INSTITUTO DE HIGIENE E
MEDICINA TROPICAL
DESDE 1902

Universidade Nova de Lisboa

Instituto de Higiene e Medicina Tropical

Modeling of the Spatiotemporal Distribution Patterns and Transmission Dynamics of Dengue, for an Early Warning Surveillance System

Donald Salami

DISSERTAÇÃO PARA A OBTENÇÃO DO GRAU DE DOUTOR EM SAÚDE PÚBLICA GLOBAL

JULY 2021

INSTITUTO DE HIGIENE E
MEDICINA TROPICAL
DESDE 1902

Universidade Nova de Lisboa

Instituto de Higiene e Medicina Tropical

Modeling of the Spatiotemporal Distribution Patterns and Transmission Dynamics of Dengue, for an Early Warning Surveillance System

Donald Salami

Orientador: Prof. Dra. Carla A. Sousa

Coorientador: Prof. Dr César Capinha

Co-orientadora: Prof. Dra. Maria Rosário Oliveira Martins

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de de doutor em saúde pública global

Apoio financeiro de Fundação para a Ciência e Tecnologia, através da bolsa de doutoramento-PD/BD/128084/2016, e parte do financiamento de WARDEN project- PTDC/SAU-PUB/30089/2017

List of Publications

This thesis is based on three articles listed below, which will be referred to throughout the text by their specific citation:

1. Salami D, Capinha C, Martins MdRO, Sousa CA (2020) Dengue importation into Europe: A network connectivity-based approach. PLoS One 15(3): e0230274. https://doi.org/10.1371/journal.pone.0230274

2. Salami, D., Sousa, C.A., Martins, M.d.R.O. et al. Predicting dengue importation into Europe, using machine learning and model-agnostic methods. Sci Rep 10, 9689 (2020). https://doi.org/10.1038/s41598-020-66650-1

3. Salami D, Capinha C, Sousa CA, Martins MdRO, Lord C (2020) Simulation models of dengue transmission in Funchal, Madeira Island: Influence of seasonality. PLoS Neglected Tropical Diseases 14(10): e0008679. https://doi.org/10.1371/journal.pntd.0008679

Other publications achieved during the period of the PhD program, but not related to this thesis:

4. Salami D, Shaaban AN, Oliveira Martins MDR (2019). Africa Rising, a Narrative for Life Expectancy Gains? Evidence from a Health Production Function. Ann Glob Health. 2019;85(1):63. doi: https://doi.org/10.5334/aogh.2307

List of Communications

- Salami D, Capinha C, Martins MdRO, Sousa CA (2019). 'Dengue importation into Europe: A network connectivity-based approach'. Encontro Ciência, Lisboa, 2019. Poster Presentation.

- Salami D, Capinha C, Sousa CA, Martins MdRO, Lord C (2020). 'Simulation models of dengue transmission in Funchal, Madeira Island: Influence of seasonality'. 2nd AIM-COST Annual Conference of Aedes Invasive Mosquitoes (AIM-COST). Accepted for oral presentation.

“*Nothing is impossible the word itself says 'I'm possible'*”

Audrey Hepburn

Acknowledgements

Firstly, I want to express my profound gratitude to God the creator of life and love, for every divine set up and miracle, that has brought me thus far in this journey of life.

I would like to thank the Fundação para a Ciência e a Tecnologia for granting me the scholarship to make this dream come through. I also express gratitude to all the faculty members of the International Global Public Health Doctoral Programme (from NOVA University and the University of Porto), for allowing me to be part of the PhD program; having been admitted again into the 2nd edition of the program, after I could not make the first admission of the 1st edition, due to travel visa issues. I would also like to thank the Institute of Hygiene and Tropical Medicine (IHMT) for hosting me and providing me with the enabling environment to develop this thesis work. Likewise, I would like to express my gratitude to the WARDEN project for giving me the opportunity to work within the protocol of the project and for the funding support of all publications from this thesis.

I would like to thank my supervisors, Carla A Sousa, César Capinha, and Maria Rosário Oliveira Martins for all the support and freedom they have given me over the past three and half years. This thesis would not have been possible without your sage advice and guidance from the formative stages to its completion.

My thanks also go out to my fellow PhD colleagues from the 2nd Edition of the International Global Public Health Doctoral Programme, i.e. (Catarina, Daniel, Joana, Manuel, Maria, Mina, Nabil, Sandra, and Sousan – the adopted member of the team). You all were my first family here in Portugal, thanks for all the memorable times mostly during our first year and subsequently. Nabil, I am forever in your debt for all you have done for me in these past 4years (my true African brother); Mina, thanks for being a supportive friend when I needed it; Manu, Daniel, Joana, Maria & Sandra thanks for your valuable friendship through these years; Sousan, thanks for being there in my darkest time during this PhD.

I would like to thank Cynthia Lord at the University of Florida for allowing me to undergo a research exchange at the Florida Medical Entomological Laboratory. The training and knowledge

I gleaned from my 6 months in Florida and the subsequent months working with you on my 3rd paper are undoubtedly a fundamental piece to shaping this thesis and future career path.

I want to greatly appreciate Dominic Freienstein for his assistant in accessing the international air travel association, passenger intelligence services (IATA-PaxIS) data at "NO COST". You gave me the biggest miracle of this PhD journey; this thesis would not be possible without your great act of kindness.

Thanks to Pedro Leite, for not just being my flat-mate and friend in the past 3years, but for becoming my Papi P. Thanks also go to my extended Portuguese family, (i.e., Carlos, Miguel, Leila, Sofia, Claudia, Puppa, Fatima, Jose Beleza, Mami, Tatiane, Ilda, Joao Pontes, Isabel, Jorge, Teresa, Jose Luis, Aryadne, Eloisa, Alessandra, Niko) for all the lunch and dinner parties, you guys gave my life the social balance it needed. Thanks also go to my friends from 'The Lucian group' now turned family, i.e., Adeoluwa, Temitee, Bryn, Jide, Yommi, Dami, Debo, Urszula, Dosienka, our shared moments gave my life joy through these past years.

I would like to thank my family (Allison, Beatrice, Christopher, Fredrick, Musa, Sultan, Mariam, Abdul-Rahman, and Abdul-Raheem) and all my friends and for their unconditional love and support, I am a sum of your love and faith in me. Most importantly, Mrs. Christiana Imagbejor (my lovely mother), your forever love and support for me is what gives my life joy, meaning, and purpose. All my achievements are dedicated to you.

Finally, I would like to thank my inamorata (Ibitola), who against all odds continues to love and support me. You are the epitome of love, thanks for loving all my perfect imperfections.

Abstract

Emerging vector-borne diseases pose a significant global public health challenge. In recent times, outbreaks of diseases, such as dengue and chikungunya fever, have increased in frequency. This is facilitated by globalization, increase in trade and travel, and the spread of invasive vectors into new areas. In Europe, this is exemplified by the recent introduction and establishment of *Aedes* mosquito species and subsequent outbreaks of diseases like dengue. With the increasing spread of dengue worldwide, the European region has also experienced increase in reported cases - majority being travel related. Likewise, there has been an increase in sporadic events of autochthonous dengue transmission, in areas with established vector presence and favourable environmental conditions. Europe is currently faced with the challenge of assessing its importation risk of viraemic cases of dengue, and the probability of local transmission.

This thesis aims to study the dynamics of viraemic cases importation and virus transmission of dengue fever in Europe, namely in Madeira Island. This is achieved by establishing an importation and transmission modelling framework. The framework combines three sub-models: (i) a network connectivity importation model (ii) a machine learning predictive model and, (iii) a compartmental vector-host transmission model. The network connectivity and machine learning model were both parameterized using a historical dengue importation data for 21 countries in Europe, and indices that characterize important parameters for dengue importation: (i) the air passenger traffic, (ii) dengue activity and seasonality, (iii) incidence rate, (iv) geographical proximity, (v) epidemic vulnerability, and (vi) wealth of a source country. The transmission model was calibrated using empirical parameters for the mosquito life history traits, viral transmission, and temperature seasonality of Funchal, Madeira Island.

The results of the network connectivity and machine learning models demonstrate a higher importation risk of a viraemic case from source countries with high passenger traffic, high incidence rates, lower economic status, and geographical proximity to a destination country. The machine learning model achieved high predictive accuracy with an AUC score of 0.94. The transmission model demonstrates the potential for summer and autumn season transmission of dengue in Funchal, with the arrival date of the infectious person significantly affecting the distribution of the timing and peak size of the epidemic. Likewise, seasonal temperature variation

dramatically affects the epidemic dynamics, with warmer starting temperatures producing large epidemics with peaks occurring more rapidly.

The modelling framework described in this thesis has the potential to serve as an integrated early warning surveillance tool for dengue in Europe. This work provides practical guidance to assist public health officials in preventing outbreaks of dengue and reducing the risk of local transmission in areas with vectors presence. This framework could be applied to other *Aedes*-borne diseases such as chikungunya and yellow fever.

Keywords: Dengue, Spatiotemporal modelling, Early warning system, Europe.

Resumo

As doenças emergentes transmitidas por vetores representam um desafio significativo para a saúde pública global. Nos últimos tempos, os surtos de doenças como a dengue e a febre de chikungunya, aumentaram em frequência. Tal é facilitado pela globalização, pelo aumento do comércio e das viagens, e pela dispersão para novas áreas dos seus vetores invasores. Na Europa, este facto é exemplificado pela recente introdução e estabelecimento de espécies de mosquitos do género *Aedes* com a subsequente ocorrência de surtos de doenças como a dengue. Com a crescente disseminação da dengue em todo o mundo, a região europeia também tem vindo a registar um aumento de casos - a maioria destes relacionados com viagens. Da mesma forma, tem havido um aumento de eventos esporádicos de transmissão autóctone de dengue em áreas onde ocorre o vetor sob condições ambientais favoráveis. Assim, atualmente, a Europa enfrenta o desafio de avaliar o risco de importação de casos virémicos de dengue e a probabilidade de ocorrência de transmissão local deste vírus.

Esta tese visa contribuir para a compreensão dos fatores relacionados com a importação do vírus da dengue na Europa e a sua transmissão neste território, nomeadamente na ilha da Madeira. Para tal foi implementado uma estrutura integrada de modelos computacionais da importação e transmissão da doença. A estrutura combina três submodelos: (i) um modelo explicativo de importação da doença assente em teoria de redes (ii) um modelo preditivo de aprendizagem automática e, (iii) um modelo compartimental de transmissão vetor-hospedeiro. Os modelos de teoria de redes e de aprendizagem automática foram parametrizados com recurso a dados históricos referentes a estimativas de casos importados de dengue em 21 países na Europa e índices que caracterizam parâmetros com relevância na importação da dengue: (i) tráfego de passageiros aéreos, (ii) atividade e sazonalidade da dengue, (iii) taxa de incidência, (iv) proximidade geográfica, (v) vulnerabilidade à epidemia, e, (vi) contexto económico do país de origem. O modelo compartimental de transmissão foi calibrado com parâmetros empíricos referentes ao ciclo de vida do mosquito, à transmissão viral e à variação anual de temperatura do Funchal, na ilha da Madeira.

Os resultados dos modelos de teoria de redes e aprendizagem automática demonstram um maior risco de importação de casos virémicos de países com elevado tráfego de passageiros,

elevadas taxas de incidência, situação económica débil e com maior proximidade geográfica em relação ao país de destino. O modelo de aprendizagem automática alcançou elevada performance preditiva, com uma pontuação AUC de 0,94. O modelo compartimental de transmissão demonstra a existência de um potencial de transmissão da dengue no Funchal nos períodos de verão e outono, com a data de chegada da pessoa infeciosa a afetar significativamente a distribuição no tempo e tamanho do pico da epidemia. Da mesma forma, a variação sazonal da temperatura afeta dramaticamente a dinâmica da epidemia, em que temperaturas iniciais mais quentes levam a surtos de maiores proporções, com o pico de casos a ocorrer mais cedo.

A estrutura de modelação descrita nesta tese tem o potencial de servir como uma ferramenta integrada de vigilância de alerta precoce para a ocorrência de surtos de dengue na Europa. Este trabalho fornece orientação prática para auxiliar as autoridades de saúde pública na prevenção de surtos de dengue e na redução do risco de transmissão local, em áreas onde ocorrem os vetores. Essa estrutura, com os devidos reajustamentos, pode ser aplicada a outras doenças transmitidas por *Aedes*, como chikungunya e febre amarela.

Palavras-chave: Dengue, Modelação espaciotemporal, Sistema de alerta precoce, Europa.

Table of Contents

Contents

List of Figures

List of Tables

Abbreviations

AIC	Akaike Information Criterion
AUC	Area under the receiving operating characteristic curve
CEPII	Centre d'Études Prospectives et d'Informations Internationales
CV	Cross-validation
DENV	Dengue virus
DENV-1	Dengue virus serotype 1
DENV-2	Dengue virus serotype 2
DENV-3	Dengue virus serotype 3
DENV-4	Dengue virus serotype 4
DHARMa	Diagnostics for HierArchical Regression Models
ECDC	European Centre for Disease Prevention and Control
EU/EEA	European Economic Area member states
GDP	Gross Domestic Product
GLMM	Generalized Linear Mixed Model
IATA	International Air Transport Association
IHME	Institute for Health Metrics and Evaluation
LHS	Latin Hypercube Sampling
LIME	Local Interpretable Model-Agnostic Explanations
PaxIS	Passenger Intelligence Services
PDP	Partial Dependence Plots
QOI	Quantities of Interest
ROC	Receiving operating characteristic curve
SEIR	Susceptible, Exposed, Infectious, and Recovered
SIR	Susceptible, Infectious, and Recovered
SMOTE	Synthetic Minority Over-Sampling Technique
TESSy	The European Surveillance System
VBD	Vector-borne diseases
VHF	Viral Haemorrhagic Fever

WARDEN	WARning system for DENgue
WHO	World Health Organization
WNV	West Nile virus

1 Introduction

1.1 Emerging vector-borne diseases

Emerging vector-borne diseases poses a significant public health challenge, due to their increasing global spread and the complexity of their transmission patterns (1, 2). Vector-borne diseases are diseases caused by parasites, viruses and bacteria transmitted to humans via the means of other organisms, namely vectors (such as mosquitoes, sandflies, triatomine bugs, blackflies, ticks, tsetse flies, mites, snails, and lice.) (3, 4). Major global vector-borne diseases of humans, in terms of estimated or reported annual number of cases, includes malaria, schistosomiasis, dengue, lymphatic filariasis, onchocerciasis, leishmaniasis, Chagas disease, chikungunya, Zika, yellow fever and Japanese encephalitis (Table 1). Combined these major vector-borne diseases account for around 17% of the estimated global burden of communicable diseases and are responsible for more than 700 000 deaths annually (5). More than 80% of the world's population live in areas at risk from at least one major vector-borne disease, with more than half at risk from two or more (Figure 1).

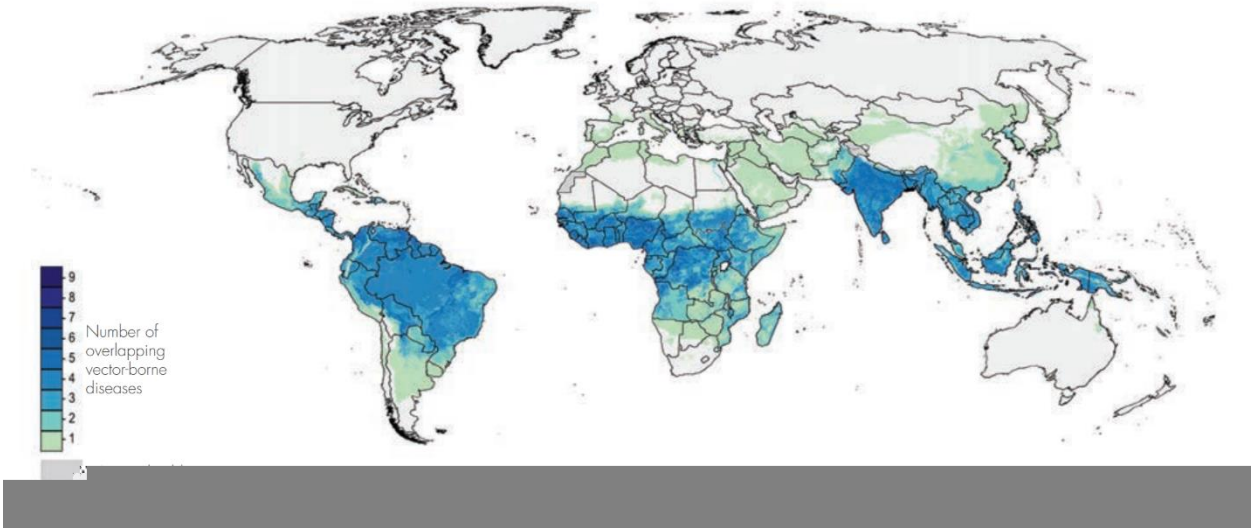

Figure 1-1. Overlapping global distribution of nine major vector-borne diseases (VBD) as of 2016.
Map shows countries/regions where the population is at risk of at least one of the following VBD: malaria, lymphatic filariasis, dengue, leishmaniasis, Japanese encephalitis, yellow fever, Chagas disease, human African trypanosomiasis, or onchocerciasis. Colours indicate the number of vector-borne diseases that pose a risk at each 5 × 5 km grid cell.
Source: https://apps.who.int/iris/bitstream/handle/10665/259205/9789241512978-eng.pdf?sequence=1.

increases in trade and transport networks, urbanization, changes in climatic and environmental factors (7, 8).

Some vector-borne diseases are now classified as emerging (or re-emerging), having met one of these criteria: (i) previously known disease, but now spreading to new geographic areas, or new populations, or reappearing after having been eradicated; or (ii) a disease occurring as a new variance of an already existing pathogen; or (iii) a disease that was previously unrecognized and unknown (4, 9). Examples of recent global outbreaks of vector-borne diseases include the emergence and rapid spread of West Nile virus (WNV) in North & South America and Europe (10); the unprecedented widespread epidemic of Zika fever starting in Brazil to other parts of South and North America (11); the re-emergence of the yellow fever virus in Angola and its rapid spread within several African countries, in South America and the Asia regions (12); and the global spread of chikungunya fever and dengue fever across the continents – Africa, Asia, North America, South America, Europe – increasingly becoming endemic in new regions (13-15). Each of these diseases has varying biological and epidemiological drivers, however, a commonality is that they are all transmitted by the same type of vector, i.e., mosquitoes.

Mosquito-borne diseases are diseases in which mosquitoes act as the primary transmitting vector from one host to another (16). Figure 2 below shows the significant resurgences of five major mosquito-borne diseases globally as of September 2019. The map displays all reported outbreaks, imported and autochthon cases of malaria, dengue fever, yellow fever, chikungunya fever, and Zika fever between 2017 and 2019. This figure clearly demonstrates the expanding spatial distribution of these diseases beyond the tropical countries/region. Notably some of the cases reported outside the tropics were imported cases, however some countries with established vector presence did report autochthon cases. The increasing spread of these diseases is primarily made possible following the introduction and successful establishment of invasive mosquitoes in new environments (17).

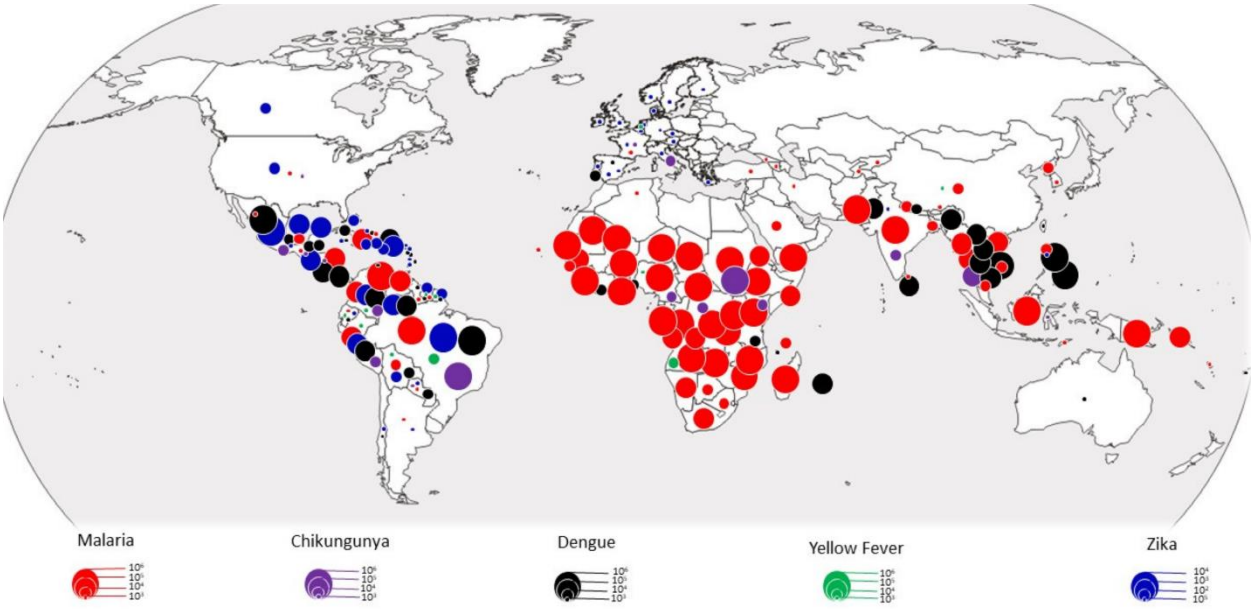

Figure 1-2. Spatial distribution of five mosquito-borne diseases globally.

Circle represents the reported number of cases (i.e. both autochthon and imported) between 2017 and 2019. Source: Handi Dahmana and Oleg Mediannikov. Mosquito-Borne Diseases Emergence/Resurgence and How to Effectively Control It Biologically. Pathogens 2020, 9(4), 310; https://doi.org/10.3390/pathogens9040310

1.2 Invasive mosquito vectors

Most invasive mosquito species belongs to the genus *Aedes*, particularly *Aedes albopictus* (Skuse, 1894), commonly referred to as the Asian tiger mosquito and *Aedes aegypti* (Linnaeus in Hasselquist, 1762), also referred to as the yellow fever mosquito (18, 19). Previously, *Aedes albopictus* and *Aedes aegypti* were geographically restricted to Southeast Asia and the African continent, respectively (19, 20), but currently, they have invaded almost all continents (21). These vectors have spread into new areas due to globalization and trade, notably transported in used tyres and plants (17, 22). Figures 3 and 4 shows the global distribution of both species as of 2015.

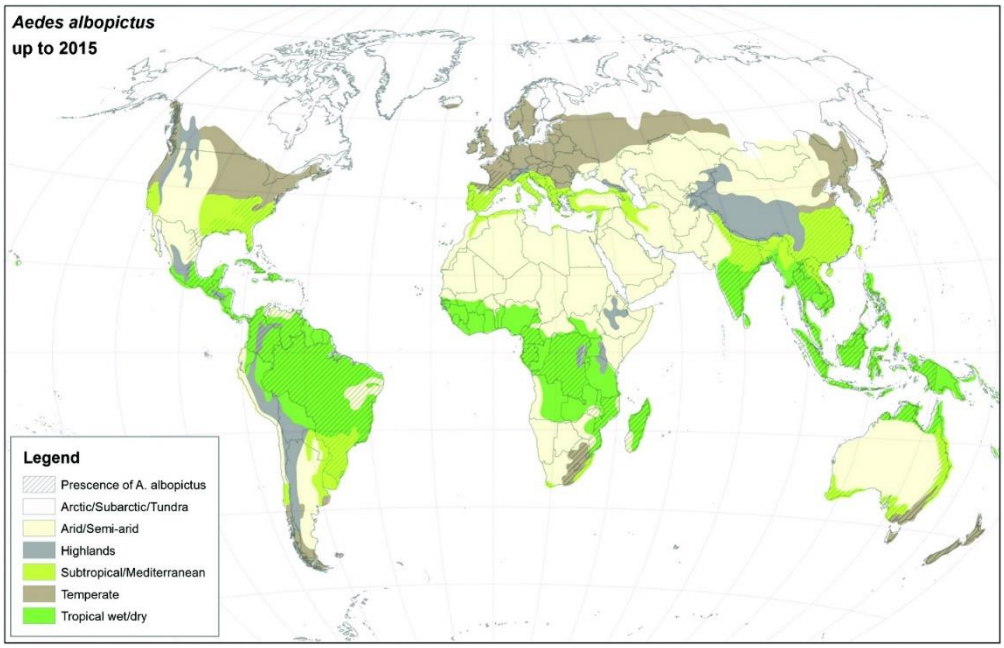

Figure 1-3. Global distribution of *Aedes albopictus* from climatic and surveillance data collected up to 2015.

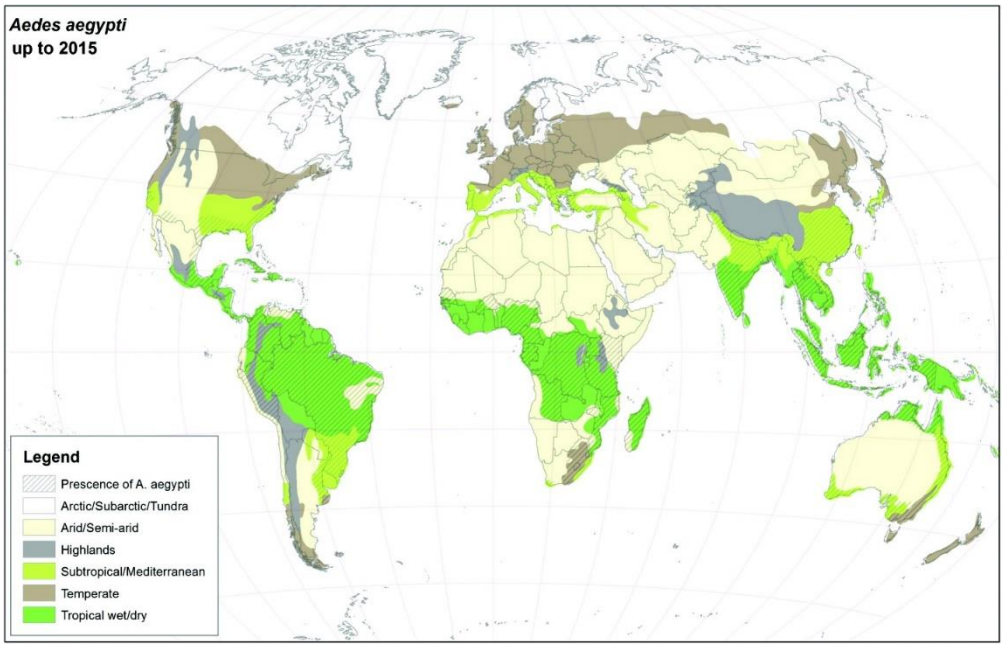

Figure 1-4. Global distribution of *Aedes aegypti* from climatic and surveillance data collected up to 2015.

Source for figures 3 & 4: Olivia Wesula Lwande, Vincent Obanda, Anders Lindström, Clas Ahlm, Magnus Evander, Jonas Näslund, and Göran Bucht.Vector-Borne and Zoonotic Diseases.Feb 2020.71-81. http://doi.org/10.1089/vbz.2019.2486

The European region exemplifies the introduction, spread and establishment of this species. *Aedes albopictus*, has greatly expanded its range over recent decades and is now firmly established in the Mediterranean Basin (Figure 5). Predictive models suggest that its geographical spread will continue to expand northwards due to climate change and the species ability to survive in regions with an average winter isotherm of 10ºC (23, 24). On the other hand, *Aedes aegypti*, has become established in Madeira (Portugal) and around the Black Sea (Georgia and the Russian Federation) (Figure 6). *Aedes aegypti* are closely adapted to human habitats and found in urban areas. The risk of its spread to northern Europe is low, due to its inability to survive in places with an average winter isotherm below 10ºC, however it could be re-established in southern Europe (23).

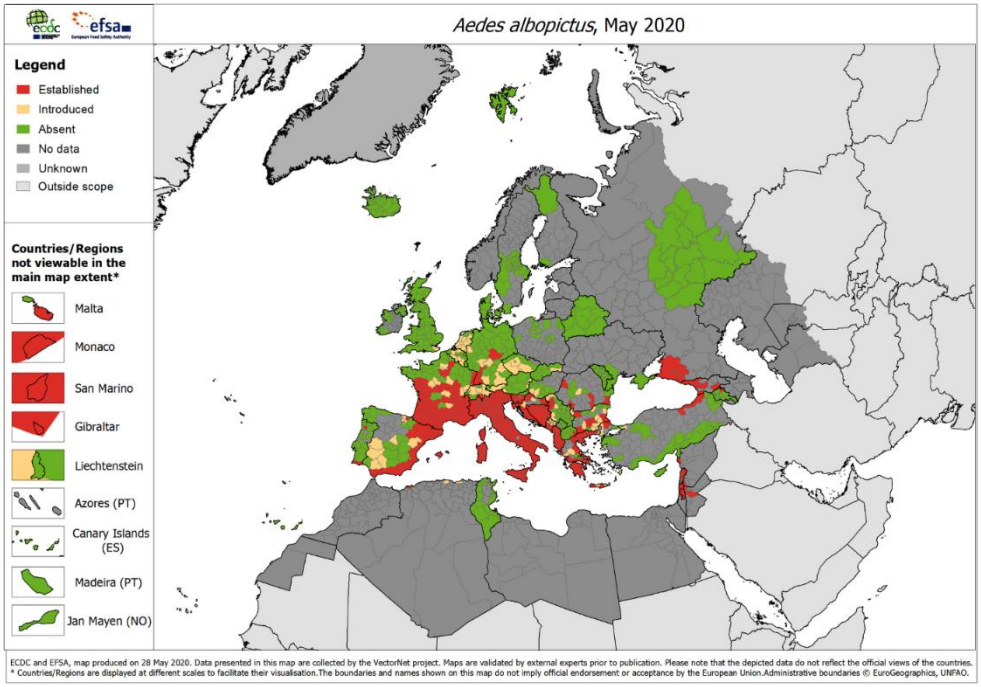

Figure 1-5. Current known distribution of *Aedes albopictus* in Europe (May 2020).

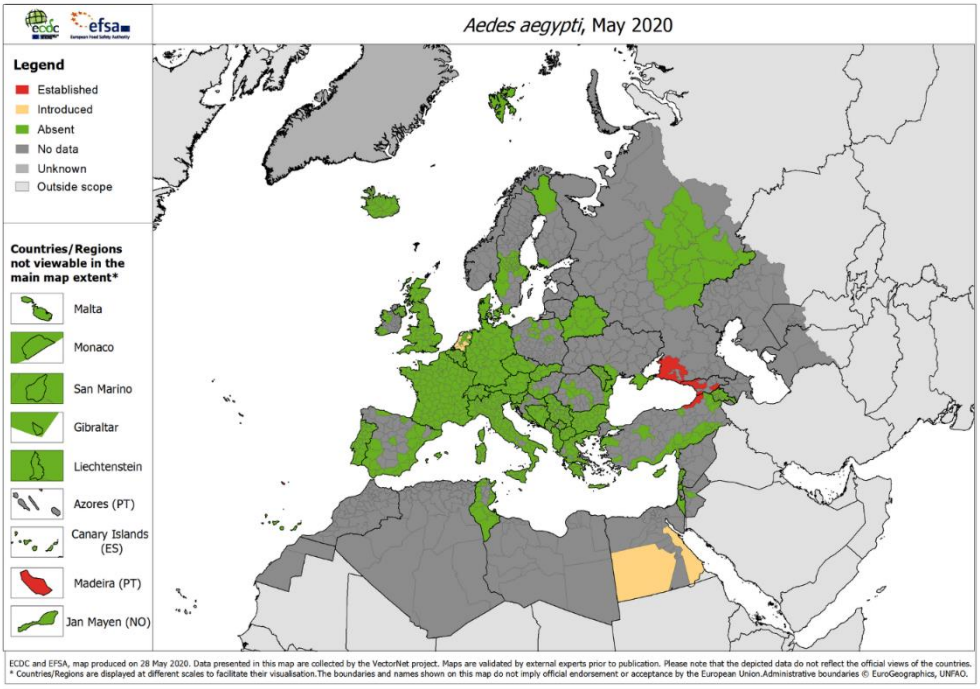

Figure 1-6. Current known distribution of *Aedes aegypti* in Europe (May 2020).

Source for Figures 5 & 6: European Centre for Disease Prevention and Control and European Food Safety Authority. Mosquito maps (internet). Stockholm: ECDC; 2020. Available from: https://ecdc.europa.eu/en/disease-vectors/surveillance-and-disease-data/mosquito-maps

1.3 Arboviral (Arthropod-borne Viral) diseases

Arboviral disease is a term used to describe diseases caused by a group of viruses which are transmitted to humans via the bite of infected arthropods (insects) such as mosquitoes (25). *Aedes albopictus* and *Aedes aegypti* are the main vectors for some arboviral (Arthropod-borne Viral) diseases. Common arboviral diseases transmitted by these species includes dengue, Zika, yellow fever, and chikungunya (25). Amongst these arboviral diseases, dengue has reported a rapid spread in the world, with a 30-fold increase in global incidence over the past 50 years (14) and an estimated 50 – 100 million dengue cases reported worldwide per year (26). Dengue poses serious concerns for global public health, given its increasing burden in endemic areas like in Africa, Asia, and South America (14, 27). Likewise, due to the risk of importation and subsequent autochthonous transmission in non-endemic areas, with established vector presence like in Europe

(28). With climate change, it is anticipated that novel and unusual outbreaks of dengue will continue with the progressive risk of endemicity in previously non-endemic areas (29).

1.4 Transmission of mosquito-borne arboviral disease

Mosquito-borne arboviral diseases are transmitted to human through the bite of an infected mosquito (serving as the disease vector). These vector-borne viruses (example dengue, Zika, yellow fever, or chikungunya) are transmitted through a human-to-mosquito-to-human cycle of transmission (Figure 7). The transmission cycle of the virus in the vector starts after an adult female mosquito feeds on the blood of an infected individual. Adult female mosquitoes require blood feeding for egg production, a key component of their life cycle as shown in the left panel of Figure 7 (30). Transmission of the virus from the human-to-mosquito occurs during the period of viremia, i.e., when the infected person has high levels of the virus in their blood steam. Following the blood feeding, the virus multiples and spread through the mosquito's body over a period of days, otherwise called the incubation (extrinsic) period (31). The duration of this period differs for different viruses, figure 8 exemplifies the incubation period for dengue. After this period, the infected mosquito can then transmit the virus to another human while feeding. The mosquito will remain infected with the virus for its entire life and will continue to transmit every time its bites a human host (30, 31).

Similarly, the susceptible human host undergoes an incubation (intrinsic) period after being bitten by an infected mosquito (Figure 8). Within this period (usually a couple of days) the individual will develop viremia, and eventually symptom of the disease. Usually, the human host would show no symptoms at the beginning of the viremia period but are infectious and can transmit the pathogen. The length of the infectious period is dependent on the specific virus, after which the human host is no longer able to transmit the infection during this current cycle (31).

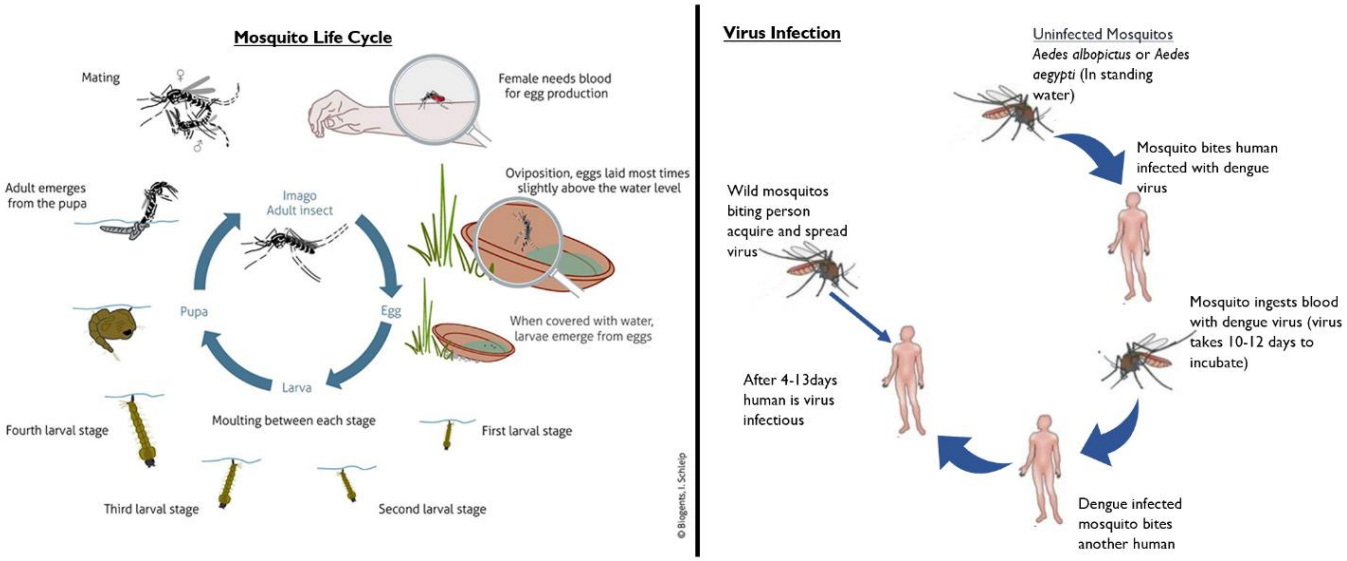

Figure 1-7. Transmission of a mosquito-borne disease (example of dengue).
Left panel of figure shows the mosquito life cycle, with two distinct stages, i.e., the immature stage (i.e. the eggs, the four larval stages, and metamorphosed pupas) and the adult stage. Right panel shows the virus infection/human-to-mosquito-to-human cycle of transmission.

Source: adapted and modified from https://eu.biogents.com/aedes-aegypti-yellow-fever-mosquitoes/ and Whitehead, S., Blaney, J., Durbin, A. et al. Prospects for a dengue virus vaccine. Nat Rev Microbiol 5, 518–528 (2007). https://doi.org/10.1038/nrmicro1690.

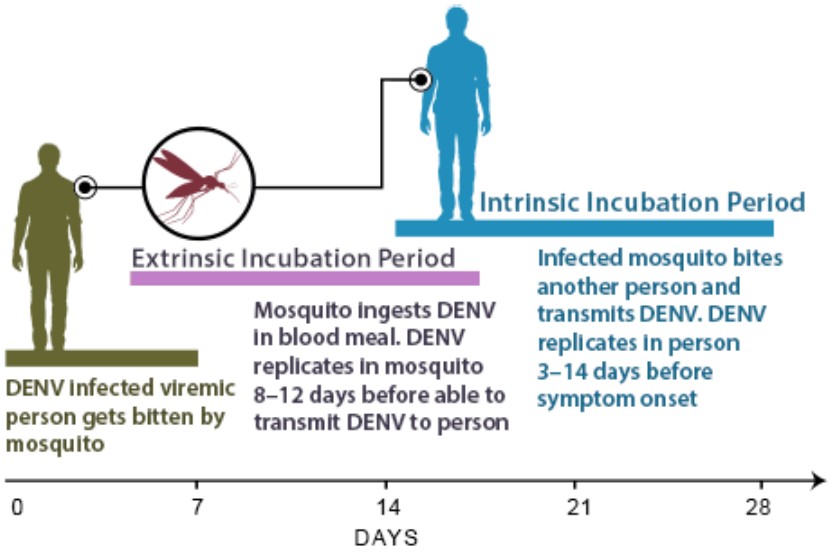

Figure 1-8. Duration of incubation period of dengue.

Source: https://www.cdc.gov/dengue/training/cme/ccm/page45915.html

The understanding of the multiple factors that contributes to the transmission cycle, is essential in the reduction and prevention of mosquito-borne diseases. A key component of this is the estimation of the vectorial capacity, which is defined as the daily rate at which future inoculations arise from a currently infective case, provided that all female mosquitoes biting that case becomes infected (32). The vectorial capacity combines different parameters that contributes to the ability of the mosquito to successfully transmit pathogens to the human host. These parameters includes the following: (i) vector density – in relation to the density of human host; (ii) human-biting rate – Average number of mosquito bites received by the human host in a unit time ; (iii) vector competence - an evaluation of the vector's capability (mechanical or biological) to transmit a pathogen; (iv) survival or mortality rates – of mosquitos per time; and (v) extrinsic incubation period – lag time between when the mosquito ingests the virus to when it becomes infectious (32) (Figure 9).

Changes in any of these components of the vectorial capacity can significantly affects the chances of transmission. For example, the biting and morality rates of mosquitoes are significantly affected by temperature changes; an increase in temperature can result in corresponding increase in the biting rate and a decrease in mortality rates or vice-versa. Other climatic (e.g., rainfall), environmental and anthropogenic factors can have significant effect on the components of the vectorial capacity and successively the rates of transmission. The understanding of the vectorial capacity and transmission of mosquito-borne diseases is fundamental for assessing local transmission risk, prevention of spread and vector control in areas with established vector presence.

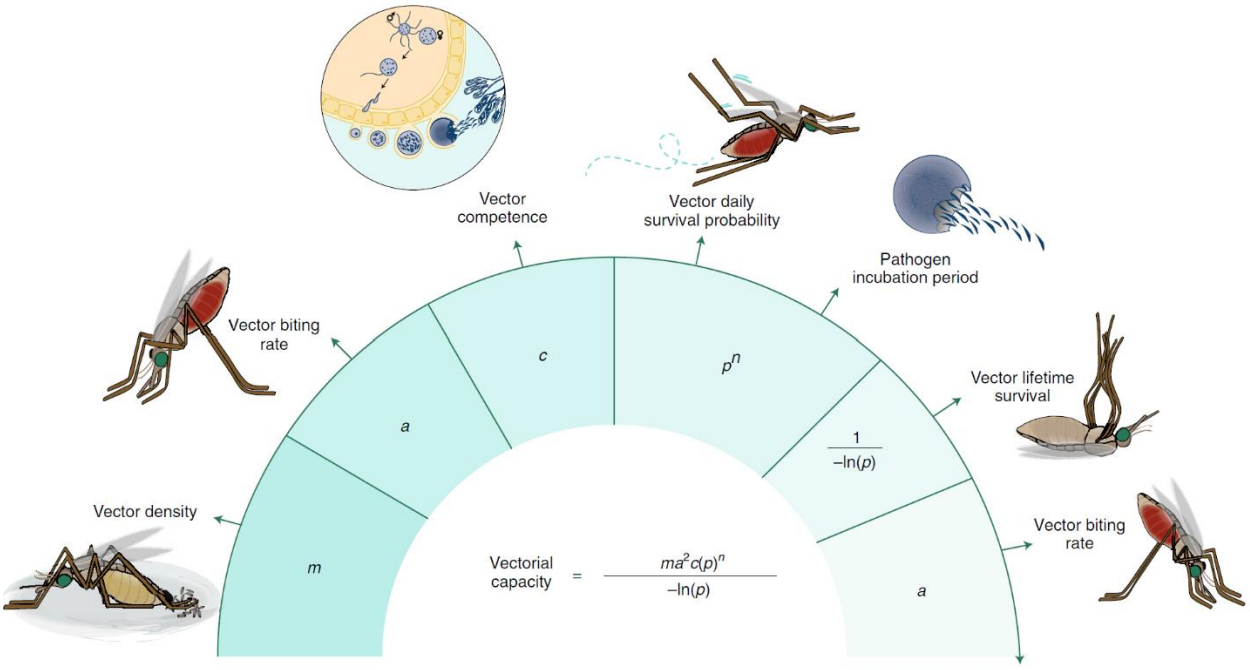

Figure 1-9. Components of vectorial capacity.

Source: Shaw, W.R., Catteruccia, F. Vector biology meets disease control: using basic research to fight vector-borne diseases. Nat Microbiol 4, 20–34 (2019). https://doi.org/10.1038/s41564-018-0214-7

1.5 Mosquito-borne disease risk in Europe

Currently, mosquito-borne diseases are not endemic to Europe, their main risk is travel-related, i.e., imported cases. Though over the past decade there have been some locally transmitted epidemics of dengue, chikungunya, and Zika in Europe (33-35).

Historically, dengue fever was endemic in southern European countries, where the vector *Aedes aegypti* was present. In the years 1927 and 1928, Greece experienced several outbreaks, with high incidence (over 1 million cases) and mortality (3000 deaths) (36, 37). After which the vector and the disease were extinct from the Europe region, but in recent years, the threat of dengue fever has re-emerged (28). Since 2010, the Europe has experienced incidence of sporadic autochthonous transmission in areas with established vector presence (of *Aedes albopictus* or *Aedes aegypti*) and favourable environmental conditions (33). Autochthonous cases have been

reported in Croatia in 2010 (38), France in 2010, 2013, 2014, 2015 and 2019 (39-42), Madeira Portugal' autonomous region, in 2012/2013 (43, 44), and Spain in 2018 and 2019 (41, 45-47). The outbreak in Madeira was the first sustained transmission since the 1920s; resulting in 1,080 confirmed cases and 78 cases reported as imported in 14 other European countries (43).

Similarly, Europe has reported recent autochthonous transmission of Zika and chikungunya. Three autochthonous cases of Zika virus were reported in France in 2019 (35). Likewise, local outbreaks of chikungunya virus were reported in Italy in 2007 and 2017 (48, 49) and France in 2010, 2014 and 2017 (39, 50, 51). These sporadic events of local Zika and chikungunya virus transmissions were in areas where *Aedes albopictus* is established (34).

These autochthonous events highlight the risk of the spread of dengue, chikungunya, and Zika in Europe, in areas with established vector presence and favourable environmental conditions. Consequently, the risk of introducing a viraemic case and triggering a local transmission is heightened with the ever-increasing global air travel traffic.

1.6 Dengue fever in Madeira Island

A pragmatic example of the introduction and establishment of *Aedes aegypti* mosquito, leading to local transmission of dengue; was demonstrated in Madeira Island, an autonomous region of Portugal. Madeira Island is an archipelago situated in the North Atlantic Ocean, around 400 kilometres to the north of the Canary Islands, 520 kilometres west of Morocco, and 1,000 kilometres from the European continent (Figure 10) (52). The island has an estimated population of 289,000 (53), and its capital city is Funchal, located on the south coast. The island's population is mostly distributed along the coastal line due to the mountainous interior of the island. Madeira Island presents a range of contrasting bioclimates because of its heterogeneous landscape and strong influence from the Gulf Stream and Canary current (Figure 10) (54). The southern coastal regions of the island (including Funchal), at low altitudes, have higher annual mean temperatures in comparison to the northern coastal regions or inland regions with higher altitudes and lower annual mean temperatures (Figure 10) (54, 55).

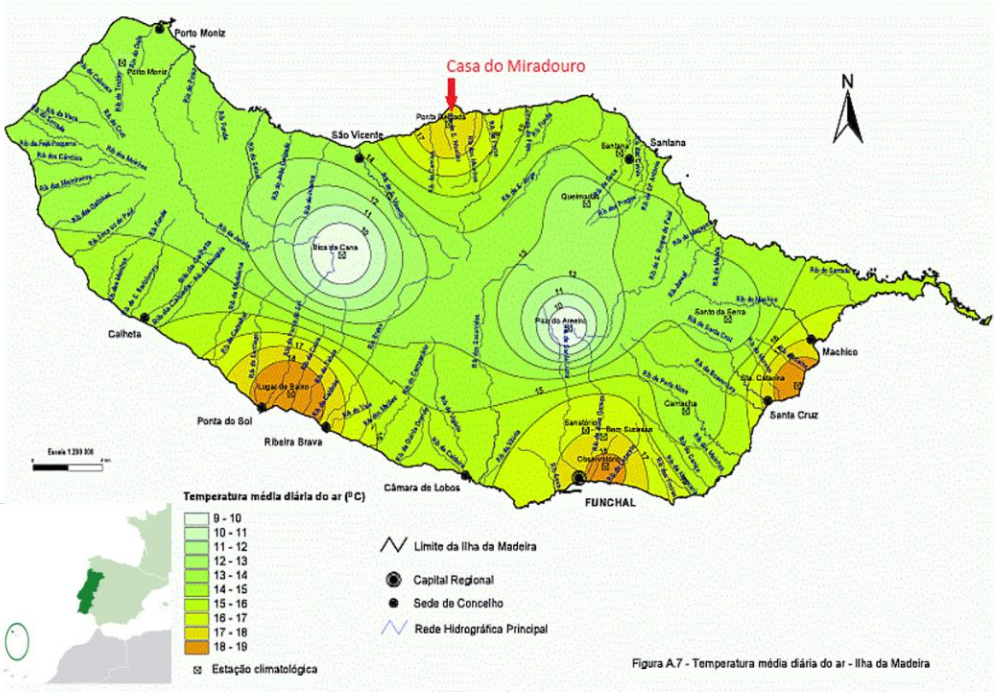

Figure 1-10. The temperature map of Madeira Island.
The map shows the annual average temperature of the island and the various municipalities of island (distributed along the coastal line). Box in the left bottom corner shows the location of Madeira Island (circled in dark green) in relation to Morocco (in grey) and Europe continental (light green, Portugal in dark green).
Source: https://www.madeiracasa.com/weather.php

The presence of *Aedes aegypti* mosquito was first detected in Funchal in 2005 and its presence was then recorded along the southern coast of the island (56). The mixture of densely populated urban areas (e.g., 45% of the island's population lives in Funchal), and an established habit of potting small plants and flowers, provides potential breeding sites and an ecological niche for the establishment of the mosquito on the island. In October 2012, the island reported its first dengue outbreak; the outbreak constituted a significant public health event because of the following: (i) it was the first sustained transmission of dengue in the European region since the 1920s; (ii) the total number of confirmed cases – 1080 and the subsequent 78 imported cases reported in 14 other European countries in travellers returning from Madeira (33, 43). Madeira island has not reported any locally transmitted dengue cases since the outbreak in 2012, however, it is at risk of another outbreak, in the event of an introduction of viraemic cases, given the presence of the *Aedes aegypti* mosquito and suitable environmental conditions. With an increase in co-

circulation of all dengue serotypes worldwide (57, 58), and Madeira's increasing lure as a popular year-round travel destination, the risk of another outbreak with a likely spread to continental Europe appears high.

1.7 Surveillance systems for dengue

With the increasingly global spread of dengue fever and other mosquito-borne diseases, the risk for Europe is heightened (22, 59). This risk can be attributed to increases in travel-related cases and changes in climatic and environmental factors providing more suitability for the spread and establishment of the vectors (i.e., *Aedes aegypti* and *Aedes albopictus*) (60). It is therefore imperative to develop integrated systems to monitor importation risk of viraemic cases and reduce the risk of their transmission in areas where invasive mosquitoes have become firmly established (61). To achieve this its crucial to understand two major events in the spread of dengue in Europe: (i) the importation pathways of viraemic cases and factors influencing it and (ii) the subsequent transmission dynamics of autochthonous dengue cases, in the event of a successful introduction of a viraemic case.

1.7.1 Dengue importation modelling

To understand the importation pathways of dengue, traditional epidemiology studies have relied on the modelling (regression analysis) of factors that affect the variability of the number of imported cases to provide insights into the burden and spread. However, in recent times, spatial statistics is gaining popularity in epidemiology studies, due to the development and application of novel computational techniques allowing for the analysis of large spatiotemporal data (62, 63). Spatio-temporal modelling involves building theories, testing them against available data, quantifying the uncertainties remaining, and informing about subsequent modelling and measurement requirements (62). This modelling framework is fitting to disease data collected across time and space, with spatial and temporal properties. Example, an imported case of dengue in a spatiotemporal dataset describes a spatial and temporal event of an imported viraemic case of

dengue at a certain time (t) and location (x). The inclusion of spatial variations helps quantify the uncertainties in spatial and spatio-temporal distributions in relation to the imported case, thereby, providing evidence of patterns of dependence and level of noise in data, that otherwise would not be captured in a simple regression.

One notable challenge in this kind of analysis is accounting for temporal and spatial correlations, which adds significant complexity to the data analysis process (64). In the instance of modelling importation of dengue, these correlations will be introduced when considering the unpredictable pattern of human movement, transportation networks and geographical (or spatial) variations to evaluate the importation pathways for different locations. Likewise, the influence of collocated neighbouring spatiotemporal objects (i.e., neighbouring locations) on one another would introduce some level of correlations. Hence to adequately model importation pathways of viraemic dengue cases, accounting for the correlations from the connectivity of the global transport network and human mobility is an integral first step.

Previous studies have utilized a common framework to explain importation; this framework assumes the probability of an imported case is directly correlated with the number of arriving air passengers (59, 65). However, a simple correlation between imported cases and arriving passengers' statistics is insufficient to explain the importation pathways of viraemic cases, as it does not account for the spatiotemporal correlations and the co-dynamics of the air transport network that influences importation (65).

