



École Polytechnique Fédérale de Lausanne

Adaptive Mitigation:
Identification of the Dynamic Drivers of Effective Policy during the
COVID-19 Pandemic

by Thierry Bossy

Master Thesis

Prof. Martin Jaggi
Thesis Advisor

Dr. Mary Anne Hartley
Thesis Supervisor

École Polytechnique Fédérale de Lausanne
(EPFL)
Machine Learning and Optimization Lab
(MLO)
August 17, 2020

Acknowledgments

Two weeks before starting my thesis, I was picking tomatoes in the northern part of Quito in Ecuador. That being said, I would like to thank COVID-19 that sent me back to Switzerland and gave me a thesis subject before I have time to wonder.

I would also and foremost like to thank Annie who gave me the opportunity to start this project in her highly contagious playground. Thank to Martin as well who gave valuable Professor's insight and follow-up throughout the project.

Thank to my friend Lucas who brought the esteemed help of a freshly graduated data scientist and Prakhar who added his experience to the cake.

I will not conclude without thanking Kimia and Farheen for their contribution to the project from Iran and Pakistan despite the fact that the pandemic canceled their internship travels to Switzerland.

Lausanne, August 17, 2020

Thierry Bossy

Abstract

Introduction: The unprecedented speed and scale of the *COVID-19* pandemic necessitated the rapid implementation of untested public health measures to mitigate the consequences of viral spread. In the 8 months that have passed since the first recognized case, an enormous amount of research has been published evaluating the efficacy of the various policies implemented by different countries, however the majority of these studies focus on a specific region or time, over-representing high-income countries during periods of extreme transmission or “peak events”. The aim of this study is to provide a more general analysis that considers a global scope of the pandemic and the dynamic drivers that build effective policies to mitigate human interactions and slow the spread of disease.

Methods: We collected a range of information regarding epidemic trends, weather, demographics, government response and mobility reports across the globe. We then built an hybrid Neural Network that combined an LSTM layer with a multilayer perceptron to infer the reproduction number from various non-epidemiological factors. The model was designed to predict the reproduction number (R -value) on 93 countries with available data and compare it to our ground truth estimate obtained from officially reported epidemiological data. Finally, we used an alternative model to assess the impact of public health measures on the epidemic.

Findings: From the available features, we obtained the best performances using demographics combined with mobility features. The sanitary indices (beds/thousand, diabetes prevalence, ...) did not help the prediction and, more interestingly, the pressure indicator of historical weather forecast improved the prediction of the reproduction number by about 4.5%. This optimized model predicted the reproduction number with a mean absolute error of 0.254 across the 93 countries over the time of the epidemic. For many countries (Switzerland, United Kingdom, South Africa, ...) this error passed below 0.17. An alternative version of the model allowed us to estimate the impact of policies in terms of average reduction in reproduction number, and more importantly, allowed us to compare these trends between countries. For instance, we observe that the model showed that no policy had a positive impact in India as opposed to Switzerland, where most of them are associated to improved epidemic control.

Conclusion: Understanding these complex interactions may allow individuals and policy makers to better adapt mitigation strategies to optimize the efficacy of the implemented policies.

Contents

Acknowledgments	2
Abstract	3
1 Introduction	7
1.1 Context	7
1.2 Challenges	8
1.3 Motivation	8
1.4 Objectives	9
2 Background	10
2.1 Lockdowns and public health measures	10
2.1.1 The Oxford COVID-19 Government Response Tracker (OxCGRT)	11
2.1.2 Stringency index	13
2.1.3 Effectiveness of policies	13
2.2 Basic and effective reproduction number	16
2.2.1 Interpretations and advantages	17
2.2.2 Estimating reproduction numbers of COVID-19	18
2.2.3 Estimating reproduction number using daily new cases	18
2.2.4 Estimating reproduction number using daily new deaths	21
2.3 Mobility	22
2.4 Long short-term memory neural network	22
2.5 Our work	23
3 Methods	24
3.1 Causality graph	24
3.2 Data sources	25
3.2.1 “Official” COVID-19 reports	25
3.2.2 Government measures	27
3.2.3 Weather	27
3.3 Delays between events	28
3.3.1 FOPH data delays	28
3.4 Effective reproduction number estimation	28

3.4.1	Adjustments	28
3.4.2	Consequences of under reporting	29
3.5	Modeling	29
3.5.1	Features	29
3.5.2	Cross correlations	32
3.5.3	Leave-one-out training	32
3.5.4	Predictions and error	33
3.5.5	Linear model for mobility prediction	33
3.5.6	LSTM Neural Network	33
3.5.7	Country selection	35
3.5.8	Baseline prediction	35
3.5.9	Feature selection	35
3.5.10	Time window selection for the LSTM	36
3.5.11	Ultimate reproduction number prediction	36
3.5.12	Policy effect	36
4	Results and interpretation	38
4.1	Delays between events	38
4.1.1	Consequences of under reporting	39
4.2	Summarizing plots	40
4.3	Cross correlations	42
4.4	Mobility predictions	44
4.4.1	Feature importance of linear model	45
4.5	R prediction with an LSTM model	48
4.5.1	Baseline using a unique simple feature	48
4.5.2	Feature selection	51
4.5.3	LSTM time window selection	52
4.5.4	Ultimate reproduction number prediction	52
4.5.5	Model performance in the face of missing data	59
4.5.6	Reproduction number prediction from policies	60
4.5.7	Policy effect	62
5	Discussion	71
5.1	Limitations	72
6	Future work	74
7	Conclusion	76
References		77

A Modeling	80
A.1 Considered countries in cross correlations	80
A.2 Considered countries in feature selection	81
A.3 Errors of mobility predictions with linear model	83
A.4 LSTM feature selection process	85
A.5 LSTM time window selection	91
A.6 Ultimate reproduction number prediction	92
A.7 R prediction from policy features	95

Chapter 1

Introduction

1.1 Context

Beginning in Wuhan — the capital of Hubei province in China — the COVID-19 outbreak took less than 4 months to acquire the status of *pandemic* by the World Health Organization (WHO) on March 11th 2020 [24] [25]. COVID-19 (**CoronaVirus Disease 2019**) is caused by a novel coronavirus (SARS coronavirus-2, or SARS-CoV-2), to which the world's population is universally susceptible. Combining this with its high transmission rate and significant mortality, has made it the deadliest pandemic to emerge in the last 100 years. After reaching the European continent, it quickly spread around the globe and affected nearly all countries in the world. To mitigate transmission, most countries responded by implementing emergency public health laws to establish physical distancing in their population. By the beginning of April 2020, half of humanity had been recommended or required to stay at home [29]. In addition to staying at home, further policies to limit human interaction (such as travel restrictions, limitations in public gatherings or school closures) were variably instituted across regions and the efficacy of their timing, duration, stringency or target has been the subject of much debate.

Due to the unprecedented scope and urgency of their application, the effectiveness of these policies in establishing and maintaining viral containment across various social and economic settings was unknown. Governments and individuals adjusted their policies and behaviors to optimize a trade-off between lives and livelihoods according to ethical calculations that were not universally shared.

To facilitate research, some private actors released information that could be used as proxies of the population's habits. Among them, Google and Apple published mobility reports showing anonymized occupancy by various means of transportation or in various locations, collected from Android and iOS smartphones around the world [15] [14].

1.2 Challenges

As the first pandemic in the age of big data, there is enormous potential to generate precious knowledge about epidemics in general and understand how to adapt responses to their settings. However, the collected data is not standardized, and variations can be caused by the conditions of their setting. For instance, the economic status or political system of a country could impact the quality or transparency of their case reporting, as well as their ability or willingness to implement/follow a certain policy. In addition, social and cultural contexts determine the perceived disruption of a policy and may affect public adherence. Thus, it is essential to consider policy efficacy as a heterogeneous phenomenon that is dependent on evolving interdisciplinary factors.

The urgency of COVID-19 has pushed the scientific community to publish work on the subject at an unprecedented pace. Between January and April 2020, the number of publications collected from just 4 sources — namely *arxiv*, *pubmed*, *biorxiv* and *medrxiv* — related to COVID-19 doubled every 14 days, following an exponential growth curve to reach about 3000 in early April [21] [7]. This is a notable shift away from the quality standards that are usually ensured at a slower pace of careful peer-review. A factor which makes it even more challenging to synthesize findings or draw meaningful comparisons.

1.3 Motivation

The motivation behind this study is to produce an adaptive global analysis of the COVID-19 mitigation strategies that may evolve with incoming data and continue to be relevant in the longer term. This is in contrast to studies with limited geographic/temporal scopes that may fail to be relevant outside the exceptional peak events they describe. In a context where global warming is suspected to make epidemics increasingly more likely [18], it is essential to understand how to adapt mitigation strategies to suit the geographic, economic and cultural setting as well as when and how to adjust these policies according to the reported epidemiological disease dynamics. For COVID-19, the lack of understanding on these key issues has led countries to take sometimes opposing strategies despite seeking similar objectives. For public health policy, it is not only its content and target but also its timing. While inexperience sometimes created unintended delays in implementation, some policies were purposefully withheld or limited in temporal scope according to varying motivations or metrics. Policy timing is particularly consequent during epidemics, where the number of infected people — and thus the number of deaths — grows exponentially.

The COVID-19 pandemic has also exacerbated the already significant inequities in scientific attention. For instance, in northern countries, the extensive research has permitted more informed responses in contrast to low income countries that might not have had been able to adjust their strategies according to reliable data. For this reason, a global study using a

standardized framework for comparative analyses may allow us to identify, quantify and compare the dynamic drivers of effective policies in various contexts and allow individuals and policy makers to better adapt their behaviors and decisions.

1.4 Objectives

To conclude the introduction, we show a list of objectives that were set at the beginning of this study:

1. Quantify and visualize the epidemic trends from official/governmental sources integrated with the public health policies and mobility data using normalized metrics across countries.
2. Describe and classify the policies in characteristics and application scope as well as quantify human rights infringement using expert labeling.
3. Build a model to identify effective policies based on changes in epidemic trends and mobility data.
4. Quantify and visualize social media sentiment and topic trends in a COVID-19 twitter dataset to explore associations between effective policy using normalized metrics.
5. Estimate the under-reporting to normalize official reporting for better comparison based on proxies for infection that are independent of testing capacity (death and self-report).

Chapter 2

Background

This chapter provides the background information and the basic knowledge that the reader may require in order to understand the concepts on which our study will be based.

2.1 Lockdowns and public health measures

The world's first COVID lockdown was instated in Wuhan on January 29th 2020, one month after declaring its first case of SARS-CoV2. All cities of the Hubei province including Wuhan were placed under mandatory quarantine by a *stay-at-home order* imposed by the Chinese national government. By restricting all movements in public spaces, stay-at-home orders, by definition, imply the imposition of several other restrictions of activities occurring outside of the home. When applied in the strictest manner, as was the case for this first lockdown, it also implies a freezing of most of "non-essential" activities which are usually protected by human rights laws, such as schooling, religious gatherings and "luxury" commerce. In European countries, the optimal stringency of lockdowns (or even the necessity of imposing one) was the subject of much debate. While France, Italy and Spain imposed dramatic quarantines in response to explosive case numbers, Sweden, the United Kingdom and the Netherlands attempted to adopt a strategy involving herd immunity. The supporters of this approach argued that without any cure or vaccine, mitigation strategies were simply postponing consequences and diversifying harm via indirect economic destruction. The United Kingdom and Netherlands finally changed their strategy in response to public pressure as hospitalizations neared saturation of the health system. They then adopted an approach to *flatten the curve* which aimed to protect the National Health Services.

Between the herd immunity strategy chosen by Sweden and the strict lockdown strategy adopted by Spain, each country adopted a strategy that lay somewhere in between these extremes in terms of stringency. Some countries recommended people to stay home without

legal requirement as it was the case in Switzerland, while others allowed a variable number of exceptions to lockdown requirement. Alongside lockdowns, many other policies have also been adopted in different countries such as school closure, public transport withdrawal or restrictions on the maximum gathering size.

2.1.1 The Oxford COVID-19 Government Response Tracker (OxCGRT)

To centralize information related to government responses, the *Oxford COVID-19 Government Response Tracker* (OxCGRT) has been published by *University of Oxford* [26]. They continuously collect publicly available information associated with measures taken against coronavirus in each country. To obtain comparable data, they classified different measures into 17 indicators permitting their quantification. Each of them is evolving over time and takes different values according to the strictness of its application. The OxCGRT has been used in our study as the reference data source for government response.

Indicators and levels

Indicators are divided into 3 categories, displayed in this section (as of July 12th 2020). Indicators of the first category only are detailed as the other ones are not used in our study [28] [27]:

Note: Level 0 is associated to no restrictions or measures.

