Mosquito habitat and dengue risk potential in Kenya: alternative methods to traditional risk mapping techniques

David F. Attaway[1,2], Kathryn H. Jacobsen[3], Allan Falconer[1], Germana Manca[1], Lauren Rosenshein Bennett[2,4], Nigel M. Waters[1,5,6]

[1]Department of Geography and GeoInformation Science, George Mason University, Fairfax, USA; [2]ESRI, Redlands, USA; [3]Department of Global and Community Health, George Mason University, Fairfax, USA; [4]Center for Information Systems and Technology, Claremont Graduate University, Claremont, USA; [5]GIS Center of Excellence, George Mason University, Fairfax, USA; [6]Institute of Public Health, Faculty of Medicine, University of Calgary, Calgary, Canada

Abstract. Outbreaks, epidemics and endemic conditions make dengue a disease that has emerged as a major threat in tropical and sub-tropical countries over the past 30 years. Dengue fever creates a growing burden for public health systems and has the potential to affect over 40% of the world population. The problem being investigated is to identify the highest and lowest areas of dengue risk. This paper presents "Similarity Search", a geospatial analysis aimed at identifying these locations within Kenya. Similarity Search develops a risk map by combining environmental susceptibility analysis and geographical information systems, and then compares areas with dengue prevalence to all other locations. Kenya has had outbreaks of dengue during the past 3 years, and we identified areas with the highest susceptibility to dengue infection using bioclimatic variables, elevation and mosquito habitat as input to the model. Comparison of the modelled risk map with the reported dengue epidemic cases obtained from the open source reporting ProMED and Government news reports from 1982-2013 confirmed the high-risk locations that were used as the Similarity Search presence cells. Developing the risk model based upon the bioclimatic variables, elevation and mosquito habitat increased the efficiency and effectiveness of the dengue fever risk mapping process.

Keywords: dengue, geographical information system, risk mapping, medical geography, Kenya.

Introduction

Dengue fever (DF) is a viral infection transmitted by *Aedes* mosquitoes. Infection with any of the four strains (DEN-1 to 4) may cause febrile illness and joint pain and has the potential for serious complications, including dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS). Nearly 40% of the world's population lives in areas where dengue is endemic (Guzman and Istúriz, 2010), and 50 to 100 million infections occur globally each year (WHO, 2012). Dengue virus infection is not yet vaccine-preventable or curable by drugs, so public health efforts must focus on mosquito control.

Recent research suggests that the burden of dengue in Africa is lower than that in Asia but comparable to that reported in the Americas (Franco et al., 2010;

Bhatt et al., 2013). The first confirmed outbreak in Kenya occurred in 1982 (Johnson et al., 1982b). More recent studies from this country have reported cases from both the subtropical coastal regions and the arid north-eastern part of the country (Sang, 2007). Table 1 and Fig. 1 highlight the dengue reports from Kenya. Several cross-sectional surveys show that a sizeable proportion of Kenyans have serologic evidence of past dengue infection, and this suggests that the reported dengue incidence may be significantly underestimated. Dengue transmission in Kenya is facilitated by the humid and rainy climate along the coast, where the vector mosquitoes thrive, as well as by expanding road and rail transportation networks (Gubler, 2004).

National *Aedes* data are continuously compiled by the Kenya Medical Research Institute (KEMRI) and collaborating agencies (Sang and Dunster, 2001; Sang, 2007). Additional sources of climatic and other geographic information are readily available for integration into spatial models. Improved visualization of dengue risk is needed and should be of use to the Kenya Ministry of Health (MoH), where it is handled by the Division of Disease Surveillance and Response (DDSR), the Kenyan Field and Laboratory training Programme

Corresponding author:
David F. Attaway
Department of Geography and GeoInformation Science
George Mason University, 5300 Columbia Pike, Apt 1014
Arlington, VA 22204, USA
Tel. +1 214 681-612; Fax +1 703 993-9299
Email: dattaway@gmu.edu

(FELTP), KEMRI and their partners in infectious disease prevention and control (Akhwale, 2013).

The goal of the analysis presented in this paper was to create a more accurate and robust map of dengue risk in Kenya based on bioclimatic layers, elevation data, high resolution population data and information about the range of the *Aedes* vector with the aim of integrating this information into a localised high-resolution format (1 x 1 km). The techniques used for the creation of this new risk map are also applicable to other infectious diseases and their settings.

Materials and methods

Spatial data

Administrative boundary data for Kenya were obtained from the Kenya Central Bureau of Statistics (2005) and verified against the Global Administrative Areas (GADM), which is a spatial database of the location of the world's administrative areas (or boundaries) (Global Administrative Areas, 2012). The datasets were set to a custom projection using a projected coor-

Table 1. Seroprevalence and dengue outbreak reports from Kenya, 1982-2013.

