

## **Electronic Supplementary Material**

### **1. Supplementary Methods**

Each of the 31 predictor variables was analyzed at a spatial resolution of 30 arc-seconds, which is approximately 1 km<sup>2</sup> in Sankuru (longitude: 21° 40' 10" – 25° 11' 41" E, latitude: 1° 41' 30" - 5° 57' 1" S).

#### **1.1. Climatic data**

Climatic variables were selected for inclusion in this study because previous work hypothesized that precipitation was important for predicting monkeypox incidence (Levine et al. 2007). Our climatic data included annual means of precipitation and temperature as well as seasonal extremes, all of which are important determinants of biodiversity (Nix 1986; Farber and Kadmon 2003) (see Table 4 of this document). The climatic data was constructed via interpolation from a 50 year time series of weather station records with a thin-plate spline that incorporates elevation, latitude, and longitude as covariates (Hijmans et al. 2005). The weather stations include such sources as the Global Historical Climatology Network (GHCN), the United Nations Food and Agricultural Organization (FAO), the World Meteorological Organization (WMO), the International Center for Tropical Agriculture (CIAT), and additional country-based station networks (Buermann et al. 2008). These climate data have been applied previously to predict the geographic distributions of other viruses (e.g. Calvete et al. 2008, 2009; Schafer and Lundstrom 2009; Ward et al. 2009).

#### **1.2. Human population data**

Our analysis included human population density in order to test the hypothesis that population is a proxy for what people eat and hunt and the rate of contact between people and the wildlife reservoirs of monkeypox. We hypothesized that per capita monkeypox would go up with

decreasing population density because the rate of contact between people and wildlife would be greater in rural areas.

We incorporated estimates of Sankuru's population based on the LandScan model, which utilizes satellite data from the year 2008 (such as night-time lights) to identify human settlements and infers population density at the 1 km<sup>2</sup> resolution by extrapolating from the best available census data (Dobson et al. 2000; Vijayaraj et al. 2007). For details on acquiring the LandScan data set, see <http://www.ornl.gov/sci/landscan/>. These population data have seen frequent use in modelling the geographic distributions of infectious diseases (e.g. Tatem and Hay 2004; Peterson 2009). LandScan estimates the number of people that travel through an area in a given day in addition to the area's night-time residents. As a result, the population estimates used here may be somewhat larger than models that are only based on night-time residents. Nevertheless, LandScan is considered one of the most accurate and widely-utilized global population data sets (Potere et al. 2009).

Population density was not included in our regression model insofar as the first three principal components did not assign a large weight to population. Thus, we could not quantitatively assess the hypothesis that per capita monkeypox is greater in areas with low population density. However, we incorporated population into the analysis qualitatively by creating a map of the population density of Sankuru overlaid on a map of the probability that site is in the ecological niche of the virus (Figure 5(c) of the main text).

### **1.3. Vegetation data**

We utilized seven variables from a variety of satellite sensors that measure the properties of forest vegetation including canopy closure (Hansen et al. 2002), canopy roughness (Long et al. 2001), deciduousness, greenness (Huete et al. 2006), leaf area (Myneni et al. 2002), and

normalized difference vegetation index (see Table 5 of this document). Previous work has demonstrated a high correlation between ground-based measurements of forest properties and these remote-sensed variables (for details, see Saatchi et al. 2007; Buermann et al. 2008). Our vegetation data were constructed from satellite images in publicly-available databases such as the NASA Warehouse Inventory Search Tool (WIST; <https://wist.echo.nasa.gov/~wist/api/imswelcome/>) and the NASA Scatterometer Climate Record Pathfinder (<http://www.scp.byu.edu/>). The vegetation variables and the other predictor variables used in this analysis are available upon request.

Sites that were predicted to be in the rope squirrel's ecological niche had a median canopy closure of 70% and high values for the leaf area index (5.5) and radar backscatter (-7.6 dB), indicating that typical rope squirrel habitat is comprised of dense secondary forest. Sites predicted to be in the ecological niche of the monkeypox virus had significantly greater forest canopy cover, leaf area index, and radar backscatter than sites not predicted to be suitable for monkeypox ( $t_{\text{canopy cover}}=-337.37$ ,  $t_{\text{leaf area index}}=-264.24$ ,  $t_{\text{backscatter}} = -139.3$ , in each case  $p < 2.2 \times 10^{-16}$ ). Sites with high canopy cover, leaf area index, and backscatter generally represent primary forests or relatively dense secondary forests.

#### **1.4. Monkeypox reservoir data**

We constructed ecological niche models for four rodents hypothesized to be monkeypox reservoirs: rope squirrels (*Funisciurus* species, family Scuridae), African dormice (two species: *Graphiurus crassicaudatus* and *G. lorraineus*, family Gliridae), and giant pouched rats (*Cricetomys* species, family Nesomyidae). These taxa were selected due to the fact that molecular tests have confirmed that these four taxa were infected with the monkeypox virus during the US outbreak (Hutson et al. 2007) and for the reasons elaborated in the Introduction of

the main text, such as the goal of assessing whether terrestrial rodents are important reservoirs in West Africa but not Central Africa. We obtained data on the occurrences of the four reservoirs from the Mammal Networked Information System (MaNIS) (Stein and Wieczorek 2004; Wieczorek et al. 2004) and a recent monograph on *Graphiurus* dormice (Holden and Levine 2009). These occurrences represent geo-referenced specimens from museum collections.

