THE ROLE OF THE GLOBAL AIR TRAVEL NETWORK IN VECTOR-BORNE DISEASE CONNECTIVITY AND SPREAD

By

ZHUOJIE HUANG

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2013

To my Mom and Dad

ACKNOWLEDGMENTS

Firstly I would like to express my gratitude to my primary collaborator, my committee chair and my mentor, Dr. Andrew Tatem. I feel lucky to have landed in his group two years ago as he provided me invaluable opportunities to study the cutting-edge geospatial methods in epidemiology and medical geography. I am looking forward to continuing our collaborations in the future. I would also feel indebted and grateful for the prolong support from my co-chair and my friend, Dr. Timothy Fik. Special thanks are given to Dr. Liang Mao, Dr. Youliang Qiu and Dr. Xiaohui Xu, without them my dissertation will not be successful. Also, I would like to acknowledge the rest of our group: Deepa Pindolia and Andres Garcia, as well as my collaborators from the computer science department: Aniruddha Das and Udayan Kumar. Their suggestions and advice shaped my thinking on geography and helped me design my research. Additionally, I would like to thank to my friends and colleagues: Renee Bullock, Erin Bunting, Jaclyn Hall, Ian Kracalik, Jing Sun, Sam Schramski, Ying Wang, Andrea Wolf and Yang Yang. They gave me inspiration and valuable advice on my analysis. Special thanks to Xiao Wu who gave me important support.

I would like to acknowledge the support from the geography department. Many thanks are given to Dr. Jane Southworth and Dr. Peter Waylen. Six years ago their decisions have changed my life trajectory. I would also like to thank Dr. Jason Blackburn, Dr. Michael Binford and Dr. Corene Matyas. In addition, I would like to thank Dr. Michael Daniels in the Division of Statistics and Scientific Computation at the University of Texas at Austin for his valuable advices for my model building process. Their support has been important for my graduate research.

This research would also not be possible without support from the Malaria Atlas Project at Oxford University. In particular, I would like to thank Dr. Pete Gething and Dr. Simon Hay. They provided wonderful data which helps form the foundation of my graduate dissertation.

Final acknowledgement should be dedicated to both my parents, Mr. Zhicheng Huang and Mrs Meifang Liang. Though they never go to a college, they always give me the best support they can afford. Also I want to thank my grandmother, Ms. Yinghua Guan. I was so grieved when she passed away while I could not come back to China to see her.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CED	Climatic Euclidean Distance
CEDt	Inverse Climatic Euclidean Distance scaled by passenger volumes (or traffic, *t*)
CEDtr	Inverse Climatic Euclidean Distance scaled by passenger volumes and 'risk' in terms of predicted disease endemicity or probability of vector presence
GIS	Geographic Information System
HTML	Hyper Text Markup Language
IATA	International Air Transport Association
JSON	JavaScript Object Notation
MAP	Malaria Atlas Project
MVC	Model-View-Controller
OAG	Official Airline Guide
SQL	Structured Query Language
TDD	Test Driven Development
VBD-AIR	Vector-borne Disease Airline Importation Risk tool

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

THE ROLE OF THE GLOBAL AIR TRAVEL NETWORK IN VECTOR-BORNE DISEASE CONNECTIVITY AND SPREAD

By

Zhuojie Huang

May 2013

Chair: Andrew J Tatem
Cochair: Timothy J Fik
Major: Geography

Over the past century, the size and complexity of the air travel network has increased dramatically. Nowadays, there are 29.6 million scheduled flights per year and around 2.7 billion passengers are transported annually. The rapid expansion of the network increasingly connects regions of endemic vector-borne disease with the rest of the world, resulting in challenges to health systems worldwide in terms of vector-borne pathogen importation and disease vector invasion events. To face these challenges, multidisciplinary approaches that incorporate spatial and network information are advocated for better understanding the role of the air travel network, specifically its flows and architecture, on the spread of vector borne disease. The research provides a series of innovative geospatial analysis and complex network analysis aiming to 1) Create a Web-GIS for visual analytics on the risk of vector borne disease importation via air travel. 2) Provide a modeled traveler flow matrix that describes how the air travel network connects airports all around the world. 3) Examine how network metrics and communities relate to malaria exportation and importation.

CHAPTER 1
INTRODUCTION

We travel faster and farther than any time in history, and so do diseases.

Everyday, the air travel network transports over 7 million people to more than 40,000 destinations all around the world [1]. 35,000 direct scheduled routes are established on the air travel network, with 865 new routes added in 2011[1,2]. Total passenger numbers will reach 1.6 trillion in 2016 accompanied by an annual growth rate of 5%[2].These represent just small facets of the continuous expansion of the air travel network in the last century. Expansion of the airline network has profound impacts on the epidemiological landscape. Studies have shown that the connectivity of the air travel network can facilitate the global spread of directly- transmitted diseases such as influenza and severe acute respiratory syndrome (SARS) [3–6]. This connectivity also greatly affects the global propagation of vector-borne diseases [7]. Endemic areas and non-endemic areas of vector borne diseases are more connected than ever. Air highways are provided for an infected person or vector carrying pathogens travelling between spatially distant locations with speeds of 600 miles per hour. Today, vectors and vector-borne diseases are moving between different regions at exceptional rates, resulting adverse consequences, such as disease importation [7,8] and species invasion [5,9,10]. Traditional ‘drawbridge strategies’ that rely on spatial barriers to stop the transmission of diseases now play a less important role and the challenge now exists to improve global disease management and health surveillance [7]. To face these challenges, multidisciplinary approaches, that incorporate spatial and network information with spatially referenced disease data and models can provide a useful

platform for more evidence-driven surveillance design and mitigation planning [7,8,11,12].

Substantial evidence now exists of the role the air travel network plays in carrying both vector-borne diseases and their vectors between distant locations [7,9]. This has resulted in imported cases [13–15], local outbreaks[16–19] in non-endemic areas, disease resurgence and re-emergence in endemic areas[20,21] and the spread of drug and insecticide resistance [22]. Up to 8% of travelers to the developing world become ill enough to seek healthcare upon returning home, with a significant proportion of these suffering from vector-borne infections [23]. Travel clinical studies have suggested that dengue infections are now considered to be the most common cause of fever in returning travellers [14,21,24]. Around 10,000 cases of imported malaria in malaria-free high income countries are reported each year, but the true number of cases is likely to be over 25,000 [25]. Due to infrequent encounters, physicians in non-endemic countries often have difficulties in diagnosing and treating travelers that suffer from vector-borne infections when they come back from endemic areas [23,26]. Evidence has shown that the flow of people via air travel from endemic areas may increase the risks of re-emergence or resurgence [20,21,27–30] of disease in previously vector borne disease free or low transmission areas [31]. With imported vector-borne infections placing a financial and operational burden on health systems in both endemic and non-endemic countries, as well as the risk of onward transmission and even establishment, geospatial methods for quantifying the spatiotemporal risks of importation of both vector-borne diseases and the vectors that carry them can be valuable for strategic

planning for vector-borne disease elimination and control, and the allocation of limited control and surveillance resources.

The availability of global maps of vector borne disease presence and prevalance [7,19,32–34] and vector distributions [35–40] now provide evidence bases for understanding vector-borne disease risks across the world. Incorporating these high resolution maps with air travel network data offers great potential for infection importation risk assessments and vector-borne disease spread modeling [7]. This dissertation presents an initial effort towards this integration for analyzing the risks of vector borne disease transportation. Two innovative approaches are proposed. The first is a visual analytic approach that utilizes a web based GIS: the Vector-Borne Disease Airline Importation Risk Tool (VBD-AIR). This tool aims to help better define the roles of airports and airlines in the transmission and spread of vector-borne diseases. In this dissertation, an interactive web portal is created allowing users to quantify seasonally changing risks of vector and vector-borne disease importation and spread by air travel to a certain airport, with decision support for mitigation strategy planning. Also, a centralized geographical database is set up, storing world-wide geographic information on disease distribution, vector distribution, climatic information, population density, air travel connectivity and network capacity which lays out foundations for the research in the following chapters.

The second method used is complex network analysis, which further consists of two parts: flow estimation and network investigation. Firstly, a global modeled air travel database is constructed to quantify the volume of passengers on the air travel network between two airports within two stops of airline transportation. In this database, airport

(node) characteristics such as degree, centrality, city population, and local area GDP are utilized in a spatial interaction model framework to predict the air transportation flows between node pairs. This database provides an evidence-base for scientists and researchers who are aiming to understand the complex spatial interaction between origin and the destination airports. Secondly, using these predicted flows on the air network, a malaria weighted network is created to describe and quantify the patterns that exist in passenger flows weighted by malaria prevalences. In addition, the connectivity within and to the Southeast Asia region where artemisinin drug resistance emergence threats appear highest, was examined to identify and highlight risk routes for its spread.

This dissertation contributes to an on-going initiative, the human mobility mapping project (www.thummp.org), aimed at better modeling human and disease mobility, and will form part of the long-spatial scale aspect of continued multi-modal assessments of vector borne movements, and assessment of malaria elimination strategies [41,42].

CHAPTER 2
WEB-BASED GIS: THE VECTOR-BORNE DISEASE AIRLINE IMPORTATION RISK (VBD-AIR) TOOL[1]

Chapter Summary

The rapid expansion of the network increasingly connects regions of endemic vector-borne disease with the rest of the world, resulting in challenges to health systems worldwide in terms of vector-borne pathogen importation and disease vector invasion events. Here we describe the development of a user-friendly Web-based GIS tool: the Vector-Borne Disease Airline Importation Risk Tool (VBD-AIR), to help better define the roles of airports and airlines in the transmission and spread of vector-borne diseases.

Spatial datasets on modeled global disease and vector distributions, as well as climatic and air network traffic data were assembled. These were combined to derive relative risk metrics via air travel for imported infections, imported vectors and onward transmission, and incorporated into a three-tier server architecture in a Model-View-Controller framework with distributed GIS components. A user-friendly web-portal was built that enables dynamic querying of the spatial databases to provide relevant information.

The VBD-AIR tool constructed enables the user to explore the interrelationships among modeled global distributions of vector-borne infectious diseases (malaria. dengue, yellow fever and chikungunya) and international air service routes to quantify seasonally changing risks of vector and vector-borne disease importation and spread by

[1] A version of this chapter has been published as “Huang Z, Das A, Qiu Y, Tatem AJ (2012) Web-based GIS: the vector-borne disease airline importation risk (VBD-AIR) tool. Int J Health Geogr 11: 33. doi:10.1186/1476-072X-11-33.”

air travel, forming an evidence base to help plan mitigation strategies. The VBD-AIR tool is available for testing at www.vbd-air.com.

VBD-AIR supports a data flow that generates analytical results from disparate but complementary datasets into an organized cartographical presentation on a web map for the assessment of vector-borne disease movements on the air travel network. The framework built provides a flexible and robust informatics infrastructure by separating the modules of functionality through an ontological model for vector-borne disease. The VBD-Air tool is designed as an evidence base for visualizing the risks of vector-borne disease by air travel for a wide range of users, including planners and decisions makers based in state and local government, and in particular, those at international and domestic airports tasked with planning for health risks and allocating limited resources.

Background

Air travel has changed the epidemiological landscape of the world, providing routes from one side of the Earth to the other that can be traversed by an infected person in significantly shorter times than the incubation period of the majority of infectious diseases. This epidemiological impact has led to a rethink in global disease management [11], with pandemic control being less reliant on conventional spatial barriers as the global air network continues to expand. Today, vector-borne diseases and vectors are moving between different regions at unprecedented rates, resulting in adverse ecological, economic and human health consequences [7,9,44]. Reducing these problems requires examination of how humans facilitate the movement and establishment of diseases and their vectors in new areas. Moreover, the speed of air travel has meant that rapid reporting and surveillance now play an important role in preventing the spread of diseases. Finally, the cost of surveillance makes sampling

design and the development of cost effective monitoring and testing approaches vital in effective early-warning systems [45]. While work on these factors is becoming more sophisticated for directly-transmitted infections[4,46,47], our understanding of the role of air travel in global vector-borne disease epidemiology remains relatively incomplete[7].

Significant evidence exists that documents examples of both vector-borne diseases and the vectors that carry them being transported between distant locations via air travel [9,23]. The global air network enables many of the world's most isolated and diverse ecosystems to become connected and aids the movement of organisms, including disease vectors, to new habitats where they can become damaging invasive species, economically and health-wise[9,40,48]. Mosquitoes can survive moderately high atmospheric pressures aboard aircraft [49] and can be transported alive between international destinations, even in wheel bays [50]. However, air travel likely plays a much more significant role in moving the vector-borne disease (via infected passengers) than in moving the vector. It provides rapid and wide-reaching connections between outbreaks or high levels of endemicity and susceptible vector populations elsewhere in the world. Up to 8% of travellers to the developing world become ill enough to seek healthcare upon returning home, with a significant proportion of these suffering from vector-borne infections [23]. Around 10,000 cases of imported malaria to high income countries are reported each year, but the true figure may be over 25,000 [25]. With imported vector-borne infections placing a financial and operational burden on health systems in non-endemic countries (e.g. [13]), as well as the risk of onward transmission and even establishment, the development of tools for quantifying the spatiotemporal risks of importation of both vector-borne diseases and the vectors that carry them can

be valuable for assessing and guiding the allocation of limited control and surveillance resources.

Recent efforts in the field of global mapping of vector-borne diseases (e.g.[32,34]) and the vectors that carry them (e.g.[36,37,51]), generally through using sample data of known disease or vector presence in combination with environmental covariates, now provide strong evidence bases for determining vector-borne disease risks across the world. The linkage of these disease and vector distribution maps with air travel network data offers great potential for infection importation risk assessment and the modelling of vector-borne disease spread. Such distribution maps, however, represent static pictures of relatively long term (>1 year) disease prevalence and vector presence. If importation and onward spread risks are to be accurately quantified, the substantial climate-driven seasonal fluctuations in disease risk and vector densities need to be accounted for [52]. If a vector or an infected individual arrives in a new location via air travel, the risk of the vector establishing or the infection being passed on to local vector populations is often dependent upon the month of arrival. The arrival of an individual infected in a mosquito-borne disease outbreak occurring in January in the southern hemisphere (e.g. Sao Paulo) into a city in the northern hemisphere (e.g. New York) will present little risk of onward transmission, due to the cold January climate being not conducive for mosquito activity. However, a similar arrival from e.g. India (where climatic conditions may be suitable for year-round transmission) in July presents a much greater risk [52]. By utilizing gridded climate data to measure climatic similarity between origin and destination locations with known presence of a suitable vector, and adjusting for flight

passenger numbers as an additional measure of risk, these factors can be accounted for [9,10].

Here we introduce the Vector-Borne Disease Airline Importation Risk tool (VBD-AIR; www.vbd-air.com), which brings together global vector-borne disease and vector distribution maps, climate data and air network traffic information to inform on the spatiotemporal risks of disease and vector importation. The VBD-AIR tool takes the form of an interactive online interface and is targeted at users with interests in specific airports or regions, and the risks to those locations of vector-borne disease importation and onward spread, or exotic vector importation and establishment.

Data

VBD-AIR utilizes an entity-relationship model to integrate data sources from airport locations and air routes, disease and vector distributions, global climate data, and global land-based travel time data. All of these datasets are described below.

Air Travel Data

Information on a total of 3,632 airports across the world, together with their coordinate locations was obtained using Flightstats (www.flightstats.com) and is mapped in Figure 2-1a. Information on the airport name, IATA code, city and country are all stored in the VBD-AIR database that was constructed for the tool (see methods). Flight schedule and seat capacity data for 2010 and 2011 were purchased from OAG (Official Airline Guide, www.OAG.com). These included information on origin and destination airports, flight distances, and passenger capacity by month of each year. All routes used in the tool are mapped in Figure 2-1b.

Disease Distribution Datasets

The current version of VBD-AIR at the time of writing focuses on four vector-borne diseases and their related vectors, chosen due to the availability of spatially-referenced data for map production and rates of importation by air travel: Malaria (*Plasmodium falciparum* and *Plasmodium vivax*), dengue, yellow fever and chikungunya, all transmitted by mosquitoes. The methods behind the construction of each of these datasets are described briefly here, with an example map that is shown in Figure 2-2a, and the remaining output maps and mapping process presented in Appendix A.

Plasmodium falciparum is a protozoan parasite, one of the species of *Plasmodium* that cause malaria in humans. *P. falciparum* is the most dangerous of these infections as *P. falciparum* (or malignant) malaria has the highest rates of complications and mortality, while *P. vivax* is the most frequent and widely distributed cause of recurring (tertian) malaria. Non-endemic countries often see many cases of imported malaria each year through travellers or returning migrants. The geographical distribution of predicted *P. falciparum* malaria endemicity in 2010 was obtained from the Malaria Atlas Project (www.map.ox.ac.uk) and the methods behind the construction of the map are presented in Gething *et al* [32]. In brief, 22,212 community prevalence surveys were used in combination with model-based geostatistical methods to map the prevalence of *P. falciparum* globally within limits of transmission estimated by annual parasite incidence and satellite covariate data [53]. The mapping and modeling framework for *P. vivax* is presently not as advanced as for *P. falciparum*, however, the limits of transmission and stable/unstable endemicity have been mapped [54], and this dataset was used within VBD-AIR to define areas of *P. vivax* malaria endemicity.

Dengue fever is an infectious tropical disease caused by the dengue virus. The incidence of dengue fever has increased dramatically since the 1960s, with around 50–100 million people infected yearly, and imported cases through air travel to non-endemic regions are on the rise [55]. Global dengue distribution was mapped in three different ways: (i) a map of environmental 'suitability' for transmission, (ii) the same map, but only with countries/regions of known recent transmission shown, and (iii) a map of environmental similarity to recent outbreaks. For construction of the first map, thousands of geographically-located data points on dengue transmission over the past decade were combined with climatic and environmental covariates within a boosted regression tree mapping tool [56], following the approaches of Sinka *et al* [36,37,51,56], to produce a global map of suitability for dengue transmission (Appendix A). To refine this map to be focused solely on regions of recent confirmed transmission, the second dengue map was constructed with dataset masked so that only regions cited by the CDC Yellow Book (http://wwwnc.cdc.gov/travel/yellowbook/2012/chapter-3-infectious-diseases-related-to-travel/dengue-fever-and-dengue-hemorrhagic-fever.htm) as having transmission in 2010 were left (Appendix A). Finally, to produce an alternative, more contemporary dataset focused on regional suitability for significant outbreaks, geographical data on outbreaks occurring since 2008 were extracted from Healthmap (www.healthmap.org) and again combined with climatic and environmental covariates in a boosted regression tree tool [56] to map predicted dengue fever outbreak risk (Appendix A).

Yellow fever is an acute viral hemorrhagic disease [57]. The yellow fever virus is transmitted by *Aedes aegypti* (and other species) and is found in tropical and

subtropical areas in South America and Africa, but not in Asia. Since the 1980s, the number of cases of yellow fever has been increasing, making it a re-emerging disease, and case numbers imported through air travel to non-endemic areas have been rising [57]. Global yellow fever distribution was mapped in two different ways: (i) a map of environmental suitability for transmission and (ii) a map of environmental similarity to recent outbreaks. Hundreds of geographically-located data points on yellow fever occurrence over the past twenty years were combined with climatic and environmental covariates within a discriminant analysis mapping framework to produce a global map of suitability for yellow fever transmission. The datasets and methods used are described in Rogers *et al* [34]. The risk map of environmental similarity to recent outbreaks was produced using the same methods as for the dengue outbreak-similarity map described above, but using yellow fever outbreak data from Healthmap (Appendix A).

Chikungunya virus is an insect-borne virus, of the genus alphavirus that is transmitted to humans by virus-carrying *Aedes* mosquitoes, principally *Ae.aegypti* and *Ae.albopictus*. Large sporadic outbreaks have occurred since around 2005, with spread associated with both the movement of infected air travelers and the spread of the range of *Ae. albopictus* [19]. An outbreak risk map was produced using the same methods as outlined above for the dengue outbreak-similarity map, but using chikungunya outbreak data since 2008 from Healthmap (Figure 2-2a).

Vector Distribution Datasets

VBD-AIR includes three global vector distribution maps: *Aedes aegypti, Aedes albopictus* and the dominant *Anopheles* vectors of malaria. An example map is shown in Figure 2-2b for *Ae. albopictus* and the remaining maps are shown in Appendix A, with the mapping process described briefly here. More complete details on the mapping

process are provided in the VBD-AIR user guide, available through the online tool (www.vbd-air.com).

The yellow fever mosquito, *Aedes aegypti* is a mosquito that can spread dengue fever, chikungunya and yellow fever viruses, as well as other diseases. The mosquito originated in Africa but is now found in tropical and subtropical regions throughout the world. Hundreds of geographically-located data points on field-caught *Ae. aegypti* occurrence over the past 10 years [58,59] were combined with climatic and environmental covariates within a boosted regression tree mapping framework [56] following the approaches of Sinka *et al* [36,37,51], to produce a global map of predicted *Ae. aegypti* presence (Appendix A).

Ae. albopictus is of medical and public health concern because it has been shown in the laboratory to be a highly efficient vector of 22 arboviruses, including dengue, yellow fever, and West Nile fever viruses [60]. In the wild, however, its efficiency as a vector appears to be generally low, although it has been implicated in recent dengue fever and chikungunya outbreaks in the absence of the principal vector, *Ae. aegypti.* From its Old World East Asian distribution reported in 1930, *Ae. albopictus* expanded its range first to the Pacific Islands and then, within the last 20 years, to other countries in both the Old World and the New World, principally through ship-borne transportation of eggs and larvae in tires [9,60]. Here, hundreds of geographically-located data points on field-caught *Ae. albopictus* occurrence over the past 10 years [39,58,60] were again combined with climatic and environmental covariates within a boosted regression tree mapping framework [36,37,51,56] to produce a global map of predicted Ae. albopictus presence (Figure 2-2b).

Anopheles is a genus of mosquito. There are approximately 460 recognized species: while over 100 can transmit human malaria, only 30-40 commonly transmit parasites of the genus *Plasmodium*, which cause malaria in humans in endemic areas. *Anopheles gambiae* is one of the best known, because of its predominant role in the transmission of the most dangerous malaria parasite species - *Plasmodium* falciparum. Thousands of geographically-referenced data points on the presence of the dominant Anopheles vectors of malaria have been gathered and used, in combination with environmental covariates, expert opinion maps and regression tree tools, to produce global maps of Anopheles distributions by the Malaria Atlas Project (www.map.ox.ac.uk) [36,37,51,56]. Here, these maps were combined to produce a global map of dominant malaria-vector presence (Appendix A).

Climate and Other Datasets

The principal climatic constraints to vector survival (mosquitoes in this case), development and the vector-borne disease life cycle within them are temperature, rainfall and humidity. Monthly gridded global data on each of these were obtained from the CRU CL 2.0 gridded climatology datasets (http://www.cru.uea.ac.uk/cru/data/hrg/) [61] . The climate information from these gridded data was extracted for the locations of each of the airports.

A dataset depicting travel time to the nearest major settlement (with population size > 50,000) was obtained (further details here: http://bioval.jrc.ec.europa.eu/products/gam/index.htm), to provide contextual information on (i) airport disease accessibility at the origin and (ii) the potential ease of disease spread upon arrival. The risk of a disease being imported to a new location should not only be quantified by the level of predicted risk at the location of the airport,

since travellers often travel long distances to get to an airport, often coming from more rural regions where disease risk may be higher. Therefore, the land-based travel-time dataset described above was used in combination with each disease distribution map to extract the maximum predicted disease risk within two hours and fifty kilometers of each airport (see Appendix A for more detail). There exists no globally comprehensive information on travel distances to airports, thus we used this simple assumption here. The maximum level of disease risk within these travel times and distances were assigned to each

Method

VBD-AIR is designed to be a flexible tool that combines multiple geospatial datasets to inform on the relative risks between differing airports, flight routes, times of year, diseases, and their vectors, in promoting the movement of passengers infected by vector-borne diseases and the vectors that spread these diseases. In general, the tool relies on the assumptions that the levels of imported vector and vector-borne disease risk via air travel are related to (i) the presence of flight routes connecting to endemic regions (promoting the movement of people, pathogens and vectors), (ii) the level of traffic between origin and destination (increasing the probability of infected passenger and vector carriage), and (iii) the monthly climatic similarity between origin and destination (since vector activity is required at both locations to firstly provide infected passengers, and secondly prompt onward transmission or vector establishment at the selected destination).

Combing Datasets and Creating Risk Metrics

VBD-AIR utilizes an entity-relationship model to integrate the data sources from airport locations and air routes, disease and vector distributions, global climate data,

and the global land-based travel time data. VBD-AIR can be identified as a “produser” for knowledge dissemination between authoritative data producers and expert users, as it consumes data from a spatial data infrastructure and produces analytical output [62]. To generate attributes for each airport from the spatial datasets, airports were setup as the reporting agents and implemented with an object-oriented notation for extracting field data properties.

A set of climate-related indices that have been outlined elsewhere [48,63] was used here as metrics for imported vector and disease establishment risk. These rely on the assumption that the climatically-sensitive disease vectors require similar climatic conditions in their new locations to that which they experienced and survived in at their previous home locations in order to successfully establish. Moreover, for imported cases of vector-borne diseases to result in onward transmission in new locations, climatic conditions that promote vector activity, similar to the origin location where the disease was contracted, are required. Three simple indices were calculated:

1. Climatic Euclidean Distances (CEDs) are a measure of similarity in climatic regime between one location and another, and in this case were calculated through obtaining measures of rainfall (*r*), temperature (*t*) and humidity (*h*) for each airport location for each month from the gridded climatic datasets described previously. The CED between airport i and j was then calculated by $1/\sqrt{((r_i - r_j)^2 + (t_i - t_j)^2 + (h_i - h_j)^2)}$ [48,63].

2. Climatic Euclidean Distance scaled by passenger volumes (or traffic, *t*) (CEDt), provides a more relevant relative measure of insect/disease introduction risk and consequent establishment/spread by route, through taking into account not only climatic suitability between regions, but the level of traffic on connecting flight routes. CEDt is calculated as in CED above, but the resulting values are normalized to the 0-1 range and multiplied by the traffic levels on the route in question, which have been normalized through dividing by the maximum traffic value in the database [48,63].

3. Climatic Euclidean Distance scaled by passenger volumes and 'risk' in terms of predicted disease endemicity or probability of vector presence (CEDtr) at the flight origin, provides a more relevant relative measure of insect/disease

introduction risk to non-endemic/vector-free regions and consequent establishment/spread by route, through accounting for not only volume of traffic, but disease/vector prevalence at origin locations. CEDtr is calculated as in CED above, but the resulting value is normalized and multiplied by the normalized traffic levels on the route in question and by predicted disease endemicity or vector presence probability at the origin location, again normalized to the 0-1 range [48,63].

Risk Assessments

Three specific 'risk assessment' groups of functions have been built into the VBD-AIR tool and the following paragraphs describe the rationale and content of each of these assessments, provide examples on getting user specified outputs and documents the caveats and limitations of each assessment.

Imported vector-borne disease case risk assessment: These are aimed at providing estimates for the relative risks between scheduled flights of incoming air passengers carrying cases of the user selected disease. Two simple measures are calculated for the selected airport, disease and month: (i) scheduled passenger capacity for 2011 on all routes coming from endemic or outbreak risk regions of the selected disease; (ii) these passenger capacity numbers normalized through dividing by the maximum traffic value in the database and rescaled by the disease risk value in the region of origin. The first metric provides a simple measure of the maximum number of passengers arriving each month from areas of the world where transmission of the selected disease is either known to be endemic, predicted to occur or has occurred in the past. Comparing these between routes and months provides a first pass measure of risk route and timing prioritization. The second one provides a refinement to the simple measure in (i) that incorporates information on disease endemicity (on a 0 - 1 scale) at the origin.

The risk values calculated are based solely on scheduled incoming flight routes in 2011, the traffic capacity on these routes and the estimated endemic disease risk at the origin airport region. These estimates do not take into account additional risk-modifying factors such as actual passenger numbers, traveler activities and prophylaxis use, seasonal variations in disease transmission, chartered flights or multiple stopovers. Further details can be found in the VBD-AIR user guide and the user-generated reports available from the online tool (www.vbd-air.com).

Onward transmission risk assessments are aimed at providing basic estimates for the relative risks between scheduled routes of incoming flights bringing infected passengers, and those passengers coming into contact with active, competent vectors to facilitate onward transmission. Two simple measures are calculated for the user-selected airport, disease and month: (i) Flight capacities rescaled by climatic similarity between origin and destination regions for flights originating in disease endemic or outbreak prone regions (CEDt). This metric is based on the assumption that the risk of infected passenger arrival and onward disease spread is related to the amount of traffic between locations (increasing the probability of disease carriage) and also the similarity of the climate at the destination to that of the origin, since vector activity is required at both locations to firstly provide infected passengers, and secondly prompt onward transmission at the selected destination. (ii) The previous metric rescaled by the disease endemicity/risk value, r, at the origin region (CEDtr) [9,10]. This provides additional refinement of the previous metric to account for spatial variations in disease risk across the world. Finally, the opportunity to overlay a map depicting travel time to the nearest major settlement is available, to provide contextual information on (i)

airport disease accessibility at the origin and (ii) the potential ease of disease spread upon arrival.

The risk values calculated are based on scheduled incoming flight routes in 2011, the traffic capacity on these routes, the climatic similarity to origin regions and the predicted presence of the disease at the origin airport region and competent vector at the destination airport region. These estimates do not take into account additional risk-modifying factors such as passenger numbers, vector preferences, passenger activities, seasonal variations in disease vector population sizes, or vector control measures in place.

Imported vector risk assessments is aimed at providing basic estimates for the relative risks between scheduled routes of incoming flights bringing exotic disease vectors and their consequent establishment. Two simple measures are calculated for the user-selected airport, vector and month: (i) Flight capacities rescaled by climatic similarity between origin and destination regions for flights originating in vector presence regions. This metric is based on the assumption that the risk of exotic vector arrival and establishment is related to the amount of traffic between locations (increasing the probability of carriage) and also the similarity of the climate at the destination to that in its home range, accounting for seasonal variations. (ii) The previous metric rescaled by the vector suitability value, r, at the origin region (CEDtr) [7,9,10]. This provides additional refinement of the previous metric to account for spatial variations in vector suitability and abundance across the world.

