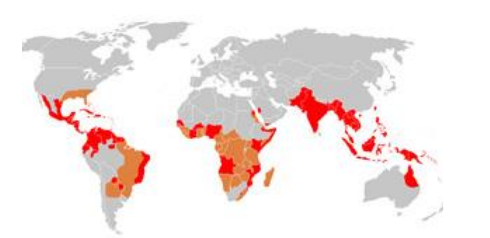
The Role of Human Mobility in Dengue Fever Dynamics in Mexico

Yang Liu (liux3204@umn.edu)[1], Matteo Convertino, Ph.D.[1, 2, 3, 4]
[1]Environmental Health Sciences Division, School of Public Health, University of Minnesota TC;
[2]Institute on the Environment, University of Minnesota TC;
[3]Institute for Engineering in Medicine, University of Minnesota TC; [4]Public Health Informatics Program, University of Minnesota TC

Background

Dengue Fever (DF) is a mosquito-borne infectious disease. The overall global burden of disease is 22,000 deaths and 390 cases annually. The transmission of DF primarily relies on the female *Aedes aegypti*.


Source: Nature

The majority of the existing literature on modeling of DF dynamics is focuses on utilizing environmental information (e.g. rainfall and temperature) as they are the key variables affecting vector activities. Under such framework, DF related public health essentially is an ecosystem service. Understanding the local ecosystem dynamics will provide valuable information to DF management.

However, there are relatively few studies that consider both environmental and other perspectives, such as the network of populations. As an infectious disease, DF at one location may largely be the result of the disease dynamics in its surroundings. Thus, the social and environmental connectivity information may be able to further improve the quality of DF models. In this study, we plan on using information theory and other spatial analysis techniques to understand the role of human mobility in the transmission of DF.

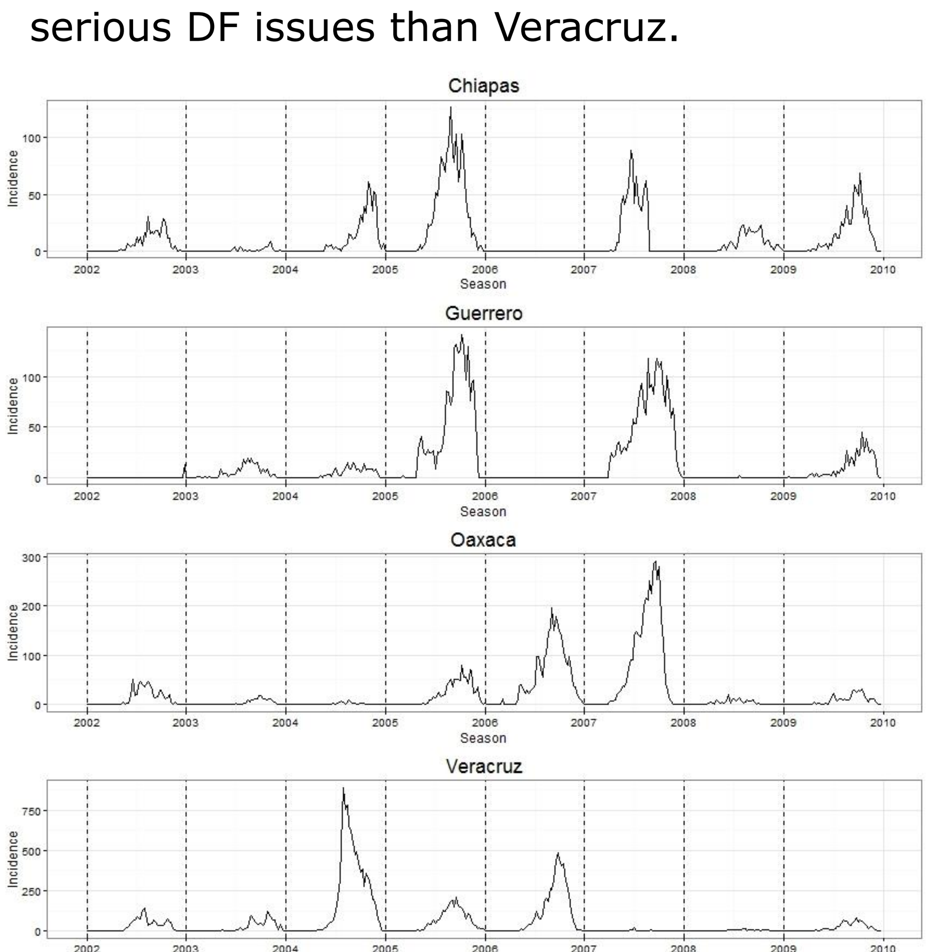
Research Question (1)
How does human mobility contribute to the mutual variability of DF activities at distinctive locations?

