region prior to this date. The highest abundances predicted around this

date were only moderate, and were south of this region (Figure 7, circle 1). Several national and international agencies issued warnings for potential RVFV activity in the 2015-2016 period because of an El Niño event in the study area [56], but we did not predict high abundances during this time period (Figure 6, right panel) and no virus was detected. Model predictions for a total of 224 weeks from 1 September to 25 January between 2002 and 2016 are available in the supplementary material (S1 Movie File).

Although our model can predict high mosquito abundances accurately, limitations exist. Satellite measurements of environmental variables aggregate values across a landscape to a single spatial resolution, even though the potential exists for local variations or for larger errors in temperature measurements in semi-arid regions because of high reflectance of land surfaces and possible aerosols in the atmosphere [57]. Our data consisted of repeat sampling at locations across Kenya, but sampling was sparse in some areas. Our model predicts abundances for one primary RVFV vector, but additional primary vector species likely exist: several species were identified for further investigation during the 2006-2007 epizootic [3]. RVFV surveillance may benefit from applying our modelling framework to those species.

Our model is the first to use predicted vector abundances as a means of predicting potential RVFV emergence in this region, providing a powerful tool to help inform current RVFV monitoring systems. Additionally, our framework reveals new information about primary vector activity during inter-epizootic years that may contribute to a better understanding on RVFV ecology. The potential exists to extend our framework to multiple vectors and disease systems. Our approach also allows exploration of the effects of climate change on *Ae. mcintoshi*. We recommend incorporating our approach into the existing RVFV surveillance framework and extending our methods to additional vectors.

Acknowledgements

The views expressed are those of the authors and should not be construed to represent the position of the US Army Medical Research Directorate – Kenya, the Walter Reed Army Institute of Research, US Department of the Army, US Department of Defense or KEMRI.

References

1. Martin V, Chevalier V, Ceccato P, Anyamba A, De Simone L, et al. (2008) The impact of climate change on the epidemiology and control of Rift Valley fever. *Revue Scientifique Et Technique-Office International Des Epizooties* 27: 413-426.
2. Rich KM, Wanyoike F (2010) An assessment of the regional and national socio-economic impacts of the 2007 Rift Valley fever outbreak in Kenya. *Am J Trop Med Hyg* 83: 52-57.
3. Sang R, Kioko E, Lutomiah J, Warigia M, Ochieng C, et al. (2010) Rift Valley fever virus epidemic in Kenya, 2006/2007: The entomologic investigations. *Am J Trop Med Hyg* 83: 28-37.
4. World Health Organization (2010) Rift Valley fever. Fact sheet N°207.
5. Pepin M, Bouloy M, Bird BH, Kemp A, Paweska J (2010) Rift Valley fever virus (Bunyaviridae: Phlebovirus): An update on pathogenesis, molecular epidemiology, vectors, diagnostics and prevention. *Vet Res* 41:61.
6. Chengula AA, Mdegela RH, Kasanga CJ (2013) Socio-economic impact of Rift Valley fever to pastoralists and agro pastoralists in Arusha, Manyara and Morogoro regions in Tanzania. *Springerplus* 2.
7. Britch SC, Binopal YS, Ruder MG, Kariithi HM, Linthicum KJ, et al. (2013) Rift Valley fever risk map model and seroprevalence in selected wild ungulates and camels from Kenya. *PLoS One* 8: e66626.

8. Anyamba A, Linthicum KJ, Mahoney R, Tucker CJ, Kelley PW (2002) Mapping potential risk of Rift Valley fever outbreaks in African savannas using vegetation index time series data. *Photogramm Eng and Rem S* 68: 137-145.
9. Anyamba A, Linthicum KJ, Tucker CJ (2001) Climate-disease connections: Rift Valley Fever in Kenya. *Cad Saude Publica* 17 Suppl: 133-140.
10. Linthicum KJ, Anyamba A, Tucker CJ, Kelley PW, Myers MF, et al. (1999) Climate and satellite indicators to forecast Rift Valley fever epidemics in Kenya. *Science* 285: 397-400.
11. Linthicum KJ, Bailey CL, Davies FG, Tucker CJ (1987) Detection of Rift-Valley Fever Viral Activity in Kenya by Satellite Remote-Sensing Imagery. *Science* 235: 1656-1659.
12. Linthicum KJ, Bailey CL, Tucker CJ, Angleberger DR, Cannon T, et al. (1991) Towards Real-Time Prediction of Rift-Valley Fever Epidemics in Africa. *Prev Vet Med* 11: 325-334.
13. Linthicum KJ, Bailey CL, Tucker CJ, Mitchell KD, Logan TM, et al. (1990) Application of Polar-Orbiting, Meteorological Satellite Data to Detect Flooding of Rift-Valley Fever Virus Vector Mosquito Habitats in Kenya. *Med Veterinary Entomol* 4: 433-438.
14. Anyamba A, Chretien JP, Formenty PBH, Small J, Tucker CJ, et al. (2006) Rift Valley fever potential, Arabian Peninsula. *Emerg Infect Dis* 12: 518-520.
15. Anyamba A, Chretien JP, Small J, Tucker CJ, Formenty PB, et al. (2009) Prediction of a Rift Valley fever outbreak. *Proc Natl Acad Sci U S A* 106: 955-959.
16. Anyamba A, Chretien JP, Small J, Tucker CJ, Linthicum KJ (2006) Developing global climate anomalies suggest potential disease risks for 2006-2007. *Int J Health Geogr* 5: 60.

17. Anyamba A, Linthicum KJ, Small J, Britch SC, Pak E, et al. (2010) Prediction, assessment of the Rift Valley fever activity in East and Southern Africa 2006-2008 and possible vector control strategies. *Am J Trop Med Hyg* 83: 43-51.
18. Bicout DJ, Sabatier P (2004) Mapping Rift Valley Fever vectors and prevalence using rainfall variations. *Vector-Borne Zoonot* 4: 33-42.
19. Murithi RM, Munyua P, Ithondeka PM, Macharia JM, Hightower A, et al. (2011) Rift Valley fever in Kenya: history of epizootics and identification of vulnerable districts. *Epidemiol Infect* 139: 372-380.
20. Tchouassi DP, Bastos AD, Sole CL, Diallo M, Lutomiah J, et al. (2014) Population genetics of two key mosquito vectors of rift valley Fever virus reveals new insights into the changing disease outbreak patterns in kenya. *PLoS Negl Trop Dis* 8: e3364.
21. Linthicum KJ, Davies FG, Kairo A (1985) Rift Valley fever virus (family Bunyaviridae, genus *Phlebovirus*). Isolations from Diptera collected during an inter-epizootic period in Kenya. *J Hyg, Camb* 95: 197-209.
22. Rosmoser W, Oviedo M, Lerdthusne E, Patrican L, Turell M, et al. (2011) Rift Valley fever virus-infected mosquito ova and associated pathology: possible implications for endemic maintenance. *Research and Reports in Tropical Medicine* 2: 121 - 127.
23. Linthicum KJ, Davies FG, Bailey CL, Kairo A (1984) Mosquito species encountered in a flooded grassland dambo in Kenya. *Mosquito News* 44: 228-232.
24. Lutomiah J, Bast J, Clark J, Richardson J, Yalwala S, et al. (2013) Abundance, diversity, and distribution of mosquito vectors in selected ecological regions of Kenya: public health implications. *J Vector Ecol* 38: 134-142.

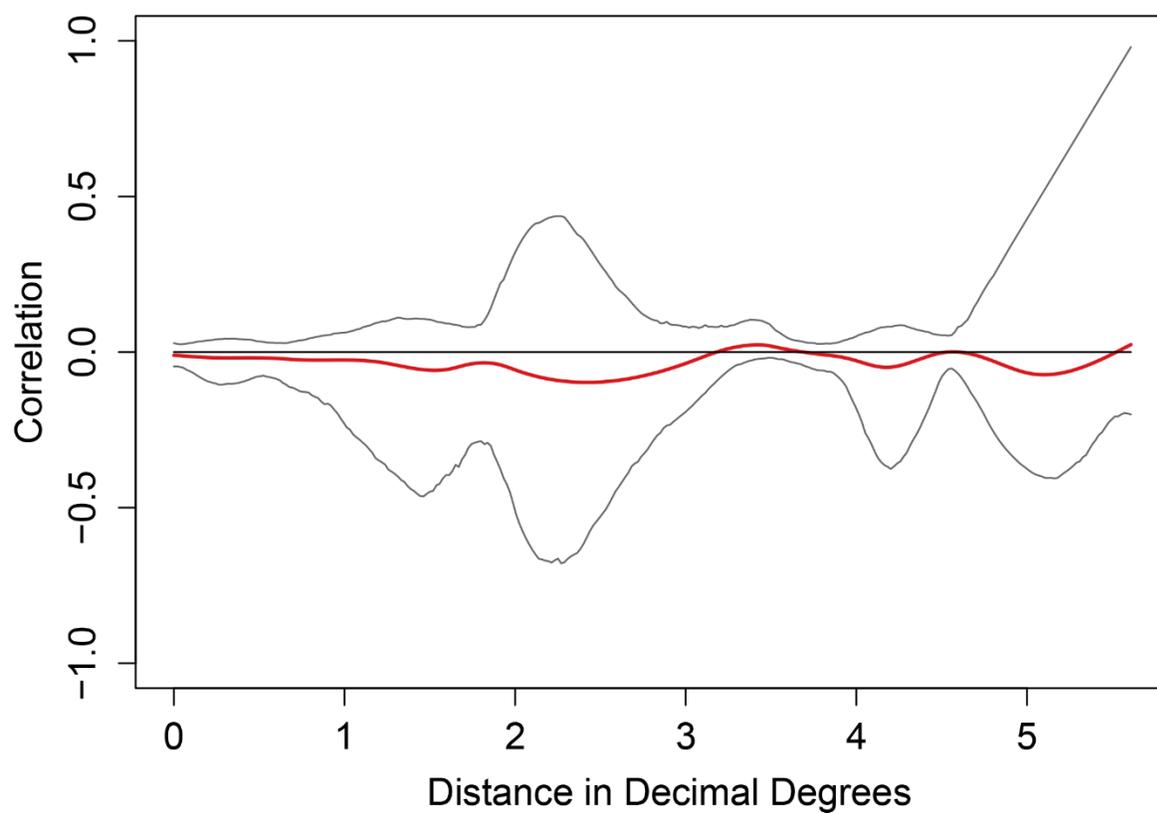
25. Linthicum KJ, Bailey CL, Davies FG, Kairo A (1985) Observations on the dispersal and survival of a population of *Aedes-Lineatopennis* (Ludlow) (Diptera, Culicidae) in Kenya. *B Entomol Res* 75: 661-670.
26. Camberlin P, Wairoto JG (1997) Intraseasonal wind anomalies related to wet and dry spells during the "long" and "short" rainy seasons in Kenya. *Theor Appl Climatol* 58: 57-69.
27. Camberlin P, Janicot S, Pocard I (2001) Seasonality and atmospheric dynamics of the teleconnection between African rainfall and tropical sea-surface temperature: Atlantic vs. ENSO. *Int J Climatol* 21: 973-1005.
28. Beven K, Kirkby M (1979) A Physically-based Variable Contributing Area Model of Basin Hydrology. *Hydrol Sci Bull* 24: 43-69.
29. Shanguan W, Dai Y, Duan Q, Liu B, Yuan H (2014) A Global Soil Data Set for Earth System Modeling. *J Adv Model Earth Syst* 6: 249-263.
30. Hijmans R (2015) raster: Geographic Data Analysis and Modeling. R package version 2.5-2 <https://CRAN.R-project.org/package=raster>.
31. Ridout M, Demetrio G, Hinde J (1998) Models for Count Data with Many Zeros. Proceedings of XIXth International Biometric Society Conference Cape Town, South Africa International Biometric Society. Washington, D.C., USA. pp. 179 - 192.
32. Zuur AF (2009) Mixed effects models and extensions in ecology with R. New York, NY: Springer. xxii, 574 p. p.
33. Lambert D (1992) Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics* 34: 1-14.
34. Yau KKW, Wang K, Lee AH (2003) Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal* 45: 437-452.

