Yellow fever in South America: The role of environment and host on transmission dynamics

Arran Timothy Patrick Hamlet

Department of Infectious Disease Epidemiology

Imperial College London

Thesis submitted for PhD examination

January 2020

Declaration of originality

I declare this work to be my own, produced under the supervision of Dr Tini Garske and Prof. Neil Ferguson. Any participants involved in collaborative work have been acknowledged in the relevant sections of this thesis.

Arran Hamlet

January 2020

Copyright declaration

Acknowledgements

I would like to thank my supervisor's Dr Tini Garske and Professor Neil Ferguson, for their guidance and encouragement. The support and flexibility coupled with their enthusiasm for the field has led to many interesting and exciting opportunities that would not have been possible without them.

To Dr Tini Garske in particular, who first took me on as an MSc student project on yellow fever, then a Research Assistant and then a PhD student. I am deeply indebted to her for bringing me into the world of yellow fever and shaping my entire experience of early career science into one that has been resoundingly positive.

I also owe much thanks to Dr Katy Gaythorpe, who as the yellow fever postdoc for the last couple years has been an unwavering source of advice, and a great soundboard for methodological conversations, paper reviews, and discussions on how best to geospatially map bigfoot. I look forward to continuing to work together on both YF and nonsensical side projects.

To UG12, who I have shared an office with for 4 years. I've learnt a lot throughout the years about the scientific process and successful networking (Monkey Puzzle/Rowans/Karaoke), your support and encouragement has been tremendously uplifting, and your professional advice invaluable. A special mention also goes to Patrick Walker, who made the foolish mistake of offering to proof read my thesis and then was taken up on it.

My colleagues in DIDE, who have made the department such a great place to learn and work in, I am grateful to everyone who has provided help, advice, a conversation or a laugh. I look forward to continuing to enjoy your company after this PhD.

I would like to thank the Medical Research Council UK for funding this work and for all the opportunities they have provided me

Finally, and most importantly, I would like to thank my partner Alison and my family who have been fully supportive and provided substantial encouragement. Alison, it's finally time for me to start paying you back in holidays and wine.

Relevant publications and presentations

Publications

1. **Published**
    - **a.** Hamlet A, Jean K, Yactayo S, Benzler J, Cibrelus L, Ferguson N, et al. POLICI: A web application for visualising and extracting yellow fever vaccination coverage in Africa. Vaccine. 2019;37(11):1384-8.
    - **b.** Hamlet A, Jean K, Perea W, Yactayo S, Biey J, Van Kerkhove M, et al. The seasonal influence of climate and environment on yellow fever transmission across Africa. PLoS Neglected Tropical Diseases. 2018;12(3):e0006284.
    - **c.** Dorigatti I, Hamlet A, Aguas R, Cattarino L, Cori A, Donnelly CA, et al. International risk of yellow fever spread from the ongoing outbreak in Brazil, December 2016 to May 2017. Euro Surveill. 2017;22(28).
2. **Under review**
    - a. Seasonal and inter-annual drivers of yellow fever transmission in South America. PLoS Neglected Tropical Diseases
        - i. Hamlet A, Gaythorpe KAM, Garske T, Fergusion, NM.
    - b. Seasonality of agricultural exposure as an important predictor of seasonal yellow fever spillover in Brazil. Nature Communications
        - i. Hamlet A, Ramos DG, Gaythorpe KAM, Romano APM, Garske T, Ferguson NM.
    - **c.** Eliminating yellow fever epidemics in Africa: vaccine demand forecast and impact modelling. PLoS Neglected Tropical Diseases

i. Jean K, Hamlet A, Benzler J, Cibrelus L, Gaythorpe KAM, Sall A, Ferguson NM, Garske T.

Presentations

1. **Oral**

a. International Conference on One Medicine One Science

i. Chang Mai, Thailand, 2019, *Land-use, vegetation and habitat fragmentation as drivers of yellow fever transmission in South America.*

b. American Society of Tropical Medicine and Hygiene (ASTMH)

i. National Harbour, USA, 2019, *Seasonality of agricultural exposure more important than seasonality of climate for predicting yellow fever transmission in Brazil.*

c. Outbreak Analysis and Modelling for Public Health

i. Bogota, Colombia, 2019, *Statistical and mathematical modelling of Yellow fever in South America*.

d. Seasonality of climate on Arboviruses Workshop

i. London, UK, 2019, *Seasonality of yellow fever in Africa and Brazil*

e. Yellow fever forecasting: Embedding modelling in lessons learn exercises

i. Brasilia, Brazil, 2018, *Yellow fever in Brazil – Modelling as a tool to inform outbreak response and public health policy.*

f. International Conference on Emerging Infectious Diseases (ICEID)

i. Atlanta, USA, 2018, *Habitat fragmentation and land-use change as drivers of yellow fever outbreaks in South America*.

2. Poster

a. American Society of Tropical Medicine Annual Meeting

   i. Baltimore, USA, 2017, *Yellow fever in South America and the potential for expansion of the endemic zone.*

   ii. New Orleans, USA, 2018, *Habitat fragmentation and land-use change as drivers of yellow fever outbreaks in South America*

b. Epidemics

   i. Sitges, Spain, 2017, *Yellow fever in South America and the potential for expansion of the endemic zone*

c. Impact of Environmental Changes on Infectious Diseases (IECID)

   i. Trieste, Italy, 2017, *The seasonal influence of climate and environment on yellow fever transmission across Africa*

Abstract

Yellow fever (YF) is an arbovirus that affects both humans and non-human primates (NHPs). Despite a longstanding recognition of YF as a significant public health problem, many aspects of its underlying transmission and maintenance remain unknown. These knowledge gaps continue to exist even with the increasing availability of data, techniques and recent large-scale outbreaks in South America and Africa.

Using several statistical and machine learning methods, I investigate the role of climate, environment and host in predicting the suitability of YF across South America, as an average, seasonally and inter-annually. Following I examine the role of seasonality of agriculture, as a proxy for exposure, on human and NHP YF reports in Brazil. To contextualise these predictions and recommendations, I calculate and describe population-level YF vaccination coverage estimates (1940-2050) across Africa and South America, as well as the interactive web-based platform these are published on. Finally, I predict the distribution and density of NHP genera across the South-East of Brazil, and use this in a stochastic multi-species, age structured, meta-population model to explore the role of NHP genera on viral maintenance, and the potential for the establishment of endemicity in the state of Rio de Janeiro.

Overall this thesis describes several key aspects necessary to understand the enigma of YF transmission in South America. A greater understanding of climate and environment allows for the possibility of forecasting periods of heightened transmission which could inform pro-active surveillance and vaccination, supported through our vaccination coverage estimates. Finally, by providing insights into the role of NHP genera on maintenance and critical community sizes for YFV transmission, I can highlight the potential for endemicity to be established.

TABLE OF CONTENTS

List of Figures

List of Tables

List of Appendices

List of Abbreviations and Acronyms

YF	Yellow fever
YFV	Yellow fever virus
WHO	World Health Organization
PAHO	Pan American Health Organization
NHP	Non-human primate
AUC	Area under the Curve
ROC	Receiver operating characteristic
ENM	Ecological Niche Modelling
EVI	Enhanced Vegetation Index
BRT	Boosted Regression Tree
RF	Random forest model
GLM	Generalised Linear Model
DRC	Democratic Republic of the Congo

Chapter 1 - Introduction and literature review

1.1 Background and aims of the thesis

Yellow fever (YF) is a disease caused by the yellow fever virus (YFV), a mosquito transmitted arbovirus of the genus *Flavivirus* affecting both humans and non-human primates (NHPs) (Monath and Vasconcelos, 2015). YF is considered endemic in 34 countries of Sub-Saharan Africa, and 13 countries in South America (Jentes et al., 2011), though around 90% of the global burden is estimated to occur in Africa where it causes estimated 130,000 severe cases and 78,000 deaths annually (Garske et al., 2014). Due to the presence of NHP's as a sylvatic reservoir, eradication is unlikely to occur.

Though YF has a substantially lower burden in South America, compared with Africa, it still represents a significant public health threat (Barnett, 2007, Garske et al., 2014). Despite this significance, much about the underlying epidemiology of YFV in South America is only partially understood, particularly in topics such as:

- The association between the climate and environment and YF transmission.
- The seasonality of YF, transmission in NHPs and human spillover.
- The role and importance of different NHP species in transmission maintenance and amplification.

These questions exist despite relatively robust public health systems, long established surveillance systems and the ability to use NHP epizootic events as additional indicators of YF transmission, as opposed to Africa where NHP's do not show severe or symptomatic illness (4). Since 1942, in South America, the sylvatic cycle has accounted for almost all human cases of YF with, until recently, the majority of these

confined to Amazonia regions (Johansson et al., 2012). However, in the past 20 years YF has rapidly expanded its endemic zone, and has seen numerous countries re-draw the lines of what is considered “endemic”, as well as seen the return of urban YF to the continent (Chaves et al., 2018, Romano et al., 2014, Johansson et al., 2012, Rezende et al., 2018). The recent inclusion of some of South America’s most densely populated metropolises into areas of transmission, and the re-introduction and proliferation of *Aedes aegypti,* the urban vector of YF, to much of South America is a significant cause for alarm (Vasconcelos et al., 1999, Cunha et al., 2019).

Given this rapidly developing situation it is imperative that further investigation into the underlying epidemiology of YFV transmission and expansion occurs – and that this is utilised in a manner that allows for informed and rational public health planning. To quantify associations between YF, the climate and environment, and how these are changing could offer the ability to act proactively, improving surveillance efforts and directing vaccination activities in at risk areas before human cases occur. This information, combined with a more mechanistic understanding of how YF transmission in NHPs is maintained, and how this relates to spillover, can then allow for more informed forecasts of potential endemic expansion and epidemic occurrence. In order to provide light on these particular aspects of YF transmission, this thesis aims to address four aspects:

1. To describe the role of host, climate and environment on the distribution of recent YF reports, both spatially and temporally.
2. To investigate the role of seasonality in YF transmission.
3. To estimate the population-level vaccination coverage across Africa and Latin America, both historically and future predicted.

4. To evaluate the role of different non-human primate species on the transmission and maintenance of YF.

1.2 Clinical disease in humans

YFV infection has a range of pathologies, occurring after the initial intrinsic incubation period of between 4.3 and 5.6 days (Johansson et al., 2010). Following this initial period, an estimated 55% of infections remain asymptomatic, 33% exhibit only mild illness and the remaining 12% severe disease (Johansson et al., 2014). YFV infection leads to lifelong immunity (Monath and Vasconcelos, 2015).

Symptomatic infection beings with numerous non-specific symptoms such as fever, headache, and general myalgia and nausea (Monath, 2001) (see Figure 1.2.1). This period generally lasts 3-6 days and is when individuals are infectious to mosquitoes (Monath, 2001). Generally, mild disease is not identified as YFV infection due to the non-specific nature of the symptoms and the presence of numerous, and more common, conditions that share these. This, in combination with a large proportion of asymptomatic infections, leads to the clinical presentation of YF only representing a minority of YFV infections and contributes to the substantial underestimations of the true YF burden (Johansson et al., 2014).

Following mild symptomatic disease there is a period of remission, in which symptoms abate and the virus is cleared from the blood. Those who do not develop severe disease recover with no further complications. In patients who continue to develop severe disease, there is a relapse and intensification of symptoms within 24 hours. During this severe stage of disease, patients are not infectious. In addition to the previously described symptoms, increased vomiting, jaundice, renal failure and

haemorrhaging can occur. This intensified period of illness lasts 3-8 days and is fatal in ~47% of cases (Johansson et al., 2014).

Estimates of the proportion of asymptomatic, mild, severe cases and the case fatality ratio was estimated by (Johansson et al., 2014), which identified 11 studies with detailed information on YF prevalence and disease outcomes. Studies were restricted to those who had shown the detection of YF viral antibodies or detection of YF virus by culture or PCR. This data was then integrated into a Bayesian model to estimate the probabilities of these outcomes. Given the quality of the data used, and the robustness of the method these estimates are likely to be accurate.

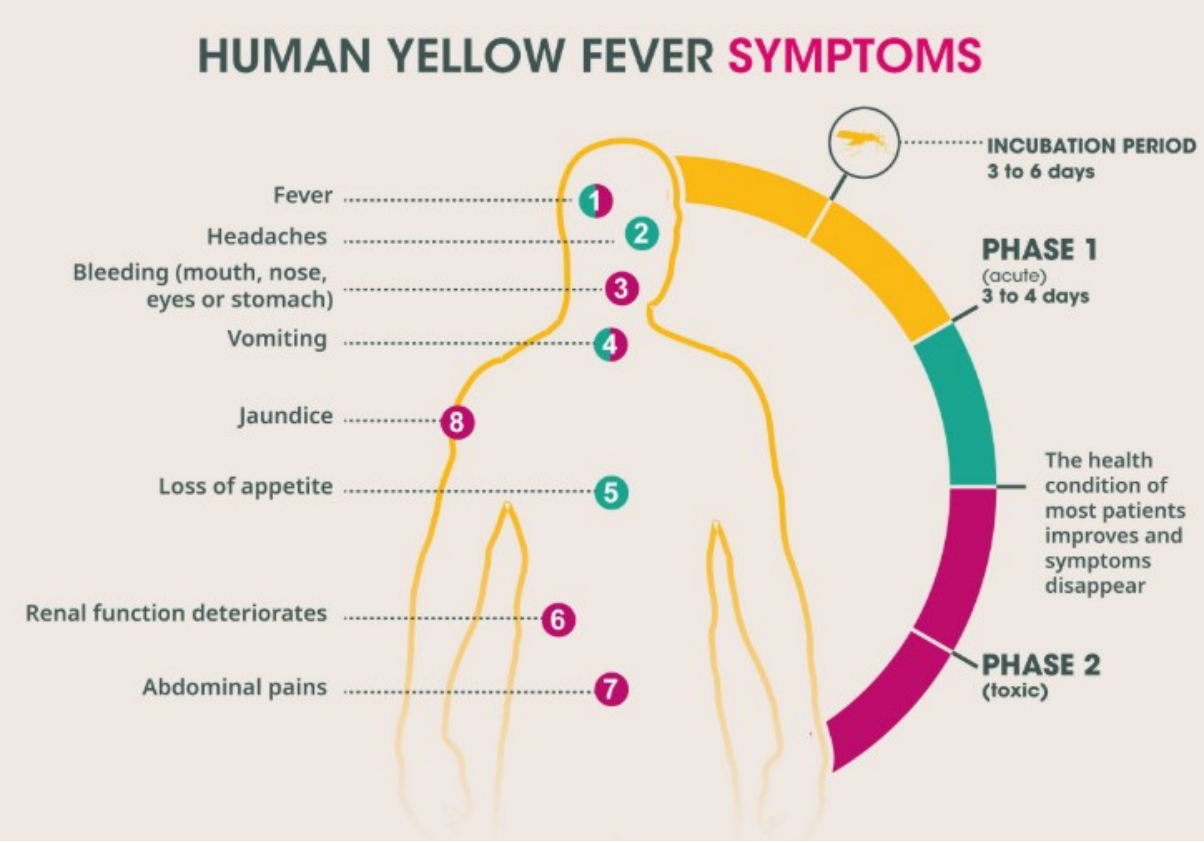


Figure 1.2.1. Stages of YF infection, showing major clinical features of the disease. Taken from https://www.sanofi.com/en/your-health/vaccines/yellow-fever.

1.3 Clinical tests for the detection of the YFV

There are several clinical tests for YF, which broadly fall into one of two categories, virological or serological. When these are used depends on availability of the test and the current timing in the course of infection. Molecular diagnosis is generally focused on the initial viremic phase (other than histopathological analysis which occurs post mortem), and serological diagnosis is only applicable after the viremic phase (Pan American Health Organization, 2018).

RT-PCR can be used to detected viral RNA during the viremic phase, which occurs during, approximately, the first 10 days of infection, or in severe cases longer. This is the gold standard of diagnosis as it is not susceptible to cross-reactivity with other *Flavivirus* infections as other serological methods are. Histopathological analysis of the liver in fatal YF cases additionally offers a confirmed diagnosis of YF as infection causes characteristic damage to liver cells.

Serological diagnosis can be carried out by either Enzyme-linked immunosorbent assay (ELISA) tests, or Plaque Reduction Neutralization Tests (PRNT), however both of these have substantial cross-reactivity with other flaviviruses such as dengue or zika, and previous YF immunization and so only indicate a probable infection.

1.4 History of the YFV – origin to present

YFV evolved in West Africa, initially as a zoonotic pathogen before becoming an important viral infection of humans. This history is reflected in the increased genetic variation in West African strains of YFV as opposed to East Africa, or South America (Lepiniec et al., 1994, Wang et al., 1996), and the resistance of Africa NHP's to fatal infection. While African NHP's are important contributors to viral maintenance (Monath and Vasconcelos, 2015), they generally do not exhibit symptomatic infection (Jentes

et al., 2011) – a likely indication of a longer period of virus and host co-evolution compared to humans (Bryant et al., 2007, Romano et al., 2014).

The spread of the YFV outside of Africa coincided with the European exploration of West Africa, and the subsequent movement of goods, and slaves. The slave trade is thought to have led to the introduction of the YFV to South America where it spread via competent vectors and established itself in susceptible NHP populations, becoming endemic around 300 to 400 years ago (Bryant et al., 2007).

While new endemic regions were not established outside of South America, YF continued to cause large scale epidemics across much of the Western Hemisphere. The USA suffered numerous and recurrent epidemics in coastal towns such as Boston, New York, Baltimore and New Orleans (Strode et al., 1951). Europe also suffered outbreaks in France, Italy, the United Kingdom and Spain, with an outbreak of in Cadiz, Spain causing an estimated 51,000 deaths in 1800 (Waddell, 1990). These frequent outbreaks often had substantial economic and social impacts (Patterson, 1992), such as influencing the establishment of Haiti as an independent country, and foiling initial attempts to build the Panama Canal (Chippaux and Chippaux, 2018).

During this period of time, YF was thought to be caused by “atmospheric miasmata” (Staples and Monath, 2008), an idea that was only challenged in 1881 when Carlos Finlay, a Cuban doctor, postulated about the transmission of YF through Aedes mosquitoes. This theory was further investigated and eventually proven by Walter Reed, under a US army commissioned report (Carter et al., 1931). Though these accreditations have been called into question (Clements and Harbach, 2017). The identification of YFV’s vector led to a wave of vector-control campaigns, initially in

Cuba, then Panama and the rest of the Americas, with the aim of eliminating YF. While there was substantial initial success in stopping the transmission of the YFV in numerous countries across the region, the presence of a yet undiscovered sylvatic cycle made elimination impossible.

It was not till the late 1920's when infection studies revealed the susceptibility of NHP's to infection, and their ability to contribute to onward transmission of the virus (Davis, 1930), that this sylvatic cycle was first recognised (Burke, 1937). This, in conjunction with the occurrence of YF outbreaks in the absence of Aedes aegypti (Soper et al., 1933), led to the discovery that a South American sylvatic genus of mosquitoes, Haemagogus, could act as a vector of the virus (Antunes and Whitman, 1937). With these revelations, it became accepted that sylvatic YF would remain endemic and act as a permanent source of potential re-infection for urban locations (Soper, 1937).

While the limitations of vector-control were being realised, research into the immunological response of YFV infection in humans and NHP's led to the development of two YF vaccines in the 1930's, the French neurotropic vaccine and the American YF-17D (Roukens and Visser, 2008, Theiler and Whitman, 1935). Both theses vaccines were utilised to great effect in the 1940's in both South America and French West Africa, drastically reducing outbreaks (Durieux, 1956, Smithburn and Mahaffy, 1945), with the French neurotropic vaccine being particularly useful in mass vaccination campaigns due to its application through small, superficial scratches to the skin (scarification), rather than subcutaneous injection like the YF-17D vaccine (Frierson, 2010). However, following these large campaigns it was discovered that the French vaccine had severe adverse side effects, most significantly encephalitis, and was eventually discontinued in 1982 (Monath, 1996). This, in addition to the withdrawal of France from West Africa led to a cessation of large-scale vaccination activities in

many of the YF endemic countries of Africa. As populations grew in the absence of routine vaccination, the incidence of YF increased with 34 large outbreaks recorded from 1985 to 2005, the vast majority in West Africa (Barrett and Higgs, 2007).

From 1990, the YF vaccine has been increasingly adopted into routine infant immunization, covering 22 African and 12 South American countries in 2016. Additionally, a global stockpile of 6 million doses a year has been maintained since 1997 for outbreak response, and, since 2006, the YF Initiative (YFI) has conducted large scale vaccination campaigns in 14 countries in West and Central Africa with the highest burden (World Health Organization, 2010b). However, demand has rapidly outstripped production and stockpiles are proving insufficient to deal with current needs. This was particularly evident during the 2015-2016 outbreak in Angola and the Democratic Republic of the Congo (DRC) which exhausted 12 million doses in its first five months (Barrett, 2016) and 30 million by the end of the epidemic. In response the WHO sanctioned the use of fractionated doses, which was shown to be efficacious at $1/10^{th}$ dilutions (Monath et al., 2016), and led to the vaccination of almost 8 million people using a $1/5^{th}$ dose in Kinshasa, the capital of the DRC.

The re-emergence of urban YF at a large scale, and the importation of cases from Angola to China, resulted in the WHO creating a new global and comprehensive long term (2017-2026) strategy to deal with eliminating yellow fever epidemics by 2026, the EYE strategy (World Health Organization, 2016a). This strategy has three strategic objectives;

1. Protect at risk populations.
2. Prevent the international spread of the disease.
3. Contain outbreaks rapidly.

While the impact of this strategy still remains to be seen, there has been progress made with the implementation of large-scale vaccination campaigns in Nigeria and the Republic of Congo (World Health Organization, 2019b, World Health Organization, 2018), as well as an increasing refinement of the methodologies behind “risk classification” of areas for endemic and epidemic YFV transmission.

1.5 Transmission cycles

The zoonotic component of YFV transmission introduces additional transmission cycles to the maintenance of yellow fever in the form of the sylvatic (jungle), intermediate (savannah) and the urban cycle.

In the sylvatic cycle, transmission is maintained between NHP’s through sylvatic mosquito species. In Africa this is thought to be primarily due to the sylvatic *Aedes africanus* (Mutebi and Barrett, 2002), and in South America mosquitoes of the *Haemogogus* and *Sabethes* genera (Barrett and Higgs, 2007, Kumm, 1950). In South America this cycle has accounted for almost all cases since 1942 (Johansson et al., 2012). Humans are considered incidental hosts in this cycle and do not contribute to onward transmission. Due to this, human cases of sylvatic YF are usually limited in number and geographic location.

The intermediate cycle is only found in Africa and acts as a “bridging point” between NHP’s and humans, facilitated by peri-domestic species of the *Aedes* genera (Mutebi and Barrett, 2002). This cycle typically occurs in semi-rural areas where there is potential for both humans and NHP’s to contribute to the maintenance of transmission, and if the virus establishes in the diurnal and domestic *Aedes aegypti*, then it leads to the urban cycle.

The urban cycle involves sustained transmission between humans, via *Aedes aegypti*, in the absence of an NHP reservoir. Historically this cycle has caused the largest and most devastating outbreaks, both within and outside of its endemic zone (World Health Organization, 2017b, Monath and Vasconcelos, 2015).

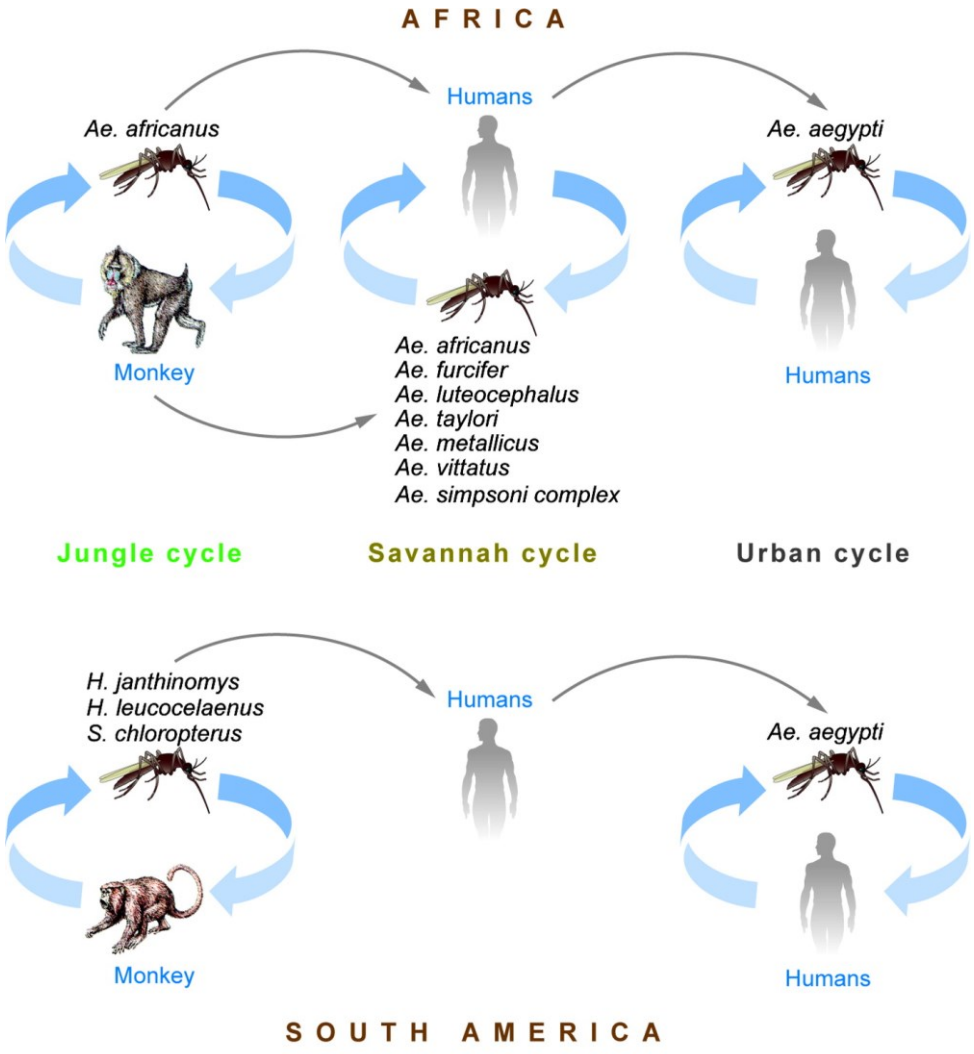


Figure 1.5.1. Transmission cycles of YF. Taken from Barrett and Higgs (2007) (Barrett and Higgs, 2007).

1.6 Role of non-human primates

Initially thought to be purely a human disease spread by *Aedes aegypti*, the role of NHP's in YF transmission was first suggested in 1914 (Balfour, 1914, Christophers,

1960), but not confirmed till the late 1920's. This led to widespread and large-scale serological studies aimed at ascertaining the susceptibility of numerous New World NHP species to infection, often involving thousands of NHPs (Kumm and Laemmert, 1950). These studies revealed widespread and silent transmission across much of South and Central America (Kumm and Laemmert, 1950, Trapido and Galindo, 1955, Vargas-Mendez and Elton, 1953) and suggested a wave-like spread of YF, with epizootic events occurring every 7-14 years following sufficient replenishment of susceptible NHP populations (Vasconcelos et al., 2004). Although both *Aedes* and *Haemogogus* species have shown evidence of vertical transmission of YF, they do so at very low rates (1:597 and 1:5745 respectively) (Beaty et al., 1980, Dutary and Leduc, 1981).

The recognition of the role NHPs play in transmission, and the susceptibility of certain species, has resulted in the adoption of epizootic surveillance as an integral part of YF prevention (Almeida et al., 2014). The use of both passive surveillance, testing sick or dead NHPs, and active surveillance, testing NHPs regardless of illness or death, is recommended by the Pan American Health Organization (PAHO) as an integral part of monitoring YF (Pan American Health Organization, 2013). However, despite the importance of NHPs to transmission and surveillance, many issues regarding the roles that species susceptibility, density, and abundance play in transmission are relatively unknown.

South American primates are severely threatened by habitat loss, habitat fragmentation, hunting and the illegal pet trade, with around half of primate species vulnerable or endangered (IUCN, 2017). Small, fragmented populations are especially vulnerable to extinction, and there is a real possibility that the regional extinctions of susceptible NHPs will be driven, in part, by YF (Holzmann et al., 2010). This risk of

localised extinction, driven by YF, was evaluated for brown howler monkeys in north eastern Argentina by Moreno et al., (2015) (Moreno et al., 2015). By constructing a demographic model based on specific population parameters they evaluated population dynamics, in the context of recurrent YF epizootics, through an SEIR model over 100 years. The findings suggested both a high probability of population decline, and localised extinction rising substantially with the frequency of outbreaks.

1.7 Vaccination

While there is no cure, the YF vaccine (YF-17D) is highly efficacious, has few adverse reactions and provides life-long protection with a single dose (Jean et al., 2016). Due to the presence of a sylvatic reservoir YF cannot be eradicated but, by achieving a sufficient population level vaccination coverage, outbreaks can be prevented. A coverage of 80% or higher appears adequate to suppress outbreaks in even the highest transmission settings in West Africa (Monath and Vasconcelos, 2015). The current vaccine was first produced in the 1930's and has changed little since its inception. A product of this developmental stagnation is that the capacity to ramp up production in response to a sudden demand is limited due to antiquated production techniques (Roukens and Visser, 2008).

This inability has had severe consequences with responding to epidemics (Monath et al., 2016), maintaining sufficient stockpiles, and sustaining routine infant immunization. The current global vaccine shortage has resulted in endemic South American countries only receiving around 50% of their estimated requirements (World Health Organization, 2016b). Furthermore, globally, no new countries have implemented routine immunization in their national programmes since 2008 despite circulation of YFV within countries outside of this subset, and current levels are sub-

optimal in many countries (World Health Organization, 2016b). Without substantial increases in vaccine production, the global stockpile shortage will only worsen due to an ever-increasing disparity between stockpiles and populations in areas suitable for transmission. This may lead to a resurgence in cases, as seen in West Africa during the 1980's and 1990s, and increase the chance of exportation outside of traditional endemic zones (Barrett and Higgs, 2007).

The authorisation by the WHO of a fractional dose represents one approach to improving the global stockpile (World Health Organization, 2017a). Previously published clinical data had provided evidence of the minimum dose of the vaccine required to provide adequate efficacy (Martins et al., 2013), as well as suggesting that "sub dosing" could produce an equivalent immunological response to administration of the full vaccine (Campi-Azevedo et al., 2014). Follow up studies have found that, compared to administering the full dose, fractional doses have similar safety profiles and confer an equivalent level of immunity, at least in the short-term (Roukens et al., 2018, Nzolo et al., 2018).

1.8 Yellow fever in South America

Historically, much of Latin America has reported outbreaks of YF. However, since the latter half of the 20th century transmission has been confined to South America, Panama and Trinidad and Tobago (Barnett, 2007) (Figure 1.8.1). Despite a lower burden than in Africa, partly due to higher vaccination coverage, YF remains a substantial public health risk (Barnett, 2007). Since 1960 there have been 7772 cases and 2963 deaths confirmed to be due to YF in Latin America (World Health Organization, Pan American Health Organization, 2017), see Figure 1.8.2. This is likely a substantial underestimation of the true burden of disease as only severe cases

are picked up, due to non-specific symptoms and poor surveillance masking the actual number of infections. Though these issues are shared with YF in Africa, surveillance and transmission detection is likely to be of higher quality in South America due to more robust healthcare systems, availability of surveillance systems and the use of NHP epizootics as “sentinel events” (Almeida et al., 2014).

Figure 1.8.1. Areas with risk of yellow fever virus transmission in South America, 2010. Figure reproduced from Jentes et al., (2011)(Jentes et al., 2011).

In South America the sylvatic cycle has accounted for almost all cases since 1942 (Johansson et al., 2012), and has historically been confined to Amazonian regions and, primarily, young males who enter these forested regions for economic activities (Gardner and Ryman, 2010, Pan American Health Organization, 2017). However, over the past 20 years YF has rapidly expanded its endemic zone to include areas that previously had not reported human or NHP cases, such as the re-emergence of urban YF in South America during an outbreak in Asuncion, Paraguay in 2008. In Brazil this has resulted in 5 re-assessments of areas considered endemic for YF since 2000, with the latest update in 2018 including the entire country (Chaves et al., 2018, Romano et al., 2014). Recent outbreaks in Brazil's South-East Atlantic forest in 2016-2019 have been the largest recorded in the country, for both human and NHPs, with cases reported in states that have previously been absent of YF in the last 80 years (Rezende et al., 2018).

The reasons for these expansions are unknown, and while it may be related to the re-infestation of the continent by *Aedes aegypti*, which was removed from much of South America in the early 20th century through highly effective vector control campaigns (Soper et al., 1943, Barnett, 2007), this is uncertain. An *Aedes aegypti* driven expansion of the endemic zone would likely result in significant urban transmission, and centred around urban areas where it thrives, as opposed to the more rural and isolated spillover events observed. The expansion is more likely driven by increasing human interactions with endemic areas and wildlife reservoirs, in combination with improved transport which has allowed infected hosts and vectors to be dispersed far from the original site of infection.

These factors, alongside rampant urbanisation, and the increasing ease of intra- and international travel, highlight the need for a more complete understanding of the epidemiology of YF in order to guard against future expansions and urban outbreaks.

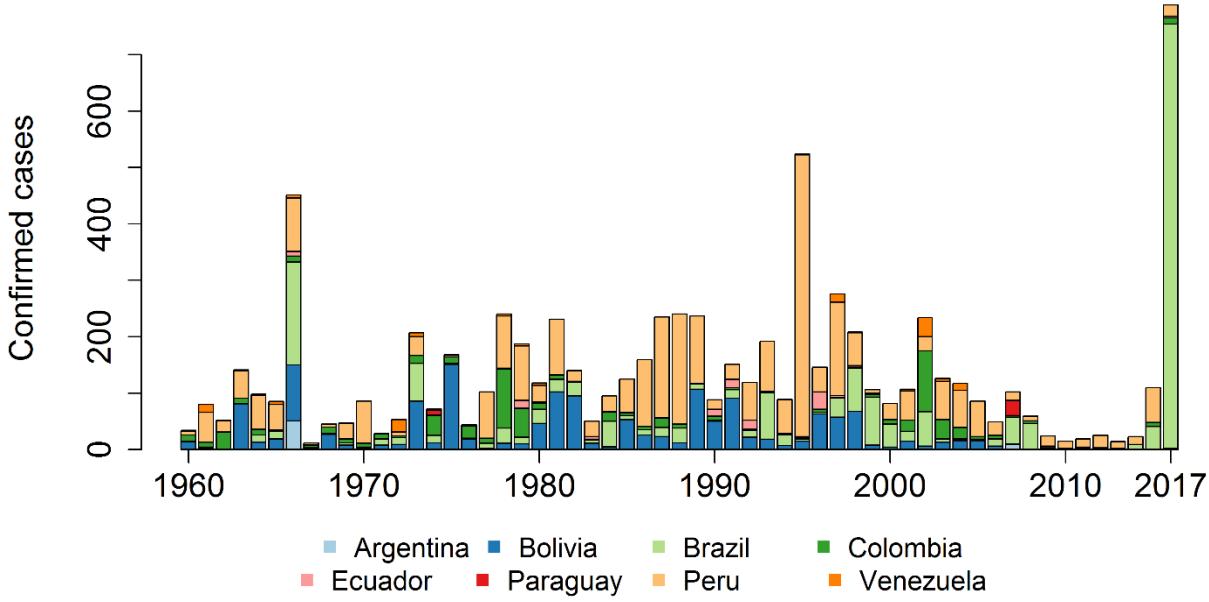


Figure 1.8.2. YF cases 1960-31/05/2017 in the 8 most affected countries in Latin America. Plot generated from PAHO web sources and reports (Pan American Health Organization, 2017, AMARELA, 2017, World Health Organization).

1.9 YF and the risk of transmission in Asia

Despite suitable climatic conditions (Rogers et al., 2006), competent vectors (Kraemer et al., 2015, van den Hurk et al., 2011), and susceptible populations, YF transmission has never been reported in Asia. Several theories, of varying credibility, have been suggested to explain its absence such as differences in vector competence, competition between *Aedes aegypti* and *Aedes albopictus* and cross immunity from other flaviviruses, such as dengue (Gubler, 2004, Amaku et al., 2011).

However, these explanations appear insufficient. Differences in African/South America and Asian vector competence are unlikely to be substantial, with modelling studies showing a wide range of vector competences that could result in endemic transmission (Amaku et al., 2011). Furthermore *Aedes albopictus* has shown the ability to transmit YF in laboratory settings (Amraoui et al., 2016), and both dengue and YF and *Aedes aegypti* and *Aedes albopictus* coexist within Africa (Kraemer et al., 2015, Bhatt et al., 2013). While the absence of YF may be due to a combination of these factors, it may also be explained by a lack of large-scale population movement between Africa and Asia in the past.

YF was introduced into Europe and the Americas following large scale population movement from Africa in the form of the West African slave trade (Cathey and Marr, 2014). Despite a longer duration, the East African slave trade with Asia was of a far smaller magnitude, which in combination with East Africa's lower burden of YF could explain the historical absence of YF in Asia (Ellis and Barrett, 2008, Cathey and Marr, 2014). If so, then this absence is unlikely to be maintained in the light of increasing air travel and trade between endemic regions and Asia (Baliamoune-Lutz, 2011), with several cases exported to China from Angola in 2016 (Li et al., 2016). As the threat of introduction and establishment grows, and YF could conceivably follow in the footsteps of similar arboviruses such as Zika's or Chikungunya's recent global expansion (Wikan and Smith, 2016, Zeller et al., 2016), the results of which would be devastating (Wasserman et al., 2016).

1.10 The influence of climate and seasonality on arboviruses

Mosquitoes, like all insects, are poikilothermic ectotherms and as such their internal temperature is linked to the ambient temperature. This has a wide variety of implications with regards to transmission. Higher temperatures, below temperature induced mortality, increase the speed of pupation (Mohammed and Chadee, 2011), the frequency of blood meals (Paaijmans et al., 2013) and reduce the extrinsic incubation period (Johansson et al., 2010). The combination of all these factors result in higher temperatures, generally, increasing the competence of the vector for disease transmission (Figure 1.10.1).

Temperature is neither the sole determinant in competence, nor a standalone factor. Sufficient availability of standing water is required for the aquatic life-stages of mosquitoes, as well as humidity affecting survival of adult stages (Yamana and Eltahir, 2013). Additionally, the interplay between precipitation and temperature may be more important than either variable alone (Hamlet et al., 2018). Both optimal temperatures with insufficient rainfall and insufficient temperatures with optimal rainfall would be unfavourable for the mosquito, and viral replication within the mosquito, therefore limiting the potential for transmission. Furthermore, land cover type is significantly associated with vector species prevalence and density and can influence the transmission potential even in the presence of suitable climates (Johnson et al., 2008, Gleiser and Zalazar, 2010, Vanwambeke et al., 2007). The combination of these factors mean that the sustained presence of vector-borne diseases is intrinsically related to the climate, land cover and land use, a fact that can be exploited in modelling the presence and burden of vector and virus (Bhatt et al., 2013, Rogers et al., 2006).

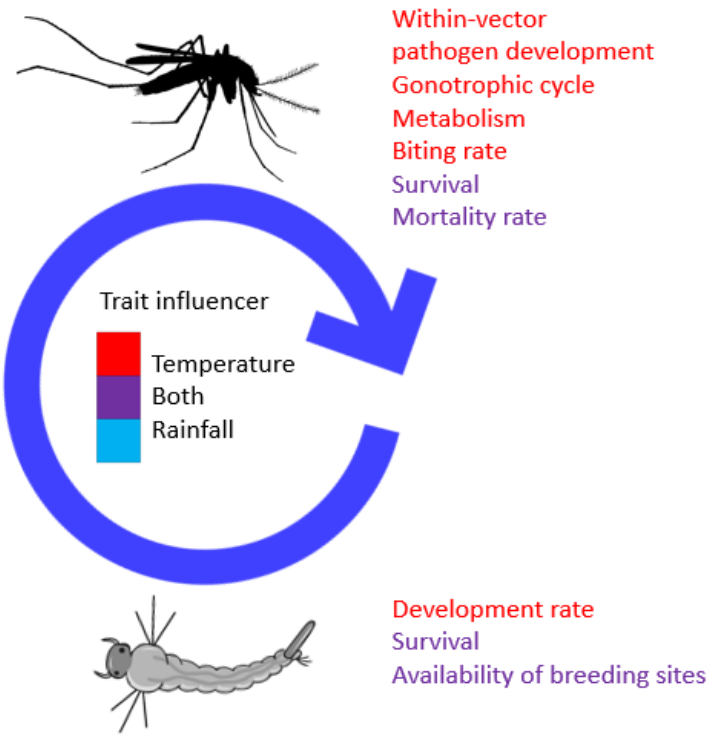


Figure 1.10.1. Examples of the mosquito life histories affected by the temperature, rainfall and their interaction.

While static models of vector-borne disease transmission offer an important snapshot of the average transmissibility, they may fail to highlight regions suitable for transmission during certain periods of the year only. In areas demarcated as inhospitable using annual means, periods of optimal conditions could allow for the seasonal transmission of numerous vector-borne diseases. This has been observed in Southern Europe, which has recorded autochthonous, locally acquired, cases of Dengue, West Nile and Chikungunya (Queyriaux et al., 2008, Bagnarelli et al., 2011). Even within endemic areas, seasonal fluctuations in disease transmission are observed (Hoshen and Morse, 2004, Kumm, 1950, Do et al., 2014).

Fluctuations in temperature and rainfall can increase or decrease an area's suitability for transmission through modifying vector population sizes (Kumm, 1950, Wee et al., 2013), metabolisms (Lambrechts et al., 2011, Johansson et al., 2010) or behaviour

(Paaijmans and Thomas, 2011). Furthermore, seasonal changes in human behaviour may facilitate disease transmission (Gardner and Ryman, 2010). Establishing and quantifying the relationship between transmission and temporally varying covariates is an important step in accurately establishing both the geographic extent of transmission risk and highlighting periods of heightened activity. Such results are important for forecasting disease occurrence, and optimising intervention strategies (Cairns et al., 2012, Johansson et al., 2009).

1.11 Ecological niche modelling

Mapping the occurrence of disease has been an important step in combating transmission for hundreds of years (Koch and Denike, 2009), and while the importance has not changed, the techniques have (Carpenter, 2011). Establishing the presence of vector-borne diseases has moved away from simple spatial frameworks, which fail to account for poor data and biological realities, and towards geospatial models (Escobar and Craft, 2016). These are based on the theory that covariates influencing disease transmission are heterogeneously located in areas where disease transmission occurs and where it is absent. Geospatial models have been used to great effect in establishing the distribution, prevalence and burden of numerous vector-borne diseases such as Malaria (Hay et al., 2009), Japanese encephalitis (Miller et al., 2012), Dengue (Bhatt et al., 2013) and YF (Garske et al., 2014, Hamrick et al., 2017, Rogers et al., 2006).

By utilising the environmental and climatic constraints of vectors and observed disease occurrence, we are able to determine the potential distribution and prevalence of vector-borne diseases through ecological niche models (Peterson, 2008), see Figure 1.11.1. These models correlate spatially explicit presence data with corresponding

environmental covariates to predict potential distribution. By doing so these techniques have an advantage over spatial interpolation methods, as spatial interpolation models may attribute low risk to an area without data (Escobar and Craft, 2016). This is particularly important when modelling diseases which are found in resource poor settings, or in areas with weak disease surveillance systems, both of which are true for YF. Niche models avoid this issue by characterising covariate values and relationships in areas of confirmed disease occurrence and subsequently identifying similar combinations (Escobar and Craft, 2016).

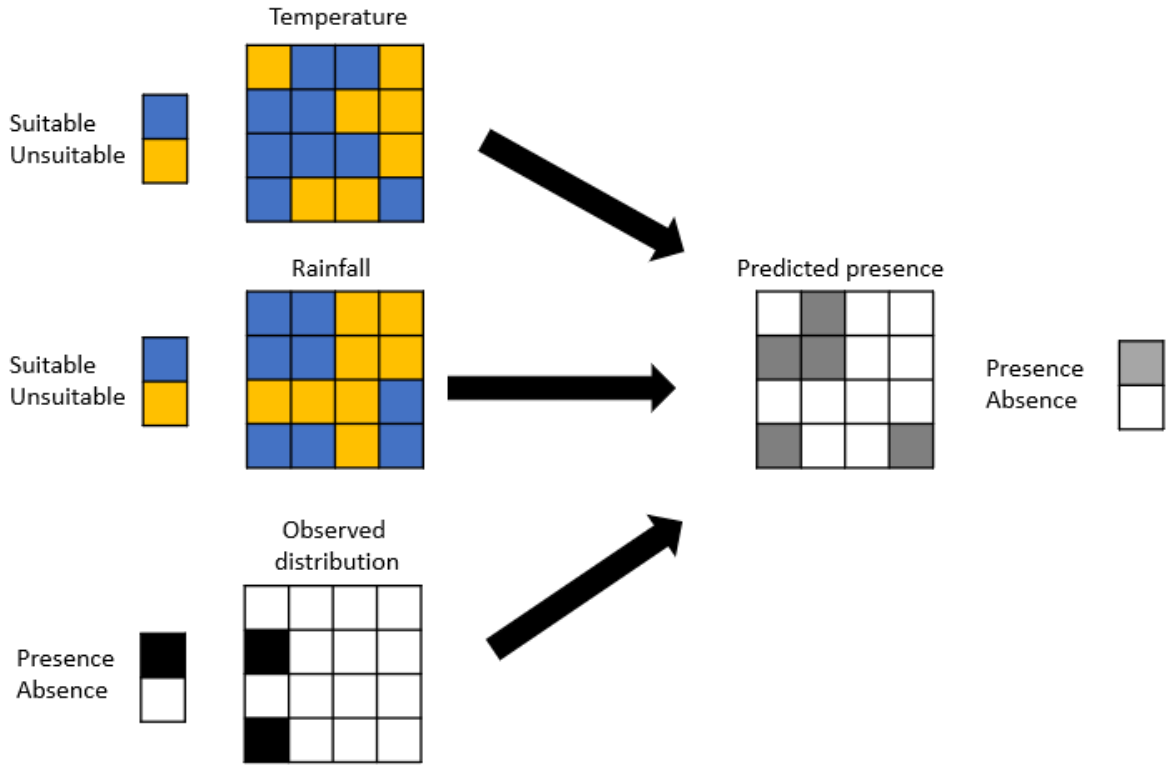


Figure 1.11.1. A simplified illustration of environmental niche modelling with 2 climate variables and observed distribution of species. Adapted from Van et al., (2010) (Van et al., 2010).

The utilisation of presence data can be problematic if true presence and absence are not accurately recorded. Improper classifications of presence, due to cross-reactive

diagnostic tests of imported cases, and absence, due to insufficient surveillance, may incorrectly portray a pathogen's niche. In order to disentangle issues with detection and true presence/absence, absence or pseudo-absence points can be used in addition to presence points. However, these are often of dubious quality in resource poor settings which lack adequate surveillance or diagnosis (Peterson et al., 2014). Incorrect descriptions of presence may lead to a clustering of reports around diagnostic centres, rather than where the pathogen is transmitted, leading to an inaccurate description of its niche (Auchincloss et al., 2012). Another caveat of this modelling technique is that it is important to exercise restraint in covariate selection. The inappropriate and overutilization of covariates in niche models can lead to overfitting and thereby reduce the predictive performance of the model (Escobar and Craft, 2016, Estrada-Pena et al., 2015). Furthermore, it is important to have a sufficient number of presence points to accurately quantify the covariate relationships and, generally, the addition of presence points creates a more accurate model. However, it is important to balance the need for quantity with the need for quality. The inclusion of points which are poorly located, or over-represent particular conditions could lead to predictions reflecting bias in surveillance rather than the true niche (Escobar and Craft, 2016).

Despite these provisions, when applied correctly and to data of a sufficient quality, ecological niche modelling is a relevant and useful tool for determining the potential distributions of vector-borne disease.

1.12 Mechanistic modelling of disease transmission

Mathematical modelling, the use of a set of dynamical equations in order to mechanistically explain systems has significant utility in understanding infectious

disease. By explaining the spread of disease through mathematical equations, modelling can offer insights into such questions as

1. The mechanisms of disease transmission.
2. The likely future severity and duration of an epidemic.
3. The potential for sustained endemic transmission following introduction of disease.
4. The impact and cost of different control measures on disease burden.

These are questions that may not be able to be answered by traditional experimental or statistical approaches due to a financial, logistical or ethical constraint, as well as the inability to accurately simulate large outbreaks of disease in a laboratory setting or conduct repeated experiments during outbreak settings (Li, 2010). Furthermore, with statistical models it is very difficult to capture non-linear, rapidly changing dynamics as well as threshold behaviour in terms of the basic reproduction number and critical community size. Rapid changes in R0 can lead to substantially different outcomes in epidemic diseases such as YF. Models of disease transmission have been utilised in responding to numerous outbreaks of disease such as HIV, bovine spongiform encephalopathy (BSE), SARS, pandemic influenza and Ebola (Heesterbeek et al., 2015, Team, 2015b, Aylward et al., 2014).

While these models offer a highly useful and impactful avenue of research in the study of infectious disease, their use and utility should be carefully considered. Due to limitations in knowledge of infection and transmission, assumptions and simplification are often made in modelling disease. While these are often inevitable, and accurate simulations of the transmission can still be achieved in their absence, it is important to take these into account when constructing and interpreting model findings. With these considerations, it is often more useful to have a simple model with clearly understood

limitations, than a complicated model which replicates more aspects of transmission but has behaviour which is impossible to understand. In order to produce a modelling framework with applicable findings to policy makers, it is important to strike a balance between under and oversimplification. Furthermore, models are only useful given the correct approach, to useful data in the right context. Modelling of disease should be about optimising the data available, not substituting for it.

When these caveats are taken into consideration, the potential benefits of modelling far outweigh its limitations. This is reflected in its continuing, and expanding, role as an important tool for furthering our knowledge of disease transmission processes and helping inform those involved in public health policy (Calder et al., 2018).

1.13 Previous modelling of YF

While mathematical and statistical modelling have had a wide and varied application to numerous arboviruses (Bhatt et al., 2013, Morin et al., 2015, Althouse et al., 2015), their application to YF has been relatively minor.

The first, published, application of geospatial modelling applied to YF was Rogers et al. (2006) (Rogers et al., 2006), aimed to capture the probability of suitability for YF, and dengue, using publicly available presence data and satellite derived environmental data. The findings of this study suggested substantial heterogeneity in transmission across, as well as predicting additional areas of high suitability outside, the contemporary boundaries of the endemic zone it also highlighted the influence of specific environmental covariates such as the Normalized Difference vegetation Index (NDVI). The next significant application came in 2012, when the CDC utilised mathematical modelling to estimate the probability of the global spread of YFV from

the 2008 outbreak of YF in Asuncion, Paraguay, that signalled the return of urban YF to South America – previously absent since 1940 (Johansson et al., 2012).

In 2014, that Garske et al., (2014) conducted the first systematic and statistically rigorous estimate of the burden of YF (Garske et al., 2014). In addition to providing estimates of the YF burden in Africa, this research was conducted in order to evaluate the impact of mass vaccination campaigns conducted as part of the Yellow Fever Initiative (YFI). Garske et al., (2014) estimated that 130,000 severe cases of YF occurred annually across Africa, leading to 78,000 deaths. Furthermore, they estimated that YF vaccination campaigns had reduced the burden of YF in Africa by 27%, and by up to 82% in the countries targeted by the YFI.

Following this, relatively few studies were published until the 2015-2016 outbreak of YF in Angola and the Democratic Republic of the Congo, the largest seen in decades and involving rapid, large-scale urban transmission of the virus – including exportation of cases inter-nationally and outside of the endemic zone (World Health Organization, 2017b). This outbreak also saw the WHO sanction the use of “fractionated doses” of the YF vaccine for outbreak response (World Health Organization, 2017a). Following this outbreak there was a flurry of modelling exercises, looking at the impact of fractional dosing (Wu et al., 2016), the spread of the virus during the outbreak and the impact of vaccination (Kraemer et al., 2017, Zhao et al., 2018). These developments, in conjunction with the largest outbreak Brazil had seen since 1940 occurring in 2017, and continuing till 2019, led to a continued interest and generation of important papers. These included those evaluating the seasonality (Hamlet et al., 2018), the international risk of spread from the Brazilian outbreak (Dorigatti et al., 2017, Brent et al., 2018, Sakamoto et al., 2018), the regional and global suitability for transmission (Hamrick et

al., 2017, Shearer et al., 2018, de Almeida et al., 2019) and those on the estimated population-level vaccination coverage (Shearer et al., 2017, Hamlet et al., 2019).

Despite this recent upsurge in modelling efforts, aimed at diverse aspects of YF epidemiology, there remain many avenues of research where modelling could provide substantial input. Further analysis of the potential for spread of YFV, the role of NHP's in epidemic and endemic transmission, the association of landcover and land conversion with YFV are among numerous avenues that should be further investigated.

1.14 Structure of the PhD

This body of work is divided into five main parts.

In chapter 2, I aim to quantify the statistical associates of YF transmission and environmental, climate and host demographics through the use of a series of algorithms, which are evaluated for relative predictive accuracy, for modelling the presence and absence of YF across South America (2003-2016).

Chapter 3 builds on the lessons learnt from chapter 2 but seeks to additionally explain temporal heterogeneity, as well as geographic, in both seasonal and inter-annual patterns of YF reporting across the time period through employing negative binomial regressions fit to the number of months reporting YF.

Chapter 4 takes an in-depth look at the role of seasonality for reporting confirmed YF in both humans and NHPs. By focusing on Brazil at the 2nd administrative division, and the relative roles of the seasonality of vegetation/climate and agriculture (planting/harvesting), which potentially represents exposure, I seek to disentangle further intricacies in the dynamics of sylvatic spillover.

Chapter 5 describes the collation of vaccination coverage activities in Africa and South America from 1940 to 2050, which were applied to a demographic model to generate age disaggregated, population-level estimates of YF vaccination coverage. Furthermore, these estimates were published in a public, interactive, online tool to increase their utility and aid interpretation for those without specialist programming knowledge or software called POLICI.

Chapter 6 focuses on the ecological niche modelling of various NHP genera in the South East Atlantic forest of Brazil, with the aim of classifying areas of presence and predicting densities for each genus. This information is used as the basis for the "realistic" population data input into the model described in chapter 7.

Chapter 7 depicts the development of a stochastic meta-population model for YFV transmission in 4 genera of NHP's in Brazils South-East Atlantic states. By attempting to replicate the underlying ecology of YFV in a realistic environment, I seek to understand the critical community size of YFV, the role of different NHP's on the maintenance of transmission and the relationship of CFR and R0. The model will then be applied to the landscape of South-East Brazil, using satellite derived data and realistic NHP sizes to estimate the probability of the establishment of endemicity following the introduction of YFV in the 2016-2019 outbreaks in the South-East Atlantic States.

Chapter 2 - The role of climate, environment and host in the presence of YF across South America

2.1 Summary

While YF represents a significant public health threat across tropical South America, many of the environmental and climatic drivers are unknown.

Here I have collated information on YF reports across South America (2003-2016) in order to assess the association of several environmental, climatic and host demographic covariates with these reports. This is carried out initially using a variety of algorithmic techniques, before the best performing technique, generalised linear models (GLM), are used to create an ensemble model of YF prediction across South America.

Model findings depict large heterogeneity in the distribution of YF risk across South America, with much of Peru, Bolivia, Colombia, Venezuela and Brazil suitable for the occurrence of YF. Additionally, discrepancies between model findings and recorded data highlight areas at potential risk for YF transmission, in the absence of currently reported YF.

These findings provide a geographical framework for surveillance of YF. By highlighting areas at risk of potential occurrence, and the climatic and environmental factors associated with transmission, proactive surveillance and vaccination activities can be targeted at areas at highest risk of transmission.

2.2 Introduction

The use of geospatial modelling of disease has long been used to highlight regions of risk and derive a greater understanding of the climatic and environmental correlates of disease transmission (Bhatt et al., 2013, Allen et al., 2017, Shearer et al., 2018). While these techniques have been used to some success in modelling YF, they have primarily been applied to climatic, vector or host explanations for presence of YFV, with minimal exploration of the relationship of landcover and disease spillover (Shearer et al., 2018, Hamrick et al., 2017). Landcover type, and changes in landcover, have been implicated in changes in disease occurrence, particularly of zoonotic infections (Patz et al., 2004, Faust et al., 2018, Burkett-Cadena and Vittor, 2018). While the exact mechanism is unknown, this is potentially due to increased interactions between sylvatic reservoirs and humans, as well as changes in vector dynamics (Burkett-Cadena and Vittor, 2018).