| Event year	Location	Study design	Sample size	Result	Reference
1982	Malindi	Case report	1	Tourist visiting Malindi positive for DEN-2	Johnson et al. (1982a)
1982-1983	Kilifi	Serosurvey of outpatients	96	Seroprevalence: increased from 7% in 1982 to 57% in 1983 (DEN-2)	Johnson et al. (1982a,b, 1990)
1987	Kilifi and Malindi	Serosurvey of outpatients	1,059	Seroprevalence: 1.3% (DEN-2)	Johnson et al. (1990)
1999	Kilifi	Case report	1	Child positive for DEN-2	Sang and Dunster (2001)
2000-2003	Msambweni district	Serosurvey of pregnant women	419	Seroprevalence rates: 63% (DEN-1), 67% (DEN-2), 55% (DEN-3), 44% (DEN-4)	Sutherland et al. (2011)
2004	Malindi	Population-based cross-sectional survey	442	Seroprevalence: 34.2% (DEN-2)	Mease et al. (2011)
2011	Mandera	Case series	5,000	5,000 cases and 10 dengue deaths reported between September and October	ProMED-mail (2011)
2010-2011	Western Kenya	Cross-sectional survey	354	Seroprevalence: 1.1%; annual incidence: 8.5/1,000 persons	Blaylock et al. (2011)
2010-2011	Busia and Trans-Nzoia counties (Western Kenya)	Serosurvey of outpatients	422	Seroprevalence: 8.5% by indirect ELISA and 1.2% by PRNT testing (DEN-2)	Awando et al. (2013)
2013	Mandera	Case series	300	300 cases and 3 dengue deaths between January and February	ProMED-mail (2013c,f)
2013	Mombasa (Kwale and Kilifi counties)	Case series	140	KEMRI identified more than 56% (140 cases) of dengue cases as positive for dengue virus: DEN-1 (69%), DEN-2 (28%), and DEN-3 (3%). One sample positive for both DEN-1 and DEN-2.	ProMED-mail (2013a,b,d,e)
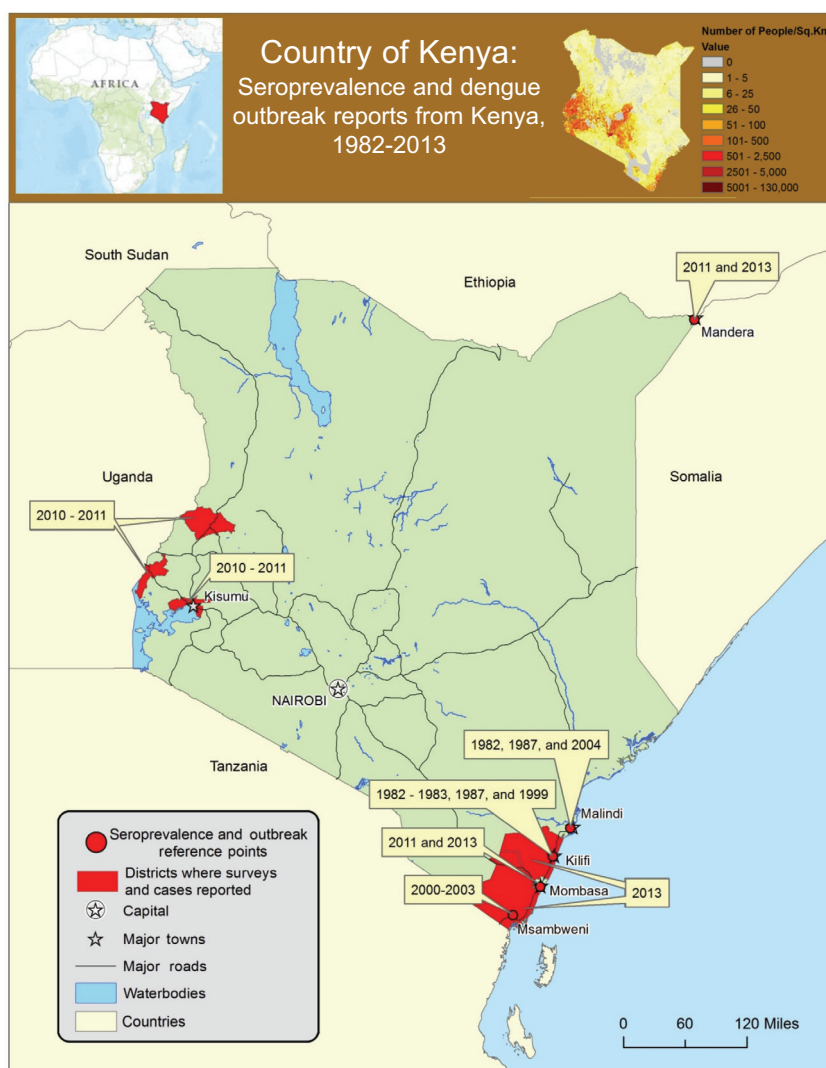
Fig. 1. Map of seroprevalence and dengue outbreak reports from Kenya, 1982-2013.

dinate system (World_Equidistant_Cylindrical) centred within the country to limit the distortion of spatial data on the edges of the map. Potential border issues were addressed by analysing the data with a 100-mile (161 km) buffer around Kenya and retesting the analysis based on the Similarity Search (ESRI, 2013a) described below.

Population density data

Human population density data were obtained from the LandScan 2011 Global Population Database, which was developed by Oak Ridge National Laboratory (ORNL) for the United States Department of Defense (Bright et al., 2011). Data obtained were in a 1 x 1 km gridded format using the ORNL methodology to derive population counts (Bright et al., 2011). Incorporating human population density provides an

essential factor in calculating risk (Rogers et al., 2006; Moffett et al., 2007; Hassan et al., 2012; Bhatt et al., 2013).

Dengue data

Information about areas in which epidemics occurred were obtained from ProMed (a website where individuals worldwide can submit reports of infectious disease outbreaks) (ISID, 1999), PubMed (a journal abstract database maintained by the U.S. National Institutes of Health) (NCBI, 1996), and the Weekly Epidemiological Report published by the World Health Organization (WHO). Kenya's infectious disease system does not compile and release reliable dengue incidence data, since diagnosis is expensive and the health information system is not set up to easily collect case reports from across the country.

Once the geographic locations of dengue cases were identified, a 1-km buffer was used to select the cells in which dengue had occurred.

Mosquito data

MosquitoMap is a repository of information that identifies locations where *Aedes aegypti* mosquitoes can survive (Foley et al., 2010; Sang et al., 2010; Nyari and Foley, 2011; Bast, 2012). In addition to the data from MosquitoMap, places within Kenya where cases had been reported in the files included in Table 1 were noted as areas that were suitable environments for *Aedes*. A MosquitoMap data portal search (http://www.mosquit-omap.org) in Kenya from 1982 through 2013 resulted in 365 survey collection sites obtained from records held and accessed through MosquitoMap from the Biosystematics Unit, Division of Entomology, Walter Reed Army Institute of Research (WRAIR), US Army Medical Research Unit-Kenya (USAMRU-K) and KEMRI (Bast, 2008; Foley and Wilkerson, 2008; Sang et al., 2010). The 365 collection sites were combined into a single geospatial layer, which resulted in 167 1 x 1 km cells where *Aedes* were collected within Kenya. There was no testing of the mosquitoes to determine if they were positive for any of the DENV strains. Previous research has demonstrated the value of museum mosquito collection records for understanding mosquito biogeography and ecology and for planning mosquito surveys (Foley et al., 2008).