One of the spatiotemporal frameworks currently used to address spatiotemporal correlations is the generalized linear mixed model (GLMM) (66). This model extends the generalized linear model to include a random effect to account for spatial dependence and to provide optimal spatial predictions, in additional to the fixed effects (66). The GLMM models can help study the spatial the relationship between imported case of dengue, temporal, and spatial correlations, in the context of other factors influencing importation (e.g., incidence rate, seasonality, number of passengers, geographical proximity), either by controlling for these factors or considering interactions with them. However, in addition to the GLMM framework, a dynamic network autocorrelation that explicitly incorporates the influence of the air transport network topology is required to adequately reconstruct the importation pathways of viraemic cases of dengue. The inclusion of network topology will provide insights to how the positioning of

countries within the network influence the propensity for imports. (67). This can be achieved by incorporating a network autocorrelation the GLMM framework to address the issue of covariance driven by network structure.

1.7.2 Transmission modelling

On the other hand, disease transmission modelling framework are mechanistic in nature. These models make it possible to describe the time evolution of an epidemic in mathematical terms, by connecting the individual stages of transmission to understand disease outcomes (add citation). Mathematical modelling of disease was pioneered by Kermack and McKendrick in 1927 (68, 69); they formulated a compartmental model, which are governed by a system of differential equations to describe the dynamics of disease transmission. The base compartmental model, categorize individuals into three main compartments i.e. Susceptible (S) - individuals not yet infected with the disease at time (t), or those susceptible to the disease of the population; Infected (I) - individuals of the population who have been infected with the disease and are capable of spreading the disease to those in the susceptible class; or Recovered and immune (R) - individuals of the population who have been infected and then removed from the disease, either due to immunization, recovered or death (70). The mathematical formulation of these compartments and their dependencies, analysing all the dynamic processes that contribute to disease transmission in each class. These models are largely dependent on the essential processes involved in transmission of the disease and the parameters that are most influential.

To model the transmission of a vector-borne disease like dengue, the basic SIR model, needs to be modified to include the transmission cycle of the vector and interactions with the human. The modified model will be a SEI-SEIR transmission model. The SEI component will describes the vector population, represented as susceptible (S_v), exposed (E_v), and infectious (I_v), while the SEIR component will describes the human population represented as susceptible (S_h), exposed (infected but not infectious) (E_h), infectious (I_h), and recovered (immune) (R_h). Governed by differential equations this model will adequately describe the dynamics of dengue transmission in both the vector and human host compartments. This modelling framework will be suited for understanding of the dynamics of autochthonous cases in areas with vector presence in Europe.

The model domain will be parametrized to a suitable spatial area with defined processes and conditions favouring transmission of dengue, which includes vector abundance, ecological, climatic, and other anthropogenic factors. For suitable areas, the introduction of a viremia case (via importation) increases the probability of triggering a sustained local transmission (for example in Madeira Island). Likewise, the redefinition of average annual mean temperatures and anomalies due to climate change might increase the environmental and climatic suitability for the vector (*Aedes aegypti*) (57, 71). The compartmental modelling approach will allow for the integration of empirical and mechanistic parameters to simulate a potential vector-host transmission, and the resulting epidemic dynamics (72, 73).

The combination of both modelling approaches (i.e., importation and transmission) described above would pave the way for the development of an integrated surveillance framework, that extends the understanding of dengue importation and transmission in non-endemic regions, like Europe. This framework will utilize a combination of air transport network data, entomological surveillance data, epidemiological surveillance data, and risk factor identification data. Thereby, setting the foundational framework for an early warning system to forecast and monitor outbreaks of dengue in Europe.

1.8 Aim and objectives

This thesis aims to add to the understanding of the importation patterns and transmission dynamics of dengue fever in Europe. To achieve this aim, we integrate statistical network modelling, machine learning algorithms, and deterministic transmission modelling approaches, to facilitate the generation of a novel modelling framework for assessing importation risk of a viraemic case and the probability of local transmissions of dengue.

The main objectives of this thesis are:

1. To develop a statistical network model, to analyse the importation patterns of dengue in Europe, the different mediating factors, and the influence of the air transport network.
2. To develop a prediction model for the importation risk of a viraemic case of dengue into Europe, with intuitive explanations for prediction.
3. To develop a compartmental vector-host deterministic transmission model, to explore the local transmission and epidemic dynamics of dengue, through the influence of seasonality.

Each objective was addressed in an individual published paper, with their respective background, methods, results, and discussion sections. The published papers are delineated in the result chapter of this thesis. The next section summarizes the methodological approach utilized in addressing each objective.

1.9 Summary of methodological approach

1.9.1 Objective 1

Objective 1 was achieved via the implementation of a refined network connectivity approach, which incorporates network autocorrelation within a generalized linear mixed model (GLMM), to examine factors mediating the importation of dengue. The methodological steps to this approach include: First, connectivity indices between a source country and a destination country in Europe were developed, using different factors that potentially mediate the importation risk of a viraemic case of dengue. The indices were decomposed into two components the source strength' and the transport or importation potential. Firstly, the source strength for all indices was modelled to represent the endemicity of dengue and the risk of dengue infection in a source country. Secondly, the transport and importation potential were modelled to characterize seasonal dengue activity, incidence rates, geographical proximity, epidemic vulnerability, air passenger volume, population size, and wealth of a source country as mediating risk factors. The indices were developed using histological data for 6 years, i.e., 2010 – 2015.

Second, 72 weighted directed networks were constructed using the monthly air passenger's data, to represent the monthly flow of air passengers into Europe, from 2010 to 2015. The network for each month was denoted by $G_m = (V_G, E_G)$, where V_G is a set containing all the nodes (or vertices), while E_G contains all the edges, with m indicating the month (m = 1, 2, 3… 72, covering the years of 2010–2015). Nodes represented all countries worldwide, while edges represent the flow of passengers from a source country to a destination country in Europe. Four different network centrality measures were then calculated to analyse the network and quantify the capacity of a source node to influence transportation of dengue or a destination node's propensity to receive an imported case of dengue, by virtue of their connection topology within the network.

Finally, a generalized linear mixed effect model (GLMMs) was fitted to quantify the variation in dengue importation as explained by the connectivity indices and network centrality measures. The GLMM was fitted with logit link functions for binomial errors and fixed effects of the connectivity indices, network centrality measures, time, and crossed random effects

(intercepts) of the destination country. The goodness of fit was evaluated by calculating the marginal and conditional GLMM R^2 and DHARMa residual diagnostics for hierarchical models.

The detailed paper (background, methods, results, and discussion) is outlined in the section 2.1 of the results chapter.

1.9.2 Objective 2

Objective 2 was achieved via the application of machine learning algorithms to develop a predictive model for the importation risk of a viraemic case of dengue into Europe. A diverse set of machine learning algorithms were trained, using historical data of dengue importation into Europe, connectivity indices, and network centrality measures. The choice of the suite of algorithms tested was a trade-off between, meta-algorithm that fits the classification problem and with built-in feature selection. Other algorithmic and systematic features that were considered include regularization (to handle the effects of multicollinearity), hyperparameter optimization (model tuning capabilities), and efficient computation time. The training model compared four widely used classifiers algorithms in machine learning: Partial least squares; Lasso and elastic-net regularized generalized linear models; Random forests and Extreme gradient boosting.

Candidate models were evaluated using the area under the receiving operating characteristic curve (AUC), sensitivity (true positive rate), and specificity (false positive rate). The final candidate model was selected based on the receiving operating characteristic curve (ROC) threshold, which maximizes the trade-off between sensitivity and specificity.

Model-agnostic methods were then applied to provide an interpretation of the predictions made by the model achieving higher predictive performance. Model-agnostic methods work by extracting post-hoc explanations from an original machine learning model. Variable importance and partial dependence plots were used to provide global interpretation (i.e., the modelled relationship and distribution of the predicted target outcome based on the input variables); while local surrogate models, called ‘Local interpretable model-agnostic explanations (LIME)’, were used to provide local interpretations (understand model predictions for a single instance i.e. a single unit of observation).

The detailed paper (background, methods, results, and discussion) is outlined in the section 2.2 of the results chapter.

1.9.3 Objective 3

Objective 3 was achieved via the implementation of a standard compartmental vector-host (SEI-SEIR) deterministic model for the transmission of dengue by Aedes aegypti mosquitoes. With the model domain set at the spatial area of Funchal, Madeira Island; the model integrates empirical and mechanistic parameters for mosquito life history traits, virus transmission, and seasonality of Funchal. The SEI component describes the vector population (i.e., *Aedes aegypti* mosquitoes), represented as susceptible (S), exposed (E), and infectious (I); while the SEIR component describes the human population represented as susceptible (S), exposed (infected but not infectious) (E), infectious (I), and recovered (immune). The model is governed by differential equations with time-varying entomological parameters and seasonality. The models assume a dynamic life-history trait of the mosquitos.

The adult female mosquitoes' population is recruited into the model by a Gaussian distribution curve. This is fitted to reflect the seasonal pattern of mosquitoes in Madeira Island as observed from the *Aedes aegypti* mosquito entomological surveillance data. Seasonal forcing (seasonality) was then incorporated into the framework, via the sinusoidally forcing, the daily mean temperature was modelled as a cosine curve with a period of 365 days. Temperature-dependence was introduced into the model by using fitted thermal response curves for some of the mosquito life-history traits, i.e., biting rate, extrinsic incubation period, and mortality rates. Constant parameters values were utilized for human components of the transmission cycle since they are not seasonal. The model parameter value/range was then based on scientific literature on empirical studies or lab trials and expert knowledge. Overall, model starting conditions and parameter values were set based on their permissibility for transmission.

Each model simulation is then analysed for the following quantities of interest (QOI): the epidemic peak size; time to peak infection in humans; and the final epidemic size, which represents a measure of epidemic suitability. The model parameters were then characterized based on their influence on the quantities of interest, via variance-based global sensitivity analysis, using a

combination of Latin hypercube sampling (LHS) and a multi-model inference on regression-based models.

The detailed paper (background, methods, results, and discussion) is outlined in the section 2.3 of the results chapter.

1.10 References

1. Kilpatrick AM, Randolph SE. Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. The Lancet. 2012;380(9857):1946-55. doi: https://doi.org/10.1016/S0140-6736(12)61151-9.

2. Swei A, Couper LI, Coffey LL, Kapan D, Bennett S. Patterns, drivers, and challenges of vector-borne disease emergence. Vector-Borne and Zoonotic Diseases. 2019;20(3):159-70. doi: https://doi.org/10.1089/vbz.2018.2432.

3. World Health Organization. A global brief on vector-borne diseases. Geneva, Switzerland: 2014 Contract No.: WHO/DCO/WHD/2014.1. Available from: http://apps.who.int/iris/bitstream/10665/111008/1/WHO_DCO_WHD_2014.1_eng.pdf

4. World Health Organization. Vector-borne diseases-fact sheets 2020 (cited 29 June 2020). Available from: https://www.who.int/en/news-room/fact-sheets/detail/vector-borne-diseases.

5. World Health Organization. Global vector control response 2017–2030. Geneva: 2017. Available from: https://apps.who.int/iris/bitstream/handle/10665/259205/9789241512978-eng.pdf?sequence=1.

6. Weaver SC, Reisen WK. Present and future arboviral threats. Antiviral Research. 2010;85(2):328-45. doi: https://doi.org/10.1016/j.antiviral.2009.10.008.

7. Metcalf CJE, Walter KS, Wesolowski A, Buckee CO, Shevliakova E, Tatem AJ, et al. Identifying climate drivers of infectious disease dynamics: Recent advances and challenges ahead. Proceedings of the Royal Society B: Biological Sciences. 2017;284(1860):20170901. doi: https://doi.org/10.1098/rspb.2017.0901.

8. Rogers DJ, Randolph SE. Climate change and vector-borne diseases. In: Hay SI, Graham A, Rogers DJ, editors. Advances in parasitology. 62: Academic Press; 2006. p. 345-81.

9. European Centre for Disease Prevention and Control. Emerging and vector-borne diseases programme 2020 (cited 29 June 2020). Available from: https://ecdc.europa.eu/en/about-us/who-we-are/disease-programmes/emerging-and-vector-borne-diseases-programme.

10. Kilpatrick AM. Globalization, land use, and the invasion of west Nile virus. Science (New York, NY). 2011;334(6054):323-7. doi: https://doi.org/10.1126/science.1201010.

11. Duffy MR, Chen T-H, Hancock WT, Powers AM, Kool JL, Lanciotti RS, et al. Zika virus outbreak on Yap Island, Federated states of Micronesia. New England Journal of Medicine. 2009;360(24):2536-43. doi: https://doi.org/10.1056/NEJMoa0805715.

12. Bryant JE, Holmes EC, Barrett ADT. Out of Africa: A molecular perspective on the introduction of yellow fever virus into the Americas. PLoS Pathogens. 2007;3(5): e75. doi: https://doi.org/10.1371/journal.ppat.0030075.

13. Pialoux G, Gaüzère B-A, Jauréguiberry S, Strobel M. Chikungunya, an epidemic arbovirosis. The Lancet Infectious Diseases. 2007;7(5):319-27. doi: https://doi.org/10.1016/S1473-3099(07)70107-X.

14. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013;496(7446):504-7. doi: https://doi.org/10.1038/nature12060.

15. Endy TP, Weaver SC, Hanley KA. Dengue virus - past, present and future. In: Hanley KA, Weaver SC, editors. Frontiers in dengue virus research. Texas, USA: Caister Academic Press; 2010. p. viii + 304.

16. World Health Organization. Mosquito-borne diseases 2020 (cited 29 June 2020). Available from: https://www.who.int/neglected_diseases/vector_ecology/mosquito-borne-diseases/en/.

17. Medlock JM, Hansford KM, Schaffner F, Versteirt V, Hendrickx G, Zeller H, et al. A review of the invasive mosquitoes in Europe: Ecology, public health risks, and control options. Vector borne and zoonotic diseases (Larchmont, NY). 2012;12(6):435-47. doi: https://doi.org/10.1089/vbz.2011.0814.

18. Bonizzoni M, Gasperi G, Chen X, James AA. The invasive mosquito species aedes albopictus: Current knowledge and future perspectives. Trends in Parasitology. 2013;29(9):460-8. doi: https://doi.org/10.1016/j.pt.2013.07.003.

19. Lwande OW, Obanda V, Lindström A, Ahlm C, Evander M, Näslund J, et al. Globe-trotting *Aedes aegypti* and *Aedes albopictus*: Risk factors for arbovirus pandemics. Vector-Borne and Zoonotic Diseases. 2019;20(2):71-81. doi: https://doi.org/10.1089/vbz.2019.2486.

20. Mousson L, Dauga C, Garrigues T, Schaffner F, Vazeille M, Failloux A-B. Phylogeography of *Aedes (stegomyia) aegypti (l.)* and *Aedes (stegomyia) Albopictus (skuse) (diptera: Culicidae)* based on mitochondrial DNA variations. Genetical Research. 2005;86(1):1-11. doi: https://doi.org/10.1017/S0016672305007627.

21. Kraemer MUG, Sinka ME, Duda KA, Mylne AQN, Shearer FM, Barker CM, et al. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. Albopictus*. eLife. 2015;4:e08347-e. doi: https://doi.org/10.7554/eLife.08347.

22. Caminade C, McIntyre KM, Jones AE. Impact of recent and future climate change on vector-borne diseases. Annals of the New York Academy of Sciences. 2019;1436(1):157-73. doi: https://doi.org/10.1111/nyas.13950.

23. World Health Organization. Dengue: guidelines for diagnosis, treatment, prevention, and control -- New edition. A joint publication of the World Health Organization (WHO) and the Special Programme for Research and Training in Tropical Diseases (TDR) (20 Aug 2020). Available from: https://www.who.int/tdr/publications/documents/dengue-diagnosis.pdf.

24. Fischer D, Thomas SM, Niemitz F, Reineking B, Beierkuhnlein C. Projection of climatic suitability for *Aedes albopictus skuse (culicidae)* in Europe under climate change conditions. Global and Planetary Change. 2011;78(1):54-64. doi: https://doi.org/10.1016/j.gloplacha.2011.05.008.

25. Leta S, Beyene TJ, De Clercq EM, Amenu K, Kraemer MUG, Revie CW. Global risk mapping for major diseases transmitted by *Aedes aegypti* and *Aedes albopictus*. International Journal of Infectious Diseases. 2018; 67:25-35. doi: https://doi.org/10.1016/j.ijid.2017.11.026.

26. Stanaway JD, Shepard DS, Undurraga EA, Halasa YA, Coffeng LE, Brady OJ, et al. The global burden of dengue: An analysis from the global burden of disease study 2013. Lancet Infect Dis. 2016;16(6):712-23. doi: https://doi.org/10.1016/s1473-3099(16)00026-8.

27. Messina JP, Brady OJ, Golding N, Kraemer MUG, Wint GRW, Ray SE, et al. The current and future global distribution and population at risk of dengue. Nature Microbiology. 2019. doi: https://doi.org/10.1038/s41564-019-0476-8.

28. Schaffner F, Medlock JM, Bortel WV. Public health significance of invasive mosquitoes in Europe. Clinical Microbiology and Infection. 2013;19(8):685-92. doi: https://doi.org/10.1111/1469-0691.12189.

29. Ryan SJ, Carlson CJ, Mordecai EA, Johnson LR. Global expansion and redistribution of aedes-borne virus transmission risk with climate change. PLoS Negl Trop Dis. 2019;13(3): e0007213. doi: https://doi.org/10.1371/journal.pntd.0007213.

30. Whitehead SS, Blaney JE, Durbin AP, Murphy BR. Prospects for a dengue virus vaccine. Nature Reviews Microbiology. 2007;5(7):518-28. doi: https://doi.org/10.1038/nrmicro1690.

31. Carrington LB, Simmons CP. Human to mosquito transmission of dengue viruses. Frontiers in immunology. 2014; 5:290-. doi: https://doi.org/10.3389/fimmu.2014.00290.

32. Dye C. Vectorial capacity: must we measure all its components? Parasitol Today. 1986 Aug;2(8):203-9. doi: https://doi.org/10.1016/0169-4758(86)90082-7.

33. European Centre for Disease Prevention and Control. Autochthonous transmission of dengue virus in EU/EEA, 2010-2020 2020 (20 Aug 2020). Available from: https://www.ecdc.europa.eu/en/all-topics-z/dengue/surveillance-and-disease-data/autochthonous-transmission-dengue-virus-eueea.

34. European Centre for Disease Prevention and Control. Autochthonous transmission of chikungunya virus in EU/EEA, 2007–2020 2020 (20 Aug 2020). Available from: https://www.ecdc.europa.eu/en/all-topics-z/chikungunya-virus-disease/surveillance-and-disease-data/autochthonous-transmission.

35. European Centre for Disease Prevention and Control. Epidemiological update: Third case of locally acquired zika virus disease in Hyères, France. 2019. https://www.ecdc.europa.eu/en/news-events/epidemiological-update-third-case-locally-acquired-zika-virus-disease-hyeres-france.

36. Rosen L. Dengue in Greece in 1927 and 1928 and the pathogenesis of dengue hemorrhagic fever: New data and a different conclusion. The American Journal of Tropical Medicine and Hygiene. 1986;35(3):642-53. doi: https://doi.org/10.4269/ajtmh.1986.35.642.

37. Kuno G. Emergence of the severe syndrome and mortality associated with dengue and dengue-like illness: Historical records (1890 to 1950) and their compatibility with current hypotheses on the shift of disease manifestation. Clinical microbiology reviews. 2009;22(2):186-201. doi: https://doi.org/10.1128/CMR.00052-08.

38. Gjenero-Margan I, Aleraj B, Krajcar D, Lesnikar V, Klobučar A, Pem-Novosel I, et al. Autochthonous dengue fever in Croatia, August–September 2010. Eurosurveillance. 2011;16(9):19805. doi: https://doi.org/10.2807/ese.16.09.19805-en.

39. Florian Franke, Sandra Giron, Amandine Cochet, Charles Jeannin, Isabelle Leparc-Goffart, Henriette de Valk, et al. Emergences of dengue and chikungunya in metropolitan France, 2010-2018. Bull Epidemiol Hebd. 2019;19-20:374-82. doi: http://beh.santepubliquefrance.fr/beh/2019/19-20/2019_19-20_2.html.

40. Succo T, Leparc-Goffart I, Ferré J-B, Roiz D, Broche B, Maquart M, et al. Autochthonous dengue outbreak in Nîmes, south of France, July to September 2015. Eurosurveillance. 2016;21(21):30240. doi: https://doi.org/10.2807/1560-7917.ES.2016.21.21.30240.

41. European Centre for Disease Prevention and Control. Rapid risk assessment: Autochthonous cases of dengue in Spain and France. Stockholm: ECDC2019 (20 Aug 2020). Available from: https://www.ecdc.europa.eu/en/publications-data/rapid-risk-assessment-autochthonous-cases-dengue-spain-and-france.

42. Sante Publique France - Agence régionale de santé - Occitanie. Press release autochthonous case of dengue in the Hérault deparment Montpellier: ARS Occitanie2020 (20 Aug 2020). Available from: http://www.herault.gouv.fr/content/download/34737/233309/file/2020_07_31_CP-Dengue-Cessenon-sur-Orb-34_V2-dm.pdf.

43. European Centre for Disease Prevention and Control. Dengue outbreak in Madeira (2012-13) 2013 (17 Jul 2019). Available from: https://ecdc.europa.eu/en/dengue-fever/threats-and-outbreaks/madeira-outbreak-2012.

44. European Centre for Disease Prevention and Control. Update on autochthonous dengue cases in Madeira, Portugal. Stockholm: European Centre for Disease Prevention and Control, 2012. Available from: https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/dengue-madeira-risk-assessment-update.pdf.

45. European Centre for Disease Prevention and Control. Autochthonous cases of dengue in Spain and France. Stockholm: ECDC, 2019 1 Oct 2019. Report No. Available from: https://www.ecdc.europa.eu/sites/default/files/documents/RRA-dengue-in-Spain-France_1Oct2019.pdf.

46. European Centre for Disease Prevention and Control. Local transmission of dengue fever in France and Spain – 2018 — 22 October 2018. 2018 (cited 20 Aug 2020). Available from: https://www.ecdc.europa.eu/sites/portal/files/documents/08-10-2018-RRA-Dengue-France.pdf.

47. Ministerio de Sanidad Consumo Y Bienestar Social. Dengue autóctono en España 2ª actualización 2019 (cited 20 Aug 2020). Available from: https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/analisisituacion/doc/ERR_Dengue_autoctono_mayo2019.pdf.

48. Angelini R, Finarelli AC, Angelini P, Po C, Petropulacos K, Silvi G, et al. Chikungunya in north-eastern Italy: A summing up of the outbreak. Weekly releases (1997–2007). 2007;12(47):3313. doi: https://doi.org/10.2807/esw.12.47.03313-en.

49. Rezza G, Nicoletti L, Angelini R, Romi R, Finarelli AC, Panning M, et al. Infection with chikungunya virus in Italy: An outbreak in a temperate region. The Lancet. 2007;370(9602):1840-6. doi: https://doi.org/10.1016/S0140-6736(07)61779-6.

50. Delisle E, Rousseau C, Broche B, Leparc-Goffart I, L'Ambert G, Cochet A, et al. Chikungunya outbreak in Montpellier, France, September to October 2014. Eurosurveillance. 2015;20(17):21108. doi: https://doi.org/10.2807/1560-7917.ES2015.20.17.21108.

51. Venturi G, Di Luca M, Fortuna C, Remoli ME, Riccardo F, Severini F, et al. Detection of a chikungunya outbreak in central Italy, August to September 2017. Eurosurveillance. 2017;22(39):17-00646. doi: https://doi.org/10.2807/1560-7917.ES.2017.22.39.17-00646.

52. Bowler J. Wildlife of Madeira and the Canary Islands.: Princeton University Press; 2018. 224 p.

53. Instituto Nacional de Estatística. Census - final results. Portugal - 2011 2012 (20 Aug 2020). Available from: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=73212469&PUBLICACOESmodo=2.

54. Fernandopullé D. Climatic characteristics of the Canary Islands. In: G. K, editor. Biogeography and ecology in the Canary Islands: Springer, Dordrecht; 1976. p. 185-206.

55. Santos FD, Valente MA, Miranda PMA, Aguiar A, Azevedo EB, Tomé AR, et al. Climate change scenarios in the Azores and Madeira Islands. World Resource Review. 2004;16(4):473-91. http://idlcc.fc.ul.pt/pdf/SantosEtalWRR2004.pdf.

56. Margarita Y, Grácio AS, Lencastre I, Silva A, Novo T, Sousa C. First record of *aedes (stegomyia) aegypti (linnaeus, 1762) (diptera, culicidae)* in Madeira Island - Portugal (in Portguese) Acta parasitológica portuguesa. 2006; 13:59-61.

57. Messina JP, Brady OJ, Pigott DM, Brownstein JS, Hoen AG, Hay SI. A global compendium of human dengue virus occurrence. Scientific Data. 2014; 1:140004. doi: https://doi.org/10.1038/sdata.2014.4.

58. Guo C, Zhou Z, Wen Z, Liu Y, Zeng C, Xiao D, et al. Global epidemiology of dengue outbreaks in 1990-2015: A systematic review and meta-analysis. Frontiers in cellular and infection microbiology. 2017; 7:317-. doi: https://doi.org/10.3389/fcimb.2017.00317.

59. Semenza JC. Prototype early warning systems for vector-borne diseases in Europe. International Journal of Environmental Research and Public Health. 2015;12(6). doi: https://doi.org/10.3390/ijerph120606333.

60. Semenza JC, Suk JE. Vector-borne diseases and climate change: A European perspective. FEMS microbiology letters. 2018;365(2): fnx244. doi: https://doi.org/10.1093/femsle/fnx244.

61. Berg Hvd, Velayudhan R, Ejov M. Regional framework for surveillance and control of invasive mosquito vectors and re-emerging vector-borne diseases 2014–2020. Copenhagen World Health Organization, 2013. Available from: http://www.euro.who.int/__data/assets/pdf_file/0004/197158/Regional-framework-for-surveillance-and-control-of-invasive-mosquito-vectors-and-re-emerging-vector-borne-diseases-20142020.pdf?ua=1.

62. Meliker JR, Sloan CD. Spatio-temporal epidemiology: Principles and opportunities. Spatial and Spatio-temporal Epidemiology. 2011;2(1):1-9. doi: https://doi.org/10.1016/j.sste.2010.10.001.

63. Chen Q, Han R, Ye F, Li W. Spatio-temporal ecological models. Ecological Informatics. 2011;6(1):37-43. doi: https://doi.org/10.1016/j.ecoinf.2010.07.006.

64. Columbia Public Health. Spatiotemporal analysis. Population health methods. 2019. https://www.publichealth.columbia.edu/research/population-health-methods/spatiotemporal-analysis.

65. Meslé MMI, Hall IM, Christley RM, Leach S, Read JM. The use and reporting of airline passenger data for infectious disease modelling: A systematic review. Eurosurveillance 2019;24(31):1800216. doi: https://doi.org/10.2807/1560-7917.ES.2019.24.31.1800216.

66. Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, et al. Generalized linear mixed models: A practical guide for ecology and evolution. Trends in Ecology & Evolution. 2009;24(3):127-35. doi: https://doi.org/10.1016/j.tree.2008.10.008.

67. Silk MJ, Croft DP, Delahay RJ, Hodgson DJ, Weber N, Boots M, et al. The application of statistical network models in disease research. Methods in Ecology and Evolution. 2017;8(9):1026-41. doi: https://doi.org/10.1111/2041-210X.12770.

68. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics--ii. The problem of endemicity.1932. Bull Math Biol. 1991;53(1-2):57-87. doi: https://doi.org/10.1007/bf02464424.

69. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics--i. 1927. Bull Math Biol. 1991;53(1-2):33-55. doi: https://doi.org/10.1007/bf02464423.

70. Kretzschmar M, Wallinga J. Mathematical models in infectious disease epidemiology. Modern Infectious Disease Epidemiology: Concepts, Methods, Mathematical Models, and Public Health. 2009:209-21. doi: https://doi.org/10.1007/978-0-387-93835-6_12.

71. Messina JP, Brady OJ, Pigott DM, Golding N, Kraemer MU, Scott TW, et al. The many projected futures of dengue. Nat Rev Microbiol. 2015;13(4):230-9. doi: https://doi.org/10.1038/nrmicro3430.

72. Mordecai EA, Caldwell JM, Grossman MK, Lippi CA, Johnson LR, Neira M, et al. Thermal biology of mosquito-borne disease. Ecology Letters. 2019;0(0). doi: https://doi.org/10.1111/ele.13335.

73. Huber JH, Childs ML, Caldwell JM, Mordecai EA. Seasonal temperature variation influences climate suitability for dengue, chikungunya, and zika transmission. PLoS Negl Trop Dis. 2018;12(5): e0006451. doi: https://doi.org/10.1371/journal.pntd.0006451.

2 Results

2.1 Dengue importation into Europe: A network connectivity-based approach.

Salami D, Capinha C, Martins MdRO, Sousa CA (2020). Dengue importation into Europe: A network connectivity-based approach. PLoS One 15(3): e0230274. https://doi.org/10.1371/journal.pone.0230274

PLOS ONE

RESEARCH ARTICLE

Dengue importation into Europe: A network connectivity-based approach

Donald Salami[1]*, César Capinha[2], Maria do Rosário Oliveira Martins[1], Carla Alexandra Sousa[1]*

1 Global Health and Tropical Medicine, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisboa, Lisbon, Portugal, **2** Centro de Estudos Geográficos, Instituto de Geografia e Ordenamento do Território, Universidade de Lisboa, Lisboa, Lisbon, Portugal

* donald.salami@gmail.com (DS); casousa@ihmt.unl.pt (CS)

Citation: Salami D, Capinha C, Martins MdRO, Sousa CA (2020) Dengue importation into Europe: A network connectivity-based approach. PLoS ONE 15(3): e0230274. https://doi.org/10.1371/journal.pone.0230274

Editor: Irene Sendiña-Nadal, Universidad Rey Juan Carlos, SPAIN

Received: October 23, 2019

Accepted: February 25, 2020

Published: March 12, 2020

Data Availability Statement: We the authors do not have the legal rights to share the air travel data used in this study, as the data is a third-party data owned by the International Air Travel Association (IATA). The same data can be purchased for use by any other researcher by contacting the International Air Travel Association (IATA)-Passenger Intelligence Services (PaxIS) (https://www.iata.org/services/statistics/intelligence/paxis/Pages/index.aspx). Similarly, the disease (dengue) data, are third-party data from the European Centre for Disease Prevention and Control (ECDC). This data is available to any other researcher, upon

Abstract

The spread of dengue through global human mobility is a major public health concern. A key challenge is understanding the transmission pathways and mediating factors that characterized the patterns of dengue importation into non-endemic areas. Utilizing a network connectivity-based approach, we analyze the importation patterns of dengue fever into European countries. Seven connectivity indices were developed to characterize the role of the air passenger traffic, seasonality, incidence rate, geographical proximity, epidemic vulnerability, and wealth of a source country, in facilitating the transport and importation of dengue fever. We used generalized linear mixed models (GLMMs) to examine the relationship between dengue importation and the connectivity indices while accounting for the air transport network structure. We also incorporated network autocorrelation within a GLMM framework to investigate the propensity of a European country to receive an imported case, by virtue of its position within the air transport network. The connectivity indices and dynamical processes of the air transport network were strong predictors of dengue importation in Europe. With more than 70% of the variation in dengue importation patterns explained. We found that transportation potential was higher for source countries with seasonal dengue activity, high passenger traffic, high incidence rates, high epidemic vulnerability, and in geographical proximity to a destination country in Europe. We also found that position of a European country within the air transport network was a strong predictor of the country's propensity to receive an imported case. Our findings provide evidence that the importation patterns of dengue into Europe can be largely explained by appropriately characterizing the heterogeneities of the source, and topology of the air transport network. This contributes to the foundational framework for building integrated predictive models for bio-surveillance of dengue importation.

Introduction

During the last few decades, dengue fever has rapidly spread into new geographical regions with a resultant increase in its global incidence [1, 2]. This global spread has notably been

request to the European Centre for Disease Prevention and Control (ECDC) (https://www.ecdc.europa.eu/en/publicationsdata/european-surveillance-system-tessy). All other relevant data, used in the study are open source, and readily available. Sources of these data are referenced within the manuscript

Funding: DS has a PhD grant from the Fundação para a Ciência e a Tecnologia, Portugal (https://www.fct.pt/), grant number: PD/BD/128084/2016. This work was also partially funded by Fundação para a Ciência e a Tecnologia(https://www.fct.pt/) grant awarded to CAS, grant number: PTDC/SAU-PUB/30089/2017 The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist

linked to increasing human mobility, particularly air travel [3–5]. Global aviation network has increased in volume by almost eight-fold in the past 40years, enabling human movement across long distances in a relatively short time [6–8]. Thus, creating a mobility network for the spread of infectious diseases like dengue [9–11]. Infected air travelers have contributed significantly to the importation of dengue to non-endemic areas [12–14].

As human mobility and connectivity continue to advance, dengue spread via importation will continue to increase at unpredictable rates [4, 15]. The complexity of the air transport network poses a substantial challenge in the understanding of the dynamics of dengue spread and importation [16, 17]. How to effectively tackle dengue spread mediated via the complex air transport network is a priority of vector-borne disease surveillance and control [18–20]. In this context, understanding the dynamics by which dengue fever is transported, across the complex and dynamic mobility network, is an important first step [5].

Previous work on network-mediated epidemic often assumes a common framework, that the probability of an imported infection is directly correlated with the number of arriving air passengers [13, 21]. However, a simple correlation between imported cases and crude travel statistics is insufficient to explain the transmission pathways [21]. Such an approach does not allow to differentiate imported cases arriving from countries with higher infection risk and transport potential, or if the variance in the number of cases is also mediated by other socio-economic and anthropogenic factors. Neither does this capture the connectivity patterns of the mobility network that influence or constrains the dynamics of importation.

Recent studies have applied a range of social network modeling approaches to understanding the transmission pathways of a network-mediated epidemic. From a general perspective, these methods combine the derived attributes of the epidemic source country and the topology of the transporting network to explain importation dynamics [17, 22]. Epidemic source attributes are characterized by the heterogeneities in the passenger's air travel volume, socio-economic and anthropogenic factors, that mediate the risk of infection and transport of the disease [23]. While the network topology is characterized by centrality measures of the nodes (countries) in the air transport network. [24, 25]. To our knowledge, there are no studies that apply this explanatory power of social network analysis to characterize dengue importation patterns into Europe.

Here, we adopt a refined network connectivity approach to analyze data on imported dengue cases from 21 European countries, within a 6-year period (2010–2015). Specifically, our approach is outlined in the following: (1) We integrate a source-to-destination country combination to construct a network connectivity for dengue importation; (2) We then examine connectivity measures accounting for factors mediating the transport and importation potential from the source country; (3) Lastly, we investigate how the topology of the air transport network influences the importation risk from a source and the propensity of a destination country to receive an imported case.

Methods

Conceptual framework

Our proposed network connectivity framework adapts techniques from previous work on dispersal connectivity, spatial autocorrelation and network modelling [25–28] to capture the dynamics of dengue importation. The inputs for our analysis consist of the dengue importation data, air travel data and the underlying air transport network structure.

In the sections below, we describe the various inputs and our modeling approaches. In the first section, we describe the disease and air travel data, with their respective sources. Next, we describe the connectivity indices representing factors that potentially facilitate the transport

and importation of dengue from a source country into Europe. We then describe the generalized linear mixed effect (GLMM) modelling approach for quantifying the variation in dengue importation as explained by the indices.

The next section introduces the concept of dependency network analysis to account for the influence of the air transport network structure. Firstly, we describe the construct of the weighted directed network from the air passenger's data and then define centrality measures to characterize each individual node's (i.e. country) influence within the network. The centrality measures of the source countries were then added to the GLMM model of the connectivity indices to account for the influence of network structure. Finally, we describe an extension of our analysis to model the propensity of a country in Europe to receive an imported case, by virtue of its network topology (i.e. centrality measures).

Disease data

We analyzed imported cases of dengue fever reported in Europe for the period of 2010–2015. Dengue fever data was obtained from the European Centre for Disease Prevention and Control (ECDC) [29]. Routine (weekly) Europe-wide infectious disease surveillance data is collected from European Economic Area member states (EU/EEA) countries by the ECDC. Data is collected and managed through The European Surveillance System (TESSy) database; a database provided by the ECDC national focal points for surveillance [30].

Here, we considered confirmed cases of dengue, according to the 2012 EU case definition for viral haemorrhagic fever (VHF) which defines a confirmed case as any person meeting the clinical and the laboratory criteria [31]. The subset of imported cases or travel-associated cases are categorized as persons having been outside the country of notification during the incubation period of the disease. Place/country of infection was defined as the place the person was during the incubation period of the disease. A total of 21 EU/EEA countries reported data on imported dengue, within the period of 2010–2015 (inclusive of zero reporting). For our analysis, we considered each imported case, its country of infection (as source country), the reporting country (as destination country) and the reporting month.

Air travel data

To describe the flow of individuals into Europe, we obtained the passenger air travel data for 2010–2015, from the International Air Transport Association (IATA) [32]. IATA Passenger Intelligence Services (PaxIS) data, is the most comprehensive airline passenger's data available today. Data includes the complete passenger itineraries, true origins and destinations, route segments and connecting points. The data contained over 11,996 airports in 229 different countries and their territorial dependencies, with calibrated passenger travel volumes for each route at a monthly timescale. The passenger volumes were available at the country level, i.e. the total number of passengers traveling from each country worldwide. We used these data to construct a monthly directional passenger flow from all countries worldwide with a final destination in Europe (also accounting for all connecting flights). This passenger flow was inclusive of flow in-between European countries.

Connectivity indices

We assembled seven indices representing factors potentially mediating the importation risk of dengue from a source country. These indices are decomposed into components representing the 'source strength' (the risk of dengue infection); and the transport and importation potential (the connection between the source country and the potential destination country in Europe).

Indices 1 and 2, characterize dengue monthly activity and annual seasonality in the source country. These indices were created using worldwide dengue outbreak notifications data from DengueMap (a unified data collection tool, that brings together disparate dengue reports of local or imported dengue cases from official, newspapers and other media sources globally) [33]. **Index 1** is defined as having notification of one or more confirmed cases in a given month (January-December) in the years of 2010 through 2015, **index 2** is defined as having dengue activity (i.e. notification of one or more confirmed cases) in a given month in two or more years from 2010 through 2015 [34]. **Index 3** is the annual dengue incidence estimates of source country adjusted for the country's population, obtained from the Institute for Health Metrics and Evaluation (IHME) [35, 36]. **Index 4** is the geographical distance between centroids of source country and destination country, often modelled as a proxy for travel time and predictor for epidemic arrival times [37]. This assumes that proximity to an endemic source increases transport and importation potential to destination. **Index 5** is the epidemic vulnerability of the source country, represented by the recent infectious disease vulnerability index from RAND cooperation [38]. This index identifies countries' vulnerabilities to control potential disease outbreaks by assessing a confluence of seven broad country level factors: demographic, health care, public health, disease dynamics, political-domestic, political-international and economic [38]. This index assumes that most vulnerable countries might pose a higher risk of infection and transportation. **Index 6** is the income per capita (GDP) of source countries; poor countries with weak economies are associated with poor health outcomes, lesser abilities to detect, prevent and respond effectively to infectious disease. Hence, we assume greater importation risk from poorer source countries. **Index 7** is the total arriving passengers from source country to destination country (i.e. accounting for both direct and connecting flights) which has often been correlated with disease importation [13], with an implicit assumption that infection risk is equal for all source countries. Hence an increase in passengers, in turn, increases the transport and importation potential.

Source strength for all indices was determined by the endemicity of dengue in the source country, while transport and importation potential were modelled based on each unique factor, as detailed in Table 1 below.

To quantify the variation in dengue importation as explained by our proposed connectivity indices, we fitted a generalized linear mixed effect model (GLMMs) [39] Unit of analysis was a source-destination-month combination, with a binary response variable coded to indicate a reported case of imported dengue (1) or not (0). The GLMM was fitted with logit link functions for binomial errors and fixed effects of the connectivity indices and crossed random effects (intercepts) of the destination country. We also included a fixed effect for time to control for the following: specific global aviation traffic effects, i.e. increase or decrease in traffic to a specific country, due to a major event (e.g. traffic increase to London, during the 2012 summer Olympics), and to also account for variation in reporting rates among countries in specific months. The time effect was specified by a set of dummy variables, one for each month. To improve normality, the continuous fixed effects variables were transformed using the $\log(x+1)$ function, then centered on zero, and standardized to unit variance, before model fitting. Model fits were evaluated by calculating the marginal and conditional GLMM R^2 [40]. Likewise, model post-estimation using model diagnostic measures and residuals plots were evaluated using the DHARMa residual diagnostics for hierarchical models [41].

Dependency network

Next, to explore the influence of the air transport network structure on the variation of dengue importation, we incorporate the dependency network approach. This is a system level analysis

Table 2-1 Connectivity indices of a focal destination country for the importation of dengue fever.

Table 1. Connectivity indices ($S_{j,f}$) of a focal destination country (j) for the importation of dengue fever (f).

Index	Meaning	Connectivity ($S_{j,f}$)	Source strength	Transport and importation
1	Dengue activity	$\sum_{i \neq j} \theta_{i,f} A_i$	$\theta_{i,f}$	*1*
2	Dengue seasonality†	$\sum_{i \neq j} \theta_{i,f} S_i$	$\theta_{i,f}$	*1*
3	Incidence estimates of source country	$\sum_{i \neq j} \theta_{i,f} IR_i$	$\theta_{i,f}$	*1*
4	Geographical distance	$\sum_{i \neq j} \theta_{i,f} / \ln d_{ij}$	$\theta_{i,f}$	$1/\ln d_{ij}$
5	Epidemic vulnerability‡	$\sum_{i \neq j} \theta_{i,f} / V_i$	$\theta_{i,f}$	$1/V_i$
6	Source country's wealth (GDP)	$\sum_{i \neq j} \theta_{i,f} / \ln g_i$	$\theta_{i,f}$	$1/\ln g_i$
7	Total air passengers from a source country	$\sum_{i \neq j} P_{i \to j}$	*1*	$P_{i \to j}$

i = source country; j = destination country; f = *dengue fever*; $\theta_{i,f}$ = endemic to dengue fever, infection is constantly maintained at a baseline level in country without external inputs (0 or 1); IR = Incidence rate; d = geographical distance between centroids of countries i and j (in kilometers); V = infectious disease vulnerability; P = total air passengers; g = Gross domestic product.

† = coded as a binary variable, with a value 0 indicates 'no activity or seasonal pattern' and 1 indicates 'activity or seasonal pattern' respectively.

‡ = infectious disease vulnerability index is summarized as a normed factor score for each country ranging between 0 and 1, from most vulnerable with the lowest score to least vulnerable with the highest score, hence we specify the inverse.

Logarithms were used in the equations to decrease variability in raw input data and to improve the normality of the indices.

https://doi.org/10.1371/journal.pone.0230274.t001

of the activity and topology of the air transport network to investigate the capacity of a country to influence or be influenced by, other countries within the network by virtue of their connections [42–44]. The specificity of network influence on dengue importation risk is that the connectivity between a source country *(i)* and a destination county *(j)* is not analyzed in isolation, but in consideration for the effects of *z*, where *z*, is the other countries within the network structure. The implication of this network structural view is that the connection between *i* and *j* is also dependent on the connections between *i* and *z* and between *j* and *z*. Therefore, the distributed heterogeneity in the connections is characterized by interdependencies within the network and must be accounted for statistically [25, 45].

Using IATA passengers air travel data, we constructed 72 weighted directed networks, to represent the monthly flow of air passengers into Europe, from 2010 to 2015. For each network, countries are represented by a node while edges represent the flow of passengers between pairs. The network graph is denoted by $G_m = (V,E)$, where V_G is a set containing all the nodes (or vertices), while E_G contains all the edges, with $m = 1, 2, \ldots 72$ indicating the month (representing the study period of 2010–2015). Edges are denoted as $e_{i,j}$, where *i* is the source and *j* is the destination of a travel route represented by the edge. For each connected pair of nodes *i* and *j*, the edge was weighted with the total number of passengers from *i* to *j* given by W_{ij}.

The role of a node in the network and its likelihood to influence the transport and importation potential of dengue was then characterized by the following centrality measures (Table 2):

Degree centrality. The number of links or connections that a node has, this assigns a score based on the number of nodes within the network, that an individual node is connected to [42, 45]. The higher a node's degree, the more it associates with neighboring nodes, potentially increasing its transport potential or increases its vulnerability to importation (in the case of a destination country).

Table 2-2 Centrality measures for a central node.

Table 2. Centrality measures for a central node (adapted from [45]).

	Centrality measure	Characteristics of a central node	Equation
1	Degree	Connected directly with many other nodes	$DC_i = s_i = \sum_{j \neq i} W_{ij}$
2	Betweenness	Lies on many shortest paths linking other pairs. The probability that communication from p to g will go through i	$BC_i = \sum_{p \neq i, p \neq q, q \neq i} \frac{g_{pq}(i)}{g_{pq}}$
3	Closeness	Short communication path to other nodes, a minimal number of steps to reach others	$CC_i = \frac{N}{\sum_j l_{ij}}$
4	Eigenvector	Connected (directly and/or indirectly) to many other nodes and/or to other high-degree nodes	$EC_i = \frac{1}{\lambda} \sum_j A_{ij} v_j$

i = source country; j = destination country; s = node strength (i.e. the sum of all edge weights attached to a node); W_{ij} = weight of the edge between nodes i and j; g_{pq} = the number of shortest-paths between nodes p and q; $g_{pq}(i)$ = the number of shortest-paths between nodes p and q (other random nodes in the network), which pass through i; N = number of nodes in a network; l_{ij} = distance of the shortest between nodes i and j, here weighted distance is given by the inverse of the number of passengers in the corresponding edge; A = adjacency matrix of the weighted network; λ = leading eigenvalue of A; v = leading eigenvector of A. All centrality measures were computed based on the weighted air transport network

https://doi.org/10.1371/journal.pone.0230274.t002

Betweenness centrality. Measures the number of times a node lies on the shortest path between other nodes in the network [42]. A node with high betweenness metric is expected to have a higher transport potential, as it bridges between different nodes that are not directly connected. Alternatively, destination countries with high betweenness are at higher risk of importation.

Closeness centrality. Measures how close (in terms of topological distance) a node is with respect to all other nodes [46]. The topological distance between any pair of connected nodes is given by the inverse of the number of passengers in the corresponding edge. Therefore, the higher the number of passengers in a given edge, the shorter the distance between them, the faster an infection gets transported or imported.

Eigenvector centrality. The basic idea of the eigenvector centrality is that a node's centrality is determined by the combination of its connections and that of its neighbors [47, 48]. A node will have high eigenvector centrality if it has strong connections with other highly connected nodes. In our context, if a node is connected to other high degree nodes, the higher its transport potential or the higher its vulnerability to importation from a random source in the network.

To quantify the influence of the air transport network topology, we employ two different modelling approaches. Firstly, we account for the network structure in our connectivity framework, by including the centrality measures of the source country. We tested whether GLMMs based on connectivity indices and the air network centrality measures outperformed the base model based on the connectivity indices alone. GLMMs were fitted with the same random effect structure as above. All possible combinations of centrality measures were considered, resultant models were compared based on their fixed effects ΔAIC and marginal GLMM R^2.

Secondly, we model the network topology of the focal destination country as a predictor of dengue importation. This assumes that dengue importation into a focal destination country could be directly related to the country's position in the air transport network. For this model, centrality measures of the destination countries were fitted as fixed effects, with crossed random effects (intercepts) of each time step (i.e. each month). As above, all possible models were fitted and compared based on their fixed effects ΔAIC and marginal GLMM R^2.

All analysis and plots were performed in R software, version 3.5.2 [49], using the following libraries: glmmtmb [50], igraph [51], tnam [52], DHARMa [41], sjstats [53], ggplot2 [54], circlize [55] and their varying dependencies.

Results

Among the 7,277 imported dengue cases reported in Europe from 2010 to 2015, 4,112 (57.0%) cases had known travel history, i.e. source country of infection. The cases with the known source were imported into 21 European countries, from 99 different countries distributed across all global regions. 62.1% of reported cases originated from South-East Asia Region, with Thailand, Indonesia, India, and Sri Lanka being the major source of import; 18.0% originated from Region of the Americas, with most cases coming from Brazil; 12.1% from the Western Pacific Region, where Vietnam and the Philippines were the major import source. The data showed that imported cases were most frequently reported in Germany, Sweden, United Kingdom, Italy and Norway (Fig 1).

An annual average of 436 million passengers entered Europe from other countries worldwide from 2010 to 2015. Of the total number of passengers arriving from regions outside of Europe, 44% originated from region of the Americas, with higher traffic from the United states, Canada and Brazil; 19% from the Eastern Mediterranean, with higher traffic from Morocco, United Arab Emirates and Tunisia; 15% from the Western Pacific with higher traffic from Thailand, India and Hong Kong; 12% from the African region, with higher traffic from Algeria, South Africa, Nigeria; 10% from South-East Asia and. A country-level passenger inflow from WHO Member States regions is shown in Fig 2. High traffic inflow was most common in the United Kingdom, France, Germany, Italy, and Spain. Most of the identified dengue importation hot spots were characterized by a high influx of air passengers, with a tendency for an increase in importation risk. However, the high influx of passengers does not necessarily lead to a high number of imported cases of dengue, as total number of arriving passengers was weakly correlated with the number of imported cases (Spearman's $\rho = 0.13$, $p = <0.01$, S1 Fig).

Connectivity indices and dengue importation

Our first model investigated the relationship between dengue importation and the connectivity indices. Among the connectivity indices, GDP and epidemic vulnerability displayed high correlations amongst themselves (Spearman's $\rho = 0.88$, $p < .001$, S1 Fig). Utilizing a threshold collinearity of >0.70, we deemed this pair to be redundant and added each variable in separate models. In a comparative-fit test, the model including epidemic vulnerability had a slightly better fit, than the model with GDP (details shown in S1 Appendix in Table 1). Table 3 presents the results of the GLMM model with estimated coefficients, odds ratios and 95% confidence intervals for each connectivity index. The estimated coefficients represent the relative influence of each connectivity index on the risk of dengue importation between source–destination country combinations. In this model, dengue importation was significantly associated with the following connectivity measures: dengue activity, seasonality, incidence rates, geographical distance, the epidemic vulnerability of source country and arriving passengers. Evaluating the model independently of random effects variance component, these connectivity indices account for 51% (marginal GLMM R^2 of 0.507) of the variation in dengue importation patterns. Overall the GLMM model accounted for 66% (conditional GLMM R^2 of 0.656) of the variation in dengue importation into Europe. The model fit was adequate, the simulated residual diagnostics test for overall uniformity showed no evidence of model misspecification.

Influence of air transport network topology

The resulting air transport network from the passenger's data was a directed graph structure with 229 nodes and 10261 edges. We took the mean of each centrality measure across the 72 weighted networks to obtain a single value for each node. Based on the mean centrality measures, the following source countries were most central in respect to all metrics: Canada,

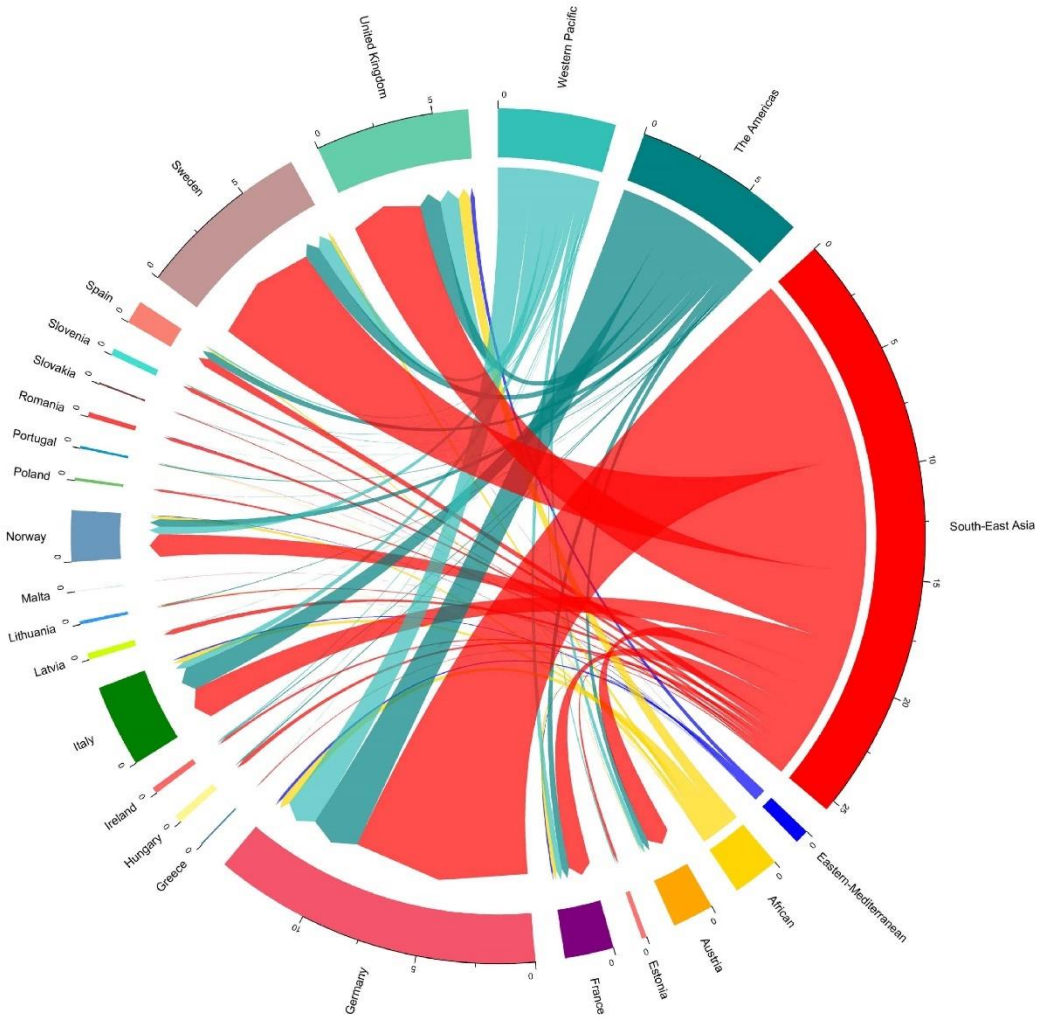

Fig 1. Distribution of imported dengue cases by the destination country and the source region. Destination countries are represented by 21 European countries. Source region represents, 99 different countries distributed across all global. Source countries were grouped into regions for visual representation, with region grouping defined by the WHO Member States region definition. WHO Member States are grouped into 6 WHO regions: African Region, Region of the Americas, South-East Asia Region, European Region (not included in source countries), Eastern Mediterranean Region, and Western Pacific Region [56].

https://doi.org/10.1371/journal.pone.0230274.g001

United States, Australia, United Arab Emirates, China and Brazil. Fig 3 shows the air transport network connections, with the average degree centrality measures of the source countries for period of 2010–2015. This figure highlights the most important connections and hide the noise of rare connections. The network metrics displayed notable relationships, with

Figure 2-1 Distribution of imported dengue cases by the destination country and the source region.

