

Spatiotemporal Fluctuations and Triggers of Ebola Virus Spillover

John Paul Schmidt, Andrew W. Park, Andrew M. Kramer, Barbara A. Han,
Laura W. Alexander, John M. Drake

Because the natural reservoir of Ebola virus remains unclear and disease outbreaks in humans have occurred only sporadically over a large region, forecasting when and where Ebola spillovers are most likely to occur constitutes a continuing and urgent public health challenge. We developed a statistical modeling approach that associates 37 human or great ape Ebola spillovers since 1982 with spatiotemporally dynamic covariates including vegetative cover, human population size, and absolute and relative rainfall over 3 decades across sub-Saharan Africa. Our model (area under the curve 0.80 on test data) shows that spillover intensity is highest during transitions between wet and dry seasons; overall, high seasonal intensity occurs over much of tropical Africa; and spillover intensity is greatest at high ($>1,000/\text{km}^2$) and very low ($<100/\text{km}^2$) human population densities compared with intermediate levels. These results suggest strong seasonality in Ebola spillover from wild reservoirs and indicate particular times and regions for targeted surveillance.

Emerging infectious diseases, a persistent threat to global public health, are often linked to rapid environmental change and increasing human mobility (1,2). Notable for its unprecedented size and geographic extent, the 2013–2015 West Africa Ebola epidemic was also the first major human Ebola outbreak outside central Africa and underscored the need for improved methods to forecast emergence in novel regions. Because the natural reservoir of the Ebola virus has not been identified (3) and spillovers present an irregular pattern (4,5), it remains unclear how the probability of Ebola virus disease (EVD) in human populations varies in space and time. Particularly, whether EVD follows a seasonal pattern (6,7) and which historically unaffected geographic regions may also be at risk for EVD outbreaks (8) are 2 important questions that remain largely unanswered. Likewise, how expanding human activities, changing settlement patterns,

and increasing population density affect the probability of spillovers remains poorly resolved. Despite the absence of an obvious explanation for the timing and location of past EVD outbreaks, a set of associated social and environmental conditions that anticipate viral spillover may be broadly identifiable. Identifying the environmental correlates that bring us closer to forecasting when and where EVD risk is elevated is critical for improving surveillance and rapid response to future spillovers.

Research on Ebola during the past 2 decades has investigated spatiotemporal disease probability by using conventional time series analysis (9) and geostatistical models (10). By using time series of satellite imagery, multiple studies have suggested that Ebola spillover to humans is more likely to occur at the onset of the dry season (7,11,12). Noting that this pattern is not universal, Lash et al. (13) analyzed patterns in the time series of vegetation greenness and land surface moisture (by using a normalized difference vegetation index) for 5 spillover events and found anomalies (i.e., extreme climatic fluctuations) at a temporal scale of 20 days preceding this subset of spillover events. More recently, species distribution models have been used to map the potential geographic extent of disease probability, as in the work of Pigott et al. (4), who used these models to identify spatial covariates that associate with the occurrence of Ebola virus infection in humans, primates, and bats.

Despite these advances, notable technology gaps remain. For example, we know that spatiotemporal variation and seasonality are key characteristics of EVD regions, but we lack integrative models that reliably incorporate spatiotemporally varying indicators of interannual and intraannual fluctuations into the calculation of spillover probability. Further, although socioeconomic factors are believed to be important drivers of spillover for numerous zoonotic diseases, including Ebola (14,15), the relationship between human population growth and the increasing frequency of EVD outbreaks since the early 1990s remains largely unexplored. We also know that the biology of this region is strongly influenced by climatic seasonality (e.g., the timing of fruit and forage availability and animal migrations). Although such seasonality is widely suspected to affect viral amplification and transmission from wild reservoirs, time-series of climate or vegetation have not been investigated across the region of documented EVD events. In this study, we

Author affiliations: University of Georgia Odum School of Ecology, Athens, Georgia, USA (J.P. Schmidt, A.W. Park, A.M. Kramer, J.M. Drake); University of Georgia Center for the Ecology of Infectious Diseases, Athens (J.P. Schmidt, A.W. Park, A.M. Kramer, J.M. Drake); Cary Institute of Ecosystem Studies, Millbrook, New York, USA (B.A. Han); University of California–Berkeley Department of Ecology, Berkeley, California, USA (L.W. Alexander)

DOI: <http://dx.doi.org/10.3201/eid2303.160101>

combine spatial data on changing human population density and distribution during the past 4 decades, satellite-derived estimates of monthly rainfall for most of the same period, and summary measures of climate in a statistical model that dynamically captures the timing of past EVD events. Our model predicts human EVD outbreaks with an estimated accuracy of 80% and shows how EVD risk shifts seasonally as a function of environmental triggers and has varied over the last 3 decades because of increases in human population and changing settlement patterns.

Methods

Ebola Spillover Origin Points and Dates

We compiled a table of all known Ebola epizootics and human outbreaks from primary sources and filtered the entries to isolate primary dates and precise locations of distinct spillover events. For human Ebola spillovers, we began with chronological lists compiled by the World Health Organization and the Centers for Disease Control and Prevention. Key sources were Lahm et al. (16) and Leroy et al. (17), who compiled reports of wildlife mortality in Gabon and the Democratic Republic of the Congo; reports by ethnologists observing great ape populations in other regions; coordinates of locations from Mylne et al. (18); and locations and associated information from Kuhn's compendium (5).

To divide incident reports into discrete spillover events, we separated primary spillovers from secondary occurrences on the basis of widely accepted chronological, geographic, or genetic distances. For example, where viral sequence data indicated that multiple spillover events had

occurred, we considered them as such even if they overlapped spatially or temporally. Most events were reported as points. When events were reported as polygons (3 cases), we used polygon centroids as point locations. In contrast to Pigott et al. (4), we excluded data derived from sampling of healthy bats not associated with a spillover event. Because we were seeking to identify potential climatologic triggers, the timing of the spillover was taken to be the earliest report (often unconfirmed) of either human or animal disease rather than the first date of confirmed infection in either humans or animals. Following this procedure, a primary list of 66 spatiotemporal candidate spillover points was reduced to a final list of 44 spillover events (Figure 1; online Technical Appendix 1, <https://wwwnc.cdc.gov/EID/article/23/3/16-0101-Techapp1.xlsx>).

Spatial Predictors

To exclude arid and semi-arid regions, which are unlikely to harbor potential Ebola reservoir species and differ sharply in climate from locations where human EVD has occurred, we defined the region of interest as the portion of Africa receiving >500 mm rainfall annually. For this region, we assembled spatial data that capture the major sources of variation in climate and landcover. Following Pigott et al. (4), we chose an enhanced vegetation index (19) and potential evapotranspiration (20) to represent composite axes of coarse environmental variation.

Candidate Triggers

To characterize spatiotemporal variation at seasonal, inter-annual, and decadal scales, we compiled 3 datasets. First,

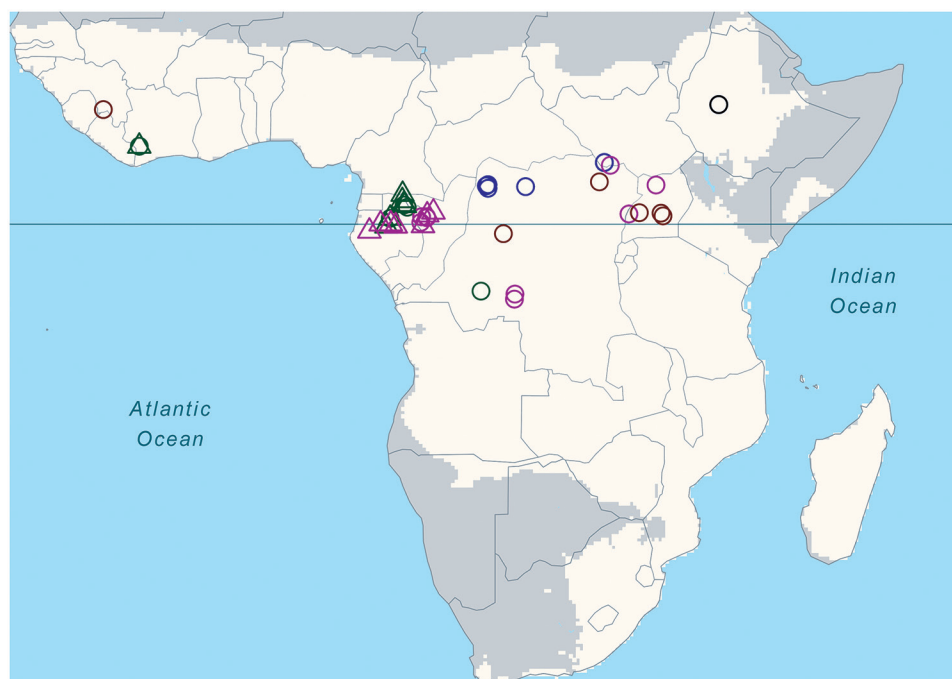


Figure 1. Locations of known Ebola virus spillover events, Africa, 1960–2010. Light-shaded area indicates the focal region in Africa of annual rainfall >500 mm. Open circles indicate human spillovers, open triangles infection/mortality in nonhuman primates or in other mammals. Yellow, blue, green, magenta, and black indicate the 5 respective decades during 1960–2010. Solid horizontal line marks the equator. No known Ebola spillovers occurred in the 1980s.

we compiled population count grids for Africa for 1960, 1970, 1980, 2000, 2005, 2010, and 2015 at 2.5 arc-minutes scale ($\approx 25 \text{ km}^2$ at the equator) from the Gridded Population of the World version 3, produced by the Columbia University Center for International Earth Science Information Network (<http://sedac.ciesin.columbia.edu/data/collection/gpw-v3>). After linearly interpolating counts by grid cell for intervening years, we \log_{10} -transformed values of human population and created 3 population bins according to $x \leq 10^2$, $10^2 < x < 10^3$, and $x \geq 10^3$.

Second, we aggregated monthly rainfall from daily rainfall estimates obtained from the Rainfall Estimator (27). This data product was developed in 1998 by the Climate Prediction Center at the National Oceanic and Atmospheric Administration and is available at high (0.1°) spatial resolution for January 1983 to the present.

Third, in addition to actual monthly rainfall, we created a rainfall anomaly index as a means of incorporating the potential importance of relative rainfall. For the time series of 384 monthly rainfall rasters, we divided the value of each month-location by the maximum value for that location to create a set of 384 scaled raster images corresponding to the original monthly rainfall raster images.

Model Fitting and Validation

We restricted analysis to the 37 (80.5%) of 44 EVD events occurring since 1982, the period for which monthly rainfall estimates across Africa were available. We sampled 100,000 random background points from within the portion of Africa receiving $>500 \text{ mm}$ rainfall annually and randomly assigned each to 1 of 384 months during 1983–2014. Spillover occurrence points and background points were divided into 2/3 training and 1/3 test sets. Because actual monthly rainfall at month and site of EVD outbreak varied considerably, we stratified by rainfall, first ranking points by rainfall amount and then assigning every third point to the test set.

Model Testing

We modeled Ebola spillover intensity, the average density or expected number of points per unit area and time, using bagged logistic regression models with main effects only (22). Bagging (bootstrap aggregating) is a machine learning approach that uses the predictive power generated from ensembles of models based on small subsets of the data (23). By using all 5 predictors described, we fit 1,000 models in which we randomly sampled 10 of the 22 outbreaks in the training dataset and 100 of 100,000 training background points. We predicted each of the 1,000 fitted models on both the training and test datasets. Taking the mean of predicted spillover intensity across model iterations, we compared average predicted spillover intensity for training and test points with labels at each point (known EVD event vs. otherwise) in each dataset to gauge the accuracy of the models.

Risk Mapping

The set of known EVD events represent a spatiotemporal point process. Point processes are described by an intensity function (i.e., the average density or expected number of points per unit area and time). Therefore, after validation, we used the complete dataset (37 spillover and 100,000 background points) to retroactively predict Ebola spillover intensity across the entire portion of Africa receiving $>500 \text{ mm}$ rainfall annually for all 384 months for which gridded rainfall data were available (January 1983–December 2014) using human population estimates for 2015. We then averaged the resulting 384 monthly rasters to map seasonal shifts in predicted Ebola spillover intensity across Africa. To map the change in spillover intensity as a function of changes in human population size and distribution across 4 decades, we averaged predicted intensity across all months of 1975 and 2015, then took the difference between annual spillover intensity in 2015 and annual spillover intensity in 1975 across the region of Africa receiving $>500 \text{ mm}$ rainfall annually.

Detailed methods are provided in online Technical Appendix 2 (<https://wwwnc.cdc.gov/EID/article/23/3/16-0101-Techapp2.pdf>), and the R code used is provided in online Technical Appendix 3 (<https://wwwnc.cdc.gov/EID/article/23/3/16-0101-Techapp3.pdf>). All data and code are available online (https://figshare.com/articles/ebola_spillover_intensity_final_Rmd/4234280).

Results

Predictive accuracy of the bagged model of EVD intensity trained on the 2/3 training dataset was high. Area under the receiver-operator curve was 0.83 when evaluated on the training dataset and 0.80 when evaluated on 1/3 of the data that were withheld from model training. Overall accuracy ($[\text{true positives} + \text{true negatives}] / \text{total points}$) was 53% for prediction on the test set. Mean annual Ebola spillover intensity was highest where the enhanced vegetation index is highest in the wettest portions of tropical Africa. For locations within the humid tropics of Africa, predicted spillover intensity was generally, but not always, lowest in dry months (rainfall $<50 \text{ mm}$) (Figure 2; online video, <https://wwwnc.cdc.gov/EID/article/23/3/16-0101-V1.htm>). Across sites, modeled spillover intensity in months of intermediate rainfall (100–250 mm) was equal to or exceeded that in high ($>250 \text{ mm}$) rainfall months (Figure 3). Whereas central Africa exhibits relatively constant spillover intensity throughout the year (particularly within the narrow equatorial region of 15° – 30° longitude), we found spillover intensity to be highly seasonal in southern Africa and somewhat variable in West Africa. These results extend the proposed potential range of spillover far beyond the locations of past outbreaks. Compared with previously published spatial models of spillover that did not include temporally varying predictors, our results suggest that a

much larger area of Africa is at moderate to high risk for spillover during some months of the year, including much of East Africa, Madagascar, and south central Africa (e.g., Angola and Zambia) and a large portion of West Africa (online video). Although Ebola spillover intensity in seasonally at-risk regions peripheral to central Africa is much lower than in high-intensity central Africa itself, predicted spillover intensity at sites in Angola, Mozambique, and Ethiopia is comparable to that predicted at known spillover locations in South Sudan and Gabon. Predicted spillover intensity over a large portion of Madagascar is similar to that of central Africa (Figure 3). Thus, within the African tropics, the potential for Ebola spillover appears to be geographically widespread.

The effect of human population on Ebola spillover intensity is much smaller than climatic or seasonal effects. The change in average annual spillover intensity did not change markedly for much of Africa as population increased during 1975–2015 (Figure 4), whereas spillover intensity exhibited striking shifts with climate and seasonality (Video). Nevertheless, our model does show that spillover intensity differs by human population density. Mean annual spillover

intensity was lowest where population size per 25 km² grid cell was intermediate ($10^2 < x < 10^3$) and highest where population density was low ($x \leq 100$) (online Technical Appendix 2 Figure 1). Large changes in spillover intensity ($\pm 5\%$) during 1975–2015 appear to result mainly from population increases. In comparing 2015 to 1975 population density, shifts from intermediate-to-high population densities have generated increased Ebola spillover intensity, particularly in West Africa and the region surrounding Lake Victoria, and shifts from low-to-medium population densities have reduced spillover risk. Similarly, settlements along transportation corridors have increased in population to intermediate densities, leading to substantial declines in predicted spillover intensity. However, as a result of population consolidation over large areas of central Africa, some remote districts have declined in population, typically increasing predicted spillover intensity (Figure 4; Figure 5).

