



Published in final edited form as:

Spat Spatiotemporal Epidemiol. 2021 February ; 36: 100380. doi:10.1016/j.sste.2020.100380.

Modelling airport catchment areas to anticipate the spread of infectious diseases across land and air travel

Carmen Huber^{a,*}, Alexander Watts^a, Ardath Grills^b, Jean Hai Ein Yong^a, Stephanie Morrison^{b,c}, Sarah Bowden^{b,c}, Ashleigh Tuite^a, Bradley Nelson^b, Martin Cetron^b, Kamran Khan^{a,d,e}

^aBlueDot, 207 Queens Quay West #801b, Toronto, Ontario, Canada

^bDivision of Global Migration and Quarantine, National Center for Emerging and Zoonotic Infectious Diseases, 1600 Clifton Road, Atlanta, GA, United States of America (USA)

^cEagle Medical Services, LLC, 1826 Loop 1604 W, Suite 336-D, San Antonio, TX, United States of America (USA)

^dDepartment of Medicine, Division of Infectious Diseases, University of Toronto, 1 King's College Circle, Toronto, Ontario, Canada

^eLi Ka Shing Knowledge Institute, St. Michael's Hospital, 38 Shuter St, Toronto, Ontario, Canada

Abstract

Air travel is an increasingly important conduit for the worldwide spread of infectious diseases. However, methods to identify which airports an individual may use to initiate travel, or where an individual may travel to upon arrival at an airport is not well studied. This knowledge gap can be addressed by estimating airport catchment areas: the geographic extent from which the airport derives most of its patronage. While airport catchment areas can provide a simple decision-support tool to help delineate the spatial extent of infectious disease spread at a local scale, observed data for airport catchment areas are rarely made publicly available. Therefore, we evaluated a probabilistic choice behavior model, the Huff model, as a potential methodology to estimate airport catchment areas in the United States in data-limited scenarios.

We explored the impact of varying input parameters to the Huff model on estimated airport catchment areas: distance decay exponent, distance cut-off, and measures of airport attractiveness. We compared Huff model catchment area patterns for Miami International Airport (MIA) and

*Corresponding author at: BlueDot Inc., 207 Queens Quay W #801b, Toronto M5J 1A7, ON, Canada. carmen@bluedot.global (C. Huber).

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the United States Centers for Disease Control and Prevention. Use of trade names and commercial sources is for identification only and does not imply endorsement by the United States Centers for Disease Control and Prevention.

Declaration of Competing Interest

KK is the founder of BlueDot, a social enterprise that develops digital technologies for public health. CH, AW, JHEY, and AT received employment or consulting income from BlueDot during this research.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.sste.2020.100380.

Harrisburg International Airport (MDT). We specifically compared our model output to observed data sampled for MDT to align model parameters with an established, observed catchment area.

Airport catchment areas derived using the Huff model were highly sensitive to changes in model parameters. We observed that a distance decay exponent of 2 and a distance cut-off of 500 km represented the most realistic spatial extent and heterogeneity of the MIA catchment area. When these parameters were applied to MDT, the Huff model produced similar spatial patterns to the observed MDT catchment area. Finally, our evaluation of airport attractiveness showed that travel volume to the specific international destinations of interest for infectious disease importation risks (i.e., Brazil) had little impact on the predicted choice of airport when compared to all international travel.

Our work is a proof of concept for use of the Huff model to estimate airport catchment areas as a generalizable decision-support tool in data-limited scenarios. While our work represents an initial examination of the Huff model as a method to approximate airport catchment areas, an essential next step is to conduct a quantitative calibration and validation of the model based on multiple airports, possibly leveraging local human mobility data such as call detail records or online social network data collected from mobile devices. Ultimately, we demonstrate how the Huff model could be potentially helpful to improve the precision of early warning systems that anticipate infectious disease spread, or to incorporate when local public health decision makers need to identify where to mobilize screening infrastructure or containment strategies at a local level.

Keywords

Airport catchment area; Infectious disease spread; Public health; Huff model

1. Introduction

Air travel is an increasingly important conduit for the worldwide spread of infectious diseases, enabling pathogens to spread across vast distances, transported by infected travellers. Outbreaks of infectious diseases can have severe impacts on human health, animal health, and the economy (Golnar et al., 2016). The rapid, global spread of COVID-19 from Wuhan, China (Bogoch et al., 2020) clearly demonstrates the dramatic influence our global transportation networks have on the domestic and international spread of disease. Similarly, the spread of SARS into 29 countries within a six-month period, the importation of Ebola virus into the United States, the spread of Zika virus into and across the Americas, and the importation of Yellow Fever to multiple countries in Africa, Asia, Europe, and South America were largely driven by domestic and international air travel (Bilotkach et al., 2012; Brent et al., 2018; Cliff and Haggett, 2004; European Centre for Disease Prevention and Control, 2018; Golnar et al., 2016; Hatcher et al., 2012; Kilpatrick et al., 2006; Powers, 2015; WHO, 2020). In effect, evaluating the potential for local infectious disease spread in a destination connected to an outbreak for global infectious disease prediction could substantially improve subnational accuracy and precision for early warning systems and scenario planning. However, rarely are data available regarding which origin airports infected travellers might use to initiate travel, and where travellers might go to after

reaching their final airport destination (Bogoch et al., 2015; Bogoch et al., 2016a,b; Tatem et al., 2012; Tuite et al., 2019). This knowledge gap can be addressed by modelling airport catchment areas.

Airport catchment areas are defined as the geographic extent that captures the area in which an airport derives most of its patronage (Hernandez et al., 2004). Within an airport's catchment area, probability to use the airport may vary spatially (Fuellhart, 2007). Airport catchment areas are commonly estimated using deterministic methods, such as circular buffers or Thiessen polygons (Arino and Khan, 2015; Brent et al., 2018; Semenza et al., 2014). These approaches grossly over-simplify the representation of catchment areas. Circular buffers assume that individuals travel a defined distance, such as 200 km, to or from airports (Bilotkach et al., 2012; McLay and Reynolds-Feighan, 2006; Wang, 2000; Wittman, 2014). The circular buffers approach often only includes proximity to or from an airport, and does not consider the attractiveness of each airport. Service frequency, quality of service, and airfares are examples of factors that affect a traveller's choice, and contribute to overall airport attractiveness (Leon, 2011; Lin et al., 2016). The Thiessen polygons approach assumes that individuals will always travel the shortest possible distance to reach an airport (Hernandez et al., 2004). While this approach can incorporate a representation of airport attractiveness using *weighted* Thiessen polygons (Boots, 1980), it does not account for overlap among multiple airport catchment areas.