Research Question (2)
Is there evidence of functional connectivity between distinctive states based on information theory?

Data

The study sites in this study are four southern coastal Mexican states: Chiapas, Guerrero, Oaxaca and Veracruz.

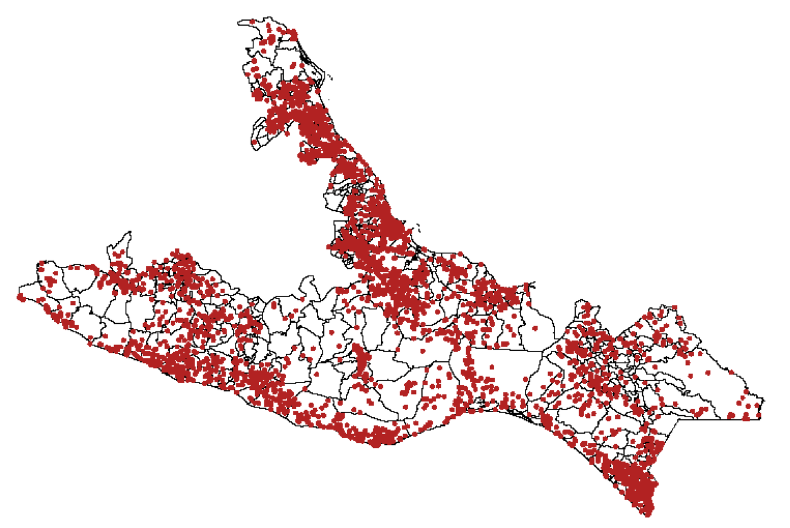
weekly incidence falls to zero outside of disease seasons. Between 2002 and 2009, Veracruz had the most total cases but is also the most populous state. After we take sizes of population into consideration, all other three states seemed to have more serious DF issues than Veracruz.



DF seasons are seen more frequently during the second half of the year, which coincides with the rainy season in Mexico. However, epidemics do not occur in all states all the time. In most seasons, only 2-3 states experienced outbreaks with some temporal synchrony.

Data

2005 was the only season in which all four states experienced large-scale outbreaks. Among the 10 districts with the highest DF cumulative incidence proportion, four were from Oaxaca. Only one of these district was from Chiapas.



Method

The association between two time series are estimated by two measurements developed in information theory, which is founded around the concept of entropy:

$$H(X) = \sum_{x \in X} p(X) \log p(X)$$

where X is a random variable, and the probability mass function of X is p(X). It describes the information contains in each message. Extending this concept further, the first measurement we rely on is mutual information (MI) is:

$$I(X,Y) = \sum_{x} \sum_{y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

Where X and Y are two random variables. MI is the amount of information that one variable contains about the other. It is a unit-less and symmetric measurement. While working with time series data, MI can be calculated at different lag times. To make MI more standardized, comprehensible and intuitive, we convert MI to statistical distance:

$$d(X,Y) = e^{-I(X,Y)}$$

There is a second measurement of dependency: Conditional Entropy. CE measures the amount of uncertainty reduced about random variable Y due to the knowledge of X:

$$H(Y|X) = -\sum_{x} \sum_{y} p(x,y) \log p(y|x)$$

Beside the existence of an association, CE can also describe the strength of the association.

