# **TITLE:** FLIRT: a web application to predict spread of infected goods and people over air travel networks validated against Zika Virus cases in the United States

**ABSRACT:**

# **INTRODUCTION**

Zika Virus was first isolated in 1947 in Uganda from a captive rhesus monkey (Dick *et al.*, 1952). 5 years later, the first evidence of humans contracting the virus was reported in Uganda and the United Republic of Tanzania (Smithburn, 1952). Over the next 50 years, the virus spread throughout Africa and Asia (Kindhauser et al., 2016), though only 14 human cases were reported during this time period (Duffy et al., 2009). The first major human outbreak of Zika occurred in 2007, in the Federated States of Micronesia on the island of Yap, where approximately three quarters of the 7,391 population contracted the virus (Duffy et al., 2009). The most likely source of this outbreak was the inadvertent import of a mosquito vector or via an infected traveler (Duffy *et al.* 2009). Duffy *et al*. (2009) warned that Zika might easily spread via air travel, citing a medical volunteer who traveled back to the United States after contracting Zika in Yap. The ongoing Zika outbreak officially started on May 7, 2015, when Brazil’s National Reference Laboratory confirmed the virus had been isolated in the Americas (Kindhauser et al., 2016). There is active transmission of the Zika Virus in 31 countries in the Americas, with 4,160 confirmed cases and 175,636 suspected cases (CDC, 2016; PAHO WHO, 2016).

The limited public health and biosurveillance resources available to prepare against Zika must be focused in the locations where the virus is most likely to spread. Developing accurate models to anticipate Zika spread is one of the first steps further spread and transmission of the disease. Air traffic data, and simulations that use them, has improved biosurveillance capabilities (Hickeyet al., 2014; Hwang *et al*., 2012). In the 2009, influenza A H1N1 pandemic, direct and indirect flight data from the country of origin significantly improved models that predicted when the virus would be detected in other countries (Hosseini et al*.*, 2010). To prevent the further spread of infectious disease outbreaks, software that predicts where infected travelers will travel to can be useful to prevent the further spread of infectious diseases. In this manuscript, we validate a web based application against

# **METHOD**

## *FLIRT Software*

FLIRT, a biosurveillance application developed under the open source Apache 2.0 license (*available at* www.apps.eha.io), is designed to predict where infected travelers will likely travel. Conversely, FLIRT also predicts where infected travelers may originate. FLIRT contains a database of flight schedule information that is updated monthly (Innovata 2016) and visualizes passenger flow data over flight networks. Flirt contains two distinct modes: (1) “explore mode” displays nonstop flight paths from a selected airport to all possible destinations and ranks destinations by summed seat count; and, (2) “analyze mode” displays the results of a Monte Carlo simulation based on passenger layover and transfer probabilities. The simulation was created to mirror real life air traveler behavior.

### *Data*

FLIRT uses a database of flights scheduled by airlines, sourced from (Innovata 2016). This dataset consists of all flights scheduled by over 800 airlines. Each record in the database represents a scheduled flight route. For each route, the following data are available: (1) operating carrier and call sign; (2) origin and arrival airports; (3) schedule of flights; (4) effective and discontinued dates for flights; and, (5) number of seats available on the scheduled flights. Flights repeat weekly on denoted days, at start and end times. When splitting flights up by interstitial stops, the relative distance of legs to impute arrival and departure time of each leg is used. The data include current, past, and planned services, extending as far back as October 1, 2014 and continuing into the future, as far as 2018.

### *Nonstop Flight Analysis (FLIRT Explore Mode)*

FLIRT assumes that airlines will optimize their planned schedules to meet demand and fill available seats. Thus, FLIRT uses the number of seats available as a proxy for the number of passengers traveling on that route. 1 degree of edge travel, for a given time period, is calculated by summing all of the seats between two destinations (nodes) for selected time period.

### *Multi-Leg Simulation (FLIRT Analyze Mode)*

To estimate the global distribution of risk from an outbreak in a location, FLIRT simulates passenger behavior, given these assumptions: (1) infected travelers behave the same as all travelers; (2) the total number of seats scheduled between two locations (nodes) is directly proportional to the number of passengers traveling between those locations (nodes); (3) some travelers take journeys consisting of multiple flights (edges); (4) the probability distribution of the number of edges per trip for all journeys worldwide is the same as that for U.S. domestic flights. (**Table XXXX**); (5) travelers on multi edge trips do not double back (i.e., subsequent destinations in a trip that would leave a passenger closer to their origin than to their current location are calculated); (6) transfers occur in a temporal window after a passenger arrives at an airport (node) weighted according to a Poisson distribution with λ = 2 hours.