To predict the geographic distribution of the reservoir species, we utilized eleven predictor variables that measured climate, land cover, and topography. We selected these predictor variables because they were used in a previous study that estimated the geographic distribution of African dormice (Holden and Levine 2009) and we wanted to validate our models through comparison with that study. Six of these variables were also used to predict monkeypox occurrence: annual mean temperature, elevation, maximum temperature of the warmest month, mean diurnal range, minimum temperature of the coldest month, and total annual precipitation. However, five of the variables were not in the set used to model the ecological niche of monkeypox: compound topographic index (CTI), flow accumulation, flow direction, land cover, and topographic aspect. The predictor variables and the species' occurrence data served as the input for a machine-learning algorithm, Maxent (Phillips et al. 2006; Phillips and Dudik 2008), which predicted the geographic distribution of the four rodent reservoirs in Sankuru at the 1 km<sup>2</sup> scale. We selected Maxent because there were few occurrences ( $n \leq 32$ ) for each reservoir, and Maxent has been shown to perform well with little data (Wisiz et al. 2008; Williams et al. 2009). For *Funisciurus* rope squirrels and *Cricetomys* pouched rats, we modeled the distribution of the genus because the number of available occurrences of any one species in the genus was insufficient to develop an accurate model of the species' niche. We applied the workflow shown in Figure 4 of the Electronic Supplementary Material to each taxon one at a time. Table 3 of the

Electronic Supplementary Material lists the most important environmental variable for determining the ecological niche of each monkeypox reservoir. Following established protocols, the accuracy of the distributional models was assessed by using 75% of the occurrences to predict the ecological niche of the reservoir, then measuring the AUC of the resulting model on the withheld 25% of the occurrences (Pawar et al. 2007).

### **1.5. Further analysis of the second principal component (PC2)**

Our best regression model explains monkeypox occurrences as a function of rope squirrel habitat suitability and forest cover (PC2). Since PC2 is a combination of forest cover and rope squirrel habitat suitability, this raises the questions of whether rope squirrel habitat might be a significant predictor of monkeypox even without accounting for forest density and whether rope squirrel habitat is more important than dormouse or pouched rat habitat. To address these questions, we repeated the logistic regression using only rope squirrel habitat suitability, dormouse suitability, and pouched rat suitability as predictors of human monkeypox cases. In this comparison, only the rope squirrel was significant for predicting human monkeypox cases (rope squirrel:  $\chi^2 = 5.18$ ,  $p = 0.023$ ; dormouse *G. crassicaudatus*:  $\chi^2 = 0.073$ ,  $p = 0.79$ ; pouched rat:  $\chi^2 = 1.35$ ,  $p = 0.25$ ). We report a comparison between only one dormouse and the rope squirrel because the two dormice were collinear with each other but uncorrelated with the rope squirrel ( $-0.03 \leq$  Pearson's  $r \leq -0.06$ ). However, comparing both dormice to the rope squirrel yielded similar results. Thus, the rope squirrel remains a more important predictor of human monkeypox than the African dormouse or the pouched rat independent of the effects of forest cover.

PC2 explains less of the variance in the independent variables than PC1 (14.8% vs. 45%). Nevertheless, a regression model comprised of PC2 explains more of the variance in the dependent variable, human monkeypox cases, than a model comprised of PC1.

### **1.6. Validation of the regression model on new data**

The logistic regression model was validated using the library Design 2.3-0 in R 2.9 (Harrell 2009). We split the data into a test and training set then measured the accuracy of the model that was constructed from the training set by assessing how well it performed on the test set. In particular, we computed the absolute value of the difference between the predicted probabilities and the observed data (hereafter “absolute error”) with a standard technique for validating logistic regression models (Miller et al. 1991). The absolute error on the test set in 1,000 bootstrap replicates was 7%, which is sufficiently small that we can conclude that our model of monkeypox risk in Sankuru is accurate.

### **1.7. Comparison of the Maxent and logistic regression predictions about human monkeypox in Sankuru.**

We implemented a standard technique for comparing two spatial data sets, the weighted kappa test (Cohen 1960; Siegel and Castellan 1988; Richards and Jia 2006), to quantify the similarity between the logistic regression and Maxent predictions. For each of the two models, we divided the predicted probabilities of monkeypox incidence into four classes based on quantiles. Next, we constructed a confusion matrix that listed the number of pixels that were assigned to the same class by both Maxent and logistic regression. The matrix was analyzed with a weighted kappa test in which the null hypothesis is that there is no agreement between the logistic regression and Maxent predictions. The data are compatible with the alternative hypothesis of significant

agreement between the Maxent and logistic regression models (for the test statistic and  $p$ -value, see the Results section of the main text).

## Electronic Supplementary Material Figures

### Figure 1. Scatterplot of the monkeypox data in the space of the first and third principal

**components.** Figures 1 and 2 of the Supplementary Material depict the weights assigned to each variable in each PC. PC1, which is labelled “Factor 1” in the figure, represents precipitation in lowlands that contain habitat for the African dormouse. PC3 represents temperature. B1 = annual mean temperature, B2 = mean diurnal range, B3 = isothermality, B4 = temperature seasonality, B5 = maximum temperature of the warmest month, B6 = minimum temperature of the coldest month, B7 = temperature annual range, B8 = mean temperature of the wettest quarter, B9 = mean temperature of the driest quarter, B10 = mean temperature of the warmest quarter, B11 = mean temperature of the coldest quarter, B12 = annual precipitation, B13 = precipitation of the wettest month, B14 = precipitation of the driest month, B15 = precipitation seasonality, B16 = precipitation of the wettest quarter, B17 = precipitation of the driest quarter, B18 = precipitation of the warmest quarter, B19 = precipitation of the coldest quarter, CRASSI = the probability that a 1 km<sup>2</sup> site is in the ecological niche of the African dormouse *Graphiurus crassicaudatus*, CRICETO = the probability that a site is in the ecological niche of *Cricetomys* giant pouched rats, EL = elevation, FUNI = the probability that a site is in the ecological niche of *Funisciurus* rope squirrels, LOR = the probability that a site is in the ecological niche of the African dormouse *Graphiurus lorraineus*, NDVIMX = maximum normalized difference vegetation index, NDVIGR = normalized difference vegetation index of wet months, LAIMX = maximum leaf area index, POP = human population density, QSM = mean radar backscatter, which is a measure of forest canopy structure, QSSD = standard deviation of radar backscatter, VCF = forest canopy closure.



