Dengue dynamics under climate drivers

Analysis with ecological and epidemiological frameworks.

IFT - UNESP
INSTITUTO DE FÍSICA TEÓRICA

Rafael Lopes Paixão da Silva

IFT-T.007/2022

Silva, Rafael Lopes Paixão da
S586d Dengue dynamics under climate drivers: analysis with ecological and epidemiological frameworks / Rafael Lopes Paixão da Silva. – São Paulo, 2022
105 f.

Tese (doutorado) – Universidade Estadual Paulista (Unesp), Instituto de Física Teórica (IFT), São Paulo
Orientador: Roberto André Kraenkel

1. Análise de séries temporais. 2. Causalidade (Física). 3. Epidemiologia – Modelos estatísticos. I. Título

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Física Teórica (IFT), São Paulo. Dados fornecidos pelo autor(a).

Dengue dynamics under climate drivers

Analysis with ecological and epidemiological frameworks.

This thesis has received financial support from CNPq and CAPES (Brazil)

Dengue dynamics under climate drivers

Analysis with ecological and epidemiological frameworks.

THESIS

submitted to the Institute of Theoretical Physics
São Paulo State University, Brazil

as a partial fulfillment of the requirements for the degree of

Doctor of Science

October 31st 2022

Rafael Lopes Paixão da Silva

Roberto André Kraenkel (supervisor)

Examination Committee

Prof. Dr. Roberto André Kraenkel (supervisor)
Instituto de Física Teórica, Universidade Estadual Paulista 'Júlio de Mesquita Filho'

Prof. Dr. Fernando Fagundes Ferreira
Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo

Prof. Dr. Roberto Andrade
Instituto de Física, Universidade Federal da Bahia

Prof. Dr. Americo Barbosa da Cunha Junior
Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro

Prof. Dr. Alexandre Souto Martinez
Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo

Instituto de Física Teórica - UNESP
R. Dr. Bento Teobaldo Ferraz 271, bloco II
01140-070 São Paulo, Brasil

This thesis is dedicated to too many people, but mostly to myself . . .

“ Apenas nos pusimos en dos pies
Comenzamos a migrar por la sabana
Siguiendo la manada de bisontes
Más allá del horizonte
A nuevas tierras, lejanas
Los niños a la espalda y expectantes
Los ojos en alerta, todo oídos
Olfateando aquel desconcertante paisaje nuevo, desconocido

Somos una especie en viaje
No tenemos pertenencias sino equipaje
Vamos con el polen en el viento
Estamos vivos porque estamos en movimiento
Nunca estamos quietos, somos trashumantes
Somos padres, hijos, nietos y bisnietos de inmigrantes
Es más mío le que sueño que lo que toco

Yo no soy de aquí
Pero tú tampoco
Yo no soy de aquí
Pero tú tampoco
De ningún lado del todo

De todos lados un poco

Atravesamos desiertos, glaciares, continentes
El mundo entero de extremo a extremo
Empecinados, supervivientes
El ojo en el viento y en las corrientes
La mano firme en el remo
Cargamos con nuestras guerras
Nuestras canciones de cuna
Nuestro rumbo hecho de versos
De migraciones, de hambrunas
Y así ha sido desde siempre, desde el infinito
Fuimos la gota de agua viajando en el meteorito
Cruzamos galaxias, vacío, milenios
Buscábamos oxígeno, encontramos sueños

Apenas nos pusimos en dos pies
Y nos vimos en la sombra de la hoguera
Escuchamos la voz del desafío
Siempre miramos el río
Pensando en la otra rivera

Somos una especie en viaje
No tenemos pertenencias sino equipaje
Vamos con el polen en el viento
Estamos vivos porque estamos en movimiento
Nunca estamos quietos, somos trashumantes
Somos padres, hijos, nietos y bisnietos de inmigrantes
Es más mío le que sueño que lo que toco

Yo no soy de aquí
Pero tú tampoco
Yo no soy de aquí
Pero tú tampoco
De ningún lado del todo y
De todos lados un poco

Lo mismo con las canciones, los pájaros, los alfabetos
Si quieres que algo se muera, déjalo quieto

Jorge Drexler, *Movimiento ”*

Acknowledgments

It is an unfair task to try to acknowledge all the persons that have had a participation or contributed to this thesis, I start this acknowledgments by saying sorry to the ones that I might forget.

Foremost, this study was financed in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) – Brazil (grant number: 141698/2018-7), and another part, the exchange period abroad, by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

The first person to thank is my Ph.D. advisor, Roberto André Kraenkel, who has accepted me as student and who has guided me since the first day of my master's in 2016 and whom I have developed a great relationship of learning and friendship. Without Roberto's sharp mind for the important questions, the theme of this thesis would have never been raised to my attention.

I would also have to thank Otavio Tavares Ranzani, who has embraced my desire to learn and understand more of epidemiology and who has been guiding me on this new field. Without Otavio's boldness to investigate the persistent questions in details and with accuracy, some of the new tools that I've learned would never have been present on this thesis.

On my time in Barcelona, I would like to thank Xavier Basagaña too, he has always been receptive and encouraging to me. As well as the friends I made there, some of

them were just kind to me a couple of times, this is important, being a newer person on a newer country could be tough, and their welcoming has helped to smoothing this period.

I have to thank Renato Mendes Coutinho and Paulo Inácio L. K. Prado, both of them although not been formally my advisors, has helped me on many parts of this thesis, sometimes directly and sometimes keeping me in pace with a high quality science that can be done. Another person that has to be remembered here is Brenno Troca Cabella, he taught me the starting point of many results on this thesis, thank you for your patience and dedication to me.

I want to thank my colleagues at São Paulo State University and Barcelona Institute of Global Health, their company on the hard-working days, coding discussion and beer, is the part not seen but very essential on the process of writing this thesis.

I want to thank the health workers and the surveillance system workers, who are responsible for constructing and inputting the data I used on this thesis, without it the thesis were not possible.

Lastly, I am very thankful to my family and my partner, Mariana, who were a constant source of support, encouragement and love.

All of you, mentioned or not here, know your role on this thesis, and I am very thankful for this time and for the help I have received on this process. It is good to know I have a great network of support.

Abstract

Dengue has been a disease present in the Brazilian ecological and epidemiological scenarios since the mid-twentieth century. From the 1990s, the cases that, before, were sporadic,and distributed without any clear association to population and territory variables, began to exhibit a new dynamic. In 2001, by a law that made notification of dengue cases mandatory, notifications of the disease exploded and dengue began to be seen as a public health problem.

Dengue is a disease transmitted by the mosquito *Aedes Aegypti*. It has a clear seasonal component. However, even if it is understood that certain seasonality is derived from climate, we still lack better quantitative analyses of the effects of climate on dengue epidemics.

In this thesis, time-series of dengue cases and essential climate variables (ECV), such as temperature and precipitation, are analysed using *convergent cross-mapping (CCM)*. CCM is based on the result of the *Takens* theorem which states that the attractor of a dynamical system can be reconstructed from the time-series of one of the variables of the system, and the time series of this same variable with time delays. Based on this result, methods for determining causal relations have been developed. We use these methods and show that, in the case of the city of Rio de Janeiro, precipitation is the most important ECV relent for dengue epidemics..

Dengue, in a less common way, can lead to hospitalization according to previous infection history and weather conditions. Using time-series of hospitalized dengue cases and temperature series, we explore which are the risks of developing a more severe dengue condition given temperature exposure. To this end, we use distributed lag non-linear models (DLNM), which, by using delays of a series of an exposure factor in a generalized linear model with predictors of the response series, in our case the series of dengue hospitalizations, provides a statistical association between these factors. This allows us to account for relative risks for each temperature benchmark. In a further development , starting at the municipalities level, we perform a meta-analysis of this association, first by macro regions of Brazil and second for the whole country, from which we derive a functional form for the relative risk that each temperature percentile has in making a dengue-case a possible hospitalization for dengue.

Keywords: Time Series Analysis; Causality Criteria; Causility; Causal Inference; Empirical

Dynamic Modelling; EDM; Cross-Convergent Mapping; CCM; Dengue; Dynamical System; Complex Systems; Distributed Lag Non-Linear Models; DLNM; Statistical Inference; Environmental Epidemiology; Statistical Modeling; Mathematical Modeling; Computational Epidemiology;

.

Knowledge Area: Physics; Non-linear Systems; Complex Systems; Mathematical Modeling; Statistical Modeling; Mathematical Ecology; Mathematical Epidemiology; Computational Epidemiology; Environmental Epidemiology.

Resumo

A dengue tem sido uma doença constante nos cenários ecológico e epidemiológico brasileiros desde os meados do século XX. A partir da década de 1990 os casos que, antes, eram esporádicos, e se distribuíam sem qualquer clara associação a variáveis populacionais e do território, começaram a ter uma nova dinâmica. Em 2001, por uma lei que tornou obrigatória a notificação de casos de dengue, as notificações da doença explodem e a dengue passa a ser vista como um problema de saúde pública.

A dengue é uma doença transmitida pela picada do mosquito *Aedes Aegypti*. Ela tem uma componente sazonal. Porém, ainda que se entenda que essa sazonalidade seja derivada do clima, ainda nos faltam melhores análises quantitativas dos efeitos do clima sobre as epidemias de dengue.

Nesta tese, as séries temporais de casos de dengue e de variáveis climáticas essenciais (VCE), como temperatura e precipitação, são analisadas através do *mapeamento cruzado convergente (CCM)*. O CCM se baseia no resultado do teorema de *Takens* que diz que o atrator de um sistema dinâmico pode ser reconstruído a partir das séries temporais de uma das variáveis do sistema, e as séries temporais desta mesma variável com atrasos temporais. A partir deste resultado desenvolvem-se métodos de determinação de relações causais. Utilizamos este método e mostramos que, no caso da cidade do Rio de Janeiro, a precipitação é a VCE mais importante em relação a epidemias de dengue.

A dengue, de forma menos comum, pode levar a uma hospitalização conforme histórico de infecção anterior e devido às condições climáticas. Lançando mão das séries-temporais de casos de dengue hospitalizados e de séries de temperatura, exploramos quais são os riscos de se desenvolver um quadro mais grave de dengue dada a exposição à temperatura. Para tal utilizamos modelos não-lineares de atrasos distribuídos (DLNM), que ao colocar atrasos de uma série de um fator de exposição num modelo generalizado linear com preditores da série resposta, no nosso caso a série de hospitalizações por dengue, fornece uma associação estatística entre esses fatores com a qual podemos contabilizar riscos relativos para cada referencial de temperatura. Em um desenvolvimento posterior e com todas as associações feitas ao nível municipal, fazemos uma meta-análise dessa associação, primeiro por macro regiões do Brasil e segundo para o país todo, do qual retiramos uma forma funcional para o risco relativo de cada percentil de temperatura tem em tornar um caso de dengue uma possível hospitalização por dengue.

Palavras Chaves: Análise de Séries Temporais; Crtiério de Causalidade; Causalidade; Inferência Causal; CCM; Sistemas Dinâmicos; Sistemas Complexos; DLNM; Dengue; Inferência Estatística; Epidemiologia Ambiental; Modelos estatísticos; Modelagem Matemática; Epidemiologia Matemática; Epidemiologia Computacional; Epidemiologia de dados.

Contents

List of Figures

List of Tables

Part I

Historical Background and data sources

1 Introduction

Dengue Fever, dengue disease or simply dengue is a vector-borne infection caused by four different serotypes of a *Flaviridae* know as dengue virus (DENV) and it is transmitted mainly by *Aedes Aegypti* & *Aedes Albopictus* mosquitoes [1, 2, 3]. Dengue epidemics are a result of a series of very complex phenomena that emerge from an infection caused by a virus and transmitted by a mosquito, but involves a series of other factors as well. The climate and the human component on this system plays a major role not yet completely understood and unveiled. To study and to understand how these factors interact and give rises to dengue epidemics as a complex phenomenon is the main endeavor of this thesis.

Dengue has been first described since the last years of the 19th century and the first decades of the 20th century [4]. Before this, a severe, debilitating fever has been documented in Asia and was probably due to dengue[4]. The *Aedes Aegypti* mosquito was eradicated in large parts of the Americas in the 1950s, but was reintroduced in the decade 1970s [4]. Campaigns against *Aedes Aegypti* infestation, a common experience for large parts of the Brazilian population.

Globally, the most recent estimates of dengue burden puts it into a level of more than ~ 390 millions infections per year, with about ~ 96 millions with some level of severity[5]. This estimate is probably not updated to the reality of the burden of dengue nowadays. The World Health Organization (WHO) states that dengue cases have increased 8-fold during the last two decades. Dengue has been a great problem in Low-Income and Middle-Income countries (LMIC) during much of the 20th century, [5, 6, 7, 8, 9, 10, 11], particularly, and continues to be a global health problem demanding specific strategies and guidelines from WHO to enforce its mitigation [12, 13, 14].

A. Aegypti is the main vector to the dengue disease, specially in Brazil[15]. It is also a vector of several other viral diseases, (e.g. Chikungunya, Zika, Yellow Fever) commonly called arbovirus, and it is a widely and globally distributed mosquito [15, 16] and it is the dominant mosquito species in the Americas[15, 16].

A. Aegypti is known to be a very efficient vector and is well adapted to urban areas. A very wide range of studies has shown not only that urban adaptation has impact on its efficiency for transmission but that also climate itself, more specifically temperature and precipitation, plays an important role in the transmission dynamics [17, 18, 19, 20, 21, 22].

Dengue infections are caused by viruses of the *Flaviviridae* family[3, 2]. Dengue has four serotypes which are present in different regions of with different endemicities [3, 23]. In particular to Brazil, the four serotypes has been present on the last decades [24, 25]. All the four serotypes of dengue virus can cause infections ranging from asymptomatic to severe cases, the former requiring, sometimes, complex health care [26].

Dengue symptomatology is characterized by recurrent fever, acute pain on the joints, behind eye pain and a prolonged lethargy.

Understanding how the mosquito's cycles, as well as those of the disease itself, are governed is nowadays a scientific problem which has given rise to a great amount of research on potentials to account and mitigate the disease spread [27, 28].

A vaccine is still a hope to curb the burden of dengue, specially on LMIC, but none of the candidates of trial vaccines being tested has been shown to be a feasible and commercial vaccine in a short-term [27, 29, 30, 31]. Our strategies of mitigation and control of dengue still rely on strategies of vector control, with regular application of larvicide and mosquitocide through the territory just before the known dengue season.

This chapter is divided into four sections, a first section with a historical background on dengue data in Brazil and briefly the history of surveillance built to record these data. A second section is on the problem of modelling dengue and its relation with climate, and how and why some climate variables are chosen to understand this problem. A third section is about the kind of methods used on this thesis to study dengue disease and why they are suited to each type of problem studied. Finally, the chapter ends with a brief section on the organization of the thesis and its parts.

1.1 Historical Background

Brazil is one of the most affected countries by dengue [5], encompassing about ~ 20.8 million notifications in the last 20 years, from 2001 to 2020. Brazil has a history during the 20th century in relation with dengue disease. At the beginning of the 20th century, dengue cases were described over different locations of the country, although it was not immediately recognized as a public health problem. After this early phase, dengue disease practically disappeared, with the local elimination of the vector, *A. Aegypti* [4]. Unfortunately, the disease has been re-established as an endemic disease[4, 17] as a consequence of the reintroduction of its vectors in the 1970's. Dengue is a disease of mandatory notification in Brazil since 2000, due to a federal decree, and has a specific routine to its notification through a dedicated system [32]. The national System for Information of Notification of Grievances, or in Portuguese, *Sistema de Informação de Agravos de Notificação* (SINAN) is the system where dengue cases are notified. The module of the SINAN dedicated to the reporting and recording of the dengue cases is sometimes called SINAN-Dengue. This, guarantees, since 2000, historical records of dengue cases, that were reported to the notification system, in the Brazilian territory, keeping the epidemiological surveillance of dengue in pace with the guidelines by the WHO[33].

Since the implementation of the surveillance system, the routines to notification have been revised to keep the notifications closer to the reality of the disease incidence, as new tools have been developed to visualize and understand the data generated by the notification

system [34, 35].

1.1.1 Before 2001

The recorded data can generate historical time-series of dengue cases registered in Brazil encompassing more than 11.6 million cases confirmed, Table 2.1, with several classifications to the final diagnostics, with the possibility of being cases closed by diagnostics given by clinical criteria, laboratory criteria and/or epidemiological criteria[32]. Although it looks like a huge number of cases registered, this time-series certainly does not include the totality of infections due to the dengue virus. The disease has different clinical manifestations, going from totally asymptotically cases to severely debilitating symptomatically cases[2, 2, 36]. In some periods of low incidence of dengue cases, estimates suggest that only 1 in 40 dengue infections are reported in SINAN-Dengue and in high incidence periods the reporting relation to the probable infections occurring is close to 1 reported case for every 20 infections occurring [37].

Since 2001, Brazil has a surveillance system in line with the WHO guidelines[33] to report suspected cases and subsequently trigger a clinical investigation that results into a diagnosis of the case, which is then confirmed on the system. Moreover, there are studies suggesting that before 2001 we had a greater incidence of dengue than the reported in the historical registers [1]. The same problem happens for data concerning hospitalization by dengue virus infections. SINAN-Dengue is a surveillance system suited to report symptomatic cases of dengue, it captures cases that range from mild to oligosymptomatic cases to cases with some severity. For the specific case of severe hospitalized cases, the most appropriate reporting system is the SIH, *Sistema de Informação de Hospitalizações* [34]. 49% of the suspected severe cases that are reported on the SIH system and are confirmed as a dengue severe cases are not reported on the SINAN-Dengue. This is specific to severe cases.

From the publicly available records of dengue cases, we can generate time series since the beginning of the 20th century. However, with very low confidence on numbers before 2000, due to the lack of definition of cases symptomatology, mandatory notification and a well established system.

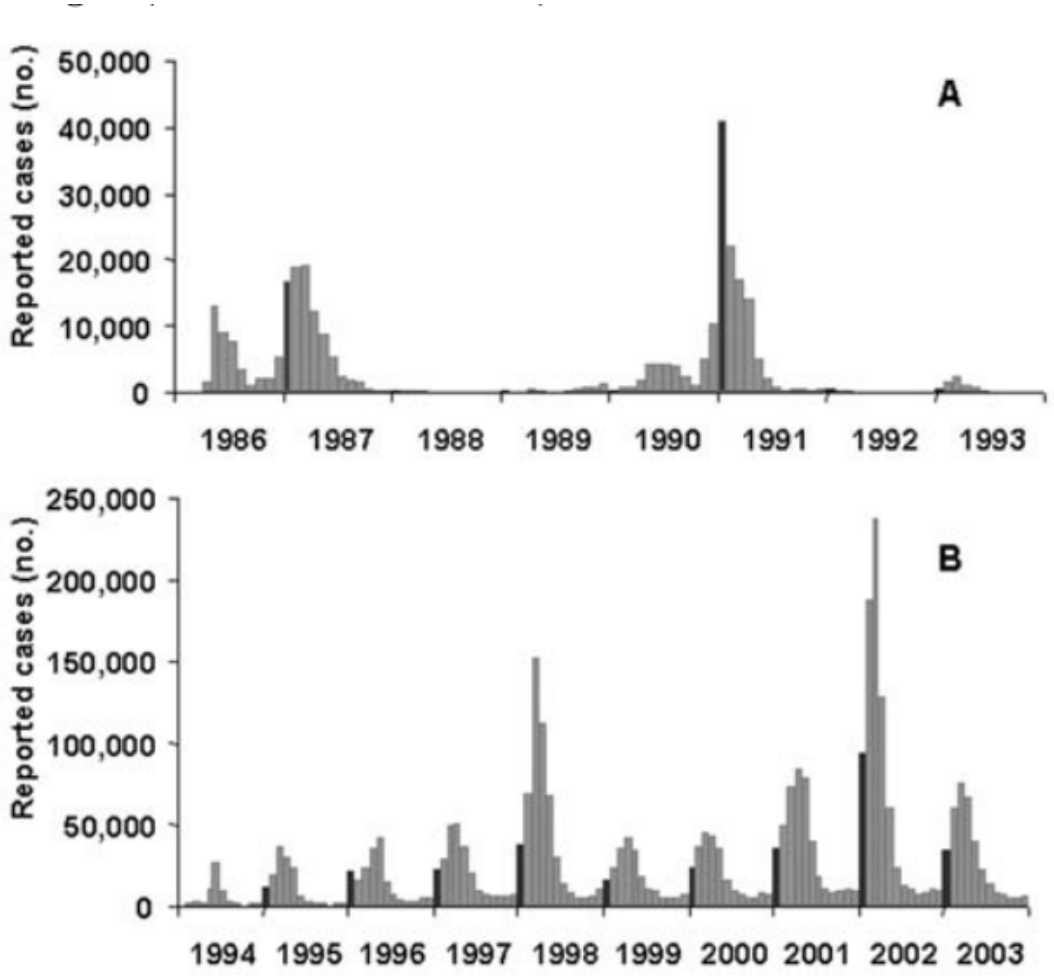

Figure 1.1: Time-series of dengue cases recorded before 2001, in Brazil. Extracted from: [1]. A) Cases reported from 1986 to 1993. B) Cases reported from 1994 to 2002.

As we can see, dengue it is a seasonal phenomenon even before the surveillance system that can monitor the disease properly was established. From the figure above, as in *Siqueira et al.* [1], we note that dengue has a raising presence throughout the two decades before 2001 in Brazil, more specifically with a clear increase of incidence since 1994. Reasons regarding how this has happened, after the local elimination of the vector, as well as with the dengue virus in past times, are not clearly understood [4] and are still a matter of present research [38].

Besides the lack of confidence on those numbers concerning cases before 2001, it is a well established and well known fact that since mid-years of the 1980s decade dengue has been increasing its presence in Brazil, with probable imports of propagules from Africa and/or Asia [37, 1]. Historical epidemics of dengue can be divided in two distinct periods, *Siquiera et al.* [1]. From 1986 to 1993, dengue was an episodic disease with localized epidemics, with strict outbreaks at only some locations. After 1994 and until the end of the period analyzed in the article (2002) by *Siqueira et al.* [1], the characteristics of the epidemics has shifted to a recurrent epidemic, with the new establishment of the disease in Brazil. More recent

definitions give to dengue in Brazil a status of *Hyperendemic*, an endemic disease with very stiff peak right after the rainy season, over large parts of the Brazilian territory, establishing dengue as a reoccurring health problem of national scale[39, 8, 17].

1.1.2 After 2001

Since 2001, Brazil has established a national surveillance system, allowing for a better surveillance on dengue cases. The dengue surveillance (SINAN-Dengue) system is part of the National surveillance system to monitor and survey the mandatory notification diseases. *Sistema de Informação de Agravos de Notificação* (SINAN) that counts and keep the information for more than 48 health grievances, ranging from the most present infectious disease in Brazil, like, Dengue, Zika, Measles, Yellow Fever, HIV, to grievance like Domestic Violence, Transit accidents, etc.[40].

Dengue is included on this list of health grievances which are to be reported on the national system, in a compulsory way, since 2001. This inclusion established, after a decade of monitoring[1] and the clearly increasing of the presence of the disease, a national surveillance system, that captures mild and symptomatic cases of the disease[34, 41]. With this inclusion, and now with a history of the reporting of the disease, we can figure out better on how dengue has been one of the most incident infectious disease[33, 38] in Brazil.

From 2001, the pattern of dengue epidemics over different regions of the country could be better observed, with varying incidence for different years on different region, with dengue expanding over the territory [17, 42]. This can be visualized through the time series of in Figure 1.2 the dengue cases for each Macro Regions of Brazil (North, Northeast, Center-West, Southeast and South).

Figure 1.2: Daily time-series of dengue cases recorded after 2001 by each Macro Region.

A first conclusion that can be drawn from Figure 1.2 is that the peaks of dengue cases appear to be approximately synchronized over all regions. Works investigating this phenomenon have been produced [42]. Another conclusion that can be drawn from the above Figure 1.2 is the clear decrease of the number of cases during the years of 2017 and 2018. It was suggested by Brito et al.[43] that this decrease is in part explained due to the surge of Zika virus. Zika virus has been introduced probably in years before 2016 and was in cryptic transmission, [44, 43], until outbreaks of 2017 and 2018.

Another point worth of attention is the expansion of dengue virus transmission over regions that had not experienced endemicity of the disease before[38, 21]. The Southern region presented a relatively low number of cases of dengue before 2010. After that year, the region has seen a greater number of cases, as shown in Figure 1.2. In a recent work, *Lee et al.*[21] it was shown that the increase of temperature, essential for the transmission, over Brazil has favored conditions for dengue epidemics in Southern Brazil, and more specifically the state of Rio Grande do Sul, the southernmost state. This agrees with similar studies, although through different modelling approaches, that have shown that there is a relation between R_0 of the dengue and temperature, with an optima for the transmission[18, 19, 22]. The change on the average mean temperature to more suitable temperatures for the transmission can be the explanation to this expansion of the range where dengue epidemics occur. Right after the

recrudescence of Zika, the Southern region has experienced its record epidemics, even with a probable sub-notification on the numbers due to the Covid-19 pandemic[38]. From InfoDengue[35], a national platform for monitoring of dengue activity, the southernmost state, Rio Grande do Sul, has been experiencing a very high number of cases in 2022.

In order to visually capture these different moments of the dengue epidemiological history since the establishment of the surveillance system, we calculate the incidence of dengue cases per 100,000 inhabitants, captured by *SINAN-Dengue*, over the population of each municipality of Brazil, by year. In summary, the mathematical formulation takes the form as follows:

$$I_{i,t} = \frac{C_{t,i}}{N_{t,i}} 10^5 \tag{1.1}$$

Where $I_{i,t}$ it is the incidence at municipality i at year t, $C_{i,t}$ is the amount of cases on the municipality i at year t and $N_{i,t}$ is the population projection for the municipality for that year. Here were used the estimates for resident population by municipality along the period of twenty years (2001-2020) taken from the data service of SUS, dataSUS. Those estimates for the resident population can be found here

We plot the incidences calculated binned by quantiles to help with the visualization, shades of red represents greater position in the incidence quantiles, shades of blue represents lower position in the quantiles distribution:

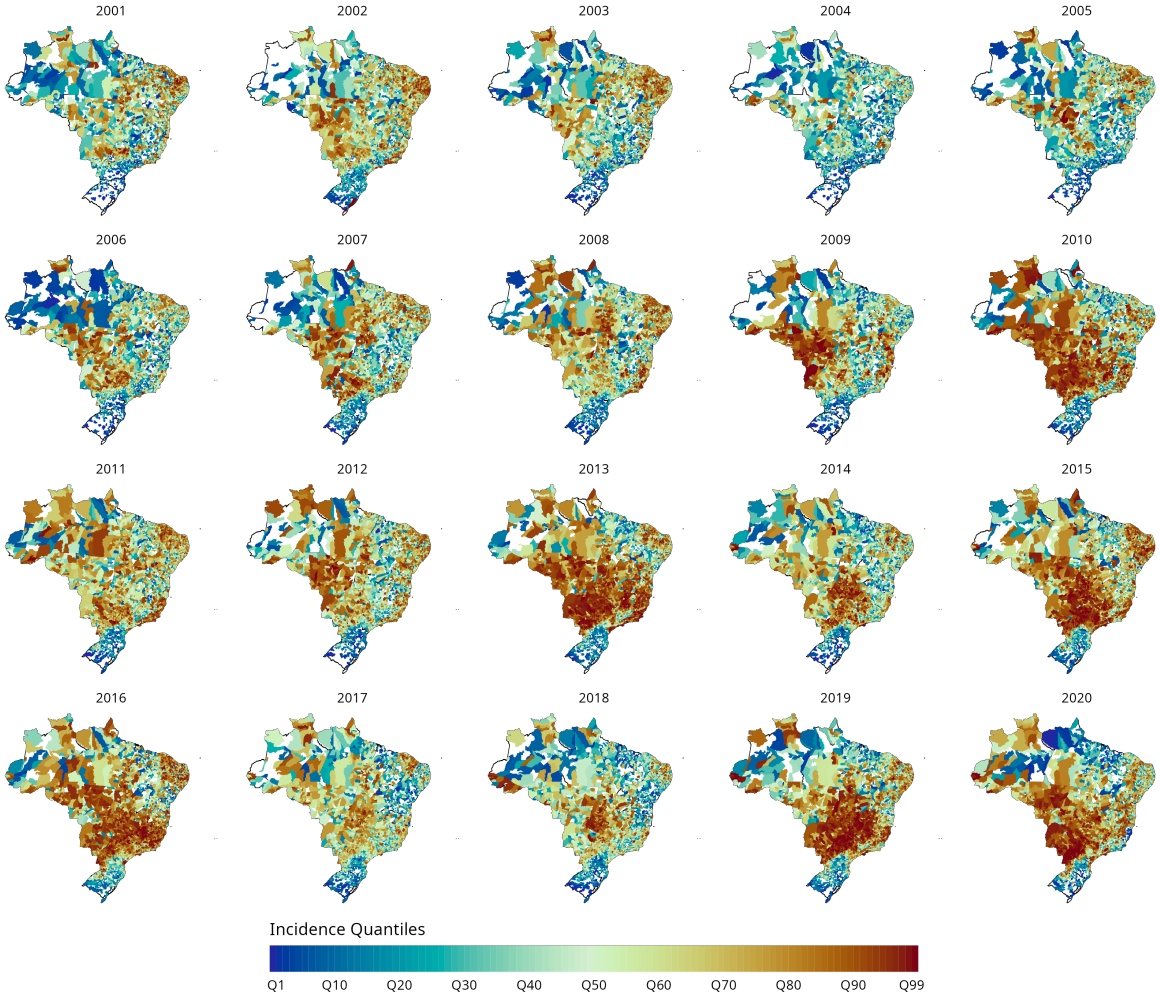

Figure 1.3: Quantiles of dengue incidence over municipalities, by year, for the period of 2001 to 2020. Blank parts are municipalities without any registered cases for the year.

A fact that can be assessed straightforwardly from these maps, Figure 1.3, is that dengue has an incidence varying greatly through geographical and temporal scales. This introduces complexity issues to any modelling approach which will require integration with data not captured by the notification system, as would be the case of climatic data.

A consequence of this wide spectrum of dengue in Brazil is the necessity of, at least, two levels of strategies to control the disease: local and regional. Local strategies have to rely on understanding the factors that drive and influence dengue at the level of a community. The surveillance service has to be aware of the potential of the spreading of outbreaks even of small scales, and should take action to decrease risks. Strategies on the regional level have to make use of the very capillary nature of the public health system that Brazil relies on, and scale up to implement strategies that connect effects on a level beyond of the community.

The results shown in Fig.Figure 1.3 raise the question of what drives the spatial variability and the expansion of dengue in Brazil. Climate change has been invoked, as suitable temperatures

for dengue transmission are shifting, as pointed out by *Lee et al.*[21]. However, the combined effects of variability of precipitation and droughts[20] are factors that might also impact the dynamics and the patterns of dengue in Brazil.

The societal component to this system can also be a variable influencing this new cycles of the disease. Brazil has passed through recent transformations in its economic profile, with particular emphasis in the decades of 00s and 010s[45], and this can have an effect of accelerating and changing social processes which influences dengue incidence.

1.2 Dengue and Climate

It is well-know that the mosquito life cycle is influenced strongly by the variability of temperature and by the abundance of water available in the ambient[3, 2, 18, 41, 19, 20, 21, 46, 47]. Higher temperature make the life cycle of mosquito shorter and egg-laying time decreases, enhances suitability of off-springs. In turn, this creates more viable females, which increase the number of bites and contribute to virus circulation [48]. Temperature modulates also the suitability of viral particles and increases the infectious bites during the hotter seasons [18, 19]. It can also modify human behaviors, increasing encounters with the mosquito population, and consequently, the number of humans being infected [49]. Precipitation has similar impacts on mechanisms that give rise to outbreaks and epidemics of dengue. With the higher occurrence of rains and its changing intensity, a larger amount of breeding sites of the mosquito population can be created [50]. This has positive impacts on mosquito population and can play a role on seasonality as well as on the extent of dengue outbreaks[18, 22]. In dry regions, where availability of water is season-dependent, lack of rainfall frequently makes people store clean and drinkable water, which could favor the increase of the mosquito population[20].

Temperature and precipitation are the climate variables that are most frequently studied among other possible climate factors that have effects on incidence and variability of dengue. These factors can possibly play some influence on mosquito life cycle, human behavior or the interconnected conditions between these two populations, both directly and indirectly [51]. Variables like relative humidity, atmospheric pollution, environmental degradation due to human exploitation are variables that could have some effect on the Dengue cycle by the exposing individuals to areas or situations with augmented risk of infection [51] but these effects are less well understood.

The purpose of this thesis is to understand and assess quantitatively how temperature and precipitation affect the number of dengue cases in certain circumstances that we will describe later. We adopt a so-called ecological approach, by working with data aggregated at a populational level. We have worked with time-structured data - time-series - and employ methods more commonly used in disease ecology and environmental epidemiology.