35. Desouhant E, Debouzie D, Menu F (1998) Oviposition pattern of phytophagous insects: on the importance of host population heterogeneity. *Oecologia* 114: 382-388.
36. Martin TG, Wintle BA, Rhodes JR, Kuhnert PM, Field SA, et al. (2005) Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecol Lett* 8: 1235-1246.
37. Yesilova A, Kaydan MB, Kaya Y (2010) Modeling Insect-Egg Data with Excess Zeros Using Zero-Inflated Regression Models. *Hacet J Math Stat* 39: 273-282.
38. Akaike H (1981) Likelihood of a model and information criteria. *J Econometrics* 16: 3-14.
39. Burnham KP, Anderson DR, Burnham KP (2002) Model selection and multimodel inference : a practical information-theoretic approach. New York: Springer. xxvi, 488 p. p.
40. Schwarz G (1978) Estimating Dimension of a Model. *Ann Stat* 6: 461-464.
41. Zeileis A, Kleiber C, Jackman S (2008) Regression models for count data in R. *J Stat Softw* 27: 1-25.
42. Bjornstad O (2015) ncf: spatial nonparametric covariance functions. R package version 11-6 <https://CRAN.R-project.org/package=ncf>.
43. Bjornstad ON, Falck W (2001) Nonparametric spatial covariance functions: Estimation and testing. *Environ Ecol Stat* 8: 53-70.
44. Zuur A, Saveliev A, Ieno E (2012) Zero Inflated Models and Generalized Linear Mixed Models with R. Newburgh, United Kingdom: Highland Statistics, Ltd.
45. Shmueli G (2010) To Explain or to Predict? *Statistical Science* 25: 289-310.
46. Sober E (2002) Instrumentalism, parsimony, and the Akaike framework. *Philos Sci* 69: S112-S123.

47. Turell MJ (1989) Effect of environmental temperature on the vector competence of *Aedes fowleri* for Rift Valley fever virus. *Res Virol* 140: 147-154.
48. Turell MJ, Rossi CA, Bailey CL (1985) Effect of Extrinsic Incubation-Temperature on the Ability of *Aedes-Taeniorhynchus* and *Culex-Pipiens* to Transmit Rift-Valley Fever Virus. *Am J Trop Med Hyg* 34: 1211-1218.
49. El Vilaly AE, Arora M, Butterworth MK, El Vilaly MAM, Jarnagin W, et al. (2013) Climate, environment and disease: The case of Rift Valley fever. *Prog Phys Geog* 37: 259-269.
50. Centers for Disease C, Prevention (2007) Rift Valley fever outbreak--Kenya, November 2006-January 2007. *MMWR Morb Mortal Wkly Rep* 56: 73-76.
51. Munyua P, Murithi RM, Wainwright S, Githinji J, Hightower A, et al. (2010) Rift Valley fever outbreak in livestock in Kenya, 2006-2007. *Am J Trop Med Hyg* 83: 58-64.
52. World Health Organization (2007) Outbreaks of Rift Valley fever in Kenya, Somalia and United Republic of Tanzania, December 2006-April 2007. *Wkly Epidemiol Rec* 82: 169-178.
53. Nderitu L, Lee JS, Omolo J, Omulo S, O'Guinn ML, et al. (2011) Sequential Rift Valley fever outbreaks in eastern Africa caused by multiple lineages of the virus. *J Infect Dis* 203: 655-665.
54. Nguku PM, Sharif SK, Mutonga D, Amwayi S, Omolo J, et al. (2010) An investigation of a major outbreak of Rift Valley fever in Kenya: 2006-2007. *Am J Trop Med Hyg* 83: 5-13.
55. ProMed-mail (2002) Rinderpest - Kenya: OIE, suspected. ProMed-mail 2002 01 Nov: 20021101.5682.
56. FAO, OIE, WHO (2015) Africa - El Niño and increased risk of Rift Valley fever -- Warning to countries. *EMPRES WATCH* 34.

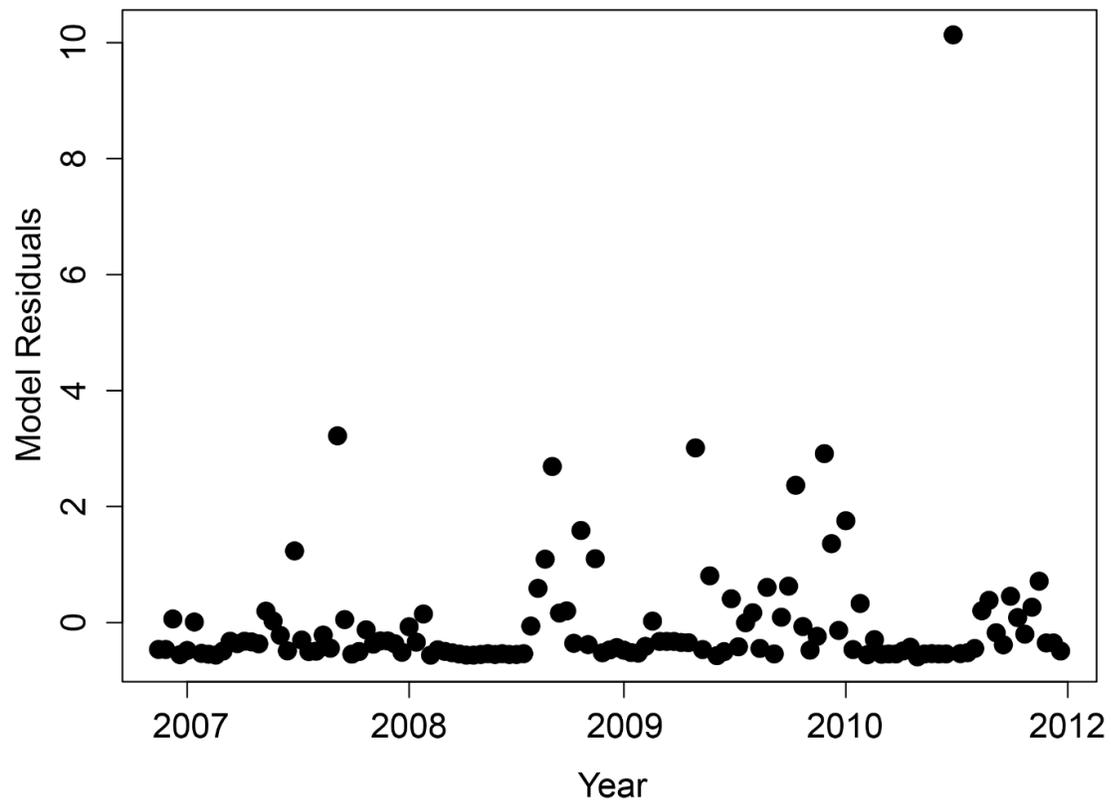
57. Li ZL, Tang BH, Wu H, Ren HZ, Yan GJ, et al. (2013) Satellite-derived land surface temperature: Current status and perspectives. *Remote Sens Environ* 131: 14-37.

Supporting Information



S1 Fig. Spline correlogram of distance between locations and model residuals for the lowest-AIC model.

S1 Table. Pearson's correlation matrix. Only one land surface temperature variable, precipitation variable, and wetness index variable were included in a candidate model at the same time; CLAY



S2 Fig. Lowest-AIC model residuals vs. sampling date. No obvious temporal autocorrelation in model residuals.

S2 Table. Candidate models: AIC and BIC rankings, scores, differences, and weights for all candidate models run in this analysis.

AIC Rank	BIC Rank	Intercept	Intercept_w	wi500max	wi500max_w	wi500min	wi500min_w	P_0_14	P_0_14_w	P_14_18	P_14_18_w
1	1	-11.71803	-8.435849868	0	0	1.76324	1.269362506	0.0119	0.008566851	0	0
2	3	-11.83287	-0.773463372	0	0	1.8639	0.121835056	0	0	0	0
3	6	-12.46989	-0.717812937	0	0	1.98572	0.114305379	0.01298	0.000747177	0	0
4	8	-12.83558	-0.686298776	0	0	2.04286	0.109228591	0.01041	0.000556607	0	0
5	4	-11.76567	-0.45847604	0	0	2.12129	0.08266088	0.01053	0.000410325	0	0
6	11	-11.89308	-0.397684422	0	0	1.90066	0.063554847	0	0	0	0
7	9	-12.81398	-0.125685006	0	0	2.09006	0.020500204	0	0	0	0
8	17	-12.90293	-0.064594966	0	0	2.13052	0.010665862	0	0	0	0
9	18	-12.91214	-0.047908599	0	0	2.09918	0.007788699	0	0	-0.00114	-4.23E-06
10	12	-12.03811	-0.041239623	0	0	2.24751	0.00769942	0	0	0	0
11	13	-12.29031	-0.037909556	0	0	2.13263	0.006578114	0	0	0	0
12	16	-11.85248	-0.023796754	0	0	2.1969	0.004410814	0	0	0.00045	9.03E-07
13	22	-12.43212	-0.024505082	0	0	2.17107	0.004279419	0	0	0	0
14	23	-12.37322	-0.014336898	0	0	2.13735	0.002476556	0	0	-0.00103	-1.19E-06
15	2	-13.28303	-0.009172587	0	0	1.99759	0.001379434	0.01138	7.86E-06	0	0
16	5	-14.26117	-0.002386272	0	0	2.18927	0.000366323	0.01016	1.70E-06	0	0
17	7	-13.77613	-0.001454998	0	0	2.1544	0.000227542	0.0125	1.32E-06	0	0
18	10	-13.65629	-0.000899539	0	0	2.09511	0.000138005	0	0	-0.00462	-3.04E-07
19	14	-14.255	-0.000334965	0	0	2.27295	5.34E-05	0	0	0	0
20	15	-14.34316	-0.00027045	0	0	2.2502	4.24E-05	0	0	-0.00173	-3.26E-08
21	24	-1.74639	-2.79E-05	0	0	0	0	0.01347	2.15E-07	0	0
22	25	1.5894	1.89E-05	-0.21648	-2.57E-06	0	0	0.01232	1.46E-07	0	0
23	19	-13.64186	-9.46E-05	0	0	2.3212	1.61E-05	0	0	0	0
24	20	-13.65357	-5.99E-05	0	0	2.29052	1.00E-05	0	0	-0.00134	-5.88E-09
25	21	-14.09733	-5.42E-05	0	0	2.15099	8.27E-06	0	0	0	0
26	26	3.57367	7.66E-06	-0.31914	-6.84E-07	0	0	0	0	0	0
27	29	3.50983	4.26E-06	-0.3379	-4.10E-07	0	0	0	0	-0.00571	-6.94E-09
28	30	3.81549	3.32E-06	-0.33005	-2.87E-07	0	0	0	0	0	0

29	27	-1.59325	-1.23E-06	0	0	0	0	0	0	0	0	-0.00481	-3.72E-09
30	28	-1.25657	-7.59E-07	0	0	0	0	0	0	0	0	0	0
31	35	5.03905	5.59E-09	-0.34912	-3.87E-10	0	0	0.01484	1.65E-11	0	0	0	0
32	32	-0.30405	-2.85E-10	0	0	0	0	0.01658	1.56E-11	0	0	0	0
33	38	5.06151	2.41E-09	-0.36334	-1.73E-10	0	0	0.00786	3.73E-12	0	0	0	0
34	33	6.25018	2.86E-09	-0.43693	-2.00E-10	0	0	0	0	0	0	0	0
35	34	-0.44105	-1.48E-10	0	0	0	0	0.00992	3.33E-12	0	0	0	0
36	42	5.88262	1.18E-09	-0.41718	-8.36E-11	0	0	0	0	0	0	0	0
37	43	6.26593	1.14E-09	-0.42624	-7.73E-11	0	0	0	0	0	0.00239	4.33E-13	0
38	39	-0.48525	-3.96E-11	0	0	0	0	0	0	0	0	0	0
39	40	-0.13617	-9.92E-12	0	0	0	0	0	0	0	0.00373	2.72E-13	0
40	41	8.25615	4.32E-10	-0.50617	-2.65E-11	0	0	0	0	0	0	0	0
41	45	8.44681	1.88E-10	-0.52166	-1.16E-11	0	0	0	0	0	0	0	0
42	47	8.23568	1.64E-10	-0.51328	-1.02E-11	0	0	0	0	0	-0.00163	-3.24E-14	0
43	31	-0.62761	-7.80E-12	-0.09418	-1.17E-12	0	0	0.0126	1.57E-13	0	0	0	0
44	44	0.51079	2.95E-12	0	0	0	0	0	0	0	0	0	0
45	46	0.53762	2.84E-12	0	0	0	0	0	0	0	-0.00034	-1.80E-15	0
46	48	10.64772	2.22E-11	-0.35296	-7.36E-13	0	0	0.01244	2.59E-14	0	0	0	0
47	36	1.36291	2.04E-12	-0.22016	-3.30E-13	0	0	0	0	0	-0.00598	-8.97E-15	0
48	37	1.50731	1.56E-12	-0.20341	-2.10E-13	0	0	0	0	0	0	0	0
49	49	12.71456	2.82E-12	-0.46311	-1.03E-13	0	0	0	0	0	0.00718	1.59E-15	0
50	53	13.4491	1.42E-12	-0.49925	-5.29E-14	0	0	0	0	0	0	0	0
51	50	-1.43455	-2.32E-15	-0.03968	-6.43E-17	0	0	0.00939	1.52E-17	0	0	0	0
52	51	1.41156	1.97E-15	-0.15189	-2.12E-16	0	0	0.01581	2.21E-17	0	0	0	0
53	52	0.21244	1.77E-16	-0.13579	-1.13E-16	0	0	0	0	0	0	0	0
54	54	1.85696	1.23E-15	-0.21244	-1.40E-16	0	0	0	0	0	0.00231	1.53E-18	0
55	55	5.35775	1.74E-16	-0.35728	-1.16E-17	0	0	0	0	0	0	0	0
56	56	5.1983	1.61E-16	-0.35299	-1.09E-17	0	0	0	0	0	-0.00175	-5.43E-20	0