Given a time interval and an origin airport or set of airports, FLIRT simulates trips at random times within that interval, with the above described behavior. With the stated assumptions in mind, the aggregated number of passengers arriving at airports (nodes) should be directly proportional to the rate of arrival of imported disease cases from an outbreak.

### *Interface*

FLIRT’s “explore” tab lets users select an airport (node), a start and end date, and then displays the one-degree connectedness (i.e., nonstop flights) for the selected airport. Additionally, users can view direct flights from multiple airports by selecting multiple airports by: (1) searching by name; (2) selecting the airport from a list; (3) automatically grouping airports together within 50 miles; or, (4) drawing a rectangle on a map to select all the airports within the box.

*Validation, Verification, Evaluation of FLIRT*

FLIRT’s explore and analyze modes were used to assess records and future schedules of flights departing from five chosen origin airports traveling to the continental U.S. over three time periods (Table 1). Origin airports (nodes) were selected based on the number of suspected and confirmed cases per country. As of 02 February 2016, news reports indicated that Brazil, Colombia, El Salvador, Venezuela, and Honduras had the most suspected human cases of Zika Virus. An international airport in each of these origin countries was chosen based on highest passenger traffic. The final airport selections were Guarulhos International Airport (GRU) in Sao Paulo, Brazil; El Dorado International Airport (BOG) in Bogota, Colombia; Monseñor Óscar Arnulfo Romero International Airport (SAL) in San Salvador, El Salvador; Simón Bolívar International Airport (CCS) in Maiquetia, Venezuela; and Ramón Villeda Morales International Airport (SAP) in San Pedro Sula, Honduras.

Using FLIRT’s explore mode, individual network maps were generated for each of the five origin airports using counts of seats traveling from selected the origin airports to all possible connected global destinations in each of the three time periods. Then, only the U.S. destination results were extracted and the number of seats from each origin were aggregated to determine the total connectedness between all five origin airports and each possible U.S. destination. Using the analyze mode, five global simulations were generated for each time range (20,000 passengers per simulation), and each simulation yielded nearly identical results. Each simulation accounts for all five origin airports to scale for differences in air traffic volume between the five airports. The results of these five simulations were summed to produce the final simulation results.

To validate FLIRT, FLIRT’s output was compared to the locations of actual U.S. imported Zika cases was collected and compared to the geographic distribution of FLIRT’s outputs. Two time ranges were used to assess FLIRT’s ability to predict the rate of imported Zika cases to the U.S during the 2015 Zika Virus epidemic, and one future time range was considered to make future Zika distribution predictions (Table 1). For each time range, nonstop flight paths and simulation output based on flight records and schedules, departing the selected origins to continental U.S. airports, were exported from FLIRT (Table 1).

Case count data were obtained daily from news reports using Google Alerts and searching the Internet from January 11, 2016 to March 11, 2016 using the search terms: (1) new U.S. Zika case; (2) U.S. Zika cases; (3) Zika Virus U.S., and, (4) searches by each state (e.g., Florida Zika Cases). Information about all confirmed and suspected Zika cases and their location was collected; however, detailed geographic information beyond the state level was not always available in the news reports. Collected case data was compared to the CDC’s case count information (*available at* http://www.cdc.gov/zika/geo/united-states.html). Overall, this study’s case data collection matched the CDC’s state level information with two generalizable differences. The CDC had higher case counts for states that contain many cases of Zika (e.g., FL 49 vs. 34). This is partially explained by the longer time frame for which the CDC reported its data (January 1, 2015 - March 9, 2016). Secondly, the data set that was created in this study reported on outbreaks in 6 states the CDC had not reported (AZ, KY, ME, NE, UT, WV). This is most likely because outbreaks occurred in these states after the March 9th cutoff date of the CDC’s available data at the time this study was conducted. While each dataset has its differences, this analysis placed a higher priority on predicting future outbreaks in new locations *a priori* therefore this study’s prospective data were used. Additionally, this data set frequently contained more detailed information (e.g., county level spatial data) which allowed for higher accuracy in associating specific airports within the airport regional analysis.

For comparison with actual case data, and for future predictions of Zika distribution, FLIRT data was exported for the two validation time ranges and grouped by state and airport/metro region. For the state level analysis, all nonstop flight airport seat counts and simulation results within a state were aggregated at the state level for incoming U.S. flights from Zika affected areas. In the airport/metro regional analyses, each airport code was kept unique, unless the airports were within 60 miles of each other (often representative of large metropolitan regions). In this grouping, JFK – LGA - EWR - HPN, IAD - DCA - BWI, MIA - FLL, and SJC – OAK –SFO were grouped together and all nonstop flights and simulation results for each of these airports were aggregated.