1.3 Methods

Dengue involves a very broad spectrum of scales, from the viral scale up to the planetary one, since it is influenced by the climate variables through many types of mechanisms[3, 2]. It is crucial to use adapted methods to obtain results at the required scale of the analysis. We will work with cases by municipalities of Brazil. In one study, we used the framework of Empirical Dynamic Modelling (EDM)[52, 53, 22], which comes from the theory of dynamical systems and from ideas connected to attractor reconstruction. In a second study, we studied dengue in the prism of epidemiology, and we make use of the framework of Distributed Lag Non-linear Models (DLNM) to assess dengue hospitalizations and the effects of climate on it, again at the level of municipalities [54, 55, 56, 57, 58, 59].

The Empirical Dynamic Modelling (EDM) was initially developed for the study of ecological interactions [60, 53, 61, 62], where it was applied to study causal drivers from time-series data. EDM assumes that the observed dynamics is generated by a dynamical process described by differential equations. It uses Taken's theorem to reconstruct the attractor of the system, [63, 64, 52]. It has been successfully applied to study climate drivers of dengue, malaria and influenza [22, 65, 66]. The aim of applying EDM on dengue time-series and climate variables is primarily inspecting which climate variables are the main drivers of dengue cases from a perspective that goes beyond purely correlational studies.

Distributed Lag Non-linear Models (DLNM) are used in epidemiology as a framework to study the effects of a given climate variable on an outcome from epidemiological interest[54, 55, 56, 57, 58, 59]. DLNM was mostly applied to quantify the effects of temperature on mortality [67, 68, 69]. Nowadays, DLNM has been widely used in epidemiological studies, with a larger variety of climate variables being questioned about their relation to outcomes of interests, for example, density of particulate matter in air, relative humidity, and others. The idea behind the framework is to give to any statistical model, like GAM, GLM, and some GNMs, a grid basis of the climate variable whose effects on the outcome are being investigated [70, 71, 72, 67]. To construct this grid basis, DLNM uses a bi-dimensional combination of splines, to account for the effects of the climate variable itself, the doses-response structure, and the lag effects of this climate variable [54, 55]. This grid basis is used to generate co-variables to be passed to the model, which are now explanatory variables for the outcome. It is thus possible to compute not only the association of the climate variable with the variable of epidemiological interest, but also the effects after some time of exposure, a lagged effect. To understand the association of severe dengue infection with the climate itself, DLNM is the suited tool for it. When planning public health policies, policymakers can rely on the results of how temperature can aggravate an ongoing infection of dengue, making it possible to plan strategies to mitigate the burden and the pressure of the disease on the health system.

1.4 Thesis Organization

This thesis is divided into four parts, and each of them is organized in chapters. After the conclusions chapter, there are two appendices, Appendix A and Appendix B, each of them developing the theoretical frameworks used on the core chapters of the thesis. Those two

appendices are part of the of whole body of the thesis, they were put on the end to help on the fluidity of the main text and the core of the thesis, which is the chapter 3 and chapter 4. After these two appendixes, there are another two appendixes with supplementary material to the chapters 3 and 4.

The first chapter is the present one, an introduction to the subjects of the thesis. The next one, chapter 2, is divided into two parts, the first one about the historical background of dengue disease in Brazil and its surveillance system. The second part of this chapter concerns issues of organizing the databases of dengue cases records and climate databases. A discussion is presented on how this work has been developed and how relevant it is to the subsequent research. It refers to the compilation and organization of the databases of dengue cases and hospitalizations in Brazil during a period of 10 years, from 2010 to 2019. Although we collected data of dengue that extend to more than 20 years, from 2000 to 2020, we focused on studying the last 10 years.

The third chapter 3 is dedicated to dengue ecology and its dynamics under climate drivers. We use data from the city of Rio de Janeiro, both epidemiological and environmental. We analyzed how the climate variability drives the dengue cases time-series, and which drivers are more important. The chapter is divided in sections for methods, results and discussion. To this chapter, we add an appendix with a broader explanation to the Empirical Dynamic Modelling (EDM) framework, Appendix A. Another appendix, C, contains supplementary material to this study.

The chapter 4 is dedicated to the study of the association of temperature and the dengue hospitalizations in the framework of Distributed Lag Non-linear Models (DLNM). As in the previous chapter, the fourth chapter is divided in sections for methods, results and discussion. Again, we restricted the study to a period of 10 years, from 2010 to 2019. There is an appendix to this chapter, Appendix B, where the theoretical development of the DLNM framework is presented and explanations and references on the DLNM framework are given. The Appendix D presents more materials relevant to this part of research.

The fifth and last chapter, chapter 5, is dedicated to the conclusions and discussions of the results, their interpretation in a wider context of dengue as a complex system and the implications to the public health measures. Finally, the thesis closes with some perspectives, presenting future avenues of research.

2

Data Wrangling of Epidemiological data and Climate data

Data are multidimensional objects that partially reflect the dynamics or the state of an object of interest, whose characteristics may involve unknown factors. For instance, epidemiological data come from a system which is comprised by medical, societal and environmental factors, which are not explicitly described. Incidence and prevalence, to cite an example, are common epidemiological indicators, but they result from a complex set of interactions which are not usually known quantitatively.

Every time we choose to wrangle data in any direction, aspect, or perform any operation that reduces the multidimensional nature of data into a more flexible lower dimensional object, as a graph, a plot or a time-series, we make a choice that poses constraints on what and how much information can be leveraged through this new object. This process is basically the process of data wrangling, which can be understood as the process of re-modeling, summarizing, cutting, selecting, filtering data which, as a result, give a new object as an output, amenable to analysis. Any conclusion about specific data is always drawn from a reduction of data to a form that allows evaluations and calculations. This process is now usually called data analytics.

One of the first difficulties when studying dengue epidemics is its data sources. Dengue is an infection with a wide spectrum of symptoms and levels of severity [36]. Before any study, we have to select which aspect is being discussed, and choose relevant data accordingly. The same occurs with climate. Climate is a very complex system, and the science of climatology produces a wide variety of measurements and estimations [73]. It is important to understand which are the adequate sources of climate data when studying dengue. Research on climate drivers of dengue needs careful preparation of data.

In this chapter, we present the sources of dengue data, cases and hospitalizations, and data sources of climate variables as a part of the work itself.

This chapter is dedicated to this work of preparation of data and has the following organization: a first section is dedicated to dengue epidemiological data sources, its particularities and our approach when organizing it for analysis. The second section is about the climate variables used in this thesis and their compilation, storage and organization. We made use of publicly available data sources for dengue data and climate variables. Dengue

data is anonymous and do not keep any identifiers of any individual.

2.1 Epidemiological Data Sources

As mentioned above, dengue has a proper notification system that covers the whole Brazilian territory. Since 2001, every symptomatic case of dengue which is identified should be notified to SINAN-Dengue [32, 14, 40]. This creates a surveillance system for the incidence of the disease and a historical time-series of the case counts. Due to common problems with notifications systems, not all cases are communicated and the complete extension and burden of the disease on the Brazilian population [34] remain uncertain.

Dengue can manifest different symptoms that can be confounded with a variety of others diseases, but it can sometimes show no symptoms - depending on previous exposure. Also, it can manifest severe symptoms [36].

These are some limitations of the dengue data records in the SINAN-Dengue system, but they do not invalidate the use of the data when researching dengue. Limitations decrease the power of conclusions that can be obtained from data, but being aware of them helps to better gauge the scope of the analysis.

2.1.1 SINAN-Dengue data

We obtained and organized data for dengue from the public accessible official records of dengue available at SINAN (Sistemas de Informaçaão de Agravos de Notificação). SINAN has modules for each of the diseases listed on the national mandatory notification grievance list [40]. These modules are sometimes called as SINAN and the name of disease it is specific for, like SINAN-Dengue. From here on we will refer to any data on SINAN about dengue cases as SINAN-Dengue, and sometimes we use only SINAN, being understood that it is about dengue data, unless it is explicit stated otherwise. The publicly available data for dengue can be downloaded in several ways: it can be ordered by the national service for information to the citizen (e-SIC)[74], it can be downloaded directly from the file transfer protocol (FTP) service from the SUS data service (DATASUS) [75] or can be downloaded through the wrapper functions present on the R package 'microdataSUS' [76]. The data used on this thesis are an equalization between data collected from e-SIC order and data collected from the FTP service, as well as by the 'micradataSUS' R package. This work of organizing all this data into an easily usable form is made available on our repository and can be accessed here: https://github.com/rafalopespx/Dengue_time_series

The notification criteria on SINAN-Dengue comprises any person alive and that has traveled in the last 14 days to areas of dengue transmission and that has/had fever, usually between 2 and 7 days, and presents two or more of the following symptoms; nausea, vomiting, rash, myalgias, headache, retroorbital pain, petechiae or positive loop test and leukopenia[77]. Any person who matches this definition and has sought for medical care, is reported as suspected case of dengue. Afterwards, the case is investigated to give a definitive diagnosis, discarding or confirming it as a dengue infection case. The investigation is done by several different means. A case can be investigated through laboratory exams, clinical investigation or epidemiological

investigation. At the end of the process, the case has to be discarded or confirmed [78, 32, 78].

This case definition fulfills the definition of cases from WHO [33]. The capillarity of the universal health system (SUS) in Brazil helps SINAN-Dengue to capture symptomatic cases of dengue [79]. This does not mean that there is no sub-notification on cases, as big part of dengue cases can be oligosymptomatic or assymptomatic, which poses a difficulty for the notification system to capture those cases. Since the establishment of SINAN-Dengue, this has been an issue to the notification system.

We have accessed the totality of suspected cases notified in SINAN-Dengue. We identified the case in categories by its final classification, being confirmed as 'Classic Dengue', 'Dengue Hemorrhagic Fever', 'Dengue shock syndrome', 'Dengue with alarming signals', 'Discarded' and 'Severe Dengue'. This is done due to the always revising definition of severe cases and its symptoms. To better capture those cases, the definition has being changing since the establishment of the SINAN-Dengue. This encompasses more than 20.8 millions registers from 2001 until 2020. From those cases, 52% are confirmed cases of dengue infection with any kind of severity, the rest are classified as 'Discarded' explicitly or do not have the information about the final classification, which are considered as 'Discarded' and are filtered out. The comparable dengue cases numbers are those classified as 'Classic Dengue'. The cases classified as 'Dengue Hemorrhagic Fever', 'Dengue Shock Syndrome' are classifications that used to exist in the system until 2011, after this the system only allows classification as 'Dengue', 'Dengue with alarming signals' and 'Severe Dengue'[77].

The cases were aggregated by the municipality of residence and by date of first symptoms, creating time-series for the 5570 municipalities of Brazil over a period of 21 years, from 2001 to 2020. This is the first product of data wrangling over the databases of SINAN-Dengue, and it is stored in our online repository https://github.com/rafalopespx/Dengue_time_series. Below, we show a table characterizing the typical data that is contained in our aggregation of all these databases.

Table 2.1: Characteristics of notified cases of dengue, 2001 to 2020.

Characteristics Variables	Brazil, N = 11,060,996	Region				
		North, N = 731,561 (6.6%)	Northeast, N = 2,222,171 (20%)	Center-West, N = 1,712,121 (15%)	Southeast, N = 5,781,636 (52%)	South, N = 613,507 (5.5%)
Age	31 (19, 46)	27 (16, 41)	27 (16, 42)	31 (19, 45)	32 (20, 47)	34 (21, 50)
Age in Categories						
0-1	218,382 (2.0%)	19,038 (2.6%)	67,926 (3.1%)	36,196 (2.1%)	87,045 (1.5%)	8,177 (1.3%)
1-9	784,185 (7.1%)	69,473 (9.5%)	229,259 (10%)	117,248 (6.8%)	334,779 (5.8%)	33,426 (5.4%)
10-17	1,511,998 (14%)	111,217 (15%)	335,665 (15%)	227,790 (13%)	761,003 (13%)	76,323 (12%)
18-39	4,766,032 (43%)	336,685 (46%)	963,253 (43%)	759,275 (44%)	2,461,554 (43%)	245,265 (40%)
40-59	2,749,271 (25%)	149,595 (20%)	460,406 (21%)	426,089 (25%)	1,538,847 (27%)	174,334 (28%)
60-79	936,741 (8.5%)	41,522 (5.7%)	148,082 (6.7%)	132,733 (7.8%)	545,222 (9.4%)	69,182 (11%)
80+	94,387 (0.9%)	4,031 (0.6%)	17,580 (0.8%)	12,790 (0.7%)	53,186 (0.9%)	6,800 (1.1%)
Self-Reported Race						
Black	414,943 (5.8%)	23,501 (4.4%)	74,484 (5.6%)	56,599 (4.8%)	237,806 (6.7%)	22,553 (4.0%)
Brown	3,213,398 (45%)	394,560 (74%)	936,463 (71%)	612,034 (52%)	1,165,690 (33%)	104,651 (19%)
Indigenous	27,054 (0.4%)	4,963 (0.9%)	5,719 (0.4%)	6,617 (0.6%)	9,032 (0.3%)	723 (0.1%)
White	3,410,059 (48%)	101,103 (19%)	286,163 (22%)	475,940 (41%)	2,123,302 (59%)	423,551 (76%)
Yellow	87,235 (1.2%)	7,053 (1.3%)	16,751 (1.3%)	18,384 (1.6%)	39,639 (1.1%)	5,408 (1.0%)
(Missing)	3,908,307	200,381	902,591	542,547	2,206,167	56,621
Sex						
Female	6,189,428 (56%)	394,800 (54%)	1,285,603 (58%)	947,284 (55%)	3,216,515 (56%)	345,226 (56%)
Male	4,871,568 (44%)	336,761 (46%)	936,568 (42%)	764,837 (45%)	2,565,121 (44%)	268,281 (44%)
Confirmation criteria						
Clinical-Epidemiological	6,903,974 (63%)	492,680 (68%)	1,580,799 (72%)	1,091,593 (64%)	3,428,098 (59%)	310,804 (51%)
By Laboratory exam	4,044,384 (37%)	232,909 (32%)	588,293 (27%)	608,206 (36%)	2,314,157 (40%)	300,819 (49%)
Under investigation	78,352 (0.7%)	2,505 (0.3%)	39,126 (1.8%)	10,227 (0.6%)	26,075 (0.5%)	419 (<0.1%)
(Missing)	34,286	3,467	13,953	2,095	13,306	1,465
Case classification						
Classic Dengue	10,874,371 (98%)	722,355 (99%)	2,186,425 (98%)	1,668,328 (97%)	5,690,819 (98%)	606,444 (99%)
Dengue Hemorrhagic Fever	23,675 (0.2%)	2,321 (0.3%)	8,699 (0.4%)	3,587 (0.2%)	8,764 (0.2%)	304 (<0.1%)
Dengue shock syndrome	939 (<0.1%)	138 (<0.1%)	320 (<0.1%)	117 (<0.1%)	335 (<0.1%)	29 (<0.1%)
Dengue with alarming signals	155,066 (1.4%)	6,543 (0.9%)	25,438 (1.1%)	38,383 (2.2%)	78,629 (1.4%)	6,073 (1.0%)
Severe Dengue	6,945 (<0.1%)	204 (<0.1%)	1,289 (<0.1%)	1,706 (<0.1%)	3,089 (<0.1%)	657 (0.1%)
Delay of Notification	3.0 (1.0, 6.0)	4.0 (2.0, 6.0)	3.0 (1.0, 7.0)	3.0 (1.0, 6.0)	3.0 (1.0, 6.0)	3.0 (1.0, 5.0)
(Missing)	15	0	3	2	8	2
Outcome						
Death by Dengue	8,722 (<0.1%)	542 (<0.1%)	2,364 (0.1%)	1,667 (<0.1%)	3,723 (<0.1%)	426 (<0.1%)
Death by other causes	1,691 (<0.1%)	167 (<0.1%)	302 (<0.1%)	323 (<0.1%)	715 (<0.1%)	184 (<0.1%)
Death under investigation	710 (<0.1%)	14 (<0.1%)	205 (<0.1%)	175 (<0.1%)	309 (<0.1%)	7 (<0.1%)
Discharged[1]	11,049,873 (100%)	730,838 (100%)	2,219,300 (100%)	1,709,956 (100%)	5,776,889 (100%)	612,890 (100%)

[1] Missing are considered as Discharged

From the table above, 2.1, we have a general view of the typical characteristics of dengue disease in Brazil over the period studied. From all notified cases in SINAN-Dengue, the median age of individuals for a dengue case is 30 years old. But this is not homogeneous throughout Brazil, with lower values in North and North-East regions.

Data about self-reported race are largely affected by lack of reporting. About 4 millions cases have no self-reported race information, constituting a chronic problem for the notification system. We can also observe differences due to specific demographics of Brazilian regions. We can see differences on the cases notified per self-reported race at North, Northeast and Center-West regions different of the rest of the country, as these regions have a majority of dengue notified cases on individuals that are self-reported as browns. With about 74% percentage on North and 71% Northeast regions and 52% percent at Center-West region, this differs from the whole country status for self-reported race, which is at 45% percent. For the South and Southeast regions, individuals self-identified as white are majorities in each region of the notified dengue cases. For the South and Southeast regions, the second most notified self-reported race on the notifications are individuals self-identified as browns.

Incidence on female individuals is 56% of the notified cases at country level. This pattern presents little difference between regions.

Notifications always present some delay with respect to the onset of symptoms. We propose a metric to understand the delay of notification of cases. Knowing the date of onset of first symptoms and the date of notification at the SINAN-Dengue system, we calculate, in days, the delay between those dates. This comprises delays coming from the fact that patients seek health services only after some days of symptoms onset to delays on the inputting process of the data. We have a delay between date of onset of first symptoms and date of notification of 3 days, in average, with a variance from 1 to 6 days. In all regions, except the North region, delay in the notification follows almost the same distribution. North region has the greatest delays for notification, with 4 days for the median of the distribution of days passed between date of first symptoms until date of notification.

Finally, the last line of 2.1 gives us the cases that resulted in death. All notifications with an explicit code filled for the outcome of the case as a death are counted as a case that evolved to deceased. Most cases evolve to a discharged outcome. The whole country has, in general, less than 0.1% of the total number of dengue cases notified with an outcome of death. There are some regional differences, but most regions follow the same pattern for the outcome of cases as in the whole country level. This, however, does not reflect the true mortality or lethality by dengue infection, necessarily, as the ability of the notification system to capture all the dengue cases varies greatly over the country. Some studies point to sub-notification of mortality due to dengue [34].

2.1.2 SIH data

Due to the problem with capturing the severe cases of dengue in databases, when we investigate other specific outcomes and more severe manifestations of dengue, we have to look to another notification system. For the results presented in chapter 4 we made use of Hospital Information System, (SIH). SIH is a system developed initially for administration purposes, but it can be used for epidemiological interests too. The SIH is a notification system

which notifies hospitalizations that are covered by the SUS. Any hospital facility can notify a hospitalization on the system and from it, it can apply to SUS for payment due to this hospitalization, if the disease is of mandatory notification. Dengue is the case, so from SIH total notification on all hospitalizations, we can filter out dengue hospitalizations, and we can get a better view of hospitalizations due to dengue [34].

We present at chapter 4 a table 4.1 of the notification data on SIH which was used to study of dengue hospitalizations and the temperature effects on it. This table is constructed by the cumulative hospitalization notifications for each region of Brazil. It is known that SIH is more sensitive to capture the more severe cases of dengue. From the totality of notifications on the system, we filter out only those with specified codes for different severe outcomes related to dengue [34].

2.2 Climate Data Sources

Throughout this thesis, we made use of climate data sources. Here we give a more detailed view on this type of data sources and how they were managed and compiled in each of the analysis chapters, 3 and 4. There is a very broad and a very diverse range of products from climate models and climate automatic stations that collect data about several climatic measures. The use of these products in another fields of study, out of climatology and meteorology, has been growing through the recent years. Together with this increasing interest in other sciences to make use of climate products, there is also a crescent sort of newer climate data products more suitable for each area of application. In environmental epidemiology this matters and the use of these products gives momentum to the field in advancing a better understanding of climate change impacts and its consequences for public health [80].

We will use ERA5-Land reanalysis products, which is the state-of-the-art in reanalysis products from the European Center for Medium-Range Weather Forecast (ECMWF)[81]. ERA5-Land has the narrower resolution of (0.1° X 0.1°) among the width used in reanalysis products from (ECMWF), ERA5 and ERA5-Land [81, 82, 73].

Shortly, reanalysis products are constructed from available data from automatic historical records throughout the world and with this data the climate model is initialized. The model produces gridded measures for any essential climate variable (ECV) of interest [81, 82, 73, 83, 84]. These products have been validated and have been applied beyond climatology and meteorology. The studies of environmental epidemiology have used these results extensively [72].

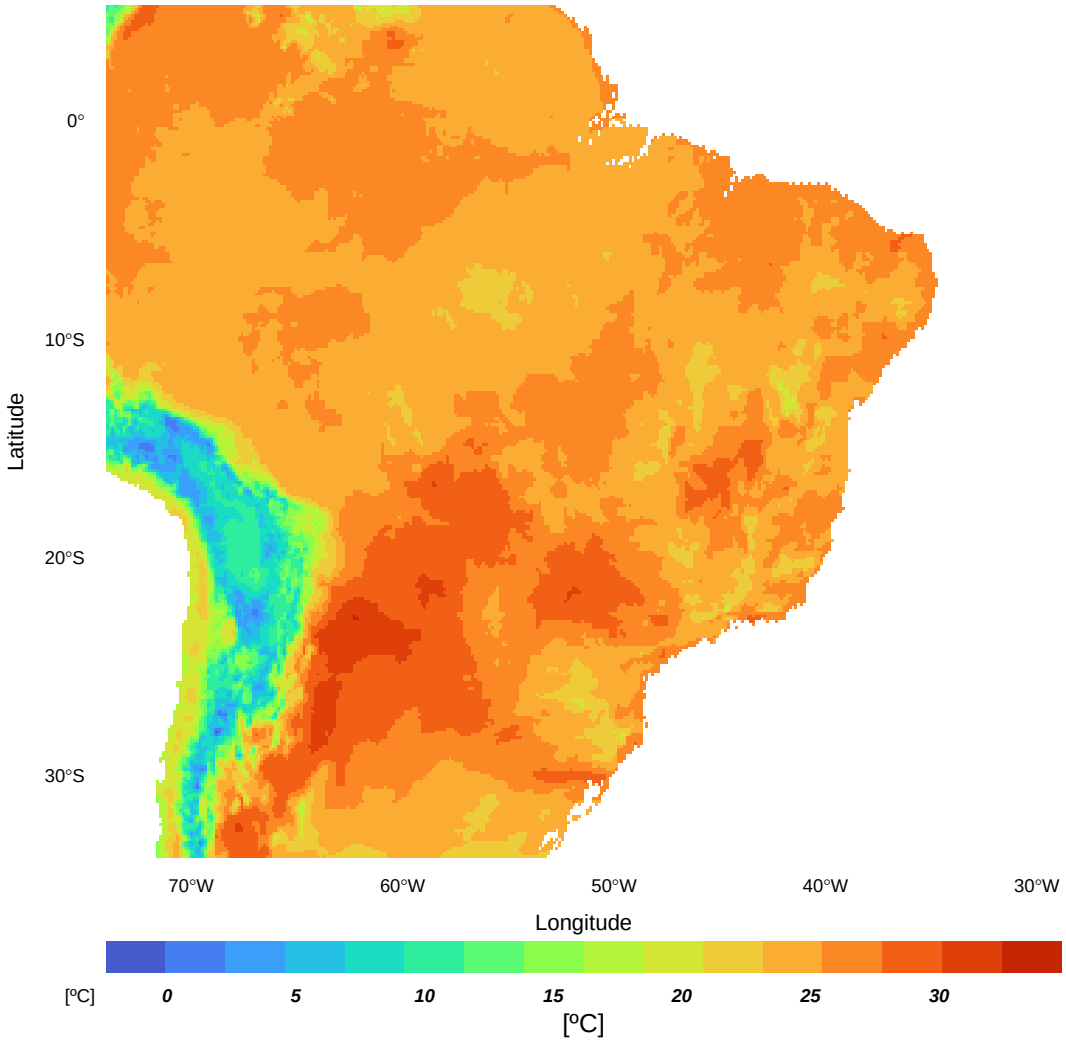

Figure 2.1: Raw Gridded data for temperature at 2 meters height, ERA5-Land, at January 1st, 2020

For the purposes of our studies, we construct, from the raw gridded data, aggregated data that can be related to the time-series of dengue cases. Our studies made use of epidemiological data aggregated at the municipality level and ERA5-Land dataset for temperature and precipitation. The reanalysis products are generated at a scale of 0.1° x 0.1°, which is finer than the typical area of a municipality, as seen in Figure 2.1. By constructing weighted mean areas on the extent of each municipality, we can build a dataset of the weighted mean area climate variable over Brazil. The raw data accessed from the ECMWF comes in .nc format, which is a specific format for geospatial temporal data, as is the case for most climate variables. The ECMWF publishes data in several formats and aggregations and can be assessed freely at https://cds.climate.copernicus.eu/.

We choose to use the data at an hourly basis, and we construct weekly weighted mean areas time-series for all municipalities of Brazil on the period ranging from 2010 to 2019, which is the analysis period of both studies presented in the chapters 3 and 4. To construct the time-series of weighted mean areas by municipalities, we made use of the functions presented on the R package 'exactextractr' [85]. We created new functions to calculate weighted mean areas over the extension of each municipality of Brazil according to the 2010 census [86]. To do so we get the files with the shapes of each municipality of Brazil which are available at the IBGE FTP server through the R package 'geobr'[87]. The result from this weighted mean areas for the creation of time series over the extent of municipalities, is an array in daily resolution by each year over the period used, from 2010 to 2019, for the climate variables. They are organized and are available on the online GitHub repository: https://github.com/rafalopespx/climate_aggreg/.

The result is a dataset of time-series of climate variables aggregated by each municipality for the 10 years period of the studies. To give a glimpse and to visualize part of this product, we plot, for the month of October 2020, the mean temperature above 2 meters from the ground. This visualization is basic to the work of harmonization produced with the compilation and aggregation done with the weighted mean areas, ERA5-Land data and 'shapefiles' from 'geobr'. Below, we present a possible visualization of such files.

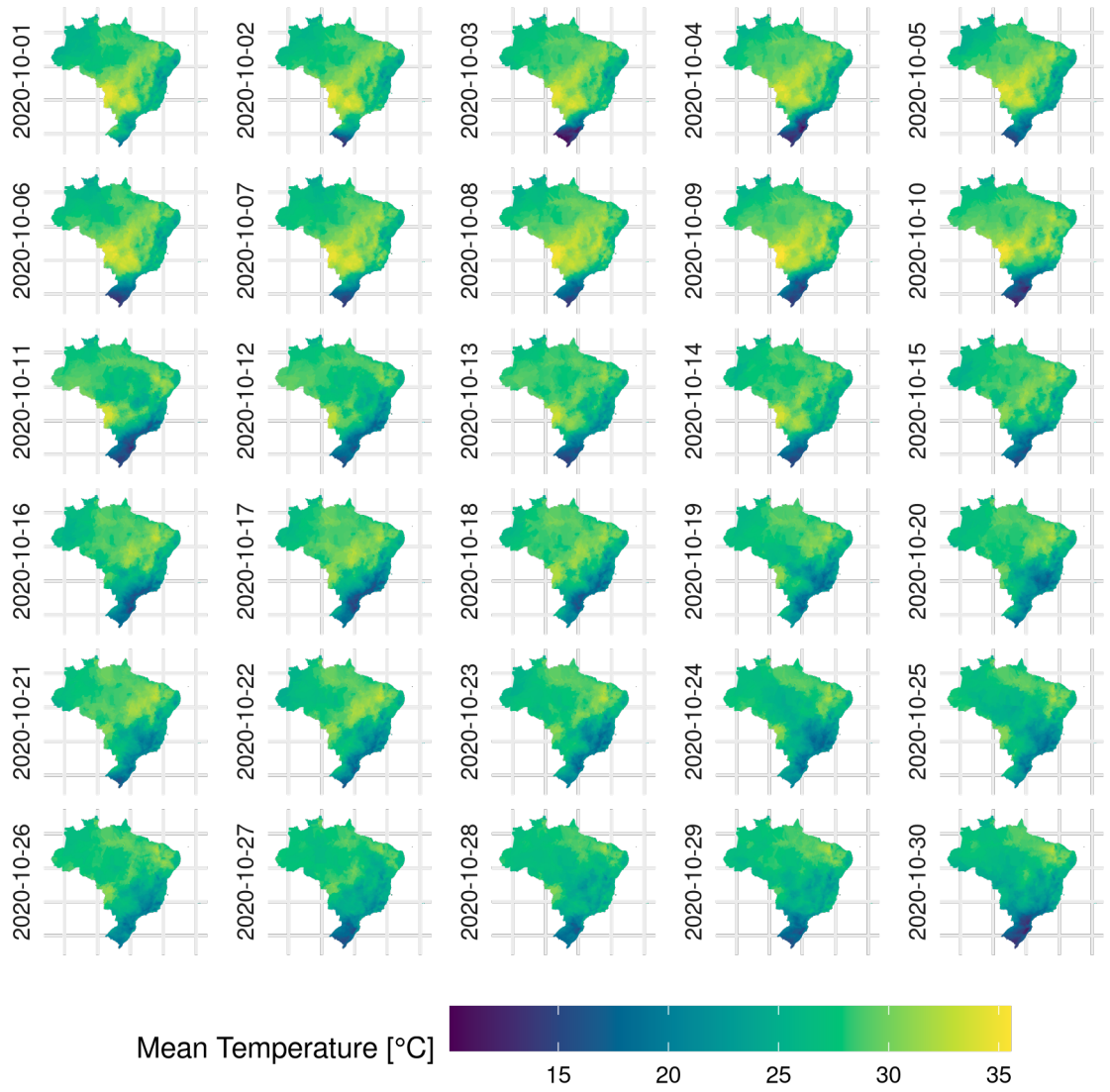

Figure 2.2: Mean temperature at 2 meters height above ground by weighted mean area in Brazil, October 2020

In this thesis, we have made use of temperature and precipitation measures that were aggregated to produce the weighted mean area of interest. The temperature is a typical meteorological measure for the two-meters above the ground column of air temperature. We have built the weighted mean area average as described above for each municipality of Brazil over the period from 2010 until 2019. To the precipitation, which is the measure of any kind of water precipitation, which could be ice precipitation, water precipitation, namely rainfall, we produced the weighted mean areas for each municipality. For precipitation, we have only daily measures and from it, we generate time-series for each municipality of the minimum amount of rainfall of the week, by taking the minimum value of daily precipitation. Analogously we generate the weekly maximum precipitation, by considering the maximum amount of precipitation of the week, and finally we generate the mean precipitation, by simply taking the arithmetic mean of daily precipitations of the week.

Part II

Dengue epidemics and climate

3

Climate drivers of Dengue at Rio de Janeiro city, 10 years period, from 2010 to 2019, an empirical dynamic modelling study

In collaboration with: Brenno Caetano Troca Cabella, Renato Mendes Coutinho[1], Roberto André Kraenkel[2]

3.1 Introduction

Dengue epidemics may present a multi-factors and multivariate combination of explanatory variables, and it is sometimes seen as related to a complex dynamical system. It has been considered as a hyper-endemic disease[3, 5, 39, 21, 28, 88, 8, 7, 13, 36], which adds more dimensions to the dynamical system. Climatic drivers have been connected to dengue epidemics [19, 22, 47, 50, 89], but still pose challenges in dengue modelling studies. In this chapter, we will study climatic dengue-epidemics drivers from the point of view of dynamical systems, along the lines of [60, 53, 90, 66]. This implies that we will consider dengue-cases times series as emerging from a deterministic dynamical system that can be described by a set of differential equations, which is subject to noise and external drivers. However, contrary to the usual approach of postulating a set of equations, we will use the dengue-cases time-series to reconstruct the attractor of an unknown system via the Taken's theorem. [63, 52]. For this study, we choose the city of Rio de Janeiro, the second major city of Brazil, which has a history of recurrence of tropical diseases, particularly arboviruses, and a high incidence of dengue cases.

Dengue cases are usually supposed to be, at least partially, driven by climate factors related to the mosquito life cycle [3, 2, 5, 89, 20, 91, 92]. However, little is known of which specific climate drivers and their relative importance in explaining the time-series of dengue-cases. We will use the framework of Empirical Dynamic Modelling (EDM) [53, 52], to provide a

[1]Centro de Matemática, Computação e Cognição, Universidade Federal do ABC, Santo André, Brazil
[2]Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, Brazil

quantification of the effects of climate drivers on the dengue cases curve.

The main idea behind EDM, is that we can reconstruct the attractor of a dynamical system from only one variable of the system, via Takens theorem, [63, 64], which allows for state-space Reconstructions (SSR): with only one time series, named as observation or observation function of the system, and a finite amount of delayed time-series of this observation of the system, we recover the attractor underlying the original dynamical system section A.1.