P 14 28	P 14 28 w	LST_1K_8	LST_1K_8 w	LST_1K_16	LST_1K_16 w	LST_1K_24	LST_1K_24 w	log_ftheta	log_ftheta w
0	0	0	0	0	0	0.07269	0.05232978	-0.73747	-0.530907175
0	0	0	0	0	0	0.06999	0.004574943	-0.81191	-0.053071034
0	0	0.04768	0.002744637	0	0	0	0	-0.81695	-0.04702666
0	0	0	0	0.04798	0.002565417	0	0	-0.83934	-0.044878222
0	0	0	0	0	0	0	0	-0.86304	-0.033630313
-0.00247	-8.26E-05	0	0	0	0	0.0677	0.002263773	-0.79455	-0.026568404
0	0	0	0	0.05108	0.000501015	0	0	-0.89429	-0.008771579
-0.00258	-1.29E-05	0	0	0.04894	0.000245005	0	0	-0.87588	-0.004384852
0	0	0	0	0.05268	0.000195461	0	0	-0.89282	-0.003312678
-0.00339	-1.16E-05	0	0	0	0	0	0	-0.90143	-0.003088079
0	0	0.02753	8.49E-05	0	0	0	0	-0.90985	-0.002806439
0	0	0	0	0	0	0	0	-0.92888	-0.001864954
-0.00336	-6.62E-06	0.02861	5.64E-05	0	0	0	0	-0.88224	-0.001738992
0	0	0.02962	3.43E-05	0	0	0	0	-0.90761	-0.001051651
0	0	0	0	0	0	0.06368	4.40E-05	-0.92463	-0.000638503
0	0	0	0	0.0538	9.00E-06	0	0	-1.03391	-0.000173001
0	0	0.0453	4.78E-06	0	0	0	0	-1.00514	-0.00010616
0	0	0	0	0	0	0.07108	4.68E-06	-0.99777	-6.57E-05
-0.00249	-5.85E-08	0	0	0.05296	1.24E-06	0	0	-1.07509	-2.53E-05
0	0	0	0	0.0575	1.08E-06	0	0	-1.1015	-2.08E-05
0	0	0	0	0	0	0.18248	2.92E-06	-1.08414	-1.73E-05
0	0	0	0	0	0	0.1804	2.14E-06	-1.05564	-1.25E-05
-0.00332	-2.30E-08	0.02682	1.86E-07	0	0	0	0	-1.07406	-7.45E-06
0	0	0.02928	1.28E-07	0	0	0	0	-1.11332	-4.88E-06
-0.00106	-4.08E-09	0	0	0	0	0.06213	2.39E-07	-1.67657	-6.45E-06
0	0	0	0	0	0	0.18405	3.95E-07	-1.08847	-2.33E-06
0	0	0	0	0	0	0.19726	2.40E-07	-1.06743	-1.30E-06

-0.00155	-1.35E-09	0	0	0	0	0	0	0	0	0.18439	1.61E-07	-1.07853	-9.39E-07
0	0	0	0	0	0	0	0	0	0	0.19936	1.54E-07	-1.114	-8.62E-07
-0.00109	-6.58E-10	0	0	0	0	0	0	0	0	0.18867	1.14E-07	-1.12378	-6.79E-07
0	0	0.15057	1.67E-10	0	0	0	0	0	0	0	0	-1.29172	-1.43E-09
0	0	0.15327	1.44E-10	0	0	0	0	0	0	0	0	-1.34214	-1.26E-09
0	0	0	0	0	0.16326	7.76E-11	0	0	0	0	0	-1.36814	-6.50E-10
0	0	0	0	0	0.17088	7.81E-11	0	0	0	0	0	-1.37314	-6.28E-10
0	0	0	0	0	0.16438	5.52E-11	0	0	0	0	0	-1.42683	-4.79E-10
0.00228	4.57E-13	0	0	0	0.16885	3.38E-11	0	0	0	0	0	-1.37654	-2.76E-10
0	0	0	0	0	0.16406	2.97E-11	0	0	0	0	0	-1.37304	-2.49E-10
0.00303	2.47E-13	0	0	0	0.17342	1.42E-11	0	0	0	0	0	-1.45182	-1.19E-10
0	0	0	0	0	0.16607	1.21E-11	0	0	0	0	0	-1.44766	-1.05E-10
0	0	0.1471	7.70E-12	0	0	0	0	0	0	0	0	-1.36136	-7.13E-11
-0.00209	-4.65E-14	0.15178	3.38E-12	0	0	0	0	0	0	0	0	-1.34844	-3.00E-11
0	0	0.15192	3.02E-12	0	0	0	0	0	0	0	0	-1.35708	-2.70E-11
0	0	0	0	0	0	0	0	0	0.18954	0	2.36E-12	-1.20624	-1.50E-11
-0.00175	-1.01E-14	0.15773	9.12E-13	0	0	0	0	0	0	0	0	-1.43441	-8.29E-12
0	0	0.15425	8.15E-13	0	0	0	0	0	0	0	0	-1.44382	-7.63E-12
0	0	0	0	0	0	0	0	0	0	0	0	-1.50126	-3.13E-12
0	0	0	0	0	0	0	0	0	0.20675	0	3.10E-13	-1.22189	-1.83E-12
-0.00096	-9.91E-16	0	0	0	0	0	0	0	0.19341	0	2.00E-13	-1.24354	-1.28E-12
0	0	0	0	0	0	0	0	0	0	0	0	-1.54144	-3.42E-13
0.00018	1.91E-17	0	0	0	0	0	0	0	0	0	0	-1.55036	-1.64E-13
0	0	0	0	0	0.20109	3.26E-16	0	0	0	0	0	-2.05654	-3.33E-15
0	0	0.16611	2.32E-16	0	0	0	0	0	0	0	0	-1.57735	-2.20E-15
0.00398	3.32E-18	0	0	0	0.19952	1.67E-16	0	0	0	0	0	-2.0676	-1.73E-15
0	0	0	0	0	0.19345	1.28E-16	0	0	0	0	0	-1.84664	-1.22E-15
-0.00162	-5.25E-20	0.16622	5.38E-18	0	0	0	0	0	0	0	0	-1.62496	-5.26E-17
0	0	0.16739	5.19E-18	0	0	0	0	0	0	0	0	-1.64023	-5.09E-17

CLAY Z	CLAY Z_w	log_lik	AIC	AIC_diff	AICw	BIC	BIC_diff	BICw	RMSE
-0.05162	-0.037161415	-506.1313	1026.26259	-0.00041	0.71990342	1046.17191	-9.00E-05	0.367841895	412.1902761
-0.05044	-0.003297044	-509.53042	1031.06083	4.79783	0.065365661	1048.12596	1.95396	0.138466471	599.8820496
-0.05077	-0.002922509	-508.65752	1031.31504	5.05204	0.057563694	1051.22435	5.05235	0.029412896	433.9448457
-0.04985	-0.002665403	-508.73132	1031.46264	5.19964	0.053468466	1051.37195	5.19995	0.027320387	513.6466175
-0.05084	-0.001981096	-510.04769	1032.09538	5.83238	0.03896727	1049.1605	2.9885	0.082546206	538.7416379
-0.05004	-0.001673253	-509.20071	1032.40142	6.13842	0.033438304	1052.31073	6.13873	0.017085723	603.5937495
-0.04891	-0.00047973	-511.42717	1034.85434	8.59134	0.009808428	1051.91946	5.74746	0.020777657	668.7166994
-0.04845	-0.000242552	-511.09973	1036.19946	9.93646	0.005006225	1056.10877	9.93677	0.002557994	673.0760558
-0.04875	-0.00018088	-511.39929	1036.79857	10.53557	0.003710353	1056.70789	10.53589	0.001895842	674.4307857
-0.04902	-0.000167931	-512.47909	1036.95818	10.69518	0.003425756	1054.0233	7.8513	0.00725694	668.6987929
-0.04952	-0.000152745	-512.58402	1037.16804	10.90504	0.003084508	1054.23316	8.06116	0.006534058	653.2198959
-0.04976	-9.99E-05	-513.0134	1038.0268	11.7638	0.002007745	1055.09192	8.91992	0.0042531	662.8624697
-0.04897	-9.65E-05	-512.03182	1038.06363	11.80063	0.00197111	1057.97295	11.80095	0.001007159	665.0102822
-0.04943	-5.73E-05	-512.56311	1039.12622	12.86322	0.001158704	1059.03553	12.86353	0.000592054	657.5470808
0	0	-514.08068	1040.16136	13.89836	0.000690549	1047.53811	1.36611	0.185778148	464.5719414
0	0	-515.49822	1042.99644	16.73344	0.000167327	1050.37319	4.20119	0.045015785	533.78235
0	0	-515.95834	1043.91669	17.65369	0.000105617	1051.29343	5.12143	0.028414319	473.6145498
0	0	-516.43049	1044.86097	18.59797	6.59E-05	1052.23773	6.06573	0.017720873	624.7641977
0	0	-517.46125	1046.9225	20.6595	2.35E-05	1054.29925	8.12725	0.006321669	672.9455335
0	0	-517.68135	1047.36271	21.09971	1.89E-05	1054.73945	8.56745	0.005072751	678.5527402
-0.06977	-1.11E-06	-517.84698	1047.69395	21.43095	1.60E-05	1064.73908	18.58708	3.38E-05	374.6723502
-0.0705	-8.38E-07	-517.1427	1048.28541	22.02241	1.19E-05	1068.19471	22.02271	6.07E-06	404.1728726
0	0	-518.68197	1049.36395	23.10095	6.93E-06	1056.74069	10.56869	0.001865004	661.9967031
0	0	-519.13945	1050.27889	24.01589	4.39E-06	1057.65565	11.48365	0.001180317	660.0983331
0	0	-519.27151	1050.54303	24.28003	3.84E-06	1057.91977	11.74777	0.001034298	609.571648
-0.06865	-1.47E-07	-519.85568	1051.71136	25.44836	2.14E-06	1068.77648	22.60448	4.54E-06	587.005376
-0.06791	-8.25E-08	-519.42374	1052.84749	26.58449	1.21E-06	1072.75679	26.58479	6.21E-07	605.0370151
-0.0683	-5.95E-08	-519.75661	1053.51323	27.25023	8.71E-07	1073.42253	27.25053	4.45E-07	590.2350541

