Zika in Brazil
how weather, population growth, and sanitation have impacted the development of microcephaly cases linked to Zika

Mario Javier Carrillo.[*]

JULY 11, 2016

[*]Affiliation: General Assembly. Email: majacaci00@gmail.com

1 Background

Zika, discovered in 1947 in Uganda is a flavivirus transmitted by the genus *Aedes* that has been found as the cause of human infections across Africa and Asia. In 2007 a massive outbreak of Zika was reported on the Island of Yap (Micronesia), from where the virus migrated to Southeast Asia, landing in the French Polynesia in 2013-14. In South America, more specifically in Brazil, the first reports of locally transmitted infection were reported in May 2015, and in October 2015, the Brazilian Government reported their first case of microcephaly related to Zika (Calvet et al., 2016; Kindhauser, Allen, Frank, Santhana, & Dye, 2016).

By February 2016 the infection moved rapidly through Latin America, and the rapid rate of increase of microcephaly cases related with Zika in Brazil prompted WHO to declare zika infections associated with microcephaly and other neurological disorders a "public health emergency of international concern" (WHO, 2016).

Although researchers have so far not proven a definite link between microcephaly cases in Brazil and Zika, the extent of which it has been studied included changes on temperature, deforestation and more. (E. Petersen et al., 2016). Currently, there is no available treatment or vaccine for zika, therefore; disease control is restricted to mosquito management via insecticides or larva breeding destruction(Yakob & Walker, 2016). Moreover, the lack of detail and credible information on how individuals, primarily adults, acquired the virus makes the prevention and treatment a challenge that has open the door for multiple theories and explanations.

This case study tries to explain how weather conditions from January 2015 to May 2016, projected 2015 and 2016 total population of men and women within a reproductive age (15-44), prevalence of microcephaly cases, growth rate of microcephaly, and sanitation and demographic characteristics of the 27 Brazilian states have influenced the increase of microcephaly confirmed reported cases linked to zika from February 2016 to May 2016. To describe and report variables/features with greater emphasis on microcephaly, the study uses linear regression, lasso and ridge methods, regression trees, random forest regression and gradient boosting regressor.

2 Data

Perhaps one of the most challenges points of this case study has to do with the construction of the multiple datasets, where the quality of the collected data (especially the reported microcephaly cases), and lack of complimentary datasets at the state level, have the potential to introduce error or bias into the analysis.

The final dataset contains the microcephaly confirmed cases collected by the CDC epidemic prevention initiative GitHub repository (CDC, 2016), which were validated against the weekly reports of Zika from the Minesterio da Saude of Brazil (Ministério da Saúde - Brazil, 2016). I complemented this dataset with climate data from the Iowa Environmental Mesonet (The Iowa Environmental Mesonet, 2016), and urban characteristics of household surroundings from the Instituto Brasileiro de Geografia e Estatistica (Instituto Brasileiro de Geografia e Estatística, 2016).
Here I am providing the links to the websites from where the datasets come from:

1. CDC epidemic prevention initiative
2. Ministério da Saúde - Brazil
3. Mesonet
4. Instituto Brasileiro de Geografia e Estatística - projected populations
5. Instituto Brasileiro de Geografia e Estatística - states areas
6. Instituto Brasileiro de Geografia e Estatística - urban characteristics