For functional network inference, we further rely on Transfer Entropy (TE). It is a measurement of uncertainty reduced about random variable Y due to the past values of Y as well as the past values of the other random variable X.

$$T_{X \to Y} = H(Y^{\tau}|Y^{t-\tau}) - H(Y^{\tau}|Y^{t-\tau}, X^{t-\tau})$$

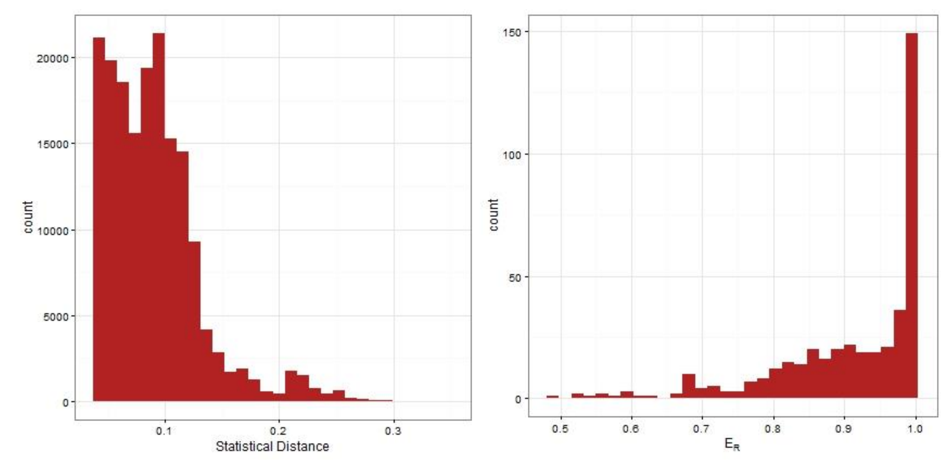
The information of mobility in Mexico is more difficult to obtain. There is no existing high-quality dataset to our knowledge. Therefore, this information will be reconstructed from the classic gravity mobility model:

$$M_{i,j} = \frac{C * P_i^{\alpha} * P_j^{\gamma}}{f(D_{i,j})}$$

where $M_{X,Y}$ is the magnitude of mobility between the origin (donor) of travelers X and the destination (recipient) of travelers Y. P_X and P_Y are population at the two locations and $D_{X,Y}$ is the Euclidian distance between the two locations. The function $f(D_{X,Y})$ is a transportation friction function.

Results

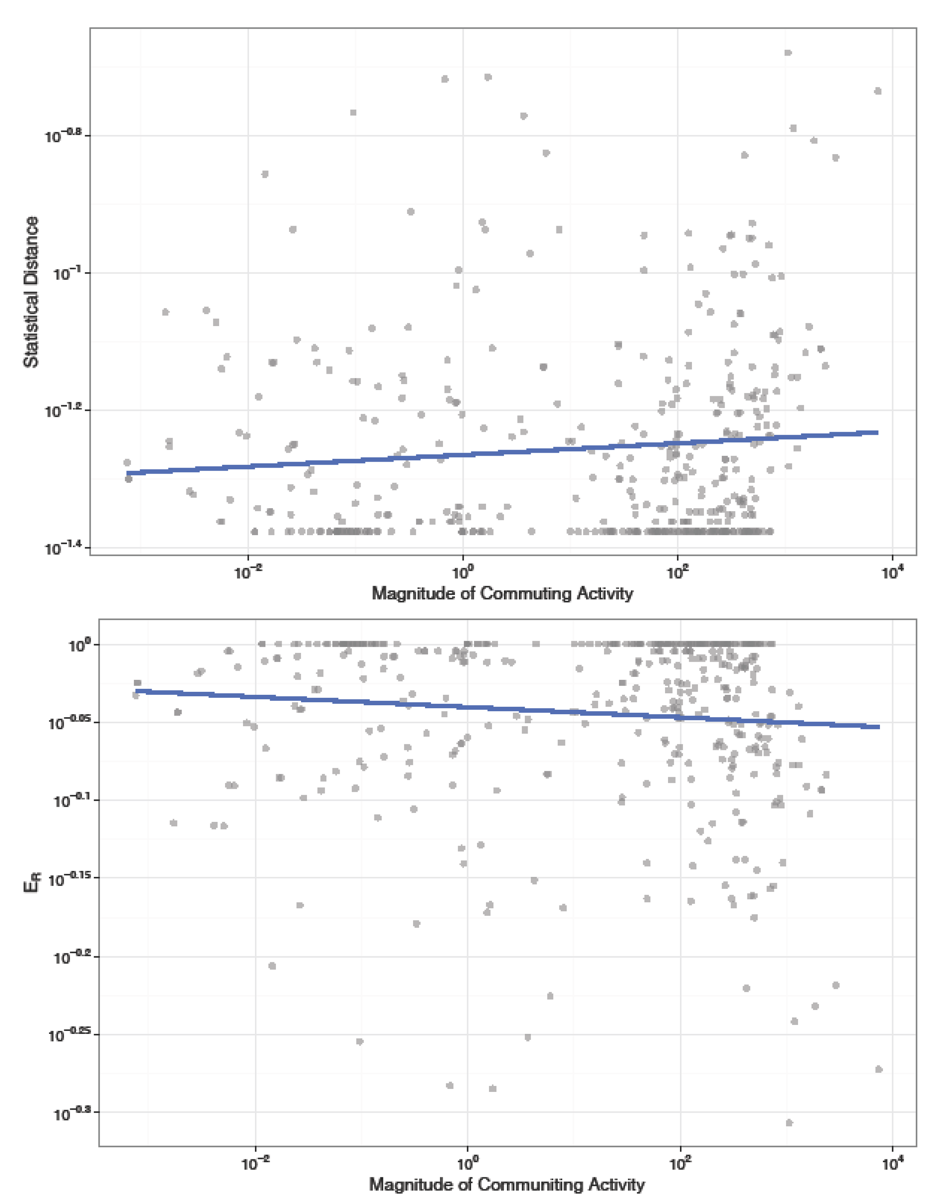
We calculated the *d(X,Y)* and E_R (Entropy Reduction parameter) for 173,472 pairs of DF incidence between 417 sub-municipal districts given 4 weeks lag time. That is to say, for values in the time series of Y at time t, t-4~t are examined for the time series X.