Zika cases were then assigned geographic locations based on known location information from news report. Geographic information was available for all Zika cases at least at the state level. Because our simulation algorithms simulate passenger transfer behaviors, which often include transfers to regional domestic flights after international travel, FLIRT’s simulation output includes both regional and international airports as destination results. For analysis of geographic case data against simulation results, we associated each case with whichever airport (regional or international) was closest to the known case location (based on google maps road distance calculated in miles). If only county level information was known, then the largest airport (regional or international) in or most nearby the county center was selected. When geographic information beyond the state level was not known, cases were associated with the highest traffic international state airport. If the nearest airport to a case was within a metro area (airports grouped because they were within 60 miles of each other), the case was associated with the whole metropolitan group (e.g., JFK - LGA - EWR).

Because FLIRT’s direct flight analysis (Explore mode) displays individual international carrier flight paths, the results returned almost exclusively international airports. Therefore, for this Zika Virus distribution comparison, the case was associated with the nearest international airport. The Zika case was then grouped with the airport to which it was geographically closest. If only state level geographic information was known about the case, the case was associated with the largest of the two chosen state airports. To assess whether the rank order of FLIRT’s predictions corresponded with the rank order of imported Zika caseloads, we computed Kendall’s τ for the same six permutations of data.

### *Generalized Linear Models*

This study assumed that Zika cases over time at a location would be proportional to the number number of flights from Zika-affected areas. We tested this with univariate Gaussian general linear models (GLMs) by regressing imported Zika cases against FLIRT’s estimates. We ran these models on all permutations for FLIRT prediction type (one-degree connectedness and multi-degree simulation), time period (restricted to early data and all data), and aggregation level (state and airport region). Before running GLMs, all input variables were standardized by dividing by twice the standard deviation (Gelman et al. 2009).

To obtain more concretely interpretable coefficients, the 100,000 passenger simulation using actual passenger data for the source airports of interest was rescaled (see Supplemental Information) [Karissa FAA wiki]. The rescaled coefficients indicate total passengers per year, so they were divided by two to obtain a rough estimate of outgoing passengers, assuming that: (1) layover passengers are a negligible portion of passengers; and, (2) overestimating the number of outgoing passengers would bias the effect estimates towards the null hypothesis. Multiplying FLIRT’s simulated passenger estimates by the total outgoing passengers over total simulated passengers converts the estimates to passengers per year. The simulations were run on flight data, and matched with cases, for a period of 61 days, this was multiplied by 61 / 365 to obtain a rough estimate of the number of passengers traveling from the selected airports in the time period of observation. We divided this result by 100,000 to obtain a measure of Zika cases per 100,000 passengers from selected airports.

*Future Predictions of Imported Zika Cases*

FLIRT was used to calculate nonstop flight paths (1 network edge) and multi-degree passenger simulations using projected airline schedule data from 11 March 2016 2016 to 06 March 2016. The results of both outputs were compared and states and airport regions were ranked according to their relative air traffic. Global FLIRT results were analyzed to create a rank list to assess the risk of Zika case distribution globally and where within these ranks U.S. destinations are located.

# 

# **RESULTS**

# FLIRT’s explore mode produced direct flight paths for each time range and each origin airport in a geographic visualization and responsive table. FLIRT’s analyze Mode output visualized results of five 20,000 passenger simulations per time range (15 total). Each simulation was run with all origin airports included to account for the differences in passenger traffic of each origin airports. All FLIRT results are exportable to JSON, CSV, XML, and XLSX formats (*available at* apps.eha.io).

## *Predicted Locations of Imported Zika Virus Cases*

FLIRT’s simulation mode (analyze), which more closely mimics real passenger flow, identified MIA - FLL, JFK - LGA - EWR, and IAH as the top destination airports for simulated traffic from the selected origins, and Florida, Texas, and New York as the destination states with most simulated traffic for the 01 February 2016 to 01 April 2016 (Table 2). The metro region simulation and state results were consistent (11 January 2016 to to 11 March 2016).

Between 01 February 2016 – 01 April 2016, FLIRT’s explore mode (nonstop flights) identified MIA - FLL, IAH, and ATL - BHM as the airport regions with most passengers arriving from the selected origin airports. On the state level, Florida, Texas, and California were identified as the states with the highest passenger flow from the selected origins. These highest flow metro areas and states were consistently ranked in the same order in the 11 January 2016 to 11 March 2016 (Table 3).

*Future Predictions of Geographic Distribution of Imported Zika Cases*

The future FLIRT evaluation (Table 1) identified MIA - FLL, JFK - EWR - LGA, and IAH as having both highest traffic flight levels and top simulation results from the chosen origins. Consequently, Florida, Texas and New York had the highest state direct flight traffic and simulation values. The simulation rankings were consistent with the earlier time period ranks. Explore mode results differed slightly from previous time period direct flight rankings where MIA - FLL, IAH and ATL - BHM were the top identified metro regions, and Florida, Texas and California were top state.