The risk values calculated are based on scheduled incoming flight routes in 2011, the traffic capacity on these routes, the climatic similarity to origin regions and the

predicted presence of the disease at the origin airport region and competent vector at the destination airport region. These estimates do not take into account additional risk-modifying factors such as passenger numbers, vector preferences, passenger activities, seasonal variations in disease vector population sizes, or vector control measures in place.

VBD-AIR Tool Architecture

VBD-AIR adopts a three tier design: (i) a web server tier, (ii) a data server tier and (iii) a map service tier (Figure 2-3). The web server for VBD-AIR is constructed in an asp.net Model-View-Controller (MVC) framework adopting agile development principles. The MVC framework provides facilities for separating the modules and functionalities defined by the knowledge integration method via the MVC framework [64]: the model reacts and responds to the commands from the controller, queries the database and transfers data to the application domain (usually a data view is generated), and alters the application state if instructions are received from the controller. The view formulates a web form with structural data presentation to the user. The controller receives inputs from the users via html form and sends it to the model.

In the VBD-AIR web server (Figure 2-3), the controller is responsible for the page actions from the user (also known as 'Http Post' and 'Http Get'), as the user clicks on a button or a dropdown list, the collection of the user’s operations are sent as an object to the server. The controller receives actions from the user and selects an appropriate model for building a data view. The data view, which is usually a projection of a data model, is rendered in an HTML form or a JSON (JavaScript Object Notation) format for data exchange between the server and the client. In VBD-AIR particular, the knowledge graph is rendered in JSON. The model in VBD-AIR is a repository of intermediate data,

which is a collection of data entities and their relationships generated from the data server. This data assimilation process includes an object-oriented mapping procedure and add-ons of analytical logics. The controller also has the responsibilities to call the map server for the overlays of base maps and raster spatial dataset layers.

The data server is an SQL database, employed in a snow-flake structure, as the airport table is assigned as a fact table in the database. This table contains geographical information on the 3,682 airports. All disease, vector, route capacity and climate spatial data are mapped into this relational database management system. The data have been rectified for occasional misalignment error using ArcGIS v10. This schema simplifies the data structure for queries on airports and facilitates the data exchange process and interoperation.

The communication between the web server and the data server is largely based on a repository pattern with a predefined data model. The repository pattern is a middleware container between the study objects and the data. It is capable of aggregating data collections using precompiled queries to the database [65]. In general, the repository pattern provides a simple querying structure for intermediate data and facilitates the data exchanges.

The map service is composed of two service providers: “base map services” and “web mapping services for raster spatial datasets”. The base map services provide an overlay of the world map from either Google Maps or OpenStreetMap, which illustrate the viewing extents for the significant air routes. Furthermore, it provides functions such as an auto zoom to the viewing extent when a map is generated or an airport is selected. The web mapping services for raster spatial datasets host the map files for disease and

vector distributions on a GeoServer. This service is able to stream the raster dataset to the client's browser via Web Mapping Services (WMS) and the user can then see the raster dataset overlay in their browser.

Once the user opens the VBD-AIR website (www.vbd-air.com), the Openlayer library with preset visualization functions is loaded to the users' browser. The base map is generated from the base map services and a raster dataset overlay is generated from the Geoserver. Feature layers for airports and air routes are populated from the JSON string streaming from the controller, and this JSON string is generated according to the user input on the html form of input choices, which contains airport information, disease/vector probabilities, flight passenger capacity information and climatic conditions. Finally, these feature layers are visualized and overlaid on the base map. Supported 'mouse events' include mouse-driven map navigation and automated air route re-rendering. Users can navigate the mouse pointer to an airport or an air route to obtain more detailed information and select different rendering scenarios on imported risks or onward transmission. When a metric is chosen, the routes are colored to distinguish between where the metric is larger than the average for all routes shown, and where it is smaller. The top 10 routes ranked by the user's chosen metric are also displayed.

As the development of VBD-AIR has and will continue to involve multidisciplinary effort, we have implemented a test-driven design routine to utilize the MVC framework, so as to minimize the cost for communication. Test Driven Development (TDD) permits flexibility for a more feature driven development. A feature can be treated as a separable function in a project [66]. TDD helps in separating the concerns of various

implementations of features, for instance, the addressing of raster data overlay and PDF report generation should rely on different programming libraries, and thus, they have totally different input and output. TDD can state the objectives of these two features by dependency injection, which uses an interface as a contract to declare functionalities. The contracting interfaces could be substituted with workable classes to implement the functionality. In the implementation, different programming resources can be imported to fulfill the requirements. TDD enables the design of reusable features that can be used across multiple development efforts.

Results

The test version of VBD-AIR is available at www.vbd-air.com. In the current version, VBD-AIR enables users to explore the interrelationships between the global distributions of the four major vector-borne infectious diseases, seasonal climatic changes and seasonally changing air traffic capacities, and the full set of user inputs available in outlined in Figure 2-4.

The visualization of the air travel network within VBD-AIR follows Schneiderman's principle for data visualization: Overview first, zoom and filter, then details-on-demand [67]. The web interface provides a map view of the flight connections from endemic disease or vector presence areas to a user-selected airport, with a predicted disease risk or vector presence map overlaid (Figure 2-5). The user can view and navigate through the flight routes from each of the airports within the endemic or vector presence regions. As the user selects their choice of risk assessment type (imported disease risks, onward transmission risks or imported vector risks – see Figure 2-4), the view supports an automatic coloring scheme to render the origin airports and the routes based on the user input and the objectives for risk assessment. Moreover,

when the user clicks on an airport, detailed location information for that airport is provided. The interface is also well equipped with a comprehensive user guide, a short tutorial and integrated pop-up help windows so that the user can get a better understanding of the tool functionality, datasets and outputs.

For both the “direct flight” and “indirect flight” options (Figure 2-4), users are required to input the name or city of their airport of interest, disease of interest, and their related choice of disease distribution or vector presence map. The direct flight options provide “imported disease risks” and “onward transmission risks” functions if the user selects a disease map, and “imported vector risks” if the user selects a vector map. The imported disease risk option provides the top 10 routes by route capacity from disease endemic areas and the top 10 routes by risk-scaled traffic from disease endemic areas. The onward transmission risks and imported vector risks provide top 10 routes by climatic similarity scaled traffic and top 10 routes by climate similarity and disease/vector risk scaled traffic. The option to produce a more detailed PDF report output is provided to enable users to examine output statistics in greater detail and view longitudinal data. An option is also available to view indirect flights to the airport of interest, which highlights the possible routes from disease endemic areas via a single flight transfer, and a list of intermediate airports.

A sample user input is presented in Figure 2-5. In this example, the user is interested in the risks of dengue importation to Miami airport in January by direct flights. Once the user selects Miami airport through the auto-complete text box, the map zooms to show the region (Figure 2-5a). The user then selects dengue from the list of diseases available and chooses the spatial representation of dengue risk they are interested in

(see Data section above for information on each map), and can click on the question mark icon to get help in choosing, which in turn links to the user guide to obtain further information. The user then selects January from the drop-down month list and chooses direct flights, before choosing to display their selections. The next screen (Figure 2-5b) shows the user's chosen dengue risk map with all direct flights from endemic areas to Miami overlaid and colored by whether the traffic on the route is greater or less than the average traffic on all routes to Miami from endemic areas. The user can then select metrics from the imported disease risks box in the top left, choosing whether to view the top 10 routes by traffic from endemic areas or by traffic scaled by disease risks at the origin location. In each case, the top 10 risk routes are displayed in the box in the top-right. For more detailed quantification of imported disease risks metrics, the user has the option to view a customized PDF format report. Finally, the user can view all those flights that are connected to Miami from endemic areas by a single flight change through returning to the data selection screen and selecting the 'one transfer' option (Figure 2-5c).

Discussion

With no apparent end in sight to the continued growth in global air travel, we must expect the continued appearance of disease-spreading vector invasions and vector-borne disease movement. Approaches that can inform decision makers on the risk factors behind these importations and onward spread risks can be used to focus surveillance and control efforts more efficiently. This paper and the VBD-AIR tool it describes show that multiple datasets on many aspects relating to the risk of movement of insect vectors and vector-borne diseases through the global air network can be compiled to provide such information through a user-friendly web tool. The principal

function of the VBD-AIR tool is to provide an evidence base for assessing the role of air travel in the spread of vector-borne diseases and their vectors through available spatial data. The VBD-AIR tool is designed with a wide range of potential users in mind. These include planners and decisions makers based in state and local government, in particular, those at international and domestic airports tasked with planning for and dealing with health risks and allocating limited resources. It is clear from exploration of outputs from VBD-AIR that each region, airport and flight route has a differing risk profile in terms of disease and vector importation, determined largely by the structure of the air network and its congruence with infectious disease distributions and outbreaks and vector distributions and seasonality, yet this is rarely quantified and used when control methods and surveillance are considered.

The VBD-AIR tool shows that a multi-disciplinary approach, which draws on a variety of spatial data on factors known to influence the spread of vectors and the diseases they carry, offers potential for assessing the risk of disease importation. A range of uncertainties and limitations do still exist in the datasets and outputs presented however, and users are made aware of these within the full user guide and throughout the information boxes within the tool. Firstly, VBD-AIR considers only direct flights and their capacities within metric calculations, rather than actual passenger numbers or stopovers, and users are made aware of the uncertainties that this entails [68]. Within the disease and vector distribution modeling processes, uncertainties are inherent throughout [69], particularly in those regions with little field data to inform predictions. Moreover, we have treated the vector distributions as single homogenous types of mosquito, yet competition, competence, adaptation and preferences can vary widely

across their global distributions [36,37,51,60]. Accurate data on outbreak locations and sizes, as with many diseases, are also difficult to obtain to be sure of comprehensive assessments of risk, however, improvements in global surveillance and the rapid availability of data are improving (e.g. [70]). Also, the distance between two airports and the population size of these airports have not been explicitly incorporated in the risk assessments here. Proximity to endemic area plays an important role in vector-borne disease importation [71], while city population size can be utilized to estimate the rate of disease movement between pairs of airports [72]. Further the use of climatic similarity measures may not be appropriate to certain contexts, such as those where arid climatic conditions prompt increased water storage, leading to rises in vector-borne disease risks, rather than the decreases that may be indicated by CEDs. Finally, how to interpret and act upon the kind of relative risks identified in VBD-AIR is a challenge yet to be overcome, but various approaches to mitigating risks are presented within the tool text boxes and user guide [7].

Future updates are planned to VBD-AIR that will expand its capabilities, as the MVC framework is designed to be flexible for robust expansion. These will include: (i) Regular updates of the disease and vector distribution maps, as new survey data and outbreak reports become available; (ii) updates to the flight data as new information on flight capacities and routes becomes available; (iii) additional scenario-related functionality will be built into the tool based on a set of control and mitigation options. This will provide users with guidance on approaches to limiting imported cases and vectors, and mitigating the effects of vector establishment or onward disease transmission; (iv) the interactive incorporation of the accessibility datasets. At present

this represents a simple visualization of access to provide context, and upcoming extensions will focus on building in measures of access to better capture airport catchment areas and estimate likely regions impacted by imported cases or vectors; (v) the incorporation of extra vector-borne diseases and vectors. The choice of additional diseases and vectors will depend upon availability of sufficient spatial data for mapping or validated global maps. Candidate diseases include leishmaniasis, Rift Valley fever and chagas disease.

It is envisioned that future research beyond the simple updates and tool expansions described above will build upon VBD-AIR to continue to improve quantification of these aspects, drawing on newly-developed spatial datasets and mathematical models of transmission, to provide an evidence base to enable airports, airlines, and public health officials to assess the appropriateness and efficacy of current control, surveillance and treatment practices, and tailor strategies to these differing risk profiles for each disease, route and airport. Three specific areas of research should be examined:

(i) *Constructing geospatial information databases* on global endemic disease distributions, and building a framework for the rapid inclusion of outbreak reporting data from surveillance databases such as HealthMap (www.healthmap.org) [70,73]. Increasingly, spatial information on the prevalence of directly-transmitted and insect-borne diseases are being made available, and approaches for using these data to build distribution maps and dynamic transmission models are following. The potential of combining such data with air traffic data for forecasting disease movements has been

shown for a handful of diseases in specific locations, but this potential has yet to be realized at global scales.

(ii) *Increasing the sophistication of flight passenger movement data and models*. Existing models of disease movement over air networks are generally driven by flight capacity and direct flight data [48], missing valuable information on stopovers, actual passenger numbers and lengths of stay. Sample datasets on ticket sales and flight occupancy (e.g. US Transtats T100 and DB1B data http://www.transtats.bts.gov) should be utilized to derive models that can better replicate realistic passenger flows, for integration with the disease risk data[68]. Also, the incorporate of proximity measure (such as geographical distance or flight time between pairs of airports) and population information which airports serve is likely to facilitate the estimations of the actual travel flows between two airports [71,72].

(iii) The *development of stochastic analogues of existing deterministic approaches* to modeling of disease movement through air networks that are capable of handling input parameter distributions rather than simple mean values, and provide measures of uncertainty with output forecasts. The process of disease importation is a stochastic process, and, depending upon the disease, each relevant variable (e.g. seasonal variations in transmission, passenger numbers, and infection risk) can exhibit substantial variations from the mean and include uncertainty in the way they are measured. By simulating risks of importation from literature-derived probability distributions for each variable, improved and more informative model outputs could be produced that would enable the user to better understand and manage the uncertainties inherent in forecasts (e.g. [74]).

Conclusions

Increases in global travel are happening simultaneously with many other processes that favor the emergence of disease [75,76]. Air travel is a potent force in disease emergence and spread, and the speed and complexity of modern aviation makes both geographical space and the traditional ‘drawbridge’ strategy of disease control and quarantine increasingly irrelevant [45]. With no apparent end in sight to the continued growth in global air travel, we must expect the continued appearance of disease vector invasions and vector-borne disease movement. Approaches that can inform decision makers on the risk factors behind vector-borne disease importation and onward spread risk can be used to focus surveillance and control efforts more efficiently. The VBD-AIR tool shows that multiple datasets on many aspects relating to the risk of movement of vector-borne diseases and their vectors through the global air network can be compiled to provide such information. VBD-AIR is available at www.vbd-air.com.

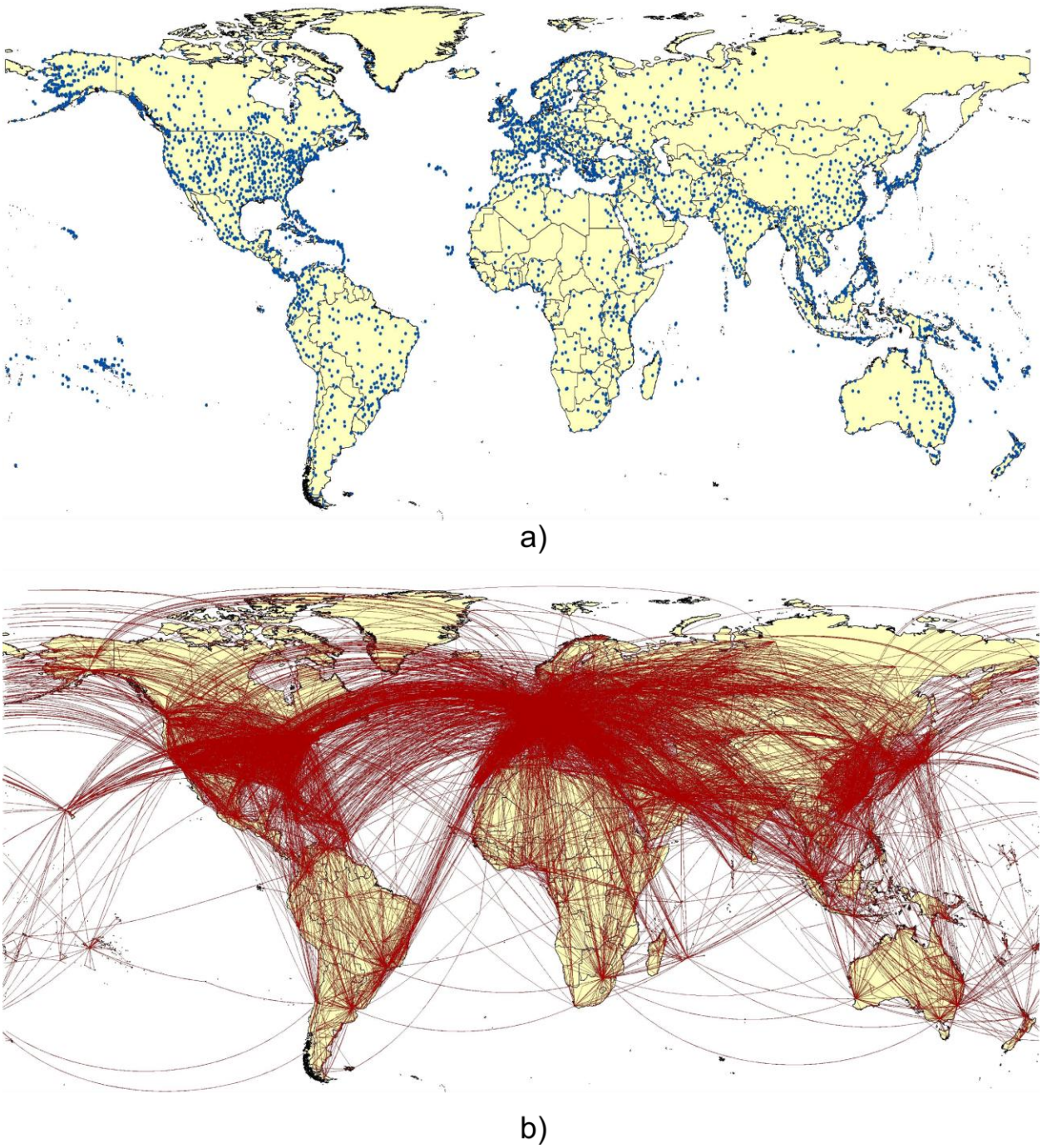
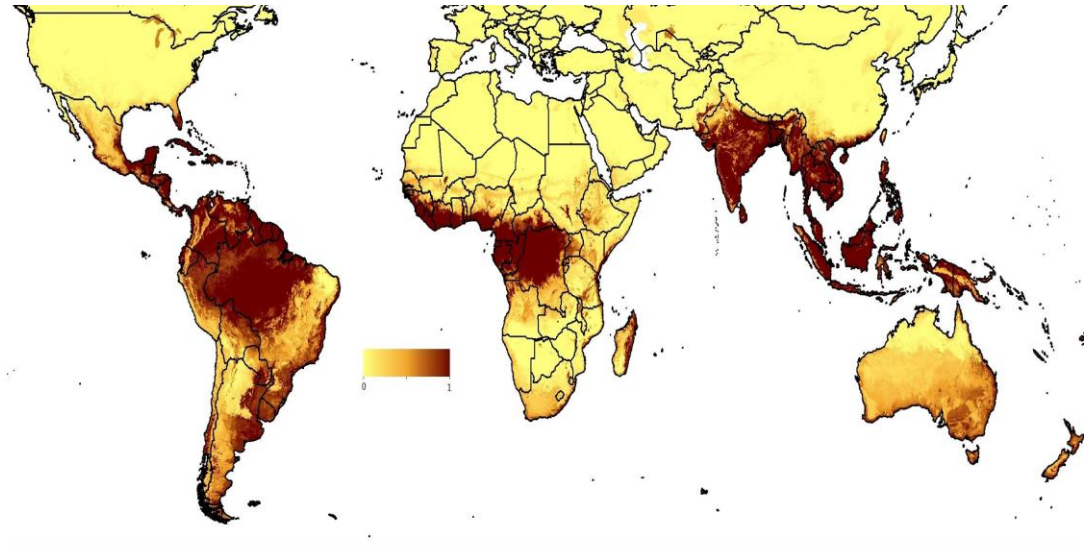
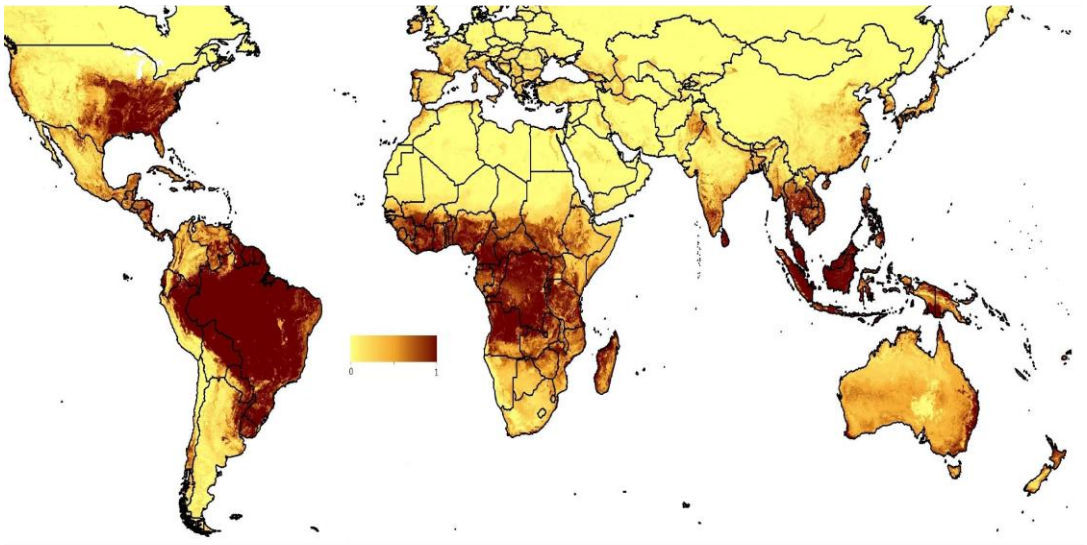


Figure 2-1. 2011 Air network data used in VBD-AIR. (a) The location of the 3,632 airports across the world; (b) the flight routes for 2011 captured within the VBD-AIR tool.



a)



b)

Figure 2-2. Example disease and vector distribution maps from the VBD-AIR tool. The predicted distribution of (a) chikungunya outbreak risk based on geolocated data on recorded outbreaks since 2008 combined with satellite-derived environmental covariates within a boosted regression tree species distribution prediction model. The color scale shows predicted unsuitable to suitable conditions for outbreaks as a continuous scale from yellow to dark blue. (b) climatic and environmental suitability for *Aedes albopictus* presence based on field survey data combined with satellite-derived environmental covariates within a boosted regression tree species distribution prediction model. The colour scale shows predicted unsuitable to suitable conditions as a continuous scale from yellow to dark blue.

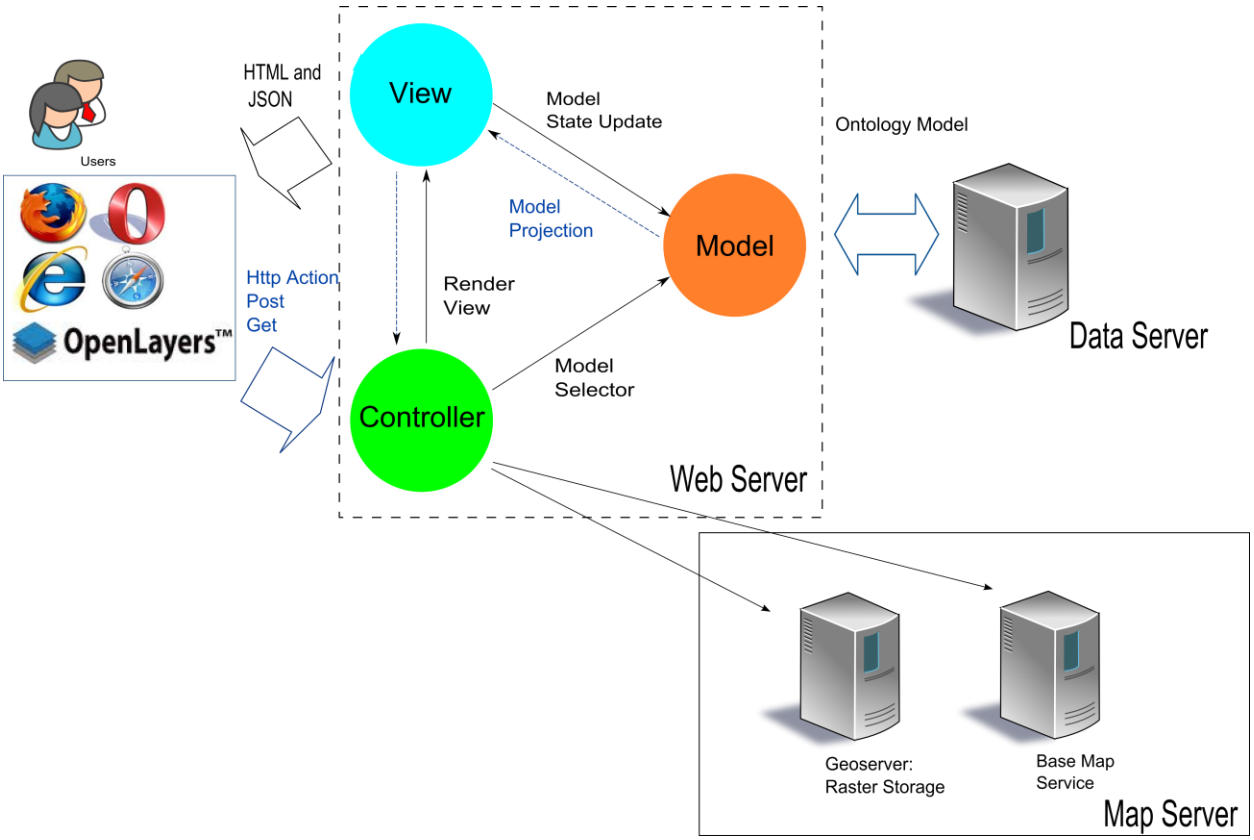


Figure 2-3. The architecture of the VBD-AIR tool. VBD-AIR adopts a three tier design: Web Server Tier, Data Server Tier and a Map Service Tier. The web interface follows AJAX standards. The web server implements a Model-View-Controller (MVC) framework. The data server is an SQL database, employed in a snow-flake structure and interpolated in the web server as data entities. The Web Mapping Services hosts the coverage file for predicted disease and vector distributions on a GeoServer. The base map service retrieves maps from Google and OpenStreetMap.

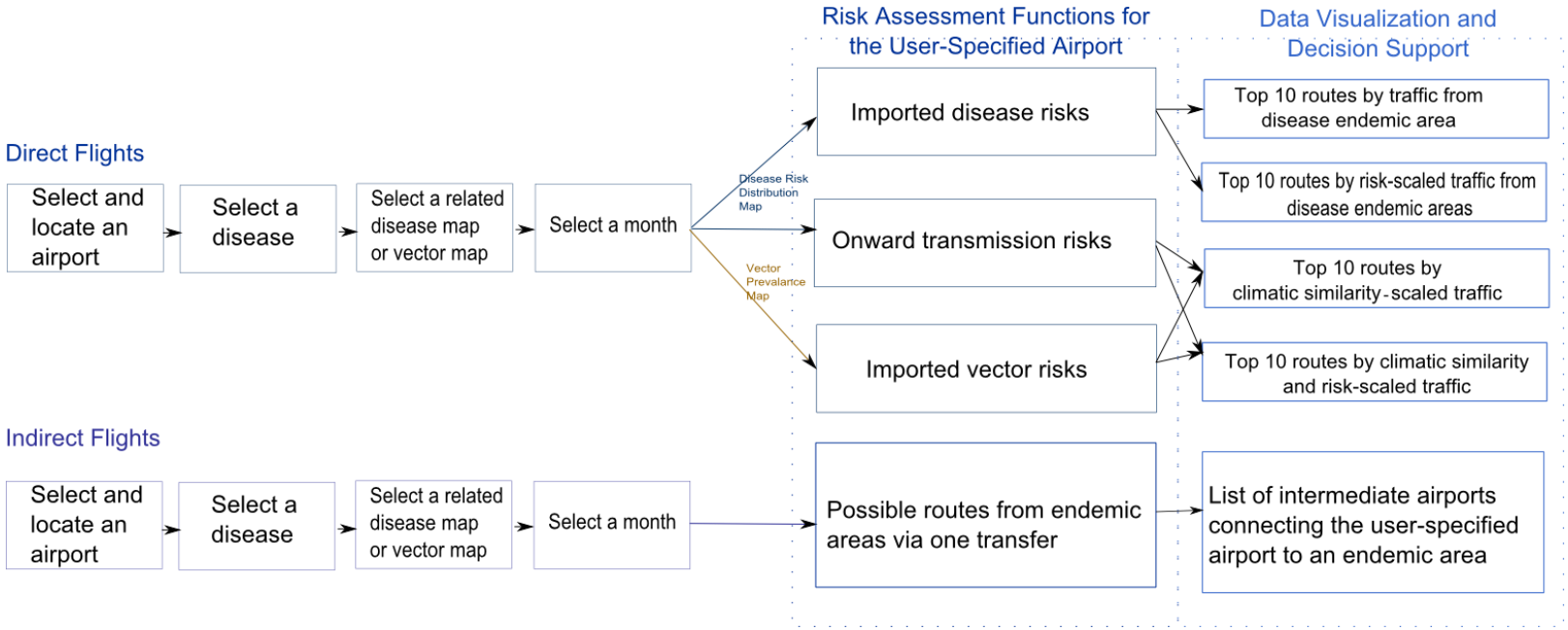


Figure 2-4. Flow of user input for the VBD-AIR tool. Risk assessment functions for the VBD-AIR tool. “Imported disease risks”, “onward transmission risks” and “imported vector risks” are provided for the direct flight route scenario and the possible routes from endemic area via one transfer are provided for the indirect flight scenario.