These results quantify a spatiotemporal pattern in the risk for Ebola spillover in 2 specific ways: first, as raw estimates by the model algorithm that can be directly compared between months and across locations (Figure 3); and second, as percentile ranks of these estimates or relative

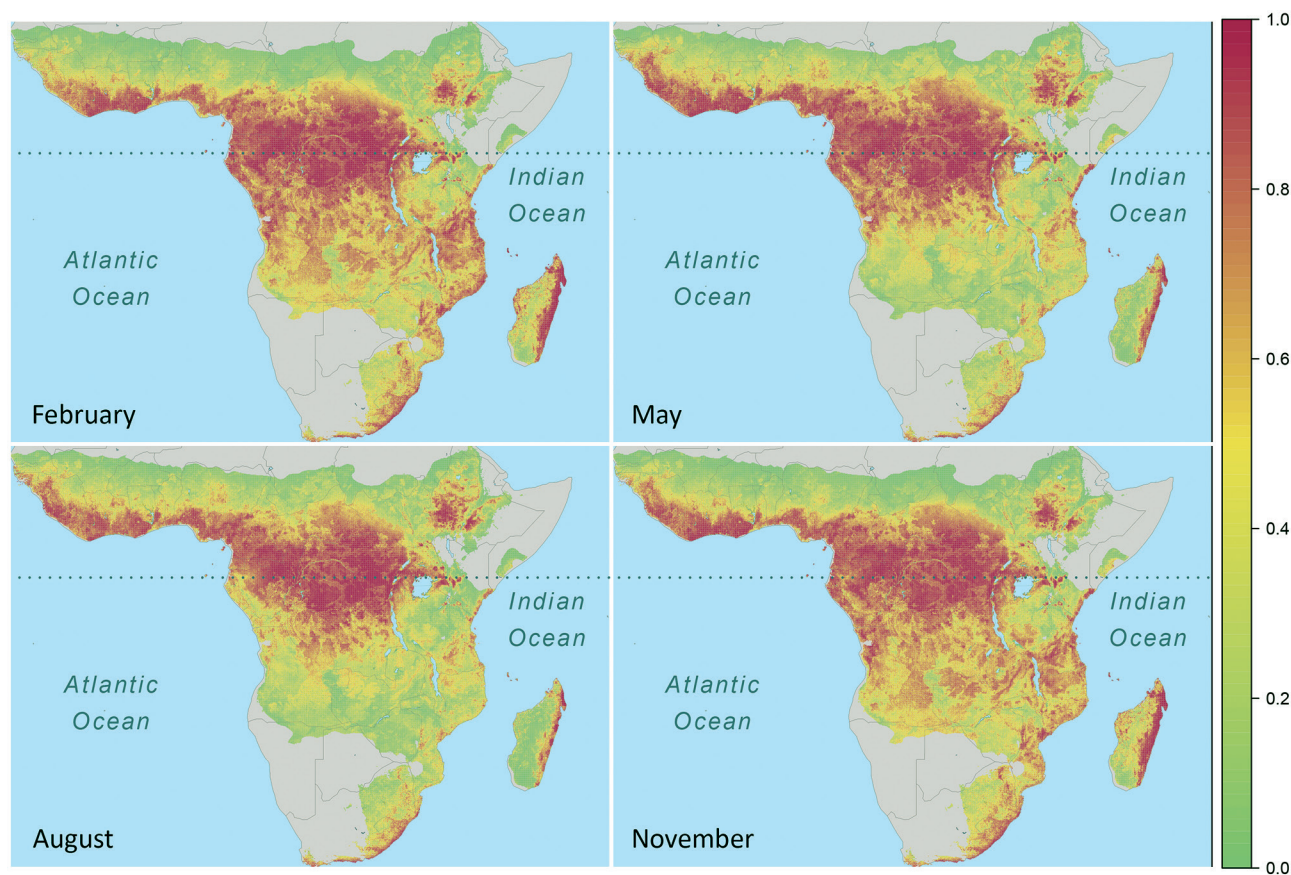


Figure 2. Seasonal spatiotemporal dynamics of Ebola virus spillover intensity (i.e., average density or expected number of points per unit area and month) as percentile values ranking predicted intensities at all grid cell locations within the region of Africa where annual rainfall was >500 mm for all months from January 1983 through December 2014. Panels capture shifts in the geographic pattern of spillover intensity seasonally. Dotted horizontal line marks the equator.

spillover intensity (online video). Percentile ranking adjusts for model miscalibration because some spillovers may not have been observed and because of the overrepresentation of spillover events in base logistic regression models. As such, percentile ranking preserves discriminability (i.e., classification accuracy as measured by area under the receiver-operator curve performance), even when probabilities are not well calibrated. Whether raw or ranked, spillover intensities are a measurement of risk, with values proportional to the probability of a spillover that changes as a function of environmental conditions based on the best information available on the location and timing of unique spillover events. Although the transmission, dynamics, and possibly the seasonality of different viral strains may differ, our approach, constrained by the small number of spillovers, properly considers EVD as a syndrome caused by all known strains of the Ebola virus. By constructing models to compare the covariates associated with this set of known spillover events to the background possibilities from which they might have been drawn, which we accurately approximate by using a sample of 100,000 random points, we have robustly determined how the intensity of Ebola spillovers changes with observable covariates.

Discussion

These results indicate that 1) there is a geographic gradient of annual Ebola spillover intensity that peaks in central Africa but extends during at least some months of the year through a large portion of tropical Africa not previously considered to be at high risk (4,24), including the tropical/subtropical forest/woodland regions of Ethiopia, Angola, Zambia, East Africa, and Madagascar; 2) there is substantial seasonal fluctuation in the spatial pattern of Ebola spillover intensity; 3) there is a temporal gradient in spillover intensity in which the driest months show the lowest intensity and intensity peaks or plateaus in months of intermediate rainfall; and 4) increases in human population density may increase Ebola spillover risk in West and central Africa. Ebola spillover intensity is greatest when regions that are typically very wet make the transition to or from dry periods. This result corroborates the finding from previous studies (7,11) linking EVD events to preceding dry-to-wet transitions through time series analysis of data from a normalized difference vegetation index. Within predominantly or seasonally wet climate zones in particular, our results show Ebola spillover intensity to be highest in moderately dry months and lowest in extremely dry months.

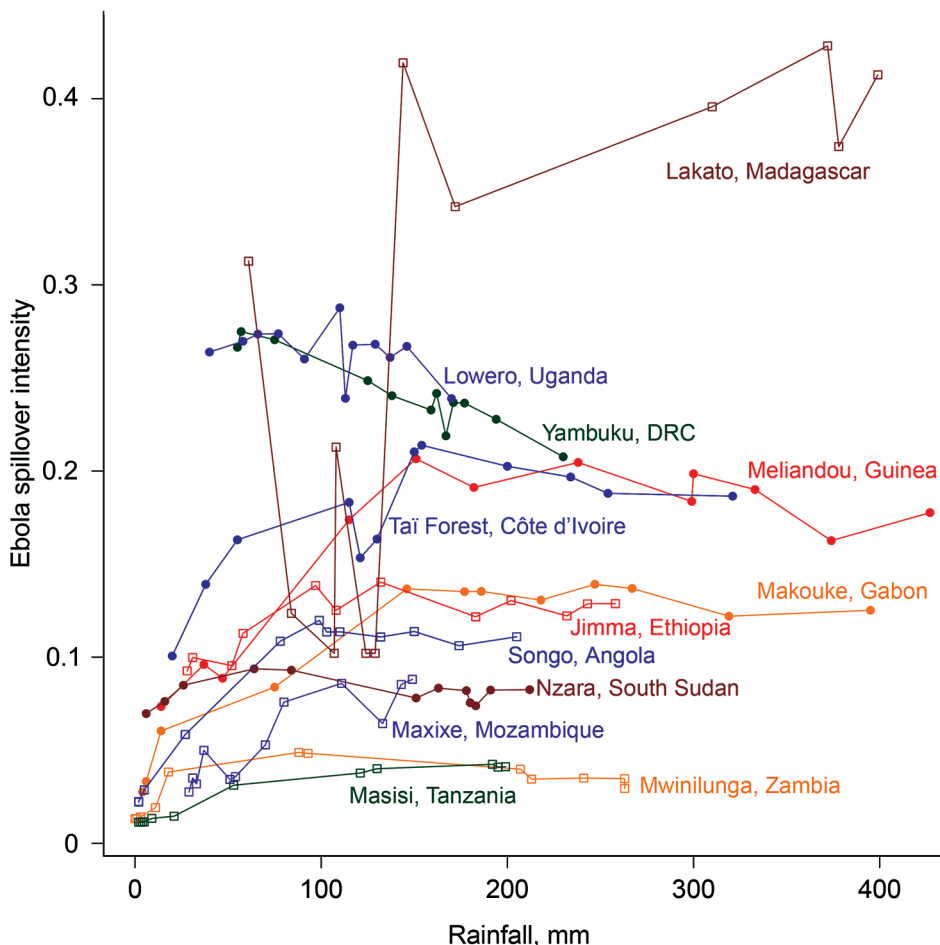


Figure 3. Phase graph showing the relationship between mean monthly rainfall and raw Ebola spillover intensity (defined as average density or expected number of points per unit area and/or time) for known Ebola virus disease locations in West and Central Africa (closed circles) and locations in northeastern or southern Africa where model results indicate moderate to high Ebola spillover intensity seasonally (open squares). Points are ordered by least to greatest monthly rainfall at each site. Dotted horizontal line marks the equator. DRC, Democratic Republic of the Congo.

Seasonal dynamics in spillover intensity are most pronounced where rainfall seasonality is greatest (i.e., outside the less seasonal and wetter rainforest biome of central Africa, where EVD events have been most frequent and spillover intensity is most steady throughout the year). Strong seasonal patterns may be related not only to seasonal drivers, such as rainfall, but to migration patterns and seasonal competence of wildlife reservoirs. Seasonal effects on resource availability may drive migrations or other changes in movement patterns that, in turn, may affect population density, social behaviors, and contact rates among hosts (25). Seasonal changes may also alter the frequency of host encounters with infective agents or material in the environment, and host immune defenses can shift with annual reproductive cycles (26). Seasonality is also likely to alter human behavior, including hunting effort, level of bushmeat consumption, or, more generally, the degree and kind of contact with wildlife.

Our model finds that Ebola spillover intensity varies temporally as a function of climate variables without explicitly incorporating sociocultural dimensions, such as land use, which was not available as a time series, or biotic features, such as the ranges of suspected reservoir hosts. Therefore, the degree of human disease intensity at

locations far from documented EVD events may also depend on whether the range of a necessary reservoir also extends to these points. In recent work, species distribution models were used to predict the ranges of potential mammal reservoirs and the degree of overlap of predicted ranges with Ebola and Marburg spillovers to suggest likely mammalian reservoirs (27). Among the taxa that overlapped with all EVD sites were the sun squirrel genus (*Heliosciurus*) and the straw-colored fruit bat (*E. helvum*), both of which had predicted ranges covering nearly all of tropical Africa (with the exception of Madagascar, where *E. dupreanum* is present), where our models predicted high Ebola spillover intensity at least seasonally. Thus, our predictions across continental Africa may adequately reflect the biotic component of risk. However, an important next step would be to assess whether the presence of suitable animal hosts or cultural or socioeconomic factors in Madagascar and East Africa make this region a priority for surveillance.

Our model was trained by using great ape and human EVD events. Great ape spillover events (usually observations made by primatologists and wildlife researchers within reserves, in this dataset restricted to Gabon and the Democratic Republic of the Congo) are associated with low human population density, whereas our model associates

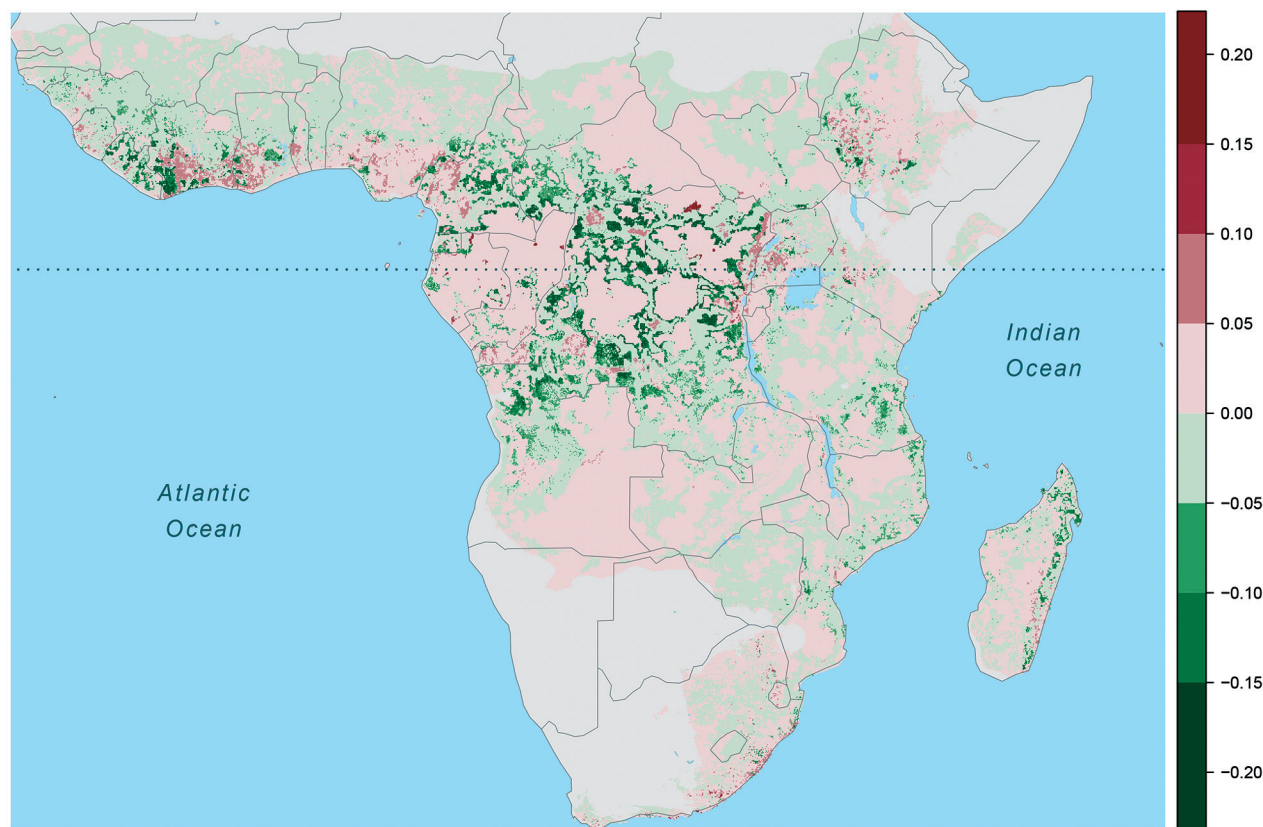


Figure 4. Change in annual Ebola spillover intensity (defined as average density or expected number of points per unit area and time), Africa, 1975–2015. Warm colors indicate increased spillover intensity; cool colors indicate decreased spillover intensity. Dotted horizontal line marks the equator.