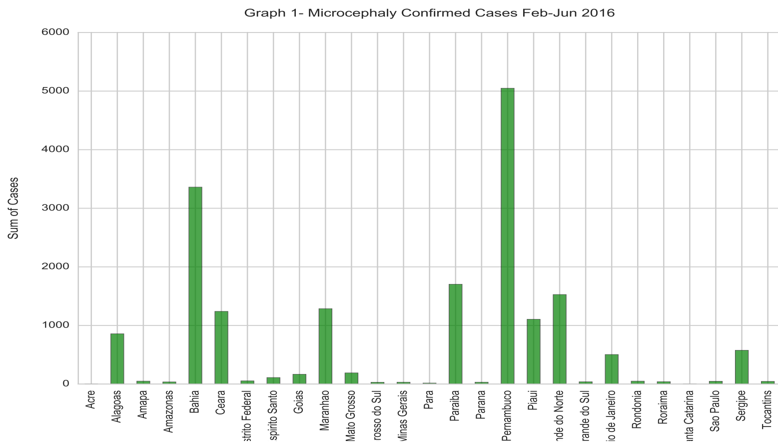
3 Methods

3.1 Microcephaly data

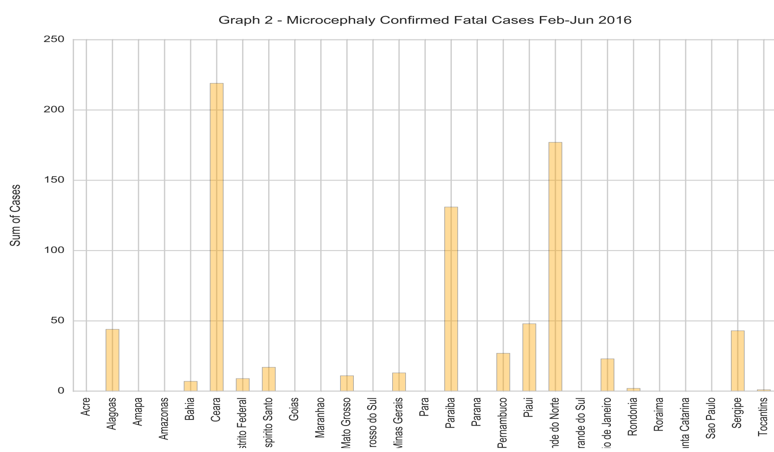
The lack of uniform and standardized definition for microcephaly has challenged the accurate monitoring of the disease; therefore, the Center for Disease and Control and Prevention (CDC) has suggested defining microcephaly as an "occipitofrontal circumference below the third percentile for gestational age and sex" (Petersen, Jamieson, Powers, & Honein, 2016).

Following this definition, the Minesterio da Saude of Brazil has been collecting and reporting microcephaly confirmed and fatal confirmed cases at the state level. From the February 2016 to June 2016 reports I found that the average number of microcephaly confirmed cases were 39.56, and the average number of microcephaly fatal confirmed cases were 1.68.

The Brazilian states located in the Northeast region presented the highest number of microcephaly confirmed cases from February 13 to to June 4. These are Pernambuco (5050 cases), Bahia (3362 cases), Paraiba (1704 cases), Rio Grande do Norte (1528 cases), Maranhao(1287 cases), Ceara (1240 cases), Piaui (1106 cases), Alagoas (858 cases), and Sergipe (576 cases). Rio de Janeiro and Sao Paulo reported a total of 504 and 48 cases respectively. The states that reported the lowest number of cases were Acre (1 case) and Santa Catarina (4 cases), which are part of the North and South regions respectively (graph 1).



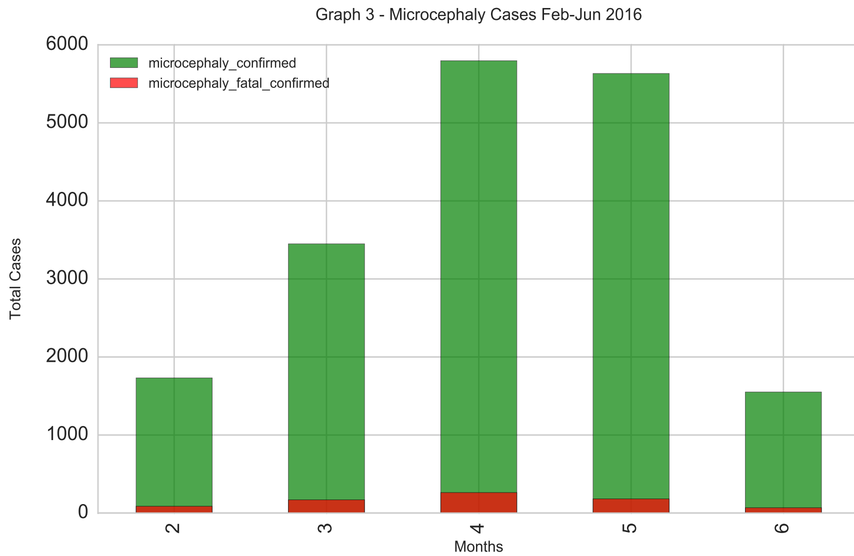
The state of Ceara, Rio Grande do Norte, and Paraiba reported the highest number of fatal microcephaly cases 219, 177 and 131 respectively (graph 2). April was the month with the highest number of microcephaly confirmed and microcephaly fatal confirmed cases, followed by May (graph 3).