Elevation data

The WorldClim Elevation database contains additional high-resolution information useful for predicting dengue range, which allowed for environmental analysis at a resolution of 1 x 1 km (Hijmans et al., 2005). Elevation data were incorporated to determine whether risk of dengue, as reported in Colombia, is limited above 1,500 m above the mean sea level (Hagenlocher et al., 2013).

Bioclimatic layers data

The WorldClim Bioclimatic variables database was used to derive monthly temperature and rainfall values, which are known to be related to mosquito distribution and dengue transmission (Hijmans et al., 2005; Rogers et al., 2006; Brady et al., 2012; Simmons et al., 2012; Bhatt et al., 2013). From these data, interpolated global climate surfaces were produced using the ANUSPLIN-SPLINA software (Hutchinson, 1995). The bioclimatic variables used for this analysis were selected based on published literature about dengue risk factors (Nyari et al., 2011) and previous dengue models (Rogers, 2006; Rogers et al., 2006; Brady et al., 2012; Simmons et al., 2012; Bhatt et al., 2013).

Table 2 summarises all of the Bioclimatic variables considered for inclusion in the model. Based on previous analyses by Foley and Nyari, the variables selected for this analysis were BIO1, BIO2, BIO4, BIO5, BIO6, BIO9, BIO12, BIO13, BIO14, BIO15, BIO19 and altitude (Foley et al., 2008, 2010, 2012; Nyari et al., 2011). We included the LandScan population data to enhance the model by integrating population data into the Similarity Search equation (Bright et al., 2011). This produced a more localised analysis at a higher resolution than previous studies, which used a more global scale in their analysis of dengue risk (Rogers, 2006; Rogers et al., 2006; Brady et al., 2012; Simmons et al., 2012; Bhatt et al., 2013).

Table 2. Definitions of the 19 bioclimatic variables (Hijmans et al., 2005).

Layer	Description	Layer	Description
BIO1	Annual mean temperature	BIO10 *	Mean temperature of warmest quarter
BIO2	Mean diurnal range	BIO11 *	Mean temperature of coldest quarter
	(mean of monthly (max temp - min temp))	BIO12	Annual precipitation
BIO3 *	Isothermality (BIO2/BIO7) (* 100)	BIO13	Precipitation of wettest month
BIO4	Temperature seasonality (standard deviation *100)	BIO14	Precipitation of driest month
BIO5 *	Max temperature of warmest month	BIO15	Precipitation seasonality (coefficient of variation)
BIO6	Min temperature of coldest month	BIO16 *	Precipitation of wettest quarter
BIO7 *	Temperature annual range (BIO5-BIO6)	BIO17 *	Precipitation of driest quarter
BIO8 *	Mean temperature of wettest quarter	BIO18 *	Precipitation of warmest quarter
BIO9	Mean temperature of driest quarter	BIO19	Precipitation of coldest quarter

* Not used in Similarity Search index.

Analysis

The various types of data listed in the previous sections were transformed from 1 x 1 km grid cells to a point layer so as to bring all the data together into a single dataset, rather than multiple raster layers. This was done so that we could use the newly developed Similarity Search tool in ArcGIS 10.2 software (ESRI, 2013a). Similarity Search allows users to find points of interest by resolving a similarity query such as for example "find the 10 most similar cities to Los Angeles." ESRI's Similarity Search categorises the results in an index from most similar to least similar (ESRI, 2013a). The different variables used in the analysis were standardised prior to the analysis. Standardising the attribute values utilised a z-transform where each value (X) is subtracted from the mean for all values ($\bar{X}$) and divided by the standard deviation for all values ($\bar{X}$) (ESRI, 2013b):

$$Z\text{-}transform = \frac{X - mean\ (\bar{X})}{Standard\ deviation\ (\bar{X})}$$

(equation 1)

The analysis then sums the differences between each feature for all variables against the composite for matching all features (MPX). Then these sums are ranked. Models were run with the 11 Bioclimatic (BioClim) variables listed above, population density and elevation (ALT) (Hijmans et al., 2005).

As noted in the methods section, the presence cells are locations where *Aedes* mosquitos were collected through surveys, or where cases of human dengue had been reported. The 365 mosquito points in the MosquitoMap database were reduced to 167 cells after adjusting for the duplicate points within the same 1 x 1 km cells. Additional presence cells were marked in the four regions (Mombasa, Mandera, Malindi and Kilifi) of the country where major dengue outbreaks had been reported in articles indexed by PubMed (Table 1). In total, 287 presence cells (out of 681,442) were identified as representing areas at risk for dengue. The geospatial dengue outbreak parameters included a 1-km buffer around the urban cities where dengue was reported in the PubMed sources. From these cells, the spatial parameters were modified to extract pertinent information to the centroid of the cell so the adjusted data set could be used within the Similarity Search process. The regions most susceptible to dengue were identified using the Similarity Search equation:

$$cosine\ similarity\ index = \frac{\sum_{i=1}^{n} A_i\ B_i}{\sqrt{\sum_{i=1}^{n}\ (A_i)^2}\ \sqrt{\sum_{i=1}^{n}\ (B_i)^2}}$$

(equation 2)

This equation calculates two vectors of attributes *A* and *B* using cosine similarity mathematics, *cos(θ)*. The cosine similarity index measures the increasing difference as the index (similarity) decreases. Using this method, two different analyses were used to create the maps in Figs. 2 and 6. The purpose of the first analysis was to determine how risk could be visualised with limited dengue data. This analysis created an average based on the entire set of presence cells for each of the variables and then compared those average values as the composite target feature (MPX) (Table 3). The second analysis visualised more detailed risk by incorporating isolated presence cells locations and the unique variable values at each location into the final analysis. This analysis involved running the Similarity Search against each of the presence cells, iteratively and then creating an average rank based on the rank value of each cell for each iteration.