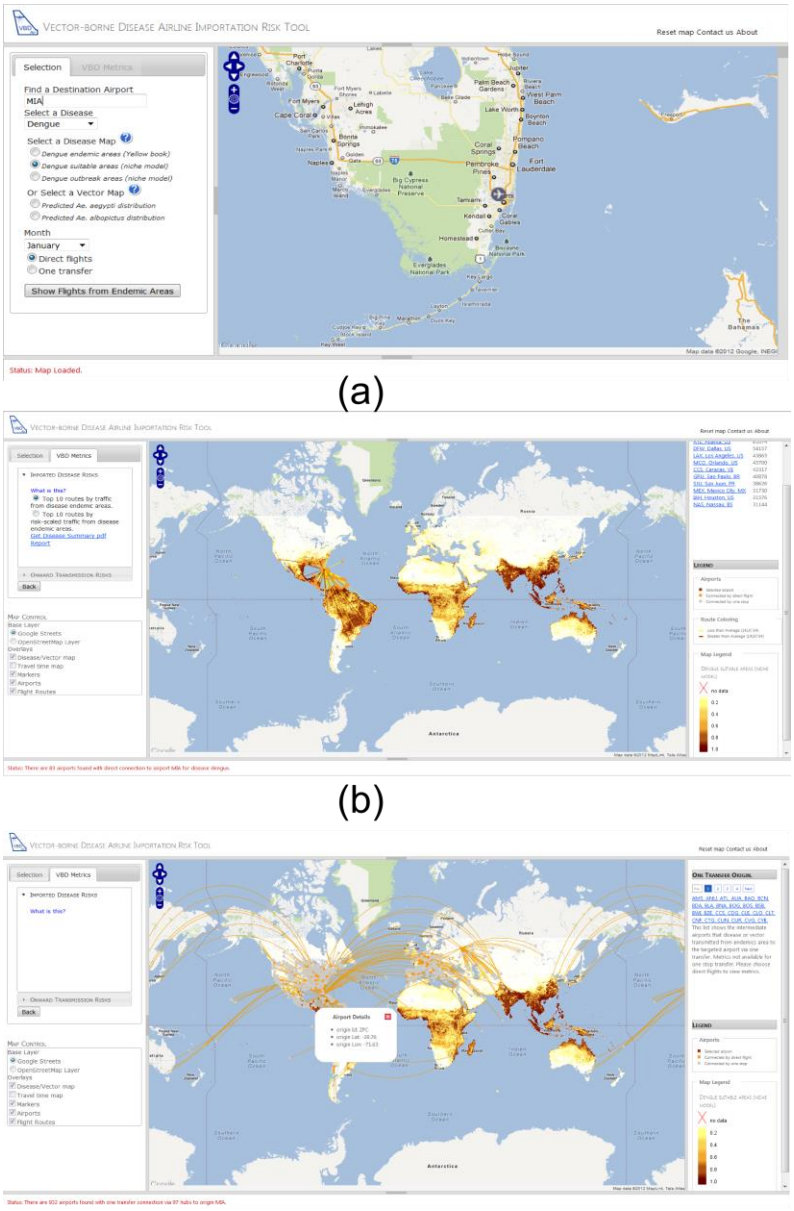


Figure 2-5. User input example for the VBD-AIR tool.A sample user input for the VBD-AIR tool for a user interested in imported dengue infection risks to Miami in January. a) The user selects their airport, disease and month of interest; b) The result shows the direct flight routes from dengue suitable areas to Miami, and the user can select risk assessments that include imported disease risk and onward transmission risks; c) The result shows the direct and one-transfer flight routes from dengue suitable areas to Miami, and the user can navigate through all the results for more detailed information.

CHAPTER 3
AN OPEN-ACCESS MODELED PASSENGER FLOW MATRIX FOR THE GLOBAL AIR NETWORK IN 2010[2]

Chapter Summary

The expanding global air network provides rapid and wide-reaching connections accelerating both domestic and international travel. To understand human movement patterns on the network and their socioeconomic, environmental and epidemiological implications, information on passenger flow is required. However, comprehensive data on global passenger flow remain difficult and expensive to obtain, prompting researchers to rely on scheduled flight seat capacity data or simple models of flow. This study describes the construction of an open-access modeled passenger flow matrix for all airports with a host city-population of more than 100,000 and within two transfers of air travel from various publicly available air travel datasets. Data on network characteristics, city population, and local area GDP amongst others are utilized as covariates in a gravity model framework to predict the air transportation flows between airports. Training datasets based on information from various transportation organizations in the United States, Canada and the European Union were assembled. A log linear model controlling the random effects on origin, destination and the airport hierarchy was then built to predict passenger flows on the network, and compared to the results produced using previously published models. Passenger flows between 1491 airports on 644,406 unique routes were estimated and analyses showed that the model presented here produced improved predictive power and accuracy compared to

[2] A version of this chapter was submitted to *Plos One* as "Huang Z, Wu X, Garcia AJ, Fik TJ, Tatem AJ (2013) An open-access modeled passenger flow matrix for the global air network in 2010. *PloS One"*

previously published models, yielding the highest successful prediction rate in terms of prediction power and accuracy at the global scale. The airport node characteristics and estimated passenger flows are freely available as part of the Vector-Borne Disease Airline Importation Risk (VBD-Air) project at: www.vbd-air.com/data

Introduction

Demand for travel has boosted the growth of the global air travel network at an unprecedented rate. In the past 20-30 years, the network has expanded dramatically with a steady growth rate of 4-5% per year [1], accompanied by a nearly 9% annual growth rate of passenger and freight traffic [77]. In 2011, the worldwide international and domestic passenger kilometers transported reached a record-high of 5.2 trillion kilometers [43]. The large volumes of air traffic, result in profound impacts on commodity trade[78], regional development[79], cultural communication[80], disease importation[7,8] and species invasion [5,9,10]. As humans and commodities are transported at exceptional rates through aviation compared to other modes of transportation, how these patterns impact the socioeconomic, environmental and epidemiological landscape is of significant interest[5,7,9,81].

Quantifying the volume of passengers on the air travel network is critical to understanding the complicated spatial interaction between origin and the destination cities [7,8]. Previously, studies from a range of fields [3,5,9,10,82–84] have made use of data from the International Air Transport Association (IATA) or the International Civil Aviation Organization (ICAO). These data are often restricted to scheduled flight plus seat capacity information on routes. However, not all commercial flights operate at full capacity, thus, such data often overestimate the passenger numbers on routes [7]. Moreover, capacity data provide information on only point-to-point connection, thus,

travel patterns that require a stopover and transfer of planes are not captured [68]. Although Origin-Destination data derived from air ticket sales are available (e.g. http://www.iata.org/ps/intelligence_statistics/paxis/pages/index.aspx), such data are expensive for research purposes, running to many tens of thousands of dollars, and can require significant legal and confidentiality agreements for data usage. Other databases of international flow by pair-wise airports are held by private companies (e.g. Marketing Information Data Transfer, http://ma.aspirion.aero/midt) with costs, again, often high to obtain, with payment required repeatedly to maintain the latest data. Here we aim to outline a model framework to produce open access estimates of global air traffic flow for research purposes that can be updated regularly.

Spatial interaction models have been utilized to estimate the volume of passengers given an origin and destination city where data are lacking [3,72,78,82,83,85–90] . The most common model used is the gravity model, which incorporates drivers such as the site characteristics of origin and destination areas, and functions of “locational separation” to depict the interaction between the origin and destination to estimate the flow value. As Grosche et al [90] summarized, the drivers in the spatial interaction model to estimate the air traffic include 1) socio-economic characteristics of origin, destination, such as population, income, GDP, urban infrastructure, education level , and 2) service-related factors such as the quality (such as flight frequency, plane size and prices) and the market demand of airline service. The locational separation is usually calibrated by the distance or travel time separating origins and destinations. The gravity model provides a solid theoretical and practical

background on understanding the movement of populations since it explicitly captures the absolute and relative spatial relationship of the origin and the destination [86].

The utilization of network characteristics sheds light on the identification of air service factors in the gravity model for flow estimation, since 1) the layout of the global air travel network follows the “hub-and-spoke” network model and 2) heterogeneities in the network topologies are indicated by the demands of air travel of the area where the airports serves. Firstly, large air travel companies in mature air travel markets adapt a hub-and-spoke model to achieve the balance of travel time for customer and the efficiencies in transportation infrastructures. In this model, a single airport is assigned to a single hub or multiple hubs to form a regional inter-connected community [91–93] , where “stop over” and “feeder” routes exist connecting the small airports with low degree connections to a larger degree hub [94] . The locations of airport hubs are selected as the optimum locations that satisfy the inter-regional travel demands and minimize the total transportation cost [91,92]. Moreover, the hub-and-spoke layout can be reflected on the “small-world” and the “scale-free” characteristics on the network. Guimera et al [95] studied the “small-world” feature and showed that most airports can be reached from every other with only a small number of connections. They also identified how central nodes with low degree connectivity play an important role for inter-regional and intra-regional communication. The “scale-free” feature that the degrees of the air travel network follow a power-law distribution is suggested by the nodal structure of flows clusters [96], as described by the hierarchical span for the major airports in the United States [84].

Secondly, the degree and centrality of airports in the air travel network can act as indicators for air travel demand, since the local measurement of air passenger volume, population, and the level of economic activities at the periphery of the hub are highly correlated [97–99]. Empirical research [100–102] have observed the increment of air passengers and increased passenger flow based on economic growth. Liu et al. [103] quantified the marginal effects of population in a given metropolitan area on the air travel market, indicating that the odds of having a 'major' air traffic market increase 41% per 100,000 population growth. Wang et al [104] studied the air travel network in China and found that cities in the more urbanized area of East China had a higher centrality score and a higher number of air passenger volumes compared to the more rural West China. These studies indicate the mutual correlation of network centralities and urban development level, which reflect the spatial agglomeration of economic activities and unequal air travel service demands.

To study the movement of vector-borne disease on the air travel network , Johansson et al [68,74] modeled the actual passengers counts between 141 airports worldwide with epidemic significance. Utilizing the air travel itineraries of the United States as a training set, they constructed a generalized linear model with a Poisson link to estimate worldwide passenger flows using nodes and routes characteristics as model covariates. Their results demonstrated good fits for the flow prediction on true origin-destination travel. Our research follows the general modeling framework used in Johansson et al [68,74], but extends to a global model which includes: 1) all the nodes with a host-city population of more than 100,000; 2) the routes between all airports that are within 0, 1 or 2 stops on the air travel network.

Materials and Methods

Airport Locations and Scheduled Routes

Information on a total of 3,416 airports across the world, together with their coordinate locations was obtained using Flightstats (www.flightstats.com) for 2010. The connectivity and scheduled air travel network routes were defined by a 2010 scheduled flight capacity dataset purchased from OAG (www.oag.com). These included information on direct links (if a commercial flight is scheduled) of origin and destination airports, flight distances, and passenger capacity by month for 2010. Directly connected airports pairs were utilized to construct a graph for the air travel network in 2010 with 3416 nodes and 37674 edges. The average degree of the network was 22.06, with the maximal degree recorded as 476 for Frankfurt Airport (IATA code: FRA). The topology of the graph exhibits both small-world and scale-free properties as already observed in similar global or regional air travel dataset analyses [95,105,106]. The coefficients of the power law function fitting the scaled-degree distribution was 1.01±0.1, which concords with a previous study [95]. The average path length is 4.11, measured as the average number of steps travelling from one node to another node, while the diameter of this network was 14, which indicates the shortest path between the two most remote airports. Based on the network created by the flight statistics assembled, we calculated the degree, centrality and strength for each node and use these measurements as covariates at the modeling stage.

Gross Domestic Product and Population Information

Generally, socio-economic variables at a global scale are difficult to obtain. The G-Econ data (http://gecon.yale.edu/) provide indices representing both market exchange rates (MER) and purchasing power parity (PPP) at a 1-degree longitude by 1-

degree latitude resolution at a global scale. Due to the large geographical coverage of the grid cells, we extracted the closest PPP value for an airport and calculated the PPP value per capita in 2005 by dividing the purchasing power parity by the population value in each grid cell. These data are utilized as local economic measurements for each airport.

Given computing power limitations on the modeling and matrix sizes, we selected the airports serving a city population number more than 100,000. To select these airports, a web crawler built on the WolframAlpha API (http://products.wolframalpha.com/api/) was used to extract the city populations for each airport. Wolfram alpha is a knowledge engine which is capable of computing population information from various sources including: U.S census data, United Nations urban agglomeration and City Population (http://www.citypopulation.de/) data. These data capture the most recent city population estimates from these data sources (for cities in United States, the US Census 2010 data are utilized). In our database, there are 1491 airports satisfying these criteria.

Actual Travel Passenger Flow

Data on passenger origins and destinations on the air travel network were obtained from a variety of sources:

- The DB1B market data from the Airline Origin and Destination Survey (DB1B) provides a 10% sample of U.S. domestic passenger tickets from reporting carriers, including information such as the reporting carrier, origin and destination airports, prorated market fare, number of market coupons, market miles flown, and carrier change indicators. To create a training dataset, these data were aggregated annually by the origin and the destination airport code with the sum of counts of itineraries. To protect the US air travel industry, the reported international Origin-Destination data by the U.S carriers is strictly restricted to U.S citizens, and requires detailed statements on the use of the data. Hence, the research presented here did not take into account the international portion of the Origin-Destination data from DB1B.

- The Canadian transport department provides statistics relating to the movement of aircraft, passengers and cargo by air for both Canadian and foreign air carriers operating in Canada (http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2703&lang=en&db=imdb&adm=8&dis=2). This survey provides estimates of the number of passengers traveling on scheduled domestic commercial flights by directional origin and destination city pairs. In this survey, significant numbers of Canada-U.S trips were reported. The city pairs were matched to the airport pair that had the shortest routes defined by the OAG database with the passenger number obtained from the above data source. For example, passenger numbers between Toronto and New York City were matched to the direct route of YYZ to JFK, since it is the shortest route between these two cities.

- Detailed route data for passenger numbers from EuroStat (http://epp.eurostat.ec.europa.eu/portal/page/portal/transport/data/database). This database presents passenger numbers between the main airports of reporting countries and their main partner airports in the European Union.

- All of these flow statistics were utilized to create a global prediction O-D matrix, described below.

Data Processing

Cities are situated in a complex hierarchical network and the flows between cities are either constrained or facilitated by this hierarchical structure [78,107]. We defined three levels of economic activity for each city per capita based on the 33% quartile of the distribution of PPP per capita. Thus, nine types of economic links are identified (low-low, low-medium, low-high, etc.) to reflect the type of flow within/across the economic hierarchies. Similarly, we defined four levels of hierarchy based on the degree distribution of the airports, and sixteen types of flows are identified to reflect the type of flow within/across the air service hierarchies.

A prediction dataset framework for routes was constructed based on the adjacency matrix defined by the OAG dataset. For each airport, destination airports via first-order connection, second-order connection and third-order connections on the air travel network were identified. Along these routes, information on the minimum number

of stopovers and the maximum seat capacity were calculated. Moreover, following approaches outlined in Bhadra's research [97] we defined a categorical variable for distance classes to separate the markets by stage lengths, with 1 for short-haul (2000 kilometers or less), 2 for medium-haul (between 2000 and 3500 kilometers) and 3 for longer hauls (3500 or more kilometers). We excluded routes less than 200 km since passengers are believed to have more efficient and effective land-based methods to travel such small distances. Note that only 3842 possible routes (<0.001%) are less than 200 km. Finally, an origin-destination (OD) pair list with 1,295,752 rows was created.

For analytical purposes, the global OD pair list was constructed following these assumptions:

- The sum of passenger counts between an airport pair only represents the number of arrivals at a final destination. If the maximum capacity of a route is greater than the actual passenger number found in our datasets, the maximum capacity variable will be replaced by the actual passenger number, which is the sum of passengers from all possible routes (including directed and undirected).
- Passengers always take the shortest path to their destination city, and they don't stop at the connecting city. The data used for modeling is itinerary data which represents the minimum number of stops from one airport to another. Hence, passenger numbers in our database represent the flows for the first order, the second order and the third order of network connections. We assume that passengers choose the first shortest path found by a breadth-first search algorithm, as the route is found by iterating all the neighboring nodes until a path from the origin and the destination is identified. If both the origin and the destination cities have multiple airports, the passengers are assumed to take the shortest path from all possible routes between these airport pairs, which usually resulted in the path between the two largest airports in terms of capacity. This is supported by Button et al [108]'s research that passengers tend to choose a larger hub for their travel.
- Passengers do not choose routes with more than two stops. We used the number of stops as a categorical variable rather than a numeric variable since it is considered to be a measure of hierarchical accessibility. In fact, for the air travel network in 2010, all of the possible calculated routes within two stops covered 83% of all the possible connections. Also, multiple-stops (more than two

stops) are comparatively rare as a share of total passengers in our actual travel flow datasets. In DB1B domestic datasets, there are no itineraries for travels between cities with a population size more than 100,000 within two stops.

Network characteristics were calculated using the igraph (http://igraph.sourceforge.net/) library in R (http://www.r-project.org/). A summary of variables included in the model is presented in Table 3-1.

Model

The model takes the form of the general gravity model:

$$P_{ij} = f\left(Node_i,\ Node_j,\ Route_{ij}, Interactions_{ij}\right) \quad (2\text{-}1)$$

where *Pij* is the annual aggregated passenger flow between Node i and j. $Node_i$ and $Node_j$ denote the collection of node characteristics, which are considered to be drivers of the size of the flow. $Route_{ij}$ denote the collection of route characteristics, which are considered to be the proximity measurements. $Interactions_{ij}$ denotes the collection of two-way interaction effects between categorical variables such as stops, country, degree link type, economic link type and haul type with other node and route characteristics.

For the purpose of better estimation and thus prediction, we tested four model forms which include 1) a lognormal model for main effects only. This model adopts the general gravity model framework as the one described in Balcan et al [109]. To utilize this model, a logarithm transformation is performed for each quantitative variable. The main effects include both node and route characteristics. 2) A generalized linear model for main effects and interactions with Poisson distribution and a log link. This model is the model utilized by Johansson et al [68,74] for predictions of the traffic flows between epidemiologically significant cities. 3) A generalized linear model for main effects and

interactions with a negative binomial distribution and a log link. This model is similar to model 2 except that it utilizes a negative binomial distribution to account for the possible over-dispersion in the data. 4) A lognormal mixed model with main effects, interactions, and random effects on origin and destination city (note that a logarithm transformation is performed for each quantitative variable as well). This model assumes that the passenger flows are independent between different degree link type but correlated within the same degree link type, while model 1-3 make the assumption that all passenger flows are independent of each other, which is very strong and unrealistic in practice. Random effects are thus included to account for the dependence among passenger flows and the possible heterogeneity between levels of air travel services. More detailed model descriptions can be found in the Appendix B.

Apart from model fitting on the whole dataset, cross-validation is performed to evaluate how accurately each model will predict in practice: firstly, the dataset is randomly partitioned into 10 subsets, each consisting of 10% of the observations. Then on each of the subsets (the testing set), we validate the analysis performed on the remaining data (the training set). Lastly, the validation results are averaged over the rounds. Three criteria are chosen for model evaluation: 1) the coverage rate of the 95% prediction intervals, which measures the percentage of the observations that fall into the corresponding 95% prediction intervals; 2) the coverage rate of the ±30% observation intervals, which measures the percentage of the predictions that fall into the ±30% intervals of the corresponding observations; 3) the successful prediction rate, which measures the percentage of predictions that fall into the same magnitude category as the corresponding observations. These magnitude categories are defined by dividing

the passenger flow numbers into five groups: 10^2 and under, 10^2-10^3, 10^3-10^4, 10^4 -10^5, and 10^5+, each group represents one category.

Results

Model Comparison

For each model, most coefficients are significant at the 0.05 significance level as the percentages of the significant coefficients are about 90%, 100%, 96%, 95% respectively for model 1 to 4. For the purpose of prediction, we keep all the covariates in the model instead of removing the non-significant ones. Not surprisingly, most of the interactions between node and route characteristics play an important role in model estimation as we treated the number of stops as a categorical variable. The interaction between haul types and inverse distance is also significant, which agrees with previous work [97].

Both model 1 and model 2 provide narrow confidence intervals for predictions, while model 3 and model 4 provide wider intervals to accommodate variation in the data. All of these models have at least 68% successful prediction rates for predicting the magnitude of passenger flow. According to the results presented in Table 3-2 and Table 3-3, model 4 provides the most accurate prediction.

For each of the models, we calculated the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). RMSE is a frequently used measure of the differences between estimate values and the values actually observed. A smaller RMSE suggests a better model fit. MAE is the average of the absolute value of the prediction errors, which serves the same purpose as RMSE and is believed to be more robust in many situations. As shown in Table 3-3, model 4 yields the lowest RMSE and MAE for the majority of the data points except for extremely large observations. For the largest

observed passenger value category, model 2 gives the lowest RMSE and MAE, while model 4 gives the second lowest RMSE and MAE.

Figure 3-1 presents the prediction and diagnostic plots for Model 4. Panel a) shows that most of the prediction values are close to the y=x (prediction=observation) line. Panel b) shows that most of the residuals scatter along the y=0 (residual=0) line, yielding no obvious pattern. Both plots indicate that Model 4 is a plausible model for the passenger flows. However, the prediction seems poor at the lower tail. This is expected, given likely randomness in the smaller amount of passenger exchanges between airports [68]. Diagnostic plots for other models are presented in the Appendix B.

Alternative diagnostics for testing the model fit were performed for model 4 as well. Firstly, a multilevel model described in Snijders et al [110] and implemented in the SAS code written by Recchia et al [111] to calculate r-squared measures for the fourth model was utilized. The first level of the model was found to explain 84.0% of the variance in the data and the second level explained 98.7% of the variance, indicating a good model fit. Secondly, for the directly connected flights, we compare both the predicted value from model 4 and the capacity data from OAG to the observed passenger flows on a log scale using the paired t-test. The results show evidence of difference between the mean predicted passenger number and mean observed passenger number, and between the mean capacity number and mean observed passenger number, both at the 0.05 significance level. However, the geometric mean ratio of log (predicted value) to log (passenger number) was 1.01(panel c) in Figure 3-1), while the geometric mean ratio of log (capacity) to log (passenger number) was 1.08(panel d) in Figure 3-1). The predicted value shows more agreement on the

observed value, while the capacity data represents a significant overestimation of flows between two directly connected airports. Hence, our predicted values provide a closer approximation of the traffic flows on the air travel network compared to the maximum seat capacity metric for the directly connected cities, as used in previous studies [3,5,9,10,82–84].

In summary, our model (Model 4) outperformed the lognormal gravity model (Model 1) used in Balcan et al [109] and the Poisson model (Model 2) used in Johansson et al [68,74]. Moreover, for direct flights, our estimates show more homogenous agreements with observed passenger numbers compared to simple seat capacity data.

Prediction and Interpretation of the O-D Passenger Flow Matrix

Model 4 was applied on the estimation dataset to predict passenger flows. We have identified the over-dispersed predictions that exceeded the maximum capacity on the routes (3% of the data) and replaced them with the product of the maximum capacity on the routes. According to the training dataset, the maximum numbers of itineraries for one stop and two stop connections were 140,086 and 8,060. Since these data are generated from the mature air travel market and constrained by the network structure, we considered them as the upper limits of the data distribution. Thus we adjusted the prediction of the first-order connection and the second-order connection flights scaled by these two maximum numbers. Following this, we removed all the predictions with less than 1 person. Finally, 644,406 routes with origin, destination airport codes, number of stops and predicted passenger number were produced.

As described before, the passenger counts were categorized into five categories as a test of successful prediction rate in magnitude: $1\text{-}10^2$,$10^2\text{-}10^3$,$10^3\text{-}10^4$,$10^4\text{ -}10^5$and

10^5 and more. The first two categories present small numbers of passenger exchanges, implying random flows between two airports, and the fourth and fifth categories indicate a higher probability representing steady flows between airports. Figure 3-2 a) shows all the flows with more than 10^5 predicted passengers.

Secondly, given an origin/destination, the dataset produced through the research outlined here can estimates the endpoints and starting points with passenger flows on the air travel network. Figures 3-2 b)-d) illustrate the passenger flows and number originating from Atlanta, categorized by number of transfer. Figure 3-2 e) shows the distribution of airports with incoming passenger numbers over 5,000,000. This reflects the mature air markets of the United States and Europe, though noticeable concentrations of airports can be observed in the emerging markets such as India and China as well.

Discussion

With continuing growth of the global air travel network, we must expect continued socioeconomic, environmental, cultural and epidemiological impacts. This research shows how network characteristics combined with multiple datasets on various perspectives relating to the movements of passengers of passenger flow on the global air network can be compiled to provide estimates that are more accurate than previous modeling efforts. Such a dataset provides a valuable resource for scientists and decision makers to measure the global flow of air traffic and its potential influences.

In the database outlined here, 644,406 unique routes spanning 1,491 airports serving city populations of more than 100,000 are modeled based primarily on publicly available datasets. Our model has outperformed similar research at the global scale and can explain 98% of the variance in the data. Within the database, 23,785 routes follow a

direct connection, 291,745 routes are one-stop connections and 328,876 routes are two-stop connections. Using this route and airport information, anyone can construct flow matrices to describe the global air traffic flow and assess its multiple impacts.

Due to data constraints, a range of uncertainties and limitations exist in the output modeled datasets. The first inconsistence comes with internal uncertainties within the DB1B dataset. To construct the DB1B dataset, Transtat only requires US carriers to report O-D pair data, hence the O-D data is likely to be inaccurate in markets served with a significant number of foreign carriers (e.g., New York, Washington D.C., Chicago, and Los Angeles). If there is more than one airport in a city, each of the airports is treated as a separate node. This may well result in overestimates of the flow to secondary airports in a city.

The second set of inconsistency is the population data. Due to data availability, only city population data were utilized, when it is sometimes the case that people in neighboring metropolitan area can access the airport in question through other ground-based transportation methods (for example, people in Gainesville FL are often likely to drive two hours to Jacksonville or Orlando to take a plane, rather than utilize the Gainesville regional airport which is 10 miles away from the city center). As a result, our predictions may overstate the markets for small airports.

The third source of uncertainty stems from the fact that the data we utilized for the training datasets were only from the United States, Canada and the European Union. Thus, international flights are less well represented in our dataset and most of the flight data describes the flows between airports in high income countries. Additionally, long haul international flights with more than three stops are absent.

The topology of air travel network is likely to vary at the regional level. Wang et al [104] found that in terms of topological measurements, the Chinese air travel network is similar to the Indian one, but different than that of the US. As current air travel networks in low income countries usually feature point-to-point connections between city pairs [112], high income countries are increasingly prompted to utilized a hub-and-spoke system due to their mature air travel markets. On the other hand, it is observed that some companies (such as Southwest Airlines and Jet Blue in United States) in high income countries also adopt spoke-to-spoke models to connect hot spots of air travel demand [97]. This heterogeneity may affect the flow estimation country-wise and overestimate the driving factor of hubs in both high and low income countries.

The demand for air travel are heterogeneous and "largely determined by the spending capacity of customers" [113]. Hence, it could be anticipated that the demand for air travel in each country varies and is correlated to GDP. Also, the demographic profile of passengers on the air travel network is likely different between countries. Under a regional context, this may affect the prediction of domestic passenger numbers, while international heterogeneities in traffic flows may be attributed to differing visa policies between countries [114]. Visa restrictions may reduce traffic flows substantially between countries [115]. Moreover, cultural differences at a country level could represent indicators of attraction and drivers of population movements [80,116].

The potential limitations discussed above arise through the constraints of the data sources used. These may be alleviated through incorporation of more publicly accessible data in future work, including 1) More detailed economic indicators (such as GDP, income etc.) at the city level: such measures could further describe drivers in the

gravity model. 2) Itineraries from low income regions of the world: such data would enlarge our training and testing databases to avoid sampling errors. 3) Hub characteristics (such as the number of enplanements, transfers and deplanements): these measures could help explain the function of the hubs in controlling network flows. Alternatively, transportation forecasting models [117,118] and radiation models [119] could be utilized to estimate the global O-D matrix based on the traffic counts on nodes and edges.

Conclusion

The research presented here has documented the generation of a world-wide Origin-Destination matrix of passenger flows in 2010 for airports with host city populations of more than 100,000. Results show that the modeled dataset improves substantially on the accuracy of datasets used in previous studies. The datasets are freely accessible for academic use and are published as part of the Vector-Borne Disease Airline Importation Risk (VBD-Air) project at www.vbd-air.com/data/ .

Table 3-1. Descriptions of covariates used in the modeling process

Variables	Descriptions
Node characteristics	
Pop_i	The population of the origin city
Pop_j	The population of the destination city
$PPP2005_i$	The purchasing power index where the origin airport serves
$PPP2005_j$	The purchasing power index where the destination airport serves
$PDA2005_i$	The purchasing power per capita index where the origin airport serves
$PDA2005_j$	The purchasing power per capita index where the destination airport serves
$Strength_i$	The sum of the edge weights of the adjacent edges for each vertex for the origin city
$Strength_j$	The sum of the edge weights of the adjacent edges for each vertex for the destination city
$Degree\_Out_i$	The degree number of the origin city on the air travel network
$Degree\_In_j$	The degree number of the destination city on the air travel network
$Closeness\_Centrality_i$	The mean geodesic distance between a given node and all other nodes with paths from the given node to the other node. This variable is calculated according to the origin city.
$Closeness\_Centrality_j$	The closeness centrality measure for the destination city.
$Betweeness\_Centrality_i$	The number of shortest paths going through a specific airport. In a weighted network, betweenness centrality is a useful local measure of the load placed on the given node in the network as well as the node's importance to the network than just relying or connecting.
$Betweeness\_Centrality_j$	This is the calculation of betweeness centrality for the destination airport.
Route characteristics	
Inverse Distance	Inverse great circle distance between the origin and the destination airport
Country	Indicates whether the origin and the destination are in the same country.
Alternative	Number of alternative routes to the destination
Stops	Number of stops on the shortest route from the origin to the destination
MaxC	The maximum capacity along the shortest path
Degree Link Type	This variable identifies the types of flows between different hierarchies of airports defined by the air travel services level.
Economic Link Type	This variable identifies the types of flows between different hierarchies of airports
Haul Type	This variable differentiates the effect of long haul flights. 1 for short-haul (2000 kilometers or less), 2 for medium-haul (between 2000 and 3500 kilometers) and 3 for longer hauls (3500 or more kilometers).