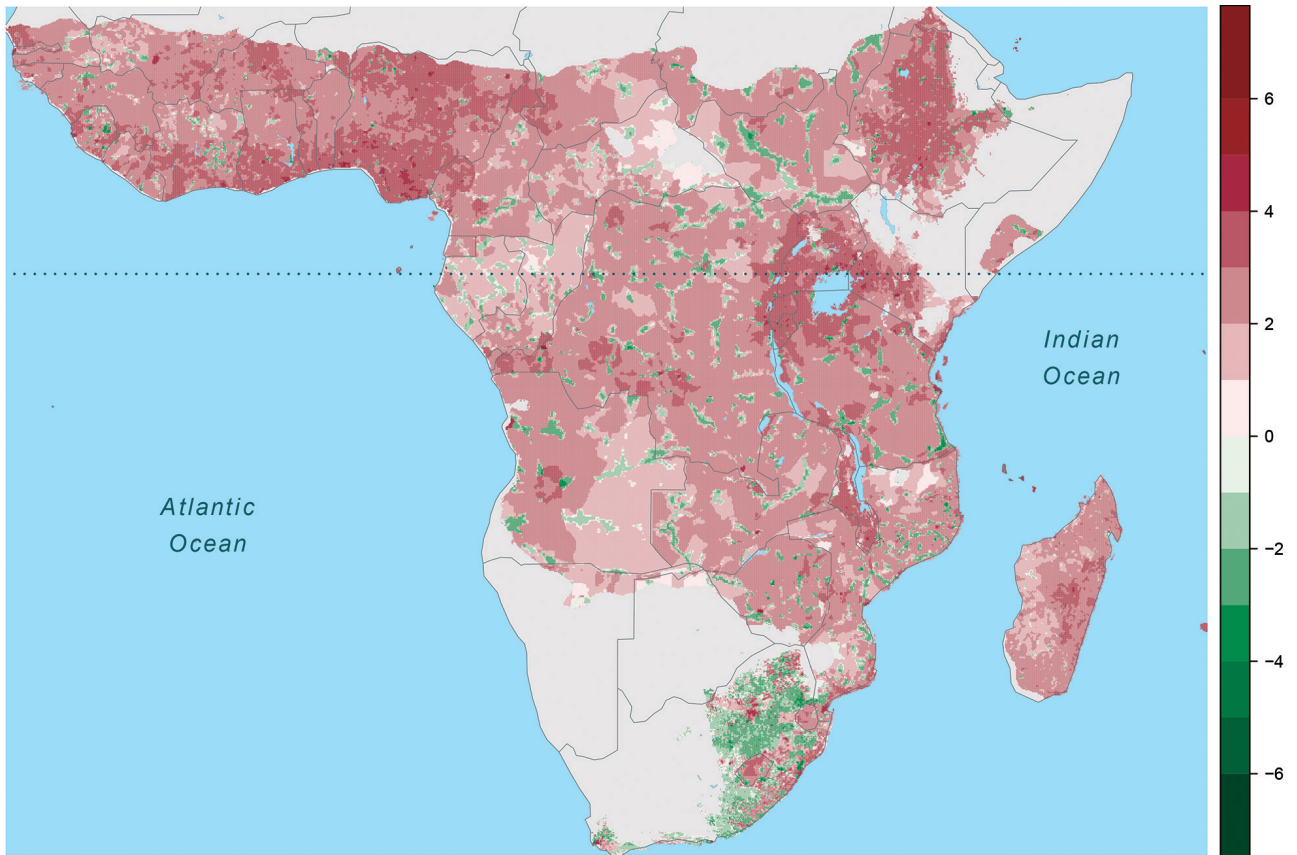


Figure 5. Change in human population size ($\log_{10}/25 \text{ km}^2$ grid cell), Africa, 1975–2015. Warm colors indicate increased population size, cool colors population declines. Dotted horizontal line marks the equator.

human spillovers with high human population densities ($>10^3$ persons/ 25 km^2). At low population densities, epidemic spread is less likely, and deaths in remote outposts may go unreported. The link between human population density and Ebola spillover intensity could be simply a function of increased reporting at high population densities (and some locations with very low population densities). Alternatively, increased contact with or consumption of wildlife as population density increases, or perhaps the increased abundance of reservoir or bridge reservoir species at either high or low human population density (or both) could drive the relationship. Substantially increased raw Ebola spillover intensity ($>5\%$) as a result of population increases is most apparent in areas of West Africa but could eventually include central Africa if urbanization or population consolidation continues there. We note that human population was a much less important predictor than variables capturing climate and seasonality.

In conclusion, we developed a model that predicts a pattern of widespread but seasonally very dynamic Ebola spillover intensity in savannah and humid tropical regions of Africa from the set of known spatiotemporal EVD points ($n = 37$ since 1990) and spatially and temporally high-

resolution rainfall and population data for Africa. Ebola virus, though not the strain that led to the recent outbreak, was known to be circulating in West Africa before 2014 (28–30). However, the potential for a major human outbreak, by far the most deadly Ebola outbreak to date, was not foreseen. Answering the need for improved forecasting, surveillance, and preparation for rapid response, our model uses the best available spatiotemporal predictors and an ensemble modeling approach to accurately identify geographic regions and seasons of elevated Ebola spillover intensity, and suggests that the socio-ecologic conditions that triggered the initial spillover in Guinea may prevail over a much larger area and at a higher temporal frequency. A key public health policy implication is that some level of Ebola surveillance should be extended to regions outside of central and West Africa. Furthermore, the spatiotemporal pattern of Ebola spillover intensity we report could be used as an early warning system to inform the design of surveillance activities.

Acknowledgments

We especially thank Nicholas Novella for his help in accessing rainfall data for Africa.

A.W.P. acknowledges funding from the National Science Foundation under a Rapid Response Research grant (award no. 1515194). J.M.D. acknowledges funding from the Research and Policy for Infectious Disease Dynamics program of the Science and Technology Directorate of the US Department of Homeland Security and from the Fogarty International Center of the National Institutes of Health. This work was also supported by the National Institute of General Medical Sciences of the National Institutes of Health (award no. U01GM110744).

Dr. Schmidt holds a PhD in ecology and is currently an assistant research scientist at the Odum School of Ecology at the University of Georgia. He is interested in the application of data science methods to ecological questions.

References

- Smith KF, Guégan JF. Changing geographic distributions of human pathogens. *Annu Rev Ecol Syst.* 2010;41:231–50. <http://dx.doi.org/10.1146/annurev-ecolsys-102209-144634>
- Daszak P, Cunningham AA, Hyatt AD. Emerging infectious diseases of wildlife—threats to biodiversity and human health. *Science.* 2000;287:443–9. <http://dx.doi.org/10.1126/science.287.5452.443>
- Heeney JL. Ebola: hidden reservoirs. *Nature.* 2015;527:453–5. <http://dx.doi.org/10.1038/527453a>
- Pigott DM, Golding N, Mylne A, Huang Z, Henry AJ, Weiss DJ, et al. Mapping the zoonotic niche of Ebola virus disease in Africa. *Elife.* 2014;3:e04395. <http://dx.doi.org/10.7554/eLife.04395>
- Kuhn JH. Filoviruses: a compendium of 40 years of epidemiological, clinical, and laboratory studies. New York: Springer; 2008.
- Groseth A, Feldmann H, Strong JE. The ecology of Ebola virus. *Trends Microbiol.* 2007;15:408–16. <http://dx.doi.org/10.1016/j.tim.2007.08.001>
- Pinzon JE, Wilson JM, Tucker CJ, Arthur R, Jahrling PB, Formenty P. Trigger events: enviroclimatic coupling of Ebola hemorrhagic fever outbreaks. *Am J Trop Med Hyg.* 2004;71:664–74.
- Changula K, Kajihara M, Mweene AS, Takada A. Ebola and Marburg virus diseases in Africa: increased risk of outbreaks in previously unaffected areas? *Microbiol Immunol.* 2014;58:483–91. <http://dx.doi.org/10.1111/1348-0421.12181>
- Linthicum KJ, Anyamba A, Tucker CJ, Kelley PW, Myers MF, Peters CJ. Climate and satellite indicators to forecast Rift Valley fever epidemics in Kenya. *Science.* 1999;285:397–400. <http://dx.doi.org/10.1126/science.285.5426.397>
- Anyamba A, Chretien JP, Small J, Tucker CJ, Formenty PB, Richardson JH, et al. Prediction of a Rift Valley fever outbreak. *Proc Natl Acad Sci U S A.* 2009;106:955–9. <http://dx.doi.org/10.1073/pnas.0806490106>
- Tucker CJ, Wilson JM, Mahoney R, Anyamba A, Linthicum K, Myers MF. Climatic and ecological context of the 1994–1996 Ebola outbreaks. *Photogramm Eng Remote Sensing.* 2002;68:147–52.
- Pinzon E, Wilson JM, Tucker CJ. Climate-based health monitoring systems for eco-climatic conditions associated with infectious diseases. *Bull Soc Pathol Exot.* 2005;98:239–43.
- Lash RR, Brunzell NA, Peterson AT. Spatiotemporal environmental triggers of Ebola and Marburg virus transmission. *Geocarto Int.* 2008;23:451–66. <http://dx.doi.org/10.1080/10106040802121010>
- Bausch DG, Schwarz L. Outbreak of Ebola virus disease in Guinea: where ecology meets economy. *PLoS Negl Trop Dis.* 2014;8:e3056. <http://dx.doi.org/10.1371/journal.pntd.0003056>
- Wolfe ND, Dunavan CP, Diamond J. Origins of major human infectious diseases. *Nature.* 2007;447:279–83. <http://dx.doi.org/10.1038/nature05775>
- Lahm SA, Kombila M, Swanepoel R, Barnes RFW. Morbidity and mortality of wild animals in relation to outbreaks of Ebola haemorrhagic fever in Gabon, 1994–2003. *Trans R Soc Trop Med Hyg.* 2007;101:64–78. <http://dx.doi.org/10.1016/j.trstmh.2006.07.002>
- Leroy EM, Telfer P, Kumulungui B, Yaba P, Rouquet P, Roques P, et al. A serological survey of Ebola virus infection in central African nonhuman primates. *J Infect Dis.* 2004;190:1895–9. <http://dx.doi.org/10.1086/425421>
- Mylne A, Brady OJ, Huang Z, Pigott DM, Golding N, Kraemer MU, et al. A comprehensive database of the geographic spread of past human Ebola outbreaks. *Sci Data.* 2014;1:140042. <http://dx.doi.org/10.1038/sdata.2014.42>
- Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens Environ.* 2002;83:195–213. [http://dx.doi.org/10.1016/S0034-4257\(02\)00096-2](http://dx.doi.org/10.1016/S0034-4257(02)00096-2)
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol.* 2005;25:1965–78. <http://dx.doi.org/10.1002/joc.1276>
- Novella NS, Thiaw WM. African rainfall climatology version 2 for famine early warning systems. *J Appl Meteorol Climatol.* 2013;52:588–606. <http://dx.doi.org/10.1175/JAMC-D-11-0238.1>
- Valentini G, Dietterich TG. Low bias bagged support vector machines. In: *Proceedings of the 20th International Conference on Machine Learning (ICML-2003)*, Washington DC; 2003. p. 752–9 [cited 2014 Dec 15]. <http://www.aaai.org/Papers/ICML/2003/ICML03-098.pdf>
- Breiman L. Bagging predictors. *Mach Learn.* 1996;24:123–40. <http://dx.doi.org/10.1007/BF00058655>
- Peterson AT, Bauer JT, Mills JN. Ecologic and geographic distribution of filovirus disease. *Emerg Infect Dis.* 2004;10:40–7. <http://dx.doi.org/10.3201/eid1001.030125>
- Altizer S, Bartel R, Han BA. Animal migration and infectious disease risk. *Science.* 2011;331:296–302. <http://dx.doi.org/10.1126/science.1194694>
- Altizer S, Dobson A, Hosseini P, Hudson P, Pascual M, Rohani P. Seasonality and the dynamics of infectious diseases. *Ecol Lett.* 2006;9:467–84. <http://dx.doi.org/10.1111/j.1461-0248.2005.00879.x>
- Peterson AT, Papeş M, Carroll DS, Leirs H, Johnson KM. Mammal taxa constituting potential coevolved reservoirs of filoviruses. *J Mammal.* 2007;88:1544–54. <http://dx.doi.org/10.1644/06-MAMM-A-280R1.1>
- Hayman DT, Yu M, Cramer G, Wang LF, Suu-Ire R, Wood JL, et al. Ebola virus antibodies in fruit bats, Ghana, West Africa. *Emerg Infect Dis.* 2012;18:1207–9. <http://dx.doi.org/10.3201/eid1807.111654>
- Formenty P, Boesch C, Wyers M, Steiner C, Donati F, Dind F, et al. Ebola virus outbreak among wild chimpanzees living in a rain forest of Côte d'Ivoire. *J Infect Dis.* 1999;179(Suppl 1):S120–6. <http://dx.doi.org/10.1086/514296>
- Formenty P, Hatz C, Le Guenno B, Stoll A, Rogenmoser P, Widmer A. Human infection due to Ebola virus, subtype Côte d'Ivoire: clinical and biologic presentation. *J Infect Dis.* 1999;179(Suppl 1):S48–53. <http://dx.doi.org/10.1086/514285>

Address for correspondence: John Paul Schmidt, University of Georgia, Odum School of Ecology, 140 E Green St, Athens, GA 30602, USA; email: jps@uga.edu

Spatiotemporal Fluctuations and Triggers of Ebola Virus Spillover

Technical Appendix 2

Methods

Ebola Spillover Origin Points and Dates

We compiled a spatiotemporally indexed table of all known EVD events from existing sources and filtered the entries to isolate primary dates and locations of distinct spillover events (Technical Appendix 2 Figure 1). For human spillovers, we began with chronological lists compiled by the World Health Organization and U.S. Centers for Disease Control and Prevention. Key sources were Lahm et al. (2007) (1) and Leroy et al. (2004) (2) who compiled reports of wildlife mortality in Gabon and the Democratic Republic of Congo, reports by ethnologists observing great ape populations in other regions, coordinates of locations from Mylne et al. (2014) (3) and Kuhn's compendium (4).

To divide incident reports into discrete spillover events, we separated incidents into primary spillovers and secondary occurrences on the basis of widely accepted chronological, geographic or genetic distances. For example, where sequence data indicated that multiple spillover events had occurred, we considered them as such even if they overlapped spatially or temporally. Most events were reported as points. When reported as polygons we used polygon centroids as point locations. In contrast to Pigott et al. (2014) (5), we excluded data from sampling of healthy bats not associated with a spillover event. Critically, because we were seeking to identify potential climatological triggers, the timing of the spillover was taken to be the earliest report (often unconfirmed) of human or animal disease rather than the first date of confirmed infection in either humans or animals. Following this procedure, a primary list of 66 spatiotemporal candidate spillover points was reduced to a final list of 44 spillover events (online Technical Appendix 1, <https://wwwnc.cdc.gov/EID/article/23/3/16-0101-Techapp1.xlsx>).

Spatial Covariates

To exclude arid and semi-arid regions which are unlikely to harbor potential Ebola reservoir species and differ sharply in climate from locations where EVD has occurred, we defined the region of interest as the portion of Africa receiving >500 mm rainfall annually. For this region we assembled spatial data that capture the significant sources of variation in climate and land cover. Following Pigott et al. (2014) (5), enhanced vegetation index (EVI) and potential evapotranspiration (PET) were chosen to represent composite axes of coarse environmental variation. EVI, an optimized index derived from satellite data for vegetation monitoring, is an enhanced measurement of reflected light in the visible and near-infrared spectrum obtained by removing spectral noise caused by canopy effects and atmospheric influences. EVI is computed as:

$$\text{EVI} = G \cdot (\text{NIR} - \text{red}) / (\text{NIR} + C1 \cdot \text{red} - C2 \cdot \text{blue} + L)$$

where NIR (near-infrared), red, and blue are atmospherically-corrected for Rayleigh and ozone absorption; surface reflectance, L is the canopy background adjustment that addresses nonlinear, differential NIR and red radiant transfer through a canopy; and C1 and C2 are coefficients of aerosol resistance, using the blue band to correct for aerosol influences in the red band. The coefficients adopted in the MODIS-EVI algorithm are: L = 1, C1 = 6, C2 = 7.5, and G (gain factor) = 2.5 (6). EVI has been adopted by NASA as a standard product of the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors. EVI values in our analyses are drawn from a raster of mean values for all months over the period 2002–2014. Mean values for EVI are served by the U.S. Geological Survey MODIS Land Processes Distributed Active Archive Center (LandDAAC) via the IRI/LDEO Climate Data Library at Columbia University ([http://iridl.ldeo.columbia.edu/SOURCES/.USGS/.LandDAAC/.MODIS/.version_005/.EAF/.EVI/\[X+Y+T+\]average](http://iridl.ldeo.columbia.edu/SOURCES/.USGS/.LandDAAC/.MODIS/.version_005/.EAF/.EVI/[X+Y+T+]average)). The original spatial resolution of MODIS EVI data are 250 m, but mean values for EVI are available at 4 km resolution.

Annual PET was obtained as 30 arc-second geospatial rasters from the CGIAR Consortium for Spatial Information (<http://www.cgiar-csi.org/data/global-aridity-and-pet-database>) and is a measure of the ability of the atmosphere to remove water through evapotranspiration, which is strongly correlated with climate and vegetation type. PET data were modeled using the WorldClim Global Climate Data (6) as input parameters. The WorldClim

data, based on a high number of climate observations and SRTM topographical data, is a high-resolution global geo-database of monthly average data (1950–2000) for precipitation, and mean, minimum and maximum temperature. PET was calculated as

$$PET = 0.0023 \cdot R \cdot (T_{\text{mean}} + 17.8) \cdot \text{Trange}^{0.5} \text{ (mm / day)},$$

where T_{mean} is mean monthly temperature, Trange is mean monthly temperature range, and R is mean monthly extra-terrestrial radiation, and validated by comparison to data from climate stations in Africa and South America. For use in statistical models to predict EVD spillover intensity, EVI and PET rasters were rescaled by subtracting the mean and dividing by the standard deviation within the rainfall >500 mm masked region.