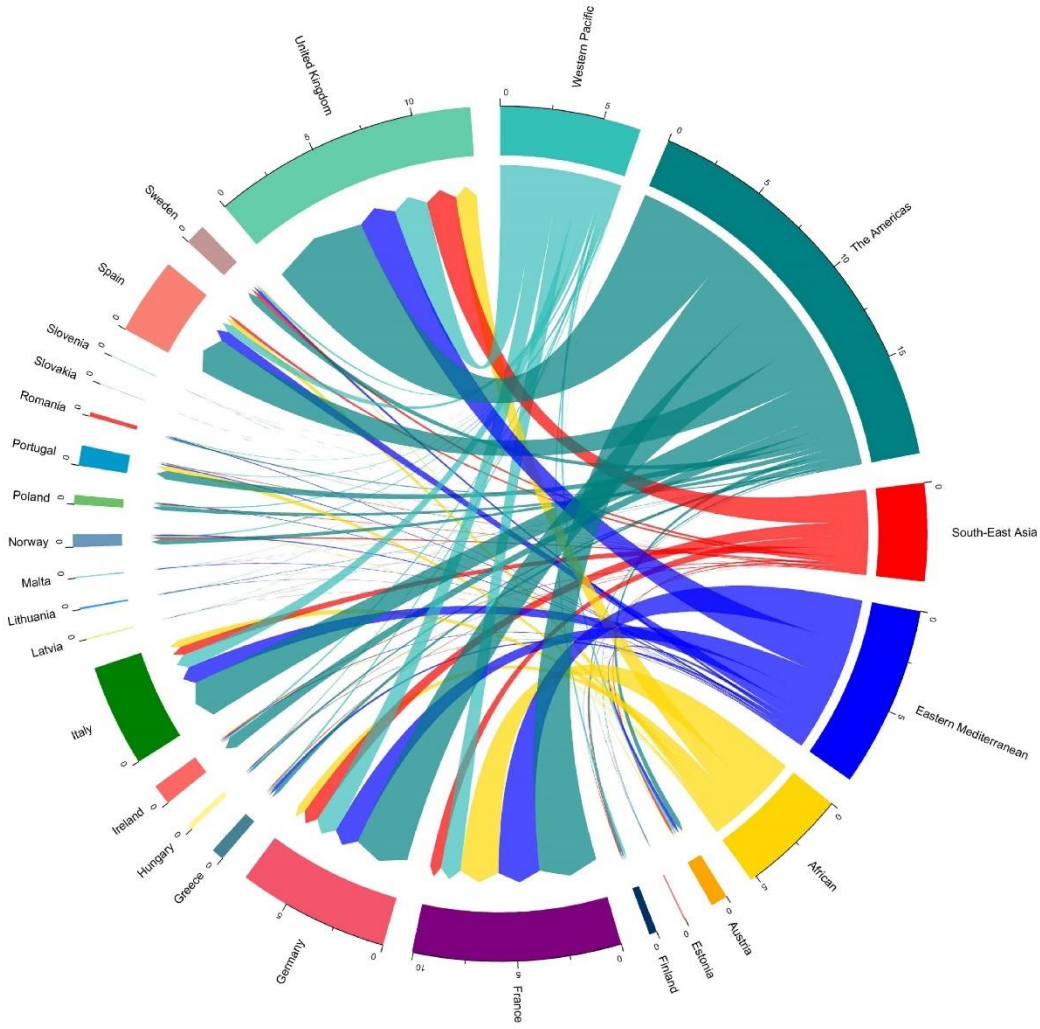

Fig 2. Distribution of air passengers arriving into Europe from WHO regions in 2010–2015. Destination countries are represented by 21 European countries. Source region represents, 99 different countries distributed across all global. Source countries were grouped into regions for visual representation, with region grouping defined by the WHO Member States region definition. WHO Member States are grouped into 6 WHO regions: African Region, Region of the Americas, South-East Asia Region, European Region (not included in source countries), Eastern Mediterranean Region, and Western Pacific Region [56].

https://doi.org/10.1371/journal.pone.0230274.g002

moderate-to-high correlations amongst themselves (S2 Fig). Betweenness and Eigenvector centrality pairs were the most highly correlated (Spearman's $\rho = 1.00$, $p < .001$). Hence to avoid redundancy in model fitting, we added the correlated pairs in separate models. All the network metrics examined were significant predictors of dengue importation. Compared with the base model of the connectivity indices alone, GLMM fits were substantially improved by

Figure 2-2 Distribution of air passengers arriving into Europe from WHO regions in 2010–2015.

Table 2-3 Comparison of generalized linear mixed models predicting dengue importation in Europe

Table 3. Comparison of generalized linear mixed models predicting dengue importation in Europe, using the connectivity indices and network autocorrelation (i.e. centrality measures).

	Model 1		Model 2		Model 3	
	Coefficient [95% CI]	Odd ratio [95% CI]	Coefficient [95% CI]	Odd ratio [95% CI]	Coefficient [95% CI]	Odd ratio [95% CI]
Connectivity Index						
Dengue activity	0.31 [0.18, 0.44]	1.36 [1.19, 1.55]	0.29 [0.16, 0.42]	1.33 [1.17, 1.52]	0.27 [0.13, 0.40]	1.30 [1.14, 1.49]
Dengue seasonality	0.12 [-0.01, 0.25]	1.13 [0.99, 1.29]	0.09 [-0.04, 0.23]	1.10 [0.96, 1.25]	0.03 [-0.11, 0.16]	1.03 [0.90, 1.18]
Incidence estimates	1.44 [1.33, 1.55]	4.23 [3.79, 4.72]	1.74 [1.61, 1.87]	5.68 [4.98, 6.49]	1.78 [1.65, 1.91]	5.92 [5.20, 6.75]
Geographical distance	-0.52 [-0.65, -0.38]	0.60 [0.52, 0.69]	-0.92 [-1.08, -0.77]	0.40 [0.34, 0.46]	-1.04 [-1.20, -0.88]	0.35 [0.30, 0.42]
Epidemic vulnerability	0.19 [0.10, 0.27]	1.20 [1.10, 1.32]	0.37 [0.27, 0.46]	1.44 [1.31, 1.58]	0.43 [0.33, 0.52]	1.53 [1.40, 1.68]
Total Air passengers	2.01 [1.90, 2.13]	7.49 [6.68, 8.39]	2.03 [1.89, 2.18]	7.63 [6.60, 8.81]	1.78 [1.61, 1.95]	5.91 [5.20, 6.75]
Centrality Measures						
Degree	—	—	—	—	0.39 [0.22, 0.55]	1.47 [1.25, 1.74]
Betweenness	—	—	—	—	0.96 [0.82, 1.11]	2.62 [2.26, 3.02]
Closeness	—	—	0.05 [-0.07, 0.17]	1.05 [0.94, 1.19]	—	—
Eigenvector	—	—	0.97 [0.81, 1.12]	2.63 [2.25, 3.06]	—	—
Model Fit						
$R^2_{GLMM(m)}$	0.507		0.535		0.531	
$R^2_{GLMM(c)}$	0.656		0.676		0.673	
Fixed effects ΔAIC	166.2		36.6		0	

Model 1: the based model of connectivity indices; **Model 2**: Base model with closeness and eigenvector centrality measures; **Model 3:** Base model with the degree and betweenness centrality measures.

$R^2_{GLMM(m)}$ = Marginal R^2; $R^2_{GLMM(c)}$ = Conditional R^2; ΔAIC = Change in Akaike Information Criterion.

https://doi.org/10.1371/journal.pone.0230274.t003

including the effects of the network topology (Fig 4). The best-fitting model was the model with degree and betweenness centrality measure pair, alongside the connectivity indices (Table 3), being the model with the lowest fixed effect AIC.

Network topology as a predictor of importation

Focal destination countries that were most central in the network were the United Kingdom, Germany, Spain, Italy, and France (Fig 5). Therefore, they are most vulnerable to dengue importation from a random source country. The centrality measures of the destination countries were highly correlated amongst themselves, hence it was practically redundant to include all measures in a single model (S2 Fig). We added each centrality measure in separate models, to capture the influence of the unique aspect of each nodal centrality. Each centrality measure was a significant predictor of dengue importation (Table 4). Surprisingly, the eigenvector centrality measure accounted for 70% of the variance in dengue importation (marginal GLMM R^2 of 0.706).

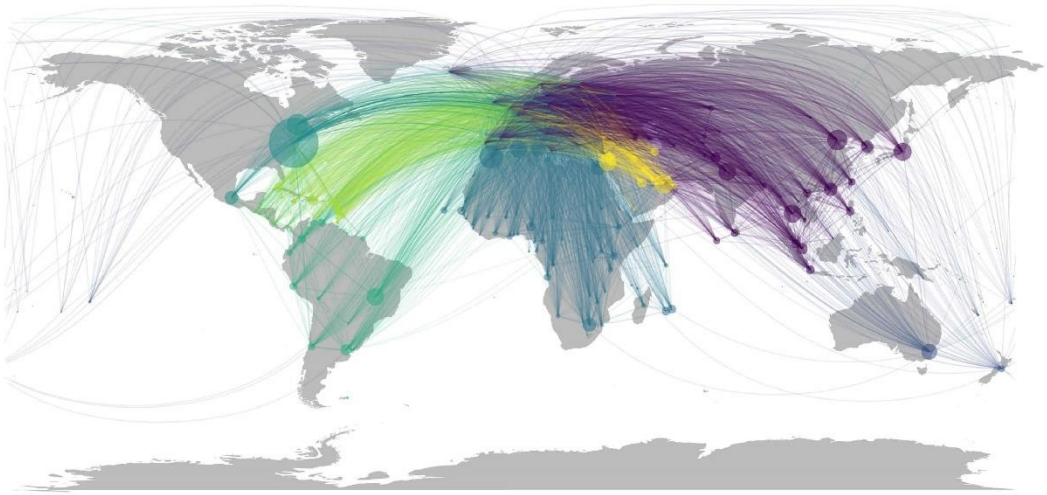

Fig 3. Air transport network and the average degree centrality measure of source countries. The links shows the connections within the air transport network, highlighting the most important pattern and hide the noise of connections with low passengers' volume. Each node represents a source country in the air transport network, and the size of a node is proportional to its mean degree score (averaged from the 72 weighted network), which is weighted by the average number of passengers with a final destination in Europe. Node color is categorized by region.

https://doi.org/10.1371/journal.pone.0230274.g003

Discussion

The importation of dengue into non-endemic regions is primarily initiated by global human mobility, hence it is critical to understand the dynamics of the transmission pathways based on the mobility networks. Here, we applied a refine network connectivity approach to model the importation of dengue into Europe. Our analysis accounted for factors that mediate the risk of importation from a source country through their effects on source strength and transport potential. In addition, we considered the influence of the air transport network topology on the importation risk from a source and the propensity of a destination country to receive an imported case. Our analysis demonstrated that the co-dynamics of the connectivity indices and the network topology explained more than 70% of the variance in dengue importation patterns. Likewise, the topologically positioning of a focal destination county in the network, played a key role in the importation patterns. These results contribute to our understanding of the transmission pathways of dengue importation and the role of the dynamical processes of the air transport network.

A major focus of our study was the understanding of importation patterns of dengue in Europe, with consideration to source strength and the characterized air transport network. Air passengers' data and transport connections have been used to infer relationships with imported cases of dengue [13, 57]. However, with the growing complexity of global mobility and transportation links, it is increasingly difficult to model importation solely on the crude aggregate statistics of air traffic [16, 58]. Air traffic connections between source and destination account for a fraction of required modeling parameters for general patterns of importation, other epidemiological and anthropogenic parameters need to be accounted for. A fundamental understanding of other mediating factors and transportation dynamics is required to achieve a more reliable predictive model.

Figure 2-3 Air transport network and the average degree centrality measure of source countries.

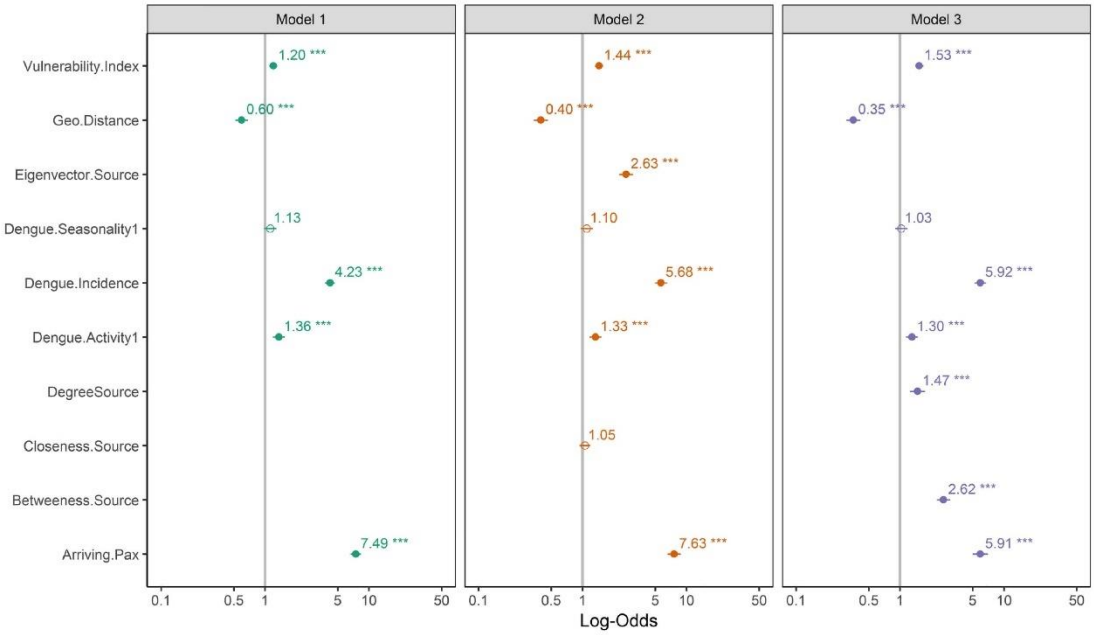

Fig 4. Forest plot comparing the base model and the network autocorrelation models for explaining dengue importation into Europe. Models are as represented in Table 3. Plots of the exponential transformed coefficients estimates, i.e. as an odd ratio. The "no effect" line is set at 1 and denoted by the grey line. Asterisks indicate the significance level of estimates "***" = p<0.001; "**" = p<0.01; "*" = p<0.05.

https://doi.org/10.1371/journal.pone.0230274.g004

Our analysis addresses this by endowing each source with a specific pattern of connectivity that mediates, its strength (risk of infection), transport and importation potential. Source strength was the risk of infection derived from the endemicity of dengue in a source country, distinguishing endemic and non-endemic countries. Heterogeneities in source strength were further characterized by dengue activity and seasonality patterns in the source country. As expected, an ongoing activity and seasonal pattern of dengue in a source country significantly increase the transport and importation potential. Transport and importation potential were also modelled uniquely for each source by the confluence of the various connectivity indices, accounting for other mediating factors. This modelling approach is corroborated by other similar studies [13, 23, 59]. With the results matching our a priori expectations, i.e. higher transport potential for source countries with high passenger traffic, high incidence rates, lower economic status, and geographical proximity to a destination country.

A new feature of interest in our analysis was the utilization of the infectious disease vulnerability index as an epidemiological factor. The vulnerability index presents a robust tool that identifies a country's ability to limit the spread of outbreak-prone diseases [38]. The combined multifarious nature of this index offers an intuitive understanding of the indigenous vulnerability of a source country. For our analysis, we modelled source strength, transport, and importation potential to increase with higher vulnerability, as expected most vulnerable countries poses a greater risk (model 3). Although our study was focused on the earliest stages of a network mediated epidemic, the inclusion of this index for focal destination country could

Figure 2-4 Forest plot comparing the base model and the network autocorrelation models for explaining dengue importation into Europe

Table 2-4 GLMMs modeling centrality measures of the destination countries as a predictor of dengue importation.

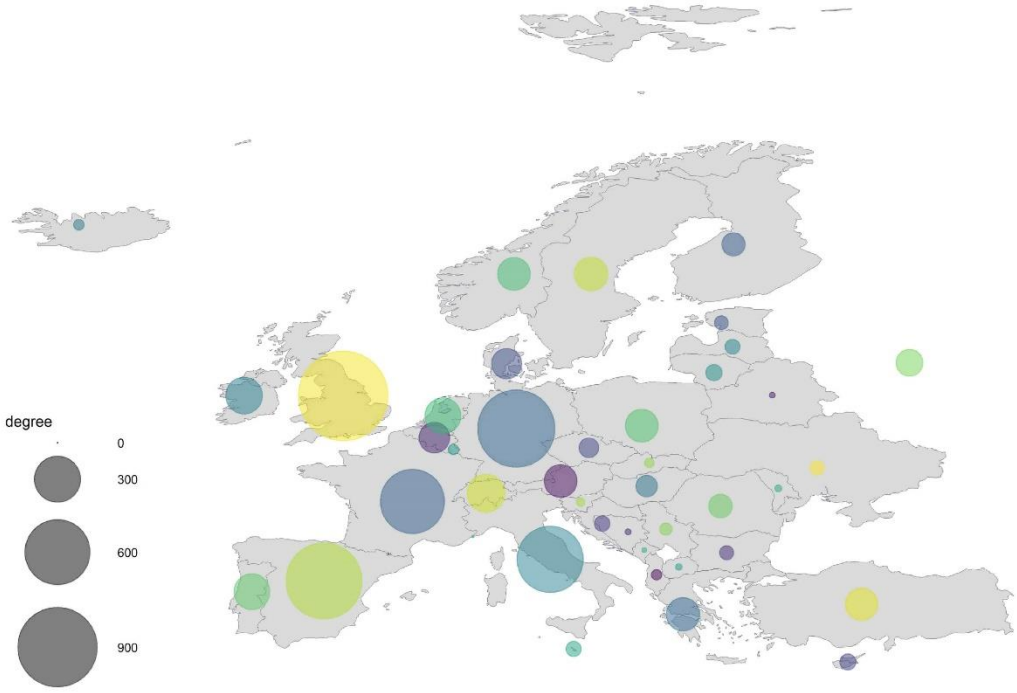

Fig 5. Degree centrality measure of the focal destination countries in Europe. Each node represents a country in Europe, and the size of a node is proportional to its degree, which is weighted by the average number of passengers. For a visual representation, the actual degree score was scaled down by a factor of 10^4.

https://doi.org/10.1371/journal.pone.0230274.g005

provide insights for modeling the establishment potential of an imported case of dengue. With the assumption, that recipient countries with higher vulnerability might present greater dissemination and establishment risk from an imported case, assuming vector presence and other ecological and climatic factors mediating transmission [60].

Our analysis went further to incorporate the dependency network approach to account for the influence of the air transport network topology on dengue importation. Utilizing centrality

Table 4. GLMMs modeling centrality measures of the destination countries as a predictor of dengue importation.

Network Metrics	Degree	Closeness	Betweenness	Eigenvector
Coefficient [95% CI]	1.12 [1.06, 1.18]	1.32 [1.22, 1.42]	1.25 [1.18, 1.32]	2.88 [2.65, 3.10]
Odd ratio [95% CI]	3.08 [2.90, 3.27]	3.73[3.37, 4.13]	3.49 [3.26, 3.74]	17.73 [14.12, 22.27]
Model Fit				
$R^2_{GLMM(m)}$	0.295	0.339	0.313	0.706
$R^2_{GLMM(c)}$	0.270	0.357	0.342	0.719
Fixed effect ΔAIC	794.7	90.2	311.5	0.0

$R^2_{GLMM(m)}$ = Marginal R^2; $R^2_{GLMM(c)}$ = Conditional R^2; ΔAIC = Change in Akaike Information Criterion.

https://doi.org/10.1371/journal.pone.0230274.t004

Figure 2-5 Degree centrality measure of the focal destination countries in Europe.

measures to quantify the influence of connection topology and the dynamical processes of the network to influence the importation of dengue [24]. We applied two different network analysis modelling approach, to quantify the unique contributions of each topological descriptor to dengue importation. The first approach incorporates network autocorrelation from the source country's centrality measures, within the GLMM framework of the connectivity indices. The addition of the centrality measures within the modeling framework addresses the issue of covariance driven by the network structure. All the network descriptors were significant predictors of dengue importation, however the combined effect of degree and betweenness centrality were the most influential. This result suggests that source countries that are highly connected (having multiple air routes into Europe) and act as connecting links to others countries (large airport hubs connecting other countries), intuitively have higher transport and importation potential, as they have the capacities to quickly connect with the wider network.

The second approach investigates if there is evidence of a correlation between the centrality measures of a destination country and its propensity to receive an imported case. We applied this approach as a valuable measure of a direct relationship between the air transport network structure and dengue importation into Europe [25]. Similar to the above results, all the network metrics were strong predictors of the variation in dengue importation. However, the eigenvector centrality, was the most fitting single predictor, explaining over 70% variance in dengue importation. These results suggest that the risk of dengue importation for a country (in Europe) can be largely explained by its position in the air transport network. Meaning countries have higher tendency for an imported case as a result of having more direct 'one hop' connections with high passengers traffic (as measured by degree centrality); having large airport hubs, bridging other countries (betweenness centrality); being effectively closer to other countries because of large passenger traffic (closeness centrality); and having multiple direct and indirect connections to other higher connected countries (eigenvector centrality). Overall, these results are particularly valuable in identifying countries in Europe that needs to prioritize investment in real-time surveillance systems, as a health security measure [35, 61], due to their increase propensity to receive an imported case.

Although our study was focused on the early stages of transport and introduction, we understand that the risk assessment of receptivity for dengue in European countries with established vector presence is essential. In general, most of Europe is considered unsuitable for the autochthonous transmission of dengue, as the established presence of the main vector (i.e. *Aedes aegypti*) is currently limited to the far eastern part of Europe, around the Black Sea [64, 65]. However, with an established presence of *Ae. aegypti* in Madeira (an autonomous region of Portugal), its recent introduction in the Netherlands, and the extending spread of *Aedes albopictus* on the continental, the risk of receptivity, establishment and autochthonous transmission is increasing [66, 67]. This risk was not considered in our study, as a larger panel of variables (which was not available to us at this time) would be required to adequately capture the interactions between the temporal and spatial dynamics of the vector population, ecological and climatic factors favoring autochthonous transmission and establishment. Nonetheless, we believe that the connectivity indices discussed here as well as the methodological approach provides a valuable way to assess importation risk, which can be easily expanded to model the risk of autochthonous transmission and establishment in areas in Europe with vector presence and favorable conditions. Likewise, in the context of the European region, where dengue is not endemic, our approach is highly relevant for countries with high passengers' traffic, to assess travel routes with higher risk of importation. Thereby providing useful information to help public health managers and decision-makers to plan resources for improve detection and prompt response to prevent and control the introduction of cases. We do believe that dengue

surveillance efforts, targeted at the introductory pathways would yield great benefits for European countries.

In summary, our paper presents a refined approach to the modelling of a network-mediated epidemic for dengue fever. The connectivity indices presented here captures the variation in source strength, transport, and importation potential, by accounting for other mediating socio-economic and anthropogenic factors. These indices are an amenable representation of real-world risk factors but offer a different approach for analyzing the connectivity dynamics of network-mediated importation of dengue. Our analysis went further to characterize the role of the air transport network topology in the dynamics of dengue importation into Europe. By investigating the network autocorrelations influencing transport potential from source countries and the network positioning of the destination country as a predictor of importation. Our analyses show that the connectivity indices and dynamical processes of the air transport network are strong predictors of dengue importation in Europe. Therefore, the network connectivity modelling approach could be useful in predicting source countries, importation patterns and destination countries with greater risk of dengue. Thereby, allowing for preemptive strategies to mitigate the impacts of imported cases in a timely, accurate and cost-effective manner [62, 63]. Finally, this modeling approach could serve as a pivotal prerequisite for the development of an early warning surveillance system to monitor and forecast the spread of dengue fever.

Supporting information

S1 Fig. Spearman correlations between dengue importation (numbers) and connectivity indices (variables). Red colors indicate strong positive correlations, dark blue indicates strong negative correlations, and yellow implies no empirical relationship between the variables.
(TIF)

S2 Fig. Spearman correlations between network properties of the source and destination countries within the air transport networks. Red colors indicate strong positive correlations and yellow indicates no empirical relationship between the centrality measures.
(TIF)

S1 Appendix. Alternative specifications of GLMM model.
(DOCX)

Acknowledgments

We greatly appreciate Dominic Freienstein for his assistant in accessing the international air travel association (IATA), passenger intelligence services (PaxIS) data.

Author Contributions

Conceptualization: Donald Salami, César Capinha, Carla Alexandra Sousa.

Data curation: Donald Salami.

Formal analysis: Donald Salami.

Methodology: Donald Salami, César Capinha.

Software: Donald Salami.

Supervision: César Capinha, Maria do Rosário Oliveira Martins, Carla Alexandra Sousa.

Visualization: Donald Salami.

Writing – original draft: Donald Salami.

Writing – review & editing: Donald Salami, César Capinha, Maria do Rosário Oliveira Martins, Carla Alexandra Sousa.

References

1. Messina JP, Brady OJ, Scott TW, Zou C, Pigott DM, Duda KA, et al. Global spread of dengue virus types: mapping the 70 year history. Trends Microbiol. 2014; 22(3):138–46. https://doi.org/10.1016/j.tim.2013.12.011 PMID: 24468533
2. Messina JP, Brady OJ, Golding N, Kraemer MUG, Wint GRW, Ray SE, et al. The current and future global distribution and population at risk of dengue. Nat. Microbiol. 2019. https://doi.org/10.1038/s41564-019-0476-8.
3. Vitaly Belik TG, Dirk Brockmann. Natural human mobility patterns and spatial spread of infectious diseases. Phys Rev X. 2011; 1(1). https://doi.org/10.1103/PhysRevX.1.011001.
4. Tian H, Sun Z, Faria NR, Yang J, Cazelles B, Huang S, et al. Increasing airline travel may facilitate co-circulation of multiple dengue virus serotypes in Asia. PLoS Negl Trop Dis. 2017; 11(8):e0005694. https://doi.org/10.1371/journal.pntd.0005694 PMID: 28771468
5. Brockmann D. Global connectivity and the spread of infectious diseases. Nova Acta Leopold. 2017; 419:129–36. http://rocs.hu-berlin.de/papers/brockmann_2017b.pdf.
6. Wilson ME. The traveller and emerging infections: Sentinel, courier, transmitter. J Appl Microbiol. 2003; 94 Suppl:1s–11s. https://doi.org/10.1046/j.1365-2672.94.s1.1.x PMID: 12675931
7. Schuttenhelm R. Global air travel increased 8 fold in 4 decades–and it's an accelerating trend. Bit of Science, 2016 [cited 2018 June 25]. Available from: http://www.bitsofscience.org/graph-global-air-travel-increase-6848/.
8. IATA. IATA annual review, 2019. International air transport association, 2019, [cited 1 July 2019]. Available from: https://www.iata.org/publications/Documents/iata-annual-review-2019.pdf.
9. Tatem AJ, Rogers DJ, Hay SI. Global transport networks and infectious disease spread. Adv Parasit. 2006; 62:293–343. https://doi.org/10.1016/s0065-308x(05)62009-x.
10. Nunes MR, Palacios G, Faria NR, Sousa EC Jr, Pantoja JA, Rodrigues SG, et al. Air travel is associated with intracontinental spread of dengue virus serotypes 1–3 in Brazil. PLoS Negl Trop Dis. 2014; 8(4): e2769. https://doi.org/10.1371/journal.pntd.0002769 PMID: 24743730
11. Sorichetta A, Bird TJ, Ruktanonchai NW, Zu Erbach-Schoenberg E, Pezzulo C, Tejedor N, et al. Mapping internal connectivity through human migration in malaria endemic countries. Sci Data. 2016; 3:160066. https://doi.org/10.1038/sdata.2016.66 PMID: 27529469
12. Buonsenso D, Barone G, Onesimo R, Calzedda R, Chiaretti A, Valentini P. The re-emergence of dengue virus in non-endemic countries: a case series. BMC Res Notes. 2014; 7:596. https://doi.org/10.1186/1756-0500-7-596 PMID: 25186647
13. Semenza JC, Sudre B, Miniota J, Rossi M, Hu W, Kossowsky D, et al. International dispersal of dengue through air travel: importation risk for Europe. PLoS Negl Trop Dis. 2014; 8(12):e3278. https://doi.org/10.1371/journal.pntd.0003278 PMID: 25474491
14. Vasquez V, Haddad E, Perignon A, Jaureguiberry S, Brichler S, Leparc-Goffart I, et al. Dengue, chikungunya, and zika virus infections imported to Paris between 2009 and 2016: characteristics and correlation with outbreaks in the french overseas territories of Guadeloupe and Martinique. Int J Infect Dis. 2018; 72:34–9. https://doi.org/10.1016/j.ijid.2018.05.007 PMID: 29782922
15. Findlater A, Moineddin R, Kain D, Yang J, Wang X, Lai S, et al. The use of air travel data for predicting dengue importation to China: a modelling study. Travel Med Infect Di. 2019. https://doi.org/10.1016/j.tmaid.2019.07.002.
16. Brockmann D, Helbing D. The hidden geometry of complex, network-driven contagion phenomena. Science. 2013; 342(6164):1337. https://doi.org/10.1126/science.1245200 PMID: 24337289
17. Lawyer G. Measuring the potential of individual airports for pandemic spread over the world airline network. BMC Infect Dis. 2016; 16:70. https://doi.org/10.1186/s12879-016-1350-4 PMID: 26861206
18. Shi Y, Liu X, Kok SY, Rajarethinam J, Liang S, Yap G, et al. Three-month real-time dengue forecast models: an early warning system for outbreak alerts and policy decision support in Singapore. Environ Health Perspect. 2016; 124(9):1369–75. https://doi.org/10.1289/ehp.1509981 PMID: 26662617
19. Lee JS, Carabali M, Lim JK, Herrera VM, Park IY, Villar L, et al. Early warning signal for dengue outbreaks and identification of high risk areas for dengue fever in Colombia using climate and non-climate datasets. BMC Infect Dis. 2017; 17(1):480. https://doi.org/10.1186/s12879-017-2577-4 PMID: 28693483

20. Wang L, Wu JT. Characterizing the dynamics underlying global spread of epidemics. Nature communications. 2018; 9(1):218. https://doi.org/10.1038/s41467-017-02344-z PMID: 29335536

21. Meslé MMI, Hall IM, Christley RM, Leach S, Read JM. The use and reporting of airline passenger data for infectious disease modelling: a systematic review. Euro Surveill 2019; 24(31):1800216. https://doi.org/10.2807/1560-7917.ES.2019.24.31.1800216.

22. Lawyer G. Stochasticity in pandemic spread over the world airline network explained by local flight connections. ArXiv. 2015. https://doi.org/10.1186/s12879-016-1350-4.

23. Gardner LM, Bota A, Gangavarapu K, Kraemer MUG, Grubaugh ND. Inferring the risk factors behind the geographical spread and transmission of zika in the Americas. PLoS Negl Trop Dis. 2018; 12(1): e0006194. https://doi.org/10.1371/journal.pntd.0006194 PMID: 29346387

24. Lana RM, Gomes M, Lima TFM, Honorio NA, Codeco CT. The introduction of dengue follows transportation infrastructure changes in the state of Acre, Brazil: a network-based analysis. PLoS Negl Trop Dis. 2017; 11(11):e0006070. https://doi.org/10.1371/journal.pntd.0006070 PMID: 29149175

25. Silk MJ, Croft DP, Delahay RJ, Hodgson DJ, Weber N, Boots M, et al. The application of statistical network models in disease research. Methods Ecol Evol. 2017; 8(9):1026–41. https://doi.org/10.1111/2041-210X.12770.

26. Moilanen A, Nieminen M. Simple connectivity measures in spatial ecology. Ecology. 2002; 83(4):1131–45. https://doi.org/10.1890/0012-9658(2002)083[1131:SCMISE]2.0.CO;2.

27. Chapman DS, Makra L, Albertini R, Bonini M, Paldy A, Rodinkova V, et al. Modelling the introduction and spread of non-native species: international trade and climate change drive ragweed invasion. Glob Change Biol. 2016; 22(9):3067–79. https://doi.org/10.1111/gcb.13220.

28. Chapman D, Purse BV, Roy HE, Bullock JM. Global trade networks determine the distribution of invasive non-native species. Glob Ecol Biogeogr. 2017; 26(8):907–17. https://doi.org/10.1111/geb.12599.

29. ECDC. The european surveillance system (tessy): European centre for disease prevention and control 2019 [cited 2018 June 15]. Available from: https://ecdc.europa.eu/en/publications-data/european-surveillance-system-tessy.

30. Nichols GL, Andersson Y, Lindgren E, Devaux I, Semenza JC. European monitoring systems and data for assessing environmental and climate impacts on human infectious diseases. Int J Environ Res. 2014; 11(4):3894–936. https://doi.org/10.3390/ijerph110403894.

31. European Union. Commission implementing decision of 8 august 2012 amending decision 2002/253/ec laying down case definitions for reporting communicable diseases to the community network under decision no 2119/98/ec of the european parliament and of the council (notified under document c(2012) 5538) text with eea relevance, Pub. L. No. 2012/506/EU(2012). Available from: http://data.europa.eu/eli/dec_impl/2012/506/oj.

32. IATA. Passenger intelligence services (paxis): International Air Transport Association; 2019 [cited 2019 1 July]. Available from: https://www.iata.org/services/statistics/intelligence/paxis/Pages/index.aspx.

33. Denguemap [Internet]. 2017 [cited 11 July 2017]. Available from: https://www.healthmap.org/dengue/en/index.php.

34. Newman LP, Bhat N, Fleming JA, Neuzil KM. Global influenza seasonality to inform country-level vaccine programs: An analysis of who flunet influenza surveillance data between 2011 and 2016. PloS One. 2018; 13(2):e0193263. https://doi.org/10.1371/journal.pone.0193263 PMID: 29466459

35. Stanaway JD, Shepard DS, Undurraga EA, Halasa YA, Coffeng LE, Brady OJ, et al. The global burden of dengue: an analysis from the global burden of disease study 2013. Lancet Infect Dis. 2016; 16 (6):712–23. https://doi.org/10.1016/S1473-3099(16)00026-8 PMID: 26874619

36. Global burden of disease study 2016 results [Internet]. Institute for Health Metrics and Evaluation. 2016 [cited 25 Jun 2018]. Available from: http://ghdx.healthdata.org/gbd-results-tool.

37. Gardner L, Sarkar S. A global airport-based risk model for the spread of dengue infection via the air transport network. PloS One. 2013; 8(8):e72129. https://doi.org/10.1371/journal.pone.0072129 PMID: 24009672

38. Moore M, Gelfeld B, Okunogbe A, Paul C. Identifying future disease hot spots: infectious disease vulnerability index. Rand Health Q. 2017; 6(3):5. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5568150/. PMID: 28845357

39. Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, et al. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol. 2009; 24(3):127–35. https://doi.org/10.1016/j.tree.2008.10.008 PMID: 19185386

40. Nakagawa S, Schielzeth H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol Evol. 2013; 4(2):133–42. https://doi.org/10.1111/j.2041-210x.2012.00261.x.

41. Hartig F. Dharma: Residual diagnostics for hierarchical (multi-level/mixed) regression models CRAN: CRAN; 2019. Available from: http://florianhartig.github.io/DHARMa/.

42. Freeman LC. Centrality in social networks conceptual clarification. Soc Netw. 1978; 1(3):215–39. https://doi.org/10.1016/0378-8733(78)90021-7.

43. Borgatti SP. Centrality and network flow. Soc Netw. 2005; 27(1):55–71. https://doi.org/10.1016/j.socnet.2004.11.008.

44. Borgatti SP, Everett MG. A graph-theoretic perspective on centrality. Soc Netw. 2006; 28(4):466–84. https://doi.org/10.1016/j.socnet.2005.11.005.

45. Oldham S, Fulcher B, Parkes L, Arnatkevičiūtė A, Suo C, Fornito A. Consistency and differences between centrality measures across distinct classes of networks. PloS One. 2019; 14(7):e0220061. https://doi.org/10.1371/journal.pone.0220061 PMID: 31348798

46. Katz L. A new status index derived from sociometric analysis. Psychometrika. 1953; 18(1):39–43. https://doi.org/10.1007/BF02289026.

47. Bonacich P. Factoring and weighting approaches to status scores and clique identification. J Math Sociol. 1972; 2(1):113–20. https://doi.org/10.1080/0022250X.1972.9989806.

48. Andreotti J, Jann K, Melie-Garcia L, Giezendanner S, Abela E, Wiest R, et al. Validation of network communicability metrics for the analysis of brain structural networks. PloS One. 2014;(1932–6203 (Electronic)). https://doi.org/10.1371/journal.pone.0115503.

49. R-Core-Team. The R project for statistical computing: CRAN; 2019. Available from: https://www.r-project.org/.

50. Mollie E. Brooks. Kasper Kristensen, Koen J. van Benthem, Arni Magnusson, Casper W. Berg, Anders Nielsen, et al. Glmmtmb balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J. 2017; 9:378–400. https://journal.r-project.org/archive/2017/RJ-2017-066/index.html.

51. Gabor C, Tamas N. The igraph software package for complex network research. InterJournal. 2006; Complex Systems:1695. http://igraph.org.

52. Leifeld P, Cranmer SJ. Tnam: Temporal network autocorrelation models. 2017. https://cran.r-project.org/package=tnam.

53. Lüdecke D. Sjstats: Statistical functions for regression models (version 0.17.5). 2019. https://doi.org/10.5281/zenodo.1284472.

54. Wickham H. Ggplot2: Elegant graphics for data analysis: Springer-Verlag New York; 2016. Available from: https://ggplot2.tidyverse.org.

55. Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in R. Bioinformatics. 2014; 1367–4811:2811–2. https://doi.org/10.1093/bioinformatics/btu393.

56. World Health Organization. Alphabetical list of who member states 2019 [15 July 2019]. Available from: https://www.who.int/choice/demography/by_country/en/.

57. Gardner LM, Fajardo D, Waller ST, Wang O, Sarkar S. A predictive spatial model to quantify the risk of air-travel-associated dengue importation into the United states and Europe. J Trop Med. 2012; 2012:103679. https://doi.org/10.1155/2012/103679 PMID: 22523497

58. Huang Z, Wu X, Garcia AJ, Fik TJ, Tatem AJ. An open-access modeled passenger flow matrix for the global air network in 2010. PloS One. 2013; 8(5):e64317. https://doi.org/10.1371/journal.pone.0064317 PMID: 23691194

59. Liebig J, Jansen C, Paini D, Gardner L, Jurdak R. A global model for predicting the arrival of imported dengue infections. ArXiv. 2018. https://arxiv.org/abs/1808.10591.

60. Bogoch II, Maxim T, Acosta H, Bhatia D, Chen S, Huber C, et al. Potential plague exportation from madagascar via international air travel. Lancet Infect Dis. 2018; 18(3):247–8. https://doi.org/10.1016/S1473-3099(18)30077-X PMID: 29485085

61. Wilder-Smith A, Gubler DJ, Weaver SC, Monath TP, Heymann DL, Scott TW. Epidemic arboviral diseases: Priorities for research and public health. Lancet Infect Dis. 2017; 17(3):e101–e6. https://doi.org/10.1016/S1473-3099(16)30518-7 PMID: 28011234

62. Johansson MA, Powers AM, Pesik N, Cohen NJ, Staples JE. Nowcasting the spread of chikungunya virus in the Americas. PloS One. 2014; 9(8):e104915. https://doi.org/10.1371/journal.pone.0104915 PMID: 25111394

63. Heesterbeek H, Anderson RM, Andreasen V, Bansal S, De Angelis D, Dye C, et al. Modeling infectious disease dynamics in the complex landscape of global health. Science. 2015; 347(6227):aaa4339. https://doi.org/10.1126/science.aaa4339 PMID: 25766240

64. Rogers DJ, Suk JE, Semenza JC. Using global maps to predict the risk of dengue in Europe. Acta Tropica. 2014; 129:1–14. https://doi.org/10.1016/j.actatropica.2013.08.008 PMID: 23973561

65. ECDC. Aedes aegypti—current known distribution: July 2019. European centre for disease prevention and control 2019 [cited 2020 Feb 22]. Available from: https://www.ecdc.europa.eu/en/publications-data/aedes-aegypti-current-known-distribution-july-2019

66. Gossner CM, Ducheyne E, Schaffner F. Increased risk for autochthonous vector-borne infections transmitted by Aedes albopictus in continental Europe. Euro Surveill. 2018; 23(24):pii = 1800268. https://doi.org/10.2807/1560-7917.ES.2018.23.24.1800268

67. ECDC. Aedes albopictus—current known distribution: January 2019. European centre for disease prevention and control 2019 [cited 2020 Feb 22]. Available from: https://www.ecdc.europa.eu/en/publications-data/aedes-albopictus-current-known-distribution-january-2019

2.1.1 Supporting Information

S1 Fig. https://doi.org/10.1371/journal.pone.0230274.s001

S2 Fig. https://doi.org/10.1371/journal.pone.0230274.s002

S1 Appendix. https://doi.org/10.1371/journal.pone.0230274.s003

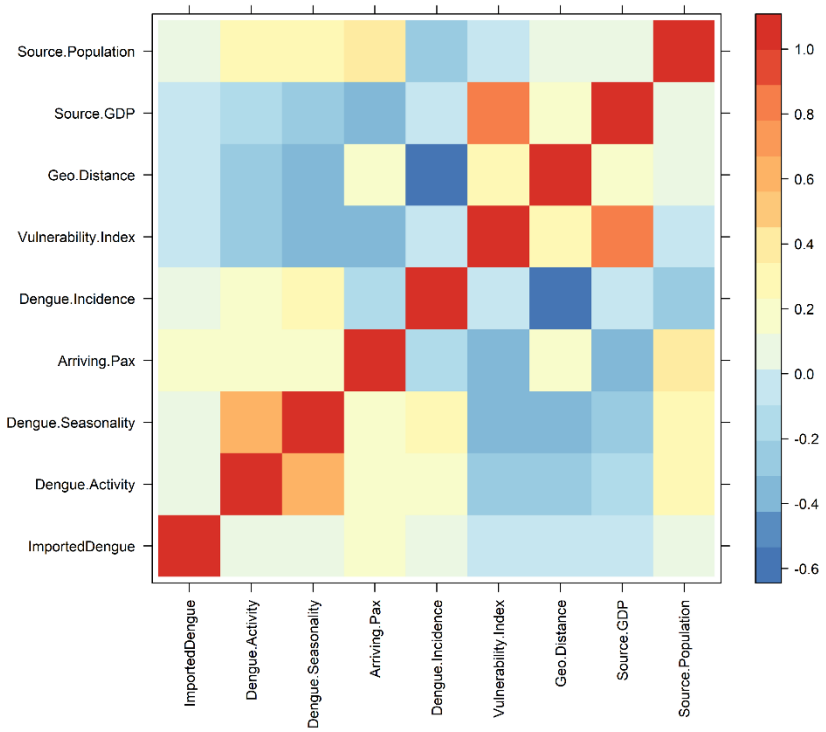

Figure 2-6 S1 Fig. Spearman correlations between dengue importation (numbers) and connectivity indices (variables).

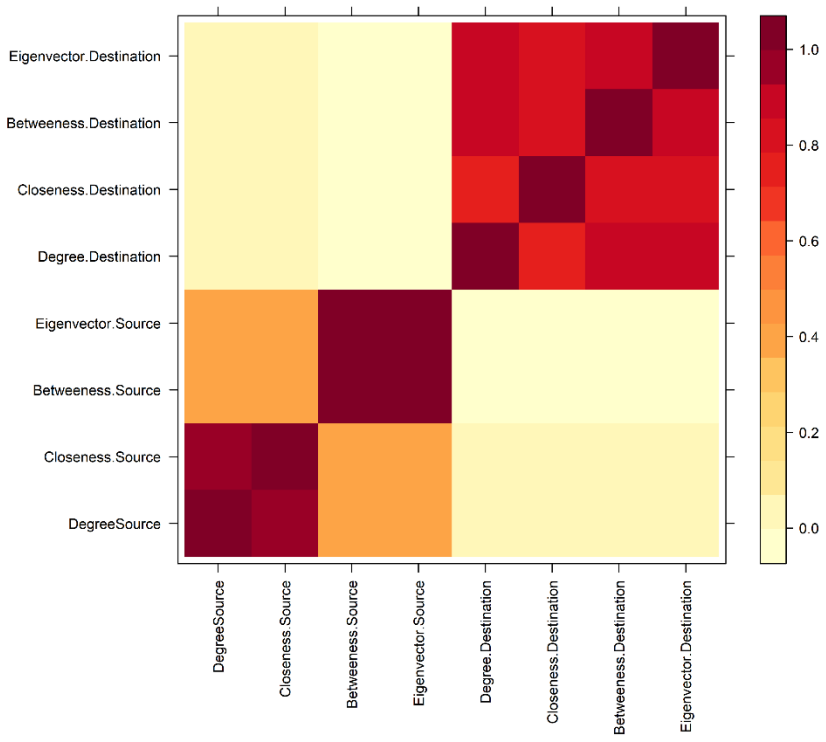

Figure 2-7 S2 Fig**.** Spearman correlations between network properties of the source and destination countries within the air transport networks**.**

Alternative specifications of the GLMM models, for correlated connectivity indices

Alternative scenarios were examined to assess the robustness of our results in the face of the high degree of correlation between the connectivity indices. As discussed in the main text, GDP and epidemic vulnerability displayed high correlations amongst themselves (Spearman's $p = 0.88$, $p < .001$). We consider scenarios where either connectivity index was included in separate models. In both instances we consider the goodness of fit, using the calculated marginal GLMM R2, to estimate the proportion of variation explained by the fixed effects and the resultant difference in their fixed effects Akaike information criteria (AIC).

Comparing both scenarios, most of the estimated coefficients (and odds ratio) of the connectivity indices are broadly stable and similar. However, the comparative fit test shows that the model with epidemic vulnerability had a slightly better fit than the model with GDP (S1 Table).

Table 2-5 S1 Table. Comparative-fit test for GLMM model including epidemic vulnerability and GLMM model including GDP**.**

	Model with epidemic vulnerability		**Model with GDP**	
	Coefficient (95% CI)	Odd ratio (95% CI)	Coefficient (95% CI)	Odd ratio (95% CI)
Connectivity Index				
Dengue activity	0.31 (0.18, 0.44)	1.36 (1.19, 1.55)	0.30 (0.17, 0.43)	1.35 (1.19, 1.54)
Dengue seasonality	0.12 (-0.01, 0.25)	1.13 (0.99, 1.29)	0.14 (0.01, 0.27)	1.15 (1.01, 1.31)
Incidence estimates	1.44 (1.33, 1.55)	4.23 (3.79, 4.72)	1.41 (1.29, 1.52)	4.08 (3.65, 4.56)
Geographical distance	-0.52 (-0.65, -0.38)	0.60 (0.52, 0.69)	-0.55 (-0.69, -0.42)	0.58 (0.50, 0.66)
Total Air passengers	2.01 (1.90, 2.13)	7.49 (6.68, 8.39)	2.03 (1.92, 2.14)	7.61 (6.81, 8.49)
Epidemic vulnerability	0.19 (0.10, 0.27)	1.20 (1.10, 1.32)		
GDP			0.26 (0.18, 0.33)	1.29 (1.20, 1.40)

Model Fit		
$R^2_{GLMM(m)}$	0.507	0.502
$R^2_{GLMM(c)}$	0.656	0.650
Fixed effects ΔAIC	0.0	26.7

Alternative specifications of the GLMM models, using count data

We did consider an alternative specification of our GLMM models, using count data (i.e. the actual imported case numbers), to explore any potential loss of information in the transformation of our response variable to a binary one. To do this, we first ran the preliminary analysis using a Poisson model on the count data to determine the prima facie dispersion status of the model and subsequently ran a negative binomial model, with log link functions. After which we ran the binomial model with logit link function, with the response variable as a binary response. In all instance we consider the goodness of fit, using the calculated marginal GLMM R^2, to estimate the proportion of variation explained by the fixed effects.

We did a comparative-fit test for both models (i.e. the count-data and the binary response), with the basic assumptions that either will lead to similar results in terms of estimated coefficient and their respective incidence rate ratio or odds ratio. The results of the negative binomial model, with log link functions (S2 Table), was similar to that of the binomial model with logit link function (reported in Table 3 of the main text) in terms of the strength, direction and statistical significance of the estimated coefficients.

However, the marginal GLMM R^2, of the binomial model with a logit link function was higher, indicating a better fit for the fixed effects (S3 Table). So, contrary to the transformation of the response variable captured more of the variation of the fixed effects as opposing a loss of information. It is also worth mentioning the logit link function had better convergence properties in comparison to the log link. Hence, we opted for the GLMM logit link function, as best fitting to our data structure.

Below is a table with the comparative-fit test for both modeling approaches, likewise we have included a table with the estimated coefficient and incidence rate ratio for the count-data

modeling approach to show the similarities with the logit-link results. If necessary, we can include this in the supporting information.

Table 2-6 S2 Table. Estimated coefficient and Incidence rate ratio for GLMM log link model (i.e. count data)

	Model 1		**Model 2**		**Model 3**	
	Coefficient (95% CI)	Incidence rate ratio (95% CI)	Coefficient (95% CI)	Incidence rate ratio (95% CI)	Coefficient (95% CI)	Incidence rate ratio (95% CI)
Connectivity Index						
Dengue activity	0.36 (0.22, 0.50)	1.44 (1.25, 1.65)	0.30 (0.16, 0.44)	1.35 (1.18, 1.56)	0.28 (0.13, 0.40)	1.32 (1.15, 1.52)
Dengue seasonality	-0.33 (-0.01, 0.25)	0.97 (0.85, 1.11)	-0.06 (-0.20, 0.08)	0.94 (0.82, 1.08)	-0.14 (-0.27, 0.00)	0.87 (0.76, 1.00)
Incidence estimates	1.49 (1.38, 1.61)	4.46 (3.99, 4.98)	1.97 (1.83, 2.11)	7.17 (6.22, 8.27)	2.01 (1.87, 2.15)	7.44 (6.48,8.55)
Geographical distance	-0.33 (-0.45, -0.20)	0.72 (0.64, 0.82)	-0.93 (-1.08, -0.78)	0.39 (0.34, 0.46)	-1.09 (-1.25, -0.94)	0.34 (0.29, 0.39)
Epidemic vulnerability	0.10 (0.01, 0.19)	1.11 (1.01, 1.21)	0.39 (0.29, 0.48)	1.47 (1.34, 1.62)	0.46 (0.37, 0.56)	1.59 (1.44, 1.74)
Total Air passengers	2.31 (2.20, 2.43)	10.09 (8.99, 11.32)	2.31 (2.17, 2.46)	10.12 (8.74, 11.71)	2.00 (1.83, 2.18)	7.41 (6.23, 8.82)
Centrality Measures						
Degree	—	—	—	—	0.54 (0.37, 0.72)	1.72 (1.45, 2.05)
Betweenness	—	—	—	—	1.09 (0.95, 1.23)	2.96 (2.57, 3.42)
Closeness	—	—	0.14 (0.01, 0.27)	1.15 (1.01, 1.30)	—	—
Eigenvector	—	—	1.10 (0.95, 1.25)	3.01 (2.59, 3.49)	—	—
Model Fit						

$\mathbf{R^2_{GLMM(m)}}$	0.443	0.492	0.490
$\mathbf{R^2_{GLMM(c)}}$	0.589	0.625	0.619
Fixed effects ΔAIC	269.7	59.1	0

Model 1: the base model of connectivity indices; **Model 2**: Base model with closeness and eigenvector centrality measures; **Model 3**: Base model with the degree and betweenness centrality measures. Log-link utilizes the count data, i.e. actual number of imported dengue cases; while the Logit-link utilizes a binary response variable, indicating an imported dengue case or not.