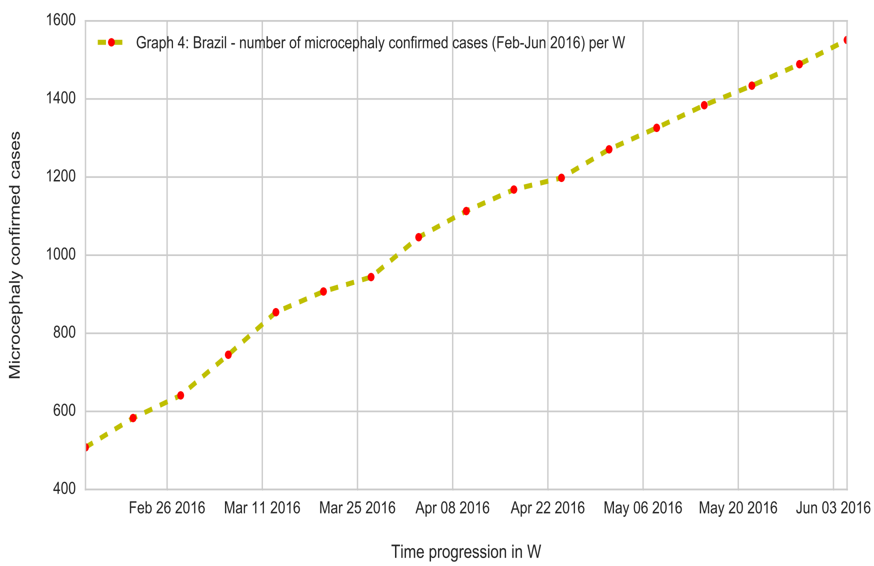


Furthermore, at the aggregate level, the number of weekly reported microcephaly confirmed cases continues to growth (graph 4). *More details on how the data was cleaned can be found in the Brazil_zika.ipynb notebook.*

3.2 Climate data

Various reports have linked the rising of temperatures and deforestation with an increase of the mosquitoes that carry the Zika virus. This suggests that the warmer it gets, the more efficiently the Aedes aegypti mosquito (the one that carries Zika, dengue fever, and other


Graph 3 - Microcephaly Cases Feb-Jun 2016
microcephaly_confirmed
microcephaly_fatal_confirmed
Total Cases
Months
0
1000
2000
3000
4000
5000
6000
2
3
4
5
6


Graph 4: Brazil - number of microcephaly confirmed cases (Feb-Jun 2016) per W
Microcephaly confirmed cases
Time progression in W
400
600
800
1000
1200
1400
1600
Feb 26 2016
Mar 11 2016
Mar 25 2016
Apr 08 2016
Apr 22 2016
May 06 2016
May 20 2016
Jun 03 2016