d(X,Y) is an indicator of how likely two time series are associated with each other. The lower *d(X,Y)* is, the closer the distribution of two time series are. In this analysis, over 66% pairs we tested have *d(X,Y)* below 0.1. E_R can verify the results. Entropy reduction process is achieved using MatLab MIDER Toolbox developed by Villaverde et al. The results show that only 0.2% of the total edges are significant associations.

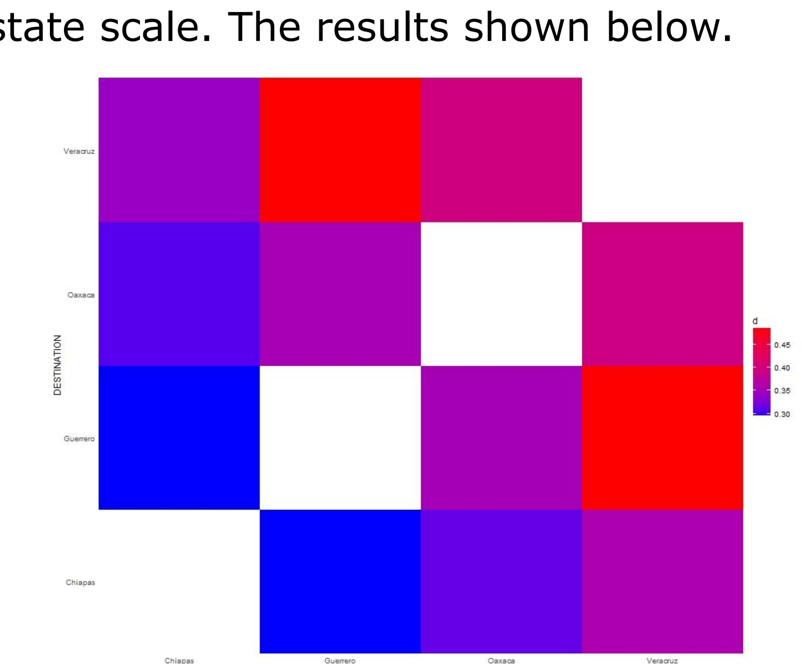
In this study, we assume that the transportation friction function to be an exponential-based. We borrow the parameters from Balcan et al.; these have already been calibrated and validated using data from 40 countries around the world. It assumes a threshold at 300km.