Table 2-7 S3 Table. General comparative-fit test for GLMM log-link (count data) & GLMM logit link (binary response) models.

Model Fit	**Model 1**		**Model 2**		**Model 3**	
	Log-link	**Logit-link**	**Log-link**	**Logit-link**	**Log-link**	**Logit-link**
$\mathbf{R^2_{GLMM(m)}}$	0.443	0.507	0.492	0.535	0.490	0.531
$\mathbf{R^2_{GLMM(c)}}$	0.589	0.656	0.625	0.676	0.619	0.673

Model 1: the base model of connectivity indices; **Model 2**: Base model with closeness and eigenvector centrality measures; **Model 3**: Base model with the degree and betweenness centrality measures. Log-link utilizes the count data, i.e. actual number of imported dengue cases; while the Logit-link utilizes a binary response variable, indicating an imported dengue case or not.

2.2 Predicting dengue importation into Europe, using machine learning and model-agnostic methods.

Salami, D., Sousa, C.A., Martins, M.d.R.O. et al. Predicting dengue importation into Europe, using machine learning and model-agnostic methods. Sci Rep 10, 9689 (2020). https://doi.org/10.1038/s41598-020-66650-1

SCIENTIFIC REPORTS
nature research

OPEN

Predicting dengue importation into Europe, using machine learning and model-agnostic methods

Donald Salami[1]✉, Carla Alexandra Sousa[1]✉, Maria do Rosário Oliveira Martins[1] & César Capinha[2]✉

The geographical spread of dengue is a global public health concern. This is largely mediated by the importation of dengue from endemic to non-endemic areas via the increasing connectivity of the global air transport network. The dynamic nature and intrinsic heterogeneity of the air transport network make it challenging to predict dengue importation. Here, we explore the capabilities of state-of-the-art machine learning algorithms to predict dengue importation. We trained four machine learning classifiers algorithms, using a 6-year historical dengue importation data for 21 countries in Europe and connectivity indices mediating importation and air transport network centrality measures. Predictive performance for the classifiers was evaluated using the area under the receiving operating characteristic curve, sensitivity, and specificity measures. Finally, we applied practical model-agnostic methods, to provide an in-depth explanation of our optimal model's predictions on a global and local scale. Our best performing model achieved high predictive accuracy, with an area under the receiver operating characteristic score of 0.94 and a maximized sensitivity score of 0.88. The predictor variables identified as most important were the source country's dengue incidence rate, population size, and volume of air passengers. Network centrality measures, describing the positioning of European countries within the air travel network, were also influential to the predictions. We demonstrated the high predictive performance of a machine learning model in predicting dengue importation and the utility of the model-agnostic methods to offer a comprehensive understanding of the reasons behind the predictions. Similar approaches can be utilized in the development of an operational early warning surveillance system for dengue importation.

The geographical spread of dengue fever is a global public health concern. This spread, particularly to non-endemic areas, has been largely facilitated by an increase in global trade and human mobility[1–3]. The expansion and connectivity of the global air transport networks in recent years, has played a key role in this spread[3]. In Europe, where dengue is not endemic, the number of travel-related cases of dengue, demonstrates how the air transport network has facilitated the spread of the disease. In the past decade, the European region has reported a significant number of imported dengue cases from epidemic/endemic tropical and subtropical countries[4]. Sporadic autochthonous transmissions have also been triggered by imported cases in areas with suitable environmental conditions and an established presence of the mosquito vector[5]. Recent examples include the autochthonous cases reported in France and Spain, which were linked to having originated from an imported case[6].

The mitigation of the continuous spread of dengue in Europe lies in part in the ability to effectively predict importation risk. However, a notable challenge in achieving this, is the complexity of global air transport networks, due to the dynamic nature and heterogeneity underlying the connections[1,7,8]. In recent times a range of modelling approaches, from the field of social network analysis, have been applied to understand the connection topology of the air transport network and their role in disease importation[9]. Unlike conventional statistical modelling approaches, these methods account for the co-dynamics of the network structure and how they interact with other risk factors to mediate the importation of dengue[10–12]. Our previous work[13] integrates this modelling approach and offers a foundational understanding of the importation patterns of dengue in Europe.

[1]Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Global Health and Tropical Medicine, Lisbon, 1349-008, Portugal. [2]Centro de Estudos Geográficos, Instituto de Geografia e Ordenamento do Território - IGOT, Universidade de Lisboa, 1600-276, Lisboa, Portugal. ✉e-mail: donald.salami@gmail.com; casousa@ihmt.unl.pt; cesarcapinha@campus.ul.pt

Conversely, an increasing number of studies are employing the use of machine learning algorithms to develop robust predictive models for dengue[14,15]. Machine learning algorithms are an applied extension of artificial intelligence. These algorithms build a mathematical model base, to automatically learn data patterns, adjust and perform inference, without explicit instructions[16]. Several studies have demonstrated the powerful predictive capabilities of machine learning models and their superiority over conventional statistical methods[17–19]. To this effect, some studies have applied them in the development of predictive models for dengue incidence[14,20–22]. A recent study by Chen *et al.*[15], utilized machine learning algorithms to develop a real-time model to forecast dengue in Singapore. Despite the high predictive performance of machine learning algorithms, they are not widely popular in epidemiological studies. This is likely to be in part because they are considered, to be "black-box" models with low interpretability, due to their complex inner workings. We argue different, that though machine learning models could fit complex relationships, several recent advancements have been made to aid the interpretation of these models[23]. To the best of our knowledge, machine learning algorithms have not been applied in modelling the risk of dengue importation for Europe.

Here, we aim to apply machine learning algorithms to develop a predictive model for dengue importation risk in Europe. To do so, we train a diverse set of machine learning algorithms, with historical data of dengue importation into Europe, connectivity indices of factors potentially mediating importation risk and centrality measures characterizing the air transport network. We then evaluate the predictive performance of the different models on a hold-out dataset to determine an optimal model. Finally, we employ the use of practical model-agnostic methods to interpret the optimal model's predictions.

Methods

Dengue data. We obtained monthly data for imported cases of dengue in Europe, for 2010–2015, from the European Centre for Disease Prevention and Control (ECDC)[24]. Here, we utilized confirmed dengue cases (as defined the European Union generic case definition for viral haemorrhagic fevers) with known travel history[25]. A total of 21 European Union/European Economic Area (EU/EEA) countries reported data on imported dengue, from a total of 98 different source countries between 2010 and 2015 (inclusive of zero reporting). The monthly level case counts were aggregated by country of infection (as source country) and the reporting country in Europe (as destination country). We transform the absolute count data into a binary response variable, that indicates whether there was an imported case of dengue (1) or not (0) in a destination country, in a month.

Air passenger's data. Comprehensive air passengers travel data for 2010–2015, was obtained from the International Air Transport Association (IATA)[26]. The data included true origin, connecting points and final-destination airports for all routes in the world and their corresponding passengers' volume. The data contains over 11,996 airports in 229 different countries and their territorial dependencies. The passengers' travel volume for each route worldwide was available at the country level and at a monthly timescale. This data was used to construct a monthly passenger flow from all countries worldwide with a final destination in Europe (accounting for all connecting flights) between 2010 and 2015. The data also included the passengers' flow between European countries.

Connectivity indices between a source and destination country. Drawing on the underlining concept of spatial interaction modelling, that inflow between two locations is a function of the attributes of the source and destination and their corresponding interaction[27]. Connectivity indices between a source country and a destination country in Europe were previously developed[13], using different factors that potentially mediate dengue importation risk. The indices were decomposed into components describing the source strength' (the risk of dengue infection) and the transport or importation potential (the connection between a source country and a potential destination country in Europe). Source strength for all indices was modelled to represent the endemicity of dengue in a source country. While transport and importation potential were modelled to characterize seasonal dengue activity, incidence rates, geographical proximity, epidemic vulnerability, air passenger volume, population size and wealth of a source country as mediating risk factors. The connectivity indices and their descriptions are listed in Table 1.

Centrality measures of the air transport network. Using the monthly air passenger's data, we constructed a weighted directed network. The network for each month was denoted by $G_m = (V, E)$, where V_G is a set containing all the nodes (or vertices), while E_G contains all the edges, with m indicating the month (m = 1, 2, 3… 72, covering the years of 2010–2015). Nodes represented all countries worldwide, while edges represent the flow of passengers from a source country to a destination country in Europe. Four different centrality measures were used to analyse the network and quantify the capacity of a source node to influence transportation of dengue or a destination node's propensity to receive an imported case of dengue, by virtue of their connection topology within the network. Centrality measures and their descriptions are listed in Table 1.

Feature (variable) engineering. One fundamental step in building machine learning models is the process of feature engineering, i.e. using domain-specific knowledge to create new features (i.e. variables) or transform and encode existing original data into a more informative format[28,29]. For this analysis, we created an additional variable based on our a priori knowledge of the data. The reporting rate for dengue data was heterogeneous across the EU/EEA member countries, as some countries were not consistent in monitoring and reporting to the ECDC (we considered "zero reporting", which designate that no imported case was recorded during the reporting period, but a report was submitted to the ECDC). Hence to account for this variability in reporting rates, we created a time effect variable – coded as a set of dummy variables, 1 for a given month and 0 for all other months. This variable not only controls for reporting rates, but for other potential time-specific effects that might be

Table 2-8 Descriptions of the variables in the dataset.

Dataset	Variable name	Description	Data source
Connectivity indices	Dengue activity[a]	Notification of one or more confirmed cases in the given month (January-December) in the source country.	HealthMap[30]
	Dengue seasonality[b]	Notification of one or more confirmed cases (dengue activity) in a given month, for two or more years from 2010 through 2015, in the source country.	HealthMap[30]
	Dengue Incidence	Annual dengue incidence estimates of the source country.	IHME[31]
	Geo Distance	Geographical (great circle) distance between centroids of the source country and destination country (spatial connectivity).	CEPII[32]
	Vulnerability index	The epidemic vulnerability of the source country.	RAND[33]
	Source GDP	Gross domestic product (wealth) of the source country.	World Bank[34]
	Source Population	The total population of the source country.	World Bank[34]
	Arriving Pax	Total air passengers from source country to a destination country in Europe.	IATA[26]
Centrality measures[c]	Degree	The number of links or connections that a node has.	Calc.[d]
	Betweenness	The number of times a node lies on the shortest path (geodesics) between other nodes in the network.	Calc.
	Closeness	The average number of steps required to access every other node from a given node.	Calc.
	Eigenvector	The combined measure of how many connections a node has (i.e. its degree) and the centrality of the other nodes that it is connected to.	Calc.
Created variable	Time effect[e]	Set of dummy variables to account for time effects (monthly timescale).	

Table 1. Descriptions of the variables in the dataset. [a]Coded as a binary variable, indicating an activity (1) or not (0); [b]Coded as a binary variable, indicating a seasonal pattern (1) or not (0); [c]Centrality measures for source and destination countries were added as separate variables in the dataset; [d]Calc= calculated from the air passengers' data; [e]Set of 72 dummy variables, coded as 1 for a given month and 0 for all other months. Detailed mathematical equations for all variables can be found in Salami *et al.*[13].

restricted to a given period. Time events that might increase or decrease passengers' traffic to a specific country, and in turn affect dengue importation (e.g. the introduction and/or discontinuation of an airline carrier or route).

Data pre-processing and splitting. The dataset used to build our machine learning models consists of the connectivity indices and centrality measures of the air transport network. The single unit of analysis is a source-destination country pair, at a monthly timescale, and a binary response variable coded to indicate an imported dengue case (1) or not (0).

Before model training, we performed the following pre-processing analyses to the full dataset. First, we examine the correlation between our predictor variables, by using Spearman's correlation to rank the statistical dependencies. Several pairs of continuous variables displayed moderate-to-high pairwise correlations. Figure 1 shows the correlation matrix between the continuous predictors in our dataset. Most of the centrality measures (for source and destination countries) were highly correlated, example the betweenness and eigenvector centrality for source countries had a Spearman's $\rho = 0.99$ This is not unusual, as on average, centrality measures are highly correlated in a network[35]. The highly correlated pairs are practically redundant in conventional regression modelling and the heuristic approach to dealing with this is to exclude one. However, in our case, we did not manually exclude variables, as we focus on the predictive power of the entire bundle of variables as oppose the estimated coefficients of individual variables. Likewise, the suite of algorithms compared, each utilizes a combination of inbuilt feature selection and penalization functions to exclude redundant variables in their ensembling and mitigate the effect of multicollinearity.

Next, we randomly split the dataset into two sets, 70% into a training subset and 30% into a testing subset. This split was done based on the distribution of our outcome variable (i.e. binary response of an imported case of dengue (1) or not (0)), with sampling occurring within each category, thereby preserving the overall class distribution of our data. Our full dataset contained a total of 2, 055 unique country pairs (i.e. source - destination country pair), with a corresponding total of 147,960 monthly observations. An imported case was recorded in 1,937 observations, i.e. 1% of the total observations had an imported case. These observations were split into 1366 for the training subset and 573 for the testing subset, at a 70:30 ratio (a similar split was done for the observations with no imported cases). The training subset is used to build and tune the various machine learning models, while the test subset is used to evaluate the predictive performance of the models.

Due to the imbalance class distribution of the outcome variable, we model this as a rare event, using a post hoc sampling approach to attenuate the effects of the imbalance during model training[36]. The synthetic minority over-sampling technique (SMOTE), was used to subsample the training subset data to create a roughly equal distribution within the classes. SMOTE utilizes a hybrid of either up-sampling, to synthesize new data points in the minority class or down-sampling, to down-size the majority class[37]. Our training dataset was balanced to a 3:4 ratio of an imported case to no imported case. The testing subset was maintained to reflect the original imbalance as a quality assurance of the predictive model performance. Lastly, before training, we apply a data transformation on all continuous variables in the dataset, by centring and scaling them. This transformation ensures that

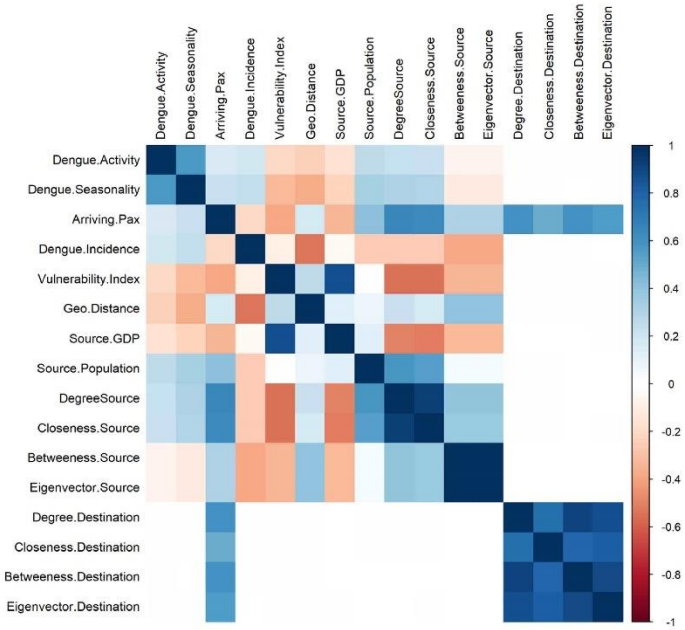

Figure 1. Spearman correlation matrix of continuous variables. Correlation is computed from the full dataset and coloured according to magnitude. Blue colours indicate strong positive correlations, red indicates strong negative correlations, and white implies no empirical relationship between the variables. Figure was generated using R programming language version 3.6.1[39].

the variables have a zero mean and a common standard deviation of one, thereby improving normality and the numerical stability of the model calculations[38].

Model selection. Our training model is a classification-based model, to predict the probability that a case of dengue is imported into a country in Europe. There are several classification techniques (or classifiers) employed in machine learning models. The choice of the suite of algorithms we tested, was a trade-off between, meta-algorithm that fits our classification problem and those with built-in feature selection. Other algorithmic and systematic features that were considered include regularization (to handle the effects of multicollinearity), hyperparameter optimization (model tuning capabilities), and efficient computation time. To build our predictive model, we compare four widely used classifiers algorithms in machine learning, as listed below:

Partial least squares (*pls*) implements a supervised version of principal component analysis, using a dimension reduction technique. This technique first summarizes the original variables into a few new variables called principal components (PCs), as supervised by their relationship to the outcome variable. These components are then used to fit a linear regression model[40]. For classification problems, the partial least squares discriminant analysis variant is fitted. This method has an embedded feature selection and regularization[41].

Lasso and elastic-net regularized generalized linear models (*glmnet*) implement a logistic generalized linear model via penalized maximum likelihood. The addition of a penalty shrinks the coefficients of the less contributive variables toward zero (L2 ridge penalty) or absolute zero (L1-Lasso penalty)[42]. The *glmnet* implements a combination of both L1 & L2 penalties (otherwise called elastic net penalty), for its regularization and simultaneous feature selection.

Random forest (*randomForest*) is a bootstrap aggregated (or bagged) decision tree-based ensemble technique. The algorithm constructs multiple decision trees by repeat resampling of the training dataset and outputs the mode of the classes as a consensus prediction. The trees are created independently from a random vector distribution; hence each tree is heterogeneous with high variance and casts a unit vote for the most popular class[43]. By averaging several decision trees, it intuitively avoids overfitting and performs an embedded feature selection.

Extreme gradient boosting (*xgboost*) implementation of a gradient boosted decision trees ensemble technique. The gradient boosting framework iteratively refines its model, to create a strong classifier by combining multiple weak classifiers in a stage-wise manner to minimize the loss function[44,45]. The *xgboost* algorithm is a commonly preferred classifier, because it utilizes parallelization and distributed computation for implementation, thereby ensuring high efficiency in computation time and resources[45–47].

Figure 2-8 Spearman correlation matrix of continuous variables.

Model tuning and validation. Machine learning models can be prone to overfitting, to mitigate this we implemented a model building approach that encompasses model tuning and repeated evaluation during training. We use a methodological resampling technique of the training dataset, i.e. five repeats of 10-fold cross-validation (CV). The 10-fold CV randomly partitions the training dataset into 10 sets of roughly equal size, one set retained, and the others used to fit a model. The retained set is used to estimate model performance. The first set is then returned to the training set and the procedure iterated until each set has been used for validation. This whole process is repeated five times before results are aggregated and summarized. This procedure automatically chooses tuning parameters associated with optimal model performance.

Candidate models were evaluated using the following performance metrics: area under the receiving operating characteristic curve (AUC), sensitivity (true positive rate), and specificity (false positive rate). Our final candidate model was selected based on the receiving operating characteristic curve (ROC) threshold, which maximizes the trade-off between sensitivity and specificity[48]. The ROC curve evaluates the class probabilities across a continuum of thresholds, with an arbitrary (algorithmically set) "optimal" cut point for determining what percentage of probability is accepted in classifying an imported case of dengue.

Model interpretability. We utilized the model-agnostic approach to provide an interpretation of our optimal model. Model-agnostic methods work by extracting post-hoc explanations from an original machine learning model[23]. This involves training an interpretable model on the predictions of the original model[49] and/or by changing the inputs of the original model and measure the changes in the prediction output[50]. We employ the use of recent model-agnostic tools[51,52] with both global and local scale interpretability functions. Global interpretation helps to understand the modelled relationship and distribution of the predicted target outcome (i.e. dengue importation) based on the input variables, while local interpretation zooms in, to help understand model predictions for a single instance (i.e. a single unit of observation or analysis).

We obtained global interpretations of our final candidate model through the following, variable importance, and partial dependence plots[49,53]. Variable importance measures the contribution of each input variable, by calculating the increase in the model's prediction error after permuting the variable[54]. While the Partial dependence plots (PDP) are graphical renderings of the prediction function that helps visualize the relationship between the variables and predicted outcome[46,55,56]. The relative importance of each variable is normalized to have a maximum value of 100, with higher scores indicating the most influential variable. We note that it may not be feasible to explore in detail the relationship of all variables in our model. Hence, we set an arbitrary cut-off on the variable importance measures at a value >50, to determine a subset of variables to focus on.

Local interpretation of our model was implemented via the use of local surrogate models, otherwise called-Local interpretable model-agnostic explanations (LIME)[23,50]. The underlining assumption of LIME is that complex black box models are linear on a local scale, hence a simple (surrogate) model can be fitted for an individual observation that mimics the behaviour of the global model at this locality. The simple model and its variable weights are then used to explain the individual predictions locally. To demonstrate the LIME technique, we selected 10 single observations from our initial testing subset. These observations were sampled methodologically to include both classes (i.e. imported case [1] or not [0]) and representative of countries with a high and low frequency of dengue importation. We set the number of variables to best describe the predicted outcome, as the 5 most influential. The resulting weights for these variables are plotted to explain the local behaviour of the model. The plots delineate if a variable increase or decreases the predicted probability of an imported case of dengue (detailed vignette for the LIME techniques can be found here[57,58].

Statistical software. All statistical analyses were performed with R Programming Language version 3.6.1[39]. For uniformity in our model build, we utilized the classification and regression training (*caret*) R package, this is an interface to a vast amount of available machine learning algorithms[38]. The package streamlines the process of building and validating predictive models by using a set of intuitive call functions. Supporting packages for specific functions includes: *pls*[41], *glmnet*[42], *randomForest*[59], *xgboost*[60], *plyr*[61], *doSNOW*[62], *DMwR*[63], *pROC*[64], *pdp*[56], *iml*[51], *lime*[52] and their various dependencies.

Results

Model prediction performance. We compared the prediction performance of the different classifiers' algorithms, in their ability to predict an imported case of dengue. Models were evaluated with the testing dataset, utilizing the area under the receiver operating characteristic curve (AUC) as the quantitative measure for performance comparisons. All four models performed comparably well, with AUC scores above 0.80 (Table 2). AUC score using *pls* was 0.88 (95% CI, 0.86 to 0.90); *glmnet* was 0.89 (95% CI, 0.87 to 0.91); *randomForest* was 0.97 (95% CI, 0.96 to 0.98); and *xgboost* was 0.97 (95% CI, 0.96 to 0.98). Performance metrics for each model are depicted in Table 2. Figure 2 also shows the ROC curve plots for the different models.

The AUC score indicates that predictions from the *randomForest* and *xgboost* models were better fitted to the dataset, outperforming the *pls* and *glmnet* models (with the pls being the least fitted). The *randomForest* and *xgboost* models had similar performance across the metrics with nearly negligible differences (Table 2). However, they had a distinction in their ROC curves effective threshold (Fig. 3). The ROC threshold appropriately maximizes the trade-off between sensitivity and specificity. With the best threshold cut-off at 68% (i.e., only probabilities greater than 0.68 were classified as an imported case of dengue, Fig. 3a) the *xgboost* model outperforms the *randomForest*, (cut-off at 0.64, Fig. 3b) in a competitive comparison of prediction accuracy. With a true positive rate of 0.88 and a false positive rate of 0.12, the *xgboost* model was selected as the optimal model for our dataset. Hence, our final predictive model was able to predict 88% of dengue importation cases in our test dataset accurately (Fig. 3a).

Table 2-9 Comparison of the prediction performance of the different models.

Model	AUC (95% CI)[a]	Sensitivity (95% CI) [a]	Specificity (95% CI) [a]
pls	0.88 (0.86–0.89)	0.75 (0.71–0.78)	0.84 (0.83–0.84)
glmnet	0.89 (0.87–0.90)	0.77 (0.73–0.80)	0.84 (0.83–0.84)
randomForest	0.94 (0.93–0.95)	0.79 (0.75–0.82)	0.92 (0.91–0.92)
xgboost	0.94 (0.94–0.95)	0.79 (0.76–0.82)	0.93 (0.92–0.93)

Table 2. Comparison of the prediction performance of the different models. **AUC** = area under the ROC curve; **Sensitivity** = rate of an imported case predicted correctly; **Specificity** = rate that non-imported cases are predicted correctly. [a]95% Confidence interval.

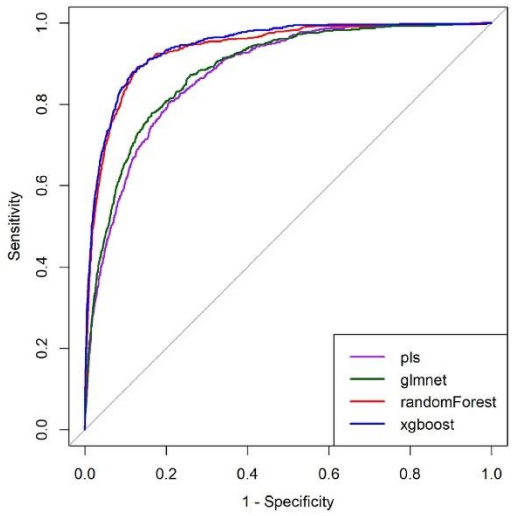

Figure 2. Comparison of the receiver operator characteristic (ROC) curve for the different models. Curves characterize the trade-off between the sensitivity (true positive rate) and 1minus specificity (false positive rate). The *y*-axis = sensitivity and the *x*-axis = 1 minus specificity.

Model interpretability. The best performing model (i.e. *xgboost*), was examined further for interpretation. Figure 4 illustrates the variable importance scores for the10 most influential variables, from our optimal (i.e. *xgboost)* model. With an arbitrary cut-off at >0.50, the subset of our 'most important' variables included the following: Source country's dengue incidence rate; population; the number of arriving passengers; betweenness, closeness and degree centrality measures of the destination country. Figure 5 illustrates a visual representation of the relationship between this subset of variables and the predicted response while accounting for the average effect of the other predictors in the model. These plots demonstrate that the probability of an imported case of dengue increases on average for source countries with higher incidence rates, large population size and higher passenger traffic. Likewise, the probability increases for destination countries with higher betweenness, closeness and degree centrality measures.

Note: to guard against over-interpreting the partial responses, we added rug displays to the plots, i.e. tick marks indicating the minimum, maximum and deciles of the variable distribution. Our interpretation of the partial response was limited to the regions within the minimum and maximum.

In addition to providing global explanations of our optimal model, we provide a local explanation for individual predictions given for 10 single observations, i.e. a single unit of analyses, source–destination-month combination (Table 3). Figure 6 gives a visual representation of the first four single observations in local subset data, each plot shows the predicted probability of each observation, being an imported case of dengue. Likewise, it shows the five most influential variables that best explain the model's prediction at the local region of the single observation, and whether these variables increase (supports) or decrease (contradicts) the probability of an imported case of dengue. With these results, we can infer that for case 1 (i.e. Indonesia-to-Germany, for February 2010), the local model-predicted probability of being an imported case, was 94%. The top five variables influencing this probability were: closeness and betweenness centrality measures of Germany, the incidence rate of dengue and closeness centrality measures of Indonesia and the geographical distance between both countries. Conversely, case 3 (i.e. Tanzania to the United Kingdom, for December 2014), had a similar set of variables as most influential,

Figure 2-9 Comparison of the receiver operator characteristic (ROC) curve for the different models

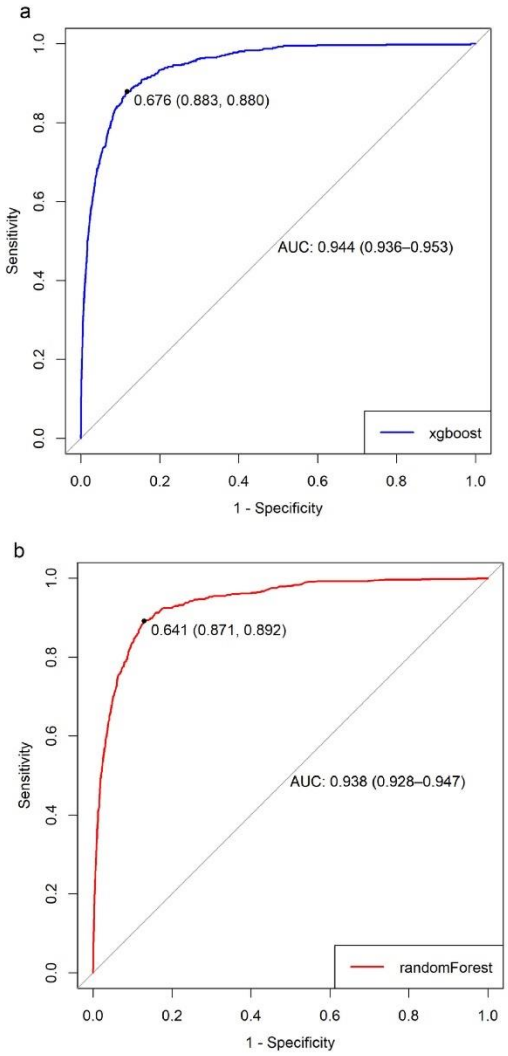

Figure 3. Comparison of the receiver operator characteristic (ROC) curves for extreme gradient boosting and random forest models. The dot on both plots indicates the value corresponding to the "best" cut-off point threshold for each model that appropriately maximizes the trade-off between sensitivity and specificity. The numbers in parentheses are (specificity, sensitivity). Extreme gradient boosting (**a**) cut-off was at 68% (i.e., probabilities greater than 0.68 are classified as an imported case of dengue), delivering a specificity of 0.883, sensitivity of 0.880, while random forest (**b**) cut-off was at 64%.

however, the incidence rate of Tanzania decreases the probability of having an imported case. This demonstrates how variables influencing predictions for a single observation can differ at the local scale. Finally, to ensure the trustworthiness of the local model, we compared the predicted probability of the local model to that of the global optimal model for each observation. There was no difference in the predicted probabilities of the global and local model (analytical comparison not shown). Overall, the local interpretation provides insights into the variations of the individual predictions and provides an important aspect to assuring trust of the model.

Figure 2-10 Comparison of the receiver operator characteristic (ROC) curves for extreme gradient boosting and random forest models.

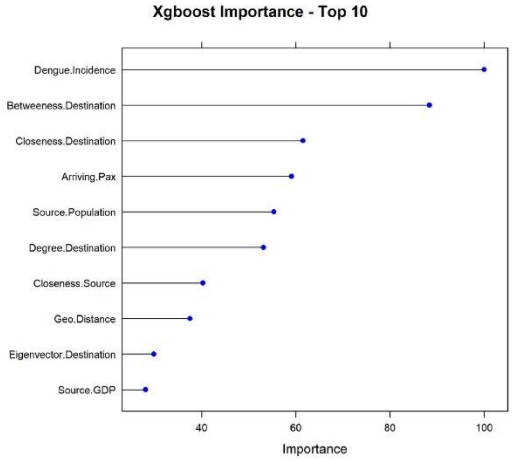

Figure 4. Variable importance plots. Top 10 most influential variables from the extreme gradient boosting model. The relative importance of each variable is normalized to have a maximum value of 100, with higher scores indicating the most influential variable.

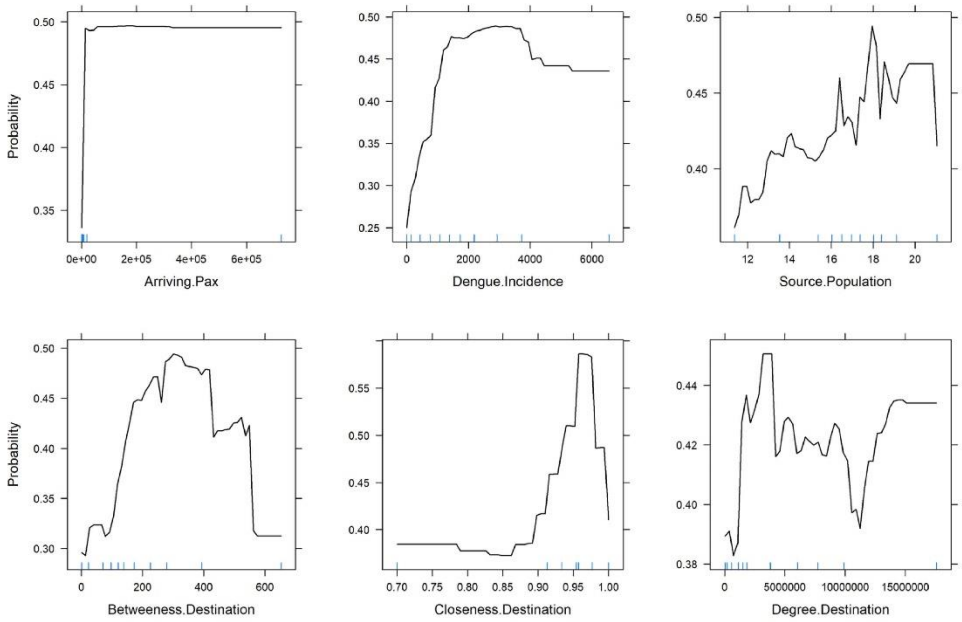

Figure 5. Partial dependence plots for a sub-set of the most influential variables in the optimal model predicting the probability of an imported case of dengue. The optimal model is the *xgboost* model. Sub-set variables represent variables with a variable importance ranking score >50. *Y*-axis is set on a probability scale since our model was a classification model; Blue rug marks at the inside bottom of plots indicate the min/max and deciles of the variable distribution. Top plots show the 3 most influential connectivity indices, while bottom plots show the 3 most influential network centrality measures.

Figure 2-11 Variable importance plots. Top 10 most influential variables from the extreme gradient boosting model.

Table 2-10 Ten selected individual observations for LIME model

Case #	Source country	Destination Country	Month[a]	True Class[b]	Predicted Class[c]	Probability[d]	Top Five variables from LIME[e]
1	Indonesia	Germany	2	1	1	0.94	Supports: Closeness. Destination; Geo Distance; Betweenness. Destination; Closeness. Source; Dengue Incidence
2	Brazil	Norway	61	1	1	0.80	Supports: Closeness. Destination; Geo Distance; Source GDP; Dengue Incidence Contradicts: Betweenness. Destination
3	Tanzania	United Kingdom	60	1	1	0.91	Supports: Closeness. Destination; Geo Distance; Betweenness. Destination; Source Population Contradicts: Dengue Incidence
4	India	Sweden	39	1	1	0.92	Supports: Closeness. Destination; Geo Distance; Source GDP; Dengue Incidence Contradicts: Betweenness. Destination
5	Thailand	Italy	34	1	1	0.90	Supports: Closeness. Destination; Geo Distance; Betweenness.Destination; Source GDP Contradicts: Dengue Incidence
6	Vietnam	France	20	1	1	0.94	Supports: Closeness. Destination; Geo Distance; Source Population; Source GDP Contradicts: Dengue Incidence
7	Brazil	Portugal	43	0	0	0.53	Supports: Geo Distance; Dengue Incidence Contradicts: Closeness. Destination; Betweenness. Destination; Source GDP
8	Columbia	Spain	14	0	0	0.52	Supports: Closeness. Destination; Geo Distance; Betweenness. Destination; Source GDP Contradicts: Dengue Incidence
9	Philippines	Austria	47	0	1	0.77	Supports: Closeness. Destination; Geo Distance; Betweenness. Destination; Source GDP; Source Population
10	Venezuela	Ireland	66	0	0	0.13	Supports: Betweenness. Destination; Dengue Incidence; Closeness. Source Contradicts: Closeness. Destination; Geo Distance

Table 3. Ten selected individual observations for LIME model (unit of source–destination-month combination). [a]Month case was reported in the destination country, 1–72 months covering the years of 2010–2015, e.g. month 1 = January 2010; [b]Original classifications of the case, in the test dataset, imported case [1] or not [0]; [c]Predicted class in LIME model, imported case [1] or not [0]; [d]Prediction probability of an imported case of dengue from LIME model; [e]Top five most influential variables, delineated by if variable increases (supports) or decreases (contradicts) the probability of an imported case of dengue.

Discussion

Our study demonstrates the use of machine learning modelling approach to predict the probability of having an imported case of dengue in Europe. Using historical dengue importation data, we trained and evaluated four machine learning classifiers algorithms, to develop an optimal predictive model. Our best-performing model was the extreme gradient boosting model with an AUC score of 0.94. Our choice of best performing model was not just based on the AUC results, as this score does not necessarily guarantee the best classifier. Given that our prediction target is the probability of having an imported case of dengue, we expected that our final model performs better in classifying the true positive cases (i.e. maximizing sensitivity). This was achieved by the effective probability threshold of the ROC curve, which offered a trade-off between sensitivity and specificity. Utilizing this we were able to maximize the sensitivity of our optimal model, to correctly predict the probability of an imported case of dengue with an 88% accuracy rate. As a first attempt to train a machine learning model for dengue importation in Europe, we can safely state that our model provides a benchmarking result for predictive performance. Given this predictive performance, our model holds great potential as a forecasting tool which can markedly improve dengue surveillance in Europe.

A limitation of other previous machine learning models is that they deliver high predictive accuracy without explaining why certain predictions are made[14,65]. In this paper, we posit for both accuracy and interpretability of our final model. We focus on demonstrating the practical explanation of our model predictions using recent model-agnostic approaches[23]. Our model included 17 predictor variables, which broadly captures the importation risk factors (as presented by the connectivity indices) and the influence of the air transport network (i.e. the centrality measures). These variables were chosen to reflect the factors known or hypothesized to be relevant to the importation dynamics. So firstly, we provided an overall quantification of the relationship between our model variables and the predicted outcome, by ranking them in terms of their importance. With a further exploratory analysis (via the PDPs visualizations) concentrating on a sub-set of the most influential variables. On average our model predicted a higher probability of an imported case of dengue, from a source country with high dengue incidence rates, large population, and high air passenger volume. These findings support a priori expectation for these factors to increase the importation risk of dengue and are consistent with other studies[12,13,66]. Also, our model predicts a higher probability of an imported case for destination countries with high connectivity within the network (as measured by degree centrality), with a putative connection hub to other countries (betweenness centrality) and connects to other countries in a relatively short amount of time (closeness centrality). Intuitively, this is expected, given that the network centrality of the destination country was modelled to act as a proxy for a country's propensity to receive an imported case. Hence, destination countries with higher connectivity within

Figure 2-12 Partial dependence plots for a sub-set of the most influential variables in the optimal model predicting the probability of an imported case of dengue.

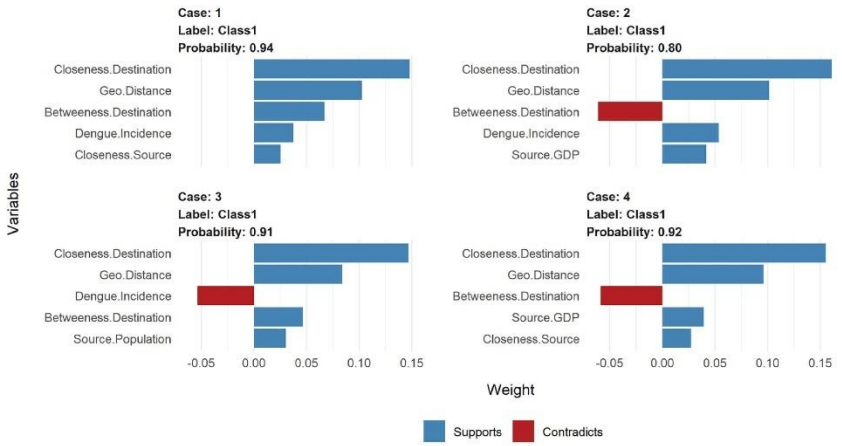

Figure 6. LIME model plots explaining individual predictions. Cases 1–4, as presented in Table 3. Each plot shows the 5 most influential variables that best explain the prediction in the local region. The blue bars represent variables that increase the predicted probability (supports), the red bars represent variables that decrease the probability (contradicts).

the network have an increased risk of importation. These findings are similar to other previous studies that characterize the role of the air transport network structure in mediating epidemic spread[10,11] and collaborates the results of our previous work[13].

The above explanations capture the relative contribution of the input variables in predicting the importation of dengue at an aggregated level for Europe. However, the risk probabilities will differ at the country level due to changes in the dynamic attributes of the different country pairs. For example, the dengue activity or seasonality in a source country can vary relative to time (oblivious to high or low incidence rates). Hence, it will be expected that importation risk based on seasonality will have temporal differences between source countries. Also, the volume of air passengers between country pairs are heterogeneous, which can be largely determined by different monthly traffic flows. Also, the topological profile of the air travel network will differ across countries with changes in passenger volume. These heterogeneities may affect the prediction of dengue importation at the country level, with certain variables supporting or contradicting depending on the country pairs in consideration. Using the local interpretable model-agnostic explanations, we were able to assess which variables are most influential on the predictions at a temporal and country-pair level. As illustrated from the examples in Table 3, the probability risk prediction for different country pairs was increased or decreased by different variable combinations at discrete points in time. For example, the predicted probability of a case of dengue from Indonesia (case 1 in Table 3) and Tanzania (case 3 in Table 3), were similar, but vary in the risk factors mediating this prediction. The lower dengue incidence rates in Tanzania (relative to Indonesia), decreases the probability of an importation, however other variables pose an increased risk for an imported case of dengue. So, the local explanations provide insights into the heterogeneities of importation risk at the different country pairing levels. This type of specific country-pair results can be useful in profiling importation risk from a specific source country or region; similar to the route-level risk assessment discussed by Gardner *et al.*[10]. This kind of information will be useful in guiding the implementation of targeted surveillance and public health preparedness in destination countries. Also, it can guide policy decisions at the European regional level, on how to effectively appropriate surveillance resources for countries at higher risk of dengue importation.

Overall, our work offers the following contributions. First, to the best of our knowledge, this is a first attempt at applying machine learning algorithms to model the risk of dengue importation into Europe. Second, we implement and demonstrate how to apply model-agnostic approaches for obtaining both aggregate (global) and individual (local) level explanations. Finally, we identify and interpret the factors influencing dengue importation through air travel at a temporal and country-pair level. When combined, these contributions can assist public health practitioners looking to develop a reliable, cost-effective, and scalable early warning surveillance system for dengue importation. Although we mainly focus on the theoretical framework of the model using historical data, the results demonstrate that the model can be applied for real-time prediction, assuming the availability of real-time data. However, there are some limitations to this work, that is worth noting for improvement in future research: (1) Dengue incidence rate for source countries was aggregated at a yearly scale, due to paucity of surveillance data at a similar scale to dengue case data (i.e. monthly). This may have overestimated or underestimated the actual effect of incidence and potentially impact the predicted risk probability from a source country. Our approach compensated for this limitation by the inclusion of the dengue activity and seasonality variables. Even though this does not necessarily capture the variability of a finer scale but serves as a proxy. (2) We only evaluated our prediction as a binary outcome (i.e. the probability of an imported case or not) and not a numeric outcome

Figure 2-13 LIME model plots explaining individual predictions.

like other similar models for dengue incidence[14,20]. A numeric outcome prediction can be achieved by modifying our model training approach from a classification model to a regression model. Even though, the additional benefit (if any) of predicting a discrete number versus a probability estimate is subjective. However, we do submit that while our approach could serve as a benchmark, we encourage alternative exploration for improved performance and accuracy.

In conclusion, our study demonstrates the efficient and powerful predictive capabilities of machine learning models in predicting the importation of dengue in Europe. Using historical dengue importation data, connectivity indices and air transport network centrality measures, we trained and evaluated a classification model to predict the probability of an imported case of dengue. Then applying recent model-agnostic interpretability approaches we provided an in-depth explanation of the model's predictions. With the predictive model and model-agnostic interpretability tools at hand, this can be applied at a regional or country level to develop a forecasting tool for dengue importation. Assuming the availability of real-time data, the methods described in this paper can be explored as a technique for developing a real-time early warning surveillance system for dengue importation.

Data availability

The air travel data used in this study, cannot be shared publicly because of a nondisclosure agreement with the International Air Travel Association (IATA). The same data can be purchased for use by any other researcher by contacting the International Air Travel Association (IATA)- Passenger Intelligence Services (PaxIS) (https://www.iata.org/services/statistics/intelligence/paxis/Pages/index.aspx).

The disease (dengue) data are available by request from the European Centre for Disease Prevention and Control (ECDC) (https://www.ecdc.europa.eu/en/publicationsdata/european-surveillance-system-tessy). All other relevant data sources are referenced in the article.

Received: 29 November 2019; Accepted: 21 May 2020;
Published online: 16 June 2020

References

1. Vitaly Belik, T. G. Dirk Brockmann. Natural human mobility patterns and spatial spread of infectious diseases. *Phys. Rev. X* **1**, https://doi.org/10.1103/PhysRevX.1.011001 (2011).
2. Tian, H. *et al.* Increasing airline travel may facilitate co-circulation of multiple dengue virus serotypes in Asia. *PLoS Negl. Trop. Dis.* **11**, e0005694, https://doi.org/10.1371/journal.pntd.0005694 (2017).
3. Tatem, A. J., Rogers, D. J. & Hay, S. I. Global transport networks and infectious disease spread. *Adv. Parasitol.* **62**, 293–343, https://doi.org/10.1016/s0065-308x(05)62009-x (2006).
4. European Centre for Disease Prevention and Control. *Dengue, in: ECDC Annual epidemiological report for 2017* https://www.ecdc.europa.eu/sites/default/files/documents/dengue-annual-epidemiological-report-2017.pdf (2019).
5. European Centre for Disease Prevention and Control. *Autochthonous transmission of dengue virus in EU/EEA, 2010-2019*, https://www.ecdc.europa.eu/en/all-topics-z/dengue/surveillance-and-disease-data/autochthonous-transmission-dengue-virus-eueea (2019).
6. European Centre for Disease Prevention and Control. *Autochthonous cases of dengue in Spain and France*, https://www.ecdc.europa.eu/sites/default/files/documents/RRA-dengue-in-Spain-France_1Oct2019.pdf (2019).
7. Brockmann, D. Global connectivity and the spread of infectious diseases. *Nova. Acta. Leopoldina.* **419**, 129–136, http://rocs.hu-berlin.de/papers/brockmann_2017b.pdf (2017).
8. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science.* **342**, 1337, https://doi.org/10.1126/science.1245200 (2013).
9. Silk, M. J. *et al.* The application of statistical network models in disease research. *Methods Ecol. Evol.* **8**, 1026–1041, https://doi.org/10.1111/2041-210X.12770 (2017).
10. Gardner, L. M., Bota, A., Gangavarapu, K., Kraemer, M. U. G. & Grubaugh, N. D. Inferring the risk factors behind the geographical spread and transmission of zika in the Americas. *PLoS Negl. Trop. Dis.* **12**, e0006194, https://doi.org/10.1371/journal.pntd.0006194 (2018).
11. Lana, R. M., Gomes, M., Lima, T. F. M., Honorio, N. A. & Codeco, C. T. The introduction of dengue follows transportation infrastructure changes in the state of Acre, Brazil: a network-based analysis. *PLoS Negl. Trop. Dis.* **11**, e0006070, https://doi.org/10.1371/journal.pntd.0006070 (2017).
12. Liebig, J., Jansen, C., Paini, D., Gardner, L. & Jurdak, R. A global model for predicting the arrival of imported dengue infections. *PLoS One.* **14**(12), e0225193, https://doi.org/10.1371/journal.pone.0225193 (2019).
13. Salami, D., Capinha, C., Martins, Md. R. O. & Sousa, C. A. Dengue importation into Europe: a network connectivity-based approach. *PLoS One.* **15**, e0230274, https://doi.org/10.1371/journal.pone.0230274 (2020).
14. Shi, Y. *et al.* Three-month real-time dengue forecast models: an early warning system for outbreak alerts and policy decision support in Singapore. *Environ. Health Perspect.* **124**, 1369–1375, https://doi.org/10.1289/ehp.1509981 (2016).
15. Chen, Y. *et al.* Neighbourhood level real-time forecasting of dengue cases in tropical urban Singapore. *BMC Med.* **16**, 129–129, https://doi.org/10.1186/s12916-018-1108-5 (2018).
16. Sammut, C. & Webb, G. I. *Encyclopedia of machine learning and data mining.* (Springer, 2017).
17. Beam, A. L. & Kohane, I. S. Big data and machine learning in health care. *JAMA.* **319**, 1317–1318, https://doi.org/10.1001/jama.2017.18391 (2018).
18. Miguel-Hurtado, O., Guest, R., Stevenage, S. V., Neil, G. J. & Black, S. Comparing machine learning classifiers and linear/logistic regression to explore the relationship between hand dimensions and demographic characteristics. *PLoS One.* **11**, e0165521, https://doi.org/10.1371/journal.pone.0165521 (2016).
19. Singal, A. G. *et al.* Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. *Am. J. Gastroenterol.* **108**, 1723–1730, https://doi.org/10.1038/ajg.2013.332 (2013).
20. Guo, P. *et al.* Developing a dengue forecast model using machine learning: a case study in China. *PLoS Negl. Trop. Dis.* **11**, e0005973, https://doi.org/10.1371/journal.pntd.0005973 (2017).
21. Siriyasatien, P., Chadsuthi, S., Jampachaisri, K. & Kesorn, K. Dengue epidemics prediction: a survey of the state-of-the-art based on data science processes. *IEEE Access.* **6**, 53757–53795, https://doi.org/10.1109/ACCESS.2018.2871241 (2018).
22. Mustaffa, Z., Sulaiman, M. H., Emawan, F., Yusof, Y. & Mohsin, M. F. M. Dengue outbreak prediction: hybrid meta-heuristic model in *19th IEEE/ACIS International conference on software Engineering, artificial intelligence, networking and parallel/distributed computing (SNPD)*, 271-274, https://doi.org/10.1109/SNPD.2018.8441095 (2018)

23. Molnar, C. *Interpretable machine learning: a guide for making black box models explainable*, https://christophm.github.io/interpretable-ml-book/index.html (2019).
24. European Centre for Disease Prevention and Control. *The European surveillance system (TESSy)*, https://ecdc.europa.eu/en/publications-data/european-surveillance-system-tessy (2019).
25. European Union. *Commission Implementing Decision of 8 August 2012 amending Decision 2002/253/EC laying down case definitions for reporting communicable diseases to the Community network under Decision No 2119/98/EC of the European Parliament and of the Council (notified under document C(2012) 5538) Text with EEA relevance*, http://data.europa.eu/eli/dec_impl/2012/506/oj (2012).
26. International Air Transport Association. *Passenger intelligence services (PaxIS)*, https://www.iata.org/services/statistics/intelligence/paxis/Pages/index.aspx (2019).
27. Rodrigue, J.-P. In *The geography of transport systems* Ch. Chapter 10, 440 (Routledge, 2017).
28. Domingos, P. A few useful things to know about machine learning. *Commun. ACM.* **55**, 78–87, https://doi.org/10.1145/2347736.2347755 (2012).
29. Max, K. & Kjell, J. *Applied predictive modeling.* (Springer-Verlag, New York, 2013).
30. HealthMap. *DengueMap*, https://www.healthmap.org/dengue/en/index.php (2019).
31. Institute for Health Metrics and Evaluation. *Global Burden of Disease Study Results*, http://ghdx.healthdata.org/gbd-results-tool (2019).
32. Centre d'Etudes Prospectives et d'Informations. *CEPII Database*, http://www.cepii.fr/cepii/en/bdd_modele/bdd.asp (2019).
33. Moore, M., Gelfeld, B., Okunogbe, A. & Paul, C. Identifying future disease hot spots: infectious disease vulnerability index. *Rand health quarterly.* **6**, 5 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5568150/ (2017).
34. The World Bank. *World bank open data*, https://data.worldbank.org/ (2019).
35. Oldham, S. *et al.* Consistency and differences between centrality measures across distinct classes of networks. *PLoS One.* **14**, e0220061, https://doi.org/10.1371/journal.pone.0220061 (2019).
36. Ling, C. X. & Li, C. Data mining for direct marketing: problems and solutions, *in proceedings of the fourth international conference on knowledge discovery and data mining*, 73–79, https://www.aaai.org/Papers/KDD/1998/KDD98-011.pdf (1998).
37. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res* **16**, 321–357, https://doi.org/10.1613/jair.953 (2002).
38. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, https://doi.org/10.18637/jss.v028.i05 (2008).
39. R-Core-Team. *The R Project for statistical computing*, https://www.r-project.org/ (2019).
40. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **58**, 109–130, https://doi.org/10.1016/S0169-7439(01)00155-1 (2001).
41. Mevik, B.-H. & Wehrens, R. The pls Package: Principal component and partial least squares regression in R. *J. Stat. Softw* **18**, https://doi.org/10.18637/jss.v018.i02 (2007).
42. Friedman, J. H., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, https://doi.org/10.18637/jss.v033.i01 (2010).
43. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32, https://doi.org/10.1023/A:1010933404324 (2001).
44. Kearns, M. & Valiant, L. Cryptographic limitations on learning Boolean formulae and finite automata. *J. ACM.* **41**, 67–95, https://doi.org/10.1145/174644.174647 (1994).
45. Valiant, L. G. A theory of the learnable. *Commun. ACM.* **27**, 1134–1142, https://doi.org/10.1145/1968.1972 (1984).
46. Goldstein, A., Kapelner, A., Bleich, J. & Pitkin, E. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. *J. Comput. Graph. Stat,* **24**(1), 44–65, https://doi.org/10.1080/10618600.2014.907095 (2015).
47. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. in *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794, https://doi.org/10.1145/2939672.2939785 (2016).
48. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874, https://doi.org/10.1016/j.patrec.2005.10.010 (2006).
49. Baehrens, D. *et al.* How to explain individual classification decisions. *J. Mach. Learn. Res.* **11**, 1803–1831, http://www.jmlr.org/papers/volume11/baehrens10a/baehrens10a.pdf (2010).
50. Ribeiro, M. T., Singh, S. & Guestrin, C. Model-agnostic interpretability of machine learning. Preprint at https://arxiv.org/abs/1606.05386 (2016).
51. Molnar, C., Bischl, B. & Casalicchio, G. iml: An R package for interpretable machine learning. *J. Open Source Softw.* **3**, 786, https://doi.org/10.21105/joss.00786 (2018).
52. Pedersen, T. L. & Benesty, M. *lime: Local interpretable model-agnostic explanations*, https://CRAN.R-project.org/package=lime (2019).
53. Sanchez, I., Rocktaschel, T., Riedel, S. & Singh, S. Towards extracting faithful and descriptive representations of latent variable models in *AAAI Spring syposium on knowledge representation and reasoning (KRR): integrating symbolic and neural approaches*, http://terraswarm.org/pubs/482.html (2015).
54. Fisher, A., Rudin, C. & Dominici, F. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res*, **20**(177), 1–81. http://jmlr.org/papers/v20/18-760.html (2019).
55. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **29**, 1189–1232, http://www.jstor.org/stable/2699986 (2001).
56. Greenwell, B. M. pdp: An R package for constructing partial dependence plots. *The R Journal.* **9**, 421–436, https://journal.r-project.org/archive/2017/RJ-2017-016/index.html (2017).
57. Pedersen, T. L. & Benesty, M. *Understanding lime*, https://cran.r-project.org/web/packages/lime/vignettes/Understanding_lime.html (2019).
58. UC Business Analytics R Programming Guide. *Visualizing ML models with LIME*, http://uc-r.github.io/lime (2019).
59. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R News.* **2**, 18–22, https://www.r-project.org/doc/Rnews/Rnews_2002-3.pdf (2002).
60. Chen, T. *et al. xgboost: Extreme gradient boosting*, https://CRAN.R-project.org/package=xgboost (2019).
61. Wickham, H. The split-apply-combine strategy for data analysis. *J. Stat. Softw.* **40**, https://doi.org/10.18637/jss.v040.i01 (2011).
62. Microsoft Corporation & Weston, S. *doSNOW: Foreach parallel adaptor for the 'snow' package*, https://CRAN.R-project.org/package=doSNOW (2019).
63. Torgo, L. *Data Mining with R, learning with case studies*. (Chapman and Hall/CRC, 2010).
64. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* **12**, 77, https://doi.org/10.1186/1471-2105-12-77 (2011).
65. Seltenrich, N. Singapore success: new model helps forecast dengue outbreaks. *Environ. Health Perspect.* **124**, A167–A167, https://doi.org/10.1289/ehp.124-A167 (2016).
66. Semenza, J. C. *et al.* International dispersal of dengue through air travel: importation risk for Europe. *PLoS Negl. Trop. Dis.* **8**, e3278, https://doi.org/10.1371/journal.pntd.0003278 (2014).