The EDM uses Takens Theorem uses the consequences of Takens theorem: if we can reconstruct an attractor from a time-series of a single variable, if we have two time-series, and they pertain to the same dynamical systems, then the attractor obtained by each of the time-series has to be the same, meaning that there should be a correspondence between them, a 1-to-1 mapping between both of them. A longer description of EDM framework, from the theorem itself and how the framework is constructed based on it, is given in the Appendix A. The city of Rio de Janeiro has a population of over 6.7 millions inhabitants, distributed over 1,200 km^2[93].

Rio de Janeiro has been suffering throughout its history with different epidemics[4], many of them related to the cycle of urbanization of the city[88]. Rio de Janeiro is one of the most important commercial hubs in Latin America and has a great part of its population in high incidence areas of dengue [49]. dengue presents seasonal characteristics in the city, but the number of cases has high variability from year to year C.1.

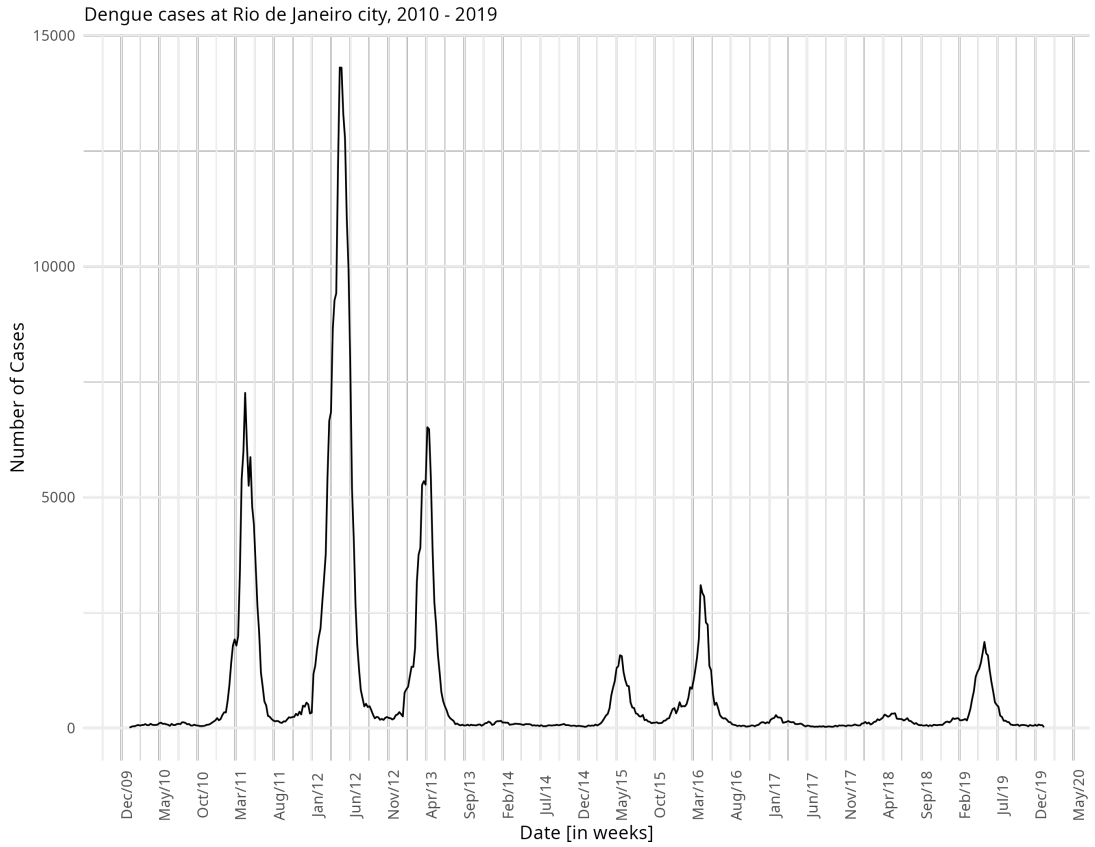

Figure 3.1: Time series of weekly cases of dengue notified in the city of Rio de Janeiro, from 2010 to 2019

We implemented Empirical Dynamic Modelling (EDM) framework to quantify and understand the effects of Temperature (Minimum, Mean and Maximum) and Total Precipitation (Minimum, Mean and Maximum) on dengue cases and how these factors can drive the variability of seasonal outbreaks of dengue. Our analysis spans a ten years period, from 2010 to 2019.

3.2 Materials and Methods

3.2.1 Data Sources and Definitions

Dengue Cases: We accessed the notifications of dengue cases in Rio de Janeiro city by filtering out the cases from the Brazil national database of for mandatory grievances notified diseases, SINAN-Dengue[32, 94]. We build weekly time-series by the date of onset of symptoms, filtering those cases with municipality of residence at Rio de Janeiro city C.1. We included only those cases classified as having any of the final diagnostics that matches dengue classification characteristics [14, 36, 88]. This resulted into a time-series of 10 years long for the period of 2010 to 2019.

Climate variables: The climate variables were assembled from the gridded (0.1° x 0.1°) reanalysis products presented on the Climate Data Store for the ERA5-Land, which are freely available by the European Center for Medium Range Weather Forecasts. The data is available at [73]. The data in its original format were in hourly measure over the extension of the municipality of Rio de Janeiro. We produced weighted mean areas averages using the R package *exactectractr* [85]. A more detailed explanation on how those weighted mean areas averages are produced can be found in the section 2.2

3.2.2 Data Analysis

We study climate drivers on the dengue time-series of cases through the EDM framework by dividing the analysis in three parts. The first part sets out to determine the optimal embedding dimension for the SSRs by using the simplex projection procedure Simplex Projection, which is the prediction of the time-series on itself for a series of tentative embedding dimensions. The embedding dimension which reduces the error between prediction and observations is set as the optimal embedding dimension E^* and which is chosen for the SSR. For each time-series analyzed, we calculated this optimal embedding dimensions.

In the second part, we study the significance of the driving relations found in the first part. To do so, we perform mapping between two different SSRs, called the Convergent Cross Mapping (CCM) procedure section A.3. CCM is the mapping between each SSR generated by a time series and an optimal amount of delays, chosen by the simplex projection. After having the optimal dimension of each time series and its SSRs, we perform cross mappings between this SSRs to discover driving relations. If the cross-mapping converges to a certain value, as we increase the number of points in the time-series, we say that this represents a driving relation. Convergent Cross-Mapping are viewed as a signature of causality, in the sense that the different time-series come as projections of a same attractor, and therefor have mutual

influences on each other. We can go further and give a directional sense to these mappings. CCM has to be interpreted in a kind of inverted direction: the time-series of the number of dengue cases contains information on the climate variable, and therefore dengue cases time-series can "predict"the climate time-series through cross-mappings, [63, 53, 60]. The section section A.3 presents a mathematical discussion of these procedures.

The next part of the analysis is a procedure to determine if the driver/causal relations found are not just an expression of seasonal drivings - which are known to be important. To do so, we create surrogates time series that have the same seasonal patterns of the original time series but different residuals parts. Our surrogates time-series are basically the shuffle of years in an order that is not chronologically correct. This gives the same seasonal pattern but puts on different patterns beyond the seasonality. If the driving relation is just because of seasonality, we expect not to see significance on CCM between the original time-series in contrast to the surrogates time-series. For those mappings that are statistical significant over the distribution of surrogates mappings, we attribute causal relation by the variability of the time-series itself and not only by the seasonality.

The third part implements a model to estimate the strength of the driving relations found in the previous step. It uses a mapping called S-Map. S-Map is a locally weighting estimate of the Jacobian elements on a multivariate SSR, constructed with the cases time-series and the drivers time-series that were attributed significant driving relations. We can infer how much each of the climate drivers are impacting the dengue cases curve. A more complete description is given at section A.4.

3.3 Analysis

3.3.1 Empirical Dynamic Analysis

Simplex projection: For each of the time-series of dengue cases at Rio de Janeiro municipality and the time0series of climate drivers - mean, minimum, maximum of temperature and precipitation - we performed a simplex projection for predicting a library part of the time-series on a prediction part of the time-series.The expression for the simplex projection is:

$$\hat{X}(t+n) = \sum_{i=1}^{E+1} w_i X_i(t+n) \quad , \tag{3.1}$$

where E is a tentative embedding dimension, up to 10. The embedding dimension that minimizes error between the observed and prediction is the optimal embedding dimension, E^* which is described as a correlation between the observed and predicted, constructed as in Equation A.9.

Seasonality extraction and Surrogates testing: After defining an optimal embedding dimension E^* for each of the drivers time-series being tested and the cases time-series, we can produce cross-mappings between driver time-series and cases time-series, to seek for causal relations. To give a statistical significance to this cross-mappings, we first perform this

cross-mappings between cases time-series and surrogate drivers time-series. The surrogate time-series were constructed with the same seasonality patterns of the original drivers time-series, but randomized residuals. We choose to only shuffle the residuals yearly, basically producing a time series that has the same variability of the original time series but with chronological order shuffled. If the original time series is statistically significant in the distribution of prediction skills ρ of all surrogate cross-mappings, we attribute it a status of causal driver to the cases time-series. On the other hand, if the original time series is not significant over the distribution of ρ of surrogates cross-mappings, we can not say more than the driving relation it is due to the seasonality pattern of the climate time-series into the cases time-series. To extract the seasonality from the drivers, we fit a smooth cubic spline to each driver time series by year over the whole period. We construct 500 surrogates time series for each of the six climate drivers being inspected, so that we have produced 3000 cross-mappings with the 500 surrogates, in total. We summarized the surrogates constructed with the below figure.

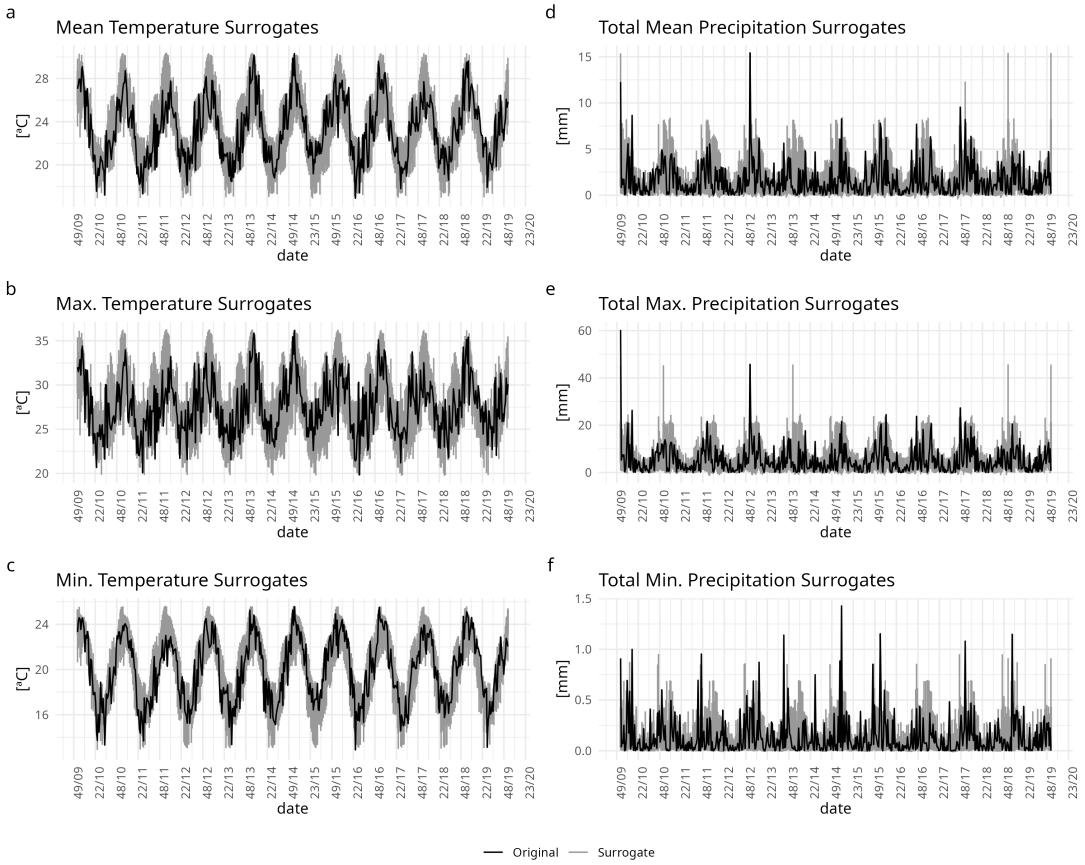

Figure 3.2: Surrogates time series, gray shades, and the original climate time series, in black.

In this figure, gray shades are the surrogate time-series and the black line is the original climate driver time-series. We see, by this, that the variability over the seasonality pattern is not big. This is actually expected, for we are only shuffling the years to produce a series that has different chronological order on residuals with respect to the original time-series.

Having constructed surrogates for CCM test of the significance of a driver, we look for the time-lag with the highest ρ. This is done by repeating the CCMs between surrogates time-series of the climate drivers to cases time-series, with different target of time for prediction tp. For different time for prediction tp, we produce new cross-mappings, but now with the target of prediction being different from simultaneous time. We do not go further than $tp = 17$ to avoid mixing patterns from one epidemic into another, as 17 is the first 0 of the auto-correlation function (ACF) for the dengue cases. A compilation of the auto-correlation for each of the drivers can be found at Figure C.2 and the auto-correlation for the dengue cases time series can be seen in Fig. 3.3

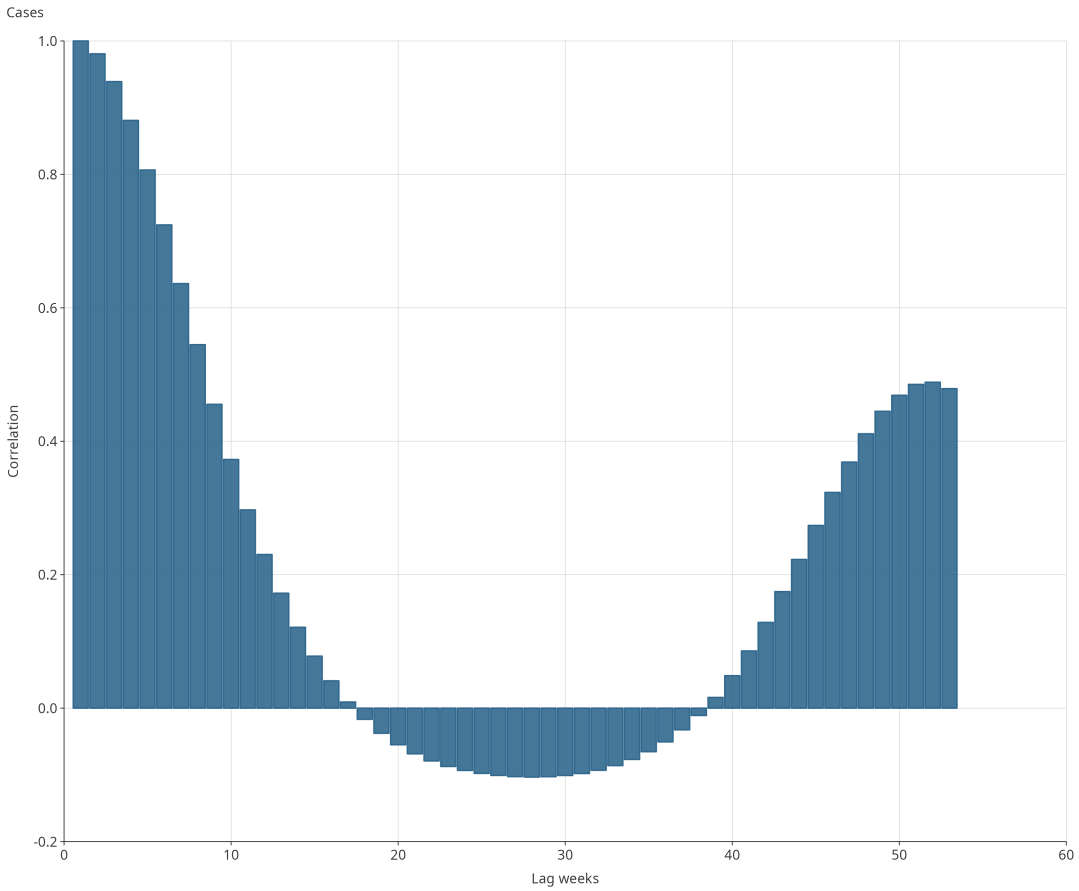

Figure 3.3: Auto-correlation for dengue Cases time series.

Instead of only performing the cross-mappings at simultaneous time, i.e., the climate drivers at time t predicts the cases time series at time t, we test, too, the CCM for climate drivers at time $t - \tau$ predicting the case time series at time t.

For each time for prediction tp we performed 3000 CCMs, in total for all drivers. We have a total of $i \times N_{surr} \times N_{tp}$, where i is the number of climate drivers being tested, here 6 (*i.e.* min., mean, max. of temperature and precipitation), N_{surr} is the number of surrogate time-series created (500) with the same seasonality pattern of the original driver and N_{tp} until 17 weeks of lag. In total, we perform $6 \times 500 \times 17 = 51,000$ CCMs. After performing all of those CCMs we get a final distribution which is a distribution of ρ per time for prediction tp for each climate driver tested, which are summarized this in Figure 3.4.

All the CCMs are performed through the same equation, which is the prediction of cases SSR by the climate drivers SSR, adapting the equation given in the equation A.10, leading to:

$$\hat{C}(t)|\mathbf{M_i} = \sum_{i=1}^{E+1} w_i C_i(t) \tag{3.2}$$

The index i runs over all drivers being tested at their optimal lags tp^* and optimal embedding dimensions E^*. For each climate driver, we choose those where the CCM is significant over the distribution of ρ, which is the CCM with the highest ρ over all time for predictions tested Figure 3.4. The optimal time-lag is understood as the highest lag effect of a climate driver onto the cases time-series. We used the time-lag that has the greater forcing on the cases time-series. The climate drivers with the highest ρ at its optimal lags are $Precip_{min}(t-7), Precip_{max}(t-2)$. This means the temperatures do not drive the cases time series by its variability but probably by the seasonality, as the original CCM were not significant over the surrogates distributions. For the precipitations, we can attribute that the highest lags have a significant causal link, so the precipitations, around those lags, drives the cases time-series by its variability and not only by its seasonality.

S-Map and non-linearity of the dynamics: After having found a set of climate drivers that drive the cases time-series, we can test quantitatively how these drivers force the dynamics of dengue cases. A way to test this in the EDM framework is to produce new SSRs that will we use to predict the cases time-series. The idea is to use S-Map, as described in the section Sequentially locally weighted Map (S-Map)[60], and perform a sequential locally weighted estimation in order to estimate the Jacobian elements at each state of SSR[61, 95], but now instead of only using the cases time-series and its lags, we substitute some of these lags by the climate drivers that we found as causal factors, in our case, the precipitations, $Precip_{min}(t-7), Precip_{max}(t-2)$. As in the equation A.12, we are determining the C_i which are understood as forces of the interaction of this climate drivers on the dengue cases time-series. Adapting the equation in the section A.4, we have:

$$Cases(t) = C_0 + \sum_{i=1}^{I} C_i(t) x_i(t) \tag{3.3}$$

where the sum is over the significant climate drivers.The x_i term is the vector of climate driver's time-series at a tp that has the highest ρ. In summary, our S-Map is of the form:

$\langle Precip_{min}(t-7), Precip_{max}(t-2)\rangle$. Before inputting the time-series to the S-Map we have to cut and normalize the time series, this guarantees all the forces of interactions are measured as relative effects on the dengue cases time-series and are more easily interpreted. The coefficients C_i have the form of an exponential decaying weight to the distance of the point we were predicting from its nearest neighbors. As given in the equation Equation A.13, we determine the θ from this equations by picking the θ that gives the minimum Mean Absolute Error (MAE) between the predicted and observed time series, predicted by this SSR of precipitations to predict the cases time-series. In Appendix C we give the plot for θ for this model of precipitations predicting cases, Figure C.3. S-Map is basically the Singular Value Decomposition (SVD) solution of the Equation 3.3 of the form $\mathbf{B} = \mathbf{A} \cdot \mathbf{C}$, with $\mathbf{B}$ as the weighted future values of the $Cases(t)$ and $\mathbf{A}$ the matrix of dimension $n \times I$ entries[61]. A mathematical description of the S-Map is given at Sequentially locally weighted Map (S-Map). S-Map estimates, at each time point, the locally weighted Jacobian Element for the prediction of the cases time-series from the climate drivers state-space reconstruction. We can give a complete model for this prediction that S-Map gives locally estimates, as, for example, in the following equation:

$$Cases(t) = C_0 + \frac{\partial Cases(t)}{\partial Precip_{max}(t-2)} Precip_{max}(t-2) + \frac{\partial Cases(t)}{\partial Precip_{min}(t-7)} Precip_{min}(t-7) \quad (3.4)$$

The term C_0 is a constant coefficient and all others are coefficients for the effects of each time series inputted on the S-Map model to predict the target time-series, the cases time-series. This procedure estimates the effects of the climate driver on the cases time-series, the partial derivatives elements in the Equation 3.4.

3.4 Results

Cases data description: In Table 3.1 we give a descriptive summary of some characteristics of dengue incidence in the municipality of Rio de Janeiro. In the "Overall" column, a general view is presented which is in line with the basic patterns for Brazil given in Table 2.1. We can notice a higher incidence in women, which can only be partially accounted for by the skewed sex-raio of the country (52% females). The case-fatality rate is, in general, small, as expected.

Table 3.1: Characteristics of dengue Cases for Rio de Janeiro city, 2010 to 2020.

Characteristics Variables	**Overall**, N = 424,938[1]	**0-1**, N = 9,543[1]	**1-9**, N = 35,805[1]	**10-17**, N = 69,906[1]	**18-39**, N = 175,652[1]	**40-59**, N = 99,873[1]	**60-79**, N = 31,232[1]	**80+**, N = 2,927[1]
Age, Median (p25, p75)	29 (16, 44)	0 (0, 1)	6 (4, 8)	14 (12, 16)	28 (22, 33)	48 (44, 53)	65 (62, 70)	83 (81, 86)
Self-Reported Race								
Black	19,659 (13%)	250 (8.6%)	1,310 (11%)	3,182 (13%)	8,513 (13%)	4,870 (13%)	1,427 (12%)	107 (11%)
Brown	70,654 (45%)	1,219 (42%)	5,465 (47%)	11,809 (48%)	29,994 (46%)	17,012 (44%)	4,810 (39%)	345 (35%)
Indigenous	555 (0.4%)	13 (0.4%)	43 (0.4%)	96 (0.4%)	228 (0.3%)	127 (0.3%)	48 (0.4%)	0 (0%)
White	62,944 (40%)	1,416 (48%)	4,767 (41%)	8,887 (36%)	25,859 (39%)	15,714 (41%)	5,773 (47%)	528 (53%)
Yellow	2,240 (1.4%)	23 (0.8%)	134 (1.1%)	398 (1.6%)	948 (1.4%)	565 (1.5%)	162 (1.3%)	10 (1.0%)
(Missing)	268,886	6,622	24,086	45,534	110,110	61,585	19,012	1,937
Sex								
Female	231,079 (54%)	4,656 (49%)	17,599 (49%)	34,068 (49%)	93,987 (54%)	59,877 (60%)	19,103 (61%)	1,789 (61%)
Male	193,859 (46%)	4,887 (51%)	18,206 (51%)	35,838 (51%)	81,665 (46%)	39,996 (40%)	12,129 (39%)	1,138 (39%)
Delay to notification, Median (p25, p75)	3.0 (1.0, 5.0)	3.0 (1.0, 5.0)	2.0 (1.0, 5.0)	2.0 (1.0, 4.0)	2.0 (1.0, 5.0)	3.0 (1.0, 5.0)	3.0 (1.0, 6.0)	3.0 (1.0, 7.0)
Outcome								
Death	125 (<0.1%)	7 (<0.1%)	3 (<0.1%)	12 (<0.1%)	23 (<0.1%)	37 (<0.1%)	31 (<0.1%)	12 (0.4%)
Discharged	424,813 (100%)	9,536 (100%)	35,802 (100%)	69,894 (100%)	175,629 (100%)	99,836 (100%)	31,201 (100%)	2,915 (100%)

[1]Median (IQR); n (%)
[2] Missing are considered as Discharged

Climate data description: The climate data were aggregated for the entire municipality area of Rio de Janeiro, and we generated a weighted mean area time-series for each climate variable of interests. The raw climate data is a product from a global climatological gridded model, with a pixel definition of (0.1°X0.1°), covering the whole land surface of the municipality. From this data, we produced the weighted mean area time-series by taking inside the municipality extend the area of each pixel and weighting the area it covers on the extent of the municipality. This was done by the R package 'exactextractr'[85].

Surrogate testing for seasonality forcing

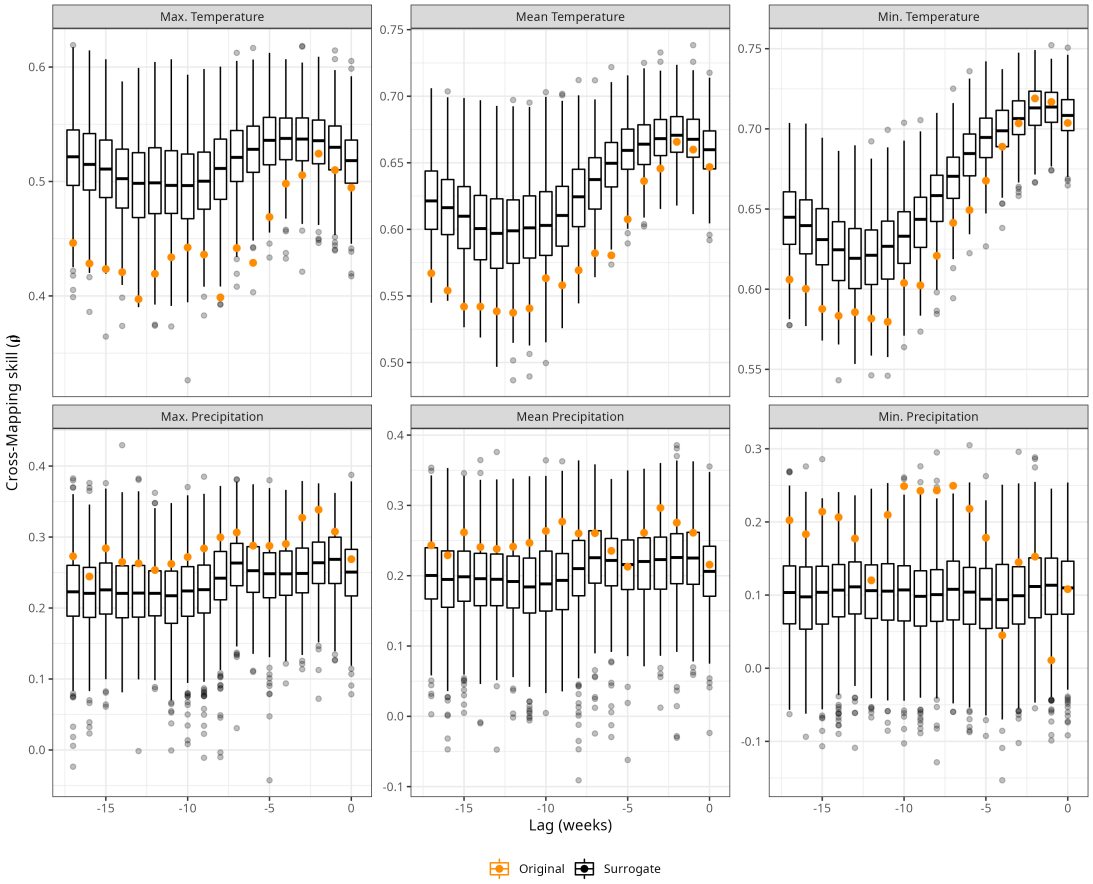

Figure 3.4: Boxplot for the ρ distribution per climate driver by time for prediction *tp*. The orange points are the CCM between original climate driver and cases time series.

The Figure 3.4 gives the cross-mapping skill distribution for the prediction of each climate driver tested by lag length of the mapping. In other words, at each specific time for prediction *tp*, the x-axis, there is a boxplot which is the distribution of the 500 CCMs of each surrogate time series and the cases time series, the orange points are the CCMs for the original driver, at each *tp* tested. The title of each panel identifies the climate driver tested. By examining this distribution, we can obtain an insight of how the original CCMs compares with the surrogates CCMs. If the original CCM does not have a clear better cross-mapping skill than the surrogates ones, this is a sign of the forcing by this climate driver is due to the seasonality of the driver and not due to its variability. To make a statistical test for this, we assign for the CCMs of the original climate time-series a significance only if the cross-mapping skill is outside the 95th percentile of the ρ_{CCM} distribution of all other CCMs for the surrogates time-series. When the CCM passes this significance test, we say that the climate driver has a forcing on the cases time-series that is not only due to the seasonality but by the variability itself, and that we have found a lagged relationship between the climate driver and the cases time series. To make it clearer, we draw the Figure 3.4 now coloring the CCMs for the original time-series of the climate by the level of significance of the ρ_{CCM}. The next figure shows this result.

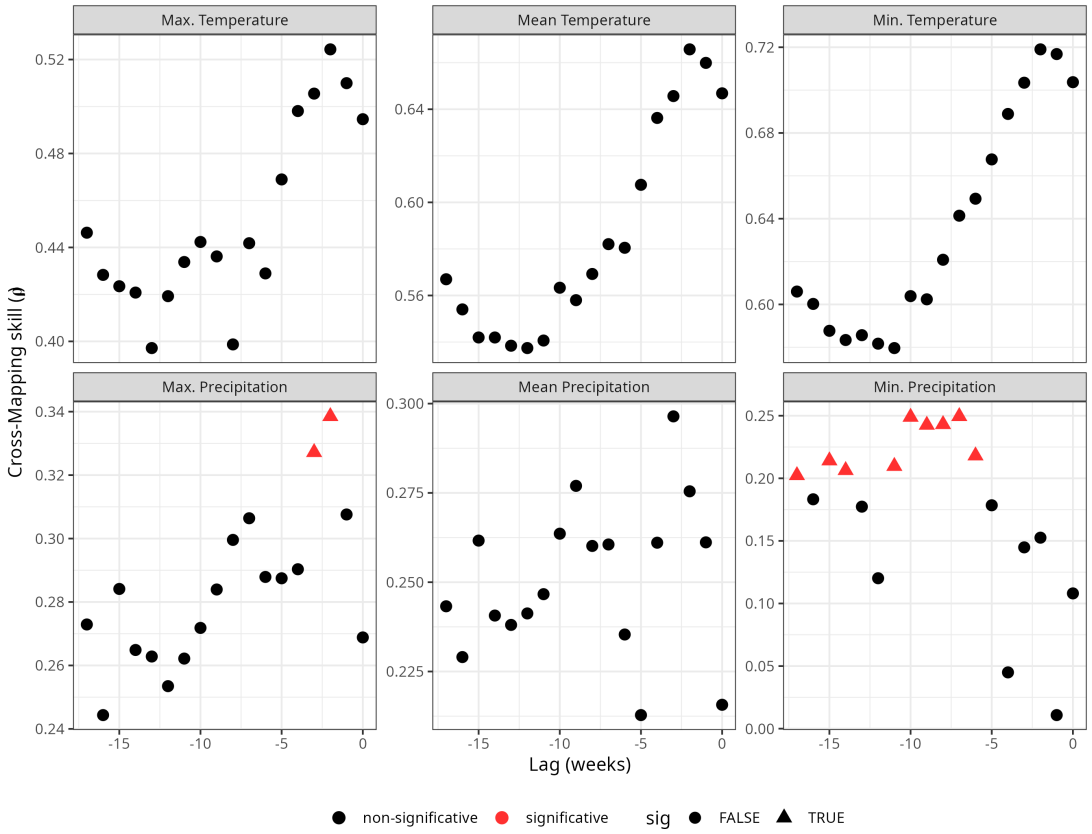

Figure 3.5: Boxplot for the ρ_{CCM} distribution per climate driver by *tp*. With significance test.

From the above figure, we can infer the minimum precipitation with a delay of 7 weeks and maximum precipitation with a delay of 2 weeks as the climate drivers causally forcing on the cases time-series. The temperature has no significant forcing for the variability, so probably the forcing of temperature on the cases is just a seasonality effect.

S-Map Models: After determining the climate drivers and their delays, we can draw S-Map Models to estimate the force of interaction of these drivers on the cases time-series. As in the Equation 3.4 we compute the locally weighted Jacobian elements at each point of time of the SSR constructed to predict the cases time series. As described in the section A.4, an S-Map is a model to predict a time-series from other time-series that are causally driving the one that is being predicted. Differently from the CCM, we are not inferring the causal link, but using those that were found before to build a new state-space, which is mixed, with the cases and the climate drivers in the same state-space. From this new state-space, we employ a model to estimate the cases from the climate drivers.

The panel below shows, the results generated by the S-Map algorithm. To help with the interpretation of the results, we use different colors for each month.