In this study I assessed the in- and out-of-sample predictive accuracy of three modelling algorithms, generalised linear regression (GLM), random forest (RF) and boosted regression trees (BRT). Following this, I have utilised the best performing algorithm to fit all permutations of covariates to predict the presence/absence of YF across South America. The best performing models were aggregated into an ensemble model, the predictions of which were mapped across South America, and the contributing covariates analysed. I then used these models to investigate and quantify the relative influences of covariates and highlight areas of potentially heightened YF transmission based on their environment and climate.

2.3 Methods

1.1.1 Reports of YF

In order to investigate the relationship between covariates and human YF, I had to collate reports of YF. Here, reports of YF cases in humans were assembled from various World Health Organization sources published in the Weekly Epidemiological Record (World Health Organization), Disease Outbreak News (World Health Organization), and the Pan American Health Organization (Pan American Health Organization, 2017). While these sources differ in the exact detail of their reporting, generally they generally report the number of cases and the range of months where they occurred, often only at the 1st administrative division in a country. Annual summaries of the number of YF cases in Africa and South America are reported in the Weekly Epidemiological Record, the following year which provide the number of cases reported, but often not the month they were reported in.

Cases reported in these sources are occasionally laboratory confirmed, though due to issues relating to the capacity for formal diagnosis the majority of YF cases are probable rather than confirmed (Waggoner et al., 2018). The WHO definition of a probable YF case is that of a patient with an acute febrile illness that develops jaundice within 14 days of symptom onset, and with the presence of YF IgM antibody in the absence of YF immunisation, or a positive post-mortem liver histopathology or an epidemiological link to a confirmed case or outbreak (World Health Organization, 2010a). While the inclusion of probable rather than purely confirmed cases may add false positives to the dataset which in turn may lead to a weakened or erroneous association of YF reporting and covariates, the relative sparsity of confirmed cases compared to probable cases was used as an argument in favour of this approach.

In this initial chapter I was investigating the presence and absence of YF probably/confirmed reports in relation to climate, environmental and host covariates without any temporal relationship for any report during the period 2003-2016. However, as in subsequent chapters I will be looking at these temporal trends, only reports where the month of symptom onset was recorded were included (823 of the original 1073 reports) to homogenise datasets. Furthermore, only reports that could be geo-located to the first sub-national administrative level, here termed province were considered.

2.3.1 Covariates

In total, 19 covariates were considered (Table 2). These were selected based on knowledge of the biology and distributions of vector species, host dynamics, inferences from the role of land-cover change and vegetation heterogeneity and the epidemiology of yellow fever in South America (Kay et al., 1997, Hamrick et al., 2017, Hamlet et al., 2018, Alencar et al., 2010).

The average values of these covariates over the time period (2003-2016, chosen based on the availability of data) were utilised. Here I have initially looked at only the presence/absence of a report over the time-period and have not considered the roles of seasonality or inter-annual variation. While this approach will de-emphasise the role of abnormal weather patterns and climate cycles such as El Niño Southern Oscillation, which have been found to affect vector-borne disease transmission (Fuller et al., 2009, Anyamba et al., 2001), and the long-term roles of landcover conversion (Burkett-Cadena and Vittor, 2018), it provides an insight into the general suitability of provinces for the transmission of YFV. Furthermore, it provides a “baseline”, to which future

investigations of temporal variations in reporting (intra- and inter-annually) can be compared.

2.3.2 Host population data

Country and year specific human population sizes were obtained from the UN World Population Prospects (WPP) (United Nations, 2016) and averaged over the study period to obtain average population sizes. These were then used to produce province-level estimates of population were obtained by disaggregating this data by using LandScan 2015 (Bright et al., 2016) population estimates with a 1/120 latitude and longitude degree resolution to calculate the proportion of the national population within each province. This approach to use the disaggregated UN WPP data, rather than just LandScan 2015, was chosen based on previous publications looking at YF transmission (Garske et al., 2014, Hamlet et al., 2018). The mean natural logarithm of human population over the time-period was used. In addition, the relative change in the human population over the 14-year time period was also used as a covariate (defined as the natural logarithm (final population/initial population)). These covariates were included by default in all models in order to account for the previously linked relationship between population size and YF surveillance (Garske et al., 2014).

Data on NHP species distribution were obtained through distribution maps of mammals in the western hemisphere (Patterson et al., 2007). This data was available as shapefiles of presence/absence of a species, which were geo-located to the province level. This was used to calculate the number of NHP species present in each province. This covariate was chosen based on a known importance of NHPs in the maintenance of the sylvatic cycle of YF.

2.3.3 Climate, vegetation and vegetation heterogeneity data

Datasets, 2003-2016, for temperature (Garske et al., 2013), enhanced vegetation index (EVI) (NASA) and rainfall (Joyce et al., 2004) were aggregated to the administrative unit 1 level from their original resolutions (of between 1/120 and 1/12 degree) by calculating population-weighted means, based on the population distribution provided by LandScan 2015 (Bright et al., 2016). These covariates were chosen based on their previous associations with YF (Garske et al., 2014, Hamlet et al., 2018, Rogers et al., 2006), and other vector-borne diseases (Huber et al., 2018, Do et al., 2014, Garske et al., 2013). Spatial heterogeneity in vegetation was assessed by evaluating the standard deviation of EVI at its original 1/120 degree resolution within an administrative unit. The average amplitude of these variables over a year (over the time period) was also calculated in order to account for the impact of seasonal variation.

2.3.4 Land-cover data

Land-cover was provided by the MODIS dataset (Friedl et al., 2010), which characterises the dominant land-cover type, at a grid resolution of 0.8333° globally.

Here I looked at several different types of landcover type based on their previous associations with YF or vector-borne diseases, such as forest and savanna cover where the sylvatic reservoir of YF is located (Monath and Vasconcelos, 2015), potential sites of interaction between humans and the sylvatic cycle in cropland/natural vegetation mosaic and cropland cover (Shah et al., 2019), and urban areas with higher human populations (Hamrick et al., 2017). This information was aggregated to the province level and the proportion of the province area occupied by each land-cover type calculated.

In order to additionally assess how these changes in land-cover over time may be associated with YF transmission, I looked at the temporal change by taking the values of the first year were subtracted from the final year of study (2003 and 2016 respectively). Previous research has shown that habitat transformation is often associated with zoonotic disease occurrence (Allen et al., 2017, Faust et al., 2018).

Table 2.3.1. Table of classification types and the covariates included.

COVARIATE	CLASSIFICATION	DESCRIPTION	REFERENCE
NHP SPECIES	Host population	The number of monkey species	(Patterson et al., 2007)
LOGARITHM OF HUMAN POPULATION	Host population	The logarithm of human population	(United Nations, 2016)
CHANGE IN LOGARITHM OF HUMAN POPULATION		The change in the logarithm of human population from 2003 to 2016	
ENHANCED VEGETATION INDEX (EVI)	Vegetation	A vegetation index designed to improve sensitivity in high biomass regions	(NASA)
EVI AMPLITUDE	Vegetation	The amplitude of the EVI	
DAY TEMPERATURE	Climate	The day temperature	(Garske et al., 2013)
DAY TEMPERATURE AMPLITUDE	Climate	The amplitude of the day temperature	
RAINFALL	Climate	The rainfall	(Joyce et al., 2004)
RAINFALL AMPLITUDE	Climate	The amplitude of the rainfall	
FOREST COVER	Land-cover	The proportion of the administrative unit covered by any forest type	(NASA)

SAVANNA COVER	Land-cover	The proportion of the administrative unit covered by any savanna type
CROPLAND/NATURAL VEGETATION MOSAIC COVER	Land-cover	The proportion of the administrative unit covered by cropland/natural vegetation mosaic
URBAN COVER	Land-cover	The proportion of the administrative unit covered by urban areas
CROPLAND COVER	Land-cover	The proportion of the administrative unit covered by cropland
FOREST COVER TEMPORAL CHANGE	Land-cover change	The final year (2016) – the first year (2003) landcover value
SAVANNA COVER TEMPORAL CHANGE	Land-cover temporal change	
NATURAL VEGETATION/CROPLAND MOSAIC COVER TEMPORAL CHANGE	Land-cover temporal change	
URBAN COVER TEMPORAL CHANGE	Land-cover temporal change	
CROPLAND COVER TEMPORAL CHANGE	Land-cover temporal change	
VEGETATION HETEROGENEITY	Vegetation heterogeneity	This is the standard deviation of the EVI at a 1x1km resolution within the administrative unit
VEGETATION HETEROGENEITY TEMPORAL CHANGE	Vegetation heterogeneity temporal change	The final year (2016) – the first year (2003) vegetation heterogeneity

Covariates were standardised to facilitate comparison through the following formula,

$$z = \frac{x - \mu}{\sigma},$$

where z, is the standardised value, x, the pre-standardised value, μ, the mean of the pre-standardised values and σ, the standard deviation of the values.

2.3.5 Models and modelling algorithms

Following initial covariate selection and processing, a list of covariates identified as relevant to YFV transmission were considered, with log of human population and the fractional change in logarithm human populations included in every model, as previously discussed. An exhaustive combination of all 19 covariates was explored.

These were fit using 3 techniques, GLM, RF and BRT.

GLM models have long been used in ecology and the modelling of disease in order to map the presence/absence (Peterson, 2008), and work by establishing a statistical relationship between a dependent variable (presence or absence) and a series of independent covariates. Here I have utilised logistic regression which models the binary outcome of presence/absence on the log-odds scale to determine the probability (0-1) of presence given the covariates.

RF and BRT algorithms are machine learning (ML) methods which are similar to traditional statistical regression methods in that they use covariates to explain patterns in the data, but instead rely on the application of decision trees to explore non-linear relationships with the outcome and the data. Decision trees represent an exploration of the data where a series of tests (nodes) are applied to data, and the outcomes of these tests (branches) are associated with an outcome (Figure 2.3.1). These are utilised in similar, but differing, approaches for RF and BRT.

RFs exploit decision trees by randomly selecting covariates at each node, which are then used to split the data and the best split calculated with this subset, with each tree independent of the others. This is repeated over many decision trees, a forest, and the

mean values taken in order to produce an overall model that is applied to the data. Here trees are created independently of each other, and as such reduces the likelihood of overfitting (Genuer et al., 2008). BRT follows a similar approach, but differs in that it produces the trees sequentially – with each new tree building on the findings of the previous model, an approach which may improve model accuracy – at the cost of potential overfitting (Elith et al., 2008).

In addition to accounting for nonlinearity, both RF and BRT methods have been shown to have substantial improvements in predictive accuracy over traditionally GLM (Elith et al., 2008, Marmion et al., 2009). Model fit was measured using the area under the receiver operating characteristic curve (AUC), a commonly used metric for assessing the predictive accuracy of models which predict a binary outcome (Marmion et al., 2009).

All data processing and analysis was carried out using the statistical programming language R (Team, 2015a), GLM models implemented using the package glmmTMB (Brooks et al., 2017), RF the package ranger (Wright and Ziegler, 2017) and BRT, the package gbm (Greenwell et al., 2019) .

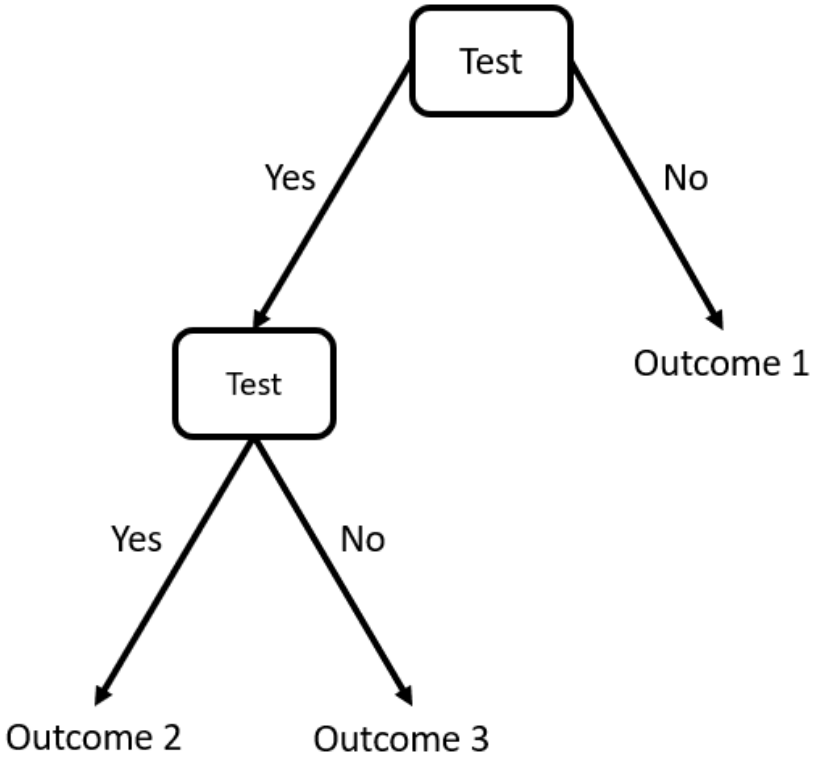


Figure 2.3.1. Simple example of a decision tree.

2.3.6 Out-of-sample validation

To assess the out-of-sample predictive ability of our models I carried out a form of out-of-sample validation called spatial block bootstrapping. The best performing models in-sample as defined by those with an in-sample area under the curve (AUC) value above the 95% quintile, for each of the GLM, RF and BRT models were tested using this spatial block bootstrapping technique.

This was done by overlaying a grid of 5° x 5° longitude of latitude over the study area and assigning provinces to a point based on their centroid coordinates. Following this, random sampling with replacement was used to build a training set of 60-70% of the points that contain an assigned province, the remaining unselected points were assigned to the validation dataset. This was repeated 200 times to generate 200 training sets and 200 validation sets.

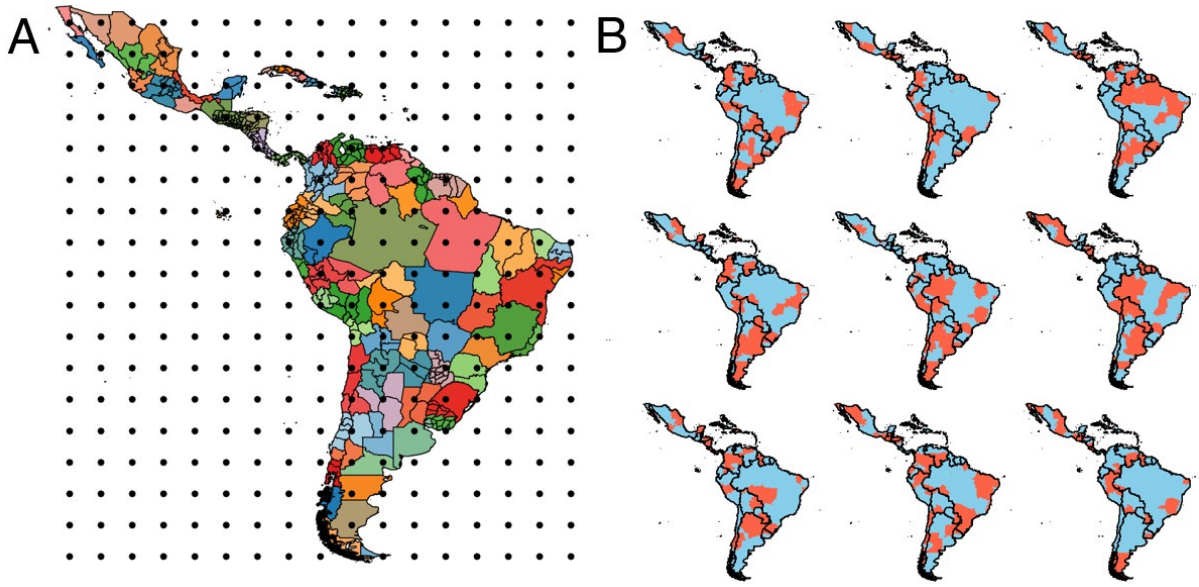


Figure 2.3.2. The grid of 5° x 5° longitude of latitude with provinces assigned and colour coded by the grid point closest to their centroid coordinates. B) Examples of the training (blue) and validation (red) datasets as chosen by random sampling of grid points.

2.3.7 GLM ensemble model calculation

The best performing models fit with GLM's were ranked based on their Akaike Information Criterion (AIC) and those with an AIC within 3 of the best performing model, as defined as the model with the lowest AIC value, were combined using Akaike weights (Wagenmakers and Farrell, 2004). To do so, the relative differences in AIC are calculated by,

$$\Delta_i = AIC_i - \min(AIC) \; [1]$$

and this is used to obtain an estimated relative likelihood of model, *i*, in proportion to the other models, *k = 1 …K*, included through,

$$w_i = \frac{exp\left\{-\frac{1}{2}\Delta_i\right\}}{\sum_{k=1}^{K} exp\left\{-\frac{1}{2}\Delta_k\right\}} \; [2].$$

These product of each of these model specific weights, w_i, and their corresponding model specific predicted values, p_i, are summed to generate a single set of weighted predictions, p_A,

$$p_A = \sum_{k=1}^{K} w_k p_k \ [3].$$

These model specific weights are additionally used to create the ensemble coefficient values, c_A, through,

$$c_A = \sum_{k=1}^{K} w_k c_k \ [4],$$

where c_k refers to the model coefficient values.

2.4 Results

2.4.1 In and out sample performance of different techniques

Through considering all combination of the 19 covariates that varied in selection, 524,288 models were produced.

Figure 2.4.1 shows the maximum in and out sample AUC values of the best performing models (as defined as the top 5% of models by in sample AUC). In-sample, the BRT model out-performs the GLM and RF models, however when tested out of sample using the spatial block bootstrapping cross validation method, both the GLM and RF implementations outperform the BRT models, substantially but with overlapping confidence intervals of the out-of-sample AUC.

The superiority of the GLM models in out of sample prediction, an indication of true predictive performance, led to their utilisation over BRT and RF models in further stages of analysis.

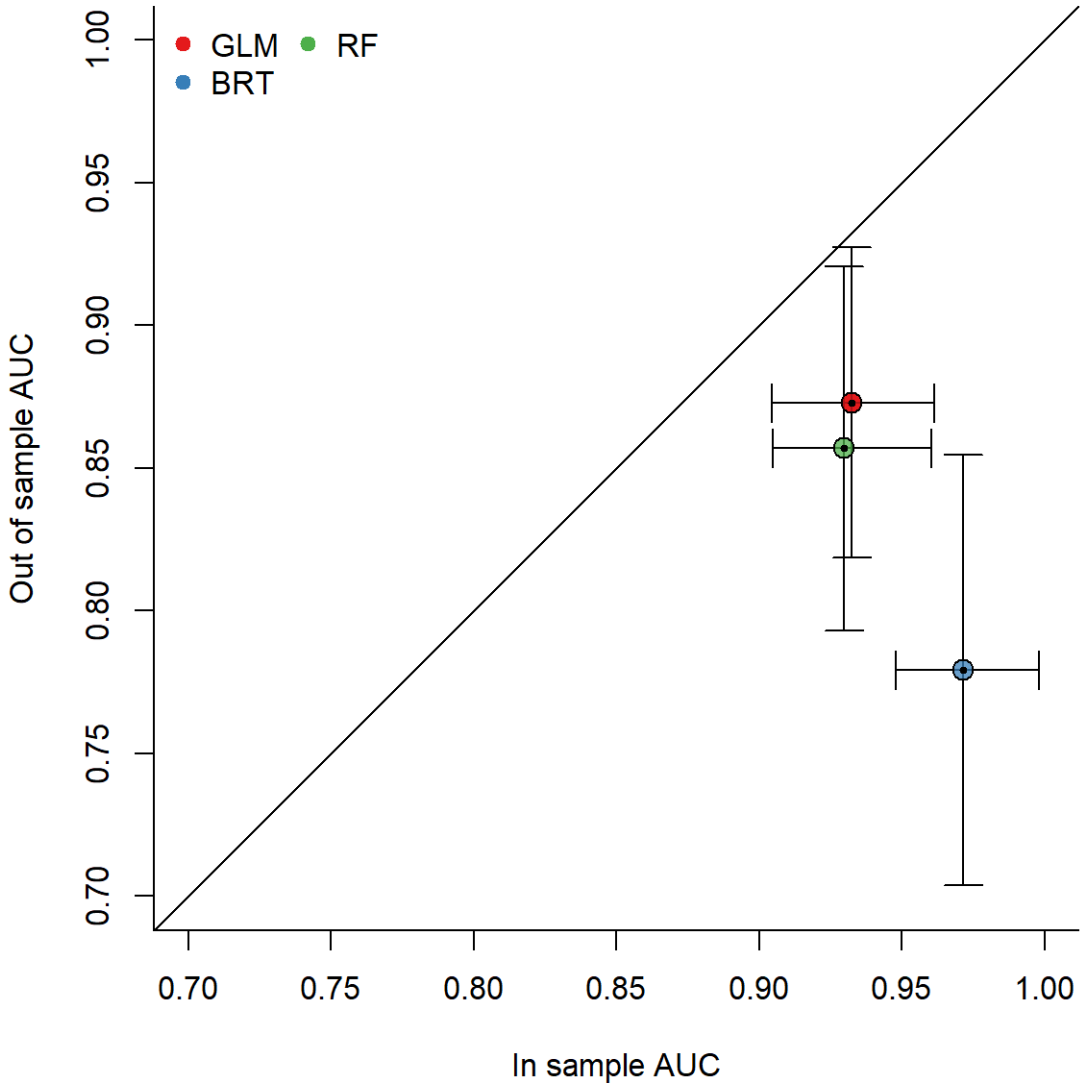


Figure 2.4.1. Comparison of the maximum in sample and out of sample area under the curve (AUC) values for models fit with generalised linear models (GLM), random forest (RF) and boosted regression trees (BRT). Horizontal and vertical bars represent the 95% confidence intervals of the in and out of sample AUC values respectively.

2.4.2 Model predictions and mapping

Our literature search and subsequent restriction based on geo-location and availability of month of report found 823 reports during the time-period (2003-2016) which related to 49 unique provinces (Figure 2.4.2A).

YF reports show a high degree of heterogeneity in the countries they are found in, with the percentage of provinces in a country reporting YF ranging from 4.16% in Ecuador and Argentina, to 42.3% in Peru and 55.6% in Bolivia. Reports primarily concentrate in Amazonian regions, East and South of the Andes in Colombia, Peru and Bolivia, and throughout much of the country, except the North-East, in Brazil.

Ensemble model predictions reproduce the data to a high degree of accuracy (in sample AUC 0.929 (95% CI: 0.899 – 0.959), and out sample 0.815 (95% CI: 0.754 – 0.875)). Regions of high report probability are found around Amazonian and adjacent forested regions, and the South-East coast of Brazil (Figure 2.4.2B). In Peru and Bolivia, high predictions of YF reports around throughout much of the country East of the Andes. Colombia shows relatively low predictions in Amazonian regions, with most of the predicted distribution focused on the North-Eastern Pacific coast.

While model predictions accurately capture the general picture of YF distribution across South America, there are discrepancies between the data and model predictions (Figure 2.4.2C). These are most notable in Brazil, Colombia and Venezuela. In Brazil, the South-east of Brazil is substantially underpredicted, apart from the states of Rio de Janeiro and Bahia – which are overpredicted. Colombia shows large variation in over and under-prediction across the country, which is mirrored in Venezuelan states which border Colombia.

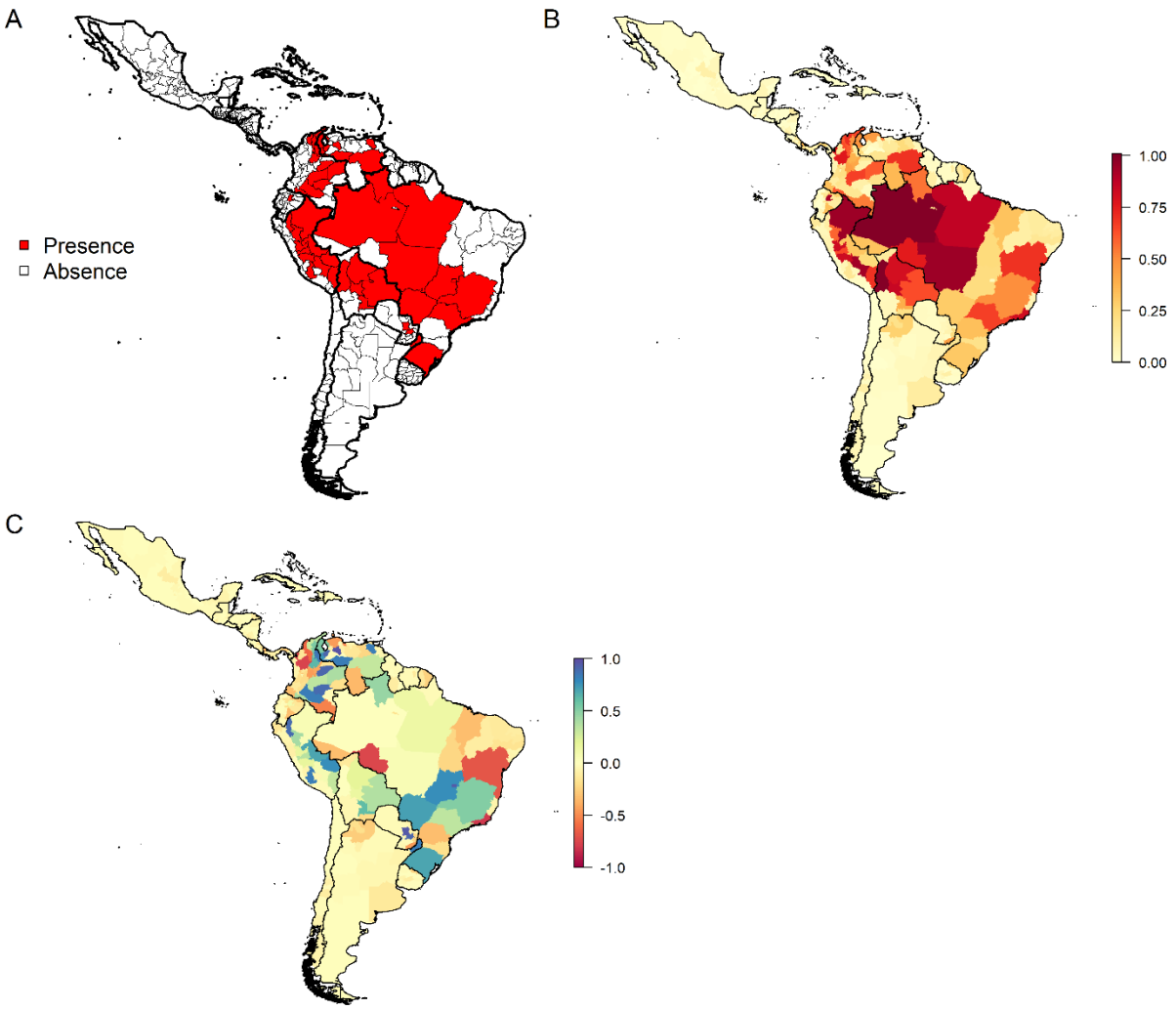


Figure 2.4.2. (A) The presence/absence of YF reports (2003-2016), (B) ensemble model predictions of the presence/absence of YF, (C) the difference in the data and model predictions.

2.4.3 Covariate influence and significance

Several covariates were found to have a significant and substantial impact of ensemble model predictions, though the majority of covariates were found to be insignificant.

Savanna cover and the temporal change in urban cover were found to have a significant, protective relationship against the probability of a YF report, with odds ratios (OR) of 0.276 (95% CI: 0.110 – 0.693) and 0.378 (0.147 – 0.968) respectively. These covariates were also found in all of the 275 models that contributed to the ensemble model. The number of NHP species was found to increase the likelihood of a YF report, with an OR of 1.973 (1.388 – 2.806), along with rainfall, 5.504 (2.128 – 14.237), and most substantially, the vegetation heterogeneity at 9.730 (3.912 – 24.197).

While other covariates were present in the best performing models, their effects were not found to be significant – even in the case of some covariates found in the majority of best performing models such as cropland and forest cover.

Table 2.4.1. Scaled ORs of all covariates tested in the models, along with the % of models the covariate was found in that contributed to the ensemble model, and whether the value was significant. Significant covariates are emboldened.

Covariate	**Odds Ratio**	**% found in the best performing models**	**Significant**
Savanna cover	**0.276 (95% CI: 0.110 - 0.693)**	**100**	**Yes**
Urban cover temporal difference	**0.378 (95% CI: 0.147 - 0.968)**	**100**	**Yes**
Number of NHP species	**1.973 (95% CI: 1.388 - 2.806)**	**100**	**Yes**
Rainfall	**5.504 (95% CI: 2.128 - 14.237)**	**100**	**Yes**
Vegetation heterogeneity	**9.73 (95% CI: 3.912 - 24.197)**	**100**	**Yes**
Log of human population	32.858 (95% CI: 0.907 - 1190.574)	100	No
Fractional change in population	736.762 (95% CI: 0.534 - 1016637.897)	100	No
Cropland cover	0.451 (95% CI: 0.178 - 1.145)	90.2	No
Forest cover	0.753 (95% CI: 0.535 - 1.060)	64.7	No
EVI	3.414 (95% CI: 0.741 - 15.732)	59.3	No
Forest cover temporal difference	1.278 (95% CI: 0.914 - 1.788)	55.3	No
Cropland cover temporal difference	1.571 (95% CI: 0.901 - 2.737)	54.2	No
Cropland/natural vegetation mosaic cover	0.816 (95% CI: 0.584 - 1.14)	46.5	No
Savanna cover temporal difference	0.016 (95% CI: 0.000 - 4.565)	42.5	No
Urban cover	0.781 (95% CI: 0.506 - 1.203)	21.1	No
Day temperature amplitude	0.985 (95% CI: 0.945 - 1.026)	13.8	No
EVI amplitude	0.357 (95% CI: 0.021 - 6.106)	12.7	No
Rainfall amplitude	0.985 (95% CI: 0.946 - 1.026)	12.7	No
Day temperature	1.001 (95% CI: 0.995 - 1.008)	8	No
Cropland/natural vegetation mosaic cover temporal difference	0.993 (95% CI: 0.041 - 24.043)	6.5	No
Vegetation heterogeneity temporal difference	0.999 (95% CI: 0.986 - 1.013)	5.8	No

2.5 Discussion

Here I have assessed the relative in and out sample predictive performance, using the AUC values, of various modelling algorithms applied to the presence and absence of YF reports across South America in 2003-2016. I have shown that in this circumstance, GLM models allow for greater predictive performance compared to other techniques considered. Following this finding, I have combined the best performing GLM models, as defined by their AIC value, into an ensemble model that predicts the distribution of YF reports across the region to a high degree of accuracy, with an in sample AUC of 0.929 (95% CI: 0.899 – 0.959), and out sample of 0.815 (95% CI: 0.754 – 0.875)). Differences between the predictions and data also highlight regions that appear highly suitable during this time period (Figure 2.4.2).

While the use of machine learning algorithms such as RF and BRTs allow for non-linear associations between the response and the covariates, as well as considering interactions, their under-performance in out of sample predictive ability compared to GLM led to their discontinuation from further analysis. This finding is at odds with several previously published research papers, which have found these machine learning techniques to substantially improve accuracy (Biau, 2012, Olden et al., 2008). However, this accuracy is often assessed in sample, and where cross-validation is utilised it does not consider spatial bias to a high enough extent. Administrative units that are geographically related are likely to share climatic and environmental covariates. By utilising the spatial-block bootstrapping cross validation method to produce spatially dissimilar training and validation datasets, I am more accurately assessing a model and algorithms out of sample predictive performance. The finding, that in this specific modelling exercise, that GLMs offers a better out of sample

predictive performance over more “sophisticated” machine learning techniques is one of worthwhile note, and that the data in question should be carefully considered when selecting modelling algorithms, rather than simply defaulting to the newest techniques.

This body of work provides insight into the geographic distribution of YF across South America, with findings highlighting substantial heterogeneity in transmission potential and many regions of the traditional “endemic zone of YF” have been highlighted, such as Amazonian regions of Brazil, Peru, Bolivia and Venezuela. In addition to these expected zones, this model has also highlighted much of the densely population South Eastern States of Brazil, which until 2017 had not experienced widespread transmission of YF. This suggests that the ecological conditions in these states are highly suitable for transmission – even in the absence of prior widespread transmission. While work into classifying the ecological niche of YF in South America has been conducted before (Shearer et al., 2018, Hamrick et al., 2017), it has not previously taken into account the role of landcover and landcover change to the same extent.

Five covariates provided significant contributions to the GLM ensemble model. Given the sylvatic nature of YFV in South America, and the forest habitat that NHPs primarily reside in, it is unsurprising that an increase in savanna cover, and urban cover over time were associated with a reduced probability of a YF report. In the same vein, higher numbers of NHP species in a province, increase the probability of a YF report - likely indicates suitability for the sylvatic hosts, with previous research finding similar findings for sylvatic YF spillover (Hamrick et al., 2017, de Almeida et al., 2019). Increasing levels of rainfall, are associated with increased YFV vector populations in South America (Kumm, 1950), and increased vector densities associated with increased flavivirus transmission (Kramer and Ebel, 2003). Vegetation heterogeneity was found

to be the covariate most highly associated with increased probability of YF reports. Increased heterogeneity in vegetation may act as a proxy for habitat fragmentation. Fragmentation, or heterogeneity, may disrupt natural sylvatic vector and host populations and behaviours, leading to increased interaction, or opportunities for interaction, with susceptible human populations (Faust et al., 2018). These findings are in line with previous modelling work across South America, with the additional identification of vegetation heterogeneity as an important predictor of YF.

As this study was carried out at the a fairly low spatial resolution, the province level, due to limitations in the reporting of YF cases and the data available, I may be not fully capturing the relationship between these covariates and YF spillover at the local or individual level (Steel and Holt, 1996). This may go some way to explaining why numerous covariates included in these models which would be thought to have a relationship with YF reporting, such as forest cover, deforestation and changes in human population, were not found to be significant. In order to further explore the relationship between these covariates and YF transmission, modelling exercises at a higher spatial and temporal resolution should be undertaken.

The 49 unique provinces where YF reports were gathered from publicly available data sources, and while I am confident of our exhaustive search of the literature and geolocation, I cannot discount the possibility of case misidentification or omission from the public domain. Given that around 88% of YFV infection is asymptomatic, or produces on a mild non-specific illness, it is likely that the vast majority of YF transmission in humans remains undetected (Johansson et al., 2014). Furthermore, due to the dominance of the sylvatic cycle, human cases primarily occur in highly rural locations where healthcare and surveillance systems may not be as robust (Escobar and Craft, 2016, Barrett and Higgs, 2007). However, these surveillance issues are

more likely to substantially effect counts of cases, by modelling presence/absence this approach is more robust to these issues, as it only takes a single report of YF during the time period to be classed as a province with YF presence. The caveat to this approach is that by only utilising simple presence/absence and may not be capturing the fully intensity and distribution of transmission. In this approach, a province with one reported occurrence of YF during this period is weighted the same as an area that has experienced numerous outbreaks. While our model replicates the data to a high degree of accuracy (AUC of 0.929 (95% CI: 0.899 – 0.959)), the findings likely do not capture the true distribution of the intensity of transmission. Future work into examining this should consider further metrics of the magnitude of transmission, such as cases of YF over time, or the number of unique introductions to a province. Additionally, the incorporation of temporal variation would give a more nuanced approach and allow models to consider this intensity of transmission as well as further elaborate the geographic and temporal drivers of YF in South America.

The identification of covariates, and corresponding provinces, of higher probabilities of YF reporting is an important first step in more formally quantifying the drivers of YF transmission in South America. While using average covariates offers a “snapshot” of transmission, it does not provide information on temporal variations in YF transmission. Both inter- and intra-annual variation in disease transmission can vary substantially, with periodic YF transmission noted in literature (Barrett and Higgs, 2007), and highly seasonal dynamics previously identified (Hamlet et al., 2018). Future work, if these temporal dynamics were highlighted, could support the development of more sophisticated forecasting systems, highlighting periods of heightened transmission based on environmental and climatic variables. This could prove

invaluable, with substantial public health value, in a context of endemic zone expansion and limited vaccine stocks.

Chapter 3 - The effect of seasonal and inter-annual variation in climate and environment on the transmission of YF in South America

3.1 Summary

While static maps of disease presence have substantial utility in combatting disease, by neglecting the temporal relationship of environmental and climate with disease transmission we may fail to detect the true correlates of transmission. This is likely to have particular effects on areas that are not generally suitable for transmission, but in specific periods, seasonally or inter-annually, are vulnerable.

Here, building on the datasets and lessons learnt in the previous chapter, I have applied a series of hierarchical negative-binomial regressions to two distinct temporally varying datasets, one looking at intra-annual (seasonal), and one inter-annual (annual) variation in YF reports.

The resulting models accurately assess the intra- and inter-annual trends in YF reporting across the continent, and highlight differential drivers between the two applications, with intra-annual variation primarily associated with EVI and climate, and the inter-annual, EVI, vegetation heterogeneity and land-cover. By further quantifying what drives intra- and inter-annual transmission this work lends itself to applications in forecasting, where predicted data on the relevant variables may be applied to highlight periods of heightened transmission both seasonally, and in years to come.

3.2 Introduction

Disease transmission is influenced by both intra- and inter- annual variations in weather and the environment, particularly for vector-borne (Muturi, 2013, Craig et al., 1999). These deviations may be relatively short in duration, due to seasonal changes in weather (Hamlet et al., 2018) or phenomena such as El Niño (Fuller et al., 2009), or they may represent a more persistent alteration, due to climate change (Lyon et al., 2017) or alterations to land-cover (Bauch et al., 2015). While climate change is likely to alter both the distribution and intensity of a number of diseases (Curto de Casas and Carcavallo, 1995, Astrom et al., 2012, Artzy-Randrup et al., 2010), this process takes place over a substantially longer period of time than anthropogenic land conversion which can completely change large swathes of habitat in a few years (Warren-Thomas et al., 2018, Jokar Arsanjani, 2018). Rapid habitat change is often associated with disease occurrence (Patz et al., 2004), especially of zoonotic infections (Allen et al., 2017), potentially due to an increased interaction between sylvatic reservoirs and humans, expressly at intermediate levels of transformation (Faust et al., 2018).

I investigated the drivers of YF transmission both intra- and inter-annually, using temporally varying covariates related to climate, land-cover, vegetation and human and NHP demographics. While the influence of habitat fragmentation in YFV transmission has previously been postulated (Bicca-Marques and de Freitas, 2010, Faust et al., 2018, de Almeida et al., 2019), detailed research into the role habitat fragmentation and land-cover play in YFV sylvatic spillover is absent. By utilising two model structures, I investigated the differential drivers of seasonal, and interannual variation in YF incidence. These models were fit to the number of months reporting YF

at each administrative level 1 geographic unit, using exhaustive model fitting in hierarchical negative-binomial regressions. The best performing models, as defined by the Akaike Information Criterion (AIC) were weighted and combined using Akaike weights to produce an ensemble model (Wagenmakers and Farrell, 2004). Model robustness was confirmed using spatial block bootstrapping.

3.3 Materials and Methods

3.3.1 YF data

Following the same approach detailed in the previous chapter, YF reports were obtained from several WHO sources, and by only considering cases that were able to be geo-located at the first administrative division and with a start month, I was able to produce two databases of reports to look at the intra- and inter-annual trends in YF reporting.

Here I was looking at the number of months that report YF over the time period (2003-2016) for each province. For the intra-annual dataset looking at seasonality, this was calculated for each province by looking at the number of reports found, for each month for a total of 14-month reports per month (the number of years). For the inter-annual report dataset, I looked at the number of reports a province had in an entire year, for a maximum of 12-month reports per year (the number of months in a year). An example of how report months are calculated for each dataset is found in Table 3.3.1.

By utilising more than just simple binary presence/absence I am capturing more complex relationships between the covariates are transmission. While a chance event may result in YF being reported in a province, this may not represent a true suitability to transmission. By including the number of months YF is reported in, I am addressing this issue and, in theory, capturing a truer representation of underlying transmission.

This approach was taken rather than using the number of cases due to the large uncertainties in reporting and detection of YF cases. Due to the presence of asymptomatic infection, and non-specific symptoms in mild cases, in addition to the rural locations and issues related to diagnosis, case numbers represent just the tip of the iceberg, and potentially do not indicate the magnitude of actual transmission, particularly in endemic settings (Monath, 2006). Furthermore, the differences between regions in their surveillance are likely to be inconsistent. However, by modelling the presence/absence of a YF report in a month, this approach is more robust to these issues, as it only takes a single report of YF during the time period to be classed as a province with YF presence.

These relationships were investigated using this approach, rather than simultaneously in a continuous model over the time-period, due to the sparsity of the data. If this approach considers using month reports, then it is likely the sparsity of data would negatively affect our ability generate meaningful statistical links with the covariates. While this approach may be utilised for case data, large uncertainties in reporting and detection of YF cases mean that the vast majority of YF transmission goes unreported, and so case numbers are likely substantially lower than recorded, and potentially not indicative of the magnitude of actual transmission, particularly in endemic settings (Monath, 2006).

Table 3.3.1. Example of how report months are calculated for each province for both the seasonal and annual datasets.

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	INTRA-ANNUAL REPORT MONTHS
JANUARY	1		1					1							**3**
FEBRUARY	1				1							1			**3**
MARCH	1	1												1	**3**
APRIL															**0**
MAY				1									1		**2**
JUNE															**0**
JULY															**0**
AUGUST															**0**
SEPTEMBER		1											1		**2**
OCTOBER														1	**1**
NOVEMBER		1		1		1									**3**
DECEMBER											1			1	**2**
INTER-ANNUAL REPORT MONTHS	**3**	**3**	**1**	**2**	**1**	**1**	**0**	**1**	**0**	**0**	**1**	**1**	**2**	**3**	**19**

3.3.2 Covariates

Reports and covariates are described in full in the previous chapter (Table 2.3.1), and were considered based on knowledge of their role, or potential role, with vector species, host dynamics and the general ecology of YF. Here, in relation to temporal trends investigated in the model, covariates are differentially calculated.

Covariates included in the intra-annual model, that do not vary intra-annually, are given the mean value over the time-period, and covariates which vary intra-annually are given their mean value in a year for inter-annual models. This is done to give a generalised value in the absence of specific values at the temporal resolution considered. The temporal resolution covariates are found in at are depicted in Table 3.3.2. This approach generated an intra-annual dataset with 12 rows (corresponding to the months of the year) for each province, and for the inter-annual dataset, 14 rows (corresponding to each year in the time period) for each province.

Table 3.3.2. Temporal resolution available for covariates

TEMPORAL RESOLUTION AVAILABLE	COVARIATE
STATIC	Monkey species (count)
	Logarithm of human population
	Change in logarithm of human population
INTRA/INTER-ANNUAL	Enhanced Vegetation Index (EVI) (0-1)
	EVI amplitude
	Vegetation heterogeneity
	Vegetation heterogeneity temporal change
	Day temperature (°C)
	Day temperature amplitude (°C)
	Rainfall
	Rainfall amplitude
INTER-ANNUAL	Forest cover
	Savanna cover
	Cropland/natural vegetation mosaic cover
	Urban cover
	Cropland cover
	Forest cover temporal change
	Savanna cover temporal change
	Natural vegetation/cropland mosaic cover temporal change
	Urban cover temporal change
	Cropland cover temporal change

3.3.3 Models

Following initial covariate selection and processing, a list of covariates identified as relevant to YFV transmission were considered, with log of human population and the fractional change in logarithm human populations included in every model, as previously discussed. The fractional change in the logarithm of human population is

calculated by using UN World Population Prospects data which takes into account changes in population size over time (United Nations, 2016). An exhaustive combination of all 19 covariates was explored.

Negative binomial regressions implemented using the glmmTMB package (Brooks et al., 2017), were used to associate the covariates with report months of YF. Negative binomial regressions are a generalisation of Poisson regressions, and offers a useful approach for modelling count variables – particularly over-dispersed outcome variables such as report months of YF (Hilbe, 2007).

Here I have also utilised a multi-level (hierarchical) approach to our models. Conceptually, hierarchical models are similar to running a standard regression where each row in the dataset refers to an administrative location and a time point (month or year depending on the model structure). By utilising a hierarchical structure however, parameters can vary between administrative location to avoid introducing biases that arise from treating temporally varying covariates within a location as independent (Gelman and Hill, 2006). This was implemented by allowing a varying intercept between administrative locations. The best performing models were ascertained through their Akaike Information Criterion (AIC) value, and model predictions and covariate coefficient values calculate as specified in the previous chapter.

Model predictive ability was assessed using the coefficient of determination (R^2) which is a measure of the proportion of variance in the dependent variable predicted from the independent variable. Out-of-sample predictive ability was assessed using Spatial-Block bootstrapping detailed in the previous chapter.

All data processing and analysis was carried out using the statistical programming language R (Team, 2015a).

3.3.4 Variable importance

Here, in addition to the % of models that a covariate was found in, I have utilised a measure of variable importance. Here variable importance refers to how a measurement score decreases when a feature is not available. Initially the full model with the initial dataset is fit, and the R2 value calculated. Then the model is refit to a modified initial dataset, where the variable of interest within the dataset been assigned the mean value of that covariate. This is done to produce "dummy data" which creates a covariate that does not provide any useful information. The R2 is then calculated from this refit, and the variable importance for variable, *i*, calculated through,

$$variable\ importance_i = 1 - \frac{refit\ R2_i}{original\ R2_i}.$$

This provides a measure of the feature importance with respect to the original R2.

3.4 Results

3.4.1 Geographical, seasonal and interannual heterogeneities in YF reports

I identified 397 unique months with a report of YF, hereby termed report months (defined spatiotemporally by the administrative unit and month), for the period 2003-2016, in 432 level 1 administrative units across 8 countries (Figure 1). Peru, Colombia and Brazil accounted for 79% of all report months, with Peru alone accounting for 39% (Figure 1A and Figure 2). Within countries, report months show substantial spatial heterogeneity, with a notable clustering in Amazonian regions of Brazil, eastern Peru and Northern Bolivia. States in the South-East Atlantic coast of Brazil have also recorded large numbers of report months (Figure 1B)(Romano et al., 2014).

The frequency of report months was relatively stable and high during 2003-2008, after which numbers fell, then plateaued until 2015, when they dipped to the lowest level seen with only 1 reported event (Figure 1A). It should be emphasised that report months are a presence/absence indicator and not a proxy for infection incidence. Throughout the endemic zone, YF follows highly seasonal patterns. At the continent scale, transmission is highest from December to February, before dropping to a relatively low level over June to September, and a period of minimal occurrence in October and November (Figure 2). However, this pattern varies slightly by country and latitude.

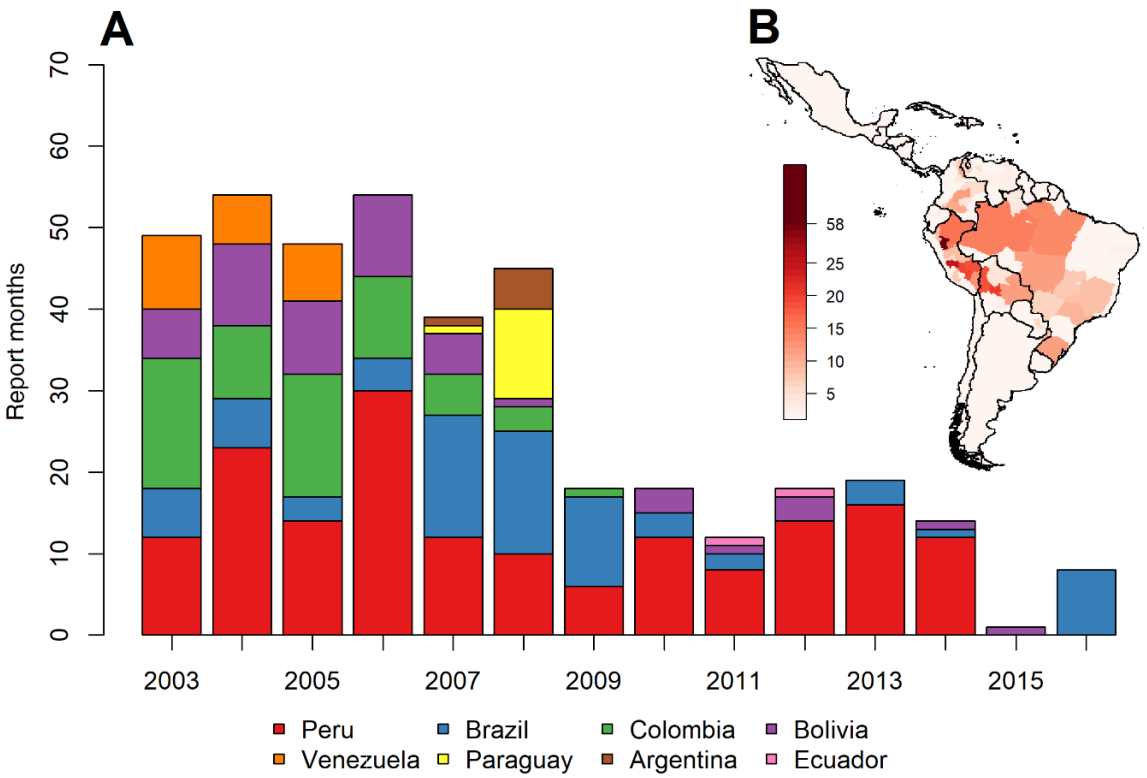


Figure 3.4.1. (A) Number of yellow fever report months over time (2003-2016) by country. (B) Total number of yellow fever reports by province (2003-2016) across South America.

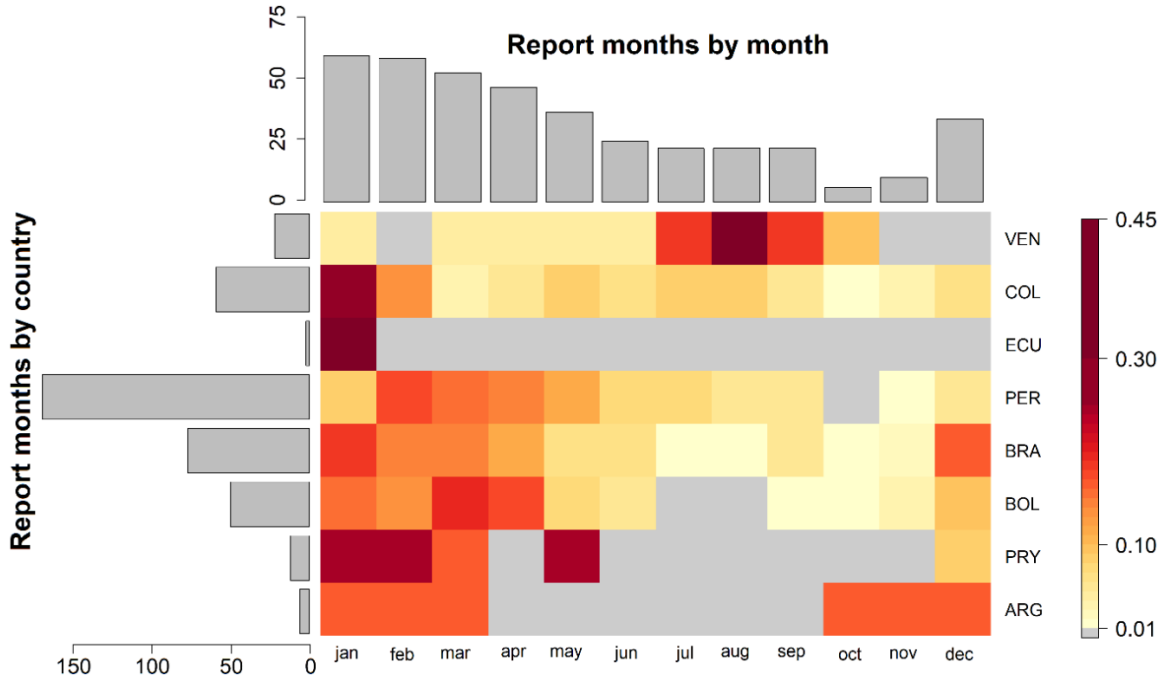


Figure 3.4.2. Yellow fever reports by country and month. The heatmap shows the proportion of reports in a country by calendar month, the bar chart on the left-hand side shows the total number of reports by country and the bar chart above shows the total number of months reporting cases by month. Countries are ordered by latitude.

3.4.2 Geographic distributions of model predictions

The ensemble interannual and seasonal models accurately approximate spatiotemporal heterogeneities in YF reports, with coefficient of determination (R^2) values of 0.43 (95% CI 0.41 – 0.45) for the interannual and 0.66 (95% CI 0.64 – 0.67) for the seasonal ensemble predictions. Due to the additional rigour of using spatial block bootstrapping compared with using an entirely random validation set (See SI), out-of-sample prediction R^2 values were lower at 0.31 (95% CI 0.28 – 0.34) for the interannual model and 0.45 (95% CI 0.44 – 0.48) for the seasonal model.

Model predictions were summed over time for each model to facilitate visual comparison with the data (Figure 3A and B vs Fig 1B). Both models reproduce the

observed geographic distributions of reports well, though the aggregate ensemble seasonal model predictions give a better fit to the data (Figure 3.4.3). Differences between the ensemble interannual model predictions and the data range from -4.35 to +3.08. The model over-predicts reports for much of Eastern Peru and the North-West of Brazil and predicts fewer reports than observed for Rio Grande do Sul in Brazil, and Misiones province in Argentina. There is additionally a cluster of lower than observed predictions on the Colombian/Venezuelan border. Ensemble seasonal model predictions showed deviations from the data an order of magnitude smaller than seen for the interannual model. The seasonal model slightly underpredicts YF reports, with only Brazilian states in the Amazon, Rio de Janeiro, and the East/North-East of the country predicted as having more reports than observed.

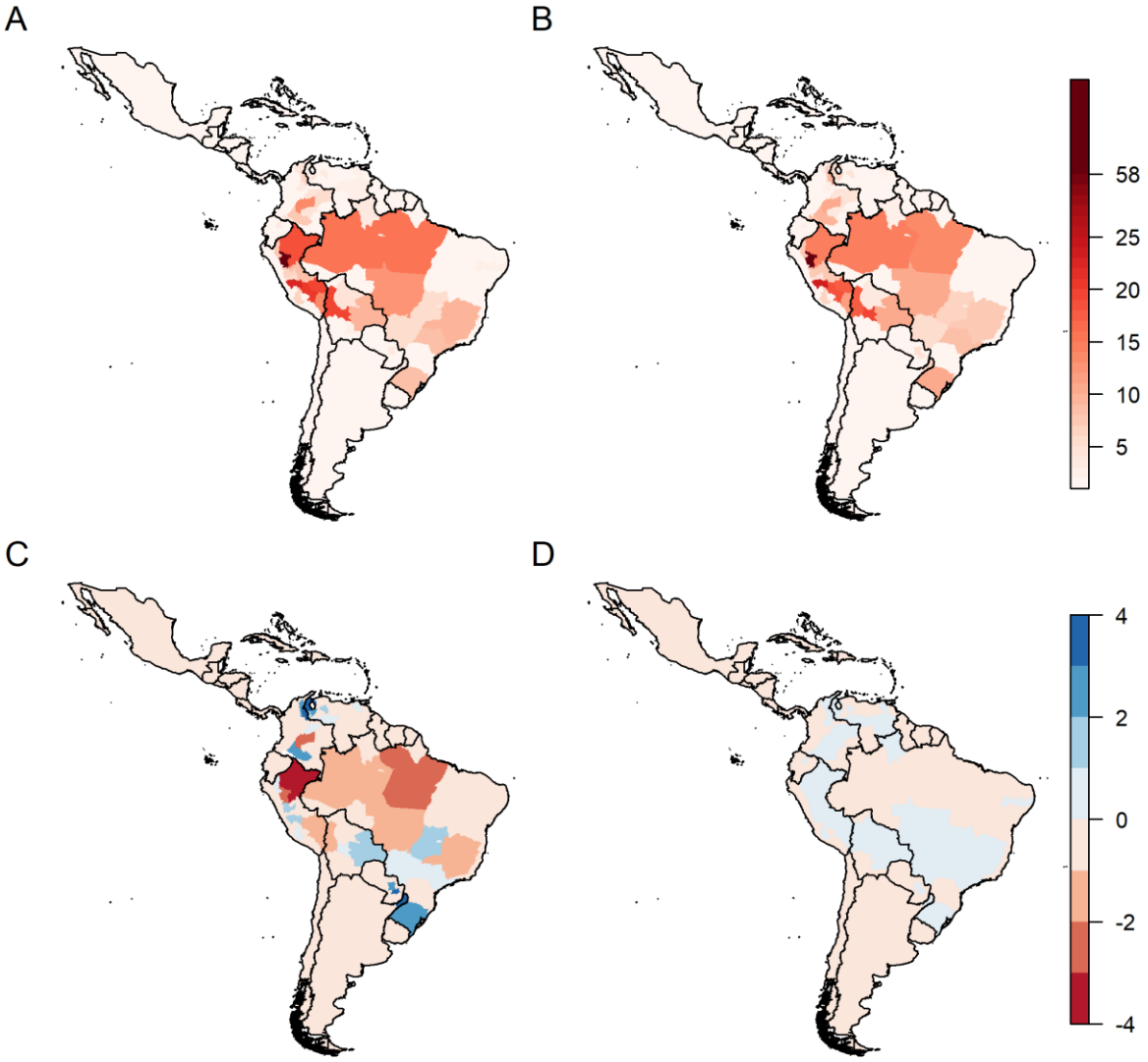


Figure 3.4.3. Ensemble model predictions of the number of YF report months for the (A) interannual model and the (B) seasonal model. (C) and (D) show the differences between these predictions and the data for the interannual model and the seasonal model, respectively.