Acknowledgements
This work was partially funded by Fundação para a Ciência e a Tecnologia (FCT), Portugal (GHTM – UID/Multi/04413/2013 and Project Warden - PTDC/SAU-PUB/30089/2017). DS has a PhD grant from FCT, Portugal (PD/BD/128084/2016). CC acknowledges support from FCT (CEECIND/02037/2017, UIDB/00295/2020 and UIDP/00295/2020). We greatly appreciate Dominic Freienstein for his assistant in accessing the international air travel association, passenger intelligence services (IATA-PaxIS) data.

Author contributions
Conceptualization, D.S., C.S. and C.C.; Data Curation, D.S.; Formal Analysis, D.S.; Methodology, D.S. andC.C.; Supervision, C.S., M.M. and C.C.; Visualization, D.S.; Writing—Original Draft Preparation, D.S.;Writing—Review & Editing, D.S., C.S., M.M. and C.C. All authors read and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to D.S., C.A.S. or C.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2.3 Simulation models of dengue transmission in Funchal, Madeira Island: Influence of seasonality

---

Salami D, Capinha C, Sousa CA, Martins MdRO, Lord C (2020) Simulation models of dengue transmission in Funchal, Madeira Island: Influence of seasonality. PLoS Neglected Tropical Diseases 14(10): e0008679. https://doi.org/10.1371/journal.pntd.0008679

PLOS NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

Simulation models of dengue transmission in Funchal, Madeira Island: Influence of seasonality

Donald Salami[1]*, César Capinha[2], Carla Alexandra Sousa[1]*, Maria do Rosário Oliveira Martins[1], Cynthia Lord[3]*

1 Global Health and Tropical Medicine, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisbon, Portugal, 2 Centro de Estudos Geográficos, Instituto de Geografia e Ordenamento do Território, Universidade de Lisboa, Lisbon, Portugal, 3 Florida Medical Entomology Laboratory, Institute of Food and Agricultural Sciences, University of Florida, Vero Beach, Florida, United States of America

* donald.salami@gmail.com (DS); casousa@ihmt.unl.pt (CS); clord@ufl.edu (CL)

Citation: Salami D, Capinha C, Sousa CA, Martins MdRO, Lord C (2020) Simulation models of dengue transmission in Funchal, Madeira Island: Influence of seasonality. PLoS Negl Trop Dis 14(10): e0008679. https://doi.org/10.1371/journal.pntd.0008679

Editor: Christopher M. Barker, University of California, Davis, UNITED STATES

Received: October 28, 2019

Accepted: August 4, 2020

Published: October 5, 2020

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: DS has a PhD grant from the Fundação para a Ciência e a Tecnologia (FCT), Portugal (https://www.fct.pt/), grant number: PD/BD/128084/2016 This work was also partially funded by FCT grants awarded to CS, grants number: Project Warden - PTDC/SAU-PUB/30089/2017 and GHTM–UID/Multi/04413/2020. CC acknowledges support from FCT, grants number: CEECIND/

Abstract

The recent emergence and established presence of *Aedes aegypti* in the Autonomous Region of Madeira, Portugal, was responsible for the first autochthonous outbreak of dengue in Europe. The island has not reported any dengue cases since the outbreak in 2012. However, there is a high risk that an introduction of the virus would result in another autochthonous outbreak given the presence of the vector and permissive environmental conditions. Understanding the dynamics of a potential epidemic is critical for targeted local control strategies.

Here, we adopt a deterministic model for the transmission of dengue in *Aedes aegypti* mosquitoes. The model integrates empirical and mechanistic parameters for virus transmission, under seasonally varying temperatures for Funchal, Madeira Island. We examine the epidemic dynamics as triggered by the arrival date of an infectious individual; the influence of seasonal temperature mean and variation on the epidemic dynamics; and performed a sensitivity analysis on the following quantities of interest: the epidemic peak size, time to peak, and the final epidemic size.

Our results demonstrate the potential for summer and autumn season transmission of dengue, with the arrival date significantly affecting the distribution of the timing and peak size of the epidemic. Late-summer arrivals were more likely to produce large epidemics within a short peak time. Epidemics within this favorable period had an average of 11% of the susceptible population infected at the peak, at an average peak time of 95 days. We also demonstrated that seasonal temperature variation dramatically affects the epidemic dynamics, with warmer starting temperatures producing large epidemics with a short peak time and vice versa. Overall, our quantities of interest were most sensitive to variance in the date of arrival, seasonal temperature, transmission rates, mortality rate, and the mosquito population; the magnitude of sensitivity differs across quantities.

Our model could serve as a useful guide in the development of effective local control and mitigation strategies for dengue fever in Madeira Island.

02037/2017, UIDB/00295/2020 and UIDP/00295/2020. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

The presence of *Aedes aegypti* mosquitoes in Madeira Island had recently caused the first local outbreak of dengue in Europe. The island is at risk of another local transmission if triggered by the introduction of the dengue virus by an infected person. Using a mathematical model for the transmission of dengue, we examine the dynamics of a potential epidemic triggered by the arrival of an infected person on the island. We also examine the impact of seasonal temperature variation on the epidemic dynamics. Our results show the potential for summer and autumn season transmission of dengue on the island, and that the arrival date of an infectious person affects the distribution of the timing and peak size of the epidemic. Late summer arrivals were more likely to produce a rapid epidemic. We also show that seasonal temperature variation dramatically affects the epidemic dynamics. With warmer starting temperatures, epidemics peak faster, producing a large final epidemic size. Our model could be useful to estimate the risk of an epidemic outbreak and as a guide for local control and mitigation strategies for dengue on the island.

Introduction

Dengue is notably the most important mosquito-borne viral disease, with approximately half the world's population at risk of infection [1]. This arboviral disease, caused by a virus of the Flaviviridae family, has gained renewed global attention due to its wide geographical spread and increased burden in recent years. The spread of the disease is in concordance with the geographical expansion of its primary vector (i.e. *Aedes aegypti*), characterized by the presence of a suitable climate and increase in global trade and travel [2, 3].

A paradigmatic example of the recent spread of dengue and its epidemic potential was demonstrated in Madeira Island, an autonomous region of Portugal (S1 Fig), with an estimated population size of 270,000. *Aedes aegypti* was first detected in Funchal, the capital city of Madeira Island, in 2005 and by October 2012 the island reported its first autochthonous case of Dengue serotype 1 (DENV 1) [4]. The importance of this epidemic is demonstrated by three main reasons: (1) It was the first sustained autochthonous transmission of dengue in the European Union since the 1920s [5]; (2) its size, with 1080 confirmed cases (of the 2168 probable cases reported) and 78 cases reported in 13 other European countries in travelers returning from Madeira [6]; and (3) the rapid time course of the epidemic, the epidemic peaked within a month after the official report of the first case in October [7] with a rapid decline thereafter with no cases recorded after 3 March 2013.

To understand the complexities of this outbreak, Lourenco and Recker [8] developed an ento-epidemiological mathematical model to explore the ecological conditions and transmission dynamics. Their findings suggest the virus was introduced over a month before cases were reported (in August). In their model, asymptomatic circulation occurred before the two initial autochthonous cases were reported in October, maintaining the virus in the population. Furthermore, their findings indicate that the transmission dynamics and eventual epidemic burnout was driven predominantly by the influence of temperature on the life-history traits of the mosquitoes (incubation period, mosquito mortality and aquatic developmental rates). The seasonal drop in autumn temperatures led to a reduction of the vectorial capacity and effectively stopped the virus propagation.

Their findings are consistent with previous experimental work addressing the strong influence of temperature on the life-history traits of the mosquitoes and arbovirus transmissions [9–12]. Likewise, many existing mechanistic transmission models have integrated the effects of

temperature on mosquito traits to understand how it influences the probability and magnitude of dengue transmissions [13–17]. Other recent work examined the influence of seasonal variation in temperature on the epidemic magnitude and duration [18, 19]. These models emphasize the strong, nonlinear (often unimodal) influence of temperature and seasonality on dengue transmission and epidemic dynamics.

Expanding on previous work, we here explore the impact of seasonal temperature variations on the potential epidemic dynamics on Madeira Island. It is imperative to explore the influence of seasonality, as Madeira Island presents with a range of contrasting bioclimates because of its heterogeneous landscape and strong influence from the Gulf Stream and Canary current [20]. The southern coastal regions of the island (including Funchal), at low altitudes, have higher annual temperatures in comparison to the northern coastal regions or inland regions with higher altitudes and lower annual temperatures [20, 21]. Although the island has not reported any dengue cases since the outbreak in 2012, new introductions would likely result in local transmission given the presence of the vector and permissive environmental conditions. A recent vector competence study with *Aedes aegypti* from Madeira reported virus transmission potential (virus in saliva) from 2 Madeira populations, although the proportions transmitting was low (transmission rate of 18%, 14 days post-infection) [22]. This demonstrates the potential risk of local transmission of a different serotype of dengue, if introduced on the island. With an increase in co-circulation of all dengue serotypes (DENV 1–4) worldwide [23, 24], and Madeira's increasing lure as a popular year-round travel destination, a potential introduction is likely [25].

We incorporate a standard deterministic SEI-SEIR transmission model parameterized from existing literature and available field data. The main goals of this model are: (1) to examine the epidemic dynamics in Funchal, Madeira as triggered by the arrival of an infectious individual at different time points during the year; and (2) to examine the influence of seasonal temperature mean and variation on epidemic dynamics. To do this, we employed a different modeling framework from that of [8], in that our model explicitly accounts for seasonality and temperature dependence in the transmission dynamics. Likewise, we explore epidemiologically relevant outcomes of epidemic size, peak incidence, and time to peak, rather than basic reproduction number or vectorial capacity, which are more complex measures of epidemic dynamics.

Methods

Model framework

We adopt a deterministic compartmental vector-host transmission model exploring chikungunya virus invasion in Florida, United States with *Aedes aegypti* and *Aedes albopictus* mosquitoes, from [26]. This was adapted into a standard SEI-SEIR transmission model with one vector, similar to others used in modeling dengue transmission (e.g. in [27–30]).

The SEI component of our model describes the vector population, represented as susceptible (S_v), exposed (E_v), and infectious (I_v). Our models explicitly consider a single vector and a single life stage, i.e. adult females of *Aedes aegypti*. Mosquitoes enter the susceptible class through a recruitment term, based on observed seasonality patterns from field data in Funchal [31]. The recruitment term does not explicitly model the aquatic (eggs, larvae, and pupae) stage of mosquitoes and is not linked to the current population size, but explicitly includes seasonality in recruitment to the adult female population. A susceptible mosquito moves into the exposed class (E_v), after biting an infectious human and becoming infected with the dengue virus. After a temperature-dependent extrinsic incubation period, surviving mosquitoes get transferred to the infectious class (I_v). They remain in the infectious class until death, due to

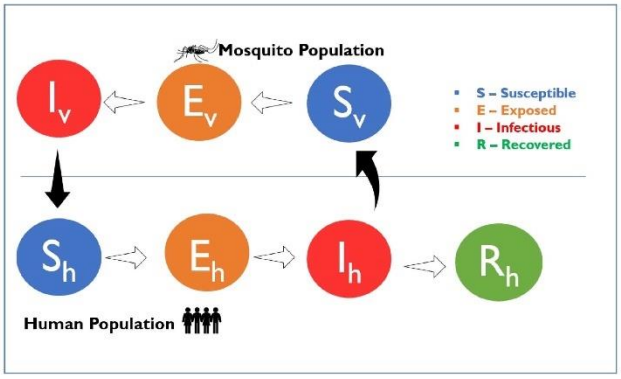

Fig 1. Schematic representation of the model. S_v, E_v, and I_v represent the susceptible, exposed, and infectious compartments of the mosquito population. S_h, E_h, I_h, and R_h represent the susceptible, exposed, infectious, and recovered compartments of the human population, respectively. The outline arrows are the transition from one compartment to the next, and the black filled arrows are the direction of transmission.

https://doi.org/10.1371/journal.pntd.0008679.g001

the assumption of the absence of immune response. Mosquitoes leave the system through a temperature-dependent mortality function. We assume the virus infection does not affect the lifespan of the mosquitoes and that there is no significant vertical transmission.

The SEIR component of our model describes the human population represented as susceptible (S_h), exposed (infected but not infectious) (E_h), infectious (I_h), and recovered (immune) (R_h). Our model assumes the human population (N_h) to be constant, not subject to demography as we considered a single outbreak with a duration in the order of a year. A susceptible individual enters the exposed class (E_h) after being successfully infected, by an infectious mosquito bite. We do not explicitly account for repeated biting, interrupted feeds, or alternative host preferences. We also assume that not every infectious bite leads to successful human infection. Once a human individual is exposed, they enter an intrinsic incubation period, until they become infectious. They then move to the infectious class (I_h) and can transmit the virus back to a susceptible mosquito. We assumed a single infectious class and did not further discriminate between the symptomatic or asymptomatic individuals as previous evidence demonstrates that asymptomatic individuals can transmit the virus [32]. We assume that humans stay infectious for a period after which they recover. Once a human individual enters the recovered/immune class (R_h), we assume a lifelong immunity, as multiple co-circulating serotypes of dengue virus were not considered. A resulting schematic representation of the model is shown in Fig 1.

Model equations

Our model is defined by the following ordinary differential equations:

$$\frac{dS_v}{dt} = \rho - \frac{a\,\beta_{h\to v}}{N_h} I_h S_v - \mu_v S_v \tag{1}$$

$$\frac{dE_v}{dt} = \frac{a\,\beta_{h\to v}}{N_h} I_h S_v - \gamma_v E_v - \mu_v E_v \tag{2}$$

Figure 2-14 Schematic representation of the model.

Table 2-11 State variables for the model.

$$\frac{dI_v}{dt} = \gamma_v E_v - \mu_v I_v \tag{3}$$

$$N_v = S_v + E_v + I_v \tag{4}$$

$$\frac{dS_h}{dt} = -\frac{a\,\beta_{v\to h}}{N_h} S_h I_v \tag{5}$$

$$\frac{dE_h}{dt} = \frac{a\,\beta_{v\to h}}{N_h} S_h I_v - \gamma_h E_h \tag{6}$$

$$\frac{dI_h}{dt} = \gamma_h E_h - \eta_h I_h \tag{7}$$

$$\frac{dR_h}{dt} = \eta_h I_h \tag{8}$$

$$N_h = S_h + E_h + I_h + R_h \tag{9}$$

Here, the coefficient ρ is the mosquito recruitment term (expanded below); a is the biting rate; $\beta_{h\to v}$ and $\beta_{v\to h}$ are the human-to-mosquito and mosquito-to-human transmission rates, respectively; $1/\gamma_h$ and $1/\eta_h$ represent the intrinsic incubation and human infectivity periods; $1/\gamma_v$ and μ_v represent the extrinsic incubation period and mortality rate for mosquitoes. The biting rate, EIP and mortality rates are temperature-dependent, detailed below. The state variables and parameters used for our model are displayed in Tables 1 and 2, respectively. Preliminary exploration of parameter values was performed to determine their permissibility for transmission and to define model starting conditions (see S1 File).

Mosquito recruitment. The coefficient (ρ) in Eq (1) above is the daily recruitment term of susceptible mosquitoes. The total number of female mosquitoes recruited into the population over the year is divided into two main phases: baseline, year-round recruitment, and a unimodal peak season recruitment. This is based on the seasonal pattern of mosquitoes on the island observed from the *Aedes aegypti* mosquito entomological surveillance by the Institute of health administration, IP-RAM, and the Funchal natural history museum [31, 47]. Weekly mosquito trap data is geo-processed and spatial analyzed to identify areas with mosquito

Table 1. State variables for the model.

Variables	Description
S_v	Number of susceptible mosquitoes
E_v	Number of exposed mosquitoes
I_v	Number of infectious mosquitoes
N_v	Total mosquito population size
S_h	Number of susceptible humans
E_h	Number of exposed humans
I_h	Number of infectious humans
R_h	Number of recovered humans
N_h	Total human population size

https://doi.org/10.1371/journal.pntd.0008679.t001

Table 2-12 Definitions and ranges of the model's parameters.

Table 2. Definitions and ranges of the model's parameters.

Parameter	Description	Default Value	Range	Reference
r_{δ}	Number of mosquitoes per host in the unimodal peak phase	2	2–6	[8, 33, 34]
p_{base}	Proportion of mosquitoes in baseline recruitment	0.05		[8, 35]
iv	The interval between baseline pulses (days)[1]	5	2–7	[35]
q	Timing of unimodal peak (days)[1]	220	209–230	[35]
σ	The variance of the unimodal peak (days)[1]	29		[35]
R_{tot}	Total mosquito recruitment (population size)[2]			Calculated based on default value and range for r_{δ} above. Used in the sensitivity analysis.
a	Mosquito biting rate[3]	$0.0043T + 0.0943$		[15, 36]
$\beta_{h\rightarrow v}$	Rate of transmission from human to mosquito	0.33	0.1–0.75	[28, 33, 37]
$\beta_{v\rightarrow h}$	Rate of transmission from mosquito to human	0.33	0.1–0.75	[28, 33, 37]
μ_v	Mortality rate of mosquitoes under optimal temperatures[4]	0.04	0.02–0.06	[13, 17]
μ_{vs}	The slope of mortality function	0.05		[13, 26]
γ_v	Virus extrinsic incubation rate (at 22.5˚C)	0.04	0.04–0.07	[17, 38]
γ_{vs}	Slope of extrinsic incubation function	0.008		[38]
$1/\gamma_h$	Intrinsic incubation period (days)	6		[39, 40]
$1/\eta_h$	Human infectious period (days)	4		[8, 41–43]
N_h	The population size of humans (constant)	30000		[44]
t_{crit}	Date of arrival of infectious human (days)[5]	181	1–365	
T_{mean}	Mean annual temperature (˚C)	20	15–30	[45, 46]
T_{range}	Temperature range (˚C)[6]	5	0–15	[45, 46]

[1] Default value and range are estimates to reflect the observed seasonal activity, not an explicit fit to the island's entomological data;

[2] See Eq 15, also note this parameter was used in the sensitivity analysis to allow estimation of the relative importance of the total recruitment;

[3] Optimal temperature for biting rate: $(21^{\circ}C > T < 32^{\circ}C)$;

[4] Mosquitoes optimal survival range: $(15^{\circ}C > T < 30^{\circ}C)$;

[5] Where $t_{crit} = 1$, calendar date is February 15; hence t_{crit} of 181 is August 15; while 365 is February 14 (i.e. a complete annual period).

[6] Temperature range is given by $(T_{max} - T_{min})$.

https://doi.org/10.1371/journal.pntd.0008679.t002

activity based on the presence/absence of eggs, the number of eggs, and adult mosquitoes captured [31].

Visual inspection of the entomological data for the years 2012–2019 (weekly time-series graphs of the cumulative number of eggs and adult mosquitoes, see S2–S5 Figs) [35] demonstrate a unimodal seasonal pattern for mosquito activity during the entomological season. The unimodal peak starts around June through to early December, with other months of the year (i.e. January to May) having little to no activity. Using a Gaussian curve, we estimated the timeline for our unimodal peak recruitment to reflect this seasonal pattern as best possible.

Actual quantifications of mosquito population density remain a grey area. However, the majority of dengue models [8, 28, 33, 34, 48] have adopted the use of a mosquito-to-human ratio of 2:1. We adopted this framework for setting the estimated number of mosquitoes recruited during the unimodal peak season (r_{δ}). From the derived number of mosquitoes recruited during the peak season, we assumed an additional 5% to be added in the baseline, year-round recruitment (R_{base}). The recruitment of susceptible mosquitoes is specified by the

following equations:

$$\rho(t) = \rho_b(t) + R_{peak}\delta(t) \tag{10}$$

Where,

$$\delta(t) = \frac{1}{2\pi\sigma_k} exp\left(\frac{1(q_k - t)^2}{2\sigma_k^2}\right) \tag{11}$$

$$\rho_b(t) = \frac{R_{base}}{iv} \tag{12}$$

$$R_{peak} = r_\delta N_h \tag{13}$$

$$R_{base} = R_{peak} p_{base} \tag{14}$$

$$R_{tot} = R_{peak} + R_{base} \tag{15}$$

Here, $\rho(t)$ represents the number of adult female mosquitoes recruited per time step; $\delta(t)$ is a Gaussian distribution for the number of mosquitoes added to the population daily, during the unimodal peak season; $\rho_b(t)$ is the number of mosquitoes added to the population at intervals (iv) during the all-year-round recruitment. R_{peak} is the total number of mosquitoes recruited during the unimodal peak season; R_{base} is the total number of mosquitoes recruited during the baseline, year-round recruitment; R_{tot} is the total number of mosquitoes recruited into the population over the year.

Seasonal forcing. To introduce seasonality in our model, we allowed the temperature to vary over time by sinusoidally forcing [49]. The daily mean temperature was modeled as a cosine curve with a period of 365 days as specified below:

$$T(t) = \frac{T_{max} - T_{min}}{2} * \left(-\cos\left(\frac{2\pi(t-\omega)}{365}\right)\right) + T_{mean} \tag{16}$$

Here, T is temperature in Celsius; T_{max}, T_{mean} and T_{max} are the average daily maximum, mean, and minimum temperatures over the year, (t) is time measured in days, and (ω) is the phase shift to align the cosine function with the seasonal factors in Funchal. For our temperature model calibration, we extracted a 10 year (2008–2018) historical daily from Weather Underground [45]. We then calculated the 10-year historical average of the mean, minimum, and maximum for each day of the year (excluding February 29), before estimating temperature seasonality. This provided a range of the average daily temperature over a calendar year for the period. Utilizing the mean, minimum, and maximum annual average temperature, we set estimates for our seasonality parameters. To fit the sinusoidal model, we considered the occurrence of the coldest and hottest day across the years, to determine a good starting point. The occurrence of the coldest day had fewer fluctuations mostly occurring on February 15; hence we set our start point and phase shift to this day, reflecting the long-term average conditions in Funchal, Madeira (T_{max} = 22˚ C, T_{mean} = 20˚C, T_{min} = 17˚C). S6 Fig shows a fit of the sinusoidal curve to the historical temperature seasonality.

Temperature-dependent parameters. The biting rate (a) of the adult female mosquitoes is modeled to be temperature dependent as previously estimated by [15] and [36]. The biting rate increases linearly with temperature as specified by a linear equation for an optimal temperature range of $21^\circ C < T > 32^\circ C$ and a defined temperature threshold (i.e. if $T \leq 12.4$ *or*

$T \geq 37^{\circ}C$, then $a = 0$). The extrinsic incubation period (γ_v) is modeled to be temperature-dependent, modified from [50]. Using a linear temperature function, specified by a slope (γ_{vs}), the rate near the midpoint of the plausible temperatures range for vectorial capacity (set at 22.5˚C), and a defined lower temperature threshold (i.e. if $T \leq 10^{\circ}C$ then $\gamma_v = 0$). Mortality rate (μ_v) for mosquitoes were modeled as a function of temperature using a mechanistic thermal response curve as described in [13]. They fitted their data as a complex polynomial resulting in a basin-shaped curve, with optimal temperature for mosquito's survival set at a range of $15^{\circ}C > T < 30^{\circ}C$. Based on preliminary exploration of changes to the mortality function, we modified the fitted polynomial to a piecewise linear curve with fixed, minimal mortality in the same temperature range. Mortality rate then increases quickly and linearly at temperatures outside this lower and upper bound, as specified by the function slope (μ_{vs}).

Quantities of interest (QOI)

We simulated the model under default starting conditions (see S1 File) and given that a simulation evolves into an epidemic (defined as ($I_h > 2$), after the virus introduction), we analyze for the following quantities of interest (QOI): the epidemic peak size (maximum human infected ($maxI_h$) at any given point during the simulation); time to peak infection in humans (time from introduction to $maxI_h$, as $tmaxI_h$); the final epidemic size, which represents a measure of epidemic suitability (cumulative proportion of humans infected, $cumI_h/N_h$, at the final time step).

Initial introduction and epidemic dynamics

Firstly, we examine the variability in epidemic dynamics, as a function of arrival dates of an infectious human in a susceptible population. To do this, we ran simulations for introductions on all 365 calendar days of the year, with all other parameters fixed at their default values (Table 2) and default starting conditions (S1 File). Each simulation was started on February 15 (the phase shift in the annual cycle) and ran for 2 years.

Seasonal variance and epidemic dynamics

Next, we examine the epidemic dynamics as a function of seasonal temperature variation. Utilizing the same compartmental framework with default starting conditions and parameter values, we ran sets of simulations under two different temperature regimes [18]. First, we simulated a set of temperature regimes as observed in the historical decadal data for Funchal, Madeira. The mean temperature varied from 19.0 ˚C to 21.0 ˚C in increments of 0.2 ˚C, while the temperature range (i.e. $T_{max} - T_{min}$) varied from 4.0 ˚C to 6.0 ˚C, in the same increments of 0.2 ˚C, resulting in 121 simulation runs. The variability in epidemic dynamics is then examined as a function of starting temperature. Starting temperature is defined as the temperature on the day of introduction of the virus (t_{crit}), as derived from the temperature curves.

Thereafter, we simulate a wider set of temperature regimes to examine plausible future forcing scenarios based on near-term projections of seasonal temperature changes in the region of Madeira Island. Near term projection refers to projected annual mean temperature changes for the period 2016–2035 relative to a reference period 1986–2005 [46, 51]. Mean temperature was varied between 15.0 ˚C to 30.0 ˚C in increments of 0.2 ˚C, while range varied from 0.0 ˚C to 15.0 ˚C in increments of 0.2 ˚C (i.e. a total of 5776 simulation runs). It is worth noting that most of the temperatures in this regime are outside projected changes for the island's region, and very unlikely to occur in Funchal. However, by simulating a wider set of temperature regimes, we can characterize uncertainties and probable outcomes of a local epidemic across other regions on the island. This also allows us to simulate similar extreme temperature

conditions already recorded on the Funchal in recent times (for example, high temperatures of 37.8˚C on August 5, 2016, and 26.5 ˚C on December 5, 2018 [52, 53]).

Model sensitivity analysis

To characterize the model parameters exerting the most influence on our quantities of interest, we performed a variance-based global sensitivity analysis, using a combination of Latin hypercube sampling (LHS) and a multi-model inference on regression-based models (See S2 File for details). The sensitivity analysis considered the main effects and pairwise (first-order) interactions of input parameters on our quantities of interest. LHS sampling was programmed in MATLAB [54], while the multi-model inference analysis was done using the *glmulti* R package [55, 56].

Results

Initial introduction and epidemic dynamics

We examined the timing and size of the epidemic peak as a function of different arrival dates of an infectious human into the susceptible population. An epidemic outbreak occurred only for simulated arrival dates in July to November; with the first epidemic occurring on July 14. Most simulations responded unimodally with peaks occurring few weeks after the arrival of an infectious human, this was typical for summer and early autumn arrivals (i.e. July to September) (Fig 2A and 2B). However, some simulations responded bimodally, i.e. an initial small outbreak, then a prolonged low-level transmission until another outbreak occurs (Fig 2C and 2D). This prolonged transmission was typical for arrival dates in mid to late autumn season (i.e. arrival dates in October and November).

We examine this behavior further by comparing the disease progression in the human population with that of the mosquito population (Fig 2D). Simulations with an arrival date of the infectious human in October and November had a lower chance to evolve into a unimodal outbreak, because of the depletion of the mosquito population. Following the initial outbreak, a low transmission is maintained until the next seasonal peak recruitment of the mosquitoes before another outbreak. This suggests that the virus can remain viable within the population at low rates, until the next favorable season for transmission. This also means the dynamics of the epidemic is also modulated by the temporal dynamics of the mosquito population, however, this is not the focus of our analysis.

Fig 3 shows our QOI, for all simulation dates resulting in an epidemic. Epidemic peak size was bimodally distributed as a function of date of arrival, with a dip in early October. In contrast epidemic peak time monotonically decreases as a function of the date of arrival, until mid-September, and reverses to an increase with a steep spike in early October. In part, these discontinuities are reflections of scoring each run for a single peak and corresponds to the model behavior shifts from the classical rapid unimodal epidemic, to the bimodal and prolonged epidemic. Arrival dates in the summer season produced epidemics with a much faster peak time, while arrivals in autumn had a longer transmission period.

Overall, the shifts in epidemic dynamics are driven by the seasonal change (summer and autumn seasons) and its effect on transmission dynamics. The shortest epidemic peak time simulated was 93 days, with 11.4% of the susceptible population infected, with an arrival date of August 29 and a starting temperature of 22.4˚C (Fig 2A). In contrast, the longest epidemic peak time simulated was 411 days, with 3.4% of the susceptible population infected, with an arrival date of October 7 and a starting temperature of 21˚C (Fig 2C). The final epidemic sizes were 99.99% and 99.80% of the human population infected, respectively for both simulations.

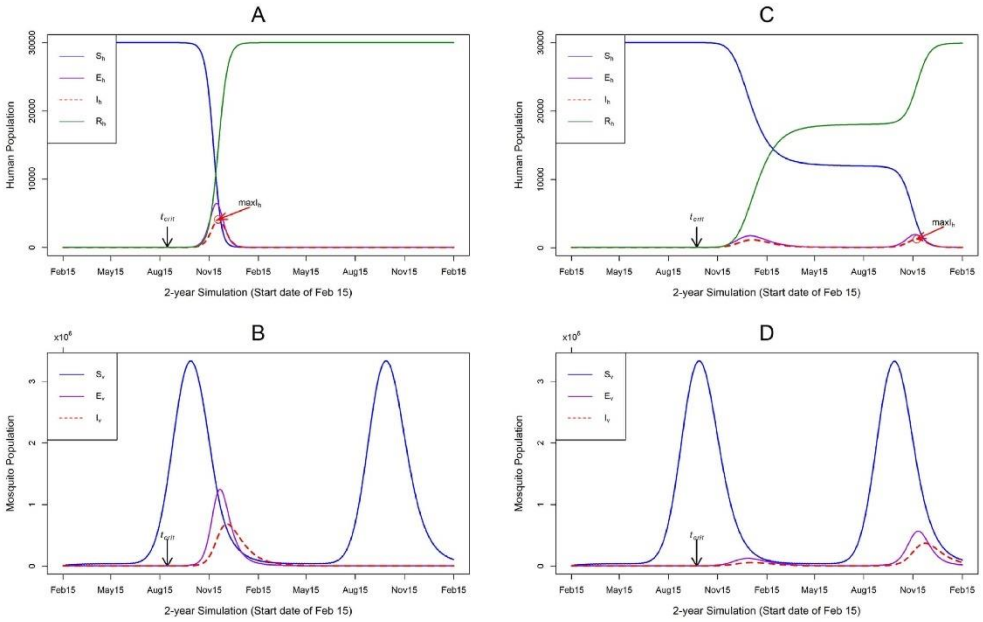

Fig 2. The epidemic progression in the human and mosquito populations. The y-axis is the number of humans or mosquitoes in the simulation and the x-axis is the 2-year simulation dates with a start date of February 15. t_{crit} = arrival date of infectious human; $maxI_h$ = the epidemic peak size, the maximum human infected at any given point during the simulation. (A) indicates the disease progression in the human population, for an arrival date of an infectious human on August 29. (C) indicates disease progression for an arrival date of an infectious human on October 7. (B) and (D) show the disease progression in the mosquito population for the respective dates of arrival. (A) shows a classical rapid epidemic, with a unimodal response, with the peak occurring few weeks after virus introduction; while (C) shows a prolonged period of lower-level transmission, resulting in a bimodal response. Default parameters from Table 2 were used for these simulations, except for dates of arrival of infectious human.

https://doi.org/10.1371/journal.pntd.0008679.g002

These results show that under the conditions used here, an epidemic potential exists within the summer and autumn seasons, with the most favorable season for a rapid outbreak being the end of the summer season. Epidemics occurring within this favorable period had an average epidemic peak size of 11% of the susceptible population infected, with a time to peak of 95 days. Arrival dates of an infectious human, in the autumn season, can dramatically affect the length of the epidemic, as these outbreaks are also modulated by the mosquito seasonal dynamics.

Seasonal variance and epidemic dynamics

To examine the epidemic dynamics as a function of the seasonal temperature variance, we simulated two different sets of temperature regimes, with a fixed arrival date in mid-summer (August 15). In the first set of temperature regimes (historical; T_{mean} varied from 19.0˚C to 22.0˚C and T_{range} varied from 4.0˚C to 6.0˚C, both in increments of 0.2˚C), all simulations resulted in an epidemic, with a classical unimodal peak. The timing and magnitude of the epidemic peak both responded differently as a function of starting temperature (calculated for August 15 from T_{mean} and T_{range} using Eq 16) with time to peak being less sensitive. Epidemic peak size increases monotonically with an increase in starting temperature, i.e. warmer

Figure 2-15 The epidemic progression in the human and mosquito populations.

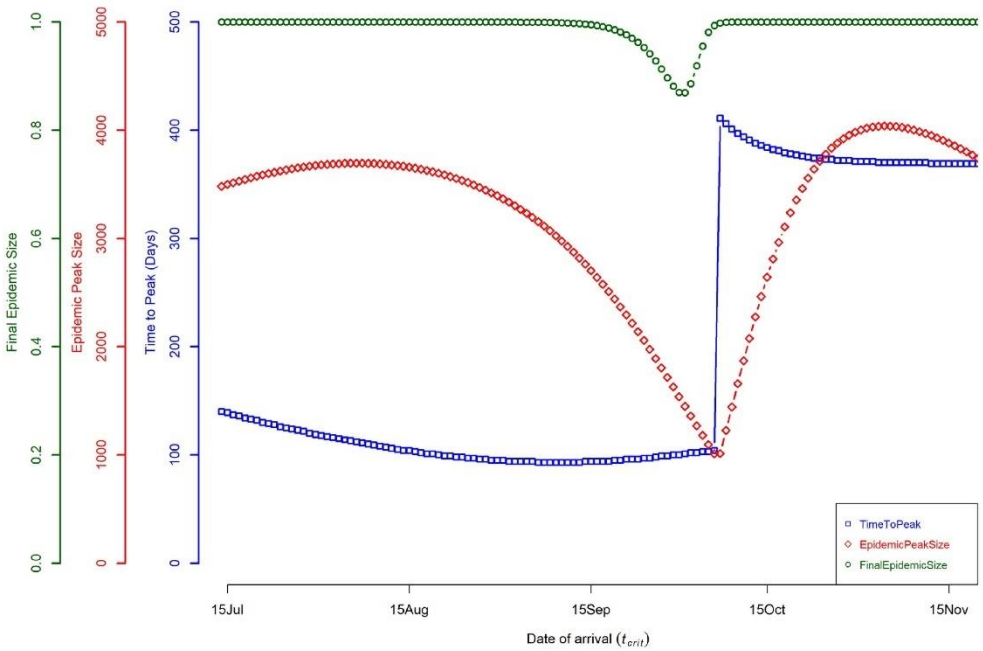

Fig 3. Quantities of interest (QOI) as a function of arrival date. QOI vs date of arrival that resulted in an epidemic by calendar date. The x-axis is the date of arrival of an infectious human, the simulation start date of February 15. The blue square points represent the time to peak infection in humans (in days). The red diamond points represent the maximum number of humans infected at any given point during the simulation. The green circle points represent the final (or cumulative) epidemic size at the end of the simulation; this is represented as the proportion of humans infected (rather than number). Simulations for arrivals dates in July to September had a classical unimodal peak, while other dates resulted in a prolonged low-level transmission, with bimodal peaks.

https://doi.org/10.1371/journal.pntd.0008679.g003

temperatures at onset produce larger epidemics within a shorter peak time and vice versa (Fig 4A). The final epidemic size was insensitive to starting temperatures, as all simulations within this regime produced final epidemic sizes of 100% of the population infected.

In the second set of temperature regimes (future; T_{mean} varied from 15.0˚C to 30.0˚C and T_{range} varied from 0.0˚C to 15.0˚C, both in increments of 0.2˚C), epidemic peak time and size show similar inverse variation as a function of starting temperature, although the overall behavior is different. Epidemic peak size had a unimodal distribution with its peak at ~28 ˚C and steadily declines afterward. Conversely the time to peak starts to increase at ~17 ˚C to ~19 ˚C, then steadily decreases until ~28 ˚C. At this peak of ~28 ˚C, the epidemic peak size was at 19% of the susceptible population infected, within a short peak time of 73 days. On the other hand, the final epidemic size steeply increases as a function of starting temperature, and plateaus at ~21 ˚C to ~30 ˚C (with 100% of the population infected), before steeply declining (Fig 4B). Extending the temperature regimes demonstrates the nonlinear influence of the interaction between mean temperature and seasonal variance on the epidemic dynamics.

To further examine this, Fig 5 shows the variation in the final epidemic size across the annual temperature bands. At a low mean annual temperature of 15˚C to 17˚C and range of 0˚C to ~7.5˚C, no epidemic occurred; however, this temperature band becomes suitable for

Figure 2-16 Quantities of interest (QOI) as a function of arrival date.

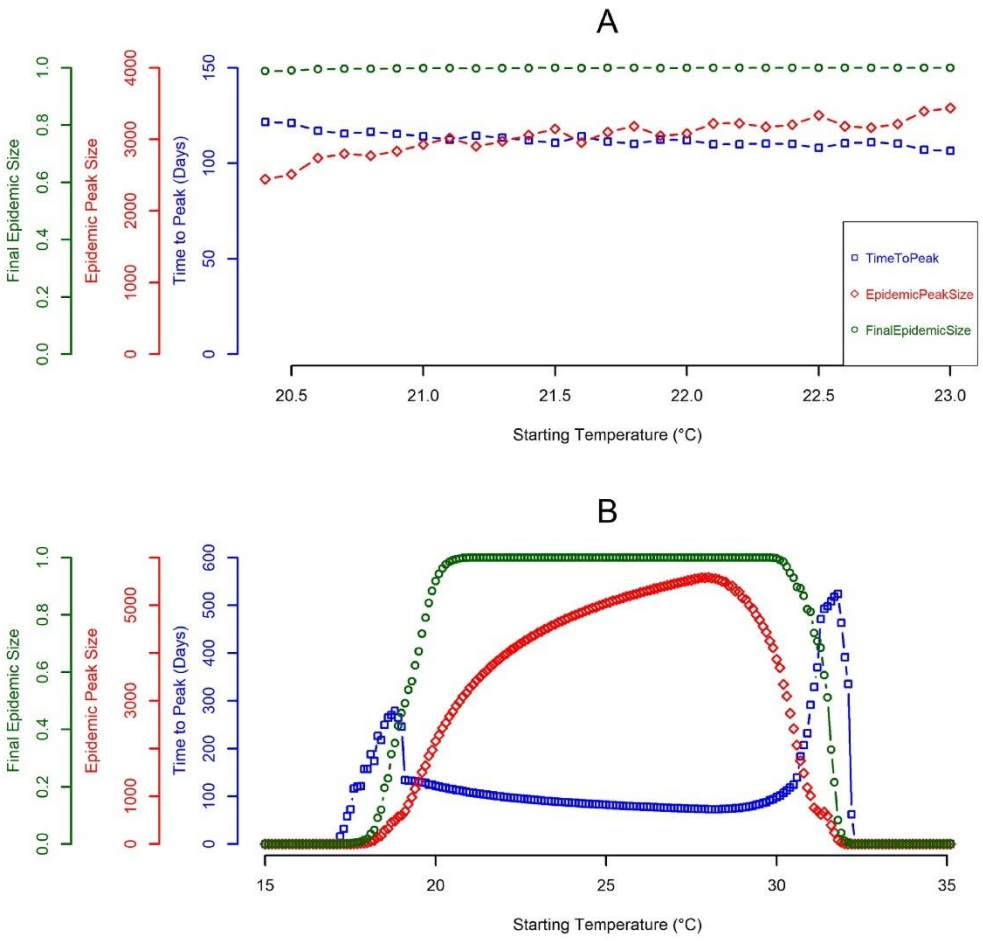

Fig 4. Quantities of interest (QOI) as a function of starting temperature. The x-axis is the starting temperatures within each set of temperature regimes and the y-axis is the associated value of the QOI being considered. Starting temperature is the temperature on August 15 calculated from T_{mean} and T_{range}. The blue square points represent the time to peak infection in humans (in days). The red diamond points represent the maximum number of humans infected at any given point during the simulation. The green circle points represent represents the final (or cumulative) epidemic size at the end of the simulation; this is represented as the proportion of humans infected (rather than number). (A) represents the historical temperature regimes, given by T_{mean} varied from 19˚C to 22˚C and T_{range} varied from 4˚C to 6˚C (both in increments of 0.2˚C), a total of 121 simulations. (B) represents the second set of temperature regimes, given by T_{mean} varied from 15˚C to 30˚C and T_{range} varied from 0˚C to 15.0˚C, a total of 5776 simulations. Due to the fixed starting date, multiple combinations of T_{mean} and T_{range} had the same starting temperature. As no other parameters were varied in these simulation sets, QOI and model behavior was identical for simulations with the same starting temperatures and overlapping points are not visible on the graphs. The default parameters from Table 2 were used for these simulations.

https://doi.org/10.1371/journal.pntd.0008679.g004

transmission as the temperature range increases beyond this point. Mean annual temperature of ~18˚C to ~26˚C, supports epidemic transmission at both low and high-temperature ranges. Epidemics introduced within these temperature regimes had the highest epidemic suitability (using final epidemic size as a measure of epidemic suitability). Lastly, mean annual

Figure 2-17 Quantities of interest (QOI) as a function of starting temperature.

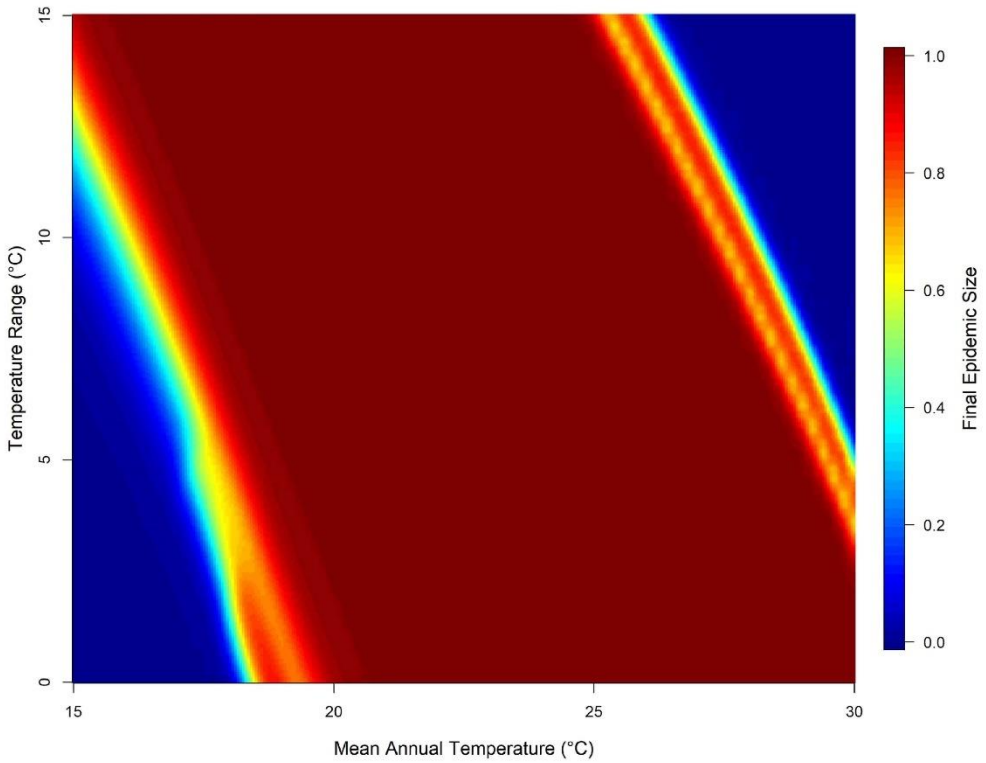

Fig 5. Final epidemic size across different seasonal temperature regimes. Heat map of final epidemic size (represented as the proportion of humans infected rather than number) as a function of mean annual temperature and temperature range. Temperature regimes here is given by T_{mean} varied from 15˚C to 30˚C and T_{range} varied from 0˚C to 15˚C, a total of 5776 simulations. Default parameters from Table 2 were used for these simulations, except for temperature parameters.

https://doi.org/10.1371/journal.pntd.0008679.g005

temperature bands of ~27˚C to 30˚C and low range of 0˚C to ~5.0˚C, transmission was supported with high epidemic suitability, this steadily diminishes as temperature range increases. In general, under the conditions used here, the dynamics of the epidemic are largely driven by the seasonal variation in temperature.

Model sensitivity analysis

Model sensitivity was characterized using 500 simulations runs for dengue parameter ranges as given in Table 2. All parameters were uniformly distributed for the LHS sampling. The conditions here were permissive for epidemics, with ($I_h > 2$) in 256 of the 500 simulated parameter sets. The epidemic progression in these parameter sets was consistent with the general model behavior seen in Fig 2A and 2C (i.e. a combination of unimodal and bimodal responses). The distribution of the QOI had a wider range of values (i.e. $MaxI_h$ = 3 to 7595 humans infected; $tmaxI_h$ = 31 to 512 days from t_{crit}; $cumI_h$ = 0.01% to 100% of the population infected), thus providing adequate variation for the sensitivity analysis.

Figure 2-18 Final epidemic size across different seasonal temperature regimes.

Table 2-13 Sensitivity analysis of the model's quantities of interest (QOI).

Table 3. Sensitivity analysis of the model's quantities of interest (QOI).

Parameter	$MaxI_h$		$tmaxI_h$		$cumI_h$	
	Coefficient	Relative Importance	Coefficient	Relative Importance	Coefficient	Relative Importance
t_{crit}	0.40	**1.00**	0.43	**1.00**	0.38	**1.00**
T_{mean}	0.51	**1.00**	0.14	**1.00**	0.16	**1.00**
T_{range}	-0.20	**1.00**	-0.02	0.45	-0.13	**1.00**
$\beta_{h\to v}$	0.53	**1.00**	0.08	**1.00**	0.23	**1.00**
$\beta_{v\to h}$	0.32	**1.00**	-0.02	0.46	0.13	**1.00**
μ_v	-0.65	**1.00**	-0.05	0.85	-0.24	**1.00**
γ_v	0.05	0.38	0.00	0.25	0.01	0.33
iv	0.01	0.28	-0.01	0.35	0.00	0.27
q	0.01	0.27	0.01	0.38	0.00	0.27
R_{tot}	0.43	**1.00**	0.02	0.49	0.13	**1.00**

Relative importance and model-averaged parameter coefficient from best-supported models (confidence set of models, see S2 File for details). QOI are denoted by: $maxI_h$ = the epidemic peak size; $tmaxI_h$ = time to peak; $cumI_h$ = final epidemic size. Before model fitting, parameters were centered on zero and scaled to unit variance, to normalize parameters within the same range, hence coefficient estimates are within a range of 0.0 to 1.0; likewise the relative importance of parameters is on a scale between 0.00 and 1.00. The model-averaged coefficients effects are summarized by their direction (+|−), positive indicates the QOI increases as the parameter increases, while negative indicates QOI decreases as the parameter increases. Relative importance ≥0.90 appears in bold, an arbitrary cut-off for visualizing most important parameters. Notations and descriptions of parameters are denoted in Table 2.

https://doi.org/10.1371/journal.pntd.0008679.t003

Table 3 shows the relative importance of the input parameters to our QOI, using uniform LHS distributions for dengue parameters (Table 2). Considering the main effects only, our QOI were all sensitive to the arrival date of the infectious human and mean annual temperature. Epidemic size (i.e. peak and final) were both sensitive to temperature range, while time to peak was less sensitive. Other influential parameters on all the QOI were transmission rates, mortality rate, and mosquito recruitment. The time to epidemic peak was less sensitive to the mosquito life-history trait parameters in comparison to other QOI (Table 3).

All the QIO were sensitive to the interaction term between the arrival date of an infectious human and mean annual temperature and the interaction term between mean annual temperature and temperature range. Another interactive term that was influential to all QOI is the interaction between mean annual temperature and mosquito recruitment. Interactions between transmission rates and other parameters had high relative importance for time to peak, indicating that transmission rates may alter the effect of other parameters (see S2 File, S7 Table for full details).

Discussion

We extend a previous model for arbovirus transmission with multiple vectors [26] to a deterministic compartmental model for dengue fever in *Aedes aegypti* in Madeira Island. This SEI-SEIR model explores dengue epidemic dynamics as triggered by the arrival of an infected person in Funchal, Madeira, and the effects of seasonally varying temperature on transmission. Our analysis focused on three quantities of interest: time to epidemic peak, epidemic peak size, and the final size of the epidemic. We then used a global sensitivity analysis to determine which input parameters were most important to quantities of interest.

Our model demonstrates that the date of arrival of an infected human in the susceptible human population dramatically affects the overall epidemic dynamics. Arrivals in mid-summer (July) to early-autumn (September) produce epidemics more quickly after introduction with a resulting large final epidemic size. These transient epidemics are indicative of a higher

transmission potential at this point of the year and are consistent with the scenario of the 2012 outbreak [6]. Our results are also consistent with the conclusions of [8], regarding the time point in the year with the highest epidemic risk and when local control strategies should be intensified. The overall model behavior in response to the timing of virus introduction within the susceptible population is similar to other previous dengue models [27, 57].

Furthermore, our results show the interaction between seasonal temperature mean and temperature range on the epidemic dynamics. The historical temperature seasonality for Funchal, Madeira, demonstrated a highly suitable thermal environment for the vectorial capacity of *Aedes aegypti* mosquitoes and arbovirus transmission. As expected, given a favorable day of introduction, all our model simulations for this temperature regime evolved into an epidemic, with warmer temperatures at onset producing high prevalence and faster peaks. Our subsequent extension of the temperature regimes demonstrates the nonlinear influence of temperature seasonality on the epidemic dynamics. Likewise, it allowed us to examine plausible future forcing scenarios based on near-term projections of seasonal temperature changes. Overall, the thermal responses from both temperature regimes were similar to those discussed in the works of [17] and [18], which investigated the effects of temperature on dengue transmission.