For the first analysis, the attributes (consisting of 13 variables) for each candidate cell were compared to the vector of standardised attributes for the composite target feature based on the 287 presence cells (Fig. 3). All variables in the composite target feature table (Table 3) are weighted equally even though the resulting values are different (as seen in Fig. 3). Once all of the 681,442 candidate cells had been processed, candidates were ranked from the smallest cosine similarity index (most similar) to the largest index (least similar). The tabular graphic displays how two points (PT1 and PT2) are compared to the variable values on which Similarity Search is based (MPX) for all 13 variables values and not on one specific value (Fig. 4). This method is not based on the magnitude of the variables but focuses on the relationships between the attributes once they have been compared to the target feature attributes (ESRI, 2013b). The resulting similarity ranges from –1 (which suggests that the candidate cell is perfectly, inversely related to the profile of presence cells) to +1 (which suggests that the candidate cell is perfectly, directly related to the profile of presence cells) with zero usually indicating independence and in-between values indicating intermediate similarity or dissimilarity. Fig. 5 shows the methodological steps used for developing the dengue risk maps for Kenya.

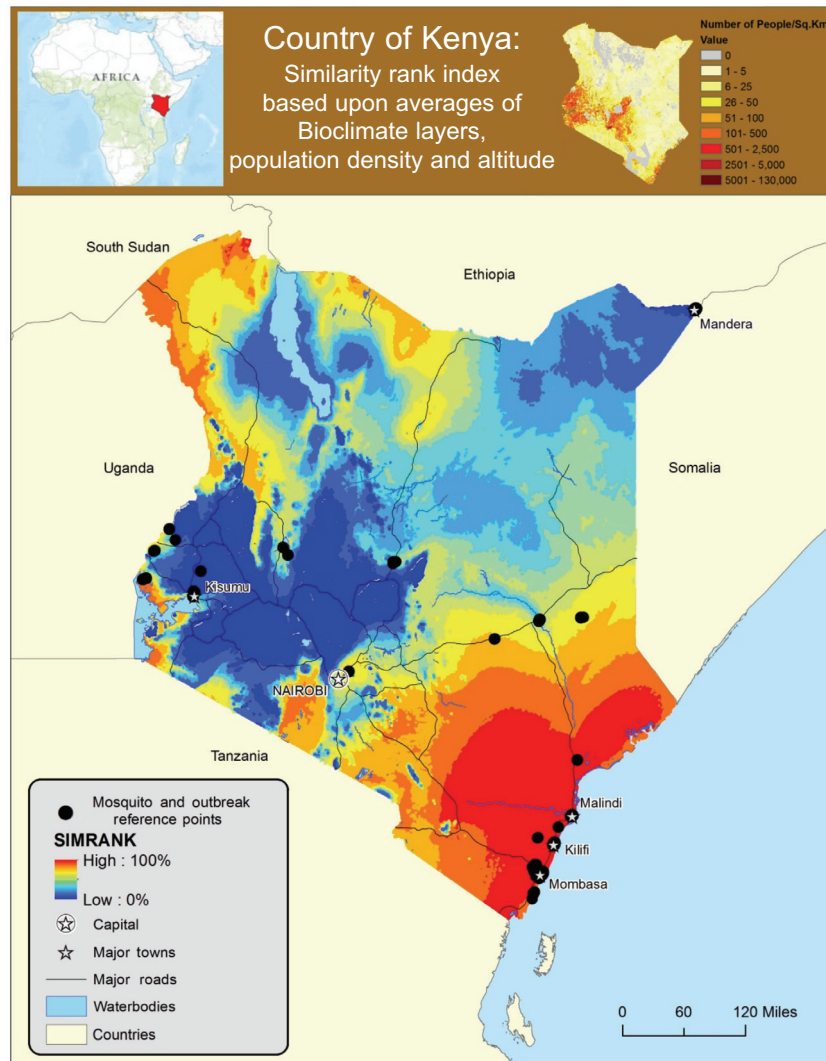The second analysis (Fig. 6) was run using each indi-

Fig. 2. Risk map created using the averages of all 287 reference points comparing bioclimatic variables, population density and altitude.

Table 3. Calculated results from the composite for matching all features when performing Similarity Search (287 cells).

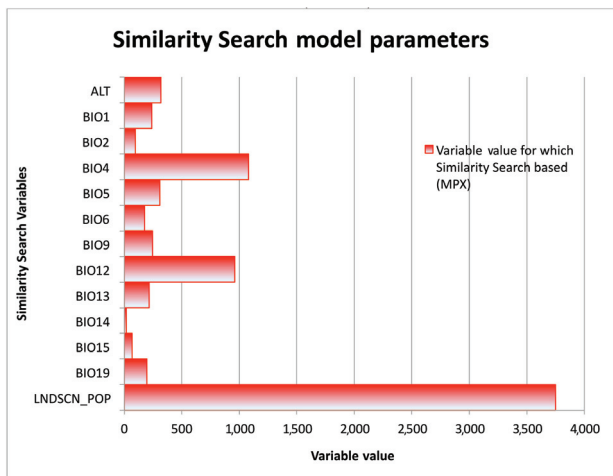| Field	Min	Max	Mean	Variable value used to calculate the SimRank for all cells
ALT	0	1,621	494.4	318.2
BIO1	0	292	68.8	238.8
BIO2	0	169	40.5	96.3
BIO4	0	1,500	425.2	1080.0
BIO5	0	401	87.9	308.5
BIO6	0	213	55.1	176.3
BIO9	0	285	71.6	245.0
BIO12	0	1,949	432.7	960.2
BIO13	0	327	91.9	215.8
BIO14	0	69	16.3	17.2
BIO15	0	125	28.3	68.0
BIO19	0	551	115.0	196.3
LNDSCN_POP	0	56,657	8246.0	3749.7

Fig. 3. Visual results of the composite for matching all features when performing Similarity Search (287 cells).