Accounting for overlap among multiple airport catchment areas is particularly relevant in cases where multiple airports are located within close proximity and serve common populations (Fuellhart, 2007; Kim and Ryerson, 2018; Lieshout, 2012; Paliska et al., 2016). For example, John F. Kennedy International, La-Guardia, and Newark Liberty International are major airports in the U.S. states of New York and New Jersey that are less than 45 km from each other, and Dulles International, Baltimore-Washington International, and Ronald Reagan Washington International are major airports in the states of Virginia, Maryland, and District of Colombia less than 95 km from each other. There is likely overlap between the catchment areas for airports within these clusters. Furthermore, both deterministic methods assume there is no variation in the likelihood to use an airport within the catchment area (Hernandez et al., 2004; Wittman, 2014).

By comparison, probabilistic approaches do not assume that individuals choose an airport solely based on proximity, but estimate the *likelihood* that a population would use one or more airports for travel (Hernandez et al., 2004; Lin et al., 2016). Therefore, using a probabilistic approach may improve the accuracy in approximating the airports an individual departing an outbreak-affected country is likely to use, or conversely the regions in destinations potentially receiving passengers arriving from outbreak-affected countries. More complex probabilistic methods have been applied to define catchment areas for a single airport or airport region (Lieshout, 2012), but these methods require numerous input parameters that make it impractical to estimate catchment areas for large geographic areas with a large number of airports. In contrast, the Huff model is a simple, but probabilistic method that offers an alternative to deterministic approaches and requires only a few input parameters. Based on the Huff model, the probability that someone would use an airport is defined largely by two main factors: 1) distance to each surrounding airport, and 2) a

measure of attractiveness for each surrounding airport, such as flight or traveller volume (Huff, 2003). While the Huff model was originally developed to predict consumer patronage patterns for retail (Huff, 1963), and has been applied to transportation scenarios (Lin et al., 2016), to our knowledge there is limited literature on its application in defining airport catchment areas for public health.

Given the lack of established catchment areas for most airports within the United States and internationally, a dearth in openly-accessible individual-level mobility data (Golnar et al., 2016), and few corresponding modelling techniques, we examine a proof of concept for the Huff model as a potential decision-support tool to help estimate spread of dangerous infectious diseases at a local level. We qualitatively explored the sensitivity of the Huff model to varying input parameters on estimated airport catchment areas: distance decay exponent, distance cut-off, and measures of airport attractiveness. We examined the potential application of the Huff model using two case studies: Miami International Airport (MIA) and Harrisburg International Airport (MDT), respectively. We specifically compared our model output to observed data sampled for MDT from a previous study to align model parameters with an established, observed catchment area.

2. Methods

We applied the Huff model to estimate the probability that residents living in each census tract within the United States would use any airport within the United States, when departing for an international destination. Results were calculated at the census tract level, a commonly used spatial boundary for public health policy decisions (Krieger, 2006). This probability of individuals at census tract i to choose airport j is described by (Huff, 2003):

$$P_{ij} = \frac{S_j^\alpha / D_{ij}^\beta}{\sum_{i=1}^n S_j^\alpha / D_{ij}^\beta}$$

where S_j is a measure of attractiveness for airport j , D_{ij} is the distance from census tract i to airport j , α is an airport attractiveness exponent, and β is the distance decay exponent.

Distance between each census tract (i) and each airport (j) was measured using Euclidean (straight line) distance. To optimize the algorithm, we applied a distance cut-off to define a subset of airports (referred to as the choice set) considered to be viable alternatives for each census tract. The distance cut-off should reflect the maximum distance an individual would plausibly travel to reach an airport (Huff and Black, 1997). The Huff model uses a distance decay exponent (β) to represent the impact of distance on the choice of airport; a smaller exponent means that distance has a lower impact on individuals' choice of airport and is usually used for goods or services that are not widely available, where the opposite is true for a larger exponent (Esri, 2016; Huff, 1963). The exponent typically ranges between 1 and 3 in the Huff model literature (Esri, 2016; Hernandez et al., 2004). The attractiveness variable (S) is a measure of attractiveness of each airport (j) to the potential traveller. We assumed that the total volume of outbound international travellers at each airport was a reasonable proxy for airport attractiveness (see supplementary materials). The attractiveness exponent (α) represents how attractiveness affects the individual's choice of

airport; a smaller exponent means that attractiveness has a weaker draw on an individual's selection of airport, where the opposite is true for a larger exponent.

While the Huff model most directly estimates probability to use an airport from a given census tract based on outbound travel, we assume symmetric airport use whereby travellers who use a specific airport to travel outbound will use the same airport to return home after arrival in their home country.

2.1. Case study 1: Miami International Airport

Although the Huff model was used to calculate probabilities for every United States airport, we focused on Miami International Airport (MIA) to assess model outputs. We selected MIA because it was identified in the early stages of the 2016 Zika epidemic as a likely place for viral importation from Brazil into the United States, given the high volume of travellers from Brazil (Bogoch et al., 2016). Furthermore, public health officials could have benefited from understanding where travellers were most likely to disperse by land after arriving at MIA, given that surrounding areas of Florida are suitable for local transmission of Zika virus by mosquitoes (Kraemer et al., 2015). In fact, in July 2016, a local outbreak of Zika virus was reported in Miami (Centers for Disease Control and Prevention, 2017).