-0.06646	-5.14E-08	-520.8751	1053.7502	27.4872	7.73E-07	1070.81532	24.64332	1.64E-06	578.6064075
-0.0669	-4.04E-08	-521.12255	1054.2451	27.9821	6.04E-07	1071.31022	25.13822	1.28E-06	570.1593673
-0.07569	-8.40E-11	-526.42215	1066.8443	40.5813	1.11E-09	1086.75361	40.58161	5.67E-10	480.6855217
-0.07561	-7.09E-11	-527.58977	1067.17954	40.91654	9.38E-10	1084.24466	38.07266	1.99E-09	416.9705383
-0.08032	-3.82E-11	-527.27003	1068.54007	42.27707	4.75E-10	1088.44937	42.27737	2.43E-10	687.9661028
-0.0781	-3.57E-11	-528.30904	1068.61807	42.35507	4.57E-10	1085.6832	39.5112	9.68E-10	790.327177
-0.08108	-2.72E-11	-528.61672	1069.23343	42.97043	3.36E-10	1086.29856	40.12656	7.12E-10	634.0325079
-0.07863	-1.58E-11	-528.13334	1070.26669	44.00369	2.00E-10	1090.17599	44.00399	1.02E-10	772.1069027
-0.0783	-1.42E-11	-528.23361	1070.46722	44.20422	1.81E-10	1090.37653	44.20453	9.26E-11	765.4593599
-0.07905	-6.46E-12	-530.03107	1072.06214	45.79914	8.17E-11	1089.12726	42.95526	1.73E-10	756.1167382
-0.07861	-5.73E-12	-530.14495	1072.2899	46.0269	7.29E-11	1089.35502	43.18302	1.54E-10	736.8199862
-0.07455	-3.90E-12	-530.47534	1072.95068	46.68768	5.24E-11	1090.0158	43.8438	1.11E-10	736.1117201
-0.07377	-1.64E-12	-530.33067	1074.66134	48.39834	2.23E-11	1094.57065	48.39865	1.14E-11	793.6033924
-0.07421	-1.47E-12	-530.44481	1074.88963	48.62663	1.99E-11	1094.79893	48.62693	1.01E-11	766.3367856
0	0	-531.91361	1075.82722	49.56422	1.24E-11	1083.20397	37.03197	3.34E-09	497.6587416
-0.07337	-4.24E-13	-532.67916	1077.35832	51.09532	5.78E-12	1094.42344	48.25144	1.22E-11	697.4808929
-0.0741	-3.92E-13	-532.76885	1077.53771	51.27471	5.28E-12	1094.60282	48.43082	1.12E-11	679.3333001
-0.08765	-1.83E-13	-533.69859	1079.39717	53.13417	2.09E-12	1096.4623	50.2903	4.42E-12	669.522061
0	0	-534.02843	1080.05685	53.79385	1.50E-12	1087.43361	41.26161	4.03E-10	667.5210125
0	0	-534.40139	1080.80279	54.53979	1.03E-12	1088.17953	42.00753	2.78E-10	650.3416983
-0.08619	-1.91E-14	-535.94029	1083.88058	57.61758	2.22E-13	1100.9457	54.7737	4.70E-13	723.3653779
-0.08599	-9.11E-15	-536.67871	1085.35742	59.09442	1.06E-13	1102.42254	56.25054	2.24E-13	745.3779342
0	0	-540.85917	1093.71834	67.45534	1.62E-15	1101.09509	54.92309	4.36E-13	755.1790595
0	0	-541.00803	1094.01606	67.75306	1.40E-15	1101.39281	55.22081	3.75E-13	579.659515
0	0	-541.52131	1095.04262	68.77962	8.35E-16	1102.41937	56.24737	2.25E-13	828.4315581
0	0	-541.75586	1095.51171	69.24871	6.61E-16	1102.88847	56.71647	1.78E-13	855.5369701
0	0	-544.77137	1101.54273	75.27973	3.24E-17	1108.91949	62.74749	8.71E-15	756.0212713
0	0	-544.81498	1101.62997	75.36697	3.10E-17	1109.00671	62.83471	8.34E-15	749.301561

S3 Table. Model averaged coefficient values for AIC and BIC models corresponding to a cumulative weighted value of 0.994 (11 models for AIC and 19 models for BIC).

Variable	Model Averaged Coefficient (AIC 11 models)	Model Averaged Coefficient (BIC 19 models)	Sum of AIC _w	Sum of BIC _w
Intercept	-11.858	-12.363	0.994	0.994
Wetness Index Max	0.000	0.000	0.000	0.000
Wetness Index Min	1.825	1.926	0.994	0.994
Precip 0 to 14 days	0.010	0.009	0.870	0.766
Precip 14 to 18 days	0.000	0.000	0.004	0.029
Precip 14 to 28 days	0.000	0.000	0.042	0.027
Land Surface Temp 8	0.003	0.003	0.061	0.160
Land Surface Temp 16	0.004	0.006	0.072	0.109
Land Surface Temp 24	0.060	0.051	0.819	0.727
log theta				
Intercept Zero-inflation	1.202	0.532	0.994	0.994
% Clay	-0.051	-0.036	0.994	0.706

S Movie File. Animated plots between 1 September and 25 January from 2002 to 2016.

CHAPTER 3

Landscape genetics of *Aedes mcintoshi*, an important vector of Rift Valley fever virus in eastern Kenya

Rift Valley fever virus (RVFV) is a medically important, mosquito-borne, zoonotic pathogen that affects humans, wild ungulates, and domestic livestock in Africa and the Arabian Peninsula (Martin et al. 2008, Rich and Wanyoike 2010). RVFV transmission occurs through mosquito bites or close contact with infected body fluids (Sang et al. 2010). Symptoms often go unnoticed in humans, although, in a small percentage of cases, the virus can lead to fatal hemorrhagic fever (World Health Organization 2010). However, RVFV severely impacts domestic livestock, resulting in spontaneous abortions referred to as “abortion storms” and high mortality rates in newborn ruminants (Pepin et al. 2010). These factors, combined with bans on the export of livestock, result in major economic losses and food insecurity in affected regions (Chengula et al. 2013).

RVFV exhibits inter-epizootic phases, when little to no virus activity is detected, and epizootic phases, which are defined by widespread occurrence detected in domestic livestock (Pepin et al. 2010). Over the past 80 years, epizootics have occurred in more than 30 countries. East Africa, and specifically Kenya, appear particularly vulnerable, with 11 epizootics reported in Kenya between 1951 and 2007 (Murithi et al. 2011). Strong associations exist between epizootics in East Africa and extreme weather events, such as El Niño (Anyamba et al. 2001, Anyamba et al. 2006), which promote massive mosquito emergences through above-average precipitation during the October-to-December “short rains” season (Davies et al. 1985, Camberlin et al. 2001).

The mosquito species *Aedes mcintoshi*, subgenus *Neomelaniconion*, is an important primary vector for RVFV in Kenya (Pepin et al. 2010, Tchouassi et al. 2014). Evidence for transovarial transmission in this species suggests that adults can emerge infected with the virus, which is then transmitted locally to wild ungulates and/or nearby livestock (Linthicum et al. 1985b, Sang et al. 2010, Rosmoser et al. 2011). Emergence of *Ae. mcintoshi* adults appears to be associated with flooding of “dambos”: low-lying ephemeral pool habitats that become inundated during precipitation events (Linthicum et al. 1984, Linthicum et al. 1985a, Lutomiah et al. 2013). Linkages among environmental variables, mosquito abundance, and RVFV foci within Kenya have been assessed previously Campbell et al. (in review). However, understanding the genetic structure of *Ae. mcintoshi* could provide further information regarding the potential for RVFV to spread across East African landscapes.

Here, we investigate whether genetic structure exists in Clade IV (described by Tchouassi et al., 2014) of *Ae. mcintoshi*, using gene sequence data for cytochrome oxidase subunit 1 (CO1) and nuclear 18S ribosomal internal transcribed spacer. We focus on Clade IV, as this clade is distributed across the eastern portion of Kenya (Tchouassi et al., 2014), where RVFV is considered to be endemic, and where the last major epizootic in Kenya began in 2006 (Nguku et al. 2010). Retrospective predictions of localized changes in mosquito abundances (Campbell et al. in review), coupled with the low dispersal distances observed in *Ae. mcintoshi* (Linthicum et al. 1985a), suggest strongly the potential for heterogeneity in connectivity between mosquito habitats across the landscape within the Clade IV range (Figure 1). Here, we investigate patterns of genetic diversity and differentiation in the Clade IV area, and potential associations with these patterns and climate and topographic variables. Given that the study area is an RVFV-endemic region, understanding connectivity between *Ae. mcintoshi* subpopulations can provide new

information regarding the genetic diversity and dispersal of this species, which could help to inform vector control and vaccination programs.

Methods

Cytochrome oxidase subunit 1 (CO1) sequences for 95 *Ae. mcintoshi* individuals from 9 locations, and nuclear 18S ribosomal internal transcribed spacer (ITS) data for 39 individuals from 8 locations in eastern Kenya were downloaded from GenBank (data were generated by Tchouassi et al. 2014: see Appendix for summary of all sequences analyzed). Numbers of individuals from each site ranged from 9 to 13 for CO1, and 3 to 6 for ITS. The nine CO1 sample locations (Figure 1) correspond to collection sites within the Clade IV geographic area, outlined by Tchouassi et al. (2014).

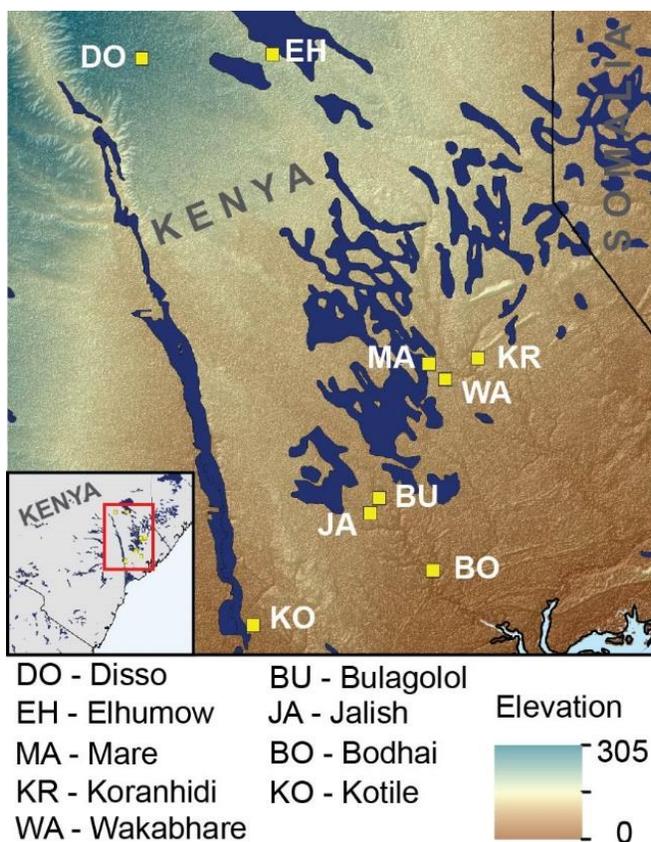


Fig. 1. *Aedes mcintoshi* sample site locations in eastern Kenya.

We aligned sequences for each locus using the program `MUSCLE`, accessed online (Edgar 2004), and inspected alignments visually using the software program Geneious version 9.1.5 (Kearse et al. 2012) (<http://www.geneious.com>). We generated a maximum likelihood tree through MEGA v6.0.6 (Tamura et al. 2013), with $n = 500$ bootstrap replicates, a GTR+ Γ +I nucleotide substitution model (four gamma categories), and nearest-neighbor-interchange, with a very strong branch swap filter, as the ML heuristic. We used an NJ/BioNJ tree as the initial ML tree for the analysis. Nodes with bootstrap support $>70\%$ were considered well supported, following Hillis and Bull (1993)). We generated a median-joining haplotype network (Bandelt et al. 1999), using PopART (Leigh and Bryant 2015). For the combined dataset, we calculated Tajima's D (Tajima 1989), number of haplotypes, average nucleotide diversity (π), and number of segregating sites in `dnaSP v5` (Librado and Rozas 2009). For individual sample locations, we calculated Tajima's D in `Arlequin v3.5.2.2`, using 10,000 simulated samples (Excoffier and Lischer 2010); Watterson's estimate of theta ($\hat{\theta}_w$) in the `pegas` package (Paradis 2010) within R (R Core Team 2015), and nucleotide diversity using `dnaSP v5`. `dnaSP` was also used to calculate the γ_{st} between individual sample locations.