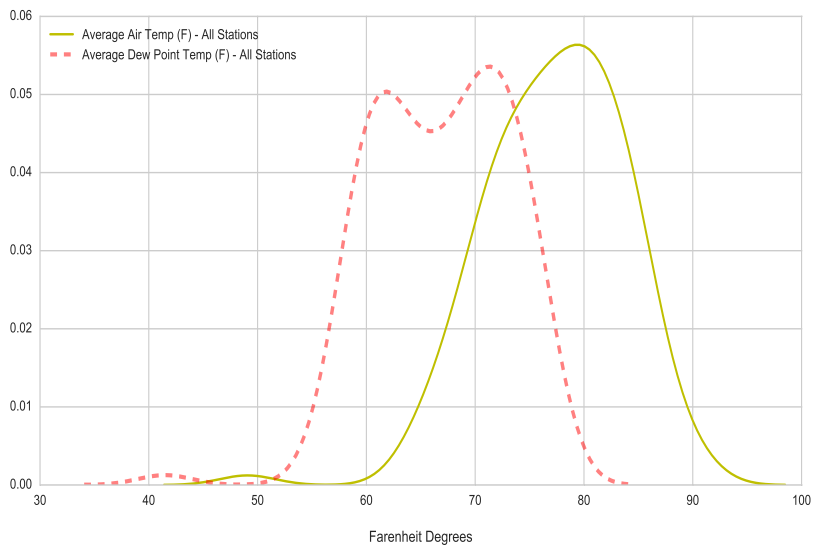
diseases) can transmit the illness; therefore, rapidly spreading the disease (Associated Press, 2016; Carlson, Dougherty, & Getz, 2016). Moreover, climate change and mosquito-borne illness have been linked with temperatures ranging from 61F and 100F (Githeko, Lindsay, Confalonieri, & Patz, 2000).

To capture the effect of temperature changes on the development of microcephaly confirmed cases related to Zika, and considering that the collected microcephaly reported cases are from February 2016 to the first week of June 2016, I collected weather variables from at least 9 months before February 2016. In doing this, the possibility of pregnant women being infected with Zika - due to climate changes - before and during their pregnancy period is being addressed.

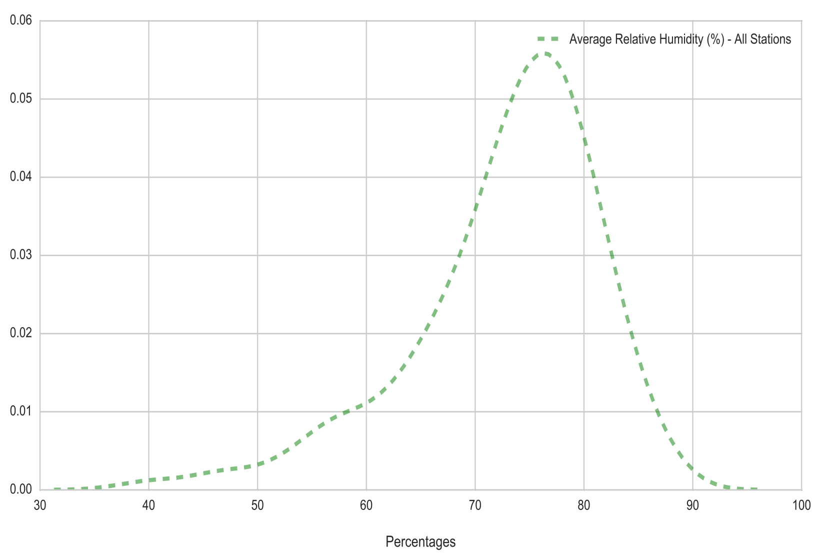
In this manner, the weather data comes from 126 stations from January 1st, 2015 to June 6, 2016. From each station the following measurements were selected:
tmpf = Air Temperature in Fahrenheit, typically @ 2 meters
dwpf = Dew Point Temperature in Fahrenheit, typically @ 2 meters (dew point is a true measurement of the atmospheric moisture)
relh = Relative Humidity in %
drct = Wind Direction in degrees from north
sknt = Wind Speed in knots
alti = Pressure altimeter in inches

These stations are not evenly located among the states; in fact, some stations are positioned within states' borders, which makes it difficult to assign them (by location) to a state. To address this and to have a better approximation of weather measurements, I aggregated each station measurement by the mean and month. Graph 5 shows the distribution of the average temperature, where at around 80 degrees Fahrenheit we see the highest temperature pick. Similarly, the average dew point presents its global maxima at around 72 degrees Fahrenheit (graph 5), and the average relative humidity presents its peak at around 77% (graph 6). To make the weather measurements more accurate I use the geopy package and calculated the distance between all the stations and the different states in 100 kilometers terms. This calculation is the previous step to facilitated the computation of the decayed distance between the sates and the stations (1/ e ˆ(distance between the states)). The decayed distance is a weight that shrinks a value (distance) the further away the station and the state are, which I use to adjust the contribution of a weather measurement from a station to the state. *More details on how the data was cleaned can be found in the following notebooks: Brazil_weather_master.ipynb and brazil_geo_data.ipynb*