Candidate Triggers

To characterize spatiotemporal variation at seasonal, inter-annual, and decadal scales, we compiled the following datasets.

1) Population count grids for Africa for 1960, 1970, 1980, 2000, 2005, 2010, and 2015 at 2.5 arc-minutes scale ($\sim 25 \text{ km}^2$ at the equator) from the Gridded Population of the World version 3 (7), produced by the Columbia University Center for International Earth Science Information Network (CIESIN) and available from the Socioeconomic Data and Applications Center (SEDAC), one of the Distributed Active Archive Centers (DAACs) in the Earth Observing System Data and Information System (EOSDIS) of the U.S. National Aeronautics and Space Administration (NASA). Data for each decade 1960–2000 reflect national or subnational input administrative units of varying resolution depending on original reporting, whereas 2005, 2010, and 2015 grids were extrapolated by CIESIN based on a combination of sub-national growth rates from census dates and national growth rates from United Nations statistics. We linearly interpolated counts by cell for intervening years. Guided by exploratory analyses of the data using boosted regression trees that identified a nonlinear relationships between Ebola spillover intensity and population size, we log10-transformed values of human population and grouped into three bins according to $x \leq 10^2$, $10^2 < x < 10^3$, $x \geq 10^3$.

2) Monthly rainfall was aggregated from daily rainfall estimates obtained from the Rainfall Estimator (RFE) (8), a data product developed in 1998 by the Climate Prediction Center (CPC) at the National Oceanic and Atmospheric Administration (NOAA) to support accurate monitoring of large-scale and climatic trends with a high (0.1°) spatial resolution by blending

gauge and satellite information on a near-real time basis. This dataset provides daily rainfall estimates over the African continent for the time period January 1983-present (<http://www.cgiar-csi.org/data/global-aridity-and-pet-database>).

3) In addition to actual monthly rainfall, as a means of incorporating the potential importance of relative rainfall, we created a rainfall anomaly index as follows. For the series of 384 monthly rainfall rasters, we divided the value of each month-location by the maximum value for that location to create a set of 384, scaled rasters corresponding to the original monthly rainfall rasters.

Model Fitting and Validation

Analysis was restricted to the 37 out of 44 (80.5%) EVD events occurring since 1982, the period for which monthly rainfall estimates for Africa were available. These were divided into 2/3 for training and 1/3 for testing. Because actual monthly rainfall at month and site of EVD spillovers varied considerably, we stratified by rainfall, first ranking points by rainfall amount and then assigning every third point to the test set. A set of 100,000 random background points was sampled from the $1,115,874 \times 384$ location-time combinations within the >500 mm rainfall mask (30 arc second resolution). Background points were similarly divided into 2/3 for training and 1/3 for testing. For spillover and background points, values from each covariate raster extracted to presence and background points. To test the sensitivity of model results to the choice of background points, we ran the models again on a second set of random points. For each set of background points, receiver operator characteristic (ROC) curves (Technical Appendix 2 Figure 2) visualize classification error for 1) prediction on training and test data for models derived from analyses using all spillover points, and 2) prediction on training data for models derived from analyses using human spillovers only. Area under the ROC curve, a metric of model performance, was similar for the models using human spillover points only (0.84 versus 0.83 in one iteration, 0.83 in both cases for the second set of random background points) when predicted on training data. Too few human spillover points were available to hold out a subset for testing. Overall, comparison of model results from two separate sets of random background points indicates little sensitivity to choice of background points (Technical Appendix 2 Figure 2), suggesting that the range of spatiotemporal variability is adequately captured by randomly sampling 100,000 background points over the 384 months for the masked region of Africa.

Model Performance Testing

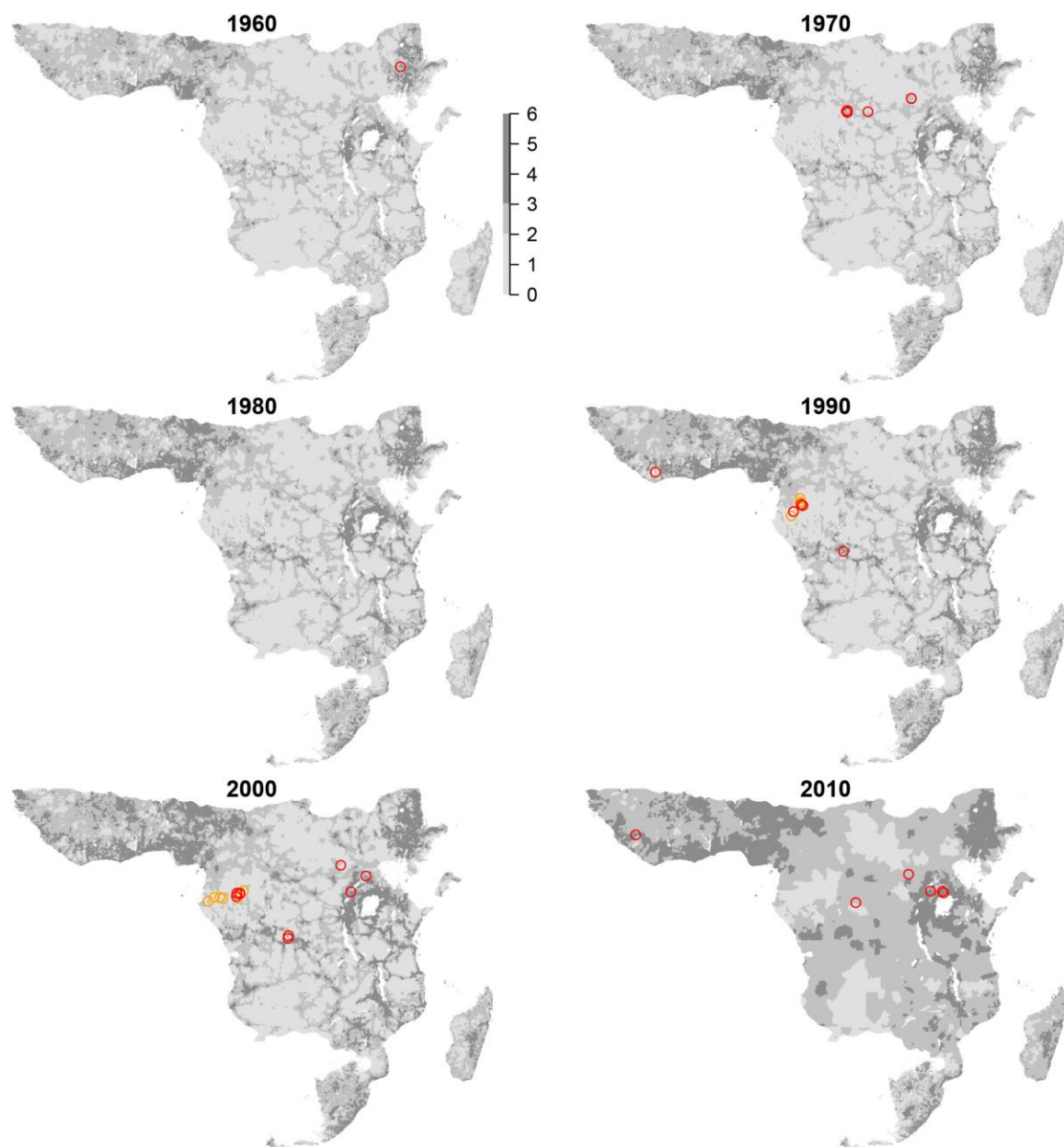
Exploratory analysis showed that conventional, flexible machine learning algorithms, such as boosted regression trees had a strong tendency to overfit to the data, very likely because of the very small number of points available for fitting. Because “bagged” low-variance models have previously been shown to yield reasonable generalizability (9), we reasoned that bagged ensembles of more rigid models might yield superior predictive performance. Bagging (bootstrap aggregating) is a machine learning approach that makes use of the predictive power generated from ensembles of so-called “weak” learners, i.e., weakly tuned models based on small subsets of the data (10). For this reason, we modeled EVD spillover intensity using bagged logistic regression models with main effects only. Using all 5 predictors, we fit 1000 models in which we randomly sampled 10 of the 22 outbreaks in the training dataset and 100 of 100,000 training background points. We predicted each of the 1000 fitted models on both the training and test data. We compared mean predicted values for training and test points to EVD versus background labels in each dataset to calculate AUC, the area under the receiver operating characteristic curve (ROC). ROC curves plot the performance of a binary classifier system as its discrimination threshold varies. By plotting the true positive rate as a function of the false positive rate, AUC provides an index of classification accuracy such that an AUC score of 0.5 indicates a classifier that performs no better than random, and an AUC score of 1 indicates perfect discrimination.

Risk Mapping

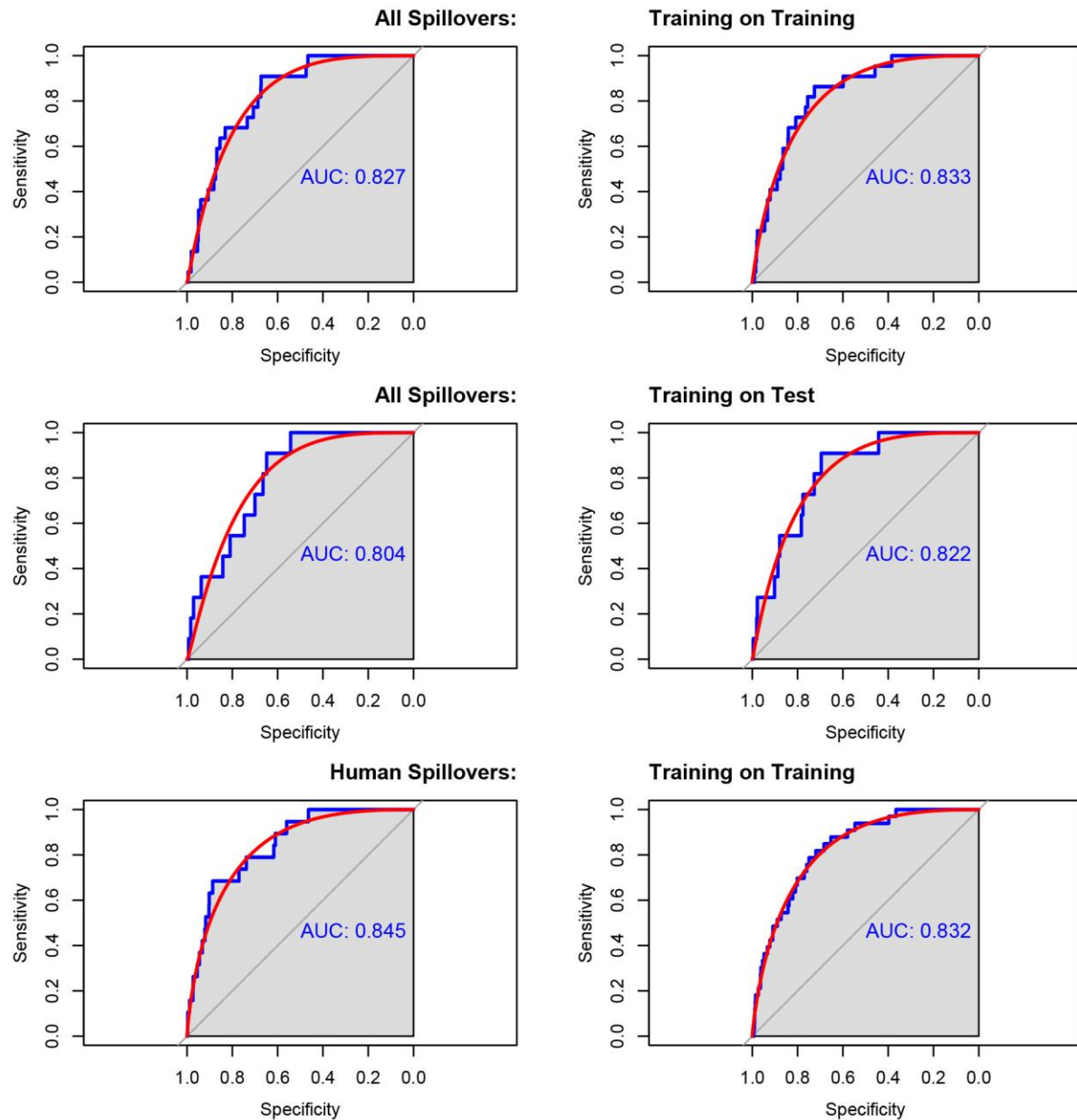
Following the bagging procedure above, we used the complete dataset (37 spillover, 100,000 background points) to predict EVD spillover intensity across the entire masked (>500 mm rainfall) region of Africa for all 384 months (January, 1983-present for which gridded rainfall data was available from NOAA) using human population estimates for 2015. We then averaged the resulting 384 spillover intensity rasters by month to create a mapped visualization of seasonal shifts in spillover intensity across Africa. This procedure was repeated using human population estimates for 1975. To map changes in EVD spillover intensity as a function of changes in human population across 4 decades, we averaged predicted spillover intensity across all months for 1975 and 2015 and took the difference of the 2 resulting grids.

References

1. Lahm SA, Kombila M, Swanepoel R, Barnes RFW. Morbidity and mortality of wild animals in relation to outbreaks of Ebola haemorrhagic fever in Gabon, 1994-2003. *Trans R Soc Trop Med Hyg.* 2007;101:64–78. [PubMed http://dx.doi.org/10.1016/j.trstmh.2006.07.002](http://dx.doi.org/10.1016/j.trstmh.2006.07.002)
2. Leroy EM, Telfer P, Kumulungui B, Yaba P, Rouquet P, Roques P, et al. A serological survey of Ebola virus infection in central African nonhuman primates. *J Infect Dis.* 2004;190:1895–9. [PubMed http://dx.doi.org/10.1086/425421](http://dx.doi.org/10.1086/425421)
3. Mylne A, Brady OJ, Huang Z, Pigott DM, Golding N, Kraemer MU, et al. A comprehensive database of the geographic spread of past human Ebola outbreaks. *Sci Data.* 2014;1:140042. [PubMed http://dx.doi.org/10.1038/sdata.2014.42](http://dx.doi.org/10.1038/sdata.2014.42)
4. Kuhn JH. *Filoviruses: a compendium of 40 years of epidemiological, clinical, and laboratory studies.* New York: Springer; 2008.
5. Pigott DM, Golding N, Mylne A, Huang Z, Henry AJ, Weiss DJ, et al. Mapping the zoonotic niche of Ebola virus disease in Africa. *eLife.* 2014;3:e04395. [PubMed http://dx.doi.org/10.7554/eLife.04395](http://dx.doi.org/10.7554/eLife.04395)
6. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol.* 2005;25:1965–78. <http://dx.doi.org/10.1002/joc.1276>
7. Center for International Earth Science Information Network. Gridded Population of the World, Version 3 (GPWv3): Population Count Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center [cited 2014 Sep 1]. <http://dx.doi.org/10.7927/H4639MPP>
8. Herman A, Kumar VB, Arkin PA, Kousky JV. Objectively determined 10-day African rainfall estimates created for famine early warning systems. *Int J Remote Sens.* 1997;18:2147–59. <http://dx.doi.org/10.1080/014311697217800>
9. Valentini G, Dietterich TG. Low bias bagged support vector machines. In: *Proceedings of the 20th International Conference on Machine Learning (ICML-2003)*, Washington DC; 2003. p. 752–9 [cited 2014 Dec 15]. <http://www.aaai.org/Papers/ICML/2003/ICML03-098.pdf>
10. Breiman L. Bagging predictors. *Mach Learn.* 1996;24:123–40. <http://dx.doi.org/10.1007/BF00058655>
11. Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG. Overview of the radiometric and biophysical performance of the MODros Inf Serv. vegetation indices. *Remote Sens Environ.* 2002;83:195–213. [http://dx.doi.org/10.1016/S0034-4257\(02\)00096-2](http://dx.doi.org/10.1016/S0034-4257(02)00096-2)



Technical Appendix 2 Figure 1. Population density and Ebola spillover locations for the masked region of Africa (>600 mm annual rainfall) for each decade during 1960–2010. Legend numbers represent human population ranges as powers of 10 per 25 km². Red circles mark human spillovers, orange circles mark nonhuman primate and other mammal spillovers.