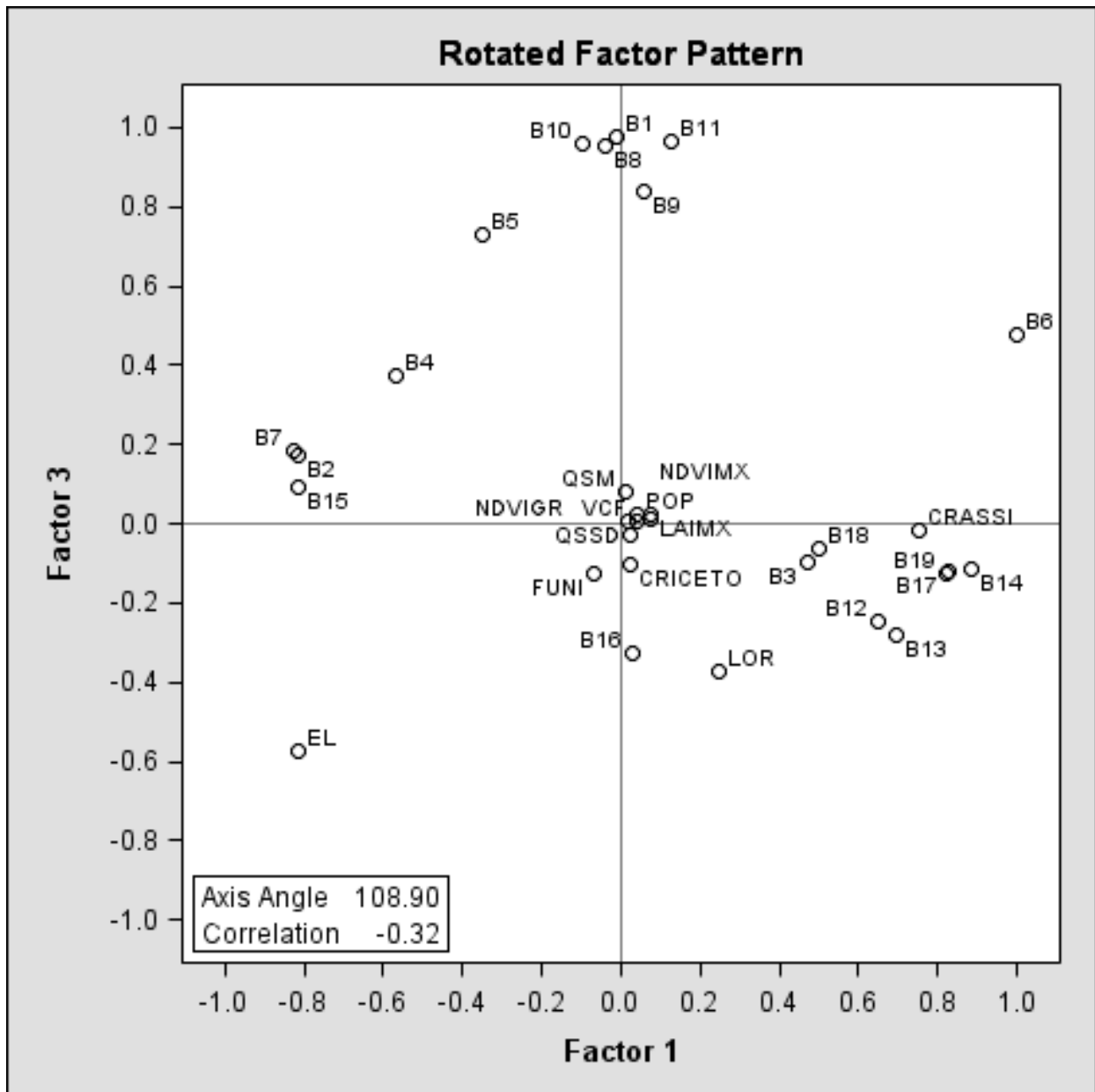
**Figure 2. Scatterplot of the monkeypox data in the space of the first and second principal components.** PC1 represents precipitation in lowlands that contain habitat for the African dormouse. PC2 represents forest vegetation in Sankuru that is in the ecological niche of *Funisciurus* rope squirrels. The abbreviations listed in the scatterplot are defined in the caption of Figure 2 of the Electronic Supplementary Material. PC1 is highly correlated with annual precipitation ( $r = 0.85$ ) and anti-correlated with elevation ( $r = -0.71$ ). PC2 correlates with the measures of forest vegetation: percent closure of the forest canopy ( $r = 0.94$ ), mean radar backscatter ( $r = 0.94$ ), leaf area index ( $r = 0.96$ ), and NDVI ( $r \geq 0.92$ ). In addition, there is a high correlation between PC2 and the probability that a site is in the ecological niche of *Funisciurus* rope squirrels ( $r = 0.88$ ). PC3 correlates with annual mean temperature ( $r = 0.99$ ).