Parameters	<=300km	>300km
α	0.46	0.64
γ	0.35	0.37
$f(D_{X,Y})$	$e^{\beta * D_{X,Y}}$ where β=0.01	



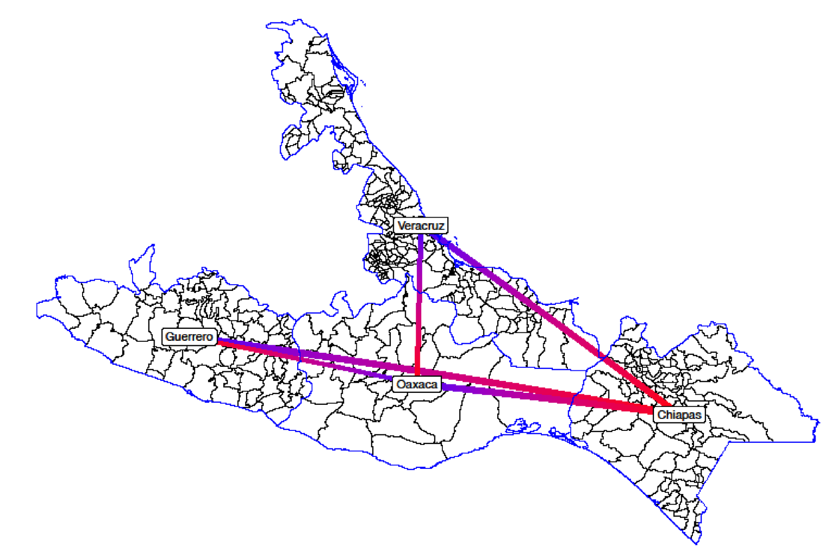
There are statistically significant scaling relationships between the magnitude of mobility and statistical distance. There is positive scaling between *d(X,Y)* and mobility. It implies that between districts that have more travelers back and forth, the probability distribution of DF weekly incidence are further away from each other and they are less likely to contain information about each other. There are also statistically significant relationship between magnitude of commuting activity and E_R. Negative scaling is observed. The more smaller the mobility is, the less likely the pair of DF incidence can explain each other.

We did the same calculation for DF on the state scale. The results shown below.



Results

We noticed that *d(X,Y)* is the lowest between Guerrero and Chiapas. They are not geographic neighbors. Veracruz and Guerrero, which are in fact much closer to each other, are much less associated with each other. This may reveal preferential connections between states in relation to directionality of mobility of other migration fluxes from bordering states.



There is a functional network that can be derived from the time series using MIDER, shown above. Chiapas is the biggest source nodes and Oaxaca and Veracruz are the biggest sink nodes. The three connections coming out of Chiapas are stronger than the other two.

Conclusions

Research Question (1)
When mobility is low between two districts, both statistical distance and entropy reduction parameters indicate that their DF weekly incidence are more likely to be associated with each other. It is likely that they are simultaneously affected by variation of the region's natural environment. When mobility is high between two regions, the association between DF dynamics is smaller. This is likely a result of complex interactions between population characteristics (e.g. immunity), the environment and the surrounding disease conditions. When we increase the scale of analysis, relationships disappear.

Research Question (2)
From the data, if we consider multiple time lags, there are significant functional connections between states, which indicates that at least part of the variability in DF at the state level can be explained by the DF activity in another state. Human mobility is most likely the reason as mosquitoes are not capable of traveling long-distance.

References

Villaverde, A. F., Ross, J., Morán, F., & Banga, J. R. (2014). MIDER: Network inference with mutual information distance and entropy reduction. *PLoS ONE*, *9*(5), 1–15.

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., & Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(51), 21484–21489.

Davidson, M. W., Haim, D. a, & Radin, J. M. (2015). Using networks to combine "big data" and traditional surveillance to improve influenza predictions. *Scientific Reports*, *5*(8154), 1–5.

SEDAC, & CIESIN. (2015). Gridded Population of the World Version 4. *Center for International Earth Science Information Network*, (November), 1–21.

Liu Y., Convertino M., The Role of Human Mobility in the Transmission of Dengue Fever in Southern Mexico. (Work in Progress.)

Acknowledgement:
Y.L. thanks the support from the Discovery Grant (PI: Convertino). Funding from the "Programa de Matemáticas Aplicadas y Computación" at the Instituto de Matemáticas UNAM (PI: JX Velasco-Hernandez) are also acknowledged. Data are from Jose Ramos Castaneda at National Institute of Public Health of Mexico (INSP).