- Containment and closure policies
 - **School closure**
 - * Level 1: Recommend closing
 - * Level 2: Require closing (only some levels or categories, e.g. just high school, or just public schools)
 - * Level 3: Require closing at all levels
 - **Workplace closing**
 - * Level 1: Recommend closing (or recommend work from home)
 - * Level 2: Require closing (or work from home) for some sectors or categories of workers
 - * Level 3: Require closing (or work from home) for all-but-essential workplaces (e.g. grocery stores, doctors)
 - **Cancel public events**
 - * Level 1: Recommend canceling
 - * Level 2: Require canceling
 - **Restrictions on gatherings**

- * Level 1: Restrictions on very large gatherings (the limit is above 1000 people)
 - * Level 2: Restrictions on gatherings between 101 and 1000 people
 - * Level 3: Restrictions on gatherings between 11 and 100 people
 - * Level 4: Restrictions on gatherings of 10 people or less
- **Close public transports**
 - * Level 1: Recommend closing (or significantly reduce volume/route/means of transport available)
 - * Level 2: Require closing (or prohibit most citizens from using it)
 - **Stay-at-home requirements**
 - * Level 1: Recommend not leaving house
 - * Level 2: Require not leaving house with exceptions for daily exercise, grocery shopping and 'essential' trips
 - * Level 3: Require not leaving house with minimal exceptions (e.g. allowed to leave once a week, or only one person can leave at a time, etc)
 - **Restrictions on internal movement**
 - * Level 1: Recommend not to travel between regions/cities
 - * Level 2: Internal movement restrictions in place
 - **International travel controls**
 - * Level 1: Screening arrivals
 - * Level 2: Quarantine arrivals from some or all regions
 - * Level 3: Ban arrivals from some regions
 - * Level 4: Ban on all regions or total border closure
- Economic policies
 - **Income support**
 - **Debt/contract relief**
 - **International support**
 - Health system policies
 - **Public information campaigns**
 - **Testing policy**
 - **Contact tracing**
 - **Emergency investment in healthcare**
 - **Investment in vaccines**
 - Miscellaneous policies (record policy announcements that do not fit anywhere else)

2.1.2 Stringency index

In addition to the detailed indicators described above, the *OxCGRT* provides overall indices ranging from 0 to 100 that comprise a linear combination of the measures. Among them, the *stringency index* is a combination of all *containment and closure policies* plus the public information campaign indicator. It can be viewed as a proxy of policy impact on the population. Fig. 2.1 shows how the stringency level of global policies changed between March 10th and July 10th of 2020.

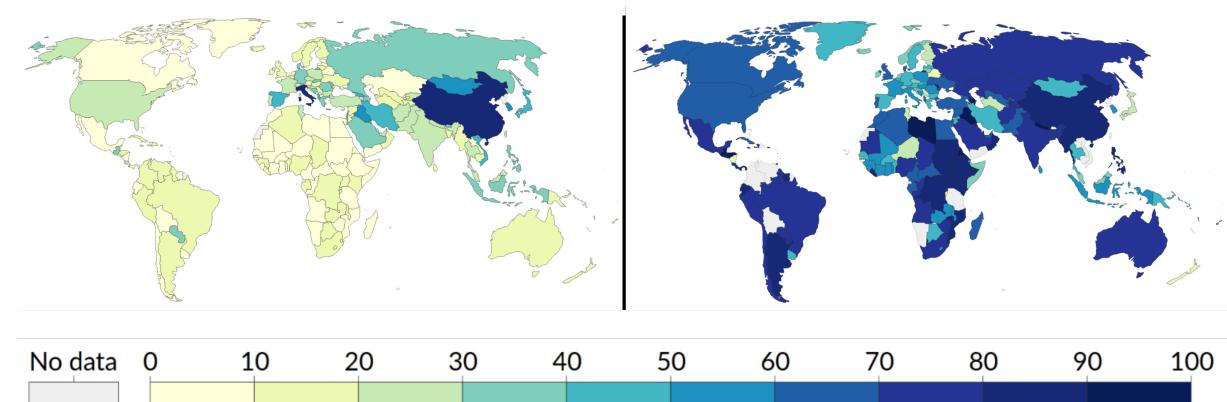


Figure 2.1: Comparison of the stringency level in the world between March 10th (left) and July 10th (right) (Source: [26])

2.1.3 Effectiveness of policies

Due to the unprecedented and rapid evolution and scale of the crisis, governments were forced to blindly implement policies based on assumptions of their effect. Several studies have retrospectively assessed their impact and in this section, we will report the methods and the results of these key studies on policy effectiveness.

1. An early study published on April 17th 2020 observed the effects of non-pharmaceutical interventions on both the COVID-19 and the influenza virus in the region of Hong Kong [8]. By estimating the daily effective reproduction numbers R_t (see Section 2.2), they concluded that all policies together were responsible for a reduction in the transmission of both outbreaks. However, no distinction could be made between the measures that took place concomitantly or within a narrow time frame.

This study also reported that the proportion of people wearing masks in public increased from 74.5% to 97.5% between the first survey on January 20-23 and the second on February 11-14. Face masks are indeed a critical item of personal protective equipment in airborne disease. However, this is rarely reported and there is currently no simple/validated way

to gauge mask usage (albeit that efforts to implement intelligent mask detection is underway [31]), and thus the contribution of mask usage to the overall policy effectiveness in this Hong Kong study (and other studies) is unknown.

2. A cross-country analysis undertaken at ETH Zurich compared the effects of the policies using reported cases from 20 countries (EU-15 countries, Switzerland, Norway, United States Australia and Canada) [4]. Using a Bayesian hierarchical model, they estimated the reduction in number of cases attributed to each non-pharmaceutical policy. Their results are reported in the following table:

Policy	Reduction	95% CI
Venue closure	36%	20% – 48%
Gathering ban	34%	21% – 45%
Border closure	31%	19% – 42%
Work ban	31%	16% – 44%
Event ban	23%	8% – 35%
School closure	8%	0% – 23%
Lockdown	5%	0% – 14%

This statistical model follows the underlying assumption that any policy has a positive impact, i.e. a non-negative reduction of the number of cases. The results that they provided are global (country-independent), although they also made an estimation of the daily number of cases that have been avoided due to the application of each of the 7 policies for the United States and the United Kingdom. The reductions in the United States are shown in Fig. 2.2:

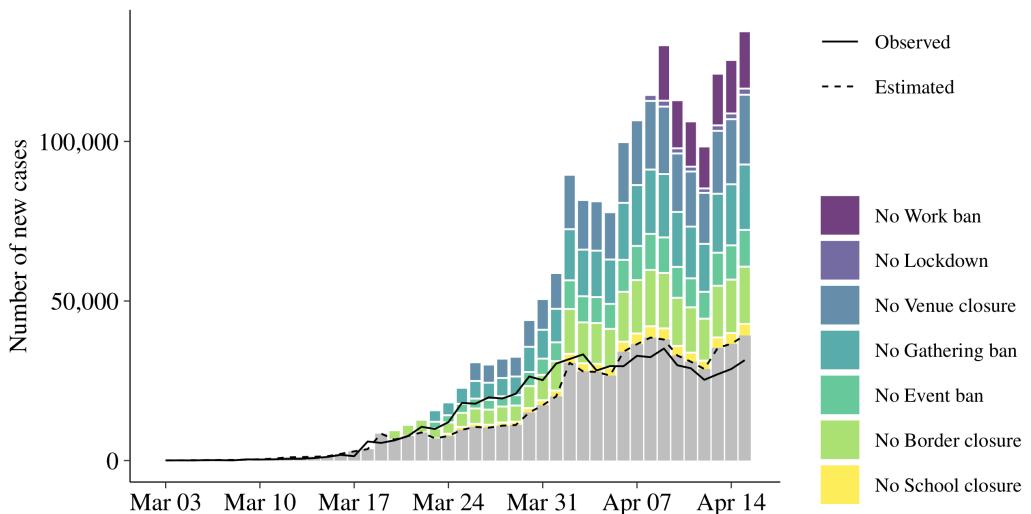


Figure 2.2: Estimated reductions on COVID-19 cases attributed to each policy (United States) [4]

The results of this study are clear and easy to interpret. However, the variables representing each policy are not inter-dependent, and thus each policy is assessed independently. One

could argue that depending on which other policies are in place, the impact of a certain policy is variable. Furthermore, the assumption that each policy has an effect that is decreasing over time may be problematic in the long-term, where new peaks could appear in a more stable context.

3. In a comprehensive study coming from a collaboration between *Central Bank of Chile* and *Humboldt University of Berlin* [2], researchers applied a regression on their estimation of the effective reproduction number (R_t) (see Section 2.2). Using data coming from 14 European countries, they made a regression on the logarithm of R_t using a set of binary indicators representing 5 policies (lockdown, public event cancellation, school closure, self isolation and social distancing). To overcome the fact that “obtaining credible counterfactuals in [that] empirical context is extremely challenging” [2, p. 14], they added other inputs that might improve the quality of the results. At first, they performed a new regression after adding a *days-since-outbreak* variable corresponding to the number of days elapsed since a country reached a total number of 100 COVID-19 cases. They then made two other regressions using — in addition to the policy indicators and the *days-since-outbreak* variable — some mobility and testing controls. It comprises mobility data (see Subsection 2.3) and testing data (number of COVID-19 tests performed) as control variables, assuming that they may confound the reproduction number.

However, the results that they obtained by sequentially adding the three control variables are divergent. It is difficult to tell which of the regressions best fit the setting and it is also difficult to differentiate the impact of each variable in the statistical model. As we will describe later (see Section 3.1), some variables can reduce the capacity of a model to provide specific information. The lack of qualitative metrics of each of the four methods makes the results difficult to interpret. For example, the model using no control variable suggests that the lockdown is the policy that has by far the biggest impact, but when the three control variables are included, the model suggests that the effect of lockdown was negligible.

4. The last key study that we will discuss was undertaken by researchers at *University of Cincinnati* [12]. They assessed the effectiveness of policies by using data relative to 12 countries collected until April 10th and for which sufficient information about testing was available: Austria, Belgium, Estonia, Finland, France, Iceland, Italy, Japan, Malaysia, South Africa, South Korea and the United Kingdom. Testing data were used to obtain the COVID-19 test-positivity rate by country. They then estimated the growth rate of this value as well as the one derived from the absolute number of daily cases and considered physical distancing as a unique event supposed to reduce growth rates. They obtained that, in 5 countries (Austria, Belgium, Italy, Malaysia, and South Korea) physical distancing conferred an average reduction of 52.37% (std. dev. 13.37%) in the growth rate of confirmed cases, as well as a decrease of 23.22 (CI95% 17.87 – 28.57) cases per 100 tests. In contrast, they found that the growth rate of the other countries (Estonia, Finland, France, Iceland, Japan, South Africa and the United Kingdom) did not have significant changes after implementation of non pharmaceutical interventions.

As stated by the authors, this model is valid for short-term predictions only. An important simplification was made by considering all measures as being a unique event occurring at a certain point in time. Again this does not allow a detailed evaluation considering the variety in non-pharmaceutical interventions or in the stringency with which it may be applied in different countries.

In conclusion, the different studies presented in this section address the same question with different approaches. However, none of them use more than 20 countries as a reference point, which makes the analysis more subject to bias. Importantly, these studies also have limited capacity to infer causality, either diluting the effect of individual policies by pooling them together or assuming all policies have a positive impact. While some assumptions are plausible, coherence, consistency and analogy between settings (which would better allow us to assess a causal link between policy implementation and effect) is limited or not considered. Additionally, the limited time-frame on which the studies were undertaken have not yet assessed the temporality or reversibility of the assumed effect.

2.2 Basic and effective reproduction number

In epidemiology, the basic reproduction number R_0 represents the expected number of cases that are directly generated by an infectious individual when everyone is susceptible to infection [19]. This value is an important parameter in an epidemic, because it gives an insight of its transmission rate. Although this value is unique and relates to the initial state of the epidemic where only one person has been infected, we can consider its time evolving variant $R_0(t)$ that can evolve in the course of an epidemic. $R_0(t)$ can be seen as the expected number of infections that an infectious case *would have produced* if everyone was susceptible at time t . The basic reproduction number can be decomposed into three factors [19]:

1. The number of contacts each individual has per time unit
2. The probability of transmission per contact between an infectious case and a susceptible person
3. The mean duration of infectiousness

These factors make clearer the propensity of $R_0(t)$ to change. While all these factors are determined by individual behavior, the last one is also subject to the "natural" evolution of the epidemic (e.g. virus mutations) or advances in medical interventions (e.g. medication intake to reduce viral shedding). The focus of our study refers therefore to the first two points. Note that this decomposition of $R_0(t)$ is a simplification that does not consider passive transmissions (e.g. transmission through *fomites* or inert objects/surfaces that do not involve inter-personal contacts).

Another value that can be considered in epidemics is the effective reproduction number $R(t)$, which is similar to the basic reproduction number. The difference is that it considers the actual setting of the epidemic, meaning that it is dependent on the number of recovered cases that are not subject to infections. It decreases naturally as the number of susceptible cases decreases, and it is linked to the time-varying version of the basic reproduction number $R_0(t)$ by the following formula:

$$R(t) = R_0(t) \times \lambda \quad (2.1)$$

where λ is the proportion of susceptible individuals in the population.

When dealing with epidemics over heterogeneous settings, we need to be careful about these concepts. Indeed, they represent averaged forms of irregular phenomena, meaning that the more variance we have in a population regarding number of contacts or probability of transmission, the less precise we are by defining an epidemic through reproduction numbers.