Our sensitivity analysis further characterizes the variability of the epidemic dynamics to other parameters in our model. Beyond the arrival date of an infected human and temperature seasonality, the quantities of interest were also sensitive to transmission rates, mosquito mortality rate, and mosquito population size. This is consistent with other modeling studies that highlight the importance of the mosquito dynamics on epidemic outcomes [28, 58, 59]. Our sensitivity analysis went a step further to characterize the interaction effects of our model parameters. An example is the interaction term between mean temperature and the mosquito population size as demonstrated in other previous models (e.g. [13, 19, 60]). Though this was not the focus of our analysis, it can be useful in improving understanding of the complex processes that interact to determine epidemic dynamics. Likewise, from a control and mitigation perspective, the interaction term effects are useful to understand how the implementation of a specific control strategy can have dramatic effects on the mosquito life-history traits and in turn the overall epidemic dynamics. A cautionary note, our sensitivity analysis varied only the mosquito portion of the transmission cycle and does not account for the variance in the human transmission and its combined influence on the epidemic dynamics.

Putting this all together, our model is a useful mathematical tool to study different epidemic scenarios and shifts in epidemic suitability for other regions on the island, other than the south coast region, where Funchal is located. The model temperature seasonality was parameterized using the observed temperature of Funchal; however, the results can be extrapolated to other regions on the island. The near term projection (i.e. for 2016–2035) for seasonal temperature variation for the region of Madeira island (i.e. the North Atlantic, Europe, and Mediterranean region), predicts a warming of ~2˚C in annual mean temperature, with summer months of June, July and August (JJA) temperatures warming up to ~2.9˚C [46, 51]. Given our results, we can speculate that the north coast and inland regions of the island, where current annual mean temperatures suggests limited or no epidemic suitability (e.g. Santana, a municipality in the north region with annual temperatures of 16.7˚C) could have an increased potential to support transmission in the near future. The south coast region would continue to support transmission, with possible higher epidemic suitability. To reiterate, most of the simulated temperatures are very unlikely to occur on the island, except for sporadic extreme summer temperatures and heatwaves. Such sporadic extreme events have been documented, with recent abrupt changes in temperatures on the island [52, 53]. Hence our simulations provide an estimate of epidemic suitability in the likelihood of these extreme temperature conditions. Also, our results can be applicable to other islands, within the Macaronesia ecoregion with

similar climatic conditions as Madeira. For example, the island of Santiago, located in the archipelago of Cape Verde, with similar *Aedes aegypti* populations. Our results can provide insights into epidemic suitability shifts and a potential outbreak of dengue on this island. It is important to note that in real life, these projected shifts in epidemic suitability will also be influenced by other factors like rainfall, humidity, and anthropogenic activities. Thus, seasonal temperature variation must be considered jointly with these factors.

Our model did not explicitly address the scenario of multiple serotypes of dengue co-circulating in the population, and the possibility of prior exposure leading to immune interactions and increased risk of severe disease (a possibility for Funchal, given the circulating serotype for the 2012 outbreak was DENV I). However, we made the following assumption when setting the constants for the human components of the transmission cycle. We consider that infection by a serotype produces permanent immunity to it, and temporal immunity to other serotypes [42, 61, 62]. Thus, we assumed that individuals infected from the 2012 outbreak, while they have permanent immunity to the DENV-1 serotype they remain susceptible to other serotypes. Hence the constants for the intrinsic incubation and infectious periods were set to reflect an introduction of a different serotype other than DENV-1. These constants reflect mean values for DENV-2 and DENV-4 as reported in previous literature [39–43]. The analysis presented here is relevant for the introduction of a new serotype on the island, which is timely in the light of the work of [22]. Their study had demonstrated the ability of the *Aedes aegypti* populations from Madeira to transmit DENV-2, and the potential risk for the local transmission if introduced to Madeira. Our results suggest that the epidemic dynamics would be significantly impacted by the introduction of a different serotype. This emphasizes the crucial need for the island to intensify its control measures to prevent local transmission while preparing effective mitigation strategies in the event of another outbreak.

In summary, the model presented here is relevant for the introduction of a new dengue serotype into Funchal, Madeira Island, and the interaction between mean temperature and seasonal variation to drive the epidemic dynamics. Our results demonstrate the potential for summer and autumn season transmission of dengue, with varying levels of epidemic suitability, following an introduction of the virus. Overall, we demonstrated that epidemic dynamics are strongly influenced by variation in the date of arrival, seasonal temperature, transmission rates, mortality rate, and the mosquito population. The model sensitivity analysis provides insight into the relative importance of these parameters and their interactive effects as mechanistic drivers of an epidemic. These results can be a useful guide in the development of effective local control and mitigation strategies for dengue fever in Madeira Island.

Supporting information

S1 Fig. Location of Madeira Island. Map of Europe in light blue, Portugal and Azorean islands in dark blue, Madeira island circled in red.
(TIF)

S2 Fig. Location of Ovitraps in Madeira Island. The distribution of ovitrap location across the island, as reported in the Entomological panel bulletins (PEnt_RAMW52/2019, in Portuguese) of the Institute of Health Administration, IP-RAM (IASAÚDE IP-RAM). The colored dots on the map, represents the cumulative number of eggs in each location for the week 10, 2018 to week 52, 2019. Blue = 0%, Dark green = 1%-10%, Light green = 11%–20%, Yellow = 21%–30%, Orange = 31%–40%, Red = 41%–100%. Cumulative number of eggs for all location = 37766. Adapted from Institute of Health Administration, IP-RAM [34].
(TIF)

S3 Fig. Cumulative results of Ovitraps in Madeira Island. Annual cumulative number of eggs reported across the island from all the ovitraps, as reported in the Entomological panel bulletins (PEnt_RAMW52/2019, in Portuguese) of the Institute of Health Administration, IP-RAM (IASAÚDE IP-RAM). Entomological year starts from week 10. Graphs shows cumulative number of eggs for week 10, 2012 to week 52, 2019, with each entomological year represented in a different line color. Adapted from Institute of Health Administration, IP-RAM [34].
(TIF)

S4 Fig. Location of BG-traps in Madeira Island. The distribution of BG-traps location across the island, as reported in the Entomological panel bulletins (PEnt_RAMW52/2019, in Portuguese) of the Institute of Health Administration, IP-RAM (IASAÚDE IP-RAM). The colored dots on the map, represents the cumulative number of adult mosquitoes in each location for the week 10, 2018 to week 52, 2019. Blue = 0%, Dark green = 1%-10%, Light green = 11%–30%, Yellow = 31%–50%, Orange = 51%–60%, Red = 61%–100%. Cumulative number of adult mosquitoes for all location = 3464. Adapted from Institute of Health Administration, IP-RAM [34].
(TIF)

S5 Fig. Cumulative results of BG-traps in Madeira Island. Annual cumulative number of adult mosquitos reported across the island from all the BG-traps, as reported in the Entomological panel bulletins (PEnt_RAMW52/2019, in Portuguese) of the Institute of Health Administration, IP-RAM (IASAÚDE IP-RAM). Entomological year starts from week 10. Graphs shows cumulative number of eggs for week 40, 2012 to week 52, 2019, with each entomological year represented in a different line color. Adapted from Institute of Health Administration, IP-RAM [34].
(TIF)

S6 Fig. Seasonal temperature pattern. The temperature seasonality pattern as observed from the historical decadal data for Funchal, Madeira. Average of minimum (green line), mean (blue line) and maximum (red line) temperatures per day between 2008 and 2018. Black dashed line is daily mean temperature of our model as calculated from sinusoidal curve (Eq 16 in the main article). The long-term average conditions in Funchal, Madeira $T_{max} = 22^{\circ}C$, $T_{mean} = 20^{\circ}C$, $T_{min} = 17^{\circ}C$. Average mean temperatures for the spring months (Mar–May) was 18˚C; Summer months (Jun–Aug) was 22˚C; Autumn months (Sep–Nov) was 21˚C and Winter months (Dec–Feb) was 17˚C.
(TIF)

S1 File. S1 Appendix. Model starting conditions.
(DOCX)

S2 File. S2 Appendix. Model sensitivity analysis and multi-model inference analysis.
(DOCX)

Acknowledgments

We appreciate Joe Pohedra who assisted in the programming of the initial transmission model and Jebidiah Light who assisted with the adaption to the current SEI-SEIR model.

Author Contributions

Conceptualization: Donald Salami, César Capinha, Carla Alexandra Sousa, Maria do Rosário Oliveira Martins, Cynthia Lord.

Formal analysis: Donald Salami.

Methodology: Donald Salami, Cynthia Lord.

Software: Cynthia Lord.

Supervision: César Capinha, Carla Alexandra Sousa, Maria do Rosário Oliveira Martins, Cynthia Lord.

Visualization: Donald Salami.

Writing – original draft: Donald Salami.

Writing – review & editing: Donald Salami, César Capinha, Carla Alexandra Sousa, Maria do Rosário Oliveira Martins, Cynthia Lord.

References

1. Messina JP, Brady OJ, Golding N, Kraemer MUG, Wint GRW, Ray SE, et al. The current and future global distribution and population at risk of dengue. Nat Microbiol. 2019. https://doi.org/10.1038/s41564-019-0476-8.
2. Kraemer MUG, Hay SI, Pigott DM, Smith DL, Wint GRW, Golding N. Progress and challenges in infectious disease cartography. Trends Parasitol. 2016; 32(1):19–29. https://doi.org/10.1016/j.pt.2015.09.006. PMID: 26604163
3. Leta S, Beyene TJ, De Clercq EM, Amenu K, Kraemer MUG, Revie CW. Global risk mapping for major diseases transmitted by *Aedes aegypti and Aedes albopictus*. Int J Infect Dis. 2018; 67:25–35. https://doi.org/10.1016/j.ijid.2017.11.026. PMID: 29196275
4. Margarita Y, Grácio AS, Lencastre I, Silva A, Novo T, Sousa C. First record of *Aedes* (*stegomyia*) *aegypti* (Linnaeus, 1762) (Diptera, culicidae) in Madeira Island—Portugal (in portguese) Acta Parasitol Port. 2006; 13:59–61.
5. Medlock JM, Hansford KM, Schaffner F, Versteirt V, Hendrickx G, Zeller H, et al. A review of the invasive mosquitoes in Europe: ecology, public health risks, and control options. Vector Borne Zoonotic Dis. 2012; 12(6):435–47. https://doi.org/10.1089/vbz.2011.0814. PMID: 22448724
6. ECDC. Dengue outbreak in madeira (2012–13) 2013. European Centre for Disease Prevention and Control [17 Jul 2019]. https://ecdc.europa.eu/en/dengue-fever/threats-and-outbreaks/madeira-outbreak-2012.
7. ECDC. Update on autochthonous dengue cases in Madeira, Portugal. Stockholm: European Centre for Disease Prevention and Control, 2012. https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/dengue-madeira-risk-assessment-update.pdf.
8. Lourenco J, Recker M. The 2012 Madeira dengue outbreak: epidemiological determinants and future epidemic potential. PLoS Negl Trop Dis. 2014; 8(8):e3083. https://doi.org/10.1371/journal.pntd.0003083. PMID: 25144749
9. Watts DM, Burke DS, Harrison BA, Whitmire RE, Nisalak A. Effect of temperature on the vector efficiency of *Aedes aegypti* for dengue 2 virus. Am J Trop Med Hyg.1987; 36(1):143–52. https://doi.org/10.4269/ajtmh.1987.36.143. PMID: 3812879
10. Rueda LM, Patel KJ, Axtell RC, Stinner RE. Temperature-dependent development and survival rates of *Culex quinquefasciatus* and *Aedes aegypti* (Diptera: Culicidae). J Med Entomol. 1990; 27(5):892–8. https://doi.org/10.1093/jmedent/27.5.892 PMID: 2231624
11. Eisen L, Monaghan AJ, Lozano-Fuentes S, Steinhoff DF, Hayden MH, Bieringer PE. The impact of temperature on the bionomics of *Aedes* (*stegomyia*) aegypti, with special reference to the cool geographic range margins. J Med Entomol. 2014; 51(3):496–516. https://doi.org/10.1603/ME13214 PMID: 24897844
12. Couret J, Dotson E, Benedict MQ. Temperature, larval diet, and density effects on development rate and survival of *Aedes aegypti* (Diptera: Culicidae). PloS One. 2014; 9(2):e87468. https://doi.org/10.1371/journal.pone.0087468. PMID: 24498328

13. Yang HM, Macoris ML, Galvani KC, Andrighetti MT, Wanderley DM. Assessing the effects of temperature on the population of *Aedes aegypti*, the vector of dengue. Epidemiol Infect. 2009; 137(8):1188–202. https://doi.org/10.1017/s0950268809002040. PMID: 19192322

14. Johansson MA, Powers AM, Pesik N, Cohen NJ, Staples JE. Nowcasting the spread of chikungunya virus in the Americas. PloS One. 2014; 9(8):e104915. https://doi.org/10.1371/journal.pone.0104915. PMID: 25111394

15. Liu-Helmersson J, Stenlund H, Wilder-Smith A, Rocklöv J. Vectorial capacity of *Aedes aegypti*: effects of temperature and implications for global dengue epidemic potential. PloS One. 2014; 9(3):e89783. https://doi.org/10.1371/journal.pone.0089783. PMID: 24603439

16. Morin CW, Monaghan AJ, Hayden MH, Barrera R, Ernst K. Meteorologically driven simulations of dengue epidemics in San Juan, PR. PLoS Negl Trop Dis. 2015; 9(8):e0004002. https://doi.org/10.1371/journal.pntd.0004002. PMID: 26275146

17. Mordecai EA, Cohen JM, Evans MV, Gudapati P, Johnson LR, Lippi CA, et al. Detecting the impact of temperature on transmission of zika, dengue, and chikungunya using mechanistic models. PLoS Negl Trop Dis. 2017; 11(4):e0005568. https://doi.org/10.1371/journal.pntd.0005568. PMID: 28448507

18. Huber JH, Childs ML, Caldwell JM, Mordecai EA. Seasonal temperature variation influences climate suitability for dengue, chikungunya, and zika transmission. PLoS Negl Trop Dis. 2018; 12(5):e0006451. https://doi.org/10.1371/journal.pntd.0006451. PMID: 29746468

19. Mordecai EA, Caldwell JM, Grossman MK, Lippi CA, Johnson LR, Neira M, et al. Thermal biology of mosquito-borne disease. Ecol Lett. 2019; 0(0). https://doi.org/10.1111/ele.13335.

20. Fernandopullé D. Climatic characteristics of the Canary Islands. In: K G., editor. Biogeography and ecology in the Canary Islands: Springer, Dordrecht; 1976. p. 185–206.

21. Santos FD, Valente MA, Miranda PMA, Aguiar A, Azevedo EB, Tomé AR, et al. Climate change scenarios in the Azores and Madeira Islands. World Resour Rev. 2004; 16(4):473–91. http://idlcc.fc.ul.pt/pdf/SantosEtalWRR2004.pdf.

22. Seixas G, Jupille H, Yen P-S, Viveiros B, Failloux A-B, Sousa CA. Potential of *Aedes aegypti* populations in Madeira Island to transmit dengue and chikungunya viruses. Parasit Vectors. 2018; 11(1):509. https://doi.org/10.1186/s13071-018-3081-4. PMID: 30208974

23. Messina JP, Brady OJ, Pigott DM, Brownstein JS, Hoen AG, Hay SI. A global compendium of human dengue virus occurrence. Sci Data. 2014; 1:140004. https://doi.org/10.1038/sdata.2014.4. PMID: 25977762

24. Guo C, Zhou Z, Wen Z, Liu Y, Zeng C, Xiao D, et al. Global epidemiology of dengue outbreaks in 1990–2015: A systematic review and meta-analysis. Front Cell Infect Microbiol. 2017; 7:317-. https://doi.org/10.3389/fcimb.2017.00317. PMID: 28748176

25. Ryan SJ, Carlson CJ, Mordecai EA, Johnson LR. Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. PLoS Negl Trop Dis. 2019; 13(3):e0007213. https://doi.org/10.1371/journal.pntd.0007213. PMID: 30921321

26. Lord CC, Lounibos LP, Pohedra JJ, Alto BW. Effects of mosquito biology on modeled chikungunya virus invasion potential in Florida. Viruses 2020, 12(8), 830; https://doi.org/10.3390/v12080830.

27. Otero M, Solari HG. Stochastic eco-epidemiological model of dengue disease transmission by Aedes aegypti mosquito. Math Biosci. 2010; 223(1):32–46. https://doi.org/10.1016/j.mbs.2009.10.005. PMID: 19861133

28. Manore CA, Hickmann KS, Xu S, Wearing HJ, Hyman JM. Comparing dengue and chikungunya emergence and endemic transmission in A. aegypti and A. albopictus. J Theor Biol. 2014; 356:174–91. https://doi.org/10.1016/j.jtbi.2014.04.033. PMID: 24801860

29. Karl S, Halder N, Kelso JK, Ritchie SA, Milne GJ. A spatial simulation model for dengue virus infection in urban areas. BMC Infect Dis. 2014; 14:447. https://doi.org/10.1186/1471-2334-14-447. PMID: 25139524

30. Ferreira CP, Godoy WAC. Ecological modelling applied to entomology. Switzerland: Springer; 2014.

31. Institute of Health Administration, IP-RAM. Mosquito on Madeira Island (in Portuguese). Institute of Health Administration, IP-RAM, Regional Secretariat of Health, Autonomous Region of Madeira; 2019 [17 Jul 2019]. http://doc.iasaude.pt/mosquito/index.php/mosquito/mosquito-na-ilha-da-madeira.

32. Duong V, Lambrechts L, Paul RE, Ly S, Lay RS, Long KC, et al. Asymptomatic humans transmit dengue virus to mosquitoes. Proc Natl Acad Sci. 2015; 112(47):14688–93. https://doi.org/10.1073/pnas.1508114112. PMID: 26553981

33. Newton EA, Reiter P. A model of the transmission of dengue fever with an evaluation of the impact of ultra-low volume (ulv) insecticide applications on dengue epidemics. The Am J Trop Med Hyg. 1992; 47(6):709–20. https://doi.org/10.4269/ajtmh.1992.47.709. PMID: 1361721

34. Menach AL, McKenzie FE, Flahault A, Smith DL. The unexpected importance of mosquito oviposition behaviour for malaria: non-productive larval habitats can be sources for malaria transmission. Malar J. 2005; 4(1):23. https://doi.org/10.1186/1475-2875-4-23.

35. Institute of Health Administration, IP-RAM. Entomological panel bulletins(in Portuguese). IASAÚDE IP-RAM, Sanitary Engineering Unit: Institute of Health Administration, IP-RAM, Regional Secretariat of Health, Autonomous Region of Madeira, 2019. http://doc.iasaude.pt/mosquito/index.php/boletins/entomologicos.

36. Scott TW, Amerasinghe PH, Morrison AC, Lorenz LH, Clark GG, Strickman D, et al. Longitudinal studies of aedes aegypti (diptera: Culicidae) in thailand and puerto rico: blood feeding frequency. J. Med. Entomol. 2000; 37(1):89–101. https://doi.org/10.1603/0022-2585-37.1.89. PMID: 15218911

37. Paupy C, Ollomo B Fau—Kamgang B, Kamgang B Fau—Moutailler S, Moutailler S Fau—Rousset D, Rousset D Fau—Demanou M, Demanou M Fau—Herve J-P, et al. Comparative role of Aedes albopictus and Aedes aegypti in the emergence of dengue and chikungunya in central Africa. Vector Borne Zoonotic Dis. 2010; 10(3). https://doi.org/10.1089/vbz.2009.0005.

38. Lord CC. The effect of multiple vectors on arbovirus transmission. Isr J Ecol Evol. 2010; 56(3–4):371–92. https://doi.org/10.1560/ijee.55.3-4.371. PMID: 23741205

39. Focks DA, Barrera R. Dengue transmission dynamics: Assessment and implications for control. CiteSeer: World health organization on behalf of the special programme for research and training in tropical diseases, 2007. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.612.689.

40. Nishiura H, Halstead SB. Natural history of dengue virus (DENV)-1 and DENV-4 infections: reanalysis of classic studies. J Infect Dis. 2007; 195(7):1007–13. https://doi.org/10.1086/511825. PMID: 17330791

41. Chan M, Johansson MA. The incubation periods of dengue viruses. PloS One. 2012; 7(11):e50972. https://doi.org/10.1371/journal.pone.0050972. PMID: 23226436

42. Lourenco J, Recker M. Dengue serotype immune-interactions and their consequences for vaccine impact predictions. Epidemics. 2016; 16:40–8. https://doi.org/10.1016/j.epidem.2016.05.003. PMID: 27663790

43. Gubler DJ, Suharyono W, Tan R, Abidin M, Sie A. Viraemia in patients with naturally acquired dengue infection. Bull World Health Organ. 1981; 59(4):623–30. https://www.ncbi.nlm.nih.gov/pubmed/6976230.

44. Vaughn DW, Green S, Kalayanarooj S, Innis BL, Nimmannitya S, Suntayakorn S, et al. Dengue viremia titer, antibody response pattern, and virus serotype correlate with disease severity. J Infect Dis. 2000; 181(1):2–9. https://doi.org/10.1086/315215. PMID: 10608744

45. Santo António Parish Council. History of Santo António parish council (in Portuguese): junta de freguesia de Santo António- an autonomous region of Madeira; 2018 [17 Jul 2019]. http://www.jf-santoantonio.pt/historia.

46. Weather Underground. Weather history for Funchal Madeira (2008–2018) 2019 [01 Mar 2019]. https://www.wunderground.com/history/daily/pt/funchal/LPMA.

47. IPCC. Climate change 2013: The physical science basis. Contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, United Kingdom, and New York, NY, USA: 2013; [1535]. https://www.ipcc.ch/report/ar5/wg1/.

48. Funchal city hall. Funchal natural history museum: Funchal city hall (Câmara Municipal do Funchal) 2019. http://www.cm-funchal.pt/pt/servi%C3%A7os/ci%C3%AAncia/museu-de-hist%C3%B3ria-natural-do-funchal/sobre-o-museu.html.

49. Andraud M, Hens N, Marais C, Beutels P. Dynamic epidemiological models for dengue transmission: a systematic review of structural approaches. PloS One. 2012; 7(11):e49085. https://doi.org/10.1371/journal.pone.0049085. PMID: 23139836

50. Kiraly A, Janosi IM. Stochastic modeling of daily temperature fluctuations. Phys Rev. 2002; 65(1539–3755 (Print)):051102 1–6. https://doi.org/10.1103/PhysRevE.65.051102.

51. Lord CC, Day JF. Simulation studies of st. Louis encephalitis virus in south Florida. Vector Borne Zoonotic Dis. 2001; 1(4):299–315. https://doi.org/10.1089/15303660160025921. PMID: 12653129

52. Seneviratne SI, Field CB, Barros V, Stocker TF, Dahe Q, Dokken DJ, et al. Managing the risks of extreme events and disasters to advance climate change adaptation. Cambridge: Intergovernmental panel on climate change 2012; [582]. https://www.ipcc.ch/report/managing-the-risks-of-extreme-events-and-disasters-to-advance-climate-change-adaptation/.

53. Madeira Island News. Madeira heatwave continues. Madeira Island News, 2016 [17 Jul 2019]. https://www.madeiraislandnews.com/2016/08/heatwave-continues.html.

54. Madeira Island News. Hottest December day for 150 years. Madeira Island News, 2018 [17 Jul 2019]. https://www.madeiraislandnews.com/2018/12/hottest-december-day-for-150-years.html.

55. MathWorks. MATLAB and Statistics Toolbox Release 2016a. Natick, Massachusetts: The MathWorks Inc.; 2016.

56. Calcagno V, de Mazancourt C. Glmulti: an R package for easy automated model selection with (generalized) linear models. 2010. 2010; 34(12): J Stat Softw. https://doi.org/10.18637/jss.v034.i12.

57. R-Core-Team. The R project for statistical computing: CRAN; 2019. https://www.r-project.org/.

58. Erickson RA, Presley SM, Allen LJS, Long KR, Cox SB. A dengue model with a dynamic Aedes albopictus vector population. Ecol Model. 2010; 221(24):2899–908. https://doi.org/10.1016/j.ecolmodel.2010.08.036.

59. Ellis AM, Garcia AJ, Focks DA, Morrison AC, Scott TW. Parameterization and sensitivity analysis of a complex simulation model for mosquito population dynamics, dengue transmission, and their control. Am J Trop Med Hyg. 2011; 85(2):257–64. https://doi.org/10.4269/ajtmh.2011.10-0516. PMID: 21813844

60. Ndii MZ, Hickson RI, Allingham D, Mercer GN. Modelling the transmission dynamics of dengue in the presence of Wolbachia. Math Biosci. 2015; 262:157–66. https://doi.org/10.1016/j.mbs.2014.12.011. PMID: 25645184

61. Yang HM, Macoris ML, Galvani KC, Andrighetti MT, Wanderley DM. Assessing the effects of temperature on dengue transmission. Epidemiol Infect. 2009; 137(8):1179–87. https://doi.org/10.1017/s0950268809002052. PMID: 19192323

62. Wearing HJ, Rohani P. Ecological and immunological determinants of dengue epidemics. Proc Natl Acad Sci. 2006; 103(31):11802–7. https://doi.org/10.1073/pnas.0602960103. PMID: 16868086

2.3.1 Supporting Information

S1 Fig. https://doi.org/10.1371/journal.pntd.0008679.s001

S2 Fig. https://doi.org/10.1371/journal.pntd.0008679.s002

S3 Fig. https://doi.org/10.1371/journal.pntd.0008679.s003

S4 Fig. https://doi.org/10.1371/journal.pntd.0008679.s004

S5 Fig. https://doi.org/10.1371/journal.pntd.0008679.s005

S6 Fig. https://doi.org/10.1371/journal.pntd.0008679.s006

S1 File. https://doi.org/10.1371/journal.pntd.0008679.s007

S2 File. https://doi.org/10.1371/journal.pntd.0008679.s008

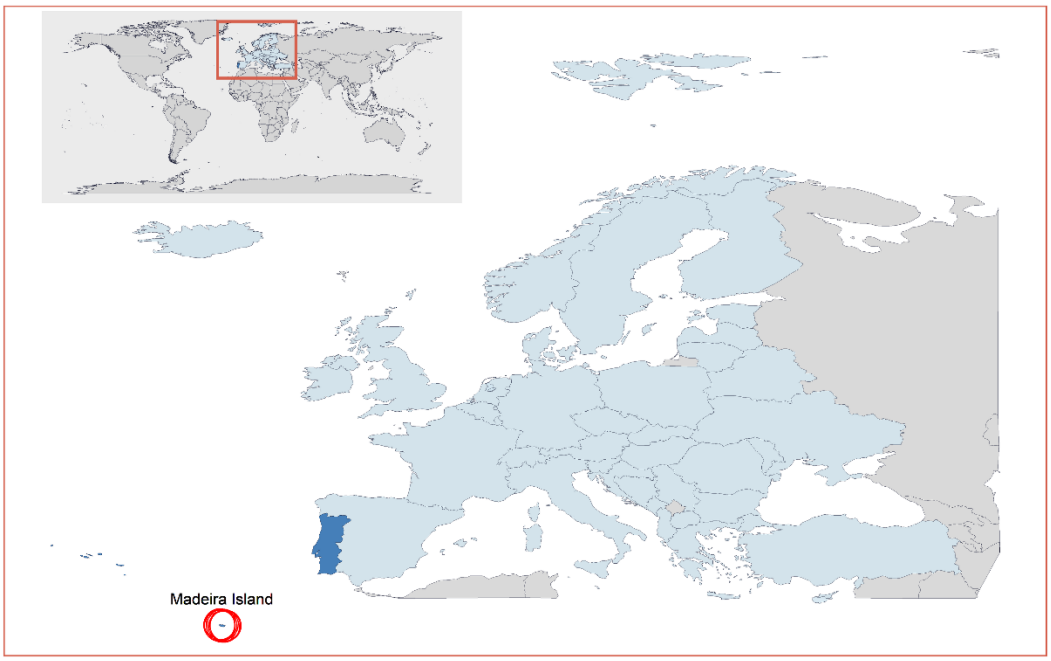

Figure 2-19 S1 Fig. Location of Madeira Island.

https://doi.org/10.1371/journal.pntd.0008679.s001

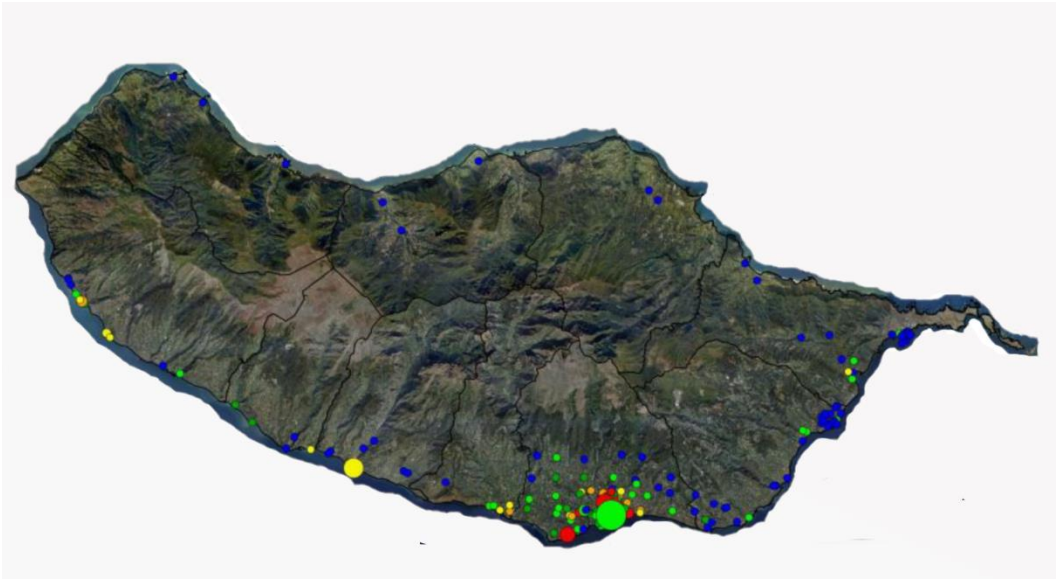

Figure 2-20 S2 Fig. Location of Ovitraps in Madeira Island.

https://doi.org/10.1371/journal.pntd.0008679.s002

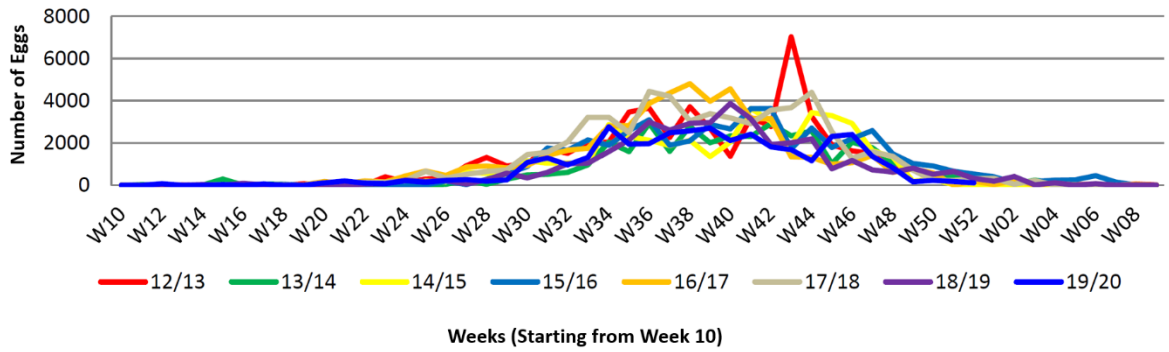

Figure 2-21 S3 Fig. Cumulative results of Ovitraps in Madeira Island.

https://doi.org/10.1371/journal.pntd.0008679.s003

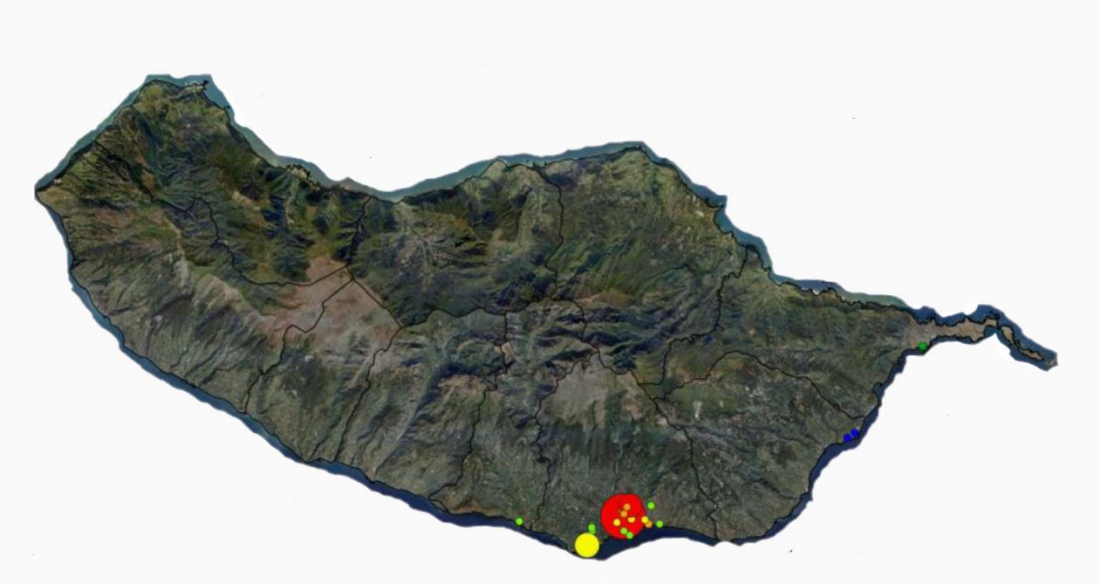

Figure 2-22 S4 Fig. Location of BG-traps in Madeira Island.

https://doi.org/10.1371/journal.pntd.0008679.s004

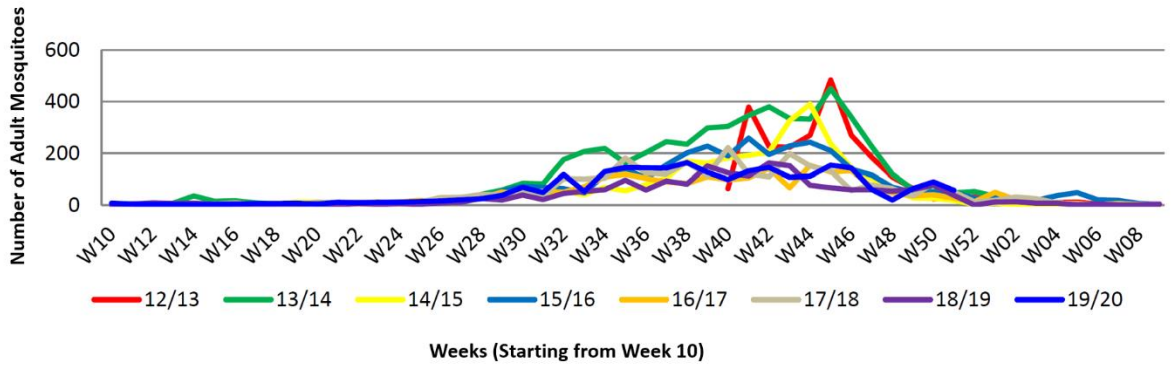

Figure 2-23 S5 Fig. Cumulative results of BG-traps in Madeira Island.

https://doi.org/10.1371/journal.pntd.0008679.s005

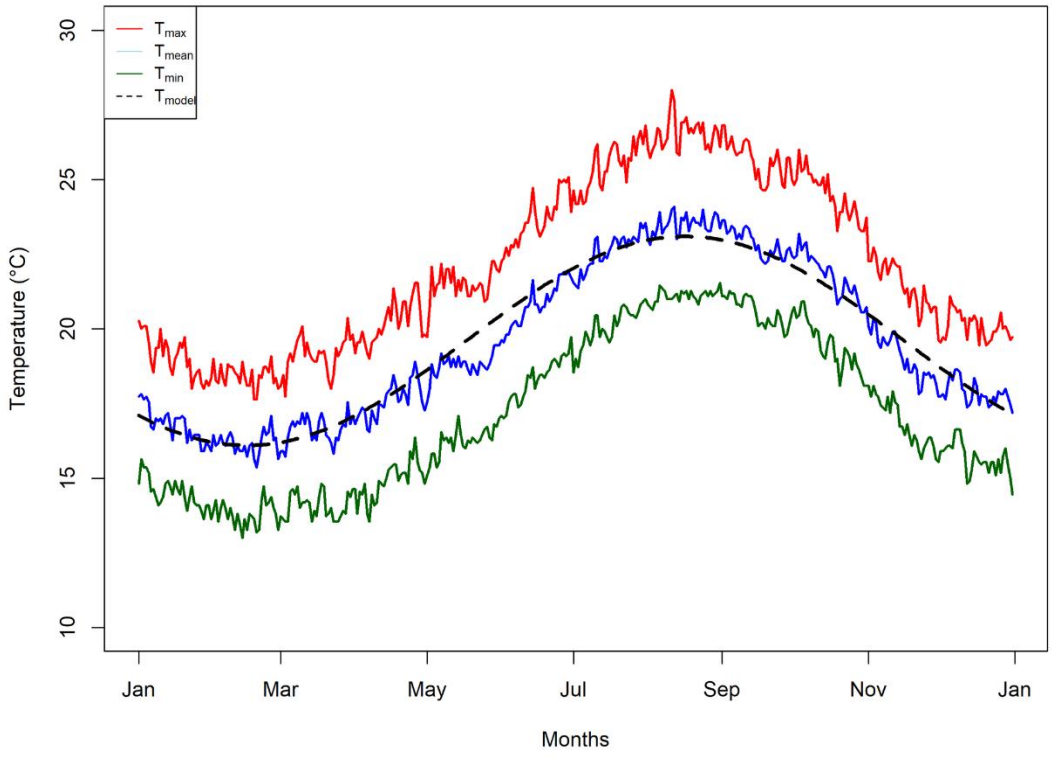

Figure 2-24 S6 Fig. Seasonal temperature pattern.

https://doi.org/10.1371/journal.pntd.0008679.s006

S1 Appendix. Model starting conditions

The model requires an estimate of each state variable (humans and mosquitoes in each class), along with estimates or values chosen from distributions for each parameter. Our parameter values are based on multiple citations from reviewed literature of previous empirical studies or lab trials and expert opinion (Table 2, of the main article). We emphasize previous studies within the context of Funchal, Madeira Island (since our main objective was to parameterize the model for the island) and chose parameter ranges to reflect conditions in Funchal.

The model assumes a homogeneously mixed population, with a total human population set at a constant 30,000 (representative of the population of the most populous civil parish –*Santo António* – in the municipality of Funchal and the island) (1). Since the human components of the transmission cycle are not seasonal, we set the intrinsic incubation ($1/\gamma_h$) and the infectious period ($1/\eta_h$) to constants of 6 and 4 days respectively. For this model, we considered only a single dengue serotype. However, these constants reflect mean values for DENV-2 and DENV-4 as reported in previous literature (2-6). Based on our mosquito recruitment term, the initial susceptible mosquito population is set as the number of females ($\rho_b(t)$) added to the population at intervals (iv), as calculated in equation 12 in the main article. With all infectious classes for both human and mosquitoes set to zero, an infection is triggered by the arrival of one infectious human on a specified day (t_{crit}) into the fully susceptible population. The default initial conditions are thus: ($S_h;\ E_h;\ I_h;\ R_h;\ S_v;\ E_v;\ I_v$) = ($N_h;\ 0;\ 0;\ 0;\ p_b(t);\ 0;\ 0$) where $N_h = 30{,}000$.

Simulations were set to start at the coldest day in the annual cycle (i.e. February 15) and ran for 730 days thereafter (allowing for simulation with t_{crit} later in the year). We set an arbitrary cut-off value for the exposed and infectious classes (i.e. human and mosquitoes): if (E_h, I_h, E_v, I_v, all <0.5) the simulation is terminated and restarted with the classes set to zero. This cutoff is necessary otherwise extremely low levels of infection may persist for long periods; in the natural system, there would be a high probability that the virus would go extinct (7).

We performed a preliminary exploration of parameter values to determine their effects on transmission. The choice of the final parameter values was based on permissibility for

transmission. The preliminary exploration also informed the initial conditions (described above) and parameter ranges for sensitivity analyses (Table 2). Model simulations were performed using the governing systems of differential equations of MATLAB's inbuilt routine "ode45" (8). Simulation outputs were analyzed in the R Programming Language version 3.5.3 (9).

References

1. Santo António Parish Council. History of Santo António parish council (in Portuguese): junta de freguesia de Santo António- an autonomous region of Madeira; 2018 (17 Jul 2019). Available from: http://www.jf-santoantonio.pt/historia.

2. Lourenco J, Recker M. Dengue serotype immune-interactions and their consequences for vaccine impact predictions. Epidemics. 2016; 16:40-8. doi: https://doi.org/10.1016/j.epidem.2016.05.003.

3. Nishiura H, Halstead SB. Natural history of dengue virus (DENV)-1 and DENV-4 infections: reanalysis of classic studies. J Infect Dis. 2007;195(7):1007–13. doi: https://doi.org/10.1086/511825.

4. Chan M, Johansson MA. The incubation periods of dengue viruses. PloS One. 2012;7(11): e50972. doi: https://doi.org/10.1371/journal.pone.0050972.

5. Gubler DJ, Suharyono W, Tan R, Abidin M, Sie A. Viraemia in patients with naturally acquired dengue infection. Bull World Health Organ. 1981;59(4):623-30. https://www.ncbi.nlm.nih.gov/pubmed/6976230

6. Vaughn DW, Green S, Kalayanarooj S, Innis BL, Nimmannitya S, Suntayakorn S, et al. Dengue viremia titer, antibody response pattern, and virus serotype correlate with disease severity. J Infect Dis. 2000;181(1):2-9. doi: https://doi.org/10.1086/315215

7. Lord CC, Day JF. Simulation studies of st. Louis encephalitis and west nile viruses: the impact of bird mortality. Vector Borne Zoonotic Dis. 2001;1(4):317-29. doi: https://doi.org/10.1089/15303660160025930.

8. MathWorks. MATLAB and Statistics Toolbox Release 2016a. Natick, Massachusetts: The MathWorks Inc.; 2016.

9. R-Core-Team. The R project for statistical computing: CRAN; 2019. Available from: https://www.r-project.org/.

S2 Appendix. Model sensitivity analysis

To characterize the model parameters exerting the most influence on our quantities of interest, we performed a variance-based global sensitivity analysis, using a combination of Latin hypercube sampling (LHS) and a multi-model inference on regression-based models. LHS is a stratified Monte Carlo sampling technique, where specified parameter distributions are divided into (*M)* equiprobable intervals, and then sampled, here (*M*) is the sample size. The entire range of each parameter is explored, by sampling each interval for each parameter only once without replacement. Parameters values are then randomly resorted into sets to use for simulation. The LHS method assumes that the sampling is performed independently for each parameter, thereby allowing for an unbiased estimate (1, 2). Parameter ranges for sampling were derived from existing literature, expert opinion, and field-based data (Table 2 of the main article), we assumed a uniform distribution for all parameter values.

Multi-model inference on a generalized least square (GLS) regression was used to estimate the relative importance of the input parameters. Multi-model inference uses the information-theoretic approach to offer a more objective way to assess the relative importance of input variables by inferring all possible models from a defined candidate set (3, 4). A vector of the input parameters (x_p) were fitted into all possible unique models and then ranked from best to worst, based on Akaike information criterion (AIC) values. An estimate of the relative importance of a single parameter (p) was then calculated by summing the Akaike weights (w) across all fitted models where parameter (p) occurs. Akaike weights are normalized, such that the sum over all models considered is 1. The relative importance of parameter (p) was quantified by the sum (w) for the parameter. The larger the sum of the weight (between 0 and 1 by definition) the more important the parameter is, relative to the other parameters (4). Input parameters were then ranked in terms of their importance to the quantities of interest.

The sensitivity analysis considered main effects and pairwise (first-order) interactions of input parameters on our quantities of interest.

Multi-model inference analysis

We used multi-model inference with *glm* function from the *glmulti* R package (5), to estimate the relative importance of our input variables on our QOIs. Our first analysis considers the main effects of the input variables on the QOIs. Utilizing the default method of the *glmulti* package (exhaustive screening of all candidate models), we built all possible unique models based on the 11 input variables. Thus, the candidate set of models for each QOIs main effect analysis contained a total of 2048 models. We defined the confidence set of models as those models within two Akaike information criterion (AIC) unit difference (ΔAIC) from the best model (4) However, the *glmulti* function by default returns a confidence set of 100 models and delineates the number of models within 2 ΔAIC from the best model.

Our second analysis considers the pairwise interactions between our input parameters. We utilized the genetic algorithm approach of the *glmulti* package, to fit a sub-set of all possible unique models (due to the extremely large number of possible candidate models). The genetic algorithm approach randomly explores a subset of all possible models, with a bias towards better models to fit the best fitting model. Our confidence set is as defined above, i.e. within 2 ΔAIC of the best model. However, due to a large number of possible candidate models, the size of the returned confidence set was increased to 500 models (from the default of 100).

Note, before the multi-model inference fitting, we normalized the input parameters (by centering on zero and scaled to unit variance) to allow comparison of resulting model-averaged estimates on a common scale.

S1- S3 Tables (main effects) and S4 – S6 Tables (pairwise interactions), presents the top five models in the confidence sets for each our QOIs, their log likelihood values, number of fitted parameters (K), Akaike information criterion (AIC), AIC weights, and the goodness of fit (R^2).

S7 Fig to S9 Fig (main effects) and S10 Fig to S12 Fig (pairwise interactions), shows the AIC profile of all fitted models.

S7 Table presents the model averaged coefficients and relative-importance weights for the parameters from the best model with the pairwise interactions. The definition and description of the parameter symbols are as presented in Table 2 of the main article.

Table 2-14 S1 Table. Top five models in the confidence set exploring the main effects of input variables on $\boldsymbol{MaxI_h}$.

The confidence set had 5 models within 2 ΔAIC from the best model.

Model Order	Model	K	Log Likelihood	AIC	AIC weight	R2
1	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \boldsymbol{\beta_{v\to h}} + \mu_v + R_{tot}$	7	-990.96	1999.92	0.33	0.36
2	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \beta_{v\to h} + \mu_v + \gamma_v + R_{tot}$	8	-990.45	2000.91	0.20	0.36
3	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \boldsymbol{\beta_{v\to h}} + \mu_v + iv + R_{tot}$	7	-990.91	2001.83	0.13	0.36
4	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{v\to h} + \beta_{v\to h} + \mu_v + q + R_{tot}$	8	-990.96	2001.92	0.12	0.36
5	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \boldsymbol{\beta_{v\to h}} + \mu_v + q + R_{tot}$	8	-990.40	2001.92	0.12	0.36

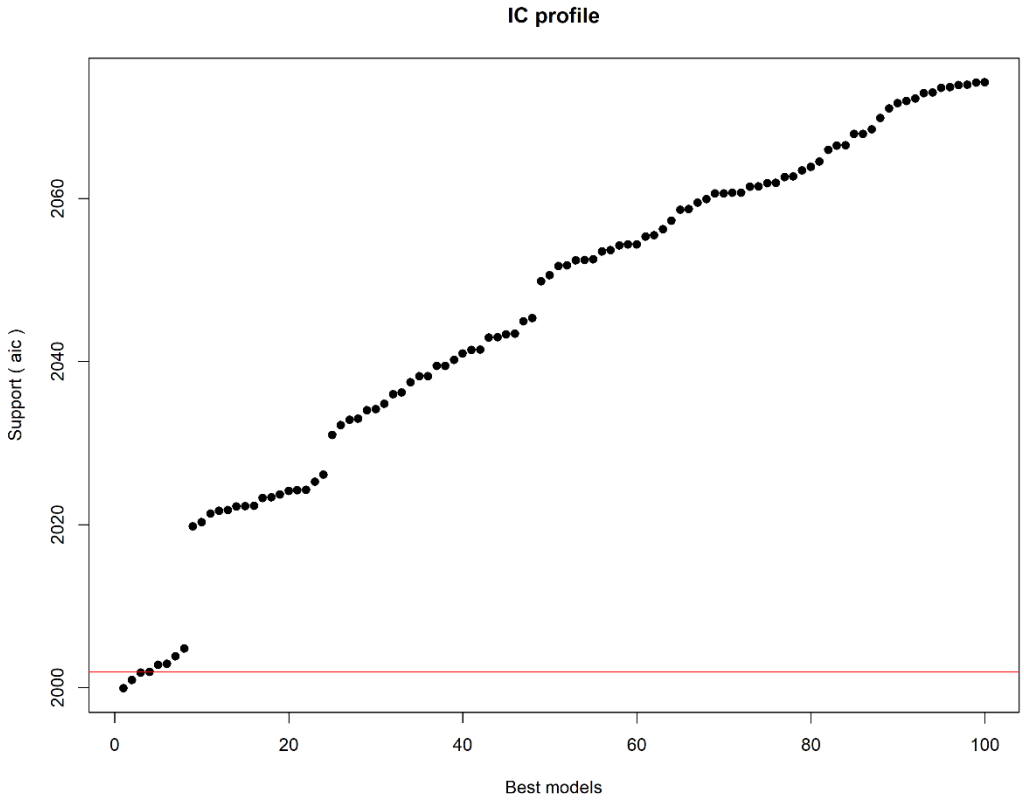

Figure 2-25 S7 Fig. The AIC profile of models exploring the main effects of input variables on $\boldsymbol{MaxI_h}$.

AIC values from the best to the worst model from the 100 returned models, horizontal line delineates models that are within 2 ΔAIC from the best model. Best AIC: 1999.92; Worst AIC: 2074.25

Table 2-15 S2 Table. Top five models in the confidence set exploring the main effects of input variables on $\boldsymbol{tmaxI_h}$.

The confidence set had 26 models within 2 ΔAIC from the best model.

Model Order	Model	K	Log Likelihood	AIC	AIC weight	R2
1	~ 1 + t_{crit} + T_{mean} + $\beta_{h\to v}$ + μ_v	4	1295.92	2603.85	0.04	0.31
2	~ 1 + t_{crit} + T_{mean} + $\beta_{h\to v}$ + μ_v+ R_{tot}	5	1294.95	2603.91	0.04	0.31
3	~ 1 + t_{crit} + T_{mean} + $\beta_{h\to v}$ + $\boldsymbol{\beta_{v\to h}}$ + μ_v	5	1295.08	2604.15	0.03	0.31
4	~ 1 + t_{crit} + T_{mean} + T_{range} + $\beta_{h\to v}$ + μ_v	5	1295.14	2604.27	0.03	0.31
5	~ 1 + t_{crit} + T_{mean} + $\beta_{h\to v}$ + $\boldsymbol{\beta_{v\to h}}$ + μ_v+ R_{tot}	6	1294.14	2604.29	0.03	0.31

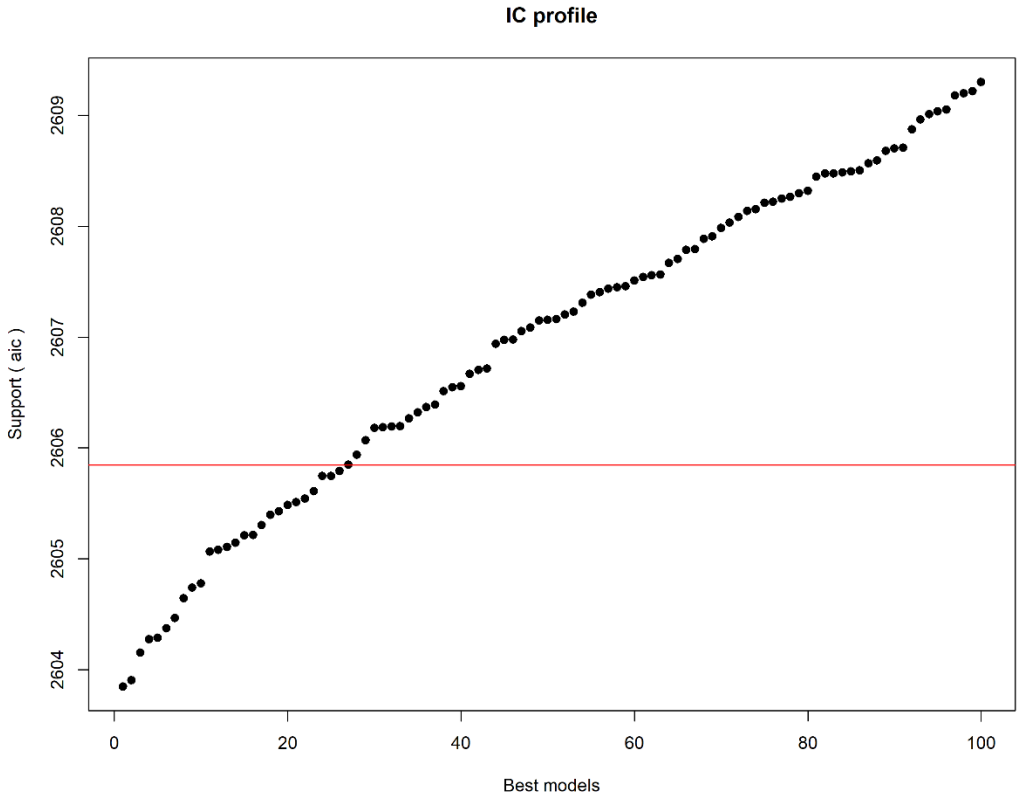

Figure 2-26 S8 Fig. The AIC profile of models exploring the main effects of input variables on $\boldsymbol{tmaxI_h}$.