We acquired the 1900-2014 `CenTrends` October to February seasonal precipitation anomaly dataset through the Climate Hazards Group for each sample location at a 10 km spatial resolution (Funk et al. 2015a). A finer spatial resolution dataset of mean monthly precipitation values for a 25-year period (1980 to 2005) was obtained from the Climate Hazards Group `clim` data for October, November, and December (5 km spatial resolution). This period corresponds to the "short rains" season, associated with massive *Ae. mcintoshi* emergences (Sang et al. 2010, Funk et al. 2015b).

Compound topographic wetness index values (Beven and Kirkby 1979) were derived from Shuttle Radar Topology Mission (SRTM) version 4.0 data at a 90 m spatial resolution, accessed through the Consultative Group on International Agricultural Research Consortium for Spatial Information (<http://srtm.csi.cgiar.org/>). Percent clay in the soil at each sample location was acquired from the Global Soil Data Set for Earth System Modeling Soils at a 1 km spatial resolution (Shangguan et al. 2014). All environmental values were extracted for sample site locations using ArcGIS 10.1.

We generated pairwise geographic (km) and genetic distance matrices (γ_{st}) to test for isolation by distance using a Mantel test with 100,000 permutations in the *ade4* package in R (Dray and Dufour 2007). We used a multivariate statistical approach to explore associations between genetic diversity ($\hat{\theta}_w$) and differentiation (γ_{st}) and environmental variables. We explored significance of global and local patterns between genetic variance and patterns of spatial autocorrelation using spatial principal components analysis (sPCA) in the *ade4* package in R (Jombart 2008, Jombart et al. 2008, Kierepka and Latch 2016).

We performed partial redundancy analyses (RDA) with $\hat{\theta}_w$ values as the response variable, and distanced-based partial redundancy analyses (dbRDA) with the pairwise γ_{st} distance matrix as the response variable. All models were conditioned on location, and were run using the *rda* and *capscale* functions in the R-package *vegan* (Oksanen et al. 2016). γ_{st} and $\hat{\theta}_w$ values served as response variables, environmental variables served as constrained variables, and UTM Zone 37N projected coordinates provided location information for the conditional portion of each model. Because of the relatively small number of sample sites, all models were run with only one constrained environmental variable to maintain greater statistical power. Model outputs

included total explained variance, partitioned into contributions from constrained environmental variables, conditional location variables, and unconstrained variables. A permutation test with 9999 iterations determined the significance of each environmental variable, and ordination plots indicated the direction of significant associations. In addition to dbRDA and RDA analyses, we performed a Procrustes test on distance matrices, comparing similarities between the γ_{st} distance matrix and the environmental distance matrices (Peres-Neto and Jackson 2001). We expected to find positive associations between genetic diversity and precipitation, soil clay content, and wetness index values because of their contributions to suitable habitats that promote higher *Ae. mcintoshi* abundances, which could lead to greater diversity. We expected to find a negative association between genetic differentiation and precipitation, soil clay content, and wetness index values because of the potential for greater connectivity between locations under environmentally suitable conditions.

Results

The 95 CO1 sequences had an aligned length of 1448 base pairs. Each sequence was unique, with 181 segregating sites present over the 95 haplotypes. Large gaps in the ITS aligned data, low Bayesian posterior probabilities supporting monophyly of Clade IV individuals (Tchouassi et al. (2014), and highly variable γ_{st} values (Supplementary Table 1), led us to exclude the ITS data from further analysis.

Our maximum likelihood tree indicated that a number of individuals from the sites Kotile and Disso fell in Clade II rather than Clade IV (Figure 2). This phylogenetic heterogeneity was also reflected by higher nucleotide diversity, $\hat{\theta}_w$, and γ_{st} values in comparison to the remaining Clade IV sample locations (Figure 3). Therefore, we excluded Kotile and Disso from subsequent

statistical analyses; results including these locations are provided in Supplementary Tables 2 and

3.

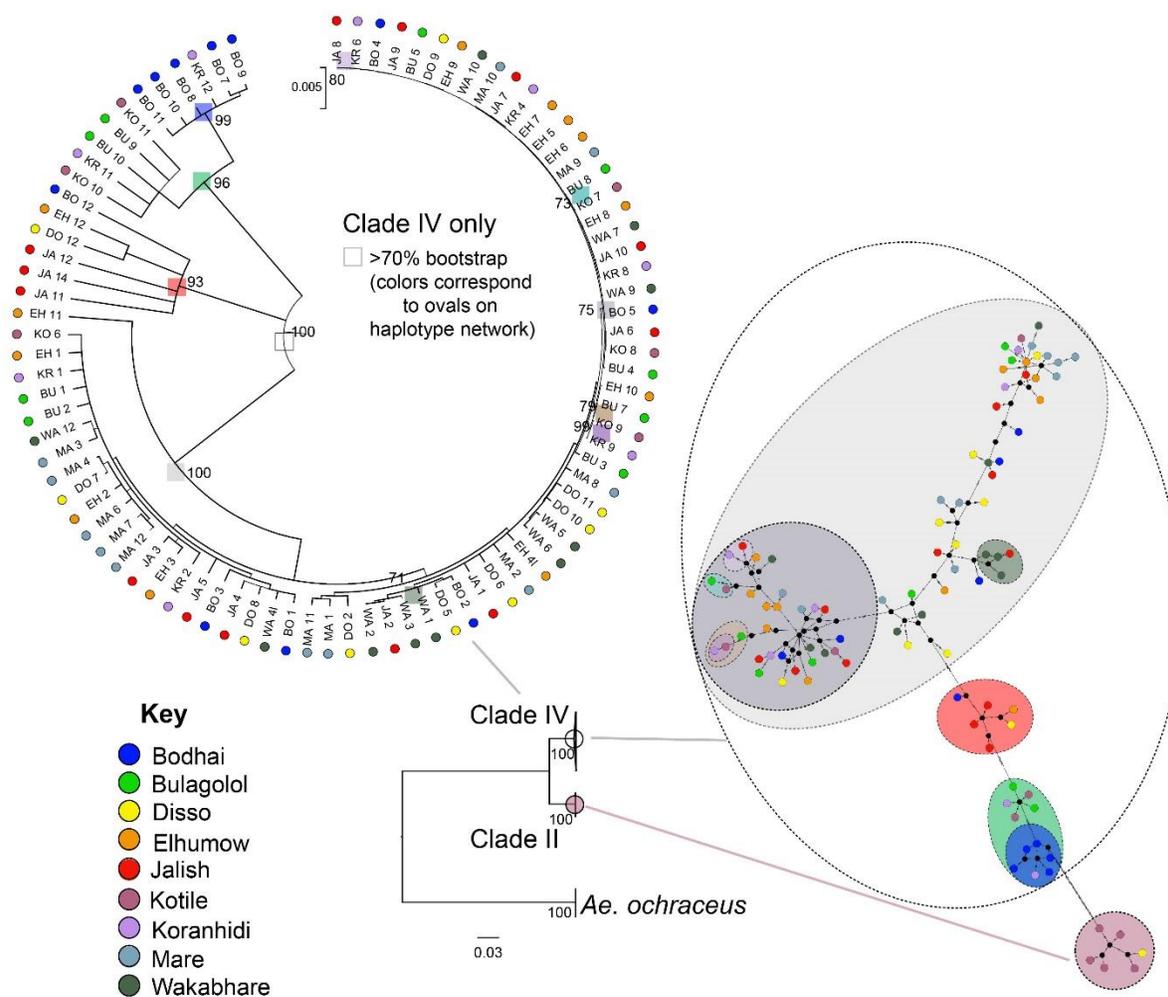


Figure 2. Maximum likelihood tree (bottom) showing relationship of Clade II and Clade IV *Aedes mcintoshi* CO1 sequences with respect to *Ae. ochraceus*. Scale bar gives relative number of substitutions per site. Median joining haplotype network (right) showing relationships of Clade II and Clade IV sequences. Shaded ovals in Clade IV correspond to nodes with high

support (>70% bootstrap) on the Clade IV ML tree (upper left), which are indicated by shaded squares.

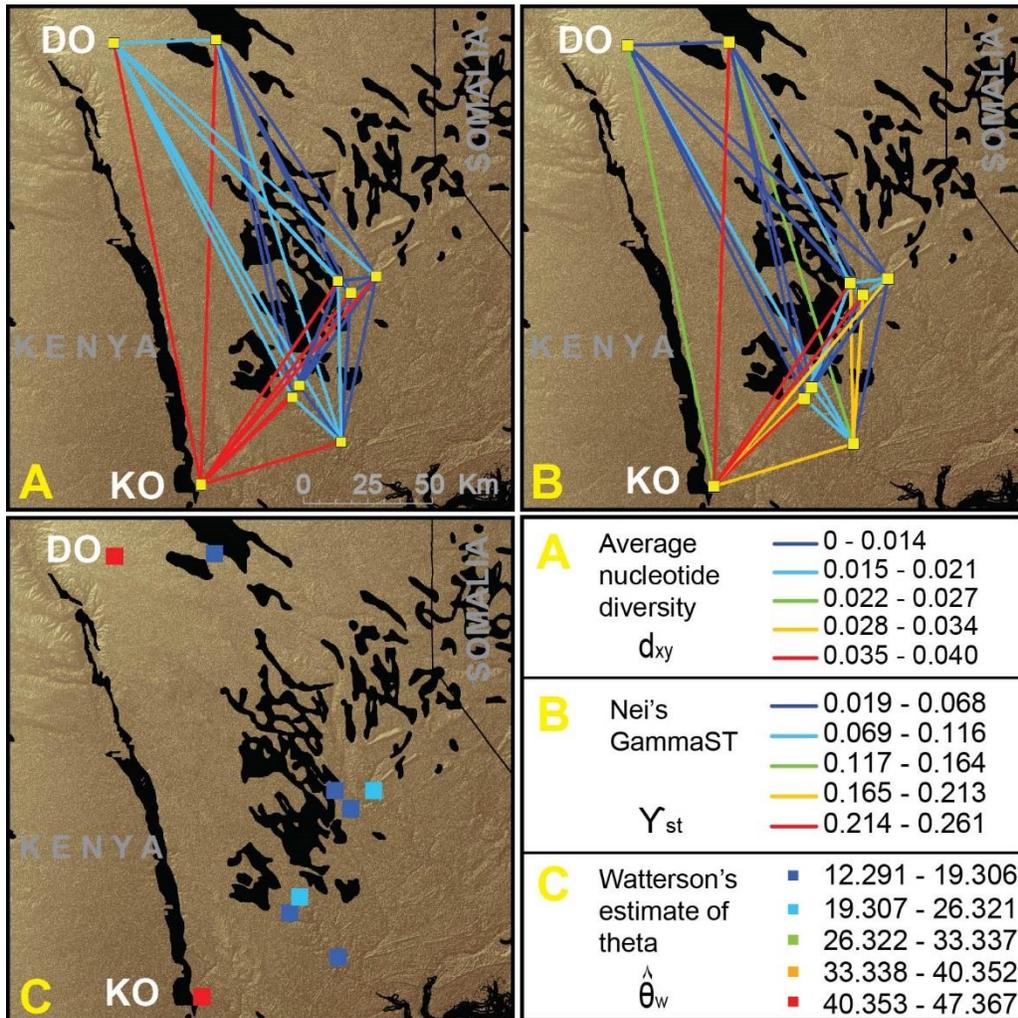


Figure 3. (A and B) Networks of average nucleotide diversity (d_{xy}) and γ_{st} between sample locations (A, B). Yellow squares are the sampling locations, with edge colors representing differentiation values. (C) Watterson's estimate of theta ($\hat{\theta}_w$). Sampling locations correspond to Figure 1.

After excluding Kotile and Disso, 111 segregating sites were present over the 74 remaining Clade IV sequences (all unique haplotypes). Average nucleotide diversity was 0.0123, less than that previously reported for *Ae. mcintoshi* more broadly, including samples from other clades (0.041), but greater than reported for *Ae. ochraceus* (0.0061), the sister species of *Ae. mcintoshi* (Tchouassi et al. 2014). We found no evidence of deviation from neutrality in our combined data set (Tajima's $D = -0.771$, $p > 0.10$), or at individual locations (Supplementary Table 4). We found significant population differentiation between several locations, summarized in Table 1.