Technical Appendix 2 Figure 2. Receiver operator characteristic curves showing performance of models resulting from analyses using all Ebola virus disease spillover points, predicted on 2/3 training data (top), and predicted on holdout 1/3 test data (middle), Africa, 1983–2015. Bottom row shows performance of models resulting from analyses using human spillover points only predicted on training data. Left and right columns compare results from 2 different sets of background points. Filled polygons represent the area under the receiver operator characteristic curve; red arcs are smoothed curves.

Spatiotemporal Fluctuations and Triggers of Ebola Virus Spillover

Technical Appendix 3

R Code Used for Risk Mapping

```
---
title: "ebola.final.spillover.Rmd"
output: word_document
---

First, we load index points of unique Ebola spillover events as determined by Laura Alexander.
```{r}
index.pts <- as.data.frame(read.table("index.pts.csv",header = T, sep=",",na.strings=c("na","NA")))
```

Subset Ebola spillovers occurring since 1983, and that were the earliest reported location for each spillover event.
```{r}
index.pts2 <- index.pts[which(index.pts$year > 1983 & index.pts$initial == "TRUE"),]
```

Load R packages needed for all the data development and analysis. (If these packages are not currently installed,
they will need to be.)
```{r}
library(raster)
library(sp)
library(maptools)
library(rgdal)
library(gdalUtils)
library(dismo)
library(rworldmap)
library(grid)
library(latticeExtra)
```

Get lat-long extent for Africa.
```{r}
africa.extent<-extent(c(-17.30253,55.84747,-34.83332,16.05834))
```

Load rainfall in mm from WorldClim 1.4 from http://www.worldclim.org/current. Values are averages over 1960-1990 for
each month.
```{r}
jan.rf<-raster("prec_1.bil")
feb.rf<-raster("prec_2.bil")
mar.rf<-raster("prec_3.bil")
apr.rf<-raster("prec_4.bil")
may.rf<-raster("prec_5.bil")
jun.rf<-raster("prec_6.bil")
jul.rf<-raster("prec_7.bil")
aug.rf<-raster("prec_8.bil")
sep.rf<-raster("prec_9.bil")
oct.rf<-raster("prec_10.bil")
nov.rf<-raster("prec_11.bil")
dec.rf<-raster("prec_12.bil")
```

Crop monthly rainfall rasters to Africa.
```{r}
af.rf.jan<-crop(jan.rf,africa.extent)
af.rf.feb<-crop(feb.rf,africa.extent)
af.rf.mar<-crop(mar.rf,africa.extent)
af.rf.apr<-crop(apr.rf,africa.extent)
af.rf.may<-crop(may.rf,africa.extent)
af.rf.jun<-crop(jun.rf,africa.extent)
af.rf.jul<-crop(jul.rf,africa.extent)
af.rf.aug<-crop(aug.rf,africa.extent)
af.rf.sep<-crop(sep.rf,africa.extent)
af.rf.oct<-crop(oct.rf,africa.extent)
af.rf.nov<-crop(nov.rf,africa.extent)
af.rf.dec<-crop(dec.rf,africa.extent)
```

Create stack of average rainfall over Africa by month.
```



```

```{r}
af.rf.stack<-
stack(af.rf.jan,af.rf.feb,af.rf.mar,af.rf.apr,af.rf.may,af.rf.jun,af.rf.jul,af.rf.aug,af.rf.sep,af.rf.oct,af.rf.nov,af.
rf.dec)
af.rf.brick<-brick(af.rf.stack)
writeRaster(af.rf.brick,filename="af.rf.brick.grd",overwrite=TRUE)
af.rf.brick<-brick("af.rf.brick.grd")
```

Sum stack to get average annual rainfall for Africa.
```{r}
af.rf.annual<-sum(af.rf.stack)
```

Create mask region where annual rainfall >500 mm in Africa.
```{r}
m<-c(0,500,NA, 500,4559,1)
rf.mask.500<-reclassify(af.rf.annual,m)
writeRaster(rf.mask.500,"rf.mask.500.tif")
rf.mask.500<-raster("rf.mask.500.tif")
```

Set random seed and generate random background points from within the rainfall mask. To test the effect of the choice
of background points, we also generated a second set of background point using set.seed(45).
Set random seed.
```{r}
randomseed<-seq.int(1,528,1)
set.seed(55)
```

Get 100k random background points from with the >500 mm rainfall mask for Africa.
```{r}
bg<-randomPoints(mask, 100000)
```

Set id for each point.
```{r}
id<-seq.int(44,100000,1)
```

Join id to coordinates of the point.
```{r}
bg<-cbind(bg,id)
```

Create sequence of years from 1983 through 2014, and months 1-12.
```{r}
years<-seq.int(1983,2014,1)
months<-seq.int(1,12,1)
```

Generate a random year and month (1970-2014 or 1983-2014) for each background point.
```{r}
year<-NULL
month<-NULL
for (i in 1:100000) {
y<-sample(years,1)
m<-sample(months,1)
year<-rbind(year,y)
month<-rbind(month,m)
}
```

Set presence = 0 for all background points.
```{r}
pres<-rep(0,100000)
all.bg<-cbind(bg,year,month,pres)
```

Set column headings for coordinates, id, year, month and presence.
```{r}
colnames(all.bg)<-c("x","y","id","year","month","pres")
```

Combine spillover locations with background points.
```{r}
all.pts<-rbind(index.pts2[,1:6],all.bg)
```

Take daily data from the Rainfall Estimator and aggregates it to monthly to create a stack of rainfall grids by month
for the period 1983-2014.
Source: "In 1998, the Climate Prediction Center (CPC) developed the Rainfall Estimator (RFE) (Herman et. al. 1997) in
response to the need for higher resolution operational daily rainfall estimates to support the humanitarian aid
programs of USAID / Famine Early Warning Systems Network (FEWS - NET). The RFE hascontinued to provide an accurate
monitoring of large-scale and regional climatic and hydrological trends. It is a unique product compared to other
satellite rainfall estimators because of its high, 0.1° gridded spatial resolution, and its ability to blend gauge and
satellite information on a near-real time basis to provide daily (06Z-06Z) rainfall estimates over the African
continent" (http://www.cpc.ncep.noaa.gov/products/fews/AFR\_CLIM/AMS\_ARC2a.pdf). The daily data above was aggregated
to derive monthly rainfall estimates at the same spatial resolution. A number of days in the early period 1983-1985
were missing data for many days.
Loop through all of the daily rasters to create a raster aggregating daily values to monthly values.
```{r}
year<-
c("1983","1984","1985","1986","1987","1988","1989","1990","1991","1992","1993","1994","1995","1996","1997","1998","1999",
"2000","2001","2002","# month<-c("01","02","03","04","05","06","07","08","09","10","11")
day<-
c("01","02","03","04","05","06","07","08","09","10","11","12","13","14","15","16","17","18","19","20","21","22","23","2
4","25","26","27","# monthsum<-stack()
for (y in 1:10){
for (m in 1:12){

```

```

for (d in 1:31){
if (exists(paste("africa_arc.", year[y], month[m],day[d], ".tif", sep="")))
dia<-raster(paste("africa_arc.", year[y], month[m],day[d], ".tif", sep=""))
monthsum<-sum(stack(monthsum,dia))
}
writeRaster(monthsum, paste("/Google Drive/ebola/monthyear/africa_prpc", year[y], month[m],
".tif",sep=""),overwrite=TRUE)
}
}
...

Create a raster stack of rainfall grids by month for the period 1983-2014.
```{r}
# setwd("~/Google Drive/ebola/monthyear")
# f <- list.files()
# r <- raster(f[1])
# rf.stack <- stack(r)
...

Bring in population count grids for Africa for 1960, 1970, 1980, 2000, 2005, 2010, and 2015 (the last 3 estimated)
from http://sedac.ciesin.columbia.edu/data/set/grump-v1-population-density/maps?facets=region:africa.
```{r}
pop60<-raster("afpop60.tif")
crs(pop60)<-"+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0"
pop70<-raster("afpop70.tif")
crs(pop70)<-"+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0"
pop80<-raster("afpop80.tif")
crs(pop80)<-"+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0"
pop90<-raster("afpop90.tif")
crs(pop90)<-"+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0"
pop00<-raster("afpop00.tif")
crs(pop00)<-"+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0"
pop05<-raster("afpop05.tif")
crs(pop05)<-"+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0"
pop10<-raster("afpop10.tif")
crs(pop10)<-"+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0"
pop15<-raster("afpop15.tif")
crs(pop15)<-"+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0"
...

Mask out no data areas which are calced to 0 in the orginial files so that 0 and no data are distinct values.
```{r}
# pop60<-mask(pop60,mask)
# pop70<-mask(pop70,mask)
# pop80<-mask(pop80,mask)
# pop90<-mask(pop90,mask)
# pop00<-mask(pop00,mask)
# pop05<-mask(pop05,mask)
# pop10<-mask(pop10,mask)
# pop15<-mask(pop15,mask)
...

Crop grids to the extent of the >600 mm rainfall mask.
```{r}
pop60<-crop(pop60,mask)
pop70<-crop(pop70,mask)
pop80<-crop(pop80,mask)
pop90<-crop(pop90,mask)
pop00<-crop(pop00,mask)
pop05<-crop(pop05,mask)
pop10<-crop(pop10,mask)
pop15<-crop(pop15,mask)
...

Change 0 values for population to 1 to avoid generating NAs with log transformation.
```{r}
# pop60<-calc(pop60,fun=function(x){ x[x==0] <- 1; return(x)})
# pop70<-calc(pop70,fun=function(x){ x[x==0] <- 1; return(x)})
# pop80<-calc(pop80,fun=function(x){ x[x==0] <- 1; return(x)})
# pop90<-calc(pop90,fun=function(x){ x[x==0] <- 1; return(x)})
# pop00<-calc(pop00,fun=function(x){ x[x==0] <- 1; return(x)})
# pop05<-calc(pop05,fun=function(x){ x[x==0] <- 1; return(x)})
# pop10<-calc(pop10,fun=function(x){ x[x==0] <- 1; return(x)})
# pop15<-calc(pop15,fun=function(x){ x[x==0] <- 1; return(x)})
...

Log 10 transform population values of population rasters.
```{r}
logpop60<-calc(pop60, fun=function(x){log10(x)})
logpop70<-calc(pop70, fun=function(x){log10(x)})
logpop80<-calc(pop80, fun=function(x){log10(x)})
logpop90<-calc(pop90, fun=function(x){log10(x)})
logpop00<-calc(pop00, fun=function(x){log10(x)})
logpop05<-calc(pop05, fun=function(x){log10(x)})
logpop10<-calc(pop10, fun=function(x){log10(x)})
logpop15<-calc(pop15, fun=function(x){log10(x)})
...

Create population density stack.
```{r}
# popstack<-stack(logpop60,logpop70,logpop80,logpop90,logpop00,logpop05,logpop10,logpop15)
# crs(popstack)<-"+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0"
...

Create population density dataframe from the raster stack.
```{r}

```

```

pop.df<-as.data.frame(popstack,xy=TRUE)
```
Name columns by year.
```{r}
colnames(pop.df)<-c("x","y",1960,1970,1980,1990,2000,2005,2010,2015)
```
For all years between censuses, linearly interpolate values for population density for each cell.
Get the difference in years between each population estimate
```{r}
pop.estimates <- c(1960,1970,1980,1990,2000,2005,2010,2015)
d <- diff(pop.estimates)
```
Get the year columns from the dataframe and linearly interpolate between years.
```{r}
pop.df2<-pop.df[,3:10]
tvseq <- function(...)t(Vectorize(seq.default)(...))
L <- lapply(seq(d), function(i) tvseq(from=pop.df2[,i], to=pop.df2[,i+1], length.out=d[i]+1)[-1])
```
Combine columns for interpolated years to census years.
```{r}
pop.df3 <- cbind(pop.df2[,1],do.call(cbind, L))
```
Leave out 2015 and add coordinates to dataframe.
```{r}
pop.df4<-cbind(pop.df2[,1:2],pop.df3[,1:55])
```
Convert population dataframe back to a raster stack by year for each cell.
Turn dataframe into a gridded dataframe.
```{r}
coordinates(pop.df4) <- ~ x + y
gridded(pop.df4) <- TRUE
```
Create a raster stack of log10-transformed population counts for every year from 1983 through 2014, the years for
which gridded monthly rainfall are available.
```{r}
popstack<-raster(pop.df4[,15])
for (i in 24:ncol(pop.df4)){
pop<-raster(pop.df4[,i])
popstack<-addLayer(popstack,pop)
print(nlayers(popstack))
}
crs(popstack)<-"+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0"
popbrick<-brick(popstack)
writeRaster(popbrick,filename="popbrick.grd",overwrite=TRUE)
popbrick<-brick("popbrick.grd")
```
Extract rainfall data to presence and background points with year and month values between January 1983 to September
2014, the period for which gridded rainfall data is available. Also, derive the scaled rainfall value (relative to
the maximum at a cell) for each month-site.
```{r}
library(matrixStats)
years<-seq.int(1983,2014,1)
months<-rep(seq(1,12,1),32)
yrs<-sort(rep(years*100,12))
m.yrs<-colSums(rbind(months,yrs))
num<-seq(1,384,1)
lookup<-as.data.frame(cbind(m.yrs,num))
yr.m.table<-all.pts[,3:4]
yr.m.table<-unique(yr.m.table)
rf.yr.m.cols<-NULL
for (i in 1:nrow(yr.m.table)){
yr<-yr.m.table[i,1]
yr.num<-(yr - 1982)
xy.coords<-all.pts[i,1:2]
m<-yr.m.table[i,2]
yr.m<-as.numeric(yr*100+m)
ind<-which(lookup[,1]==yr.m, arr.ind=TRUE)
name <- paste("rainfall/africa_prpc", yr.m, "_Resample_R.tif",sep = "")
r<-raster(name)
yr.pts<-subset(all.pts,year==yr)
yr.month.pts<-subset(yr.pts,month==m)
use coordinates to extract values at points
rf.ym<-extract(r,yr.month.pts[,1:2])
rf.ym.log<-log10(rf.ym+1)
now get scaled rainfall
extract all values for a cell
rf.site<-extract(rf.stack,yr.month.pts[,1:2])
rf.max<-rowMaxs(rf.site)
rf.rescaled<-rf.ym/rf.max
now put it all together in a row
rf.yr.m<-cbind(yr.month.pts,rf.ym.log,rf.rescaled)
now concatenate rows
rf.yr.m.cols<-rbind(rf.yr.m.cols,rf.yr.m)
print(i)
}
```
Get population value at each site for the year corresponding to the point.
```{r}

```



```

yr.table<-unique(yr.table)
pd.yr.row<-NULL
rf.pd.cols<-NULL
for (i in 1:32){
yr<-years[i]
yr.num<-(yr - 1982)
yr.pts<-rf.yr.m.cols[which(rf.yr.m.cols$year==yr),]
yr.raster<-subset(popbrick,yr.num)
pd.yr<-extract(yr.raster,yr.pts[,1:2])
pd.yr.row<-cbind(yr.pts,pd.yr)
rf.pd.cols<-rbind(rf.pd.cols,pd.yr.row)
}
...

Create 3 bins of log10 population, 0-2, 2-3, and 3-6.
```{r}
# pd.yr.binned<-as.numeric(cut(rf.pd.cols[,9], breaks = c(0,2,3,6)))
...

Function to rescale a raster by subtracting mean and dividing by standard deviation.
```{r}
rescale.raster <- function(data0, data1=NULL){
if(missing(data1)){
raster <- (data0-cellStats(data0,'mean'))/cellStats(data0,'sd')
}else{
raster <- (data1-cellStats(data0,'mean'))/cellStats(data0,'sd')
}
return(raster)
}
...