We applied the Huff model using a range of estimates for inputs and qualitatively assessed the validity of the results. To restrict the choice set, we excluded airports based on distance cut-offs of 200, 500, and 1000 km from the centroid of each census tract to approximate both specific and sensitive scenarios of airport catchment areas. A recent study assumed that most users of an airport travel from within 200 km of that airport (Brent et al., 2018). We used this value as the lower estimate of our distance cut-off. In contrast to circular buffers, the distance cut-off incorporated in the Huff model does not define the catchment area itself, but simply restricts the airports included in the choice set. Hence, we wanted to include distance cut-off values that were larger than the distance one would generally travel to an airport. Additionally, we assumed that the maximum distance individuals are willing to travel to reach an airport likely varies across the United States. For example, an individual living in a very rural area would likely be willing to travel further to reach an airport if there are few options nearby, as compared to an individual living in New York City where there are multiple airport options. Therefore, a relatively small cut-off might exclude airports from the choice set that should not be eliminated. For these reasons, we tested the larger distance cut-offs of 500 km and 1000 km. We ran the model with three different distance decay exponents: 1, 1.5, and 2, selected based on existing literature (Esri, 2016; Hernandez et al., 2004). To represent airport attractiveness, we calculated total outbound international travellers at each airport using 2016 passenger flight itinerary data from the International Air Transport Association (IATA) (IATA, 2017b). IATA is the trade association for the world's airlines and represents 83% of the world's total air traffic (IATA, 2017a). We then tested the model using traveller volume only to the destination of Brazil, to understand whether travel destination has a notable impact on an airport's catchment area. Outbound traveller volume was aggregated based on the initial origin and final destination of each travel itinerary only. Therefore, indirect itineraries (with layovers) were included but not treated differently than

direct itineraries. To simplify the model, we assumed an attractiveness exponent of 1 and controlled this exponent throughout the analysis.

2.2. Case study 2: Harrisburg International Airport

While openly-accessible empirical data for airport catchment areas is limited, observational data on travellers using Harrisburg International Airport (MDT) were available from an earlier study (Fuellhart, 2007). In that study, the researchers determined place of residence for a sample of travellers to MDT, by obtaining five-digit ZIP code level data from approximately 42,000 vehicle drivers that paid for parking at MDT between March 30th and May 17th 2001 (Fuellhart, 2007). We re-ran the Huff model to incorporate the specific characteristics of the sample such as the geographic unit and year the input data was reported for, and then transformed results to be comparable to data presented by Fuellhart (i.e. to represent *proportion* of travellers expected to travel using MDT by ZIP code area, see supplementary materials). We transformed Huff model results to the five-digit ZIP code level instead of the census tract level, to ensure that they could be easily compared with the observed data. We represented airport attractiveness using all outbound air traveller volume to any destination (both domestic and international) because the sample in Fuellhart's study was not restricted to a specific travel origin or destination. We applied combinations of inputs for the distance cut-off and distance decay exponent, informed by the MIA test case.

We assessed the relative agreement between Huff model estimates and the observed, published data by visualizing them spatially on a map, producing Bland-Altman plots (Bland and Altman, 1986), and calculating the intra-class correlation coefficient using the two-way mixed effects model. More in-depth statistical analysis was limited by our access to the observed data. We did not have access to the raw data representing the observed catchment area of MDT, but only to a visualization of the data in the form of a map; values for percent of total MDT travellers by ZIP code were aggregated into five classes. Thus, statistical analysis could only be conducted using these classes, rather than the raw data itself.

3. Results

3.1. Miami International Airport

The largest airport catchment areas and highest probabilities of travel to Miami International Airport (MIA) were observed with larger distance cut-offs and smaller distance decay exponents (Figs. 1 and 2).

Smaller distance decay exponents increased the number of census tracts with higher probabilities of using MIA. With a distance cut-off of 500 km and the lowest distance decay exponent of 1, most census tracts across the predicted catchment area had greater than a 20% probability of using MIA (Figs. 1D and 2D). Travellers in census tracts that were less likely to use MIA were those in close proximity to other major airports. At the largest distance decay exponent of 2, only census tracts within approximately 300 km north of MIA had a higher than 20% probability of travellers using it (Figs. 1F and 2F).

The most obvious impact of changes to the distance cut-off was the geographic range of census tracts with some probability of using MIA (Figs. 1 and 2). Varying the distance

cut-off also resulted in changes to the estimated probability of using MIA for many census tracts. At a cut-off of 200 km, most census tracts had a greater than 40% probability of using MIA (Figs. 1G–I and 2G–I). At a cut-off of 500 km, probability of using MIA decreased in many of the census tracts within 200 km of the airport (that had greater than 40% probability with the 200 km cut-off) (Figs. 1D–F and 2D–F). Additional airport options were included with the largest cut-off of 1000 km, but the probability for census tracts within 500 km did not notably differ from when the 1000 km cut-off was used (Figs. 1A–C and 2A–C).

Many census tracts had a higher probability of using MIA when attractiveness was represented by traveller volume to Brazil, relative to all international destinations (seen when comparing Figs. 1 and 2). This difference was most notable with a distance cut-off of 200 km (Figs. 1G–I and 2G–I) or 1000 km (Figs. 1A–C and 2A–C). There were only minor differences with a distance cut-off of 500 km.

Overall, input combinations produced varying results. Variability was most prominent for census tracts with a probability to travel to MIA of 40% or less. The distance decay exponent of 1 and distance cut-off of 1000 km estimated that even areas extremely far from MIA would have high probability to use the airport. For example, Figs. 1A and 2A–C show that census tracts on the edge of the cut-off are 20–40% likely to travel to MIA, despite being within close proximity to Hartsfield-Jackson Atlanta International Airport (not shown in figures). The 200 km distance cut-off did not allow for gradual change in probability over distance. Thus, it appeared that the combination of a distance cut-off of 500 km and a distance decay exponent of either 1.5 or 2 provided outputs that we determined to be the most reasonable among outputs produced by our choices of parameters (Figs. 1E, F and 2E, F). While airport attractiveness based on the specific travel destination of Brazil produced differences in some outputs, there was no notable difference when the highlighted distance cut-off (500 km) and distance decay exponents (1.5 and 2) were used (Figs. 1E, F and 2E, F).

3.2. Harrisburg International Airport

When applied to estimate percent of total Harrisburg International Airport (MDT) travellers by ZIP code, Huff model estimates using the distance decay exponents of 1.5 and 2 had a high level of agreement with the previously published, observed data for the same variable. The intra-class correlation coefficient suggested that this level agreement was slightly higher when the distance decay exponent of 2 was applied ($ICC = 0.76, p < 0.0001$), as compared to the distance decay exponent of 1.5 ($ICC = 0.72, p < 0.0001$). Additionally, using the distance decay exponent of 2 our model predicted that 84% of travellers to MDT fell within 120 km of the airport, very similar to the value of 85% observed from the published data (Fuellhart, 2007). When the distance decay exponent of 1.5 was tested the model only predicted that 69% of expected MDT users fell within that radius.