2.2.1 Interpretations and advantages

Let us revisit the definition of the effective reproduction number $R(t)$. If this number is high, each infected individual will infect a high number of susceptible people who will in turn infect newly susceptible ones at the same rate. Therefore, if $R(t) > 1$, an exponential growth of the number of positive cases occurs. Thus, an “epidemic” is defined as an infection with an $R(t)$ above 1. Indeed, if each infected person goes on to infect less than one other individual on average, the outbreak stops as each infection is “replaced” by less than one new infection. As mentioned above, when the basic reproduction number R_0 is greater than 1, the effective reproduction number $R(t)$ will naturally be reduced to less than one as the population approaches herd immunity. Herd immunity is reached when roughly $1 - (1/R_0)$ of the population is considered immune (assuming that it is impossible for an infected person to be susceptible again in the future, which is still uncertain in the case of COVID-19). However, as shown in Equation 2.1, $R(t) \leq R_0(t)$. Therefore, by reducing $R_0(t)$ to a value below 1, it is possible to stop the epidemic regardless of the proportion of susceptible λ . This is the purpose of non-pharmaceutical interventions that are intended to reduce contacts or probability of transmission.

Reproduction numbers are used because they are more stable values than other observable values, such as the number of cases, which has exponential growth. It *describes the state of the epidemic at a point in time rather than tracking individuals over time* [11]. The time-varying form of the basic reproduction number $R_0(t)$ depends neither on the past behavior of the epidemic nor on the number of susceptible individuals. Knowing the properties of an epidemic, this value is an indicator of the general behavior of a population and their capacity to increase or decrease transmission. Assuming the virus remains unchanged, an unchanged behaviour in a population (accumulating to net unchanged infectious contacts) implies a constant $R_0(t)$.

2.2.2 Estimating reproduction numbers of COVID-19

In the context of COVID-19, basic reproduction number R_0 quickly gained popular interest. However, the variable estimates that have emerged has revealed that its calculation is not trivial. In a review of 12 different studies published before the 7th of February [20] [30], Liu *et al.* reported that the estimations predicted values between 1.5 and 6.68. Although the real R_0 is subject to variations depending on the population observed, such a variance in the estimations shows how difficult it is to observe this value in practice. The heterogeneity in contacts or probability of transmission can be a cause of the variations in estimations. Another cause is that as behavior changes (either through individual initiative or by force of following public health measures), $R_0(t)$ decreases, making the retrospective observation of R_0 impossible.

A second matter of interest is the continuous estimation of the reproduction numbers. Instead of providing inherent properties about the epidemic, it gives a continuous notion of the tendency of a population to fight against it.

Knowing the parameters of the epidemic, the effective reproduction number $R(t)$ can be estimated from different observed outcomes. We may note that if the proportion of susceptible λ is close to 1, $R_0(t) \approx R(t)$ (see Equation 2.1). It is naturally the case in the beginning of an epidemic. We will describe here two approaches that have been developed to estimate $R(t)$.

Research performed by Lemaitre Joseph. C. *et al.* published at the end of May 2020 provides estimations of the reproduction numbers in cantons of Switzerland [6]. To produce them, they created an augmented SEIR model with Markov states representing normal hospitalization, intensive care units (ICU), as well as specific states for infectious persons that develop severe symptoms (usually before hospitalization). They fitted this model as a *Hidden Markov Model* (HMM) using the data coming from the *Federal Office of Public Health* (FOPH) in Switzerland. They used hospitalization and deaths data only, that they combined to an assumed infection fatality ratio. From the obtained parameters of their model, they were able to obtain an estimation of the effective reproduction number. Fitting such a detailed model was possible thanks to the Swiss data reporting each case individually with complete information about hospitalization.

In cross-country analysis, we usually do not have access to such high resolution data. Therefore, we may want to estimate $R(t)$ from more basic data that are available in a majority of countries.

2.2.3 Estimating reproduction number using daily new cases

In 2007, Christopher Fraser derived an expression of the effective reproduction number $R(t)$ as a function of the incidence rate $I(t)$ and the infectiousness $\omega(t)$ [11]. In our context, the incidence rate $I(t)$ is the derivative of the number of persons that have been infected, in its continuous form. It can be seen as the continuous-time version of the number of daily new infections. The

infectiousness $\omega(t)$ is defined as the distribution of infection events of an individual as a function of time since the infection. In other words, it shows the degree of infectiousness, or the likelihood to infect someone after being infected. It is bounded by the fact that a person cannot transmit the virus before being infected and by the fact that an infection lasts a finite amount of time:

$\omega(t) = 0 \forall t < 0$ and $\omega(t) = 0 \forall t > t'$, with t' finite. The infectiousness is normalized as $\int_0^{t'} \omega(t) dt \equiv 1$.

Using those definitions, the author defines an estimator $\hat{R}(t)$ of the effective reproduction number

$$\hat{R}(t) = \frac{I(t)}{\int_0^{t'} I(\tau) \omega(\tau) d\tau} \quad (2.2)$$

In practice, the discrete forms of $I(t)$ and $\omega(t)$, respectively I_i and ω_i can be used. This variant of the continuous estimator is shown below (Equation 2.3), where n is the number of days after which an infected person is not infectious anymore:

$$\hat{R}_i = \frac{I_i}{\sum_{j=0}^n I_{i-j} \omega_j} \quad (2.3)$$

On April 2020, another study done in the context of COVID-19 [13] added asymptotic confidence intervals to the same estimator $R_i(t)$. With q being the $(1 - \frac{\alpha}{2})$ -quantile of the standard normal distribution, the $(1 - \alpha)$ -confidence interval for $R(t)$ is:

$$\left[\hat{R}_i - q \sqrt{\frac{\hat{R}_i}{\sum_{j=0}^n I_{i-j} \omega_j}}, \hat{R}_i + q \sqrt{\frac{\hat{R}_i}{\sum_{j=0}^n I_{i-j} \omega_j}} \right] \quad (2.4)$$

The study also set an infectiousness $\omega(t)$ of COVID-19 that starts at day 2, linearly increases to reach its maximum at day 4 for 3 days, and then decreases to reach 0 at day 11.

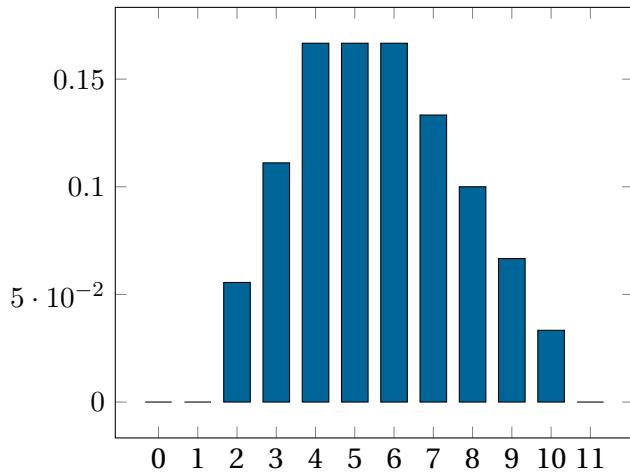


Figure 2.3: Infectiousness $\omega(t)$ of COVID-19 as a function of time (in number of days) since infection

Infection to case delay

We saw that our estimator uses I_i as an indicator of infections, as reproduction numbers are conceptually reflecting a number of infections. However, each positive case is reported a certain amount of time after the infection. This is inevitable due to incubation time (time from infection to symptom onset), testing delay and reporting delay. For our estimator \hat{R}_i , we want I_i to represent the infections at day i and not the number of reported cases that we will call C_i . If we consider a constant delay δ from infection to reported case, we have that $I_i = C_{i+\delta}$. It means that if we know δ , we should use the delayed version of the cases $C_{i+\delta}$ in our estimator to be more correct:

$$\hat{R}_i = \frac{I_i}{\sum_{j=0}^n I_{i-j} \omega_j} = \frac{C_{i+\delta}}{\sum_{j=0}^n C_{i+\delta-j} \omega_j} \quad (2.5)$$

Another and maybe simpler way to compute this estimator is to use the estimator based on daily cases C_i . By looking at Equation 2.3 we see that an estimator using C_i — let us call it \bar{R}_i — is in fact a delayed version of our estimator \hat{R}_i :

$$\hat{R}_i = \bar{R}_{i+\delta}, \text{ with } \bar{R}_i = \frac{C_i}{\sum_{j=0}^n C_{i-j} \omega_j}$$

We give more details and estimations of the delay δ in Section 3.3.

2.2.4 Estimating reproduction number using daily new deaths

Instead of using daily cases to estimate $R(t)$, we can use the exact same method on the daily deaths provoked by COVID-19. Indeed, while a high under-reporting rate can occur in the counting of cases, the counting of deaths is less subject to this obstacle. A large proportion of infected people may go undetected due to the fact that asymptomatic cases or people that are not at risk are unlikely to be tested. Unreported deaths are, on the other hand, less common. More concretely, it means that some reported deaths due to COVID-19 are related to individuals that were not reported as infected cases.

Let us look now at the differences between $\hat{R}(t)$ and the estimator based on daily deaths that we will denote $\hat{R}_d(t)$. Although it may be difficult to interpret the value estimated by $\hat{R}_d(t)$, we can argue that COVID-19 deaths are events that occur on a certain proportion of infected cases. This proportion is called the infection fatality rate and we will call it α . Let us denote by D_i the number of deaths reported on day i . If we assume the infection fatality rate to be constant, we can consider the number of deaths as being a better indicator of the real number of infected cases than the actual number of reported cases C_i . In other terms, $\alpha \times D_i$ better represents the real number of cases than C_i . If we now go back to the formula of the discrete estimator (Equation 2.3), we see that it is insensitive to a constant scaling factor in the incidence rate I_i . Specifically, if we want to estimate the reproduction number $R(t)$ using the number of cases derived from deaths ($\alpha \times D_i$), we can simply use D_i and obtain the same estimation. It implies that we do not need to know the infection fatality rate to estimate $R(t)$ from deaths.

However, this approach has two major flaws. Firstly, the infection-to-death delays have a much larger variance among cases than the infection-to-case delays (see Fig. 4.1). As we want to obtain a representative of infections at each point in time, this highly varying delay makes D_i a bad candidate to infer $R(t)$. Secondly, considering a constant infection fatality rate α may lead to incorrect conclusions especially when we are dealing with worldwide data. As a matter of fact, deaths are not evenly distributed among cases, just as the infections are not evenly distributed in the population. As a consequence, a wave of infections in a population at risk will be interpreted differently by the estimator than a wave in a population that is not at risk, due to the fact that the two situations will lead to different growth in the number of deaths.

In this section we have seen two methods to estimate $R(t)$. The first one may be more precise, but requires high resolution individual-level data that are usually private and not reported in most countries. Conversely, the 2nd approach allows an estimation by making use of the basic reports of positive cases. Both estimations are always dependent on the quality of the reports that can be more or less appropriate depending on the testing and reporting systems. We will see the implications of under-reporting in Section 3.4.

2.3 Mobility

Considering the contacts in the population as being one of the main factors influencing the curve of the epidemic (see Section 2.2), we may be interested in knowing the movements of the population. Indeed, contacts are not easy to observe from a global point of view, and the mobility of people is a good proxy to get information about it.

Concretely, Apple Inc. and Google Inc. collected movements from devices running iOS and Android respectively [14] [15]. As they represent together a majority of the market share of mobile devices, they were able to get representative information in most countries of the world. They published since April 2020 daily trends regarding mobility. Apple data contain daily trends of mobility usage in different transportation means, as Google data include trends of affluence in different places. The daily information is relative to a reference point prior to the outbreak and the gathering is made at the country, province or city level. The provided categories are the following:

Google reports	Apple reports
Retail and recreation	Driving
Groceries and pharmacies	Walking
Parks	Transit (public transports)
Transit stations	
Workplaces	
Residential	

2.4 Long short-term memory neural network

Recurrent Neural Networks (RNN) are a type of Artificial Neural Network commonly used when dealing with temporal sequences. Their structure allows them to keep memory of past values, which has many applications in grammar learning, speech recognition or handwriting recognition among others. An RNN often used in *Deep Learning* is the backbone of LSTM (*long short-term memory*). LSTM's allow long term dependencies — if they exist — to be more easily learned by the model in practice. It is facilitated by the internal design, in contrast to basic RNN's in which older inputs tend to be diluted over time.

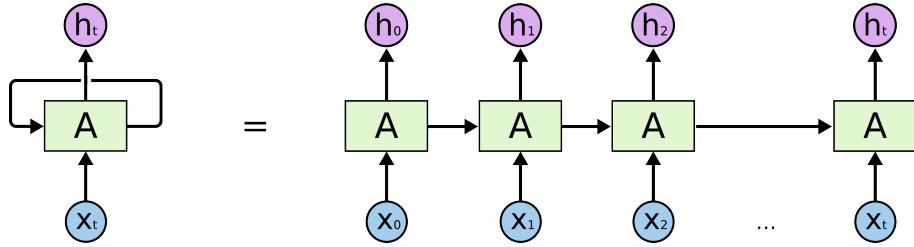


Figure 2.4: Information passing in recurrent neural network’s structure, where x_i ’s and h_i ’s represent inputs and outputs respectively, as A is the state of the neural network. (Source: [22])

2.5 Our work

In this study, we retrospectively assess the capacity of policies to predict epidemic trends across heterogeneous global settings.