Global airports at risk were also ranked between 11 March 2016 – 11 June 2016. According to the simulations, Tocumen International Airport in Panama City, Panama, had the highest probable incoming passengers from the five origin airports (GRU, BOG, SAP, SAL, CCS). U.S. airports were not on the top 10 ranked simulation list. However, in the direct flight global ranking Miami International Airport and George Bush Intercontinental Airport ranked in the top ten airport destinations. Toucumen International Airport, Rafael Nunez International Airport, Alfonso Bonilla Aragon International Airport, and Santiago Marino Caribbean International Airport all ranked on both direct flight and simulation top 10 lists.

### *General Linear Model*

The GLM output also shows a significant association between the number of flights from Zika-affected areas and Zika cases. For the 11 January 2016 – 11 March 2016 time period, the region-aggregated model indicates 7.24 (95% CI 6.85 – 7.62) imported Zika cases per 100,000 passengers, and the state-aggregated model suggests 11.33 (95% CI 10.80 – 11.90) imported Zika cases per 100,000 passengers.

**DISCUSSION**

Zika prevalence in the U.S. is a prime case study for FLIRT’s predictive power because most cases are imported via travel. Although there is now evidence of sexual transmission of Zika, ecological niche of vectors (A. aegypti and A. albopictus) only covers small portion of the southern continental U.S.

* In early February, authors used FLIRT data from the 2.1.2016-4.1.2016 time range to make a prediction of future Zika spread. It was predicted that Florida, Texas, New York, and California (respectively) would be at the highest risk of zika transmission, with the Miami, Houston New York, and Los Angeles metropolitan areas most at risk. News reports of Zika case occurrences since that time have confirmed FLIRT’s predictions for the 2.1.2016-4.1.2016 time range with Florida, Texas, and California respectively having the most reported Zika cases which closely matches our ranked FLIRT results. The Kendall T statistic between ranked case and simulation lists for this time range was\_\_\_\_ and the general linear model output supported the hypothesis that the correlation between actual cases and FLIRT predicted geographic locations was statistically significant.
* To expand our analysis and better determine accuracy of future predictions, we looked at all case data available (1.11.216-3.11.16 time range) and produced FLIRT simulation and Nonstop Flight results specific to this time range. The GLM results for this time range produced smaller standard errors than the 2.1.2016-4.1.2016 time range, suggesting that when more data was included and case data and FLIRT time ranges were better matched, the fit of the model improved.
* Our results for the 2.1.2016-4.1.2016 time range and the 1.11.16- 3.11.16 time range support the fact that modeling air traffic can be a powerful tool in predicting where diseases will spread next.
* Using this logic, we made a future prediction of what areas will continue to be Zika hotspots in the U.S. due to high travel volume from infected areas.
* FLIRT evaluation of future time period (3.10.16- 6.1.16) identified that MIA - FLL, JFK - EWR - LGA, IAH will continue to have the highest traffic levels, and Florida, Texas and New York will maintain the highest state traffic values. The simulation result for the future time period identified MIA - FLL, IAH, and JFK - EWR - LGA (respectively) as the highest airport region risk areas, and Florida, Texas, and New York as the highest state risk areas. As FLIRT was shown to significantly predict distribution of Zika Virus cases in the past, there should be heightened biosurveillance in these states, especially in the large metropolitan areas.
* Study Weakness: Find better geographic information about where cases are confirmed. Often, precise case locations were not publically available, and CDC only released state level information about U.S. cases
* Study Weakness: The rate of imported cases would be at least partially dependent on the prevalence of Zika in the source population. If Zika prevalence varied greatly between different airports, such that their number of cases per passenger was vastly different, it could alter the distribution of cases. However, [we could do a quick and dirty sensitivity analysis, and] it is unlikely that this would overwhelm the effects of network structure, given the magnitude of difference between cities.
* Future direction: expand study globally.

# V. [Supplemental Materials](https://docs.google.com/a/ecohealthalliance.org/document/d/1oOdmLaX3gROkAhtFxUM-U4E9ZHJ16VXYfXcDoqH691g/edit?usp=sharing)

**References**

Centers for Disease Control and Prevention. Countries and Territories in the Americas with Active Zika Virus Transmission [Internet]. Atlanta, GA: CDC; 2016 [cited 17 March 2016]. Available from: <http://www.cdc.gov/zika/geo/americas.html>

Dick GW, Kitchen SF, Haddow AJ. Zika virus. I. Isolations and serological specificity. Trans R Soc Trop Med Hyg. 1952 Sep;46(5):509–20. http://dx.doi.org/10.1016/0035-9203(52)90042-4 PMID:12995440