vidual presence cell (the 287 total presence cells, each with 13 variables) against all the other 681,442 cells. We combined all 287 iterations, each possessing 681,442 cells which were ranked from 1 to 681,442 in each iteration. In total, this constituted over 195,573,854 cells ranked based upon 2,542,460,102 variables and then combined into one layer using the union tool within ArcGIS. This analysis allowed the conditions existing at each presence location to be ranked against all other cells in the country. It also lim-
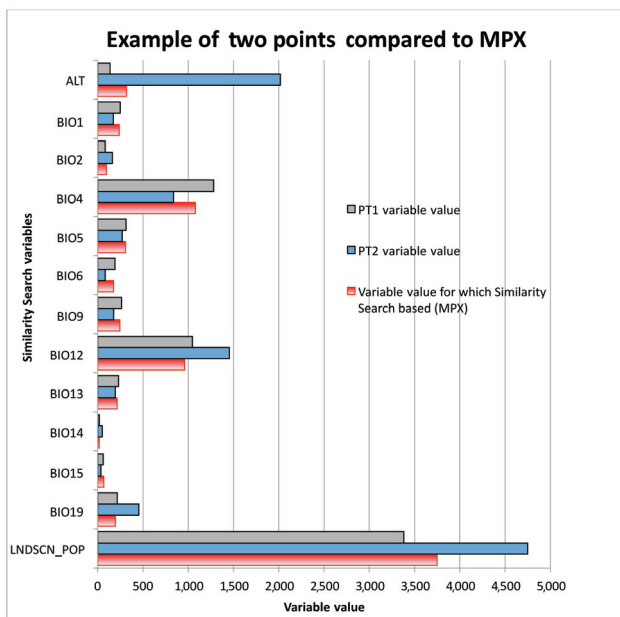
Fig. 4. Sample of processing completed to determine Similarity Search between two points (PT1 and PT2) compared to the composite for matching all features (MPX).

ited the spatial bias incurred from only using presence points (Hijmans, 2012). Performing the analysis using this methodology allowed for locations such as Mandera (upper north-eastern region of Kenya) to have a higher input into the final analysis.

Results

Fig. 7 visually compares the relationship between dengue high-risk and population density for both the average risk map (Fig. 2) and the individually combined risk map (Fig. 6). Our results indicate that not all areas with high population density in Kenya carry a high dengue risk. Many locations of low risk for dengue (including the mountainous western portion of Kenya) are also high-density population regions. The risk for dengue in major towns, such as Nairobi and Kisumu, is not high (<80%) based on the prevailing climate conditions, while locations along the Kenyan border to Uganda and South Sudan are at high risk though located in low-population regions.

The average risk map (Fig. 2) and the individually combined risk map (Fig. 6) created with this Similarity Search method indicate that dengue risk is highest along the coast and in the central areas of Kenya, which are home to nearly 15 million people and a primary location for tourism. These areas (indicated in red on the map) are the most cost-effective locations to target for dengue prevention activities such as vector control.

Table 4 shows the number of outbreak points (287 cells) within the two risk maps (Figs. 2 and 6) with <70% predicted risk range of dengue indicated by yellow-green-blue colours, 70-80% risk by orange, 80-90% risk by reddish orange and 90-100% by red.

The risk maps also show that the risk of dengue outbreaks is minimal in the mountainous areas in the centre and west of Kenya. While previous research has reported a decrease for dengue risk in Colombia above 1,500 m above the mean sea level (Hagenlocher et al., 2013), we found that an elevation >1,800 m is needed to limit dengue risk in Kenya. Determining the highest risk areas identifies the locations where more centralised and focused dengue prevention and control activities efforts can be prioritised in combating dengue.

Discussion

This paper demonstrates the utility of the Similarity Search function for ranking and visualizing locations with similar risk profiles. Using this tool, we created a risk profile for dengue fever in Kenya based on the environmental characteristics of a small number of
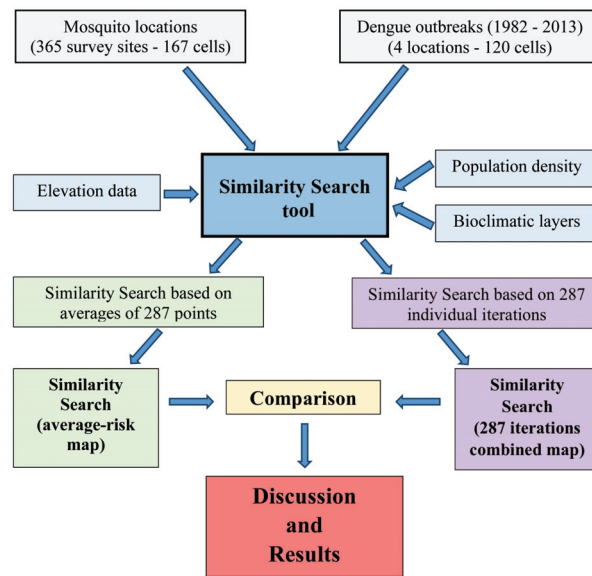
Fig. 5. Graphic view of systematic cycle completed using Similarity Search.
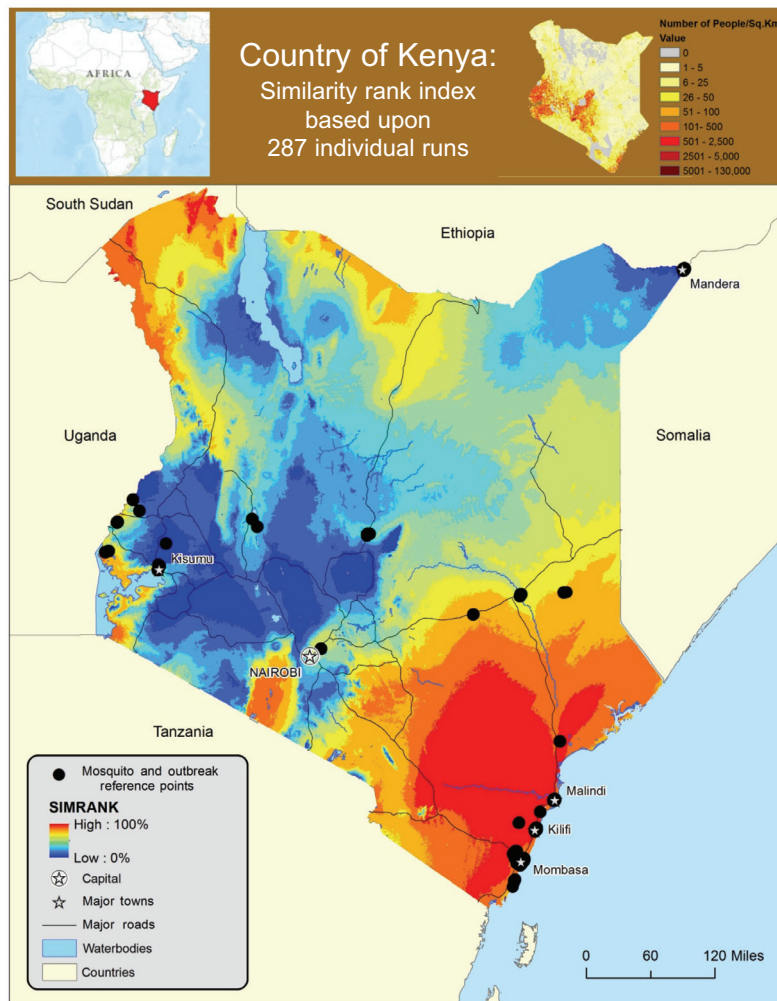