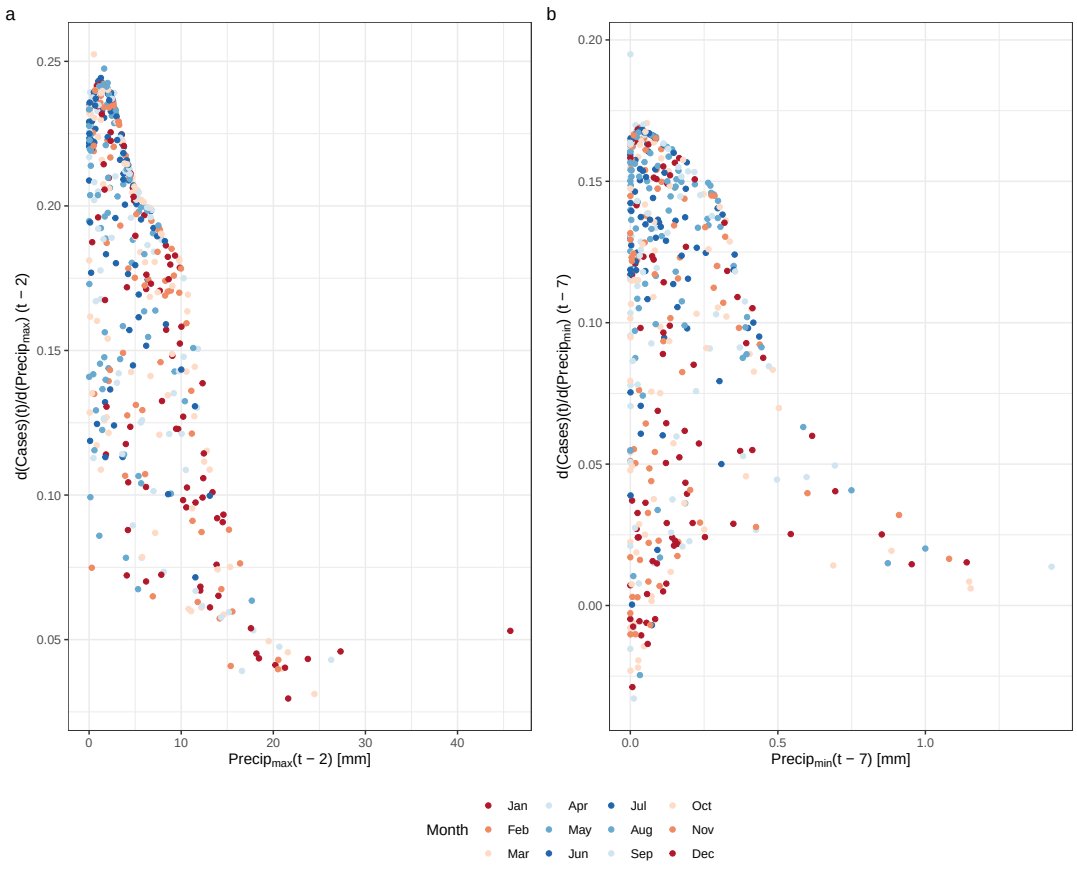

Figure 3.6: Forces of Interaction for Minimum and Maximum Precipitations, generated by the S-Map algorithm

The results show how important a driver is as a function of its magnitude. Each point is a week, and the points cover all the time-series extension. With very few exceptions, the coefficients are positive, meaning that an increase of maximum precipitation at $t-2$ and of the minimum precipitation at $t-7$ tend to increase the number of dengue cases. However, the magnitude of this driving effect is variable. In the panels a and b, the effect of precipitation on the number of cases is shown. In both cases, we see a saturation effect: for higher values of the precipitation, the effect on the number of cases is small. On the other hand, for low values of the precipitation, the effect depend strongly on the season of the year. This can distinctively be seen in the panel b: in the months of December, January and February, the effect of $\text{Precip}_{\text{min}}(t-7)$ tends to be small, but in the usually drier months of the middle of the year, the effect is strong.

3.5 Discussion

We used the Empirical Dynamical Modelling (EDM) framework to find climate drivers of the number of dengue cases in the municipality of Rio de Janeiro, on a 10-year period, from 2010 to 2019.

First, we determined the climatic drivers that showed causal association with the number of dengue cases, via CCMs. We also calculated the lag between driver variables and cases that showed the stronger associations. Finally, we characterized quantitatively these drivers, calculating their effect of the number of cases as a function of the magnitude of the driver variable.

Our main findings are that precipitations, ($\text{Precip}_{\text{min}}$ and $\text{Precip}_{\text{max}}$), forces dengue cases time-series through its variability other than only through its seasonality. Time lags between these two drivers and the number of cases were seven and two weeks respectively. Temperatures were tested and were not found as significant drivers. This does not mean temperature has no effect on the number of cases, it means that cases variability are not causally connected with variability of the temperatures, after taking seasonal effects out of the analysis. Further analysis, via S-map techniques, estimated the interaction forces of the two causal drivers and showed that the effect of increasing precipitations increase the number of dengue cases - a positive effect. This effect is however saturated for high values of precipitations and is more important in the dry months of the year.

These results are in line with recent work on climate drivers of dengue cases in San Juan, Puerto Rico [22]. These authors also found that precipitation is a causal driver of cases beyond seasonal effects, while temperature is not. For diseases that are affected by climate, as is the case of dengue, climate effects may be context dependent and may have opposing effects according to environmental factors and geographical locations [50]. It is thus remarkable that our results and that of Ref.[22] agree on the effects of precipitation and temperature on dengue cases, even in different locations.

Time-series analysis cannot provide definitive answers about the mechanisms that are actually linking cause and effect. However, checking if the relations found are biologically viable is important. It is known that standing water is habitat for larvae of *Aedes Aegytpi*, the vector of dengue infections. Rainfall obviously provides water, but the relation to mosquitoes

population dynamics is complex. While a certain minimum amount of water is necessary, too heavy rainfall has an effect of flushing, [96]. Our results capture some of these complexities. Maximum precipitation at $(t-2)$ has a very small effect on the number of cases if the maximum precipitation is above 20 mm in a week. These points correspond to the rainy season, consistent with an effect of flushing. The effect is not immediate, but its impact is in a short timescale, two weeks. We also found saturation of the impact of the minimum precipitation with a lag of seven weeks. More interesting is the region of small minimum precipitations. A clear difference exist between the drier months and usually wetter months. In dry months, the effect of minimum precipitation, at, $(t-7)$ is important, but no so in the wetter months. A possible interpretation of this fact is the importance of the inter-epidemic period. This has been discussed in [97, 66] and is linked to the maintenance of the mosquito population at inter-epidemics periods at high enough populational level.

3.5.1 Conclusion

We have shown that, beyond the seasonal effects, dengue cases are driven by maximum and minimum weekly precipitations, with a delay between the driver and the cases of, respectively, two and seven weeks. The effect is more important in dry months, indicating the importance of the inter-epidemic period. For higher precipitations in the rainy season, the effect of precipitation is saturated, possibly indicating a combination of positive and negative effect of precipitation.

3.5.2 Acknowledgments:

The authors also thank the research funding agencies: the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (Finance Code 001 to RLPS), Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brazil (grant number: 141698/2018-7 to RLPS). The funding agencies had no role in the conceptualization of the study.

3.5.3 Credit Author Statement:

Conceptualization – RLPS, RMC and RAK. **Methodology –** RLPS, RMC and RAK. **Software –** RLPS. **Validation –** RLPS. **Formal Analysis –** RLPS. **Investigation –** RLPS. **Data Curation –** RLPS. **Writing - Original Draft –** RLPS. **Writing - Review & Editing –** All authors. **Visualization –** RLPS and BCTC.

3.5.4 Data Availability Statement:

All data used in this study are publicly available. Dengue cases data is available at http://sinan.datasus.gov.br/principal/index.php, and temperature at https://cds.climate.copernicus.eu/#!/home. The code to reproduce this analysis is available at: https://github.com/rafalopespx/dengue_ccm_cities

Part III

Environmental epidemiology of dengue

4

Ambient Temperature and Dengue Hospitalisation in Brazil over a 10 years period, 2010-2019: a time series analysis

In collaboration with: Xavier Basagaña[1], Leonardo S. L. Bastos[2], Fernando A. Bozza[3], Otavio T. Ranzani[4]

4.1 Introduction

Dengue fever has been a major global seasonal endemic disease, present largely in the tropics, with an estimated burden over $\sim$ 390 million infections per year.[5, 67, 89, 2] Approximately one-quarter of these infections manifest as clinical or subclinical disease[5]. Brazil is one of the most affected countries by Dengue, that has caused over 20·9 millions cases of infections in the last 20 years, since the compulsory universal notification to the Brazilian health systems was established [98, 20]. Additionally, it is estimated that the Brazilian Universal Health system (SUS) spent more than USD 159 millions in the treatment and assistance to Dengue cases and USD 10 million on severe Dengue between 2000 and 2015[98]. The impressive burden of Dengue in Brazil is also observed in several others low-income and middle-income countries[6].

Dengue incidence is influenced by climate variables, due to mosquito life cycle and human

[1] 2 - Barcelona Institute for Global Health, ISGlobal, Barcelona, Spain; 3 - Universitat Pompeu Fabra, UPF, Barcelona, Spain; 4 - CIBER Epidemiología y Salud Pública, CIBERESP, Madrid, Spain

[2] 5 - Department of Industrial Engineering (DEI), Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, RJ, Brazil

[3] 6 - National Institute of Infectious Disease Evandro Chagas (INI), Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro, RJ, Brazil

[4] 2 - Barcelona Institute for Global Health, ISGlobal, Barcelona, Spain; 3 - Universitat Pompeu Fabra, UPF, Barcelona, Spain; 4 - CIBER Epidemiología y Salud Pública, CIBERESP, Madrid, Spain; 7 - Pulmonary Division, Heart Institute (InCor), HCFMUSP, Faculdade de Medicina da Universidade de São Paulo, Brazil

behaviour changes, particularly with rainfall and temperature.[5, 67, 20, 6] However, there is lacking evidence for the association of ambient temperature and dengue severity. The pathophysiology of Dengue severity involves dehydration and coagulation disorders, both conditions that can be aggravated by ambient temperature.[6, 26, 94]

Based on the extensive literature on the association of temperature and all-cause hospitalisations [67, 89, 98, 20, 6], we hypothesised that high ambient temperatures are associated with a higher risk of hospitalisation due to Dengue. We aimed to evaluate the association between ambient temperature and Dengue hospitalisation in Brazil, analysing the period from 2010 to 2019.

4.2 Methods:

Setting: Brazil has about 211 million inhabitants distributed over 8·5 million km² area, is divided into 26 states and the country capital, a federal district.[99] These 27 states are grouped in 5 administrative macro-regions: North, Northeast, Center-West, Southeast and South. The country is located in a tropical region and has 3 main Köppen climate types and 12 subtypes.[100] We conducted a time-stratified regression analysis to evaluate the association between ambient temperature and Dengue hospitalisation in Brazil. The unit of analysis was municipality (n=5,570) of residence.

4.2.1 Data sources and definitions:

Dengue hospitalizations: We used the Brazilian Hospital Admission System (SIH), a nationwide database that comprises individual level data of all hospitalizations covered by the Universal Healthcare System (SUS) in Brazil. We defined a dengue hospitalisation by the ICD-10 codes 'A90', 'A91', 'A97', 'A970', 'A971', 'A972', 'A979'.[34] We build daily time series by date of hospitalisation for each municipality of residence. We excluded dengue hospitalisations from three municipalities that are placed in small islands, because there was no ambient temperature for them. This filtering resulted in a loss of 153 (0.03%) cases in total, for the whole period of the 10 years. The final time series aggregation by municipalities encompasses a total of 579,703 hospital admissions due to Dengue over the 10 years period of 2010 to 2019.

Ambient temperature: We used the gridded (0.1° x 0.1°) 2 metres daily temperature taken from reanalysis products (ERA5-Land) freely available by the Coppernicus Climate Service through the Climate date store.[73] The original data was an hourly gridded dataset over the extension of Brazil and encompassing the whole period of 10 years. We estimated the daily mean temperature for each municipality by calculating the daily temperature weighted to the municipality area. All these weighted mean areas were done with the *'exactextractr'* R[101] package.[85]

4.2.2 Data Analysis:

We study the association between dengue hospitalizations counts time-series and ambient daily temperature using the Distributed Lag Nonlinear Model (DLNM) framework.[54, 55] We use the two-stage approach, first fitting generalised linear models at Brazilian State levels. In the second stage, we pooled the estimates using a meta-analysis at the country level and its five macro-regions.[57, 102, 56] At the first stage, we fitted a Conditional Poisson Model for the aggregated time series of Dengue hospitalisation counts.[103] The model strata was a dow-month-city term.[104] We report the overall Relative Risk (RR) compared to the Minimum Hospitalisation Temperature (MHT) point for each state. MHT is the temperature with minimum hospitalisations according to the model. We also reported the lagged effects at the 50th and 95th percentiles. The second-stage was one meta-analysis for the whole country and another for each of the macro administrative regions. The meta-analysis for the whole country was conducted, aggregating all the coefficients fitted by each state and their covariance matrix in a multivariate meta-analysis random effects model[57, 56]. The meta-analysis for each macro-region used the coefficients and the covariance matrix of the states within a given region. From the meta-analysis, we estimated the overall cumulative RR curve, relative to MHT and the lag-effects at 50th and 95th percentile.

4.2.3 Statistical Analysis:

First-stage: In the first-stage, we ran the following model for each state, we suppose the assumption of the Dengue hospitalisations counts being given by a *quasiPoisson* distribution, where the mean of the counts at i-th municipality at day t it is the $\lambda_{i,t}$ and ψ is the parameter to account for over dispersion[103], the whole model it is then:

$$\begin{aligned} Y_{i,t} &= quasiPoisson(\lambda_{i,t}, \psi), \\ g[E(Y_{i,t})] &= log[E(Y_{i,t})] = \xi_{s,i} + f(x_{i,t}, l) + \sum_{k=1}^{K} s_k(t) \end{aligned} \tag{4.1}$$

Where $Y_{i,t}$ is the daily count of hospitalisations on the i-th municipality, $\xi_{s,i}$ the dow-month-municipality strata term conditioned out. $f(x_{i,t}, l)$ which is the bi-dimensional exposure-lag-response distributed lag non-linear model for mean temperature by each day of delay, until 21 days of lags. The cross-basis is parametrized with natural splines, with 2 knots equally spaced on the exposure-response structure and 3 knots equally spaced on the log transformed scale for the lag-response structure. The last term is the long-term trend model choice for temperature trend along the whole period, a natural spline with 7 degrees of freedom by each year on the whole period. A more detailed description on how to parametrize and how the DLNM is constructed is given at the Appendix B.

Second-stage: Whole Country and Regions: In the meta-analysis, we did the aggregation of the data generated by the first-stage analysis on the whole country and by each of the five macro regions. We ran a multivariate meta-analysis with random effects[56], so from the i-th state-level study, $_i$, we have the following model:

$$\hat{\theta}_i|u_i \sim N_k(\theta_i + u_i, S_i),$$
$$u_i \sim N_k(0, \psi) \tag{4.2}$$

Where θ_i is the estimate for the i-th location, u_i the random effects for the coefficients of this study with S_j intra-location covariance matrix of coefficients estimates, and ψ the between-location covariance matrix. For the whole country meta-analysis, the coefficients i were taken from the 27 states, while in the macro regions meta-analysis, the coefficients i were only taken from states pertaining to the macro region. We ran two sensitivity analyses with different splines parametrization for each of the exposure-response and lag-response structures. On the first sensitivity analysis, we parameterized the cross-basis with 3 knots (one extra knot compared to the main analysis) equally spaced on the exposure-response structure and again 3 knots equally spaced on the log transformed scale for the lag-response structure. In the second sensitivity analysis we parameterized the cross-basis with 2 knots equally spaced on the dose-response structure and 4 knots placed on the days 1, 2, 7, 14 for the lag-response structure. All the results for the sensitivity analysis are compiled at the Appendix D. All the analyses were run on R statistical Software[101], version 4.1.2.

4.3 Results:

Climate data description: A descriptive table (Table D.1) of the mean daily temperature for each state, macro regions and the whole country over the ten years period is can be found in the supplementary material Appendix D. The overall temperature range across locations and years covered from -0·09° Celsius to 34·8° Celsius. Figure 4.1 gives the distribution of the mean daily temperature for each state.

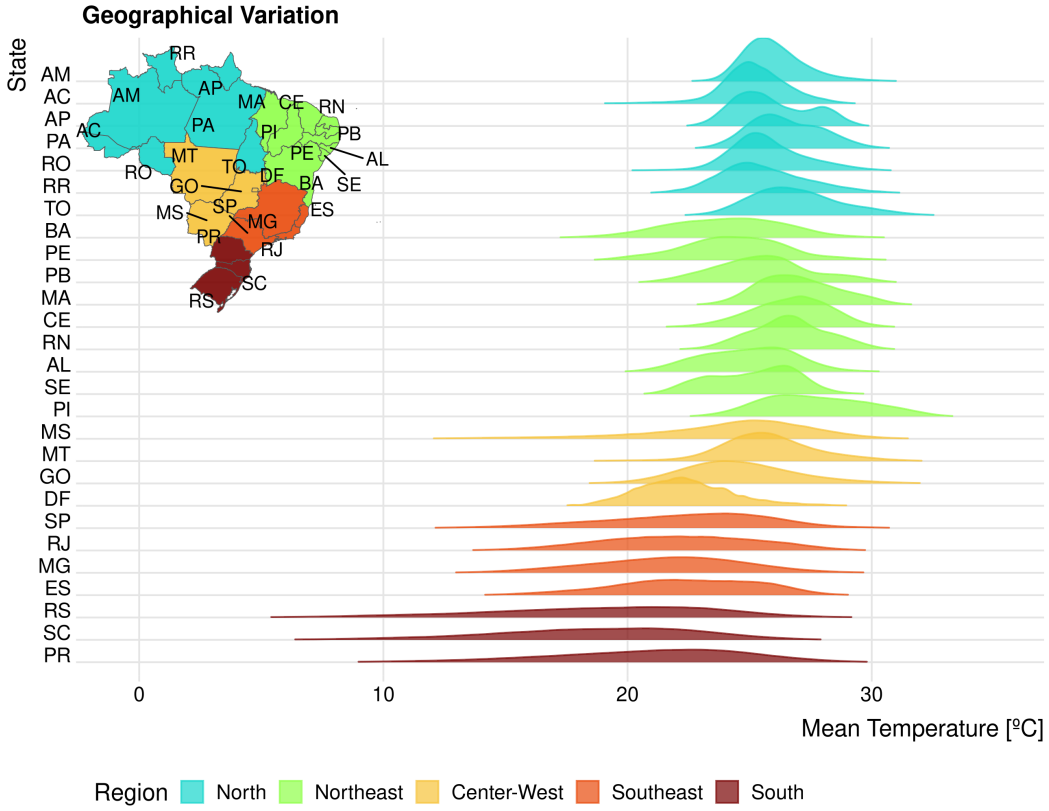

Figure 4.1: Mean temperature density distribution over each state. The colours are given by each macro administrative region.

Cases data description: Table 4.1 gives a characterization of the data summarised by each macro administrative region. In the overall data, the mean age for a hospitalisation by Dengue was 30 years old. Between regions, it can be seen a difference of 12 years old to the mean age of hospitalisation, being the mean age of hospitalisation of 37 years old for the South region and 25 years old for the Northeast region. The sex ratio in overall was typically representative of the sex ratio estimates from the national statistics. The crude in-hospital mortality in general is low, with mortality rate of 0·6% of the total hospitalizations, and there are some discrepancies between regions. The Southeast has a rate of 1·0% of the hospitalizations, with the highest death outcome rate. The North and Northeast with the lowest rate of mortality with an average of 0·4% rate of deaths to the total hospitalizations. Figure D.1 gives a visual description of the time series for Dengue Hospitalisation for each region.

First-stage Results: The cumulative overall lags RR curves to the MHT on each state level, for the whole period of analysis, 2010 to 2019 are given in the supplementary material (Figure D.2).

Second-stage Results: Figure 4.2 and Figure 4.3, presents the cumulative overall lags RR curve derived from the meta-analysis to the association with temperature, for the whole country and for each macro-region. Figure 4.2B and Figure 4.2C, presents the lag effects for the 50th (23·96° C) and 95th (28·68° C) percentiles of mean temperature distribution for Brazil. The lag effects for each macro-region is shown in supplementary material, Figure D.3, Figure D.4, Figure D.5, Figure D.6 and Figure D.7. A summary of the RR and MHT for Brazil and each macro-region is shown in Table 4.2.

Table 4.1: Characteristics of Hospitalised cases of Dengue, 2010 to 2019.

		Region				
Characteristics Variables	**Brazil**, N = 579,703	**North**, N = 75,304	**Northeast**, N = 224,085	**Center-West**, N = 100,056	**Southeast**, N = 155,405	**South**, N = 24,853
Age	30 (15, 49)	27 (15, 44)	25 (12, 45)	36 (20, 53)	34 (17, 54)	37 (21, 55)
Age in Categories						
0-1	8,752 (1.5%)	1,309 (1.7%)	4,457 (2.0%)	1,058 (1.1%)	1,726 (1.1%)	202 (0.8%)
1-9	73,507 (13%)	9,401 (12%)	37,354 (17%)	8,438 (8.4%)	16,620 (11%)	1,694 (6.8%)
10-17	91,786 (16%)	12,562 (17%)	41,991 (19%)	12,155 (12%)	22,171 (14%)	2,907 (12%)
18-39	192,819 (33%)	29,153 (39%)	72,769 (32%)	34,330 (34%)	48,009 (31%)	8,558 (34%)
40-59	126,059 (22%)	14,863 (20%)	39,640 (18%)	26,825 (27%)	38,268 (25%)	6,463 (26%)
60-79	71,395 (12%)	6,737 (8.9%)	22,314 (10.0%)	14,656 (15%)	23,494 (15%)	4,194 (17%)
80+	15,385 (2.7%)	1,279 (1.7%)	5,560 (2.5%)	2,594 (2.6%)	5,117 (3.3%)	835 (3.4%)
Self-Reported Race						
Black	12,467 (3.1%)	1,042 (2.3%)	3,946 (2.7%)	1,187 (1.9%)	5,760 (4.9%)	532 (2.6%)
Brown	244,869 (62%)	39,925 (87%)	120,476 (81%)	34,973 (55%)	45,113 (38%)	4,382 (21%)
Indigenous	1,092 (0.3%)	306 (0.7%)	116 (<0.1%)	595 (0.9%)	62 (<0.1%)	13 (<0.1%)
White	129,031 (33%)	3,893 (8.5%)	19,700 (13%)	24,396 (39%)	65,746 (55%)	15,296 (74%)
Yellow	9,165 (2.3%)	700 (1.5%)	4,003 (2.7%)	2,116 (3.3%)	1,991 (1.7%)	355 (1.7%)
((Missing)	183,079	29,438	75,844	36,789	36,733	4,275
Sex						
Female	310,308 (54%)	38,225 (51%)	121,175 (54%)	54,916 (55%)	82,550 (53%)	13,442 (54%)
Male	269,395 (46%)	37,079 (49%)	102,910 (46%)	45,140 (45%)	72,855 (47%)	11,411 (46%)
Time of Hospitalization	3.00 (2.00, 4.00)	2.00 (2.00, 3.00)	3.00 (2.00, 4.00)	2.00 (2.00, 3.00)	3.00 (2.00, 4.00)	2.00 (2.00, 3.00)
Outcome						
Death	3,436 (0.6%)	281 (0.4%)	958 (0.4%)	548 (0.5%)	1,510 (1.0%)	139 (0.6%)
Discharged[1]	576,267 (99%)	75,023 (100%)	223,127 (100%)	99,508 (99%)	153,895 (99%)	24,714 (99%)

[1] Missing are considered as Discharged

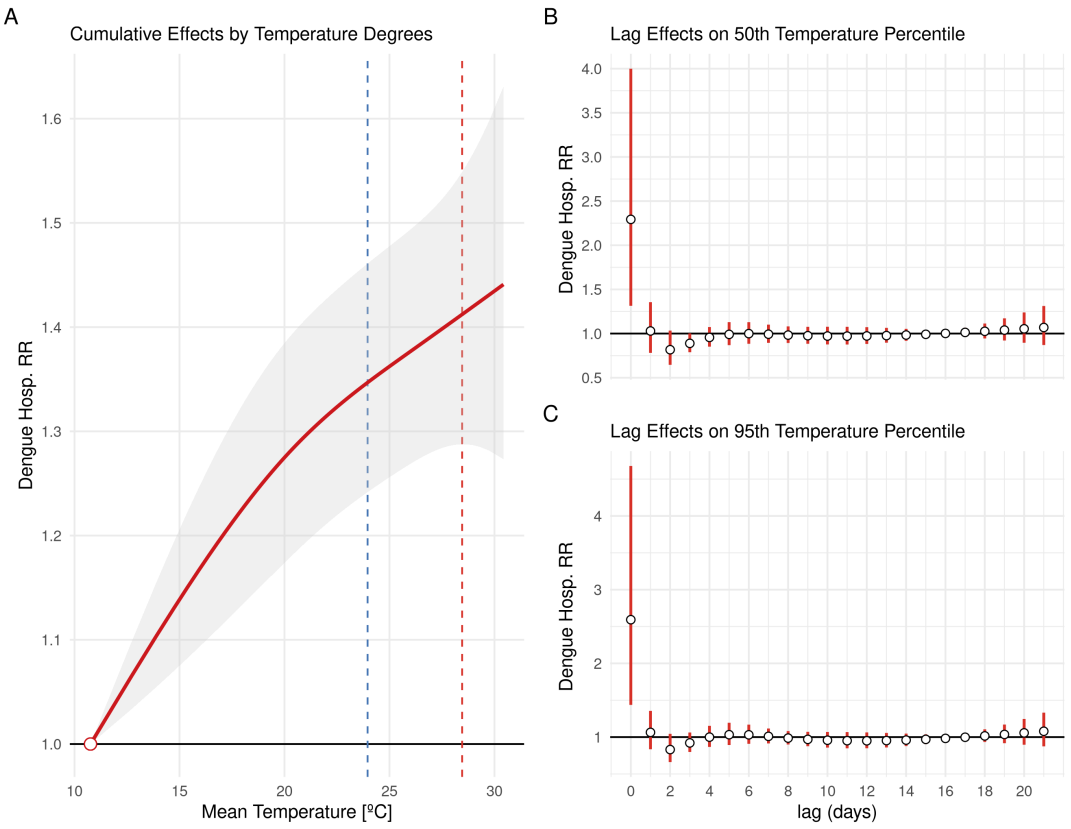

Figure 4.2: A) Cumulative overall lags RR for a Dengue hospitalisation to the MHT association to the mean temperature distribution. Vertical traced lines mark the 50th (Blue) and 95th (Red) percentile of the temperature distribution. The grey shade is a 95% confidence interval derived from the meta-analysis. B) Lag effect to the RR to the MHT of Hospitalisation due to Dengue on the 50th (23·96° C) percentile of temperature. C) Lag effect to the RR to the MHT of hospitalisation due to Dengue on the 95th (28·68° C) percentile of temperature.

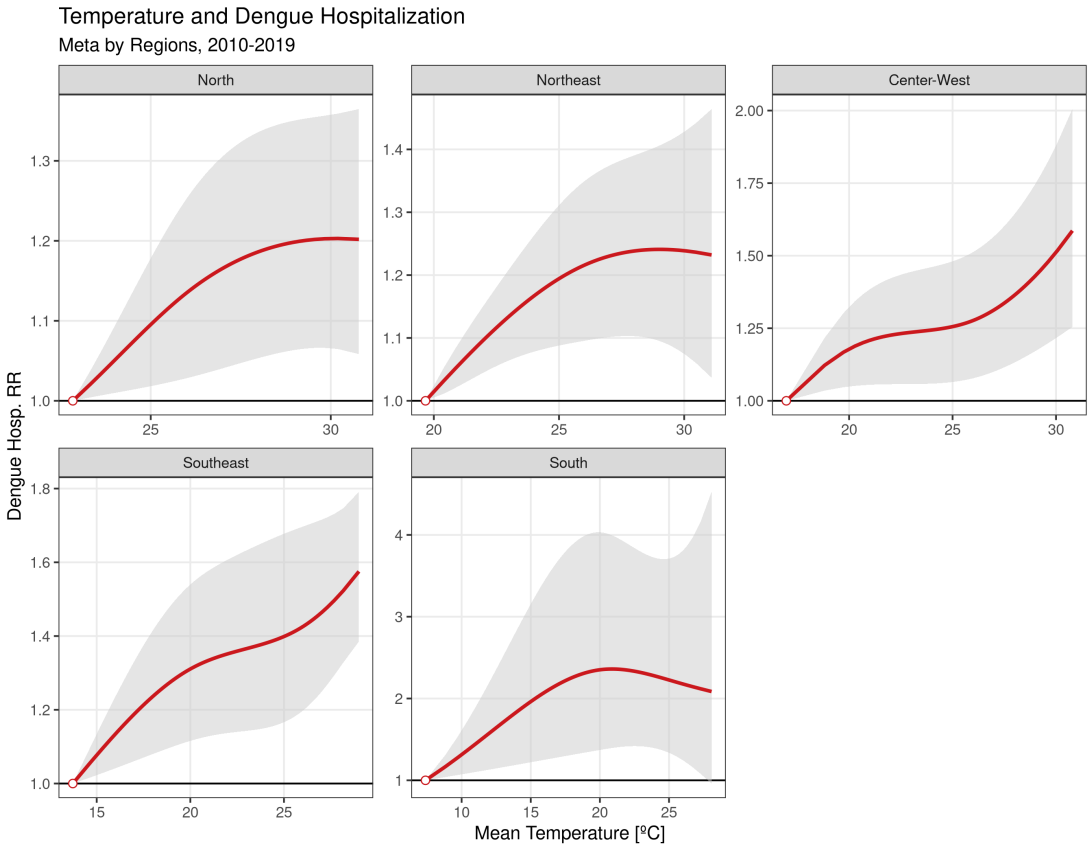

Figure 4.3: Cumulative overall lags RR to the MHT curves to temperature by region. Each panel is given by a meta-analysis model run over all the states coefficients and covariance matrices that pertains to a given macro administrative region. Order by region is given by latitude extent.

Unit	MHT (°C)	RR (IC 95%) at 50th (23·96° C)	RR (IC 95%) at 95th (28·68° C)
Brazil	10.8	1.404 (1.274-1.548)	1.491 (1.337-1.664)
North	22.8	1.153 (1.026-1.295)	1.256 (1.039-1.518)
Northeast	19.7	1.196 (1.091-1.312)	1.255 (1.059-1.488)
Center-West	17	1.303 (1.093-1.553)	1.460 (1.198-1.779)
Southeast	13.7	1.353 (1.171-1.563)	1.498 (1.298-1.729)
South	7.4	2.723 (1.447-5.125)	2.508 (1.362-4.618)

* Relative Risk to the Minimum Hospitalisation Temperature of hospitalisation due to Dengue
** Minimum Hospitalisation Temperature, on Celsius degree

Table 4.2: Minimum Hospitalisation Temperature

Sensitivity analyses Overall, the sensitivity analysis results were comparable with the main analysis. From the meta-analyses, on the 50th percentile for whole Brazil aggregation the RR was 1·629 (1·276-2·079) and 1·405 (1·274-1·549), for the first and second sensitivity analysis parametrization, respectively. On the 95th percentile, they were 1·780 (1·354-2·339) and 1·495 (1·344-1·663), respectively. These results are shown on Table D.2 and Figure D.8 and Figure D.9, in the supplementary material.

4.4 Discussion:

We ran a two-stage time stratified design study to assess the association between hospitalisation by Dengue disease and ambient temperature. Our main findings are that there is a higher risk of being hospitalised by Dengue as the temperature gets higher. In a quantitative way, in general for Brazil, after 20 degrees Celsius, the risk is 1·2x higher than for the minimum temperature observed, which coincides with the MHT. This is the cumulative effects of lags up to 21 days before the date of hospitalisation. We found a stronger immediate effect on the RR to the hospitalisation. We found the same patterns over each macro region of Brazil for the general association between RR hospitalisation and ambient temperature.

In comparison with the literature, we see that, in general, higher temperature is related to an augmented RR of mortality and hospitalisation by any causes[67, 89, 105, 106]. There is scarce literature on the association between temperatures and hospitalisation by an infectious disease. Our study is the first, to our knowledge, to measure an association between temperature and hospitalisation by dengue directly.[107, 108] We did not observe a typical U-shaped or J-shaped curves observed in the majority of studies looking for all-cause and cause-specific hospitalisations.[107, 108] Some specific features of Dengue could explain this difference. First, its progression to severe disease is highly linked with factors that would be aggravated as the temperature increases. Second, the observed days with relatively low temperature are not frequent in Brazil and the incidence of the Dengue disease, of any severity, is much lower during cold months, therefore limiting the power to observe a potential effect of cold temperatures on Dengue hospitalisations. This reflects an additional challenge of studying climate variables and vector-borne infectious diseases severity, among others.[91]

Overall in the country and in each of the macro regions, the extreme heat had an effect of increasing the severity of cases, which evolved to a necessity of a hospitalisation. Quantitatively, the South region had the highest RR to its own MHT overall regions, with 1·5x higher risk on the 95th percentile of its distribution of temperatures. This might be explained by the fact that the South region has the lowest MHT and the wider range of temperature in comparison to all other regions. The difference observed for the South could also reflect

differences in the mosquito life cycle, human behaviour and Dengue mitigation strategies[5, 21, 92]. A direct comparison between regions is not straightforward because of regional disparities. Although Brazil has a universal access health system (SUS), which has great disparities over geographical distribution[109]. Thus, we might expect that regions with more hospital beds per population, such as South and Southeast, would hospitalise milder cases compared with regions with fewer hospital beds. This is also important during massive Dengue outbreaks, where the threshold for hospitalisation might change according to beds availability.