Duffy MR, Chen TH, Hancock WT, Powers AM, Kool JL, Lanciotti RS, et al. Zika virus outbreak on Yap Island, Federated States of Micronesia. N Engl J Med. 2009 Jun 11;360(24):2536–43. http://dx.doi.org/10.1056/NEJMoa0805715 PMID:19516034

Gelman A. Scaling regression inputs by dividing by two standard deviations. Statist Med. 2008;27(15):2865–73. (<http://doi.wiley.com/10.1002/sim.3107>)

Hickey A, Wong D, Hendricks J, Stephens M, Pedersen E, Carr D, et al. Informing US Federal Public Health Preparation for Emerging Virus Pandemic Threats at Ports of Entry. Online journal of public health informatics. 2014;6(1).

Hosseini P, Sokolow SH, Vandergrift KJ, Kilpatrick AM, Daszak P. Predictive power of air travel and socio-economic data for early pandemic spread. 2010. PLoS One. 5(9)e12763.

Hwang GM, Mahoney PJ, James JH, Lin GC, Berro AD, Keybl MA, Goedecke DM, Mathieu JJ, Wilson T. A model-based tool to predict the propagation of infectious disease via airports. Travel medicine and infectious disease. 2012 Jan 31;10(1):32-42.

Kindhauser MK, Allen T, Frank V, Santhana RS & Dye C. Zika: the origin and spread of a mosquito-borne virus [Submitted]. Bull World Health Organ E-pub: 9 Feb 2016. doi: http://dx.doi.org/10.2471/BLT.16.171082

Pan American Health Organization / World Health Organization. Cumulative Zika suspected and confirmed cases reported by countries and territories in the Americas, 2015-2016 [Internet]. Washington, D.C.: PAHO/WHO; 2016 [cited 2016 March 17]. Available from: http://ais.paho.org/phip/viz/ed\_zika\_cases.asp

Smithburn KC. Neutralizing antibodies against certain recently isolated viruses in the sera of human beings residing in East Africa. J Immunol. 1952 Aug;69(2):223–34. PMID:14946416

|  |  |  |
| --- | --- | --- |
| **Analysis Time Range** | **Case Data Time Range** | **Purpose** |
| 01 Feb. 2016 – 01 Apr. 2016 | 01 Feb. 2016 – 01 Apr. 2016 | **Prediction Validation:** Validate predictions made on Feb 2 2016 forecasting areas at highest risk |
| 11 Jan. 2016 – 11 Mar. 2016 | 11 Jan. 2016- 11 Mar. 2016 | **Expanded Data:** Determine if FLIRT output and actual case locations are associated |
| 11 Mar. 2016 – 11 Jun. 2016 | N/A | **Future Forecast:** Predict locations at future risk of receiving Zika Virus cases |

**Table 1:** Three time were ranges chosen to predict and validate Zika case distribution in the continental U.S. The *Prediction Validation* time range was chosen to use up to date case data to validate an earlier future prediction made by the authors using FLIRT data. The analysis using the *Expanded Data* range uses all available case data and matching FLIRT output date ranges for a more in-depth look at FLIRT’s ability to forecast geographic distribution of Zika cases. The *Future Forecast* ranked FLIRT’s results for a future time period to show possible vulnerable areas of future Zika spread in the U.S.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **FLIRT Simulation and Zika Virus Airport Rankings Compared between 11 Jan. 2016 – 11 Mar. 2016** | | | | |
| **FLIRT Simulation Rank Predicted** | **Airport Code** | **Simulation Seats Predicted** | **Case Counts Observed** | **Case Count Rank**  **Observed** |
| 1 | MIA/FLL | 3,694 | 19 | 1 |
| 2 | JFK/EWR/LGA | 1,503 | 8 | 3 |
| 3 | IAH | 991 | 8 | 3 |
| 4 | LAX | 508 | 5 | 6 |
| 5 | IAD/DCA/BWI | 466 | 6 | 5 |
| 6 | MCO | 427 | 10 | 2 |
| 7 | DFW | 340 | 4 | 7 |
| 8 | ATL | 282 | 7 | 4 |
| 9 | ORD | 245 | 6 | 5 |
| 10 | SJC/OAK/SFO | 211 | 6 | 5 |