Table 1. Population pairwise average nucleotide differences between locations (upper triangle of matrix) and corrected p-values (lower triangle of matrix).

	BO	BU	EH	JA	KR	MA	WA
	21.078	22.154	22.407	22.448	22.281	22.331	22.280
BO	-	21.155	18.508	19.590	20.029	18.457	17.689
BU	0.164	-	16.836	17.292	19.184	15.379	15.091
EH	0.017	0.637	-	18.746	20.251	16.610	15.934
JA	0.039	0.581	0.705	-	22.592	19.155	18.566
KR	0.264	0.998	0.635	0.593	-	12.280	13.127
MA	0.001	0.048	0.114	0.084	0.053	-	13.033
WA	0.008	0.194	0.287	0.387	0.171	0.183	-

We did not detect any significant signal of isolation by distance (Mantel test: observation = -0.106, $p = 0.478$), and a plot of genetic distance vs. the natural log of geographic distance showed no visible trend (Figure 4). Results from the sPCA indicated no significant global ($p = 0.744$) or local spatial structure ($p = 0.247$), indicating that spatially lagged scores did not need to be included as a variable in the partial dbRDA analysis to account for spatial autocorrelation.

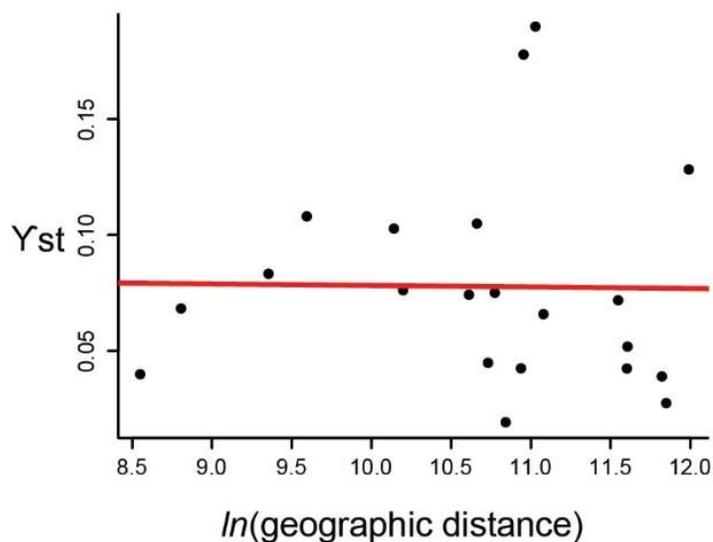


Figure 4. Genetic vs. geographic distance (in km) excluding Kotile and Disso locations.

Results from partial dbRDA models indicated a positive association between pairwise γ_{st} values and mean November precipitation (1980 - 2005) ($p = 0.048$; Table 1). We found no significant associations between pairwise γ_{st} values and mean precipitation during 1980 - 2005 for October or December, anomalous precipitation values during 1900 - 2014, wetness index values, or percent clay in the soil.

Partial RDA model results indicated a positive association between $\hat{\theta}_w$ and mean October precipitation for 1980 - 2005 ($p = 0.071$), and a negative association between $\hat{\theta}_w$ and percent clay in the soil ($p = 0.055$). No significant associations were found between $\hat{\theta}_w$ and minimum or maximum wetness index values, mean precipitation for November or December in 1980 - 2005, or mean anomalous precipitation values (1900 - 2014).

The partial RDA $\hat{\theta}_w$ model with mean October precipitation values (1980 – 2005) had the greatest proportion of explained variance out of the three significant models (constrained variance = 0.659), followed by the partial RDA $\hat{\theta}_w$ model with percent soil as the environmental variable (constrained variance = 0.647), and the partial dbRDA pairwise γ_{st} model that included mean November precipitation values for 1980 - 2005 (constrained variance = 0.530; Table 2, Figure 5). Proportions of explained variance in the conditional location variables for each model were 0.110, 0.110, and 0.233, respectively. Results from the Procrustes test were similar to dbRDA results, indicating a significant similarity between the distance matrix for mean precipitation for the month of November between 1980 and 2005 and the γ_{st} distance matrix ($p = 0.047$; Supplementary Table 5).

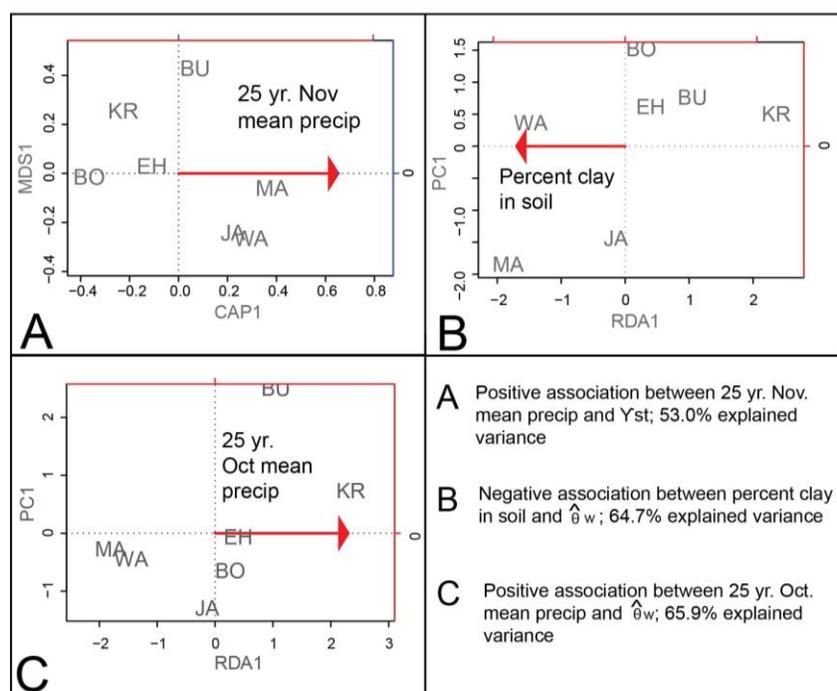


Figure 5. Plots of partial dbRDA and partial RDA models for the three models with the greatest proportions of explained variance in the constrained environmental variable. Red arrows indicate the direction of the association.

Table 2. Significant associations from partial dbRDA and partial RDA analyses, excluding Kotile and Disso sites.

<u>γst</u>										
		Proportion of explained variance					Summary statistics			
Model	Environmental Variable	Conditional (location)	Constrained (environmental)	Unconstrained	DF	Sum of squares	F	P-value	Residual DF	Residual sum of squares
Positive	Mean precipitation (Nov) 1980 to 2005	0.233	0.53	0.236	$\hat{\theta}$ 1	0.014	6.734	0.048	3	0.006
Watterson's Estimate of Theta (w)										
		Proportion of explained variance					Summary statistics			
Model	Environmental Variable	Conditional (location)	Constrained (environmental)	Unconstrained	DF	Variance	F	P-value	Residual DF	Residual variance
Positive	Mean precipitation (Oct) 1980 to 2005	0.11	0.659	0.232	1	9.121	8.52	0.071	3	3.212
Negative	Percent clay in soil	0.11	0.647	0.243	1	8.9262	7.977	0.055	3	3.371

Discussion

We investigated genetic subpopulation structure of *Ae. mcintoshi* in eastern Kenya and evaluated potential associations between environmental variables and genetic diversity and differentiation. Significant subpopulation structure existed across the study area, and significant associations existed between recent mean precipitation values and percentage clay in the soil with genetic diversity and differentiation. Of particular interest were associations with mean precipitation and anomalous precipitation at different time periods that suggest that more recent climate and weather events influence genetic structure over the study area. We expected higher precipitation values to promote connectivity among dambo habitats, reducing genetic differentiation between locations and increasing genetic diversity of subpopulations.

A significant positive association between the 25-year mean precipitation for October and genetic diversity appears to support our hypothesis that higher precipitation is associated with increased genetic diversity. October marks the beginning of the “short rains” season in East Africa. At least one massive emergence of *Ae. mcintoshi* occurred during the short rains in this 25-year time period, just prior to the 2006 and 2007 epizootic eastern Kenya (Sang et al. 2010), but additional punctuations in abundance likely occurred, most notably in October 1997, just prior to reported RVFV cases in this area in November 1997 (Murithi et al. 2011).

Although mean precipitation for October from 1980 to 2005 had a positive association with genetic diversity at locations, we did not find any association with reduced genetic differentiation between locations as we expected. In fact, contrary to our expectations, we found a positive association between genetic differentiation (γ_{st}) and mean precipitation for November over 1980 - 2005. This association did not support our hypothesis that greater precipitation would provide greater connectivity between locations, resulting in reduced genetic

differentiation. Fine-scale, limited dispersal of this species, coupled with additional variables not included in this analysis, may be responsible for this discrepancy. Alternatively, suitable habitat resulting from elevated precipitation may limit dispersal of this species by reducing the need to disperse to obtain blood meals or reproduction. Further investigation into the role of precipitation at additional spatial and temporal scales is needed to determine whether this result is indeed, an artefact.

Although we hypothesized a positive association between genetic diversity and soil clay content, model results suggested that diversity increased as percent clay decreased at sample site locations. Soils with lower percentages of clay tend to drain water more readily, which may drive *Ae. mcintoshi* to disperse greater distances to find optimal habitats for blood feeding. Linthicum et al. (1985a) noted that host density tends to decrease with distance from flooded dambo habitats because domesticated livestock and wild mammals congregate at these sites for drinking water. *Aedes mcintoshi* emerging from less-ideal habitats may be more likely to disperse to more suitable environments, and this factor could impact genetic diversity in samples collected in suboptimal habitats. Alternatively, these results may be influenced by the low number of sampling sites and relatively homogeneous soil values across study locations, warranting additional investigation to capture a broader heterogeneity of soil attributes across the geographic range of Clade IV.

Campbell et al. (in review) found that elevated minimum wetness index values within 500 m of a sampling site had a strong positive effect on mosquito abundances, but we found no significant association between wetness index values and genetic diversity or differentiation. We expected locations with higher minimum wetness index values to be on landscapes more likely to be connected during precipitation events, providing opportunities for greater genetic diversity

and reduced differentiation. Because wetness index values are derived from a static topographic variable that is relatively unchanged over time, this variable provided an opportunity to observe potential associations arising over a longer temporal period, but wetness index values were not an important variable in our models, at least at these spatial and temporal scales. In addition, we found no significant associations with mean anomalous precipitation (1900 – 2014) and genetic diversity. This variable was measured at a broader spatial scale than the clim mean precipitation data from 1980 to 2005 (i.e., 10 km vs. 5 km spatial resolution), which may have smoothed variability important to associations between precipitation and genetic variability in this region. Additionally, this variable is a mean composite derived across several months from October to February, which may have diminished associations further.

We found no evidence for isolation by distance contributing to genetic differentiation across the study area. Notably, differentiation was often low between locations separated by large geographic distances, such as Jalish and Elhumow, whereas greater γ st values were sometimes observed over close proximity, such as Bodhai and Bulagolol. However, location was an important conditioning variable, accounting for 11 - 23% of the proportion of explained variance in the significant models. These results indicated that at-location variables that were not included in the analysis may contribute to genetic diversity and differentiation in the study area. Additional variables, such as soil pH, finer-scale wetness index values, and longer-term climatic values at finer scales, may provide additional information regarding genetic structure in this area.

In addition to investigating alternative environmental variables, the combination of flooding events and the topographic structure of dambo habitats may be an important factor in determining the genetic diversity found at a location. Linthicum et al. (1984, 1988) found that the majority of *Ae. mcintoshi* larvae and pupae in flooded dambos were not located in the central,

lowest-lying area, but rather were located toward habitat edges, where sedge vegetation transitioned to grasses. Variable rainfall patterns within the study region may result in long intervals between flooding of outer dambo habitat perimeters, whereas areas slightly inward may experience more regular cycles of adult emergence. This phenomenon, coupled with the small dispersal distances observed in field studies, could introduce additional complexity into measurements of genetic diversity at sites and differentiation between locations that was not captured in this analysis.