Several mechanisms can play a role in the observed overall association and lag effects. The immediate effect generating the first peak on lag-0 is expected and observed in the majority of studies evaluating temperature and heat waves and hospitalizations and deaths.[106, 108, 68, 110, 111, 112] It reflects an immediate worsening of the clinical condition due to high temperature. This sharp increase on lag-0 reinforces that our results is a direct effect of ambient temperature on hospitalisation risk, rather than the indirect effect of temperature on mosquito activity and increased risk of dengue incidence, an association usually present in the scale of months.[20] Another mechanism is that Dengue has a natural evolution of 1 to 3 days of incubation period, and more 5 to 7 days to complete clearance.[2, 3] If an individual is exposed to higher temperatures during the viral infection evolution, this can lead to more severe infection, as seen in a potential late lag-effect on some regions. Finally, the decrease on the RR following the immediate effect could be due to a depletion of susceptible ("harvesting effect").[110, 111] This mechanism is reported for the lag between temperature and deaths, but also has been hypothesised in another study evaluating temperature and other seasonal infectious diseases such as hand, foot and mouth disease.[113] If harvesting was the only explanation for the observed association, then the increase in hospitalizations observed on the hot day would just represent an advancement of a few days, because of heat, of hospitalisations that would have occurred anyway a few days later in the absence of heat. In that case, the overall RR cumulated over all lags will be close to 1. In our study, we observe important increases in the RR cumulated over three weeks, suggesting that these represent extra hospitalisations that would have not occurred in the absence of high temperatures. Interestingly, the South and Center-West regions present significant associations at later lags, see Figure D.8 and Figure D.9. When varying points for the knots in a sensitivity analysis, lag-effects were less prominent, increasing the likelihood of collinearity explaining this finding.[114, 69, 115]

Our study has some strengths. We evaluated a nationwide database providing a 10 year time-series of Dengue hospitalizations in a LMIC exposed to a wide temperature range. Ambient temperature from reanalysis products from ERA5-Land are the state-of-the-art in climatology[110, 111], as the framework applied for data analysis, accounting for dose-response and lag-response structures and correlated daily data on temperature[104, 116]. This has great generalizability to many other LMIC countries with great burden of Dengue and provides support to public health measures.

Our study has limitations to be mentioned. First, we did not evaluate other factors that could modify the high temperature effects, such as green space, urbanisation and relative humidity. In the same extension, we did not evaluate individual factors, such as age and sex, that could show different effects of temperature and risk of Dengue hospitalization in vulnerable

populations. Second, we obtained estimates for 27 states and pooled them for Brazil and each corresponding macro-regions. States from the same macro-region will have similar climate as well as a similar Dengue incidence. This guarantees that they were likely comparing states with similar conditions for mosquito transmission as well hospitalisations conditions.[21] Although this analysis respected administrative divisions, it might not be generalizable to some municipalities. Finally, we considered hospitalisation due to Dengue as a surrogate of severe Dengue without evaluation of clinical severity using defined clinical criteria.

4.4.1 Conclusion

We found an increased RR to Dengue hospitalisation due to high temperatures in Brazil for a ten years period, 2010-2019. This RR varies over the whole extension of Brazil and between macro-regions. With the findings presented here, we hope the climate variability of Brazil can be taken into account when public health measures are designed to curb and mitigate the impacts of Dengue on the population.

4.4.2 Acknowledgments:

The authors thank Carles Milà for the advice on ERA5-Land data management. The authors also thank the research funding agencies: the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (Finance Code 001 to RLPS), Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brazil (grant number: 141698/2018-7 to RLPS). OTR is funded by a Sara Borrell fellowship (CD19/00110) from the Instituto de Salud Carlos III. We acknowledge support from the Spanish Ministry of Science and Innovation through the Centro de Excelencia Severo Ochoa 2019-2023 programme and from the Generalitat de Catalunya through the Centres de Recerca de Catalunya (CERCA) programme. The funding agencies had no role in the conceptualization of the study.

Role of the funding source: The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Contributors: RLPS, OTR and XB participated in the design and concept of the study. RLPS analysed the data. RLPS wrote the first version of the manuscript. XB, LSLB, FAB and OTR revised the manuscript with important intellectual contributions. XB and OTR supervised the study. All authors had full access to all data in the study, participated in data interpretation, revised the manuscript, and approved the final version of the manuscript for publication. RLPS and OTR verified the underlying data.

Declaration of interests: We declare no competing interests.

4.4.3 Data Availability Statement:

All data used in this study are publicly available. Hospitalisation data is available at http://sihd.datasus.gov.br/principal/index.php, and temperature at https://cds.climate.copernicus.eu/#!/home. The code to reproduce this analysis is available at: https://github.com/rafalopespx/dengue_t2m_severity_paper

Part IV

Conclusion

5 Final Remarks and Perspectives

In this thesis, after going through an overview of the situation of dengue epidemics in Brazil, we investigated how these epidemics relate to climate and environment in two distinct situations. In the first one we focused on dengue cases in the city of Rio de Janeiro and, by means of Empirical Dynamical Modeling, we examined the possible climatic drivers of dengue cases, beyond simple seasonal effects. We found that precipitation, minimum and maximum, are drivers of variability in the numbers of cases. We determined the delay between these drivers and their effect and finally, analyzed the strength of the effects, which turned out to be positive - the higher the precipitations, more case occur. Finally, we also found that this positive effects is saturated, very high precipitations have relatively smaller effects. In our second study, we addressed the relation between temperature and hospitalizations due to dengue, which can be viewed as a proxy of severity of the disease. We consistently found that higher temperature, above 20 degrees Celsius, increase the relative risk of dengue hospitalization, both considering the entire country (Brazil) and each of its macro-regions.

These results contribute to the general knowledge of the relations between climatic factors and dengue disease and can be viewed as a contribution to disease-ecology. Although extrapolation of results obtained in one region to other regions should be considered with caution, the results presented can shed some light on the possible consequences of climate change on the range and consequences of dengue epidemics. It is known, for instance, that dengue is expanding in Brazil, [38]. In particular, in the first three months of 2022, 63 municipalities have presented dengue cases for the first time, half of them in the southern region. It has been hypothesized that this "migration to the south" is connected with climate change, and in particular due to the fact that the southern Brazilian region is becoming wetter, mainly in the spring, [117]. This is coherent with ours findings in a qualitative way, but surely deserves a more specific examination.

Beyond the basic science implications of the results of this thesis, there is also a point to be made concerning public health mitigation strategies. Medium term planning and preparedness should take climate change into account. Our results are part of a larger picture of influences that drive dengue cases and affects its severity. Incorporating these results, we can build a better understanding of future perspectives on occurrence regions of dengue.

The studies presented in this thesis have some limitations. First and foremost, they are based on data of cases and climate variables. This hinders the establishment of a more mechanistic interpretation, as no data about mosquito abundance is used. This is due to the fact that time-series of proxies of mosquito population sizes - infestation indexes - are not generally available. There is a dataset for the city of Porto Alegre, which is only partially in open domain. We are unaware of a similar dataset for Rio de Janeiro, for example. We note that most studies about mosquito-borne diseases suffer the same limitation. However, our study about temperature and severity is not affected by this limitation.

Further limitations come from the fact that the results are context dependent. Strictly, they apply to the places where data was collected. Obviously, this is so for almost every study on climate and diseases incidences and severity. We note, however, that our results about the importance of precipitation go hand-in with results obtained for dengue cases in San Juan de Porto Rico, indicating that we could expect, at least to some extent, that the results reveal some more general traits of the relation between climate and dengue.

When inspecting the relation of incidence and severity with climate variables, we have to choose beforehand which variables that will be the object of the analysis. This is clearly not a systematic procedure. Our choices in this thesis were not totally *ad hoc* as they rely on knowledge of previous results in other contexts. In chapter chapter 3 we studied each climatic variable separately. It could be that a conjunction of variables be more important, as in a cooperative effect. However, we have no clues about which combination could be important, and there is no literature addressing this point clearly. In the same way, in chapter 4, we could ask whether other climate factors produce augmented risk of severity due to a dengue infection. A good candidate would be the near-ground relative humidity, which is known to be a crucial factor of dehydration, which is a mechanism affecting the severity of dengue cases.

In principle, depletion of susceptible could be an important factor that gauges the sizes of epidemic outbreaks. The usual view about this point is that the incidence is not so high as to really affects the pool of susceptible to such an extent that it would, for example, block an epidemic. As the number of susceptible is not a known or measurable quantity, discussion about this issue tends to be speculative in nature. Further, dengue has four serotypes with partial to no cross-immunity, a matter that complicates further the discussion.

The discussion about limitations naturally takes us to consider perspectives for future studies. A first point is to consider new locations and inspect relations between climate and dengue cases. However, as comprehensive study across all Brazilian regions might not viable, an interesting direction is to examine municipalities at the fringes of the regions of dengue occurrence, similar to what has been studied for malaria [65]. This is specially relevant in view of the expansion of the region where dengue occurs. An interesting case would be the city of Porto Alegre, which has some summers with dengue cases and others without. Therefore, drivers beyond seasonality come into consideration naturally, and the methods used in this thesis should be useful.

On the DLNM side, the next immediate work is to include relative humidity near ground as another variable on the modelling strategy, in much the same way we have done with temperature. If it turns out that relative humidity is related to severity of cases, an analysis considering both variables, temperature and relative humidity, at the same time, would be in order. Also, regarding the severity of dengue cases, stratification of the study into

age-dependent settings and sex-differences setting, are interesting new modelling strategies.

In a broader view, results as the ones obtained in this thesis can be incorporated in systems providing advice to policymakers. However, this is far from trivial. A first point is that alert systems do need very recent, if not real-time, data of case counts. These data are not usually open data, at least in the Brazilian case, and access to it needs to be obtained through special agreements. Building trust between modelers, epidemiologists and policymakers is also a challenge, in view of different expectations of the actors.

We hope this thesis can be a piece in the knowledge framework that will help mitigate the burden of dengue disease, and help create movement and energy in this direction.

Part V

Appendices

A Empirical Dynamic Modeling

Empirical Dynamic Modeling (EDM) is a framework developed with the scope of studying and understanding ecological relations [53]. It was first developed to infer causality relations from ecological time-series, as, for instance, population dynamics [53, 62]. EDM make uses of a set of mathematical theorems from dynamical system theory to infer, from time-series only, causal relations between two variables. The premise is that we can reconstruct the dynamics generating the variability from the time-series and lagged copies of it.

In this appendix, we give an overview of the theory behind the EDM framework with a more mathematical formulation. To do so, we first describe the Takens Theorem for state-space reconstruction (SSR) and its corollary to infer causal relations. We will, however, not give rigorous definitions and proofs, as this is beyond the scope of this thesis. Rather, we prefer to explain how empirical dynamics works in practice.

A.1 Takens Theorem

Before formulating the theorem itself, we will review some concepts that will help with the formulation. We understand any dynamical system as a rule that takes a point in a state space to another point in the same space. This is known as a succession rule or composition rule, like:

$$X_t = F^n(X_{t-1}), \tag{A.1}$$

Where F^n is a composition of rules taking the point at time $t-1$ to the value on t. F^n can have the form of non-linear function or rule, and can produce complex dynamics, like chaotic behaviors. We can decompose this rule into more intermediary steps:

$$X_t = F^n(F^{n-1}(X_{t-2})) = \ldots = F^n(X_0), \tag{A.2}$$

In summary, the above Equation A.2, F^n is in fact a function of the first state, X_0. Any dynamical system that can be defined in this way will bear the proprieties to the

reconstruction of it's state-space. As a special case, we can construct systems which are given by a set of differential equations. This is the first premise to the Takens theorem.

The above definition has a continuous-time version. In this case, of discrete-time mappings, we have ordinary differential equations. One of the most studied systems that will be of interest is the Lorenz system, well known to exhibit chaos for a set of parameters. The Lorenz system [118] has been well-studied, and is a good example to better understand the Takens theorem and its consequences. The Lorenz system is a kind of minimal system that exhibits complex dynamics. From this set of equations, we can produce for each of the variables a time-series, which can exhibit chaotic behavior or not. The Lorenz system is given by:

$$\dot{X} = -\sigma X + \sigma Y \tag{A.3}$$

$$\dot{Y} = -XY + rX - Z \tag{A.4}$$

$$\dot{Z} = XY - bZ \quad , \tag{A.5}$$

where X, Y and Z are functions of the independent variable, t, which we will call time. Notice that this a nonlinear, coupled, system of ordinary differential equations.

Let us say we are studying a system whose dynamics is given by a system of ordinary of n differential equations $\{X_i\}$. However, it is often the case that we don't know this system of equations, we don't know which variables are coupled, and we don't know n. But we have the knowledge, through observation, of one of the variables, say X_1.

Before we explain the Takens theorem in this context, we first have to discuss attractors. Attractors are a region of the state space such that (i) if the system is on the attractor at time T, it will continue on the attractor for all times $t > T$ - the attractor is *invariant*; and (ii) then exists a neighborhood of the attractor such that any point in this neighborhood will be infinitesimally close to the attractor for sufficient large times. This neighborhood is called the *basin of attraction*. For the Lorenz system, for instance, the basin of attraction is the whole state space R^3. Intuitively, the attractor is the region of the state space towards which the system tends for large times. It describes, thus, the asymptotic dynamics.

The importance of Takens theorem, in view of applications, is that we can reconstruct an attractor solely from **one** of its variables time-series. In the context of the above discussion, it means that we can use X_1 to obtain the attractor of the system that generated the time-series X_1, without any knowledge of the underlying equations. The reconstruction is built considering an m-dimensional space with time-lagged coordinates, $\{X_1(t), X_1(t+\tau), X_1(t+2\tau), ...X_1(t+(m-1)\tau\}$. By considering the time-series $X_1(t)$ for all t, we will have a sub-manifold of the m-dimensional space. This sub-manifold has a one-to-one correspondence with the original attractor of the system. Takens theorem guarantees that the dimension m is, at most, $2n+1$.

Reconstruction of the attractor in state space was an obvious application: given that we know the time series X_t up to the present, we can use the reconstructed state space to make projections in the future. This is the topic of the next section.

A.2 Simplex Projection

The simplex projection can be understood as encapsulating a point by its simplex on a given embedding dimension. In the definition of simplex, the amount of points that encapsulate a point in a space of E dimension is $E + 1$ points. The simplex projection is basically a method to follow n time steps ahead each of the simplex points encapsulating a single point, the predictee point, $\hat{X}(t + n)$. With the simplex points evolved in time n time steps ahead, we can give a weighted prediction to the future of the predictee point. The correlation between this prediction, $\hat{X}(t + n)$ and the observed future of the predictee point is the prediction skill, ρ.

In the previous section we did not discuss how the embedding dimension, m, is to be chosen. The prediction skill offers a way to determine the best embedding dimension. The optimal dimension of embedding E^* is the dimension that maximizes the prediction skill ρ. The expression for the simplex projection is as follows:

$$\hat{X}(t+n) = \sum_{i=1}^{E+1} w_i X_i(t+n) \tag{A.6}$$

Where w_i are weights that consider how each of the points nearby the predictee point are pondered to produce a projection for the future of the predictee point. On the simplex projection, these weights have the following formula:

$$w_i = \frac{u_i}{\sum u_i} \tag{A.7}$$

where u_i have the following form:

$$u_i = \exp - \frac{\|X_c - X_i\|}{\|X_c - X_i\|} \tag{A.8}$$

If we produce simplex projections for a long enough part of the time series, we can compare it with the observed, and this comparison will give us an estimate of how good are the predictions produced by the simplex projection. When varying how many delays we put into the embedding dimension of reconstructed state-space, we can choose the optimal embedding dimension for the system. This is given by:

$$\rho_{O,P} = \frac{Cov[O,P]}{\sigma_O \sigma_P} \tag{A.9}$$

The dimension of embedding that gives the highest correlation between the observed and predicted from the simplex projection is the optimal embedding dimension. One of the features of non-linearity is the decay of this correlation as we increase the embedding dimensions. This is due to the fact that we are putting too much information into the system to make predictions, we are starting to confound it.

A.3 Convergent Cross-Mapping and Causality Criteria

Let us begin by exploring the Lorenz attractor. The upper part of Fig.(A.1) shows the attractor in the state space with the original coordinates (X, Y, Z). We can reconstruct the attractor from any of these coordinates through Takens theorem. We call these attractors **shadow manifolds**. In the same Fig.(A.1) we show two shadow manifolds, one built from $X(t)$ and its delays, the other coming from $Y(t)$ and its delays. Unsurprisingly, they clearly map on one-to-one to the original attractor.

Let us now assume that we know, from observations, the time-series $X(t)$ and $Y(t)$, but we do not know the dynamical system that originated them, in the case, the Lorenz system. We want to answer the following question: are $X(t)$ and $Y(t)$ connected by a causality relation? Causality here means that both variables participate in a common dynamical system, therefore influencing one another. If we construct shadow manifolds for both of them, if they come from the same dynamical system, there must be one-to-one map between both shadow manifolds. This mapping is called a cross-mapping. Via Takens theorem, we can therefore establish a causality criterion between two time-series.

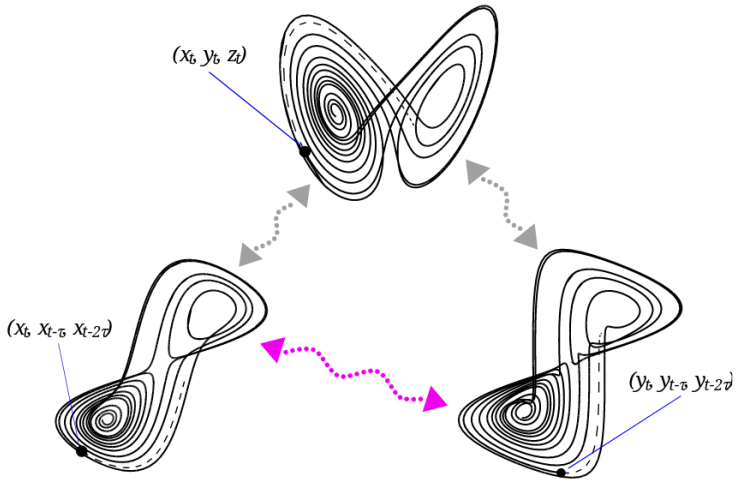

Figure A.1: State-space reconstructions from $X(t)$ and $Y(t)$ of the Lorenz system, giving similar state-spaces topology and dynamics to the original one. Figure taken from https://ha0ye.github.io/rEDM/articles/rEDM.html

We can define a cross-mapping skill in the same way as we did before, in the simplex projection. We call the cross-mapping a convergent if the cross-mapping skill converges to a fixed value when we increase the time-series size. Instead of doing projections of the future of the time-series onto the same time-series, as in the simplex projection, we can project the future of one time-series to the future of the other. The convergence of the cross-mapping skill indicates causality relations. If we had infinite time-series, we would expect the cross-mapping skill to be close to one. However, in practical applications, we always have finite time-series. For this reason, instead of using the actual value of the cross-mapping skill, we use only the convergence of the mapping as a causality criterion. This has to do with the fact that if both reconstructed state-spaces (SSR) are part of the same dynamical system, for some long enough cross-mapping, we will reach the maximum information we can gather from one SSR from another. Notice that, when considering cross-mappings, it is the effect that keeps information from its cause, so from effects time-series we can predict the causes time-series. The direction of causality is in the inverted direction of the cross-mapping. If we are cross-mapping $X(t)$ into $Y(t)$, we are trying to infer if $Y(t)$ causes $X(t)$, in the inverted direction of the mapping.

As before in the simplex projection A.6, we made weighted projection of the future steps of the predictee center, but now with the other time-series and its shadow-manifold. From the effect time-series, say X, we find on the $\mathbf{M_Y}$, the cause SSR, the points that form the simplex around the predictee center and give to them a weighed projection based on their distances in the $\mathbf{M_X}$. If we repeat this for all the points on the $\mathbf{M_X}$ on the $\mathbf{M_Y}$, we are mapping one SSR into another. To generate an estimate to $Y(t)$ from $\mathbf{M_X}$, denoted as $\hat{Y}(t)|\mathbf{M_X}$, we seek for points in the $\mathbf{M_Y}$ by its temporal indexes from points that are forming the simplex in $\mathbf{M_X}$. This is why the estimate is denoted as $\hat{Y}(t)|\mathbf{M_X}$, it is an estimate on Y given by points on $\mathbf{M_X}$. In a formula, we have:

$$\hat{Y}(t)|\mathbf{M_X} = \sum_{i=1}^{E+1} w_i Y_i(t) \tag{A.10}$$

Repeating this mapping for the complete set of points of each time series, we can have a prediction of Y given the points of X, and vice-versa. The comparison between the prediction given by the mapping and the observed points, give us a measure of how good is this mapping, how causally related these variables are. The CCM skill is the correlation between the predicted attractor and the observed attractor, given by:

$$\rho = \rho_{Y,\hat{Y}|\mathbf{M_X}} = \rho_{CCM}(L) = \langle \frac{Cov[Y,\hat{Y}|\mathbf{M_X}]}{\sigma_Y \sigma_{\hat{Y}|\mathbf{M_X}}} \rangle(L), \tag{A.11}$$

A.3.1 Surrogate statistical testing

To use the concepts developed in the previous section for systems that are periodically forced, we need an adaptation. To give a stronger statistical significance to the CCM, we can trick the algorithm, and test which is the main mechanism by which the driving relation arises. The surrogate testing asks for the relation found by CCM, if the forcing mechanism is seasonality or the variability of the series itself that drives the relation. Basically, we construct surrogate

series of the cause, e.g. the climate factors in our example of a population being driven by temperature and humidity. The surrogates are constructed in a way that reproduces the pattern of the seasonality of the climate time-series, but has different residuals. There are many algorithms to generate such kinds of surrogate time-series, and many of them are implemented at the 'rEDM' package [119]. Our algorithm is to take the seasonal component and subtract it from the time-series, and the rest, which are residuals, shuffle it in a non-chronological order. This algorithm will generate surrogates time series that have the same seasonal pattern and residuals in 'wrong' chronological order.

After creating a reasonable quantity of surrogate time-series (we used 500 surrogates as a good quantity), we calculate CCM between the time-series we are trying to estimate, which is the mechanism of forcing. In our example of a population driven by the climate variables, we will produce surrogate of the temperature and humidity time-series, and proceed to calculate CCM from each of the surrogate time series and the $y(t)$ population time series. From this, we obtain a distribution of ρ_{CCM}. If the CCM skill between the original time-series of each climate factor and the population time-series is above the 95th percentile of the distribution of ρ_{CCM}, we say the driving mechanism is by the variability of the climate time-series. This means that the residuals of the actual climate variable time-series, in the correct chronological order, are important to predict the population time-series.

A.4 Sequentially locally weighted Map (S-Map)

After discovering which are the causal relations, we can ask quantitatively how strongly a variable drives another one. On the framework of EDM, this can be done with the use of the S-Map, or the Sequentially locally weighted multivariate Mappings (S-Map). S-Map makes a local regression on each point of the SSR to estimate the Jacobian elements of the interaction on the system. In summary, the algorithm moves along a multivariate SSR, and for each point on the SSR, calculates the Jacobian element of the interaction between each of the variables that integrates the multivariate SSR.

In the pipeline of EDM, after inferring from simplex projections the optimal embedding dimension, and after performing the cross-maps to discover which are the causal relations between the variables studied, we put into the same SSR those variables that we inferred that have a causal relation. Given a set of variables, that has some inferred causal relation to another, as the climate factors driving the dynamics of a time-series, say $y(t)$ for population and temperature and humidity for the climate variables $x_i(t)$. We construct an SSR with all the climate drivers and the population time series, and try to predict the population by weighting locally the Jacobian elements of this interaction, as in:

$$y(t) = C_0 + \sum_{i=1}^{I} C_i(t) x_i(t) \quad , \tag{A.12}$$

where I is the embedding dimension of this S-Map, which runs over the number of climate drivers considered significant as defined at subsection A.3.1 plus the population time-series and some delays until we reach the embedding dimension. The C_i weights and controls how much each of the drivers considered has influence on the total effect on $y(t)$. The S-Map

estimates each of the C_i locally, meaning, on each point of the time-series the effect of the drivers are modulated by the C_i coefficients. To better understand how this is established, we have the following expression for the C_i:

$$C_i = \exp^{-\theta \frac{\|x_k - x(t^*)\|}{d}} \tag{A.13}$$

Where $\|x_k - x(t^*)\|$ is simply the Euclidean distance between two vectors in the state-space and θ is a weighting parameter that gives the shape of decaying of the points considered relevant to estimate the interaction force. When θ is equal to 1, we are simply calculating the least squares of the predicted values by the model and the observed. Notice that θ is always positive, $\theta > 0$. There is a balance between different values of θ. For too small θ we may underestimate the true variability and influence of the drivers on $y(t)$, however, if θ it is too big, only the proximal points to the predictor center will influence the prediction, or too similar states will weight the prediction. We need to define operational criteria to choose the values of θ, and the typical are Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). d is the normalized Euclidean distance under the form:

$$d = \frac{1}{n} \sum_{i=1}^{n} \|x_k - x(t^*)\| \tag{A.14}$$

The solution of equation A.12 is simply the singular value decomposition (SVD) solution of the equation $\mathbf{B} = \mathbf{A} \cdot \mathbf{C}$, with:

$$\mathbf{B} = w_k y(t_i + n) \tag{A.15}$$

which is an n-dimensional vector for the future values of $y(t_i)$ and $\mathbf{A}$ is the matrix of dimension $n \times E$ entries given by:

$$\mathbf{A_{ij}} = w_k x_i(t_i) \tag{A.16}$$

When we solve the SVD problem, we estimate the forces of interaction of the multivariate SSR of the Equation A.12, the $C_i(t)$ terms. These estimates are time-series showing how the climate driver is forcing on the $y(t)$ time-series. A way to understand the effect of a driver on another time-series is to plot the estimate for the force of interaction versus the climate driver itself.

Distributed Lag Non-linear Models

Distributed Lag Non-linear Models (DLNM) is a framework developed in statistics and epidemiology to account for the effects of climate and environmental variables on an epidemiological response of interest [54]. The framework has been widely used to infer associations relations of environmental factors, such as pollution, ambient temperature, or any other exposure that can be accounted from a time-series. The framework works by applying the summary of all contributions accounted from the DLNM framework of any environmental exposure time-series to a time-series of counts of any epidemiological factor, as disease cases counts, all causes mortality counts, hospitalizations case counts, etc. From this we can infer the impact of exposure and its delayed effects on the response of interest. DLNM advanced the field of environmental epidemiology, giving a framework and a tool capable to deal with the delayed relation of environmental exposures and responses to it. Studies to account attributable all causes mortality due to exposure to such environmental factors have been one of the main research interests and uses of the DLNM [71, 70].

The first difficulty when dealing with environmental exposure in epidemiology is its complex relations with the epidemiological time-series of expected response. Although it is well known that are diverse mechanisms that can cause or aggravate any disease, or cause an augmented mortality, these effects have relations that can be delayed and that are nonlinear. This imposes the necessity, when building models for environmental epidemiology, to explicitly model this type of relations. If this type of relations are not considered in the models, it can be a source of overestimating or underestimating the real effect of such factors on the epidemiological curve of interest, depending on the type of relation [59, 70].

This appendix is dedicated to brief overview of the issues that arise when studying environmental exposures in epidemiology and how to overcome them with the DLNM framework. To do so, we start by understanding which are the problems when not counting for delayed effects when modelling environmental exposure-outcome relationship. After this we build the framework for Distributed Lag Models (DLM), the first approach attempted to model environmental exposures. Finally we generalize it to DLNM framework. This chapter ends with a brief section on interpretation of DLNM results, with an example from the study in the chapter 4.

The derivation given here is strongly based on the seminal papers by Gasparrini et al. [54, 55]. The reader is referenced, too, to another paper which gives a slightly different but equivalent construction for the DLNM framework here [58].

B.1 Issues on environmental exposures

Environmental exposures are complex in its effects on population, and it is not a simple task to model the effects of temperature, air population, to cite the main environmental exposures studied, on a population. The first difficulty is how to weight the temperature for a given location at a given time to a specific population. As discussed in chapter 2, the climate data itself are products from an entire scientific field, sometimes dedicated not to epidemiological modelling. The climate data use on environmental epidemiology are always supposed to be treated and wrangled before any modelling application.

The second issue is that environmental exposures will have different effects in different locations and/or different time. From an epidemiological point of view, this is a basic premise when studying any of the possible effects that environment can have on health. Temperature, air pollution, rainfall, etc., vary across time and geographic location and to illustrate that we can refer to the curve of weighted mean area temperatures for each state in Brazil 4.1. Although the distribution of temperature is not used as a component to model effects of temperature, it can give us insights on how big is the variance of this exposure per geographic location. Moreover, it is an issue to consider the temporal variation of the same exposure, which in our case it is the temperature, but any other environmental exposure suffer from this same issues. The following figure shows the distribution of temperature from 2010 to 2019, for each of the macro regions of Brazil. This makes the case to consider delayed effects on an outcome when constructing models for environmental epidemiology. It can be noticed, too, that some regions have less variance on their temperature distribution than others (see how the peak of the temperature distribution moves for the same region as we run over the years).

The main difficulty when modelling environmental exposures is to build models that are capable to capturing nonlinear effects as well as effects of delayed exposure. The main known effect of this kind is the 'harvesting' effect. 'Harvesting' happens when an exposure inflicts over a population of susceptible or frail to it, and has not only immediate effects but effects of continued exposing to it[120, 121]. Air pollution is one of the principal environmental exposures to have this type of effect, as well as temperature. Differently from only a delayed effect, 'harvesting' is the augmented risk to a continuous exposure during a certain time. The idea is that due to this continued exposing to the exposure, the effect will appear after a time and after a continued period of exposure. Delayed effects and 'harvesting' effects are normally strictly linked and maybe can not be untangled. This is why it is necessary to have a framework capable to model it altogether.

Another issue when modelling environmental exposures is the necessity to consider adaptation to the exposure. A population that is exposed frequently to the same level of air pollution or temperature will be, to a certain measure, adapted to it. It is very recently known how much this kind of adaptation can reduce the effects of exposures on a population, or can signal an aggravated risk to environmental changes[122, 123, 59, 124]. Taking the example of

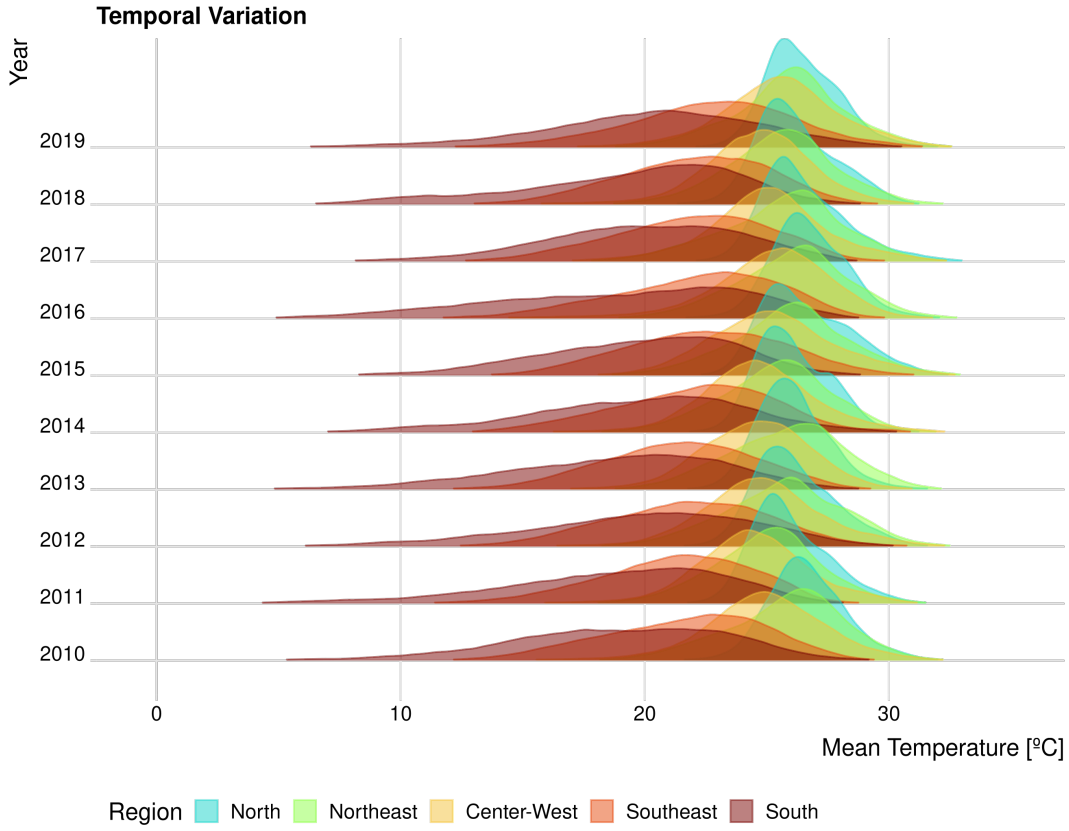

Figure B.1: Temporal variation on temperature over the years in Brazil, from 2010 to 2019.

climate change, more than the risks associated with new levels of exposure to high temperature, higher concentration of air pollution and/or extreme events, is the problem that this kind of phenomena can be far from the levels this population is adapted to it.