**Table 2:** Top 10 ranked airport with most simulated passengers compared to states ranked according to number of ZIka cases per state for date range 11 Jan. 2016 – 11 Mar. 2016.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **FLIRT Direct Flight and Zika Ranks Compared between Jan 11 2016 - Mar 11 2016** | | | | |
| **FLIRT’s Nonstop Flight Rank Predicted** | **Airport Code** | **Flight Seats** | **Case Counts**  **Observed** | **Case Count Rank**  **Observed** | |
| 1 | MIA/FLL | 910,848 | 19 | 1 | | |
| 2 | JFK/EWR/LGA | 346,552 | 12 | 2 | | |
| 3 | IAH | 216,151 | 8 | 4 | | |
| 4 | IAD/DCA/BWI | 166,628 | 6 | 6 | | |
| 5 | LAX | 159,814 | 5 | 7 | | |
| 6 | MCO | 130,464 | 10 | 3 | | |
| 7 | ATL | 74,939 | 7 | 5 | | |
| 8 | DFW | 68,318 | 4 | 8 | | |
| 9 | ORD | 57,262 | 6 | 6 | | |
| 10 | SJC/OAK/SFO | 47,670 | 6 | 6 | | |

**Table 3:** The top 10 ranked airports with highest direct flight volume compared to states ranked according to number of Zika Virus cases observed from 11 Jan. 2016 – 11 Mar. 2016 and Kendall’s τ for the various permutations of aggregation and FLIRT output type are displayed. All rankings are significantly correlated (p < 0.05).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **FLIRT Prediction Type** | **Prediction Time Period** | **Size of the Area Examined** | **Kendall’s τ** | **p value** |
| Multi-Leg Simulation | 11 Jan. 2016 – 11 Mar. 2016 | Metro Areas | 0.266 | < .001 |
| Multi-Leg Simulation | 01 Feb. 2016 – 01 Apr. 2016 | States | 0.498 | < .001 |
| Multi-Leg Simulation | 11 Jan. 2016 – 11 Mar. 2016 | Metro Areas | 0.216 | < .001 |
| Multi-Leg Simulation | 01 Feb. 2016 – 01 Apr. 2016 | States | 0.437 | < .001 |
| Nonstop Flights | 11 Jan. 2016 – 11 Mar. 2016 | Metro Areas | 0.520 | < .001 |
| Nonstop Flights | 01 Feb. 2016 – 01 Apr. 2016 | States | 0.420 | < .001 |
| Nonstop Flights | 11 Jan. 2016 – 11 Mar. 2016 | Metro Areas | 0.534 | .002 |
| Nonstop Flights | 01 Feb. 2016 – 01 Apr. 2016 | States | 0.372 | .007 |

**Table 4:** A comparison of the ordinal ranks of locations receiving imported cases of Zika Virus predicted by FLIRT, the observed time period, and the size of the region analyzed. FLIRT’s nonstop flight and multi-leg simulation were both predictive of where Zika Cases were most likely to arrive at the state and metro area levels for the observed time periods.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Estimate** | **Standard Error** | |
| **Metro Area aggregated simulation for 11 January 2016 – 11 March 2016** | | |
| Intercept | 0.66 | 0.10 | |
| Est. Passengers (100,000) | 7.13 | 0.38 | |
| **State aggregated simulation for 11 January 2016 – 11 March 2016** | | |
| Intercept | 1.39 | 0.27 | |
| Est. Passengers (100,000) | 11.34 | 0.53 | |

**Table 5:** The intercept and standard error for general linear models for state and metro area aggregated models. A comprehensive listing of linear models, including models of previous estimates with restricted date ranges, and models for Nonstop flight data, using standardized confidence intervals, are available in Supplementary Information.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metro Area Ranks** | |  | **State Ranks** | |
| **Code** | **Nonstop Seats** | **State** | **Nonstop Seats** |
| MIA/FLL | 3,033 |  | FL | 3,451 |
| JFK/EWR/LGA | 1,247 |  | TX | 1,144 |
| IAH | 880 |  | NY | 1,087 |
| LAX | 363 |  | CA | 516 |
| MCO | 354 |  | GA | 328 |
| ATL | 327 |  | NJ | 198 |
| IAD/DCA/BWI | 246 |  | DC | 188 |
| DFW | 202 |  | IL | 181 |
| ORD | 155 |  | NC | 126 |
| SJC/OAK/SFO | 121 |  | PR | 109 |
| CLT | 116 |  | MT | 87 |
| SJU | 107 |  | MA | 67 |
| GPI | 87 |  | MI | 65 |
| BOS | 67 |  | PA | 61 |
| DTW | 53 |  | NV | 60 |
| DEN | 51 |  | MD | 58 |
| LAS | 50 |  | CO | 53 |
| PHL | 45 |  | AZ | 32 |
| TPA | 32 |  | LA | 32 |
| PHX | 28 |  | OH | 30 |