AIC values from the best to the worst model from the 100 returned models, horizontal line delineates models that are within 2 ΔAIC from the best model. Best AIC: 2603.85; Worst AIC: 2609.30

Table 2-16 S3 Table. Top five models in the confidence set exploring main effects of input variables on $\boldsymbol{cumI_h}$.

The confidence set had 5 models within 2 ΔAIC from the best model.

Model Order	Model	K	Log Likelihood	AIC	AIC weight	R2
1	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \beta_{v\to h} + \mu_v + R_{tot}$	7	-1207.91	2433.81	0.36	0.46
2	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \beta_{v\to h} + \mu_v + \gamma_v + R_{tot}$	8	-1207.63	2435.25	0.17	0.46
3	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \beta_{v\to h} + \mu_v + iv + R_{tot}$	8	-1207.90	2435.81	0.13	0.46
4	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \beta_{v\to h} + \mu_v + q + R_{tot}$	8	-1207.91	2435.81	0.13	0.46
5	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \beta_{v\to h} + \mu_v + R_{tot}$	7	-1207.62	2435.81	0.13	0.46

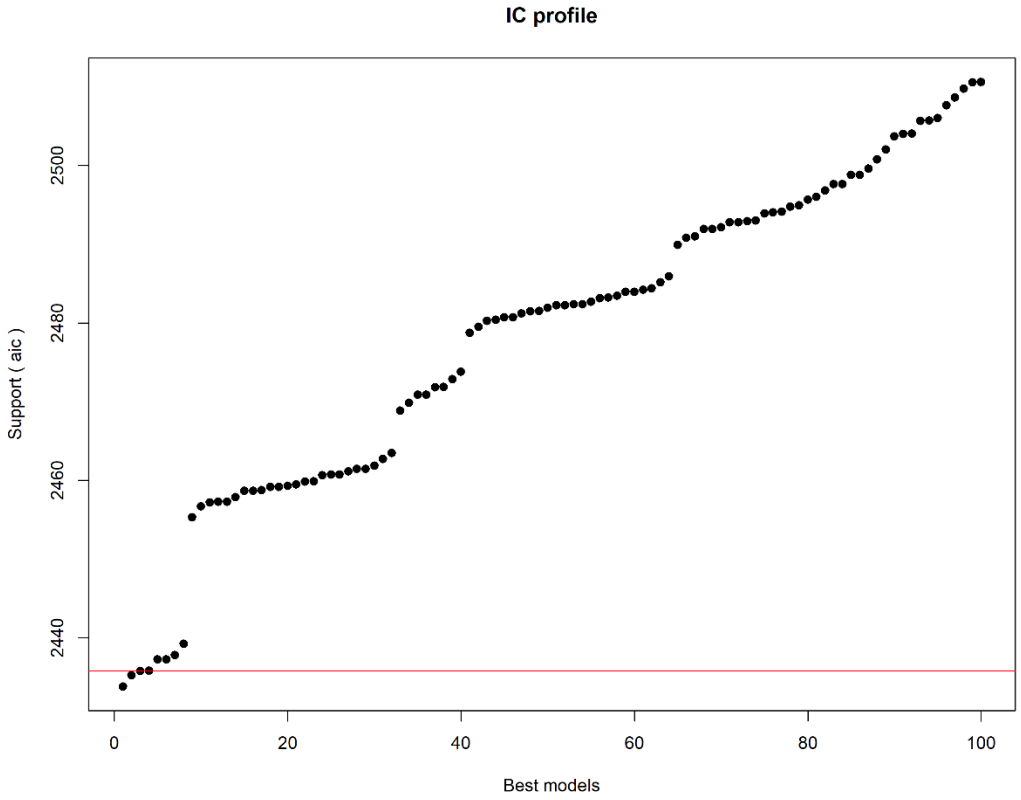

Figure 2-27 S9 Fig. The AIC profile of models exploring the main effects of input variables on $\boldsymbol{cumI_h}$.

AIC values from the best to the worst model from the 100 returned models, horizontal line delineates models that are within 2 ΔAIC from the best model. Best AIC: 2433.81; Worst AIC: 2510.59

Table 2-17 S4 Table. Top five models exploring in the confidence set the main effects and pairwise (first order) interactions of input variables on $\boldsymbol{MaxI_h}$.

The confidence set had 47 models within 2 ΔAIC from the best model.

Model Order	Model	K	Log Likelihood	AIC	AIC weight	R2
1	$\sim 1 + \beta_{h\to v} + \gamma_v + iv + t_{crit}:T_{mean} + t_{crit}:T_{range} + T_{mean}:T_{range} + t_{crit}:\beta_{v\to h} + \beta_{h\to v}:\mu_v + T_{range}:\gamma_v + \gamma_v:iv + T_{range}:q + T_{mean}:R_{tot} + \boldsymbol{\beta_{v\to h}}:R_{tot} + q:R_{tot}$	14	-958.11	1948.23	0.03	0.41
2	$\sim 1 + \beta_{h\to v} + \gamma_v + iv + t_{crit}:T_{mean} + t_{crit}:T_{range} + T_{mean}:T_{range} + t_{crit}:\beta_{v\to h} + \beta_{h\to v}:\mu_v + T_{range}:\gamma_v + \gamma_v:iv + T_{range}:q + T_{mean}:R_{tot} + \beta_{v\to h}:R_{tot} + \mu_v:R_{tot} + q:R_{tot}$	15	-957.45	1948.90	0.02	0.41
3	$\sim 1 + \beta_{h\to v} + \gamma_v + iv + t_{crit}:T_{mean} + t_{crit}:T_{range} + T_{mean}:T_{range} + t_{crit}:\beta_{v\to h} + t_{crit}:\mu_v + \beta_{h\to v}:\mu_v + T_{range}:\gamma_v + \gamma_v:iv + T_{range}:q + T_{mean}:R_{tot} + \boldsymbol{\beta_{v\to h}}:R_{tot} + q:R_{tot}$	15	-957.55	1949.11	0.02	0.41
4	$\sim 1 + \beta_{h\to v} + \gamma_v + iv + t_{crit}:T_{mean} + t_{crit}:T_{range} + T_{mean}:T_{range} + t_{crit}:\beta_{v\to h} + T_{range}:\beta_{h\to v} + \beta_{h\to v}:\mu_v + T_{range}:\gamma_v + \gamma_v:iv + T_{range}:q + T_{mean}:R_{tot} + \beta_{v\to h}:R_{tot} + q:R_{tot}$	15	-957.57	1949.13	0.02	0.41
5	$\sim 1 + \beta_{h\to v} + \gamma_v + iv + t_{crit}:T_{mean} + t_{crit}:T_{range} + T_{mean}:T_{range} + t_{crit}:\beta_{v\to h} + t_{crit}:\boldsymbol{\beta_{v\to h}} + \beta_{h\to v}:\mu_v + T_{range}:\gamma_v + \gamma_v:iv + T_{range}:q + T_{mean}:R_{tot} + T_{range}:R_{tot} + \beta_{v\to h}:R_{tot} + q:R_{tot}$	16	-957.68	1949.36	0.02	0.41

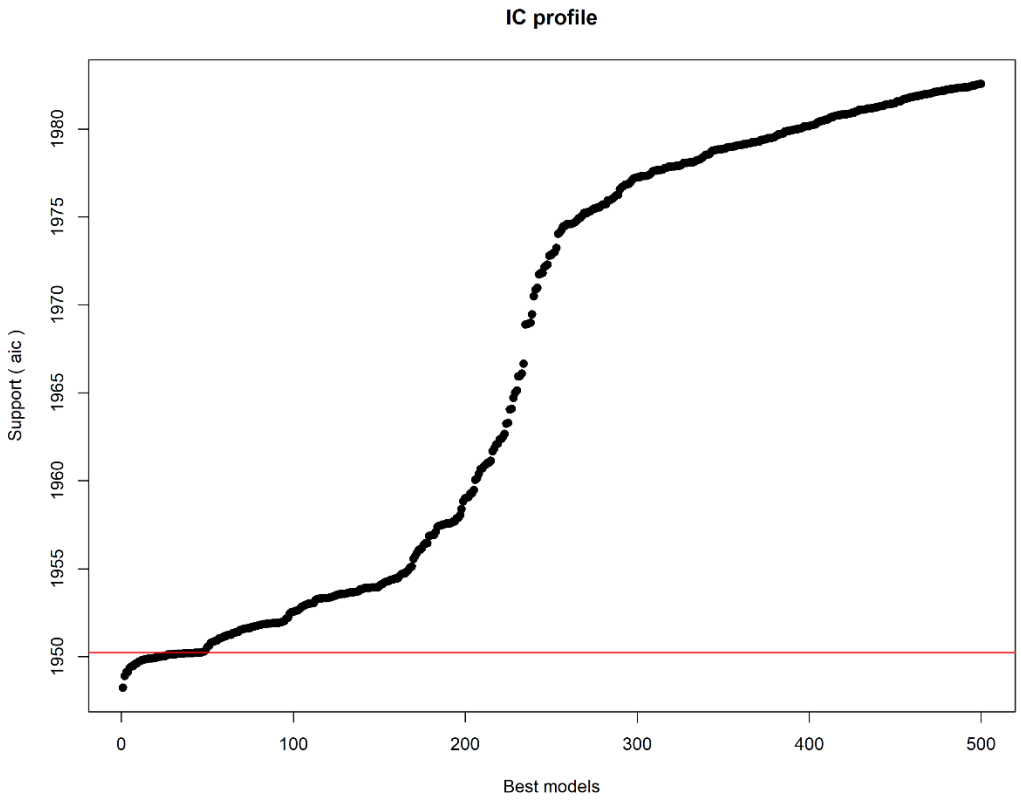

Figure 2-28 S10 Fig. The AIC profile of models exploring the main effects of input variables on $\boldsymbol{MaxI_h}$.

AIC values from the best to the worst model from the 500 returned models, horizontal line delineates models that are within 2 ΔAIC from the best model. Best AIC: 1948.23; Worst AIC: 1982.56

Table 2-18 S5 Table. Top five models in the confidence set exploring the main effects and pairwise (first order) interactions of input variables on $\boldsymbol{tmaxI_h}$.

The confidence set had 128 models within 2 ΔAIC from the best model.

Model Order	Model	K	Log Likelihood	AIC	AIC weight	R2
1	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \beta_{v\to h} + \mu_v + R_{tot} + t_{crit}:T_{mean} + t_{crit}:T_{range} + T_{mean}:T_{range} + T_{mean}:\beta_{h\to v} + T_{range}:\beta_{h\to v} + T_{mean}:\beta_{v\to h} + \beta_{h\to v}:\beta_{v\to h} + \beta_{v\to h}:\mu_v + t_{crit}:\gamma_v + \beta_{h\to v}:\gamma_v + \beta_{v\to h}:\gamma_v + \mu_v:\gamma_v + t_{crit}:iv + T_{range}:iv + \beta_{v\to h}:iv + t_{crit}:q + T_{mean}:q + iv:q + T_{mean}:R_{tot} + T_{range}:R_{tot} + \beta_{h\to v}:R_{tot} + \mu_v:R_{tot}$	29	-1243.69	2547.38	0.01	0.40
2	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \beta_{v\to h} + \mu_v + R_{tot} + t_{crit}:T_{mean} + t_{crit}:T_{range} + T_{mean}:T_{range} + T_{mean}:\beta_{h\to v} + T_{range}:\beta_{h\to v} + T_{mean}:\beta_{v\to h} + \beta_{h\to v}:\beta_{v\to h} + \beta_{v\to h}:\mu_v + t_{crit}:\gamma_v + \beta_{h\to v}:\gamma_v + \beta_{v\to h}:\gamma_v + \mu_v:\gamma_v + t_{crit}:iv + T_{range}:iv + \beta_{v\to h}:iv + t_{crit}:q + T_{mean}:q + iv:q + T_{range}:R_{tot} + \beta_{h\to v}:R_{tot} + \mu_v:R_{tot}$	28	-1244.74	2547.47	0.01	0.40
3	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \beta_{v\to h} + \mu_v + R_{tot} + t_{crit}:T_{mean} + t_{crit}:T_{range} + T_{mean}:T_{range} + T_{mean}:\beta_{h\to v} + T_{range}:\beta_{h\to v} + T_{mean}:\beta_{v\to h} + \beta_{h\to v}:\beta_{v\to h} + \beta_{v\to h}:\mu_v + t_{crit}:\gamma_v + \beta_{h\to v}:\gamma_v + \beta_{v\to h}:\gamma_v + \mu_v:\gamma_v + t_{crit}:iv + T_{range}:iv + \beta_{v\to h}:iv + t_{crit}:q + T_{mean}:q + iv:q + T_{mean}:R_{tot} + T_{range}:R_{tot} + \beta_{h\to v}:R_{tot} + \mu_v:R_{tot}$	29	-1244.75	2547.48	0.01	0.40

4	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \beta_{v\to h} + \mu_v + R_{tot} + t_{crit}:T_{mean} + t_{crit}:T_{range} + T_{mean}:T_{range} + T_{mean}:\beta_{h\to v} + T_{range}:\beta_{h\to v} + T_{mean}:\beta_{v\to h} + \beta_{h\to v}:\beta_{v\to h} + \beta_{v\to h}:\mu_v + t_{crit}:\gamma_v + \beta_{h\to v}:\gamma_v + \beta_{v\to h}:\gamma_v + \mu_v:\gamma_v + t_{crit}:iv + T_{range}:iv + \beta_{v\to h}:iv + t_{crit}:q + T_{mean}:q + iv:q + T_{mean}:R_{tot} + T_{range}:R_{tot} + \beta_{h\to v}:R_{tot} + \mu_v:R_{tot}$	29	-1244.77	2547.53	0.01	0.40
5	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \beta_{v\to h} + \mu_v + R_{tot} + t_{crit}:T_{mean} + t_{crit}:T_{range} + T_{mean}:T_{range} + T_{mean}:\beta_{h\to v} + T_{range}:\beta_{h\to v} + T_{mean}:\beta_{v\to h} + \beta_{h\to v}:\beta_{v\to h} + \beta_{v\to h}:\mu_v + t_{crit}:\gamma_v + \beta_{h\to v}:\gamma_v + \beta_{v\to h}:\gamma_v + \mu_v:\gamma_v + t_{crit}:iv + T_{range}:iv + \beta_{v\to h}:iv + t_{crit}:q + T_{mean}:q + iv:q + T_{mean}:R_{tot} + T_{range}:R_{tot} + \beta_{h\to v}:R_{tot} + \mu_v:R_{tot}$	29	-1244.79	2547.58	0.01	0.40

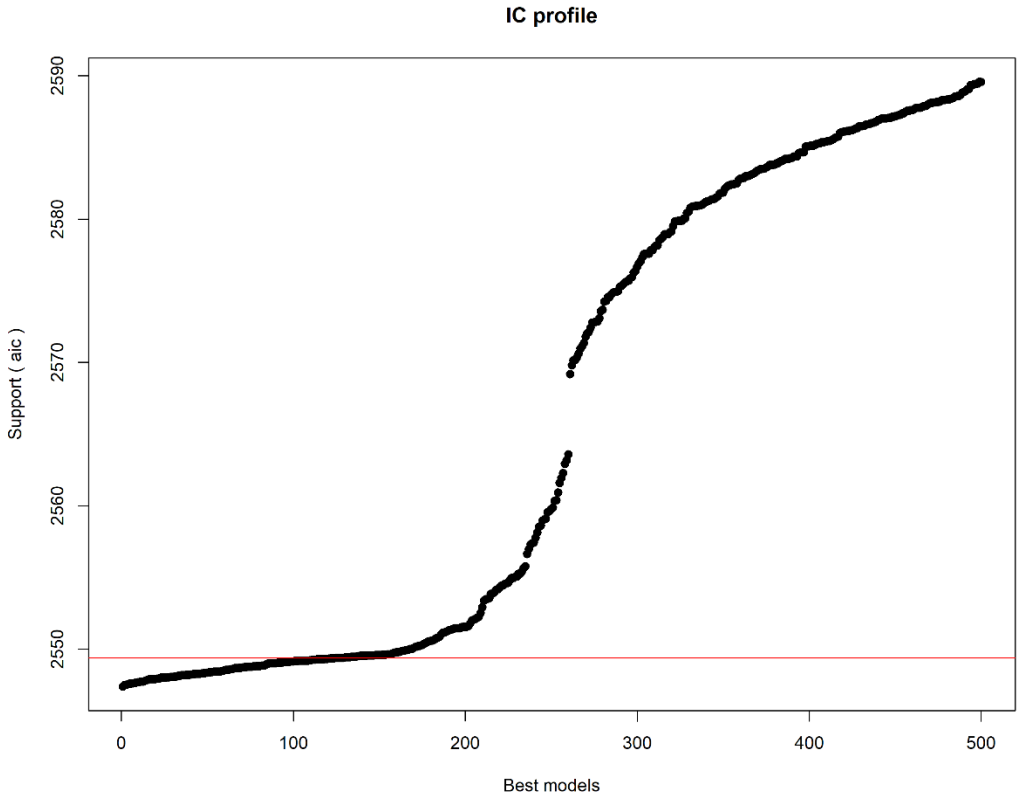

Figure 2-29 S11 Fig. The AIC profile of models exploring main effects and pairwise (first order) interactions of input variables on $\boldsymbol{tmaxI_h}$.

AIC values from the best to the worst model from the 500 returned models, horizontal line delineates models that are within 2 ΔAIC from the best model. Best AIC: 2547.38; Worst IC: 2589.56

Table 2-19 S6 Table. Top five models in the confidence set exploring the main effects and pairwise (first order) interactions of input variables on $\boldsymbol{cumI_h}$.

The confidence set had 84 models within 2 ΔAIC from the best model.

Model Order	Model	K	Log Likelihood	AIC	AIC weight	R2
1	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \beta_{v\to h} + \mu_v + R_{tot} + t_{crit}:T_{mean} + t_{crit}:T_{range} + T_{mean}:T_{range} + t_{crit}:\beta_{v\to h} + t_{crit}:\mu_v + \beta_{h\to v}:\mu_v + t_{crit}:\gamma_v + T_{range}:\gamma_v + \gamma_v:iv + t_{crit}:q + iv:q + T_{mean}:R_{tot}$	19	-1159.61	2361.21	0.02	0.53
2	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \beta_{v\to h} + \mu_v + R_{tot} + t_{crit}:T_{mean} + t_{crit}:T_{range} + T_{mean}:T_{range} + T_{range}:\beta_{v\to h} + t_{crit}:\beta_{v\to h} + t_{crit}:\mu_v + \beta_{h\to v}:\mu_v + t_{crit}:\gamma_v + T_{range}:\gamma_v + \gamma_v:iv + t_{crit}:q + iv:q + T_{mean}:R_{tot}$	20	-1158.63	2361.26	0.02	0.53
3	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \beta_{v\to h} + \mu_v + R_{tot} + t_{crit}:T_{mean} + t_{crit}:T_{range} + T_{mean}:T_{range} + t_{crit}:\beta_{v\to h} + t_{crit}:\mu_v + \beta_{h\to v}:\mu_v + t_{crit}:\gamma_v + T_{range}:\gamma_v + \beta_{h\to v}:iv + \gamma_v:iv + t_{crit}:q + iv:q + T_{mean}:R_{tot}$	20	-1158.67	2361.34	0.02	0.53

4	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \beta_{v\to h} + \mu_v + R_{tot} + t_{crit}{:}T_{mean} + t_{crit}{:}T_{range} + T_{mean}{:}T_{range} + T_{range}{:}\beta_{h\to v} + t_{crit}{:}\beta_{v\to h} + t_{crit}{:}\mu_v + \beta_{h\to v}{:}\mu_v + t_{crit}{:}\gamma_v + T_{range}{:}\gamma_v + \gamma_v{:}iv + t_{crit}{:}q + iv{:}q + T_{mean}{:}R_{tot}$	20	-1157.70	2361.41	0.01	0.53
5	$\sim 1 + t_{crit} + T_{mean} + T_{range} + \beta_{h\to v} + \beta_{v\to h} + \mu_v + R_{tot} + t_{crit}{:}T_{mean} + t_{crit}{:}T_{range} + T_{mean}{:}T_{range} + T_{range}{:}\beta_{h\to v} + t_{crit}{:}\beta_{v\to h} + t_{crit}{:}\mu_v + \beta_{h\to v}{:}\mu_v + t_{crit}{:}\gamma_v + T_{range}{:}\gamma_v + \gamma_v{:}iv + t_{crit}{:}q + iv{:}q + T_{mean}{:}R_{tot}$	20	-1159.75	2361.51	0.01	0.53

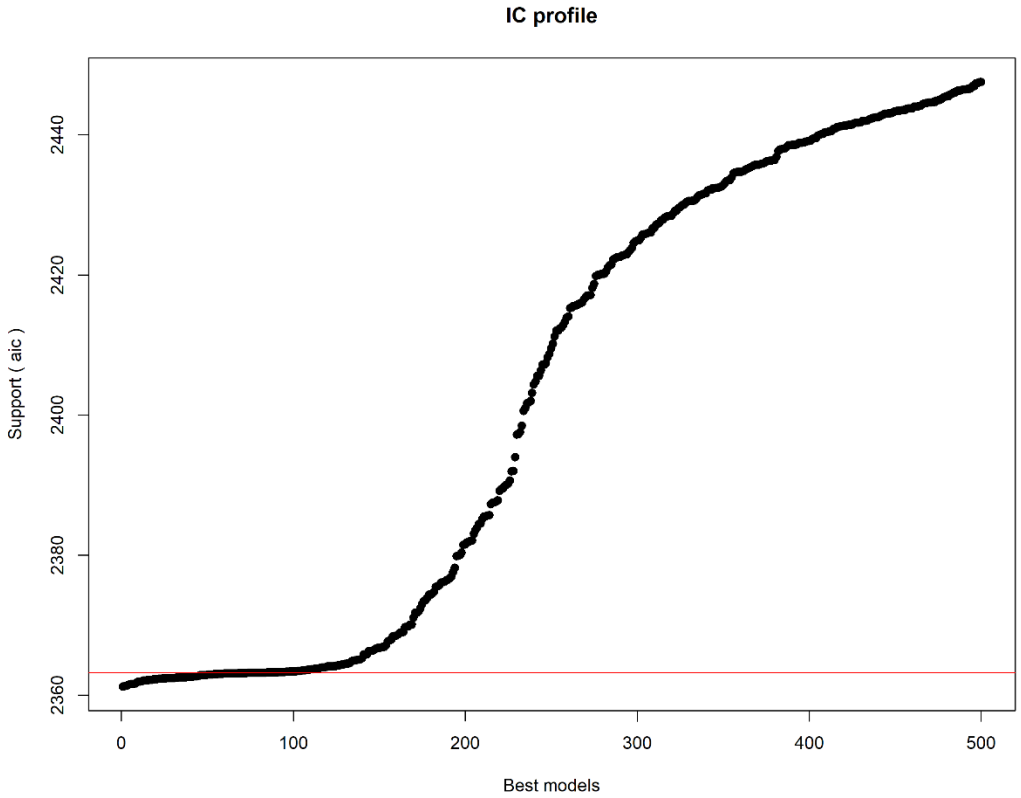

Figure 2-30 S12 Fig. The AIC profile of models exploring main effects and pairwise (first order) interactions of input variables on $\boldsymbol{cumI_h}$.

AIC values from the best to the worst model from the 500 returned models, horizontal line delineates models that are within 2 ΔAIC from the best model. Best AIC: 2361.21; Worst AIC: 2447.52

Table 2-20 S7 Table. Model averaged coefficients and relative-importance weights for dengue parameters.

From the best model with the pairwise interactions. Before model fitting, parameters were centered on zero and scaled to unit variance, to normalize parameters within the same range. Hence coefficient estimates are within the range of 0.0 to 1.0. Relative importance ≥0.90 appears in bold, an arbitrary cut-off for visualizing most important parameters.

Parameter	$MaxI_h$		$tmaxI_h$		$cumI_h$	
	Coefficient[1]	Relative Importance	Coefficient[1]	Relative Importance	Coefficient[1]	Relative Importance
t_{crit}	0.00	0.02	0.43	**1.00**	0.36	**1.00**
t_{crit}:T_{mean}	0.38	**1.00**	0.15	**1.00**	0.10	**1.00**
t_{crit}:T_{range}	-0.14	**0.96**	0.00	0.04	-0.07	**1.00**
t_{crit}:$\beta_{h\to v}$	0.00	0.03	0.00	0.01	0.00	0.02
t_{crit}:$\boldsymbol{\beta_{v\to h}}$	0.16	**0.98**	0.00	0.01	0.07	**1.00**
t_{crit}:μ_v	0.00	0.02	0.00	0.08	-0.07	**1.00**
t_{crit}:γ_v	0.00	0.02	0.05	**0.99**	0.03	0.85
t_{crit}:iv	0.00	0.01	-0.02	0.54	0.00	0.03
t_{crit}:q	0.00	0.02	0.07	**1.00**	0.04	**0.97**
t_{crit}:R_{tot}	0.00	0.02	0.00	0.02	0.00	0.02
T_{mean}	0.00	0.02	0.15	**1.00**	0.15	**1.00**
T_{mean}:T_{range}	-1.04	**1.00**	0.08	**1.00**	-0.10	**1.00**
T_{mean}:$\beta_{h\to v}$	0.00	0.02	0.07	**1.00**	0.00	0.06
T_{mean}:$\boldsymbol{\beta_{v\to h}}$	0.00	0.02	-0.05	**0.98**	0.00	0.02
T_{mean}:μ_v	0.00	0.02	0.00	0.01	0.00	0.02
T_{mean}:γ_v	0.00	0.02	0.00	0.01	0.00	0.01
T_{mean}:iv	0.00	0.02	0.00	0.02	0.00	0.02
T_{mean}:q	0.00	0.02	0.06	**0.99**	0.00	0.02
T_{mean}:R_{tot}	1.11	**1.00**	0.04	**0.94**	0.08	**1.00**
T_{range}	0.01	0.02	-0.04	**0.96**	-0.12	**1.00**

T_{range}: $\beta_{h\to v}$	-0.01	0.06	0.07	**0.99**	-0.02	0.48
T_{range}: $\boldsymbol{\beta_{v\to h}}$	0.00	0.02	0.00	0.03	0.00	0.02
T_{range}: μ_v	0.00	0.02	0.00	0.04	0.00	0.02
T_{range}: γ_v	-0.49	0.99	0.00	0.08	-0.04	**0.96**
T_{range}: iv	0.00	0.01	0.07	**1.00**	0.00	0.03
T_{range}: $\boldsymbol{q}$	1.45	**1.00**	0.00	0.01	0.00	0.03
T_{range}: R_{tot}	0.00	0.03	0.04	**0.95**	0.00	0.02
$\beta_{h\to v}$	1.15	**0.99**	0.08	**1.00**	0.23	**1.00**
$\beta_{h\to v}$: $\boldsymbol{\beta_{v\to h}}$	0.00	0.01	-0.05	**0.99**	0.00	0.01
$\beta_{h\to v}$: μ_v	-0.66	**1.00**	0.00	0.04	-0.05	**0.98**
$\beta_{h\to v}$: γ_v	0.00	0.01	-0.06	**1.00**	0.00	0.01
$\beta_{h\to v}$: iv	0.00	0.02	0.00	0.02	0.01	0.23
$\beta_{h\to v}$: $\boldsymbol{q}$	0.01	0.02	0.00	0.02	0.00	0.08
$\beta_{h\to v}$: R_{tot}	0.00	0.02	-0.05	**0.98**	0.00	0.03
$\boldsymbol{\beta_{v\to h}}$	0.00	0.02	-0.06	**0.99**	0.12	**1.00**
$\boldsymbol{\beta_{v\to h}}$: μ_v	0.02	0.15	0.05	**0.98**	0.00	0.03
$\boldsymbol{\beta_{v\to h}}$: γ_v	0.00	0.02	-0.05	**0.98**	0.00	0.05
$\boldsymbol{\beta_{v\to h}}$: iv	0.00	0.03	0.01	0.36	0.00	0.03
$\boldsymbol{\beta_{v\to h}}$: $\boldsymbol{q}$	0.00	0.02	0.00	0.01	0.00	0.02
$\boldsymbol{\beta_{v\to h}}$: R_{tot}	0.16	0.85	0.00	0.03	0.00	0.02
μ_v	0.00	0.02	-0.06	**0.99**	-0.25	**1.00**
μ_v: γ_v	0.00	0.03	0.03	0.81	0.00	0.03
μ_v: iv	0.00	0.02	0.00	0.07	0.00	0.03
μ_v: $\boldsymbol{q}$	0.00	0.03	0.00	0.02	0.00	0.01
μ_v: R_{tot}	-0.06	0.22	0.05	**0.97**	0.00	0.02
γ_v	1.27	**1.00**	0.00	0.02	0.00	0.04
γ_v: iv	-0.72	**0.98**	0.00	0.02	-0.04	**0.95**

γ_v: $\boldsymbol{q}$	0.01	0.02	0.00	0.02	0.00	0.05
γ_v: R_{tot}	0.01	0.03	0.00	0.01	0.00	0.05
$\boldsymbol{iv}$	0.53	0.65	0.00	0.02	0.00	0.01
$\boldsymbol{iv}$: $\boldsymbol{q}$	0.15	0.37	-0.05	**0.99**	-0.05	**0.98**
$\boldsymbol{iv}$: R_{tot}	0.01	0.03	0.00	0.02	0.00	0.03
$\boldsymbol{q}$	-0.21	0.16	0.01	0.21	0.00	0.02
$\boldsymbol{q}$: R_{tot}	-0.79	**0.99**	0.00	0.01	0.00	0.02
R_{tot}	0.02	0.02	0.04	**0.96**	0.13	**1.00**

References

1. Marino S, Hogue IB, Ray CJ, Kirschner DE. A methodology for performing global uncertainty and sensitivity analysis in systems biology. J Theor Biol. 2008;254(1):178-96. doi: https://doi.org/10.1016/j.jtbi.2008.04.011.

2. Wu J, Dhingra R, Gambhir M, Remais Justin V. Sensitivity analysis of infectious disease models: methods, advances, and their application. J R Soc Interface. 2013;10(86):20121018. doi: https://doi.org/10.1098/rsif.2012.1018.

3. Hamby DM. A review of techniques for parameter sensitivity analysis of environmental models. Environ Monit Assess. 1994;32(2):135-54. doi: https://doi.org/10.1007/bf00547132.

4. Burnham KP, Anderson DR. Model selection and multimodel inference: a practical information-theoretic approach: Springer, New York, NY; 2002.

5. Calcagno V, de Mazancourt C. Glmulti: An R package for easy automated model selection with (generalized) linear models. 2010. 2010;34(12): J Stat Softw. doi: https://doi.org/10.18637/jss.v034.i12.

3 General discussion

This general discussion summarizes the content and the main findings of this thesis, discusses the key contributions, and integrates finalizes with a concise conclusion. The section is outlined according to the overall modeling framework applied i.e., the importation models and the transmission model and a potential real-world application of both frameworks.

3.1 Discussion

Understanding the dispersal patterns of infectious diseases is increasingly important in the current era of fast-paced globalization. To do this, it is crucial to understand how diseases are transported, introduced, established, and spread into new geographical regions. This spread is very often initiated by human mobility; hence the role of the global transport network in facilitating importation is an integral focus in the epidemiology of infectious diseases. This thesis contextualizes this, in relation to the importation and spread of dengue fever in Europe. This thesis focused on the transport and introduction stages of viraemic cases of dengue, as these remain less studied than the establishment and transmission stages. Sections 2.1 and 2.2. of this thesis, employ two specific modelling approaches, (i.e., a connectivity model and a predictive model) towards the understanding and prediction of dengue importation into Europe. The connectivity model investigates the factors mediating the risk of importation from the source country into Europe, while the predictive model forecasts the importation risk of a viraemic case of dengue.

The model outlined in section 2.1, integrated a refined connectivity network approach, drawing on the underlining concept of spatial interaction modelling, generalized linear mixed modelling, temporal network autocorrelation modelling, and consideration of factors that may affect the likelihood of dengue transport and importation (1-4). The choice to integrate these different approaches, the detail of each modelling approach, and the main benefits/contributions are discussed here.

Firstly, as dengue importation are initiated by human movement, which is supported by transportation infrastructures, this implies that a source country must be linked to a destination country for a successful transferability (5). The quantification of this transferability is modelled under the basic assumption of spatial interaction modelling, that inflow between two locations is a function of the attributes of the source and destination, the friction of distance, and corresponding interactions (6). Section 2.1 details the development of seven simple connectivity indices specific to the transport and introduction of viraemic cases of dengue between any possible location pairs (with countries in Europe). These indices were used to characterize the specific attributes of a source country. These attributes are epidemiologic, socio-economic, and anthropogenic – characterizing the role of the dengue activity, seasonality, incidence rate, epidemic vulnerability, geographical proximity, air passenger traffic, and wealth of a source country. This modelling

structure is an improvement over standard epidemiological models, that establish only correlations between imported dengue cases and total inflow passengers. These previous methods were generally not sufficient (7, 8), because statistical descriptors of passengers' data alone, do not indicate whether imports are arriving from regions at high risk of dengue transmission or if they are correlated with other factors facilitating imports. The spatial interaction modelling approach improves the simple statistical approach and offers a more refined quantification of the attributes between each location pair and their corresponding flow. As discussed in section 2.1, each of the developed indices represents a spatial interaction, as an analogy for human mobility and dengue importation, between a discrete source/destination pair.

The next modelling approach used in the importation model is the generalized linear mixed model (9). The first analysis considered which combination of the connectivity indices was best at explaining the importation of dengue into Europe. To this, a generalized linear mixed-effects model (GLMM) with logit link functions for binomial errors and fixed effects of the connectivity indices were fitted. The choice of a GLMM method is justified by the particularity of the connectivity indices having more than one source of random variability, as the connectivity data were drawn from a hierarchy of different spatial and temporal groupings. The random effects allowed by the GLMM add to the variance component, which accounts for the random variability that relates to the hierarchal differences across the country pairs. The GLMM also allowed the estimation of the fixed effects of variation in the connectivity indices between country pairs. To this effect, the primary outputs of the first analysis in section 2.1, were quantitative parameter estimates describing how dengue importation into Europe changes as a function of the connectivity indices, and the variability among the levels of the random effects as represented in the country pairs.

Finally, the model in section 2.1 incorporated a temporal network autocorrelation to account for the co-dynamics of the air transport network structure in facilitating dengue importation (10). It is important to consider the air transport network in the understanding of the importation patterns because this network is dynamic. The dynamic nature of the air transport network has a direct influence on the interactions and flow between country pairs. To this effect, the air transport network was constructed using direct and indirect route-based data of air passengers between country pairs. As a measure of the network connections, centrality measures were calculated for

each country pair. The effect of the network autocorrelation was then incorporated into the GLMM framework (described above) to address the issue of covariance driven by the network structure. This alternative approach allowed for testing how individual connectivity indices change in the context of the dynamic and multiple temporal snapshots of the network. Offering insights into how the connectivity indices, network positioning, and the interaction of countries within the network can best explain the transport and introduction of dengue into a destination country.

Overall, the analyses show that the connectivity indices and centrality measures of the air transport network are strong indicators of dengue importation in Europe. This result is particularly valuable as it substantially adds to the understanding of the dynamics of dengue importation, by relating source countries' attributes and the air transport network structure. The outputs (i.e., the connectivity indices & network centrality measures), were then used to parameterize a predictive model.

Section 2.2 outlines the development of a predictive model for the importation risk of a viraemic case of dengue. Predictive modelling encompasses a variety of mathematical and/or statistical techniques to analyse current and historical data to make forecasts about future or otherwise unknown events (11). This thesis integrated the use of machine learning algorithms – an applied extension of artificial intelligence – for the development of the predictive model. Machine learning algorithms are built on a mathematical model base, that can automatically learn patterns in the data, and perform inference, without explicit instructions (12). Several studies have demonstrated the powerful predictive capabilities of machine learning models and their superiority over conventional statistical methods for prediction (13-15). The primary interest of our model was to optimize predictive accuracy for the importation risk of a viraemic case of dengue, with a secondary interest in providing intuitive explanations for the model's predictions.

The modelling approach compared four widely used classifiers algorithms in machine learning. In a supervised learning technique, the predictor variables (in our case, the connectivity indices and centrality measures) and an outcome of interest (imported case of a viraemic case of dengue) are already known. The machine-learning algorithm then uses a subset of the supplied data to learn the mapping function from the input to the output, allowing it to predict expected outcomes in new datasets. The choice of the final optimal model was based on, a meta-algorithm that fits our classification problem, built-in feature selection, regularization (ability to handle the

effects of multicollinearity), hyperparameter optimization (model tuning capabilities), and efficient computation time. The predictive performance was evaluated using the area under the receiving operating characteristic curve of the receiving operating characteristic curve (AUC-ROC) threshold, which maximizes the trade-off between sensitivity and specificity. The best-performing model was the extreme gradient boosting model with an AUC score of 0.94, a true positive rate of 0.88, and a false positive rate of 0.12. The final model was able to correctly predict the probability of an imported case of dengue with an 88% accuracy rate. As a first attempt to apply a machine learning model for predicting dengue importation in Europe, we can safely state that this model provides a benchmark result for predictive performance.

Beyond the predictive accuracy of the final model, we explain the rationale behind the predictions, using intuitive model-agnostic techniques. Model-agnostic methods work by extracting post-hoc explanations from an original machine learning model (16), by training an interpretable model on the predictions of the original model, and/or by changing the inputs of the original model and measure the changes in the prediction output (17, 18). The interpretability framework employed provided both global and local scale explanation to our model predictions. Global interpretability provided an understanding of the general pattern of the relationship and distribution of the predicted outcome (such as key features facilitating dengue importation in Europe), while local interpretability provided an in-depth understanding of a single predicted instance (such as why was there an imported case for a specific source to destination country pair). The former was important to capture the relative contribution of the input variables in predicting the importation of a viraemic case of dengue at an aggregated level for Europe; while the latter provided insights into the heterogeneities that affect the prediction of dengue importation at the country level.

The global interpretability utilized both variable importance and partial dependence plots to quantify the relationship between our model variables and the predicted outcome (17, 19). Generally, the model implies that on average the probability of an imported case of dengue increases for source countries with higher dengue incidence rates, larger population, and higher air passenger volume. Likewise, the propensity of a destination country to receive an imported case increases for countries with the following attributes: (i) high connectivity within the air travel network, (ii) putative connection hubs to other countries, and (iii) have a relatively short

connection time with other countries. The local interpretability utilized local surrogate models, otherwise called- Local interpretable model-agnostic explanations (LIME) – to provide an intuitive explanation of a single prediction (16, 18). The single unit of analysis in the model was a source-destination country pair, at a monthly timescale. Using the local interpretable model-agnostic explanations, the model explained the influences on the predictions at a single temporal and country-pair level. In the real-world risk, probabilities will differ based on the dynamics of attributes and the air travel network topology from one source-destination country pair to another. The local explanations provide insights into the heterogeneities of importation risk of a viraemic case of dengue at the different country pairing levels and how they change over time. This is particularly useful in profiling the importation risk of viraemic cases from a specific source country or region to guide targeted surveillance and public health preparedness for destination countries.

Section 2.3 of this thesis went further to examine the modelling of dengue transmission in a composite space-time domain under conditions of the vector presence, biological and environmental variability. This was achieved using a standard deterministic compartmental vector-host transmission model for dengue. The two components of this model were specified as follows: (i.) the vector (*Aedes aegypti* mosquitoes)- SEI and (ii) the human component - SEIR.

The general formulation of the SEI component, describes the vector population, at any given time during the transmission cycle. Adult female mosquitoes are classified into one of three classes, i.e. susceptible (S), exposed (E), and infectious (I). The model allows for the mosquitoes to bite, transmit the virus, and remain infected until they die. The human component of the transmission cycle is represented as susceptible (S), exposed (infected but not infectious) (E), infectious (I), and recovered (immune) (R). The model allows for individual humans to be in one of four classes, at any given time in the transmission cycle and did not discriminate between the symptomatic or asymptomatic infectious individuals (22).

The model was governed by state variables and parameters, which were satisfied by differential equations. Parameter values were retrieved from peer-reviewed literature on empirical studies or laboratory trials and expert knowledge. The parameter range used to set the characteristics boundaries of the model reflects the spatiotemporal conditions in Funchal, Madeira Island, as best possible. Adult female mosquitoes were introduced into the system using a fitted Gaussian curve, mimicking the observed seasonal pattern of mosquitoes on the island from the

entomological data for the years 2012 – 2019 (23, 24). Seasonality was incorporated into the compartmental modelling framework by a sinusoidal curve, with a period of 365 days. This was fitted using a decadal historical daily mean temperature in Funchal (25, 26). Mechanistic thermal response curves were then fitted for some components of life-history traits of the mosquitoes (i.e. biting rate, extrinsic incubation, and mortality). The initial starting conditions of the model assume a homogeneously mixed population, with all infectious classes set to zero, and an infection was triggered by the arrival of an infectious human on a specified day. The constants for the intrinsic incubation and infectious periods were set to reflect an introduction of a different serotype other than DENV-1 (i.e., the circulating serotype of the 2012/13 dengue outbreak in Madeira) (27). The model simulation was used to explore epidemiologically relevant outcomes of epidemic size, peak incidence, and time to peak. These outcomes were used to examine how the epidemic dynamics varied as a function of arrival dates and seasonality. Finally, the influence of the input parameters on the outcomes was characterized using a variance-based global sensitivity analysis.

The model results demonstrate a high potential for a sporadic dengue transmission on the island, driven by the seasonal change, with the arrival date significantly affecting the distribution of the timing and peak size of the epidemic. Mid-summer (July) to early-autumn (September) arrivals tend to produce epidemics with a faster peak rate and a resulting large final epidemic size. The model results went further to demonstrate the impact of variation in seasonal temperature regimes, as the model simulated both historical and futuristic temperature regimes. Annual mean temperatures of 18ºC to 26ºC were most suitable for epidemics transmission, irrespective of the seasonal temperature variation. When seasonal temperature variation was low, annual average temperatures of 27ºC to 30ºC, produced the largest epidemics, but this range shifted to cooler annual average temperatures (15ºC to 17ºC) as seasonal temperature variation increased. Beyond the arrival date of an infected human and temperature seasonality, the sensitivity analysis characterized the variability of the epidemic dynamics to other parameters. Epidemic dynamics were also sensitive to transmission rates, mortality rate, and mosquito population size, highlighting the importance of the mosquito life-history traits on epidemic outcomes.

A key aspect of the model results is the interaction between mean annual temperature and temperature range and its effect on epidemic suitability. These findings can be extrapolated to other regions on the island. With rising mean annual temperatures and increasing seasonal

temperature variation due to climate change (28, 29), these results provide insights into the likely shift in the epidemic suitability on the island. Importantly, areas on the island (i.e., the North coast regions) with previously limited or no epidemic suitability may have an increased potential for transmission in the future. On the other hand, the South Coast region with current suitability may have an increased potential for year-round transmission and higher epidemic rates.

Other key public health-related findings from the model sensitivity analysis are the insight it provides into the relative importance of parameters and their interactive effects as mechanistic drivers of an epidemic. This serves as a guide for targeted control and mitigation strategies for future outbreaks on the island. For example, reducing the mosquito population size via adulticides and larvicides will be effective as both preventive and control measures in the event of an outbreak. Given the knowledge of the most sensitivity parameters, a more effective preventive and mitigative strategy can be integrated acting over multiple parameters simultaneously.

Overall, the modelling framework presented in this thesis fits into one of the principal strategic approaches proposed by the World Health Organization (WHO) European regional framework for surveillance for re-emerging vector-borne diseases (30). The strategy calls for the development of an integrated system of vector and disease surveillance. A system with the ability to combine data sources from entomological surveillance, epidemiological surveillance, and risk factor identification, to guide effective preventive measures and epidemic response. The importation model presented in this thesis, integrates epidemiological, socio-economical, and anthropogenic data, to analyse and predict the importation risk of a viraemic case of dengue. While the dengue transmission model integrates entomologic, epidemiologic, and climatic data to examine the probability of local transmission in areas with established vector presence. The combination of both modelling techniques provides an integrated surveillance system for monitoring the importation risk of viraemic cases and assess the risk of local transmission of dengue.

Although this work focuses heavily on practical aspects of infectious diseases modelling, it provides a basic structure for an early warning system to forecast and monitor the risk of dengue in Europe. Despite this possibility, a few limitations are worth noting and that could be targeted for improvement in future implementations:

(i) The dengue incidence data used in the importation model was aggregated at a yearly scale, due to the paucity of surveillance data at a finer scale. This may have overestimated or underestimated the actual effect of incidence and potentially impact the predicted risk probability from a source country. Even though this was compensated for within the analysis −by the inclusion of the dengue activity and seasonality variables (serving as proxies). Future implementation of this framework may also benefit from data at a finer temporal resolution (e.g., monthly, or weekly), which could lead to improved performance and accuracy of dengue importation predictions.

(ii) The transmission model did not explicitly address the scenario of multiple serotypes of dengue co-circulating in the population. Considering the circulating serotype (i.e., DENV 1) of the 2012 outbreak in Madeira Island, subsequent exposure to a different serotype (i.e., DENV 2, 3, or 4) might lead to immune interactions and increased risk of severe disease and dengue haemorrhagic fever in secondary infections. Likewise, reinfection with the same serotype (DENV 1) might contribute to an increase in the number of primary and secondary dengue cases.

(iii) The transmission model varied only input parameters from the vector component of the transmission cycle and did not vary parameters in the human component. Likewise, the model did not account for the influence of ongoing vector control on the island on the overall transmission cycle. Additional modification to improve the functionality of both modelling techniques is discussed in the next section (i.e., section 3.2).

The results from the overall modelling framework represented in this thesis indicate that the importation and potential local transmission of dengue in Europe to be was quite multifaceted. As such, public health implications can be considered from multiple domains and scope. The public health implications will be discussed on the merits of each model (i.e., the importation or transmission model) and as a single early warning system that combines both models.

Risk assessment tool: The importation model provides insight into risk assessment, which can be translated into real-time policy decisions regarding resource allocation for destination-based surveillance. The ability of the model to predict specific country-pair or route-level risk profiling enables public health officials to conduct destination-based surveillance and assign resources to most at-risk routes. For example, attention can be focused on passengers arriving from Brazil into

Portugal, given a risk assessment that demonstrates: Brazil has a high incidence rate of dengue, presence of other conditions that support local transmission (e.g., ongoing dengue seasonality), and a high amount of expected air travel from Brazil into Portugal. This kind of risk assessment will help health authorities plan for potential spikes in travel-related cases of dengue, requiring clinical treatment. Likewise, it can help authorities estimate the potential consequences for travel-related cases in areas with vector presence, suitable environmental and climatic conditions for local transmissions. Similar risk assessment and profiling can be conducted in the instance of a major event that will result in an increase of arriving passengers (e.g., a major sporting event) to a specific destination country in Europe. Results from such risk assessment can assist public health officials in preparing for the scale-up of laboratory diagnostic assays and prepare physicians for effective case management.

Mitigation and control strategies: The transmission model provides insights into mitigation strategies for local transmission in areas with vector presence and suitable environmental and climatic conditions. From a mitigation perspective, the results of sensitivity analysis indicate parameters that a slight change in them would greatly affect the dynamics of the epidemics (in the event of an outbreak). The result highlight that even small changes in parameters like mosquito population and morality rates can have a significant effect on the distribution of the timing and peak size of the epidemic. Hence this can guide the implementation of local mitigation and control strategies, like reducing the vector density via adulticides and larvicides or reducing the mosquito lifespan via the use of lethal ovitraps. Likewise, the results of the pair-wise interactions in the sensitivity analysis can be considered as an effective guide in the formulation of integrated strategies that can change multiple parameters simultaneously. These integrated strategies will assist in preventing outbreaks of dengue and reduce the risk of transmission in areas where invasive mosquitoes have become firmly established in Europe.

Event-Based Surveillance: Considering both modelling techniques in a combined early warning system, it presents a convenient and efficient tool for investigation of the dynamics of dengue in Europe and its quantitative changes over time. This is mainly because the combined framework goes far beyond traditional disease-based surveillance and syndromic surveillance and offers the option of event-based surveillance. Event-based health surveillance systems are systems that can integrate public reports, stories, rumours, and other information about health events and

analysis the potential serious risk to public health. The importation model presented in this thesis offers that functionality, in the sense that it integrates the collection and analysis of data from sources beyond the health system, to enable the prediction of imported cases of dengue and/or simulations of potential local transmission. The model encompasses the air transportation network, socio-economic, environmental, ecological, climatic, and anthropogenic data, to offer a more robust type of surveillance system. The dynamic nature of the modelling framework (if implemented with real-time data) will allow the update of changes in dengue incidence, seasonality, and activity for source countries (as obtained from DengueMap) and their effect on the importation of viraemic cases of dengue into Europe. The country-pair level profiling of the modelling framework also allows patterns of infection/transmission in different countries to be compared and the factors that are responsible for these differences determined. This can guide public health officials in the implementation of an early warning and a rapid response to prevent a potential outbreak of dengue.

3.2 Real-world application of the modelling framework

The modelling framework developed in this thesis is expected to fit into the WARDEN early warning system. The WARDEN project (31) aims to develop an operational, real-time, forecasting model of the risk of outbreaks of dengue, chikungunya, and yellow fever for Madeira Island. Contextually the project aims to supplement the existing surveillance system with an early warning system providing daily forecasts of outbreak risk for dengue, chikungunya, and yellow fever, up to 16 days in advance. This will be achieved using a dynamic spatial-temporal and compartmental SEI-SEIR modelling framework. The early warning system will result from the integration of two subsystems. The first subsystem corresponds to a data assimilation module, which will collect data concerning the foreseeable status of variables (e.g., meteorological forecasts, airline, and ship arrivals) influencing transportation and the introduction of these diseases for a 16-day risk prediction. The second module corresponds to the compartmental SEI-SEIR model, based on comprehensive ordinary differential equations which can provide risk forecasts of disease transmission and epidemic dynamics.

3.2.1 The dengue importation framework

The modelling framework developed in sections 2.1 and 2.2 is expected to fit into the first subsystem of the WARDEN early warning system. However, the framework will be modified slightly to capture the contextual dynamics of Madeira Island in relation to the input variables. The following data input components will be modified accordingly:

- Dengue data: change to the historical imported cases of dengue in Madeira (for all years available).
- Air passenger's data: using the IATA comprehensive air passenger travel data, reconstruct the monthly passenger flow from all countries worldwide with a final destination in Madeira (accounting for all connecting flights).
- Connectivity indices between a source and destination: reconstitute the connectivity indices to be between a source country and Madeira Island. Example - change the

geographical distance between centroids to reflect the spatial connectivity of the source country and Madeira Island.

- Centrality measures of the air transport network: using the monthly air passenger's data, reconstructed a weighted directed network, with nodes representing all countries worldwide, while edges represent the flow of passengers from a source country into Madeira Island.
- Unit of analysis: update this to a source country to destination (Madeira Island) pair, at a monthly timescale, and a binary response variable coded to indicate an imported dengue case (1) or not (0).

The modelling technique – data pre-processing and splitting, modelling algorithm, training, tuning, validation, and evaluation – will remain the same. The model interpretability (model-agnostic) technique will remain unchanged, however, should be implemented independently from the model training component. This is necessary to ensure that when the base model inputs are updated and refined, the model-agnostic algorithms would not be affected. The ensemble of the prediction and the post-hoc explanations would first be based on a hold-out dataset (i.e., the test dataset from historical data). Subsequently, this subsystem will be linked to real-time/forecasted data to provide futuristic importation risk prediction (i.e., of viraemic cases of dengue). To achieve this, the system needs to be linked to the following real-time data source:

- DengueMap: a unified data collection tool, that brings together disparate dengue reports of local or imported dengue cases from official, newspapers, and other media sources globally (20). This will enable the categorizing of real-time dengue activity and seasonality changes around the globe.
- IATA AirportIS database: a powerful market intelligence and analysis tool for airports, that aggregates global passenger traffic flows (both historical and future schedules) through a series of advanced algorithms and several unique databases to capture 100% of traffic around the world (21). This will enable the categorizing of real-time and forecasted air passenger flow in Madeira Island and the construction of the air transport network.

Note: the following input data – dengue incidence, vulnerability index, GDP, and population – will be updated as new data becomes available. The model will be constantly updated with newly available data to improve its predictive accuracy.

Given real-time information, the model can predict the importation risk of a viraemic case of dengue for each flight route (based on projected arriving passengers) and compute the corresponding influential input variables. These forecasts can be updated every 16 days (as proposed by the WARDEN framework), based on real-time data for air-passenger flow.