Graph 5: Mean of Air Temperature and Dew Point - All Stations



Graph 6: Mean of Average Relative Humidity - All Stations



3.3 Population and urban characteristics data

The increase in urban migration and population growth, poses a risk for the expansion of the disease, especially in zones where inadequate sanitary conditions are present. Furthermore, the likelihood of getting the disease through other ways different than a mosquito bite in areas where there is a higher population density is another concern.

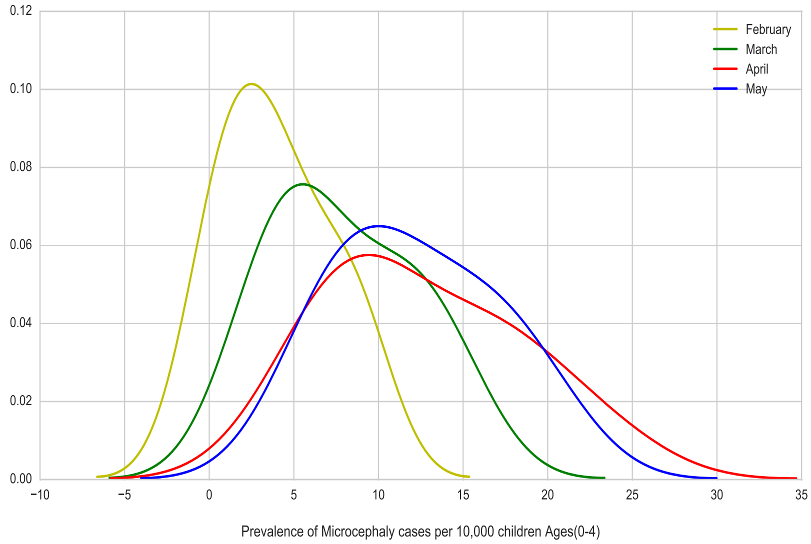
Sexually transmitted cases of Zika have been confirmed after the mosquito-borne Zika virus (ZIKV) was isolated from semen of men with Zika symptoms (E. Petersen et al., 2016). To my knowledge, there are not public records of Zika instances transmitted via sex or archives of how pregnant women acquired the disease in Brazil. The only public registers found are related to the reported number of microcephaly confirmed cases. Additionally, official numbers of newborn children per month and by states during February to June 2016 (the period of this study) are not available. These pose yet, another challenge in the construction of datasets.

I filled these gaps by using the projected population by state and month released by the Instituto Brasileiro de Geografia e Estatística. Using these numbers, I calculated the total number of men and women at a reproductive age (15-44) for 2015 and 2016, and the total population numbers for both years. Then I proceeded to calculate the population density for each year and state by dividing the total population to its correspondent state area.

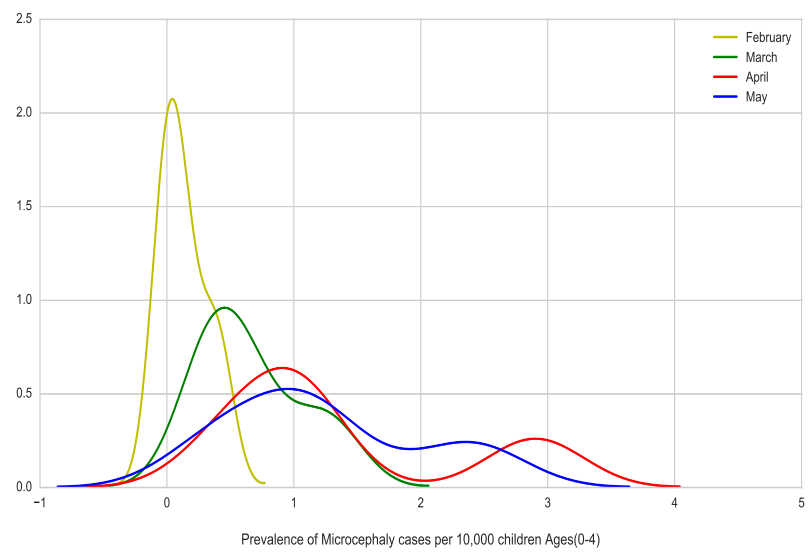
On the other hand, to control for reporting errors that could potentially be double counting or missing reported cases, I used the prevalence rate calculation of microcephaly cases for each month by 10,000 children. These rates provide a snapshot of how much microcephaly cases are present in a population group at a single point in time (Missouri Department of Health and Senior Services, 2015). Ideally, in the calculation of these rates, I will have to include the sum of newborn children in 2016 at a given month as the denominator; but there are not official numbers of newborn babies. I filled these gap by first calculating the percentage of children ages 0 to 4 for each state from the total population groups reported in the annual population projections. Then I used this percentage to obtain an estimated number of children present on each month and each state, using the total projected monthly populations (graph 9 and graph 12)