Table 3-2. Comparison of the four models with respect to prediction accuracy (in percentages)

	Coverage rate of the 95% prediction intervals	Coverage rate of the 95% prediction intervals (cross-validation)	Coverage rate of the ±30% observation intervals	Coverage rate of the ±30% observation intervals (cross-validation)	Successful prediction rate	Successful prediction rate (cross-validation)
Model 1	6.39	6.73	29.82	29.76	68.42	68.33
Model 2	4.80	4.79	31.48	30.63	69.16	68.80
Model 3	23.16	24.16	33.09	33.38	70.04	69.43
Model 4	52.11	49.86	47.79	31.17	79.72	70.41

Table 3-3. Root Mean Squared Errors and Mean Absolute Errors for all models

Measurement	Categories	Number of Records	Model 1	Model 2	Model 3	Model 4
RMSE	Observed Passenger (OP)<10^2	2379	1680	2923	1947	726
	OP in 10^2-10^3	6440	3536	5127	32405	1802
	OP in 10^3-10^4	7314	7397	8771	10639	4346
	OP in 10^4-10^5	4817	20780	23002	41585	21940
	OP >10^5	1132	163352	85897	216610	127194
MAE	Observed Passenger (OP)<10^2	2379	286	538	402	120
	OP in 10^2-10^3	6440	629	1073	1413	333
	OP in 10^3-10^4	7314	2729	3218	3140	1621
	OP in 10^4-10^5	4817	14697	14929	19689	13415
	OP >10^5	1132	115710	61305	94447	89233

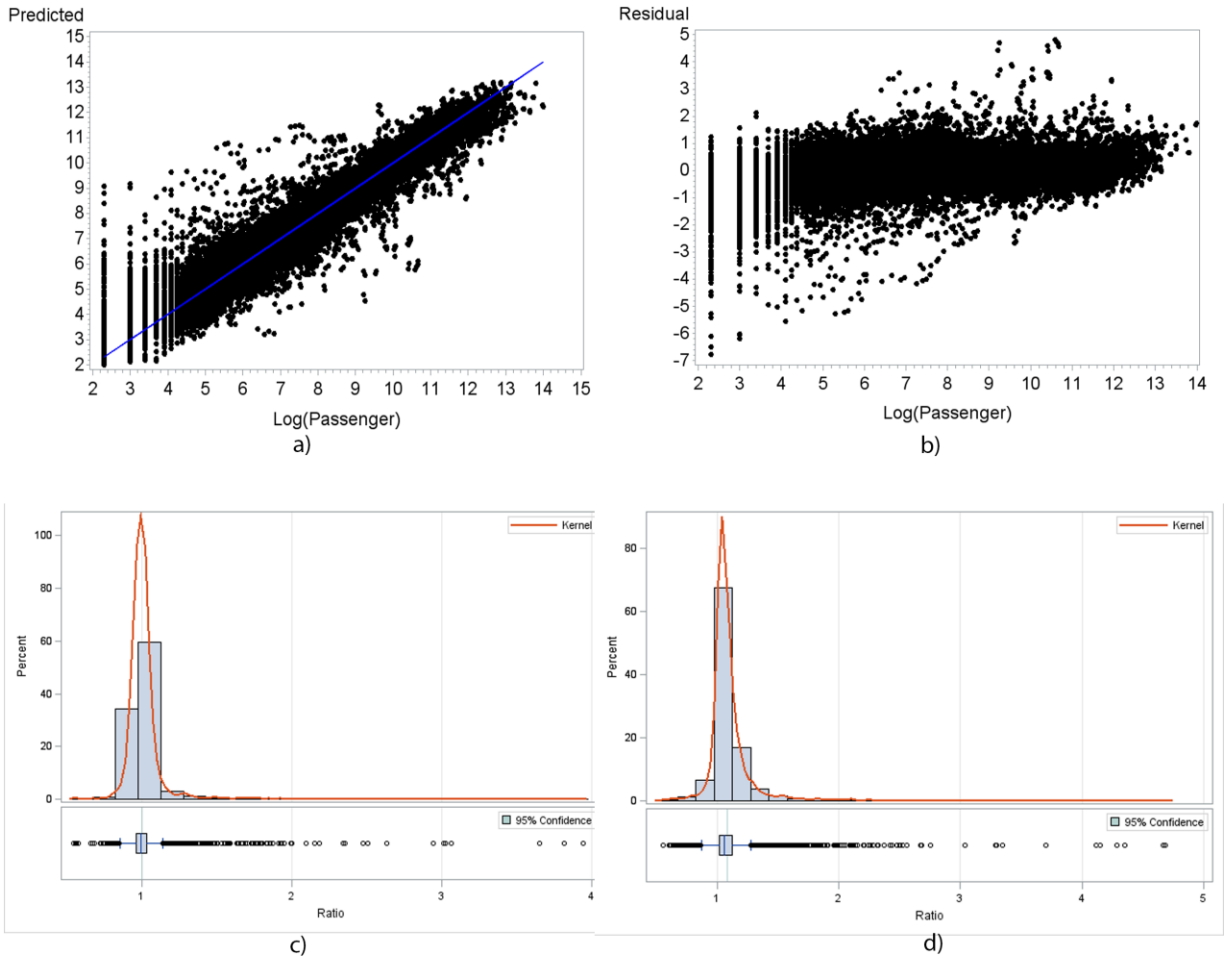


Figure 3-1. Diagnostic plots for all the models. a) Predicted vs. observed value of model 4. b) Residual vs. observed value of model 4. c) Distribution of ratio of predicted value vs. observed value in log scale with 95% confidence interval for geometric mean. d) Distribution of ratio of capacity vs. observed value in log scale with 95% confidence interval for geometric mean.

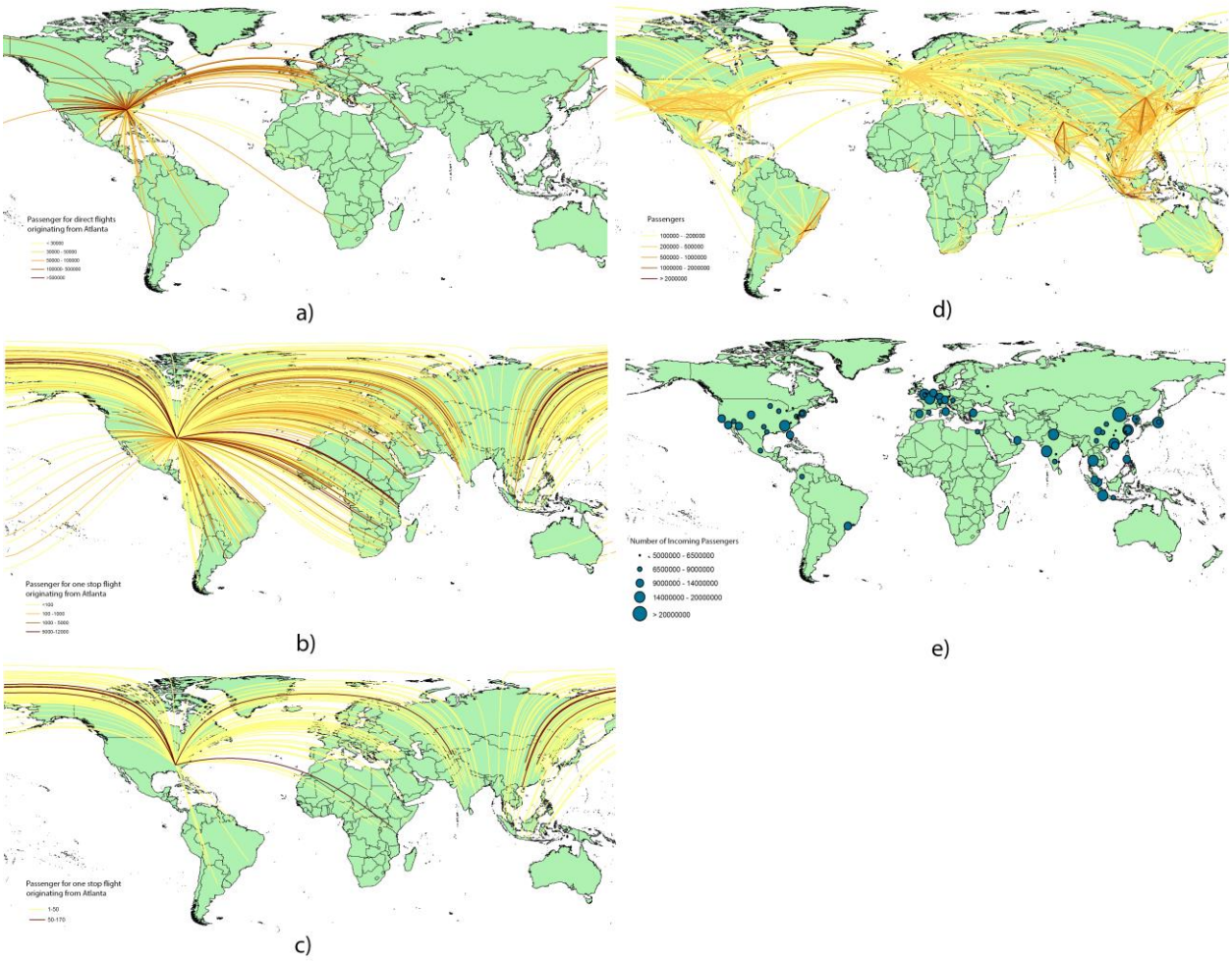


Figure 3-2. Predicted air traffic flows. a) Predicted flights with passenger flows of more than 100,000. b) All possible passenger flows through direct flights originating from Atlanta. c) All possible passengers' flows through one-stop flights originating in Atlanta. d) All possible passengers' flows through two-stop flights originating Atlanta. e) All airports with an incoming passenger numbers more than 5,000,000.

CHAPTER 4
GLOBAL MALARIA CONNECTIVITY THROUGH AIRTRAVEL[3]

Chapter Summary

Air travel has expanded at an unprecedented rate and continues to do so. Its effects have been seen on malaria in rates of imported cases, local outbreaks in non-endemic areas and the global spread of drug-resistance. With elimination and global eradication back on the agenda, changing levels and compositions of imported malaria in malaria free countries, and the threat of artemisinin resistance spreading from Southeast Asia, there is a need to better understand how the modern flow of air passengers connects each *P. falciparum* and *P.vivax* endemic region to the rest of the world.

Recently constructed global *P. falciparum* and *P.vivax* malaria risk maps along with data on flight schedules and modelled passenger flows across the air network, were combined to describe and quantify global malaria connectivity through air travel. Network analysis approaches were then utilized to describe and quantify the patterns that exist in passenger flows weighted by malaria prevalences. Finally, the connectivity within and to the Southeast Asia region where the threat of artemisinin resistance is highest, was examined to highlight risk routes for its spread. .

The analyses demonstrate the substantial connectivity that now exists between and from malaria endemic regions through air travel. While the air network provides connections to previously isolated malarious regions, it is clear that great variations exist, with significant regional communities of airports connected by higher rates of flow

[3] A version of this chapter has been submitted to *Malaria Journal* as "Huang Z. and Tatem AJ, Global Malaria Connectivity through airtravel."

standing out. The structures of these communities are often not geographically coherent, with historical, economic and cultural ties evident, and variations between *P.falciparum* and *P.vivax* clear. Moreover, results highlight how well connected the malaria endemic areas of Africa are now to Southeast Asia, illustrating the many possible routes that artemisinin resistant strains could take.

The continuing growth in air travel is playing an important role in the global epidemiology of malaria, with the endemic world becoming increasingly connected to both malaria-free areas and other endemic regions. The research presented here provides an initial effort to quantify and analyse the connectivity that exists across the malaria endemic world through air travel, and provide a basic assessment of the risks it results in for movement of infections.

Background

The worldwide air travel network has expanded at an exceptional rate over the past century. International passenger numbers are projected to rise from 1.11 billion in 2011 to 1.45 billion by 2016, with an annual growth rate of 5.3% [2]. Today, there are 35,000 direct scheduled routes on the air travel network, with 865 new routes established in 2011 [1]. Malaria endemic areas are more connected to the rest of the world than at any time in history, with the disease able to travel at speeds of 600 miles per hour within infected passengers. The growth of the air travel network results in substantial concerns and challenges to the global health system , with a need to place more emphasis on evidence-driven surveillance and reporting that incorporates spatial and network information [7,8,11,12].

Rising rates of travel between malaria free and endemic countries have led to general patterns of increased rates of imported malaria over recent decades

[15,26,28,120]. Due to infrequent encounters [7,26], imported cases can challenge health systems in non-endemic countries, with difficulties in diagnosis [120], misdiagnosis and delays in treatment [121,122], as well as significant treatment expenses [123]. Further, flights may bring infected vectors, resulting in “airport malaria”, where patients who do not have a foreign travel history become infected through being bitten in the vicinity of international airports [10,124–126]. Patterns in imported cases and airport malaria have been shown to be related to a combination of the numbers of travellers and the malaria risk at the destination [7,10], and these relationships will continue to evolve as new routes become established.

The flow of people via air travel between endemic areas may increase the risks of re-emergence or resurgence [20] in previously malaria free or low transmission areas [31]. The autochthonous malaria outbreaks in Virginia in 2002 [17], Florida in 2003 [16] and Greece in 2011 [18], for example, demonstrate the continued risks of local outbreaks following reintroduction through air travel, though such occurrences are rare [42]. Further, the examples of malaria resurgence in island nations such as Sri Lanka [127], Mauritius[128] and Madagascar [129] after control measures were relaxed reinforce the importance of vigilance and robust surveillance in terms of human movement in pre and post-elimination periods [20]. Identifying the risks of malaria movement through the air travel network can provide an evidence base through which public health practitioners and strategic planners can be informed about potential malaria influxes and their origins [7,130].

Meanwhile, growing concerns have been raised about the possible spread of artemisinin resistance from the Greater Mekong sub-region in South East Asia to other

endemic areas. Recent research has highlighted increasing numbers of patients showing slow parasite clearance rates following treatment with artemisinin-based drugs in the Cambodia-Thailand border and Thailand-Myanmar border regions [22,131,132]. Tremendous health and socio-economic costs occurred when chloroquine-resistant parasites arrived in sub-Saharan Africa from Southeast Asia and spread across the continent [133,134]. Similarly, sulfadoxine and pyrimethamine resistance emerged in Asia and spread to Africa [135,136]. The WHO reports that there is already "at least one study with a high treatment failure rate (≥ 10%) reported from six of the 23 African countries that have adopted artesunate-amodiaquine compound" [137], and fear remains over the spread of artemisinin resistance from Southeast Asia to Africa, that could undermine current control and elimination efforts, with no alternative drugs coming in the foreseeable future.

Rates of imported malaria, risks of resurgence and the spread of drug resistance are all today influenced by how the global air travel network connects up the malaria endemic regions of the world, and the numbers of passengers moving along it. Here we combine recently constructed global *P. falciparum* and *P.vivax* malaria prevalence maps with data on modelled passenger flows across the air network, to describe and quantify global malaria connectivity through air travel in 2010. We derive weighted network analysis statistics to examine (i) which regions show greatest connectivity to *P. falciparum* and *P.vivax* malaria endemic zones, (ii) where the largest estimated passenger flows from endemic areas occur, (iii) which regions form 'communities', whereby malaria infection flows within them are likely to be larger than between communities, and finally, (iv) we examine the connectivity within and to the Southeast

Asia region where the threat of artemisinin resistance is highest, to explore risk routes for the spread of resistance.

Methods

Airport Locations, Flight Routes and Passenger Flow Matrix

Information on the longitude, latitude, city name and airport code for a total of 1,449 airports which serve cities with more than 100,000 people, and a modelled actual traffic flow connectivity list with 644,406 routes amongst these airports were obtained (http://www.vbd-air.com/data) [8,138]. A connectivity matrix was created from the connectivity list, quantifying the volumes and the directionalities of the passenger flows. Within this passenger flow matrix, 23,785 routes were direct connections between two airports, 291,745 routes were one-stop connections and 328,876 routes were two-stop connections. The travel volumes on the routes were modelled based primarily on publicly available datasets under a generalized linear model framework. Full model details are provided in Huang et al [138], but in brief, to construct the matrix, topological characteristics of the air travel network, city population, and local area GDP, amongst others, were utilized as covariates. Actual travel volumes for training and validation were extracted and assembled from various transportation organizations in the United States, Canada and the European Union. A log linear model controlling for random effects on origin, destination and the airport hierarchy was then built to predict passenger flows on the network. The model outperformed existing air travel passenger flow models in terms of prediction accuracy [138].

Malaria Distribution

Global *P. falciparum* and *P. vivax* prevalence maps were obtained from the Malaria Atlas Project (www.map.ox.ac.uk) and the methods behind their construction

are presented in Gething *et al* [32,33]. In brief, 22,212 community prevalence surveys were used in combination with model-based geostatistical methods to map the prevalence of *P. falciparum* globally in 2010 within limits of transmission defined by annual parasite incidence and satellite covariate data. Similarly, 9,970 geocoded *P. vivax* parasite rate (*P.vivax* PR) surveys collected between 1985 and 2010 were utilized in a spatiotemporal Bayesian model-based geostatistical approach to map endemicity [36], under the restrictions of a mask of the stable/unstable endemicity [33] and information on the prevalence of the Duffy blood group[139] . We do not consider *P. ovale, P.malariae or P.knowlesi* here, since similar datasets on their distributions do not yet exist.

Weighted Network Analysis and Community Detection

Malaria prevalence can vary greatly in the region around airports and the cities they serve, and travellers taking flights from a specific airport may reside many kilometres from the airports in higher transmission areas than found in the vicinity of the airport. Thus, simply assigning the predicted prevalence from the malaria maps at the location of each airport could underestimate the risk and rate of infection exportation at the airport in question and under-represent its contribution to global malaria connectivity. Therefore, following Huang et al [8], local accessibility to each airport was considered by assuming that passengers would travel less than 50 km with a travel time less than two hours to access an airport to take a flight. Under this assumption, the *P. falciparum* and the *P. vivax* malaria prevalences assigned to an airport were obtained as the maximum prevalence from the malaria maps within a mask of 50 kilometres and 2 hour travel time (Figure C-1), in which the mask was generated using a global travel time map (http://bioval.jrc.ec.europa.eu/products/gam/index.htm). Likewise, an indicator that

defines whether an airport is located in the stable/unstable endemic zone was created according to the same mask, in which the indicator defines whether the majority area of the mask is located in the stable/unstable zone. Figure C-1 in Appendix C shows this travel time/distance mask with the global travel time map.

The above approach ensured that each airport had an assignment of a *P. falciparum* and *P.vivax* prevalence rate (or unstable/malaria free), which could then be used as a weighting applied to the passenger flow estimates to derive relative 'malaria flow' indices for each route that could be compared to other routes across the global network to analyse malaria connectivity. Thus, *P. falciparum* and *P. vivax* flows were calculated on each route as origin prevalence * estimated passenger volume, to produce *P. falciparum* and *P. vivax* malaria networks.

A group of weighted centrality analyses and network community partition analyses were performed on the malaria networks to quantify features of global malaria connectivity. First, the in-strength and out-strength of each connection was calculated as

$$s_i = \sum_{j=1}^{N} a_{ij} w_{ij} \quad (4\text{-}1)$$

In which a_{ij} is the airport adjacency matrix and w_{ij} is the weighted malaria flow. This metric estimates the total weight of malaria flows that airports send and receive.

Following this, weighted "betweenness" analyses were performed on the malaria flow matrices. Betweenness centrality measures the number of shortest paths going through a specific vertex [140]. In a weighted network, betweenness centrality is a useful local measure of the load placed on the given node in the network as well as the node's importance to the network other than just connectivity [106]. It is often used in

transport network analysis to provide an approximation of the traffic handled by the vertices [95]. Thus, here it provides an indication of the importance of each airport in the global flow of malaria infections via air travel – i.e. a measure of how many infections likely pass through each airport each year, relative to other airports. The betweenness centrality is calculated as:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\delta_{st}(v)}{\delta_{st}} \qquad (4\text{-}2)$$

In which δ_{st} is the total number of shortest paths from node s to node t and $\delta_{st}(v)$ is the number of those paths that pass through v. Note that on our weighted *P.falciparum/P.vivax* networks, the distance between the two nodes s and t is defined by the sum of *P.falciparum* flows or *P.vivax* flows as the cost on this path [141]. A normalized betweenness was used as

$$C_{Bnorm}(v) = \frac{C_B(v)}{\sum C_B(v)/n} \qquad (4\text{-}3)$$

where n was the number of nodes (airports) in the air travel network[95].

Communities in a network reflect the partition of nodes that are densely connected and separated from the other nodes in the network, thus these nodes "probably share common properties and/or play similar roles within the graph" [142]. By mapping communities on the malaria networks defined here, we aimed to identify groups of airports that show strong links in terms of likely movements of infections. This potentially has utility in terms of providing evidence upon which regional surveillance strategies can be designed [42]. Newman and Girvan [105] define a modularity score which measures the quality of network partitions as:

$$Q = \frac{1}{2m}\sum_{i,j}[W_{ij} - \frac{k_i k_j}{2m}]\delta(c_i, c_j) \qquad (4\text{-}4)$$

In which, W_{ij} represents the weight of the edge between i and j (here these are the *P.falciparum* and *P.vivax* flow matrices), $k_i = \sum_j W_{ij}$ is the sum of the weights of the connections attached to airport i, c_i is the community to which airport i is assigned; $\delta(c_i, c_j)$ is 1 if c_i = c_j, otherwise $m = \frac{1}{2}\sum W_{ij}$.

A multilevel algorithm for community detection [145] was implemented. This method utilizes an iterative approach that merges communities to maximize the modularity score: Firstly modularity is optimized by allowing only local changes of communities, secondly the established communities are combined together to construct a new network. These two passes are repeated iteratively until no increase of modularity is possible. The number of communities returned by this algorithm yields the maximum modularity score.

We performed a simple Wilcoxon rank-sum test [146] on the differences between "internal" and "external" degrees of a community in order to test whether the establishment of communities was significant. We defined air connections within a community as "internal" and the connections connecting the airports of a community with the rest of the network as "external". The null hypothesis of this test was that there was no difference between the number of internal and external routes incident to an airport of the community.

Results

The results of the global malaria connectivity analyses are presented in two sections: (i) analyses focussed on the connection of endemic malaria regions to each

other and to malaria-free areas, that has particular relevance to imported malaria and malaria resurgence and re-emergence; and (ii) analyses examining the connections between Southeast Asia and the rest of the malaria endemic world, which are relevant to the spread of artemisinin resistance.

Connectivity within Endemic Areas and to Non-Endemic Areas

Figure 4-1 shows the results of regional community structure analyses based on traffic flow data overlaid on the *P. falciparum/P. vivax* endemicity and stable/unstable transmission limits maps. The Wilcoxon test results show that the internal degrees for the airports within all communities are significantly different from the external degrees, with p values of < 0.01, thus the community partitions shown are significant. The maps highlight those countries that form communities linked by high levels of traffic scaled by *P. falciparum/P. vivax* prevalence at their origin endemic area. Figure C-2 describes similar analyses based on the travel network data from Huang et al [29]. The communities detected reflect the architecture of the air network, and how this relates to malaria endemicity around the World. Geographical contiguity is clearly evident, as traffic levels on shorter distance routes are generally higher than on longer distance routes, but interesting patterns relating to historical ties emerge. For instance, for *P.falciparum*, London forms part of the Nigeria community, but Paris shows stronger ties to the remainder of sub-Saharan Africa. These connections are often reflected in imported malaria statistics, with Nigeria being the main source of *P.falciparum* cases seen in the UK, but for France, the French-speaking African countries are the main origin. Similarly, UK airports also form part of the India/Bangladesh community, where historical ties exist, resulting in significant travel between the two regions, and consequent *P.falciparum* and *P. vivax* malaria importation to the UK. Ties also exist

between the western US and East Asia, which form a single *P.falciparum* community (Figure 4-2a). Figure C-3 shows a community detection analysis for airports with direct-connections, one-transfer connections and two-transfer connections from endemic areas.

To examine directional and net potential movements of people and parasites between airports in different countries, we summed up the international route weightings to identify possible "source" and "sink" airports of malaria infections (Table 4-1 and Table 4-2). Here, the weights of all possible incoming flows for airports in the non-endemic areas, and the weights of all possible outgoing flows from airports in endemic areas were summed up to define "vertex strengths" of importation and exportation (Note that we only considered the routes connecting two different countries regardless of the domestic routes). In this table, airports in the Far East and Middle Asia such as Singapore, Hong Kong, Dubai, and Sharjah display the highest importation values (Note that Singapore ranked the first in both categories). Unsurprisingly, major air hubs in Europe (such as airports in London, Paris and Frankfurt) also showed high potential incoming *P.falciparum* flows. Miami is the only airport in the United States on the importation flow top 10 lists, with its strong connections to Central and South America. In terms of exportation, the largest airports by traffic capacity and connections to the rest of the malaria endemic world were highlighted (Table 4-3 and Table 4-4). Mumbai was ranked first as the largest exporter of *P.falciparum* and *P.vivax* flows, suggesting that it likely acts as an important portal for spreading malaria to the rest of the world.

The betweenness centrality metric was utilized to inspect the connectivity from endemic areas. As the betweenness metric $C_B(v)$ is defined as the number of shortest

paths connecting any two airports that involve a transfer at airport v, high centrality airports in endemic areas provide hubs for people originating at less-accessible airports in remote places to reach the rest of the world. Table 4-3 and Table 4-4 show the top 10 highest betweenness centrality airports for transferring *P.falciparum* flow and *P.vivax* flow elsewhere. For the *P.falciparum* flow, international airports in Africa play important roles as hubs for routing infections. Some airports are observed to have small degrees (low numbers of connecting routes) and large centrality (importance as a hub), which can be considered as an abnormality [95]. These airports connect less accessible and connected airports in endemic areas to other airports in the world. For *P.vivax* flow, Asian international hubs play more important roles. Of interest is Phoenix airport, which ranked the 6th in terms of *P.vivax* centrality, suggesting that it plays an important role as a gateway in linking *P.vivax* endemic areas to the United States. Figure C-3 presents the spatial distribution of betweenness centrality scores for airports, weighted by *P.falciparum* or *P.vivax* flows. As a comparison, a similar figure based solely on the modelled passenger flow is presented in Figure C-4.

To further investigate the effects of flows from endemic zones, Figure C-5 a) and b) shows the sums of international incoming risk flows for all the airports in those 36 countries that have national policies for malaria elimination, and are closest to eliminating the disease [147]. Importation of infections threatens the success of elimination programs [31] and while air travel may not be the highest risk source for these introductions for most of these countries, it remains a potentially important source of incoming infections. From these two maps, it can be seen that China and countries in middle Asia are subjected to the greatest pressure of incoming flows, relative to other

elimination countries, due to their larger incoming traffic volumes from endemic regions elsewhere around the world. Further analyses on airport connectivity are provided in Appendix C.

Connectivity to Southeast Asia

Figure 4-2 maps out the passenger flows scaled by origin prevalence for *P.falciparum* and *P.vivax* from the greater Mekong sub-region. Significant amounts of flow exchange within Southeast Asia can be seen in the close-up subsets. For both *P.falciparum* and *P.vivax* it can be seen that the connectivity, through numbers of travellers, to Latin American endemic regions is weak, but that much stronger connections to sub-Saharan Africa and the Indian subcontinent exist. Increasing connections through trade and labour markets between Asia and Africa over the past decade is exemplified here in the strong connections between the Southeast Asian region and all of sub-Saharan Africa's major airport hubs. Table C-1 presents the top 10 risk routes spreading drug resistance of *P.falciparum* and *P.vivax* from the Greater Mekong Sub-region to non-Asian destinations, with estimated *P.falciparum* / *P.vivax* flow and the number of stops needed to travel from the origin city to the destination city shown.

Discussion

The continuing growth in air travel is playing an important role in the global epidemiology of malaria. Flight routes now connect previously isolated malaria endemic regions to the rest of the world, and travelers on these routes can carry infections to the opposite side of the world in less than 24 hours. While many endemic areas still remain relatively isolated, the malaria endemic world is becoming increasingly connected to both malaria-free areas and other endemic regions. The impacts of this can be seen in

imported cases, vector invasions and the spread of drug-resistant parasite strains. Here we present a spatial network analysis approach to demonstrate the connectivity that exists across the malaria endemic world through air travel, and provide quantitative indicators of the risks it results in for malaria movement.

Results highlight the substantial connectivity that now exists between and from malaria endemic regions through air travel. While the air network provides connections to previously isolated malarious regions, it is clear that great variations exist, with significant regional communities of airports connected by high rates of prevalence-scaled flow standing out (Figure 4-1). The structures of these networks are often not geographically coherent, with historical, economic and cultural ties evident. As new routes continue to be established, these communities will likely change, with new popular travel routes, such as those between China and Africa [7] likely altering global malaria flow routes, and new airports appearing in tables 1 and 2. These community maps (Figure 4-1) and lists of cities by likely import/export of infections (Table 4-1 and Table 4-2) and hubs for infection flow (Table 4-3 and Table 4-4) provide a quantitative picture of how malaria infections are likely moving globally through air travel, and information from which global surveillance strategy design can draw upon. Tables 4-1 to 4-4 highlight that certain airports provide significant hubs and gateways for the movement of infections and their entry into countries, and that these are widely distributed across the world. Their role in providing important nodes as both significant through-flow of infections in the network, and entry and exit gateways for cases to/from regions means that they potentially represent valuable sentinel sites for focused surveillance. Finally, Figure 4-2 provides a stark reminder of how well connected the

malaria endemic areas of Africa are now to Southeast Asia, illustrating the many possible routes that artemisinin resistant strains could take. These routes can provide a first-step quantification to support the global plan against artemisinin resistance containment [137] and design of surveillance systems [148], and should be refined with information on the locations of resistance found. Such data could also inform decisions on where and how to limit the risk of spread, for example by pre-travel or arrival screening and treatment.

A range of limitations and uncertainties exist in the analyses presented here. In terms of the quantification of malaria transmission, the use of static maps of annual average prevalence [32,33] neglects the seasonality in transmission that is common to many areas, and also the substantial changes in transmission intensity seen in a variety of locations in recent years [149]. Further, we have used parasite prevalence as our malaria metric, and while this may be an adequate measure of population prevalence at origin locations, it is not so appropriate for assessing the risk of infection acquisition for naïve travellers, and entomological-based indices are likely more appropriate here, as used in more local studies [130,144,150]. Finally, our examination of relative artemisinin resistance spread risk focuses simply on all travel from four countries, and thus does not account for any heterogeneity in resistance in the region.

Uncertainties and limitations relating to the travel data used also exist. The modelled passenger flows represent just a 2010 snapshot, and thus routes and changes since then are not captured, while inherent uncertainty due to the modelling process also exists [138]. Moreover, the types of traveller and their activities during travel and their residential location are unknown, each of which contributes to differing

malaria infection risks. Finally, overland and shipping travel flows are not considered here, which also contribute to local, regional and global malaria connectivity and flows.

This work forms the basis for future analyses on imported malaria, elimination feasibility and the risks and potential routes of artemisinin resistance spread. Rates and routes of imported malaria have been shown to be significantly related to a combination of numbers of travellers to/from endemic destinations and the prevalence of malaria there [7]. The potential thus exists to construct a model based on global malaria prevalence [32,33], transmission models for attack rate estimation [130], and traveller flow data [138], that can be used to forecast imported malaria rates, validated with imported malaria data.