The Bland-Altman plots support these findings and suggest there is a high level of agreement between Huff model estimates and the observed data when either distance decay exponent (1.5 and 2) was applied to the Huff model; in both cases most of the data points fell within $\pm$ two standard deviations of the mean difference (Fig. 3). Both plots showed

some data points that fell outside of these bounds; this reveals that there were larger differences between the estimated and observed values for some ZIP codes.

Moreover, when visualized on a map the overall spatial patterns of the Huff model estimates were similar to those of the observed data (Fig. 4). This was most notable for ZIP codes closest to MDT. Similar to the observed data, our model predicted a cluster of high proportions surrounding MDT that extends further to the southwest than in other directions. In contrast with the observed data, our model did not predict a cluster of ZIP codes with proportions $> 0.15\%$ north of the airport (Fig. 4). Patterns from our Huff model demonstrate this area had a relatively higher proportion of MDT users than the surrounding areas, but the proportion was lower than the legend class break of 0.15% . Therefore, the cluster was absorbed into the lowest class break and does not appear in the figure (see supplementary materials).

4. Discussion

To assess the risks of infectious disease spread in today's highly interconnected world, public health officials must consider how ground travel and air travel intersect. In the event of a global pandemic, air travel data has been shown to be predictive of destination cities or regions most at risk of imported cases of infectious diseases. However, the area in which arriving passengers locally disperse after arrival at a destination airport is often unknown. Our methods explore how the Huff model might more accurately identify which airports an individual departing an outbreak area may use, as well as delineate local areas more at risk of receiving imported cases after individuals arrive at an airport from an outbreak-affected country. Indeed, adapting geospatial methods that generalize the catchment areas for any given airport could improve the geographic accuracy by which containment measures or targeted communication campaigns are executed. Furthermore, results could be leveraged to improve the precision of early warning systems that anticipate infectious disease spread. Our study specifically examined the Huff model, which we used to estimate catchment areas for two commercial airports in the United States. We analyzed results for Miami International Airport (MIA) because of its presumed role as the introduction source for Zika virus to the Miami area in 2016. We also studied Harrisburg International Airport (MDT), and compared our findings with those from an earlier study (Fuellhart, 2007).

Our evaluation of model sensitivity to changes in inputs parameters demonstrated that parameter choice should be considered carefully, specifically for the distance decay exponent and the distance cut-off. First, based on our case studies, applying a distance decay exponent of 2 resulted in a more reasonable representation of the effect of distance on individuals' willingness to travel to an airport as compared to the other selected inputs. For the MIA test case, exponents of 1.5 and 2 resulted in the expected decrease in probability to travel from census tracts with increasing distance from MIA, or with increasing proximity to other major airports. Furthermore, results with a distance decay exponent of 2 had the highest level of agreement with the sample of observed MDT users.

Second, the distance cut-off of 500 km appeared most reasonable based on the relative findings from our MIA and MDT scenarios. In the MIA sensitivity analysis, the 500 km

cut-off did not extend outside of Florida, but also did not restrict the catchment area to only a small area of southern Florida (both of which seemed unrealistic). When paired with the distance decay exponent of 2, the 500 km cut-off allowed for a gradual decrease in probability over distance; 200 km resulted in a drastic decrease from greater than 40% to 0% at the cut-off, while 1000 km resulted in an *increase* in probability after approximately 650 km (when considering the travel destination of Brazil). Moreover, when a 500 km cut-off was applied in the MDT scenario, model outputs were very similar to the observed data.

Third, we found that using travel volume to the specific international destination of Brazil (versus all international travel) had little impact on the predicted choice of airport when the highlighted distance decay exponent and distance cut-off (2 and 500 km, respectively) were applied. However, we expect that the choice of airport might be different for domestic versus international travel; there would be dramatic differences in probability in cases where airports had only domestic flights.

There are notable differences between our estimated catchment areas produced by the Huff model and those produced by other methods that have been applied in existing literature (i.e. circular buffers and using Primary Statistical Areas as a proxy). In comparison to a 200 km buffer, some similarities exist. For example, in most outputs the area our model estimated to have greater than 40% probability to use MIA was similar in extent to a 200 km buffer. However, our model's results may be more realistic than the 200 km buffer because they reflect lower probability to use MIA for census tracts close to other attractive airports, such as Fort Lauderdale International Airport (FLL) located north of MIA, which aligns with findings from multiple studies that suggest that distance/access time has a notable impact on airport choice (Başar and Bhat, 2004; Hess and Polak, 2005; Ishii et al., 2009; Leon, 2011; Lieshout, 2012; Pels et al., 2003; Suzuki, 2007). Moreover, a 200 km buffer would likely over-estimate the catchment area for smaller, less attractive airports. In contrast, it would likely under-estimate the catchment area for highly attractive airports, especially for those that have few other attractive airport options nearby.

Additionally, for our test cases the Huff model predicted larger airport catchment areas than would be predicted by using Primary Statistical Area (PSA) boundaries (Supplementary Fig. 1). The United States Census Bureau divides the country into PSAs associated with a variety of census data tailored to each region (Wittman, 2014) (see supplementary materials). For example, the overall catchment area estimated by the Huff model for MDT is larger than the PSA boundary for this area, though most areas that the Huff model estimated to have a high proportion of MDT users would fall within it. Also, as a deterministic method the PSA boundary assumes that within its boundary there is no variation in probability for travellers to use MDT; our model accounts for such variation.

Our study has a number of limitations. First, we limited our evaluation of the Huff model to only two airports (MIA and MDT) within the United States. Other airports in the United States or internationally may differ in their catchment area parameters, depending on the number of feasible airport options, population characteristics, or varying access to transportation infrastructure. For example, access to a personal vehicle or to rail infrastructure would likely influence willingness to travel further distances to reach

an airport. Second, the results of our model are more applicable to residents than to tourists. We assumed that residents will deplane and travel to their place of residence, but various factors will influence where a tourist will travel and disperse from a given airport. It is possible that a separate model is required to estimate catchment areas for tourists. Third, we used itinerary-level traveller volume as a proxy for all factors that represent the attractiveness of an airport but including multiple factors to represent attractiveness (such as airfares or flight frequency) might improve results. The trade-off between possible increases in model predictive power and model parsimony requires further evaluation. Fourth, we measured distance using Euclidean distance; road network distance or drive-time may be more accurate but requires greater computational resources, especially for a large study area such as the United States. Relative differences between Euclidean distance and road-network based measures would be most drastic where a physical barrier, such as a mountain or a water body, was present. Finally, calibration of parameter inputs and validation of results were restricted by the limited observed data. We did not have access to observed data for MIA to validate the estimated catchment area based on our selected inputs to the Huff model. Observed data for MDT was available from an existing study, so validation was conducted on results for this airport. Data for MDT was based on the parked cars of a sample of MDT users, which had its own limitations (see supplementary materials).