**Figure 3. Using the epidemiological triangle to analyze human monkeypox cases.**

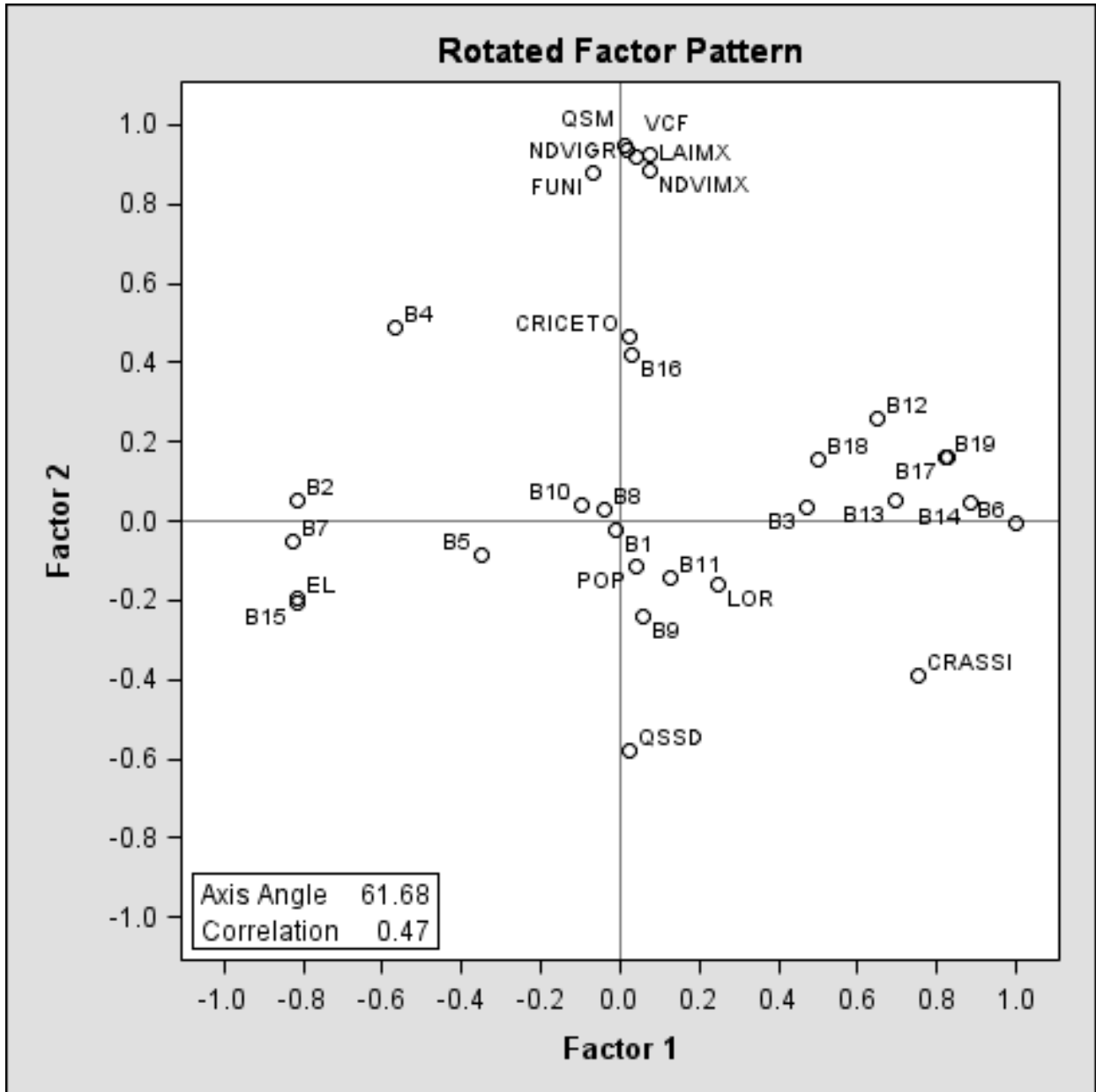
**Figure 4. Workflow for constructing a model of the ecological niche of a reservoir of monkeypox in Sankuru.** The gray box lists the 11 environmental variables that were used in the construction of a niche model for each reservoir species. “Aspect” refers to the direction that a slope faces, which affects the incident solar energy available for photosynthesis. Compound topographic index (CTI) is a measure of the tendency of a site to pool water. Flow accumulation is an estimate of how much water will flow into a 1 km<sup>2</sup> site based on a hydrological model. Flow direction measures the direction in which a river or stream would flow upon leaving a site. The environmental variables were used along with reservoir occurrence data from MaNIS and Holden & Levine (2009). The environmental variables and the occurrences of the reservoir became the input for Maxent, which constructed a model of the ecological niche of the reservoir in Sankuru at the 1 km<sup>2</sup> scale.

**Figure 5. Ecological niche models of potential monkeypox reservoir species.** These models were constructed using the workflow described in Figure 4. The models of distributions of the African dormice *G. crassicaudatus* and *G. lorraineus* are concordant with previously-published models of the ecological niches of these species in Africa (Holden and Levine 2009). Sites shown in black are predicted to be in the ecological niche of the reservoir, whereas white sites are not predicted to be part of the niche.