In a first step we use the official statistics published by national authorities to derive the time-varying reproduction number R_0 from the daily cases. This permits us to look at the evolution of transmissions through a bounded indicator that is easier to interpret as opposed to the number of daily cases that follows an exponential evolution. We then make use of the list of indicators provided by the *University of Oxford* (see 2.1.1) to visualize the evolution of the reproduction number in relation to the timing and stringency of implemented mitigation policies. These indicators are thus standardized across heterogeneous settings, and while they may be influenced by variable collection of data or adherence to policies, the changes can still be meaningful in relation to different time points within a given setting and/or in relation to changes in other settings. We also combine those visualizations with a certain number of variables coming from mobility reports to complete our overall view of the evolution of the pandemic country by country.

In a second step, we estimate the average reporting delay using the detailed records of the Swiss *Federal Office of Public Health* (FOPH).

We then add various types of data such as demography, sanitary as well as weather-related sources to the mobility and policy values in order to build an LSTM (long short-term memory) Neural Network. This model has the objective of predicting the reproduction number in a country without knowing the number of cases from which it is initially derived. We output a prediction of the effective reproduction number using leave-one-out cross validation. Specifically, the prediction for a country is done after training on all other countries.

Finally, we use this model to estimate the difference that some specific changes in the government responses would have produced in the resulting reproduction number.

Chapter 3

Methods

In this chapter, we present the analysis that we made and the methods used. To display the results that we obtained in most of the countries of the world, we selected 6 countries that are good representative of three continents through which we will show our results. They also share the particularity of testing and reporting relatively well the positive cases of COVID-19. Those countries are Switzerland, United Kingdom, India, South Africa, United States and Brazil.

3.1 Causality graph

As an overview of the phenomena taking place in the matter of our study, we provide the causality graph below (Fig. 3.1). It shows a number of selected occurrence that we consider relevant to understand the structure of our study. The causality relations are shown by arrows linking a cause to its effect(s). We differentiate by blue frames the elements that are observable and from which we make our perception of reality, from the ones that we do not directly observe.

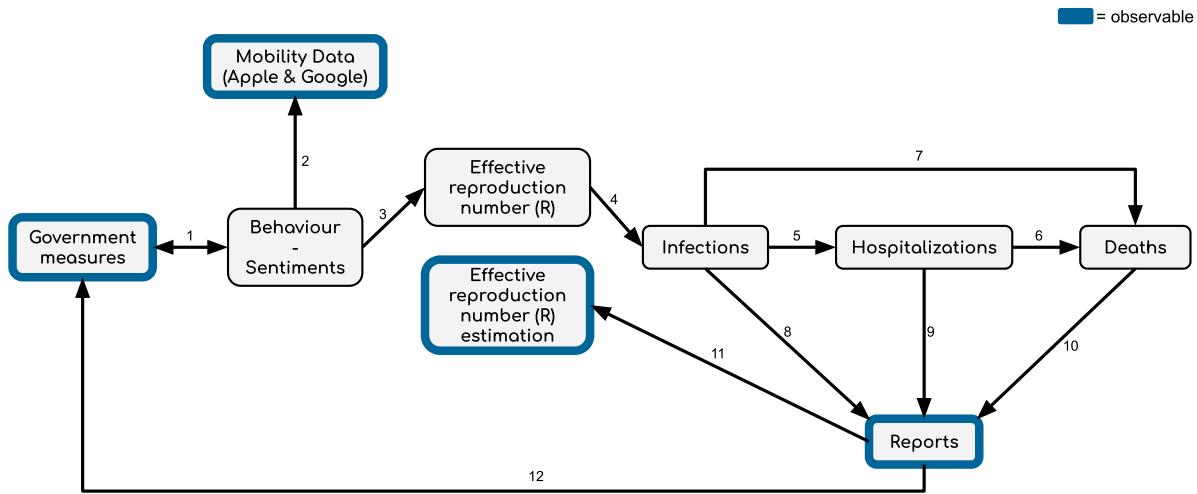


Figure 3.1: Causality graph of the pandemic-related phenomena

The elements framed in blue do not establish an exhaustive subset of observable phenomena but is rather a characterization of the data that we used in our study. We make a distinction between the real infections, hospitalization and deaths and the reported values of these events, knowing that a gap between reporting and reality exists and it varies over time and from country to country.

In the next sections we will sometimes refer to the links shown in this graph.

3.2 Data sources

The data sources corresponding to the blue items of our causality graph 3.1 are described in this section. Unless otherwise stated, the data were **collected up to July 14th**. It means that the consequent analysis do not take into account any event that may have occurred after this date.

3.2.1 “Official” COVID-19 reports

Most countries reported publicly the positive cases detected from tests made by accredited authorities (mostly hospitals). In Switzerland, the *Federal Office of Public Health* (FOPH) also provided detailed data of individual cases. We describe in the following subsections the information provided by both sources that are relevant for our purpose.

FOPH

The data provided by *Federal Office of Public Health* is private and reserved to scientist that accept to share the results of their work with them. It collects information about every positive case detected in a Swiss hospital.

The provided fields related to each positive tested patient are the following:

- **Gender** (male, female or *unknown*)
- **Canton**: the canton where the patient lives
- **Case date**: this date corresponds to the first date between these dates (date of declaration (reception by FOPH), date of test, date of sample). It corresponds almost always to the date of the test.
- **Date of symptom onset**: the date corresponding to the apparition of the first symptoms.
- **Report date**: the date on which the case was received and reported by FOPH.
- **Date of hospitalization** if the patient has been hospitalized
- **Date of death** if the patient died from COVID-19

In addition to patients details, FOPH provides the number of positive and negative tests performed each day.

We used the database shared by FOPH on May 15 containing all cases reported up to that day. Note that due to the varying delay from test to report, the tests made in the last days (2 to 3 days) before May 15th may not have all been reported.

Our World In Data

OurWorldInData is an organization specialized in collecting and reporting data related to global issues concerning living conditions, environment and wars. In the context of COVID-19, they provide a daily updated database [5] bringing together numerous indicators in every country of the world. They get the number of confirmed COVID-19 cases and deaths from *European Centre for Disease Prevention and Control* (ECDC) and the other fields are collected from various sources. Their database contains the following fields, reported each day in each country:

- | | | | |
|-----------------------|---------------------|----------------------------|--------------------------|
| • total cases | • new cases | • total cases / 1M | • new cases / 1M |
| • total deaths | • new deaths | • total deaths / 1M | • new deaths / 1M |

- **total tests**
- **new tests**
- **total tests / 1K**
- **new tests / 1K**
- **population**
- **population density**
- **continent**
- **location**
- **median age**
- **older than 65 (%)**
- **older than 70: (%)**
- **gdp per capita**: the gross domestic product per capita
- **diabetes prevalence**
- **hospital beds / 1K**
- **life expectancy** (in years)

We note that some information, such as the number of hospital beds per 1000 inhabitants is unavailable in most of the countries.

3.2.2 Government measures

As quantifiable data for government response, we used the *Oxford COVID-19 Government Response Tracker* (OxCGRT) described in Section 2.1.1. From this source, we used the indicators from *containment and closure policies* class only. The *economic policies* and *health system policies* classes were not taken into account because we assumed that their impact to the epidemic would be indirect. In other words, their impact would be the result of too many unobserved processes that would appear as noise in the context described by our causality graph (see Fig. 3.1).

3.2.3 Weather

To detect a possible correlation between the epidemic trend and the weather, we used data provided by the *World Weather Online API* [32]. The daily averaged version of their data has been chosen to match the epidemiological data. A simplification has been made in the fact that we took the capitals as representative cities for each country. Here is the list of temporal attributes that were collected:

- **the maximum temperature of the day (in °C)**
- **the minimum temperature of the day (in °C)**
- **the sensation of temperature (in °C)**
- **the humidity (in %)**
- **the pressure (in mb)**

3.3 Delays between events

In this section, we use the very precise Swiss data obtained from FOPH (section 3.2.1) to estimate the time delay between the infection and the report of COVID-19 cases. This delay is used to shift our estimation of the effective reproduction number.

3.3.1 FOPH data delays

We consider 4 events for which a date may be reported for a proportion of the entries: 1) symptom onset, 2) declaration of a case (generally date of test), 3) reporting of a case (entry in the database) and 4) death. We assume that these events usually occur in this order. The obtained time-delays are displayed in the *results* section (Fig. 4.1).

3.4 Effective reproduction number estimation

This section corresponds to the link numbered 11 of our causality graph (Fig. 3.1). To estimate the effective reproduction number in each country, we used the estimator defined in Equation 2.5, based on the daily cases shifted by a delay of $\delta = 11$. We considered the infectiousness defined in Section 2.2.3 (see Fig. 2.3).

3.4.1 Adjustments

In many countries, a weekly pattern appears in the curves of reported cases. Those are most likely not due to any epidemiological effect but rather to irregularities in testing and reporting routine. As an example, we see in the picture below a pattern in the United States where local maximums appear either on Fridays or Saturdays. To overcome this effect, we smoothed the number of cases with a moving average of 7 days before estimating the reproduction number.

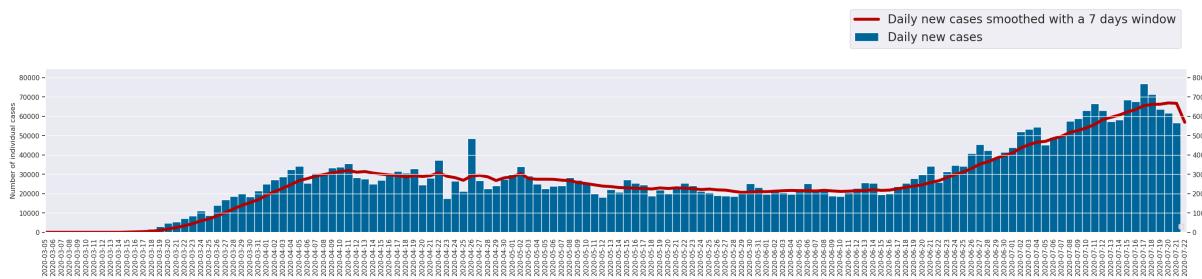


Figure 3.2: Daily reported cases of COVID-19 in United States and its smoothed curve

When the number of daily cases is low, small inaccuracies in the number of cases have a big impact on the relative growth and thus on the reproduction number's estimation. Additionally, a very low number of cases does not really reflect the reproduction number as it is possible to keep the few detected cases under control. For this reason, we consider that the estimation is not valid when the number of cases is below 5.

Another artifact due to low initial case numbers, is an excessively high R value at the beginning of the epidemic. The R value is often artificially inflated in the beginning of the outbreak due to the rapid expansion of testing capacity and detection strategies. To reduce this effect we simply discard the estimations of R that are greater than 4, which is epidemiologically implausible according to the transmission dynamics of the virus reported until now.

3.4.2 Consequences of under reporting

As we use the number of reported cases to estimate R_i , the ratio of detected cases among all infections can affect our estimation. We have seen in Section 2.2.4 that our estimator was invariant to a constant scaling of I_i . We see now how the estimation of R_i behaves in practice with different under-reporting settings. To do so, we take the daily cases in the United States and consider that they represent the total number of real cases. We then apply different detection ratios on it and look at the differences in both estimations of R_i . Increasing rates of detection are a good representation of the reality where testing capacity, detection methods and contact tracing improve over time. The reproduction number estimations resulting from 3 generated time series of under-reported cases are shown in the *results* section (see Section 4.1.1).

3.5 Modeling

In this section we explore how to combine the various complex epidemic parameters listed below to find patterns that may be impossible to identify visually. The advantage of complex modeling is also its scalability, in that we can combine the data of many countries in the hope of finding patterns, analogies and coherence to better support causality.

3.5.1 Features

The model we use is based on a set of features which are either directly or indirectly describe the setting and epidemic context. In our setting, the spatial and temporal features are essential and are defined at a resolution of country (place) and day (time).

Below is a complete list of features that have been considered in the design of the model. Those which proved to be invariant, were poorly correlated or heterogeneously collected across

countries were removed.

List of features

Category	Feature	Description	Remark
Epidemic	relative cases	number of daily new COVID-19 cases relative to the population	
	relative deaths	number of daily new COVID-19 deaths relative to the population	
	relative tests	number of daily new COVID-19 tests relative to the population	◊poor
	R estimation	Estimation of the effective reproduction number as described in Section 3.4	
	outbreak progression	time evolution of the outbreak, in days, since the country has reported a total of 15 cases per 1M inhabitants	
Demography	population density		*constant
	median age		
	aged 65 or older	proportion of the population older than 65 years old	
	aged 70 or older	proportion of the population older than 70 years old	
	gdp per capita	gross domestic product per capita	
Sanitary	diabetes prevalence		*constant
	hospital beds	number of hospital beds per 1000 inhabitants	*constant ◊poor
	life expectancy		*constant

Continent	Africa	binary indicators equal to 1 if it belongs to the continent, 0 otherwise	*constant
	Europe		
	Asia		
	North America		
	South America		
	Oceania		
Mobility	retail & recreation	In %, relative to a specific day before the outbreak	(Google reports)
	grocery & pharmacy		
	parks		
	transit stations		
	workplaces		
	residential		
	driving		(Apple reports) ◊poor
	transit		
	walking		
Weather	max temperature	maximum temperature over the day in °C	monitored in the capital of each country
	min temperature	minimum temperature over the day in °C	
	perceived temperature	average perceived temperature over the day in °C	
	humidity	humidity averaged over the day in %	
	pressure	pressure averaged over the day in mb	

Policies	school closing	the time-varying level of application of each policy (from 0 to 4) as defined in Section 2.1.1	
	workspace closing		
	cancel public events		
	restrictions on gathering size		
	public transports closure		
	home confinement order		
	restrictions on internal movement		
	restrictions on international travel		

The **constant* remark indicates that the feature is constant over time while the *◊poor* remark designates a feature that is poorly reported (unavailable in most countries or available for a few days only). When we select a subset of those features for a modeling task, we require the data to be complete. If one of the selected feature is not available for a country at a certain day, the sample is not considered. It implies that using a features that is not accessible in some countries also means to take the said countries out of our analysis.