Zika has been associated as part of the "diseases of poverty" that are endemic in poor regions where lack of public health infrastructure allows illnesses to spread rapidly without significance resistance (Kruskal, 2016). In Brazil, where nearly 85 percent of the population is urban, sanitation has not expanded in the magnitude related to the population growth, especially concerning sewage and waste treatment systems, and piped water (Osava, 2016). Therefore, it still common to see open water containers located in rooftop cisterns, buckets and more, as a way to store water, and sewage running on the streets. These

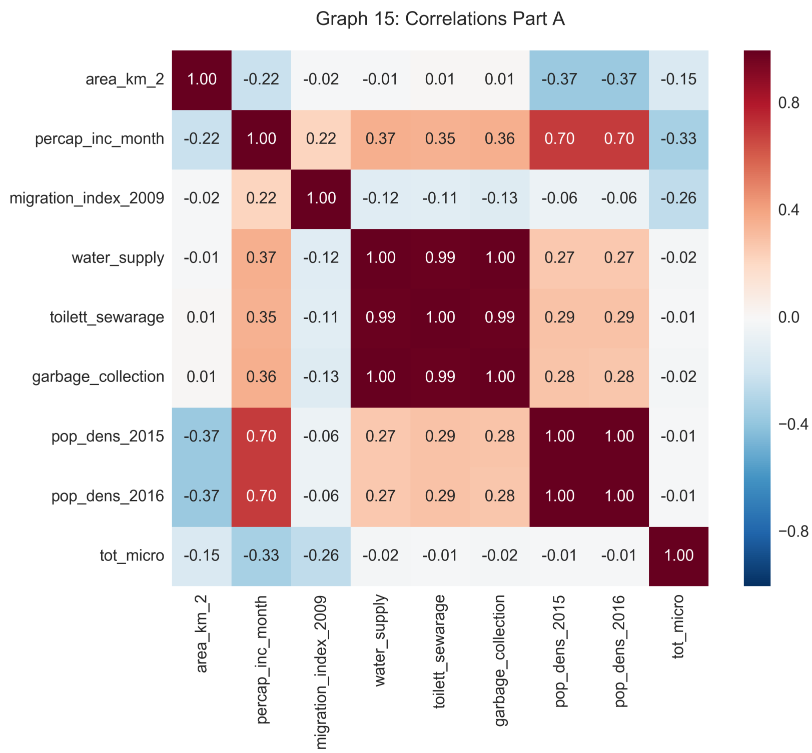
Graph 9: Prevalence of Microcephaly Northeast Region Feb-May 2016



Graph 12 : Prevalence of Microcephaly South East Region Feb-May 2016



muddy and shoddy waters coupled with old tires and other debris are the perfect habitats for the mosquito and the spread of Zika, as it occurred in the cities of Recife and Salvador in northeastern Brazil (Gillis, 2016). To control for these issues, I am using the urban population numbers that have access to piped water, sewerage system, and garbage collection from the Brazilian 2010 demographic census data. Lastly, as a measurement of wealth among the Brazilian states I use the monthly per capita income by state in US dollars (see graph 15). *More details on how the data was clean can be found in the following notebook: brazil_demographics_deforestation.ipynb*



4 Results

The method that offers a reliable explanation on the train and test data is random forest regression (see Models Performance graph), from which population density, men and women within reproductive age, and weather indicators from February 2016 have the stronger descriptive power. Although the same control variables are used on each model (see summary of statistics), the results obtained with the random forest regression model should not be considered as definitive. First of all, there are not clear medical or clinical answers on how microcephaly is linked to Zika, and how the reported cases are directly related to mothers infected with the virus. Besides, other factors like deforestation and dams should be included to have a complete picture of how the weather is related to Zika and microcephaly.