A more thorough validation of results requires more detailed traveller information or empirical human mobility data to understand ground travel to and from multiple airports. While survey data such as that collected by Fuellhart (2007) could be used for validation, conducting surveys for multiple airports would require considerable time, resources, and analysis. Newly emerging data sources such as call detail records (CDR) or online social network (OSN) data (collected when users use location-enabled smartphone applications) (Barbosa-filho et al., 2018) could help to understand population mobility at a local or regional scale. Human mobility data derived by anonymized mobile devices has recently become publicly available in some regions, through academic research agreements or through proprietary contracts, and could be leveraged to calibrate and validate the Huff model. Local human mobility data could also be used to directly define the catchment areas on an ongoing basis (e.g. accounting for seasonal variation, etc.), though this may be cost-, time-, and resource-prohibitive.

Defining airport catchment areas is important for assessing the possibility of local-level spread of infectious disease outbreaks because they could help approximate 1) airports where an infected individual is most likely to travel outbound from the epicenter of an outbreak, and 2) the geographic area where an infected traveller might go after deplaning at their destination. For example, the importation of cases of ZIKV, and subsequent locally-acquired transmission of ZIKV in Wynwood, Miami (Florida Health, 2017), represents a scenario in which Huff model-derived catchment area results could have been utilized as a decision-support tool to select census tracts more likely to receive travellers arriving from ZIKV-affected countries and territories, and help identify where strategic infrastructure could be placed.

5. Conclusions

In the absence of empirical local human mobility data and in data-limited public health scenarios, the Huff model could be used to readily estimate airport catchment areas at any level of geographic unit within a reasonable spatial scale. This helps to address the need for improved global pandemic modelling predictions. We support that the Huff model may be an acceptable alternative to commonly-applied deterministic methods such as circular buffers or Thiessen polygons because the outputs are probabilistic and provide more variability within an approximated catchment area. However, this method requires further validation using emerging sources of local population mobility data, and may not be transferrable across geographic areas where non-automobile forms of land travel predominate (e.g. areas where trains are frequently used for regional travel). While this methodology could be used to estimate airport catchment areas and such estimations could help to better anticipate the spread of infectious disease via air travel, there is more work to be done before public health officials would be able to reliably use such results to make policy-level decisions. Future studies might offer a comparison to ground truth using quantitative methods in which both model outputs as well as the outputs from other methodologies could be rigorously compared to ground truth, thereby showing that their outputs were closer to truth (“better”) than the other methods. While empirical data generated by mobile devices show significant promise for public health applications, their use is complicated by the size of the datasets, the skills required to manipulate and analyze these high spatiotemporal resolution data, questions about the representativeness of the data, and other challenges related to privacy and data security.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

This work was supported by the Centers for Disease Control and Prevention grant number USG CK000433–01. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

References

- Arino J, Khan K, 2015. Using mathematical modeling to integrate disease surveillance and global air transportation data. In: Chen D, Moulin B, Wu J (Eds.), *Analyzing and Modeling Spatial and Temporal Dynamics of Infectious Diseases* John Wiley & Sons, Inc., pp. 97–107. doi:10.1002/9781118630013.ch5.
- Barbosa-filho H, Barthelemy M, Ghoshal G, James CR, Simini F, Tomasini M, 2018. Human mobility: models and applications. *Phys. Rep* 734, 1–74. doi:10.1016/j.physrep.2018.01.001.
- Başar G, Bhat C, 2004. A parameterized consideration set model for airport choice: an application to the San Francisco Bay Area. *Transp. Res. Part B* 38 (10), 889–904. doi:10.1016/j.trb.2004.01.001.
- Bilotkach V, Clougherty JA, Mueller J, Zhang A, 2012. Regulation, privatization, and airport charges: panel data evidence from European airports. *J. Regul. Econ* 42 (1), 73–94. doi:10.1007/s11149-011-9172-1.
- Bland J, Altman D, 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1 (8476), 307–310. [PubMed: 2868172]