3.5.2 Cross correlations

As a global overview of the relationships between parameters, we computed the *Pearson correlation coefficient*: a cross correlation over all pairs of a subset of features. The coefficient is visualised in a heatmap in *results* section (see Section 4.3). The coefficient is bounded between -1 and +1 and values tending toward either pole are considered negative or positive correlations respectively.

3.5.3 Leave-one-out training

In Machine Learning, cross-validation is commonly used to assess the *generalizability* of a model. *Leave-one-out* cross validation works by splitting the data into batches (complementary subsets) and sequentially selecting them one-by-one as a validation set while the training of the model is performed on the rest of the data. In our context, this procedure has many advantages. For example, whereas batch-splitting commonly creates batches of random size, we use “country”

as the natural batch-decomposition of the global dataset. Thus, while we keep the technical advantage of having an isolated validation set, we also retain interpretable validation batches that represent the timeline of the outbreak in specific countries. By choosing such cross-validation subsets, we are able to predict the evolution of a feature by using data of other countries as training data. In practice, it can be used to predict a value that is missing in a certain country. As the subsets are not equally sized, we need to consider their sizes when we are averaging an error measure over multiple batches by weighting the error of each country by its number of samples.

3.5.4 Predictions and error

We use Machine Learning models to produce predictions of various outputs that are optimized according to the mean absolute error (MAE). When we show prediction plots, we display the ground truth (in blue), the prediction (in orange) and the absolute error (in green) over time for specific countries. It is thus visually interpretable by looking at how the prediction differs from the ground truth. We generally provide the mean absolute error from all predictions (every country), as well as the mean absolute error of our 6 selected countries. It is important to keep in mind that in some countries, noisy data or bad reporting practices can affect both the predicted value and the ground truth. For that reason, we provide the result obtained for the 6 selected countries as well.

3.5.5 Linear model for mobility prediction

We start by using a linear model to explore the relationship between policies and mobility (links numbered 1 and 2 in the causality graph 3.1). This allows us to see how well mobility can be predicted by policies (i.e. how policies may have impacted human movement/interactions).

Once the model performances have been evaluated, we fit a linear model on the whole dataset to obtain information by looking at the coefficients that define our optimal setting. As the policies are the only input of the model, we can state that the more strongly a policy is weighted by our model, the more impact it has on a specific mobility value.

We show the predictions results and the features importance in the *results* section (see Section 4.4 and 4.4.1).

3.5.6 LSTM Neural Network

After using a linear model to analyze a “direct” relationship, we need more complex modeling to analyze the value of our interest: the reproduction number. We use different subsets of our features in different settings in accordance to the results we want to obtain. As explained in

Section 2.4, LSTM networks are good to handle time series, but our features are composed of both time series and constant features (see Subsection 3.5.1). To manage this heterogeneity in feature types we create a hybrid network combining an LSTM part and a *multilayer perceptron* (MLP). This fully connected part is reserved for the constant features that are inputted with no time dimension, as the LSTM manages the time evolving features (Fig. 3.3). The memory window of the LSTM layer can be parameterized in number of days. As an example, a memory window of 3 days means that to predict the output at a day d , the times series of days d , $d - 1$ and $d - 2$ are used as inputs.

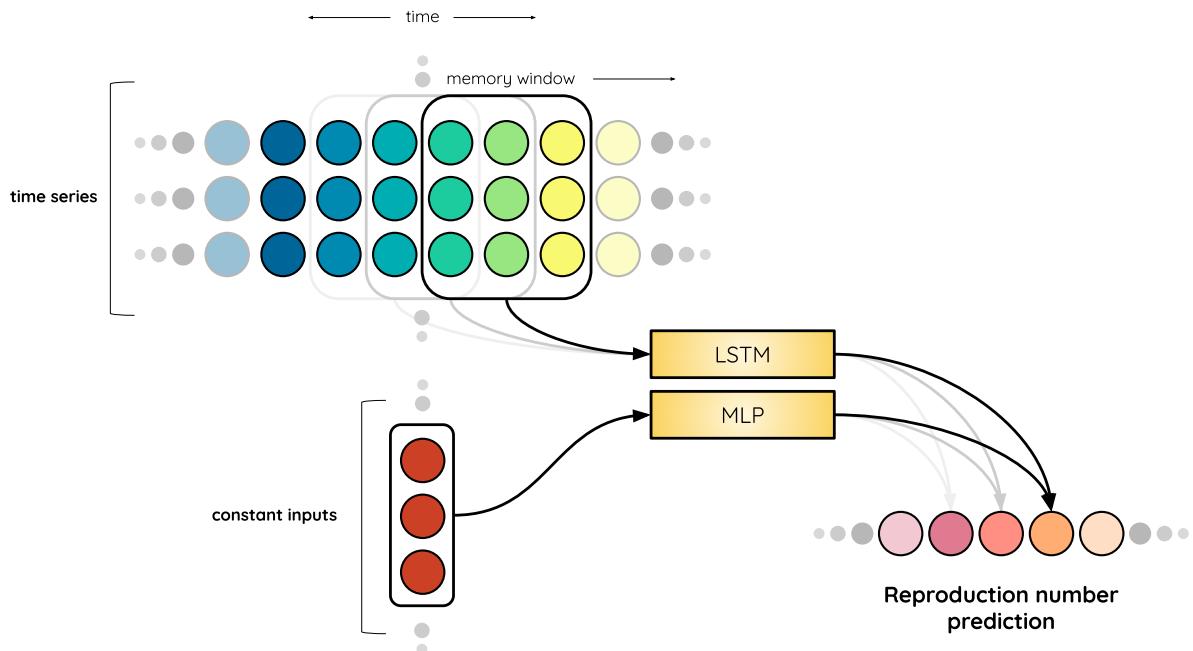


Figure 3.3: Structure of the hybrid model for R predictions, figure created by author

To avoid over-fitting, we use dropout layers that we place after each LSTM or MLP layer. The dropout probability used is 0.2. In addition, we apply early stopping that will stop the training before the maximum epochs if the loss is not decreasing for 5 epochs. In this event, the model restores the weights captured from the epoch corresponding to the best loss value. We show below a more technical scheme of the model and all its layers (Fig. 3.4), where *dense* layers correspond to the fully connected layers (MLP). All the inner layers (*lstm_139*, *dense_278* and *dense_279*) have 15 hidden nodes as outputs, while the target layer naturally outputs a uni-dimensional value (reproduction number).

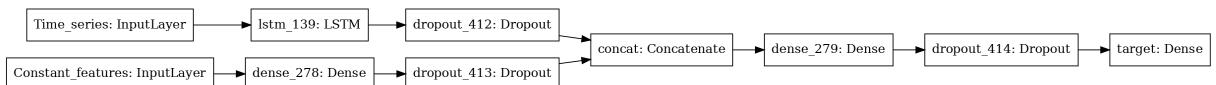


Figure 3.4: Detailed structure of the hybrid model for R predictions

3.5.7 Country selection

The countries that take part in the training are selected automatically according to various criteria. As seen in the *features* section (see Subsection 3.5.1), the availability of the input features determines the initial selection step, as a country for which one of the features is not present cannot be used in the training. A second selection step is performed to address the reliability of the reproduction number estimation. As we have seen, this estimation is subject to a lot of uncertainty across the relationships that it is based on. To avoid fitting our model to poorly relevant targets, we filtered out the countries for which the reproduction number estimation was unavailable for more than 10% of the days since the first valid estimation. Indeed, the first cause of an invalid estimation is a low number of daily cases (less than 5), in which case we assume that either the country is not subject to a significant outbreak or the cases are not reliably reported. The second cause is that the estimation goes too frequently over what we defined as the maximum plausible value (see Subsection 3.4.1). This second filter eliminates 59 countries (from 179 countries to 120 countries).

3.5.8 Baseline prediction

As a baseline for prediction and before proceeding to feature selection, we train the model on the *outbreak progression* feature only. It allows us to have a preliminary idea of what our model is capable of with the most basic input: i.e. a simple scaling of each country centered on the day when 15 cases per 1 million inhabitants were reported.

3.5.9 Feature selection

Before training our model, the features need to be carefully selected for reasons of performance but also of interest. If the goal of our model is to be able to predict the reproduction number in countries where the estimation cannot be done due to bad reporting, it would make no sense to train a model with these badly reported features. Thus, we do not consider features related to cases, deaths and tests. In addition, as it is important to keep a meaningful number of countries in the training, we remove the features that are considered as *poor* (see Subsection 3.5.1). The features concerned are *test* (that was already taken out for relevance reasons), hospital beds and the mobility features coming from Apple reports.

The second feature selection is according to model performance. As this step is computationally heavy, we concentrate the feature selection on the 6 countries that have been chosen as representative. Starting with Google mobility features (Apple reports have been eliminated), we perform a forward feature selection with sets of features grouped by category (demographic, sanitary, ...). At each step, we select the category that is associated to the best reduction in MAE. If a group is added, we apply a backward selection on the individual features to optionally elimi-

nate those that would have increased the error. We then continue with the remaining categories until our error is no longer reducing. We then continue with a forward selection on individual remaining features. The whole process is performed on the same data that corresponds to the largest subset of countries for which each selected feature is available (93 countries). The said list of countries is shown in Appendix A.2.

The model used for the feature selection has a memory window of 3 days and runs for a maximum of 50 epochs.

3.5.10 Time window selection for the LSTM

Our last model parameter to optimize is the memory time window of the LSTM layer. Again, we did this selection by evaluating the model performance on the 6 countries with time windows ranging from 1 to 8 days. To compare the errors produced by models of varying time windows, we need to consider the following effect. As an LSTM model of D -days memory needs $D - 1$ days in the past to do a prediction, the first prediction appears at the D^{th} day. If we compare the errors resulting from different time windows, we need to remove the $W - D$ first days of data, where W is the maximum window size considered in the comparison. It ensures that we are comparing outputs over the same range of days. Otherwise, we may penalize models of shorter windows by considering the noisier values at the start of the epidemic, while they are not penalizing the models of larger windows.

The model used for the time window selection also runs for a maximum of 50 epochs.

3.5.11 Ultimate reproduction number prediction

With the features and the LSTM time window selected optimally, we run final predictions on all available countries to evaluate the mean absolute error for each country, as well as a global average error. We also plot the prediction for our 6 selected countries. This time, a maximum number of 150 epochs is applied, which is not supposed to be reached because of early stopping.

3.5.12 Policy effect

To assess the impact of policies on the reproduction number, we modify our model by replacing the mobility features with policies. This is a consequent change that is based on the assumption that most variations in mobility are caused by the policies (see causality graph 3.1). Indeed, if mobility is mainly caused by the policies, a lot of information about mobility can be derived from the policies. Secondly, it is important to ensure that the other features (demography, weather, ...) are not affected by the policies, otherwise their effect could be over-shadowed by those

features. This can be ensured by the two following facts. Firstly, the constant feature values (demography, sanitary, ...) did not change since the beginning of the pandemic and thus could not be influenced by government measures. Secondly, it is fair to assume without any proof that the weather parameters do not depend on policies.

By replacing mobility information with policy inputs, we ensure that the information brought by the policies is not over-shadowed by mobility data. However, the policy information is less precise than mobility data and its relation with the reproduction number is less direct. Therefore, we expect a poorer prediction, that will however permit us to have a better understanding of the policy impact.

At first, we evaluate our modified model and look at the performance of the prediction using policies instead of mobility inputs. We then virtually remove the implementation of each policy one-by-one by modifying the inputs in order to look at the difference between predictions with and without the said policy. We will call this difference the **impact** of the policy that can be depicted over time. Finally, we report the average impact of each policy (as an average reduction of the reproduction number) in each of the 6 chosen countries. The level of application of the policies needs to be considered as well, as a more rigorously implemented policy is expected to have a greater impact. To visualize this new notion, we also plot the evolving impact in Switzerland.

Chapter 4

Results and interpretation

4.1 Delays between events

We displayed here basic statistics about the delays between events of interest. For each delay, we removed entries for which the delays were considered as implausible, i.e. a consequence of wrongly reported dates.