3.4.3 Temporal distributions of model predictions

In addition to representing geographic variation, the models also consider temporal heterogeneity in YF incidence (Figure 3.4.4). Both the interannual model and the seasonal model fit temporal trends well, as shown in their R^2 values of 0.43 (95% CI 0.41 – 0.45) for the interannual and 0.66 (95% CI 0.64 – 0.67).

At the continent level, the inter-annual and seasonal model replicate the trends, but not the overall magnitude of temporal variation in report months. In the Inter-annual model, report months are underpredicted until 2009, after which they are slightly over-predicted. In the seasonal model, the model underestimates the data until the 5th month, then over-predicts later months (Figure 3.4.4, A and B). When years and months are ranked by the number of report months, there is a high degree of concordance between predictions and the data. This is shown in the high Pearson correlation coefficient values between the predicted and actual rank of years and months, at 0.926 for the interannual model, and 0.873 for the seasonal model (Figure 4C, D).

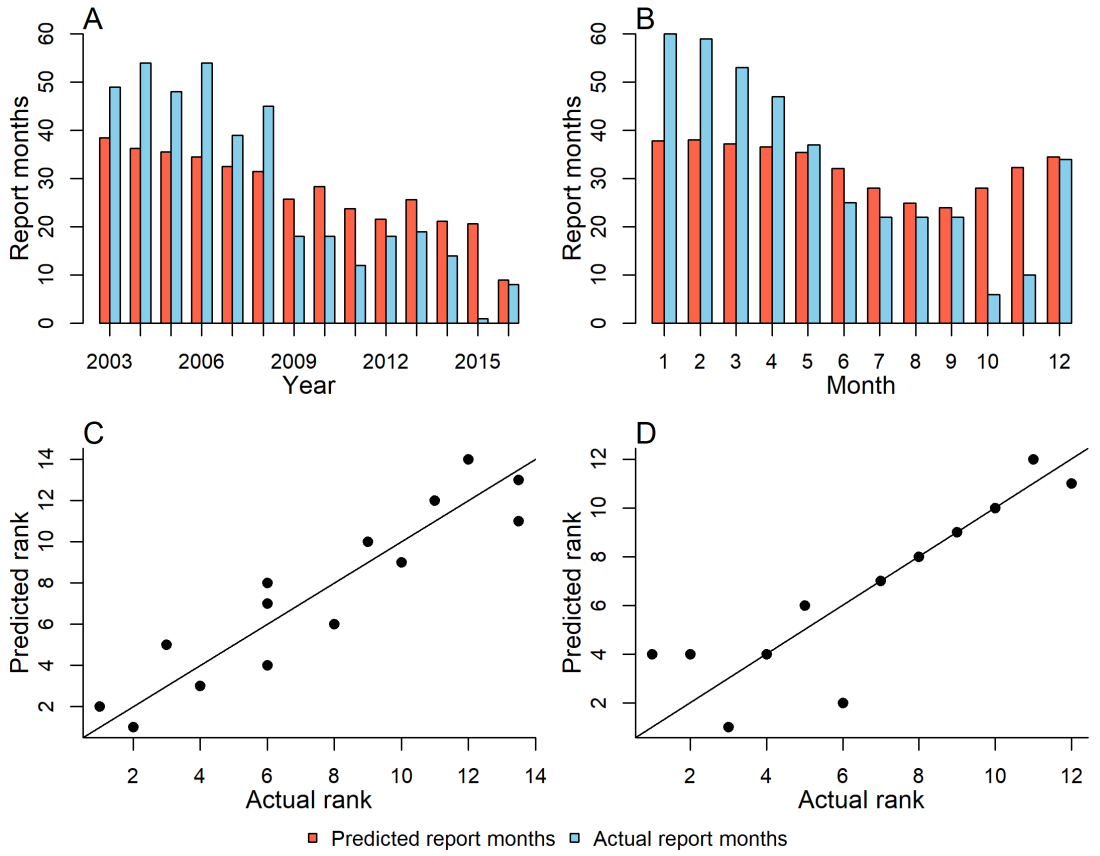


Figure 3.4.4. Summed ensemble model predictions (points) for (A) each year for the interannual model (A), and for each month for the seasonal model (B), contrast against the actual summed report months (lines) for each year or month. Yearly (C) and monthly (D) predictions ranked against the actual report months for the interannual and seasonal models, respectively (lines show predicted = actual).

3.4.4 Drivers of seasonal, annual and long-term yellow fever transmission

The interannual and seasonal ensemble models showed both similarities and differences in the predictors found to be most significant (Table 1). For both, the covariate grouping relating to host demographics were the most important, with log of human population explaining the most variance in both model sets. The number of NHP species present also had a smaller but significant contribution for each. Both demographic predictors were positively associated with YF reports. The Enhanced

Vegetation Index (EVI) was the second most important predictor for both models, again positively associated with YF reports. Other predictors differed between models, likely reflecting that the interannual model selected predictors best able to reflect long-term trends in YF reports, while the seasonal model selected those able to reproduce intra-annual seasonal patterns.

For the interannual model, landcover (cropland and savannah being negatively associated with YF) and vegetation heterogeneity (the standard deviation of EVI) were the next most important predictor groupings. Temporal changes between the current and previous year in vegetation and land-cover were significant predictors but made relatively small contributions to model fit. No climate coefficients were significant in the interannual model.

For the seasonal model, mean monthly day temperature and mean monthly rainfall amplitude (see methods for definitions) were the other significant predictors, both negatively associated with YF reports.

Significant covariates were found in all (or almost all) of the best performing models, with all covariate groupings found in interannual models except climate, and in the seasonal models only climate, vegetation and host demographics were found in the best performing models. Variable importance was highest in the EVI for both inter-annual and seasonal models, with the vegetation heterogeneity of a similar level of importance in the inter-annual model, and the number of NHP species and logarithm of human population slightly, but still important in the seasonal model. Despite the significance of the mean day temperature in the seasonal model, it was found to have an almost negligible variable importance – indicating it did not particularly contribute to predictive accuracy.

Table 3.4.1. Table of the permutation importance of different covariate groups, and individual covariates as well as standardised coefficient values. Only covariates that were significant in at least one of the model sets are shown. (A) Refers to the inter-annual model, and (B) the seasonal model.

COVARIATE GROUPINGS	% OF MODELS COVARIATE GROUP FOUND IN		COVARIATE	% OF MODELS COVARIATE FOUND IN		VARIABLE IMPORTANCE		COEFFICIENT VALUES	
	A	B		A	B	A	B	A	B
CLIMATE	0	100	**Mean day temperature**	0	99.5	0	0.01		-2.71 (95% CI: -5.10 - -0.33)
			Mean rainfall amplitude	0	100	0	0.25		-0.35 (95% CI: -0.64 - -0.06)
VEGETATION	100	100	**EVI**	100	100	0.82	0.84	6.42 (95% CI: 3.09 - 9.76)	6.21 (95% CI: 3.89 - 8.53)
			EVI amplitude	100	0	0.19	0	1.17 (95% CI: 0.33 - 2.01)	
VEGETATION HETEROGENEITY	100	0	**Vegetation heterogeneity**	100	0	0.79	0	3.34 (95% CI: 1.77 - 4.92)	
VEGETATION HETEROGENEITY TEMPORAL CHANGE	100	0	**Vegetation heterogeneity temporal change**	100	0	0.14	0	-0.39 (95% CI: -0.6 - -0.17)	
LAND-COVER	100	0	**Cropland cover**	100	0	0.21	0	-2.47 (95% CI: -4.27 - -0.66)	
			Savanna cover	100	0	0.50	0	-2.27 (95% CI: -3.73 - -0.8)	
LAND-COVER TEMPORAL CHANGE	100	0	**Savanna cover temporal change**	100	0	0	0	0.32 (95% CI: 0.10 - 0.54)	
HOST DEMOGRAPHICS	100	100	**Number of NHP species**	100	100	0.48	0.74	1.34 (95% CI: 0.75 - 1.92)	1.68 (95% CI: 0.90 - 2.47)
			Log of human population	100	100	0.24	0.66	6.89 (95% CI: 0.62 - 13.16)	13.65 (95% CI: 5.19 - 22.11)

3.5 Discussion

In this study I have described the geographic, seasonal and interannual trends in YF reports in Latin America from 2003-2016, using publicly available data. I used hierarchical negative binomial regression models to create ensemble models predicting interannual and seasonal variation in YF report months with a series of climatic, land-cover, vegetation and host demographic covariates. Our models explained a substantial amount of the observed variation, with R^2 values of 0.43 (95% CI 0.41 – 0.45) for the interannual and 0.66 (95% CI 0.64 – 0.67) for the seasonal model.

The geographic distribution of reports highlights “hotspots” for YF transmission, in Eastern Peru, North Western Peru and South Eastern Brazil (Figure 3.4.1B).The seasonal model reproduced these geographic trends more accurately than the interannual model. Continental-level interannual and seasonal trends in the data were also well-reproduced by the respective models. While at this level, the magnitude of temporal trends in report months are not fully captured, the relative ranking of years is and therefore model results can shed some light on what is associated with increased, or decreased, YF reporting in particular years and months.

Though differing covariates are important for driving interannual and seasonal changes in YF transmission, vegetation (EVI) is highly influential for both models. This has been previously highlighted as a predictor of seasonal YF transmission (Hamlet et al., 2018), and potentially acts as a proxy for the interaction of rainfall and temperature, both important for arboviral transmission, while also taking into account a more complex interaction than is captured by either covariate alone. In both the interannual and the seasonal models, the log of human population was the most

important predictor. This is not unexpected – larger populations give more opportunity for spillover, and since a report month is a month where one or more human YF cases are reported, larger populations are more likely to accumulate 1 or more cases in any one month even with a spatially invariant per-capita risk of YF. For the interannual model, landcover and heterogeneity in vegetation were also influential covariates in explaining interannual variation in YF reports. While cropland and savanna cover are negatively associated with YF reports, vegetation heterogeneity is positively associated. The heterogeneity covariate I adopted maybe acting as a proxy for habitat fragmentation. Fragmentation may affect sylvatic hosts in a number of ways, such as increasing their exposure to human contacts via modified behaviours (Gottdenker et al., 2012, Goldberg et al., 2008) or increased susceptibility to infection due to a stress-weakened immune system (Seltmann et al., 2017). Furthermore, vegetation heterogeneity may alter vector dynamics and predispose greater rates of spillover either through increased human-sylvatic cycle contact or the favouring of more anthropophilic vector species in fragmented habitats (Burkett-Cadena and Vittor, 2018). These effects have previously been suggested to affect zoonotic disease transmission, but until now had not been statistically implicated in YF emergence (Faust et al., 2018, Bicca-Marques and de Freitas, 2010).

While I have explained a substantial proportion of the seasonal and inter-annual variation in YF reporting across South America (2003-2016) (Interannual model: 0.43 (95% CI 0.41 – 0.45), seasonal model: 0.66 (95% CI 0.64 – 0.67)), this still means that, respectively, 67% and 34% of this variation is unexplained. This, in part, may be due to the spatial resolution at which the study was carried out. Due to data limitations in the reporting of YF cases, I may not have fully captured the relationship between climate and environment with YF spill over at the local or individual level. This may

explain why some covariates that may be expected to be associated with increased spillover, such as forest cover and change in forest cover, have not been found to be significant. To account for this, future modelling work should take place at a higher spatial resolution, though the trade-off between the availability and quality of data with a potentially furthered understanding should be thoroughly explored.

While climatic and landcover fluctuations both inter- and intra-annually lead to changes that can lead to increased disease transmission, they do not represent the whole picture of spillover. In order for YF to enter human populations it has to be both circulating within the NHP reservoir, and there has to be human exposure to the sylvatic cycle. Across South America (2000 – 2014), 60% of human cases of YF were in people employed in farming, hunting or fishing – highly seasonal activities (Pan American Health Organization, 2017). This changing risk of exposure is likely to account for a proportion of the temporal and spatial reporting of YF. In order to better capture these relationships with YF spillover into human populations across South America, future modelling exercises should endeavour to capture both the underlying suitability to disease transmission, and these correlates and determinants of exposure.

This analysis uses 397 months of YF report months, only included publicly available case reports (World Health Organization, Pan American Health Organization, 2017) which had a confirmed onset date and which could be geolocated to at least the province level. Due to missing data, 23% of case reports were excluded from our analysis. In addition, due to the remote locations that sylvatic YF is often found in and the non-specific symptoms many cases show, it is likely that substantial numbers of YF cases are never recorded (Johansson et al., 2014, Barrett and Higgs, 2007). Underreporting in rural areas may lead us to underestimate YF risk in those locations. However, surveillance and data quality issues affect estimation of absolute case

incidence, report months (presence/absence of cases in a specific administrative unit in a particular month) is likely to be more robust to under-ascertainment, as it only takes one reported case to be classified as YF positive. I am unable to ascertain whether the predictors of YF transmission that were identified affect sylvatic transmission or human exposure, given I have only analysed reports of human cases here. Data on NHP cases of YF across the continent are limited however, and their omission is a permissible oversight given this.

While expansion of the endemic zone is occurring, increases in population-level vaccination coverage in the endemic zones, where the majority of transmission is predicted, has precluded much of the human population from infection. This is in contrast with areas outside of this zone– where YF vaccination is either not usually necessary or not prioritised, and where spillover is more likely given the available of susceptible humans. This may go some way to explaining the decrease in report months over the time period (Figure 1).

In conclusion this body of work represents an important quantification of both the seasonality and interannual transmission of YF across South America (2003-2016). By identifying covariates, and their statistical relationship, with report months of YF, the work presented here helps support offers the ability for forecasting systems to be developed, with the aim of highlighting risks of increased transmission potential for both entire years and specific months. Such systems could have substantial public health value, in a context where the geographic range of YF is changing and vaccine stocks are still limited.

Chapter 4 - The seasonality of agriculture and climate on predicting human and NHP reports of YF in Brazil

4.1 Summary

Previous descriptions of YF's highly seasonal nature have relied on climatic explanations, with the role of the seasonality in exposure neglected – despite the high proportion of cases occurring in those involved in seasonal agricultural activities. This study investigates the role of seasonality in agriculture, as well as climate, in YF reporting for humans and NHP's.

I fit a series of random forest classification models to yellow fever (YF) occurrence in both human and NHP's (2003-2018) at the 2nd administrative division aimed at classifying the probability of presence of human reports of YF, NHP reports of YF and both human and NHP reports of YF. This was done through fitting four classes of covariates describing the seasonality of climate, the seasonality of agriculture (planting, harvesting), agricultural output and host demography. Models more accurately captured municipalities with human and human and NHP reports of YF compared to NHP reports. Furthermore, models fit to covariates including the seasonality of agriculture capture seasonal trends in YF reporting at a substantially higher level than those that simply accounted for the seasonality of vegetation/climate.

These findings illustrate the importance of the seasonality of exposure and that it is not necessarily just an increased viral transmission in zoonotic reservoirs which leads to spillover, but also an increased interaction with the sylvatic cycle. Direct application of these findings in a public health setting is possible, with 45% of YF cases in Brazil occurring in those involved in agriculture. By highlighting both the crop types involved,

and the temporal window, vaccination activities and public health outreach could effectively and pro-actively educate and vaccinate those at greatest risk of YF.

4.2 Introduction

In Brazil, since 1942, all cases of YF have been recorded as due to the sylvatic cycle, with much of this transmission confined to the North and North West of the country (Almeida et al., 2014). However, since 1998 there has been a significant expansion of the endemic zone (Almeida et al., 2014), culminating in the largest outbreak of YF since records began in 2016-17 in the densely populated South Eastern states of the country (Cunha et al., 2019). This was followed by an equally large and widespread outbreak during the following season, 2017-2018, with additional, low level transmission detected outside its endemic zone in the 2018-2019 season.

While the seasonality of YF has been previously highlighted (Hamlet et al., 2018, Kumm, 1950), there remain substantial knowledge gaps about the processes behind this. Seasonal variations in climate can lead to increased vector populations and the suitability for disease transmission, factors which have been used to explain this temporal variability – and even allowed for the forecasting of coming seasons with a high degree of accuracy (Do et al., 2014, Cairns et al., 2015, Fisman, 2007, Fuller et al., 2009, Hu et al., 2005). However, due to sylvatic transmission driving YF cases in humans in Brazil, there remains a counterpart to the seasonality of transmission, the seasonality of exposure. In Brazil around 45% of cases of YF occur in those involved in agriculture or hunting, both highly seasonal activities (Pan American Health Organization, 2017). Despite the relationship between agriculture and human disease transmission being one of considerable scientific interest (Kingsley, 2016), with numerous articles on how landscape changes can affect exposure to human populations (Gibb et al., 2017, Faust et al., 2018), changes in vector composition (Burkett-Cadena and Vittor, 2018, Zittra et al., 2017), or alter zoonotic reservoir

behaviours (Wynne and Wang, 2013), research on how disease transmission is altered by the seasonality of agriculture is lacking.

In this study I seek to investigate the drivers of seasonal YF transmission in Brazil in both humans and NHP's. I apply classification random forest models to predict occurrence of human or NHP YF using covariates related to the seasonality of climate and of agriculture. I then evaluate the relative importance of these components and identify individual crop types and agricultural activities that are related to increase YF reporting.

4.3 Methods

4.3.1 YF reports

YF case data for humans and NHP's was provided by the Brazilian Ministry of Health at the municipality level for all cases recorded between 2003 and 2018 in a personal communication. These cases, for both human and NHP were laboratory confirmed. By modelling both humans and NHPs I aimed to gain a greater understanding of what was both associated with increased sylvatic transmission, as well as the spill over of infection into human populations.

Human cases were anonymised and included a municipality (the second administrative division of Brazil) identification number, municipality name, and the date of symptom onset. There were 2423 human cases of YF in the original dataset; of these 10 did not contain a date, 18 could not be geo-located, leaving 2395 cases. These 2395 cases translated to 694 monthly occurrences of YF across 434 unique municipalities. Case data for NHP's contained a municipality identification number, municipality name and the data of epizootic event discovery. There were 3209 NHP epizootic events, of which all could be identified at the municipality level, and with a date – though 10 occurred before 2003, leaving 3199 cases. This led to 771 monthly occurrences in 409 unique municipalities.

Monthly reports of YF were aggregated over the time period 2003 – 2018 due to the relative scarcity of YF reports on an individual annual basis. Thus, the final dataset consisted of occurrence (coded as a binary 0/1 variable) of YF for each of the 12 calendar months and each municipality. Here, as I am aiming to disentangle seasonal spillover the approach of aggregating over years allows us to more strongly associate seasonal covariate dynamics with human and NHP reports.

4.3.2 Host demographics

NHP species distribution maps were obtained from NatureServe (Patterson et al., 2007). This data was available as shapefiles of distribution, which was geo-located to the municipality level and used to calculate the number of NHP species present in each municipality. Data on the human population of each municipality was provided by the Brazilian Ministry of Health.

These covariates were included based on the importance for NHPs in sylvatic transmission, and the previously discussed relationship with human populations and disease surveillance (Garske et al., 2014, Hamlet et al., 2018).

4.3.3 Seasonally varying agricultural activity

Information on agricultural activities (planting and harvesting) at the state (first administrative level) was extracted from an agricultural calendar published by Companhia Nacional de Abastecimento (CONAB) in conjunction with the Ministério da Agricultura, Pecuária e Abastecimento (MAPA) in Brazil (Companhia Nacional de Abastecimento (CONAB), 2017). This provided data on a monthly basis for 15 crops in Brazil. This information was tabulated as a dataset of monthly presence and absence (0/1) of planting and harvesting for each crop.

Of these crop types, 8 were chosen for further analysis due to the number of farms producing them, and in order to restrict the seasonally varying agricultural activity covariates to a similar number as the seasonally varying climate and vegetation covariates. Therefore, I have included the seasonality of agriculture activity for: peanuts, rice, the common bean, castor beans, corn, soya, sorghum and wheat. These 8 crop types represent 16 binary covariates of planting and harvesting.

4.3.4 Agricultural output

Information on agricultural output of Brazil at the municipality level is provided by the "2017 Agricultural, Forestry and Aquaculture Census" in a variety of formats at their portal https://censoagro2017.ibge.gov.br/templates/censo_agro/resultadosagro/agricultura.html (Instituto Brasileiro de Geografia e Estatística (IBGE), 2017). This provided the number of farms producing each of the 8 crop types that seasonal agriculture data was available for at the municipality level.

Furthermore, in order to account for the magnitude of potential agriculture related exposure, I utilised the proportion of the municipality population working in agriculture in 2017 as provided by the Brazilian Ministry of Health in a personal communication.

4.3.5 Seasonally varying climate and vegetation

As in previous chapters, data on temperature (Garske et al., 2013), vegetation (as measured by the Enhanced Vegetation Index (EVI) (NASA)), and rainfall (Joyce et al., 2004) were spatially aggregated from their original resolution, of between 1/120 and 1/12 degree) by calculating population-weighted means, based on the population distribution from LandScan 2015 (Bright et al., 2016) for municipality in Brazil. Multi-year averages (over 2003-2016) were calculated for each calendar month of the year and municipality to produce the average climate and vegetation levels for a municipality over this time period.

I additionally account for the relationship of delays in temperature and rainfall on YF reporting. This was done based on previously published research which has highlighted the effect of lagged environmental data on YF (Hamlet et al., 2018). This delay is likely related to the time required for vector populations to increase to increase,

and the development of the pathogen within the vector (Paaijmans et al., 2013, Huber et al., 2018).

4.3.6 Covariate groupings

Agricultural output covariates were included alongside their relevant seasonal variations in agricultural activity, along with covariates related to climate and vegetation, and the number of NHP species, the proportion of the population working in agriculture and the logarithm of human population. In order to increase the relevance of model findings, the dataset was ordered to follow the Brazilian YF surveillance period of July – June.

These were grouped into four classes and all possible combinations of these four classes were investigated, for a total of 15 models.

Table 4.3.1. Covariate groupings, monthly variaton and sources of data

GROUPING	COVARIATE	MONTHLY VARIATION	SOURCE
AGRICULTURAL OUTPUT	Number of peanut farms	No	(Instituto Brasileiro de Geografia e Estatística (IBGE), 2017)
	Number of rice farms	"	"
	Number of bean farms	"	"
	Number of castor bean farms	"	"
	Number of corn farms	"	"
	Number of soya farms	"	"
	Number of sorghum farms	"	"
	Number of wheat farms	"	"
	% of population working in agriculture	"	Brazilian Ministry of Health
HOST DEMOGRAPHICS	Number of NHP species	"	(Patterson et al., 2007)
	Logarithm of total population	"	Brazilian Ministry of Health
SEASONALITY OF AGRICULTURE	Wheat planting	Yes	(Companhia Nacional de Abastecimento (CONAB), 2017)
	Wheat harvesting	"	"
	Bean planting	"	"
	Bean harvesting	"	"
	Corn planting	"	"
	Corn harvesting	"	"

	Rice planting	“	“
	Rice harvesting	“	“
	Peanut planting	“	“
	Peanut harvesting	“	“
	Castor bean planting	“	“
	Castor bean harvesting	“	“
	Sorghum planting	“	“
	Sorghum harvesting	“	“
	Soya planting	“	“
	Soya harvesting	“	“
SEASONALITY OF VEGETATION/CLIMATE	Rainfall	“	(Joyce et al., 2004)
	Day temperature	“	(Garske et al., 2013)
	Night temperature	“	(Garske et al., 2013)
	Diurnal/Nocturnal temperature range	“	Calculated
	EVI	“	(NASA)
	Rainfall delay by 1 month	“	
	Day temperature delay by 1 month	“	
	Night temperature delay by 1 month	“	
	Diurnal/Nocturnal temperature range delay by 1 month	“	Calculated
	EVI delay by 1 month	“	
	Rainfall delay by 2 months	“	
	Day temperature delay by 2 months	“	
	Night temperature delay by 2 months	“	
	Diurnal/Nocturnal temperature range delay by 2 months	“	Calculated
	EVI delay by 2 months	“	

For the purpose of comparison, covariates were standardised to the have zero mean and unit standard deviation before being used in the random forest models.

4.3.7 Random forest models

As referred to in previous chapters (Models and modelling algorithms 2.3.5) , Random Forests (RF) are a machine learning ensemble regression method (Biau, 2012). These are similar to traditional regression models, in that they use covariates to explain patterns in data but work by creating a series of decision trees to explain the results.

These “trees” are then aggregated, and the mean taken produce a “forest”. These can provide substantial improvements in accuracy over traditional regressions, in addition to accounting for both interactions and non-linear relationships (Biau, 2012).

Here, I have applied a “classification” RF model, rather than the simple regression version I had utilised in Chapter 2. Fundamentally, these follow the same underlying architecture but instead of simply predicting a single outcome, they calculate the probability of each “classification” level being assigned. I used the RF models to classify municipalities into one of four categories, no reports of YF, human reports of YF, NHP reports of YF or both, for each month.

Furthermore, partial dependence plots (Molnar, 2019) were calculated for each model to assess the contribution of individual covariates to predicted YF risk. Partial dependence plots show the marginal effect of a feature on the predicted outcome, and so can show you how a covariate affects the probability of classification in each category changes over a range of the covariate’s values, assuming all other covariates remain static. This allows for the visualisation of the non-linear relationships of a covariate with the outcome.

Model fit for each classification type was assessed by the out-of-sample area under the receiver operating characteristic curve (AUC), a measure of sensitivity and specificity, and the overall model performance rank by the out-of-sample Brier score (Fielding and Bell, 1997). The Brier score is a way of modelling the accuracy of probabilistic predictions when outcomes are mutually exclusive, with the lowest score indicating the best set of predictions (Benedetti, 2010) and so serves as a way of quantifying overall model performance.

Random forest modelling was carried out using the Ranger package (Wright and Ziegler, 2017) in the statistical programming language R (Team, 2015a), version 3.5.1.

4.3.8 Out-of-sample validation

The same method as utilised in previous chapters was used to assess out-of-sample predictive ability using a spatially disaggregated form of cross-validation called spatial-block bootstrapping. A 5°x5° grid of longitude and latitude was constructed, and municipalities assigned to grid squares using their centroid coordinates. Grid squares were randomly sampled from this grid with replacement to produce a training dataset of the same size as the original but comprising of 60-70% of the municipalities. The remaining 30-40% of municipalities were used as a validation set (Figure 4.3.1). This was repeated 200 times to produce 200 different training and validation datasets

Models were fitted to the training dataset and used to predict the validation dataset, with predictions being assessed via the out-of-sample AUC. This was repeated 200 times with different block bootstrapped training and validation sets. The average AUC across all 200 samples was then taken to ascertain the out-of-sample predictive performance of the models.

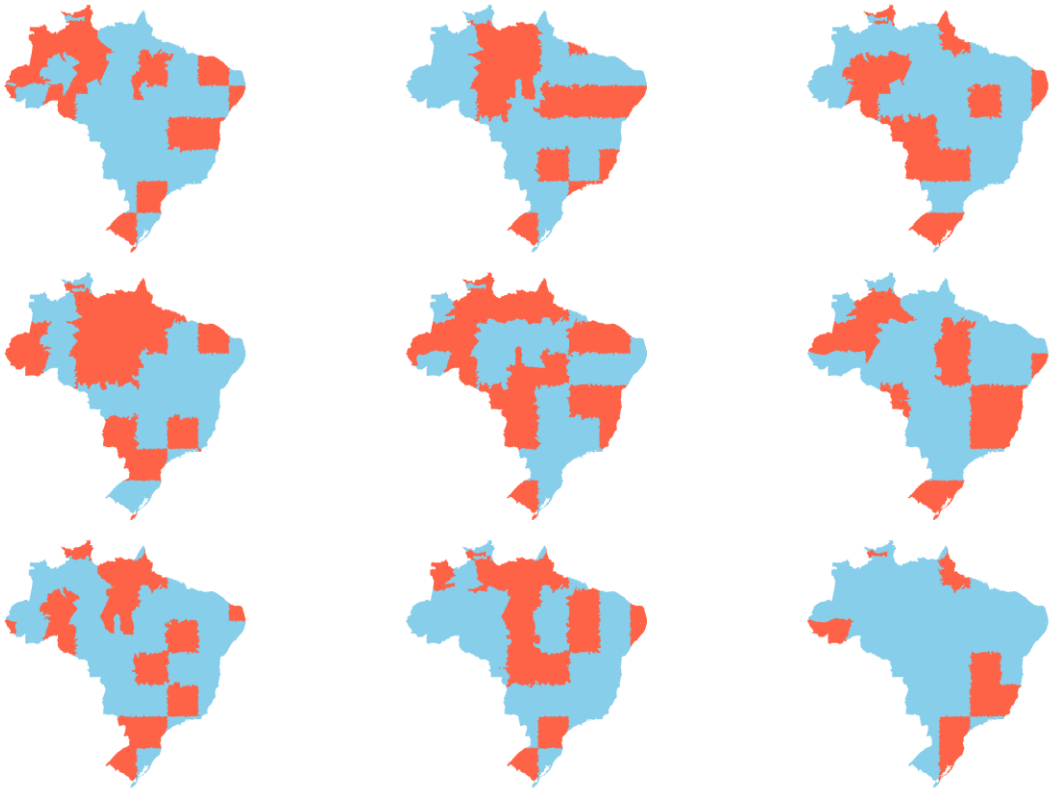


Figure 4.3.1. Examples of the training(blue) and validation (red) datasets as chosen by random sampling of districts on a grid of 5° x 5° longitude and latitude.

4.4 Results

4.4.1 Seasonality of YF reports in humans and NHP's in Brazil

YF reports were highly seasonal in both humans and NHPs, though specific patterns differed slightly (Figure 4.4.1). Human YF reports are minimal throughout much of the year, June – November, but increase rapidly in December to a peak in January before decreasing towards minimum values in May. In contrast, NHP reporting has a lower seasonality amplitude – with cases reported throughout the year at a background level. Cases increase from October, with a similarly timed peak in January. This remains stable for February and March, before descending to background levels in June

The vast majority of human reports of YF occur between -18° and -25° latitude, whereas NHP reports have a more widespread distribution, with less clustering and substantial numbers further south at -27° to -32° latitude.

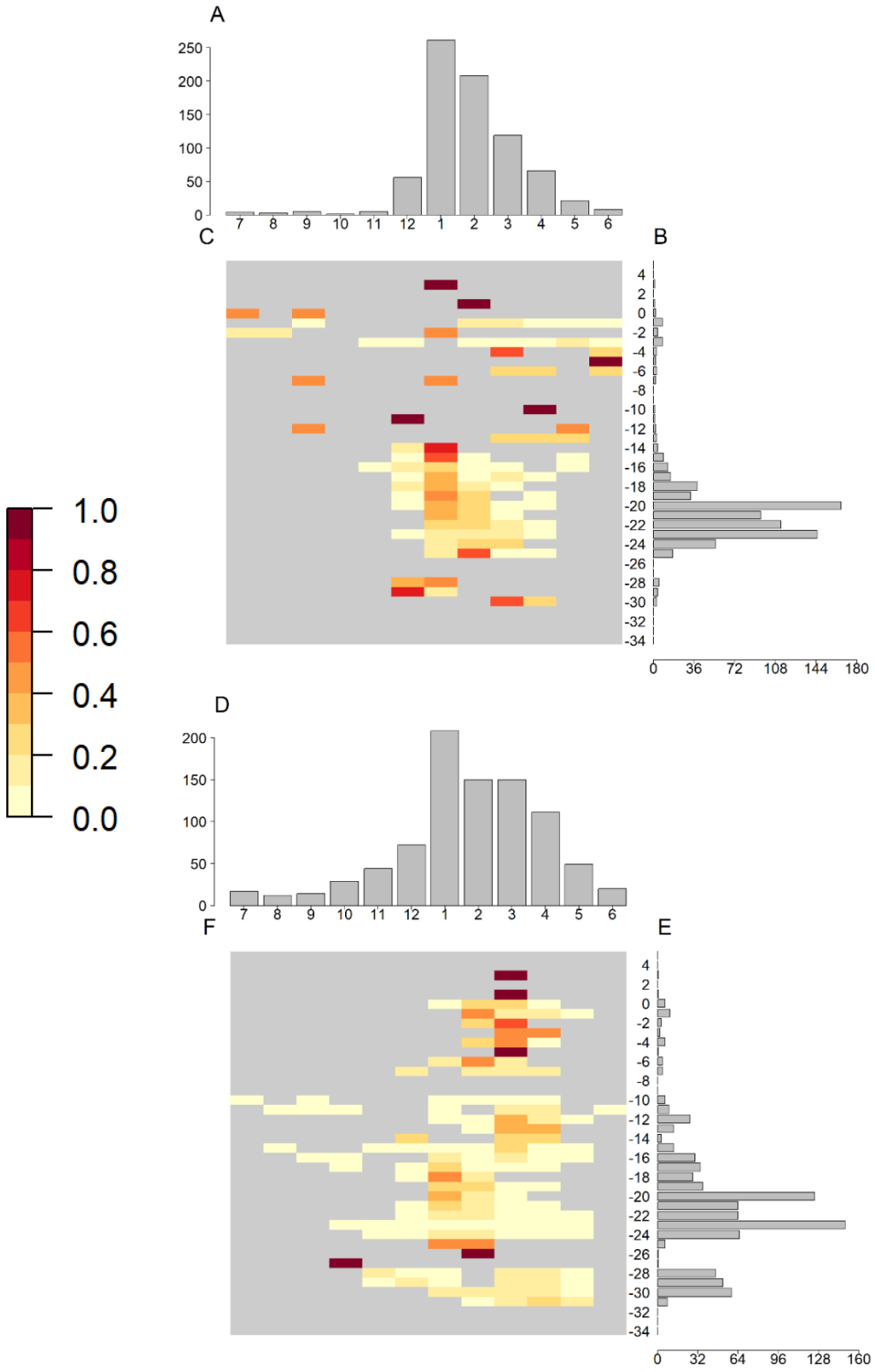


Figure 4.4.1 (A-C) The seasonality of human and (D-F) NHP YF reports by latitude in Brazil. (A, D) The number of human and NHP YF reports by month, (B, E) the number of YF reports by 1° latitude and (C, F) the proportion of cases by latitude and month. Colour scale shows the proportion of the latitudes cases that occur in a month.

4.4.2 Model fits and comparison of agricultural seasonality and climate/vegetation

Model fits varied across all covariate inclusions and classification report types. Generally, AUC scores for out-of-sample predictions of reports of human cases and of both human and NHP cases were higher than those of reports of NHP cases alone (Table 4.4.1). The best performing models, as measured by the Brier score, was 15. Model 15 contained all covariate groups. The best-fitting model which did not include agricultural seasonality was number 11, ranked 5th. The out-of-sample AUC for human reports of YF varied from 0.81 (0.74 – 0.81) in model 1, to 0.93 (0.89 – 0.93) in models 14 and 15. AUCs for NHP reports of YF were lower, ranging from 0.72 (0.66 – 0.72) in model 4 and 8, to 0.85 (0.80 – 0.85) in model 7 and 15. Municipalities that had both human and NHP reports of YF had out of sample AUCs ranging, from 0.81 (0.71 - 0.81) for model 4, to 0.95 (0.88 - 0.95) for the models 11, 14 and 15.

Out-of-sample predictive performance, as calculated using a spatial-block bootstrapping method was only slightly lower than within-sample performance for predicting the absence of reports or human-only reports, but somewhat worse for predicting only NHP or both human and NHP reports (Figure 4.4.2). Out-of-sample performance tracked within-sample performance for all models.

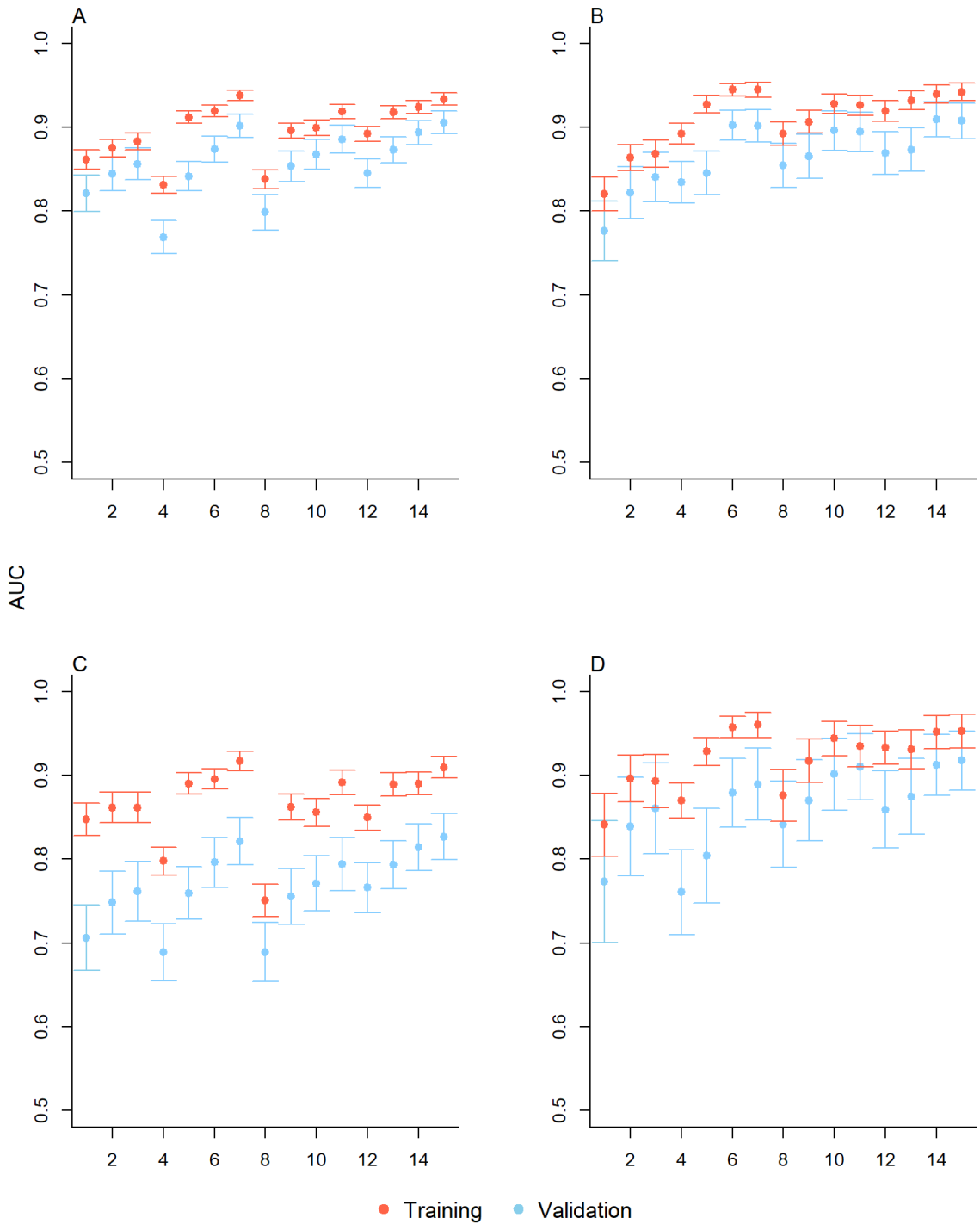


Figure 4.4.2. Comparison of the training and validation AUC values for the classification of a municipality as having no YF report (A), human YF report (B), NHP YF report (C) and human and NHP YF report (D). The x axis numbers refer to the models found in Table 4.4.1.

Table 4.4.1. Human and NHP YF report models by covariate grouping. The presence, 1, or absence, 0, of covariate groupings is shown with the corresponding out-of-sample AUC, Brier score and overall model rank. The best model according to Brier score are highlighted and emboldened.

AGRICULTURE OUTPUT	HOST	AGRICULTURE SEASONALITY	CLIMATE VEGETATION SEASONALITY	HUMAN REPORT	NHP REPORT	BOTH REPORT	BRIER SCORE	ROW	OVERALL RANK
1	0	0	0	0.81 (0.74 - 0.81)	0.75 (0.67 - 0.75)	0.85 (0.70 - 0.85)	0.019372	1	14
0	1	0	0	0.85 (0.79 - 0.85)	0.79 (0.71 - 0.79)	0.90 (0.78 - 0.90)	0.018919	2	11
1	1	0	0	0.87 (0.81 - 0.87)	0.80 (0.73 - 0.80)	0.92 (0.81 - 0.92)	0.019972	3	15
0	0	1	0	0.86 (0.81 - 0.86)	0.72 (0.65 - 0.72)	0.81 (0.71 - 0.81)	0.019138	4	12
1	0	1	0	0.87 (0.82 - 0.87)	0.79 (0.73 - 0.79)	0.86 (0.75 - 0.86)	0.018234	5	6
0	1	1	0	0.92 (0.89 - 0.92)	0.83 (0.77 - 0.83)	0.92 (0.84 - 0.92)	0.018325	6	7
1	1	1	0	0.92 (0.88 - 0.92)	0.85 (0.79 - 0.85)	0.93 (0.85 - 0.93)	0.017431	7	2
0	0	0	1	0.88 (0.83 - 0.88)	0.72 (0.65 - 0.72)	0.89 (0.79 - 0.89)	0.019236	8	13
1	0	0	1	0.89 (0.84 - 0.89)	0.79 (0.72 - 0.79)	0.92 (0.82 - 0.92)	0.018507	9	9
0	1	0	1	0.92 (0.87 - 0.92)	0.80 (0.74 - 0.80)	0.94 (0.86 - 0.94)	0.018393	10	8
1	1	0	1	0.92 (0.87 - 0.92)	0.83 (0.76 - 0.83)	0.95 (0.87 - 0.95)	0.018045	11	5
0	0	1	1	0.89 (0.84 - 0.89)	0.80 (0.74 - 0.80)	0.91 (0.81 - 0.91)	0.018588	12	10
1	0	1	1	0.90 (0.85 - 0.90)	0.82 (0.76 - 0.82)	0.92 (0.83 - 0.92)	0.017963	13	4
0	1	1	1	0.93 (0.89 - 0.93)	0.84 (0.79 - 0.84)	0.95 (0.88 - 0.95)	0.017841	14	3
1	**1**	**1**	**1**	**0.93 (0.89 - 0.93)**	**0.85 (0.80 - 0.85)**	**0.95 (0.88 - 0.95)**	**0.017422**	**15**	**1**

4.4.3 Seasonal trends in model predictions

While all covariate groupings captured the monthly seasonality of YF to a degree, they did so at differing levels of accuracy. The seasonality of human YF reports was generally better reproduced (R^2 of 0.64 in the best fit climate/vegetation seasonality model to 0.99 in the best fit including both types of seasonality), than the seasonality of NHP reports (R^2 0.64 in the climate/vegetation seasonality model to 0.98 in the models that included the agriculture of seasonality) or of reports of both human and NHP cases (0.65 in the agricultural seasonality model to 0.90 in models that included everything and climate/vegetation seasonality) (Table 4.4.2, Table 4.4.3 and Figure 4.4.3).

The best fit models that included the seasonality of agriculture provided a substantially better fit to the seasonality of human reports (Table 4.4.2, Table 4.4.3 and Figure 4.4.3). In particular they more accurately captured the magnitude of seasonality – something that the best fit model that only included the seasonality of climate/vegetation failed to account for. Models generally underpredicted reports of YF in months of heightened transmission, and marginally overpredicted during the “low season” (Figure 4.4.3).

At a national level, there is significant seasonality in YF reports (human, NHP and both), with 79.8% of all reports occurring January-March, and a minimum of 1 report in October, and 255 in January. The probability of a human report is minimal from July to October for all models and the data, while in November the best fit climate/vegetation seasonality model predicts a substantial increase in reports this is not reflected in the data or the other model predictions. A rise in actual reports, from 4 to 55, and predicted reports in December, with the climate/vegetation seasonality

model overpredicting the number of reports. January sees a significant increase in the reporting of reports, rising from 55 to 255 reports, followed by a fall to 194 in February, a trend which is accurately followed in all models' predictions, apart from the climate/vegetation seasonality model which underpredicts substantially.

NHP reports follow a weaker seasonal pattern than human reports, with comparatively higher levels of reporting seen across the year, with the minimum of 12 reports occurring in August, and the maximum of 181 in January. Model predictions follow a similar pattern to predicting to the human reports, with the climate/vegetation model predictions consistently over-predicting the "low-season" months of June-December, and under-reporting the peak of January-April, with other models generally performing well.

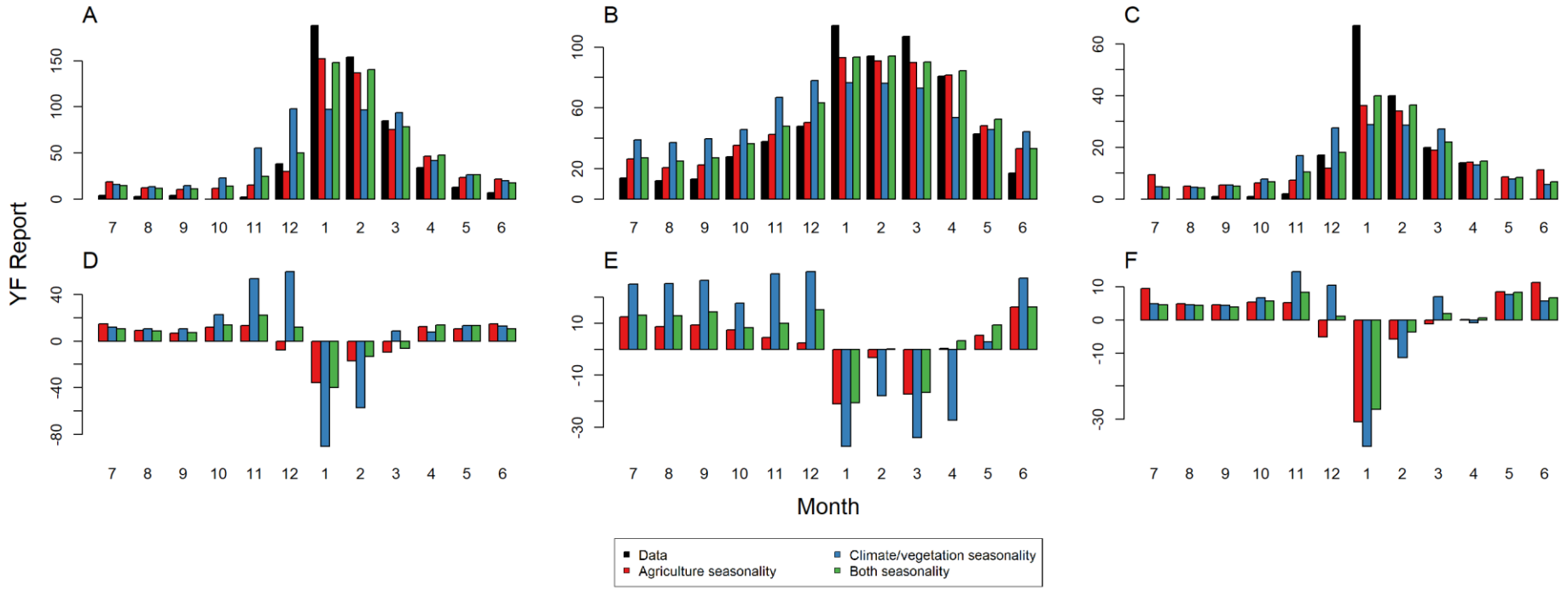


Figure 4.4.3. Total monthly YF reports and in-sample model predictions (A and D) for humans, (B and E) NHPs and (C and F) both classifications. The top row (A, B, C) depicts the overall monthly data and model predictions for each classification type. The bottom row (D, E, F) show the residuals. Results are shown for the best fit model including agricultural (but not climate) seasonality (model 7), climate (but not agricultural) seasonality (model 11) and both forms of seasonality (model 15). Models were fitted to all the data, so within-sample predictions are shown.

Table 4.4.2. Absolute total deviances between YF reports and within-sample model predictions (for models fitted to all the data) by covariate grouping. Results are shown for the best fit model including agricultural (but not climate) seasonality (model 7), climate (but not agricultural) seasonality (model 11) and both forms of seasonality (model 15).

COVARIATE GROUPINGS	MONTHLY DIFFERENCE FROM DATA (TOTAL YF REPORTS – TOTAL MODEL PREDICTIONS		
	Human	NHP	Both
EVERYTHING	171.8	140.0	76.7
AGRICULTURAL SEASONALITY	163.7	108.1	92.5
CLIMATE/VEGETATION SEASONALITY	358.8	300.0	116.8

Table 4.4.3. R^2 values comparing within-sample model predictions (for models fitted to all the data) with the data by covariate grouping. Results are shown for the best fit models including agricultural (but not climate) seasonality (model 7), climate (but not agricultural) seasonality (model 11) and both forms of seasonality (model 15).

COVARIATE GROUPINGS	MONTHLY PREDICTIONS CORRELATION		
	Human	NHP	Both
EVERYTHING	0.98	0.96	0.93
AGRICULTURAL SEASONALITY	0.99	0.98	0.90
CLIMATE/VEGETATION SEASONALITY	0.64	0.64	0.65

4.4.4 Geographical distribution of YF reports

Reports of YF in all classifications are found throughout much of the country, with the exception of the North East of Brazil (Figure 4.4.4).

Notable hotspots for human reports are seen in the South-Eastern Atlantic states of Brazil, Western and Amazonian states. NHP reports are more widely spread throughout the country, with reports in states without human cases such as Bahia and Tocantins. Municipalities with both human and NHP reports reflected the distributions of human and NHP reports, with much of Espírito Santo and large areas of Sao Paulo state recording both human and NHP reports.

The best fit model (number 15) reproduced all these patterns well. The predicted pattern of human reports largely matched the data, with the exception of predictions of higher in the North West states that constitute part of the Amazon where cases have not yet been reported.

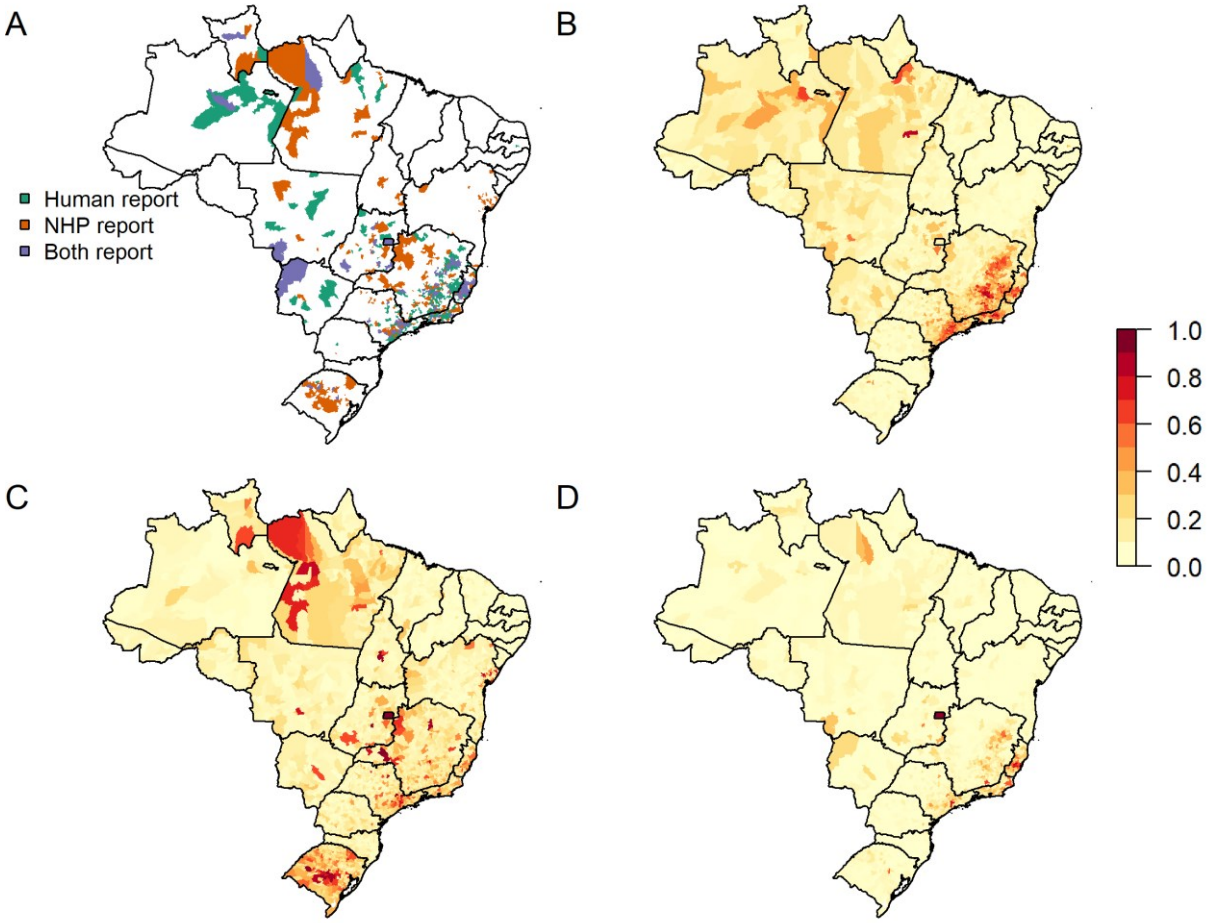


Figure 4.4.4. (A) Aggregate reports of the data for human, NHP and both reports model predictions for the probability of classifying an administrative location as (A) only having human reports, (C) only NHP reports and (D) both human and NHP reports. Model predictions are from the best fit model with all covariates (model 15).

4.4.5 Variable importance comparisons for best fitting models

I assessed variable importance ranks for the best-fitting models with only agricultural seasonality (number 7), only climate seasonality (number 11) and with both (number 15), focussing on the most important 50% of covariates in each case.

For the model which had both seasonal and vegetation/climate seasonality, model 15, 13 of the 22 most significant covariates were in the vegetation/climate grouping, including the 4 most influential covariates (Figure 4.4.5A). Agricultural covariates (both output and seasonality) made up 7 of the top 21, namely the number of wheat, corn, soya and bean farms in a municipality, planting of peanuts and harvesting of rice and the proportion of the population in agriculture. The number of NHP species present and, the logarithm of the total population were also both in the top 21.

The model including agricultural seasonality, model 7, 8 of the 14 most important covariates were related to agricultural output, including the 2 most important covariates (Figure 4.4.5B). All host demographic covariates were found, with the logarithm of the rural population and the number of NHP species particularly important. Agricultural seasonality covariates comprised 4 of the 14, the harvesting of rice and peanuts, and the planting of peanuts and beans.

Finally, the model with vegetation/climate seasonality, model 11, had vegetation/climate seasonality covariates as 11 of the top 14 covariates, including the most important 3, night temperature, day temperature and the diurnal/nocturnal temperature range (Figure 4.4.5C). Agricultural output made up 2 and Host demographics the final.