- Bogoch II, Brady OJ, Kraemer MUG, German M, Creatore MI, Brent S, Watts AG, Hay SI, Kulkarni MA, Brownstein JS, Khan K, 2016. Potential for Zika virus introduction and transmission in resource-limited countries in Africa and the Asia-Pacific region: a modelling study. *Lancet Infect. Dis* 16 (11), 1237–1245. doi:10.1016/S1473-3099(16)30270-5. [PubMed: 27593584]
- Bogoch II, Brady OJ, Kraemer MUG, German M, Creatore MI, Kulkarni MA, Brownstein JS, Mekaru SR, Hay SI, Groot E, Watts A, Khan K, 2016. Anticipating the international spread of Zika virus from Brazil. *Lancet* 387 (10016), 335–336. doi:10.1016/S0140-6736(16)00080-5. [PubMed: 26777915]
- Bogoch II, Creatore MI, Cetron MS, Brownstein JS, Pesik N, Miniota J, Tam T, Hu W, Nicolucci A, Ahmed S, Yoon JW, Berry I, Hay SI, Anema A, Tatem AJ, MacFadden D, German M, Khan K, 2015. Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 west African outbreak. *Lancet* 385 (9962), 29–35. doi:10.1016/S0140-6736(14)61828-6. [PubMed: 25458732]
- Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MUG, Khan K, 2020. Potential for global spread of a novel coronavirus from China. *J. Travel Med* 27 (2), 1–3. doi:10.1093/jtm/taaa011.
- Boots B, 1980. Weighting Thiessen polygons. *Econ. Geogr* 56 (3), 248–259. doi:10.2307/142716.
- Brent SE, Watts A, Cetron M, German M, Kraemer UG, Bogoch II, Brady OJ, Hay SI, Creatore I, Khan K, 2018. International travel between global urban centres vulnerable to yellow fever transmission. *Bull. World Health Organ* 95 (5), 343–354. doi:10.2471/BLT.17.205658.
- Cliff A, Haggett P, 2004. Time, travel and infection. *Br. Med. Bull* 69 (1), 87–99. doi:10.1093/bmb/ldh011. [PubMed: 15226199]
- Esri, 2016. How Original Huff Model Works 10.3/tools/business-analyst-toolbox/how-original-huff-model-works.htm.
- European Centre for Disease Prevention and Control, 2018. Yellow Fever Risk Assessment: High Number of Infected Travellers Highlights the Needs of Vaccination Before Travelling to Brazil European Centre for Disease Prevention and Control <https://www.ecdc.europa.eu/en/news-events/yellow-fever-risk-assessment-high-number-infected-travellers-highlights-needs>.
- Florida Health, 2017. Zika Incident Response Playbook http://www.floridahealth.gov/diseases-and-conditions/mosquito-borne-diseases/_documents/zika-playbook.pdf.
- Fuellhart K, 2007. Airport catchment and leakage in a multi-airport region: the case of Harrisburg International. *J. Transp. Geogr* 15 (4), 231–244. doi:10.1016/j.jtrangeo.2006.08.001.
- Golnar AJ, Kading RC, Hamer GL, 2016. Quantifying the potential pathways and locations of Rift Valley fever virus entry into the United States. *Transbound. Emerg. Dis* 65 (1), 85–95. doi:10.1111/tbed.12608.
- Hatcher MJ, Dick JTA, Dunn AM, 2012. Disease emergence and invasions. *Funct. Ecol* 26 (6), 1275–1287. doi:10.1111/j.1365-2435.2012.02031.x. [PubMed: 32313353]
- Hernandez T, Lea T, Bermingham P, 2004. What's in a Trade Area?
- Hess S, Polak JW, 2005. Mixed logit modelling of airport choice in multi-airport regions. *J. Air Transp. Manag* 11 (2), 59–68. doi:10.1016/j.jairtraman.2004.09.001.
- Huff DL, 1963. A probabilistic analysis of shopping center trade areas. *Land Econ* 39 (1), 81. doi:10.2307/3144521.
- Huff DL, 2003. Parameter estimation in the Huff model. *ArcUser* 34–36. www.esri.com.
- Huff DL, Black WC, 1997. The Huff model in retrospect. *Appl. Geogr. Stud* 1 (2), 83–93.
- Centers for Disease Control and Prevention, 2017. South Florida Maps. <https://www.cdc.gov/zika/intheus/florida-maps.html>
- IATA, 2017a. About us <http://www.iata.org/about/pages/index.aspx>
- IATA, 2017b. Passenger Intelligence Services (PaxIS) <http://www.iata.org/services/statistics/intelligence/paxis/Pages/index.aspx>.
- Ishii J, Jun S, Van Dender K, 2009. Air travel choices in multi-airport markets. *J. Urban Econ* 65 (2), 216–227. doi:10.1016/j.jue.2008.12.001.
- Kilpatrick AM, Daszak P, Goodman SJ, Rogg H, Kramer LD, Cedeño V, Cunningham AA, 2006. Predicting pathogen introduction: West Nile virus spread to Galápagos. *Conserv. Biol* 20 (4), 1224–1231. doi:10.1111/j.1523-1739.2006.00423.x. [PubMed: 16922238]

- Kim AM, Ryerson MS, 2018. A long drive: interregional airport passenger “leakage” in the U.S. *Tour. Manag* 65, 237–244. doi:10.1016/j.tourman.2017.10.012.
- Kraemer MUG, Sinka ME, Duda KA, Mylne AQN, Shearer FM, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Elyazar IR, Teng HJ, Brady OJ, Messina JP, Pigott DM, Scott TW, Smith DL, ... Hay SI, 2015. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *ELife* 4, 1–18. doi:10.7554/eLife.08347.
- Krieger N, 2006. A century of census tracts: health & the body politic (1906–2006). *J. of Urban Health* 83 (3), 355–361. doi:10.1007/s11524-006-9040-y. [PubMed: 16739037]
- Leon S, 2011. Airport choice modeling: empirical evidence from a non-hub airport. *J. Transp. Res. Forum* 50 (2), 5–16. doi:10.5399/osu/jtrf.50.2.2711.
- Lieshout R, 2012. Measuring the size of an airport’s catchment area. *J. Transp. Geogr* 25, 27–34. doi:10.1016/j.jtrangeo.2012.07.004.
- Lin T, Xia J, Robinson TP, Olaru D, Smith B, Taplin J, Cao B, 2016. Enhanced Huff model for estimating Park and Ride (PnR) catchment areas in Perth, WA. *J. Transp. Geogr* 54, 336–348. doi:10.1016/j.jtrangeo.2016.06.011.
- McLay P, Reynolds-Feighan A, 2006. Competition between airport terminals: the issues facing Dublin Airport. *Transp. Res. Part A* 40 (2), 181–203. doi:10.1016/j.tra.2005.06.002.
- Paliska D, Drobne S, Borruso G, Gardina M, Fabjan D, 2016. Passengers’ airport choice and airports’ catchment area analysis in cross-border Upper Adriatic multi-airport region. *J. Air Transp. Manag* 57, 143–154. doi:10.1016/j.jairtraman.2016.07.011.
- Pels E, Nijkamp P, Rietveld P, 2003. Access to and competition between airports: a case study for the San Francisco Bay Area. *Transp. Res. Part A* 37 (1), 71–83.
- Powers AM, 2015. Risks to the Americas associated with the continued expansion of Chikungunya virus. *J. Gen. Virol* 96 (1), 1–5. doi:10.1099/vir.0.070136-0. [PubMed: 25239764]
- Semenza JC, Sudre B, Miniota J, Rossi M, Hu W, Kossowsky D, Suk JE, Van Bortel W, Khan K, 2014. International dispersal of dengue through air travel: importation risk for Europe. *PLoS Negl. Trop. Dis* 8 (12). doi:10.1371/journal.pntd.0003278.
- Suzuki Y, 2007. Modeling and testing the “two-step” decision process of travelers in airport and airline choices. *Transp. Res. Part E* 43 (1), 1–20. doi:10.1016/j.tre.2005.05.005.
- Tatem AJ, Huang Z, Das A, Qi Q, Roth J, Qiu Y, 2012. Air travel and vector-borne disease movement. *Parasitology* 139 (14), 1816–1830. doi:10.1017/S0031182012000352. [PubMed: 22444826]
- Tuite AR, Watts AG, Khan K, I.I., B., 2019. Countries at risk of importation of chikungunya virus cases from Southern Thailand: a modeling study. *Infect. Dis. Model* 4, 251–256. doi:10.1016/j.idm.2019.09.001. [PubMed: 31667444]
- Wang F, 2000. Modeling commuting patterns in Chicago in a GIS environment: a job accessibility perspective. *Profess. Geogr* 52 (1), 120–133. doi:10.1111/0033-0124.00210.
- WHO, 2020. Coronavirus Disease 2019 (COVID-19), Situation Report 35 <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200224-sitrep-35-covid-19.pdf>.
- Wittman MD, 2014. An assessment of air service accessibility in U.S. metropolitan regions, 2007–2012 (Issue 5).