- Total number of reported entries: **30460**
- Number of entries with reported symptom onset: **4654** (15.28%)
- Number of entries with reported death: **1593** (5.23%)

delay between	min/max accepted values	average delay	std. dev.
symptom onset & case	0 - 90 (4536 results)	4.596	5.328
symptom onset & report	0 - 90 (4594 results)	5.965	5.362
symptom onset & death	0 - 150 (431 results)	14.411	9.499
case & report	0 - 20 (30125 results)	1.546	1.006
case & death	0 - 120 (1568 results)	9.731	7.527

We see here that the standard deviation of case-to-death delay is greater than 7 days, as the standard deviation of symptom onset to deaths is almost double that of the symptom onset to case. This is the main reason why using deaths to estimate the reproduction number is not as accurate as using reported cases, even though the total number of reported entries may be more reliable (see Section 2.2.4). We note that the noise in delays that comes naturally with deaths affects our R estimation more than the noise in amplitude that reported cases are subject to.

To estimate the mean time between the infection and the report, we consider the symptom onset to report delay whose average value is about 6 days. This number has to be added to the

incubation time (time between the infection and the symptom onset) which has been estimated to be approximately 5 days on average by the World Health Organization and an early study made in Wuhan in March 2020 [1] [23]. We thus consider a **delay of 11 days** as the most appropriate shifting value for our reproduction number estimation. This value of 11 days has also been found in the Swiss study that we mentioned earlier [6].

4.1.1 Consequences of under reporting

As under-reporting settings, we considered 3 detection ratios:

- A constant detection rate of 80% of the real cases
- A linearly increasing ratio that starts at 0% and reaches its maximum of 80% after 60 days
- An increasing ratio that starts around 0% and tends to its maximum of 80% with a logistic shape

The resulting estimations of effective reproduction numbers are shown below (Figs. 4.1, 4.2 and 4.3).

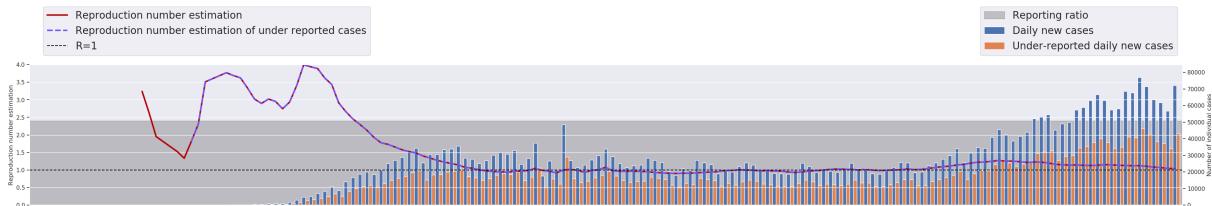


Figure 4.1: R estimation with constant under-reporting

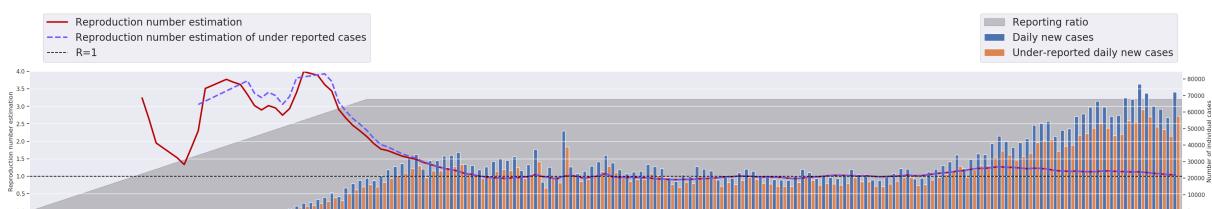


Figure 4.2: R estimation with linearly increasing under-reporting that reaches its maximum at the beginning of the epidemic

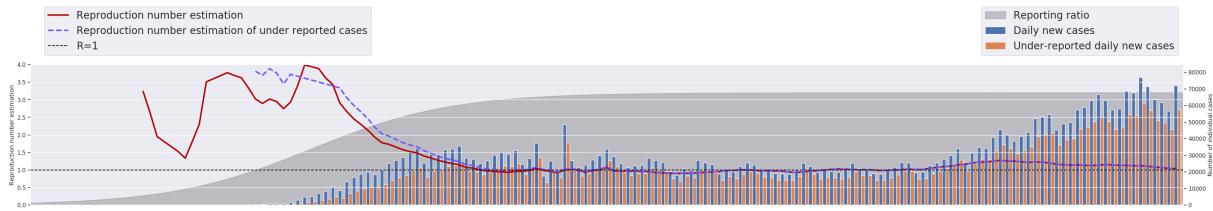


Figure 4.3: R estimation with an increasing under-reporting

Here, the blue and orange bars represent the *real* and the *under-reported* number of cases respectively. The resulting estimated reproduction numbers are shown with the red solid line (*real*) and the dashed purple line (*under-reported*) respectively. As expected, the constant reporting ratio does not affect the estimation: both lines are exactly the same. On the two plots representing increasing detection ratios, the estimation tends to be over-estimated. Indeed, an increasing detection rate makes the growth rate appear larger than it is in reality. Note that the R estimation of under-estimated cases starts later than the other one. The reason is that the under reporting makes the number of cases pass below the threshold of minimum cases for which we consider that the R estimation is valid (see Section 3.4.1).

4.2 Summarizing plots

As an overall view, we provide below a few plots that combine information of our different sources for the 6 countries selected. With bar plots, we show the daily number of cases and deaths, as the R estimation and the mobility are represented by curves. We chose to represent 3 categories of mobility: transport, driving and residential. The stringency index over time is also represented by a curve. To represent the implementation times of policies, we placed vertical dashed lines when one or more policies were implemented and when the level of a policy has been increased. We denoted each policy according to the provided dictionary (see Tab. 4.2). Their stringency level is written in parenthesis and follows the definitions in Section 2.1.1.

keyword	policy
c1	school closing
c2	workplace closing
c3	cancel public events
c4	restrictions on gathering size
c5	close public transport
c6	“shelter-in-place” and home confinement orders
c7	restrictions on internal movement
c8	restrictions on international travel

Figure 4.4: Dictionary of policies keywords

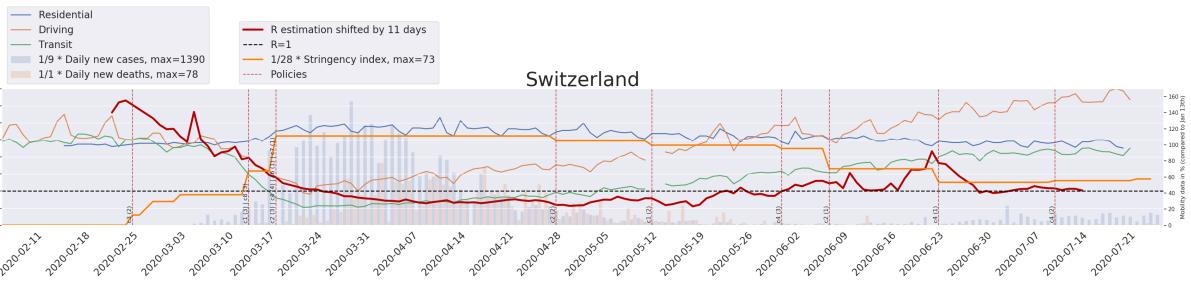


Figure 4.5: Visualization of the COVID-19 outbreak in Switzerland

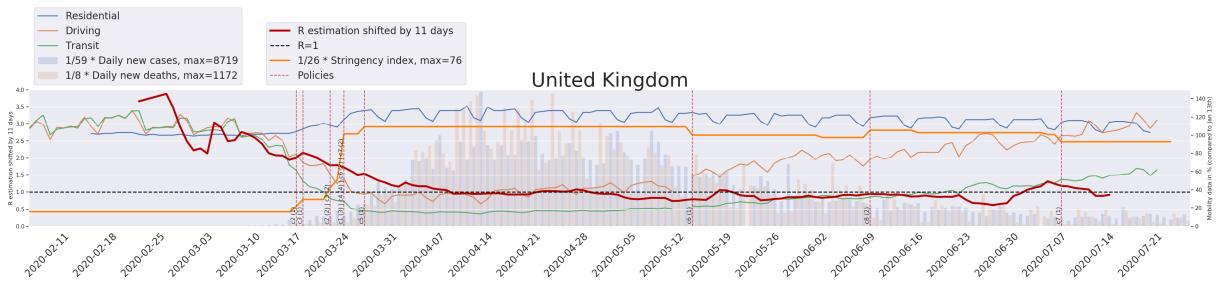


Figure 4.6: Visualization of the COVID-19 outbreak in the United Kingdom

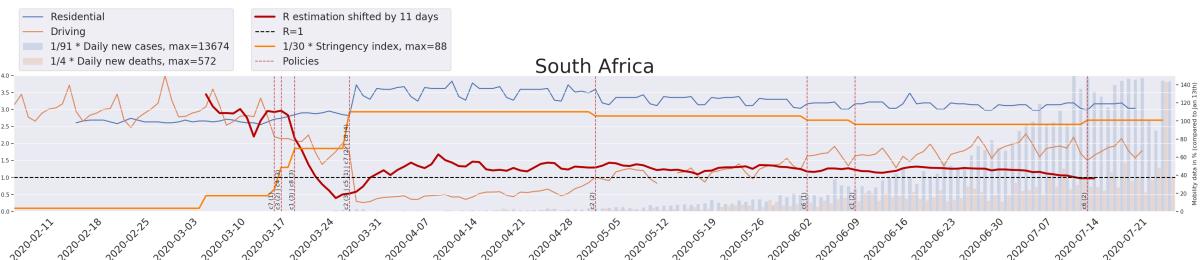


Figure 4.7: Visualization of the COVID-19 outbreak in South Africa

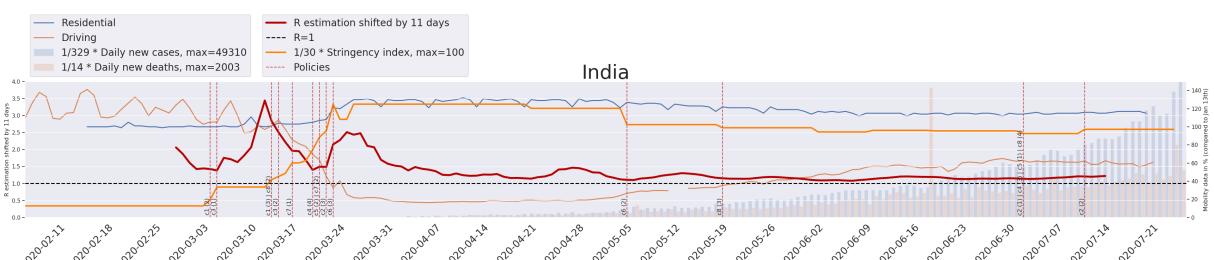


Figure 4.8: Visualization of the COVID-19 outbreak in India

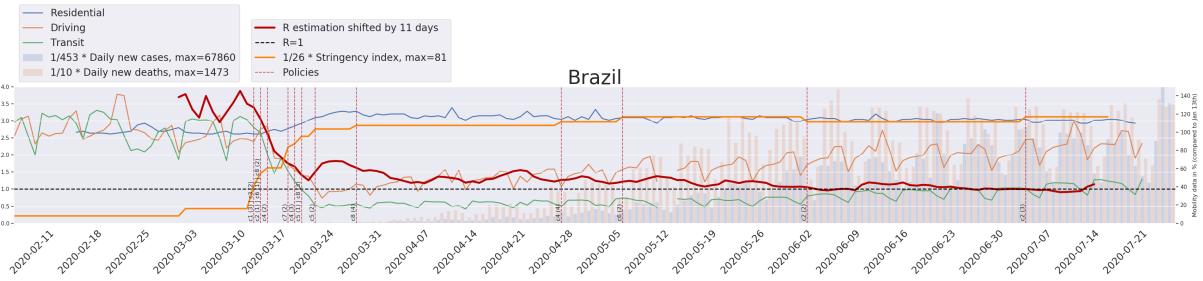


Figure 4.9: Visualization of the COVID-19 outbreak in Brazil

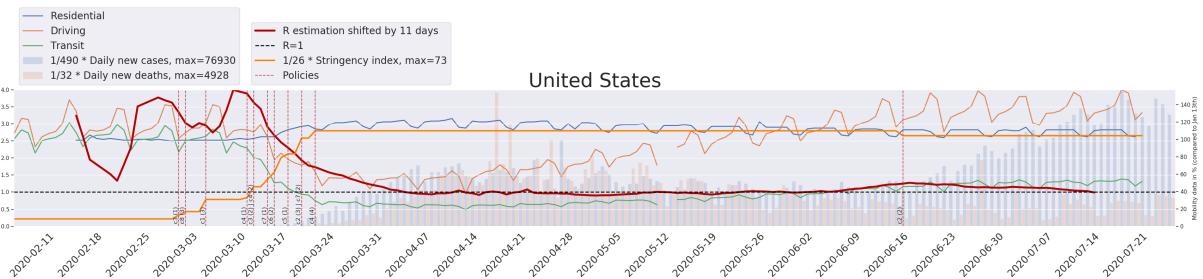


Figure 4.10: Visualization of the COVID-19 outbreak in the United States

Plots of all countries can be found at: http://go.epfl.ch/covid_plots

A first observation that we can make from these visualizations is a certain correlation between modes of mobility (public transport and driving) and the effective reproduction number estimation. We can explain it intuitively by the fact that a reduction in mobility is a good proxy of human interaction and thus viral transmission. We also see through the *residential* indicator that people spent more time at home after the implementation of the policies. Some of these are clearly correlated to policy implementation, for instance, in South Africa, the curve rises sharply when policies for workspace and public transport closing is applied.