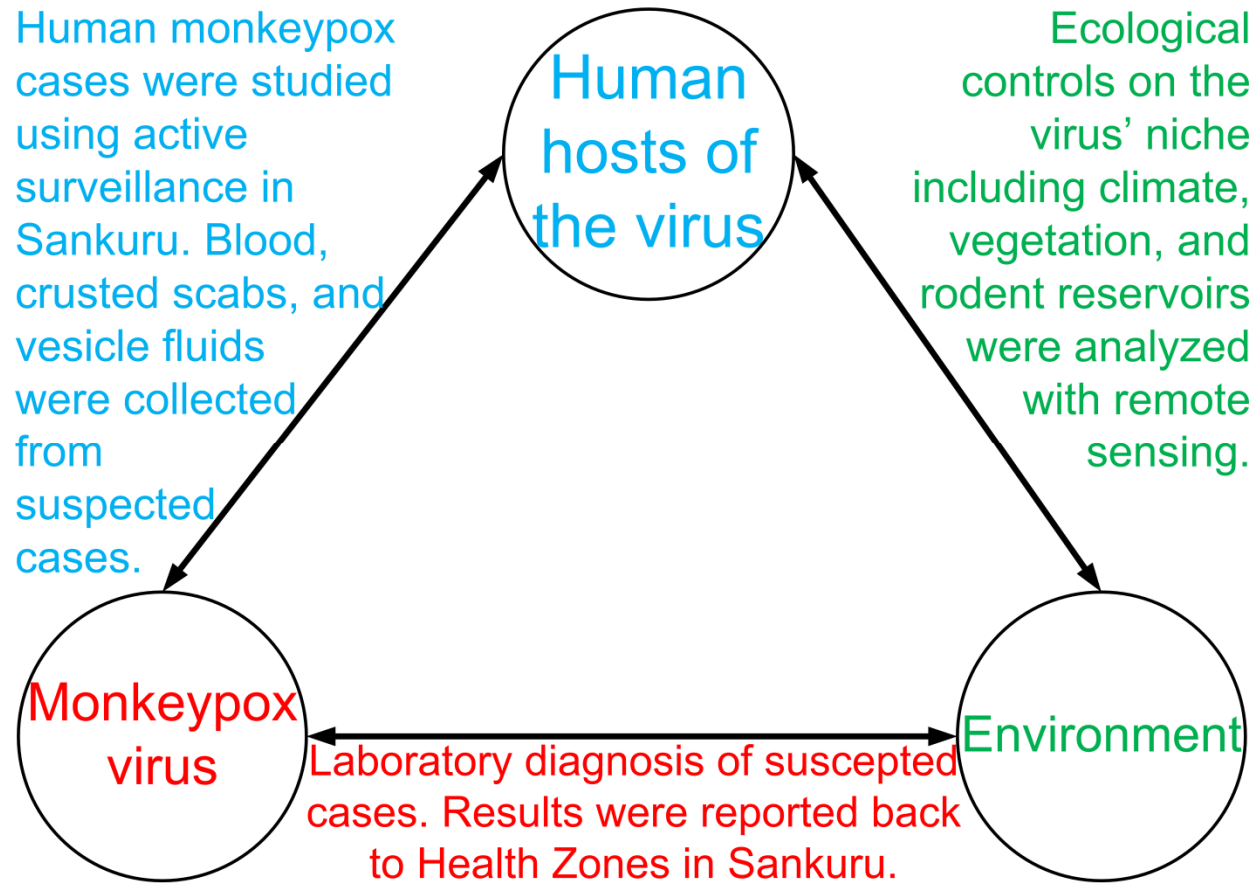
Electronic Supplementary Material Figure 1.



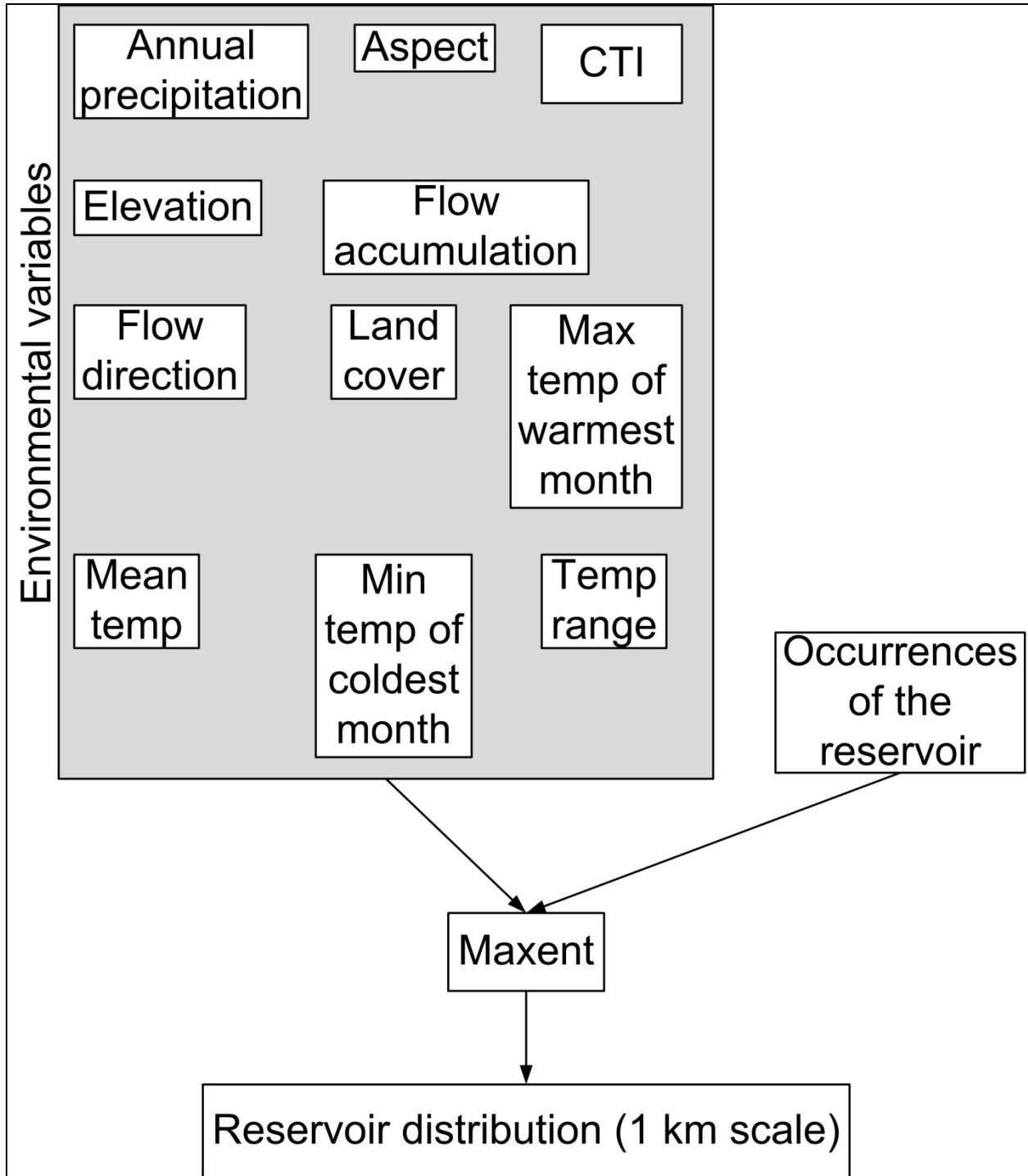
Electronic Supplementary Material Figure 2.