As nations make progress towards elimination [147], the importance of human movement and imported cases increases. The research presented here contributes to an on-going initiative, the human mobility mapping project (www.thummp.org), aimed at better modelling human and disease mobility, and will form one aspect of continued multi-modal assessments of malaria movements [31,130,143,144] and assessment of malaria elimination strategies [41,42]. Finally, the potentially disastrous consequences of the rise and spread of artemisinin resistance requires that detailed and effective planning be implemented in preparation for containing and stemming any spread [148]. We have presented a basic assessment here of prevalence-scaled travel from the four Southeast Asian countries where resistance has previously been observed, but significant refinements of these estimates and modeling methods should be undertaken. These may include improved tracking and mapping of observed resistance and human movement patterns in Southeast Asia, as is being undertaken by the TRAC project

(http://www.wwarn.org/partnerships/projects/trac), as well as scenario modeling of the risks of resistance escape to Africa or Latin America. Further, the incorporation of accessibility [151,152] and travel data [144,150] with drug use data (e.g. [153]), prevalence information [32,33] and models [154], all undertaken within a probabilistic modeling framework (e.g. [8,74]), could aid in estimation of spread routes should resistance arise elsewhere.

Table 4-1. Top 10 airports based on estimated relative malaria importation rates

Malaria importation *P.falciparum* import			*P.vivax* import		
City	Country	*P.falciparum* Flow In	City	Country	*P.vivax* Flow In
Singapore	Singapore	366132	Singapore	Singapore	175220
Bangkok	Thailand	276563	Hong Kong	Hong Kong	105261
Paris	France	253291	Dubai	United Arab Emirates	103790
Dubai	United Arab Emirates	218599	Kuala Lumpur	Malaysia	74976
London	United Kingdom	212918	Seoul	South Korea	58278
Hong Kong	Hong Kong	183314	Sharjah	United Arab Emirates	53548
Johannesburg	South Africa	167734	London	United Kingdom	48571
Casablanca	Morocco	149298	Miami	United States	47277
Kuala Lumpur	Malaysia	141161	Taipei	Taiwan	42840
Frankfurt	Germany	132597	Mexico City	Mexico	41710

Note: *P.falciparum* / *P.vivax* flow measures are calculated based on the incoming and outgoing numbers of passengers travelling internationally, scaled by the malaria prevalence at the origin of the routes in the case of importation, and at the airport listed in the case of exportation. The flows represent a relative measure of infection movement and are not designed to represent actual number of infections

Table 4-2. Top 10 airports based on estimated relative malaria exportation rates

Malaria exportation *P.falciparum* export			*P.vivax* export		
City	Country	*P.falciparum* Flow out	City	Country	*P.vivax* Flow Out
Mumbai	India	974832	Mumbai	India	214006
Ouagadougou	Burkina Faso	513636	Bangkok	Thailand	179633
Kinshasa	Congo	467295	Manila	Philippines	177760
Abuja	Nigeria	461492	Delhi	India	161249
Delhi	India	447491	Ho Chi Minh City	Vietnam	80725
Bamako	Mali	439138	Bogota	Colombia	72173
Douala	Cameroon	382278	Ahmedabad	India	70888
Manila	Philippines	377384	Panama City	Panama	70568
Lome	Togo	267054	Guatemala City	Guatemala	54788
Cotonou	Benin	248226	Phnom Penh	Cambodia	53898

Note: *P.falciparum* / *P.vivax* flow measures are calculated based on the incoming and outgoing numbers of passengers travelling internationally, scaled by the malaria prevalence at the origin of the routes in the case of importation, and at the airport listed in the case of exportation. The flows represent a relative measure of infection movement and are not designed to represent actual number of infections

Table 4-3. Top 10 *P.falciparum* betweenness centrality airports with their degrees in a sub-network that only contains direct links from airports in *P.falciparum* or *P.vivax* endemic areas.

Airport	City	Country	Normalized Betweenness Centrality	Degree
NBO	Nairobi	Kenya	47.35	80
MBA	Mombasa	Kenya	32.44	27
JRO	Kilimanjaro	Tanzania	32.39	14
BOM	Mumbai	India	30.41	104
ADD	Addis Ababa	Ethiopia	28.21	64
DEL	Delhi	India	23.16	111
JIB	Djibouti	Djibouti	19.77	15
ADE	Aden	Yemen	18.63	15
MGQ	Mogadishu	Somalia	14.45	8
HRE	Harare	Zimbabwe	14.35	20

Table 4-4. Top 10 *P.vivax* betweenness centrality airports with their degrees in a sub-network that only contains direct links from airports in *P.falciparum* or *P.vivax* endemic areas.

Airport	City	Country	Normalized Betweenness Centrality	Degree
BKK	Bangkok	Thailand	96.43	146
ICN	Seoul	South Korea	78.12	150
DEL	Delhi	India	59.55	133
BOM	Mumbai	India	34.17	116
KMG	Kunming	China	30.79	90
PHX	Phoenix	United States	28.63	91
DPS	Denpasar Bali	Indonesia	27.94	34
SJO	San Jose	Costa Rica	27.72	37
DOH	Doha	Qatar	25.91	100
TAS	Tashkent	Uzbekistan	25.85	69

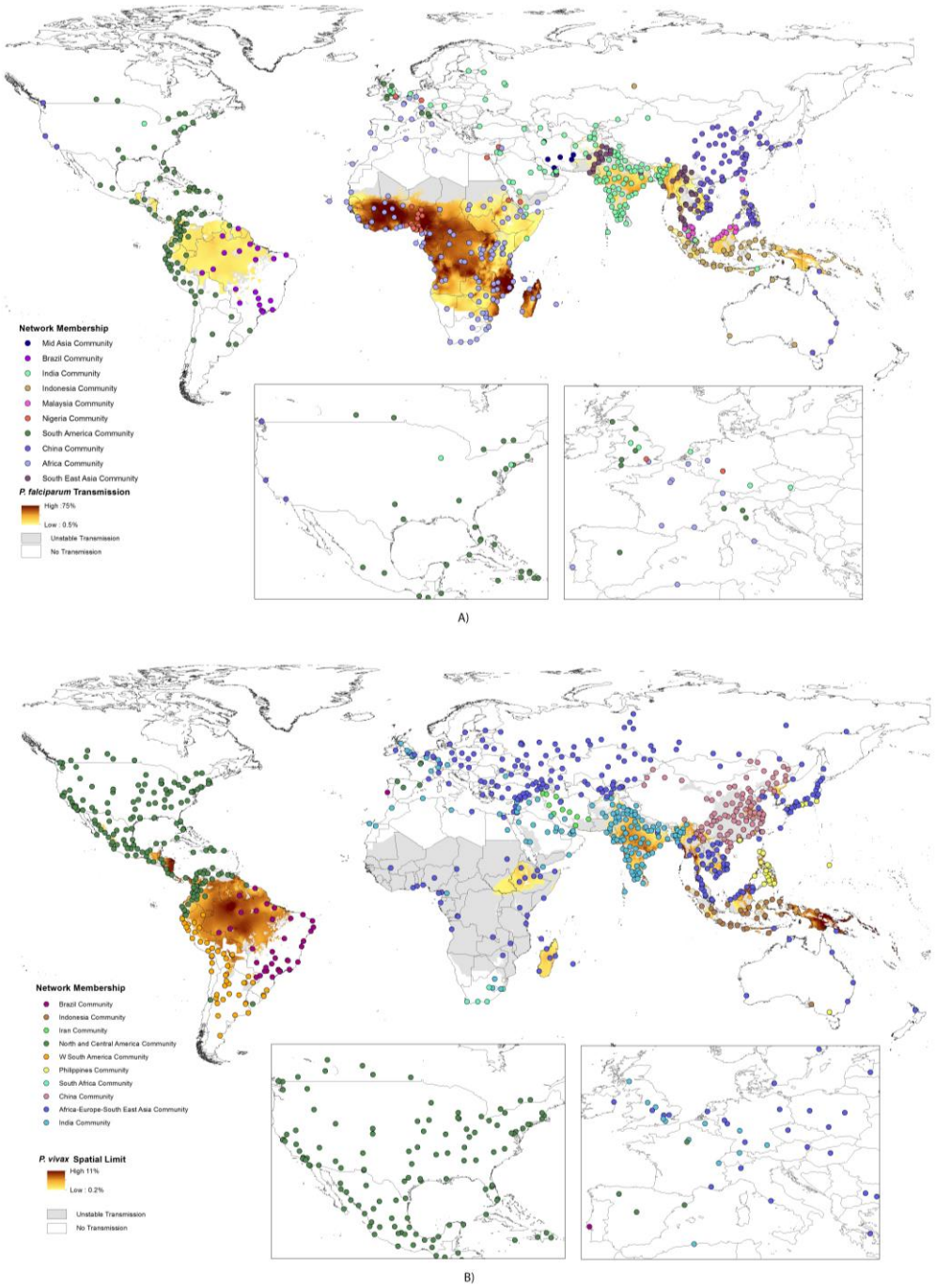


Figure 4-1. Spatial distribution of *P.falciparum/P.vivax* network communities overlaid on *P.falciparum/P.vivax* prevalence maps. A) *P.falciparum* multilevel membership. B) *P.vivax* multilevel membership. These two maps show only airports that have direct connections from endemic to non-endemic areas. The inset maps present close-up views of the United States and Western Europe. Airports with the same community membership (indicated by the same color) display stronger links in terms of likely movements of infections between them than to airports in other communities. Note that in the *P.vivax* map, communities with less than 10 airports are not shown.

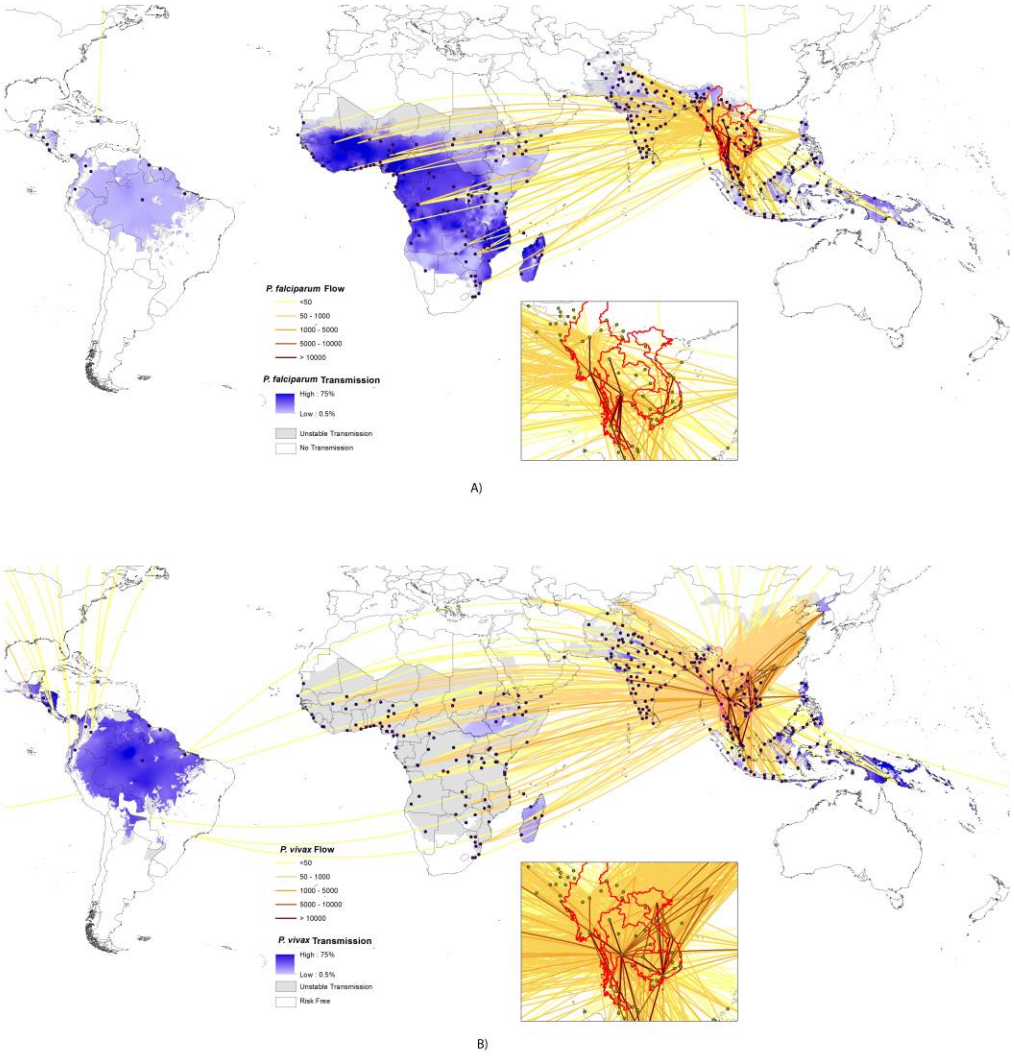


Figure 4-2. Estimated relative *P.falciparum/P.vivax* flows originating from the Great Mekong sub-region overlaid on *P.falciparum/P.vivax* prevalence maps. A) *P.falciparum* flows originating from the Great Mekong sub-region. B) *P.vivax* flows originating from the Great Mekong sub-region. The flows include estimated passenger numbers, including direct, one-transfer and two-transfer flight routes. The inset maps show close-up views for airports in the Great Mekong sub-region.

CHAPTER 5
CONCLUSION

The main purpose of this study is to examine how the interrelationships among the global distribution of vector-borne infectious diseases, locations of known outbreaks, and international air service routes interact to result in seasonally changing risks of insect-borne infectious disease transmission and spread by air travel. It presents initial efforts to integrate disparate datasets (disease/vector distribution, air travel network, and climate data etc.), to create a unified weighted network framework for vector borne diseases with spatial attributes. Up until now there has been no significant research on the complex network analysis of air travel network changes, and how it connects endemic vector-borne disease regions. This research lays out analyses and data to facilitate a better understanding of the evolution of the risk of spread of vector borne diseases on the air travel network, coupled with the changes in the topographical structure of the network, climate, and the ecological niche of vector borne diseases.

The research presented here has documented the generation of a world-wide Origin-Destination matrix of passenger flows in 2010 for airports with host city populations of more than 100,000. The datasets are freely accessible for academic use and are published as part of the Vector-Borne Disease Airline Importation Risk (VBD-Air) project at www.vbd-air.com/data/ . Researchers can utilize this open database for further analyses on disease transmission on the air travel network.

Moreover, the analyses demonstrate the substantial connectivity that now exists between and from malaria endemic regions through air travel. While the air network provides connections to previously isolated malarious regions, it is clear that great variations exist, with significant regional communities of airports connected by higher

rates of flow standing out. The structures of these communities are often not geographically coherent, with historical, economic and cultural ties evident, and variations between *P.falciparum* and *P.vivax* clear. Moreover, results highlight how well connected the malaria endemic areas of Africa are now to Southeast Asia, illustrating the many possible routes that artemisinin resistant strains could take.

The expansion of the air travel network is so far never anticipated to end, while our understanding of the role of air travel in the transmission of vector-borne diseases are still limited [7]. We must prepare for this challenge by modeling and quantifying the dynamics of human-disease interactions on the complex air travel network. This dissertation presents the first step towards better decision support for vector-borne disease control, management, and ultimately eradication. To reduce the global burden of vector-borne diseases and improve human health and wellbeing, we still have a long way to go.

APPENDIX A
SUPPLEMENTARY FIGURES FOR CHAPTER 2

This document shows the supplementary figures used in chapter 2.

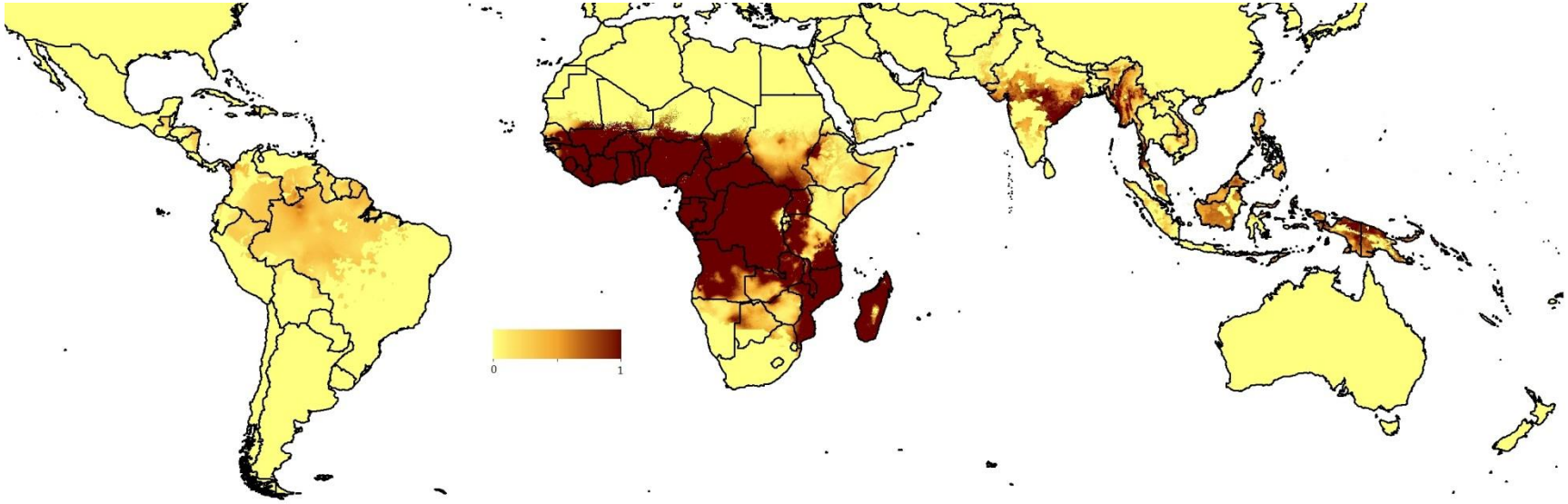


Figure A-1. *P. falciparum* prevalence: *Plasmodium falciparum* is a protozoan parasite, one of the species of Plasmodium that cause malaria in humans. It is transmitted by the female Anopheles mosquito. *P.falciparum* is the most dangerous of these infections as *P.falciparum* (or malignant) malaria has the highest rates of complications and mortality. As of 2006 it accounted for 91% of all 250 million human malarial infections (98% in Africa) and 90% of the deaths. It is more prevalent in sub-Saharan Africa than in other regions of the world; in most African countries, more than 75% of cases are due to *P.falciparum*, whereas in most other countries with malaria transmission, other Plasmodia species predominate. Non-endemic countries often see many cases of imported *P.falciparum* malaria each year through travelers or returning migrants. The Malaria Atlas Project (www.map.ox.ac.uk) recently produced a global map of *P. falciparum* malaria endemicity in 2010, based on over 24,000 community prevalence surveys. The methods used behind construction of this dataset are described in more detail here: http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.1000048

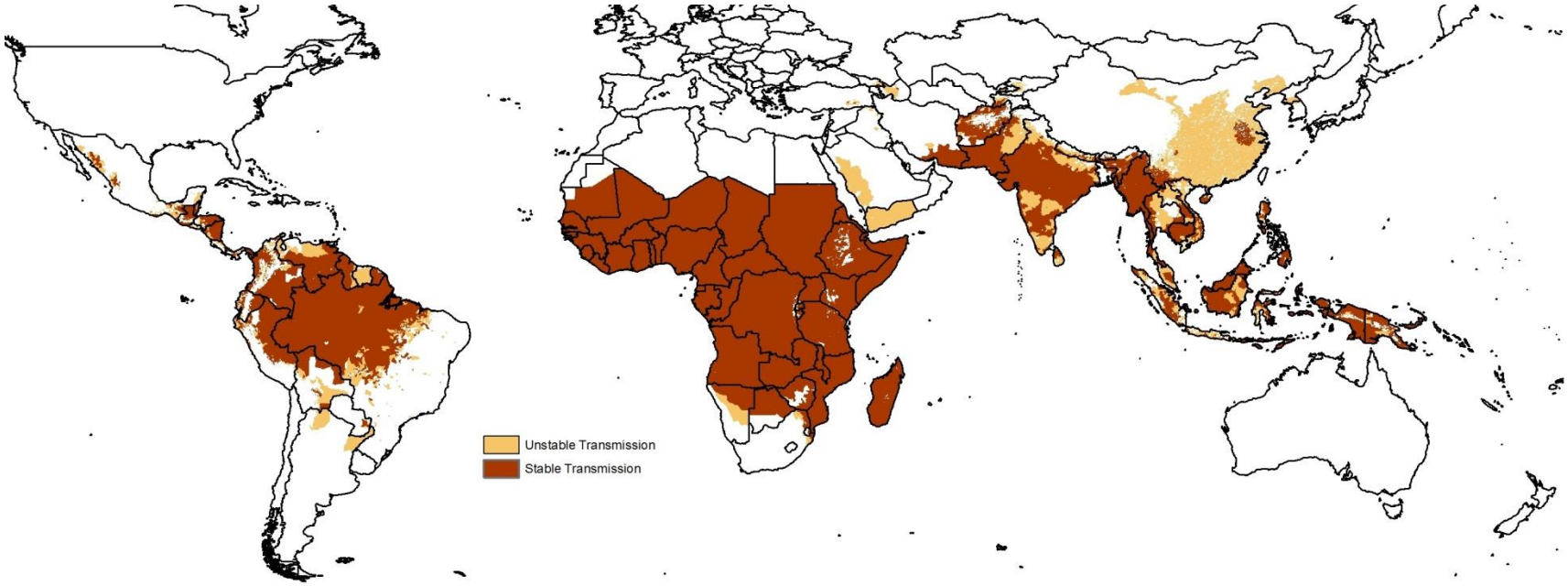


Figure A-2. *P. vivax* endemic areas: *P. vivax* is the most frequent and widely distributed cause of recurring (tertian) malaria, *P. vivax* is one of the four species of malarial parasite that commonly infect humans. It is less virulent than P. falciparum, which is the deadliest of the four, and is seldom fatal. P. vivax is carried by the female Anopheles mosquito. The Malaria Atlas Project (www.map.ox.ac.uk) recently produced a global map of the limits of *P. vivax* malaria transmission in 2009, based on thousands of reports of cases. The methods used behind construction of this dataset are described in more detail here: www.plosntds.org/article/info:doi/10.1371/journal.pntd.0000774

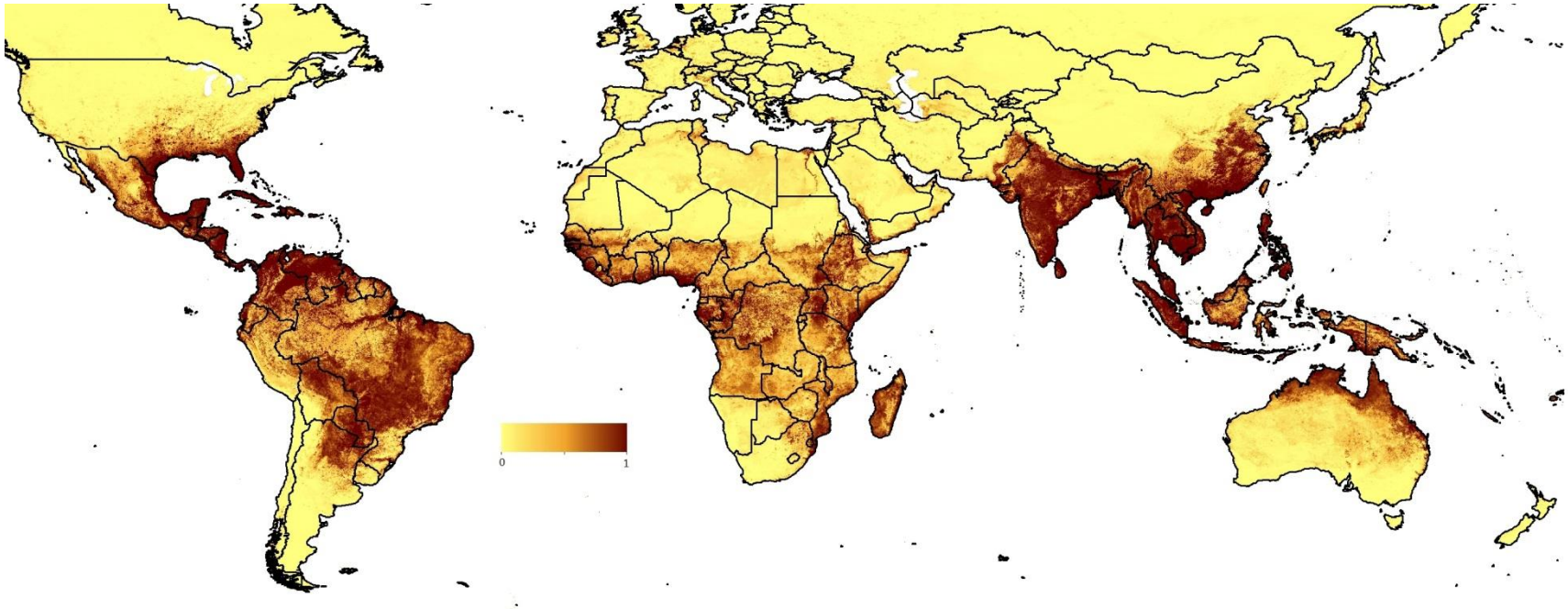


Figure A-3. Dengue suitability map. Dengue suitable areas (niche model): a map showing the predicted suitability for dengue transmission based on thousands of reports of dengue cases. Dengue fever, also known as “breakbone” fever, is an infectious tropical disease caused by the dengue virus. Dengue is transmitted by several species of mosquito within the genus *Aedes*, principally *Ae. aegypti*, but evidence also exists for *Ae. Albopictus* transmission. The virus has four different types infection with one type usually gives lifelong immunity to that type, but only short-term immunity to the others. Subsequent infection with a different type increases the risk of severe complications. The incidence of dengue fever has increased dramatically since the 1960s, with around 50–100 million people infected yearly, and imported cases through air travel to non-endemic regions are on the rise.

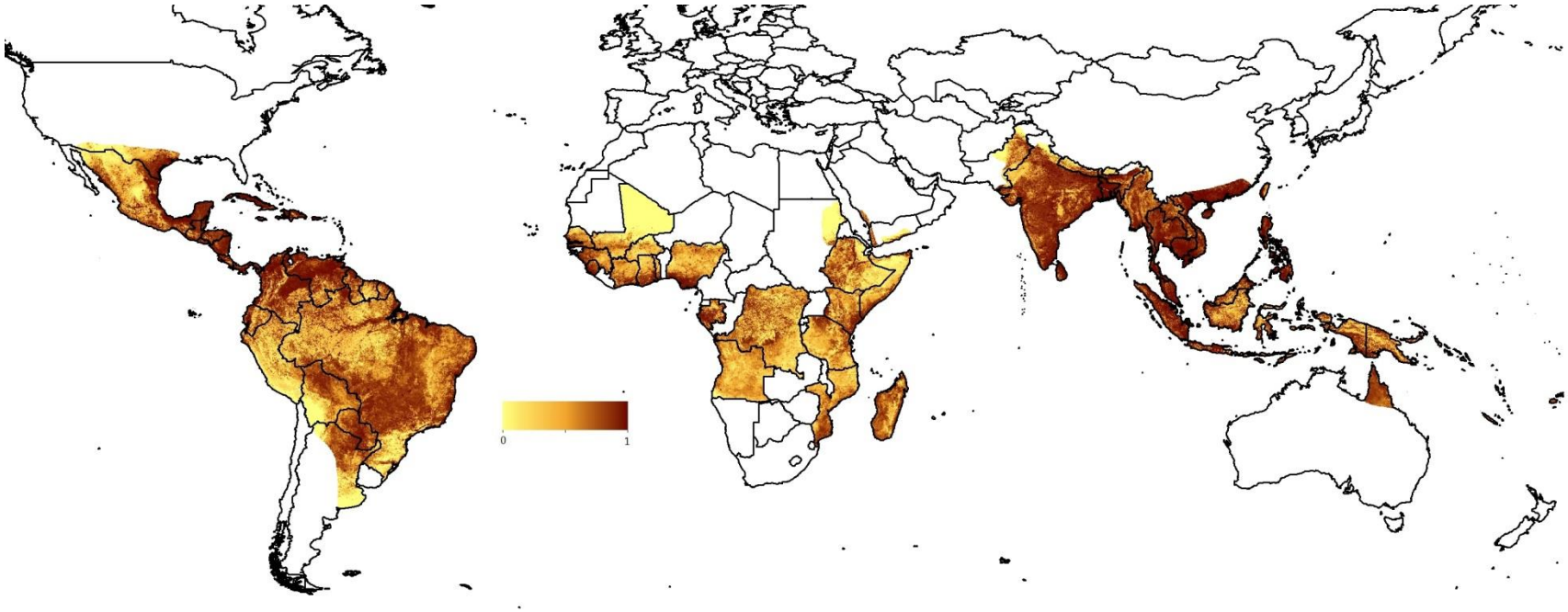


Figure A-4. Dengue suitability in areas of known outbreaks in 2010. Dengue endemic areas (Yellow book): a map showing the predicted suitability for dengue transmission based on thousands of reports of dengue cases. This map is a refinement of the dengue suitability map in focusing solely on regions of recent confirmed transmission, as it was masked so that only regions cited by the CDC as having transmission in 2010 were left.