As a second observation, we also note that the implementation times of policies usually matches with a drop of the reproduction number, with variable delay which is hardly observable from the plots.

4.3 Cross correlations

The grid displayed below (Fig. 4.11) is a heatmap where the cells are colored according to the *Pearson correlation coefficient* between two features. Deep red corresponds to a strongly positive correlation, deep blue corresponds to a strongly negative correlation while color tones between the two extremes are linearly mapped according to the value of the correlation. A low correlation

tends to a grey tint. The list of countries that have been considered for the cross-correlations was automatically selected according to the available data and can be seen in Appendix A.1.

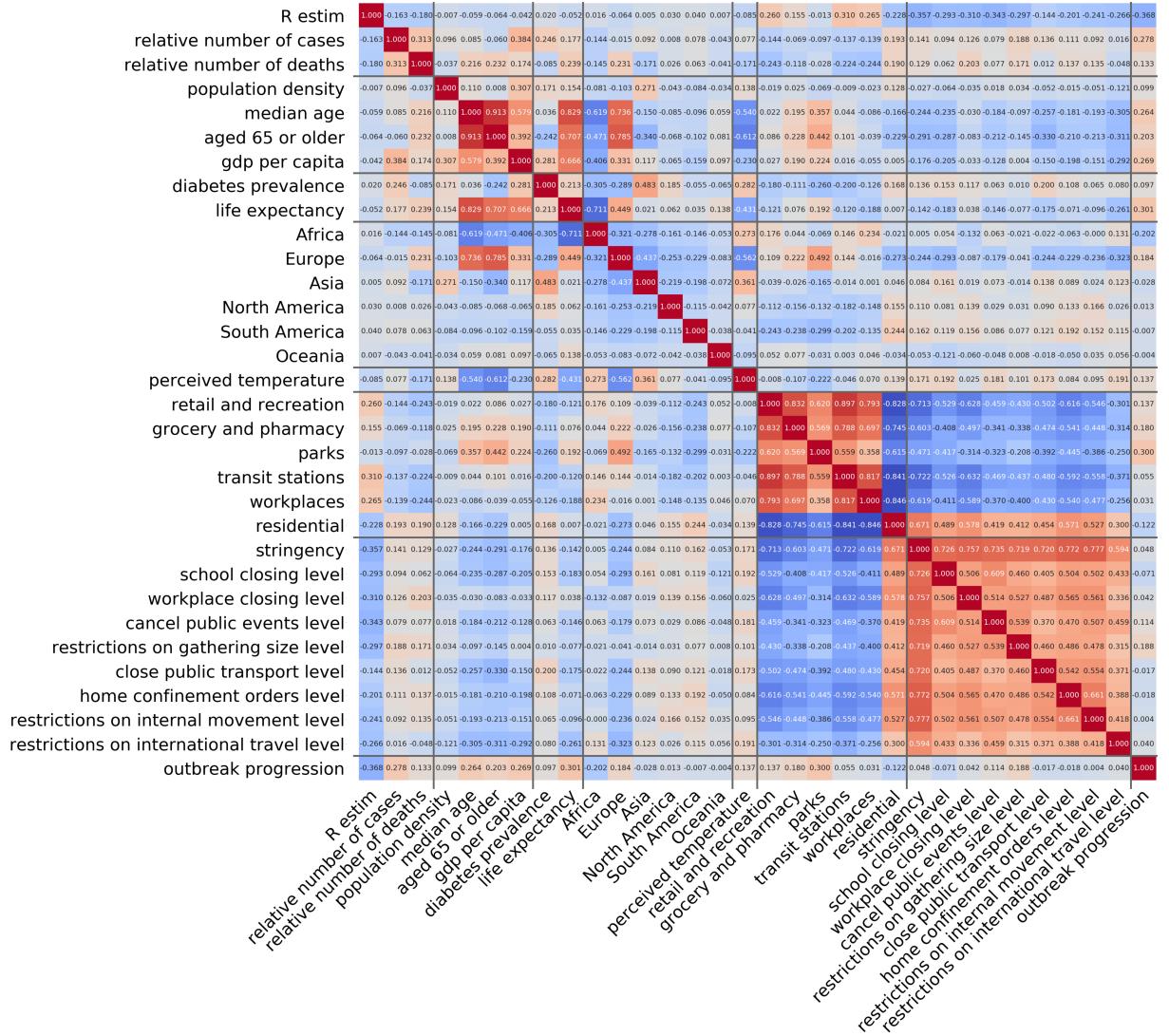


Figure 4.11: Cross correlations of a selected subset of features

The first interesting and immediately perceptible fact to note is the strong correlation between mobility and policies. Indeed, the two rectangles representing their intersection are represented by deep colors. As mobility and occupancy of public spaces is negatively correlated with the application level of policies, the housing occupancy rate is positively correlated with them. It is an expected result as government measures tend to reduce the mobility and induce or force people to stay at home. This observation confirms the relevance of the links 1 and 2 of our causality graph (see Fig. 3.1).

Another observation to make is that European countries are associated with lower application

levels of policies globally over the whole period of consideration. However, we need to be careful about the continents features as they are represented by unequal number of countries, and thus an unequal amount of data.

Regarding the reproduction number estimation, we see that it is actually negatively correlated with the level of policies, which is reassuring. This will be analyzed more in detail by the LSTM model.

4.4 Mobility predictions

By fitting our linear model on each mobility profile using the policies features only, we obtained the errors that are shown in the appendix (see Table A.3). We have seen with the cross correlations (see Section 4.3) that policies and mobility were correlated. From those results, we see that the *transit* indicator showing changes in distance covered by public transports is the most predictable (according to the ratio MAE / std. dev. of full data). We also note that the mobility in the United States can barely be predicted by policies. This phenomenon could be explained by the fact that the policy implementation is often applied at the state or county level and thus does not appear in our data, even though they affect the mobility as much as the policies applied at the country level. Sometimes, the mobility looks very noisy and does not seem to be explainable by the policies (ex.: parks occupancy in Switzerland, see Fig. 4.12). On the other hand, some indicators are really well explained by the policies (ex.: *transit* in the United Kingdom and *residential* in India, see Figs. 4.13 and 4.14).