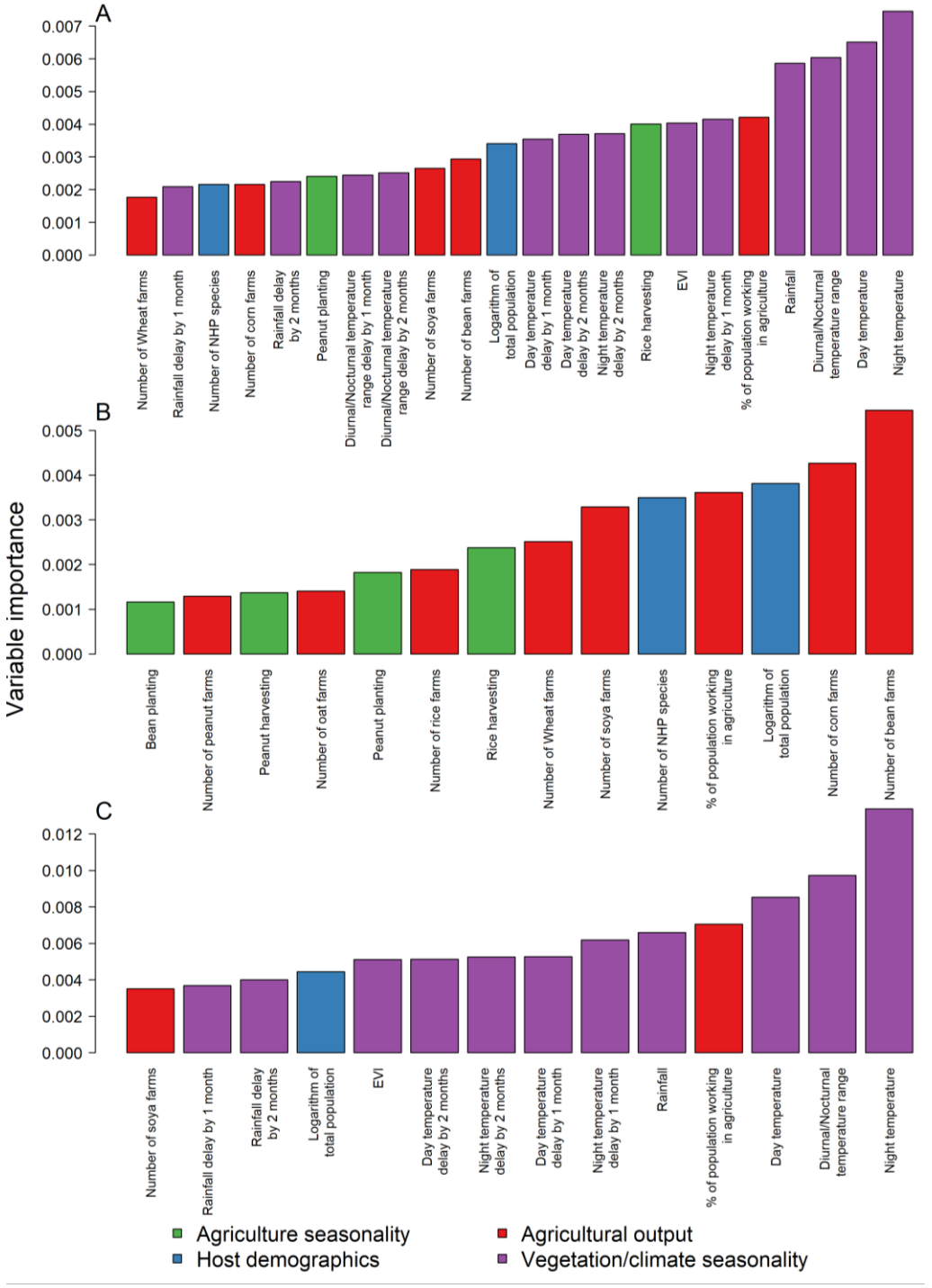


Figure 4.4.5. Variable importance values for (A) the model with both agriculture and vegetation/climate seasonality, model 15, (B) the model with agricultural seasonality but not climate seasonality, model 7, and (C) the model with vegetation/climate seasonality but not agricultural seasonality, model 11. Only covariates in the top 50% of variable importance are shown.

4.4.6 Partial dependence plots

Here I have described the partial dependence plots (PDP) of the top 50% of variable importance for the overall best performing model (model 15). Partial dependence plots show how these covariates influence the outcome across their range of values (Figure 4.4.6).

Partial dependence plots show how these importance variables influence the outcome across their range of values. Generally, following an initial dip for the number of bean and corn farms, increases in the number of farms are associated with increased probabilities of classifying a municipality has possessing a YF reports in NHP's and both human and NHP, with human reports either remaining relatively unaffected or decreasing over the range of values.

The number of NHP species is associated with large increases in the probability of classifying a municipality as reporting all report types of YF, apart from an initial dip in the probability of an area being classed as reporting both human and NHP cases, but quickly plateaus around 8 species. Higher percentages of the population working in agriculture are negatively associate with all YF reports, the logarithm of the total population is associated with a rapid increase in all types of reports, particularly human. The harvesting of rice and planting of peanuts is positively associated with all types of YF reporting.

As rainfall increases, the probability of YF reports increases, most substantially in areas that report both human and NHP reports. Day temperature's effects on the probability of a report of YF is highest between 20 and 23°C and 37 to 45°C. Increasing night temperatures reduce the probability of a human NHP report, till around 24 °C where it quickly rises. NHP and both human and NHP reports steadily rise over the

range of values. Increases in the range of diurnal/nocturnal temperature are negatively associated with all reports until around 15 when it rises.

Lower levels of EVI are more positively associated with YF reports, with a slight raise at higher values. Generally, delayed rainfall, temperatures, temperature ranges and EVI follow a similar pattern, with slight deviations between the relative influences on the report classifications, with heightened levels at low values and high values across the ranges.

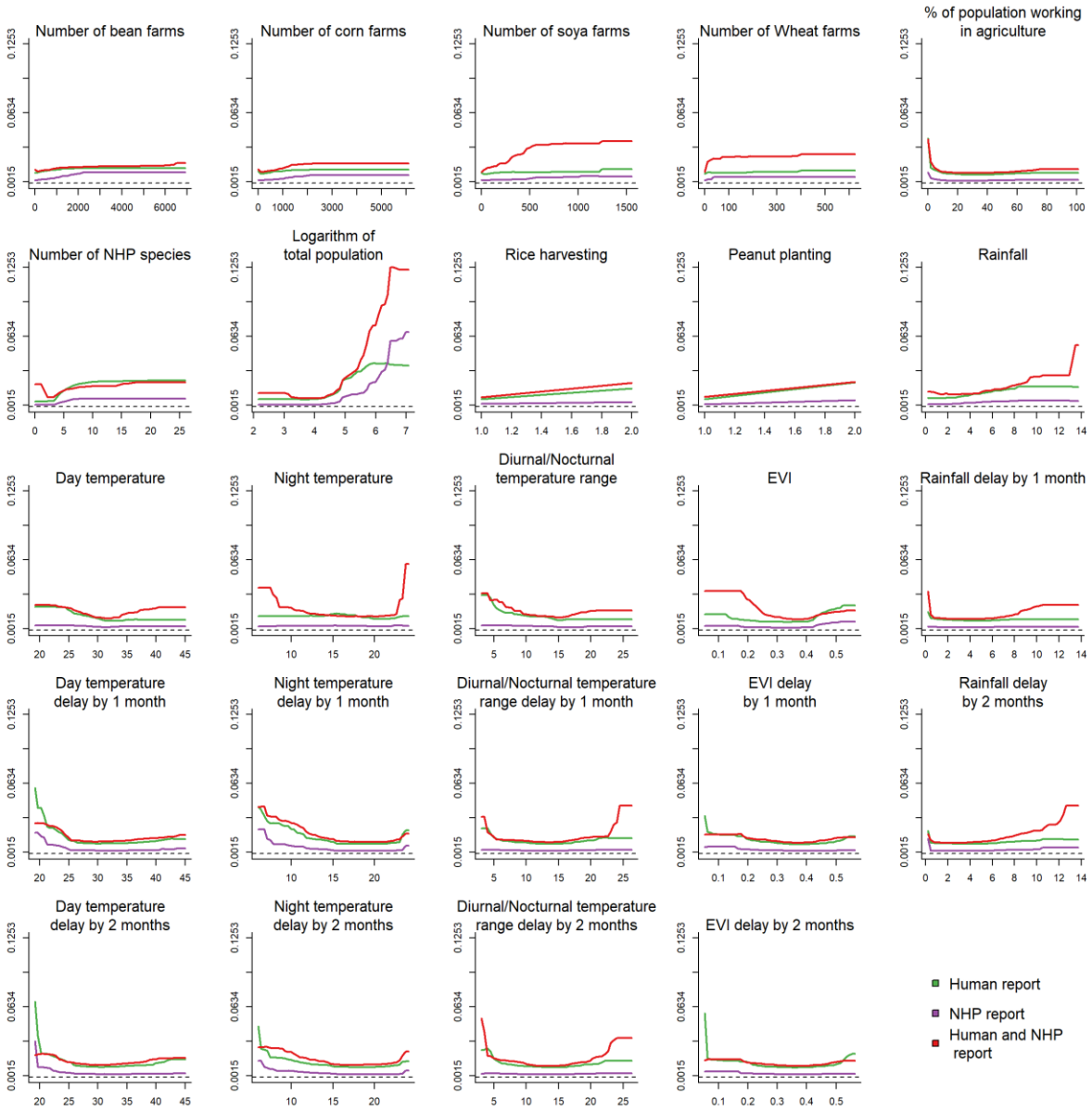


Figure 4.4.6. Partial dependence plots for the covariates in the top 50% of variable importance of the model that included agricultural seasonality but not vegetation/climate seasonality. The y axis on the right the probability of human, NHP and human and NHP reports.

4.5 Discussion

I have identified the highly seasonal nature of YF reporting in both humans and NHP's, as well as demonstrating the relative predictive power of utilising covariates related to the seasonality of climate and the seasonality of agriculture. All model fits accurately captured the seasonality of reporting in humans and NHP's, though models fit to reports of YF that included humans performed significantly better. Models that included the seasonality of agriculture had a significant and substantial improvement in their ability to predict aggregate monthly human reports (aggregate monthly R^2: 0.99 vs 0.64) (Table 4.4.1 and Table 4.4.3*)*. Our findings illustrate the importance of the seasonality of exposure, and that it is not necessarily just an increased viral transmission in zoonotic reservoirs which leads to spillover, but also an increased interaction with the sylvatic cycle. In addition to this I have highlighted the individual role of different crop types such as peanut and rice planting/harvesting on increasing the probability of YF reporting.

While the link between agriculture and disease has long been highlighted, there has been little work done on how the seasonality of exposure relates to increased disease transmission, and even less for yellow fever. The increased predictive performance of models fit to the seasonality of agriculture over climate is an important finding for predicting zoonotic spillover into human populations. Despite climate dictating the environmental suitability for vectors, and so likely increasing viral transmission within the sylvatic reservoir, it is covariates that indicate an increased exposure to the sylvatic cycle that appear more determinant. The importance of human-animal contact in zoonotic spillover has previously been highlighted as a significant determinant of spillover events (Olival et al., 2017, Plowright et al., 2017), though this has not been

explored in the context of seasonality of agriculture as a driver of exposure for vector-borne diseases.

Though further study is needed to establish a mechanistic understanding of how these agricultural activities increase exposure to the sylvatic cycle, the associations highlighted here represent an important first step. Agricultural activities in Brazil, and much of the world, are rapidly changing. Rising populations and the growing of “cash” crops, such as rice and soya, for exportation are changing the agricultural ecosystem, as well as driving deforestation and general habitat conversion (DeFries et al., 2013). These changes are likely to lead to both short term effects, and long-term changes in the epidemic and endemic potential of numerous diseases – particularly those with a zoonotic component (Allen et al., 2017, Patz et al., 2004, Faust et al., 2018). Our findings suggest that these changes, in addition to changing the overall suitability of a habitat, may even change the relative seasonality of spillover. Following the initial disruptions to the sylvatic cycles that are brought on by land conversion, the regular transformation of the landscape, through planting and harvesting, as well as the increased interaction of humans to this habitat appear to increase the risk of YF spillover for a number of crop types and activities. This “anthropogenic seasonality” may have public health consequences for surveillance and further transmission. Due to the strong seasonality typically associated with changes in rainfall and temperature (Kumm, 1950, Hamlet et al., 2018), surveillance efforts for YF are generally focused between November and April in Brazil with priorities shifted towards other diseases outside of this time period. However, if this seasonal spillover is not purely dictated by climate as was previously believed, then human transmission may be occurring undetected at higher levels than currently suspected outside of the traditional seasonal period. Undetected and unopposed spillover into humans additionally raises the risk

of establishment in *Aedes aegypti* populations – potentially sparking urban epidemics which have historically spread rapidly, been hard to contain, and exported international (World Health Organization, 2017b, Wasserman et al., 2016).

While climate and agriculture are intrinsically linked, with different activities occurring at times where climate favours growing and harvesting, they are not equivalent. In addition to the significant differences in model fit, climate and agricultural activity covariates only show moderate correlation. This is partly because agriculture is not solely decided by the climate, with anthropomorphic adaptations such as the use of irrigation, fertilisers and herb/pesticides allowing for an ever-increasing detachment from seasonal farming. This, in conjunction with the binary nature of our agriculture seasonality data, suggests that I am investigating two separate processes and not simply overfitting to climate driven data.

Though in and out-sample AUC values of models fit to human reports of YF were not significantly different, AUC values for NHP predictions showed larger differences, indicating potential overfitting. This suggests that these models do not capture the underlying transmission dynamics as well for NHP's as they do for humans, an unsurprising finding given the differences expected in human and NHP exposure to the sylvatic cycle. Additionally, this result may relate to the surveillance of NHP cases. As the number of humans entering habitats suitable for NHP's and sylvatic mosquitoes increases, then there will be an increased observation of NHP's and so the probability of detecting an NHP YF report will increase. Therefore, the relationship I have captured may be more related to a seasonal increase in surveillance, a finding which could be used to correct for seasonal variations in NHP report detection biases related to exposure to sylvatic habitats.

Reporting biases may also be introduced by inaccurate geolocation based on where the human cases were treated or diagnosed, not where infection was acquired. This is likely heightened in rural areas where healthcare resources are limited, and cases are transported to a regional hospital with increased capacity. Furthermore, this reporting bias likely changes over the year, and likely has a greater effect on NHP reporting than human reporting, due to humans being the ones who report NHP YF. Overall, for humans, this bias would likely disengage a link between agricultural activities and YF reports, in contrast with the strong relationship between the two I have found, and so the effect of this is likely minimal.

Future work may be able to expand this work through increased detail on the seasonality of agricultural activities and reporting. Here I have only been able to collate presence/absence for limited crop types at the first administrative division, which despite its limitations still offers substantial improvements to models (Table 4.4.1 and Figure 4.4.3*)*. Additional quantification of the scale of agricultural activities, additional crop types, and at a higher spatial resolution may reveal further relationships between agriculture and YF reporting.

In conclusion, our analysis represents an important first look into the relationship of the seasonality of agriculture and yellow fever, as well as other arboviruses. By identifying the types of agriculture and crops associated with YF transmission, this work has direct and immediate applicability. Through targeting vaccination campaigns and surveillance activities towards areas, and time-periods, most at risk of spillover, we can more accurately and effectively prevent human YF before it occurs. This increased understanding of YF spillover is particularly important in the context of limited resources (Barrett, 2016), and a globally changing epidemiology of YF (Makhani et al., 2019).

Chapter 5 - Population-level vaccination coverage estimates for the YF endemic zone in South America and Africa (1940-2050) and their implementation into an interactive web-platform POLICI

The work in this chapter formed the basis of: Hamlet A, Jean K, Yactayo S, Benzler J, Cibrelus L, Ferguson N, et al. POLICI: A web application for visualising and extracting yellow fever vaccination coverage in Africa. Vaccine. 2019;37(11):1384-8.

I would like to acknowledge the prior contributions of Kevin Jean and Tini Garske in developing the original framework for vaccination coverage estimations. The work undertaken here by myself substantially expanded and rewrote the underlying code and generation process, as well as the creation of the online platform which was solely developed by myself.

5.1 Summary

Recent yellow fever (YF) outbreaks have highlighted the increasing global risk of urban spread of the disease. In context of recurrent vaccine shortages, preventive vaccination activities require accurate estimates of existing population-level immunity. I present POLICI (POpulation-Level Immunization Coverage – Imperial), an interactive online tool for visualising and extracting YF vaccination coverage estimates in Africa. I calculated single year age-disaggregated sub-national population-level vaccination coverage for 1950–2050 across the African endemic zone by collating vaccination information and inputting it into a demographic model. This was then implemented on an open interactive web platform.

5.2 Introduction

Historically sporadic and scattered YF vaccination activities have led to geographic and demographic heterogeneities in population-level vaccination coverage. These heterogeneities can complicate the planning of further vaccination campaigns as it can be inefficient to target all locations and age groups within a country equally if there is substantial pre-existing immunity in some sub-populations. Due to the lifelong immunity conferred by the vaccine, all previous vaccination activities contribute to contemporary population-level immunity.

Here I have built on previous work first applied by (Garske et al., 2014), who systematically collated information on past YF vaccination activities and combined this with demographic information to estimate population-level vaccine-derived immunity coverage in the African endemic zone to provide past and future coverage estimates from 1940 to 2050, as well as expanding the work to South America. Interpreting vaccination coverage data for public health or policy often requires considering various specific stratifications (by time, location and age group) that static tabular presentations do not effectively communicate.

In order to improve the easy of access, use and dissemination I have developed the online tool POLICI (POpulation-Level Immunization Coverage – Imperial), developed in the programming language R (Team, 2015a). I showcase both disaggregated estimates of YF vaccination coverage across the endemic zone of Africa and South America, as well as a web-based interface which offers an interactive and accessible way for global and public health professionals to utilise this information without specialist knowledge or software.

5.3 Methods

The primary aim of this application is to provide an accessible, easy-to-use tool to visualise the coverage of YF vaccination across Africa through time. In order to achieve this, I placed an emphasis on interactivity which I attained by combining the R packages Shiny, Leaflet and Plotly (Beeley, 2013), to create interactive maps and graphs in order to allow the user to customise plots of population-level vaccination coverage, and to download the underlying datasets. https://shiny.dide.imperial.ac.uk/polici/

5.3.1 Demographic data

I used UN World Population Prospects (UN WPP) data (United Nations, 2016) to inform the size of any annual birth cohorts between 1940 and 2050 for African and South American countries considered endemic or at risk for YF (Jentes et al., 2011).

In order to disaggregate these country wide estimates of population and age breakdown, I aggregated gridded population estimates from LandScan 2015 (Bright et al., 2015) to the province or district (first and second sub-national administrative unit, respectively) as has been further detailed in previous chapters. This was then used to scale the national population size through time with the proportion of the population estimated for each province or district to achieve higher spatial resolution, assuming the national age distribution throughout.

5.3.2 Vaccination data

YF vaccination activities were collated from various sources on (i) historical large-scale vaccination activities 1940-1960s (Durieux, 1956), (ii) outbreak response campaigns since 1960s reported in the WHO Weekly Epidemiological Record (World Health

Organization), and WHO Disease Outbreak News (World Health Organization), (iii) routine infant YF immunization as reported by the WHO/UNICEF Estimates of National Immunization Coverage (WUENIC)(WHO, 2016), (iv) preventive mass vaccination campaigns conducted as part of the Yellow Fever Initiative (YFI)(World Health Organization, 2016c), and (v) communications with individual Ministries of Health, such as the Brazilian Ministry of Health which has provided the number of YF vaccine doses distributed at the municipality level for 1994 – 2018.

This data is provided in a vast array of formats and resolutions and requires substantial effort to homogenise and geo-locate. This was done on a case-by-case basis, where locations and the number of vaccine doses, targeted populations or various survey coverages were extracted with the year provided and the location specified. Where locations were named, they were cross-referenced with country level maps and other information located in the text to assure accurate geo-location. This also ensured that historical changes in names and administrative zones were considered when allocating vaccination activities.

I compared the resulting list of activities with data from the WHO International Coordinating Group (ICG) on Vaccine Provision, and the Brazilian Ministry of Health to resolve any discrepancies. By drawing vaccination coverage activities from a wide variety of sources, including the grey literature, and validating these with the help of the ICG, our dataset represents a very robust and inclusive catalogue of previous vaccination activities in Africa and South America.

Vaccination activities were resolved to the district level while recording the magnitude of the activity (coverage achieved, target population or doses administered), as well as the age groups targeted.

5.3.3 Estimation and visualisation of population-level vaccination coverage

The completeness and available detail of information regarding vaccination activities was highly variable. When provided, I used information on the coverage achieved in a specific area, either estimated through post-campaign surveys or administratively reported (with a preference given to the former as this is generally considered as more reliable). In the absence of information on coverage, I estimated the coverage by dividing the number of doses administered by the population size of the area targeted.

Unless provided with additional information, I assumed that all targeted age groups had an equal chance to receive vaccination, and I used a variety of approaches to combine subsequent vaccination activities targeting the same subpopulations dependent on the situation and available information. As a default I assumed no correlation between subsequent vaccination activities, so the combined coverage from two subsequent activities was calculated as

$$vc_{combined} = 1 - \frac{(1 - vc_1)}{(1 - vc_2)} \text{ [1].}$$

This approach was applied when precise geographic information on the target area and magnitude of the campaign were provided. In some cases, in particular for large preventative campaigns that were performed in stages over several years, information was only available at the scale of the overall target area, and no further information was available regarding the sequence of implementation across sub-areas. I assumed that subsequent campaigns would target those previously unvaccinated, and therefore the resulting coverage was given as

$$vc_{combined} = vc_1 + vc_2(1 - vc_1) \text{ [2].}$$

For routine infant immunization I used WUENIC data and assumed vaccination of the yearly cohorts at 1 year of age for the entire respective country without subnational heterogeneities and without prior coverage of the respective cohort.

Combining routine immunization and campaigns, I calculated the achieved vaccination coverage by age and province for each year between 1940 and 2050 by incorporating information on YF vaccination activities into our demographic model. For onward projection after 2019, I assumed that routine infant immunization coverage was maintained at the 2019 levels, with no further vaccination campaigns. For each province, time-varying population-level vaccination coverage was calculated by aggregating the age-specific vaccination coverage for all age groups.

To disseminate these estimates the online tool is made available at https://shiny.dide.imperial.ac.uk/polici/

5.4 Results

5.4.1 POLICI application

In order to aide visualisation and to disseminate these population-level vaccination coverage estimates an interactive web-based tool was created to aid sharing and utilisation. This application was termed POLICI (POpulation-Level Immunization Coverage – Imperial) and hosted on an Imperial server.

Here users are able to explore vaccination coverage from 1940 to 2050 at a subnational level and explore how this change across ages interactively. Furthermore, the maps and tables presented on the application are available for download in order to facilitate onward usage.

For the previously published paper that includes an example of how to utilise the application please see **Appendix A.**

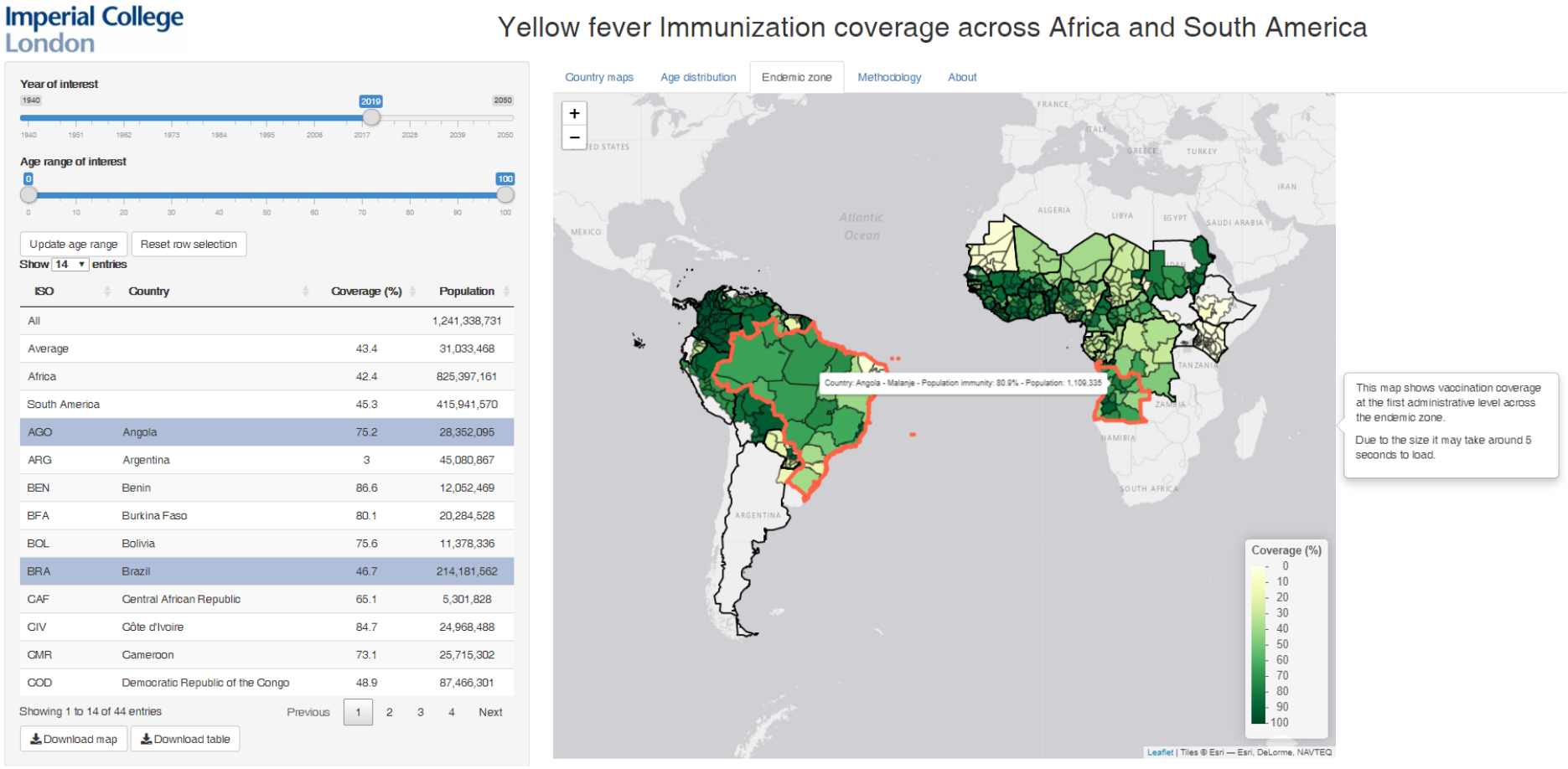


Figure 5.4.1. Example of a visualisation of the population-level YF vaccination coverage of the YF endemic zone in 2019 from the POLICI application.

5.4.2 Historical coverage of YF across the Endemic zone

YF vaccination activities began in the late 1930's and early 1940's across several countries in South America and French controlled African colonies which resulted in relatively high levels of vaccination coverage by the 1950's, however outside of these regions, reported vaccination coverage was minimal (Figure 5.4.2). In many of these countries' coverage remained high, particularly in French West Africa, till the discontinuation of the French neurotropic vaccination, due to adverse effects, and withdrawal of the French from their colonies led to a cessation of mass vaccination campaigns. In Africa, the 1980's and 1990's saw a resurgence of YF epidemics, following this reduction in population immunity. Outside of Africa, coverage remained fairly constant.

The following decades till 2005 saw differing uptakes in coverage between Africa and South America. While overall coverage reduced, West Africa, and parts of Nigeria, saw the implementation of reactive campaigns to many of the outbreaks since 1985. This resulted in an increasingly patchwork mosaic of coverage across much of West Africa. Though outside of these reactive campaigns, minimal activities occurred the adoption of routine infant immunization in many national control programmes began to slowly increase the population-level immunity. In South America, vaccination coverage substantially increased in many countries, with above 80% found in many endemic zones – with the exception of those in Brazil which had only recently become classed as endemic. Following the implementation of the Eliminate Yellow Fever (EYE) strategy, coverage quickly increased in much of West Africa, and by 2010 coverage was substantial across most of West Africa, as well as in Cameroon and the Central African Republic had raised coverage in central Africa. 2015 saw further

increases in South America, with most of the endemic areas for YF above 80% coverage, and a further improvement of the coverage in West and Central Africa.

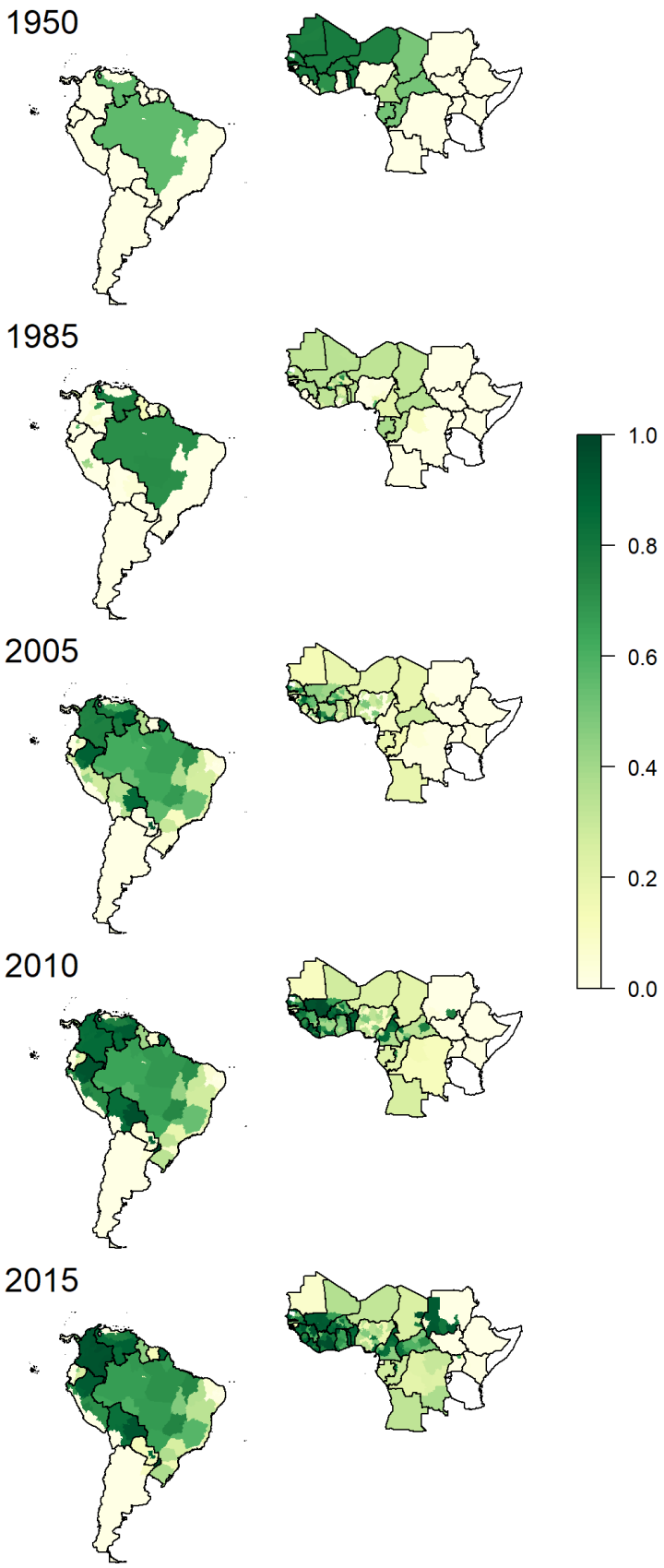


Figure 5.4.2. Population-level vaccination coverage across the YF endemic zone in 1950, 1985, 2005, 2010 and 2015.

5.4.3 Heterogeneities in coverage: Sub-nationally and by age

While coverage at the national level may appear sufficient to prevent further transmission, this is not always the case at the sub national level, or even within this there may be heterogeneities in coverage by age (Figure 5.4.3).

An example of this can be seen in the vaccination coverage of Burkina Faso in 2019. While the national average vaccination coverage is 80.1%, several states have coverage substantially below this, with the lowest found in Plateau-Central which only has a coverage of 52.8%. Further heterogeneities are seen within the age specific coverage of this state. While coverage is substantial in both those ages that were included in the mass vaccination campaigns of the 1960's and 1970's, and children who were included in routine-infant immunization, ages between these two events have very low rates of vaccination. Those between 34 and 60 years of age are predicted to have no population-level vaccination coverage, indicating a vulnerability in coverage – particularly given working age males are the primary victims of YF.

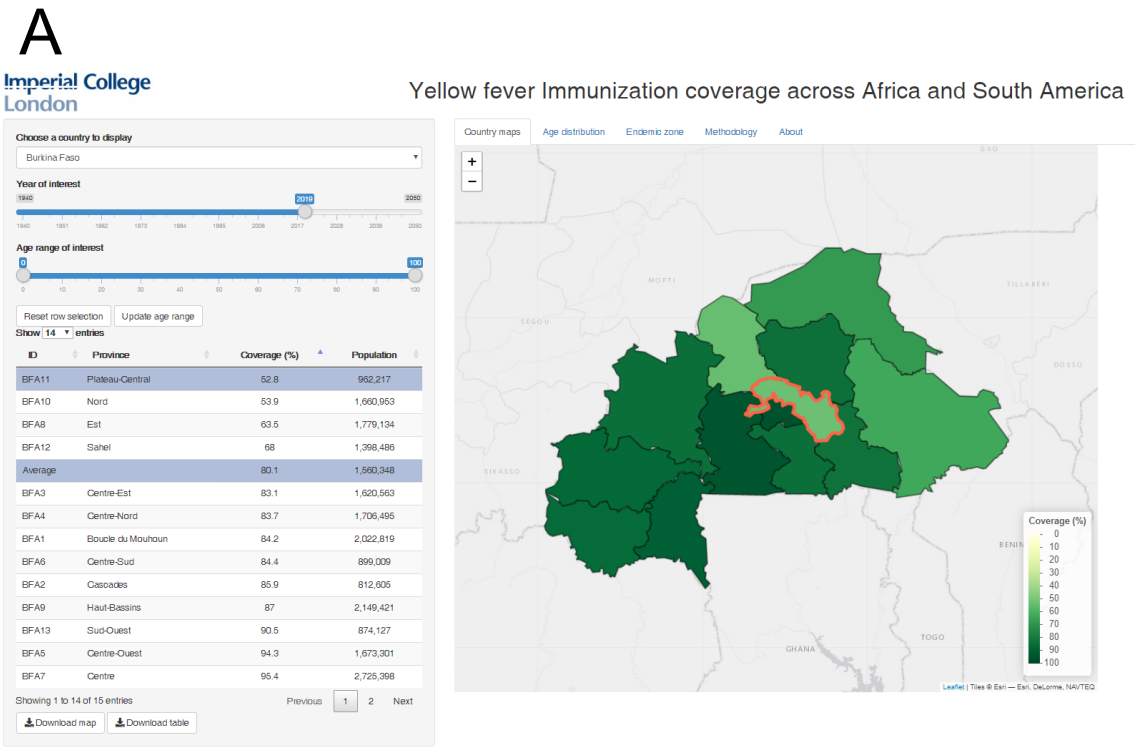


Figure 5.4.3. (A) Subnational estimates of population-level vaccination coverage for Burkina Faso in 2019 with the province with the lowest vaccination coverage (Plateau-Central) highlighted and (B) Plateau-Central age specific vaccination coverage. Figures adapted from the application POLICI https://shiny.dide.imperial.ac.uk/polici/.

5.4.4 Population-level vaccination coverage in 2019: Successes and shortfalls

Population-level vaccination coverage in Africa and South America shows substantial heterogeneity across its endemic zones (Figure 5.4.4A). While the low levels of vaccination coverage in some areas that are considered endemic but “low risk” such as East Africa, much of Ecuador, Peru, Bolivia West of the Andes as well as large parts of Argentina, there remains significant areas of high population within areas of “high risk” transmission with low population-level vaccination coverage (Figure 5.4.4B).

In Africa, a coverage of 42.4% translates to 475,000,000 unvaccinated people – of which Ethiopia contributes a substantial proportion (~109,000,000) which is in a low risk zone, but Nigeria, a country in an area of high YF risk and with a history of urban outbreaks, only has a 43.8% coverage and ~113,000,000 unvaccinated people (Table 5.4.1). South America has a similar coverage at 45.3%, and ~227,000,000 unvaccinated. In Brazil, the largest country in South America, ~114,000,000 are unvaccinated.

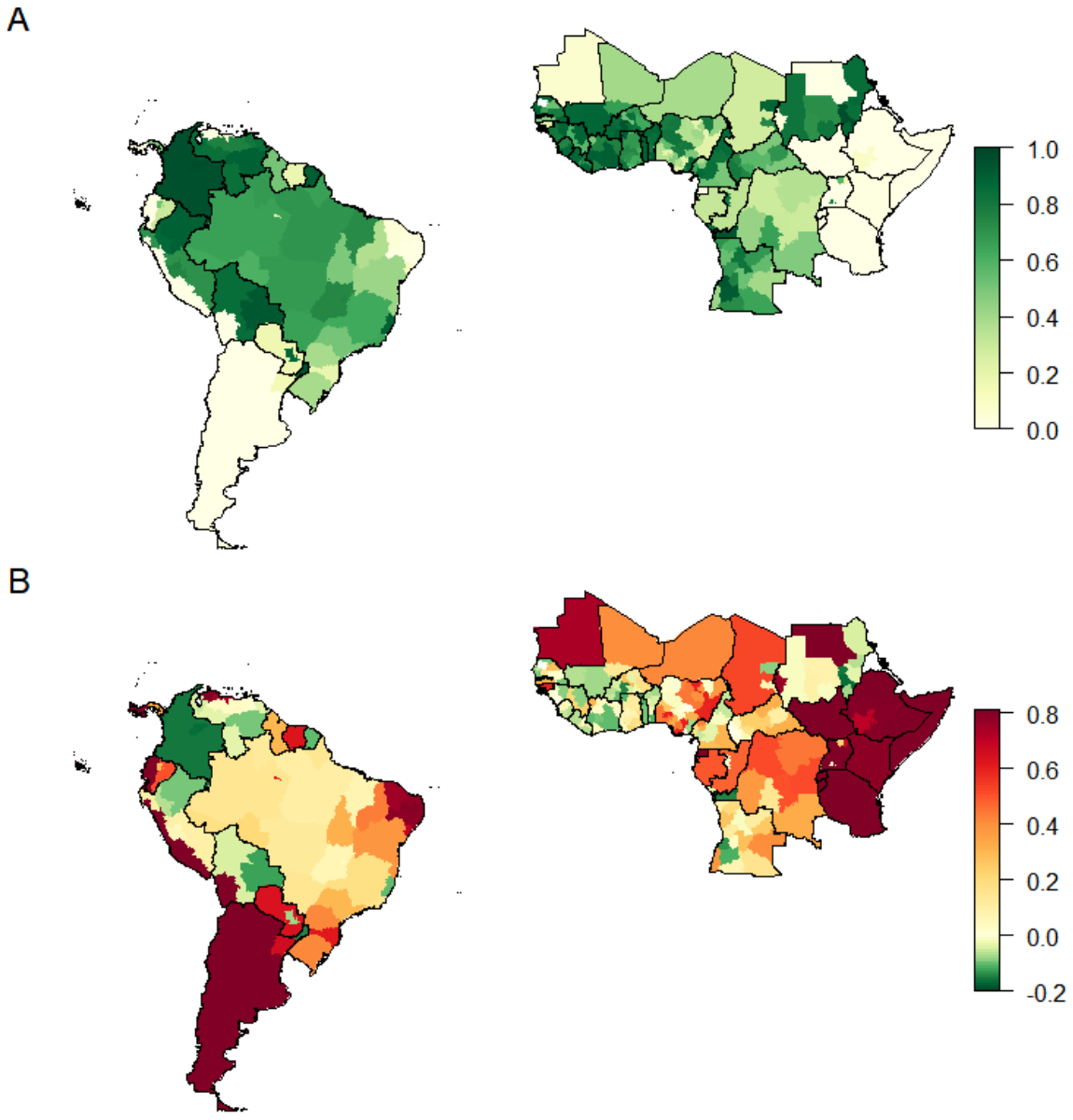


Figure 5.4.4. (A) The population-level vaccination coverage in 2019 for both endemic Africa and South America and the (B) difference in the WHO stated goal of 80% coverage and the population level vaccination coverage.

Table 5.4.1. Population and vaccination coverage across the YF endemic zone

Country	***Population***	***Vaccination coverage (%)***	***Unvaccinated***
Angola	28,352,095	75.2	7,040,262
Argentina	45,080,867	3	43,712,554
Benin	12,052,469	86.6	1,610,035
Burkina Faso	20,284,528	80.1	4,026,522
Bolivia	11,378,336	75.6	2,772,010
Brazil	214,181,562	46.7	114,160,638
Central African Republic	5,301,828	65.1	1,848,863
Côte d'Ivoire	24,968,488	84.7	3,822,337
Cameroon	25,715,302	73.1	6,905,576
DRC	87,466,301	48.9	44,652,133
Republic of Congo	5,125,793	47.2	2,708,018
Colombia	49,802,583	95.5	2,261,534
Ecuador	17,092,626	1.7	16,797,103
Ethiopia	109,402,291	1.9	107,362,422
Gabon	1,876,370	31.5	1,284,389
Ghana	29,897,823	70.1	8,926,948
Guinea	13,992,346	78	3,083,415
Gambia	2,253,865	80.5	440,048
Guinea-Bissau	2,021,012	21.8	1,580,655
Equatorial Guinea	943,571	0	943,571
French Guiana	295,953	91	26,526
Guyana	777,427	51.4	377,605
Kenya	50,920,368	1.7	50,045,861
Liberia	4,966,090	83.4	822,854
Mali	19,850,769	75.2	4,929,831
Mauritania	4,467,660	5.7	4,211,264
Niger	23,360,162	39.5	14,140,633
Nigeria	201,710,601	43.8	113,294,740
Panama	4,167,643	21.7	3,263,627
Peru	32,927,176	28	23,723,432
Paraguay	6,971,643	40.2	4,166,780
Sudan	44,243,626	58.7	18,287,266
Senegal	16,999,419	76.3	4,036,087
Sierra Leone	7,017,059	82.1	1,259,520
South Sudan	13,774,398	0.1	13,755,050
Suriname	557,659	19.1	451,189
Chad	15,929,676	32.8	10,709,011
Togo	8,089,283	89.9	815,806
Uganda	44,413,968	4.2	42,545,184
Venezuela	32,708,095	51.5	15,857,272
Africa total	**825,397,161**	**42.4**	**475,088,301**
South America total	**415,941,570**	**45.3**	**227,570,270**
Total	**1,241,338,731**	**43.4**	**702,658,571**

5.5 Discussion

Vaccine-related data such as target population, doses administered, or coverage achieved are typically available at the level of individual vaccination activities. However, the protection of populations afforded by the vaccine depends on the coverage at population level, which is a result of the history of vaccination activities a population was exposed to. I calculate this population-level coverage by consolidating this history in a demographic model. The POLICI tool provides an easy way to visualize these combined estimates, but also to the underlying estimates, interactively across the endemic zone as well as for any sub-geographies or specific age groups at any point in time.

In Africa, since the start of the YFI in 2006, coverage has increased substantially across the continent – particularly in areas considered at a heightened risk transmission, such as West Africa, which has seen population-level vaccination coverage raised above 80% in much of this region. The success of these preventive and responsive campaigns has seen large parts of West Africa achieving up to an 82% reduction in deaths (Garske et al., 2014). However, the magnitude of these successes in West Africa have not been replicated across the rest of the continent, with several countries with low vaccination coverage, reporting outbreaks of YF in recent years, such as Nigeria, Angola, Democratic Republic of the Congo, South Sudan and Ethiopia (Nigeria Centre For Disease Control (NCDC), 2018, Barrett, 2016, World Health Organization, 2019a). While many of these outbreaks have been met with response campaigns, the lack of sustained and large-scale preventive campaigns in these areas means that transmission is likely to continue, and these areas remain a potential “hotspot” for future outbreaks. Given issues with the global supply of the

YF vaccine, these areas are likely to remain under vaccinated for some time (Barrett, 2016).

In South America, areas considered as endemic for YF have historically had relatively high population-level coverage. However, while coverage in these areas has maintained – the areas considered as "endemic" have changed substantially, with the re-emergence of urban YF and epizootic and human spillover reported in areas far outside what was previously considered the endemic zone (Johansson et al., 2012, Almeida et al., 2014, Holzmann et al., 2010, Van der Stuyft et al., 1999). This has been particularly evident in Brazil. The majority of Brazil's major population centres are located in the South-East Atlantic states, areas not considered at risk for YF transmission, and have not previously required YF vaccination. However, while this may have been previously permissible – this is no longer the case. In 2016 YF reached the South-East Atlantic coast and, in this previously unexposed and unimmunized population, caused the largest YF outbreak recorded in Brazil since the 1940's (Cunha et al., 2019). While the unimmunized human population is not thought to have contributed to transmission, the level of spillover (~2000 human cases in 2016-2018) led to the classification of the whole country as "vaccine recommended" (Abreu et al., 2019, Cunha et al., 2019). This sudden expansion of the endemic zone was met with widespread responsive vaccination campaigns, but much of the country remains unvaccinated and at risk (Brazilian Ministry of Health, 2018).

These vaccination coverage estimates highlight the need for continued, and expanded, vaccination coverage across both Africa and South America. Despite a longstanding commitment to improving coverage across Africa, it remains insufficient across much of the endemic zone. Even within countries that have a sufficient national coverage, substantial heterogeneity in administrative units and across ages reveals

sub-populations at risk of transmission. In South America, similar issues exist – though these primarily exist as a consequence of an expanding endemic zone which has seen the re-emergence of the urban cycle of transmission and the inclusion of previously unafflicted large population centres.

Our vaccination coverage estimates rely on several assumptions. In the absence of reliable data available in all countries considered, within-country age structures where assumed to be homogeneous and population movements, within and between countries, were neglected. Movement, particularly from rural to urban areas, likely has a substantial impact on "diluting" population-level vaccination coverage especially in countries experiencing high rates of urbanisation, as many of those within the YF endemic zone are (Dos Santos et al., 2017). This may result in urban/rural discrepancies between the calculated and actual coverage. Coverage estimates are highly sensitive to uncertainties in population sizes, though quantification of these uncertainties is not currently available. I additionally cannot exclude that some vaccination activities were omitted, or doses wasted in implementation, which would translate to under-estimated coverages. Additionally, with the EYE strategy aiming to eliminate epidemic yellow fever by 2030, and increased funding in the area following recent outbreaks our projections are very likely to substantially underestimate future population-level coverage, as they only consider routine-infant immunization. However, these successes require a maintained commitment to vaccination - or YF could well re-emerge again as it did in the 80's and 90's across West Africa (Barrett and Higgs, 2007).

Shearer et al. recently published similar estimates of YF vaccination coverage in Africa and Latin America (Shearer et al., 2017). However, their vaccination coverage is considerably lower than our results in several West and Central African countries. It

appears that estimates by Shearer et al (2017) do not account for several recent large-scale campaigns in West and Central Africa, mostly performed as part of the YFI since 2006. Overall, these activities represent a total of nearly 31 million doses administered between 2006 and 2016 (See Appendix B). For example, accounting for these neglected campaigns led us to estimate a country level coverage of 88.5% in Benin in 2016, substantially higher than the 40% upper estimate provided by Shearer et al., (2017).

Previous research has found that within infectious disease epidemiology there is a strong interest in dynamic, interactive graphics which allow the user to understand the information, and its place in a wider context (Carroll et al., 2014). POLICI achieves this through customisable and interactive tables, maps and graphs which allow us to convey this information simply but effectively. Online hosting also allows users to connect remotely and without specialist software, and by facilitating the downloading of plots and tables it allows the information contained within this application to be utilised off-line.

POLICI has already garnered considerable interest from both scientists and decision-makers. It has been used as part of a recent study modelling the 2016 Angola outbreak (Zhao et al., 2018) and been utilised in 2016 by the World Health Organization to inform the reframed global strategy to Eliminate Yellow fever Epidemics (EYE)(World Health Organization, 2016a). Furthermore, the online, shiny-based format, allows for regular maintenance, including updates to the vaccination coverages and incorporation of new features. Development is currently underway to expand the tool to diversifying the tool to include the 10 diseases covered by the Vaccine Impact Modelling Consortium (VIMC)(Vaccine Impact Modelling Consortium, 2018).

Chapter 6 - Predicting primate presence and density in Brazil's South-East Atlantic rainforest

6.1 Summary

The Atlantic Forest represents one of the world's biodiversity hotspots and is home to 26 species of NHP. Despite the long-recognised biodiversity and importance of the habitat for conservation of numerous NHP species, quantitative information is limited. This has significant ramifications for both conservation, and modelling systems which take these into account.

Here I use random forest (RF) models fit to the presence and density of 6 NHP genera found in the Atlantic Rainforest of Brazil using a variety of environmental and climatic covariates in order to predict their distribution and density at a 1/120 degree resolution. Models reveal that there is much heterogeneity in presence and density across the range, and significant overlap of areas of heightened NHP genera biodiversity.

This information has both important conservation aspects and serves as the basis for the "realistic" population and density distributions of NHP in our meta-population modelling in the next subsequent chapter.

6.2 Introduction

The Atlantic Forest was once one of the largest rainforests in South America. Stretching from Paraguay and Argentina up to the far North East of Brazil. Extending over 3000 km, and composing more than 150 million hectares of forest (Ribeiro et al., 2009) it represents a biodiversity hotspot, which may include up to 8% of the world's total species (Myers et al., 2000). However, this area has been reduced to 11.4 – 16.0% of its original range, with the remaining areas heavily fragmented (Ribeiro et al., 2009). Within the Atlantic forest, 26 native primate species are found – of which 19 are endemic. These are often severely threatened throughout much of this range by habitat loss, fragmentation, hunting and the illegal pet trade, with around half of the primate species endangered (IUCN, 2017, Culot et al., 2019).

Despite the long-recognised biodiversity and importance of the habitat for conservation of numerous NHP species, quantitative information is limited. The little information there is has historically been collected without standardised protocols or recording of data. Furthermore, complexities which are intrinsic to the diverse topography and flora of the region often limit the geographic locations to which sampling can be conducted at (Ribeiro et al., 2009, Gardner et al., 2008). This has resulted in large areas of "unknowns" where studies have not been carried out. These gaps in information on presence and size of populations represent substantial barriers to conservation efforts. Without accurate identification of NHP populations and population densities we cannot quantify the effects of anthropogenic change or infectious diseases, such as YF, on NHP population dynamics. These key data gaps hinder conservation efforts to preserve NHP populations (Rylands et al., 2008).

In the absence of information, ecological niche modelling (ENM) offers an approach to help further our understanding of the distribution and density of primates within this region, with a proven history of successful application to mapping the range of primates in the Americas (Vidal-Garcia and Serio-Silva, 2011, Pinto et al., 2009). ENM works on the basis that factors, such as average temperature or landcover, are influential on the distribution or occurrence of a species. These factors, or predictors, allow us to predict occurrence in areas where data is sparse or absent. This has a long history of use in ecology for producing predictions of the potential distribution of species and can offer an important first step in conservation planning (Calixto-Perez et al., 2018, Borrell et al., 2019). In order to provide accurate, and useful, predictions, high quality data on presence/absence and density of a species is needed. In this study, this is given by the ATLANTIC-PRIMATES dataset, which represents the largest and most complete dataset of reports of NHPs in the Atlantic Forest (Culot et al., 2019).

I have collated an extensive dataset of covariates related to environmental, landcover, climatic and human populations to be utilised in conjunction with the ATLANTIC-PRIMATES dataset in order to predict the presence/absence and density of all 6 genera of NHP's found in the Atlantic Forest within the South East Atlantic states of Brazil. This is carried out using random forest (RF) models in order to predict the geographic extent of genera, and the predicted density of genera, at a 1/120 degree resolution. These predictions describe the geographic extent of genera, highlighting areas of genera overlap and primate density "hotspots", and contextualise these predictions in light of current conservation efforts.

6.3 Methods

6.3.1 NHP observational and density data

NHP observation and density data was provided through the ATLANTIC-PRIMATES dataset which represents the largest complete review and quantification of communities and the occurrence of primates in the Atlantic Forests of South America (Culot et al., 2019). This dataset originally had 715 georeferenced studies of primate species with quantitative information on abundance and density. This was refined to only locations which had estimates of primate density, leaving 535 points. Studies where the recorded date of survey was before 1992 were not included in further study, reducing the overall number of records to 475. These were omitted in order to discount known historical records which may not represent current NHP densities. Of these studies, 317 did not include a date of study. While many of these studies may have taken place before 1992, an omission criterion, the relative lack of data required their presence, with the caveat that predictions may better represent historical distributions and densities rather than current.

Here I have mapped the presence and the density of genera rather than species for NHPs found in the region. This decision was undertaken following exploration of the data which, despite representing the largest and most complete dataset of NHP communities in the Atlantic Rainforest, was relatively sparse. Density was calculated at the genus level, with individual species within the genus aggregated, given the scarcity of data on numerous individual species. Data availability for the 6 different NHP genus found in the Atlantic Forest region varied substantially (Table 6.3.1). Data selection steps are shown in (Figure 6.3.1).

Following this, raster maps of density data were created for the 6 genera of NHP's found in the Atlantic Forest region (Figure 6.3.3). Where the area of study was provided, this was converted directly to density points on a raster of 1/120 degree resolution. Where this was not provided the area assumed was the median area of study which was 2.3km^2. Overlapping measures of NHP density were reconciled by taking the median value of the genera.

Table 6.3.1. Number of data points by species available in the refined quantitative database on Atlantic Primates

GENUS	NUMBER OF STUDIES	DATA POINTS
ALOUATTA	92	292
BRACHYTELES	35	550
CALLICEBUS	34	195
CALLITHRIX	56	302
LEONTOPITHECUS	217	1743
SAPAJUS	41	355
TOTAL	475	3437

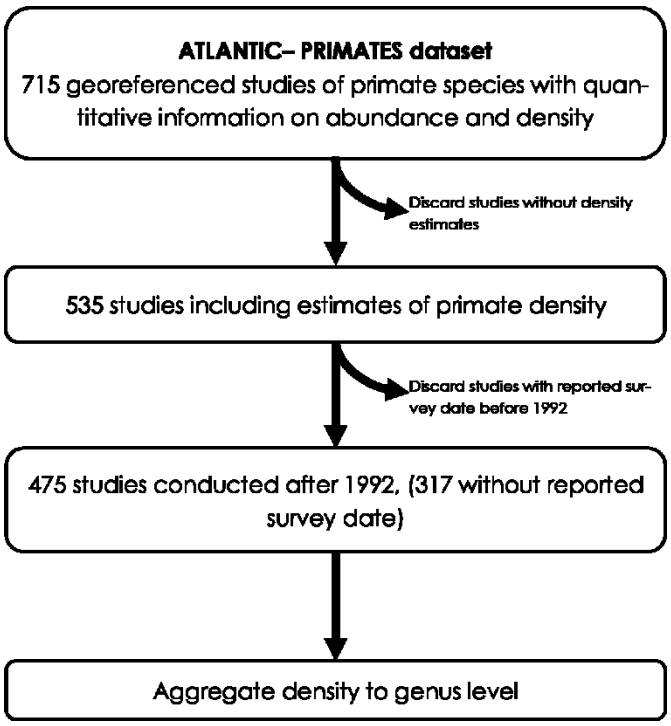


Figure 6.3.1. Flowchart of data selection and aggregation.

Figure 6.3.2 Map of States found in Brazils South East Atlantic Coast

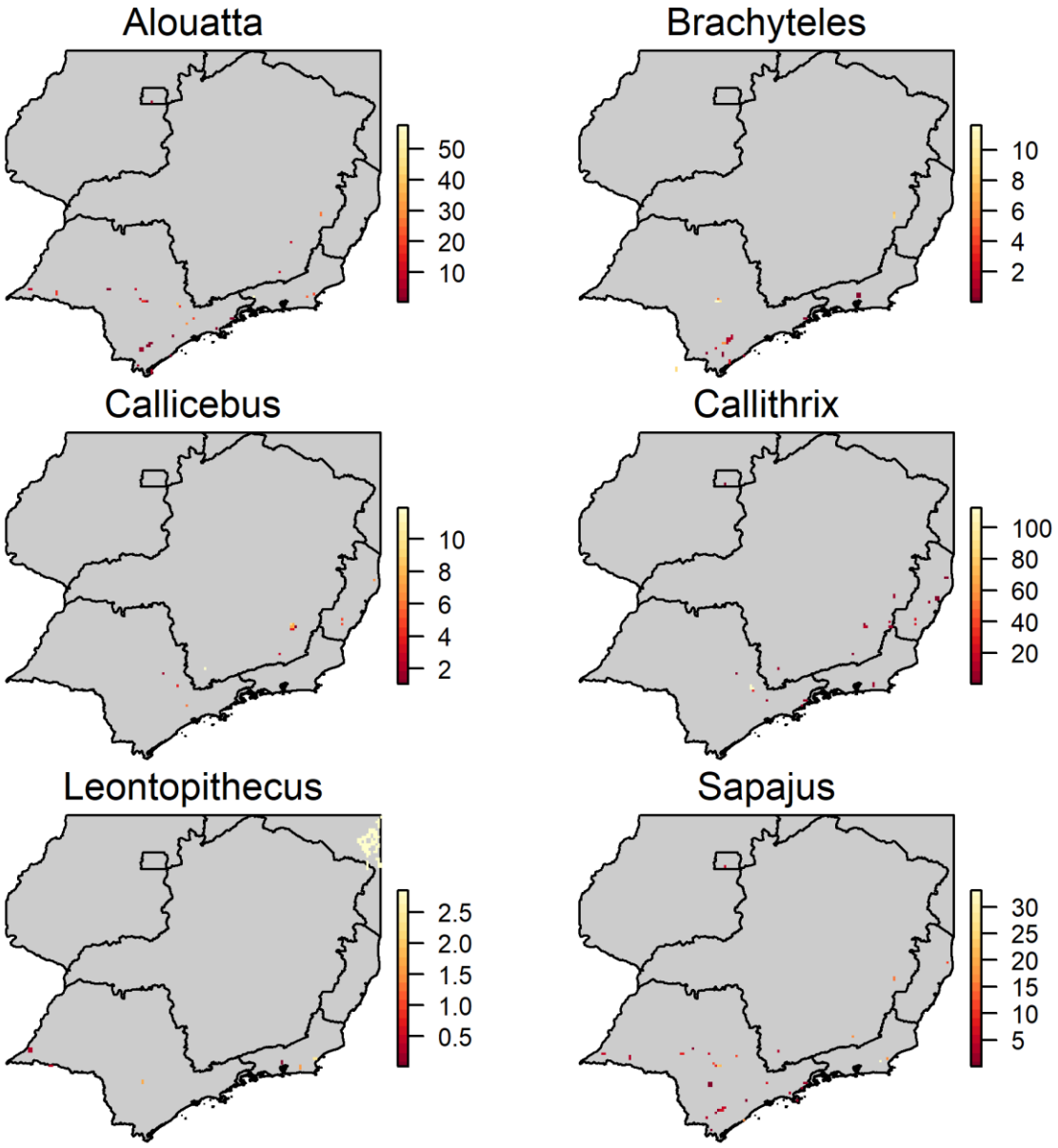


Figure 6.3.3. The recorded NHP density of each genus from the ATLANTIC-PRIMATES dataset NHP data for each genus. Pixels show the density recorded for each genus across the states considered.. To aid visualisation, pixels for the data have been enlarged from the original resolution of 1/120 to 1/12.