3.2.2 The dengue transmission framework

This transmission modelling framework (section 2.3) is expected to fit into the second subsystem of the WARDEN early warning system. The simulated model provides important insights into the input parameters required for the implementation of the SEI-SEIR module of the system, within a deterministic compartmental framework. However, it will be modified slightly to fit the proposed differential equations for risk forecasting within the system. The WARDEN system proposes three differential equations:

- Equation 1: represents the dynamics of vector populations and quantifies the average number of mosquitoes that become infected per each infectious human host in the system.
- Equation 2: quantifies the average number of humans that become infected by each infectious mosquito
- Equation 3: combines the previous quantities to estimate the basic reproduction number.

The framework described in section 2.3 address equations 1 and 2 but does not explicitly address equation 3. The following modifications can be applied to accommodate this:

- Introduce the basic reproductive number (R_o) to the quantities of interest of the model. The R_o would define the expected number of secondary infections that one infectious individual would cause throughout the infectious period in a fully susceptible population.
- Set a threshold condition for disease-free equilibrium. This can be achieved using the next generation matrix approach, as described by Van denDriessche and Watmough (32), to derive the R_o, which defines a threshold condition for when the disease-free equilibrium (the point at which no disease is present in the population) is achieved. Thus, setting R_o as a threshold parameter for the model; if $R_o < 1$, then the disease-free equilibrium is locally asymptotically stable, whereas if $R_o > 1$), then it is unstable.

- Adopting a discrete event simulation approach with space and time being represented discretely. This will represent Madeira island as a spatial unit (i.e., in grid cells), with all processes of the model occurring in the basic spatial unit of the model.
- Modify the human population mixing from homogeneous to heterogeneous population in which the vital and epidemiological parameters for an individual will not be fixed, but dependent on factors such as disease stage, spatial position, age, or behaviour. This will be introduced via the spatial-temporal unit of the WARDEN system.

The above-described modification will introduce stochasticity to the model and provide important insights into the variation of dengue transmission in Madeira Island.

3.3 General conclusions

The various spatiotemporal modelling approaches presented in this thesis provide a significant expansion to existing frameworks modelling the spread and transmission of dengue fever. Overall, this thesis offers the following main contributions. First, it identifies factors facilitating the transport and importation of dengue fever into Europe, thereby providing valuable guidance to the parameterization of future epidemiological models. Second, this is a first attempt at applying machine learning algorithms and model-agnostic approaches to predict the risk of dengue importation into Europe, thereby offering a benchmark framework for subsequent predictive models and early warning surveillance systems. Third, it integrates a deterministic SEI-SEIR dynamic model to simulate the transmission of Dengue in Madeira Island, following an introduction of infectious individuals, thereby providing an understanding of how environmental conditions and seasonality affect dengue transmission and epidemic dynamics.

With dengue fever being non-endemic in Europe, and the majority of reported cases are travel-related, the importation models presented in this thesis are of interest to public health officials and researchers within the region, allowing to assess the importation risk of a viraemic case of dengue and the drivers behind its variation. The transmission model though spatially limited to Madeira Island provides relevant insights for control and mitigation of dengue transmission and could be adapted for other areas of Europe. When combined, these contributions can assist public health practitioners to develop a reliable, cost-effective, and scalable early warning surveillance system for dengue. The results of this thesis provide strategic and evidence-based information to public health decision-makers to guide implementation for targeted vector control and effective prevention strategies. The results also provide useful information for the control and prevention of other *Aedes aegypti* borne diseases both in Madeira Island and by extension to the European region. The modelling approaches presented could also be applied to chikungunya and yellow fever, two other diseases transmitted by the same mosquito vector (i.e., *Aedes aegypti*), by re-parameterisation of the models appropriately.

3.4 References

1. Silk MJ, Croft DP, Delahay RJ, Hodgson DJ, Weber N, Boots M, et al. The application of statistical network models in disease research. Methods in Ecology and Evolution. 2017;8(9):1026-41. doi: https://doi.org/10.1111/2041-210X.12770.

2. Moilanen A, Nieminen M. Simple connectivity measures in spatial ecology. Ecology. 2002;83(4):1131-45. doi: https://doi.org/10.1890/0012-9658(2002)083(1131:SCMISE)2.0.CO;2.

3. Chapman D, Purse BV, Roy HE, Bullock JM. Global trade networks determine the distribution of invasive non-native species. Global Ecology and Biogeography. 2017;26(8):907-17. doi: https://doi.org/10.1111/geb.12599.

4. Chapman DS, Makra L, Albertini R, Bonini M, Paldy A, Rodinkova V, et al. Modelling the introduction and spread of non-native species: International trade and climate change drive ragweed invasion. Glob Chang Biol. 2016;22(9):3067-79. doi: https://doi.org/10.1111/gcb.13220.

5. F. Dormann C, M. McPherson J, B. Araújo M, Bivand R, Bolliger J, Carl G, et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. Ecography. 2007;30(5):609-28. doi: https://doi.org/10.1111/j.2007.0906-7590.05171.x.

6. Rodrigue J-P. Spatial interactions and the gravity model. The geography of transport systems. Fourth ed. New York: Routledge; 2017. p. 440.

7. Semenza JC, Sudre B, Miniota J, Rossi M, Hu W, Kossowsky D, et al. International dispersal of dengue through air travel: Importation risk for europe. PLoS Negl Trop Dis. 2014;8(12):e3278. doi: https://doi.org/10.1371/journal.pntd.0003278.

8. Meslé MMI, Hall IM, Christley RM, Leach S, Read JM. The use and reporting of airline passenger data for infectious disease modelling: A systematic review. Eurosurveillance 2019;24(31):1800216. doi: https://doi.org/10.2807/1560-7917.ES.2019.24.31.1800216.

9. Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, et al. Generalized linear mixed models: A practical guide for ecology and evolution. Trends in Ecology & Evolution. 2009;24(3):127-35. doi: https://doi.org/10.1016/j.tree.2008.10.008.

10. Leifeld P, Cranmer SJ. Tnam: Temporal network autocorrelation models. 2017. https://cran.r-project.org/package=tnam.

11. Sadeghieh T, Waddell LA, Ng V, Hall A, Sargeant J. A scoping review of importation and predictive models related to vector-borne diseases, pathogens, reservoirs, or vectors (1999–2016). PLOS ONE. 2020;15(1):e0227678. doi: https://doi.org/10.1371/journal.pone.0227678.

12. Sammut C, Webb GI. Encyclopedia of machine learning and data mining: Springer Publishing Company, Incorporated; 2017. 1335 p.

13. Beam AL, Kohane IS. Big data and machine learning in health care. JAMA. 2018;319(13):1317-8. doi: http://dx.doi.org/10.1001/jama.2017.18391.

14. Miguel-Hurtado O, Guest R, Stevenage SV, Neil GJ, Black S. Comparing machine learning classifiers and linear/logistic regression to explore the relationship between hand dimensions and demographic characteristics. PLOS ONE. 2016;11(11):e0165521. doi: http://dx.doi.org/10.1371/journal.pone.0165521.

15. Singal AG, Mukherjee A, Elmunzer BJ, Higgins PDR, Lok AS, Zhu J, et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. The American journal of gastroenterology. 2013;108(11):1723-30. doi: http://dx.doi.org/10.1038/ajg.2013.332.

16. Molnar C. Interpretable machine learning: A guide for making black box models explainable. 2019. Available from: https://christophm.github.io/interpretable-ml-book/index.html.

17. Baehrens D, Schroeter T, Harmeling S, Kawanabe M, Hansen K, M K-R, et al. How to explain individual classification decisions. J Mach Learn Res. 2010;11:1803-31. http://www.jmlr.org/papers/volume11/baehrens10a/baehrens10a.pdf.

18. Ribeiro MT, Singh S, Guestrin C. Model-agnostic interpretability of machine learning. arXiv. 2016. https://arxiv.org/abs/1606.05386.

19. Sanchez I, Rocktaschel T, Riedel S, Singh S. Towards extracting faithful and descriptive representations of latent variable models in AAAI Spring syposium on knowledge representation and reasoning (KRR): integrating symbolic and neural approaches. 2015. http://terraswarm.org/pubs/482.html.

20. Healthmap. Denguemap (Internet). 2017 (cited 11 July 2017). Available from: https://www.healthmap.org/dengue/en/index.php.

21. International Air Transport Association. Airportis -market intelligence and analysis tool for airports. IATA; 2020 (cited 29 Jun 2020). Available from: https://www.iata.org/en/services/statistics/intelligence/airportis/.

22. Duong V, Lambrechts L, Paul RE, Ly S, Lay RS, Long KC, et al. Asymptomatic humans transmit dengue virus to mosquitoes. Proceedings of the National Academy of Sciences. 2015;112(47):14688-93. doi: https://doi.org/10.1073/pnas.1508114112.

23. IP-RAM I. Mosquito on madeira island (in portguese) Institute of Health Administration, IP-RAM, Regional Secretariat of Health, Autonomous Region of Madeira; 2019 (cited 17 Jul 2019). Available from: http://doc.iasaude.pt/mosquito/index.php/mosquito/mosquito-na-ilha-da-madeira.

24. IP-RAM I. Entomological panel bulletins(in portguese) IASAÚDE IP-RAM, Sanitary Engineering Unit: Institute of Health Administration, IP-RAM, Regional Secretariat of Health,

Autonomous Region of Madeira, 2019. Available from: http://doc.iasaude.pt/mosquito/index.php/boletins/entomologicos.

25. Weather Underground. Weather history for funchal madeira (2008-2018) 2019 (cited 01 Mar 2019). Available from: https://www.wunderground.com/history/daily/pt/funchal/LPMA.

26. Kiraly A, Janosi IM. Stochastic modeling of daily temperature fluctuations. Physical Review E. 2002;65(1539-3755 (Print)):051102 1-6. doi: https://doi.org/10.1103/PhysRevE.65.051102.

27. Lourenco J, Recker M. The 2012 madeira dengue outbreak: Epidemiological determinants and future epidemic potential. PLoS Negl Trop Dis. 2014;8(8):e3083. doi: https://doi.org/10.1371/journal.pntd.0003083.

28. IPCC. Climate change 2013: The physical science basis. Contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA: 2013; (1535). Available from: https://www.ipcc.ch/report/ar5/wg1/.

29. Seneviratne SI, Field CB, Barros V, Stocker TF, Dahe Q, Dokken DJ, et al. Managing the risks of extreme events and disasters to advance climate change adaptation. Cambridge: Intergovernmental Panel on Climate Change 2012; (582). Available from: https://www.ipcc.ch/report/managing-the-risks-of-extreme-events-and-disasters-to-advance-climate-change-adaptation/.

30. Berg Hvd, Velayudhan R, Ejov M. Regional framework for surveillance and control of invasive mosquito vectors and re-emerging vector-borne diseases 2014–2020. Copenhagen World Health Organization, 2013. Available from: http://www.euro.who.int/__data/assets/pdf_file/0004/197158/Regional-framework-for-surveillance-and-control-of-invasive-mosquito-vectors-and-re-emerging-vector-borne-diseases-20142020.pdf?ua=1.

31. Fundação para a Ciência e Tecnologia. An operational early warning system for dengue and other arboviral diseases in Madeira Island 2017. Available from: https://www.fct.pt/apoios/projectos/consulta/vglobal_projecto?idProjecto=151278&idElemConcurso=11922.

32. Van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Mathematical Biosciences. 2002;180(1):29-48. doi: https://doi.org/10.1016/S0025-5564(02)00108-6.

4 Annex

This annex contains other publication achieved during the period of the PhD program, which was not related to the thesis.

4.1 Africa Rising, a Narrative for Life Expectancy Gains? Evidence from a Health Production Function

---

Salami D, Shaaban AN, Oliveira Martins MDR (2019). Africa Rising, a Narrative for Life Expectancy Gains? Evidence from a Health Production Function. Ann Glob Health. 2019;85(1):63. doi: https://doi.org/10.5334/aogh.2307

Annals of Global Health

Salami D, et al. Africa Rising, a Narrative for Life Expectancy Gains? Evidence from a Health Production Function. *Annals of Global Health*. 2019; 85(1): 63, 1–10. DOI: https://doi.org/10.5334/aogh.2307

ORIGINAL RESEARCH

Africa Rising, a Narrative for Life Expectancy Gains? Evidence from a Health Production Function

Donald Salami, Ahmed Nabil Shaaban and Maria do Rosário Oliveira Martins

Background: The narrative of Africa Rising has increasingly been called into scrutiny, not just as a debate for economic growth and development, but also as a possible link to the surge in life expectancy on the continent. Theoretically, an increase in economic development tends to result in an increase in public health spending and subsequent better health outcomes.
Objective: This paper examines the contribution of economic development and other social determinants to the health status of the African continent and to provide evidence on whether the increase in life expectancy of the past two decades can be largely attributed to the Africa Rising narrative.
Methods: We estimated an empirical health production function, with life expectancy gains as the output of the health care system, and various socio-economic, environmental and lifestyle factors as contributory factors. We fitted a generalized least squares model, using panel data from 52 African countries for the period 1995–2014.
Findings: The estimation shows that while increases in health care spending contributed to life expectancy gains, urbanization rates and improved water access were the major drivers of life expectancy gains with substantially larger impacts in the past two decades.
Conclusions: Overall, the results provide an evidence base for iterating the need to prioritize increasing funding and examine more critically how to improve the efficiency of health spending. It also illustrates potential gains that can be achieved from an inclusive health policy agenda with a broader range of social and economic development issues.

Introduction

The surge in life expectancy in the African region since 2000 is evidently the greatest since the reverse gains in the 1990s. Life expectancy at birth increased in the region by 9.4 years to 60 years, with some countries experiencing as high as 42% rise between 2000 and 2014 [1]. These remarkable gains have taken place mainly in the context of increased access to antiretrovirals therapy, progress in malaria control and improvements in child survival over time [2]. Over the same period there was a significant economic growth and rise in income across the continent, popularly referred to as 'Africa Rising'. This rapid economic growth suggests more money for governments to spend on healthcare [3]. However, there is scarce evidence to support that the economic growth has led to higher health spending and higher health spending resulting in better health outcome for the continent. Few studies have been devoted to exploring the contribution of health inputs, particularly health expenditures, as a determinant of health outcomes in Africa.

To assess this relationship, we adopted the health production function framework using aggregate macro data. The theoretical tenets of this approach views health as an output of a health care system that is produced by a set of inputs to the system [4, 5]. A relatively large number of empirical studies have adopted this approach as a way of disentangling the overall impact of the health system on the health status of the population from the relative contribution of other broader socioeconomic factors [5–7]. The information gained from an aggregate health production function helps annotate the gains from increased health care spending and indicates whether investment in alternative health and socioeconomic agendas may have a better return on health outcomes [5]. This is a central question facing health policy makers in Africa in the context of whether the Abuja Declaration call to increase health expenditure yielded sufficient and commensurate value in health outcomes [8]. Furthermore, this is consequential to achieving universal health coverage and financial risk protection on the continent [9].

The aim of this paper is to examine the relationship between health care expenditure and life expectancy gains as a health outcome in African countries over the period of 1995–2014 by estimating the relative contributions of health care expenditure and the broader socioeconomic determinants of health as inputs of an empirical health production function. Few studies have previously estimated an aggregate health production function for the Sub-Saharan Africa region [10, 11]. We seek to build

Global Public Health, Institute of Hygiene and Tropical Medicine, Nova University of Lisbon, PT

Corresponding author: Donald Salami, M.Sc., MPH (donald.salami@gmail.com)

upon these previous studies by incorporating all countries on the continent and by refining the econometric methodology.

Methods

Data and Variables

Data used for our analysis comes mainly from the World Bank open database [12]; however, some variables were sourced from other officially recognized international sources. Utilizing a longitudinal study approach, we estimate life expectancy gains between 1995 to 2014 in 52 African countries, as the dependent variable in a health production function. Explanatory variables include factors representing health systems, income, socioeconomic, lifestyle, and environment. These variables reflect key determinants of health appropriate to the circumstances of the African continent and are consistent with previous empirical health production function models. The definition and sources of all variables are listed in Table II of the appendix.

Life expectancy (LE)

Life expectancy (LE) at birth is widely used as a proxy measure of population health statue at the aggregate level. A summary indicator of mortality conditions, it represents a proxy of health outcomes and distribution at different demographic, geographic, and economic groups in the population. The dependent variable for our health production function is life expectancy (LE) at birth for the total population. It indicates the average number of years that a newborn is expected to live if current mortality rates continue to apply [13].

Health Systems Factors (HF)

Total health expenditure and out-of-pocket spending are included to represent health systems inputs. The specification of the health care resources per capita, in empirical analysis, vary slightly from study to study. We included the monetary measurement represented by the total health expenditure per capita, the sum of public and private health spending as a ratio of total population. It constitutes the provision of health services, family planning activities, nutrition activities, and emergency aid designated for health, exclusive of the provision of water and sanitation [12]. The utilization of total health spending will provide a synthesis understanding of the available health funding with the system. Our a priori expectation is a positive association between health care spending and life expectancy gains, under the assumption that increase resources imply an improvement in the level and/or quality of health care. We do not anticipate diminishing returns given the generally low levels of expenditure in Africa in comparison to other, richer regions.

The second contributory variable to the health system included in our model is the out-of-pocket payments measured as a percent of total expenditure on health. Out-of-pocket payments are defined as expenditures borne directly by individuals or private households to health care providers at the time of service use, where neither compulsory nor voluntary insurance covers the full cost [12]. Some health production function studies include out-of-pocket health spending as a main explanatory variable, however, in recent times, it has become a key proxy for access to care and financial protection under the overall framework of the universal health coverage [14, 15]. The implicit theory is high out-of-pocket payment is negatively associated with financial protection, causing individuals to incur catastrophic expenditures, and substantially reduces the health-seeking behaviors of households. The reliance on out-of-pocket payments differs considerably amongst African countries, in some countries out-of-pocket spending are as high as 50% or more of current health spending [9], therefore the relative contribution to the health status cannot be predicted a priori.

Socioeconomic Factors (SF)

Socioeconomic influences on life expectancy are represented by three variables: income per capita, education, and food availability. Income per capita as measured by GDP per capita represents the level of economic development and wealth of a society. Income has been largely linked to being a major determinant of health status, though a seemingly ambiguous causal link. For the most part, empirical studies have suggested a positive relationship between income levels and health outcomes [7, 16]. Higher incomes facilitate access to goods & services (e.g., food, housing, transportation), better working conditions and better living standard, which in turn contributes to improved health and longevity. Converse arguments state, beyond some threshold level of affluence, increasing income may no longer improve health and rather result in adverse risky behaviors and unhealthy lifestyle (e.g., eating habit, alcohol & tobacco consumption), that may negatively affect health [15, 17]. Irrespective of the ambiguity of the impact of income on health, our a priori expectation is higher income contributes to improving health, due to low average income per capita across the region [10].

The second socioeconomic variable is educational attainment, measured as the average years of schooling of the population aged 15 and over. It has been widely theorized by researchers that higher education is an investment in better health [18]. Better educated people are typically more knowledgeable about the risks and benefits of different behaviors and better positioned to understand and act accordingly [19, 20]. They are also more informed about the availability of health services resulting in greater use of services [7]. Furthermore, education may be associated with improved self-management and subsequent efficiency of medical treatment, particularly for chronic diseases [21]. Therefore, we expect higher education levels to be positively associated with life expectancy gains.

Food availability is the third socioeconomic factor, measured as the gross food production per capita. Food production is based on the sum of price-weighted quantities of different agricultural commodities, covering food crops that are considered edible and that contain nutrients (exclusive of coffee and tea as they have no nutritive value) [12]. The production index is used as a proxy for food availability and food security, which is closely interconnected with nutrition security and subsequently

health [22]. When considered at per capita level, the effect of production on health can be validated in relation to warding off famine and prevention of chronic undernutrition or malnutrition. Africa over the years had recorded the proven successes in agricultural development [23], however, the effect on health cannot be predicted a priori, because maintaining food self-sufficiency seems daunting in conflict-affected areas in the continent.

Environmental Factors (EF)

Urbanization, carbon dioxide emissions, water access, and sanitation were included as environmental variables. Urban population refers to people living in urban areas as defined by national statistical offices, a percentage of the total population. It is a proxy for a collection of potential negative and positive health-related factors [5]. High degree of urbanization is associated with adverse consequences in congestion, increase in slum settlements, poverty, ill-health and the decline in urban capital per person (e.g., access to services, water, and sanitation, etc.) [24]. However, urbanization is not inherently negative, it can be a positive determinant of health, via the gains in access to better job opportunities, public service/goods, infrastructure, and better health care. The WHO commission on macroeconomics and health affirms that investments in urban health can create major returns on economics and life expectancy of the population [25]. The marginal effect of urbanization will be dependent on the net effect of these two channels, hence the impact on life expectancy is indeterminate.

The second environmental variable is carbon dioxide emissions (metric tons per capita) as a proxy for air pollution. Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring. The negative effect of carbon dioxide (CO_2) on health and causal link of air pollution mortality have been well documented [26]. We expect to capture the negative effects of air pollution on health outcomes.

The third environmental factor is access to water and sanitation, which tracks two indicators, percentage of the population with access to improved drinking water sources and improved sanitation. Access to an improved water source refers to the percentage of the population using an improved drinking water source. While access to improved sanitation facilities refers to the percentage of the population using improved sanitation facilities. Access to water and sanitation is well documented to reduce the effects of exposure to pollution and disease, minimize contact with harmful contaminants, such as bacteria and viruses, thereby promoting health and well-being [27]. Thus, both indicators should have a positive impact on health outcomes.

Lifestyle Factors (LF)

Lastly, we included adult alcohol consumption in litres per capita as a lifestyle variable. Total alcohol consumption per capita measured in equivalent litres of pure alcohol (ethanol) consumed per capita per person ages 15+ per year. Excessive alcohol consumption has detrimental health effects. An associated risk factor for numerous chronic diseases (cardiovascular disease, liver cirrhosis, and certain cancers) as well as accidents and violent deaths [28]. A priori expectation is to capture the negative effects of high consumption on health.

Econometric Specification

Our basic analytical approach utilizes macro-level panel data to estimate an empirical aggregate health production function. A Cobb-Douglas production function is employed with all variables expressed in logarithmic form. We used panel data of 52 African countries over the period of 1995–2014.

The general form of the health production function was specified as follows:

$$\ln LE_{i,t} = \alpha_i + \beta_1 \ln H_{i,t} + \beta_2 \ln S_{i,t} + \beta_3 \ln E_{i,t} + \beta_4 \ln L_{i,t} + e_{i,t}$$

with all variables in natural logarithm form **(*ln*)** and LE_i is the life expectancy at birth for country ***i*** in period ***t***; ***α*** the country fixed effect to account for country level characteristics that are constant over the period of analysis; and ***e*** is the error term.

H is the vector of health systems factors (which includes – total health expenditure per capita constant US$ PPP, as a measure of health care spending, out-of-pocket expenditure as a percentage of total health expenditure as a proxy for access to care and financial protection).

S is the vector of socioeconomic factors (which includes – GDP per capita constant US$ PPP, as a measure of income, mean years of schooling as a measure of educational attainment, gross food production per capita US$, a proxy for food availability).

E is the vector of environmental factors (urban population as a percentage of the total population as a measure urbanization, CO_2 emissions metric tons per capita as a proxy measure for air pollution, percentage of the population with access to improved drinking water sources and improved sanitation facilities as a measure of access to water and sanitation).

L is a lifestyle factor represented by alcohol consumption in litres per capita.

Estimation Method

We fitted a Generalized Least Squares (GLS) model, which was subsequently complemented by an alternative specification to address possible endogeneity issue and lag-effects. The different empirical models are explained below.

Model 1 – Standard model, with all explanatory variables included. GLS model specified with country fixed effects (specified by country dummies), country-specific first-order autocorrelation structures for errors (AR1) and a correction for heteroscedasticity. This specification is as follows:

$$\ln LE_{i,t} = \alpha_i + \beta_1 \ln H_{i,t} + \beta_2 \ln S_{i,t} + \beta_3 \ln E_{i,t} + \beta_4 \ln L_{i,t} + e_{i,t} \quad [Model\,1]$$

Model 2: as per model 1, with 5-year lagged explanatory variables. The introduction of lagged variables considers

Table 4-1 Descriptive statistics of variables

possible delayed effects of key determinants of health on life expectancy. Also corrects for possible endogeneity, due to possible simultaneity between life expectancy and some of the explanatory variables. Choice of the 5 year lag is based on previous literature and maintaining an adequate number of observations for the analysis. Explanatory variables were lagged from 1995 to 2009, to account for delayed effects in assessing life expectancy gains for 2000 to 2014. The specification is as follows:

$$\ln LE_{i,t} = \alpha_i + \beta_1 \ln H_{i,t-5} + \beta_2 \ln S_{i,t-5} + \beta_3 \ln E_{i,t-5} + \beta_4 \ln L_{i,t-5} + e_{i,t} \quad [Model\ 2]$$

All models were fitted using the GLS function of the nlme R package [29], with the default method of maximizing the restricted loglikelihood (REML). Model accuracy and goodness of fit were based on the normalized root mean square error, and the R^2 statistic between actual and predicted values. All analyses were performed using R version 3.5.0.

Results

We estimated two panel-data regressions on the life expectancy at birth, for 52 African countries in the period of 1995–2014 on a broad range of explanatory variables (descriptive statistics are listed in **Table 1**). Data gaps in specific years were addressed utilizing a standard linear interpolation with linear extrapolation outside of the known data range. However, two countries (South Sudan and Somalia) were excluded from our analysis due to significant missing data for most variables and time-period. All variables in the model were log transformed, as such coefficients will be interpreted as elasticities.

Results from our standard model (model 1, **Table 2**), show that while adjusting for other factors, increase in health expenditure, income per capita, education, urbanization, and access to sanitation over time have a positive and statistically significant impact on life expectancy gains. Surprisingly out of pocket expenditure and carbon dioxide emission had a positive and statistically significant coefficient, which is contrary to a priori expectations. Likewise, improved water access was contrary to a priori expectations with a negative but insignificant coefficient. The strength and significance of all coefficients were notably impacted with the introduction of the lagged variables in the subsequent model. Model 2 is preferred as it considers delayed effects and partially addressed endogeneity issues.

Table 2 shows the coefficients for model 2 estimation, with the following specific findings: the coefficient of urbanization and access to water were positive and statistically significant, suggesting that a 1 percent increase in urbanization rates and access to improved water source increase life expectancy gains by about 0.24% and 0.11% respectively, other variables held at a constant. The coefficient of health expenditure, income per capita, education and sanitation access were positive and statistically significant. However, the practical significance of their impact on life expectancy at a 1 percent change level is considerably low. Their contribution to life expectancy gains over the review period is represented in **Table 3**.

The direction and significance of the food production coefficient remain unchanged even after accounting for lagged effects in model 2. Thereby suggesting the variable has a negative impact on life expectancy, a contradiction with a priori expectations.

Lastly, the result in **Table 1** indicates that a change in carbon dioxide emission and alcohol consumption may adversely impact life expectancy outcome, though not statistically significantly. However, these results reflect the impact of lag in time before these variables affect an individual's health and the relative marginal changes in both variables over time in African countries.

Table 3 presents the relative contributions of the explanatory variables to life expectancy gains over time. During the period under review, the most significant contributor was urbanization with a resultant gain of 23.1

Table 1: Descriptive statistics of variables.

	Observation	Mean (Range)
Life expectancy at birth, total (years)	1034	57.03 (31.96–75.64)
Health expenditure per capita, PPP (US$)	1032	202.06 (5.94–1768.68)
Out-of-pocket health expenditure (% of total health expenditure)	1032	39.65% (2.00–80.91)
Gross domestic product, PPP † (US$)	1021	41.16 (2.72–384.41)
Mean years of schooling population 25+ years‡	959	4.46 (0.90–10.30)
Food production index	1037	100.67 (55.77–182.51)
Urban population (% of total)	1037	39.53% (7.21–86.92)
*CO_2 emissions (metric tons per capita)**	1036	112.90 (1.57–1004.37)
Improved water source (% of population with access)	1025	68.72% (19.50–99.90)
Improved sanitation facilities (% of the population with access)	1038	37.26% (3.00–98.40)
Total alcohol consumption, per capita‡*	934	305.50 (1.00–1189.00)

† – Divided by 100; * – Multiplied by 100.

‡Total alcohol consumption and mean years of schooling had the most data gaps, with a missing analysis of 10.2% and 7.8% respectively. This was addressed using standard linear interpolation with linear extrapolation outside of the known data range.

Table 4-2 Regression coefficient of model estimations

Table 2: Regression coefficients of model estimations.

Explanatory Variables	Model 1 – Life expectancy coefficient (t-ratio)	Model 2 – Life expectancy coefficient (t-ratio)
Health Expenditure	0.0023 (2.6494)	0.0027 (2.8082)
Out-of-pocket Expenditure	0.0015 (1.9770)	0.0009 (0.8500)
Income	0.0064 (4.9465)	0.0146 (7.4467)
Education	0.0132 (5.4185)	0.0190 (7.7176)
Food Production	–0.0007 (–1.9799)	–0.0021 (–2.4424)
Urbanization	0.3537 (46.3057)	0.2385 (22.3734)
CO_2 emissions	0.0019 (2.2721)	–0.0002 (–0.2572)
Water Access	–0.0156 (–1.7724)	0.1054 (8.1903)
Sanitation Access	0.0402 (6.9397)	0.0471 (8.4221)
Alcohol Consumption	0.0000 (0.1579)	–0.0003 (–0.7965)
Intercept	2.6208 (0.0467)	2.5203 (0.0322)
Observations (countries)	1040 (52)	780 (52)
NRMSE Accuracy	0.985	0.990
R^2	0.813	0.911

Model 1 = Generalized least square regressions, with country-fixed effects, error terms following a country-specific AR(1) and correction for heteroskedasticity. Model 2 = As Per model 1, with 5-year, lagged explanatory variables. NRMSE = normalized root mean square error. R^2 = Efron's pseudo r-squared.

Table 3: Relative contributions of the explanatory variables to life expectancy gains over time: 1995 to 2009 (from estimated coefficients in Table 2).

Explanatory variables	Regression Coefficient[1]	Contribution to Life Expectancy		1995 Value	2009 Value
		%	Months		
Health Expenditure	0.0027**	0.21	1.4	123.6	264.8
Out-of-pocket Expenditure	0.0009	ns	ns	43.9	37.4
Income	0.0146**	1.11	7.7	2403.8	5154.9
Education	0.0190**	0.51	3.5	3.6	4.7
Food Production	–0.0021*	–0.02	–0.13	94.6	103.7
Urbanization	0.2385**	3.36	23.1	35.8	41.3
CO_2 emissions	–0.0002	ns	ns	0.9	1.2
Water Access	0.1054**	1.56	10.7	62.1	72
Sanitation Access	0.0471**	0.67	4.6	33.9	39.1
Alcohol Consumption	–0.0003	ns	ns	2.9	3.1
Countries Average			687.6		

1 = model 2, with 5-year, lagged explanatory variables. Regression based on 780 Observations across 52 countries.
Significant code: ** indicates significance at 1%; * indicates at 5%; 'ns' indicates not significant.

months of life expectancy. Another major contributing variable to gains in life expectancy over the two decades was improved water access, then followed by income per capita, and then access to sanitation and education. Health expenditure and income per capita had the most significant growth by 114% respectively from 1995 to 2009, but surprisingly this growth did not yield a commensurate impact on life expectancy gains. These results conclude that the changes in health expenditure over the review period have made a relatively small contribution to the increase in life expectancy in African countries.

Discussion

Utilizing a macro-perspective health production function approach, we measured the impact of the health care system inputs in conjunction with other contributory factors

Table 4-3 Relative contributions of the explanatory variables to life expectancy gains over time: 1995 to 2009

on life expectancy gains in African countries. The results of our empirical model are consistent with the conclusions of Or [30], which suggested that environmental factors are more important than health systems inputs in explaining variations in life expectancy gains. Our findings are also consistent with other empirical health production function studies: an increase in health expenditure, income, and better education have a favorable impact on life expectancy gains [7, 10, 15].

Health care expenditure contributed marginally to the gains in life expectancy in African countries over the past two decades. This is rather startling, as health expenditure per capita surged by 114% in real terms from 1995 to 2009 (from USD PPP 123.6 in 1995 to USD PPP 264.8 in 2009 in constant terms) [12]. The theoretical assumption of increased economic development in the region leading to increased health spending and substantially contributing to better health outcomes was not well supported by our results. Likewise, it contradicts the proposition of Nixon et al. [7], stating slight changes in health expenditure of developing countries would almost certainly lead to bigger impacts on health outcomes. This relatively small contribution of health expenditure might be explained by large financing gaps and low domestic investments in relation to the health needs in African regions [31, 32]. To assess the sensitivity of our coefficient estimates, in view of endogeneity and collinearity issues, we estimated an alternative specification by excluding GDP per capita. Two major variables that are highly correlated in a health production function is GDP per capita and health expenditure per capita. Theoretically, instrumental variables could be used to address this issue, though in practice it is challenging to identify a plausible instrument and results prove sensitive to the choice of instruments to provide reliable estimates [33]. Specifications excluding GDP per capita imply a larger impact of health expenditure, suggesting that when GDP is omitted income effects that are unrelated to health expenditure are captured. This scenario also assumes that the impacts of GDP are indirect, i.e. through health expenditure. Health expenditure remained statistically significant in this alternative specification (See Table I in the appendix) suggesting the coefficient estimates to be the actual contribution of health expenditure to life expectancy gains.

Urbanization and improved water access are found to be the major contributing factors to gains in life expectancy during the review period. Urbanization rates had a substantial (15.1%) change in real terms from 1995 to 2009, with a resultant gain of 23.1 months of life expectancy (**Figure 1**). Our a priori expectation for urbanization rates was indeterminate as it could inherently have a positive or negative impact on health outcomes. The positive effects of urbanization on life expectancy, as suggested by our results, are indisputable given the explosive city growth rates since the 1960s. Indeed Africa is now the second-fastest urbanizing continent [34]. This growth can be postulated to have resultant returns on the continent's economic growth, fostering better income, and better income can be linked to improved health outcomes [35]. However, we do not negate the potential negative impact of high urbanization rates on health outcomes over time, arguably from the stance of it leading to increased pollution and congestion [24]. On a cautionary note, the true marginal effect of urbanization should be considered as the net effect of both channels. Similarly, improved drinking water source had a 16% increase in real terms from 1995 to 2009, with a relative contribution of 13.0 months to life expectancy (**Figure 1**). These results are consistent with the conclusions of other previous studies assessing the impact of water access on health [36]. Improved water access has been associated with the prevention of a tenth of the global disease burden and prevention of

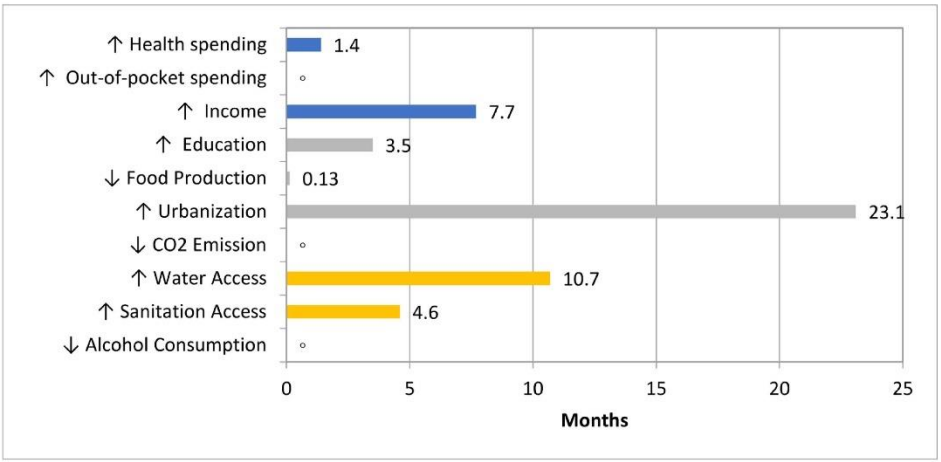

Figure 1: Life expectancy gains associated with percentage change in explanatory variables over time.
°indicates not significant.

Figure 4-1 Life expectancy gains associated with percentage change in explanatory variables over time.

Table 4-4 Appendix Table I: Alternative model specification excluding GDP per capita.

dehydration, which has a direct impact on population health outcomes. Likewise, it is also reasonably linked to improving food security, livelihood choices, poverty reduction and educational opportunities, which indirectly impacts health outcomes in African countries [37]. The gain in improved water access is an imperative prerequisite to achieving most of the dimensions of the sustainable development goals in Africa.

Our findings indicate that an increase in food production, a decrease in carbon dioxide emission and alcohol consumption, may not have contributed to the gains in life expectancy within the period. A probable reason for the negative effect of food production might be due to variations in annual production level for drought-stricken and conflict-affected areas in the continent [38]. To explore this assumption, we estimated an alternative specification of our model, with the addition of a year-specific time dummy variable (details not shown). The direction remained unchanged, but became non-significant suggesting the effects of specific time-period shocks in food production with restrictive impacts on life expectancy. The coefficient for carbon dioxide emission and alcohol consumption became negative and significant with control for specific years. As earlier stated, a plausible explanation for results of carbon dioxide (CO_2) emission and alcohol consumption is the reflection of a long lag in time before these variables affect a person's health. Furthermore, it is worth mentioning that the African region is notably still one of the lowest global CO_2 emitters [39–41]. And that the marginal effects of CO_2 emission on population health are also dependent on the duration of exposure and the balancing oxygen concentration in the atmosphere at a given time [42]. In an equivalent manner, the continent has somewhat of a low alcohol-attributable burden of disease, which is attributed to lag time and likely data paucity [43]. It would be more informative to explore alternative model specifications with longer lags on these variables. We did not pursue this, as longer lags would reduce the number of data observations in our models, which will inadvertently lead to weaker effects.

Conclusions

Our results not only provide some new evidence on the contributors to life expectancy gains in African countries over the past two decades but raises several important considerations for policymakers. The observed marginal effect of health financing gives an evidence base for iterating the need to prioritize bridging the large gaps in healthcare financing, increase domestic investments in health systems, and examine critically how to improve the efficiency of health spending in African countries as already proposed by the Abuja Declaration. The policy implications of other findings are also important, as they illustrate potential gains that can be achieved from inclusive healthcare agendas with a broader range of social and economic development issues. Our model specifications followed the best methodological practice, however, the limitations of the macro-level approach which demonstrates only aggregate effects and not country-level differences should be considered when interpreting the findings. Our study is broadly consistent with the previous health production function study for Sub-Sahara Africa countries and supports the general opinions of other similar empirical studies.

Appendix

Table I: Alternative model specification excluding GDP per capita.

Explanatory Variables	Full Model[1]	Model excluding Income[2]
Health Expenditure	0.0027 (2.8082)	0.0041 (4.6023)
Out-of-pocket Expenditure	0.0009 (0.8500)	−0.0000 (−0.0136)
Education	0.0190 (7.7176)	0.0281 (10.0765)
Food Production	−0.0021 (−2.4424)	−0.0013 (−1.4477)
Urbanization	0.2385 (22.3734)	0.2811 (26.5612)
CO_2 emissions	−0.0002 (−0.2572)	0.0011 (1.1730)
Water Access	0.1054 (8.1903)	0.0412 (5.2916)
Sanitation Access	0.0471 (8.4221)	0.0686 (12.2647)
Alcohol Consumption	−0.0003 (−0.7965)	−0.0001 (−0.6496)
Income	0.0146 (7.4467)	
Intercept	2.5203 (0.0322)	2.5710 (0.0293)
Observations (countries)	780 (52)	780 (52)
NRMSE Accuracy	0.990	0.990
R^2	0.911	0.907

[1] GLS Model, with country-fixed effects, error terms following a country-specific AR(1) and correction for heteroskedasticity, with 5-year lagged explanatory variables.

[2] As per full model specification, excluding income (GDP per capita).

Table 4-5 Appendix Table II: Definition and sources of all variables.

Table II: Definition and sources of all variables.

Variable	Definition	Aggregation Method	Source
Life expectancy at birth, total (years)	Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.	Total (years) Weighted average	World Bank
Health expenditure per capita	Total health expenditure is the sum of public and private health expenditures as a ratio of total population. It covers the provision of health services (preventive and curative), family planning activities, nutrition activities, and emergency aid designated for health but does not include the provision of water and sanitation.	PPP (constant 2011 international $) Weighted average	World Bank
Out-of-pocket expenditure	Out of pocket expenditure is any direct outlay by households, including gratuities and in-kind payments, to health practitioners and suppliers of pharmaceuticals, therapeutic appliances, and other goods and services whose primary intent is to contribute to the restoration or enhancement of the health status of individuals or population groups. It is a part of private health expenditure.	% of total expenditure on health Weighted average	World Bank
Gross domestic product – GDP, PPP (constant 2011 international $)	PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products.	PPP (constant 2011 international $) Weighted average	World Bank
Mean years of schooling population 25+ years	An average number of completed years of education of a country's population aged 25 years and older, excluding years spent repeating individual grades.	Total (years) Weighted average	UNESCO Institute for Statistics
Food Production index (Production Indices)	Food production index covers food crops that are considered edible and that contain nutrients. Coffee and tea are excluded because, although edible, they have no nutritive value.	Gross per capita Production Index Number (2004–2006 = 100): International US$	Food and Agriculture Organization of the United Nations (FAO)
Urban population (% of total)	Urban population refers to people living in urban areas as defined by national statistical offices.	% of population Weighted average	World Bank
CO_2 emissions (metric tons per capita)	Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring.	Metric tons per capita	World Bank
Improved water source (% of the population with access)	Access to an improved water source refers to the percentage of the population using an improved drinking water source. The improved drinking water source includes piped water on premises (piped household water connection located inside the user's dwelling, plot or yard), and other improved drinking water sources (public taps or standpipes, tube wells or boreholes, protected dug wells, protected springs, and rainwater collection).	% of population Weighted average	World Bank
Improved sanitation facilities (% of the population with access)	Access to improved sanitation facilities refers to the percentage of the population using improved sanitation facilities. Improved sanitation facilities are likely to ensure hygienic separation of human excreta from human contact. They include flush/pour flush (to the piped sewer system, septic tank, pit latrine), ventilated improved pit (VIP) latrine, pit latrine with slab, and composting toilet.	% of population Weighted average	World Bank
Total alcohol consumption per capita (liters of pure alcohol, projected estimates, 15+ years of age)	Total alcohol consumption per capita is based on projections for the amount of alcohol consumption (liters of pure alcohol) per person ages 15+ per year.	Liters of alcohol consumed per capita. Weighted average	World bank

Funding Information

The study was partially funded by the Fundação para a Ciência e a Tecnologia. DS and ANS both have PhD grants from Fundação para a Ciência e a Tecnologia (PD/BD/128084/2016). We express our thanks to Manuel Serrano for his comments and constructive criticisms in the preparation of this manuscript.

Competing Interests

The authors have no competing interests to declare.

References

1. **World Health Organization.** WHO methods and data sources for life tables 1990–2015; 2016. https://www.who.int/healthinfo/global_burden_disease/LT_method_1990_2015.pdf?ua=1. Accessed June 25, 2018.
2. **World Health Organization.** World health statistics 2016: monitoring health for the SDGs, sustainable development goals; 2016. apps.who.int/iris/bitstream/10665/206498/1/9789241565264_eng.pdf. Accessed June 25, 2018.
3. **Bughin J, Chironga M, Desvaux G,** et al. Lions on the move II: Realizing the potential of Africa's economies. *Mckinsey Glob Inst*; 2016. https://www.mckinsey.com/global-themes/middle-east-and-africa/lions-on-the-move-realizing-the-potential-of-africas-economies. Accessed June 25, 2018.
4. **Grossman M.** The demand for health: A theoretical and empirical investigation. New York, NY: National Bureau of Economic Research; 1972.
5. **Thornton J.** Estimating a health production function for the US: Some new evidence. *Appl Econ.* 2002; 34(1): 59–62. DOI: https://doi.org/10.1080/00036840010025650
6. **Auster R, Leveson I** and **Sarachek D.** The Production of Health: An exploratory study. *J Hum Resour.* 1969; 411–36. DOI: https://doi.org/10.2307/145166
7. **Nixon J** and **Ulmann P.** The relationship between health care expenditure and health outcomes: Evidence and caveats for a causal link. *Eur J Heal Econ.* 2006; 7: 7. DOI: https://doi.org/10.1007/s10198-005-0336-8
8. **United Nations.** Abuja Declaration on HIV/AIDS, Tuberculosis and Other Related Infectious Diseases. In: *African Summit on on HIV/AIDS, Tuberculosis and Other Related Infectious Diseases*; 2001. http://www.un.org/ga/aids/pdf/abuja_declaration.pdf. Accessed June 25, 2018.
9. **World Health Organization.** New Perspectives on Global Health Spending for Universal Health Coverage; 2018. http://www.who.int/health_financing/topics/resource-tracking/new_perspectives_on_global_health_spending_uhc.pdf. Accessed June 25, 2018.
10. **Fayissa B** and **Gutema P.** Estimating a health production function for Sub-Saharan Africa (SSA). *Appl Econ.* 2005; 37(2): 155–164. DOI: https://doi.org/10.1080/00036840412331313521
11. **Torkian E.** A panel data approach to the measurement of health technical efficiency of Sub-Saharan Africa. *Zagreb Int Rev Econ Bus.* 2015; 18(1): 1–15. DOI: https://doi.org/10.1515/zireb-2015-0001
12. **World Bank.** Indicators|Data; 2018. https://data.worldbank.org/indicator. Accessed June 25, 2018.
13. **World Health Organization.** The world health report 2006: Working together for health; 2006. https://www.who.int/whr/2006/whr06_en.pdf. Accessed June 25, 2018.
14. **Moreno-Serra R** and **Smith PC.** Broader health coverage is good for the nation's health: Evidence from country level panel data. *J R Stat Soc Ser A Stat Soc.* 2014; 178(1): 101–124. DOI: https://doi.org/10.1111/rssa.12048
15. **Chris James MD** and **Franco S.** Inclusive growth and health. *OECD Heal Work Pap.* 2017; 103. DOI: https://doi.org/10.1787/93d52bcd-en
16. **Heijink R, Koolman X** and **Westert GP.** Spending more money, saving more lives? the relationship between avoidable mortality and healthcare spending in 14 countries. *Eur J Heal Econ.* 2013; 14: 527. DOI: https://doi.org/10.1007/s10198-012-0398-3
17. **De Vogli R, Mistry R, Gnesolto R** and **Cornia GA.** Has the relation between income inequality and life expectancy disappeared? Evidence from Italy and top industrialised countries. *J Epidemiol Community Health.* 2005; 59(2): 158–162. DOI: https://doi.org/10.1136/jech.2004.020651
18. **Grossman M.** The human capital model of the demand for health. *J Polit Econ.* 1999; 04: 7078. DOI: https://doi.org/10.1007/s13398-014-0173-7.2
19. **Cutler DM** and **Lleras-Muney A.** Understanding differences in health behaviors by education. *J Health Econ.* 2009; 29(1): 1–28. DOI: https://doi.org/10.1016/j.jhealeco.2009.10.003
20. **Mackenbach JP, Stirbu I, Roskam A-JR,** et al. Socioeconomic inequalities in health in 22 European countries. *N Engl J Med.* 2008; 358: 2468–2481. DOI: https://doi.org/10.1056/NEJMsa0707519
21. **Goldman DP** and **Smith JP.** Can patient self-management help explain the SES health gradient? *Proc Natl Acad Sci.* 2002; 99(16): 10929–10934. DOI: https://doi.org/10.1073/pnas.162086599
22. **Swaminathan MS** and **Bhavani RV.** Food production & availability – Essential prerequisites for sustainable food security. *Indian J Med Res.* 2013; 138(3): 383–391. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3818607/.
23. **Spielman DJ** and **Pandya-Lorch R.** Millions Fed: Proven Successes in Agricultural Development. 2009; 1–18. DOI: https://doi.org/10.2499/9780896296619BK
24. **Self S** and **Grabowski R.** How effective is public health expenditure in improving overall health? A cross-country analysis. *Appl Econ.* 2003; 35(7): 835–845. DOI: https://doi.org/10.1080/0003684032000056751
25. **World Health Organization.** Macroeconomics and health: Investing in health for economic

development: Report of the Commission on Macroeconomics and Health; 2001. http://www.who.int/iris/handle/10665/42463. Accessed June 25, 2018.
26. **Jacobson MZ.** On the causal link between carbon dioxide and air pollution mortality. *Geophys Res Lett.* 2008; 35: L03809. DOI: https://doi.org/10.1029/2007GL031101
27. **WHO/UNICEF JMP.** Progress on Drinking Water, Sanitation and Hygiene WHO Library Cataloguing-in-Publication Data; 2017. http://www.unwater.org/publications/whounicef-joint-monitoring-program-water-supply-sanitation-hygiene-jmp-2017-update-sdg-baselines/. Accessed June 25, 2018.
28. **Rehm J** and **Shield KD.** Global alcohol-attributable deaths from cancer, liver cirrhosis, and injury in 2010. *Alcohol Res.* 2014; 35(2): 174–183. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3908708/.
29. **Pinheiro J, Bates D, DebRoy S, Sarkar D** and **RCT.** Linear and nonlinear mixed effects models [R package nlme]. 2018; 3.1-139 https://cran.r-project.org/web/packages/nlme/index.html.
30. **Or Z, Elmeskov J, Feiner M, Hurst J, Martin J** and **Scherer P.** Determinants of health outcomes in industrialised countries: A pooled, cross-country, time-series analysis. *OECD Econ Stud.* 2000; 30(1): 53–77. https://www.oecd.org/eco/growth/2732311.pdf.
31. **Mills A.** Health care systems in low- and middle-income countries. *N Engl J Med.* 2014; 6; 370(6): 552–7. DOI: https://doi.org/10.1056/NEJMra1110897
32. **Husøy Onarheim K.** Global health financing: Priority to poor people or poor countries? *Tidsskr den Nor legeforening.* 2017; 22: 1–6. DOI: https://doi.org/10.4045/tidsskr.17.0715
33. **Joumard I, André C, Nicq C** and **Chatal O.** Health status determinants: Lifestyle, environment, health care resources and efficiency. *OECD Econ Dep.* 2008; 627: 1–74. DOI: https://doi.org/10.1787/240858500130
34. **Sow M.** Foresight Africa 2016: Urbanization in the African context. The Brookings Institution. https://www.brookings.edu/blog/africa-in-focus/2015/12/30/foresight-africa-2016-urbanization-in-the-african-context/. Accessed June 25, 2018.
35. **World Health Organization.** Our Cities, Our Health, Our Future; 2008. https://www.who.int/social_determinants/resources/knus_final_report_052008.pdf. Accessed June 25, 2018.
36. **Metwally AM, Ibrahim NA, Saad A** and **Abu El-Ela MH.** Improving the roles of rural women in health and environmental issues. *Int J Environ Health Res.* 2006; 16(2): 133–144. DOI: https://doi.org/10.1080/09603120500539208
37. **United Nations.** Water and Sanitation – United Nations Sustainable Development; 2018. https://www.un.org/sustainabledevelopment/water-and-sanitation/. Accessed June 25, 2018.
38. **Food and Agriculture Organization of the United Nations.** The State of Food Security and Nutrition in the World (SOFI) Report; 2017. https://www.wfp.org/content/2017-state-food-security-and-nutrition-world-sofi-report. Accessed June 25, 2018.
39. **Boden TA, Marland G** and **Andres RJ.** Global, Regional, and National Fossil-Fuel CO_2 Emissions; 2017. DOI: https://doi.org/10.3334/CDIAC/00001
40. **British Petroleum.** BP Statistical Review of World Energy 2017. *Br Pet*; 2017. http://www.bp.com/content/dam/bp/en/corporate/pdf/energy-economics/statistical-review-2017/bp-statistical-review-of-world-energy-2017-full-report.pdf. Accessed June 25, 2018.
41. **United Nations Framework Convention on Climate Change.** National Inventory Submissions; 2017. https://unfccc.int/index.php/process-and-meetings/transparency-and-reporting/reporting-and-review-under-the-convention/greenhouse-gas-inventories-annex-i-parties/submissions/national-inventory-submissions-2017. Accessed June 25, 2018.
42. **Rice SA.** Health Effects Of Acute And Prolonged CO_2 Exposure In Normal And Sensitive Populations. In: *Second Annual Confernce on Carbon Sequesration*; 2003. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.464.2827&rep=rep1&type=pdf. Accessed June 25, 2018.
43. **World Health Organization.** Global status report on alcohol and health; 2014. https:/entity/substance_abuse/publications/global_alcohol_report/en/index.html. Accessed June 25, 2018.

How to cite this article: Salami D, Shaaban AN and Oliveira Martins MR. Africa Rising, a Narrative for Life Expectancy Gains? Evidence from a Health Production Function. *Annals of Global Health.* 2019; 85(1): 63, 1–10. DOI: https://doi.org/10.5334/aogh.2307

Published: 29 April 2019

]u[*Annals of Global Health* is a peer-reviewed open access journal published by Ubiquity Press. OPEN ACCESS