Fig. 6. Risk map created using individual iterations of all 287 reference points comparing bioclimatic variables, population density and altitude.
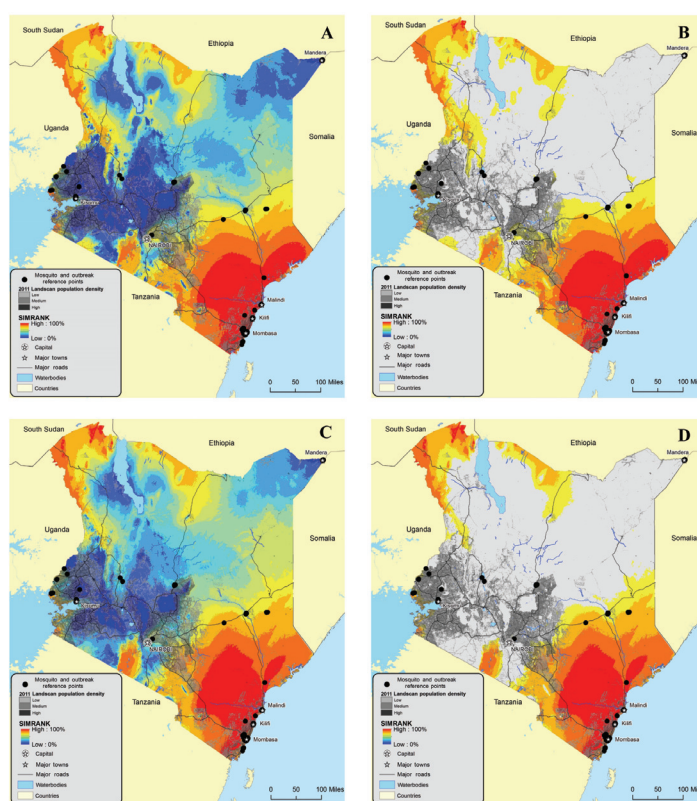
Fig. 7. Average-risk map (A) and individual iterations risk map (B) compared with landscan population density overlaid on both maps (C and D).

1 km² cells that were known to be at high risk of dengue. We then used Similarity Search to identify other cells that are likely to have high dengue risk. Our model incorporates data about human demography and health, mosquito distributions and a variety of social-environmental factors to create a more complex and larger-scale map of risk at the country level than has previously been possible.

Previous vector disease models emphasise human population density as an essential factor in calculating risk (Moffett et al., 2007). Our research reinforces that greater population density equates to a higher risk of

contracting disease as shown by numerous authors (Hales et al., 2002; Rogers, 2006; Rogers et al., 2006; Brady et al., 2012; Simmons et al., 2012; Bhatt et al., 2013). Fig. 1 confirms that outbreaks are more likely to be detected in areas characterised by high population densities and in areas close to major hospitals as opposed to rural areas that are far from health services.

Geographic location and climate are important when determining dengue risk. Additional variables such as population density provide the means for accentuating risk potential. The risk maps (Figs. 2 and 6) demonstrate how variables (including population density, pre-

Table 4. Dengue outbreak points (287 cells) that fall within the Similarity Search risk maps.

	Outbreak locations that fall within:				
Risk map	<70%	70-80%	80-90%	90-100%	Total
Average iteration (Fig. 2 based on values in Fig. 3)	0	0	0	287	287
All individual iterations (287 combined - Fig. 6)	94	0	24	169	287

cipitation, and temperature) utilising the Similarity Search calculation provide essential modelling approaches for specific regions of the world impacted by dengue. Further models can build upon these calculations for a more specific regional analysis. Since we already knew, from the literature, the various environmental factors most important to dengue risk modelling (Rogers, 2006; Rogers et al., 2006; Brady et al., 2012; Simmons et al., 2012; Bhatt et al., 2013), we were able to incorporate this information directly into the model calculations. The incorporation of the relationships between temperature, precipitation and population density through standardising all the variables in the analysis, translates into a new view of risk mapping.

The advantages of Similarity Search include the power that comes with defining a target based on the average characteristics of hundreds of presence cells, the utilisation of a method previously confined to textual searches (Bayardo et al., 2007; Adelfio et al., 2011), the provision of greater flexibility in testing variables and the ranking of the attribute values to the target features by most similar to least similar. The main limitation of Similarity Search is the lengthy computer processing required for the completion of the individual cell analysis (used to produce the map shown in Fig. 6). If a hierarchical computer process method is not used, the analysis can take days to complete. The use of an iterative process is often necessary to prevent the computer from running out of memory, but this has the disadvantage of again increasing the processing time. Nevertheless, the high-resolution local analysis allowed by Similarity Search makes this a useful tool for identifying the relationships between the attributes and the target feature.

The primary limitations for this model relate to incomplete surveillance and mosquito data available for most cities, towns and rural areas within Kenya and the need for more frequent updates of the environmental data. Human health data are dependent on an accurate diagnosis of dengue through testing, surveillance and dissemination of the data. A particular concern is that most case reports come from urban areas where surveillance and reporting systems are more developed, thus skewing the maps towards assuming higher risk in urban areas than is appropriate. Given the lack of nationwide case finding efforts, absence of data about cases from much of the country cannot be considered evidence of there being no cases in those regions. Mosquito habitats vary due to climate change, making current data essential for accurate mapping and reporting. The need for updated mosquito data will become especially important if global cli-

mate changes alter the range of *Aedes* mosquitoes (Patz et al., 1998; Hales et al., 2002). Global climate change may mean that environmental data quickly become inaccurate and this will constrain the ability to predict vector-borne disease hotspots based on similarity searches and limited human health data. Ideally, environmental data must be updated often enough to allow for adjustments to models based on floods, droughts and urbanization (or other patterns of human migration). They must also exist at a sufficient scale for use in both rural and urban areas, e.g. one risk factor for dengue is the presence of standing water which has collected in old tires or small cans (WHO, 2012), and this is neither captured by current databases nor by 1-km^2 resolution maps.