*Airport attractiveness: Outbound travel volume to any international destination (2016)

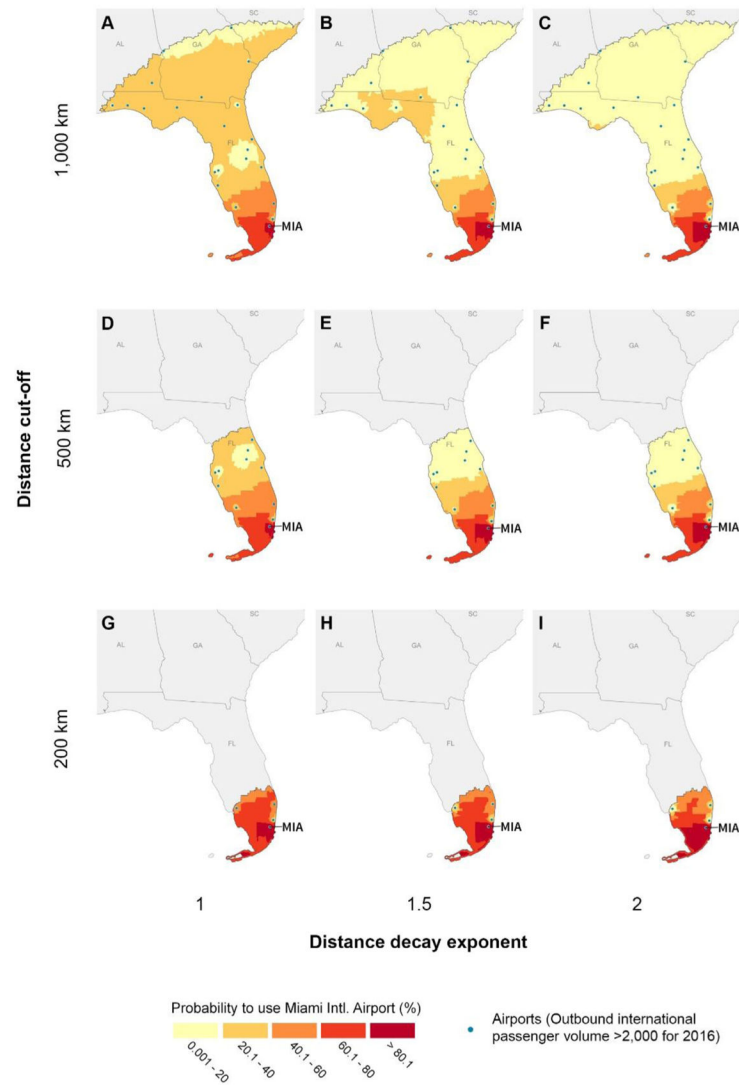


Fig. 1.

Variation in Huff model results based on changes to two key parameters: the distance decay exponent (x -axis) and distance cut-off (y -axis), where probability is represented at the census tract level and airport attractiveness is represented by international traveller volume.

*Airport attractiveness: Outbound travel volume to Brazil (2016)

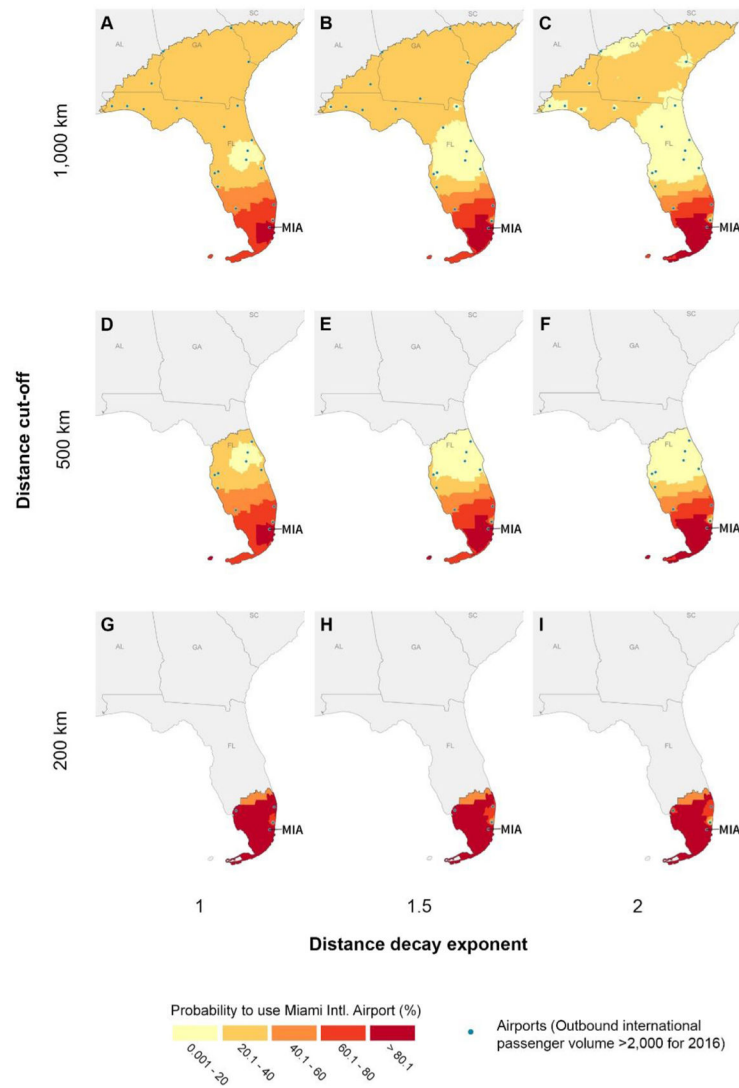


Fig. 2.