B.2 Distributed Lag Models (DLM)

One of the first approaches in environmental epidemiology to account for the effects of environmental variables into epidemiological time-series of interest was the Distributed Lag Models (DLM)[57]. It is a basic model taken from econometrics where the outcome is modelled with covariates at the same time of the outcome and with lags of the covariates, all of this is then modelled together into the model designed. Moreover, for an outcome Y_t at time t, DLM is constructed to model the environmental covariate at time, t, i.e. x_t, to account for its relation on immediate effects to the outcome and x_{t-j} for delayed effects, whose expression is:

$$\log[E[Y_t]] = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_j x_{t-j} + Covariates \quad , \tag{B.1}$$

where j runs over the time-lags considered. This is a model that can give unstable estimates to the effect of climate covariates on the outcome, because it models in an unconstrained way

the relation between outcome and exposure association. This first model does not makes use of smoothing functions.

A first improvement to DLM framework is to constrain the lagged covariate with a smoothing function, in a compact notation *Gasparrrini et al.* [54] gives the following expression for a general DLM model representation:

$$g(\mu_t) = \alpha + \sum_{j=1}^{J} s_j(x_{tj}; \boldsymbol{\beta}_j) + Covariates \tag{B.2}$$

This model is the same as in Equation B.1, g is a typical monotonic function of link and Y_t is supposed to be given by a distribution from the exponential family [125]. Here s_j is a smoothing function to relate the variable x_j and $\boldsymbol{\beta}_j$. The strategy to construct DLMs is to give a *basis functions* for the variables x_j, which are entered in the generalized linear model as *basis variables*[54]. A possible form for the s_j term is to expand the vector of variables x_t on *basis functions* (i.e., polynomials, Fourier transform, splines, etc.). This approach is the traditional one when modelling environmental exposures. It is simply given by a smooth form to the variables and constrained to avoid over and underestimating. s_j now can be written in the following form:

$$s(x_t; \boldsymbol{\beta}) = \mathbf{z}_t^T \cdot \boldsymbol{\beta} \quad , \tag{B.3}$$

where z_t is the *t-th* row of the matrix $\mathbf{Z} = f(\mathbf{x})$ constructed when applying the *basis functions* to the vector of exposures $\mathbf{x}$. If instead of inputting the exposure variables we input to the smoothing function one of the variables and its past values, x_{t-l}, we can build another *basis variables* that now model the delayed effects. This is the DLM, a predictor inserted in the model with the past values of an exposure expanded through a *basis function*, to model smoothly the relation between an outcome and an exposure and its past values. We have then:

$$s(x_t; \boldsymbol{\beta}) = \mathbf{q}_t^T \cdot \boldsymbol{\beta} \tag{B.4}$$

Here $\mathbf{q_t}$ is similar to the z_t, a matrix row, but $\mathbf{q_t}$ has the form of $\mathbf{q_t} = [x_t, \ldots, x_{t-L}]^T$. L is the maximum amount of lags considered. The matrix $\mathbf{Q}$ has the dimension of $n \times (L+1)$. It is straightforward to note that the first row of $\mathbf{Q}$ is the non-lagged vector of variables $\mathbf{x}$, $\mathbf{q_1} \equiv \mathbf{x}$. A general form to a DLM is to apply the *basis functions* to the vector of lags l:

$$s(x_t; \boldsymbol{\eta}) = \mathbf{q}_t^T \cdot \mathbf{C}\boldsymbol{\eta} = \mathbf{w}_t^T \boldsymbol{\eta} \quad , \tag{B.5}$$

where in the last passage we have written the product $\mathbf{q}_t^T \cdot \mathbf{C}$ in terms of the *basis functions* matrix elements, which is basically given by multiplying $\mathbf{Q}$ with $\mathbf{C}$, of the form $\mathbf{W} = \mathbf{Q} \cdot \mathbf{C}$. The *basis functions* matrix has the dimension of $n \times (L+1) \cdot (L+1) \times v_l \implies n \times v_l$. It is an expansion of the variables through *basis functions* v_l. This is a general case of a DLM. If we give the expression for $\mathbf{w_t}$ we define the DLM.

In a final full form we will have, for the complete DLM, the following equation:

$$g(\mu_t) = \alpha + \sum_{j=1}^{J} \mathbf{w}_{t,j}^T \boldsymbol{\eta} + Covariates \tag{B.6}$$

We can input any number of other *basis functions* modelling the relation of an environmental variable and its lags to an exposure of interest. This new DLM will now sum up over all the environmental variables as well as sum up over its lags considered.

The $\boldsymbol{\eta}$ are the unknown parameters that will be estimated. They have the following relation with the $\hat{\boldsymbol{\beta}} = \mathbf{C}\hat{\boldsymbol{\eta}}$ and by closure the variance-covariance matrices are related too, by the expression $V(\hat{\boldsymbol{\beta}}) = \mathbf{C}V(\hat{\boldsymbol{\eta}})\mathbf{C}^T$

B.3 Distributed Lag Non-linear Models

In the previous section, we defined a DLM as a model which included *basis functions* of the lagged exposure variables. We have written it in terms of the elements of a matrix that we called *cross-basis* at Equation B.2 and Equation B.5. A *cross-basis* can be generalized to a form that include smoothing functions with non-linear form and model altogether the lag-outcome relation and exposure-outcome relation is the aim of this section.

To have a more general approach while modelling environmental exposures, we want to have an expression for the second term of equation Equation B.2 that encompasses non-linear relations to the exposure variables and its lags. When we compute this new *basis functions* with simultaneous modelling forms, we build a *cross-basis functions* and produced a Distributed Lag Non-linear Model (DLNM). To do so, we basically take the product of the exposure-response structure, which expands the exposure variables by a *basis functions* $\mathbf{z}_t^T$ and the *basis functions* of the expanding shape for the lag-response structure given by the *basis functions* $\mathbf{q}_t^T$. This produces an array $n \times v_x \times (L+1)$ which gives the lagged effects of each exposure variables x_t on the outcome $g(\mu_t)$:

$$s(x_t; \boldsymbol{\eta}) = \sum_{j=1}^{v_x} \sum_{k=1}^{v_l} \mathbf{r}_{tj}^T \mathbf{c}_k \eta_{jk} = \mathbf{w}_t^T \boldsymbol{\eta} \tag{B.7}$$

Here $\mathbf{r}_{tj}^T$ are the delayed effects of the lag exposure at time t expanded on the *basis functions* j. As in the Equation B.5 we rewrite the product $\mathbf{r}_{tj} \cdot \mathbf{c}_k$ as a new variable $\mathbf{w}_t^T$, again the *cross-basis*. The $\mathbf{w}_t^T$ is obtained by applying the now *cross-basis function* $v_x \cdot v_l$ on the variables x_t. It is not straightforward to see on how the $\sum_{j=1}^{v_x} \sum_{k=1}^{v_l} \mathbf{r}_{tj}^T \mathbf{c}_k \eta_{jk}$ term transform into the compact form of the $\mathbf{w}_t^T$. To obtain the compact form in the last term of the Equation B.7, we have to write the inner term as:

$$\mathbf{A}_t = (\mathbf{1}_{v_l}^T \otimes \mathbf{R}_t) \odot (\mathbf{C} \otimes \mathbf{1}_{v_x}^T), \tag{B.8}$$

The array $\mathbf{A}_t$ can be seen as the Haddamard product, or the element-wise matrices product, $\odot$, between the tensorial products, $\otimes$, of the $\mathbf{R}_t$ matrix and the lag dimension and the $\mathbf{C}$ and the

variables dimension. Each tensorial product is basically producing the expansion of each matrix on the dimension used for the one matrices.

The term $\mathbf{1}_{v_l}^T \otimes \mathbf{R_t}$ is the expansion of the variables over the dimension of lags v_l. The term $\mathbf{C} \otimes \mathbf{1}_{v_x}^T$ is the smoothing function expanded over the dimension considered for the variables v_x. In the end of the day, this produced the expected *cross-basis*, the expansion of the variables and each of the lags considered, at the same time, over the *basis functions* for the lag and the *basis functions* for the shape over the variable dimension, this is the so called *exposure-lag-response* surface or structure. After summing up over the lag dimension, v_l, we obtain the final *cross-basis*, $\mathbf{W}$.

The main difference of this *cross-basis* in the Equation B.7 to the one on Equation B.5 is that we are expanding the variables and its lags at the same time. This is the innovation introduced by the DLNMs, by taking into account the effect of the variables and its lags at the same on the *cross-basis* function, we can reduce the terms inputted to the model and avoid unnecessary and time-consuming computations. This equation reduces to one in Equation B.5 if the *basis functions* for the variables, v_x, is a simple linear case. The innovation presented on DLNM, that it is of instead including the variables one by one and its structure modelled, we include them only by one operation and which is generalized, in the sense it can model non-linear relationship for the exposure-response and lag-response structures, now named exposure-lag-response.

Equation B.7 provideas a bi-dimensional surface, expanding on lag dimension and variable dimension, that is then inputted to the model and which for the model can produce estimates. At Figure B.2 we show how it can be visualized as the final product of the DLNM, which is the estimate for Relative Risk over two dimensions, lag-response and exposure-response. This is the result of putting into the model, as in the Equation B.5, a surface that explicitly models the relation of the exposure, the lag of this exposure through the days and from it can be associated to the risk.

B.4 Interpretation

This section briefly gives interpretations of the outputs generated from the DLNM framework for modelling environmental exposures.

DLNM, as mentioned above, builds a bi-dimensional grid to be entered to a model as predictors for the exposure-lag-response structure. This is done by expanding the exposure and its lags on *cross-basis functions,* which models simultaneously the structures of exposure-response and lag-response.

This grid could be understood as a surface of predictors that is entered to the model. When this surface is entered to the model it expresses the structure supposed for the data. From the construction of DLNM it guarantees that this surface only will have the terms that matter for the prediction and do not overfit the model or impose spurious correlations on the association. This is because the *cross-basis* provides a surface that only has the terms that survive the element-wise matrices product or, if seen by the alternative formulation for the DLNM, only those terms that have contributions to the dimensions of the bivariate function at the same time. Procedures for selecting the best shape fitted to the data can be applied. There are some

new implementations using these matrices products, as in Gasparrini et al. 2017 [126].

To visualize it, we take the surface plot from the output generated by the prediction given after the fit of a model with a *cross-basis* as predictor. This is a grid for the modelling relation between the temperature and its lags to the effects on the relative risk of hospitalizations in Brazil.

The standard output from the prediction on a *cross-basis* grid is centered on the 50^{th} percentile of the distribution, this is a good general assumption, but not all the time an environmental variable has the assumed shape. In environmental epidemiology, normally, the effects of any environmental variable are supposed to be modelled by a J-shape or U-shape, where the center of the distribution is the minimum point of effects of this variable on the outcome tested. In our study of Dengue hospitalization and temperature effects, chapter 4, this is not found. We found an increasing relation from the minimum temperature of hospitalization (MHT), which in our case is the minimum of the distribution of temperature. The prediction surface for the State of São Paulo (SP) can be visualized below:

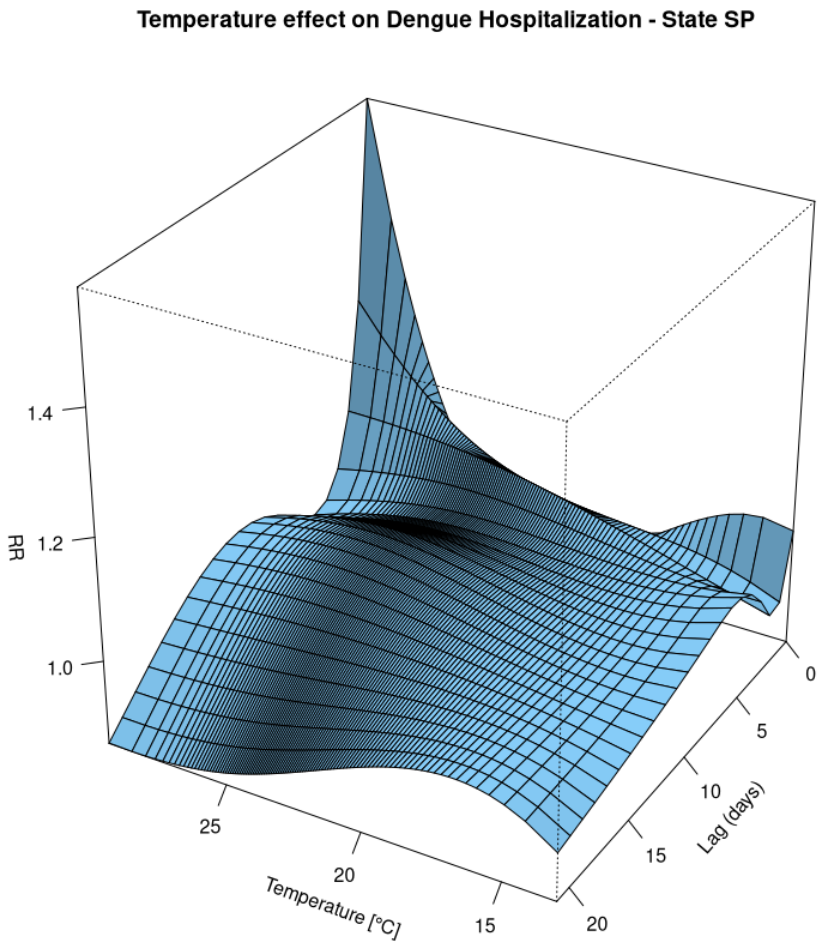

Figure B.2: Surface for the cross-basis prediction to Dengue Hospitalization RR as a function of temperature and its lags in the state of São Paulo.

This surface gives the RR to a Dengue hospitalization by temperature degrees and by days of lags. The prediction has a Credible Interval (CI), but it is not possible to give the CI representation together with the surface. We can slice this surface by specific temperature degrees and lags. When sitting on a specific lag day, we are looking for the effect of the range of temperatures with delayed effect to the amount of day we are sitting. In the contrary way, if we sit on a specific temperature degree, we are looking how this degree affects the RR through the length of lags expanded.

To illustrate this, we give a set of different lag specific views of the surface together with the CI. For the state of São Paulo, we give the lag lasting effects on 50^{th} and 95^{th} percentile of the temperature distribution and the overall effect, when summing over all the lags on the range of temperature distribution. We marked the percentiles on the overall cumulative effects plot as a traced vertical red line for the 95^{th} percentile and as a traced vertical blue line for the 50^{th} percentile.

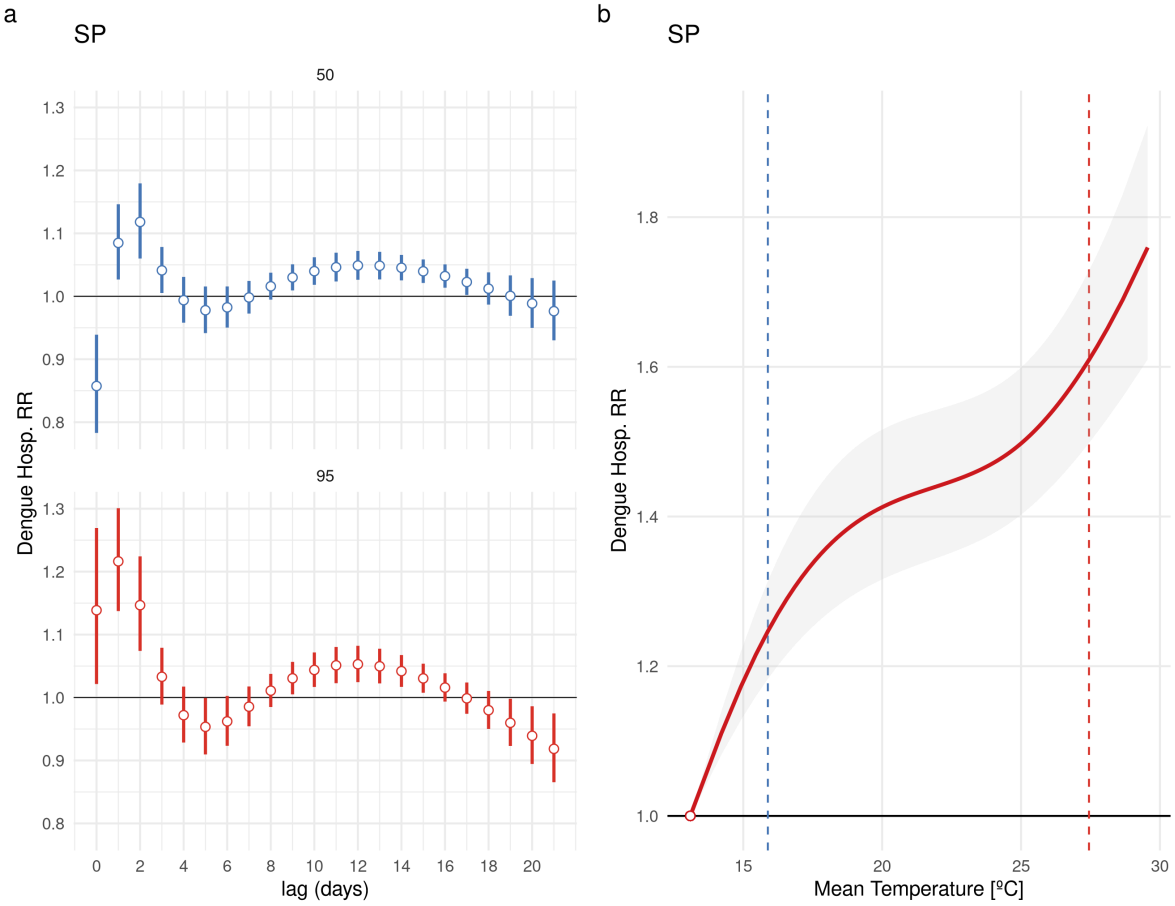

Figure B.3: Lag specific effects of temperature for the a) 50^{th} percentile and 95^{th}, and b) cumulative overall effects for all lags, in the state of São Paulo (SP) prediction.

The overall cumulative effect curve gives a measure of how much an exposure can affect an outcome associated to it. In this example, we see that for temperatures higher than the MHT, which for São Paulo state is 17.5º degrees Celsius, the RR to a Dengue hospitalization increases as the temperature increases. As said before, this is an expected but non-traditional environmental effect. After the Dengue infection started, when the exposed to a higher temperature will give more risk to evolve into a more severe Dengue infection.

Part VI

Supplementary Materials

C

Supplementary Material to the Chapter 3

Here we present the supplementary material to chapter 3, with supplementary figures and tables.

This supplementary material contains several supplementary figures: a figure showing the Dengue cases curve at Rio de Janeiro city, from 2010 to 2019, a figure of auto-correlation functions for all drivers tested, a figure for the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) produced when comparing the observed and predicted by the S-Map model in function of the value for the θ tested and a figure of the seasonality pattern generated from the spline fitting and yearly shuffle algorithm to produce surrogates.

Codes to reproduce the whole analysis are available at the repository at https://github.com/rafalopespx/dengue_ccm_cities

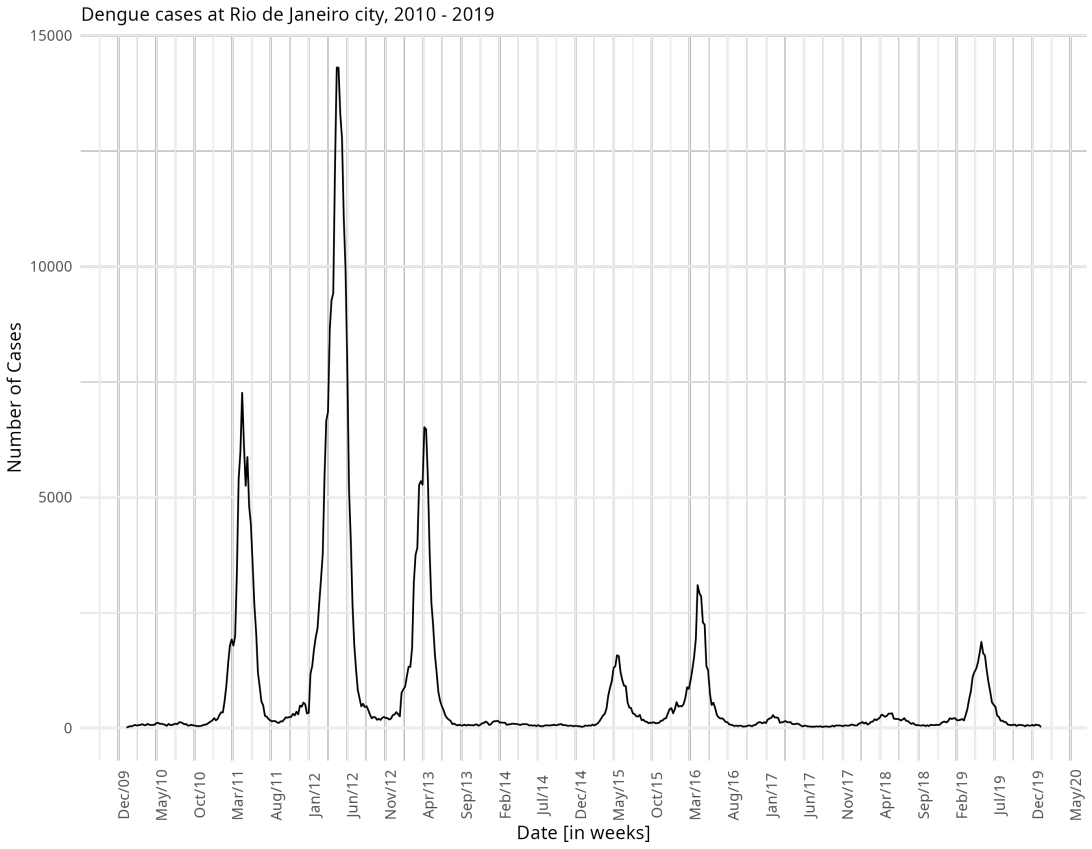

Figure C.1: Weekly cases of Dengue notified in the city of Rio de Janeiro, from 2010 to 2019

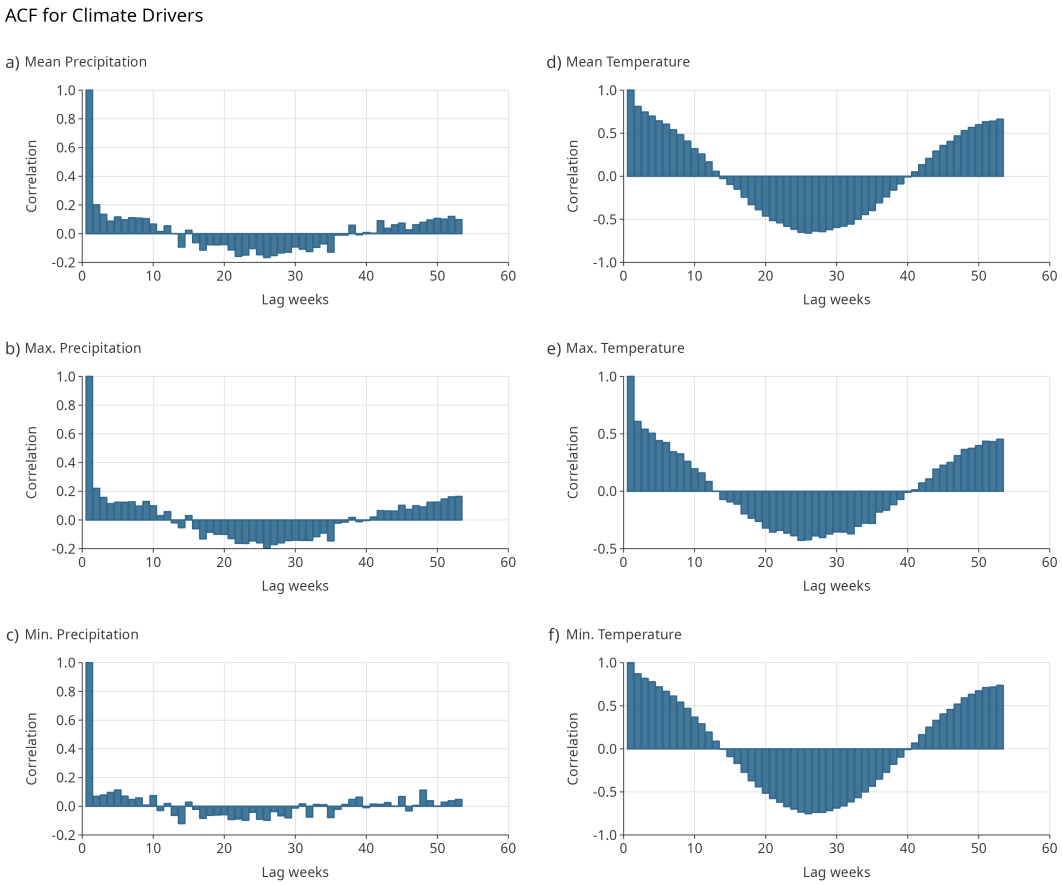

Figure C.2: Auto-correlation function for the climate drivers of the city of Rio de Janeiro

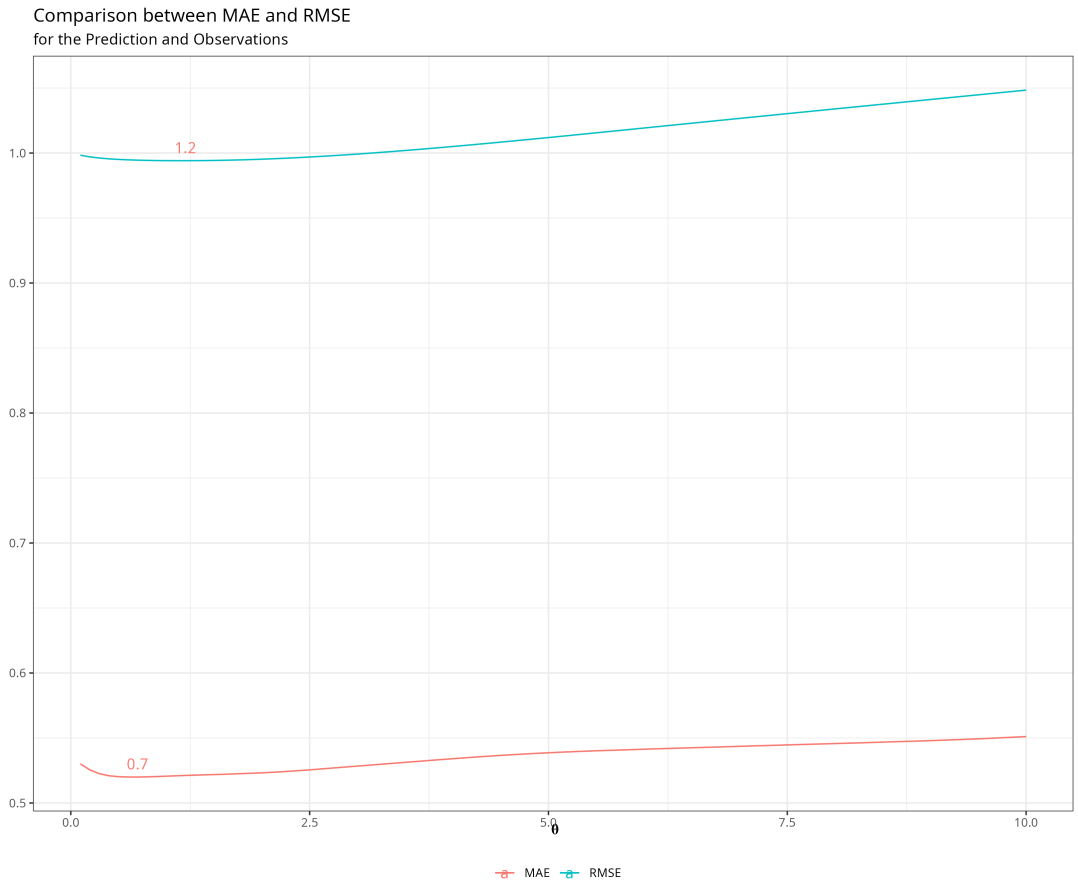

Figure C.3: Errors for the MAE and RMSE between predicted and observed on the S-Map model of precipitations only predicting cases

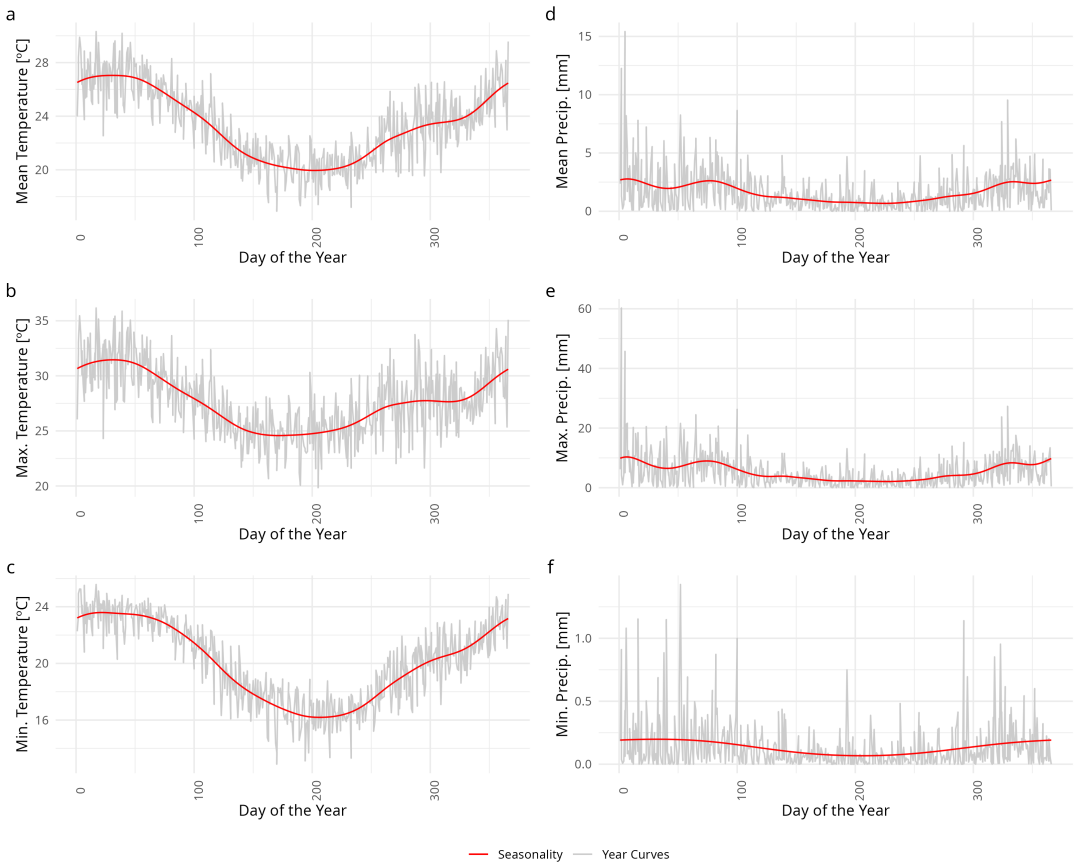

Figure C.4: Pattern of seasonality during the year length, for each of climate driver tested, grey shades are the variability around seasonality realized by year

D

Supplementary Material to the Chapter 4

This supplementary material contains supplementary tables and figures; a table with descriptive temperature for each state and region of Brazil during the analyzed period, a figure showing the Dengue hospitalizations curves by each macro region of Brazil, from 2010 to 2019, a figure of each state with the curve of RR for temperature and Dengue Hospitalisation RR, 5 figures for Dengue RR hospitalization and temperature by each macro regions, two figures with the Dengue hospitalization RR for the sensitivity analysis and a table comparing the RR found for the main analysis and the two sensitivity analysis ran.