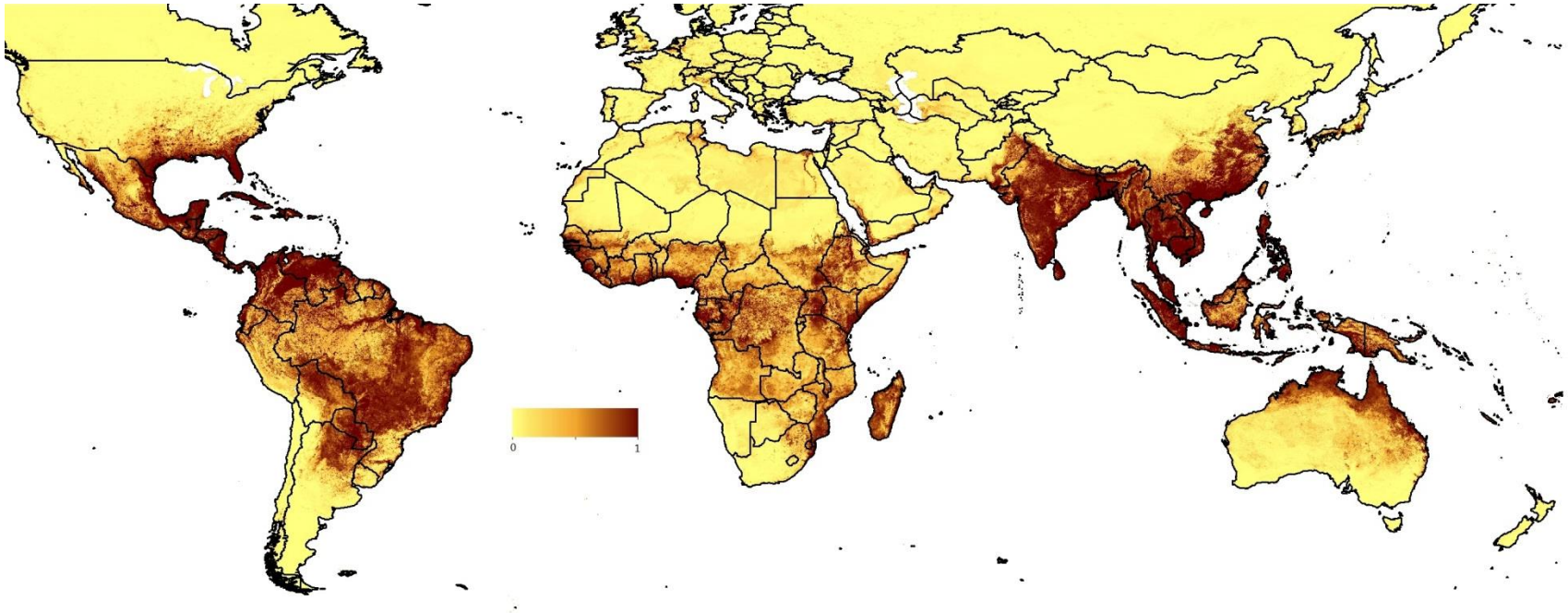


Figure A-5. Dengue outbreak prone. Dengue outbreak areas (niche model): a map showing the predicted suitability for significant dengue outbreaks based on data on outbreaks since 2008 reported on Healthmap.

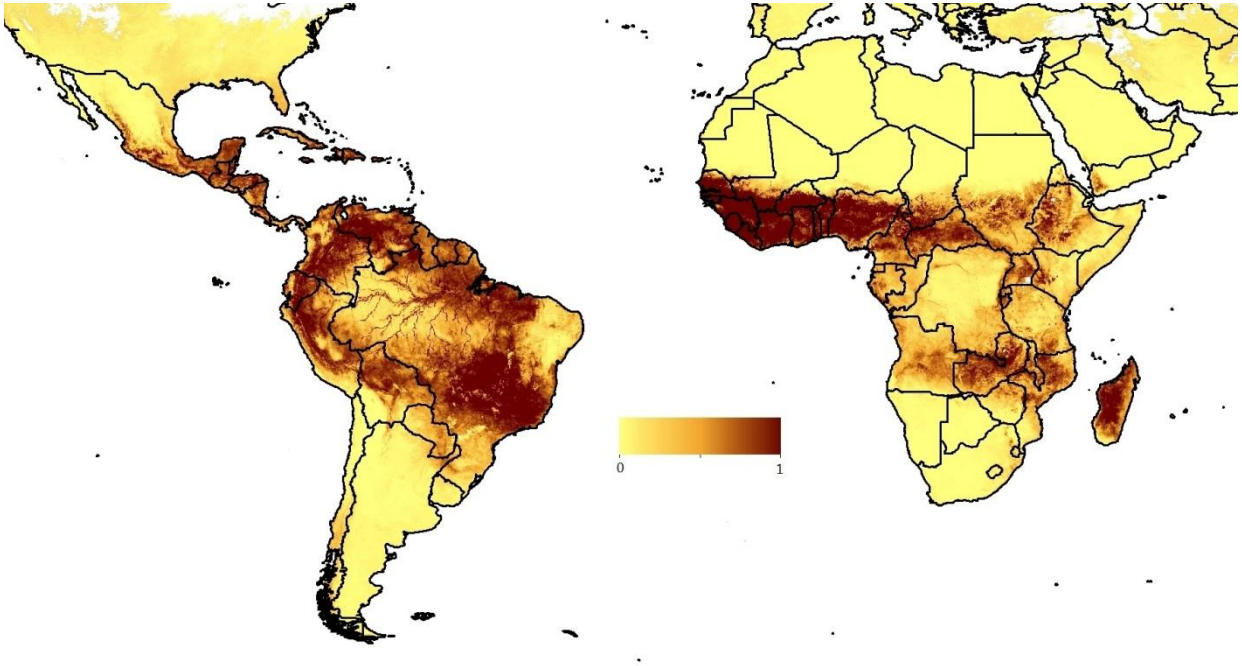


Figure A-6. Yellow fever suitability. Yellow fever suitable areas (niche model): a map showing the predicted suitability for yellow fever transmission based on hundreds of reports of yellow fever cases. Yellow fever is an acute viral hemorrhagic disease. The yellow fever virus is transmitted by the bite of female mosquitoes (the yellow fever mosquito, *Aedes aegypti*, and other species) and is found in tropical and subtropical areas in South America and Africa, but not in Asia. The WHO estimates that yellow fever causes 200,000 illnesses and 30,000 deaths every year in unvaccinated populations; around 90% of the infections occur in Africa. A safe and effective vaccine against yellow fever has existed since the middle of the 20th century and some countries require vaccinations for travelers. Since no therapy is known, vaccination programs are, along with measures to reduce the population of the transmitting mosquito, of great importance in affected areas. Since the 1980s, the number of cases of yellow fever has been increasing, making it a reemerging disease, and case numbers imported through air travel to non-endemic areas have been rising.

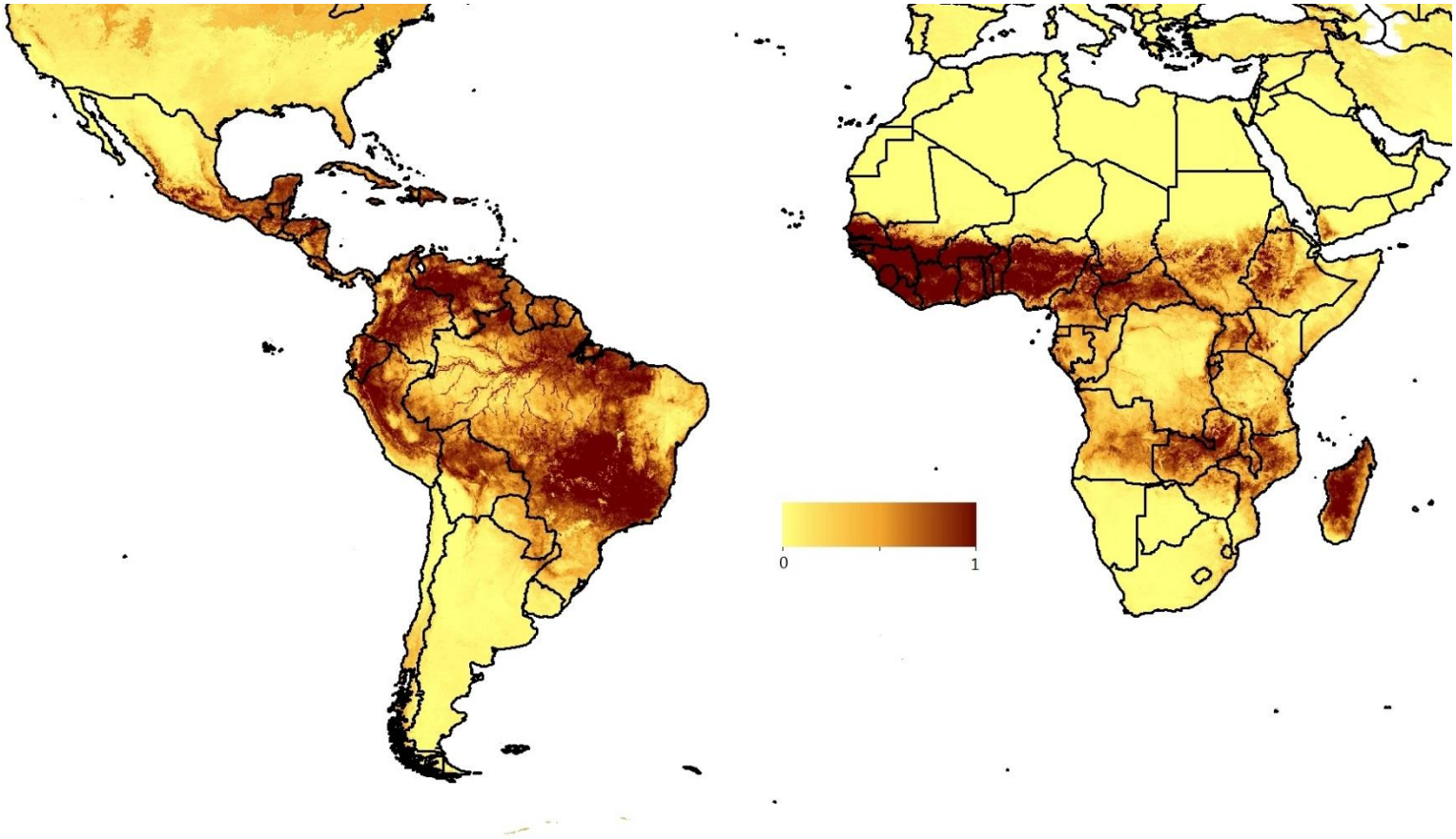


Figure A-7. Yellow fever outbreak prone. Yellow fever outbreak areas (niche model): Yellow fever outbreak prone: a map showing the predicted suitability for significant yellow fever outbreaks based on data on outbreaks since 2008 reported on HealthMap.

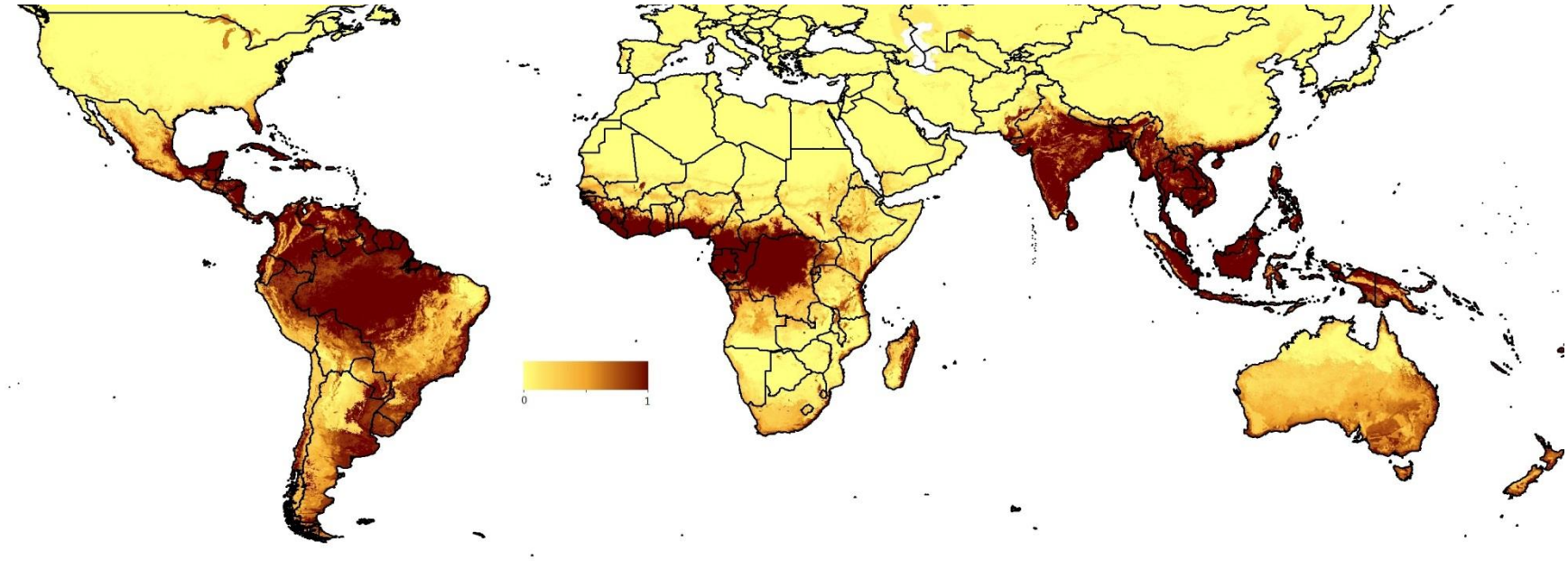


Figure A-8. Chikungunya outbreak prone. Chikungunya outbreak areas (niche model): a map showing the predicted suitability for significant chikungunya outbreaks based on data on outbreaks since 2008 reported on Healthmap. Chikungunya virus (CHIKV) is an insect-borne virus, of the genus Alphavirus, that is transmitted to humans by virus-carrying *Aedes* mosquitoes, principally *Ae.aegypti* and *Ae.albopictus*. There have been recent breakouts of CHIKV associated with severe illness. CHIKV causes an illness with symptoms similar to dengue fever. Large sporadic outbreaks have occurred since around 2005, with spread associated with both the movement of infected air travelers and the spread of the range of *Ae. albopictus*.

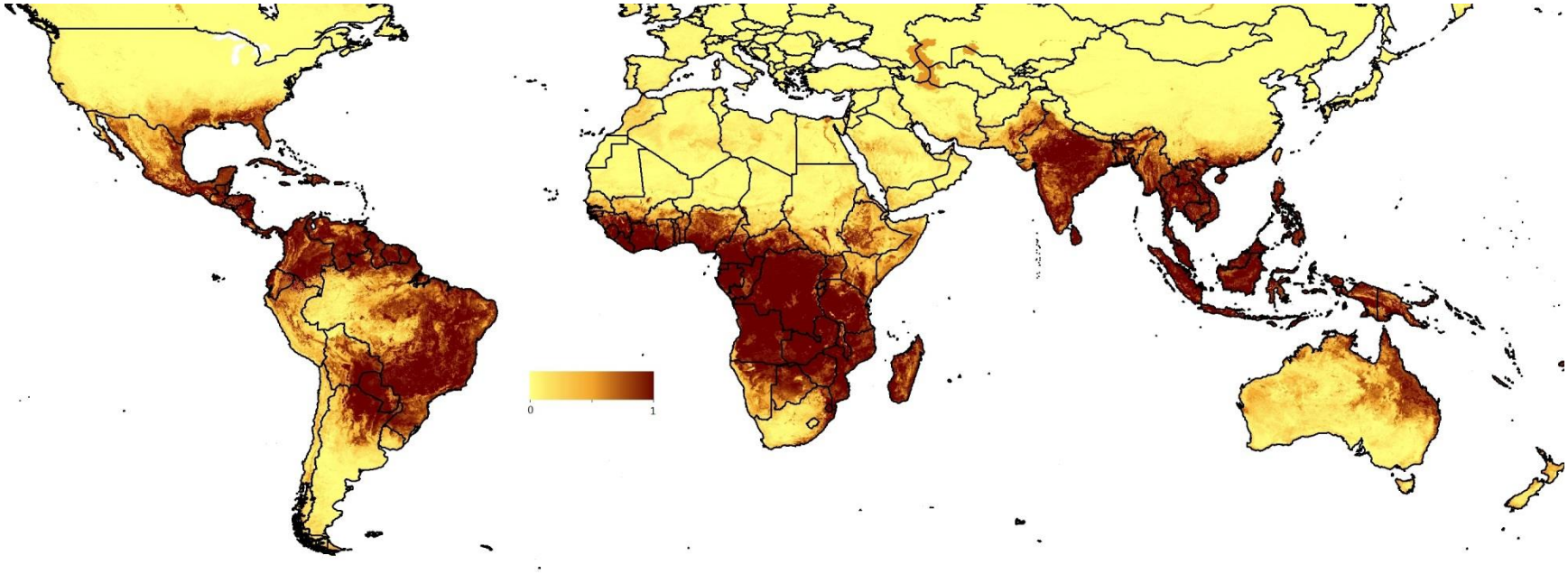


Figure A-9. *Aedes aegypti* presence: a map showing the predicted presence of *Aedes aegypti* based on hundreds of confirmed presence location data points. The yellow fever mosquito, *Aedes aegypti* is a mosquito that can spread the dengue fever, Chikungunya and yellow fever viruses, and other diseases. The mosquito can be recognized by white markings on legs and a marking in the form of a lyre on the thorax. The mosquito originated in Africa but is now found in tropical and subtropical regions throughout the world. Hundreds of geographically-located data points on field- caught *Ae aegypti* occurrence over the past 10 years were combined with climatic and environmental covariates within a regression tree mapping framework to produce a global map of predicted *Ae aegypti* presence.

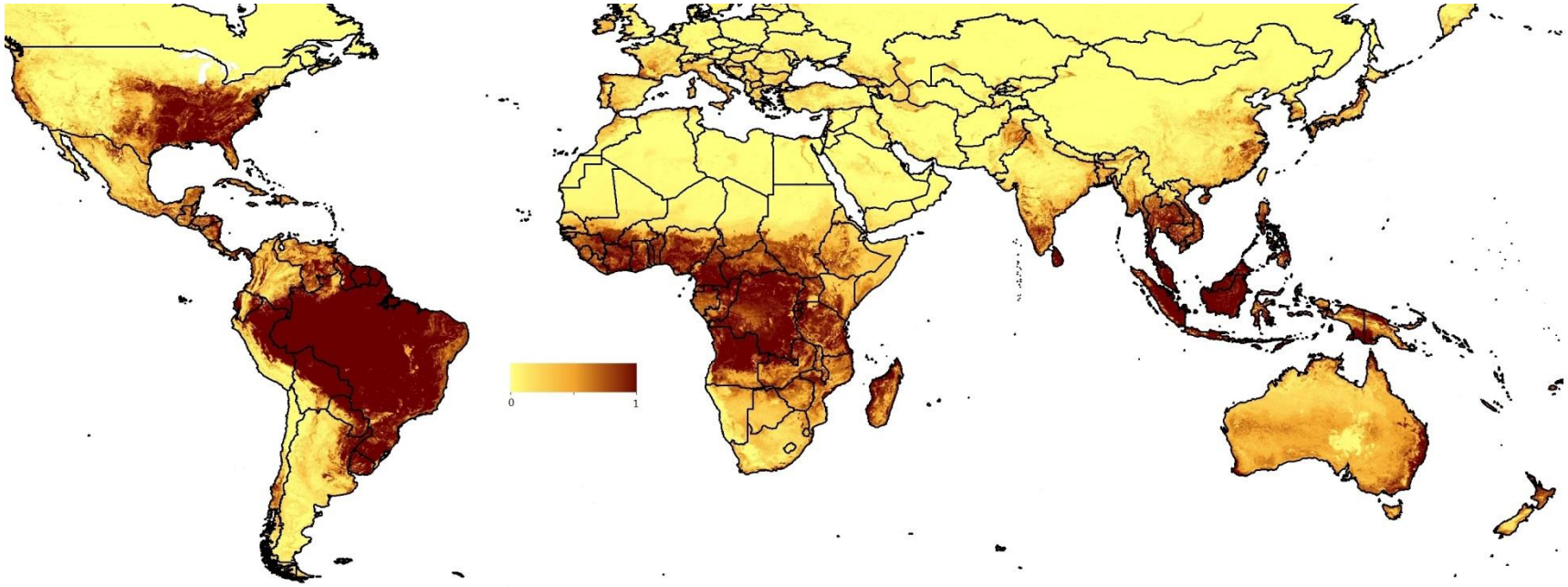


Figure A-10. Predicted *Ae. albopictus* distribution: a map showing the predicted presence of *Aedes albopictus* based on hundreds of confirmed presence location data points. *Ae. Albopictus* is of medical and public health concern because it has been shown in the laboratory to be a highly efficient vector of 22 arboviruses, including dengue, yellow, and West Nile fever viruses. In the wild, however, its efficiency as a vector appears to be generally low, although it has been implicated in recent dengue fever and chikungunya outbreaks in the absence of the principal vector, *Ae. aegypti*. From its Old World east Asian distribution reported in 1930, *Ae. albopictus* expanded its range first to the Pacific Islands and then, within the last 20 years, to other countries in both the Old World and the New World, principally through ship-borne transportation of eggs and larvae in tires. Hundreds of geographically-located data points on field-caught *Ae albopictus* occurrence over the past 10 years were combined with climatic and environmental covariates within a regression tree mapping framework to produce a global map of predicted *Ae albopictus* presence.

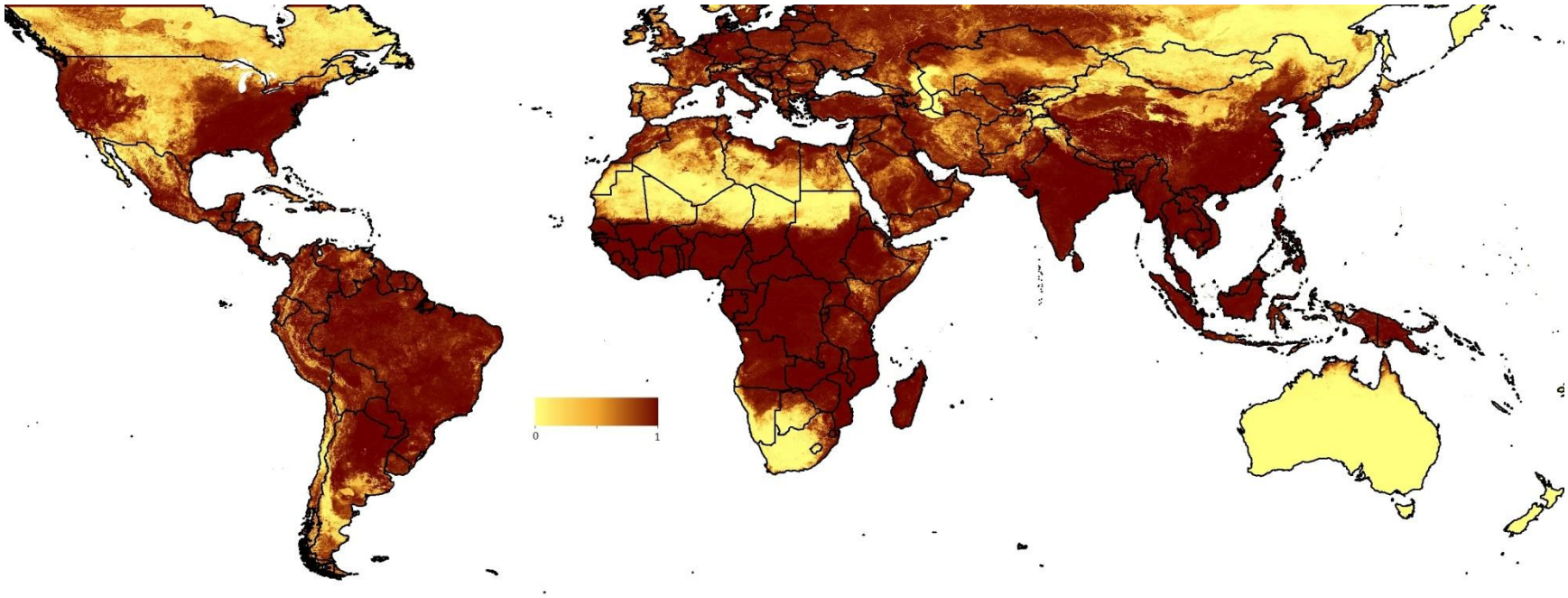


Figure A-11. *Anopheles* distribution: a map showing the predicted presence of one or more of the dominant Anopheles vectors of human malaria based on thousands of confirmed presence location data and expert opinion maps. Anopheles is a genus of mosquito. There are approximately 460 recognized species: while over 100 can transmit human malaria, only 30–40 commonly transmit parasites of the genus Plasmodium, which cause malaria in humans in endemic areas. *Anopheles gambiae* is one of the best known, because of its predominant role in the transmission of the most dangerous malaria parasite species – *Plasmodium falciparum*. Thousands of geographically-references data points on the presence of the dominant *Anopheles* vectors of malaria have been gathered and used, in combination with environmental covariates, expert opinion maps and regression tree tools, to produce global maps of *Anopheles* distributions by the Malaria Atlas Project (www.map.ox.ac.uk). Here, these maps were combined to produce a global map of dominant malaria-vector presence.

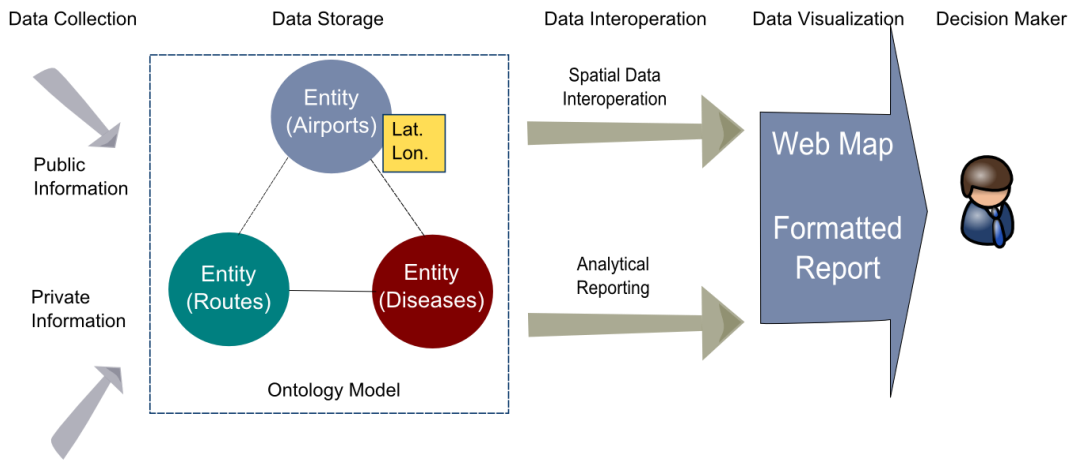


Figure A-12. Data flow in the VBD-Air. A data flow is presented here from raw data to decision maker. This data flow contains five stages: data collection allows business organization gathers information from various sources. Data storage constructs an ontology model for a knowledge graph connecting different concepts in the study domain; the location information is assigned to the vertexes as attributes. Data interoperation establishes data channels for visualization. Data Visualization provides interfaces to present spatial information on a web map or on a formatted report. The final object of this data flow is for decision making.

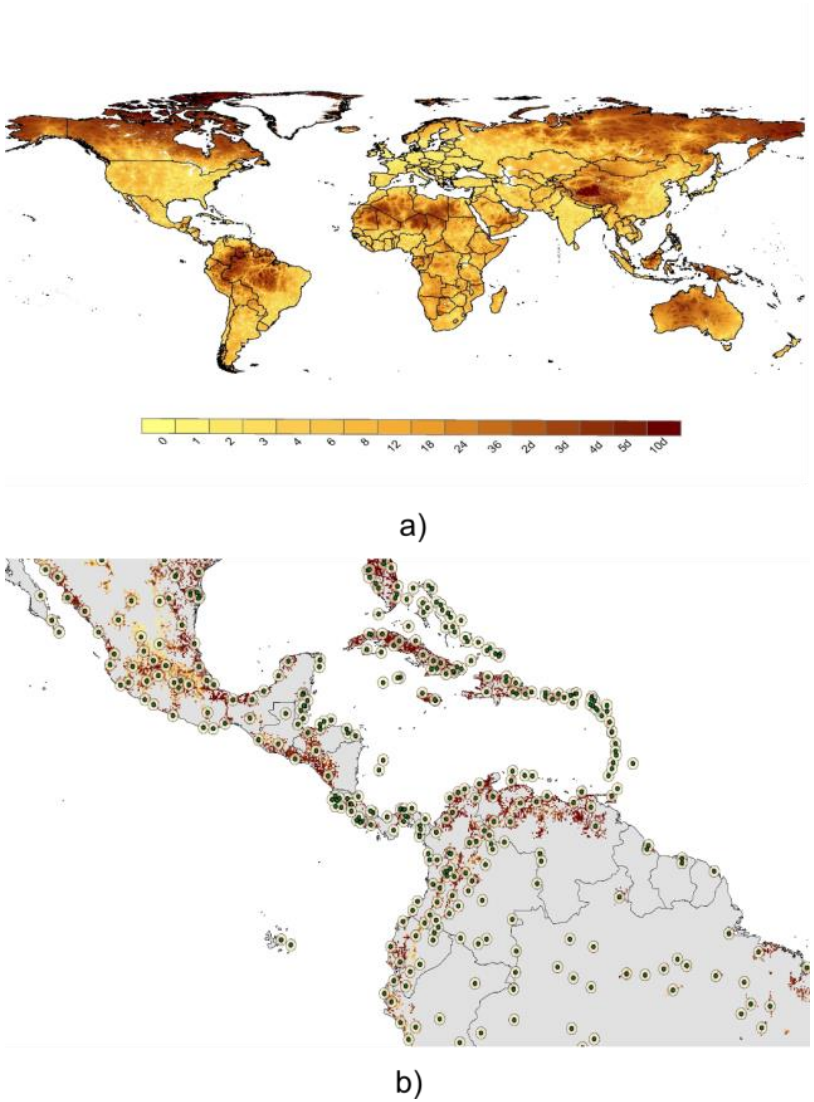


Figure A-13. Global accessibility map. a) Global Accessibility Map to the nearest major settlement (population size >50,000). b) Accessibility masks to an airport generated from the Global Accessibility Map.