6.3.2 NHP genus geographic extent

Information on the total extent of NHP species distribution was obtained from distribution maps provided by the International Union for Conservation of Nature (IUCN) (IUCN, 2017) through shapefiles for each species. All species that composed

a genus were aggregated to produce extents of genus presence/absence across the Brazilian Atlantic Forest region.

These were explicitly included as a potential covariate in density mapping, but not for presence/absence. Instead, models were predicted in the areas within these provided ranges. This was done as the aim was to predict the distribution within known ranges, rather than recreate the known range as would occur if included as a potential covariate as all data points were found within the existing range.

Sensitivity analysis of this covariate was explored in Appendix C.

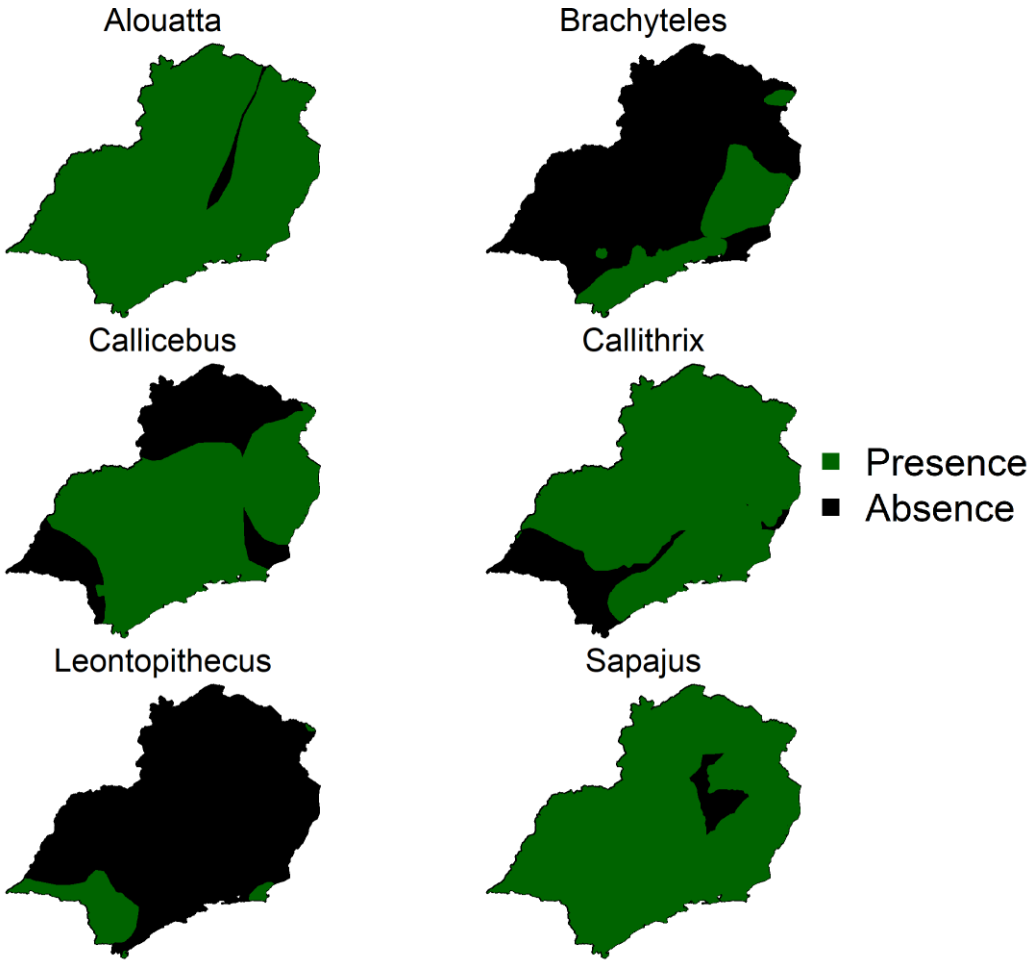


Figure 6.3.4 IUCN shapefiles of genera distribution in states of the South East of Brazil.

6.3.3 Areas of conservation

Areas of conservation in Brazil were provided by the United Nations Environment Programme World Conservation Monitoring Office (UNEP-WCMC) (UNEP-WCMC., 2019). Areas were classified as a binary covariate, with inside and outside of conservation zones. These were included in order to account for the role of, relative, protection from anthropogenic pressures that areas of conservation provide.

6.3.4 Land cover

Landcover (LC) data was provided by the European Space Agencies Climate Change Initiative which provides global LC maps at a 1/360 degrees spatial resolution for the years 1992-2015 (Climate Change Initiative, 2017). Each pixel defines the dominant LC type for each individual year during the time period out of a total of 37 individual types. The modal value of each pixel was taken to generate the mode LC type during this period. LC cover was then aggregated to 1/120 degree resolution by calculating the proportion of the area taken up by each LC type.

Land cover is an important determinant of NHP abundance, with primates of the Atlantic Rainforest requiring forested areas for habitation, and the showing differing effects of other types of land cover on populations (Kay et al., 1997, Pinto et al., 2009).

6.3.5 Bioclimatic variables

Bioclimatic variables (1-19) were provided through WorldClim Version 2 at a 1/120 degree resolution (Fick and Hijmans, 2017). While bioclimatic variables such as temperature and rainfall are not as deterministic for vertebrate endotherms as mosquito vectors, there are still known associations and their inclusion in species distribution modelling is widely utilised (Pinto et al., 2009, Vidal-Garcia and Serio-Silva, 2011).

6.3.6 Human population

LandScan 2015 (Bright et al., 2016) population estimates with a 1/120 degree resolution. Human populations have been shown to have differing effects on NHP populations (Michalski and Peres, 2005, Thorn et al., 2009, Pinto et al., 2009).

6.3.7 Random forest modelling of presence and density

Random Forests (RF) are a machine learning ensemble regression method (Biau, 2012) which takes into account non-linear relationships and interactions between covariates to produce outcomes. Due to this, they can have substantial improvements in accuracy over traditional regressions (Biau, 2012).

The total dataset for each genus, which I split into training and validation sets, is created by taking every pixel where I have data on that NHP's genera, and combining this with a dataset randomly selected of the same number of pixels outside of the NHP genus range, as defined by the ICUNs shapefiles of species distribution (IUCN, 2017), to produce a dataset that is of half data and half "pseudo-absence" points. The inclusion of pseudo-absence, which is used as controls to where I am somewhat confident the species is not found, has been shown to improve the fit of machine learning methods (Barbet-Massin et al., 2012).

Model fitting was carried out using an out-of-sample stepwise forward variable selection process. In this, I start with a single explanatory covariate, which is assessed for its out-of-sample predictive ability by dividing the dataset the training and the validation set. Due to the relative sparsity of data, the decision was made to divide the training and validation set into two sets of relatively equal size.

The model is trained on the training set and predicted on the validation set, and the SSE value calculated. This is then repeated 100 times, each to different training and validation sets, and the average SSE calculated. This is then repeated for all covariates. Once this has been done for all covariates, the covariate which produced the best out-of-sample fit (assessed by SSE) is saved and included by default in the next round of covariate selection. This continues until all covariates have been ranked by their out-of-sample SSE ability and the final formula includes all covariates possible. Following this, the covariate combination that produced the highest out of sample SSE ability is selected as the final formula for prediction. This was carried out for each genus to produce genus specific models for presence and density.

While RF models are generally thought to be relatively insensitive to over utilisation of covariates (Biau, 2012), refining the input covariates allows for a greater understanding of covariate relationships and increases the generalisability of onward predictions. Furthermore, by selecting covariates based on their out-of-sample predictability, I am increasing the predictability of our model, which is imperative given the aim of predicting presence and density across these states. The rational of using this approach was based on unpublished research on predicting the global intensity of dengue transmission, by Lorenzo Cattarino in the Department of Infectious Disease Epidemiology at Imperial College London, which found it suitable for high resolution predictions (Cattarino et al., 2020).

Random forest modelling is carried out using the Ranger package (Wright and Ziegler, 2017), which also provides the R^2 and pseudo-R^2 estimates for the presence/absence models, in the statistical programming language R (Team, 2015a), version 3.5.0.

6.3.8 Combined estimates of NHP presence/absence and density

In order to produce estimates of total NHP presence/absence across all genera, a cut off of a probability of presence of below 0.5 (50%) was assumed for the presence/absence of a species. These probabilities were summed across all species in order to produce estimates of overlapping ranges and genera diversity.

6.4 Results

6.4.1 Covariate selection

Covariate selections were similar between density and presence/absence models.

Generally, climate variables were highly important for the presence/absence of genera, occupying the majority of selected covariates. Of these, the seasonality of rainfall and temperature, along with ranges of these were found in most models. Additionally, numerous indicators of tree coverage were selected.

For the density models, genera specific range was selected as important in all models. Conservation areas were only found important for Brachyteles presence and density and Callicebus presence.

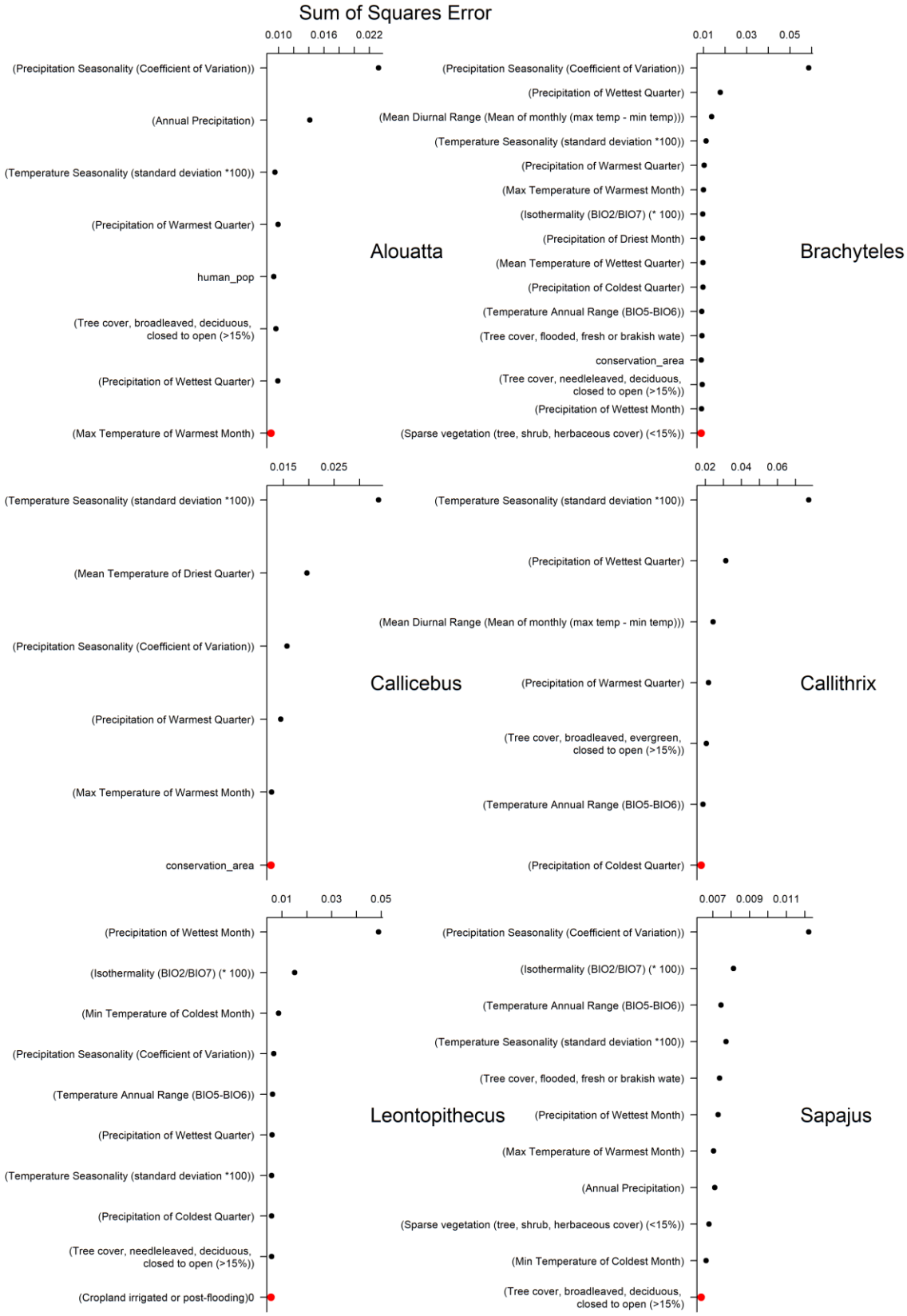


Figure 6.4.1. Covariates selected through out-of-sample forward stepwise selection for each genera model for presence. The red point indicates the covariate addition that resulted in the lowest sum of squares error.

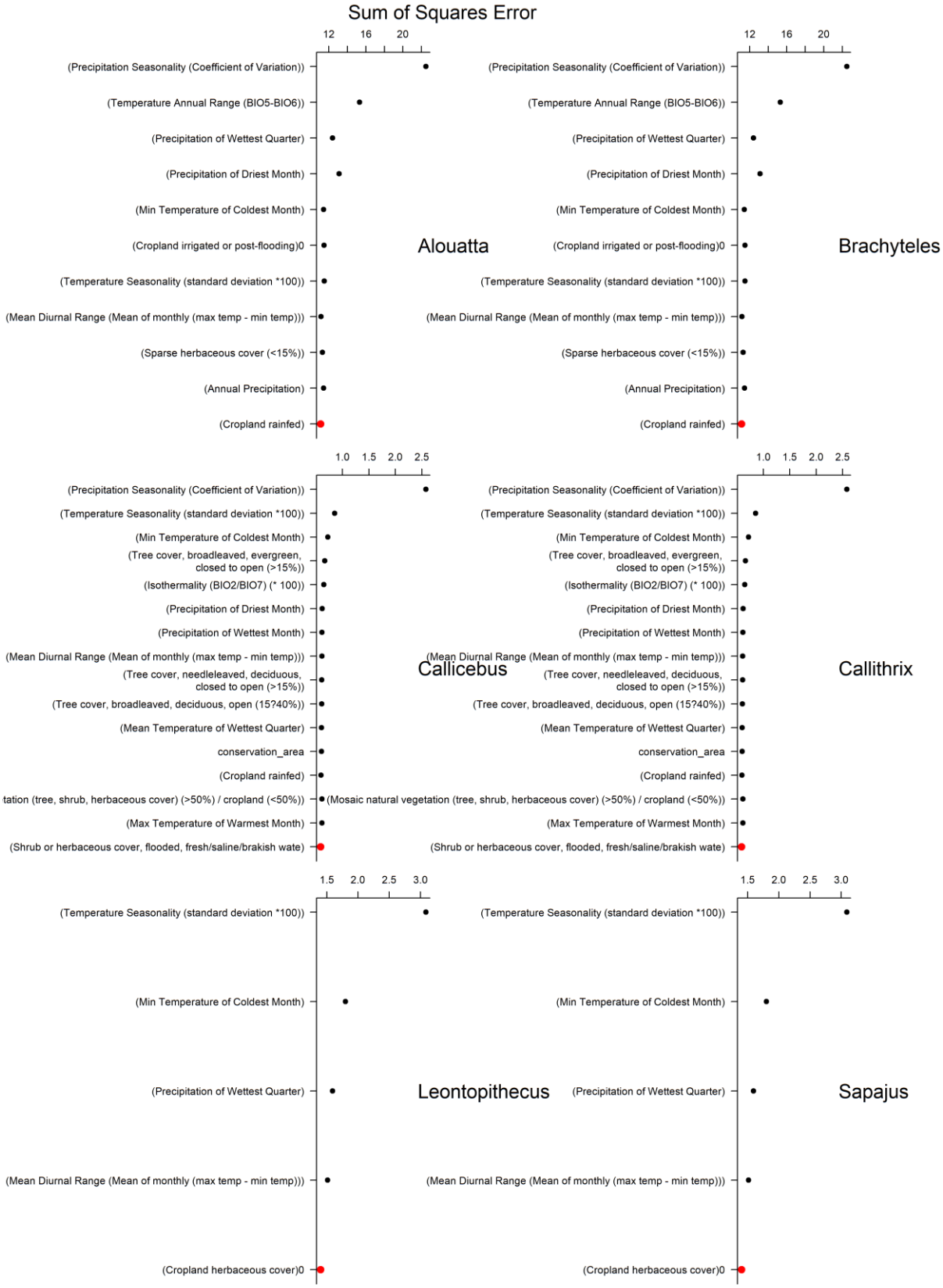


Figure 6.4.2. Covariates selected through out-of-sample forward stepwise selection for each genera model for density. The red point indicates the covariate addition that resulted in the lowest sum of squares error.

6.4.2 NHP predicted distributions

All model predictions of presence had a high degree of agreement with the data (

Table *6.4.1*).

The probability of Alouatta presence/absence is substantial throughout much of the four states, apart from the central and northern regions of Minas Gerais, where probabilities are close to 0 (Figure 6.4.3). Brachyteles distribution follows a similar pattern to Alouatta. Callicebus are limited to Southern and Western Sao Paulo, areas north of Rio de Janeiro city and the border of Espirito Santo and Minas Gerais. Callithrix shows similar distributions as Callicebus but shows further areas of heightened present in central and Western Sao Paulo. Leontopithecus and Sapajus have similar distributions, found mainly in Espiritio Santo, except the far north, much of Rio de Janeiro, Southern Minas Gerais and Eastern Sao Paulo, with Sapajus additionally found further East in Sao Paulo..

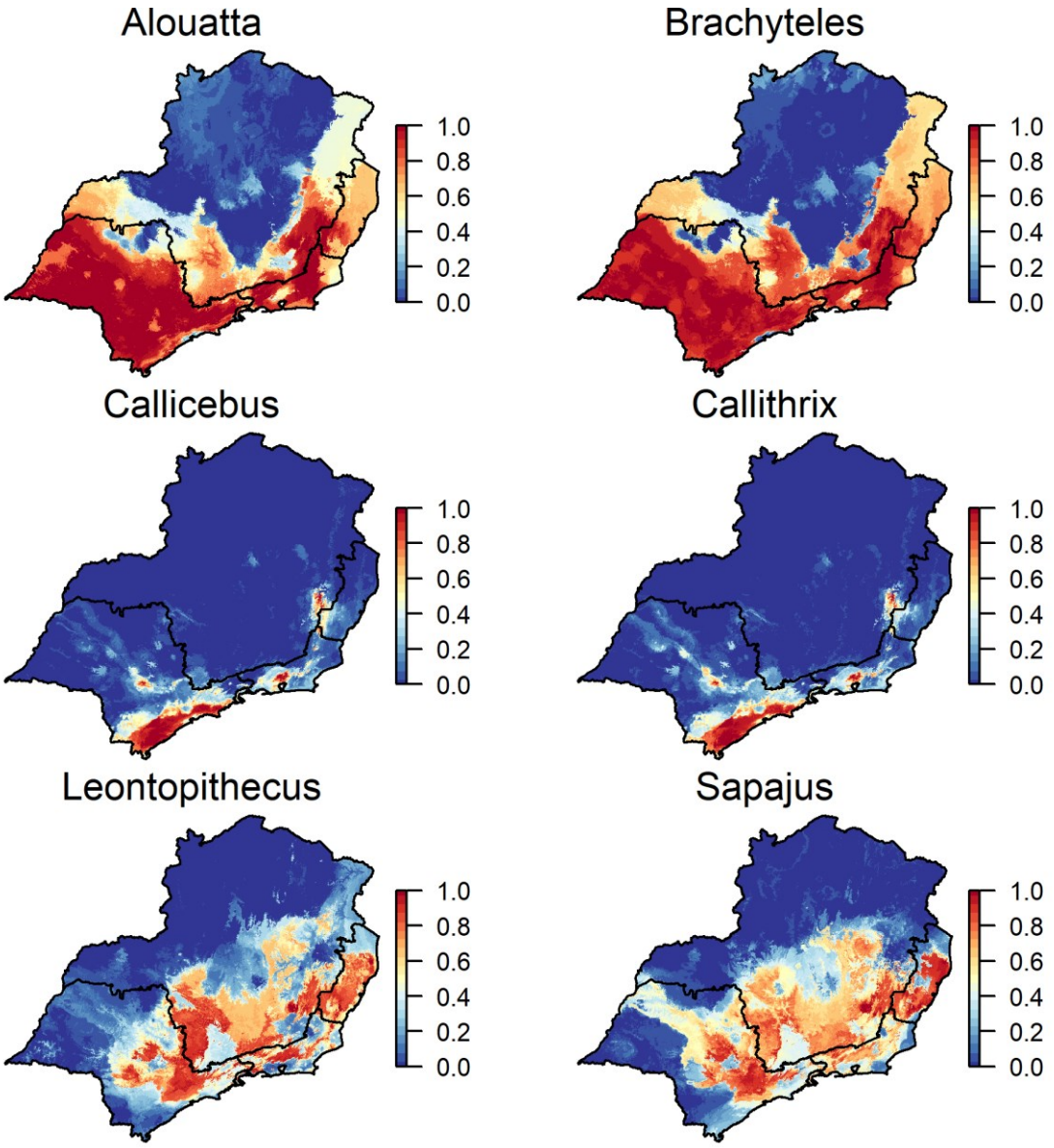


Figure 6.4.3. Random forest predictions of models created through the out-of-sample forward stepwise covariate selection method for the probability of presence for the six NHP genera found in the South-East Atlantic states of Brazil.

6.4.3 Predictions of population density

Expectedly, model fits of population density were generally not as accurate as predictions of presence/absence, though models still explained a large proportion in the variability of geographic distribution of species densities, except for Brachyteles and Sapajus which had poorer fits (Table *6.4.2*).

Alouatta has the highest predicted densities of any genera, up to a maximum of approximately 25 individuals/km^2. The highest densities are found in Sao Paulo, Rio de Janeiro and the borders of Espirito Santo and Minas Gerais. Brachyteles populations are low density, at a maximum of 4 individuals/km^2) concentrated in thin strips along Sao Paulo, Minas Gerais and Rio de Janeiro, with pockets in Southern Sao Paulo Callicebus was found throughout much of the region, though with relatively low densities of up to 5 individuals/km^2, with notable absences in Northern Minas Gerais and Western Sao Paulo. Callithrix populations were highly localised, with low levels found throughout much of the 4 states, but areas of particularly high density located in the border of Minas Gerais, Espirito Santo, parts of Rio de Janeiro and Sao Paulo.

Leontopithecus had the lowest densities predicted of all genera (maximum of 2.5 individuals/km^2), with the highest populations found in Southern Rio de Janeiro and a small fragment in North Western Minas. Heightened levels of Sapajus density were found primarily in Eastern Minas Gerais, Espirito Santo and Rio de Janeiro, and much of Sao Paulo. The highest densities of Sapajus are found in the Far east of Minas Gerais, and coastal regions of Rio de Janeiro.

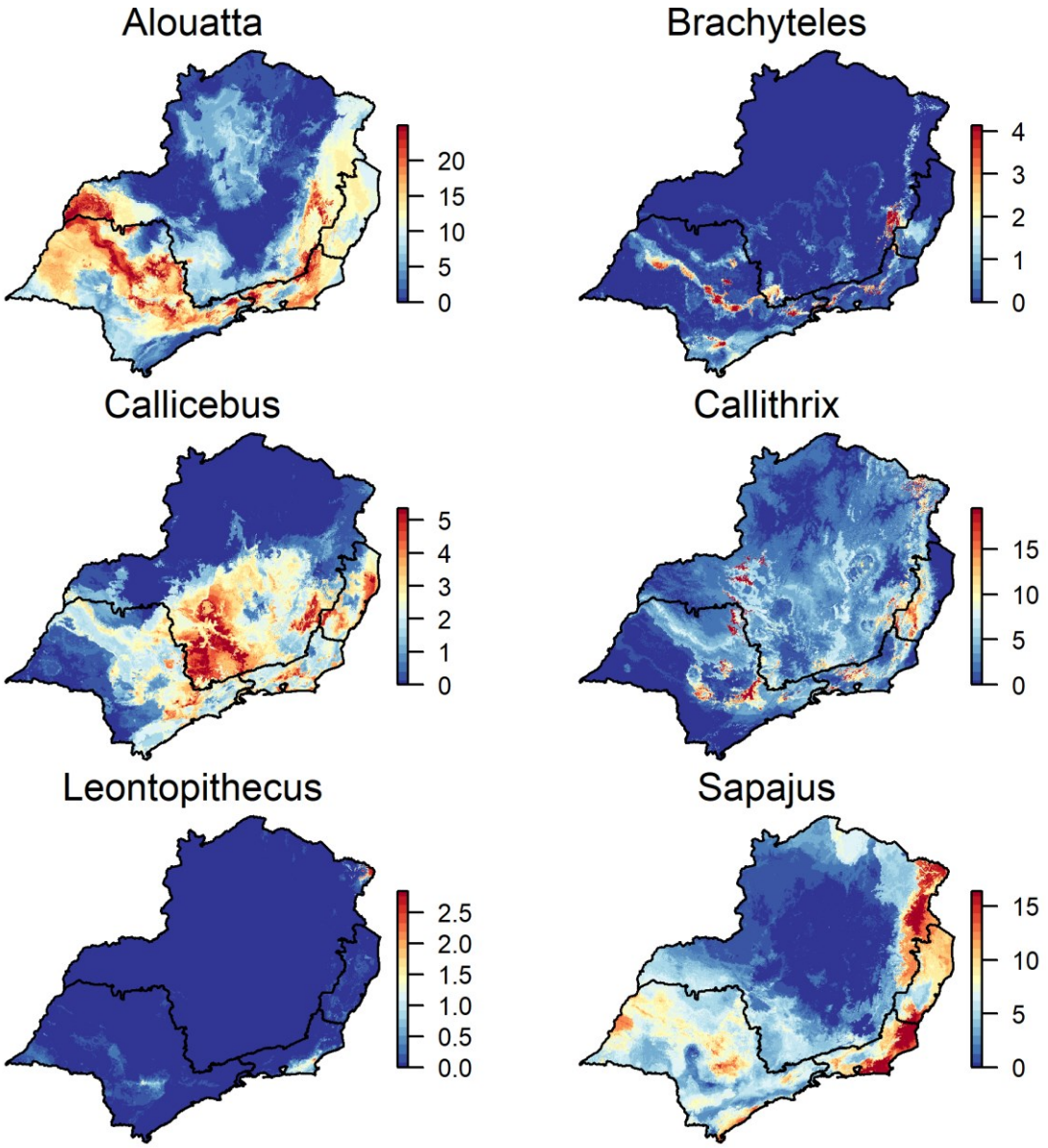


Figure 6.4.4. Random forest predictions of models created through the out-of-sample forward stepwise covariate selection method for the population density (individuals/km^2) for the six NHP genera found in the South-East Atlantic states of Brazil.

Table 6.4.1. In-sample area true and false positive/negative values for random forest models predicting the probability of presence for the six NHP genus found in South-East Atlantic states of Brazil.

GENUS	TRUE POSITIVE	FALSE POSITIVE	TRUE NEGATIVE	FALSE NEGATIVE
ALOUATTA	146	146	17	275
BRACHYTELES	90	918	0	92
CALLICEBUS	182	13	191	4
CALLITHRIX	297	5	232	70
LEONTOPITHECUS	1471	248	1719	48
SAPAJUS	393	225	48	44

Table 6.4.2. In-sample R^2 values comparing the predicted densities for areas with NHP genera data and the pseudo-absence locations with none.

GENUS	IN-SAMPLE R^2
ALOUATTA	0.681
BRACHYTELES	0.119
CALLICEBUS	0.748
CALLITHRIX	0.910
LEONTOPITHECUS	0.806
SAPAJUS	0.458

Combined estimates of NHP presence and density

By summing the number of predicted species and densities of each NHP genera, I can show there is substantial heterogeneity in the presence and number of NHP species, with notable hotspots, and areas of reduced genera diversity across the region (Figure 6.4.5).

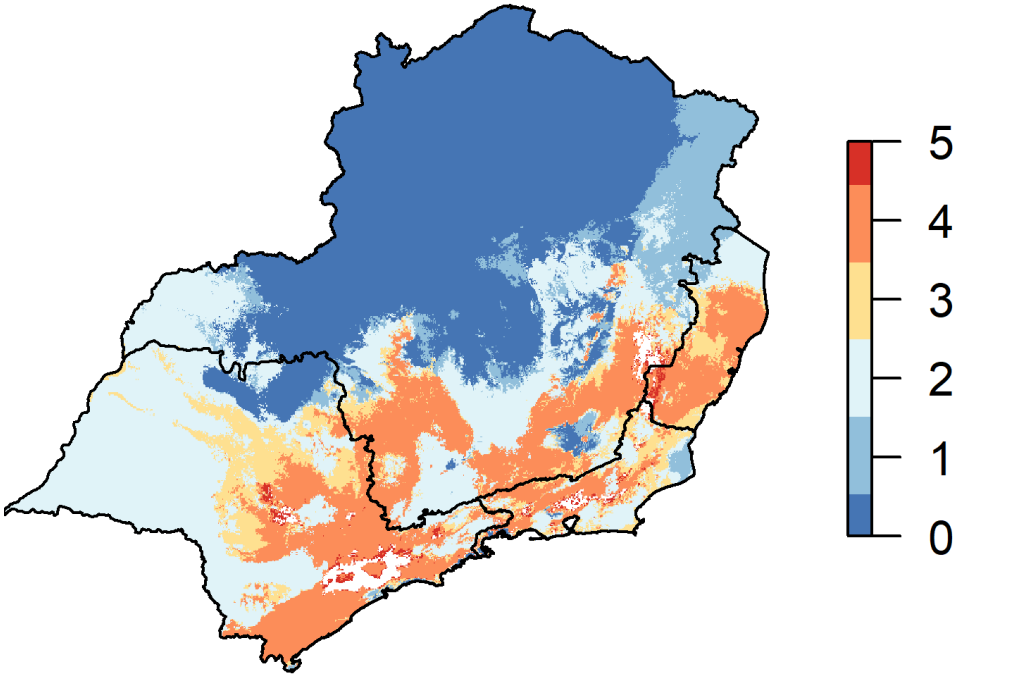
Areas with the most recorded genera are found in Southern Sao Paulo, Rio de Janeiro and Espirito Santo, as well as limited regions in Southern Minas Gerais (Figure 6.4.5A). Much of Northern and Central Minas Gerais have low levels of NHP genera (0-1), and Western Sao Paulo (2-3). In contrast, almost all regions of Espirito Santo, and the majority of Rio de Janeiro have high genera diversity (3-5).

While high genera presence is often correlated with areas of high NHP density, this is not always the case (Figure 6.4.4B). The highest levels of NHP density are found along the Eastern border of Minas Gerais and Espirito Santo, the border of Espirito Santo and Rio de Janeiro, much of Rio de Janeiro and parts of Eastern Sao Paulo, with the majority of these areas seeing densities of 50-70 individuals/km^2, and very limited areas up to 150 individuals/km^2.

Much of Minas Gerais and Western Sao Paulo are predicted to have low densities of NHP's, rarely exceeding ~25 individuals/km^2, apart from in the aforementioned regions.

While areas of conservation are associated with higher percentages of overall overlapping genera habitat, with 77.7% of the areas all 6 genera are found located within areas of conservation, the vast majority of areas of high genera diversity are found outside of these, as well as there was no difference between the median NHP individual density (Figure 6.4.6 and Table 6.4.3).

A



B

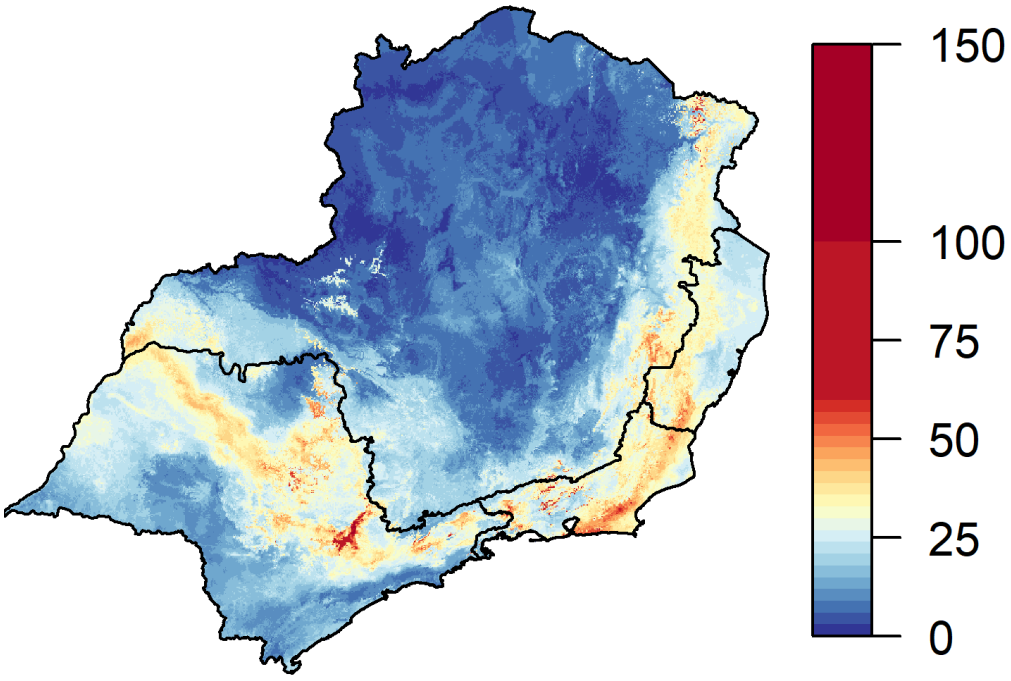
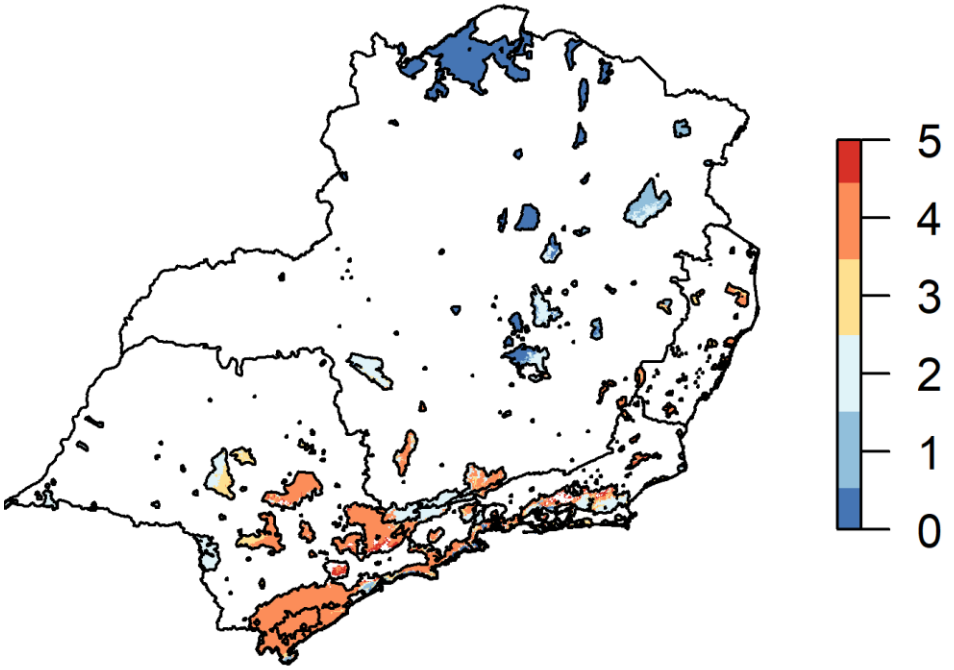


Figure 6.4.5. A) The predicted number of primate species and the B) summed predicted density (individuals km^2) of all NHP genera.

A



B

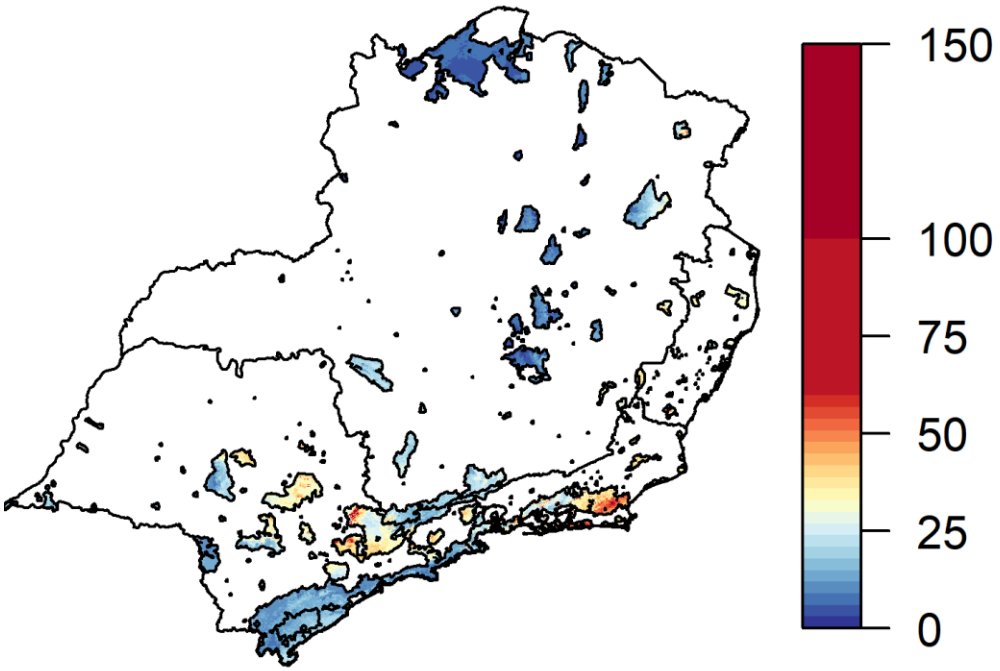


Figure 6.4.6. A) The predicted number of primate species in areas of conservation, and the B) summed predicted density (individuals km^2) of all NHP genera in areas of conservation

Table 6.4.3. Comparison of the distribution of the number of genera and density between areas of conservation and non-conservation.

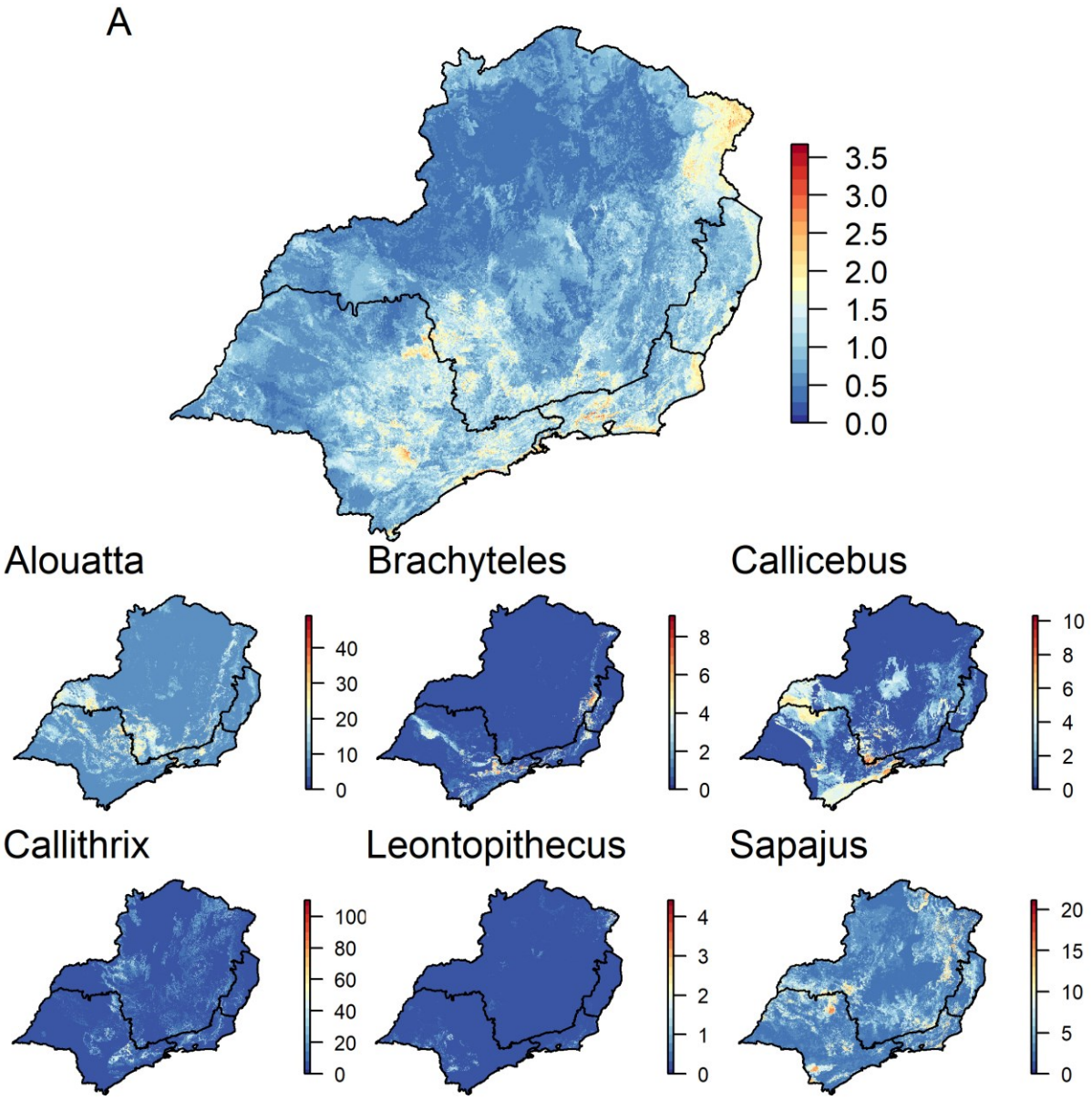
	% OF AREAS BY NUMBER OF GENERA					MEDIAN GENERA	MEDIAN NHP DENSITY	MAX NHP DENSITY
	<=2	3	4	5				
CONSERVATION AREAS	5.4	8.6	19.3	27.6	22.3	2	13.1	0
NON-CONSERVATION AREAS	94.6	91.4	80.7	72.4	77.7	1	10.8	0

6.4.4 Uncertainty in predictions

Given the relative sparsity of the data, and the complexity of the models involved, there is substantial scope for uncertainty in predictions. Here we have subset predictions to locations with data to show the differences in the upper and lower 95% confidence in presence and density predictions Figure 6.4.7.

Uncertainty in the presence of genera is highly localised, with specific areas showing the majority of variance – particularly in Brachyteles and Leontopithecus. Alouatta shows wider patterns of variance, with most of the areas it is predicted in showing a 0.2 difference in the probability of presence and absence.

This pattern is further seen in density, with the majority of differences between bounds minimal, except in specific areas where the numbers are substantial.


A
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Alouatta
40
30
20
10
0
Brachyteles
8
6
4
2
0
Callicebus
10
8
6
4
2
0
Callithrix
100
80
60
40
20
0
Leontopithecus
4
3
2
1
0
Sapajus
20
15
10
5
0

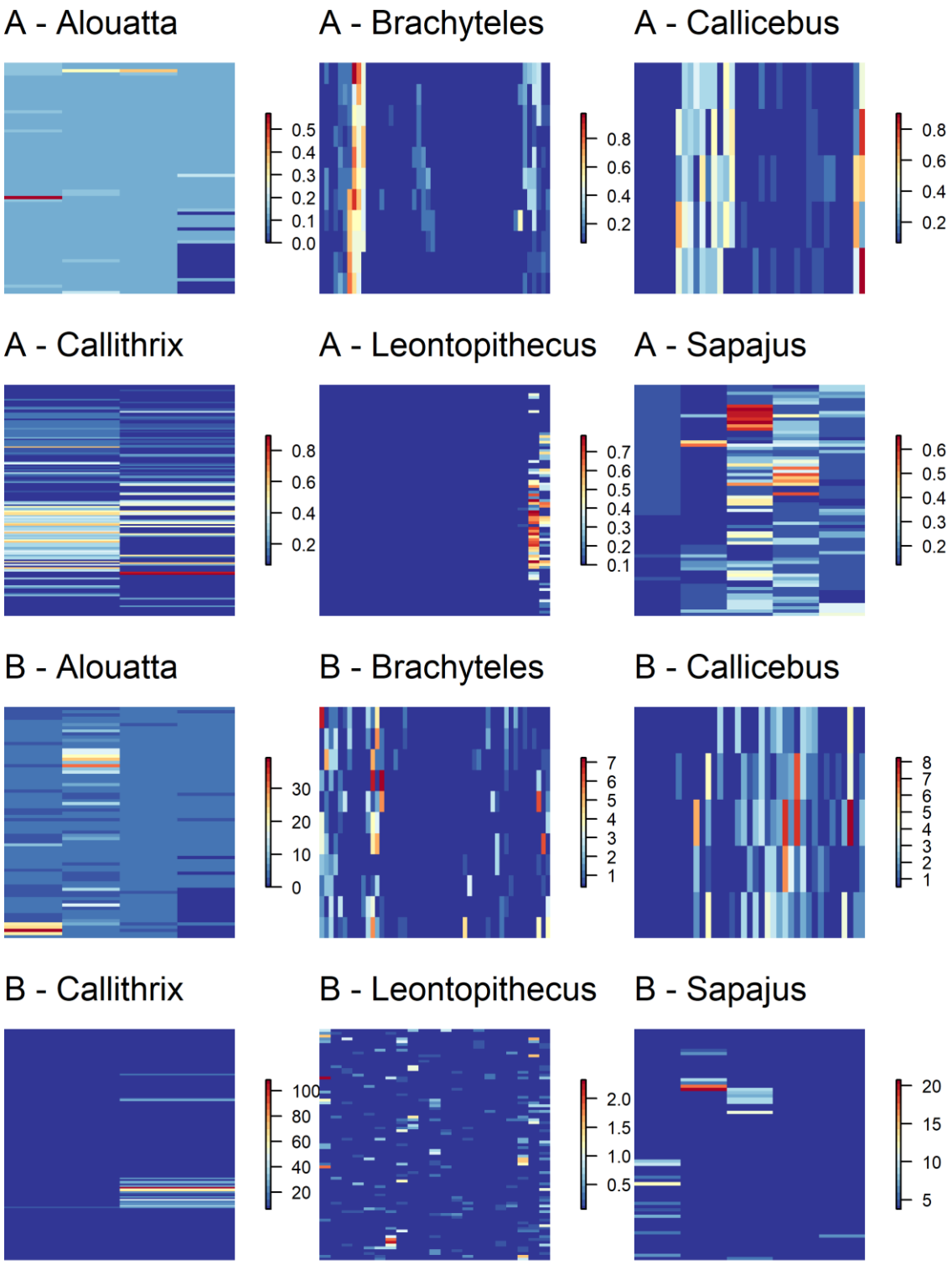


Figure 6.4.7. The difference in the upper and lower 95% confidence interval predictions for locations with data on (A) the probability of presence for each species predicted and (B) NHP density for each species. Each cell represents a different location where data on the species was available and used to fit the model.

6.5 Discussion

Here, I have utilised machine learning methods with the largest and most complete dataset of NHP primates in the Atlantic Forest region in order to produce predictions of the presence/absence and the density of NHP's across four states in South-Eastern Brazil to a relatively high degree of accuracy (

Table *6.4.1*, Figure 6.4.3 and Figure 6.4.4). These individual predictions have been combined in order to generate overall predictions of genera richness, and individual NHP density (Figure 6.4.5). While there have been previous models predicting the ecological niche and density modelling of NHP's in the Atlantic Forest, they have not previously been conducted using as stringent methodology, or as substantial a baseline dataset (Pinto et al., 2009).

While the Atlantic Forest has a substantial history of primate conservation research, the intrinsically complex nature of estimating populations and distributions over the region means that many aspects of NHP ecology within the region are still poorly understood (Culot et al., 2019). The utilisation of machine-learning methods for the identification of notable hotspots of NHP genera biodiversity, and regions of high NHP density – and the fact these do not always overlap has important considerations for conservation, and this mapping could be used to guide future surveys to areas which have the highest probability of occurrence or density. Areas that support numerous NHP genera are not always the areas that support the highest density of NHPs – even when the overlapping species compositions are the same. This is potentially an unfortunate consequence of habitat degradation and fragmentation, and that while numerous NHP genera are still present in these areas, the suitability, or lack thereof,

the habitat calls into question the long-term subsistence of these areas as genera biodiversity “hotspots”. Particularly given that 77.7% of areas that are predicted to have the highest level of NHP genera overlap are outside of areas of current conservation (Table 6.4.3).

The role of habitat loss and fragmentation is of particular concern to many NHP’s of the Atlantic Forest region, with only 12% of the original forest remaining, and over 80% of forest patches remaining less than 50 ha (0.5km^2) (Ribeiro et al., 2009). Though numerous NHP genera are resistant to fragmentation, such as Alouatta, all are negatively affected by it (Estrada et al., 2018). Highly fragmented populations are at a higher risk of local extinction through disease and anthropogenic pressures (Michalski and Peres, 2005, Bicca-Marques and de Freitas, 2010, Mbora and McPeek, 2009). In particular, these fragmented populations may be of substantial risk to outbreaks of yellow fever (YF). YF causes substantial mortality in NHP’s, with large scale epizootic events leading to thousands of deaths (Almeida et al., 2014). While YF has been historically confined to the North-West and Amazonian regions of the country, over the last 20 years it has seen a substantial expansion of its endemic zone (Cunha et al., 2019), eventually reaching the South-East Atlantic coastal states of Rio de Janeiro and Espírito Santo in 2016. This has led to several thousand human and NHP cases of YF occurring 2016-2019 (Faria et al., 2018). The introduction of this highly lethal virus into a previously unexposed population is of serious cause for concern, particularly given the number of endangered and susceptible primates that live in these areas. It is entirely possible that the additional mortality inflicted by these YF epidemics could contribute to, or hasten, the localised extinction of numerous genera, as was found through a modelling study looking at the risk of extinction on Alouatta populations in Argentina (Holzmann et al., 2010).

As previously discussed, many of these genera are experiencing rapid habitat degradation and loss, with populations of certain species experiencing dramatic declines in only a few years. Therefore, the inclusion of surveys as far back as 1992, and where there are no recorded dates of study, may lead to over-estimation of the current extents and population densities. However, given the sparsity of recordings, even in this exhaustive quantification of the available literature, this is a permissible allowance, though this limitation should be carefully considered when extrapolating from these findings. An additional limitation is found in the highly complex relationship of environment, climate, and primate distribution (Kamilar, 2009). The role of abiotic and anthropogenic factors, such as hunting, on species distribution and density likely play a substantial, and potentially even greater role, than climate for NHP's in the current remaining fragments of the Atlantic Forest. Here I have attempted to account for this by considering human populations, areas conservation, and landcover as potential covariates to identify the realised niche, rather than the potential (Bicca-Marques and de Freitas, 2010). This accounting for the anthropogenic interaction with species distribution and density, either directly through human population or indirectly through landcover types associated with human populations, possibly allows for a more "realistic" depiction of distributions and densities. Furthermore, the use of RF models, which account for non-linear relationships between the predictor and explanatory covariates can allow for interactions that more closely approximate real-life relationships (Biau, 2012). Despite this, it is likely I have not fully captured these pressures – and this depiction of presence and density likely represents an "idealised" situation, given our population and landcover caveats, rather than the actual realities encountered in the environment. A fact highlighted in our depictions of the model predictions of uncertainty in NHP distribution and density (Figure *6.4.7*).

In summary, despite the modelling and data restrictions, this body of work represents an important contribution to continuing conservation efforts in the Atlantic Forest region of South East Brazil. By identifying overlapping genera ranges, and potential hotspots of NHP density, conservation efforts can be focused. This is particularly important given the decreasing habitat ranges, and increasing habitat fragmentation, and emerging threat of YF faced by all NHP species in the region. These quantitative estimates of distribution and density also offer their use in additional fields outside of traditional conservation. Given the role of NHP's in the maintenance and spillover of YFV into human populations, and the relatively new exposure of these areas to circulation, mathematical modelling of infectious disease dynamics could offer important insights into the possibility of long-term persistence of the virus, and the consequences this has for both human and NHP populations.

Chapter 7 - Mathematical modelling non-human primate (NHP) YFV transmission in South-East Brazil: Implications for establishment of endemicity and consequences for NHP populations.

7.1 Summary

While the importance of NHP's in both endemic and epidemic YFV transmission has long been recognised, relatively little quantitative research has been undertaken in the last 80 years to further develop this knowledge.

Here we construct and parameterise a stochastic meta-population model of 4 different NHP genera, considering age and NHP movement in order to simulate a more realistic depiction of the underlying YFV transmission ecology. By accounting for different genera that are found in the South-East of Brazil, the location of recent epidemic transmission in naive populations of NHPs, we estimate the critical community size, and the composition of NHP genera, required for long-term persistence of YFV.

This model is then applied to "realistic" NHP population distributions and makeups in the state, and bordering regions, of Rio de Janeiro in attempt to answer the question of whether the establishment of endemicity is likely in this previously YFV naïve region.

7.2 Introduction

Despite a broad qualitative understanding of the importance of NHPs to transmission and surveillance, detailed research and understanding into the role different NHP's species play, along with overall population sizes and density, play in the maintenance and proliferation of the YFV are lacking. These knowledge gaps are of significant cause for concern following large-scale, and unprecedented transmission in Brazil's, most densely populated, South-Eastern states. In addition to dense human habitation, these states are an NHP biodiversity hotspot, with 26 native primate species found throughout the severely fragmented remnants of what was once one of South Americas largest rainforests, the Atlantic Forest (Ribeiro et al., 2009). These large and diverse populations of NHP's, of which several endemic and endangered species are severely threatened, pose the worry that YF may become endemic in the South East Atlantic Coast.

Here I attempt to examine the underlying ecology of South American YFV in order to address imperative issues about critical population sizes, the role of different NHP genus and the possibility for the establishment of endemicity outside of its current endemic zone. While there have been previous attempts to mechanistically model YF transmission, including that in primates (Moreno et al., 2015), they have not offered an in-depth look at the underlying ecology of transmission. Here I attempt to replicate a more realistic environment of YFV transmission by considering age and species structure, population movement, the role of mosquito vectors, seasonality in a stochastic patch metapopulation model. This is used to gain a greater understanding of the critical community thresholds required for endemic transmission, how different species and different species compositions in an area can influence the transmission

of YFV and how the seasonality of vectors impacts the seasonality of YFV transmission. This model is then applied to modelled estimates of NHP population sizes and distributions, in order to assess the thresholds for endemic and epidemic transmission of YF in the state of Rio de Janeiro.

7.3 Methods

7.3.1 Software

Modelling was carried out using “odin” (FitzJohn, 2019), a high-level language for describing and implementing stochastic model in the statistical programming language R version 3.6 (Team, 2015a).

7.3.2 Stochastic model

For simplicity, we have described the model as a discrete time model using ordinary differential equations as an approximation of continuous time stochastic model at single day time steps.

Movement through or out of compartments (S^h_{ijk}, E^h_{ijk}, I^h_{ijk}, R^h_{ijk}, S^v_i, E^v_i, I^v_i) via infection, incubation or recovery, as well mortality, movement and aging is initially decided through a binomial draw of all competing hazards of movement from a compartment. This determines the number leaving the compartment based on the probabilities of all outcomes, and a multinomial draw assigns these to the new compartments based on the individual hazards relative probabilities.