Codes to reproduce the whole analysis is available at our repository at https://github.com/rafalopespx/dengue_t2m_severity_paper

Table S1 Descriptive distribution of 2m height Mean Temperature

State, Region, Brazil	Minimum	5th	25th	50th	75th	95th	Maximum
Brazil	-0.09	15.39	20.92	23.96	26.08	28.68	34.8
North	13.99	3.91	25.15	26.13	27.39	29.32	34.32
AC	13.99	22.81	24.35	25.13	26.01	27.49	31.01
AM	17.98	24.18	25.09	25.81	26.68	28.39	33.06
AP	21.98	23.72	24.73	25.65	27.14	28.59	30.5
PA	21.7	24.24	25.38	26.31	27.53	28.99	32.59
RO	14.11	23.34	24.67	25.52	26.6	28.59	31.61
RR	20.04	22.88	24.29	25.33	26.67	28.82	32.24
TO	19.6	24.17	25.58	26.71	28.03	30.27	34.32
Northeast	14.04	21.49	24.16	25.8	27.24	29.54	34.39
AL	18.6	21.76	23.52	25.02	26.32	27.9	32.78
BA	14.04	19.99	22.31	24.08	25.73	27.8	33.99
CE	19.09	23.82	25.58	26.77	27.81	29.16	32.41
MA	21.75	24.53	25.79	26.84	28.09	29.89	33.29
PB	19.17	22.36	24.19	25.49	26.74	29.15	32.08
PE	17.59	20.77	22.95	24.35	25.76	27.84	32.71
PI	20.71	24.44	26.07	27.55	29.3	31.41	34.39
RN	20.18	24.02	25.56	26.6	27.66	29.29	31.86
SE	19.77	22.17	23.74	25.35	26.53	27.77	32.25
Center-West	6.09	20.68	23.34	24.88	26.41	28.96	34.3
DF	15.96	19.42	21.01	22.23	23.54	25.87	29.58
GO	11.3	20.99	22.97	24.4	26	28.76	34.18
MS	6.09	16.95	22.42	24.71	26.46	28.81	34.26
MT	10.6	22.55	24.46	25.61	26.89	29.31	34.3
Southeast	4.69	16.13	19.7	22.13	24.31	27.03	34.8
ES	10.78	17.22	20.39	22.49	24.66	26.82	31.46
MG	6.71	16.15	19.47	21.76	23.82	26.62	34.54
SP	4.69	15.88	19.96	22.66	24.83	27.45	34.8
RJ	9.78	16.75	19.79	22.12	24.45	27.2	32.25
South	-0.09	11.03	16.52	19.92	22.76	26.02	33.03
PR	0.67	13.25	18.19	21.27	23.78	26.78	32.42
RS	0.77	9.99	15.71	19.33	22.35	25.7	33.03
SC	-0.09	10.8	15.86	18.98	21.73	24.89	31.13

Table D.1: Descriptive Table of temperature for States and regions

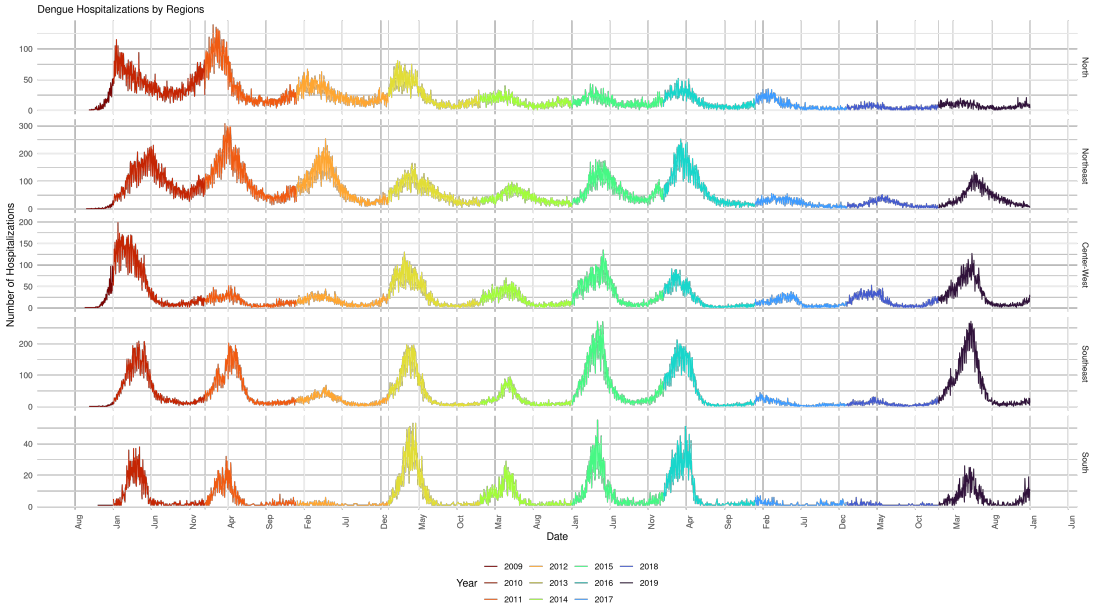

Figure D.1: Time series of Dengue hospitalisation by macro administrative region of Brazil. Colour by year, the data covers a period of 10 years, from the whole epidemiological year of 2010 to the whole epidemiological year of 2019.

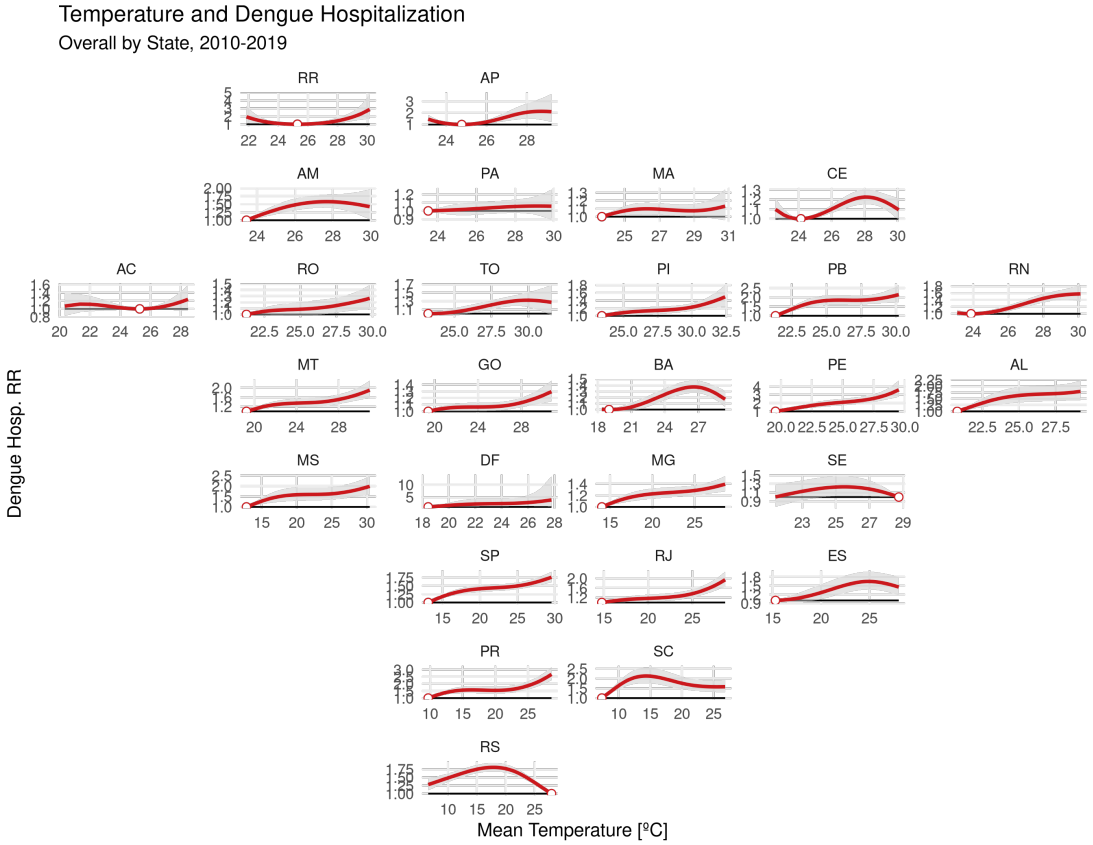

Figure D.2: Cumulative over all the lags RR to the MHT (MHT) on each state level, for the whole period of analysis, 2010 to 2019. The curve is plotted in red lines and the 95% confidence interval generated from the fitted model is given by the grey shaded ribbon around it. The title of each subplot is the abbreviations for the name of each state.

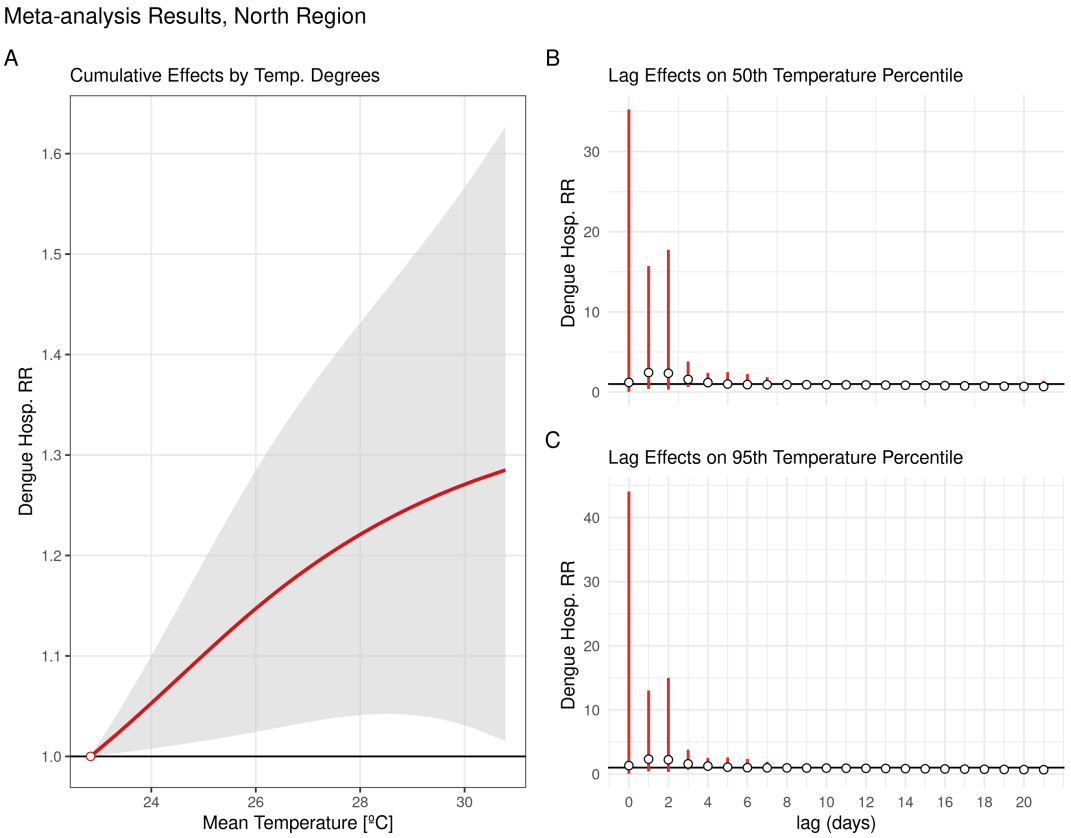

Figure D.3: A) Cumulative overall lags RR for a Dengue hospitalisation to the MHT association to the mean temperature distribution at the North Region. Vertical traced lines mark the 50th (Blue) and 95th (Red) percentile of the temperature distribution. The grey shade is 95% confidence interval derived from the meta-analysis. B) Lag effect to the RR to the MHT of Hospitalisation due to Dengue on the 50th percentile of temperature. C) Lag effect to the RR to the MHT of hospitalisation due to Dengue on the 95th percentile of temperature

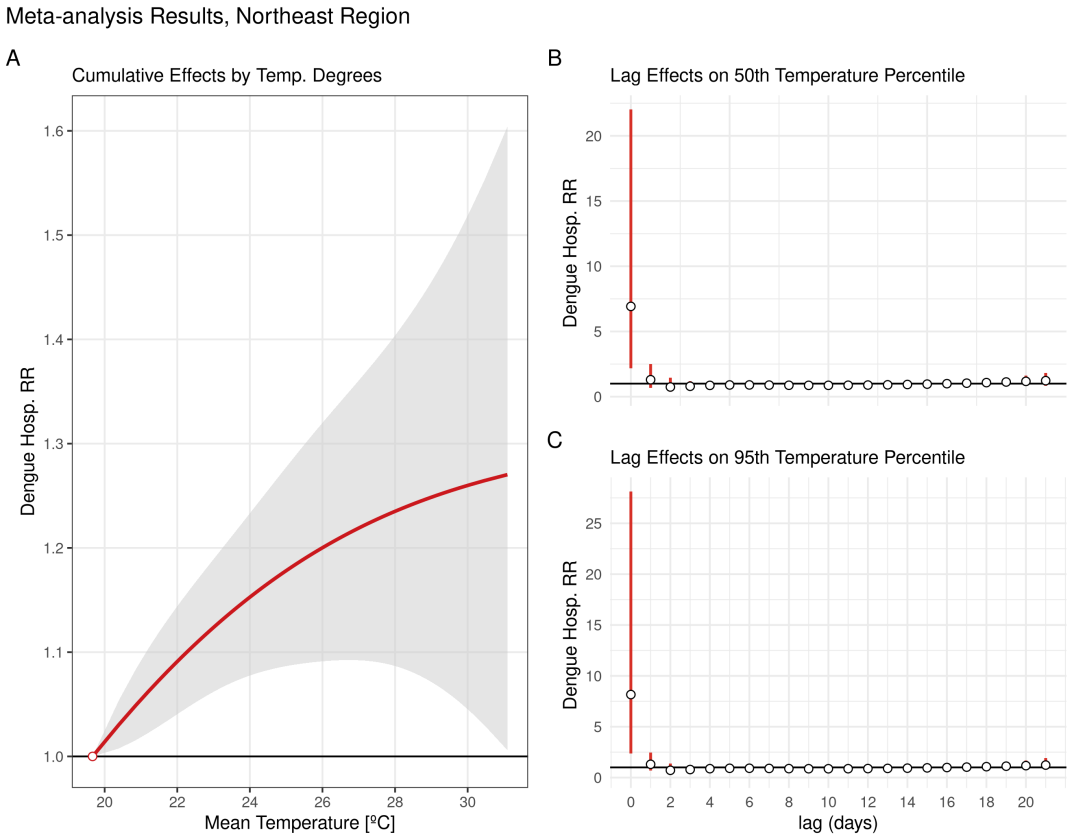

Figure D.4: A) Cumulative overall lags RR for a Dengue hospitalisation to the MHT association to the mean temperature distribution at the Northeast Region. Vertical traced lines mark the 50th (Blue) and 95th (Red) percentile of the temperature distribution. The grey shade is 95% confidence interval derived from the meta-analysis. B) Lag effect to the RR to the MHT of Hospitalisation due to Dengue on the 50th percentile of temperature. C) Lag effect to the RR to the MHT of hospitalisation due to Dengue on the 95th percentile of temperature

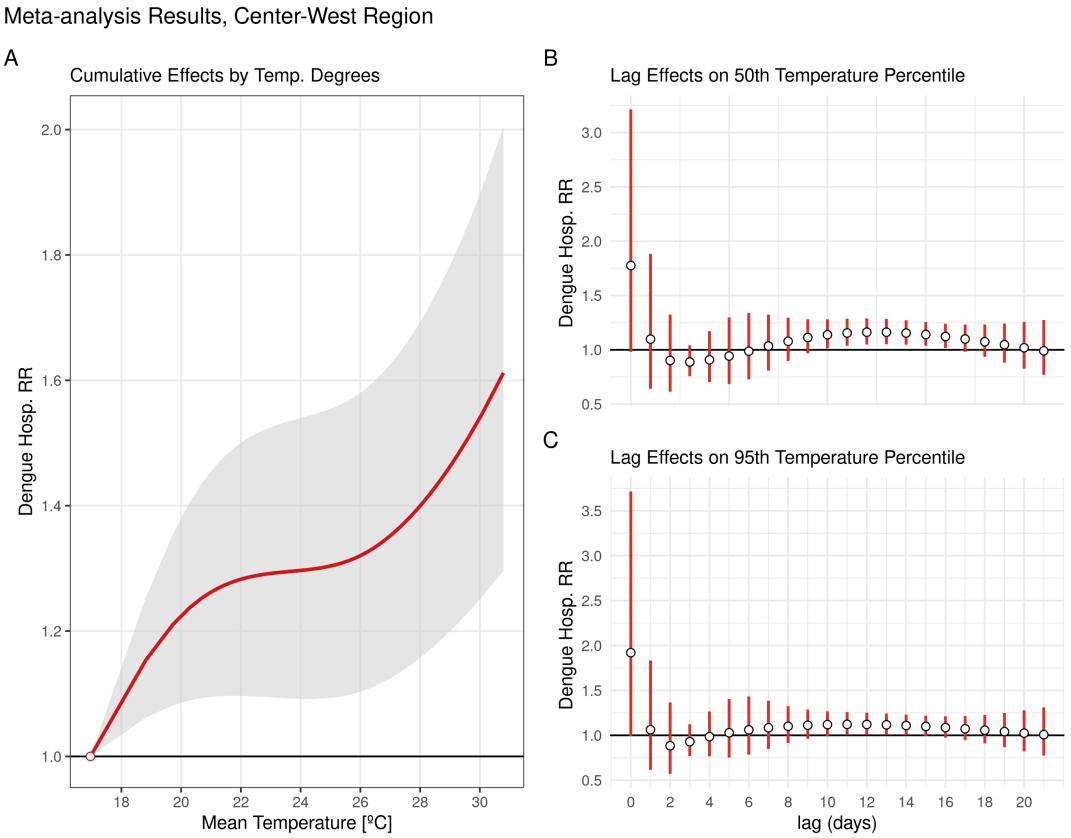

Figure D.5: A) Cumulative overall lags RR for a Dengue hospitalisation to the MHT association to the mean temperature distribution at the Center-West Region. Vertical traced lines mark the 50th (Blue) and 95th (Red) percentile of the temperature distribution. The grey shade is 95% confidence interval derived from the meta-analysis. B) Lag effect to the RR to the MHT of Hospitalisation due to Dengue on the 50th percentile of temperature. C) Lag effect to the RR to the MHT of hospitalisation due to Dengue on the 95th percentile of temperature

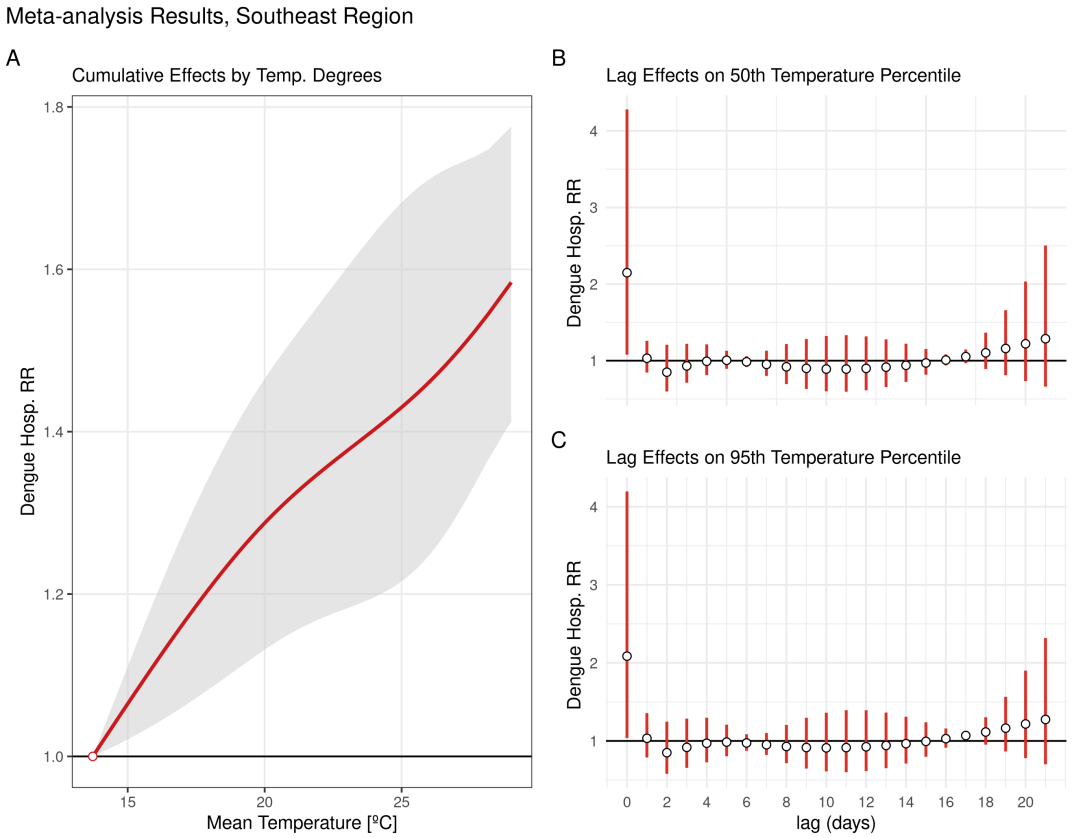

Figure D.6: A) Cumulative overall lags RR for a Dengue hospitalisation to the MHT association to the mean temperature distribution at the Southeast Region. Vertical traced lines mark the 50th (Blue) and 95th (Red) percentile of the temperature distribution. The grey shade is 95% confidence interval derived from the meta-analysis. B) Lag effect to the RR to the MHT of Hospitalisation due to Dengue on the 50th percentile of temperature. C) Lag effect to the RR to the MHT of hospitalisation due to Dengue on the 95th percentile of temperature

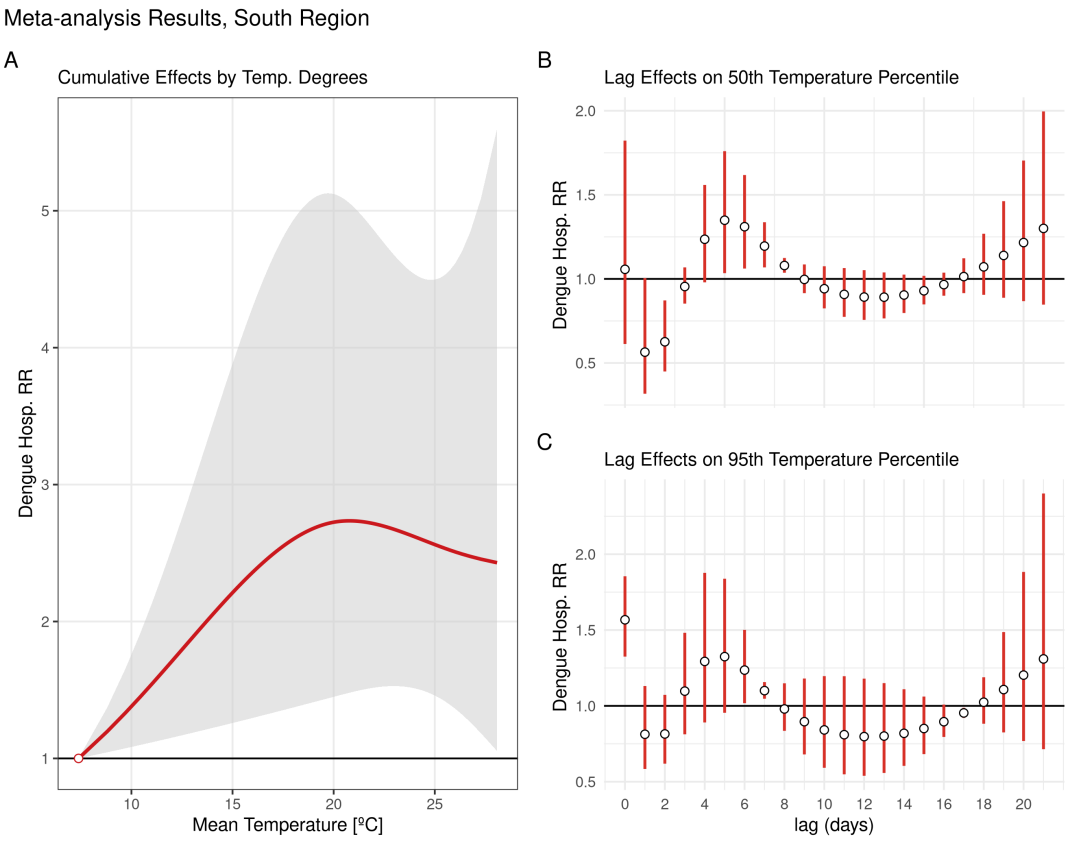

Figure D.7: A) Cumulative overall lags RR for a Dengue hospitalisation to the MHT association to the mean temperature distribution at the South Region. Vertical traced lines mark the 50th (Blue) and 95th (Red) percentile of the temperature distribution. The grey shade is 95% confidence interval derived from the meta-analysis. B) Lag effect to the RR to the MHT of Hospitalisation due to Dengue on the 50th percentile of temperature. C) Lag effect to the RR to the MHT of hospitalisation due to Dengue on the 95th percentile of temperature

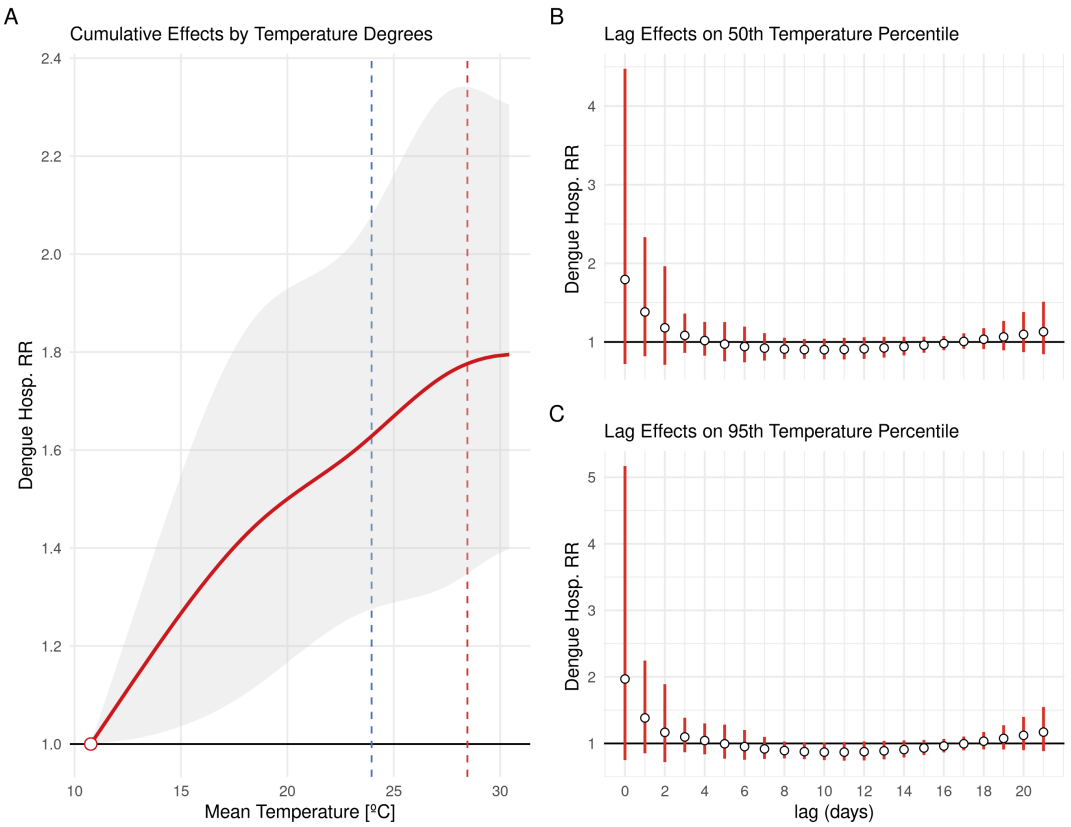

Figure D.8: A) Cumulative overall lags RR for a Dengue hospitalisation to the MHT association to the mean temperature distribution for the parametrization on Sensitivity Analysis 1 (dose-response: 3 knots equally spaced; lag-response: 3 knots equally spaced at the log-scale). Vertical traced lines mark the 50th (Blue) and 95th (Red) percentile of the temperature distribution. The grey shade is 95% confidence interval derived from the meta-analysis. B) Lag effect to the RR to the MHT of Hospitalisation due to Dengue on the 50th percentile of temperature. C) Lag effect to the RR to the MHT of hospitalisation due to Dengue on the 95th percentile of temperature

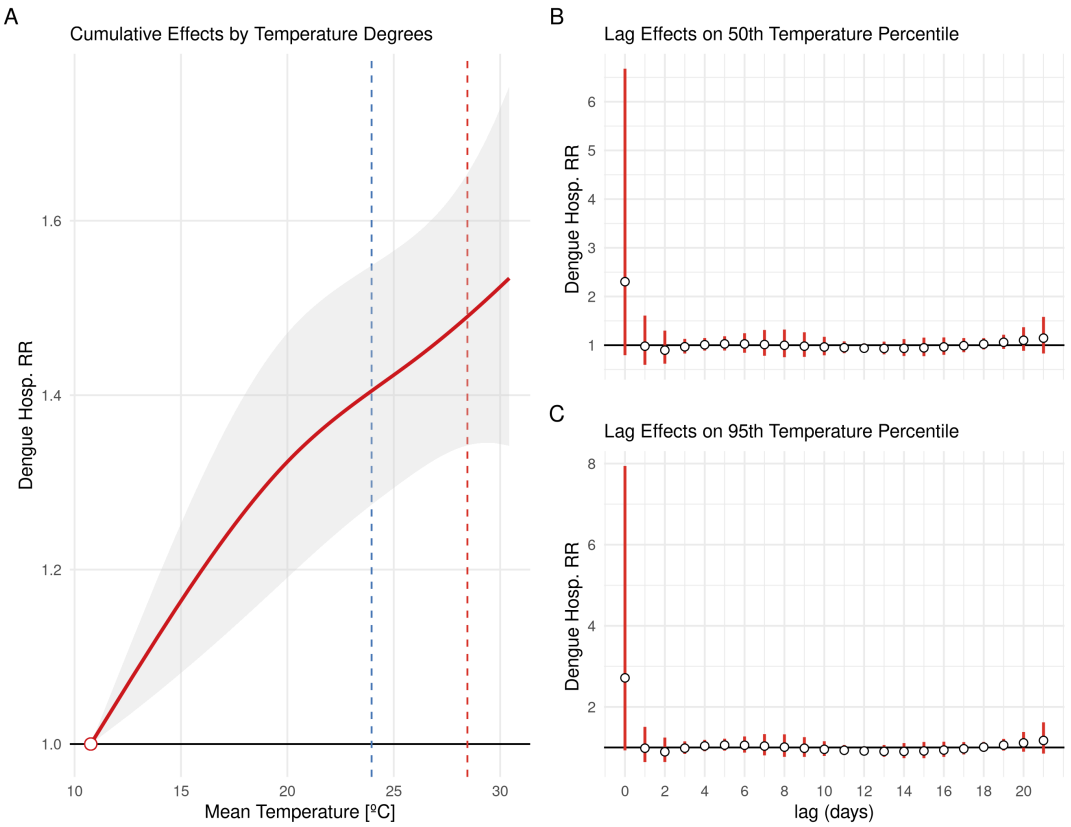

Figure D.9: A) Cumulative overall lags RR for a Dengue hospitalisation to the MHT association to the mean temperature distribution for the parametrization on Sensitivity Analysis 1 (dose-response: 3 knots equally spaced; lag-response: 3 knots equally spaced at the log-scale). Vertical traced lines mark the 50th (Blue) and 95th (Red) percentile of the temperature distribution. The grey shade is 95% confidence interval derived from the meta-analysis. B) Lag effect to the RR to the MHT of Hospitalisation due to Dengue on the 50th percentile of temperature. C) Lag effect to the RR to the MHT of hospitalisation due to Dengue on the 95th percentile of temperature

Main Analysis			
Unit	**MHT (°C)**	**RR (IC 95%) 50th**	**RR (IC 95%) 95th**
Brazil	10.8	1.404 (1.274-1.548)	1.491 (1.337-1.664)
North	22.8	1.153 (1.026-1.295)	1.256 (1.039-1.518)
Northeast	19.7	1.196 (1.091-1.312)	1.255 (1.059-1.488)
Center-West	17	1.303 (1.093-1.553)	1.460 (1.198-1.779)
Southeast	13.7	1.353 (1.171-1.563)	1.498 (1.298-1.729)
South	7.4	2.723 (1.447-5.125)	2.508 (1.362-4.618)
Sensitivity Analysis 1			
Unit	**MHT (°C)**	**RR (IC 95%) 50th**	**RR (IC 95%) 95th**
Brazil	10.8	1.629 (1.276-2.079)	1.780 (1.354-2.339)
North	22.8	1.080 (1.002-1.163)	1.077 (0.871-1.332)
Northeast	19.7	1.199 (1.094-1.315)	1.276 (1.066-1.526)
Center-West	17	1.485 (1.177-1.874)	1.652 (1.251-2.181)
Southeast	13.7	1.456 (1.234-1.717)	1.706 (1.420-2.049)
South	7.4	28.704 (4.719-174.587)	3.152 (5.102-215.429)
Sensitivity Analysis 2			
Unit	**MHT (°C)**	**RR (IC 95%) 50th**	**RR (IC 95%) 95th**
Brazil	10.8	1.405 (1.274-1.549)	1.495 (1.344-1.663)
North	22.8	1.144 (1.021-1.282)	1.251 (1.074-1.456)
Northeast	19.7	1.194 (1.089-1.309)	1.251 (1.060-1.475)
Center-West	17	1.301 (1.097-1.544)	1.454 (1.198-1.765)
Southeast	13.7	1.407 (1.217-1.628)	1.556 (1.342-1.804)
South	7.4	2.994 (1.627-5.510)	2.764 (1.617-4.723)

Table D.2: Dengue hospitalisation relative risk by Brazil and each macro-region: main and sensitivity analyses, * Relative Risk to the Minimum Hospitalisation Temperature of hospitalisation due to Dengue, ** at 50th (23·96° C) and at 95th (28·68° C)

Bibliography

[1] J. B. Siqueira Jr, C. M. T. Martelli, G. E. Coelho, A. C. da Rocha Simplício, and D. L. Hatch, "Dengue and dengue hemorrhagic fever, brazil, 1981–2002," *Emerging infectious diseases*, vol. 11, no. 1, p. 48, 2005.