Extract values of the enhanced vegetation index (EVI) available from by the U.S. Geological Survey MODIS Land
Processes Distributed Active Archive Center (LandDAAC) via the IRI/LDEO Climate Data Library at Columbia University
(http://iridl.ldeo.columbia.edu/SOURCES/.USGS/.LandDAAC/.MODIS/.version_005/.EAF/.EVI/\[X+Y+T+\].average/).
Because EVI scales over a large range, values were rescaled by subtracting the mean and dividing by the standard
deviation within the rainfall >600 mm masked region using the function above.
```{r}
# evir<-raster("evir.tif")
# evi<-extract(evir,dyn.cols[,1:2])
...

Extract values of Annual potential evapotranspiration (PET) obtained as 30 arc-second geospatial rasters from the
CGIAR Consortium for Spatial Information (http://www.cgiar-csi.org/data/global-aridity-and-pet-database).
Because PET scales over a large range, values were rescaled by subtracting the mean and dividing by the standard
deviation within the rainfall >600 mm masked region using the function above.
```{r}
petr<-raster("petr.tif")
pet<-extract(petr,dyn.cols[,1:2])
...

Put all the data together.
```{r}
# final.pts<-cbind(rf.pd.cols,pd.yr.binned,evi,pet)
# final.pts<-as.data.frame(read.table("final.pts.csv",header = T, sep="," ,na.strings=c("na","NA")))
...

Subset presence points.
```{r}
pres.pts<-subset(final.pts,pres==1)
...

Sort by rainfall amount.
```{r}
o<-order(pres.pts[,7])
pres.pts<-pres.pts[o,]
...

Set rows for training and test, stratifying for rainfall amount by choosing every 3rd point for test.
```{r}
train.rows<-c(1,2,4,5,7,8,10,11,13,14,16,17,19,20,22,23,25,26,28,29,31,32)
test.rows<-c(3,6,9,12,15,18,21,24,27,30,33)
...

Create training and test presence points.
```{r}
train.pres<-pres.pts[train.rows,]
test.pres<-pres.pts[test.rows,]
...

Subset absence points.
```{r}
abs.pts<-as.data.frame(subset(final.pts,pres==0))
...

Randomly split absence points using random seed for repeatability. "index/3" indicates randomly divide into 3/3
training and 1/3 test.
```{r}
splitdf <- function(dataframe, seed=NULL) {
  if (!is.null(seed)) set.seed(seed)
  index <- 1:nrow(dataframe)
  trainindex <- sample(index, trunc(length(index)/3))
  trainset <- dataframe[-trainindex, ]
  testset <- dataframe[trainindex, ]
  list(trainset=trainset,testset=testset)
}
...

Split using random seed = 59.
```{r}
abs.index<-splitdf(abs.pts,59)

```

```

```
Create training and test background sets.
```{r}
train.bg<-abs.index$trainset
test.bg<-abs.index$testset
```

New packages to do bagged logistic regression.
```{r}
library(parallel)
library(foreach)
library(doSNOW)
library(rlecuyer)
library(data.table)
```

Bagging function.
```{r}
bagging<-function(form.x.y,train.pres,train.bg,new.data,iterations=1000){
 predictions<-foreach(m=1:9,.combine='+') %dopar% {
 training_positions.p <- sample(nrow(training.pres),size=10)
 training_positions.b <- sample(nrow(training.abs),size=100)
 train_pos.p<-1:nrow(training.pres) %in% training_positions.p
 train_pos.b<-1:nrow(training.abs) %in% training_positions.b
 glm_fit<-glm(form.x.y,data=rbind(training.pres[train_pos.p,],training.abs[train_pos.b,]),family=binomial(logit))
 predict(glm_fit,newdata=new.data,type="response")
 }
 return(predictions)
}
```

Allocate memory.
```{r}
cores.to.use<-176
cl<-makeCluster(cores.to.use,type="MPI")
clusterSetupRNG(cl)
registerDoSNOW(cl)
```

Logistic regression formula.
```{r}
glm.formula<-as.formula("pres~rf.ym+evi+pet+rf.scaled+as.factor(pd.yr.binned)")
```

Predict bagged model on training set.
```{r}
training<-rbind(train.pres,train.bg)
output1<-
bagging(form.x.y=glm.formula,training.pres=train.pres,training.abs=train.bg,new.data=training,iterations=cores.to.use)
```

Predict bagged model on test set.
```{r}
testing<-rbind(test.pres,test.bg)
output2<-
bagging(form.x.y=glm.formula,training.pres=train.pres,training.abs=train.bg,new.data=testing,iterations=cores.to.use)
```

Now do PREDICTION on the entire area of Africa >500 mm rainfall using bagged model and full data.
Create a lattice of points from the masked region for spillover intensity prediction. Note resolution reduced by a
factor of 4 to speed computation.
First, resample the >500 mm rainfall mask.
```{r}
resampleFactor<-4
inCols <- ncol(mask)
inRows <- nrow(mask)
resampledRaster <- raster(ncol=(inCols / resampleFactor), nrow=(inRows / resampleFactor))
extent(resampledRaster) <- extent(mask)
```

Turn cells within the >500 mm rainfall portion of the new raster into points.
```{r}
coord.raster<-subset(resampledRaster,1)
coord.db<-as.data.frame(coord.raster,xy=TRUE)
coords<-coord.db[,1:2]
```

Create a raster stack of rainfall for the >500 mm of rainfall region for the 384 months of the 1983-2014 period.
```{r}
setwd("~/rainfall/")
f<-list.files()
rf.rasters <- lapply(f,raster)
```

Extract rainfall values at the resampled rainfall mask coordinates, and turn into a dataframe.
```{r}
rf.extracted <- lapply(rf.rasters,extract,coords)
rf.extracted2 <- unlist(rf.extracted)
setwd("~/")
rf.ym.df<-as.data.frame(rf.extracted2)
rf.ym.df<-as.data.frame(read.table("rf.ym.df.csv",header = T, sep=",",na.strings=c("na","NA")))
```

Generate scaled rainfall (percent of maximum value at a cell 1983-2014) dataframe from rainfall dataframe.
```{r}
library(matrixStats)
rf.ym.all<-as.matrix(rf.ym.df)
rf.maxes.all<-rowMaxs(rf.ym.all,na.rm=T)
rf.scaled.all<-rf.ym.all/rf.maxes.all
```

```

```

# rf.scaled.all<-as.data.frame(read.table("rf.scaled.all.csv",header = T, sep=",",na.strings=c("na","NA")))
```
Extract values for static variables, EVI and PET, and 2015 values of binned population. Combine into one dataframe.
```{r}
# evi.all<-extract(evir,coords)
# pet.all<-extract(petr,coords)
```
Extract 2015 population counts at the resampled rainfall mask coordinates.
```{r}
# pd.yr.all<-extract(logpop15,coords)
```
Bin population values.
```{r}
# pd.yr.binned.all<-as.numeric(cut(pd.yr, breaks = c(0,2,3,6)))
```
Combine EVI, PET, and binned population into a single dataframe.
```{r}
# evi.pet.pd.binned<-cbind(evi.all,pet.all,pd.yr.binned.all)
# evi.pet.pd.binned<-as.data.frame(read.table("evi.pet.pd.binned.csv",header = T, sep=",",na.strings=c("na","NA")))
```
Combine training and test data for new bagged logistic regression model to do prediction on >500 mm of rainfall
masked region of Africa.
```{r}
# final.p<-rbind(train.pres,test.pres)
# final.b<-rbind(train.bg,test.bg)
```
Loop to do bagging predictions on all 384 months of rainfall. Do in 12 parts because loading data takes the same
amount of time.
```{r}
# for (i in 1:12){
#   rf.scaled<-fread("rf.scaled.all.csv",select=((i*32)-31):(i*32))
#   rf.ym<-fread("rf.ym.df.csv",select=((i*32)-31):(i*32))
#   for (j in 1:32){
#     rf.scaled.step<-rf.scaled.all[,j] #Single year of data
#     rf.ym.step<-rf.ym.all[,j] #Single year of data
#     all.data.step<-as.data.frame(cbind(evi.pet.pd.binned,rf.scaled.step,log10(rf.ym.step+1)))
#     colnames(all.data.step)<-c("evi","pet","pd.yr.binned","rf.scaled","rf.ym")
#     sum<-NULL
#     chunks<-6
#     for (k in 1:chunks){
#       output<-
#       bagging(form.x.y=glm.formula,train.pres=final.p,training.bg=final.b,new.data=all.data.step,iterations=cores.to.use)
#       sum<-sum+output
#     }
#     mean.predict<-sum/(cores.to.use*chunks)
#     writeRDS(mean.predict,file=paste("Bags/predictionsSet",i*j))
#     print(i*j)
#   }
# }
```
Create sequence from 1-384 for each month.
```{r}
# jan<-seq(1,384,12)
# feb<-seq(2,384,12)
# mar<-seq(3,384,12)
# apr<-seq(4,384,12)
# may<-seq(5,384,12)
# jun<-seq(6,384,12)
# jul<-seq(7,384,12)
# aug<-seq(8,384,12)
# sep<-seq(9,384,12)
# oct<-seq(10,384,12)
# nov<-seq(11,384,12)
# dec<-seq(12,384,12)
```
Separate Ebola spillover intensity values by month.
```{r}
library(matrixStats)
# jan.risk<-NULL
# for (i in 1:length(jan)){
#   risk<-readRDS(paste("Bags/predictionsSet",jan[i],".rds",sep=""))
#   jan.risk<-cbind(jan.risk,risk)
#   jan.risk<-rowSums(jan.risk)
# }
# feb.risk<-NULL
# for (i in 1:length(feb)){
#   risk<-readRDS(paste("Bags/predictionsSet",feb[i],".rds",sep=""))
#   feb.risk<-cbind(feb.risk,risk)
#   feb.risk<-rowSums(feb.risk)
# }
# mar.risk<-NULL
# for (i in 1:length(mar)){
#   risk<-readRDS(paste("Bags/predictionsSet",mar[i],".rds",sep=""))
#   mar.risk<-cbind(mar.risk,risk)
#   mar.risk<-rowSums(mar.risk)
# }
# apr.risk<-NULL
# for (i in 1:length(apr)){

```

```

# risk<-readRDS(paste("Bags/predictionsSet",apr[i],".rds",sep=""))
# apr.risk<-cbind(apr.risk,risk)
# apr.risk<-rowSums(apr.risk)
# }
# may.risk<-NULL
# for (i in 1:length(may)){
# risk<-readRDS(paste("Bags/predictionsSet",may[i],".rds",sep=""))
# may.risk<-cbind(may.risk,risk)
# may.risk<-rowSums(may.risk)
# }
# jun.risk<-NULL
# for (i in 1:length(jun)){
# risk<-readRDS(paste("Bags/predictionsSet",jun[i],".rds",sep=""))
# jun.risk<-cbind(jun.risk,risk)
# jun.risk<-rowSums(jun.risk)
# }
# jul.risk<-NULL
# for (i in 1:length(jul)){
# risk<-readRDS(paste("Bags/predictionsSet",jul[i],".rds",sep=""))
# jul.risk<-cbind(jul.risk,risk)
# jul.risk<-rowSums(jul.risk)
# }
# aug.risk<-NULL
# for (i in 1:length(aug)){
# risk<-readRDS(paste("Bags/predictionsSet",aug[i],".rds",sep=""))
# aug.risk<-cbind(aug.risk,risk)
# aug.risk<-rowSums(aug.risk)
# }
# sep.risk<-NULL
# for (i in 1:length(sep)){
# risk<-readRDS(paste("Bags/predictionsSet",sep[i],".rds",sep=""))
# sep.risk<-cbind(sep.risk,risk)
# sep.risk<-rowSums(sep.risk)
# }
# oct.risk<-NULL
# for (i in 1:length(oct)){
# risk<-readRDS(paste("Bags/predictionsSet",oct[i],".rds",sep=""))
# oct.risk<-cbind(oct.risk,risk)
# oct.risk<-rowSums(oct.risk)
# }
# nov.risk<-NULL
# for (i in 1:length(nov)){
# risk<-readRDS(paste("Bags/predictionsSet",nov[i],".rds",sep=""))
# nov.risk<-cbind(nov.risk,risk)
# nov.risk<-rowSums(nov.risk)
# }
# dec.risk<-NULL
# for (i in 1:length(dec)){
# risk<-readRDS(paste("Bags/predictionsSet",dec[i],".rds",sep=""))
# dec.risk<-cbind(dec.risk,risk)
# dec.risk<-rowSums(dec.risk)
# }
...

Get mean for each month across the 32 year period.
```{r}
jan.mean<-jan.risk/length(jan)
feb.mean<-feb.risk/length(feb)
mar.mean<-mar.risk/length(mar)
apr.mean<-apr.risk/length(apr)
may.mean<-may.risk/length(may)
jun.mean<-jun.risk/length(jun)
jul.mean<-jul.risk/length(jul)
aug.mean<-aug.risk/length(aug)
sep.mean<-sep.risk/length(sep)
oct.mean<-oct.risk/length(oct)
nov.mean<-nov.risk/length(nov)
dec.mean<-dec.risk/length(dec)
...

Take spillover intensity values for each month and turn them into a raster for each month.
```{r}
# jan.risk.r <-resampledRaster
# values(jan.risk.r)<-jan.mean
# feb.risk.r <-resampledRaster
# values(feb.risk.r)<-feb.mean
# mar.risk.r <-resampledRaster
# values(mar.risk.r)<-mar.mean
# apr.risk.r <-resampledRaster
# values(apr.risk.r)<-apr.mean
# may.risk.r <-resampledRaster
# values(may.risk.r)<-may.mean
# jun.risk.r <-resampledRaster
# values(jun.risk.r)<-jun.mean
# jul.risk.r <-resampledRaster
# values(jul.risk.r)<-jul.mean
# aug.risk.r <-resampledRaster
# values(aug.risk.r)<-aug.mean
# sep.risk.r <-resampledRaster
# values(sep.risk.r)<-sep.mean

```



```

# oct.risk.r <-resampledRaster
# values(oct.risk.r)<-oct.mean
# nov.risk.r <-resampledRaster
# values(nov.risk.r)<-nov.mean
# dec.risk.r <-resampledRaster
# values(dec.risk.r)<-dec.mean
```
Create a stack of Ebola spillover intensity by month.
```{r}
# risk.stack<-
stack(jan.risk.r,feb.risk.r,mar.risk.r,apr.risk.r,may.risk.r,jun.risk.r,jul.risk.r,aug.risk.r,sep.risk.r,oct.risk.r,nov
.risk.r,dec.risk.r)
# risk.brick<-brick(risk.stack)
# writeRaster(risk.brick,filename="risk.brick.grd",overwrite=T)
risk.brick<-brick("risk.brick.grd")
```
Create a dataframe of mean Ebola spillover by month and convert values to percentile.
```{r}
# mean.risk.df<-
cbind(jan.mean,feb.mean,mar.mean,apr.mean,may.mean,jun.mean,jul.mean,aug.mean,sep.mean,oct.mean,nov.mean,dec.mean)
# q<-ecdf(mean.risk.df)
# cum.risk.df<-q(mean.risk.df)
```
Take cumulative spillover intensity values for each month and turn them into a raster for each month.
```{r}
# jan.cum.r <-resampledRaster
# values(jan.cum.r)<-cum.risk.df[,1]
# feb.cum.r <-resampledRaster
# values(feb.cum.r)<-cum.risk.df[,2]
# mar.cum.r <-resampledRaster
# values(mar.cum.r)<-cum.risk.df[,3]
# apr.cum.r <-resampledRaster
# values(apr.cum.r)<-cum.risk.df[,4]
# may.cum.r <-resampledRaster
# values(may.cum.r)<-cum.risk.df[,5]
# jun.cum.r <-resampledRaster
# values(jun.cum.r)<-cum.risk.df[,6]
# jul.cum.r <-resampledRaster
# values(jul.cum.r)<-cum.risk.df[,7]
# aug.cum.r <-resampledRaster
# values(aug.cum.r)<-cum.risk.df[,8]
# sep.cum.r <-resampledRaster
# values(sep.cum.r)<-cum.risk.df[,9]
# oct.cum.r <-resampledRaster
# values(oct.cum.r)<-cum.risk.df[,10]
# nov.risk.r <-resampledRaster
# values(nov.cum.r)<-cum.risk.df[,11]
# dec.cum.r <-resampledRaster
# values(dec.cum.r)<-cum.risk.df[,12]
```
Create a stack of cumulative Ebola spillover intensity by month.
```{r}
# risk.stack2<-
stack(jan.cum.r,feb.cum.r,mar.cum.r,apr.cum.r,may.cum.r,jun.cum.r,jul.cum.r,aug.cum.r,sep.cum.r,oct.cum.r,nov.cum.r,dec
.cum.r)
# risk.brick2<-brick(risk.stack2)
# writeRaster(risk.brick2,filename="risk.brick2.grd",overwrite=T)
# risk.brick2<-brick("risk.brick2.grd")
```
Figure 1 - Map of known EVD events. Brown region indicates the focal region in Africa of annual rainfall > 600 mm.
Open circles indicate human spillovers, open triangles indicate infection/mortality in non-human primates or in other
mammals. Yellow, blue, green, magenta, and black indicate decade 1960-2010. Note that no Ebola spillovers are known
from the 1980s.
Get lat-long grid to overlay on map figures.
```{r}
latlonlines<-gridlines(risk.brick[[1]],east = c(-10,0,10,20,30,40,50),norths=c(10,0,-10,-20,-30))
equator<-Line(cbind(c(-20,60),c(0,0)))
Lab.IO<-list("sp.text",c(48,-5),expression(atop(italic(Indian),italic(Ocean))),fontsize=list(text=6),font=2,
col="lightblue4")
Lab.AtO<-list("sp.text",c(-5,-15),expression(atop(italic(Atlantic),italic(Ocean))),fontsize=list(text=6),font=2,
col="lightblue4")
```
Get political map to overlay on map figures.
```{r}
countries<-getMap()
countries<-countries[!is.na(countries$continent),]
countries<-countries[countries$continent=="Africa",]
lakes<-shapefile("ne_10m_lakes.shp")
```
Separate human from non-human index points.
```{r}
index.h<-index.pts[ which(index.pts$human=="TRUE"), ]
index.nh<-index.pts[ which(index.pts$human=="FALSE"), ]
```
Set the coordinate reference to lat-long, and turn coordinates from index.pts into spatial points.
```{r}
index.h.pts<-SpatialPoints(index.h[,1:2],proj4string=crs("+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84
+towgs84=0,0,0"))