7.3.3 NHP

Four genera of NHP’s were selected to populate the model, Alouatta (Howler monkeys), Callicebus (Titis), Callithrix (Marmosets and tamarins), and Sapajus (Capuchin). These were selected based on their estimated population sizes in Brazil’s South-East Atlantic coast, diversity in life-histories and knowledge of their susceptibility to YF infection (Jones et al., 2009, Cunha et al., 2019, Kumm and Laemmert, 1950).

In order to improve the generalisability of results I have averaged the life histories of species within a genus. While there are local adaptations and diversity within different species of each genus found in the South-East Atlantic forest, the inter-genus variation is likely greater than the intra-genus variation with the sparsity of individual species data for all relevant parameters additionally required this approximation. Information on NHP rate of reproduction, the intrinsic incubation period, recovery rate, background mortality and aging were provided from various literature sources (Jones et al., 2009, Bates and Roca-Garcia, 1945) and are detailed in Table 7.3.3.

Realistic NHP distributions and population sizes for these four genera were defined as the populations and distributions provided through the modelling exercises detailed in the previous chapter. These were aggregated to a 1/5 degree resolution around the state of Rio de Janeiro. This resolution was chosen as a trade-off between increased insight into spatio-temporal patterns and computational limitations; it represents a maximum patch size of 460km^2. For distributions of different species see Figure 7.3.1.

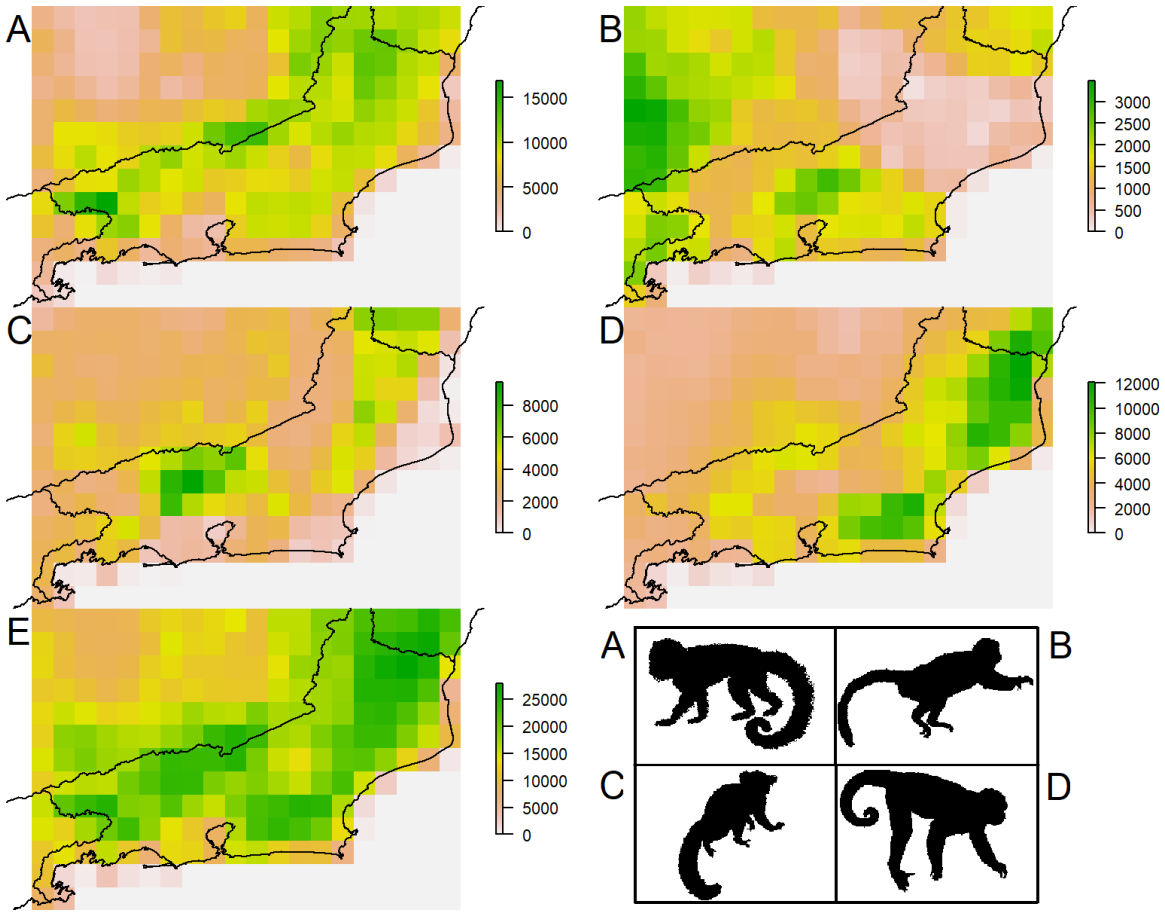


Figure 7.3.1. o The population distribution estimates provided by the previous chapters estimates of NHP population sizes for (A) Alouatta, (B) Callicebus, (C) Callithrix, (D) Sapajus, and (E) all genera, at a 1/5 degree resolution for the state of Rio de Janeiro and surrounding areas. These population estimates are used in the simulations of “realistic” population distributions

7.3.4 Vector

The primary vector of sylvatic YF in South America are mosquitoes of the *Haemagogus* genera, with additional transmission supported by those of the *Sabethes* genera. These sylvatic, diurnal mosquitoes are found throughout much of the continent in a variety of species, with *H. janthinomys* the primary species in recent outbreaks in the South-East Atlantic coast of Brazil (Abreu et al., 2019), which was used to parameterise our model.

7.3.5 Area of study

Simulations of the model using realistic depictions of NHP population size and distribution were focused on the state of Rio de Janeiro on the coast of South-Eastern Brazil and immediate border areas of Sao Paulo, Espirito Santo and Minas Gerais. This region had previously not reported YF in humans or NHP's before 2016, and so represents an introduction of the virus into a naïve location with large populations of susceptible hosts

7.3.6 Model

Figure 7.3.2 describes the compartmental model used to simulate YFV transmission within, and between, hosts and vectors, stochastically. Transmission is divided into two components, the host, NHP's, and the vector, here parameterised with *H. janthinomys* life-histories.

Infection was introduced at 10 random intervals during years 5 – 10, and the model ran for a total of 30 years (except in evaluating the long-term population dynamics of NHP genera), in order to represent successive introductions and investigate the long-term persistence of the virus. This length of 30 years was found to be suitable to look at long term persistence, as the persistence of YF transmission had stabilised before this point.

For the NHP's I consider movement between patches for the same compartment, two age categories of infant (non-reproductive) and adult (reproductive), and 4 genera of primates commonly found in the South-East Atlantic states of Brazil, Alouatta, Callicebus Callithrix and Sapajus (Culot et al., 2019).

Table 7.3.1. List of indices and definitions

Indices	*Definition*
i	Patch
j	Age (infant or adult)
k	Species (Alouatta, Callicebus, Callithrix and Sapajus)

Host populations are classified as Susceptible (S^h), Exposed (E^h), Infectious (I^h) and Recovered (R^h), and vector populations as, adult susceptible (S^v), adult exposed (E^v) and Infectious (I^v). As only female mosquitoes take a blood meal, I am only considering female mosquitoes in the model.

For hosts, each compartment is affected by the background mortality rate, μ^h_{ijk}, and movement from and into patches, ω_{ijk} and M_{ijk}, respectively. Aging occurs from infant indices to adult, at genus dependent rate φ_k where if j is 1 (infants) then the inflow is 0, and if j is 2 (adults) then outflow is 0. New susceptible infants enter the S^h compartment through births$^h_{ijk}$, and from S^h to E^h occurs due to the force of infection, λ^h_{ijk}, and progression from exposed to infected, I^h, occurs at rate τ^h_{ijk}, which represents the intrinsic incubation period. Infectious hosts recover at rate, γ_{ijk}, where they are transferred to compartment R^h relative to the case fatality ratio (CFR). The infectious period of YF relates to the initial period where symptoms are mild or asymptomatic, and there is no disease specific mortality, rather than the period of intoxication where severe disease causes substantial mortality, but no onward transmission occurs (Monath, 2001). Therefore, the effect of the CFR relates solely to the proportion of

NHP's that are recovered in a population. Upon recovery, immunity is assumed lifelong and complete.

Vectors are assumed to be immobile, and do not move between patches. Vectors enter through *births*$^{v}_{i}$, and progress to adult susceptibles, S^v, at rate ρ_i. Here, transition to exposed occurs when a susceptible mosquito feeds on an infectious host, with the force of infection, λ_i^v, and progression from E^v to infected, I^v, occurs at rate τ_i^v. Infection is assumed to be lifelong and not associated with increased mortality for the vector.

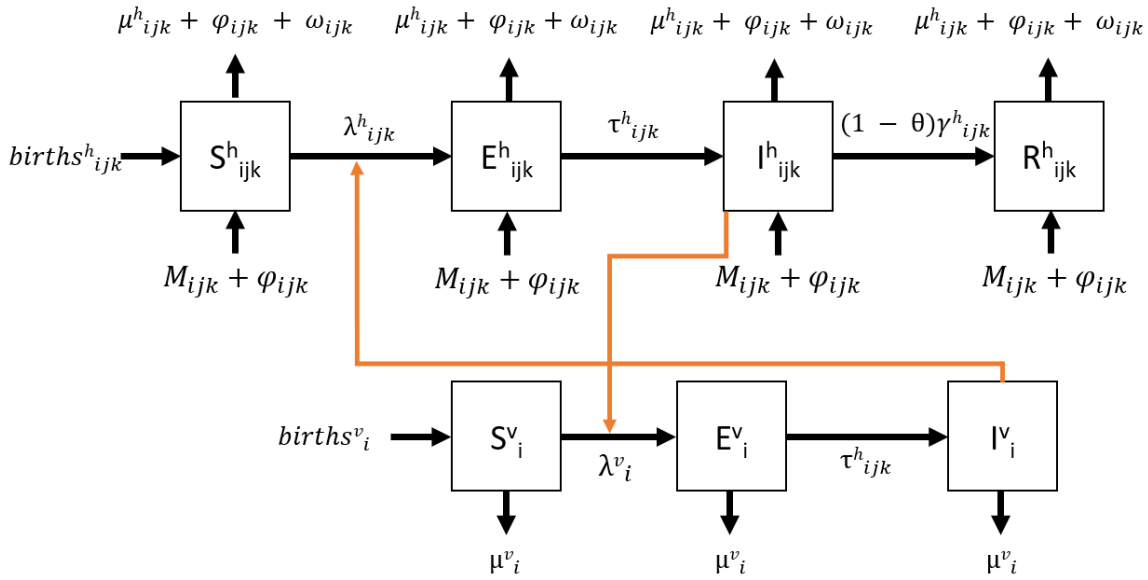


Figure 7.3.2. Diagram of the YFV transmission model where black arrows represent transition between states, and orange arrows represent interactions between host and vector. States for the host are defined as, S^h, susceptible hosts, EThe subscripts indicate, (i) the patch, (j) the age, infant or adult, and (k) the species. States and parameters with a superscript ^{h} indicate they are related to the host, and ^{v} indicate they are associated with the vector.

Hosts

$$\frac{dS_{ijk}^h}{dt} = births_{ijk}^h + \varphi_k S_{ijk}^h + M_{ijk} - \lambda_{ijk}^h S_{ijk}^h - (\omega_{ijk} + \mu_{ijk}^h + \varphi_k) S_{ijk}^h$$

$$\frac{dE_{ijk}^h}{dt} = \lambda_{ijk}^h S_{ijk}^h + \varphi_k S_{ijk}^h + M_{ijk} - \tau^h E_{ijk}^h - (\omega_{ijk} + \mu_{ijk}^h + \varphi_k) E_{ijk}^h$$

$$\frac{dI_{ijk}^h}{dt} = \tau^h E_{ijk}^h + \varphi_k S_{ijk}^h + M_{ijk} - \gamma_k I_{ijk}^h - (\omega_{ijk} + \mu_{ijk}^h + \varphi_k) I_{ijk}^h$$

$$\frac{dR_{ijk}^h}{dt} = (1 - \theta_{jk})\gamma_k I_{ijk}^h + \varphi_k S_{ijk}^h + M_{ijk} - (\omega_{ijk} + \mu_{ijk}^h + \varphi_k) R_{ijk}^h$$

Vectors

$$\frac{dS_i^v}{dt} = births_i^v - \lambda_i^v S_i^v - \mu_i^v S_i^v$$

$$\frac{dE_i^v}{dt} = \lambda_i^v S_i^v - \tau_i^v E_i^v - \mu_i^v E_i^v$$

$$\frac{dI_i^v}{dt} = \tau_i^v I_i^v - \mu_i^v I_i^v$$

Where births are calculated through,

$$births_{ik}^h = \delta_k^h A_{ik}^h,$$

$$births_i^v = \mu_i^v S_i^v + \mu_i^v E_i^v + \mu_i^v I_i^v .$$

Here, species-specific host birth rate is denoted by δ_k^h, and the total female adult population by species, A_{ik}^h. For vectors, the births are equal to deaths in order to keep a stable population size.

The population of NHP's is assigned depending on the scenario in question, while the population of vectors is a ratio of the total NHP population and the product of the basic reproductive number, R_0, which is calculated using the assumption that,

$$R_0 = \frac{ma^2bce^{-\mu EIP}}{\gamma\mu},$$

Where *m* is the vector to host ratio, *b* and *c* the probability of transmission between host and vector and vector and host respectively, and *γ* the recovery rate of the human host, *μ* the background mortality rate of the vector, *a,* the biting rate of the vector and *EIP* the extrinsic incubation period. Here EIP is a distribution, which may result in unrealistically short incubation periods. This is used to calculate the ratio of vector to host given an R_0 value through (Hartemink et al., 2015),

$$m = \frac{R_0\gamma\mu}{a^2bce\left(\frac{\frac{1}{EIP}}{\frac{1}{EIP} + \mu}\right)}.$$

Which then provides the patch specific mosquito population through,

$$N_i^v = mN_i^h .$$

The calculation for the force of infection for vector and host are taken from Suparit et al., (2018), and calculated as,

$$\lambda_{ijk}^h = \frac{\sigma^v\sigma^h N_i^v}{\sigma^v N_i^v + \sigma^h N_i^h}\beta^h \frac{I_i^v}{N_i^v},$$

$$\lambda_{ijk}^v = \frac{\sigma^v\sigma^h N_i^h}{\sigma^v N_i^v + \sigma^h N_i^h}\beta^v \frac{I_i^h}{N_i^h},$$

where σ^v refers to the number of times a vector can bite a host per unit time and σ^h the maximum number of bites a host can support (Suparit et al., 2018). N_i^v is the number of adult vectors in a patch, and $N_i^h = \sum_{jk} N_{ijk}^h$,the total number of hosts in a patch regardless of age or species. β^h is the probability of an infected vector infecting a host on feeding, and β^v, the probability of an infected host infecting a vector on

feeding. To account for seasonality, the biting rate is modulated by a seasonal oscillation in the intensity through

$$\sigma^{v} = \rho\xi\left(1 + 0.6\sin\left(\frac{2\pi(t)}{365}\right)\right),$$

Where ξ is the ratio of vectors to hosts, t the time and ρ the rate of the gonotrophic cycle. Biting rate is three times higher in the high season as opposed to the low season to simulate changes in real-life *Haemagogus* populations (Chadee et al., 1995) which is used to set the default seasonality modifier at 0.6.

Movement from a patch is based on the patch size, and genus movement where, assuming equal distribution of NHP's in a patch, the proportion of NHP's within the genus movement range of the edge of a patch is calculated through,

$$\omega_{ijk}^{h} = \frac{(\chi_{k}^{h})^{2}}{\varsigma},$$

Where χ_{k}^{h}, is daily movement in km, and ς is the patch size in km^2. Movement into patches depends on the suitability of the habitat, which is based on the initial NHP population and calculated through,

$$\psi_{i} = \frac{N_{i}^{h}}{\max\left(N_{i}^{h}\right)},$$

to produce a suitability index that ranks patches on a 0 – 1 scale based on population.

This is then used to determine the likelihood of an NHP entering patch using the relative suitability of neighbouring patches, as NHP's can move horizontally, vertically and diagonally each patch has 8 potential neighbours. The number of NHP's of entering each patch by compartment is calculated by,

$$M_{ijk} = \sum_{x \in \eta_i} \alpha_{xjk} (\omega_{xjk}^h N_{xjk}^h),$$

With α_{ijk}, the probability of entering each patch calculated by,

$$\alpha_{ijk} = \sum_{x \in \eta_i} g(\psi_{\eta_x}).$$

Where,

$$g(\psi_{\eta i}) = \frac{\psi_j}{\psi_{\eta i}} \; and \; \psi_{\eta i} = \sum_{x \in \eta_i} \psi_x \, ,$$

where η_i are the neighbours of patch *i*. This is then used to assign NHP's that leave a patch into neighbouring patches.

The carrying capacity dictates the underlying mortality rate for NHP hosts, μ_{ijk}^h, through,

$$\mu_{ijk}^h = \epsilon_{ijk}^h + \frac{(\epsilon_{ijk}^h \, \vartheta_k^h)}{(1 + \frac{c_{ijk}^h}{\sum_j N_{ijk}^h})^x}.$$

Here, ϵ_k, refers to the background mortality rate, x, the power the value is raised to which determines the slope of the effect of carrying capacity, ϑ the species-specific mortality modifier – which modulates the population replenishment below 1, is calculated for hosts by,

$$\vartheta_k^h = 0.5 \left(\frac{\delta_k^h}{\epsilon_k^h} \right)$$

where the 0.5 represents the female population (proportionally half the total), and the, patch carrying capacity, *c*, is provided through the total number of NHP's at time 0 so that at equilibrium,

$$c_{ik}^{h} = N_{ik}^{h}.$$

Model defaults

Unless otherwise stated, the model specifications including patch size, number of patches and populations are shown in Table 7.3.2.

The default number of patches (16) and population size (10,000) was chosen following initial exploration of the relationship of population size and the number of patches which found that this combination, in more than 50% of runs, resulted in endemic transmission of YF in a seasonal transmission setting. The patch size of 1km^2, which is related to the rate of movement between patches was chosen as a default in order to focus on demography as an explanation for YFV persistence. All genera were included, as much of the study area has overlapping distributions of these and to explore the role each genus plays in the maintenance of YFV. Due to a lack of quantified knowledge of CFR, other than it appears highly lethal to certain primate genera (Bates and Roca-García, 1946b), and the default R0 chosen based on a reasonable assumption following estimates in human outbreaks (Kraemer et al., 2017, Massad et al., 2003). The model was run for 10950 days and the mean of 100 runs taken for each simulation as this was found to be more than sufficient runs to capture overall trends and the probability of stochastic die-off to be minimal.

Table 7.3.2. Default values for indices and specific parameters altered in model exploration

Model specification	***Default value***
Number of patches	16
Patch size	1km^2

CFR	0.5
R0	2
Total NHP population	10,000
NHP genera included	Alouatta, Callicebus, Callithrix, Sapajus
Number of runs	100
Time model ran for	10950 days (30 years)

Table 7.3.3. Descriptions and values of all parameters used in the model.

Cycle	*Parameters*	*Definition*	*Value*	*Unit*	*Reference*
Host	A_{ik}^{h}	Adult female host population in patch i a of species k	Half of the adult NHP population in a patch	Count	
Host	σ^{h}	The number of bites a host can support per unit time	20	Count	Assumed similar to humans (Suparit et al., 2018)
Host	δ_{k}^{h}	Rate of reproduction (lifetime births/life expectancy in days) of species k	Alouatta: 19.32/8127 Callicebus: 22.87/9108 Callithrix: 46.77/7085 Sapajus: 43.63/5733	Days	(Jones et al., 2009)
Host	β^{h}	Probability of transmission host to vector	0.5	Probability	Assumed
Host	τ^{h}	The intrinsic incubation period	1/3	Days	(Bates and Roca-García, 1946b)
Host	γ_{k}	The recovery rate of species k	1/5	Days	(Bates and Roca-García, 1946b)
Host	ψ	Patch suitability	0 - 1		Calculated based on total NHP population in a patch
Host	c_{i}^{h}	Carrying capacity of patch i	Total of NHP population in a patch	Count	Assumed as initial total NHP population in a patch
Host	x	The slope of the carrying capacity effect	10		Calculated
Host	ϵ_{k}^{h}	Background mortality rate (1/life expectancy in days) of species k	Alouatta: 1/8127 Callicebus: 1/9108 Callithrix: 1/7085 Sapajus: 1/5733	Days	(Jones et al., 2009)
Host	φ_{k}	Age rate (1/age at sexual maturity) of species k	Alouatta: 1/1515 Callicebus: 1/1473 Callithrix: 1/621 Sapajus: 1/622	Days	(Jones et al., 2009)
Host	θ_{jk}	Case Fatality Ratio (CFR) of age j and species k	0.5 (0-1)	Probability	Assumed and varied
Host	χ_{k}^{h}	Daily movement assumed from their home range for species k	Alouatta: 0.31km Callicebus: 0.14km Callithrix: 0.29km Sapajus: 0.14km	Kilometers	(Jones et al., 2009)
Host	α_{k}^{h}	The probability of entering a patch		Probability	Calculated
Vector	ρ	Baseline biting rate	1/7.25	Days	(Degallier et al., 1998)

Vector	σ^v	Seasonal biting rate	$\rho\xi\left(1+0.6\sin\left(\frac{2\pi(t)}{365}\right)\right)$	Days	Calculated
Vector	τ^v	Extrinsic Incubation Period	1/12	Days	(Bates and Roca-Garcia, 1945, Bates and Roca-García, 1946a)
Vector	β^v	Probability of transmission vector to host	0.5	Probability	Assumed
Vector	R_0	Basic reproductive number	2 (1 – 5)	Days	Assumed and varied
Vector	ϵ_k^v	Background mortality rate (1/life expectancy in days)	1/25	Days	(Degallier et al., 1991)

7.4 Results

7.4.1 Transient YFV dynamics for a single patch

Within a single patch there is substantial reduction in the number of susceptible NHP's following the introduction of YFV. Here we see that from an original population of ~1000, the majority of NHP's experience infection and the number of susceptible primates falls rapidly following YFV introduction at ~2000 days. Following this the number of susceptible primates "bounces back" before falling following another wave of infection. This continues throughout the time period observed, though at a lower magnitude between peaks and troughs. This results in the population of susceptible primates being maintained around 400 individuals, before additional infections occur and reduce this.

Here infection becomes endemic following the first introduction.

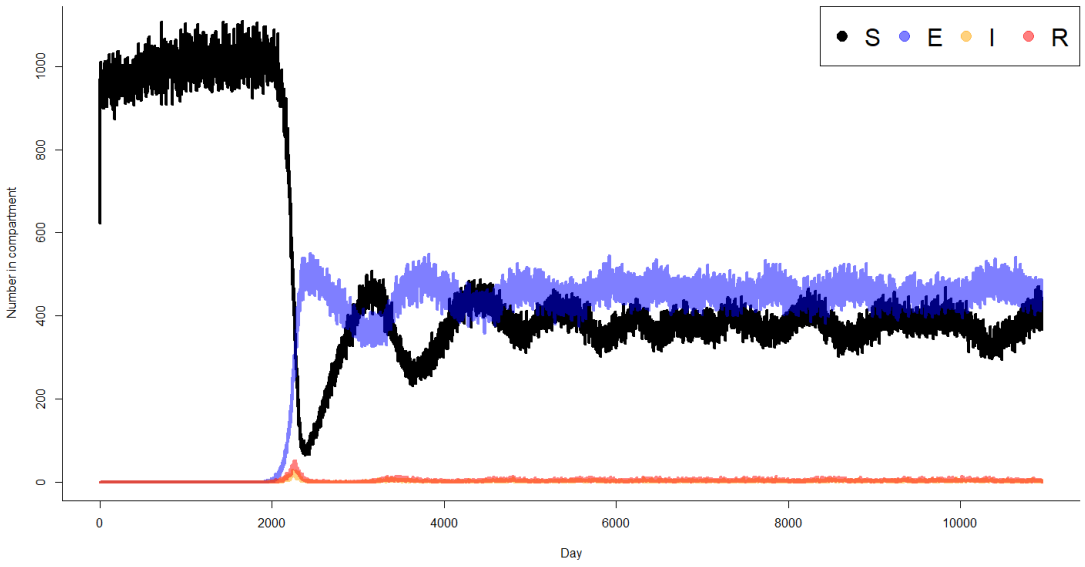


Figure 7.4.1. Transient YFV dynamics for the total population of NHP's (the sum of all genera and ages) for a single patch. This graph depicts the total population of Susceptible (S), Exposed (E), Infected (I) and Recovered (R) NHP's over 10000

days following infection at ~2000 days. Refer to Table 7.2 for default values of non-varying parameters.

7.4.2 Relationship between the number of patches and populations

Persistence of the YFV in NHP populations was related to both the number of patches and the overall population size, with larger populations or the numbers of patches associated with a lower probability of extinction (Figure 7.4.2).

Seasonality had a deleterious effect on YFV persistence. In a single patch and with a population of 2500 individuals, simulations without seasonality had a ~50% chance of YFV becoming endemic on average, while this fell to 0% with seasonality. In order to reach a patch and population combination where extinction is less than 50% in the seasonal simulations, 9 patches and 10000 individuals are required (extinction probability 47%).

At higher numbers of patches and populations, YFV in the non-seasonal model has an almost 0% chance of extinction and even with high population sizes and large numbers of patches, in the seasonal model there remains a substantial chance for extinction (10-30%).

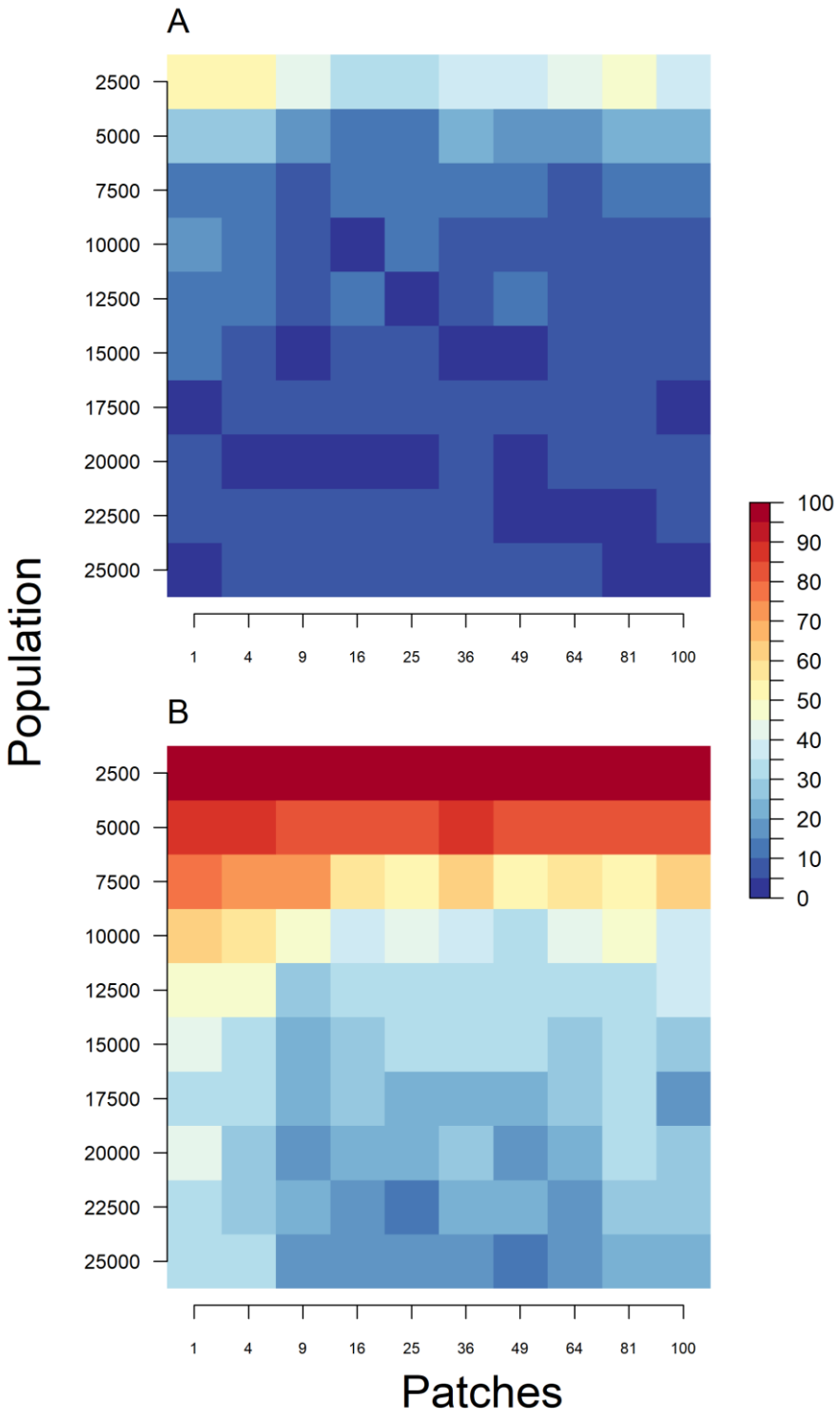


Figure 7.4.2. The mean probability, over 100 simulations, of YFV extinction given different patch and population sizes for simulations (A) without seasonality and (B) with seasonality. Refer to Table 7.3.2 for default values of non-varying parameters.

7.4.3 Relationship between CFR and R_0

While at higher levels of R_0, persistence is highly likely in both seasonal and non-seasonal simulations, there is a substantial effect of the CFR on the required R_0 to sustain transmission (Figure 7.4.3).

Higher values, up to but not including 1, of CFR increase the probability of YFV persistence in a population. Higher levels of mortality reduce the number of recovered, and so the overall "herd immunity" in a population, increasing the likelihood of persistence, as well as leaving a "gap" to be filled in the population relative to the carrying capacity, replenishing susceptibles.

Here, even in the seasonal model, if YF has a CFR of 0.9, R_0 is only required to be 1.25 for a greater than 50% chance of persistence (Figure 7.4.3). Conversely, if the CFR is low then a much higher R_0 value is required.

The absence of persistence seen with a CFR of 1 in both simulations is due to the reduction of the NHP population (with a default value of 10,000) to levels where sustained transmission is not possible.

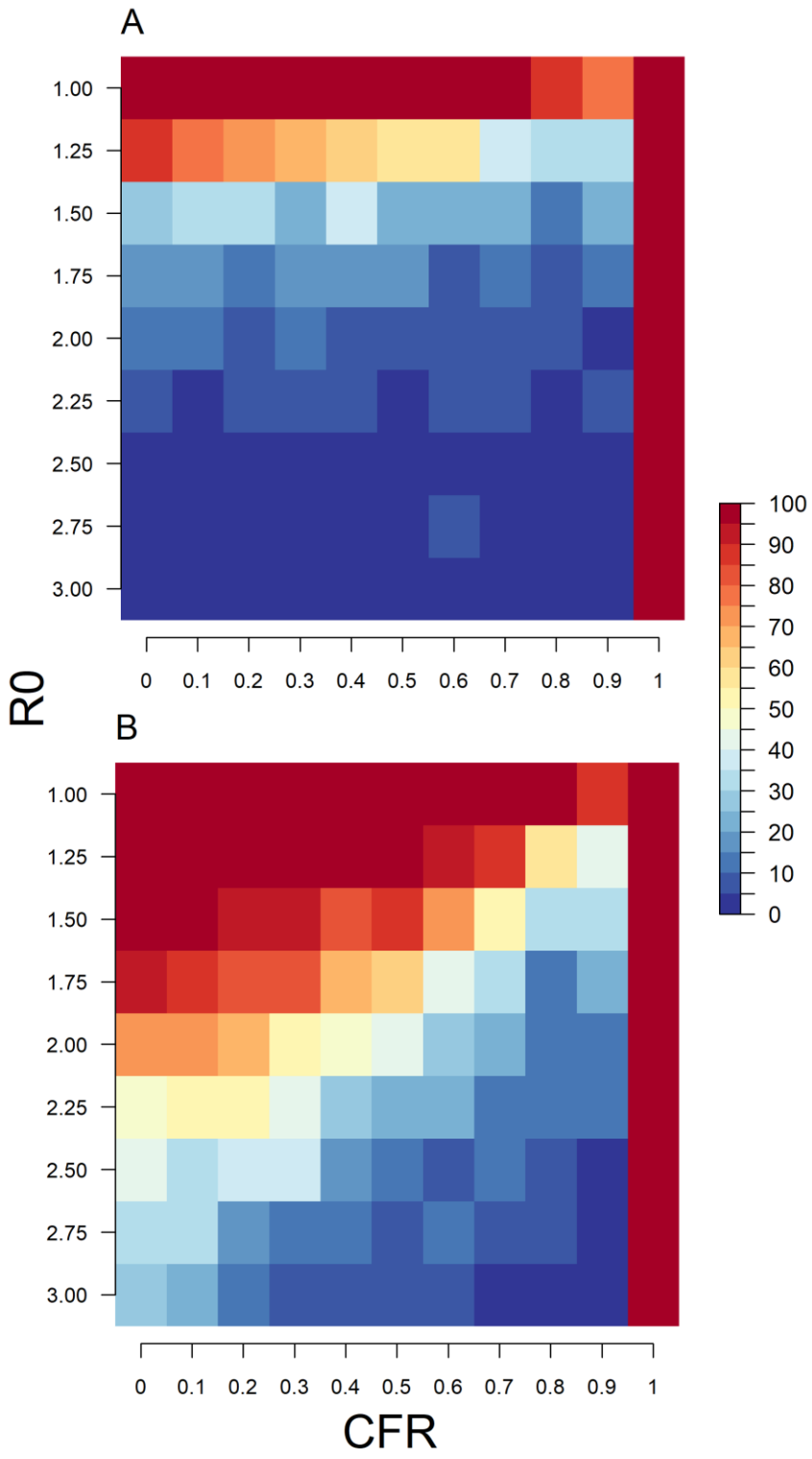


Figure 7.4.3. The mean probability, over 100 simulations, of YFV extinction given differing CFR and R0s with the probability of YFV extinction in simulations (A) without and (B) with, seasonality. Refer to Table 7.3.2 for default values of non-varying parameters.

7.4.4 Relationship of R_0 and population size

As R_0 increases, the total population size of NHP's required for persistence decreases. At an R_0 of 1.75, the probability of YFV extinction is ~50% in an overall population of 2500 NHPs, however once this increases to 5000 the probability drops to ~30%. Similarly, if the R_0 were to increase to 2, the probability of extinction would drop to ~40%, and by 3 the probability is almost 0%.

This relationship is particularly evident for seasonal simulations where higher populations and higher R_0 values are required for persistence. A population of 2500 NHP's is not able, at any value of R_0, able to sustain YFV – while a population of 10,000 NHP's is able to at a R0 of 2.

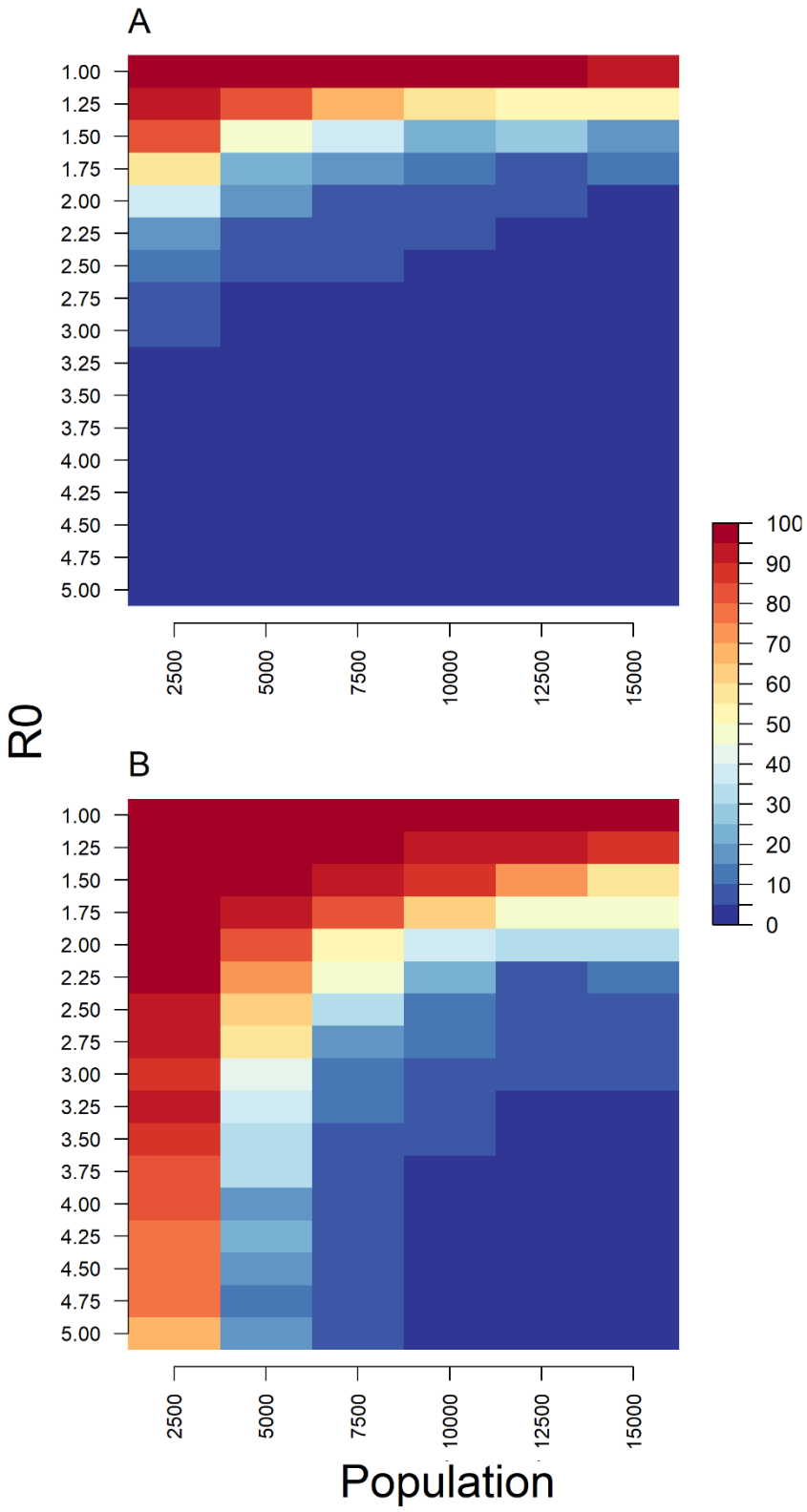


Figure 7.4.4. The mean probability, over 100 simulations, of YFV extinction given a range of R0 values and population sizes. Refer to Table 7.3.2 for default values of non-varying parameters.

7.4.5 The role of seasonality on persistence

Higher levels of seasonality have a detrimental effect on the probability of YFV persistence. At higher levels of seasonality, larger populations of NHP's are required to sustain transmission. For a population of 5000 NHP's, seasonality must be at 0.4 or lower to have a less than 50% chance of YFV extinction, while for a population of 15,000 NHP's, YFV is able to persist at seasonality levels of 0.7 in more than 50% of simulations.

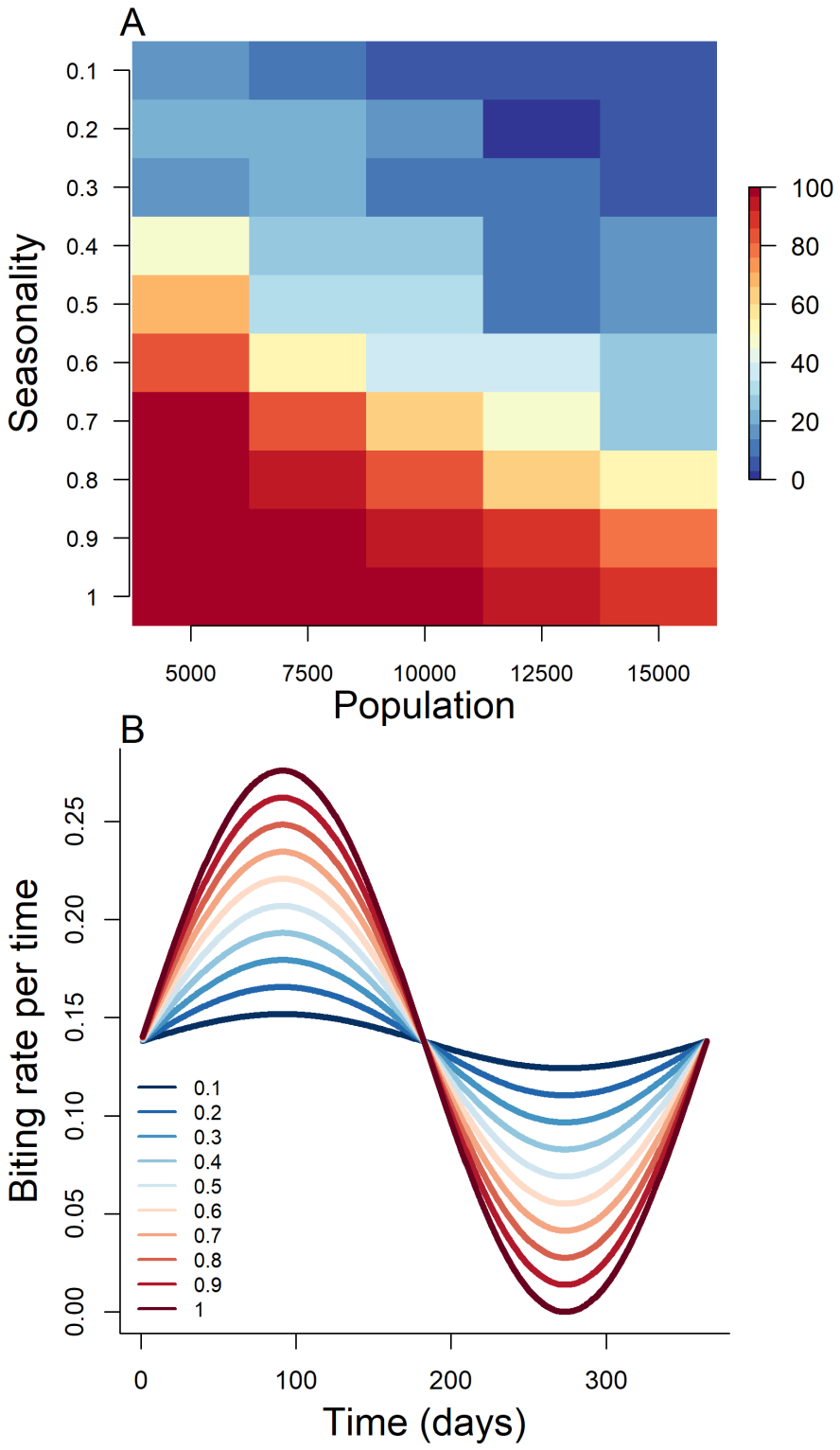


Figure 7.4.5. (A) The relationship of the level of seasonality on the population required for persistence of YFV, and (B) the corresponding biting rates over a year for each level of seasonality. Refer to Table 7.3.2 for default values of non-varying parameters.

7.4.6 Relationship of genera presence and population size

NHP genera inclusion, or exclusion, had a substantial effect on the persistence of YFV (Figure 7.4.6). Here each row corresponds to the probability of YFV persistence in populations composing of the genera specified on the Y axis. Where there are several genera included, populations are divided equally.

Alouatta and Callicebus required substantially higher populations than Callithrix and Saguinus in order to maintain transmission. In non-seasonal simulations, those which only included these genera consistently possessed a greater than 50% probability of YFV extinction – even with population size of 15,000. Conversely, Callithrix and Saguinus only simulations had low probabilities of extinction: below 20% in all non-seasonal models that included either genera and were able to consistently maintain transmission in models that included only themselves.

In simulations that contain Callithrix and or Saguinas, Callithrix act as an enhancer of transmission. Similarly, in simulations containing Alouatta and/or Callicebaus, Alouatta act as an enhancer of transmission. The enhancers allow YFV at lower population levels than otherwise possible.

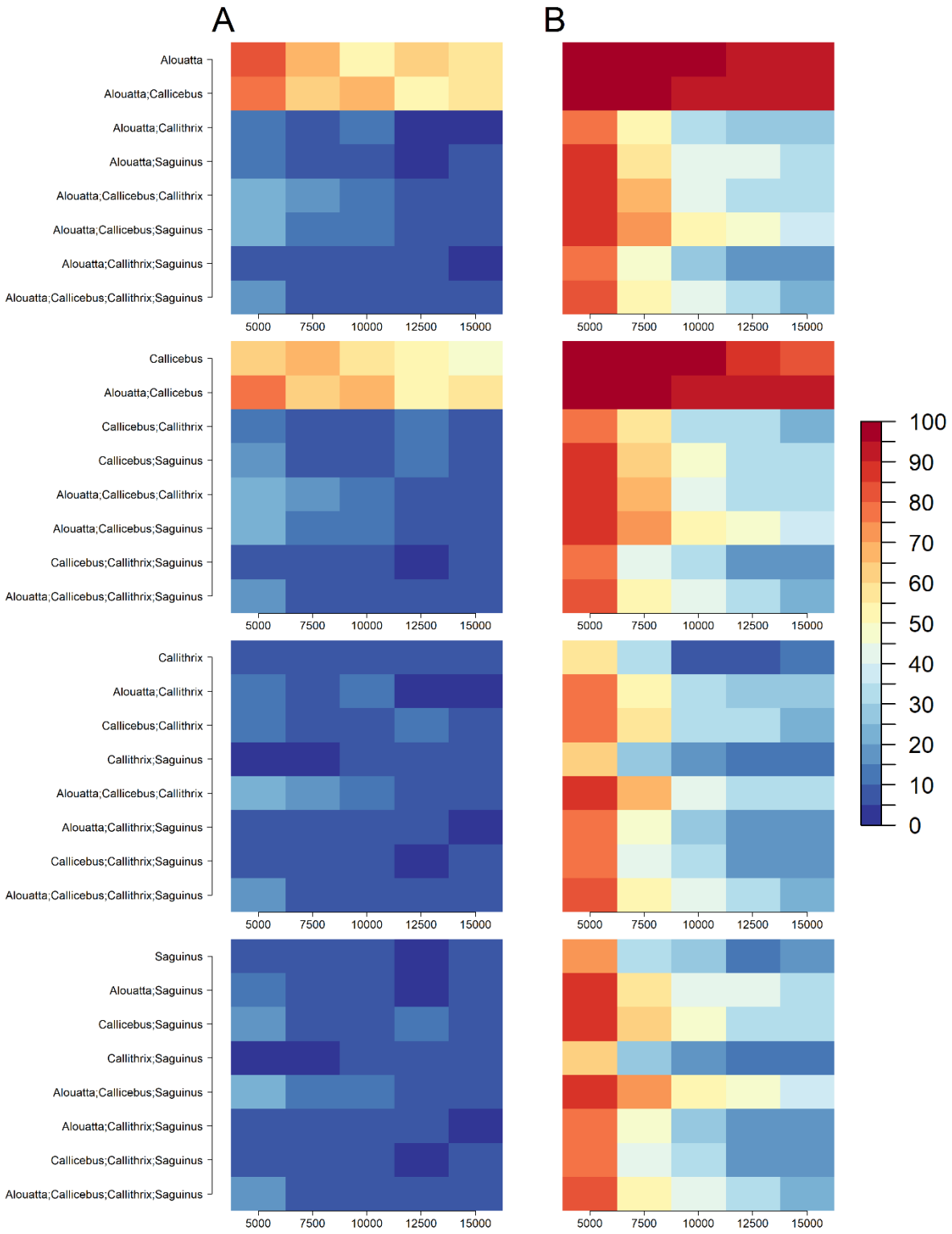


Figure 7.4.6. The relationship of model genera composition and population size on the probability of mean YFV extinction, over 100 simulations, (A) without and (B) with seasonality. Refer to Table 7.3.2 for default values of non-varying parameters.

7.4.7 YFV persistence in Rio de Janeiro – comparison of uniform and realistic distributions of NHPs

YFV persistence in Rio de Janeiro and the immediate surrounding states is highly likely even with population sizes significantly below those predicted (Figure 7.4.7).

While the realistic population sizes and distributions of NHP genera increase the probability of YFV extinction, even with population sizes 7.5% of the original predictions, and under assumptions of seasonality, persistence is likely (with only a 32% probability of extinction). For non-seasonal simulations with equal genera distributions, populations sizes need only be 2.5% of the original predictions in order to reduce the probability of YFV extinction below 10%. For non-seasonal simulations with realistic genera distributions, the population sizes must be 5.0% of the original prediction to reach the same threshold. With seasonal simulations, population sizes need to be 10%, for equal genera distributions, and 12.5%, for realistic genera distributions, to reduce the probability of YFV extinction below 10%.

Realistic distributions of NHP's lead to minor increases in prevalence in non-seasonal simulations at populations larger than 0.125 of the modelled estimates of population size, and equal them in seasonal simulations at 0.200. Prevalence increases with population size till in plateaus around 0.58% for non-seasonal realistic distributions and 0.50% for non-seasonal uniform distributions of NHPs.

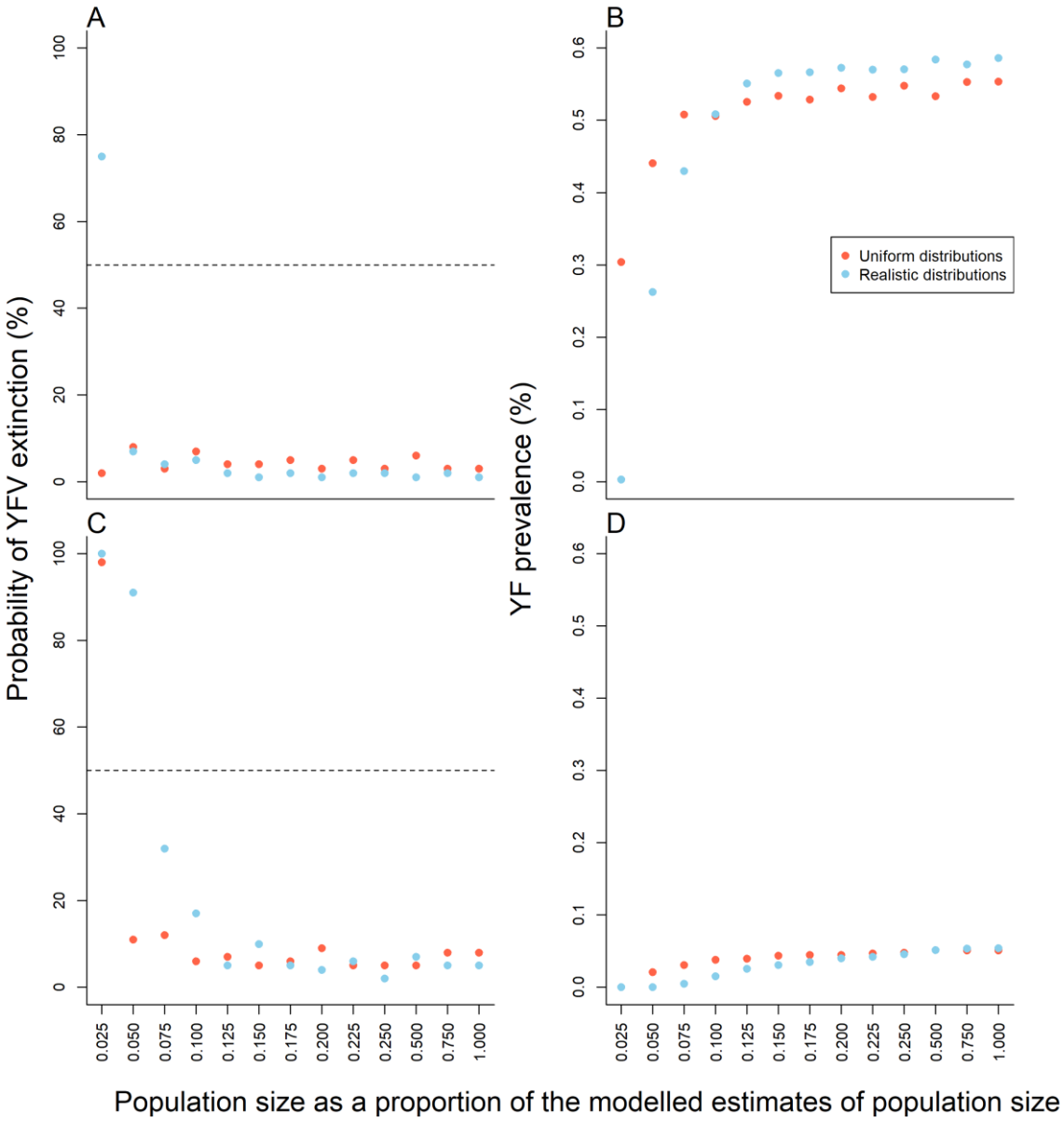


Figure 7.4.7. Comparison of the mean probability of YFV extinction for (A) non-seasonal and (C) seasonal models, and the end prevalence of YFV infection across all NHP's for (B) non-seasonal and (D) seasonal models using uniform or realistic distributions of NHP's. Refer to Table 7.3.2 for default values of non-varying parameters.

7.4.8 Effect of YFV persistence on NHP populations

Population dynamics are different for each genera, with some genera seemingly unaffected and others trending towards extinction (Figure 7.4.8).

Despite originally composing the majority of the NHP population, *Alouatta* show a marked decrease in population size after infection first appears in a population, reducing from over 700,000 individuals, to less than 25,000 within 20 years, and falling to almost 0 after 100 years. Callicebus, the rarest of NHP genera, also experiences substantial population declines from ~18,000 to almost 0 after 100 years.

Both Callithrix and Sapajus are minimally affected by YF endemicity, while there are fluctuations in response to YF transmission, both genera increase their overall population sizes.

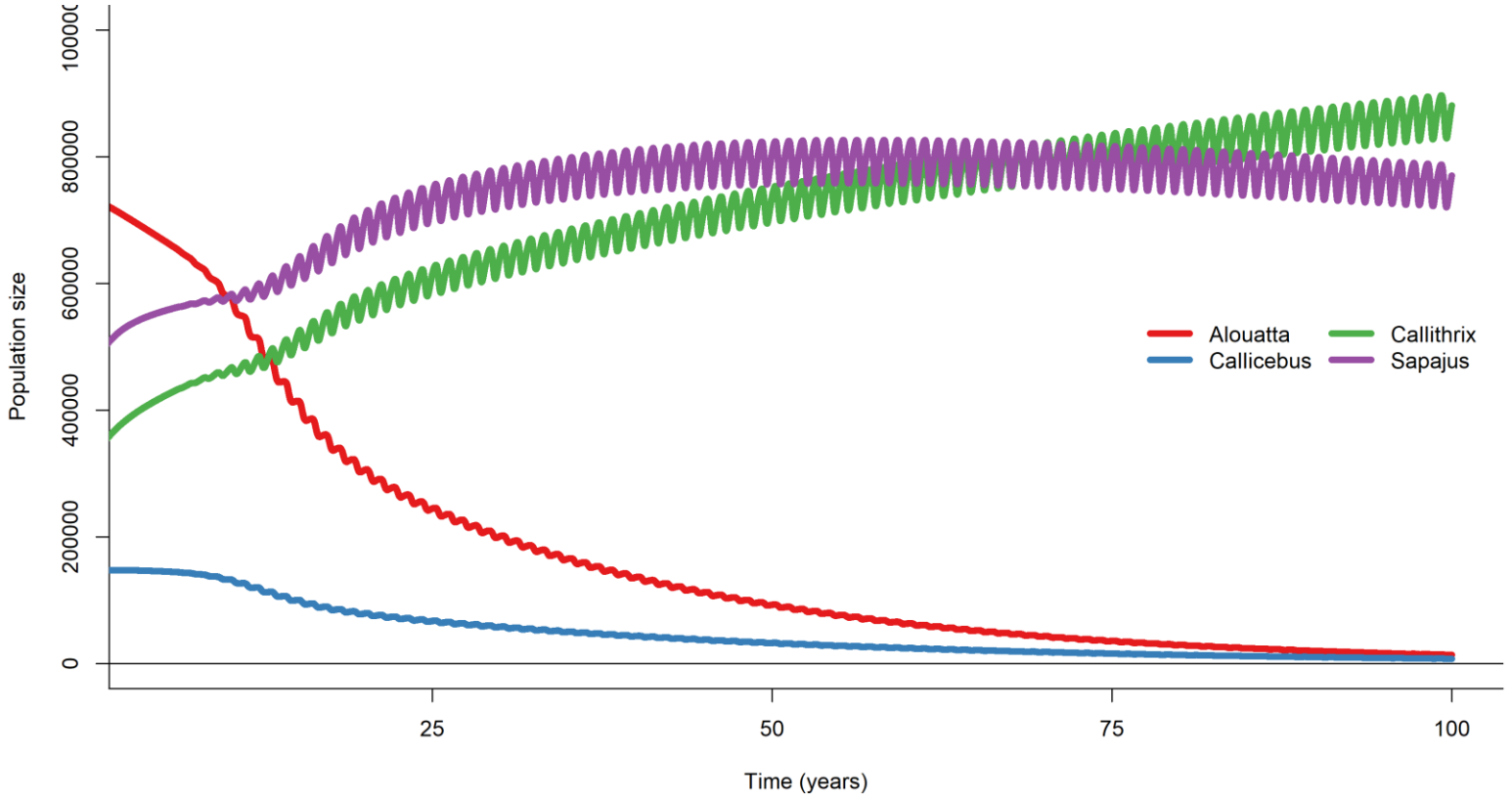


Figure 7.4.8 The mean populations for each of the NHP genera over 100 runs for 100 years. Models were run with realistic distributions of NHP's and population sizes.