Nonetheless, the results from this case study should be used as guidance for more elaborate studies.

Summary of Statistics (Brazilian States) Feb-May 2016

variable	mean	std	min	state	max	state
Microcephaly confirmed cases	615	1086	1	Acre	4687	Pernambuco
Per capita income (usd/month) 1 USD = 3.42 reals	$287.69	$148.83	$658.48	Maranhao	$658.00	Distrito Federal
Men 15-44 (2015)	1,834,976	2,155,469	129,617	Roraima	10,684,290	Sao Paulo
Men 15-44 (2016)	1,846,669	2,163,027	132,690	Roraima	10,732,980	Sao Paulo
Women 15-44 (2015)	1,824,703	2,126,346	125,483	Roraima	1.05E+07	Sao Paulo
Women 15-44 (2016)	1,835,537	2,130,803	128,466	Roraima	1.06E+07	Sao Paulo
Migration Index (2009)	0.02	0.17	-0.30	Alagoas	0.3254	Espirito Santo
Urban population	**mean**	**std**	**min**	**state**	**max**	**state**
Access to piped water	5,204,608	7,372,400	292,675	Amapa	3.68E+07	Sao Paulo
Access to sewarage system	5,372,482	7,554,111	339,633	Roraima	3.74E+07	Sao Paulo
Access to garbage collection system	5,523,294	7,489,837	332,016	Roraima	3.74E+07	Sao Paulo
Rate of Increase Microcephaly cases (%)	**mean**	**std**	**min**	**state**	**max**	**state**
March-February	95.66	96.56	0	Acre + 5 states	369.23	Maranhao
April-March	181.61	315.06	0	Santa Catarina	1400.00	Amazonas
May-April (%)	1400.00	597.97	0	Acre	3100.00	Roraima
Air Temp (F) (Jan 2015 - May 2016)	**mean**	**std**		**Relative Humidity (%**	**mean**	**std**
Dec-2015	98.66	111.07		Dec-2015	89.56	106.80
Jan-2016	97.30	109.26		Jan-2016	96.48	110.04
Feb-2016	99.54	113.08		Feb-2016	91.38	105.59
Mar-2016	97.91	110.06		Mar-2016	94.00	110.95
Apr-2016	96.99	108.77		Apr-2016	91.00	103.49
May-2016	90.96	98.37		May-2016	94.32	112.48
Prevalence rate per 10,000 children	**min**	**state**	**max**	**state**		
Feb-2016	0	Acre + 12 states	8.74	Rio Grande do Norte		
Mar-2016	0	Acre + 5 states	14.68	Pernambuco		
Apr-2016	0	Santa Catarina	22.91	Pernambuco		
May-2016	0	Acre	20.09	Pernambuco		

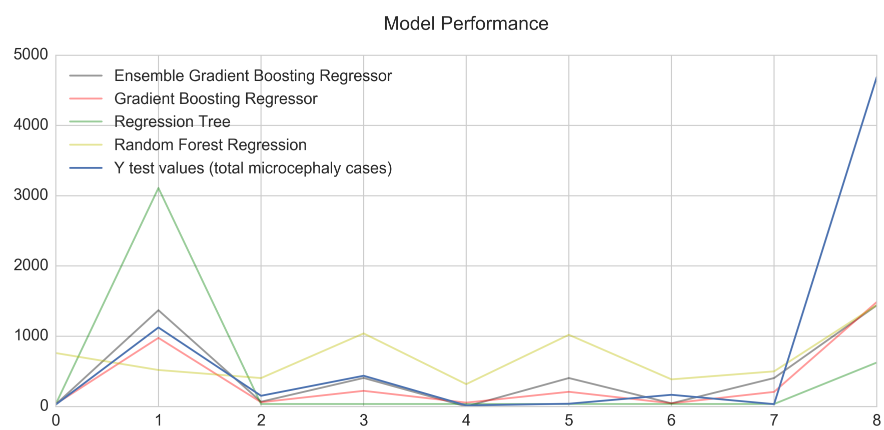


FIGURE I: Models Performance (test data)