Although this study demonstrated subpopulation genetic structure of *Ae. mcintoshi* across eastern Kenya and potential associations between this structure and a suite of environmental variables, further analyses including additional sampling of sites and better sampling of the genome will be critical to understanding in greater detail the genetic diversity and differentiation of *Ae. mcintoshi* in this region. The large number of haplotypes found in this data set indicates that much diversity remains to be detected in this area. Although mitochondrial DNA analysis has been successful in identifying mosquito subpopulation structure (Yugavathy et al. 2016), the thousands of loci characterized in next-generation sequencing approaches would provide a more robust assessment of the genetic diversity and differentiation across this region. In addition, including more sample site locations across a broader geographic area within the Clade IV range would capture more environmental heterogeneity, while providing an opportunity to investigate associations between combinations of environmental variables and genetic diversity and differentiation of *Ae. mcintoshi*. Understanding the subtle interactions between *Ae. mcintoshi* subpopulations across the region that may contribute to RVFV maintenance and dissemination in the natural environment will contribute to more informed

targeting of veterinary and public health resources and more comprehensive prevention approaches.

Acknowledgements

The authors would like to thank A. Townsend Peterson for his careful review of the manuscript prior to submission.

References

- Anyamba, A., J. P. Chretien, J. Small, C. J. Tucker, and K. J. Linthicum. 2006. Developing global climate anomalies suggest potential disease risks for 2006-2007. *International Journal of Health Geographics* **5**:60.
- Anyamba, A., K. J. Linthicum, and C. J. Tucker. 2001. Climate-disease connections: Rift Valley fever in Kenya. *Cadernos de Saúde Publica* **17**:133-140.
- Bandelt, H. J., P. Forster, and A. Rohlf. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**:37-48.
- Beven, K., and M. Kirkby. 1979. A physically-based variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin* **24**:43-69.
- Camberlin, P., S. Janicot, and I. Pocard. 2001. Seasonality and atmospheric dynamics of the teleconnection between African rainfall and tropical sea-surface temperature: Atlantic vs. ENSO. *International Journal of Climatology* **21**:973-1005.
- Chengula, A. A., R. H. Mdegela, and C. J. Kasanga. 2013. Socio-economic impact of Rift Valley fever to pastoralists and agro pastoralists in Arusha, Manyara and Morogoro regions in Tanzania. *Springer Plus* **2**:549.

- Davies, F. G., K. J. Linthicum, and A. D. James. 1985. Rainfall and epizootic Rift-Valley fever. *Bulletin of the World Health Organization* **63**:941-943.
- Dray, S., and A. B. Dufour. 2007. The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software* **22**:1-20.
- Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**:1792-1797.
- Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**:564-567.
- Funk, C., S. E. Nicholson, M. Landsfeld, D. Klotter, P. Peterson, and L. Harrison. 2015a. The Centennial Trends Greater Horn of Africa precipitation dataset. *Sci Data* **2**:150050.
- Funk, C., A. Verdin, J. Michaelsen, P. Peterson, D. Pedreros, and G. Husak. 2015b. A global satellite-assisted precipitation climatology. *Earth System Science Data* **7**:275-287.
- Hillis, D. M., and J. J. Bull. 1993. An empirical-test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* **42**:182-192.
- Jombart, T. 2008. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**:1403-1405.
- Jombart, T., S. Devillard, A. B. Dufour, and D. Pontier. 2008. Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* **101**:92-103.
- Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**:1647-1649.

- Kierepka, E. M., and E. K. Latch. 2016. Fine-scale landscape genetics of the American badger (*Taxidea taxus*): Disentangling landscape effects and sampling artifacts in a poorly understood species. *Heredity* **116**:33-43.
- Leigh, J. W., and D. Bryant. 2015. POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6**:1110-1116.
- Librado, P., and J. Rozas. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**:1451-1452.
- Linthicum, K. J., C. L. Bailey, F. G. Davies, and A. Kairo. 1985a. Observations on the dispersal and survival of a population of *Aedes lineatopennis* (Ludlow) (Diptera, Culicidae) in Kenya. *Bulletin of Entomological Research* **75**:661-670.
- Linthicum, K. J., C. L. Bailey, F. G. Davies, A. Kairo, and T. M. Logan. 1988. The horizontal distribution of *Aedes* pupae and their subsequent adults within a flooded dambo in Kenya: Implications for Rift-Valley fever virus control. *Journal of the American Mosquito Control Association* **4**:551-554.
- Linthicum, K. J., F. G. Davies, C. L. Bailey, and A. Kairo. 1984. Mosquito species encountered in a flooded grassland dambo in Kenya. *Mosquito News* **44**:228-232.
- Linthicum, K. J., F. G. Davies, A. Kairo, and C. L. Bailey. 1985b. Rift-Valley fever virus (Family Bunyaviridae, Genus *Phlebovirus*): Isolations from Diptera collected during an inter-epizootic period in Kenya. *Journal of Hygiene* **95**:197-209.
- Lutomiah, J., J. Bast, J. Clark, J. Richardson, S. Yalwala, D. Oullo, J. Mutisya, F. Mulwa, L. Musila, S. Khamadi, D. Schnabel, E. Wurapa, and R. Sang. 2013. Abundance, diversity, and distribution of mosquito vectors in selected ecological regions of Kenya: Public health implications. *Journal of Vector Ecology* **38**:134-142.

- Martin, V., V. Chevalier, P. Ceccato, A. Anyamba, L. De Simone, J. Lubroth, S. de la Rocque, and J. Domenech. 2008. The impact of climate change on the epidemiology and control of Rift Valley fever. *Revue Scientifique Et Technique Office International Des Epizooties* **27**:413-426.
- Murithi, R. M., P. Munyua, P. M. Ithondeka, J. M. Macharia, A. Hightower, E. T. Luman, R. F. Breiman, and M. K. Njenga. 2011. Rift Valley fever in Kenya: History of epizootics and identification of vulnerable districts. *Epidemiology and Infection* **139**:372-380.
- Nguku, P. M., S. K. Sharif, D. Mutonga, S. Amwayi, J. Omolo, O. Mohammed, E. C. Farnon, L. H. Gould, E. Lederman, C. Rao, R. Sang, D. Schnabel, D. R. Feikin, A. Hightower, M. K. Njenga, and R. F. Breiman. 2010. An investigation of a major outbreak of Rift Valley fever in Kenya: 2006-2007. *American Journal of Tropical Medicine and Hygiene* **83**:5-13.
- Oksanen, J., G. F. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. H. Wagner. 2016. *vegan*: Community ecology package. R package version 2.3-4. <http://CRAN.R-project.org/package=vegan>.
- Paradis, E. 2010. *pegas*: an R package for population genetics with an integrated-modular approach. *Bioinformatics* **26**:419-420.
- Pepin, M., M. Bouloy, B. H. Bird, A. Kemp, and J. Paweska. 2010. Rift Valley fever virus (Bunyaviridae: *Phlebovirus*): An update on pathogenesis, molecular epidemiology, vectors, diagnostics and prevention. *Veterinary Research* **41**:61.
- Peres-Neto, P. R., and D. A. Jackson. 2001. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* **129**:169-178.

- R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rich, K. M., and F. Wanyoike. 2010. An assessment of the regional and national socio-economic impacts of the 2007 Rift Valley fever outbreak in Kenya. *American Journal of Tropical Medicine and Hygiene* **83**:52-57.
- Rosmoser, W., M. Oviedo, E. Lerdtusne, L. Patrican, M. Turell, D. Dohm, K. Linthicum, and C. Bailey. 2011. Rift Valley fever virus-infected mosquito ova and associated pathology: Possible implications for endemic maintenance. *Research and Reports in Tropical Medicine* **2**:121 - 127.
- Sang, R., E. Kioko, J. Lutomiah, M. Warigia, C. Ochieng, M. O'Guinn, J. S. Lee, H. Koka, M. Godsey, D. Hoel, H. Hanafi, B. Miller, D. Schnabel, R. F. Breiman, and J. Richardson. 2010. Rift Valley fever virus epidemic in Kenya, 2006/2007: The entomologic investigations. *American Journal of Tropical Medicine and Hygiene* **83**:28-37.
- Shangguan, W., Y. Dai, Q. Duan, B. Liu, and H. Yuan. 2014. A global soil data set for Earth system modeling. *Journal of Advances in Modeling Earth Systems* **6**:249-263.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585-595.
- Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**:2725-2729.
- Tchouassi, D. P., A. D. Bastos, C. L. Sole, M. Diallo, J. Lutomiah, J. Mutisya, F. Mulwa, C. Borgemeister, R. Sang, and B. Torto. 2014. Population genetics of two key mosquito

vectors of Rift Valley fever virus reveals new insights into the changing disease outbreak patterns in Kenya. *PLoS Neglected Tropical Diseases* **8**:e3364.

World Health Organization. 2010. Rift Valley fever. Fact sheet N°207.

Yugavathy, N., L. Kim-Sung, S. Joanne, and I. Vythilingam. 2016. Genetic variation of the mitochondrial genes, CO1 and ND5, in *Aedes aegypti* from various regions of peninsular Malaysia. *Tropical Biomedicine* **33**:543-560.

Supplementary Table 1. Pairwise γ st ITS data, with and without Kotile and Disso sites.

Pairwise γ st ITS data including Kotile and Disso							
BO							
BU	0.16197						
DO	0.93272	0.43038					
EH	0.875	0.37857	0.36364				
JA	0.28467	0.03082	0.32819	0.27083			
KO	0.20666	0.01326	0.37966	0.33793	0.02271		
MA	0.87506	0.44452	0.37551	0.01818	0.31409	0.39565	
WA	0.16009	0.01606	0.4155	0.35069	0.02458	0.01958	0.41273
Pairwise γ st ITS data no Kotile							
BO							
BU	0.16197						
DO	0.93272	0.43038					
EH	0.875	0.37857	0.36364				
JA	0.28467	0.03082	0.32819	0.27083			
MA	0.87506	0.44452	0.37551	0.01818	0.31409		
WA	0.16009	0.01606	0.4155	0.35069	0.02458	0.41273	
Pairwise γ st ITS data no Kotile or Disso							
BO							
BU	0.16197						
EH	0.875	0.37857					
JA	0.28467	0.03082	0.27083				
MA	0.87506	0.44452	0.01818	0.31409			
WA	0.16009	0.01606	0.35069	0.02458	0.41273		

Supplementary Table 2. CO1 data yst model results with and without Kotile and Disso sites.

Model		Proportion of explained variance					Summary statistics				
Sites	Association	Environmental variable	Conditional (location)	Constrained (environmental)	Unconstrained	DF	Sum of squares	F	p-value	Residual DF	Residual sum of squares
All	No	Mean 25 yr. precip (Oct)	0.486469	0.032122	0.481408	1	0.002304	0.3336	0.816	5	0.034528
No KO	No	Mean 25 yr. precip (Oct)	0.233	0.124	0.643	1	0.004	0.771	0.505	4	0.019
No KO or DO	No	Mean 25 yr. precip (Oct)	0.233294	0.386834	0.379872	1	0.010571	3.055	0.143	3	0.010381
All	Positive	Mean 25 yr. precip (Nov)	0.486	0.198	0.315	1	0.014	3.147	0.033	5	0.023
No KO	Positive	Mean 25 yr. precip (Nov)	0.233	0.475	0.292	1	0.014	6.504	0.014	4	0.009
No KO or DO	Positive	Mean 25 yr. precip (Nov)	0.233	0.530	0.236	1	0.014	6.734	0.048	3	0.006
All	No	Mean 25 yr. precip (Dec)	0.486469	0.028078	0.485453	1	0.002014	0.2892	0.845	5	0.034818
No KO	No	Mean 25 yr. precip (Dec)	0.23323	0.054247	0.712524	1	0.001641	0.3045	0.762	4	0.021558
No KO or DO	No	Mean 25 yr. precip (Dec)	0.233294	0.09154	0.675166	1	0.002502	0.4067	0.656	3	0.01845
All	No	Anomalous precipitation	0.486469	0.060087	0.453444	1	0.00431	0.6626	0.578	5	0.032522
No KO	No	Anomalous precipitation	0.23323	0.070811	0.695959	1	0.002143	0.407	0.68	4	0.021057

No KO or DO	No	Anomalous precipitation	0.233294	0.102457	0.664249	1	0.0028	0.4627	0.583	3	0.018152
All	No	Soil	0.486469	0.041984	0.471547	1	0.003011	0.4452	0.701	5	0.033821
No KO	No	Soil	0.23323	0.078333	0.688438	1	0.00237	0.4551	0.655	4	0.02083
No KO or DO	No	Soil	0.233294	0.096383	0.670323	1	0.002634	0.4314	0.635	3	0.018318
All	No	Minimum wetness index	0.486469	0.143547	0.369984	1	0.010296	1.9399	0.147	5	0.026536
No KO	No	Minimum wetness index	0.23323	0.316162	0.450609	1	0.009566	2.8065	0.129	4	0.013634
No KO or DO	No	Minimum wetness index	0.233294	0.349337	0.417369	1	0.009546	2.511	0.183	3	0.011405
All	No	Maximum wetness index	0.486469	0.027871	0.48566	1	0.001999	0.2869	0.832	5	0.034833
No KO	No	Maximum wetness index	0.23323	0.033749	0.733021	1	0.001021	0.1842	0.902	4	0.022179
No KO or DO	No	Maximum wetness index	0.233294	0.0230315	0.7436745	1	0.000629	0.0929	0.943	3	0.020322

Supplementary Table 3. CO1 data $\hat{\theta}_w$ model results with and without Disso and Kotile sites.