```



```

xlab="Longitude",ylab="Latitude",
scales=list(x=list(draw=T,alternating=F),y=list(draw=T,alternating=F)),#layout=c(3,4),
par.settings=list(panel.background=list(col="lightblue1"),
layout.widths=list(axis.key.padding=1,ylab.right=4,ylab.left=4,right.padding=4),
layout.heights=list(top.padding=0.5,bottom.padding=0.5)),
axis.args=list(cex.axis=0.6),
sp.layout=list(list("sp.lines",latlonlines,col="lightblue3",lty=3,lwd=0.5,first=T),
list("sp.polygons",countries,fill="gray85",lwd=0),
list("sp.polygons",countries,lwd=0.1,first=F,col="gray60"),
list("sp.lines",equator,col="lightblue4",lty=3,lwd=2),
Lab.IO,Lab.AtO,
list("sp.text",c(-3,-32),months[i],fontsize=list(text=12),font=2,col="black")
)
)
print(risk.map)
grid.text("Predicted spillover intensity",x=0.98,y=0.5,rot=90,gp=gpar(fontsize=12))
}
}
...

Save the animation of Ebola intensity maps of Africa by month.
```{r}
saveGIF(gif(), movie.name="SeasonalEbolaRisk.gif", interval=1, ani.width=1000, ani.height=800)
```

Create map of average Ebola spillover intensity.
```{r}
average.risk<-mean(risk.brick2)
average.map<-spplot(average.risk,col.regions=colors,at=breaks,
colorkey=TRUE,
xlab="Longitude",ylab="Latitude",
scales=list(x=list(draw=T,alternating=F),y=list(draw=T,alternating=F)),#layout=c(3,4),
par.settings=list(panel.background=list(col="lightblue1"),
layout.widths=list(axis.key.padding=1,ylab.right=4,ylab.left=4,right.padding=4),
layout.heights=list(top.padding=0.5,bottom.padding=0.5)),
axis.args=list(cex.axis=0.6),
sp.layout=list(list("sp.lines",latlonlines,col="lightblue3",lty=3,lwd=0.5,first=T),
list("sp.polygons",countries,fill="gray85",lwd=0),
list("sp.polygons",countries,lwd=0.1,first=F,col="gray60"),
list("sp.lines",equator,col="lightblue4",lty=3,lwd=2),
Lab.IO,Lab.AtO,
list("sp.text",c(-3,-32),"Annual
average",fontsize=list(text=10),font=2,col="black")
)
)
print(average.map)
```

Add animated lines at top and right depict the marginal intensity by month.
Mask risk.brick by mean percentile rank to capture > 0.5.
```{r}
rc1<-function(x){
ifelse(x>0.5,1,
ifelse(x<=0.5,NA,NA))}
risk.mask50<-calc(average.risk,fun=rc1)
risk.brick.m<-mask(risk.brick2,risk.mask50)
```

Set figure dimensions.
```{r}
theme.novpadding <-
list(layout.heights =
list(top.padding = 0,
main.key.padding = 0,
key.axis.padding = 0,
axis.xlab.padding = 0,
xlab.key.padding = 0,
key.sub.padding = 0,
bottom.padding = 0),
axis.line = list(col = 0),
clip =list(panel="off"),
layout.widths =
list(left.padding = 0,
key.ylab.padding = 0,
ylab.axis.padding = 0,
axis.key.padding = 0,
right.padding = 0))
```

Loop to get the averages and calculate row and column averages by month.
```{r}
col.ave<-NULL
smooth.col.ave<-NULL
row.ave<-NULL
smooth.row.ave<-NULL
for (i in 1:12){
col.ave<-rbind(col.ave,colSums(risk.brick.m[[i]],na.rm=T)/colSums(!is.na(risk.brick.m[[i]])))
smooth.col.ave<-cbind(smooth.col.ave,loess(col.ave[i,]~seq(1:length(col.ave[i,])),span=0.3)$fitted)
row.ave<-rbind(row.ave,rowSums(risk.brick.m[[i]],na.rm=T)/rowSums(!is.na(risk.brick.m[[i]])))
smooth.row.ave<-cbind(smooth.row.ave,loess(row.ave[i,]~seq(1:length(row.ave[i,])),span=0.3)$fitted)
}
col.max<-max(smooth.col.ave,na.rm=T)
row.max<-max(smooth.row.ave,na.rm=T)

```

```

...
Code lines representing monthly marginal averages by latitude and longitude to gray.
```{r}
top.plot<-list()
right.plot<-list()
for(i in 1:12){
top.plot[[i]]<-xyplot(smooth.col.ave[,i]~seq(1:length(col.ave[i])),type="l",col="lightgray",
ylim=c(0.4,1),xlim=c(1,length(col.ave[i])),
ylab="",xlab="",scales=list(x=list(draw=F),y=list(draw=F,tck=c(1,0))),
par.settings=list(theme.nowpadding,axis.line = list(col = 0)),
panel=function(...){
lims <- current.panel.limits()
panel.xyplot(...)
})
right.plot[[i]]<-xyplot(smooth.row.ave[,i]~seq(1:length(row.ave[i])),type="l",col="lightgray",
ylim=c(0.4,1),xlim=c(1,length(row.ave[i])),
ylab="",xlab="",scales=list(x=list(draw=F),y=list(draw=F,tck=c(1,0))),
par.settings=list(theme.nowpadding,axis.line = list(col = 0)),
panel=function(...){
lims <- current.panel.limits()
panel.xyplot(...)
})
}
...

Create loop to generate animated figure with line corresponding to month represented in red.
```{r}
gif2<-function(){
for(i in 1:12){
risk.map<-spplot(risk.brick2[[i]],maxpixels=1000000,col.regions=colors,at=breaks,
colorkey=list(space="bottom"),
#main=list(label=months[i],cex=0.9,lines=1),
xlab=list("Longitude",cex=1.4),
ylab=list("Latitude",cex=1.4),
scales=list(y=list(draw=T,alternating=F,at=c(-40,-30,-20,-10,0,10,20)),x=
(list(draw=T,at=c(-10,0,10,20,30,40,50))),
axis.args=list(cex.axis=0.6),
#scales=list(x=list(draw=T,alternating=F),y=list(draw=T,alternating=F)),#layout=c(3,4),
par.settings=list(panel.background=list(col="lightblue1"),
layout.widths=list(axis.key.padding=0,ylab.right=0,ylab.left=0,right.padding=0,
top.padding = 0,
main.key.padding = 0,
axis.xlab.padding = 0,
xlab.key.padding = 0,
axis.ylab.padding = 0,
key.sub.padding = 0,
bottom.padding = 0),
layout.heights=list(top.padding = 0,
main.key.padding = 0,
key.axis.padding = 0,
axis.xlab.padding = 0,
xlab.key.padding = 4,
key.sub.padding = 4,
bottom.padding = 4,
axis.key.padding=4)),
axis.args=list(cex.axis=1.2),
sp.layout=list(list("sp.lines",latlonlines,col="lightblue3",lty=3,lwd=0.5,first=T),
list("sp.polygons",countries,fill="gray85",lwd=0),
list("sp.polygons",countries,lwd=1,first=F,col="gray50"),
list("sp.lines",equator,col="lightblue4",lty=3,lwd=2),
Lab.IO,Lab.AtO,
list("sp.polygons",lakes,fill="lightblue1",lwd=0),
list("sp.text",c(-3,-32),months[i],fontsize=list(text=24),font=2,col="black")
)
)
top.plot.current<-xyplot(smooth.col.ave[,i]~seq(1:length(col.ave[i])),type="l",col="maroon",
ylim=c(0,1),xlim=c(1,length(col.ave[i])),
ylab="",xlab="",scales=list(x=list(draw=F),y=list(draw=T,tck=c(1,0))),
par.settings=list(theme.nowpadding,axis.line = list(col = 0)))
right.plot.current<-xyplot(smooth.row.ave[,i]~seq(1:length(row.ave[i])),type="l",col="maroon",
ylim=c(0,1),xlim=c(1,length(row.ave[i])),
ylab="",xlab="",scales=list(x=list(draw=F),y=list(draw=T,tck=c(1,0))),
par.settings=list(theme.nowpadding,axis.line = list(col = 0)))
grid.newpage()
grid.text("Spatio-temporal dynamics of Ebola spillover intensity",x=0.45,y=0.96,gp=gpar(fontsize=28))
pushViewport(viewport(x = 0.0, y = 0,
height = 0.95, width = 1-0.15,
just = c("left", "bottom"),
name = "main"))
print(risk.map,newpage=F)
grid.text("Cumulative intensity (percentile)",x=0.53,y=0.15,gp=gpar(fontsize=16))
upViewport(1)
pushViewport(viewport(x = 0.05, y = .793,clip="inherit",
height = 0.15, width = .813,
just = c("left", "bottom"),
name = paste("t",i)))
print(top.plot[[1]]+top.plot[[2]]+top.plot[[3]]+top.plot[[4]]+top.plot[[5]]+top.plot[[6]]+top.plot[[7]]+top.plot[[8]]+top.plot[[9]]+top.plot[[10]]+top.plot[[11]]+top.plot[[12]])
y = c(0.26,0.42,0.58,0.74),gp=gpar(fontsize=10))

```

```

upViewport(1)
pushViewport(viewport(x = 0.815, y = 0.815,
height = 0.15, width = 0.59, angle = -90,
just = c("left", "bottom"),
name = paste("r", i)))
print(right.plot[[1]]+right.plot[[2]]+right.plot[[3]]+right.plot[[4]]+right.plot[[5]]+right.plot[[6]]+right.plot[[7]]+right.plot[[8]]+right.upViewport(1)
}
}
}
}

Save the animation of Ebola intensity maps of Africa by month with marginal intensity lines by latitude and longitude
at top and side.
```{r}
saveGIF(gif2(), movie.name = "Seasonal Ebola risk2.gif", interval = 1, ani.width = 1000, ani.height = 1000)
```

Figure 3 - Phase graph showing the relationship between mean monthly rainfall and raw EVD spillover intensity, the
average density or expected number of points per unit area and/or time, for known EVD locations in West and central
Africa (closed circles) and locations in northeastern or southern Africa where model results indicate moderate to
high EVD spillover intensity seasonally (open squares). Points are ordered by least to greatest monthly rainfall at
each site.
Get lat-long coordinates of locations of interest.
```{r}
library(ggmap)
Meliandou<-geocode("meliandou guinea", source="google", output="latlon")
Yambuku<-geocode("yambuku democratic republic of congo", source="google", output="latlon")
Makouke<-geocode("makouke gabon", source="google", output="latlon")
Nzara<-geocode("nzara south sudan", source="google", output="latlon")
Lowero<-geocode("lowero uganda", source="google", output="latlon")
Tai_Forest<-geocode("tai cote d'ivoire", source="google", output="latlon")
Songo<-geocode("songo angola", source="google", output="latlon")
Mwinilunga<-geocode("mwinilunga zambia", source="google", output="latlon")
Masasi<-geocode("masasi tanzania", source="google", output="latlon")
Maxixe<-geocode("maxixe mozambique", source="google", output="latlon")
Jimma<-geocode("jimma ethiopia", source="google", output="latlon")
Lakato<-geocode("lakato madagascar", source="google", output="latlon")
```

List places, names, months.
```{r}
places<-rbind(Meliandou, Yambuku, Makouke, Nzara, Lowero, Tai_Forest, Songo, Mwinilunga, Masasi, Maxixe, Jimma, Lakato)
names<-c("Meliandou", "Yambuku", "Makouke", "Nzara", "Lowero", "Tai_Forest", "Songo", "Mwinilunga", "Masasi", "Maxixe", "Jimma", "Lakato")
places<-cbind(places, names)
months<-c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec")
month.num<-c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)
```

Extract average risk and average rainfall by month for each site.
```{r}
mel.risk<-extract(risk, places[1, 1:2])
mel.rf<-extract(af.rf.brick, places[1, 1:2])
mel<-as.data.frame(cbind(t(mel.risk), t(mel.rf)))
tai.risk<-extract(risk, places[6, 1:2])
tai.rf<-extract(af.rf.brick, places[6, 1:2])
tai<-as.data.frame(cbind(t(tai.risk), t(tai.rf)))
mak.risk<-extract(risk, places[3, 1:2])
mak.rf<-extract(af.rf.brick, places[3, 1:2])
mak<-as.data.frame(cbind(t(mak.risk), t(mak.rf)))
yam.risk<-extract(risk, places[2, 1:2])
yam.rf<-extract(af.rf.brick, places[2, 1:2])
yam<-as.data.frame(cbind(t(yam.risk), t(yam.rf)))
nza.risk<-extract(risk, places[4, 1:2])
nza.rf<-extract(af.rf.brick, places[4, 1:2])
nza<-as.data.frame(cbind(t(nza.risk), t(nza.rf)))
low.risk<-extract(risk, places[5, 1:2])
low.rf<-extract(af.rf.brick, places[5, 1:2])
low<-as.data.frame(cbind(t(low.risk), t(low.rf)))
son.risk<-extract(risk, places[7, 1:2])
son.rf<-extract(af.rf.brick, places[7, 1:2])
son<-as.data.frame(cbind(t(son.risk), t(son.rf)))
mwi.risk<-extract(risk, places[8, 1:2])
mwi.rf<-extract(af.rf.brick, places[8, 1:2])
mwi<-as.data.frame(cbind(t(mwi.risk), t(mwi.rf)))
mas.risk<-extract(risk, places[9, 1:2])
mas.rf<-extract(af.rf.brick, places[9, 1:2])
mas<-as.data.frame(cbind(t(mas.risk), t(mas.rf)))
max.risk<-extract(risk, places[10, 1:2])
max.rf<-extract(af.rf.brick, places[10, 1:2])
max<-as.data.frame(cbind(t(max.risk), t(max.rf)))
jim.risk<-extract(risk, places[11, 1:2])
jim.rf<-extract(af.rf.brick, places[11, 1:2])
jim<-as.data.frame(cbind(t(jim.risk), t(jim.rf)))
lak.risk<-extract(risk, places[12, 1:2])
lak.rf<-extract(af.rf.brick, places[12, 1:2])
lak<-as.data.frame(cbind(t(lak.risk), t(lak.rf)))
```