[2] M. G. Guzman and E. Harris, "Dengue," *The Lancet*, vol. 385, no. 9966, pp. 453–465, 2015.

[3] M. G. Guzman, D. J. Gubler, A. Izquierdo, E. Martinez, and S. B. Halstead, "Dengue infection," *Nat Rev Dis Primers*, vol. 2, pp. 1–25, Aug. 2016. Number: 1 Publisher: Nature Publishing Group.

[4] T. S. Salles, T. da Encarnação Sá-Guimarães, E. S. L. de Alvarenga, V. Guimarães-Ribeiro, M. D. F. de Meneses, P. F. de Castro-Salles, C. R. Dos Santos, A. C. do Amaral Melo, M. R. Soares, D. F. Ferreira, *et al.*, "History, epidemiology and diagnostics of dengue in the american and brazilian contexts: a review," *Parasites & vectors*, vol. 11, no. 1, pp. 1–12, 2018.

[5] S. Bhatt, P. W. Gething, O. J. Brady, J. P. Messina, A. W. Farlow, C. L. Moyes, J. M. Drake, J. S. Brownstein, A. G. Hoen, O. Sankoh, M. F. Myers, D. B. George, T. Jaenisch, G. R. W. Wint, C. P. Simmons, T. W. Scott, J. J. Farrar, and S. I. Hay, "The global distribution and burden of dengue," *Nature*, vol. 496, pp. 504–507, Apr. 2013.

[6] J. A. Suaya, D. S. Shepard, J. B. Siqueira, C. T. Martelli, L. C. S. Lum, L. H. Tan, S. Kongsin, S. Jiamton, F. Garrido, R. Montoya, B. Armien, R. Huy, L. Castillo, M. Caram, B. K. Sah, R. Sughayyar, K. R. Tyo, and S. B. Halstead, "Cost of dengue cases in eight countries in the Americas and Asia: a prospective study," *Am J Trop Med Hyg*, vol. 80, pp. 846–855, May 2009.

[7] P. M. Luz, B. Grinsztejn, and A. Galvani, "Disability adjusted life years lost to dengue in brazil," *Tropical Medicine & International Health*, vol. 14, no. 2, pp. 237–246, 2009.

[8] D. C. Andrioli, M. A. Busato, and J. A. Lutinski, "Spatial and temporal distribution of dengue in brazil, 1990-2017," *PLoS One*, vol. 15, no. 2, p. e0228346, 2020.

[9] L. T. M. Figueiredo, "Dengue in brazil during 1999–2009: a review.," 2010.

[10] G. L. Ferreira, "Global dengue epidemiology trends," *Revista do Instituto de Medicina Tropical de São Paulo*, vol. 54, pp. 5–6, 2012.

[11] O. B. Dick, J. L. San Martín, R. H. Montoya, J. del Diego, B. Zambrano, and G. H. Dayan, "The history of dengue outbreaks in the americas," *The American journal of tropical medicine and hygiene*, vol. 87, no. 4, p. 584, 2012.

[12] O. J. Brady, P. W. Gething, S. Bhatt, J. P. Messina, J. S. Brownstein, A. G. Hoen, C. L. Moyes, A. W. Farlow, T. W. Scott, and S. I. Hay, "Refining the global spatial limits of dengue virus transmission by evidence-based consensus," 2012.

[13] B. Zambrano and J. L. San Martin, "Epidemiology of dengue in latin america," *Journal of the Pediatric Infectious Diseases Society*, vol. 3, no. 3, pp. 181–182, 2014.

[14] W. H. Organization, S. P. for Research, T. in Tropical Diseases, W. H. O. D. of Control of Neglected Tropical Diseases, W. H. O. Epidemic, and P. Alert, *Dengue: guidelines for diagnosis, treatment, prevention and control*. World Health Organization, 2009.

[15] M. U. Kraemer, M. E. Sinka, K. A. Duda, A. Q. Mylne, F. M. Shearer, C. M. Barker, C. G. Moore, R. G. Carvalho, G. E. Coelho, W. Van Bortel, *et al.*, "The global distribution of the arbovirus vectors aedes aegypti and ae. albopictus," *elife*, vol. 4, p. e08347, 2015.

[16] F. Ding, J. Fu, D. Jiang, M. Hao, and G. Lin, "Mapping the spatial distribution of aedes aegypti and aedes albopictus," *Acta tropica*, vol. 178, pp. 155–162, 2018.

[17] C. Barcellos and R. Lowe, "Expansion of the dengue transmission area in b razil: the role of climate and cities," *Tropical Medicine & International Health*, vol. 19, no. 2, pp. 159–168, 2014.

[18] E. A. Mordecai, J. M. Cohen, M. V. Evans, P. Gudapati, L. R. Johnson, C. A. Lippi, K. Miazgowicz, C. C. Murdock, J. R. Rohr, S. J. Ryan, *et al.*, "Detecting the impact of temperature on transmission of zika, dengue, and chikungunya using mechanistic models," *PLoS neglected tropical diseases*, vol. 11, no. 4, p. e0005568, 2017.

[19] E. A. Mordecai, J. M. Caldwell, M. K. Grossman, C. A. Lippi, L. R. Johnson, M. Neira, J. R. Rohr, S. J. Ryan, V. Savage, M. S. Shocket, *et al.*, "Thermal biology of mosquito-borne disease," *Ecology letters*, vol. 22, no. 10, pp. 1690–1708, 2019.

[20] R. Lowe, S. A. Lee, K. M. O'Reilly, O. J. Brady, L. Bastos, G. Carrasco-Escobar, R. de Castro Catão, F. J. Colón-González, C. Barcellos, M. S. Carvalho, M. Blangiardo, H. Rue, and A. Gasparrini, "Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: a spatiotemporal modelling study," *The Lancet Planetary Health*, vol. 5, pp. e209–e219, Apr. 2021.

[21] S. A. Lee, T. Economou, R. de Castro Catão, C. Barcellos, and R. Lowe, "The impact of climate suitability, urbanisation, and connectivity on the expansion of dengue in 21st century Brazil," *PLoS Negl Trop Dis*, vol. 15, p. e0009773, Dec. 2021.

[22] N. Nova, E. R. Deyle, M. S. Shocket, A. J. MacDonald, M. L. Childs, M. Rypdal, G. Sugihara, and E. A. Mordecai, "Susceptible host availability modulates climate effects on dengue dynamics," *Ecology letters*, vol. 24, no. 3, pp. 415–425, 2021.

[23] O. J. Brady and S. I. Hay, "The global expansion of dengue: how aedes aegypti mosquitoes enabled the first pandemic arbovirus," *Annual review of entomology*, vol. 65, pp. 191–208, 2020.

[24] J. P. Messina, O. J. Brady, T. W. Scott, C. Zou, D. M. Pigott, K. A. Duda, S. Bhatt, L. Katzelnick, R. E. Howes, K. E. Battle, *et al.*, "Global spread of dengue virus types: mapping the 70 year history," *Trends in microbiology*, vol. 22, no. 3, pp. 138–146, 2014.

[25] M. E. Wilson and L. H. Chen, "Dengue: update on epidemiology," *Current infectious disease reports*, vol. 17, no. 1, pp. 1–8, 2015.

[26] M. Burattini, L. Lopez, F. Coutinho, J. Siqueira-Jr, S. Homsani, E. Sarti, and E. Massad, "Age and regional differences in clinical presentation and risk of hospitalization for dengue in Brazil, 2000-2014," *Clinics*, vol. 71, pp. 455–463, Aug. 2016.

[27] M. A. Johansson, J. Hombach, and D. A. Cummings, "Models of the impact of dengue vaccines: A review of current research and potential approaches," *Vaccine*, vol. 29, no. 35, pp. 5860–5868, 2011.

[28] D. J. Gubler and M. Meltzer, "Impact of dengue/dengue hemorrhagic fever on the developing world," *Advances in virus research*, vol. 53, pp. 35–70, 1999.

[29] D. P. Webster, J. Farrar, and S. Rowland-Jones, "Progress towards a dengue vaccine," *The Lancet infectious diseases*, vol. 9, no. 11, pp. 678–687, 2009.

[30] S.-Q. Deng, X. Yang, Y. Wei, J.-T. Chen, X.-J. Wang, and H.-J. Peng, "A review on dengue vaccine development," *Vaccines*, vol. 8, no. 1, p. 63, 2020.

[31] B. Guy, J. Lang, M. Saville, and N. Jackson, "Vaccination against dengue: challenges and current developments," *Annual review of medicine*, vol. 67, pp. 387–404, 2016.

[32] D. d. V. E. Ministério da Saúde, Secretaria de Vigilância em Saúde, *Sistema de Informação de Agravos de Notificação, Normas e Rotinas*.

[33] World Health Organization, "Dengue and severe dengue."

[34] G. E. Coelho, P. L. Leal, M. d. P. Cerroni, A. C. R. Simplicio, and J. B. Siqueira, "Sensitivity of the Dengue Surveillance System in Brazil for Detecting Hospitalized Cases," *PLoS Negl Trop Dis*, vol. 10, p. e0004705, May 2016.

[35] C. Codeco, F. Coelho, O. Cruz, S. Oliveira, T. Castro, and L. Bastos, "Infodengue: A nowcasting system for the surveillance of arboviruses in brazil," *Revue d'Épidémiologie et de Santé Publique*, vol. 66, p. S386, 2018.

[36] S. B. Halstead, "Dengue virus–mosquito interactions," *Annu. Rev. Entomol.*, vol. 53, pp. 273–291, 2008.

[37] M. M. Silva, M. S. Rodrigues, I. A. Paploski, M. Kikuti, A. M. Kasper, J. S. Cruz, T. L. Queiroz, A. S. Tavares, P. M. Santana, J. M. Araújo, *et al.*, "Accuracy of dengue reporting by national surveillance system, brazil," *Emerging infectious diseases*, vol. 22, no. 2, p. 336, 2016.

[38] C. T. Codeco, S. S. Oliveira, D. A. C. Ferreira, T. I. S. Riback, L. S. Bastos, R. M. Lana, I. Almeida, V. B. Godinho, O. G. Cruz, and F. C. Coelho, "Fast expansion of dengue in Brazil," *The Lancet Regional Health - Americas*, vol. 12, p. 100274, 2022.

[39] I. Rodriguez-Barraquer, M. T. Cordeiro, C. Braga, W. V. De Souza, E. T. Marques, and D. A. Cummings, "From re-emergence to hyperendemicity: the natural history of the dengue epidemic in brazil," *PLoS neglected tropical diseases*, vol. 5, no. 1, p. e935, 2011.

[40] Ministério da Saúde, "Lista nacional de notificação compulsória de doenças, agravos e eventos de saúde pública," 2022.

[41] L. Cattarino, I. Rodriguez-Barraquer, N. Imai, D. A. Cummings, and N. M. Ferguson, "Mapping global variation in dengue transmission intensity," *Science translational medicine*, vol. 12, no. 528, p. eaax4144, 2020.

[42] L. A. Castro, N. Generous, W. Luo, A. Pastore y Piontti, K. Martinez, M. F. Gomes, D. Osthus, G. Fairchild, A. Ziemann, A. Vespignani, *et al.*, "Using heterogeneous data to identify signatures of dengue outbreaks at fine spatio-temporal scales across brazil," *PLoS neglected tropical diseases*, vol. 15, no. 5, p. e0009392, 2021.

[43] A. F. Brito, L. C. Machado, R. J. Oidtman, M. J. L. Siconelli, Q. M. Tran, J. R. Fauver, R. D. d. O. Carvalho, F. Z. Dezordi, M. R. Pereira, L. A. de Castro-Jorge, *et al.*, "Lying in wait: the resurgence of dengue virus after the zika epidemic in brazil," *Nature Communications*, vol. 12, no. 1, pp. 1–13, 2021.

[44] N. R. Faria, J. Quick, I. Claro, J. Theze, J. G. de Jesus, M. Giovanetti, M. U. Kraemer, S. C. Hill, A. Black, A. C. da Costa, *et al.*, "Establishment and cryptic transmission of zika virus in brazil and the americas," *Nature*, vol. 546, no. 7658, pp. 406–410, 2017.

[45] C. Moreno, *O Brasil Made in China.* Fundação Rosa Luxemburgo, 2015.

[46] J. H. Huber, M. L. Childs, J. M. Caldwell, and E. A. Mordecai, "Seasonal temperature variation influences climate suitability for dengue, chikungunya, and zika transmission," *PLoS neglected tropical diseases*, vol. 12, no. 5, p. e0006451, 2018.

[47] A. S. Siraj, R. J. Oidtman, J. H. Huber, M. U. Kraemer, O. J. Brady, M. A. Johansson, and T. A. Perkins, "Temperature modulates dengue virus epidemic growth rates through its effects on reproduction numbers and generation intervals," *PLoS neglected tropical diseases*, vol. 11, no. 7, p. e0005797, 2017.

[48] N. B. Tjaden, S. M. Thomas, D. Fischer, and C. Beierkuhnlein, "Extrinsic incubation period of dengue: knowledge, backlog, and applications of temperature dependence," *PLoS neglected tropical diseases*, vol. 7, no. 6, p. e2207, 2013.

[49] V. Romeo-Aznar, L. Picinini Freitas, O. Gonçalves Cruz, A. A. King, and M. Pascual, "Fine-scale heterogeneity in population density predicts wave dynamics in dengue epidemics," *Nature communications*, vol. 13, no. 1, pp. 1–9, 2022.

[50] J. M. Caldwell, A. D. LaBeaud, E. F. Lambin, A. M. Stewart-Ibarra, B. A. Ndenga, F. M. Mutuku, A. R. Krystosik, E. B. Ayala, A. Anyamba, M. J. Borbor-Cordova, *et al.*, "Climate predicts geographic and temporal variation in mosquito-borne disease dynamics on two continents," *Nature communications*, vol. 12, no. 1, pp. 1–13, 2021.

[51] M. R. Donalisio, "O dengue no espaço habitado," in *O dengue no espaço habitado*, pp. 195–195, 1999.

[52] E. R. Deyle and G. Sugihara, "Generalized theorems for nonlinear state space reconstruction," *Plos one*, vol. 6, no. 3, p. e18295, 2011.

[53] G. Sugihara, R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, and S. Munch, "Detecting causality in complex ecosystems," *science*, vol. 338, no. 6106, pp. 496–500, 2012.

[54] A. Gasparrini, B. Armstrong, and M. G. Kenward, "Distributed lag non-linear models," *Statist. Med.*, vol. 29, pp. 2224–2234, Sept. 2010.

[55] A. Gasparrini, "Distributed Lag Linear and Non-Linear Models in R: The Package dlnm," *J Stat Softw*, vol. 43, pp. 1–20, July 2011.

[56] A. Gasparrini, B. Armstrong, and M. G. Kenward, "Multivariate meta-analysis for non-linear and other multi-parameter associations," *Statist. Med.*, vol. 31, pp. 3821–3839, Dec. 2012.

[57] A. Gasparrini and B. Armstrong, "Reducing and meta-analysing estimates from distributed lag non-linear models," *BMC Medical Research Methodology*, vol. 13, p. 1, Jan. 2013.

[58] A. Gasparrini, "Modeling exposure–lag–response associations with distributed lag non-linear models," *Statistics in medicine*, vol. 33, no. 5, pp. 881–899, 2014.

[59] A. Gasparrini, Y. Guo, F. Sera, A. M. Vicedo-Cabrera, V. Huber, S. Tong, M. d. S. Z. S. Coelho, P. H. N. Saldiva, E. Lavigne, P. M. Correa, *et al.*, "Projections of temperature-related excess mortality under climate change scenarios," *The Lancet Planetary Health*, vol. 1, no. 9, pp. e360–e367, 2017.

[60] G. Sugihara and R. M. May, "Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series," *Nature*, vol. 344, no. 6268, pp. 734–741, 1990.

[61] E. R. Deyle, R. M. May, S. B. Munch, and G. Sugihara, "Tracking and forecasting ecosystem interactions in real time," *Proceedings of the Royal Society B: Biological Sciences*, vol. 283, no. 1822, p. 20152258, 2016.

[62] M. Ushio, C.-h. Hsieh, R. Masuda, E. R. Deyle, H. Ye, C.-W. Chang, G. Sugihara, and M. Kondoh, "Fluctuating interaction network and time-varying stability of a natural fish community," *Nature*, vol. 554, no. 7692, pp. 360–363, 2018.

[63] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical systems and turbulence, Warwick 1980*, pp. 366–381, Springer, 1981.

[64] F. Takens, "Detecting nonlinearities in stationary time series," *International Journal of Bifurcation and Chaos*, vol. 3, no. 02, pp. 241–256, 1993.

[65] K. Laneri, B. Cabella, P. I. Prado, R. Mendes Coutinho, and R. A. Kraenkel, "Climate drivers of malaria at its southern fringe in the americas," *PloS one*, vol. 14, no. 7, p. e0219249, 2019.

[66] M. Rypdal and G. Sugihara, "Inter-outbreak stability reflects the size of the susceptible pool and forecasts magnitudes of seasonal epidemics," *Nature communications*, vol. 10, no. 1, pp. 1–8, 2019.

[67] Martínez-Solanas and X. Basagaña, "Temporal changes in the effects of ambient temperatures on hospital admissions in Spain," *PLoS ONE*, vol. 14, p. e0218262, June 2019.

[68] D. Royé, C. Íñiguez, and A. Tobías, "Comparison of temperature–mortality associations using observed weather station and reanalysis data in 52 Spanish cities," *Environmental Research*, vol. 183, p. 109237, Apr. 2020.

[69] J. Barrera-Gomez and X. Basagana, "Using the R package collin to visualize the effects of collinearity in distributed lag models," p. 26.

[70] A. M. Vicedo-Cabrera, N. Scovronick, F. Sera, D. Royé, R. Schneider, A. Tobias, C. Astrom, Y. Guo, Y. Honda, D. Hondula, *et al.*, "The burden of heat-related mortality attributable to recent human-induced climate change," *Nature climate change*, vol. 11, no. 6, pp. 492–500, 2021.

[71] A. M. Vicedo-Cabrera, A. Tobias, J. J. Jaakkola, Y. Honda, M. Hashizume, Y. Guo, J. Schwartz, A. Zanobetti, M. L. Bell, B. Armstrong, *et al.*, "Global mortality burden attributable to non-optimal temperatures," *The lancet*, vol. 399, no. 10330, p. 1113, 2022.

[72] M. N. Mistry, R. Schneider, P. Masselot, D. Royé, B. Armstrong, J. Kyselỳ, H. Orru, F. Sera, S. Tong, É. Lavigne, *et al.*, "Comparison of weather station and climate reanalysis data for modelling temperature-related mortality," *Scientific reports*, vol. 12, no. 1, pp. 1–14, 2022.

[73] J. Muñoz-Sabater, E. Dutra, A. Agustí-Panareda, C. Albergel, G. Arduini, G. Balsamo, S. Boussetta, M. Choulga, S. Harrigan, H. Hersbach, B. Martens, D. G. Miralles, M. Piles, N. J. Rodríguez-Fernández, E. Zsoter, C. Buontempo, and J.-N. Thépaut, "ERA5-Land: a state-of-the-art global reanalysis dataset for land applications," *Earth Syst. Sci. Data*, vol. 13, pp. 4349–4383, Sept. 2021.

[74] C. G. da União, "Serviço eletrônico de informação ao cidadão," 2022.

[75] DATASUS, "File transfer protocol - datasus," 2022.

[76] R. d. F. Saldanha, R. R. Bastos, and C. Barcellos, "Microdatasus: pacote para download e pré-processamento de microdados do departamento de informática do sus (datasus)," *Cadernos de Saúde Pública*, vol. 35, no. 9, 2019.

[77] S. de Informação de Notificação de Agravos SINAN, "Ficha de investigaÇÃo dengue e febre de chikungunya," 2022.

[78] Ministério da Saúde (MS) , "Capítulo 7 - arboviroses urbanas causadas por vÍrus transmitidos pelo aedes: Dengue, chikungunya e zika," in *Guia de Vigilância em Saúde - 5ª edição*, pp. 685–768, 2020.

[79] S. Runge-Ranzinger, O. Horstick, M. Marx, and A. Kroeger, "What does dengue disease surveillance contribute to predicting and detecting outbreaks and describing trends?," *Tropical Medicine & International Health*, vol. 13, no. 8, pp. 1022–1041, 2008.

[80] T. W. Parr, A. R. Sier, R. Battarbee, A. Mackay, and J. Burgess, "Detecting environmental change: science and society—perspectives on long-term research and monitoring in the 21st century," *Science of the total environment*, vol. 310, no. 1-3, pp. 1–8, 2003.

[81] J. Muñoz-Sabater, E. Dutra, A. Agustí-Panareda, C. Albergel, G. Arduini, G. Balsamo, S. Boussetta, M. Choulga, S. Harrigan, H. Hersbach, *et al.*, "Era5-land: A state-of-the-art global reanalysis dataset for land applications," *Earth System Science Data*, vol. 13, no. 9, pp. 4349–4383, 2021.

[82] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, *et al.*, "The era5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.

[83] S. Kobayashi, Y. Ota, Y. Harada, A. Ebita, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, *et al.*, "The jra-55 reanalysis: general specifications and basic characteristics," *Journal of the Meteorological Society of Japan. Ser. II*, vol. 93, no. 1, pp. 5–48, 2015.

[84] R. Gelaro, W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. A. Randles, A. Darmenov, M. G. Bosilovich, R. Reichle, *et al.*, "The modern-era retrospective analysis for research and applications, version 2 (merra-2)," *Journal of climate*, vol. 30, no. 14, pp. 5419–5454, 2017.

[85] "Fast Extraction from Raster Datasets using Polygons."

[86] I. Censo, "Disponível em:< http://www. censo2010. ibge. gov. br/>," *Acesso em*, vol. 23, 2010.

[87] R. Pereira, C. Gonçalves, P. De Araujo, G. Carvalho, R. De Arruda, I. Nascimento, B. Da Costa, W. Cavedo, P. Andrade, A. Da Silva, *et al.*, "geobr: loads shapefiles of official spatial data sets of brazil," *GitHub repository*, 2019.

[88] D. J. Gubler, "Dengue, urbanization and globalization: the unholy trinity of the 21st century," *Tropical medicine and health*, vol. 39, no. 4SUPPLEMENT, pp. S3–S11, 2011.

[89] A. Seah, J. Aik, L.-C. Ng, and C. C. Tam, "The effects of maximum ambient temperature and heatwaves on dengue infections in the tropical city-state of Singapore – A time series analysis," *Science of The Total Environment*, vol. 775, p. 145117, June 2021.

[90] S. J. Ryan, C. J. Carlson, E. A. Mordecai, and L. R. Johnson, "Global expansion and redistribution of aedes-borne virus transmission risk with climate change," *PLoS neglected tropical diseases*, vol. 13, no. 3, p. e0007213, 2019.

[91] C. Imai, B. Armstrong, Z. Chalabi, P. Mangtani, and M. Hashizume, "Time series regression model for infectious disease and weather," *Environmental Research*, vol. 142, pp. 319–327, Oct. 2015.

[92] O. J. Brady, P. W. Gething, S. Bhatt, J. P. Messina, J. S. Brownstein, A. G. Hoen, C. L. Moyes, A. W. Farlow, T. W. Scott, and S. I. Hay, "Refining the Global Spatial Limits of Dengue Virus Transmission by Evidence-Based Consensus," *PLoS Negl Trop Dis*, vol. 6, p. e1760, Aug. 2012.

[93] "cidades ibge - rio de janeiro."

[94] G. L. Werneck, A. E. Macias, C. Mascarenas, L. Coudeville, D. Morley, V. Recamier, M. Guergova-Kuras, E. Puentes-Rosas, N. Baurin, and M.-L. Toh, "Comorbidities increase in-hospital mortality in dengue patients in Brazil," *Mem. Inst. Oswaldo Cruz*, vol. 113, July 2018.

[95] S. Cenci, G. Sugihara, and S. Saavedra, "Regularized s-map for inference and forecasting with noisy ecological time series," *Methods in Ecology and Evolution*, vol. 10, no. 5, pp. 650–660, 2019.

[96] C. M. Benedum, O. M. Seidahmed, E. A. Eltahir, and N. Markuzon, "Statistical modeling of the effect of rainfall flushing on dengue transmission in singapore," *PLoS neglected tropical diseases*, vol. 12, no. 12, p. e0006935, 2018.

[97] S. I. Hay, M. F. Myers, D. S. Burke, D. W. Vaughn, T. Endy, N. Ananda, G. D. Shanks, R. W. Snow, and D. J. Rogers, "Etiology of interepidemic periods of mosquito-borne disease," *Proceedings of the National Academy of Sciences*, vol. 97, no. 16, pp. 9335–9339, 2000.

[98] I. P. Godói, L. V. D. Da Silva, A. R. Sarker, I. Megiddo, A. Morton, B. Godman, S. Alvarez-Madrazo, M. Bennie, and A. A. Guerra-Junior, "Economic and epidemiological impact of dengue illness over 16 years from a public health system perspective in Brazil to inform future health policies including the adoption of a dengue vaccine," *Expert Review of Vaccines*, vol. 17, pp. 1123–1133, Dec. 2018.

[99] IBGE, "Brasil | Cities and States |."

[100] C. A. Alvares, J. L. Stape, P. C. Sentelhas, J. L. de Moraes Gonçalves, and G. Sparovek, "Köppen's climate classification map for Brazil," *metz*, vol. 22, pp. 711–728, Dec. 2013.

[101] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[102] D. Jackson, R. Riley, and I. R. White, "Multivariate meta-analysis: Potential and promise," *Statist. Med.*, vol. 30, pp. 2481–2498, Sept. 2011.

[103] B. G. Armstrong, A. Gasparrini, and A. Tobias, "Conditional Poisson models: a flexible alternative to conditional logistic case cross-over analysis," *BMC Medical Research Methodology*, vol. 14, p. 122, Nov. 2014.

[104] A. Gasparrini, "The case time series design," *Epidemiology (Cambridge, Mass.)*, vol. 32, no. 6, p. 829, 2021.

[105] L. d. S. V. Jacobson, B. F. A. d. Oliveira, R. Schneider, A. Gasparrini, and S. d. S. Hacon, "Mortality Risk from Respiratory Diseases Due to Non-Optimal Temperature among Brazilian Elderlies," *IJERPH*, vol. 18, p. 5550, May 2021.

[106] N. Pudpong and S. Hajat, "High temperature effects on out-patient visits and hospital admissions in Chiang Mai, Thailand," *Science of The Total Environment*, vol. 409, pp. 5260–5267, Nov. 2011.

[107] B. Wang, G. Chai, Y. Sha, Q. Zha, Y. Su, and Y. Gao, "Impact of ambient temperature on cardiovascular disease hospital admissions in farmers in China's Western suburbs," *Science of The Total Environment*, vol. 761, p. 143254, Mar. 2021.

[108] A. Vaidyanathan, S. Saha, A. M. Vicedo-Cabrera, A. Gasparrini, N. Abdurehman, R. Jordan, M. Hawkins, J. Hess, and A. Elixhauser, "Assessment of extreme heat and hospitalizations to inform early warning systems," *PNAS*, vol. 116, pp. 5420–5427, Mar. 2019. Publisher: National Academy of Sciences Section: PNAS Plus.

[109] K. V. M. d. S. Noronha, G. R. Guedes, C. M. Turra, M. V. Andrade, L. Botega, D. Nogueira, J. A. Calazans, L. Carvalho, L. Servo, and M. F. Ferreira, "Pandemia por COVID-19 no Brasil: análise da demanda e da oferta de leitos hospitalares e equipamentos de ventilação assistida segundo diferentes cenários," *Cad. Saúde Pública*, vol. 36, no. 6, p. e00115320, 2020.

[110] J. Schwartz, J. M. Samet, and J. A. Patz, "Hospital admissions for heart disease: the effects of temperature and humidity," *Epidemiology*, vol. 15, pp. 755–761, Nov. 2004.

[111] M. V. Saha, R. E. Davis, and D. M. Hondula, "Mortality Displacement as a Function of Heat Event Strength in 7 US Cities," *American Journal of Epidemiology*, vol. 179, pp. 467–474, Feb. 2014.

[112] C. Morral-Puigmal, Martínez-Solanas, C. M. Villanueva, and X. Basagaña, "Weather and gastrointestinal disease in Spain: A retrospective time series regression study," *Environment International*, vol. 121, pp. 649–657, Dec. 2018.

[113] X. Yi, Z. Chang, X. Zhao, Y. Ma, F. Liu, and X. Xiao, "The temporal characteristics of the lag-response relationship and related key time points between ambient temperature and hand, foot and mouth disease: A multicity study from mainland China," *Science of The Total Environment*, vol. 749, p. 141679, Dec. 2020.

[114] X. Basagaña and J. Barrera-Gómez, "Reflection on modern methods: visualizing the effects of collinearity in distributed lag models," *International Journal of Epidemiology*, p. dyab179, Aug. 2021.

[115] B. X, "Visualizing the effects of collinearity in the context of distributed lag models," *Environmental Epidemiology*, vol. 3, p. 23, Oct. 2019.

[116] E. de Schrijver, M. Bundo, M. S. Ragettli, F. Sera, A. Gasparrini, O. H. Franco, and A. M. Vicedo-Cabrera, "Nationwide analysis of the heat-and cold-related mortality trends in switzerland between 1969 and 2017: the role of population aging," *Environmental health perspectives*, vol. 130, no. 3, p. 037001, 2022.

[117] P. Regoto, C. Dereczynski, S. C. Chou, and A. C. Bazzanela, "Observed changes in air temperature and precipitation extremes over brazil," *International Journal of Climatology*, vol. 41, no. 11, pp. 5125–5142, 2021.

[118] E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of atmospheric sciences*, vol. 20, no. 2, pp. 130–141, 1963.

[119] H. Ye, A. Clark, E. Deyle, S. Munch, J. Cai, J. Cowles, Y. Daon, A. Edwards, O. Keyes, J. Stagge, M. Ushio, E. White, and G. Sugihara, "rEDM: Applications of Empirical Dynamic Modeling from Time Series," Dec. 2018.

[120] S. L. Zeger, F. Dominici, and J. Samet, "Harvesting-resistant estimates of air pollution effects on mortality," *Epidemiology*, pp. 171–175, 1999.

[121] J. Schwartz, "Harvesting and long term exposure effects in the relation between air pollution and mortality," *American journal of epidemiology*, vol. 151, no. 5, pp. 440–448, 2000.

[122] Q. Zhao, Y. Guo, T. Ye, A. Gasparrini, S. Tong, A. Overcenco, A. Urban, A. Schneider, A. Entezari, A. M. Vicedo-Cabrera, *et al.*, "Global, regional, and national burden of mortality associated with non-optimal ambient temperatures from 2000 to 2019: a three-stage modelling study," *The Lancet Planetary Health*, vol. 5, no. 7, pp. e415–e425, 2021.

[123] W. Lee, Y. Kim, F. Sera, A. Gasparrini, R. Park, H. M. Choi, K. Prifti, M. L. Bell, R. Abrutzky, Y. Guo, *et al.*, "Projections of excess mortality related to diurnal temperature range under climate change scenarios: a multi-country modelling study," *The Lancet Planetary Health*, vol. 4, no. 11, pp. e512–e521, 2020.

[124] È. Martínez-Solanas, M. Quijal-Zamorano, H. Achebak, D. Petrova, J.-M. Robine, F. R. Herrmann, X. Rodó, and J. Ballester, "Projections of temperature-attributable mortality in europe: a time series analysis of 147 contiguous regions in 16 countries," *The Lancet Planetary Health*, vol. 5, no. 7, pp. e446–e454, 2021.

[125] P. McCullagh and J. A. Nelder, "Generalized linear models. chapman and hall," *London, UK*, 1989.

[126] A. Gasparrini, F. Scheipl, B. Armstrong, and M. G. Kenward, "A penalized framework for distributed lag non-linear models," *Biometrics*, vol. 73, no. 3, pp. 938–948, 2017.