**Table 6**: The top 20 results from a simulation of 100,000 passengers from five origin airports between 11 March 2016 and 11 June 2016. Aggregation of all airports within a state does not change the rank order to a large degree.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metro Area Ranks** | |  | **State Ranks** | |
| **Code** | **Nonstop Seats** | **State** | **Nonstop Seats** |
| MIA/FLL | 712,843 |  | FL | 833,039 |
| JFK/EWR/LGA | 295,741 |  | TX | 299,655 |
| IAH | 233,736 |  | NY | 240,935 |
| LAX | 154,989 |  | CA | 170,949 |
| MCO | 120,196 |  | GA | 97,625 |
| ATL | 97,625 |  | DC | 62,831 |
| DFW | 65,919 |  | NJ | 54,806 |
| IAD/DCA/BWI | 62,831 |  | PR | 39,480 |
| SJU | 39,480 |  | IL | 29,279 |
| ORD | 29,279 |  | MI | 21,828 |
| DTW | 21,828 |  |  |  |
| SJC/OAK/SFO | 15,960 |  |  |  |

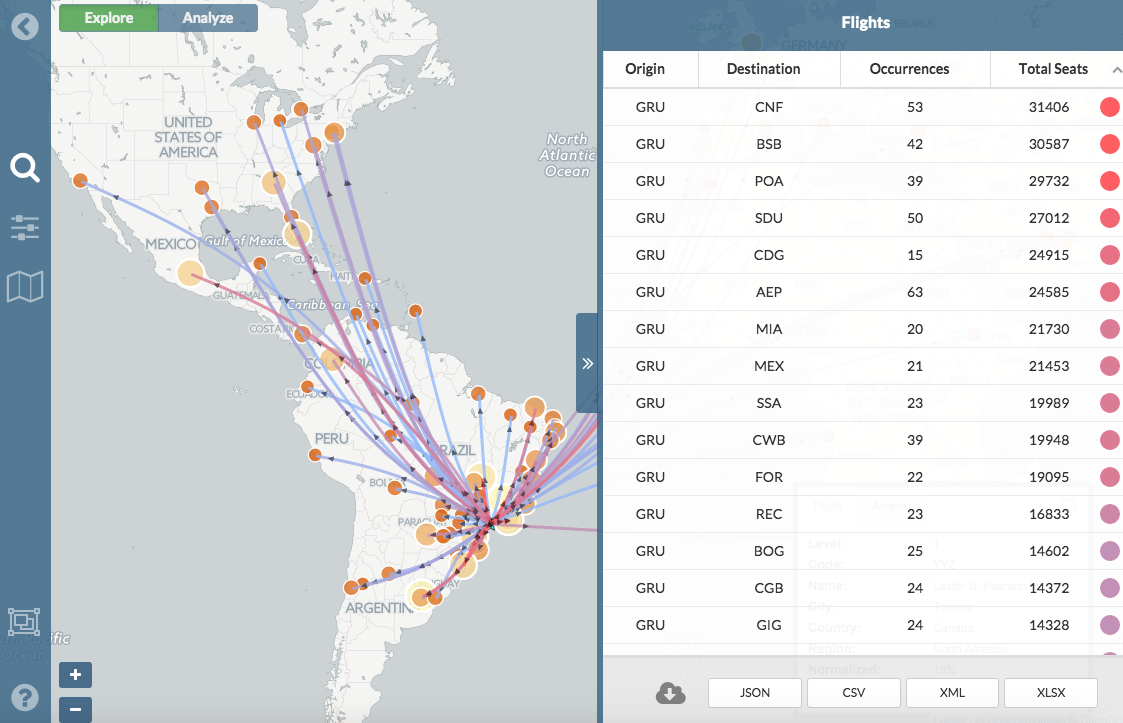
**Table 7**: Nonstop flights ranked from five origin airports for 3.11.2016- 6.11.2016 time period. The top 20 results from scheduled nonstop flights from five origin airports between 11 March 2016 and 11 June 2016. Aggregation of all airports within a state does not change the rank order to a large degree.

|  |  |  |  |
| --- | --- | --- | --- |
| **Rank** | **Code** | **Airport Name** | **Simulation Seats** |
| 1 | PTY | Tocumen International Airport, Panama City, Panama | 4,948 |
| 2 | MDE | José María Córdova International Airport, Rionegro, Colombia | 4,199 |
| 3 | CLO | Alfonso Bonilla Aragón International Airport, Palmira, Colombia | 3,598 |
| 4 | CTG | Rafael Núñez International Airport, Cartagena, Colombia | 3,155 |
| 5 | LIM | Jorge Chávez International Airport, Lima, Peru | 3,148 |
| 6 | PMV | Santiago Mariño Caribbean International Airport, Isla Margarita, Venezuela | 2,497 |
| 7 | BAQ | Ernesto Cortissoz International Airport, Barranquilla, Colombia | 2,429 |
| 8 | SCL | Comodoro Arturo Merino Benítez International Airport, Santiago, Chile | 2,086 |
| 9 | BGA | Palonegro International Airport, Bucaramanga, Colombia | 1,994 |
| 10 | SMR | Simón Bolívar International Airport, Santa Marta, Colombia | 1,964 |

**Table 8:** The top 10 global destinations from chosen origins with sustained Zika Virus transmission based on a simulation of 100,000 passengers between 11 March 2016 and 11 June 2016.