APPENDIX B
SUPPLEMENTARY INFORMATION FOR CHAPTER 3

This appendix describes and compares the models we utilized in Chapter 3. To choose the best model for the predictions of air travel number, we adopted and compared these four models:

Lognormal model:

$$Ln(P_{ij}) = Intercept + \alpha' N_i + \sigma' N_j + \beta' R_{ij} + e_{ij} \quad (1)$$

Poisson model with variable interactions:

$$Ln(P_{ij}) = Intercept + \alpha' N_i + \sigma' N_j + \beta' R_{ij} + \gamma' Interaction_{ij} + e_{ij} \quad (2)$$

$$P_{ij} \sim Poisson$$

Negative Binomial Model with variable interactions:

$$Ln(P_{ij}) = Intercept + \alpha' N_i + \sigma' N_j + \beta' R_{ij} + \gamma' Interaction_{ij} + e_{ij} \quad (3)$$

$$P_{ij} \sim Negative\ Binomial$$

Log-linear model with variable effects and random effects on origin and destination airports

$$Ln(P_{ij}) = Intercept + \alpha' N_i + \sigma' N_j + \beta' R_{ij} + \gamma' Interaction_{ij} + b'_{hub_{ij}} City_{ij} + e_{ij} \quad (4)$$

$$b_{hub_{ij}} = \begin{pmatrix} b_{1hub_{ij}} \\ b_{2hub_{ij}} \end{pmatrix} \quad City_{ij} = \begin{pmatrix} Origin_{ij} \\ Destination_{ij} \end{pmatrix}$$

In these models, N_i and N_j describe the node characteristics for the origin airport and the destination airport. R_{ij} describes the route characteristics. $Interaction_{ij}$ shows the interactions of routes and nodes characteristics. The first model utilizes a mixed model with all the variables transformed in a logarithmic scale. The second and the third model utilize the generalized linear model framework with link function of Poisson and

Negative Binomial Distribution. In model 4, $City_{ij}$ denotes the random effects on the city and $b_{hub_{ij}}$ is identified as subjects in the mixed model.

Figure B-1 and figure B-2 shows the diagnostic plots for all four models. Table 1 presents the detailed fit statistics.

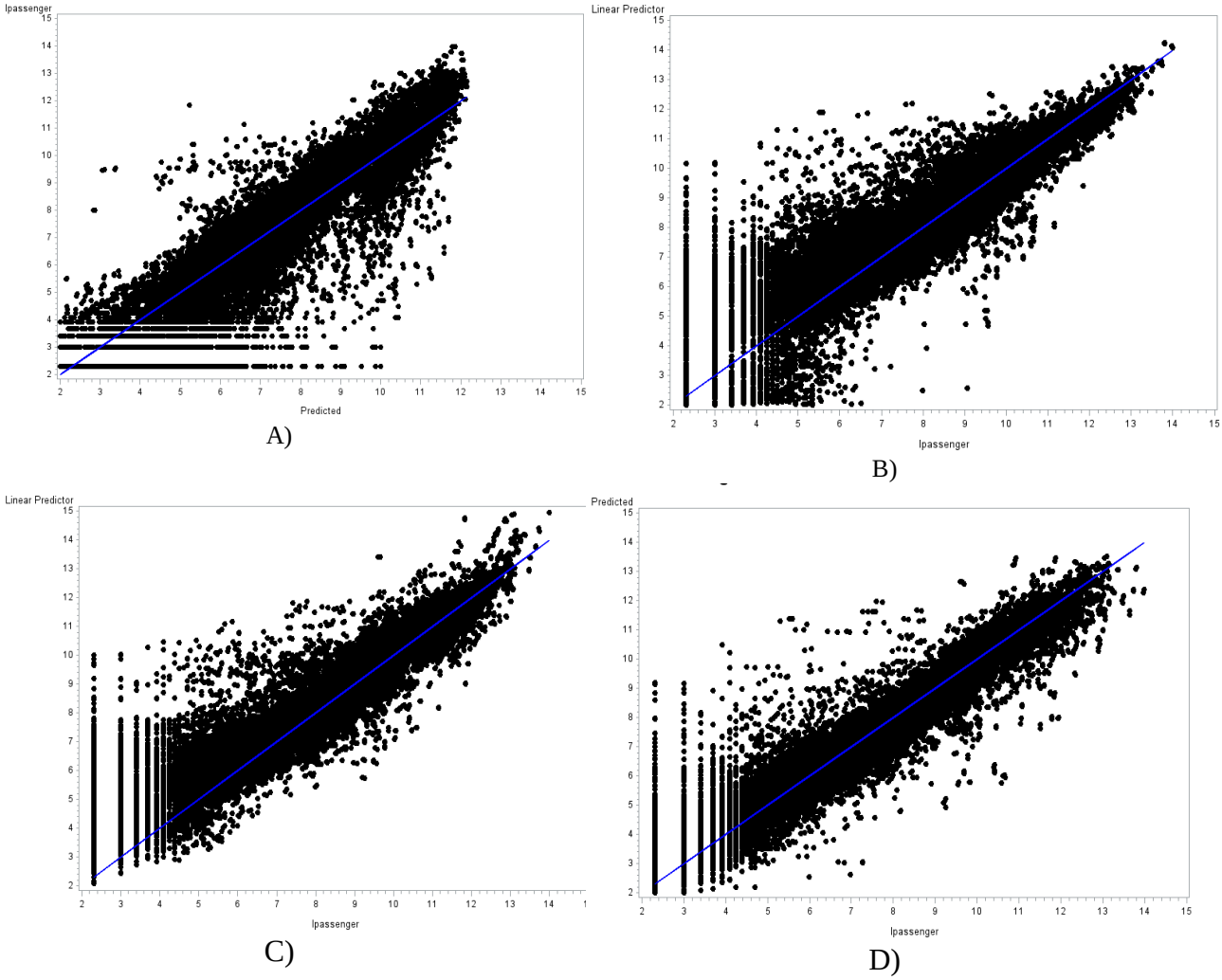


Figure B-1 Plots for predicted value vs. the predicted value at a log scale for all four models. A) Log normal model with main effects. B) Poisson prediction with variable interaction. C) Negative Binomial model with variable interaction. D) Log normal model with variable information and random effect

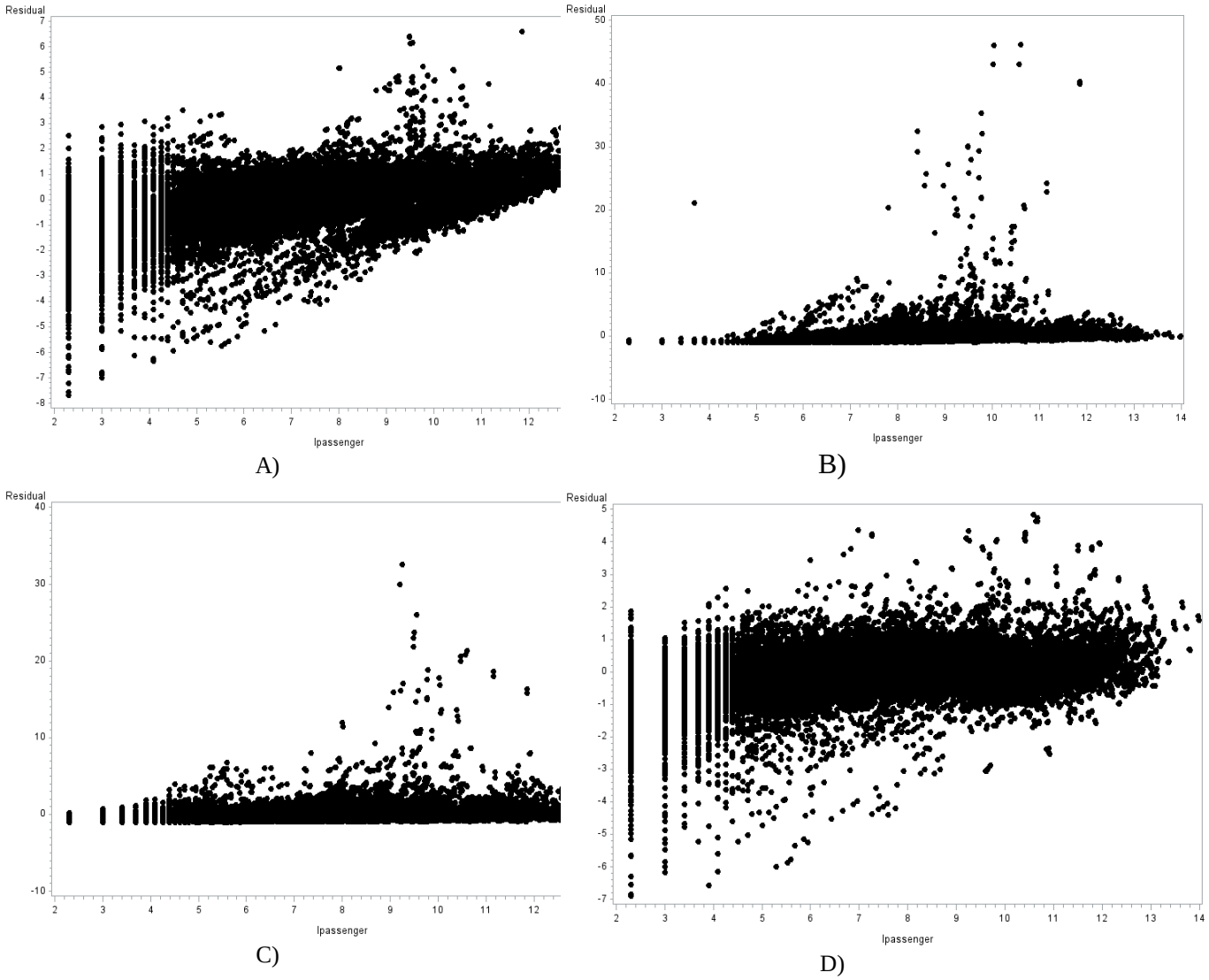


Figure B-2 Plots for residuals vs. the predicted values at a log scale for all models. A) Model 1. B) Model 2. C) Model 3. D) Model 4.

Table B-1. Fit characteristics and variable effects for model 4

Fit Statistics	
-2 Res Log Likelihood	54094.8
AIC (smaller is better)	54100.8
AICC (smaller is better)	54100.8
BIC (smaller is better)	54103.1

Table B-2 Fixed effects for variables in model 4

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
lop*stops	2	2E4	3.21	0.0403
ldp*stops	2	2E4	7.29	0.0007
lSX*stops	2	2E4	46.61	<.0001
lSY*stops	2	2E4	45.23	<.0001
lDOX*stops	2	2E4	2.24	0.1062
lDIY*stops	2	2E4	1.60	0.2024
lccX*stops	1	2E4	17.94	<.0001
lccY*stops	1	2E4	1.04	0.3073
lbcX*stops	2	2E4	2.43	0.0879
lbcY*stops	2	2E4	3.39	0.0339
li_d*stops	2	2E4	18.31	<.0001
lA*stops	1	2E4	13.68	0.0002
lPPPX*stops	2	2E4	4.90	0.0075
lPPPY*stops	2	2E4	5.81	0.0030
lPDCX*stops	2	2E4	1.15	0.3172
lPDCY*stops	2	2E4	2.07	0.1264
lop*dist	2	2E4	8.01	0.0003
ldp*dist	2	2E4	10.62	<.0001
lSX*dist	2	2E4	7.19	0.0008
lSY*dist	2	2E4	5.26	0.0052
lDOX*dist	2	2E4	2.51	0.0810

Table B-2. Continued

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
lDIY*dist	2	2E4	3.54	0.0291
lccX*dist	2	2E4	5.40	0.0045
lccY*dist	2	2E4	9.08	0.0001
lbcX*dist	2	2E4	0.78	0.4589
lbcY*dist	2	2E4	1.01	0.3652
li_d*dist	2	2E4	131.81	<.0001
lA*dist	2	2E4	11.75	<.0001
lPPPX*dist	2	2E4	2.87	0.0567
lPPPY*dist	2	2E4	3.07	0.0464
lPDCX*dist	2	2E4	0.38	0.6867
lPDCY*dist	2	2E4	1.16	0.3123
lop*hub	15	2E4	2.55	0.0008
ldp*hub	15	2E4	2.65	0.0005
lSX*hub	15	2E4	8.77	<.0001
lSY*hub	15	2E4	10.01	<.0001
lDOX*hub	15	2E4	4.01	<.0001
lDIY*hub	15	2E4	4.40	<.0001
lccX*hub	13	2E4	2.09	0.0121
lccY*hub	14	2E4	1.68	0.0517
lbcX*hub	15	2E4	2.62	0.0006
lbcY*hub	15	2E4	2.90	0.0001
li_d*hub	15	2E4	34.09	<.0001

Table B-2. Continued

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
lA*hub	15	2E4	9.11	<.0001
lPPPX*hub	15	2E4	2.03	0.0102
lPPPY*hub	15	2E4	2.39	0.0018
lPDCX*hub	15	2E4	2.75	0.0003
lPDCY*hub	15	2E4	2.93	0.0001
lop*gdpl	8	2E4	2.53	0.0095
ldp*gdpl	8	2E4	2.50	0.0104
lSX*gdpl	8	2E4	6.33	<.0001
lSY*gdpl	8	2E4	5.60	<.0001
lDOX*gdpl	8	2E4	6.59	<.0001
lDIY*gdpl	8	2E4	6.99	<.0001
lccX*gdpl	8	2E4	2.01	0.0414
lccY*gdpl	8	2E4	1.84	0.0645
lbcX*gdpl	8	2E4	2.04	0.0382
lbcY*gdpl	8	2E4	2.10	0.0321
li_d*gdpl	8	2E4	5.07	<.0001
lA*gdpl	8	2E4	8.38	<.0001
lPPPX*gdpl	8	2E4	2.39	0.0144
lPPPY*gdpl	8	2E4	1.99	0.0433
lPDCX*gdpl	8	2E4	1.33	0.2220
lPDCY*gdpl	8	2E4	1.01	0.4262
stops	2	2E4	36.85	<.0001

Table B-2. Continued

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
dist	2	2E4	16.00	<.0001
hub	14	772	4.22	<.0001
Country	1	2E4	113.04	<.0001

APPENDIX C
SUPPLEMENTARY INFORMATION FOR CHAPTER 4

This appendix presents the supplementary tables and figures for chapter 3.

Table C-1. Top 10 *P.f.* Routes and *P.v.* Routes outside of the Great Mekong Sub-Region. Values with (*) are returned as the adjusted largest flows between two connection flights.

Outgoing *P.f.* Flow						
Origin City	Destination City	Source Country	Destination Country	Flow Value	Number of Connections	Transmission Type at Destination
Yangon	Harare	Myanmar	Zimbabwe	997.2	1	Unstable
Yangon	Brazzaville	Myanmar	Congo	287.8*	2	Unstable
Yangon	Kinshasa	Myanmar	Congo	287.8*	2	Unstable
Yangon	Mombasa	Myanmar	Kenya	287.8*	2	Unstable
Yangon	Lome	Myanmar	Togo	287.8*	2	Unstable
Yangon	Maputo	Myanmar	Mozambique	287.8*	2	Unstable
Yangon	Ouagadougou	Myanmar	Burkina Faso	287.8*	2	Unstable
Yangon	Djibouti	Myanmar	Djibouti	287.8*	2	Stable
Yangon	Bamako	Myanmar	Mali	282.3*	2	Unstable
Ho Chi Minh City	Harare	Vietnam	Zimbabwe	243.2	1	Unstable
Outgoing *P.v.* Flow						
Origin City	Destination City	Source Country	Destination Country	Flow Value	Number of Connections	Transmission Type at Destination
Bangkok	Addis Ababa	Thailand	Ethiopia	1940.0	0	Unstable
Bangkok	Nairobi	Thailand	Kenya	1631.0	0	Unstable
Ho Chi Minh City	Harare	Vietnam	Zimbabwe	981.6	1	Unstable
Yangon	Harare	Myanmar	Zimbabwe	512.5	1	Unstable
Bangkok	Lusaka	Thailand	Zambia	434.1	1	Unstable

Phnom Penh	Nairobi	Cambodia	Kenya	410.8	1	Unstable
Phnom Penh	Addis Ababa	Cambodia	Ethiopia	410.8	1	Unstable
Phnom Penh	Antananarivo	Cambodia	Madagascar	363.1	1	Unstable
Siem Reap	Addis Ababa	Cambodia	Ethiopia	275.5*	1	Unstable
Siem Reap	Nairobi	Cambodia	Kenya	275.5*	1	Unstable

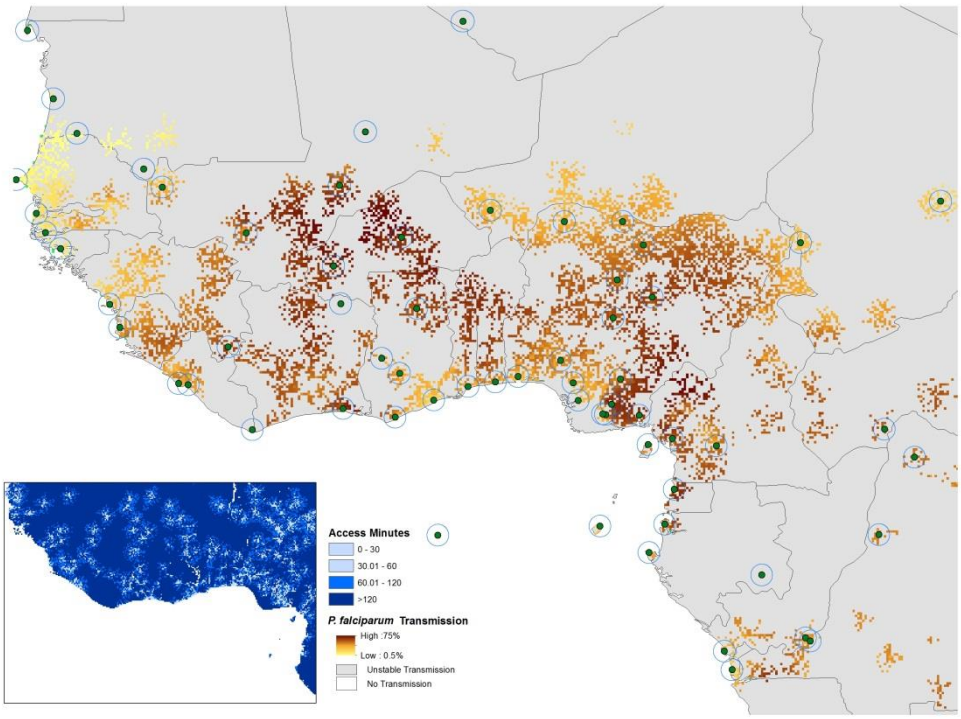


Figure C-1. The travel time/distance mask to extract the prevalent rate of *P.falciparum /P.vivax*. Inset map: travel time to the nearest major settlement (population size >50,000). Main map: each dot shows an airport location with a 50km buffer around it, the raster is the global *P.f.* relevant map clip by the global access time map with value less than 2 hour. These 2 hour and 50km thresholds were used to assign disease risks to airports (see main text).

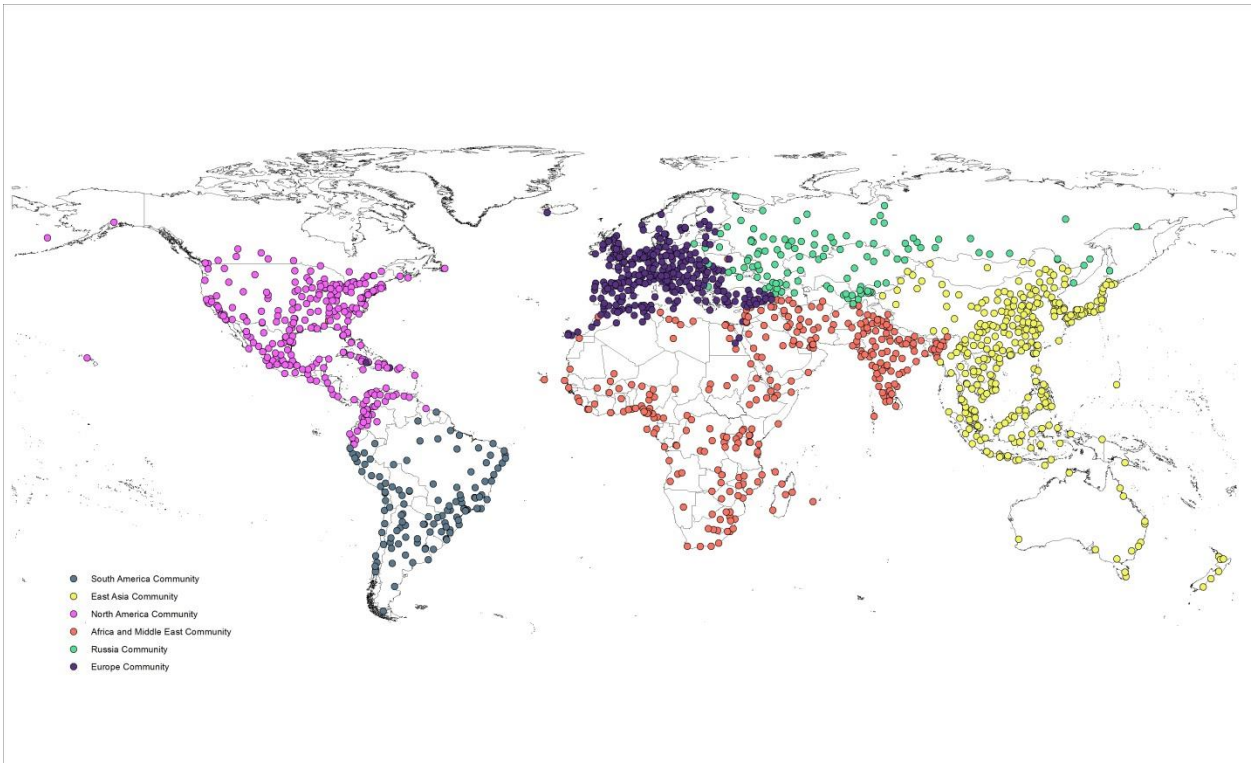
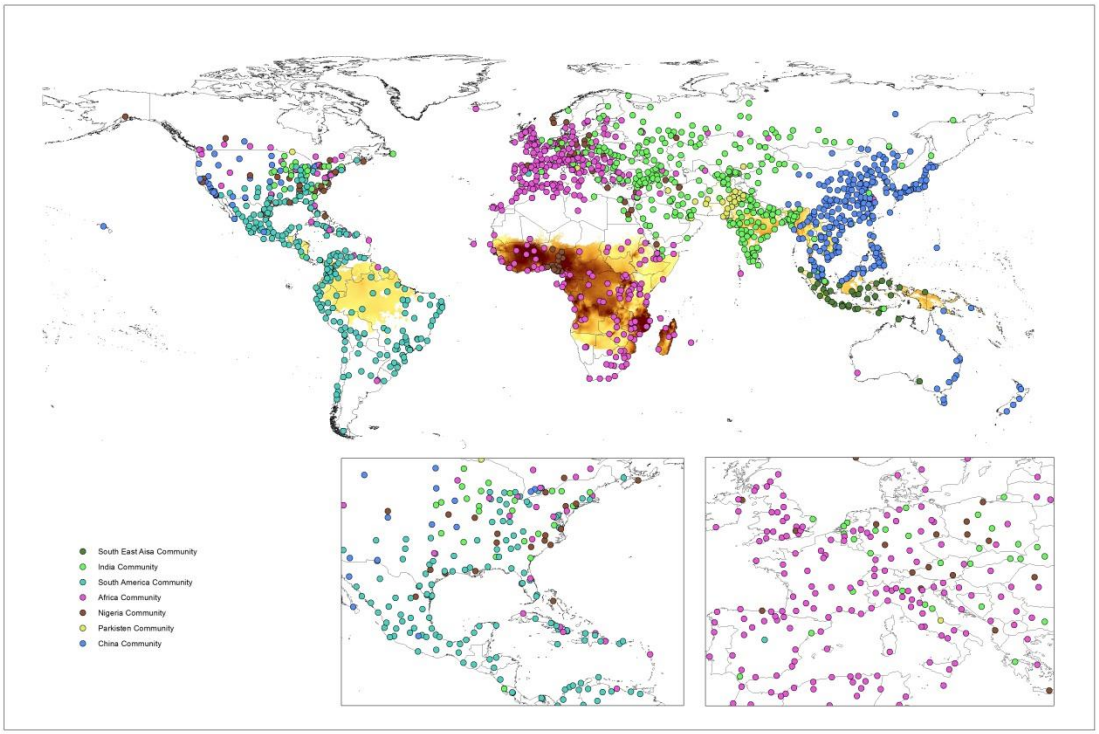
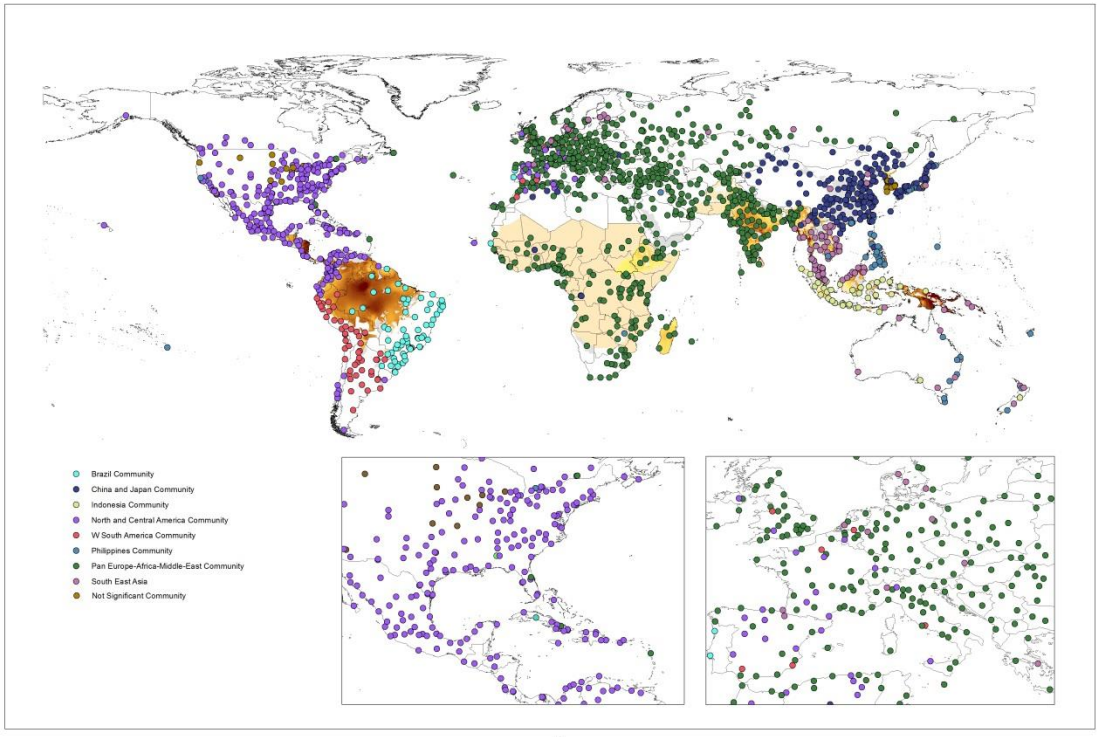


Figure C-2. Air travel network communities weighted by directed estimate flow.

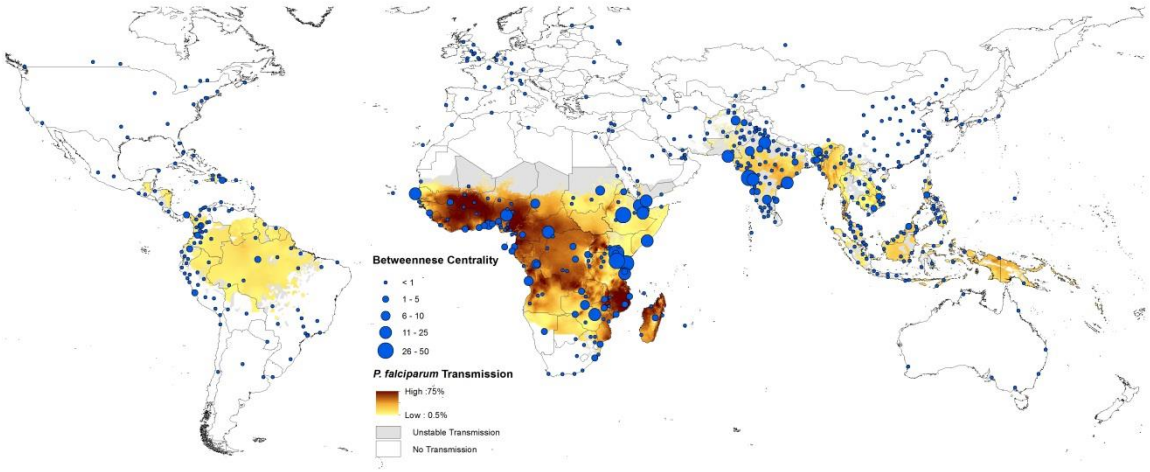


A)

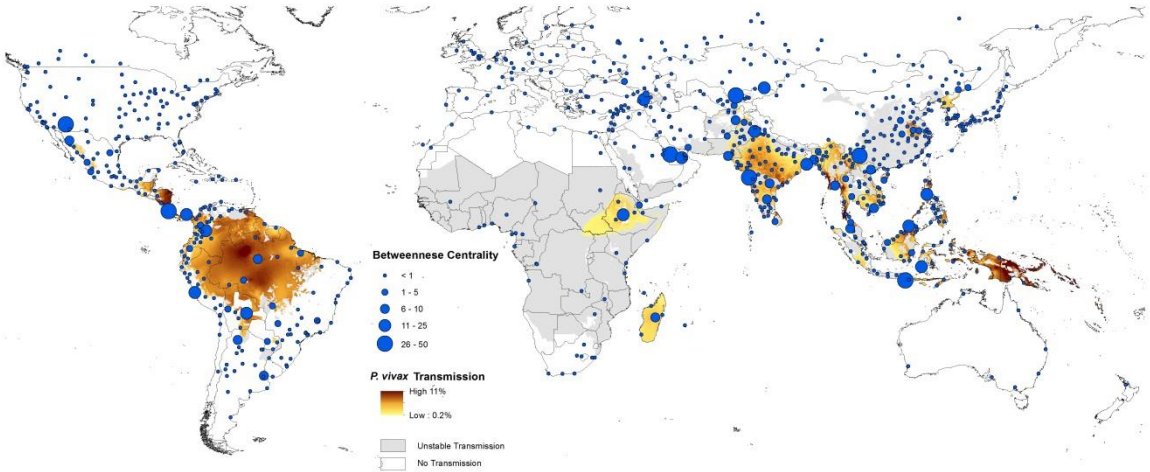


B)

Figure C-3. Communities for all possible connections originating from *P.falciparum /P.vivax.* endemic areas. A) *P.falciparum* multilevel membership. B) *P.vivax* multilevel membership. These two maps show direct-connected, one-transferred and two-transferred airports from endemic area.