7.5 Discussion

Here I have modelled the transmission and persistence of YFV in NHP populations using a stochastic meta-population model, which accounts for age and 4 of the most common NHP genera found in, and around, the Brazilian state of Rio de Janeiro.

I show that relatively small populations of NHP's are required to sustain transmission, even under the suppressive effects of seasonality. In addition to seasonal dynamics, this critical community size is highly dependent on the composition of NHP genera found in the area, with physically larger and slower to reproduce genera such as *Alouatta* and *Callicebus* requiring substantially more individuals to sustain transmission than the shorter lived and more rapidly reproducing *Callithrix* and *Saguinus*. Using modelled estimates of NHP population sizes, distributions and genera overlaps I have found that the establishment of YFV endemicity in the state of Rio de Janeiro is likely, a finding robust to substantially lower predicted population sizes.

These findings have significant ramifications for both NHP conservation and human health. YFV causes significant epizootic events across South America, with events capable of causing the die-off of thousands of NHPs (Cunha et al., 2019, Romano et al., 2014, Moreno et al., 2013). While I have shown that due to their individual life histories, certain NHP genera are unlikely to be negatively impacted in the long term, other species may be. This additional pressure of morbidity is particularly detrimental to primates in the South-East Atlantic coast whose habitat, the Atlantic Rainforest, only occupies up to 16% of its original range, and is highly fragmented (Ribeiro et al., 2009). Within the remnants of this forest, 26 species of primate are found, 19 of which are endemic, and 13 endangered (IUCN, 2017, Culot et al., 2019). Smaller, fragmented

populations are more vulnerable to extinction through human activities, increased susceptibility to infection and inbreeding depression (Michalski and Peres, 2005, Mbora and McPeek, 2009, Bicca-Marques and de Freitas, 2010, Crnokrak and Roff, 1999). Thus, there is the real possibility that the regional extinction of certain species may be driven, in part, by YF. This additional extinction pressure provided by YF in the South-East Atlantic forest highlights the need for additional conservation of the resident NHPs with larger populations, in less fragmented habitats, more resistant to these outbreaks (White et al., 2018).

The discovery of the role of particular species in YFV maintenance may be used to the advantage of YF control. While not currently implemented, the use of a NHP vaccine for halting YFV transmission has been postulated several times (Kalter and Jeffries-Klitch, 1969, Massad et al., 2018b). Though there are many logistical and economic hurdles to this approach, the accurate identification of the key maintainers of transmission could allow for a more targeted approach, reducing these obstacles.

From a human perspective, the establishment of endemicity in Rio de Janeiro has substantial consequences for public health. Following recent YF transmission in the South-East Atlantic coast (2016-2019), the recommendation of vaccination for Brazil has been expanded from the previous endemic zone to the entirety of Brazil (CDC, 2019). This has been complemented with the large-scale vaccination of individuals in many of these previously unexposed states, such as Rio de Janeiro (Possas et al., 2018). While these efforts have vaccinated tens of millions of people in a relatively short time span, there still remain millions unvaccinated. Additionally, unless routine immunization of infants is implemented, susceptible populations will quickly replenish. The presence of continued YFV transmission in NHP populations in close, often overlapping, proximity to densely populated, *Aedes* infested, areas of human

population should be a situation of significant concern (Massad et al., 2018a). While urban YFV transmission has not been detected in Brazil since the 1940's, given the rapidity and magnitude of previous urban outbreaks of YF, it is imperative that vaccination coverage is kept at levels high enough to deter transmission (World Health Organization, 2017b). This is important not just for local populations, but globally, as the exportation of the virus to seemingly suitable, but historically unexposed, populations in much of tropical Asia could be devastating (Wasserman et al., 2016).

Our model has found that at higher levels of seasonality, larger populations are needed for disease persistence, findings in line with previous research (Altizer et al., 2006). Areas with higher seasonality, are generally associated with the extremes of the YF endemic zone away from the equator, and lowered primate diversity (Jentes et al., 2011, Estrada et al., 2017). This lowered diversity, and if associated with primate population sizes, would decrease the probability of YFV persistence – though this can be compensated through highly susceptible and rapidly reproducing species. In addition to this, there are well known seasonal limitations of the transmission of vector-borne diseases (Huber et al., 2018), and so the limits of the endemic zone of the YFV are likely an interaction between host availability and the suitability of the environment for vector mediated transmission.

Another finding of this study is that the critical community sizes required to sustain endemicity in NHP's are relatively small. The critical community size, the population required to maintain YFV persistence, is a ratio of the infectious period and the lifespan/birth rate of the populations in question (Bartlett, 1960). The relatively rapid reproduction rates, and smaller lifepsans of Sapajus and Callithrix generas allow them to "lower" the required NHP population for disease persistence, compared to if the population were primarily made up of Alouatta and Callicebus. These findings are

many orders of magnitude lower than the critical community size required for human pathogens such as measles, but they are larger than estimates for swine flu in domestic pigs (Pitzer et al., 2016, Bartlett, 1960). While the density and mixing patterns of domestic pigs are likely very different to that of NHP's in the wild, there is very little quantification of critical community sizes in sylvatic disease, and so while these do not offer a direct comparison, they are useful as a qualitative evaluation (Lloyd-Smith et al., 2005).

The relationship between R0 and CFR is one of interest. At low levels of R0, a high CFR (but below 1) is associated with an increased probability of YF persistence, and at low levels of CFR, a high R0 is required. While this may be counter-intuitive, it is understandable in the context of YFV infection and intoxication. YFV is only transmissible during the acute stage of infection, where the individual is asymptomatic or has mild non-specific symptoms not associated with mortality. Following this, a patient either recovers to no further symptoms and lifelong immunity, or develops severe disease associated with mortality, at which point they are no longer capable of onward transmission. Therefore, increased mortality only serves to remove "recovered" individuals, who display lifelong immunity, from the population, in effect lowering the influence of "herd immunity". In this model, this has a substantial impact on the persistence of the virus, with an increased CFR resulting in a higher likelihood of the virus to be able to be maintained in a population. One thing to consider is whether the dynamics modelled are likely represented in nature, here we have assumed that following primate die-off due to YFV infection, mosquito populations change to reflect this. However, as *Haemaegogus* species are generalist, feeding on numerous hosts, they may be able to sustain populations by switching to biting

alternative hosts. While there is no current data on mosquito populations following these events, this may be a permissible assumption, it should be highlighted.

This is a significant finding, which likely has significant consequences for understanding YF transmission and persistence. YFV is the only Flavivirus that displays significant disease induced mortality, and compared to other Flaviviruses, YF has a lower predicted R0 value (Zhao et al., 2018, Massad et al., 2003, Towers et al., 2016). Though substantially more evidence is required to support this hypothesis, the potential trade-off between R0 and CFR may be an evolutionary compromise that allows for the long-term persistence of YFV at a lower R0 than similar Flaviviruses.

While this modelling exercise aims to simulate a more realistic scenario of YF epidemiology in South-Eastern Brazil by considering NHP movement, age structures and different NHP genera, it relies on several necessary assumptions and aggregations. Here, due to the sparsity of data on NHP life histories, I have taken the mean over NHP species within a genus in order to produce an “average” representative of the genera rather than individual species. While there are likely notable differences in life histories within genera, between genera variation is likely to be higher. Additionally, these approximations of genera have potentially offered a more generalisable interpretation of the results than would be possible if specific species had been used, and though I have focused on 4 of the 6 genera found in the South-East Atlantic coast, there are substantially more that are capable of YFV infection across South America (Bates and Roca-García, 1946b, Kumm and Laemmert, 1950, de Almeida et al., 2019). These different species and genera will have differing thresholds for the establishment of endemicity and reaction to YFV

endemic, however, the underlying dynamics are likely to remain the same. The trends seen in the relationship of CFR and R0, the ability for short lived and more rapidly reproducing NHP's to maintain YFV transmission with smaller population sizes than longer lived and more slowly reproducing NHP genera, and the long-term effects of this on the genera composition of a region are findings that are relatively agnostic to these changes.

One aspect that would prove an influential factor on the trends observed is the genera-specific suitability for infection and onward transmission. YF epizootic events, and the high mortalities associated, are often detected in populations of *Alouatta*, with several thousand found dead in the South-East Atlantic coast between 2016 and 2019 (Cunha et al., 2019). While this may be due to a genera specific suitability to infection, it likely has an additional relationship with surveillance. *Alouatta* are among the largest, and most notable NHP genera found in South America, with their characteristic calls able to be heard up to 5km away. This means that their absence, due to YFV induced mortality or otherwise, is expected to be noted by local human populations, and their size (up to 10kg in *Alouatta caraya*, compared to 256g in *Callithrix jacchus*) means detection of their remains is more likely. Additionally, immunologically, while all South American NHP's appear susceptible to YFV infection (Bates and Roca-García, 1946b, Kumm and Laemmert, 1950, de Rodaniche, 1957, Bates and Roca-Garcia, 1945), their immunological roles in maintenance and amplification are unknown. While there currently appears no robust evidence for an immunological basis on differing contributions to YFV maintenance, the habits of different NHP species may make them more likely for the acquisition of infection. *Haemagogus* is a diurnal, tree-dwelling vector, and so the feeding and residing preferences of different NHP species may lead them to interact with both the vector and other susceptible populations of NHP's

following infection at differing rates. This differential exposure based on genera may also contribute to the qualitative association of *Alouatta* with infection (Bicca-Marques and de Freitas, 2010). Our assumption of equal mixing between vectors and different genera is likely an oversimplification, however in the absence of any reliable data supporting this hypothesis, a permissible one.

Mixing behaviours are also likely related to the movement of NHP's, which I have simplified through the assumption that NHP's preferentially move towards areas with higher populations of NHP's. Populations of NHP's in an area are partially determined by resource availability, but additionally by inter-species competition, and anthropogenic pressures (Fortes et al., 2015, Paaijmans and Thomas, 2011, Rylands, 1986, Dobson and Lees, 1989). While I have considered the relationship of mortality and the carrying capacity of an environment, I have not investigated other related determinants of population dynamics such as inbreeding depression or density dependent fecundity (Dobson and Lees, 1989, Crnokrak and Roff, 1999).

In an attempt to increase the applicability of our models' findings, I have utilised population sizes, distributions and compositions modelled on the density of NHP's across the South-East Atlantic Rainforest. While these population estimates come with their own caveats, they represent the most information-based approach available with no alternative estimates of population sizes or distributions available. Furthermore, by exploring a range of different populations relative to the predicted distribution (2.5% - 100%), and simulated population sizes, I have demonstrated the robustness of these predictions.

Vector population dynamics here have been ignored for simplicity, with populations fixed at the initial size determined by the specified R0, and the larval stages omitted

entirely. While this increases the interpretability of the transmission dynamics in NHP's, it reduces the generalisability of the findings. By omitting the egg and larval stages of the vector, and assuming immediate susceptible replacement, I am artificially increasing the turnover of transmission through neglecting the delay between egg laying and development to an adult vector. Though this would primarily affect short-term trends, and likely does not substantially affect long-term persistence.

In order to account for seasonality in the force of infection, I have modified the biting rate using a sinusoidal function. While the biting rate is related to seasonal variations in climate, particularly temperature, the additional assumption of invariance in vector population sizes, mortality and EIP is erroneous. Precipitation, and the availability of suitable habitats for the larval stages of development, along with temperature interact with vector-borne transmission dynamics in a number of ways (Kumm, 1950, Garske et al., 2013). In particular, temperature's non-linear relationship with the transmission of vector-borne diseases has long been understood, with various aspects such as the biting rate (Paaijmans et al., 2013), mortality and the extrinsic incubation (EIP) related to the external temperature (Huber et al., 2018, Hamlet et al., 2019, Johansson et al., 2010). The accurate parameterisation of a model for *Haemagogus* is a significant undertaking given the sparsity of temperature variant laboratory experiments on the vector, apart from a select few from the 1930 - 1940's (Antunes and Whitman, 1937, Bates and Roca-Garcia, 1945, Bates and Roca-García, 1946a). Future developments of this work should aim to produce a more mechanistic understanding of vector dynamics in order to more accurately replicate transmission dynamics.

The focus on a single vector, *Haemagogus*, is another simplification of the sylvatic ecology of YF. While *Haemagogus* is the primary vector, additional genera such as *Sabethes*, and to a much lesser extent, *Aedes* can play a role in transmission (Abreu

et al., 2019). These vectors are likely to be differentially distributed, with variable feeding preferences, and capacities for transmission, affecting the likelihood for YFV persistence. These differential feeding preferences likely have a substantial impact, as *Haemagogus* is capable of feeding on numerous hosts (Alencar et al., 2008), several of which are dead-ends for onward transmission. However, due to the numerous uncertainties that surround this proportion, and the unquantified variation across geographical locations, the proportion of NHP blood meals as a modifier for the force of infection has been omitted from our calculations. The role of vertical transmission as a method for non-transmission based persistence of YFV has not been explored, as while *Haemagogus* is capable of vertical transmission of YFV, (Dutary and Leduc, 1981), it does so at low rates and its role in the maintenance of YFV is unlikely (Abreu et al., 2019).

In conclusion our model, and subsequent analysis, represents an important step towards the quantification of YFV transmission in NHP's. By creating a stochastic meta-population model that considers differing NHP life histories, I have shed some light on various aspects of the underlying ecology of YF, and the interaction between virus and host. This research has highlighted the strong possibility for the establishment of YFV endemicity in the state of Rio de Janeiro following repeated introducing during 2016-2019, which has long-term consequences for both human and NHP health. Future work based on this approach is needed, with more quantification of the assumptions and their consequences, in order to accurately predict the consequences of this introduction and establishment, as well as the ramifications for YFV transmission outside of this area and across all of South America.

Chapter 8 - Final discussion

8.1 Summary of findings

The overall aim of this thesis was to better quantify the environmental, climate and host dynamics that are associated with YF transmission in South America. Through a diverse range of approaches, each applied to different aspects of transmission, I have arrived at several main conclusions.

Large parts of South America are suitable for YF transmission, even outside of areas of what is currently thought of as the endemic zone, with a large part of this suitability associated with the presence of NHP species, EVI and the heterogeneity in EVI. However, this suitability is highly variable between years and months, with large differences found in geographic areas and across years and months, with differential drivers of seasonal and inter-annual transmission.

While changes in environment and climate over time account for large amounts of the variation in YF reporting, they are only one piece of the puzzle. Focusing on Brazil as a case study, I have shown that while there are strong associations of seasonal variation in these with human and NHP reports, a better prediction of these is given by looking at the seasonality of agriculture, through planting and harvesting. Models which include these instead of climate/EVI seasonality capture the distribution and magnitude of monthly reports of YF in both humans and NHPs to a much higher degree of accuracy. This increased ability of prediction highlights the importance of accounting for the seasonality of exposure as well as climate when trying to capture zoonotic spillover.

These predictions of large areas of South America that are highly suitable for YF are cause for alarm, however, there has been a long history of successful population-level vaccination coverage in South America, and recent successes in Africa. Here, our modelled estimates of population-level YFV vaccination coverage (1940-2050) have shown high coverage in large parts of the endemic zone but have additionally highlighted areas of concern. Much of Brazil's South-East Atlantic coast remains substantially unvaccinated. These findings are of an even greater anxiety if we are to believe recent findings that have found that long-term immunity in those vaccinated during infancy may be waning. While the seropositivity is not a complete determinant of protection, it has raised some alarm (Campi-Azevedo et al., 2019, Domingo et al., 2019). If so, then this would potentially reduce population level coverage substantially, as in South America those under 20 (as a rough cut-off of having received routine infant immunization) make up 35% of all vaccinated individuals.

Human population-level vaccination coverage, while important to stop human cases of YF due to the dominant role of the sylvatic transmission cycle, has no effect on the underlying transmission of YFV in South America. Instead, this is determined through NHP populations. By modelling the predicted densities and distributions of different NHP genera in the South-East Atlantic coast of Brazil, I have produced estimates of these at a 1/120 degree resolution. These estimates, while produced with the intent to be used in a mechanistic model of YFV transmission, have important conservation applications, highlighting expected areas of genera overlap abundance, in and outside of areas of current conservation.

Through constructing a stochastic meta-population model of YFV transmission, that includes population movement and different genera of NHP's, I have shown that relatively small population sizes are required to maintain endemic transmission of

YFV, particularly for certain compositions of genera in a population. When applied to realistic depictions of genera abundance and distribution I found that even if populations of NHP are 10% of what I have predicted, that the establishment of YFV as an endemic disease in the state of Rio de Janeiro is likely. A finding that is supported through Chapters 2 and 3 which had identified it as an area of high environmental and climatic suitability, and particularly worrying given its substandard vaccination coverages shown in Chapter 5.

8.2 Future work and limitations

While the work in Chapters 2 and 3 provided significant findings, there is room for improvement. The use of data at the first administrative division, while necessary due to the poor reporting of cases of YF, has restricted our ability to ascertain truer associations of covariates and reports. These administrative units can be of substantial size, and so using averaged covariates over this large geographic space may not pick up the relationship of these with YF. Future expansions of this work should seek to resolve these approaches at a higher spatial resolution in order to further explore these relationships. Additionally, by only including human cases of YF I am limiting the models ability to fully capture the underlying transmission. Though Brazil is the only country that routinely publishes this data, efforts should be made to obtain this information from other South American countries.

In Chapter 4 I have looked at the seasonality of YF in humans and NHP's in Brazil using the seasonality of agriculture. While this is an important first step, the quantification of the seasonality of agriculture has several areas where improvements could be made. The geographic resolution of the seasonality is quite large. While these planting and harvesting seasons are provided sub-nationally, even this approximates

reality. Furthermore, the binary nature of the data is again a limitation – with the intensity of agricultural activities likely not equal throughout the entire time-period of activity. The seasonality of agriculture is also not the only seasonally varying anthropogenic activity that may be associated with an increased risk, such as forestry work and eco-tourism. Future extensions of this work should seek to better quantify the intensity of agriculture and include these additional proxies for the seasonality of exposure in their analysis.

Chapter 5 focuses on the estimate of population-level vaccination coverage. While this effort represents the most complete accumulation of vaccination activities available and is utilised by the WHO in their EYE strategy, it makes several main assumptions. Firstly, the proportion of a countries population in each administrative unit has assumed to remain constant, and that no population movement has occurred. While empirically incorrect, an alternative to this assumption has not been found. Though the incorporation of population movement, and the subsequent dilution or concentration of vaccine derived YF population-level would be a method for modelling this, historical applications of this method would be difficult. These estimates have been improved in their accuracy through the provision of information by the WHO and the Brazilian Ministry of Health, which for specific periods and areas have provided the number of doses used in a municipality. In order to improve the validity and usefulness of these results, further collaborations with country level Ministries of Health should be attempted.

Chapter 6 estimated the presence and density of NHP genera in states of Sao Paulo, Rio de Janeiro, Minas Gerais and Espirito Santo. Here I have made several assumptions, and through aggregating species-specific reports to the genera level I may be confounding some of the results. While species within a genera are likely

closer in characteristics and requirements than those between genera, there are specific adaptations. Furthermore, within the same genera, different species have different population levels and conservation statuses. This was often necessary due to the sparsity of records but should be re-evaluated at a species level following additional consultation. Additionally, the input of expert information through the consultation of regional primatologists is necessary. The employment of expert knowledge has been shown to substantially improve the validity and utility of species distribution models (Calixto-Perez et al., 2018). While I have only estimated these in 4 states of Brazil, the Atlantic Rainforest extends further north to the state of Bahia and South into Argentina and Paraguay. To further improve the utility of these estimates for conservation and YFV mechanistic modelling, this work should be extended to include the entire range.

The final chapter, Chapter 7, focused on the employment of a stochastic meta-population model for YFV transmission in NHP's. Here I have made several key assumptions that should be expanded upon. The utilisation of an "average" genera occludes the roles of specific species on maintenance and may unrealistically represent the actual life-histories of the reservoir, following on from future directions of the previous chapter, efforts made to model the species rather than genera should be made. By only utilising 4 of the 6 genera of NHP's found in the South-East Atlantic coast I have not captured the full ecology of YF in its sylvatic reservoir. While this was undertaken due to computational limitations, and the relatively low populations and distributions expected of these genera, their contributions should be evaluated. In the same vein, expansion of the model to include several of the additional vectors of YF, and a more realistic modelling of their life-cycles, should be undertaken.

Implications of research

This body of work represents an important overview of the role of climate, environment and host in YF transmission across South America. By further categorising environmental and climatic determinants of YF transmission in South America, I have added to a growing understanding of the underlying ecology of YF. This has been aided by the quantification of seasonality and inter-annual transmission. By identifying covariates, and their statistical relationship with YF, the work presented here helps support offers the ability for forecasting systems to be developed, with the aim of highlighting risks of increased transmission potential for both entire years and specific months. By showing that the seasonality of agriculture is highly associated with YF in Brazil, for both human and NHP reporting, the work here may offer recommendations of increased surveillance for these periods, and increased vaccination outreach in those involved in agricultural activities to detect and prevent sylvatic spillover.

Estimates of sub-national age-disaggregated vaccination coverage highlight regions and ages at risk of YF, as well as illustrating the success of recent campaigns in raising the population-level YF vaccination coverage. By mapping NHP genera in Brazil's South-East Atlantic states and exploring the role of these in maintaining YF transmission answers some important questions regarding YF ecology. The relatively low population sizes, and the role of certain NHP genera in the maintenance of transmission supports the hypothesis that YF has the ability to become endemic in the state of Rio de Janeiro in Brazil. While this state had not reported YF in human or NHP populations before 2016, the recent epidemics and high levels of human and NHP transmission, and our findings, suggest that it is here to stay. This has significant conservation and public health ramifications and will necessitate a substantial

expansion of routine infant immunization, catch-up campaigns and a re-examination of current NHP conservation efforts, and YF public health policy in Brazil.

8.3 Conclusion

While we have long known about YFs significance as a public health threat, we still do not understand as much as we should in order to successfully combat it. This body of work answers several significant questions and provides a framework for a greater understanding of the underlying ecology of YF in South America. Through the findings described here we add to the overall knowledge of the YFV, and aid efforts to contain this ancient and deadly disease.

References

ABREU, F. V. S., RIBEIRO, I. P., FERREIRA-DE-BRITO, A., SANTOS, A., MIRANDA, R. M., BONELLY, I. S., NEVES, M., BERSOT, M. I., SANTOS, T. P. D., GOMES, M. Q., SILVA, J. L. D., ROMANO, A. P. M., CARVALHO, R. G., SAID, R., RIBEIRO, M. S., LAPERRIERE, R. D. C., FONSECA, E. O. L., FALQUETO, A., PAUPY, C., FAILLOUX, A. B., MOUTAILLER, S., CASTRO, M. G., GOMEZ, M. M., MOTTA, M. A., BONALDO, M. C. & LOURENCO-DE-OLIVEIRA, R. 2019. Haemagogus leucocelaenus and Haemagogus janthinomys are the primary vectors in the major yellow fever outbreak in Brazil, 2016-2018. *Emerg Microbes Infect,* 8**,** 218-231.

ALENCAR, J., MARCONDES, C. B., SERRA-FREIRE, N. M., LOROSA, E. S., PACHECO, J. B. & GUIMARAES, A. E. 2008. Feeding patterns of Haemagogus capricornii and Haemagogus leucocelaenus (Diptera: Culicidae) in two Brazilian states (Rio de Janeiro and Goias). *J Med Entomol,* 45**,** 873-6.

ALENCAR, J., SERRA-FRIERE, N. M., MARCONDES, C. B., SILVA, J. D., CORREA, F. F. & GUIMARAES, A. E. 2010. Influence of Climatic Factors on the Population Dynamics of Haemagogus Janthinomys (Diptera: Culicidae), a Vector of Sylvatic Yellow Fever. *Entomological News,* 121**,** 45-52.

ALLEN, T., MURRAY, K. A., ZAMBRANA-TORRELIO, C., MORSE, S. S., RONDININI, C., DI MARCO, M., BREIT, N., OLIVAL, K. J. & DASZAK, P. 2017. Global hotspots and correlates of emerging zoonotic diseases. *Nat Commun,* 8**,** 1124.

ALMEIDA, M. A., CARDOSO JDA, C., DOS SANTOS, E., DA FONSECA, D. F., CRUZ, L. L., FARACO, F. J., BERCINI, M. A., VETTORELLO, K. C., PORTO, M. A., MOHRDIECK, R., RANIERI, T. M., SCHERMANN, M. T., SPERB, A. F., PAZ, F. Z., NUNES, Z. M., ROMANO, A. P., COSTA, Z. G., GOMES, S. L. & FLANNERY, B. 2014. Surveillance for yellow Fever virus in non-human primates in southern Brazil, 2001-2011: a tool for prioritizing human populations for vaccination. *PLoS Negl Trop Dis,* 8**,** e2741.

ALTHOUSE, B. M., HANLEY, K. A., DIALLO, M., SALL, A. A., BA, Y., FAYE, O., DIALLO, D., WATTS, D. M., WEAVER, S. C. & CUMMINGS, D. A. 2015. Impact of climate and mosquito vector abundance on sylvatic arbovirus circulation dynamics in Senegal. *Am J Trop Med Hyg,* 92**,** 88-97.

ALTIZER, S., DOBSON, A., HOSSEINI, P., HUDSON, P., PASCUAL, M. & ROHANI, P. 2006. Seasonality and the dynamics of infectious diseases. *Ecol Lett,* 9**,** 467-84.

AMAKU, M., COUTINHO, F. A. & MASSAD, E. 2011. Why dengue and yellow fever coexist in some areas of the world and not in others? *Biosystems,* 106**,** 111-20.

AMARELA, C. D. O. D. E. E. S. P. S. F. 2017. Monitoramento dos casos e óbitos de Febre Amarela no Brasil - Informe 43. http://portalsaude.saude.gov.br/index.php/o-ministerio/principal/leia-mais-o-ministerio/619-secretaria-svs/l1-svs/27300-febre-amarela-informacao-e-orientacao: Ministério da Saúde Secretaria de Vigilância em Saúde.

AMRAOUI, F., VAZEILLE, M. & FAILLOUX, A. B. 2016. French Aedes albopictus are able to transmit yellow fever virus. *Euro Surveill,* 21.

ANTUNES, P. C. A. & WHITMAN, L. 1937. Studies on the Capacity of Mosquitoes of the Genus Haemagogus to Transmit Yellow Fever1. *The American Journal of Tropical Medicine and Hygiene,* s1-17**,** 825-831.

ANYAMBA, A., LINTHICUM, K. J. & TUCKER, C. J. 2001. Climate-disease connections: Rift Valley Fever in Kenya. *Cad Saude Publica,* 17 Suppl**,** 133-40.

ARTZY-RANDRUP, Y., ALONSO, D. & PASCUAL, M. 2010. Transmission intensity and drug resistance in malaria population dynamics: implications for climate change. *PLoS One,* 5**,** e13588.

ASTROM, C., ROCKLOV, J., HALES, S., BEGUIN, A., LOUIS, V. & SAUERBORN, R. 2012. Potential Distribution of Dengue Fever Under Scenarios of Climate Change and Economic Development. *Ecohealth,* 9**,** 448-454.

AUCHINCLOSS, A. H., GEBREAB, S. Y., MAIR, C. & DIEZ ROUX, A. V. 2012. A review of spatial methods in epidemiology, 2000-2010. *Annu Rev Public Health,* 33**,** 107-22.

AYLWARD, B., BARBOZA, P., BAWO, L., BERTHERAT, E., BILIVOGUI, P., BLAKE, I., BRENNAN, R., BRIAND, S., CHAKAUYA, J. M., CHITALA, K., CONTEH, R. M., CORI, A., CROISIER, A., DANGOU, J. M., DIALLO, B., DONNELLY, C. A., DYE, C., ECKMANNS, T., FERGUSON, N. M., FORMENTY, P., FUHRER, C., FUKUDA, K., GARSKE, T., GASASIRA, A., GBANYAN, S., GRAAFF, P., HELEZE, E., JAMBAI, A., JOMBART, T., KASOLO, F., KADIOBO, A. M., KEITA, S., KERTESZ, D., KONE, M., LANE, C., MARKOFF, J., MASSAQUOI, M., MILLS, H., MULBA, J. M., MUSA, E., MYHRE, J., NASIDI, A., NILLES, E., NOUVELLET, P., NSHIMIRIMANA, D., NUTTALL, I., NYENSWAH, T., OLU, O., PENDERGAST, S., PEREA, W., POLONSKY, J., RILEY, S., RONVEAUX, O., SAKOBA, K., KRISHNAN, R. S. G., SENGA, M., SHUAIB, F., VAN KERKHOVE, M. D., VAZ, R., KANNANGARAGE, N. W., YOTI, Z. & TEAM, W. E. R. 2014. Ebola Virus Disease in West Africa - The First 9 Months of the Epidemic and Forward Projections. *New England Journal of Medicine,* 371**,** 1481-1495.

BAGNARELLI, P., MARINELLI, K., TROTTA, D., MONACHETTI, A., TAVIO, M., DEL GOBBO, R., CAPOBIANCHI, M., MENZO, S., NICOLETTI, L., MAGURANO, F. & VARALDO, P. 2011. Human case of autochthonous West Nile virus lineage 2 infection in Italy, September 2011. *Euro Surveill,* 16.

BALFOUR, A. 1914. The wild monkey as a reservoir for the virus of yellow fever. *Lancet,* 1**,** 1176-1778.

BALIAMOUNE-LUTZ, M. 2011. Growth by Destination (Where You Export Matters): Trade with China and Growth in African Countries. *African Development Review-Revue Africaine De Developpement,* 23**,** 202-218.

BARBET-MASSIN, M., JIGUET, F., ALBERT, C. H. & THUILLER, W. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution,* 3**,** 327-338.

BARNETT, E. D. 2007. Yellow fever: epidemiology and prevention. *Clin Infect Dis,* 44**,** 850-6.

BARRETT, A. D. 2016. Yellow Fever in Angola and Beyond--The Problem of Vaccine Supply and Demand. *N Engl J Med,* 375**,** 301-3.

BARRETT, A. D. & HIGGS, S. 2007. Yellow fever: a disease that has yet to be conquered. *Annu Rev Entomol,* 52**,** 209-29.

BARTLETT, M. S. 1960. The Critical Community Size for Measles in the United States. *Journal of the Royal Statistical Society. Series A (General),* 123**,** 37-44.

BATES, M. & ROCA-GARCIA, M. 1945. Laboratory Studies of the Saimiri-Haemagogus Cycle of Jungle Yellow Fever1. *The American Journal of Tropical Medicine and Hygiene,* s1-25**,** 203-216.

BATES, M. & ROCA-GARCÍA, M. 1946a. The Development of the Virus of Yellow Fever in Haemagogus Mosquitoes1. *The American Journal of Tropical Medicine and Hygiene,* s1-26**,** 585-605.

BATES, M. & ROCA-GARCÍA, M. 1946b. Experiments with Various Colombian Marsupials and Primates in Laboratory Cycles of Yellow Fever*. *The American Journal of Tropical Medicine and Hygiene,* s1-26**,** 437-453.

BAUCH, S. C., BIRKENBACH, A. M., PATTANAYAK, S. K. & SILLS, E. O. 2015. Public health impacts of ecosystem change in the Brazilian Amazon. *Proc Natl Acad Sci U S A,* 112**,** 7414-9.

BEATY, B. J., TESH, R. B. & AITKEN, T. H. 1980. Transovarial transmission of yellow fever virus in Stegomyia mosquitoes. *Am J Trop Med Hyg,* 29**,** 125-32.

BEELEY, C. 2013. *Web Application Development with R Using Shiny*, PACKT publishing.

BENEDETTI, R. 2010. Scoring Rules for Forecast Verification. *Monthly Weather Review,* 138**,** 203-211.

BHATT, S., GETHING, P. W., BRADY, O. J., MESSINA, J. P., FARLOW, A. W., MOYES, C. L., DRAKE, J. M., BROWNSTEIN, J. S., HOEN, A. G., SANKOH, O., MYERS, M. F., GEORGE, D. B., JAENISCH, T., WINT, G. R., SIMMONS, C. P., SCOTT, T. W., FARRAR, J. J. & HAY, S. I. 2013. The global distribution and burden of dengue. *Nature,* 496**,** 504-7.

BIAU, G. 2012. Analysis of a Random Forests Model. *Journal of Machine Learning Research,* 13**,** 1063-1095.

BICCA-MARQUES, J. C. & DE FREITAS, D. S. 2010. The role of monkeys, mosquitoes, and humans in the occurrence of a yellow fever outbreak in a fragmented landscape in south Brazil: protecting howler monkeys is a matter of public health. *Tropical Conservation Science,* 3**,** 78-89.

BORRELL, J. S., AL ISSAEY, G., LUPTON, D. A., STARNES, T., AL HINAI, A., AL HATMI, S., SENIOR, R. A., WILKINSON, T., MILBORROW, J. L. H., STOKES-REES, A. & PATZELT, A. 2019. Islands in the desert: environmental distribution modelling of endemic flora reveals the extent of Pleistocene tropical relict vegetation in southern Arabia. *Ann Bot*.

BRAZILIAN MINISTRY OF HEALTH 2018. Epidemiological Update Yellow fever. https://portalarquivos2.saude.gov.br/images/pdf/2018/outubro/08/Informe-FA.pdf.

BRENT, S. E., WATTS, A., CETRON, M., GERMAN, M., KRAEMER, M. U. G., BOGOCH, I. I., BRADY, O. J., HAY, S. I., CREATORE, M. I. & KHAN, K. 2018. International travel between global urban centres vulnerable to yellow fever transmission. *Bulletin of the World Health Organization,* 96**,** 343-+.

BRIGHT, E. A., ROSE, A. N. & URBAN, M. L. 2015. LandScan 2014. *In:* LABORATORY, O. R. N. (ed.). http://web.ornl.gov/sci/landscan/: Oak Ridge National Laboratory.

BRIGHT, E. A., ROSE, A. N. & URBAN, M. L. 2016. LandScan 2015. *In:* LABORATORY, O. R. N. (ed.). http://web.ornl.gov/sci/landscan/: Oak Ridge National Laboratory.

BROOKS, M. E., KRISTENSEN, K., VAN BENTHEM, K. J., MAGNUSSON, A., BERG, C. W., NIELSEN, A., SKAUG, H. J., MACHLER, M. & BOLKER, B. M. 2017. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *R Journal,* 9**,** 378-400.

BRYANT, J. E., HOLMES, E. C. & BARRETT, A. D. 2007. Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas. *PLoS Pathog,* 3**,** e75.

BURKE, A. W. 1937. An Epidemic of Jungle Yellow Fever on the Planalto of Matto Grosso, Brazil1. *The American Journal of Tropical Medicine and Hygiene,* s1-17**,** 313-334.

BURKETT-CADENA, N. D. & VITTOR, A. Y. 2018. Deforestation and vector-borne disease: Forest conversion favors important mosquito vectors of human pathogens. *Basic and Applied Ecology,* 26**,** 101-110.

CAIRNS, M., ROCA-FELTRER, A., GARSKE, T., WILSON, A. L., DIALLO, D., MILLIGAN, P. J., GHANI, A. C. & GREENWOOD, B. M. 2012. Estimating the potential public health impact of seasonal malaria chemoprevention in African children. *Nat Commun,* 3**,** 881.

CAIRNS, M. E., WALKER, P. G., OKELL, L. C., GRIFFIN, J. T., GARSKE, T., ASANTE, K. P., OWUSU-AGYEI, S., DIALLO, D., DICKO, A., CISSE, B., GREENWOOD, B. M., CHANDRAMOHAN, D., GHANI, A. C. & MILLIGAN, P. J. 2015. Seasonality in malaria transmission: implications for case-management with long-acting artemisinin combination therapy in sub-Saharan Africa. *Malar J,* 14**,** 321.

CALDER, M., CRAIG, C., CULLEY, D., DE CANI, R., DONNELLY, C. A., DOUGLAS, R., EDMONDS, B., GASCOIGNE, J., GILBERT, N., HARGROVE, C., HINDS, D., LANE, D. C., MITCHELL, D., PAVEY, G., ROBERTSON, D., ROSEWELL, B., SHERWIN, S., WALPORT, M. & WILSON, A. 2018. Computational modelling for decision-making: where, why, what, who and how. *R Soc Open Sci,* 5**,** 172096.

CALIXTO-PEREZ, E., ALARCON-GUERRERO, J., RAMOS-FERNANDEZ, G., DIAS, P. A. D., RANGEL-NEGRIN, A., AMENDOLA-PIMENTA, M., DOMINGO, C., ARROYO-RODRIGUEZ, V., POZO-MONTUY, G., PINACHO-GUENDULAIN, B., URQUIZA-HAAS, T., KOLEFF, P. & MARTINEZ-MEYER, E. 2018. Integrating expert knowledge and ecological niche models to estimate Mexican primates' distribution. *Primates,* 59**,** 451-467.

CAMPI-AZEVEDO, A. C., DE ALMEIDA ESTEVAM, P., COELHO-DOS-REIS, J. G., PERUHYPE-MAGALHÃES, V., VILLELA-REZENDE, G., QUARESMA, P. F., MAIA, M. D. L. S., FARIAS, R. H. G., CAMACHO, L. A. B., FREIRE, M. D. S., GALLER, R., YAMAMURA, A. M. Y., ALMEIDA, L. F. C., LIMA, S. M. B.,

NOGUEIRA, R. M. R., SILVA SÁ, G. R., HOKAMA, D. A., DE CARVALHO, R., FREIRE, R. A. V., FILHO, E. P., LEAL, M. D. L. F., HOMMA, A., TEIXEIRA-CARVALHO, A., MARTINS, R. M. & MARTINS-FILHO, O. A. 2014. Subdoses of 17DD yellow fever vaccine elicit equivalent virological/immunological kinetics timeline. *BMC Infectious Diseases,* 14**,** 391.

CAMPI-AZEVEDO, A. C., REIS, L. R., PERUHYPE-MAGALHAES, V., COELHO-DOS-REIS, J. G., ANTONELLI, L. R., FONSECA, C. T., COSTA-PEREIRA, C., SOUZA-FAGUNDES, E. M., DA COSTA-ROCHA, I. A., MAMBRINI, J. V. D., LEMOS, J. A. C., RIBEIRO, J. G. L., CALDAS, I. R., CAMACHO, L. A. B., MAIA, M. D. D., DE NORONHA, T. G., DE LIMA, S. M. B., SIMOES, M., FREIRE, M. D., MARTINS, R. D., HOMMA, A., TAUIL, P. L., VASCONCELOS, P. F. C., ROMANO, A. P. M., DOMINGUES, C. M., TEIXEIRA-CARVALHO, A. & MARTINS, O. A. 2019. Short-Lived Immunity After 17DD Yellow Fever Single Dose Indicates That Booster Vaccination May Be Required to Guarantee Protective Immunity in Children. *Frontiers in Immunology,* 10.

CARPENTER, T. E. 2011. The spatial epidemiologic (r)evolution: a look back in time and forward to the future. *Spat Spatiotemporal Epidemiol,* 2**,** 119-24.

CARROLL, L. N., AU, A. P., DETWILER, L. T., FU, T. C., PAINTER, I. S. & ABERNETHY, N. F. 2014. Visualization and analytics tools for infectious disease epidemiology: a systematic review. *J Biomed Inform,* 51**,** 287-98.

CARTER, H. R., CARTER, L. A. & FROST, W. H. 1931. *Yellow fever; an epidemiological and historical study of its place of origin,* Baltimore,, The Williams & Wilkins company.

CATHEY, J. T. & MARR, J. S. 2014. Yellow fever, Asia and the East African slave trade. *Trans R Soc Trop Med Hyg,* 108**,** 252-7.

CATTARINO, L., RODRIGUEZ-BARRAQUER, I., IMAI, N., CUMMINGS, D. A. & FERGUSON, N. 2020. Mapping global variation in dengue transmission intensity. *Sci. Transl. Med.,* 12.

CDC. 2019. *Health Information for Travelers to Brazil* [Online]. https://wwwnc.cdc.gov/travel/destinations/traveler/none/brazil: CDC. [Accessed 07/11/19 2019].

CHADEE, D. D., GANESH, R., HINGWAN, J. O. & TIKASINGH, E. S. 1995. Seasonal Abundance, Biting Cycle and Parity of the Mosquito Haemagogus-Leucocelaenus in Trinidad, West-Indies. *Medical and Veterinary Entomology,* 9**,** 372-376.

CHAVES, T., ORDUNA, T., LEPETIC, A., MACCHI, A., VERBANAZ, S., RISQUEZ, A., PERRET, C., ECHAZARRETA, S., RODRIGUEZ-MORALES, A. J. & LLOVERAS, S. C. 2018. Yellow fever in Brazil: Epidemiological aspects and implications for travelers. *Travel Med Infect Dis,* 23**,** 1-3.

CHIPPAUX, J. P. & CHIPPAUX, A. 2018. Yellow fever in Africa and the Americas: a historical and epidemiological perspective. *J Venom Anim Toxins Incl Trop Dis,* 24**,** 20.

CHRISTOPHERS, S. R. 1960. *The Yellow Fever Mosquito,* Cambridge, University Press.

CLEMENTS, A. N. & HARBACH, R. E. 2017. History of the discovery of the mode of transmission of yellow fever virus. *J Vector Ecol,* 42**,** 208-222.

CLIMATE CHANGE INITIATIVE 2017. Land Cover 1992-2015. *In:* EUROPEAN SPACE AGENCY (ed.) 2.0 ed. http://maps.elie.ucl.ac.be/CCI/viewer/.

COMPANHIA NACIONAL DE ABASTECIMENTO (CONAB) 2017. Calendário de Plantio e Colheita de Grãos no Brasil 2017. Agricultural Observatory.

CRAIG, M. H., SNOW, R. W. & LE SUEUR, D. 1999. A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitol Today,* 15**,** 105-11.

CRNOKRAK, P. & ROFF, D. A. 1999. Inbreeding depression in the wild. *Heredity (Edinb),* 83 (Pt 3)**,** 260-70.

CULOT, L., PEREIRA, L. A., AGOSTINI, I., DE ALMEIDA, M. A. B., ALVES, R. S. C., AXIMOFF, I., BAGER, A., BALDOVINO, M. C., BELLA, T. R., BICCA-MARQUES, J. C., BRAGA, C., BROCARDO, C. R., CAMPELO, A. K. N., CANALE, G. R., CARDOSO, J. D. C., CARRANO, E., CASANOVA, D. C., CASSANO, C. R., CASTRO, E., CHEREM, J. J., CHIARELLO, A. G., COSENZA, B. A. P., COSTA-ARAUJO, R., SILVA, N. C. D., DI BITETTI, M. S., FERREIRA, A. S., FERREIRA, P. C. R., FIALHO, M. S., FUZESSY, L. F., GARBINO, G. S. T., GARCIA, F. O., GATTO, C., GESTICH, C. C., GONCALVES,

P. R., GONTIJO, N. R. C., GRAIPEL, M. E., GUIDORIZZI, C. E., ESPINDOLA HACK, R. O., HASS, G. P., HILARIO, R. R., HIRSCH, A., HOLZMANN, I., HOMEM, D. H., JUNIOR, H. E., JUNIOR, G. S., KIERULFF, M. C. M., KNOGGE, C., LIMA, F., DE LIMA, E. F., MARTINS, C. S., DE LIMA, A. A., MARTINS, A., MARTINS, W. P., DE MELO, F. R., MELZEW, R., MIRANDA, J. M. D., MIRANDA, F., MORAES, A. M., MOREIRA, T. C., DE CASTRO MORINI, M. S., NAGY-REIS, M. B., OKLANDER, L., DE CARVALHO OLIVEIRA, L., PAGLIA, A. P., PAGOTO, A., PASSAMANI, M., DE CAMARGO PASSOS, F., PERES, C. A., DE CAMPOS PERINE, M. S., PINTO, M. P., PONTES, A. R. M., PORT-CARVALHO, M., PRADO, B., REGOLIN, A. L., REZENDE, G. C., ROCHA, A., ROCHA, J. D. S., DE PAULA RODARTE, R. R., SALES, L. P., SANTOS, E. D., SANTOS, P. M., BERNARDO, C. S. S., SARTORELLO, R., SERRA, L., SETZ, E., DE ALMEIDA, E. S. A. S., SILVA, L. H. D., SILVA, P., SILVEIRA, M., SMITH, R. L., DE SOUZA, S. M., SRBEK-ARAUJO, A. C., TREVELIN, L. C., VALLADARES-PADUA, C., ZAGO, L., MARQUES, E., FERRARI, S. F., BELTRAO-MENDES, R., HENZ, D. J., DA VEIGA DA COSTA, F. E., et al. 2019. ATLANTIC-PRIMATES: a dataset of communities and occurrences of primates in the Atlantic Forests of South America. *Ecology,* 100**,** e02525.

CUNHA, M. S., DA COSTA, A. C., FERNANDES, N. C. C. D., GUERRA, J. M., DOS SANTOS, F. C. P., NOGUEIRA, J. S., D'AGOSTINO, L. G., KOMNINAKIS, S. V., WITKIN, S. S., RESSIO, R. A., MAEDA, A. Y., VASAMI, F. G. S., KAIGAWA, U. M. A., DE AZEVEDO, L. S., FACIOLI, P. A. D., MACEDO, F. L. L., SABINO, E. C., LEAL, E. & DE SOUZA, R. P. 2019. Epizootics due to Yellow Fever Virus in Sao Paulo State, Brazil: viral dissemination to new areas (2016-2017). *Scientific Reports,* 9.

CURTO DE CASAS, S. I. & CARCAVALLO, R. U. 1995. Climate change and vector-borne diseases distribution. *Soc Sci Med,* 40**,** 1437-40.

DAVIS, N. C. 1930. The Transmission of Yellow Fever : Experiments with the "Woolly Monkey" (Lagothrix Lago-Tricha Humboldt), the "Spider Monkey" (Ateleus Ater F. Cuvier), and the "Squirrel Monkey" (Saimiri Scireus Linnaeus). *J Exp Med,* 51**,** 703-20.

DE ALMEIDA, M. A. B., DOS SANTOS, E., CARDOSO, J. D., DA SILVA, L. G., RABELO, R. M. & BICCA-MARQUES, J. C. 2019. Predicting Yellow Fever Through Species Distribution Modeling of Virus, Vector, and Monkeys. *Ecohealth,* 16**,** 95-108.

DE RODANICHE, E. 1957. Survey of Primates Captured in Panamá, R. P. During the Years 1952–1956 for Protective Antibodies Against Yellow Fever. *The American Journal of Tropical Medicine and Hygiene,* 6**,** 835-839.

DEFRIES, R., HEROLD, M., VERCHOT, L., MACEDO, M. N. & SHIMABUKURO, Y. 2013. Export-oriented deforestation in Mato Grosso: harbinger or exception for other tropical forests? *Philos Trans R Soc Lond B Biol Sci,* 368**,** 20120173.

DEGALLIER, N., DAROSA, A. P. A. T., VASCONCELOS, P. F. C., GUERREIRO, S. C., DAROSA, J. F. S. T. & HERVE, J. P. 1991. Study of the Survival Rate, Relative Density, and Infection-Rate of a Population of the Sylvatic Yellow-Fever Vector Haemagogus-Janthinomys Dyar (Diptera, Culicidae), from Which the Yf Virus Was Isolated in Brazilian Amazonia. *Bulletin De La Societe De Pathologie Exotique,* 84**,** 386-397.

DEGALLIER, N., SA FILHO, G. C., MONTEIRO, H. A., CASTRO, F. C., DA SILVA, O. V., BRANDAO, R. C., MOYSES, M. & DA ROSA, A. P. 1998. Release-recapture experiments with canopy mosquitoes in the genera Haemagogus and Sabethes (Diptera: Culicidae) in Brazilian Amazonia. *J Med Entomol,* 35**,** 931-6.

DO, T. T., MARTENS, P., LUU, N. H., WRIGHT, P. & CHOISY, M. 2014. Climatic-driven seasonality of emerging dengue fever in Hanoi, Vietnam. *BMC Public Health,* 14**,** 1078.

DOBSON, A. P. & LEES, A. M. 1989. The population dynamics and conservation of primate populations. *Conserv Biol,* 3**,** 362-80.

DOMINGO, C., FRAISSINET, J., ANSAH, P. O., KELLY, C., BHAT, N., SOW, S. O. & MEJÍA, J. E. 2019. Long-term immunity against yellow fever in children vaccinated during infancy: a longitudinal cohort study. *The Lancet Infectious Diseases*.

DORIGATTI, I., HAMLET, A., AGUAS, R., CATTARINO, L., CORI, A., DONNELLY, C. A., GARSKE, T., IMAI, N. & FERGUSON, N. M. 2017. International risk of yellow fever spread from the ongoing outbreak in Brazil, December 2016 to May 2017. *Euro Surveill,* 22.

DOS SANTOS, S., ADAMS, E. A., NEVILLE, G., WADA, Y., DE SHERBININ, A., MULLIN BERNHARDT, E. & ADAMO, S. B. 2017. Urban growth and water access in sub-Saharan Africa: Progress, challenges, and emerging research directions. *Sci Total Environ,* 607-608**,** 497-508.

DURIEUX, C. 1956. Mass yellow fever vaccination in French Africa south of the Sahara. *Monograph Series World Health Organisation.* Geneva: World Health Organisation.

DUTARY, B. E. & LEDUC, J. W. 1981. Transovarial transmission of yellow fever virus by a sylvatic vector, Haemagogus equinus. *Trans R Soc Trop Med Hyg,* 75**,** 128.

ELITH, J., LEATHWICK, J. R. & HASTIE, T. 2008. A working guide to boosted regression trees. *J Anim Ecol,* 77**,** 802-13.

ELLIS, B. R. & BARRETT, A. D. 2008. The enigma of yellow fever in East Africa. *Rev Med Virol,* 18**,** 331-46.

ESCOBAR, L. E. & CRAFT, M. E. 2016. Advances and Limitations of Disease Biogeography Using Ecological Niche Modeling. *Front Microbiol,* 7**,** 1174.

ESTRADA-PENA, A., ESTRADA-SANCHEZ, A. & ESTRADA-SANCHEZ, D. 2015. Methodological caveats in the environmental modelling and projections of climate niche for ticks, with examples for Ixodes ricinus (Ixodidae). *Vet Parasitol,* 208**,** 14-25.

ESTRADA, A., GARBER, P. A., MITTERMEIER, R. A., WICH, S., GOUVEIA, S., DOBROVOLSKI, R., NEKARIS, K. A. I., NIJMAN, V., RYLANDS, A. B., MAISELS, F., WILLIAMSON, E. A., BICCA-MARQUES, J., FUENTES, A., JERUSALINSKY, L., JOHNSON, S., RODRIGUES DE MELO, F., OLIVEIRA, L., SCHWITZER, C., ROOS, C., CHEYNE, S. M., MARTINS KIERULFF, M. C., RAHARIVOLOLONA, B., TALEBI, M., RATSIMBAZAFY, J., SUPRIATNA, J., BOONRATANA, R., WEDANA, M. & SETIAWAN, A. 2018. Primates in peril: the significance of Brazil, Madagascar, Indonesia and the Democratic Republic of the Congo for global primate conservation. *PeerJ,* 6**,** e4869.

ESTRADA, A., GARBER, P. A., RYLANDS, A. B., ROOS, C., FERNANDEZ-DUQUE, E., DI FIORE, A., NEKARIS, K. A.-I., NIJMAN, V., HEYMANN, E. W., LAMBERT, J. E., ROVERO, F., BARELLI, C., SETCHELL, J. M., GILLESPIE, T. R., MITTERMEIER, R. A., ARREGOITIA, L. V., DE GUINEA, M., GOUVEIA, S., DOBROVOLSKI, R., SHANEE, S., SHANEE, N., BOYLE, S. A., FUENTES, A., MACKINNON, K. C., AMATO, K. R., MEYER, A. L. S., WICH, S., SUSSMAN, R. W., PAN, R., KONE, I. & LI, B. 2017. Impending extinction crisis of the world's primates: Why primates matter. *Science Advances,* 3**,** e1600946.

FARIA, N. R., KRAEMER, M. U. G., HILL, S. C., GOES DE JESUS, J., AGUIAR, R. S., IANI, F. C. M., XAVIER, J., QUICK, J., DU PLESSIS, L., DELLICOUR, S., THEZE, J., CARVALHO, R. D. O., BAELE, G., WU, C. H., SILVEIRA, P. P., ARRUDA, M. B., PEREIRA, M. A., PEREIRA, G. C., LOURENCO, J., OBOLSKI, U., ABADE, L., VASYLYEVA, T. I., GIOVANETTI, M., YI, D., WEISS, D. J., WINT, G. R. W., SHEARER, F. M., FUNK, S., NIKOLAY, B., FONSECA, V., ADELINO, T. E. R., OLIVEIRA, M. A. A., SILVA, M. V. F., SACCHETTO, L., FIGUEIREDO, P. O., REZENDE, I. M., MELLO, E. M., SAID, R. F. C., SANTOS, D. A., FERRAZ, M. L., BRITO, M. G., SANTANA, L. F., MENEZES, M. T., BRINDEIRO, R. M., TANURI, A., DOS SANTOS, F. C. P., CUNHA, M. S., NOGUEIRA, J. S., ROCCO, I. M., DA COSTA, A. C., KOMNINAKIS, S. C. V., AZEVEDO, V., CHIEPPE, A. O., ARAUJO, E. S. M., MENDONCA, M. C. L., DOS SANTOS, C. C., DOS SANTOS, C. D., MARES-GUIA, A. M., NOGUEIRA, R. M. R., SEQUEIRA, P. C., ABREU, R. G., GARCIA, M. H. O., ABREU, A. L., OKUMOTO, O., KROON, E. G., DE ALBUQUERQUE, C. F. C., LEWANDOWSKI, K., PULLAN, S. T., CARROLL, M., DE OLIVEIRA, T., SABINO, E. C., SOUZA, R. P., SUCHARD, M. A., LEMEY, P., TRINDADE, G. S., DRUMOND, B. P., FILIPPIS, A. M. B., LOMAN, N. J., CAUCHEMEZ, S., ALCANTARA, L. C. J. & PYBUS, O. G. 2018. Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science,* 361**,** 894-899.

FAUST, C. L., MCCALLUM, H. I., BLOOMFIELD, L. S. P., GOTTDENKER, N. L., GILLESPIE, T. R., TORNEY, C. J., DOBSON, A. P. & PLOWRIGHT, R. K. 2018. Pathogen spillover during land conversion. *Ecol Lett,* 21**,** 471-483.

FICK, S. E. & HIJMANS, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology,* 37**,** 4302-4315.

FIELDING, A. H. & BELL, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation,* 24**,** 38-49.

FISMAN, D. N. 2007. Seasonality of infectious diseases. *Annu Rev Public Health,* 28**,** 127-43.

FITZJOHN, R. 2019. odin: ODE Generation and Integration. https://github.com/mrc-ide/odin: R package version 1.0.2.

FORTES, V. B., BICCA-MARQUES, J. C., URBANI, B., FERNÁNDEZ, V. A. & DA SILVA PEREIRA, T. 2015. Ranging Behavior and Spatial Cognition of Howler Monkeys. *In:* KOWALEWSKI, M. M., GARBER, P. A., CORTÉS-ORTIZ, L., URBANI, B. & YOULATOS, D. (eds.) *Howler Monkeys: Behavior, Ecology, and Conservation.* New York, NY: Springer New York.

FRIEDL, M., SULLA-MENASHE, D., TAN, B., SCHNEIDER, A., RAMANKUTTY, N., SIBLEY, A. & HUANG, X. 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment,* 114**,** 168-182.

FRIERSON, J. G. 2010. The yellow fever vaccine: a history. *Yale J Biol Med,* 83**,** 77-85.

FULLER, D. O., TROYO, A. & BEIER, J. C. 2009. El Nino Southern Oscillation and vegetation dynamics as predictors of dengue fever cases in Costa Rica. *Environmental Research Letters,* 4.

GARDNER, C. L. & RYMAN, K. D. 2010. Yellow fever: a reemerging threat. *Clin Lab Med,* 30**,** 237-60.