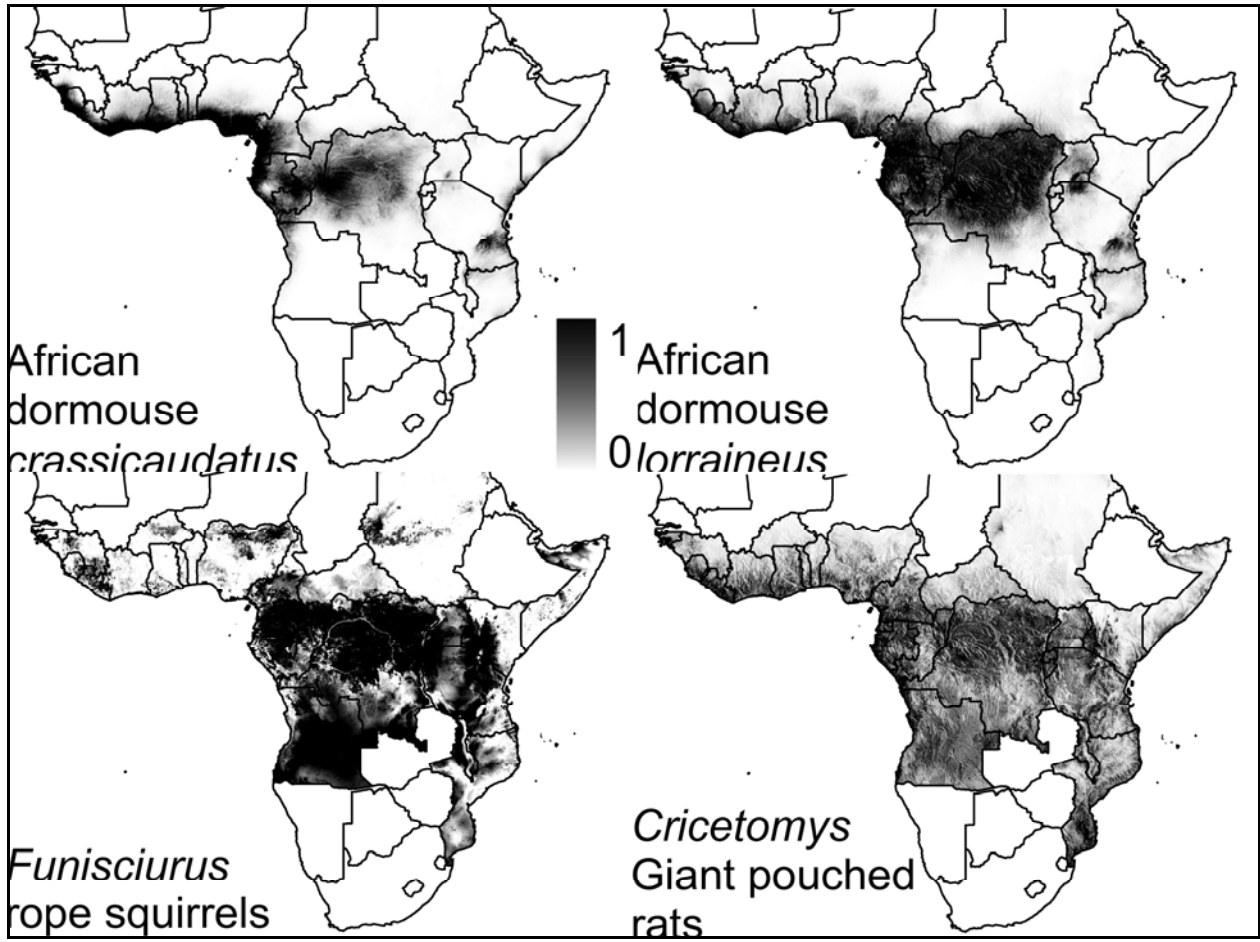
Electronic Supplementary Material Figure 3.



Electronic Supplementary Material Figure 4.



Electronic Supplementary Material Figure 5.



Electronic Supplementary Material Tables

<b>Model</b>	<b>AIC</b>	<b><math>\Delta_i</math></b>
Intercept only	1614	7
Intercept + principal component two	1607	0

**Table 1. Models analyzed during the variable selection procedure.** We utilized a standard automated variable selection procedure implemented in LOGISTIC procedure in SAS 9.2 for Windows (Montgomery et al. 2006). The procedure iteratively selects variables to be included in or removed from the regression model based on a 0.05 significance level. In the first iteration of the algorithm, only the intercept was included in the model. The second principal component was added to the model in the second iteration of the variable selection procedure. AIC = Akaike information criterion. The smaller the AIC, the better the model. The AIC selects models that are the best combination of goodness-of-fit and parsimony insofar as it penalizes models with an excessive number of terms, which may overfit the data.  $\Delta_i$  is the level of empirical support for model  $i$ , defined as the difference between the AIC of model  $i$  and the smallest AIC of any model (Burnham and Anderson 2002). When  $\Delta_i$  is less than two, there is considerable support for a model. There is considerably less support for a model when  $\Delta_i = 7$  (Burnham and Anderson 2002). Thus, the model that explains human monkeypox cases as a function of forest vegetation and habitat suitability for the rope squirrel (principal component two) has substantially more empirical support in the Sankuru data than a model comprised of only the intercept.