We used Similarity Search as a tool for a specific purpose, but the possibility of application in other areas is very open. Comparison of cities based upon social media preference (Seth et al., 2011), the geographic locality of cuisine (Zhu et al., 2013) and calculating similarity between user-generated information about locations (Scheffler et al., 2012) provide real world examples utilising cosine similarity mathematics. The prediction maps created with Similarity Search can provide critical information for Government officials, businesses, epidemiologists and others involved in planning and preparing for future events and making decisions about resource dissemination and public outreach.

Conclusion

The dengue risk maps for Kenya presented here provide additional information for health officials by indicating parts of the country where intensified dengue prevention and control activities may be most beneficial for public health interventions and, conversely, the regions where resources do not need to be deployed at this time. The use of tools like Similarity Search to identify areas at high risk for vector-borne disease and other threats to health will increase the effectiveness of public health interventions by allowing them to be more precisely targeted towards the populations that will most benefit from them.

The effectiveness of the technology employed highlights its potential value for other applications both within and beyond public health. For example, Similarity Search could be used to determine areas of a city that are more vulnerable to certain types of crime based on existing crime locations, to determine economic development by benchmarking one city against its most similar counterparts, to evaluate real estate trends

and foreclosure risks, to justify salaries by looking at other areas with, for example, a similar cost of living.

References

Adelfio MD, Nutanong S, Samat H, 2011. Similarity search on a large collection of point sets. In: GIS '11: Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, November 2011, 132-141.

Akhwale W, 2013. Dengue fever outbreak response: Kenya sets up sentinel surveillance sites. Quarterly Bulletin: The East Africa Public Health Lab Network Newsletter - June 2013. Available at: http://www.eaphln-ecsahc.org/kenya/?wpfb_dl=39 (accessed on August 2014).

Awando JA, Ongus JR, Ouma C, Mwau M, 2013. Seroprevalence of anti-dengue virus 2 serocomplex antibodies in out-patients with fever visiting selected hospitals in rural parts of western Kenya in 2010-2011: a cross sectional study. Pan Afr Med J 16, 73.

Bast J, 2008. MosquitoMap - *Aedes aegypti* - Arbovirus surveillance project. Walter Reed Biosystematics Unit (WRAIR, Division of Entomology) Suitland, MD. Available at: http://www.mosquitomap.org (accessed on August 2014).

Bast J, 2012. MosquitoMap - *Aedes aegypti* - arbovirus surveillance project. Walter Reed Biosystematics Unit (WRAIR, Division of Entomology) Suitland. Available at: http://www.mosquitomap.org (accessed on August 2014).

Bayardo RJ, Ma Y, Srikant R, 2007. Scaling up all pairs similarity search. In: Proceedings of the 16th International Conference on World Wide Web (WWW '07), New York, 131-140 pp.

Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O et al., 2013. The global distribution and burden of dengue. Nature 496, 504-507.

Blaylock JM, Maranich A, Bauer K, Nyakoe N, Waitumbi J, Martinez LJ, Lynch J, 2011. The seroprevalence and seroincidence of dengue virus infection in western Kenya. Travel Med Infect Dis 9, 246-248.

Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, Hoen AG, Moyes CL, Farlow AW, Scott TW, Hay SI, 2012. Refining the global spatial limits of dengue virus transmission by evidence-based consensus. PLoS Negl Trop Dis 6, e1760.

Bright EA, Coleman PR, Rose AN, Urban ML, 2012. LandScan 2011. Oak Ridge National Laboratory, Oak Ridge. Available at: http://www.ornl.gov/landscan/ (accessed on August 2014).

ESRI, 2013a. ArcGIS Help 10.2 - Similarity Search. Redlands: Environmental Systems Research Institute. Available at: http://resources.arcgis.com/en/help/main/10.2/index.html#/Similarity_Search/005p0000005m000000/ (accessed on August 2014).

ESRI, 2013b. ArcGIS Help 10.2 - how Similarity Search works. Redlands: Environmental Systems Research Institute. Available at: http://resources.arcgis.com/en/help/main/10.2/index.html#//005p00000059000000/ (accessed on August 2014).

Foley D, Wilkerson R, 2008. MosquitoMap - *Aedes aegypti*. Walter Reed Biosystematics Unit (WRAIR, Division of Entomology) Suitland. Available at: http://www.mosquito-map.org (accessed on August 2014).

Foley DH, Weitzman AL, Miller SE, Faran ME, Rueda LM, Wilkerson RC, 2008. The value of georeferenced collection records for predicting patterns of mosquito species richness and endemism in the Neotropics. Ecol Entomol 33, 12-23.

Foley DH, Wilkerson RC, Birney I, Harrison S, Christensen J, Rueda LM, 2010. MosquitoMap and the Mal-area calculator: new web tools to relate mosquito species distribution with vector borne disease. Int J Health Geogr 9, 11.

Foley DH, Wilkerson RC, Dornak LL, Pecor DB, Nyari AS, Rueda LM, Long LS, Richardson JH, 2012. SandflyMap: leveraging spatial data on sand fly vector distribution for disease risk assessments. Geospat Health 6, S25-S30.

Franco L, Di Caro A, Carletti F, Vapalahti O, Renaudat C, Zeller H, Tenorio A, 2010. Recent expansion of dengue virus serotype 3 in West Africa. Euro Surveill 15, 1-4.

Global Administrative Areas, 2012. GADM database of global administrative areas, version 2.0. Available at: www.gadm.org (accessed on August 2014).

Gubler DJ, 2004. The changing epidemiology of yellow fever and dengue, 1900 to 2003: full circle? Comp Immunol Microbiol Infect Dis 27, 319-330.

Guzman A, Istúriz RE, 2010. Update on the global spread of dengue. Int J Antimicrob Agents 36, S40-S42.