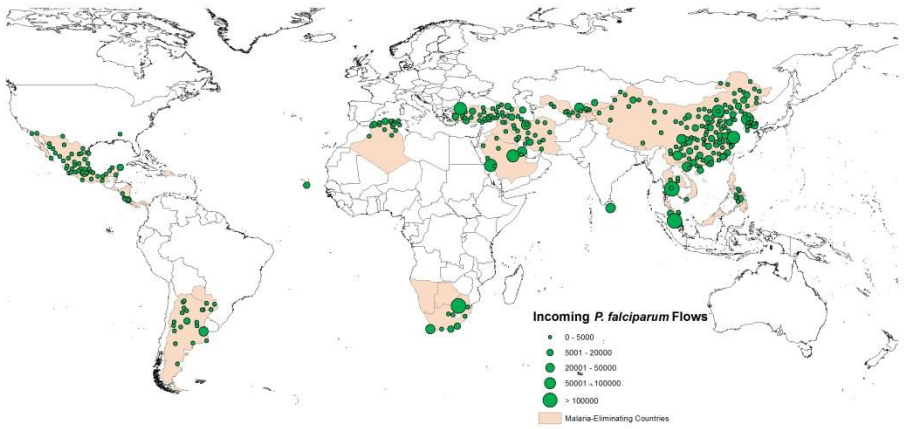


A)

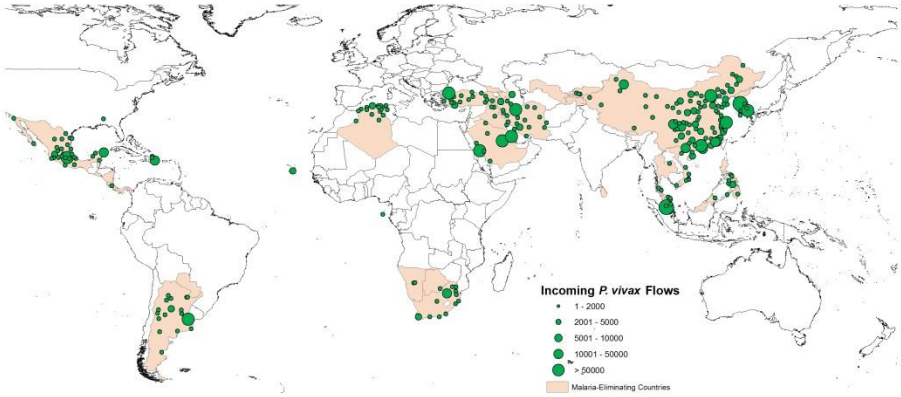


B)

Figure C-4. Spatial distributions of airports with *P.falciparum /P.vivax* betweenness centrality scores. A) Top airports with high betweenness score from *P.falciparum* endemic area, weighted by the *P.falciparum* Flow. B) Top airports with high betweenness score from *P.vivax* endemic area, weighted by the *P.vivax* Flow.



A)



B)

Figure C-5. Spatial distributions of airport nodes weighted by incoming *P.falciparum /P.vivax* flows. A) Top airports with high incoming *P.falciparum* risk flows from *P.falciparum* endemic area, weighted by the *P.falciparum.* Flow. B) Top airports with high incoming *P.vivax* risk flows from *P.vivax* endemic area, weighted by the *P.vivax* Flow.

LIST OF REFERENCES

1. IATA (2012) IATA 2012 Annual Review. Beijing: IATA.

2. IATA (2012) IATA Press Release: 50. Available: http://www.iata.org/pressroom/pr/Pages/2012-12-06-01.aspx.

3. Rvachev L, Longini IM (1985) A mathematical model for the global spread of influenza. Math Biosci 75: 3–22. doi:10.1016/0025-5564(85)90064-1.

4. Colizza V, Barrat A, Barthelemy M (2007) Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. PLoS Med.

5. Tatem AJ, Rogers DJ, Hay SI, Graham a J, Goetz SJ (2006) Global transport networks and infectious disease spread. Adv Parasit 62: 37–77. doi:10.1016/S0065-308X(05)62009-X.

6. Ruan S, Wang W, Levin S (2006) The effect of global travel on the spread of SARS. Mathematical Biosciences and ….

7. Tatem AJ, Huang Z, Das A, Qi Q, Roth J, et al. (2012) Air travel and vector-borne disease movement. Parasitology: 1–15. doi:10.1017/S0031182012000352.

8. Huang Z, Das A, Qiu Y, Tatem AJ (2012) Web-based GIS: the vector-borne disease airline importation risk (VBD-AIR) tool. Int J Health Geogr 11: 33. doi:10.1186/1476-072X-11-33.

9. Tatem AJ, Hay SI, Rogers DJ (2006) Global traffic and disease vector dispersal. Proc Natl Acad Sci 103: 6242–6247. doi:10.1073/pnas.0508391103.

10. Tatem AJ, Rogers DJ, Hay SI (2006) Estimating the malaria risk of African mosquito movement by air travel. Malaria J 5: 57. doi:10.1186/1475-2875-5-57.

11. World Health Organization (2007) The world health report 2007 - A safer future: global public health security in the 21st century. Heymann D, editor.

12. Khan K, McNabb SJ, Memish ZA, Eckhardt R, Hu W, et al. (2012) Infectious disease surveillance and modelling across geographic frontiers and scientific specialties. Lancet Infect Dis 12: 222–230. doi:10.1016/S1473-3099(11)70313-9.

13. Svihrova V, Szilagyiova M, Novakova E, Svihra J, Hudeckova H (2012) Costs analysis of the treatment of imported malaria. Malaria journal 11: 1. doi:10.1186/1475-2875-11-1.

14. Jelinek T (2008) Trends in the epidemiology of dengue fever and their relevance for importation to Europe. Euro Surveill: 1999–2001.

15. Odolini S, Parola P, Gkrania-Klotsas E, Caumes E, Schlagenhauf P, et al. (2012) Travel-related imported infections in Europe, EuroTravNet 2009. Clin Microbiol Infect 18: 468–474. doi:10.1111/j.1469-0691.2011.03596.x.

16. Centers for Disease Control and Prevention (2003) Local Transmission of Plasmodium vivax Malaria --- Palm Beach County, Florida, 2003. MMWR Morb Mortal Wkly Rep 52: 908–911.

17. Centers for Disease Control and Prevention (2002) Local Transmission of Plasmodium vivax Malaria --- Virginia, 2002. MMWR Morb Mortal Wkly Rep 51: 921–923.

18. Danis K, Baka A, Lenglet A, Van Bortel W, Terzaki I, et al. (2011) Autochthonous Plasmodium vivax malaria in Greece, 2011. Euro Surveill 16: 1–5.

19. Powers A, Logue C (2007) Changing patterns of chikungunya virus: re-emergence of a zoonotic arbovirus. Journal of General Virology.

20. Cohen JM, Smith DL, Cotter C, Ward A, Yamey G, et al. (2012) Malaria resurgence: a systematic review and assessment of its causes. Malaria J 11: 122. doi:10.1186/1475-2875-11-122.

21. Mackenzie JS, Gubler DJ, Petersen LR (2004) Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. Nature medicine 10: S98–109. doi:10.1038/nm1144.

22. Phyo AP, Nkhoma S, Stepniewska K, Ashley E a, Nair S, et al. (2012) Emergence of artemisinin-resistant malaria on the western border of Thailand: a longitudinal study. Lancet 379: 1960–1966. doi:10.1016/S0140-6736(12)60484-X.

23. Freedman DO, Weld LH, Kozarsky PE, Fisk T, Robins R, et al. (2006) Spectrum of disease and relation to place of exposure among ill returned travelers. New England Journal of Medicine 354: 119–130.

24. Stephan C, Allwinn R, Brodt HR, Knupp B, Preiser W, et al. (2002) Travel-Acquired Dengue Infection: Clinical Spectrum and Diagnostic Aspects. Infection 30: 225–228. doi:10.1007/s15010-002-3052-7.

25. Franco-Paredes C, Santos-Preciado JI (2006) Problem pathogens: prevention of malaria in travellers. The Lancet infectious diseases 6: 139–149.

26. Askling HH, Bruneel F, Buchard G, Castelli F, Chiodini PL, et al. (2012) Management of imported malaria in Europe. Malaria J 11: 328. doi:10.1186/1475-2875-11-328.

27. Charrel RN, De Lamballerie X, Raoult D (2007) Chikungunya outbreaks—the globalization of vectorborne diseases. New England Journal of Medicine 356: 769–771.

28. Checkley AM, Smith A, Smith V, Blaze M, Bradley D, et al. (2012) Risk factors for mortality from imported falciparum malaria in the United Kingdom over 20 years: an observational study. Bmj 344: e2116–e2116. doi:10.1136/bmj.e2116.

29. Kilpatrick AM (2011) Globalization, land use, and the invasion of West Nile virus. Science (New York, NY) 334: 323–327. doi:10.1126/science.1201010.

30. Shang C-S, Fang C-T, Liu C-M, Wen T-H, Tsai K-H, et al. (2010) The role of imported cases and favorable meteorological conditions in the onset of dengue epidemics. PLoS neglected tropical diseases 4: e775. doi:10.1371/journal.pntd.0000775.

31. Pindolia DK, Garcia AJ, Wesolowski A, Smith DL, Buckee CO, et al. (2012) Human movement data for malaria control and elimination strategic planning. Malaria J 11: 205. doi:10.1186/1475-2875-11-205.

32. Gething P, Patil A, Smith D (2011) A new world malaria map: Plasmodium falciparum endemicity in 2010. Malaria J.

33. Gething PW, Elyazar IRF, Moyes CL, Smith DL, Battle KE, et al. (2012) A Long Neglected World Malaria Map: Plasmodium vivax Endemicity in 2010. PLoS Negl 6: e1814. doi:10.1371/journal.pntd.0001814.

34. Rogers DJ, Wilson a J, Hay SI, Graham a J (2006) The global distribution of yellow fever and dengue. Adv Parasit 62: 181–220. doi:10.1016/S0065-308X(05)62006-4.

35. Sinka M, Bangs M, Manguin S (2010) The dominant Anopheles vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis. … Vectors.

36. Sinka ME, Bangs MJ, Manguin S, Chareonviriyaphap T, Patil AP, et al. (2011) The dominant Anopheles vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis. Parasites & Vectors 4: 89.

37. Sinka ME, Rubio-Palis Y, Manguin S, Patil AP, Temperley WH, et al. (2010) The dominant Anopheles vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis. Parasites vectors 3: 72.

38. Fuller D, Troyo A, Calderon-Arguedas O, Beier J (2010) Dengue vector (Aedes aegypti) larval habitats in an urban environment of Costa Rica analysed with ASTER and QuickBird imagery. International Journal of Remote Sensing 31: 3–11.

39. Benedict MQ, Levine RS, Hawley WA, Lounibos LP (2007) Spread of the tiger: global risk of invasion by the mosquito Aedes albopictus. Vector borne and zoonotic diseases (Larchmont, NY) 7: 76–85. doi:10.1089/vbz.2006.0562.

40. Lounibos LP (2002) Invasions by insect vectors of human disease. Annual review of entomology 47: 233–266.

41. Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW (2004) Reviews The global distribution and population at risk of malaria : past , present , and future. Lancet 4: 327–336.

42. Chiyaka C, Tatem A, Cohen J, Gething P, Johnston G, et al. (2013) The Stability of Malaria Elimination. Science in press.

43. IATA (2011) Annual Report 2011 International Air Transport Association.

44. Randolph SE, Rogers DJ (2010) The arrival, establishment and spread of exotic diseases: patterns and predictions. Nature Reviews Microbiology 8: 361–371. doi:10.1038/nrmicro2336.

45. Haggett P (2000) The geographical structure of epidemics. Oxford University Press, USA.

46. Colizza V, Barrat A, Barthe M, Vespignani A (2005) The role of the airline transportation network in the prediction and predictability of global epidemics. Proc Natl Acad Sci.

47. Coelho FC, Cruz OG, Codeço CT (2008) Epigrass: a tool to study disease spread in complex networks. Source code for biology and medicine 3: 3. doi:10.1186/1751-0473-3-3.

48. Tatem AJ, Hay SI (2007) Climatic similarity and biological exchange in the worldwide airline transportation network. Proceedings Biological sciences / The Royal Society 274: 1489–1496. doi:10.1098/rspb.2007.0148.

49. Laird M (1984) Commerce and the spread of pests and disease vectors. 15th Pacific Science Congress, Dunedin, NZ (USA), ….

50. Russell RC (1987) Survival of insects in the wheel bays of a Boeing 747B aircraft on flights between tropical and temperate airports. Bull World Health Organ 65: 659–662.

51. Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, et al. (2010) The dominant Anopheles vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic precis. Parasites vectors 3: 117. doi:10.1186/1756-3305-3-117.

52. Charrel RN, De Lamballerie X, Raoult D (2008) Seasonality of mosquitoes and chikungunya in Italy. The Lancet infectious diseases 8: 5–6. doi:10.1016/S1473-3099(07)70296-7.

53. Guerra C a, Gikandi PW, Tatem AJ, Noor AM, Smith DL, et al. (2008) The limits and intensity of Plasmodium falciparum transmission: implications for malaria control and elimination worldwide. PLoS medicine 5: e38. doi:10.1371/journal.pmed.0050038.

54. Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP, et al. (2010) The international limits and population at risk of Plasmodium vivax transmission in 2009. PLoS neglected tropical diseases 4: e774.

55. Whitehorn J, Farrar J (2010) Dengue. British medical bulletin.

56. Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. The Journal of animal ecology 77: 802–813. doi:10.1111/j.1365-2656.2008.01390.x.

57. Monath T (2001) Yellow fever: an update. The Lancet infectious diseases.

58. Moffett A, Strutz S, Guda N (2009) A global public database of disease vector and reservoir distributions. PLoS neglected tropical ….

59. Foley DH, Wilkerson RC, Birney I, Harrison S, Christensen J, et al. (2010) MosquitoMap and the Mal-area calculator: new web tools to relate mosquito species distribution with vector borne disease. Int J Health Geogr.

60. Gratz NG (2004) Critical review of the vector status of Aedes albopictus. Medical and veterinary entomology 18: 215–227. doi:10.1111/j.0269-283X.2004.00513.x.

61. Res C, New M, Lister D, Hulme M, Makin I (2002) A high-resolution data set of surface climate over global land areas. 21: 1–25.

62. Budhathoki NR, Bruce B (Chip) C, Nedovic-Budic Z (2008) Reconceptualizing the role of the user of spatial data infrastructure. GeoJournal 72: 149–160. doi:10.1007/s10708-008-9189-x.

63. Tatem AJ (2009) The worldwide airline network and the dispersal of exotic species: 2007-2010. Ecography 32: 94–102. doi:10.1111/j.1600-0587.2008.05588.x.

64. Sanderson S (2010) Pro ASP.NET MVC 2 Framework. Apress.

65. Vlissides J, Helm R, Johnson R, Gamma E (1995) Design patterns: Elements of reusable object-oriented software. Reading: Addison-Wesley.

66. Cockburn A (2007) Agile software development: the cooperative game.

67. Shneiderman B (1996) The eyes have it: A task by data type taxonomy for information visualizations IEEE. pp. 336–343.

68. Johansson MA, Arana-Vizcarrondo N, Biggerstaff BJ, Staples JE, Gallagher N, et al. (2011) On the Treatment of Airline Travelers in Mathematical Models. PLOS ONE 6: e22151. doi:10.1371/journal.pone.0022151.

69. Patil AP, Gething PW, Piel FB, Hay SI (2011) Bayesian geostatistics in health cartography: the perspective of malaria. Trends Parasitol.

70. Brownstein JS, Freifeld CC, Reis BY, Mandl KD (2008) Surveillance Sans Frontieres: Internet-based emerging infectious disease intelligence and the HealthMap project. PLoS medicine 5: e151. doi:10.1371/journal.pmed.0050151.

71. Gardner L, Fajardo D, Waller S (2012) A Predictive Spatial Model to Quantify the Risk of Air-Travel-Associated Dengue Importation into the United States and Europe. Journal of Tropical ….

72. Parker J, Epstein JM (2011) A Distributed Platform for Global-Scale Agent-Based Models of Disease Transmission. Acm T Model Comput S 22: 1–25. doi:10.1145/2043635.2043637.

73. Freifeld CC, Mandl KD, Reis BY, Brownstein JS (2008) HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. Journal of the American Medical Informatics Association 15: 150.

74. Johansson MA, Arana-Vizcarrondo N, Biggerstaff BJ, Gallagher N, Marano N, et al. (2012) Assessing the risk of international spread of yellow fever virus: a mathematical analysis of an urban outbreak in Asuncion, 2008. Am J Trop Med Hyg 86: 349–358. doi:10.4269/ajtmh.2012.11-0432.

75. Wilson M (1995) Travel and the emergence of infectious diseases. Emerging infectious diseases.

76. Wilson M (2003) The traveller and emerging infections: sentinel, courier, transmitter. Journal of applied microbiology.

77. Upham P, Thomas C, Gillingwater D, Raper D (2003) Environmental capacity and airport operations: current issues and future prospects. J Air Transp Manag 9: 145–151. doi:10.1016/S0969-6997(02)00078-9.

78. Smith D, Timberlake M (2001) World City Networks and Hierarchies, 1977-1997 An Empirical Analysis of Global Air Travel Links. Am Behav Sci 44: 1656–1678.

79. Marazzo M, Scherre R, Fernandes E (2010) Air transport demand and economic growth in Brazil: A time series analysis. Transport Res E-Log 46: 261–269. doi:10.1016/j.tre.2009.08.008.

80. Adey P, Budd L, Hubbard P (2007) Flying lessons: exploring the social and cultural geographies of global air travel. Prog Hum Geog 31: 773–791. doi:10.1177/0309132507083508.

81. Stoddard ST, Morrison AC, Vazquez-Prokopec GM, Soldan VP, Kochel TJ, et al. (2009) The role of human movement in the transmission of vector-borne pathogens. PLoS Negl 3: e481. doi:10.1371/journal.pntd.0000481.

82. Grais RF, Ellis JH, Glass GE (2003) Assessing the impact of airline travel on the geographic spread of pandemic influenza. European journal of epidemiology 18: 1065–1072.

83. Grais RF, Ellis JH (2004) Modeling the Spread of Annual Influenza Epidemics in the U.S.:The Potential Role of Air Travel. Health Care Manag Sci 7: 127–134.

84. Grubesic TH, Matisziw TC, Zook M a. (2008) Global airline networks and nodal regions. GeoJournal 71: 53–66. doi:10.1007/s10708-008-9117-0.

85. Long W (1970) Air travel, spatial structure, and gravity models. Ann Reg Sci: 97–107.

86. Haynes K, Fotheringham A (1984) Gravity and spatial interaction models.

87. Brownstein J (2006) Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. PLoS medicine 3: e401.

88. Balcan D, Gonçalves B, Hu H (2010) Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model. J Comput Sci 1: 132–145.

89. Kenah E, Chao DL, Matrajt L, Halloran ME, Longini IM (2011) The global transmission and control of influenza. PloS one 6: e19515. doi:10.1371/journal.pone.0019515.

90. Grosche T, Rothlauf F, Heinzl A (2007) Gravity models for airline passenger volume estimation. J Air Transp Manag 13: 175–183. doi:10.1016/j.jairtraman.2007.02.001.

91. O'Kelly M (1986) The location of interacting hub facilities. Transport Sci 20: 92–106.

92. Bryan DL, O'Kelly ME (1999) Hub-and-Spoke Networks in Air Transportation: An Analytical Review. J Regional Sci 39: 275–295. doi:10.1111/1467-9787.00134.

93. O'Kelly ME (2010) Activity Levels at Hub Facilities in Interacting Networks. Geogr Anal 18: 343–356. doi:10.1111/j.1538-4632.1986.tb00106.x.

94. Kuby MJ, Gray RG (1993) The hub network design problem with stopovers and feeders: The case of Federal Express. Transport Res A-Pol 27: 1–12.

95. Guimerà R, Mossa S, Turtschi a, Amaral L a N (2005) The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. Proc Natl Acad Sci 102: 7794–7799. doi:10.1073/pnas.0407994102.

96. Nystuen J, Dacey M (1961) A Graph Theory Interpretation of Nodal Regions. Pap Reg Sci 34: 853–864.

97. Bhadra D (2003) Demand for air travel in the United States: bottom-up econometric estimation and implications for forecasts by origin and destination pairs. J of Air Transp 8.

98. Taaffe E (1956) Air transportation and United States urban distribution. Geogr Rev 46: 219–238.

99. Brueckner J (2003) Airline Traffic and Urban Economic Development. Urban Stud 40: 1455–1469. doi:10.1080/0042098032000094388.

100. Jin F, Wang F, Liu Y (2004) Geographic Patterns of Air Passenger Transport in China 1980 – 1998 : Imprints of Economic Growth, Regional Inequality and Network Development: 37–41.

101. Goetz A (1992) Air passenger transportation and growth in the US urban system, 1950–1987. Growth Change 23: 217–238.

102. Irwin M, Kasarda J (1991) Air passenger linkages and employment growth in US metropolitan areas. Am Sociol Rev 56: 524–537.

103. Liu Z-J, Debbage K, Blackburn B (2006) Locational determinants of major US air passenger markets by metropolitan area. J Air Transp Manag 12: 331–341. doi:10.1016/j.jairtraman.2006.08.001.

104. Wang J, Mo H, Wang F, Jin F (2011) Exploring the network structure and nodal centrality of China's air transport network: A complex network approach. J Transp Geogr 19: 712–721. doi:10.1016/j.jtrangeo.2010.08.012.

105. Newman M (2004) Analysis of weighted networks. Phys Rev E 70: 056131.

106. Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. Proc Natl Acad Sci 101: 3747–3752. doi:10.1073/pnas.0400087101.

107. Taylor PJ, Catalano G, Walker DRF (2002) Measurement of the World City Network. Urban Stud 39: 2367–2376. doi:10.1080/00420980220080011.

108. Button K, Lall S, Stough R, Trice M (1999) High-technology employment and hub airports. J Air Transp Manag 5: 53–59. doi:10.1016/S0969-6997(98)00038-6.

109. Balcan D, Colizza V, Gonc B, Hu H (2009) Multiscale mobility networks and the spatial spreading of infectious diseases. Proc Natl Acad Sci 106: 21484–21489.

110. Snijders TAB, Bosker RJ (1994) Modeled Variance in Two-Level Models. Sociological Methods & Research 22: 342–363. doi:10.1177/0049124194022003004.

111. Recchia A (2010) R-squared measures for two-level hierarchical linear models using SAS. J Stat Softw 32: Code Snippet 2.

112. Wang J, Jin F (2007) China's Air Passenger Transport: An Analysis of Recent Trends. Eurasian Geography and Economics 48: 469–480. doi:10.2747/1538-7216.48.4.469.

113. Brons M, Pels E, Nijkamp P, Rietveld P (2002) Price elasticities of demand for passenger air travel: a meta-analysis. J Air Transp Manag 8: 165–175. doi:10.1016/S0969-6997(01)00050-3.

114. Collinson S (1996) Visa requirements, carrier sanctions, "safe third countries" and "readmission": the development of an asylum "buffer zone" in Europe. T I Brit Geogr 21: 76–90.

115. Neumayer E (2010) Visa restrictions and bilateral travel. The Professional Geographer: 37–41.

116. Crotts JC (2004) The Effect of Cultural Distance on Overseas Travel Behaviors. J Of Travel Res 43: 83–88. doi:10.1177/0047287504265516.

117. Bell M (1983) The estimation of an origin-destination matrix from traffic counts. Transport Sci 17: 198–217.

118. Yang H, Zhou J (1998) Optimal traffic counting locations for origin-destination matrix estimation. Transport Res B-Meth 32: 109–126.

119. Simini F, González MCM, Maritan A, Barabási A-L (2012) A universal model for mobility and migration patterns. Nature 484: 96–100. doi:10.1038/nature10856.

120. Abanyie FA, Arguin PM, Gutman J (2011) State of malaria diagnostic testing at clinical laboratories in the United States, 2010: a nationwide survey. Malaria J 10: 340. doi:10.1186/1475-2875-10-340.

121. Kain K, Harrington M, Tennyson S, Keystone J (1998) Imported malaria: prospective analysis of problems in diagnosis and management. Clin Infect Dis 27: 142–149.

122. Nilles EJ, Arguin PM (2012) Imported malaria: an update. Am J Emerg Med 30: 972–980. doi:10.1016/j.ajem.2011.06.016.

123. Widmer LL, Blank PR, Van Herck K, Hatz C, Schlagenhauf P (2010) Cost-effectiveness analysis of malaria chemoprophylaxis for travellers to West-Africa. BMC Infectious Diseases 10: 279. doi:10.1186/1471-2334-10-279.

124. Isaäcson M (1989) Airport malaria: a review. Bull World Health Organ 67: 737.

125. Thang H, Elsas R, Veenstra J (2002) Airport malaria: report of a case and a brief review of the literature. Neth J Med 60: 441–443.

126. Loupa C, Tzanetou K (2012) Autochthonous plasmodium vivax malaria in a Greek schoolgirl of the Attica region. Malaria J 11.

127. Wijesundera MDS (1988) Malaria outbreaks in new foci in Sri Lanka. Parasitology today (Personal ed) 4: 147–150.

128. Tatarsky A, Aboobakar S, Cohen JM, Gopee N, Bheecarry A, et al. (2011) Preventing the reintroduction of malaria in Mauritius: a programmatic and financial assessment. PloS one 6: e23832. doi:10.1371/journal.pone.0023832.

129. Lepers J, Deloron P, Fontenille D, Coulanges P (1988) Reappearnce of Falciparum Malaria in Central Highland Plateaux of Madagascar. Lancet 331: 586. doi:10.1016/S0140-6736(88)91375-X.

130. Le Menach A, Tatem AJ, Cohen JM, Hay SI, Randell H, et al. (2011) Travel risk, malaria importation and malaria transmission in Zanzibar. Scientific reports 1: 93. doi:10.1038/srep00093.

131. Fairhurst RM, Nayyar GML, Breman JG, Hallett R, Vennerstrom JL, et al. (2012) Artemisinin-resistant malaria: research challenges, opportunities, and public health implications. Am J Trop Med Hyg 87: 231–241. doi:10.4269/ajtmh.2012.12-0025.

132. Anderson TJC, Nair S, Nkhoma S, Williams JT, Imwong M, et al. (2010) High heritability of malaria parasite clearance rate indicates a genetic basis for

artemisinin resistance in western Cambodia. J Infect Dis 201: 1326–1330. doi:10.1086/651562.

133. Dondorp AM, Yeung S, White L, Nguon C, Day NPJ, et al. (2010) Artemisinin resistance: current status and scenarios for containment. Nature Rev Microbiol 8: 272–280. doi:10.1038/nrmicro2331.

134. Trape JF (2001) The public health impact of chloroquine resistance in Africa. Am J Trop Med Hyg 64: 12–17.

135. Mita T, Venkatesan M, Ohashi J, Culleton R, Takahashi N, et al. (2011) Limited geographical origin and global spread of sulfadoxine-resistant dhps alleles in Plasmodium falciparum populations. J Infect Dis 204: 1980–1988. doi:10.1093/infdis/jir664.

136. Naidoo I, Roper C (2010) Following the path of most resistance: dhps K540E dispersal in African Plasmodium falciparum. Trends Parasitol 26: 447–456. doi:10.1016/j.pt.2010.05.001.

137. World Health Organization (2010) Global report on antimalarial drug efficacy and drug resistance: 2000-2010. Geneva: WHO.

138. Huang Z, Wu X, Garcia AJ, Fik TJ, Tatem AJ (2013) An open-access modeled passenger flow matrix for the global air network in 2010. PloS One.

139. Howes RE, Patil AP, Piel FB, Nyangiri O a, Kabaria CW, et al. (2011) The global distribution of the Duffy blood group. Nat Commun 2: 266. doi:10.1038/ncomms1265.

140. Freeman L (1977) A set of measures of centrality based on betweenness. Sociometry.

141. Brandes U (2001) A Faster Algorithm For Betweenness Centrality. J Math Sociol: 37–41.

142. Fortunato S (2010) Community detection in graphs. Phys Rep.

143. Tatem AJ, Smith DL (2010) International population movements and regional Plasmodium falciparum malaria elimination strategies. Proc Natl Acad Sci 107: 12222–12227. doi:10.1073/pnas.1002971107.

144. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, et al. (2012) Quantifying the impact of human mobility on malaria. Science 338: 267–270. doi:10.1126/science.1223467.

145. Blondel V, Guillaume J (2008) Fast unfolding of communities in large networks. J Stat Mech Theor Exp: 1–12.

146. Bauer D (1972) Constructing confidence sets using rank statistics. J Amer Statist Assoc 67: 687–690.

147. UCSF Global Health Group (2011) Atlas of Malaria-Eliminating Countries. San Francisco.

148. Talisuna A, Karema C, Ogutu B (2012) Mitigating the threat of artemisinin resistance in Africa: improvement of drug-resistance surveillance and response systems. Lancet Infect Dis 12: 888–896. doi:10.1016/S1473-3099(12)70241-4.Mitigating.

149. O'Meara WP, Mangeni JN, Steketee R, Greenwood B (2010) Changes in the burden of malaria in sub-Saharan Africa. Lancet Infect Dis 10: 545–555. doi:10.1016/S1473-3099(10)70096-7.

150. Tatem AJ, Qiu Y, Smith DL, Sabot O, Ali AS, et al. (2009) The use of mobile phone data for the estimation of the travel patterns and imported Plasmodium falciparum rates among Zanzibar residents. Malaria J 8: 287. doi:10.1186/1475-2875-8-287.

151. Tatem AJ, Hemelaar J, Gray R, Salemi M (2012) Spatial accessibility and the spread of HIV-1 subtypes and recombinants in sub-Saharan Africa. AIDS.

152. Linard C, Gilbert M, Snow RW, Noor AM, Tatem AJ (2012) Population distribution, settlement patterns and accessibility across Africa in 2010. PloS One 7: e31743. doi:10.1371/journal.pone.0031743.

153. Cohen J, Woolsey A, Sabot O (2012) Optimizing Investments in Malaria Treatment and Diagnosis. Science 338: 612–614.

154. Flegg JA, Guerin PJ, White NJ, Stepniewska K (2011) Standardizing the measurement of parasite clearance in falciparum malaria: the parasite clearance estimator. Malaria J 10: 339. doi:10.1186/1475-2875-10-339.

BIOGRAPHICAL SKETCH

Mr. Zhuojie Huang is from Guangzhou, China. He did his undergrad study in his hometown and he finished his Master of Science in the department of geography, University of Florida. He can speak fluent Cantonese, Mandarin and English.