Variable	$w_+(j)$
Precipitation in lowlands	0.41
Habitat suitability for the rope squirrel/forest density	0.8
Temperature	0.28

**Table 2. Akaike weights for the predictor variables.** Eight possible models can be constructed from the three predictors variables that we analyzed: precipitation in lowlands, habitat suitability for the rope squirrel/forest density, and temperature. For each model, we computed the AIC and the Akaike weight,  $w_i = e^{-\frac{1}{2}\Delta_i} / \sum_{r=1}^R e^{-\frac{1}{2}\Delta_r}$ , where  $R$  represents the set of all models.  $w_i$  is interpreted as the relative likelihood of a model. Next, to assess the relative importance of the three predictors, we computed the Akaike weight for the  $j^{\text{th}}$  variable,  $w_+(j)$ . The Akaike weight of a variable is defined as the sum of the Akaike weights of the models in which the variable occurs. The larger  $w_+(j)$  is, the more important variable  $j$  is relative to the other variables. The variable that represents habitat suitability for the rope squirrel had the largest Akaike weight of any variable. The magnitude of this weight, 0.8, indicates that there is strong evidence that this variable is important relative to the other variables that we considered. In particular, the variable that represents rope squirrel habitat suitability had a weight that was twice as large as the weight of any other variable. This supports the claim that rope squirrel habitat suitability is more important for predicting human monkeypox cases than the others variables that we analyzed (such as temperature or precipitation). For a review of Akaike weights, see Burnham and Anderson (2002).

<b>Reservoir species or genus</b>	<b>Common name</b>	<b>Most important variable</b>
<i>Cricetomys</i> species	Giant pouched rat	Maximum temperature of the warmest month
<i>Graphiurus crassicaudatus</i>	African dormouse	Minimum temperature of the coldest month
<i>Graphiurus lorraineus</i>	African dormouse	Total annual precipitation
<i>Funisciurus</i> species	Rope squirrel	Land cover

**Table 3. Most important variables for predicting the niches of monkeypox reservoirs.**

Variable importance was assessed by measuring the percent contribution of each variable to the Maxent model (Phillips et al. 2006).

**Table 4. Climate data used to predict human monkeypox (19 variables). For details, see Section 1.1 and references therein.**

<b>Climatic parameter</b>	<b>Units</b>	<b>Minimum</b>	<b>Maximum</b>
Annual mean temperature	°C	15.8	26
Mean diurnal range (mean of monthly maximum temperature – minimum temperature)	°C	9.3	12.3
Isothermality (mean diurnal range / temperature annual range)	Dimensionless	0.7	0.9
Temperature seasonality (temperature standard deviation / mean temperature)	Dimensionless	2.42	7.46
Maximum temperature of the warmest month	°C	21.4	32.9
Minimum temperature of the coldest month	°C	9.8	20.1
Temperature annual range (maximum temperature of the warmest month – minimum temperature of the coldest month)	°C	10.8	16.8
Mean temperature of the wettest quarter	°C	16	26
Mean temperature of the driest quarter	°C	15.3	25.9
Mean temperature of the warmest quarter	°C	16.1	26.6
Mean temperature of the coldest quarter	°C	15.3	25.5
Annual precipitation (total)	mm	1115	2182
Precipitation of the wettest month	mm	149	303
Precipitation of the driest month	mm	9	78
Precipitation seasonality (precipitation standard deviation / mean precipitation)	Dimensionless	18	70
Precipitation of the wettest quarter	mm	412	804
Precipitation of the driest quarter	mm	17	443
Precipitation of the warmest quarter	mm	156	656
Precipitation of the coldest quarter	mm	17	644

**Table 5 - Vegetation data obtained from remote sensing used to predict human monkeypox (7 original variables). For details, see Section 1.3 and references therein.**

<b>Data record</b>	<b>Remote sensing (RS) sensor</b>	<b>Vegetation/landscape parameter</b>	<b>RS metrics at 1 km resolution</b>
Monthly Normalized Difference Vegetation Index (NDVI) in 2001	MODIS	Vegetation type and seasonality	NDVI-1: maximum NDVI NDVI-2: mean NDVI wet months
Monthly (2000-2004) leaf area index (LAI)	MODIS	Vegetation type, seasonality, and productivity	LAI: maximum LAI
Percent tree cover (derived from the original 500m data)	MODIS	Forest cover, heterogeneity	VCF: continuous field product
Scatterometer backscatter (horizontal polarization) Monthly composites in 2001 interpolated from the original 2.25 km data	QuikSCAT	Canopy roughness, leaf/wood density, and vegetation deciduousness	QSCAT-M: mean backscatter QSCAT-S: standard deviation of backscatter
Digital elevation (derived from the original 90 m resolution)	SRTM	Surface elevation	SRTM-HGT: mean elevation