Make phase graph of average rainfall by month (x-axis) and average Ebola spillover intensity by month.
```{r}
par(mar=c(4, 4, 1, 1))
mel <- mel[order(mel$V2), ]

```



```

plot(mel$V2,mel$V1,type='o',col="red",pch=16,xlim=c(0,420),ylim=c(0.01,0.43),xlab="Rainfall (mm)",ylab="Ebola
spillover intensity")
text(380,0.2,"Meliandou, Guinea",col="red",cex=0.85)
par(new=TRUE)
tai<-tai[order(tai$V2),]
plot(tai$V2,tai$V1,type='o',col="blue",pch=16,xlim=c(0,420),ylim=c(0.01,0.43),xlab="",ylab="",xaxt='n',yaxt='n')
text(200,0.17,"Tai Forest, Cote d'Ivoire",col="blue",cex=0.85)
par(new=TRUE)
mak<-mak[order(mak$V2),]
plot(mak$V2,mak$V1,type='o',col="orange",pch=16,xlim=c(0,420),ylim=c(0.01,0.43),xlab="",ylab="",xaxt='n',yaxt='n')
text(380,0.135,"Makouke, Gabon",col="orange",cex=0.85)
par(new=TRUE)
yam<-yam[order(yam$V2),]
plot(yam$V2,yam$V1,type='o',col="darkgreen",pch=16,xlim=c(0,420),ylim=c(0.01,0.43),xlab="",ylab="",xaxt='n',yaxt='n')
text(260,0.22,"Yambuku, DRC",col="darkgreen",cex=0.85)
par(new=TRUE)
nza<-nza[order(nza$V2),]
plot(nza$V2,nza$V1,type='o',col="brown",pch=16,xlim=c(0,420),ylim=c(0.01,0.43),xlab="",ylab="",xaxt='n',yaxt='n')
text(270,0.083,"Nzara, South Sudan",col="brown",cex=0.85)
par(new=TRUE)
low<-low[order(low$V2),]
plot(low$V2,low$V1,type='o',col="purple",pch=16,xlim=c(0,420),ylim=c(0.01,0.43),xlab="",ylab="",xaxt='n',yaxt='n')
text(210,0.25,"Lowero, Uganda",col="purple",cex=0.85)
par(new=TRUE)
jim<-jim[order(jim$V2),]
plot(jim$V2,jim$V1,type='o',col="red",pch=22,xlim=c(0,420),ylim=c(0.01,0.43),xlab="",ylab="",xaxt='n',yaxt='n')
text(277,0.115,"Jimma, Ethiopia",col="red",cex=0.85)
par(new=TRUE)
son <- son[order(son$V2),]
plot(son$V2,son$V1,type='o',col="blue",pch=22,xlim=c(0,420),ylim=c(0.01,0.43),xlab="",ylab="",xaxt='n',yaxt='n')
text(240,0.10,"Songo, Angola",col="blue",cex=0.85)
par(new=TRUE)
mwi<-mwi[order(mwi$V2),]
plot(mwi$V2,mwi$V1,type='o',col="orange",pch=22,xlim=c(0,420),ylim=c(0.01,0.43),xlab="",ylab="",xaxt='n',yaxt='n')
text(325,0.03,"Mwinilunga, Zambia",col="orange",cex=0.85)
par(new=TRUE)
mas<-mas[order(mas$V2),]
plot(mas$V2,mas$V1,type='o',col="darkgreen",pch=22,xlim=c(0,420),ylim=c(0.01,0.43),xlab="",ylab="",xaxt='n',yaxt='n')
text(160,0.025,"Masisi, Tanzania",col="darkgreen",cex=0.85)
par(new=TRUE)
max<-max[order(max$V2),]
plot(max$V2,max$V1,type='o',col="purple",pch=22,xlim=c(0,420),ylim=c(0.01,0.43),xlab="",ylab="",xaxt='n',yaxt='n')
text(195,0.06,"Maxixe, Mozambique",col="purple",cex=0.85)
par(new=TRUE)
lak<-lak[order(lak$V2),]
plot(lak$V2,lak$V1,type='o',col="brown",pch=22,xlim=c(0,420),ylim=c(0.01,0.43),xlab="",ylab="",xaxt='n',yaxt='n')
text(340,0.36,"Lakato, Madagascar",col="brown",cex=0.85)
...

Bring in brick of monthly Ebola intensity predicted on the full data set, but using 1975 rather than 2015 population.
```{r}
#risk.brick75<-brick("risk.brick75.grd")
...

Get mean Ebola spillover intensity across months for 1975 and 2015.
```{r}
mean.risk75<-mean(risk.brick75)
mean.risk15<-mean(risk.brick)
...

Get difference in average annual risk from 1975 to 2015.
```{r}
risk.dif<-mean.risk2015-mean.risk75
...

Figure 4 - Change in annual spillover intensity, the average density or expected number of points per unit area
and/or time, between 1975 and 2015. Warm colors indicate increased spillover intensity, cool colors indicate
decreased spillover intensity.
Set color palette.
```{r}
risk.colors<-colorRampPalette(c("darkgreen","white","brown"))
colors<-risk.colors(8)
breaks<-c(-0.232,-0.15,-0.10,-0.05,0,0.05,0.10,0.15,0.224)
colorkey<-list(at=c(-0.15,-0.10,0,0.10,0.15),labels=list(at=c(-0.15,-0.10,0,0.10,0.15),cex=1.2),tick.number=5)
...
```{r}
risk.dif.map<-spplot(risk.dif,maxpixels=1000000,col.regions=colors,at=breaks,
colorkey=TRUE,
xlab=list("Longitude",cex=1.4),
ylab=list("Latitude",cex=1.4),
scales=list(x=list(draw=T,alternating=F),y=list(draw=T,alternating=F)),
par.settings=list(panel.background=list(col="lightblue1"),
layout.widths=list(axis.key.padding=1,ylab.right=4,ylab.left=4,right.padding=4),
layout.heights=list(top.padding=0.5,bottom.padding=0.5)),
axis.args=list(cex.axis=0.6),
sp.layout=list(list("sp.lines",latlonlines,col="lightblue3",lty=3,lwd=0.6,first=T),
list("sp.polygons",countries,fill="gray92",lwd=0),
list("sp.polygons",countries,lwd=0.5,first=F,col="gray60"),
list("sp.polygons",lakes,fill="lightblue1",lwd=0),
list("sp.lines",equator,col="lightblue4",lty=3,lwd=2),
Lab.IO,Lab.At0)
)

```

```

print(risk.dif.map)
```
Get delta population by taking the log10 of the absolute difference between population counts per 25 km-2 grid cell 1975-2015.
```{r}
pd.delta<-risk.dif
values(pd.delta)<-as.matrix(log10(abs(10^pop15-10^pop75))*(10^pop15-10^pop75)/abs(10^pop15-10^pop75))
writeRaster(pd.delta,"pd.delta.tif")
pd.delta<-raster("pd.delta.tif")
```

```

Figure 5 - Log10 (change in human population size) per 25 km-2 grid cell 1975-2015. Warm colors indicate increased population size, cool colors indicate population declines.
Set color palette.

```

```{r}
risk.colors<-colorRampPalette(c("darkgreen","white","brown"))
colors<-risk.colors(14)
breaks<-c(-7.474,-6,-5,-4,-3,-2,-1,0,1,2,3,4,5,6,7.66)
colorkey<-
list(at=c(-7.474,-6,-5,-4,-3,-2,-1,0,1,2,3,4,5,6,7.66),labels=list(at=c(-7.474,-6,-5,-4,-3,-2,-1,0,1,2,3,4,5,6,7.66),cex=1.2),tick.number=``)
Plot map of change in population 1975-2015.
```{r}
delta.pd.map<-spplot(pd.delta,maxpixels=1000000,col.regions=colors,at=breaks,
colorkey=TRUE,
xlab=list("Longitude",cex=1.4),
ylab=list("Latitude",cex=1.4),
scales=list(x=list(draw=T,alternating=F),y=list(draw=T,alternating=F)),
par.settings=list(panel.background=list(col="lightblue1"),
layout.widths=list(axis.key.padding=1,ylab.right=4,ylab.left=4,right.padding=4),
layout.heights=list(top.padding=0.5,bottom.padding=0.5)),
axis.args=list(cex.axis=0.6),
sp.layout=list(list("sp.lines",latlonlines,col="lightblue3",lty=3,lwd=0.6,first=T),
list("sp.polygons",countries,fill="gray92",lwd=0),
list("sp.polygons",countries,lwd=0.5,first=F,col="gray60"),
list("sp.polygons",lakes,fill="lightblue1",lwd=0),
list("sp.lines",equator,col="lightblue4",lty=3,lwd=2),
Lab.IO,Lab.AtO)
)
print(delta.pd.map)
```

```

Figure S1 - Population density and Ebola spillover locations for the masked region of Africa (> 600 mm annual rainfall) for each decade 1960-2010. Legend numbers represent human population ranges as powers of 10 per 25 km2. Red circles mark human spillovers, orange circles non-human primate and other mammal spillovers.

Set population breakpoints of log10 values per cell.

```

```{r}
breakpoints <- c(0,1,2,3,4,5,6)

Set colors at breakpoints used to bin population counts in bagged regression.
```{r}
colors <- c("gray88","gray88","gray76","gray55","gray55","gray55")

Get 44 final human and non-human Ebola points and separate by decade.
```{r}
index.h60<-index.h[ which(index.h$decade == 6), ]
index.h70<-index.h[ which(index.h$decade == 7), ]
index.nh70<-index.nh[ which(index.nh$decade == 7), ]
index.h90<-index.h[ which(index.h$decade == 9), ]
index.nh90<-index.nh[ which(index.nh$decade == 9), ]
index.h00<-index.h[ which(index.h$decade == 0), ]
index.nh00<-index.nh[ which(index.nh$decade == 0), ]
index.h10<-index.h[ which(index.h$decade == 1), ]
index.nh10<-index.nh[ which(index.nh$decade == 1), ]
```

```

Make trellis map figure of population density and Ebola spillovers.

```

```{r}
par(mfrow=c(3,2))
par(mar=c(0.5,0.5,0.5,0.5),bty="n")
logpop60<-raster("logpop60.tif")
plot(logpop60,breaks=breakpoints,col=colors,ylim=c(-35,15),xlim=c(-20,50),xaxt="n",yaxt="n",legend=T)
points(index.h60$x,index.h60$y,pch=1,col="red",cex=1.5)
title("1960", line = -0.5)
par(mar=c(0.5,0.5,0.5,0.5),bty="n")
logpop70<-raster("logpop70.tif")
plot(logpop70,breaks=breakpoints,col=colors,ylim=c(-35,15),xlim=c(-20,50),xaxt="n",yaxt="n",legend=FALSE)
points(index.nh70$x,index.nh70$y,pch=1,col="orange")
points(index.h70$x,index.h70$y,pch=1,col="red",cex=1.5)
title("1970", line = -0.5)
par(mar=c(0.5,0.5,0.5,0.5),bty="n")
logpop80<-subset(popbrick,1)
plot(logpop80,breaks=breakpoints,col=colors,ylim=c(-35,15),xlim=c(-20,50),xaxt="n",yaxt="n",legend=FALSE)
title("1980", line = -0.5)
par(mar=c(0.5,0.5,0.5,0.5),bty="n")
logpop90<-subset(popbrick,8)
plot(logpop90,breaks=breakpoints,col=colors,ylim=c(-35,15),xlim=c(-55,50),xaxt="n",yaxt="n",legend=FALSE)
points(index.nh90$x,index.nh90$y,pch=1,col="orange",cex=1.5)
points(index.h90$x,index.h90$y,pch=1,col="red",cex=1.5)
title("1990", line = -0.5)
par(mar=c(0.5,0.5,0.5,0.5),bty="n")

```

```

logpop00<-subset(popbrick,18)
plot(logpop00,breaks=breakpoints,col=colors,ylim=c(-35,15),xlim=c(-20,50),xaxt="n",yaxt="n",legend=FALSE)
points(index.nh00$x,index.nh00$y,pch=1,col="orange",cex=1.5)
points(index.h00$x,index.h00$y,pch=1,col="red",cex=1.5)
title("2000", line = -0.5)
par(mar=c(0.5,0.5,0.5,0.5),bty="n")
logpop10<-subset(popbrick,28)
plot(logpop10,breaks=breakpoints,col=colors,ylim=c(-35,15),xlim=c(-20,50),xaxt="n",yaxt="n",legend=FALSE)
points(index.nh10$x,index.nh10$y,pch=1,col="orange",cex=1.5)
points(index.h10$x,index.h10$y,pch=1,col="red",cex=1.5)
title("2010", line = -0.5)
...

```

Figure S2 - ROC curves showing performance of models resulting from analyses using all spillover points, predicted on 2/3 training data (top), and predicted on holdout 1/3 test data (middle). Bottom row shows performance of models resulting from analyses using human spillover points only predict on training data. Left and right columns compare results from 2 different sets of background points. Filled polygons represent the area under the ROC curve (AUC), red arcs are smoothed curves.

Training on test: all points.

```

...{r}
library(ROCR)
library(pROC)
test <- read.table("test.csv",header = T, sep=",",na.strings=c("na","NA"))
pred <- prediction(test[,1],test[,2])
perf <- performance(pred,"sens","spec")
plot(perf)
...

Training on training: human spillovers only.
...{r}
test.h<-read.table("test.h.csv",header = T, sep=",",na.strings=c("na","NA"))
pred2 <- prediction(test.h[,1],test.h[,2])
perf2 <- performance(pred2,"sens","spec")
plot(perf2)
...

Training on training: all points.
...{r}
test.a<-read.table("test.all.csv",header = T, sep=",",na.strings=c("na","NA"))
pred3 <- prediction(test.a[,1],test.a[,2])
perf3 <- performance(pred3,"sens","spec")
plot(perf3)
...

Create paired graphs.
...{r}
par(mfrow=c(3,2))
test.a<-read.table("test.all.csv",header = T, sep=",",na.strings=c("na","NA"))
rocobj3<-roc(test.a[,2],test.a[,1])
plot(rocobj3,col="blue",print.auc=T,main=" ",print.auc.cex=1.15, auc.polygon=T)
plot.roc(smooth(rocobj3), add=TRUE, col="red")
title("All Spillovers:", line=3,adj=1)
test.a2<-read.table("test.all2.csv",header = T, sep=",",na.strings=c("na","NA"))
rocobj4<-roc(test.a2[,2],test.a2[,1])
plot(rocobj4,col="blue",print.auc=T,main=" ",print.auc.cex=1.15, auc.polygon=T)
plot.roc(smooth(rocobj4), add=TRUE, col="red")
title("Training on Training", line=3,adj=0)
test <- read.table("test.csv",header = T, sep=",",na.strings=c("na","NA"))
rocobj<-roc(test[,2],test[,1])
plot(rocobj,col="blue",print.auc=T,main=" ",print.auc.cex=1.15, auc.polygon=T)
plot.roc(smooth(rocobj), add=TRUE, col="red")
title("All Spillovers:", line=3,adj=1)
test2 <- read.table("test2.csv",header = T, sep=",",na.strings=c("na","NA"))
rocobj5<-roc(test2[,2],test2[,1])
plot(rocobj5,col="blue",print.auc=T,main=" ",print.auc.cex=1.15, auc.polygon=T)
plot.roc(smooth(rocobj5), add=TRUE, col="red")
title("Training on Test", line=3,adj=0)
test.h<-read.table("test.h.csv",header = T, sep=",",na.strings=c("na","NA"))
rocobj2<-roc(test.h[,2],test.h[,1])
plot(rocobj2,col="blue",print.auc=T,main=" ",print.auc.cex=1.15, auc.polygon=T)
plot.roc(smooth(rocobj2), add=TRUE, col="red")
title("Human Spillovers:", line=3,adj=1)
test.h2<-read.table("test.h2.csv",header = T, sep=",",na.strings=c("na","NA"))
rocobj6<-roc(test.h2[,2],test.h2[,1])
plot(rocobj6,col="blue",print.auc=T,main=" ",print.auc.cex=1.15, auc.polygon=T)
plot.roc(smooth(rocobj6), add=TRUE, col="red")
title("Training on Training", line=3,adj=0)
...

```