GARDNER, T. A., BARLOW, J., ARAUJO, I. S., AVILA-PIRES, T. C., BONALDO, A. B., COSTA, J. E., ESPOSITO, M. C., FERREIRA, L. V., HAWES, J., HERNANDEZ, M. I., HOOGMOED, M. S., LEITE, R. N., LO-MAN-HUNG, N. F., MALCOLM, J. R., MARTINS, M. B., MESTRE, L. A., MIRANDA-SANTOS, R., OVERAL, W. L., PARRY, L., PETERS, S. L., RIBEIRO-JUNIOR, M. A., DA SILVA, M. N., DA SILVA MOTTA, C. & PERES, C. A. 2008. The cost-effectiveness of biodiversity surveys in tropical forests. *Ecol Lett,* 11**,** 139-50.

GARSKE, T., FERGUSON, N. M. & GHANI, A. C. 2013. Estimating air temperature and its influence on malaria transmission across Africa. *PLoS One,* 8**,** e56487.

GARSKE, T., VAN KERKHOVE, M. D., YACTAYO, S., RONVEAUX, O., LEWIS, R. F., STAPLES, J. E., PEREA, W., FERGUSON, N. M. & YELLOW FEVER EXPERT, C. 2014. Yellow Fever in Africa: estimating the burden of disease and impact of mass vaccination from outbreak and serological data. *PLoS Med,* 11**,** e1001638.

GELMAN, A. & HILL, J. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models,* Cambridg, Cambridge University Press.

GENUER, R., POGGI, J.-M. & TULEAU, C. 2008. Random Forests: some methodological insights. 6729.

GIBB, R., MOSES, L. M., REDDING, D. W. & JONES, K. E. 2017. Understanding the cryptic nature of Lassa fever in West Africa. *Pathog Glob Health,* 111**,** 276-288.

GLEISER, R. M. & ZALAZAR, L. P. 2010. Distribution of mosquitoes in relation to urban landscape characteristics. *Bull Entomol Res,* 100**,** 153-8.

GOLDBERG, T. L., GILLESPIE, T. R., RWEGO, I. B., ESTOFF, E. L. & CHAPMAN, C. A. 2008. In Forest Fragmentation as Cause of Bacterial Transmission among Primates, Humans, and Livestock, Uganda (vol 14, pg 1375, 2008). *Emerging Infectious Diseases,* 14**,** 1825-1825.

GOTTDENKER, N. L., CHAVES, L. F., CALZADA, J. E., SALDANA, A. & CARROLL, C. R. 2012. Host Life History Strategy, Species Diversity, and Habitat Influence Trypanosoma cruzi Vector Infection in Changing Landscapes. *Plos Neglected Tropical Diseases,* 6.

GREENWELL, B., BOEHMKE, B. & CUNNINGHAM, J. 2019. gbm: Generalized Boosted Regression Models. https://CRAN.R-project.org/package=gbm.

GUBLER, D. J. 2004. The changing epidemiology of yellow fever and dengue, 1900 to 2003: full circle? *Comp Immunol Microbiol Infect Dis,* 27**,** 319-30.

HAMLET, A., JEAN, K., PEREA, W., YACTAYO, S., BIEY, J., VAN KERKHOVE, M., FERGUSON, N. & GARSKE, T. 2018. The seasonal influence of climate and environment on yellow fever transmission across Africa. *PLoS Negl Trop Dis,* 12**,** e0006284.

HAMLET, A., JEAN, K., YACTAYO, S., BENZLER, J., CIBRELUS, L., FERGUSON, N. & GARSKE, T. 2019. POLICI: A web application for visualising and extracting yellow fever vaccination coverage in Africa. *Vaccine,* 37**,** 1384-1388.

HAMRICK, P. N., ALDIGHIERI, S., MACHADO, G., LEONEL, D. G., VILCA, L. M., URIONA, S. & SCHNEIDER, M. C. 2017. Geographic patterns and environmental factors associated with human yellow fever presence in the Americas. *PLoS Negl Trop Dis,* 11**,** e0005897.

HARTEMINK, N., CIANCI, D. & REITER, P. 2015. R0 for vector-borne diseases: impact of the assumption for the duration of the extrinsic incubation period. *Vector Borne Zoonotic Dis,* 15**,** 215-7.

HAY, S. I., GUERRA, C. A., GETHING, P. W., PATIL, A. P., TATEM, A. J., NOOR, A. M., KABARIA, C. W., MANH, B. H., ELYAZAR, I. R., BROOKER, S., SMITH, D. L., MOYEED, R. A. & SNOW, R. W. 2009. A world malaria map: Plasmodium falciparum endemicity in 2007. *PLoS Med,* 6**,** e1000048.

HEESTERBEEK, H., ANDERSON, R. M., ANDREASEN, V., BANSAL, S., DE ANGELIS, D., DYE, C., EAMES, K. T., EDMUNDS, W. J., FROST, S. D., FUNK, S., HOLLINGSWORTH, T. D., HOUSE, T., ISHAM, V., KLEPAC, P., LESSLER, J., LLOYD-SMITH, J. O., METCALF, C. J., MOLLISON, D., PELLIS, L., PULLIAM, J. R., ROBERTS, M. G., VIBOUD, C. & ISAAC NEWTON INSTITUTE, I. D. D. C. 2015. Modeling infectious disease dynamics in the complex landscape of global health. *Science,* 347**,** aaa4339.

HILBE, J. M. 2007. *Negative Binomial Regression*, Cambridge University Press.

HOLZMANN, I., AGOSTINI, I., ARETA, J. I., FERREYRA, H., BELDOMENICO, P. & DI BITETTI, M. S. 2010. Impact of Yellow Fever Outbreaks on Two Howler Monkey Species (Alouatta guariba clamitans and A. caraya) in Misiones, Argentina. *American Journal of Primatology,* 72**,** 475-480.

HOSHEN, M. B. & MORSE, A. P. 2004. A weather-driven model of malaria transmission. *Malar J,* 3**,** 32.

HU, W., TONG, S., MENGERSEN, K. & OLDENBURG, B. 2005. Rainfall, mosquito density and the transmission of Ross River virus: A time-series forecasting model. *Epidemiology,* 16.

HUBER, J. H., CHILDS, M. L., CALDWELL, J. M. & MORDECAI, E. A. 2018. Seasonal temperature variation influences climate suitability for dengue, chikungunya, and Zika transmission. *PLoS Negl Trop Dis,* 12**,** e0006451.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). 2017. *Censo Agro* [Online]. https://censoagro2017.ibge.gov.br/templates/censo_agro/resultadosagro/agricultura.html IBGE. [Accessed 2019].

IUCN. 2017. *The ICUN Red List of Threatened Species. Version 2017-1.* [Online]. http://www.iucnredlist.org. [Accessed 10/08/2017].

JEAN, K., DONNELLY, C., FERGUSON, N. & GARSKE, T. 2016. A Meta-Analysis of Serological Response Associated with Yellow Fever Vaccination. *American Journal of Tropical Medicine and Hygiene,* 16.

JENTES, E. S., POUMEROL, G., GERSHMAN, M. D., HILL, D. R., LEMARCHAND, J., LEWIS, R. F., STAPLES, J. E., TOMORI, O., WILDER-SMITH, A., MONATH, T. P. & INFORMAL, W. H. O. W. G. O. G. R. F. Y. F. 2011. The revised global yellow fever risk map and recommendations for vaccination, 2010: consensus of the Informal WHO Working Group on Geographic Risk for Yellow Fever. *Lancet Infect Dis,* 11**,** 622-32.

JOHANSSON, M. A., ARANA-VIZCARRONDO, N., BIGGERSTAFF, B. J., GALLAGHER, N., MARANO, N. & STAPLES, J. E. 2012. Assessing the risk of international spread of yellow fever virus: a mathematical analysis of an urban outbreak in Asuncion, 2008. *Am J Trop Med Hyg,* 86**,** 349-58.

JOHANSSON, M. A., ARANA-VIZCARRONDO, N., BIGGERSTAFF, B. J. & STAPLES, J. E. 2010. Incubation periods of Yellow fever virus. *Am J Trop Med Hyg,* 83**,** 183-8.

JOHANSSON, M. A., CUMMINGS, D. A. T. & GLASS, G. E. 2009. Multiyear Climate Variability and Dengue-El Nino Southern Oscillation, Weather, and Dengue Incidence in Puerto Rico, Mexico, and Thailand: A Longitudinal Data Analysis. *Plos Medicine,* 6.

JOHANSSON, M. A., VASCONCELOS, P. F. & STAPLES, J. E. 2014. The whole iceberg: estimating the incidence of yellow fever virus infection from the number of severe cases. *Trans R Soc Trop Med Hyg,* 108**,** 482-7.

JOHNSON, M. F., GOMEZ, A. & PINEDO-VASQUEZ, M. 2008. Land use and mosquito diversity in the Peruvian Amazon. *J Med Entomol,* 45**,** 1023-30.

JOKAR ARSANJANI, J. 2018. Characterizing, monitoring, and simulating land cover dynamics using GlobeLand30: A case study from 2000 to 2030. *J Environ Manage,* 214**,** 66-75.

JONES, K. E., BIELBY, J., CARDILLO, M., FRITZ, S. A., O'DELL, J., ORME, C. D. L., SAFI, K., SECHREST, W., BOAKES, E. H., CARBONE, C., CONNOLLY, C., CUTTS, M. J., FOSTER, J. K., GRENYER, R., HABIB, M., PLASTER, C. A., PRICE, S. A., RIGBY, E. A., RIST, J., TEACHER, A., BININDA-EMONDS, O. R. P., GITTLEMAN, J. L., MACE, G. M. & PURVIS, A. 2009. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology,* 90**,** 2648-2648.

JOYCE, R., JANOWIAK, J., ARKIN, P. & XIE, P. 2004. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydromet,* 5**,** 487-503.

KALTER, S. S. & JEFFRIES-KLITCH, H. 1969. Yellow fever vaccination of primates. *Am J Trop Med Hyg,* 18**,** 466-9.

KAMILAR, J. M. 2009. Environmental and geographic correlates of the taxonomic structure of primate communities. *Am J Phys Anthropol,* 139**,** 382-93.

KAY, R. F., MADDEN, R. H., VANSCHAIK, C. & HIGDON, D. 1997. Primate species richness is determined by plant productivity: Implications for conservation. *Proceedings of the National Academy of Sciences of the United States of America,* 94**,** 13023-13027.

KINGSLEY, D. H. 2016. Emerging Foodborne and Agriculture-Related Viruses. *Microbiol Spectr,* 4.

KOCH, T. & DENIKE, K. 2009. Crediting his critics' concerns: remaking John Snow's map of Broad Street cholera, 1854. *Soc Sci Med,* 69**,** 1246-51.

KRAEMER, M. U. G., FARIA, N. R., REINER, R. C., JR., GOLDING, N., NIKOLAY, B., STASSE, S., JOHANSSON, M. A., SALJE, H., FAYE, O., WINT, G. R. W., NIEDRIG, M., SHEARER, F. M., HILL, S. C., THOMPSON, R. N., BISANZIO, D., TAVEIRA, N., NAX, H. H., PRADELSKI, B. S. R., NSOESIE, E. O., MURPHY, N. R., BOGOCH, II, KHAN, K., BROWNSTEIN, J. S., TATEM, A. J., DE OLIVEIRA, T., SMITH, D. L., SALL, A. A., PYBUS, O. G., HAY, S. I. & CAUCHEMEZ, S. 2017. Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015-16: a modelling study. *Lancet Infect Dis,* 17**,** 330-338.

KRAEMER, M. U. G., GOLDING, N., BRADY, O. J., MESSINA, J. P., SMITH, D. L., WINT, G. R. W. & HAY, S. I. 2015. The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *Tropical Medicine & International Health,* 20**,** 38-38.

KRAMER, L. D. & EBEL, G. D. 2003. Dynamics of flavivirus infection in mosquitoes. *Adv Virus Res,* 60**,** 187-232.

KUMM, H. 1950. Seasonal variations in rainfall: Prevalence of Haemagogus and incidence of jungle yellow fever in Brazil and Colombia. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 43**,** 673-682.

KUMM, H. W. & LAEMMERT, H. W., JR. 1950. The geographical distribution of immunity to yellow fever among the primates of Brazil. *Am J Trop Med Hyg,* 30**,** 733-48.

LAMBRECHTS, L., PAAIJMANS, K. P., FANSIRI, T., CARRINGTON, L. B., KRAMER, L. D., THOMAS, M. B. & SCOTT, T. W. 2011. Impact of daily temperature fluctuations on dengue virus transmission by Aedes aegypti. *Proc Natl Acad Sci U S A,* 108**,** 7460-5.

LEPINIEC, L., DALGARNO, L., HUONG, V. T., MONATH, T. P., DIGOUTTE, J. P. & DEUBEL, V. 1994. Geographic distribution and evolution of yellow fever viruses based on direct sequencing of genomic cDNA fragments. *J Gen Virol,* 75 (Pt 2)**,** 417-23.

LI, M. 2010. *An Introduction to Mathematical Modeling of Infectious Diseases*, Springer, Cham.

LI, R., DING, J. & LI, H. 2016. Six cases of imported yellow fever in China: March 12, 2016–March 24, 2016. *Radiology of Infectious Diseases,* 3**,** 143-144.

LLOYD-SMITH, J. O., CROSS, P. C., BRIGGS, C. J., DAUGHERTY, M., GETZ, W. M., LATTO, J., SANCHEZ, M. S., SMITH, A. B. & SWEI, A. 2005. Should we expect population thresholds for wildlife disease? *Trends in Ecology & Evolution,* 20**,** 511-519.

LYON, B., DINKU, T., RAMAN, A. & THOMSON, M. C. 2017. Temperature suitability for malaria climbing the Ethiopian Highlands. *Environmental Research Letters,* 12.

MAKHANI, L., KHATIB, A., CORBEIL, A., KARIYAWASAM, R., RAHEEL, H., CLARKE, S., CHALLA, P., HAGOPIAN, E., CHAKRABARTI, S., SCHWARTZ, K. L. & BOGGILD, A. K. 2019. 2018 in review: five hot topics in tropical medicine. *Trop Dis Travel Med Vaccines,* 5**,** 5.

MARMION, M., PARVIAINEN, M., LUOTO, M., HEIKKINEN, R. K. & THUILLER, W. 2009. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions,* 15**,** 59-69.

MARTINS, R. M., MAIA MDE, L., FARIAS, R. H., CAMACHO, L. A., FREIRE, M. S., GALLER, R., YAMAMURA, A. M., ALMEIDA, L. F., LIMA, S. M., NOGUEIRA, R. M., SA, G. R., HOKAMA, D. A., DE CARVALHO, R., FREIRE, R. A., PEREIRA FILHO, E., LEAL MDA, L. & HOMMA, A. 2013. 17DD yellow fever vaccine: a double blind, randomized clinical trial of immunogenicity and safety on a dose-response study. *Hum Vaccin Immunother,* 9**,** 879-88.

MASSAD, E., AMAKU, M., COUTINHO, F. A. B., STRUCHINER, C. J., LOPEZ, L. F., COELHO, G., WILDER-SMITH, A. & BURATTINI, M. N. 2018a. The risk of urban yellow fever resurgence in Aedes-infested American cities. *Epidemiol Infect***,** 1-7.

MASSAD, E., BURATTINI, M. N., COUTINHO, F. A. & LOPEZ, L. F. 2003. Dengue and the risk of urban yellow fever reintroduction in Sao Paulo State, Brazil. *Rev Saude Publica,* 37**,** 477-84.

MASSAD, E., MIGUEL, M. M. & COUTINHO, F. A. B. 2018b. Is vaccinating monkeys against yellow fever the ultimate solution for the Brazilian recurrent epizootics? *Epidemiology and Infection,* 146**,** 1622-1624.

MBORA, D. N. & MCPEEK, M. A. 2009. Host density and human activities mediate increased parasite prevalence and richness in primates threatened by habitat loss and fragmentation. *J Anim Ecol,* 78**,** 210-8.

MICHALSKI, F. & PERES, C. A. 2005. Anthropogenic determinants of primate and carnivore local extinctions in a fragmented forest landscape of southern Amazonia. *Biological Conservation,* 124**,** 383-396.

MILLER, R. H., MASUOKA, P., KLEIN, T. A., KIM, H. C., SOMER, T. & GRIECO, J. 2012. Ecological niche modeling to estimate the distribution of Japanese encephalitis virus in Asia. *PLoS Negl Trop Dis,* 6**,** e1678.

MOHAMMED, A. & CHADEE, D. D. 2011. Effects of different temperature regimens on the development of Aedes aegypti (L.) (Diptera: Culicidae) mosquitoes. *Acta Trop,* 119**,** 38-43.

MOLNAR, C. 2019. *Interpretable Machine Learning*, Christoph Molnar.

MONATH, T. P. 1996. Yellow fever vaccines: The success of empiricism, pitfalls of application, and transition to molecular vaccinology. *Vaccinia, Vaccination, Vaccinology***,** 157-182.

MONATH, T. P. 2001. Yellow fever: an update. *Lancet Infect Dis,* 1**,** 11-20.

MONATH, T. P. 2006. Yellow fever as an endemic/epidemic disease and priorities for vaccination. *Bull Soc Pathol Exot,* 99**,** 341-7.

MONATH, T. P. & VASCONCELOS, P. F. 2015. Yellow fever. *J Clin Virol,* 64**,** 160-73.

MONATH, T. P., WOODALL, J. P., GUBLER, D. J., YUILL, T. M., MACKENZIE, J. S., MARTINS, R. M., REITER, P. & HEYMANN, D. L. 2016. Yellow fever vaccine supply: a possible solution. *Lancet,* 387**,** 1599-600.

MORENO, E. S., AGOSTINI, I., HOLZMANN, I., DI BITETTI, M. S., OKLANDER, L. I., KOWALEWSKI, M. M., BELDOMENICO, P. M., GOENAGA, S., MARTINEZ, M., LESTANI, E., DESBIEZ, A. L. & MILLER, P. 2015. Yellow fever impact on brown howler monkeys (Alouatta guariba clamitans) in Argentina: a metamodelling approach based on population viability analysis and epidemiological dynamics. *Mem Inst Oswaldo Cruz,* 110**,** 865-76.

MORENO, E. S., SPINOLA, R., TENGAN, C. H., BRASIL, R. A., SICILIANO, M. M., COIMBRA, T. L., SILVEIRA, V. R., ROCCO, I. M., BISORDI, I., SOUZA, R. P., PETRELLA, S., PEREIRA, L. E., MAEDA, A. Y., SILVA, F. G. & SUZUKI, A. 2013. Yellow fever epizootics in non-human primates, Sao Paulo state, Brazil, 2008-2009. *Rev Inst Med Trop Sao Paulo,* 55**,** 45-50.

MORIN, C. W., MONAGHAN, A. J., HAYDEN, M. H., BARRERA, R. & ERNST, K. 2015. Meteorologically Driven Simulations of Dengue Epidemics in San Juan, PR. *PLoS Negl Trop Dis,* 9**,** e0004002.

MUTEBI, J. P. & BARRETT, A. D. 2002. The epidemiology of yellow fever in Africa. *Microbes Infect,* 4**,** 1459-68.

MUTURI, E. J. 2013. Larval rearing temperature influences the effect of malathion on Aedes aegypti life history traits and immune responses. *Chemosphere,* 92**,** 1111-6.

MYERS, N., MITTERMEIER, R. A., MITTERMEIER, C. G., DA FONSECA, G. A. & KENT, J. 2000. Biodiversity hotspots for conservation priorities. *Nature,* 403**,** 853-8.

NASA. *Land Processes Distributed Active Archieve Centre (LP DAAC) Vegetation Indices 16-Day L3 Global 1 km (13 A2)* [Online]. Sioux Falls, South Dakota: USGS/Earth Resources Observation and Science (EROS) Center. Available: http://lpdaac.usgs.gov/get_data [Accessed 13 July 2012].

NIGERIA CENTRE FOR DISEASE CONTROL (NCDC) 2018. Yellow fever outbreak in Nigeria. http://www.ncdc.gov.ng/diseases/sitreps/?cat=10&name=An%20update%20of%20Yellow%20Fever%20outbreak%20in%20Nigeria: NCDC.

NZOLO, D., ENGO BIONGO, A., KUEMMERLE, A., LUSAKIBANZA, M., LULA, Y., NSENGI, N., NSIBU NDOSIMAO, C., TONA LUTETE, G. & VAN GEERTRUYDEN, J. P. 2018. Safety profile of fractional dosing of the 17DD Yellow Fever Vaccine among males and females: Experience of a community-based pharmacovigilance in Kinshasa, DR Congo. *Vaccine,* 36**,** 6170-6182.

OLDEN, J. D., LAWLER, J. J. & POFF, N. L. 2008. Machine learning methods without tears: A primer for ecologists. *Quarterly Review of Biology,* 83**,** 171-193.

OLIVAL, K. J., HOSSEINI, P. R., ZAMBRANA-TORRELIO, C., ROSS, N., BOGICH, T. L. & DASZAK, P. 2017. Host and viral traits predict zoonotic spillover from mammals. *Nature,* 546**,** 646-650.

PAAIJMANS, K. P., CATOR, L. J. & THOMAS, M. B. 2013. Temperature-dependent pre-bloodmeal period and temperature-driven asynchrony between parasite development and mosquito biting rate reduce malaria transmission intensity. *PLoS One,* 8**,** e55777.

PAAIJMANS, K. P. & THOMAS, M. B. 2011. The influence of mosquito resting behaviour and associated microclimate for malaria risk. *Malaria Journal,* 10.

PAN AMERICAN HEALTH ORGANIZATION 2013. Recommendations for Scientific Evidence-Based Yellow Fever Risk Assessment in the Americas. *Technical Report.* Imperial College London.

PAN AMERICAN HEALTH ORGANIZATION. 2017. *YELLOW FEVER: Number of Confirmed Cases and Deaths by Country in the Americas, 1960-2015* [Online]. http://ais.paho.org/phip/viz/ed_yellowfever.asp: Pan American Health Organization,. [Accessed 15/08/2017].

PAN AMERICAN HEALTH ORGANIZATION 2018. Laboratory Diagnosis of Yellow Fever Virus infection. https://www.paho.org/hq/index.php?option=com_docman&view=download&category_slug=guidelines-5053&alias=46877-laboratory-diagnosis-of-yellow-fever-virus-infection&Itemid=270&lang=en: Pan American Health Organization.

PATTERSON, B. D., CEBALLOS, G., SECHREST, W., TOGNELLI, M. F., BROOKS, T., LUNA, L., ORTEGA, P., SALAZAR, I. & YOUNG, B. E. 2007. Digital Distribution Maps of Mammals of the Western Hemisphere, version 3.0. *In:* NATURESERVE (ed.). Arlington, Virgina, USA.

PATTERSON, K. D. 1992. Yellow fever epidemics and mortality in the United States, 1693-1905. *Soc Sci Med,* 34**,** 855-65.

PATZ, J. A., DASZAK, P., TABOR, G. M., AGUIRRE, A. A., PEARL, M., EPSTEIN, J., WOLFE, N. D., KILPATRICK, A. M., FOUFOPOULOS, J., MOLYNEUX, D., BRADLEY, D. J., WORKING GROUP ON LAND USE, C. & DISEASE, E. 2004. Unhealthy landscapes: Policy recommendations on land use change and infectious disease emergence. *Environ Health Perspect,* 112**,** 1092-8.

PETERSON, A. T. 2008. Biogeography of diseases: a framework for analysis. *Naturwissenschaften,* 95**,** 483-91.

PETERSON, A. T., MOSES, L. M. & BAUSCH, D. G. 2014. Mapping transmission risk of Lassa fever in West Africa: the importance of quality control, sampling bias, and error weighting. *PLoS One,* 9**,** e100711.

PINTO, N., LASKY, J., BUENO, R., KEITT, T. & GALETTI, M. 2009. Primate Densities in the Atlantic Forest of Southeast Brazil: The Role of Habitat Quality and Anthropogenic Disturbance. *South American Primates.* Springer Science+Business Media.

PITZER, V. E., AGUAS, R., RILEY, S., LOEFFEN, W. L., WOOD, J. L. & GRENFELL, B. T. 2016. High turnover drives prolonged persistence of influenza in managed pig herds. *J R Soc Interface,* 13.

PLOWRIGHT, R. K., PARRISH, C. R., MCCALLUM, H., HUDSON, P. J., KO, A. I., GRAHAM, A. L. & LLOYD-SMITH, J. O. 2017. Pathways to zoonotic spillover. *Nat Rev Microbiol,* 15**,** 502-510.

POSSAS, C., LOURENCO-DE-OLIVEIRA, R., TAUIL, P. L., PINHEIRO, F. P., PISSINATTI, A., CUNHA, R. V. D., FREIRE, M., MARTINS, R. M. & HOMMA, A. 2018. Yellow fever outbreak in Brazil: the puzzle of rapid viral spread and challenges for immunisation. *Mem Inst Oswaldo Cruz,* 113**,** e180278.

QUEYRIAUX, B., ARMENGAUD, A., JEANNIN, C., COUTURIER, E. & PELOUX-PETIOT, F. 2008. Chikungunya in Europe. *Lancet,* 371**,** 723-4.

REZENDE, I. M., SACCHETTO, L., MUNHOZ DE MELLO, E., ALVES, P. A., IANI, F. C. M., ADELINO, T. E. R., DUARTE, M. M., CURY, A. L. F., BERNARDES, A. F. L., SANTOS, T. A., PEREIRA, L. S., DUTRA, M. R. T., RAMALHO, D. B., DE THOISY, B., KROON, E. G., TRINDADE, G. S. & DRUMOND, B. P. 2018. Persistence of Yellow fever virus outside the Amazon Basin, causing epidemics in Southeast Brazil, from 2016 to 2018. *PLoS Negl Trop Dis,* 12**,** e0006538.

RIBEIRO, M. C., METZGER, J. P., MARTENSEN, A. C., PONZONI, F. J. & HIROTA, M. M. 2009. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation,* 142**,** 1141-1153.

ROGERS, D. J., WILSON, A. J., HAY, S. I. & GRAHAM, A. J. 2006. The global distribution of yellow fever and dengue. *Adv Parasitol,* 62**,** 181-220.

ROMANO, A. P., COSTA, Z. G., RAMOS, D. G., ANDRADE, M. A., JAYME VDE, S., ALMEIDA, M. A., VETTORELLO, K. C., MASCHERETTI, M. & FLANNERY, B. 2014. Yellow Fever outbreaks in unvaccinated populations, Brazil, 2008-2009. *PLoS Negl Trop Dis,* 8**,** e2740.

ROUKENS, A. H. & VISSER, L. G. 2008. Yellow fever vaccine: past, present and future. *Expert Opin Biol Ther,* 8**,** 1787-95.

ROUKENS, A. H. E., VAN HALEM, K., DE VISSER, A. W. & VISSER, L. G. 2018. Long-Term Protection After Fractional-Dose Yellow Fever Vaccination: Follow-up Study of a Randomized, Controlled, Noninferiority Trial. *Ann Intern Med,* 169**,** 761-765.

RYLANDS, A. B. 1986. Ranging Behavior and Habitat Preference of a Wild Marmoset Group, Callithrix-Humeralifer(Callitrichidae, Primates). *Journal of Zoology,* 210**,** 489-514.

RYLANDS, A. B., WILLIAMSON, E. A., HOFFMANN, M. & MITTERMEIER, R. A. 2008. Primate surveys and conservation assessments. *Oryx,* 42**,** 313-314.

SAKAMOTO, Y., YAMAGUCHI, T., YAMAMOTO, N. & NISHIURA, H. 2018. Modeling the elevated risk of yellow fever among travelers visiting Brazil, 2018. *Theoretical Biology and Medical Modelling,* 15.

SELTMANN, A., CZIRJAK, G. A., COURTIOL, A., BERNARD, H., STRUEBIG, M. J. & VOIGT, C. C. 2017. Habitat disturbance results in chronic stress and impaired health status in forest-dwelling paleotropical bats. *Conserv Physiol,* 5**,** cox020.

SHAH, H. A., HUXLEY, P., ELMES, J. & MURRAY, K. A. 2019. Agricultural land-uses consistently exacerbate infectious disease risks in Southeast Asia. *Nat Commun,* 10**,** 4299.

SHEARER, F. M., LONGBOTTOM, J., BROWNE, A. J., PIGOTT, D. M., BRADY, O. J., KRAEMER, M. U. G., MARINHO, F., YACTAYO, S., DE ARAUJO, V. E. M., DA NOBREGA, A. A., FULLMAN, N., RAY, S. E., MOSSER, J. F., STANAWAY, J. D., LIM, S. S., REINER, R. C., JR., MOYES, C. L., HAY, S. I. & GOLDING, N. 2018. Existing and potential infection risk zones of yellow fever worldwide: a modelling analysis. *Lancet Glob Health,* 6**,** e270-e278.

SHEARER, F. M., MOYES, C. L., PIGOTT, D. M., BRADY, O. J., MARINHO, F., DESHPANDE, A., LONGBOTTOM, J., BROWNE, A. J., KRAEMER, M. U. G., O'REILLY, K. M., HOMBACH, J., YACTAYO, S., DE ARAUJO, V. E. M., DA NOBREGA, A. A., MOSSER, J. F., STANAWAY, J. D., LIM, S. S., HAY, S. I., GOLDING, N. & REINER, R. C., JR. 2017. Global yellow fever vaccination coverage from 1970 to 2016: an adjusted retrospective analysis. *Lancet Infect Dis*.

SMITHBURN, K. C. & MAHAFFY, A. F. 1945. Immunization Against Yellow Fever. *The American Journal of Tropical Medicine and Hygiene,* s1-25**,** 217-223.

SOPER, F. L. 1937. The Newer Epidemiology of Yellow Fever. *American Journal of Public Health,* 27.

SOPER, F. L., PENNA, H., CARDOSO, E., SERAFIM JR, J., FROBISHER JR, M. & PINHEIRO, J. 1933. Yellow Fever without Aëdes aegypti. Study of a Rural Epidemic in the Valle do Chanaan, Espirito Santo, Brazil, 1932. *American Journal of Hygiene,* 18**,** 555-587.

SOPER, F. L., WILSON, D. B., LIMA, S. & ANTUNES, W. S. 1943. *The Organization of Permanent Nationwide Anti-Aedes aegypti Measures in Brazil,* New York, The Rockefeller Found.

STAPLES, J. E. & MONATH, T. P. 2008. Yellow fever: 100 years of discovery. *JAMA,* 300**,** 960-2.

STEEL, D. G. & HOLT, D. 1996. Analysing and adjusting aggregation effects: The ecological fallacy revisited. *International Statistical Review,* 64**,** 39-60.

STRODE, G., BUGHER, J., AUSTIN-KERR, K., SMITH, H., SMITHBURN, K., TAYLOR, R., THEILER, M., WARREN, A. & WHITMAN, L. 1951. *Yellow Fever,* New York, McGraw-Hill Book Company.

SUPARIT, P., WIRATSUDAKUL, A. & MODCHANG, C. 2018. A mathematical model for Zika virus transmission dynamics with a time-dependent mosquito biting rate. *Theor Biol Med Model,* 15**,** 11.

TEAM, R. C. 2015a. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

TEAM, W. E. R. 2015b. West African Ebola Epidemic after One Year - Slowing but Not Yet under Control. *New England Journal of Medicine,* 372**,** 584-587.

THEILER, M. & WHITMAN, L. 1935. Quantitative Studies of the Virus and Immune Serum Used in Vaccination against Yellow Fever1. *The American Journal of Tropical Medicine and Hygiene,* s1-15**,** 347-356.

THORN, J. S., NIJMAN, V., SMITH, D. & NEKARIS, K. A. I. 2009. Ecological niche modelling as a technique for assessing threats and setting conservation priorities for Asian slow lorises (Primates: Nycticebus). *Diversity and Distributions,* 15**,** 289-298.

TOWERS, S., BRAUER, F., CASTILLO-CHAVEZ, C., FALCONAR, A. K. I., MUBAYI, A. & ROMERO-VIVAS, C. M. E. 2016. Estimate of the reproduction number of the 2015 Zika virus outbreak in Barranquilla, Colombia, and estimation of the relative role of sexual transmission. *Epidemics,* 17**,** 50-55.

TRAPIDO, H. & GALINDO, P. 1955. The investigation of a sylvan yellow fever epizootic on the north coast of Honduras, 1954. *Am J Trop Med Hyg,* 4**,** 665-74.

UNEP-WCMC. 2019. Protected Area Profile for Brazil from the World Database of Protected Areas. https://www.protectedplanet.net/country/BRA.

UNITED NATIONS, D. O. E. A. S. A., POPULATION DIVISION, POPULATION ESTIMATES AND PROJECTIONS SECTION. 2016. *World Population Prospects: The 2016 Revision* [Online]. Available: http://esa.un.org/unpd/wpp/ [Accessed 08 July 2016].

VACCINE IMPACT MODELLING CONSORTIUM. 2018. *Vaccine Impact Modelling Consortium* [Online]. https://www.vaccineimpact.org/. [Accessed 12/02/18].

VAN DEN HURK, A. F., MCELROY, K., PYKE, A. T., MCGEE, C. E., HALL-MENDELIN, S., DAY, A., RYAN, P. A., RITCHIE, S. A., VANLANDINGHAM, D. L. & HIGGS, S. 2011. Vector competence of Australian mosquitoes for yellow fever virus. *Am J Trop Med Hyg,* 85**,** 446-51.

VAN DER STUYFT, P., GIANELLA, A., PIRARD, M., CESPEDES, J., LORA, J., PEREDO, C., PELEGRINO, J. L., VORNDAM, V. & BOELAERT, M. 1999. Urbanisation of yellow fever in Santa Cruz, Bolivia. *Lancet,* 353**,** 1558-62.

VAN, P., MANH, V. & HOANG, T. 2010. *Using environmental niche model to study the distribution of Tonkin snub-nosed monkey (Rhinopithecus avunculus) in the Northeastern Vietnam under some climate change scenarios*.

VANWAMBEKE, S. O., SOMBOON, P., HARBACH, R. E., ISENSTADT, M., LAMBIN, E. F., WALTON, C. & BUTLIN, R. K. 2007. Landscape and land cover factors influence the presence of Aedes and Anopheles larvae. *J Med Entomol,* 44**,** 133-44.

VARGAS-MENDEZ, O. & ELTON, N. W. 1953. Naturally acquired yellow fever in wild monkeys of Costa Rica. *Am J Trop Med Hyg,* 2**,** 850-63.

VASCONCELOS, P. F., BRYANT, J. E., DA ROSA, T. P., TESH, R. B., RODRIGUES, S. G. & BARRETT, A. D. 2004. Genetic divergence and dispersal of yellow fever virus, Brazil. *Emerg Infect Dis,* 10**,** 1578-84.

VASCONCELOS, P. F. C., ROSA, P. A. T. A., PINHEIRO, F. P., RODRIGUES, S. G., ROSA, E. S. T., CRUZ, A. C. R. & ROSA, J. F. S. T. 1999. Aedes aegypti, dengue and re-urbanization of yellow fever in Brazil and other South American countries–Past and present situation and future perspectives. *Dengue Bulletin***,** 55-66.

VIDAL-GARCIA, F. & SERIO-SILVA, J. C. 2011. Potential distribution of Mexican primates: modeling the ecological niche with the maximum entropy algorithm. *Primates,* 52**,** 261-70.

WADDELL, D. 1990. Yellow fever in Europe in the early 19th century - Cadiz 1819. *Rep Proc Scott Soc Hist Med***,** 20-34.

WAGENMAKERS, E. J. & FARRELL, S. 2004. AIC model selection using Akaike weights. *Psychonomic Bulletin & Review,* 11**,** 192-196.

WAGGONER, J. J., ROJAS, A. & PINSKY, B. A. 2018. Yellow Fever Virus: Diagnostics for a Persistent Arboviral Threat. *J Clin Microbiol,* 56.

WANG, E., WEAVER, S. C., SHOPE, R. E., TESH, R. B., WATTS, D. M. & BARRETT, A. D. 1996. Genetic variation in yellow fever virus: duplication in the 3' noncoding region of strains from Africa. *Virology,* 225**,** 274-81.

WARREN-THOMAS, E. M., EDWARDS, D. P., BEBBER, D. P., CHHANG, P., DIMENT, A. N., EVANS, T. D., LAMBRICK, F. H., MAXWELL, J. F., NUT, M., O'KELLY, H. J., THEILADE, I. & DOLMAN, P. M. 2018. Protecting tropical forests from the rapid expansion of rubber using carbon payments. *Nat Commun,* 9**,** 911.

WASSERMAN, S., TAMBYAH, P. A. & LIM, P. L. 2016. Yellow fever cases in Asia: primed for an epidemic. *Int J Infect Dis,* 48**,** 98-103.

WEE, L. K., WENG, S. N., RADUAN, N., WAH, S. K., MING, W. H., SHI, C. H., RAMBLI, F., AHOK, C. J., MARLINA, S., AHMAD, N. W., MCKEMY, A., VASAN, S. S. & LIM, L. H. 2013. Relationship between Rainfall and Aedes Larval Population at Two Insular Sites in Pulau Ketam, Selangor, Malaysia. *Southeast Asian Journal of Tropical Medicine and Public Health,* 44**,** 157-166.

WHITE, L. A., FORESTER, J. D. & CRAFT, M. E. 2018. Disease outbreak thresholds emerge from interactions between movement behavior, landscape structure, and epidemiology. *Proceedings of the National Academy of Sciences of the United States of America,* 115**,** 7374-7379.

WHO. 2016. *WHO/UNICEF estimates of national immunization coverage* [Online]. Available: http://www.who.int/immunization/monitoring_surveillance/routine/coverage/en/index4.html [Accessed].
WHO & UNICEF 2010. Yellow Fever Initiative - Progress report 2006 to 2009. WHO/UNICEF.
WHO & UNICEF 2011. Yellow Fever Initiative - Joint WHO and UNICEF 2010 Progress Report.
WIKAN, N. & SMITH, D. R. 2016. Zika virus: history of a newly emerging arbovirus. *Lancet Infect Dis,* 16**,** e119-e126.
WORLD HEALTH ORGANISATION 2011. Disease Outbreak News (DON). http://www.who.int/csr/don/2012_02_03b/en/.
WORLD HEALTH ORGANISATION 2013. Disease Outbreak News (DON). http://www.who.int/csr/don/2012_02_03b/en/.
WORLD HEALTH ORGANISATION. 2018. *Dashboard: International Coordinating Group (ICG) on Vaccine Provision on yellow fever* [Online]. http://www.who.int/csr/disease/icg/yellow-fever-dashboard/en/. [Accessed 07/02/2018 2018].
WORLD HEALTH ORGANISATION, U. 2009. Yellow Fever Initiative Progress 2006 to 2009. Geneva: World Health Organisation.
WORLD HEALTH ORGANISATION, U. 2010. Yellow Fever Initiative Joint WHO and UNICEF 2010 Progress Report. Geneva: World Health Organisation.
WORLD HEALTH ORGANIZATION Disease Outbreak News (DON). World Health Organisation.
WORLD HEALTH ORGANIZATION The Weekly Epidemiological Record (WER). World Health Organisation.
WORLD HEALTH ORGANIZATION 2008. The Weekly Epidemiological Record (WER). World Health Organization,.
WORLD HEALTH ORGANIZATION 2010a. Weekly epidemiological record. *Weekly epidemiological record.* https://www.who.int/wer/2010/wer8547.pdf?ua=1: World Health Organization.
WORLD HEALTH ORGANIZATION 2016a. Global Strategy to Eliminate Yellow fever Epidemics (EYE). http://www.who.int/immunization/sage/meetings/2016/october/2_EYE_Strategy.pdf World Health Organization,.
WORLD HEALTH ORGANIZATION 2016b. Global Strategy to Eliminate Yellow Fever Epidemics (EYE). *SAGE.* Geneva: World Health Organization.
WORLD HEALTH ORGANIZATION 2016c. The Yellow fever initiative: an introduction.
WORLD HEALTH ORGANIZATION 2017a. WHO position on the use of fractional doses - June 2017, addendum to vaccines and vaccination against yellow fever WHO: Position paper - June 2013. *Vaccine,* 35**,** 5751-5752.
WORLD HEALTH ORGANIZATION 2017b. Yellow fever in Africa and the Americas, 2016. *Wkly Epidemiol Rec,* 92**,** 442-52.
WORLD HEALTH ORGANIZATION 2018. Yellow fever – Republic of the Congo. *Disease Outbreak News.* https://www.who.int/csr/don/7-september-2018-yellow-fever-congo/en/: World Health Organization.
WORLD HEALTH ORGANIZATION 2019a. The Weekly Epidemiological Record (WER). World Health Organization,.
WORLD HEALTH ORGANIZATION 2019b. Yellow fever - Nigeria. *Disease Outbreak News.* https://www.who.int/csr/don/09-january-2019-yellow-fever-nigeria/en/: World Health Organization.
WORLD HEALTH ORGANIZATION, U. 2010b. Yellow Fever Initiative Joint WHO and UNICEF 2010 Progress Report. Geneva: World Health Organisation.
WRIGHT, M. & ZIEGLER, A. 2017. ranger: A fast implementation of random forests for high dimensional data in C ++ and R. *Journal of Statistical Software*.
WU, J. T., PEAK, C. M., LEUNG, G. M. & LIPSITCH, M. 2016. Fractional dosing of yellow fever vaccine to extend supply: a modelling study. *Lancet*.
WYNNE, J. W. & WANG, L. F. 2013. Bats and viruses: friend or foe? *PLoS Pathog,* 9**,** e1003651.

YAMANA, T. K. & ELTAHIR, E. A. 2013. Incorporating the effects of humidity in a mechanistic model of Anopheles gambiae mosquito population dynamics in the Sahel region of Africa. *Parasit Vectors,* 6**,** 235.

ZELLER, H., VAN BORTEL, W. & SUDRE, B. 2016. Chikungunya: Its History in Africa and Asia and Its Spread to New Regions in 2013-2014. *J Infect Dis,* 214**,** S436-S440.

ZHAO, S., STONE, L., GAO, D. & HE, D. 2018. Modelling the large-scale yellow fever outbreak in Luanda, Angola, and the impact of vaccination. *PLoS Negl Trop Dis,* 12**,** e0006158.

ZITTRA, C., VITECEK, S., OBWALLER, A. G., ROSSITER, H., EIGNER, B., ZECHMEISTER, T., WARINGER, J. & FUEHRER, H. P. 2017. Landscape structure affects distribution of potential disease vectors (Diptera: Culicidae). *Parasit Vectors,* 10**,** 205.

Appendix A

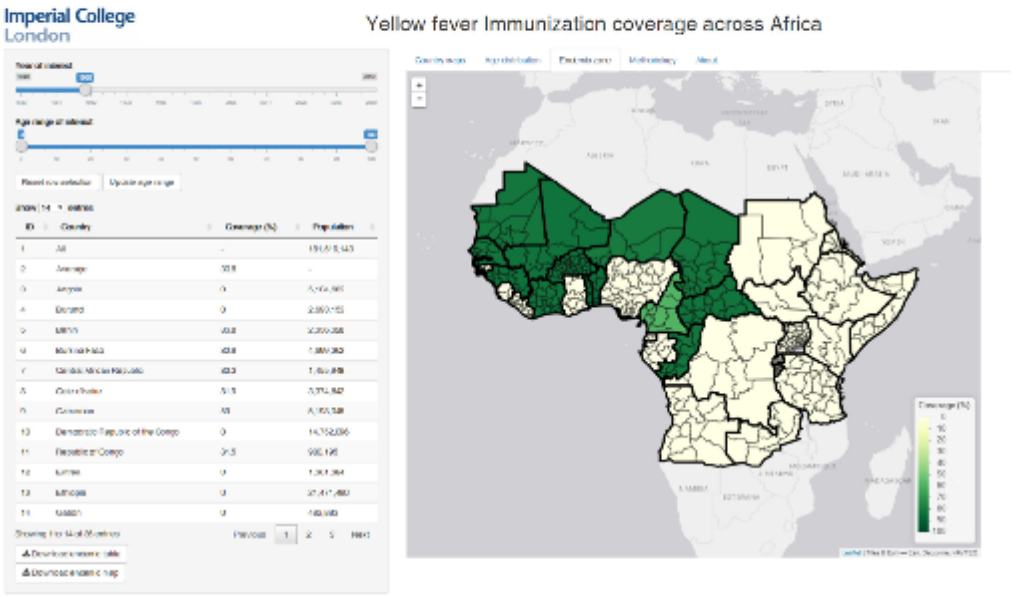


Fig. 1. "Endemic zone" tab showing a table and map of the vaccination coverages at the country level across the African endemic zone.

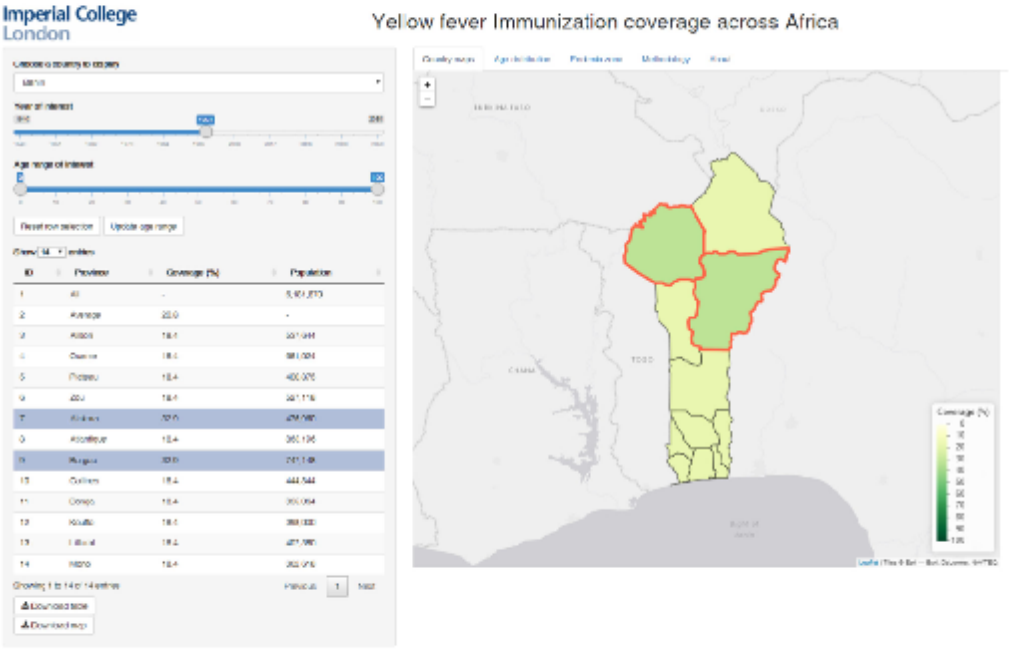


Fig. 2. "Country maps" tab showing a table and map of the vaccination coverage at the province level in Benin 1997 with two provinces highlighted, and the country average.

Appendix B

Table 1. Large scale campaigns since 2006 missing from Shearer et al., (2017)'s vaccination coverage estimates.

COUNTRY	YEAR	VACCINATED /DOSES	TYPE OF CAMPAIGN	SOURCES
TOGO	2007	3,590,000	Preventive	(World Health Organisation, 2009)
LIBERIA	2008	471,725	Responsive	(World Health Organization, 2008)
MALI	2008	5,870,000	Preventive	(WHO and UNICEF, 2010)
BENIN	2009	6,320,000	Preventive	(WHO and UNICEF, 2010)
SIERRA LEONE	2009	3,980,000	Preventive	(WHO and UNICEF, 2010)
CENTRAL AFRICAN REPUBLIC	2010	2,261,230	Preventive	(WHO and UNICEF, 2011)
LIBERIA	2011	133,033	Response	(World Health Organisation, 2010)
COTE D'IVOIRE	2011	801,070	Response	(World Health Organisation, 2010)
GHANA	2011	5,800,000	Preventive	(World Health Organisation, 2011)
SIERRA LEONE	2011	144,479	Response	(World Health Organisation, 2011)
CHAD	2013	1,000,000	Response	(World Health Organisation, 2013)

UGANDA	2016	775,670	Response	_ENREF_39(World Health Organisation, 2018)
TOTAL		**31,147,207**		

Appendix C

Alternative covariate selection process for NHP density prediction in the South-East Atlantic states of Brazil with the omission of the IUCN defined species range

Covariate selection

Here we have omitted the “genera range”, as defined as the IUCN specified species range (IUCN, 2017), as a potential covariate for the covariate selection process. This was done to explore how genera distributions would appear without this potentially artificial constraint.

Here, , covariate selection (other than the range) remains fairly constant between those selected when the genera range is present, and those when it has been omitted. The primary difference is found in modelling the density of Brachyteles which included numerous new covariates.

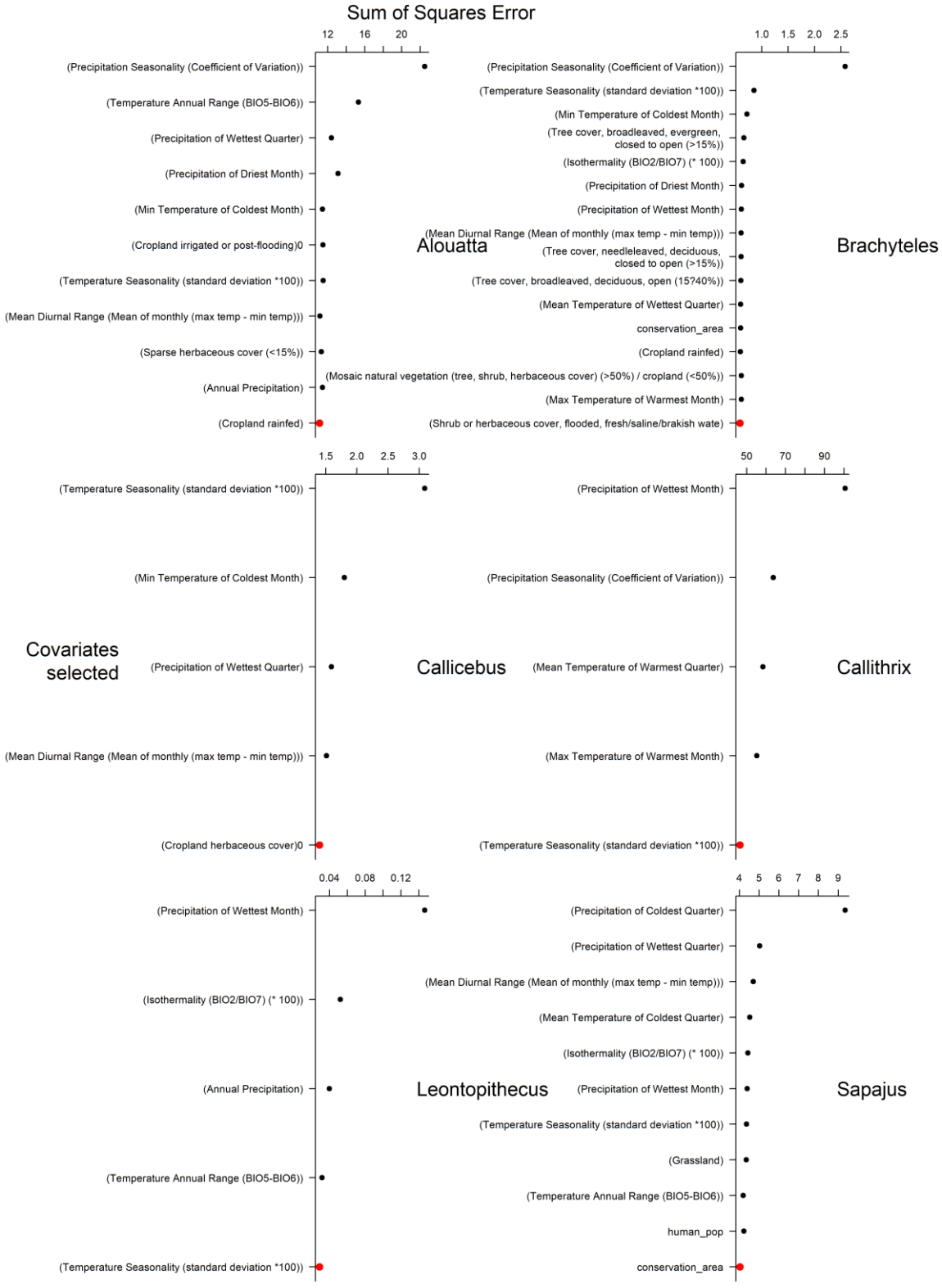


Figure 8.3.1 Covariates, with the IUCN range of genera excluded from the pool, selected through out-of-sample forward stepwise selection for each genera model for density. The red point indicates the covariate addition that resulted in the lowest sum of squares error.

NHP predicted distributions

Here, there are broad agreements between previous genera density distributions and amounts, indicating robust predictions across much of the range. However, there are notable differences in individual genera and areas.

The primary difference shared between genera predictions is the increased range in predictions, with larger areas predicted to have low or no densities of NHP's here, where in models that were allowed to select the range, they are present.This has the potential for artificially inflating the predicting distributions and populations of NHP's, as areas where the IUCN genera distribution shapefiles (a binary presence/absence) note presence, values are raised regardless of the suitability indicated by all other covariates.

This is seen particularly in Alouatta and Sapajus distributions, which in the prediction runs that include range show a presence throughout much of the mapped area, but when the range is omitted, much of Northern and Central Minas Gerais are predicted to have negligible densities.

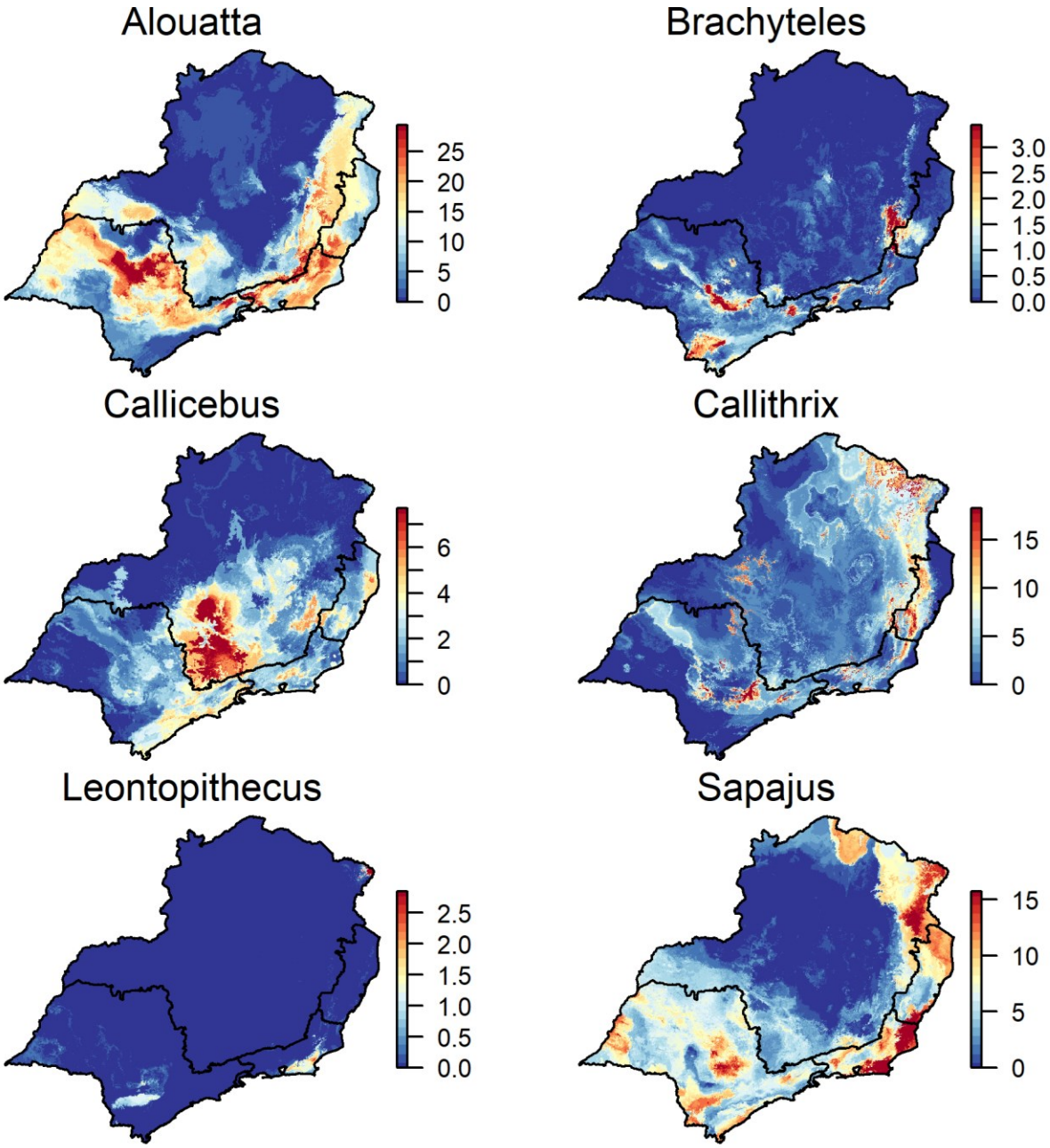


Figure 8.3.2 Random forest predictions of models created through the out-of-sample forward stepwise covariate selection method, for the density of six NHP genera found in the South-East Atlantic states of Brazil.