## References for the Electronic Supplementary Material

- Buermann W, Saatchi S, Smith TB, Zutta BR, Chaves JA, Mila B, et al. (2008). Predicting species distributions across the Amazonian and Andean regions using remote sensing data. *Journal of Biogeography* **35**:1160-1176.
- Burnham KP, and Anderson DR (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Second Edition*. Springer, New York.
- Calvete C, Estrada R, Miranda MA, Borrás D, Calvo JH, and Lucientes J (2008). Modelling the distributions and spatial coincidence of bluetongue vectors *Culicoides imicola* and the *Culicoides obsoletus* group throughout the Iberian peninsula. *Medical and Veterinary Entomology* **22**:124-134.
- Calvete C, Estrada R, Miranda MA, Borrás D, Calvo JH, and Lucientes J (2009). Ecological correlates of bluetongue virus in Spain: Predicted spatial occurrence and its relationship with the observed abundance of the potential *Culicoides* spp. vector. *Veterinary Journal* **182**:235-243.
- Cohen J (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **20**:37-46.
- Dobson JE, Bright EA, Coleman PR, Durfee RC, and Worley BA (2000). LandScan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing* **66**:849-857.
- Farber O, and Kadmon R (2003). Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. *Ecological Modelling* **160**:115-130.
- Hansen MC, DeFries RS, Townshend JRG, Sohlberg R, Dimiceli C, and Carroll M (2002). Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sensing of Environment* **83**:303-319.
- Harrell FE (2009). *Package 'Design'*. Available from <http://cran.r-project.org/web/packages/Design/Design.pdf>.

- Hijmans RJ, Cameron SE, Parra JL, Jones PG, and Jarvis A (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**:1965-1978.
- Holden ME, and Levine RS (2009). Systematic revision of Sub-Saharan African dormice (Rodentia: Gliridae: Graphiurus) Part II. *Bulletin of the American Museum of Natural History* **331**:314-355.
- Huete AR, Didan K, Shimabukuro YE, Ratana P, Saleska SR, Hutyyra LR, et al. (2006). Amazon rainforests green-up with sunlight in dry season. *Geophysical Research Letters* **33**.
- Hutson CL, Lee KN, Abel J, Carroll DS, Montgomery JM, Olson VA, et al. (2007). Monkeypox zoonotic associations: insights from laboratory evaluation of animals associated with the multi-state US outbreak. *American Journal of Tropical Medicine and Hygiene* **76**:757-767.
- Levine RS, Peterson AT, Yorita KL, Carroll D, Damon IK, and Reynolds MG (2007). Ecological niche and geographic distribution of human monkeypox in Africa. *Plos One* **2**:e176.
- Long DG, Drinkwater MR, Holt B, Saatchi S, and Bertoia C (2001). Global ice and land climate studies using scatterometer image data. *EOS, Transactions of the American Geophysical Union* **82**:503.
- Miller ME, Hui SL, and Tierney WM (1991). Validation techniques for logistic regression models. *Statistics in Medicine* **10**:1213-1226.
- Montgomery DC, Peck EA, and Vining GG (2006). *An Introduction to Linear Regression Analysis. Fourth Edition*. J. W. Wiley & Sons, New York.
- Myneni RB, Hoffman S, Knyazikhin Y, Privette JL, Glassy J, Tian Y, et al. (2002). Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment* **83**:214-231.
- Nix H (1986). *A Biogeographic Analysis of Australian Elapid Snakes*. Australian Government Publishing Service, Canberra.
- Pawar S, Koo MS, Kelley C, Ahmed MF, Chaudhuri S, and Sarkay S (2007). Conservation assessment and prioritization of areas in Northeast India: Priorities for amphibians and reptiles. *Biological Conservation* **136**:346-361.

- Peterson AT (2009). Shifting suitability for malaria vectors across Africa with warming climates. *Bmc Infectious Diseases* **9**:59.
- Phillips SJ, Anderson RP, and Schapire RE (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**:231-259.
- Phillips SJ, and Dudik M (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**:161-175.
- Potere D, Schneider A, Angel S, and Civco DL (2009). Mapping urban areas on a global scale: which of the eight maps now available is more accurate? *International Journal of Remote Sensing* **30**:6531-6558.
- Richards JA, and Jia X (2006). *Remote Sensing Digital Image Analysis: An Introduction. Fourth Edition*. Springer-Verlag, Berlin.
- Saatchi S, Houghton RA, dos Santos Alvalá RC, Soares JV, and Yu Y (2007). Distribution of aboveground live biomass in the Amazon basin. *Global Change Biology* **13**:816-837.
- Schafer ML, and Lundstrom JO (2009). The present distribution and predicted geographic expansion of the floodwater mosquito *Aedes sticticus* in Sweden. *Journal of Vector Ecology* **34**:141-147.
- Siegel S, and Castellan NJ (1988). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York.
- Stein BR, and Wiecek J (2004). Mammals of the world: MaNIS as an example of data integration in a distributed network environment. *Biodiversity Informatics* **1**:14-22.
- Tatem AJ, and Hay SI (2004). Measuring urbanization pattern and extent for malaria research: a review of remote sensing approaches. *Journal of Urban Health-Bulletin of the New York Academy of Medicine* **81**:363-376.
- Vijayaraj V, Bright EA, and Bhaduri BL (2007). High resolution urban feature extraction for population mapping using high performance computing. Pages 278-281 in Proceedings of IGARSS 2007. IEEE, Barcelona, Spain.

Ward MP, Wittich CA, Fosgate G, and Srinivasan R (2009). Environmental risk factors for equine West Nile virus disease cases in Texas. *Veterinary Research Communications* **33**:461-471.

Wieczorek J, Guo Q, and Hijmans R (2004). The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographic Information Science* **18**:745-767.

Williams JN, Seo CW, Thorne J, Nelson JK, Erwin S, O'Brien JM, et al. (2009). Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions* **15**:565-576.

Wisn MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A, et al. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions* **14**:763-773.