Hagenlocher M, Delmelle E, Casas I, Keinberger S, 2013. Assessing socioeconomic vulnerability to dengue fever in Cali, Colombia: statistical vs expert-based modeling. Int J Health Geogr 12, 36.

Hales S, de Wet N, Maindonald J, Woodward A, 2002. Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. Lancet 360, 830-834.

Hassan H, Shohaimi S, Hasim NR, 2012. Risk mapping of

dengue in Selangor and Kuala Lumpur, Malaysia. Geospat Health, 7, 21-25.

Hijmans RJ, 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology 93, 679-688.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A, 2005. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25, 1965-1978.

Hutchinson MF, 1995. Interpolating mean rainfall using thin-plate smoothing splines. Int J Geogr Inf Sci 9, 385-403.

International Society for Infectious Diseases (ISID), 1999. ProMED-mail website. Available at: http://www.promed-mail.org (accessed on August 2014).

Johnson BK, Musoke S, Ocheng D, Gighogo A, Rees PH, 1982a. Dengue-2 virus in Kenya. Lancet 2, 208-209.

Johnson BK, Ocheng D, Gichogo A, Okiro M, Libondo D, Kinyanjui P, Tukei PM, 1982b. Epidemic dengue fever caused by dengue type 2 in Kenya: preliminary results of human urological and serological studies. East Afr Med J 59, 781-784.

Johnson BK, Okath F, Tukei PM, Mugambi M, Woody JN, Morrill JC, Hyams KC, 1990. Dengue-2 virus in Kenya. Lancet 336, 1071.

Kenya Central Bureau of Statistics, 2005. Geographic dimensions of well-being in Kenya: who and where are the poor? A constituency level profile - volume 2. The Regal Press Kenya Ltd, 7-13 pp.

Mease LE, Coldren RL, Musila LA, Prosser T, Ogolla F, Ofula VO, Schoepp RJ, Rossi CA, Adungo N, 2011. Seroprevalence and distribution of arboviral infections among rural Kenyan adults: a cross-sectional study. Virol J 8, 371.

Moffett A, Shackelford N, Sarkar S, 2007. Malaria in Africa: vector species' niche models and relative risk maps. PLoS One 2, e824.

National Center for Biotechnology Information (NCBI), 1996. PubMed website. Available at: http://www.ncbi.nlm.nih.gov/pubmed (accessed on August 2014).

Nyari A, Foley DH, 2011. *Aedes aegypti* Nyari. Walter Reed Biosystematics Unit (WRAIR, Division of Entomology) Suitland. Available at: http://www.mosquitomap.org/ metadata.htm#vec31 (accessed on August 2014).

Patz JA, Martens WJM, Focks DA, Jetten TH, 1998. Dengue fever epidemic potential as projected by general circulation models of global climate change. Environ Health Perspect 106, 147-153.

ProMED-mail, 2011. Dengue/DHF update 2011 (39). 20111004.2985. Available at: http://www.promedmail.org (accessed on August 2014).

ProMED-mail, 2013a. Dengue - Kenya (02): (Coastal). 20130325.311131. Available at: http://www.promedmail.org (accessed on August 2014).

ProMED-mail, 2013b. Dengue - Kenya (03): (Coast).

20130428.1676860. Available at: http://www.promedmail.org (accessed on August 2014).

ProMED-mail, 2013c. Dengue - Kenya: (North Eastern). 20130221.304287. Available at: http://www.promedmail.org (accessed on August 2014).

ProMED-mail, 2013d. Dengue/DHF update (24): Asia, Africa. 20130325.1601484. Available at: http://www.promedmail.org (accessed on August 2014).

ProMED-mail, 2013e. Dengue/DHF update (33): Asia, Africa, Pacific. 20130428.1676860. Available at: http://www.promedmail.org (accessed on August 2014).

ProMED-mail, 2013f. Dengue/DHF update (16): Asia, Pacific, Africa. 20130225.1558046. Available at: http://www.promed-mail.org (accessed on August 2014).

Rogers DJ, 2006. Predicted probability occurrence for dengue: models for vectors and vector-borne diseases. Adv Parasitol 62, 1-35.

Rogers DJ, Wilson AJ, Hay SI, Graham AJ, 2006. The global distribution of yellow fever and dengue. Adv Parasitol 62, 181-220.

Sang R, Kioko E, Lutomiah J, Warigia M, Ochieng C, O'Guinn M, Lee JS, Koka H, Godsey M, Hoel D et al., 2010. Rift Valley fever virus epidemic in Kenya, 2006/2007: the entomologic investigations. Am J Trop Med Hyg 83, 28-37.

Sang RC, 2007. Dengue in Africa. In: Report of the scientific working group meeting on dengue. Geneva: WHO Special Programme for Research and Training in Tropical Diseases, 50-52 pp.

Sang RC, Dunster LM, 2001. The growing threat of arbovirus transmission and outbreaks in Kenya: a review. East Afr Med J 78, 655-661.

Scheffler T, Schirru R, Lehmann P, 2012. Matching points of interest from different social networking sites. In: KI 2012: Advances in artificial intelligence. Glimm B, Krüger A (eds). Berlin: Springer, 245-248 pp.

Seth R, Covell M, Ravichandran D, Sivakumar D, Baluja S, 2011. A tale of two (similar) cities: inferring city similarity through geo–spatial query log analysis. In: Proceedings of the International Conference on Knowledge Discovery and Information Retrieval. Paris: KDIR, 179-189 pp.

Simmons CP, Farrar JJ, Chau N, Wills B, 2012. Dengue. N Engl J Med 366, 1423-1432.

Sutherland LJ, Cash AA, Huang YJS, Sang RC, Malhotra I, Moormann AM, King CL, Weaver SC, King CH, LaBeaud AD, 2011. Serologic evidence of arboviral infections among humans in Kenya. Am J Trop Med Hyg 85, 158-161.

WHO, 2012. Global strategy for dengue prevention and control 2012-2020, World Health Organization Press, V-30 pp.

Zhu YX, Huang J, Zhang ZK, Zhang QM, Zhou T, Ahn YY, 2013. Geography and similarity of regional cuisines in China. PLoS One 8, e79161.