Model		Proportion of explained variance					Summary statistics				
Sites	Association	Environmental Variable	Conditional (location)	Constrained (environmental)	Unconstrained	DF	Variance	F	p-value	Residual DF	Residual Variance
All	No	Mean 25 yr. precip (Oct)	0.728	0.051	0.222	1	8.225	1.142	0.315	5	36.018
No KO	No	Mean 25 yr. precip (Oct)	0.620	0.130	0.251	1	16.045	2.073	0.221	4	30.958
No KO or DO	Positive	Mean 25 yr. precip (Oct)	0.110	0.659	0.232	1	9.121	8.520	0.071	3	3.212
All	No	Mean 25 yr. precip (Nov)	0.728	0.019	0.253	1	3.084	0.375	0.553	5	41.159
No KO	No	Mean 25 yr. precip (Nov)	0.620	0.048	0.333	1	5.929	0.577	0.482	4	41.074
No KO or DO	No	Mean 25 yr. precip (Nov)	0.110	0.173	0.717	1	2.401	0.725	0.463	3	9.932
All	Negative	Mean 25 yr. precip (Dec)	0.728	0.139	0.133	1	22.587	5.215	0.070	5	21.656
No KO	Negative	Mean 25 yr. precip (Dec)	0.620	0.234	0.146	1	28.937	6.407	0.071	4	18.065
No KO or DO	No	Mean 25 yr. precip (Dec)	0.110	0.232	0.659	1	3.210	1.055	0.389	3	9.124

All	Negative	Anomalous precipitation	0.728	0.200	0.073	1	32.444	13.749	0.016	5	11.799
No KO or DO	Negative	Anomalous precipitation	0.620	0.278	0.103	1	34.293	10.793	0.025	4	12.710
All	Negative	Anomalous precipitation	0.110	0.330	0.561	1	4.564	1.762	0.294	3	7.769
No KO or DO	Negative	Soil	0.728	0.154	0.118	1	24.992	6.491	0.075	5	19.251
No KO or DO	Negative	Soil	0.620	0.247	0.134	1	30.518	7.406	0.047	4	16.484
All	Negative	Soil	0.110	0.647	0.243	1	8.962	7.977	0.055	3	3.371
All	No	Minimum wetness index	0.728	0.002	0.272	1	0.036	0.004	0.955	5	44.207
No KO or DO	No	Minimum wetness index	0.619	0.013	0.367	1	1.632	0.144	0.750	4	45.371
No KO or DO	No	Minimum wetness index	0.110	0.017	0.873	1	0.235	0.058	0.823	3	12.098
All	No	Maximum wetness index	0.728	0.034	0.239	1	5.460	0.704	0.400	5	38.783
No KO or DO	No	Maximum wetness index	0.619	0.033	0.347	1	4.098	0.382	0.560	4	42.905
No KO or DO	No	Maximum wetness index	0.110	0.051	0.839	1	0.710	0.183	0.721	3	11.623

Supplementary Table 4. Theta-Pi for each sample location.

Theta-Pi for each sample location									
Statistics	BO	BU	EH	JA	KR	MA	WA	Mean	<u>s.d.</u>
Sample size	11	9	12	13	8	11	10	10.5714	1.71825
Sample size	56	57	50	57	57	36	37	50	9.55685
Pi	20.6909	20.7778	16.5909	18.4487	22.1786	12.1455	12.8889	17.6745	3.96362
Tajima's D	0.38938	-0.0475	0.00941	0.01974	0.04788	-0.0553	-0.0704	0.0419	0.15936
Tajima's D p-value	0.6962	0.5197	0.5427	0.5468	0.5445	0.5211	0.5045	0.55364	0.06481

Supplementary Table 5. Procrustes test for congruence between the Yst distance matrix and environmental variable distance matrices.

Response Variable	Environmental Variable	Procrustes Sum of Squares (m12 squared)	Correlation in a symmetric Procrustes rotation	Significance	Number of permutations
<u>Yst</u>	Location	0.7721	0.4774	0.543	999
<u>Yst</u>	Minimum wetness index	0.4965	0.7096	0.128	999
<u>Yst</u>	Maximum wetness index	0.8071	0.4392	0.713	999
<u>Yst</u>	Percent clay in soil	0.8568	0.3784	0.54	999
<u>Yst</u>	Mean 25 yr. precipitation (Oct)	0.6753	0.5698	0.374	999
<u>Yst</u>	Mean 25 yr. precipitation (Nov)	0.4585	0.7358	0.047	999
<u>Yst</u>	Mean 25 yr. precipitation (Dec)	0.7382	0.5117	0.424	999
<u>Yst</u>	Mean anomalous precipitation (1900 to 2014)	0.748	0.502	0.528	999

CONCLUSION

The set of comprehensive analyses presented here represents new applications of modeling approaches and technologies toward investigations of disease ecology questions. Individually, each approach addressed different research questions and made different contributions to the larger body of scientific knowledge regarding the distribution of risks from these pathogens across the globe. The first chapter delivered a map of suitable environments for *Ae. aegypti* and *Ae. albopictus* mosquitos under present and future climate conditions, while identifying potential changes in co-occurrence of these species between the present time period and 2050. The second chapter outlined a new model for identifying environmental drivers of *Ae. mcintoshi* abundances, and retrospective predictions indicated very high predicted abundances just prior to the last major epizootic in Kenya and Somalia. The finer temporal resolution of these predicted abundances, along with the potential to expand this framework to other vector species in this region, delivered a novel approach to identifying the timing and location of potential virus emergence. In the third chapter, further investigation into the genetic structure of *Ae. mcintoshi* revealed that subpopulation structure exists and that environmental factors, rather than isolation-by-distance, are contributors to this observed pattern. In addition to these results, the analyses performed here demonstrated the utility for continued future investigations of *Ae. mcintoshi* using a greater number of sampling sites and next-generation sequencing approaches.

Together, these analyses shared the common purpose to provide robust, quantitative information that may be used to help narrow broad geographic ranges for field studies, target intervention or control approaches, or help identify areas in which to further emphasize prevention programs. While each of these studies was not without limitations, the importance of ecologically-focused analyses addressing different disease system components cannot be

understated. As humans continue to interact with the natural environment in ways that increase contact with areas undisturbed previously, promote rapid changes in climate and weather, or provide new habitats for pathogen dispersal, human disease risk from vector species will continue to challenge public health efforts. In this context, ecological models provide a fundamental resource for assessing potential risk of vector-borne diseases, and further development and applications will serve to improve our understanding of these systems now and into the future.

References

- Chengula, A. A., R. H. Mdegela, and C. J. Kasanga. 2013. Socio-economic impact of Rift Valley fever to pastoralists and agro pastoralists in Arusha, Manyara and Morogoro regions in Tanzania. *Springer Plus* **2**:549.
- Hayes, E. B. 2009. Zika virus outside Africa. *Emerging Infectious Diseases* **15**:1347-1350.
- Jentes, E. S., G. Pomeroy, M. D. Gershman, D. R. Hill, J. Lemarchand, R. F. Lewis, J. E. Staples, O. Tomori, A. Wilder-Smith, T. P. Monath, and I. W. W. G. Geographi. 2011. The revised global yellow fever risk map and recommendations for vaccination, 2010: Consensus of the Informal WHO Working Group on Geographic Risk for Yellow Fever. *Lancet Infectious Diseases* **11**:622-632.
- Kalluri, S., P. Gilruth, D. Rogers, and M. Szczur. 2007. Surveillance of arthropod vector-borne infectious diseases using remote sensing techniques: A review. *PLoS Pathogens* **3**:1361-1371.
- Kilpatrick, A. M., and S. E. Randolph. 2012. Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *Lancet* **380**:1946-1955.

- Lambrechts, L., T. W. Scott, and D. J. Gubler. 2010. Consequences of the expanding global distribution of *Aedes albopictus* for dengue virus transmission. *American Journal of Tropical Medicine and Hygiene* **83**:6-7.
- Last, J. M., J. H. Abramson, and International Epidemiological Association. 1995. A dictionary of epidemiology. 3rd edition. Oxford University Press, New York.
- Linthicum, K. J., F. G. Davies, A. Kairo, and C. L. Bailey. 1985. Rift-Valley fever virus (Family Bunyaviridae, Genus *Phlebovirus*): Isolations from Diptera collected during an inter-epizootic period in Kenya. *Journal of Hygiene* **95**:197-209.
- Meade, M. S., and R. Earickson. 2000. Medical geography. 2nd edition. Guilford Press, New York.
- Murithi, R. M., P. Munyua, P. M. Ithondeka, J. M. Macharia, A. Hightower, E. T. Luman, R. F. Breiman, and M. K. Njenga. 2011. Rift Valley fever in Kenya: History of epizootics and identification of vulnerable districts. *Epidemiology and Infection* **139**:372-380.
- Ostfeld, R. S., F. Keesing, and V. T. Eviner. 2008. Infectious disease ecology : The effects of ecosystems on disease and of disease on ecosystems. Princeton University Press, Princeton, N.J.
- Pages, F., C. N. Peyrefitte, M. T. Mve, F. Jarjaval, S. Brisse, I. Iteaman, P. Gravier, D. Nkoghe, and M. Grandadam. 2009. *Aedes albopictus* mosquito: The main vector of the 2007 chikungunya outbreak in Gabon. *PloS One* **4**:e4691.
- Pavlovsky, E. N. 1966. The natural nidity of transmissible disease. University of Illinois Press, Urbana, IL.

- Pepin, M., M. Bouloy, B. H. Bird, A. Kemp, and J. Paweska. 2010. Rift Valley fever virus (Bunyaviridae: *Phlebovirus*): An update on pathogenesis, molecular epidemiology, vectors, diagnostics and prevention. *Veterinary Research* **41**:61.
- Peterson, A. T. 2014. Mapping disease transmission risk : Enriching models using biogeography and ecology. Johns Hopkins University Press, Baltimore, MD.
- Plowright, R. K., S. H. Sokolow, M. E. Gorman, P. Daszak, and J. E. Foley. 2008. Causal inference in disease ecology: investigating ecological drivers of disease emergence. *Frontiers in Ecology and the Environment* **6**:420-429.
- Reisen, W. K. 2010. Landscape epidemiology of vector-borne diseases. *Annual Review of Entomology* **55**:461-483.
- Rogers, D. J., A. J. Wilson, S. I. Hay, and A. J. Graham. 2006. The global distribution of yellow fever and dengue. *Advances in Parasitology* **62**:181-220.
- Rosmoser, W., M. Oviedo, E. Lerdthusne, L. Patrican, M. Turell, D. Dohm, K. Linthicum, and C. Bailey. 2011. Rift Valley fever virus-infected mosquito ova and associated pathology: Possible implications for endemic maintenance. *Research and Reports in Tropical Medicine* **2**:121 - 127.
- Tchouassi, D. P., A. D. Bastos, C. L. Sole, M. Diallo, J. Lutomiah, J. Mutisya, F. Mulwa, C. Borgemeister, R. Sang, and B. Torto. 2014. Population genetics of two key mosquito vectors of Rift Valley fever virus reveals new insights into the changing disease outbreak patterns in Kenya. *PLoS Neglected Tropical Diseases* **8**:e3364.