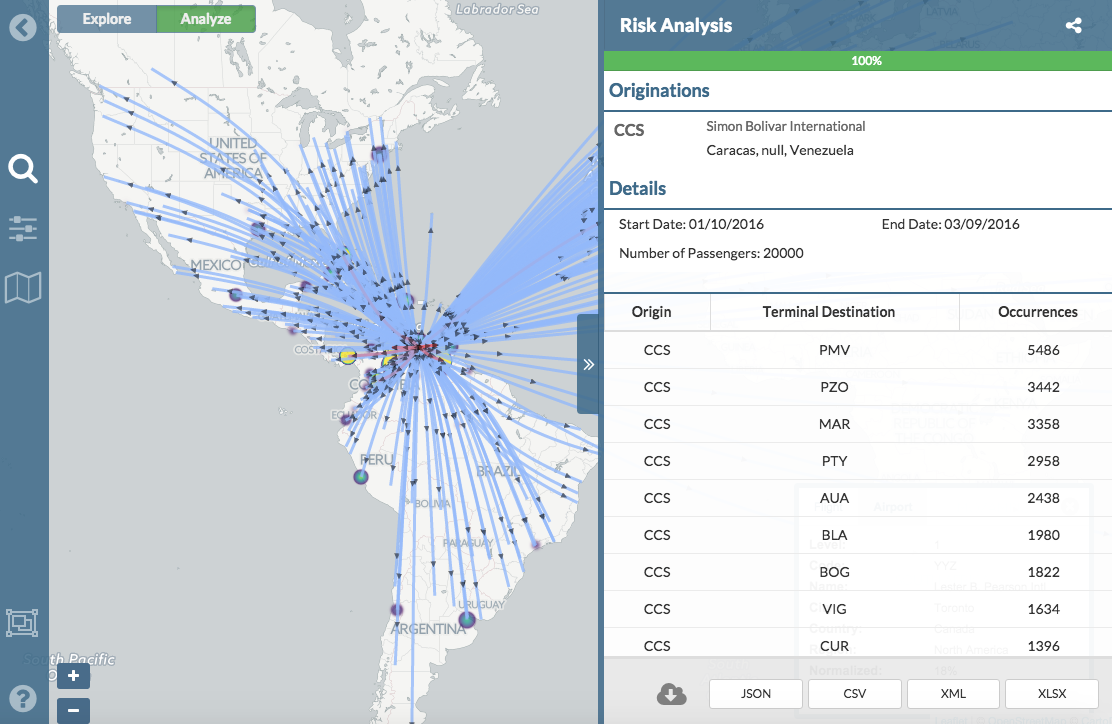
|  |  |  |  |
| --- | --- | --- | --- |
| **Rank** | **Code** | **Airport** | **Nonstop Seats** |
| 1 | MDE | José María Córdova International Airport, Rionegro, Colombia | 1,745,862 |
| 2 | CLO | Alfonso B. Aragon Aiport, Pamira Colombia | 1,381,541 |
| 3 | CTG | Rafael Núñez International Airport, Cartegena, Colombia | 1,017,339 |
| 4 | LIM | Jorge Chavez International Airport, Callao, Peru | 847,899 |
| 5 | BAQ | Ernesto Cortissoz International Airport, Barranquilla, Colombia | 788,231 |
| 6 | PTY | Tocumen International Airport, Panama City, Panama | 734,065 |
| 7 | SCL | Comodoro Arturo Merino Benítez International Airport, Santiago, Chile | 639,997 |
| 8 | PMV | Santiago Mariño Caribbean International Airport, Isla Margarita, Venezuela | 627,167 |
| 9 | BGA | Palonegro International Airport, Bucaramanga, Colombia | 622,840 |
| 10 | MIA | Miami International Airport, Miami, FL | 617,450 |

**Table 9:** The top 10 global destinations from chosen origins with sustained local Zika Virus transmission in the Western hemisphere based on the scheduled nonstop flights between 11 March 2016 and 11 June 2016.



**Figure 1**: A screenshot of FLIRT’s interface displaying a network graph based upon scheduled nonstop flights from GRU between 01 February 2016 to 01 April 2016.

## 



**Figure 2**: A screenshot of FLIRT’s interface displaying a network graph based upon the simulation of 20,000 passengers departing from CCS between 11 January 2016 to 11 March 2016.