5 References

1. Associated Press. (2016). Higher Temperatures Make Zika Mosquito Spread Disease More. Retrieved June 25, 2016, from https://weather.com/science/news/warm-temperatures-

allow-zika-spreading.

2. Calvet, G., Aguiar, R. S., Melo, A. S. O., Sampaio, S. A., de Filippis, I., Fabri, A., ? de Filippis, A. M. B. (2016). Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: A case study. The Lancet Infectious Diseases, 3099(16), 1?8. http://doi.org/10.1016/S1473-3099(16)00095-5

3. Carlson, C. J., Dougherty, E. R., & Getz, W. (2016). An ecological assessment of the pandemic threat of Zika virus. bioRxiv Beta. http://doi.org/10.1101/040386

4. CDC. (2016). CDC Epidemic Prevention Initiative, GitHub repository. Retrieved June 1, 2016, from https://github.com/cdcepi/zika

5. Gillis, J. (2016, February 20). In Zika Epidemic, a Warning on Climate Change. New York Times.

6. Githeko, A. K., Lindsay, S. W., Confalonieri, U. E., & Patz, J. A. (2000). Climate change and vector-borne diseases: a regional analysis.

7. Instituto Brasileiro de Geografia e Estatística. (2016). Projeção da População das Unidades da Federação por sexo e idade: 2000-2030.

8. Kindhauser, M. K., Allen, T., Frank, V., Santhana, R. S., & Dye, C. (2016). Zika: the origin and spread of a mosquito-borne virus. Retrieved from http://dx.doi.org/10.2471/BLT.16.170860

9. Kruskal, J. (2016, February). Zika Virus?: How Poverty and Politics Will Determine its Social Costs. International Policy Digest, 1?10. Retrieved from http://intpolicydigest.org/2016/02/19/zika-virus-poverty-politics-will-determine-social-costs/

10. Ministério da Saúde -Brazil. (2016). Informe Epidemiologico. Retrieved from http://portalsaude.saude.gov.br/index.php/o-ministerio/principal/leia-mais-o-ministerio/197-secretaria-svs/20799-microcefalia

11. Missouri Department of Health and Senior Services. (2015). Principles of Infectious Disease Epidemiology - Statistical Measures. Retrieved from http://health.mo.gov/training/epi/PrevalenceRates-b.html

12. Osava, B. M. (2016). Zika Epidemic Offers Sanitation a Chance in Brazil. Retrieved from http://www.ipsnews.net/2016/02/zika-epidemic-offers-sanitation-a-chance-in-brazil/

13. Petersen, E., Wilson, M. E., Touch, S., McCloskey, B., Mwaba, P., Bates, M., ? Zumla, A. (2016). Rapid Spread of Zika Virus in The Americas - Implications for Public Health Preparedness for Mass Gatherings at the 2016 Brazil Olympic Games. International Journal of Infectious Diseases, 44, 11?15. http://doi.org/10.1016/j.ijid.2016.02.001

14. Petersen, L., Jamieson, D., Powers, A., & Honein, M. (2016). Zika Virus. The New England Journal of Medicine, 374(16). The Iowa Environmental Mesonet. (2016). ASOS-AWOS-METAR Data.

15. WHO. (2016). WHO Director-General summarizes the outcome of the Emergency Committee regarding clusters of microcephaly and Guillain-Barré syndrome. WHO. World Health Organization.

16. Yakob, L., & Walker, T. (2016). Zika virus outbreak in the Americas: the need for novel mosquito control methods. The Lancet Global Health. http://doi.org/10.1016/S2214-109X(16)00048-6