Variation in Huff model results based on changes to two key parameters: the distance decay exponent (x -axis) and distance cut-off (y -axis), where probability is represented at the census tract level and airport attractiveness is represented by traveller volume to Brazil.

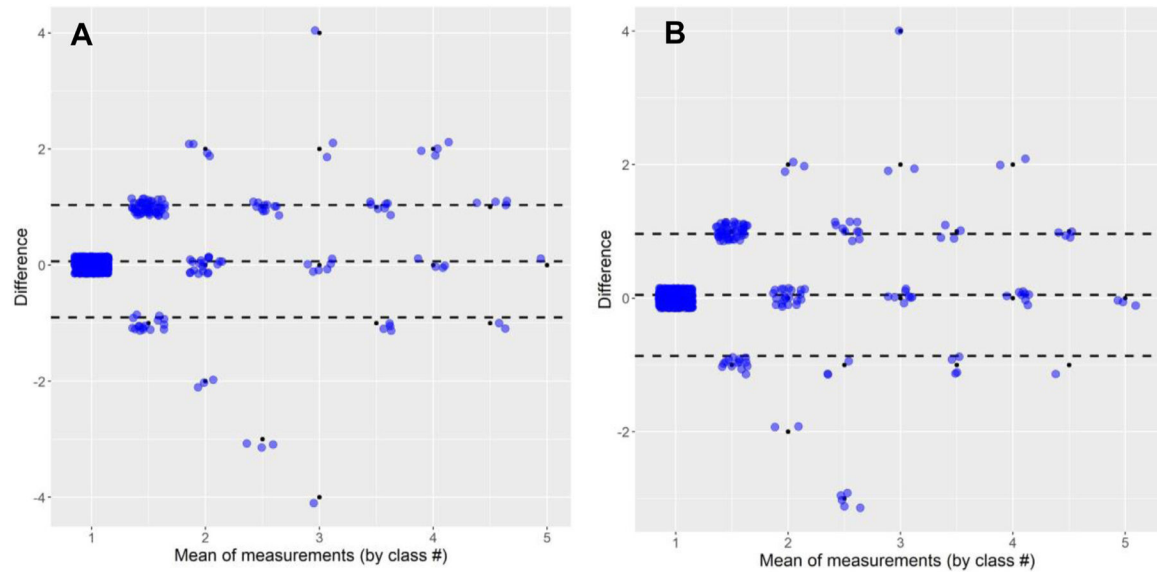


Fig. 3.

Bland-Altman plots to assess agreement between Huff model estimates and observed data adapted from Fuellhart (2007), using two different distance decay exponents applied in the Huff model. (A) Huff model parameters included a distance cut-off of 500 km and distance decay exponent of 1.5. (B) Huff model parameters included a distance cut-off of 500 km and distance decay exponent of 2. The attractiveness parameter included all outbound domestic and international travellers. Centre dashed line represents the mean difference between the Huff model estimates and observed values. Upper and lower dashed lines represent the mean difference +2 standard deviations and -2 standard deviations, respectively.

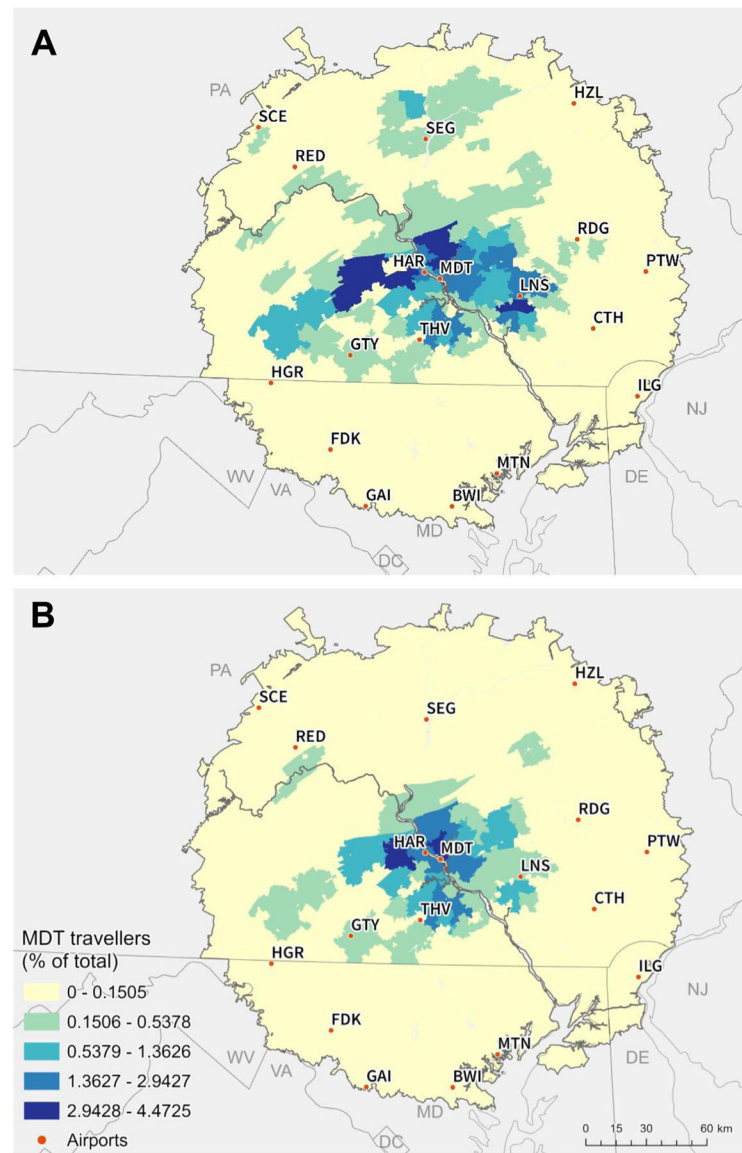


Fig. 4.

Comparison of Huff model results to observed data. (A) Travellers to Harrisburg International Airport (MDT) within 120 km by ZIP code, adapted from Fuellhart (2007). (B) Modelled travellers to MDT by ZIP code. The Huff model parameters included a distance cut-off of 500 km and distance decay exponent of 2. The attractiveness parameter included all outbound domestic and international travellers.