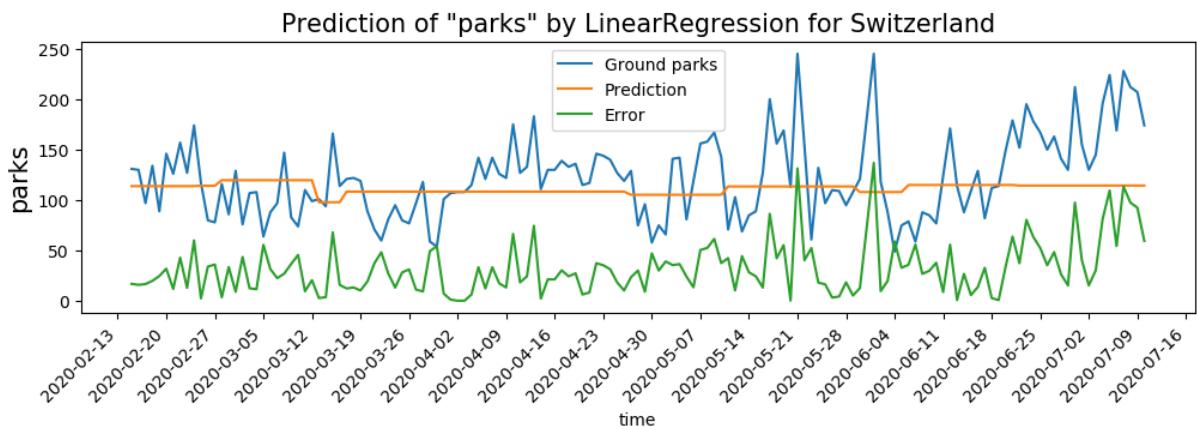


Figure 4.12: Parks occupancy prediction in Switzerland

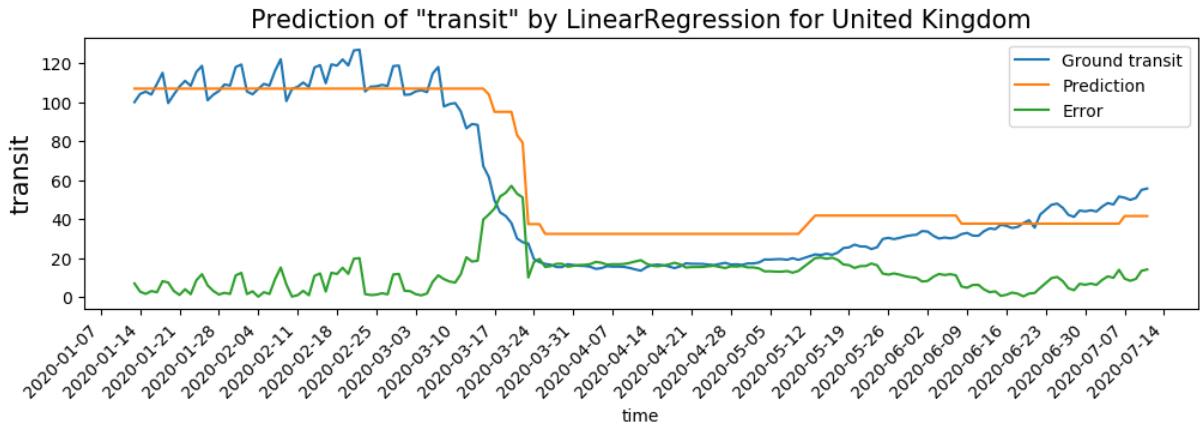


Figure 4.13: Kilometers traveled in transports in the United Kingdom

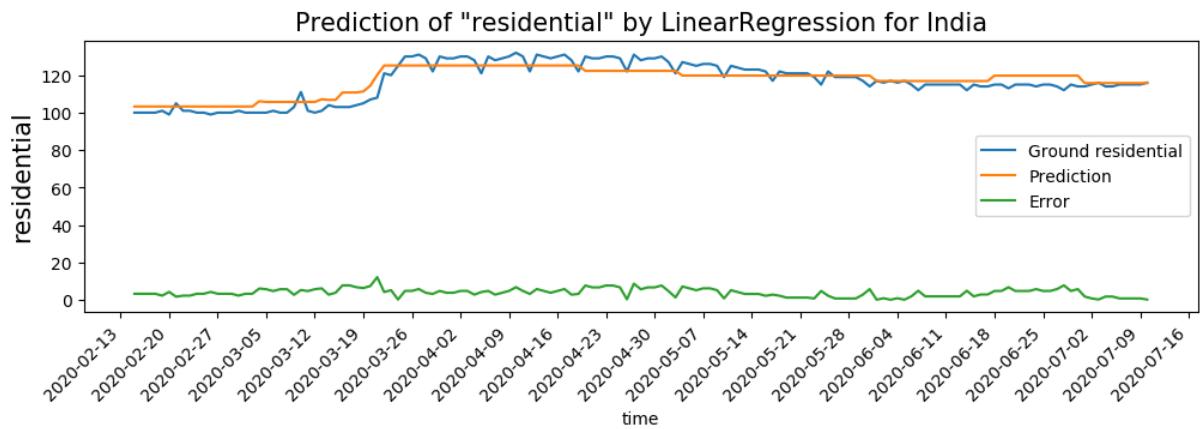
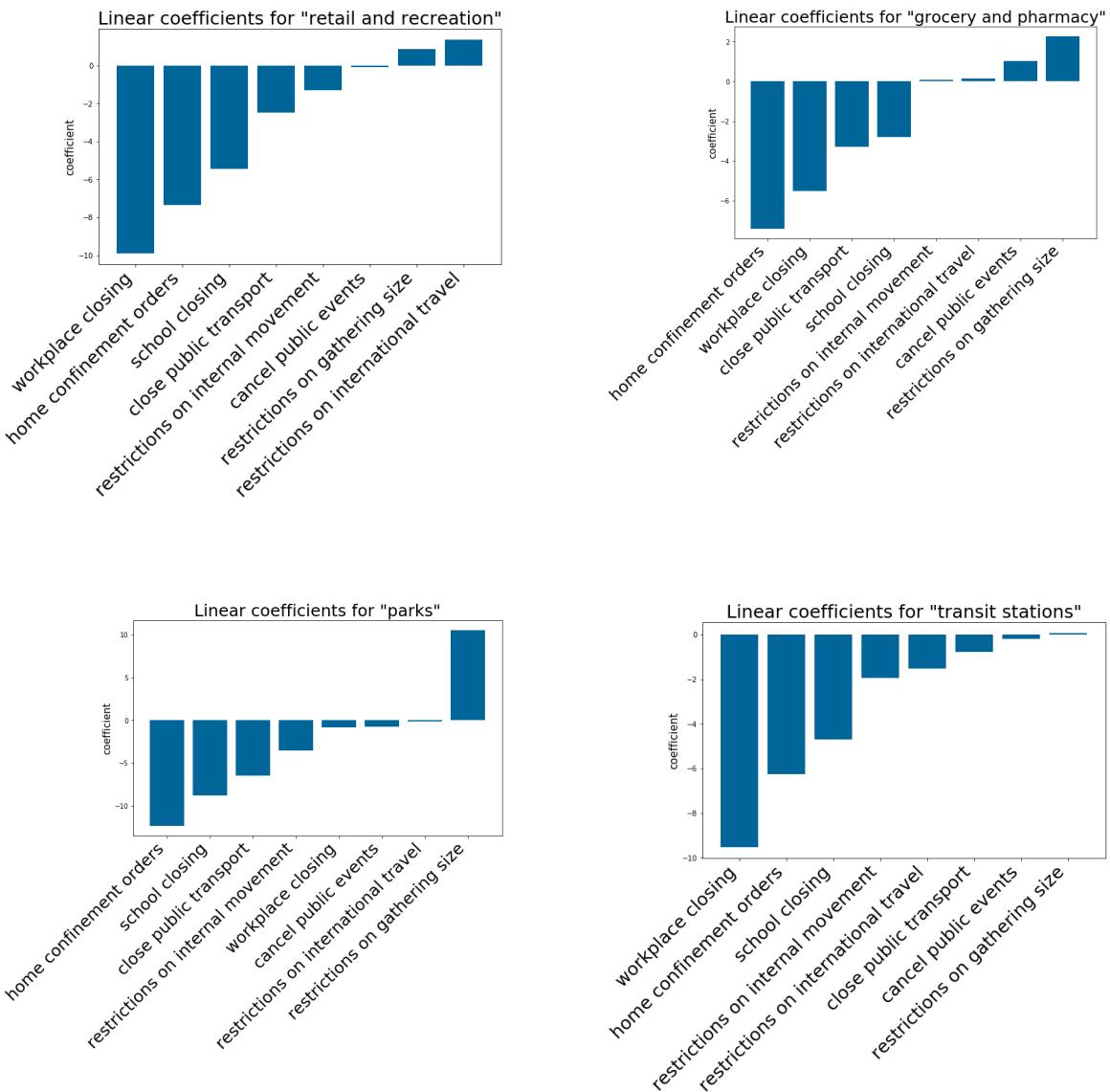
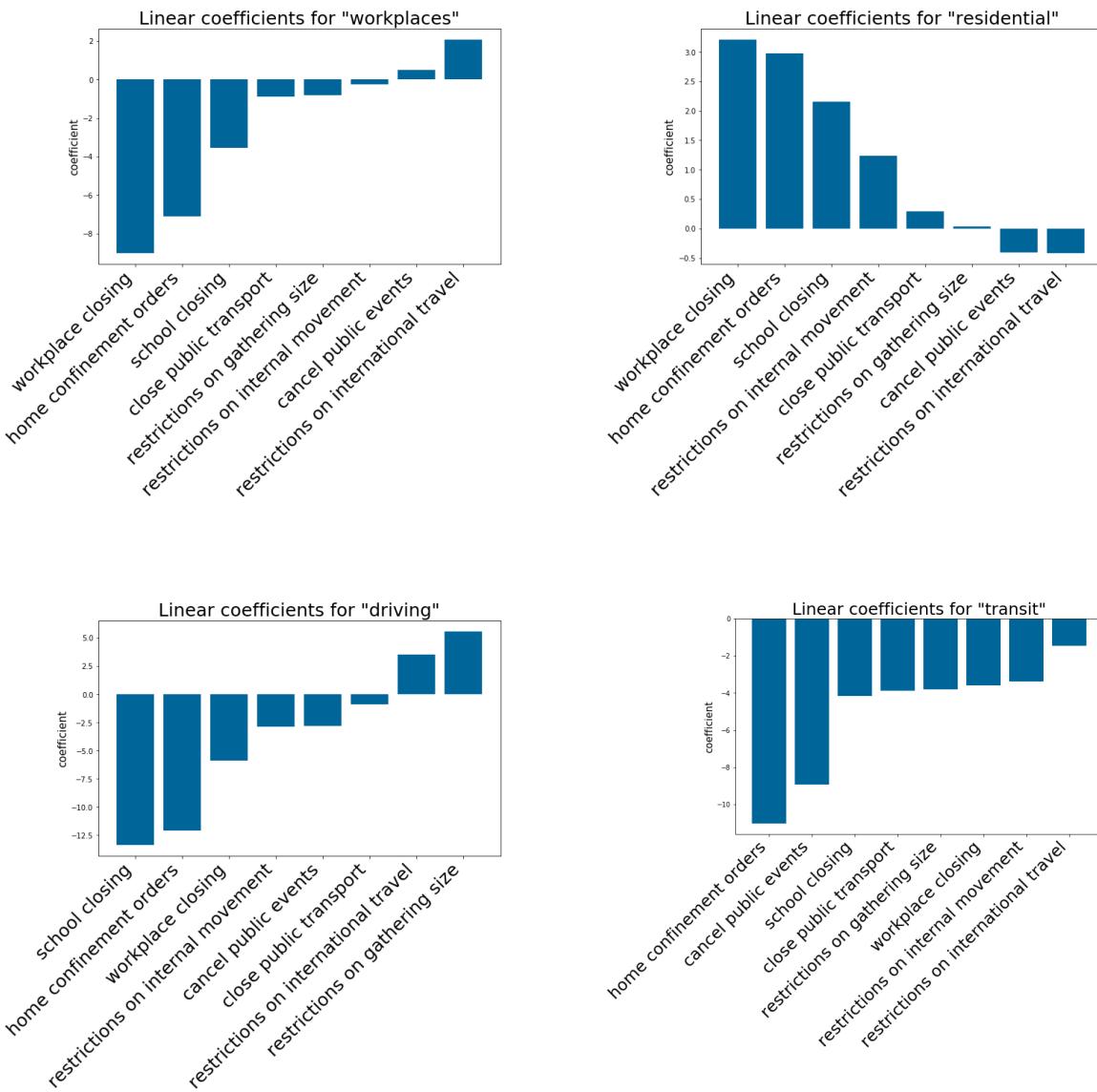


Figure 4.14: Home occupancy in India

4.4.1 Feature importance of linear model

We show below the policies ordered by their weights in the models fitted for each mobility value on the global data. We ordered them by *positive impact*, assuming that each policy tends to reduce the value of the mobility, except for the *residential* occupancy that is supposed to be increased by policies.





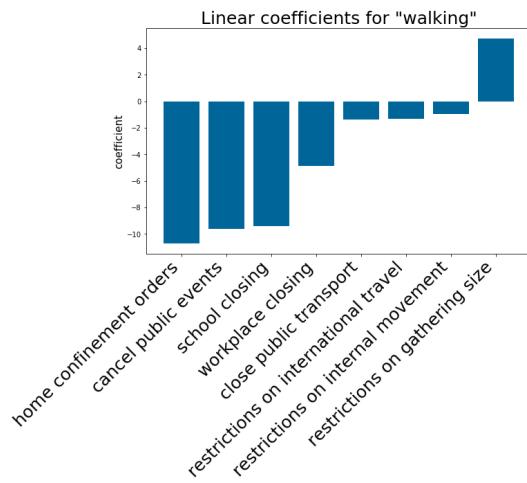


Figure 4.15: Policies ordered by their impact on each mobility category

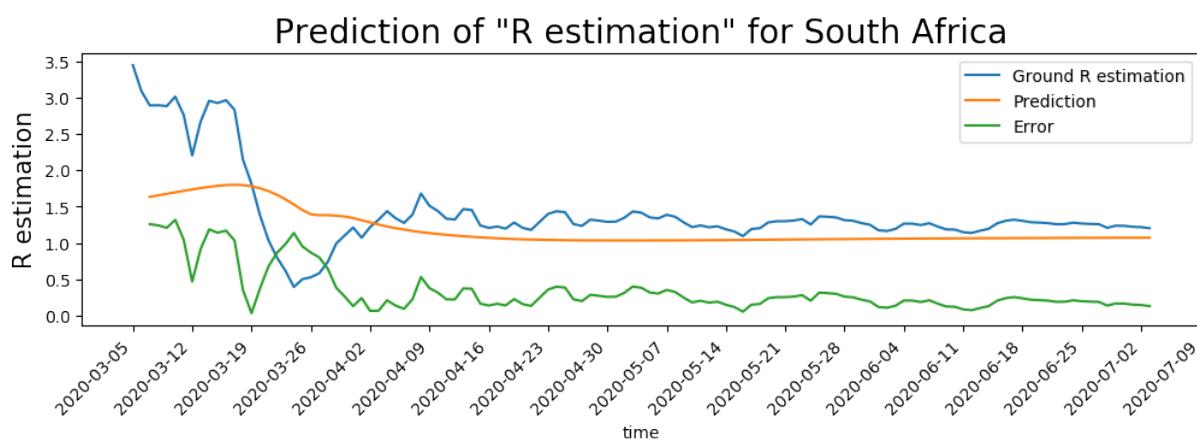
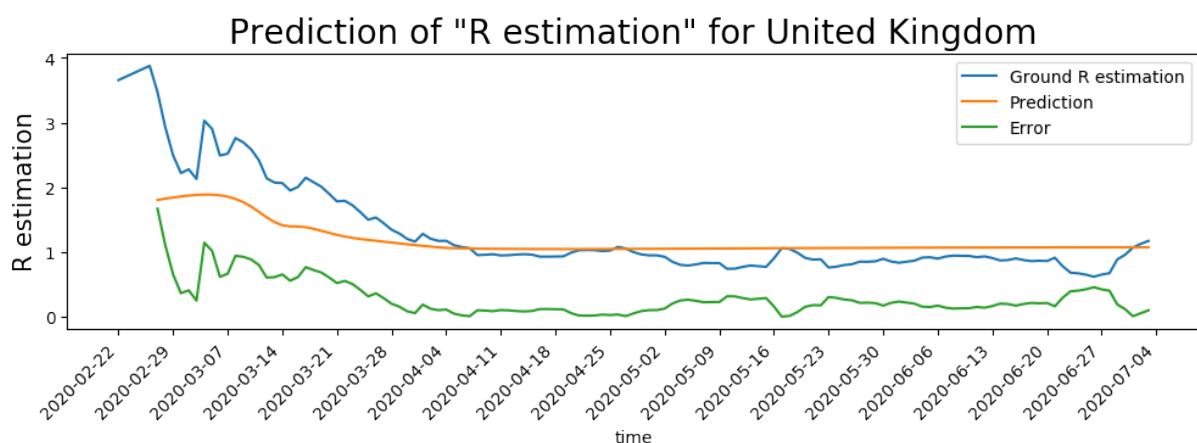
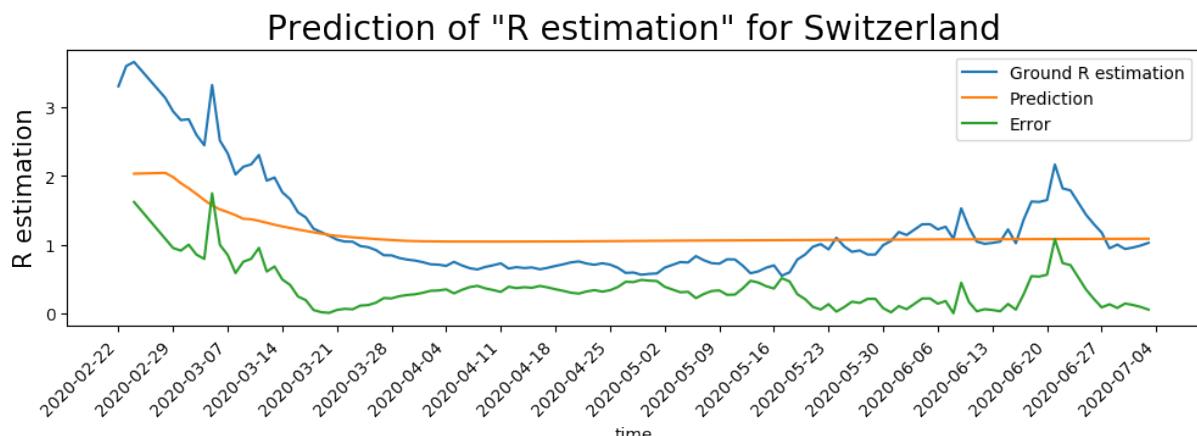
We see from those plots that *home confinement orders* and *workplaces closure* are the policies that most affect the mobility in general. While policies are usually affecting mobility in the right way, it is interesting to remark that the restrictions on gathering size increases the driving and walking mobility as well as parks occupancy.

4.5 R prediction with an LSTM model

4.5.1 Baseline using a unique simple feature

We show below the results obtained by training the model with the *outbreak progression* feature only (Table 4.5.1). The prediction plots of the 6 countries are displayed below the table (Figs. 4.16).

Zone	Mean Absolute Error
World (average of 120 countries)	0.293
Switzerland	0.367
United Kingdom	0.267
South Africa	0.331
India	0.255
Brazil	0.281
United States	0.321



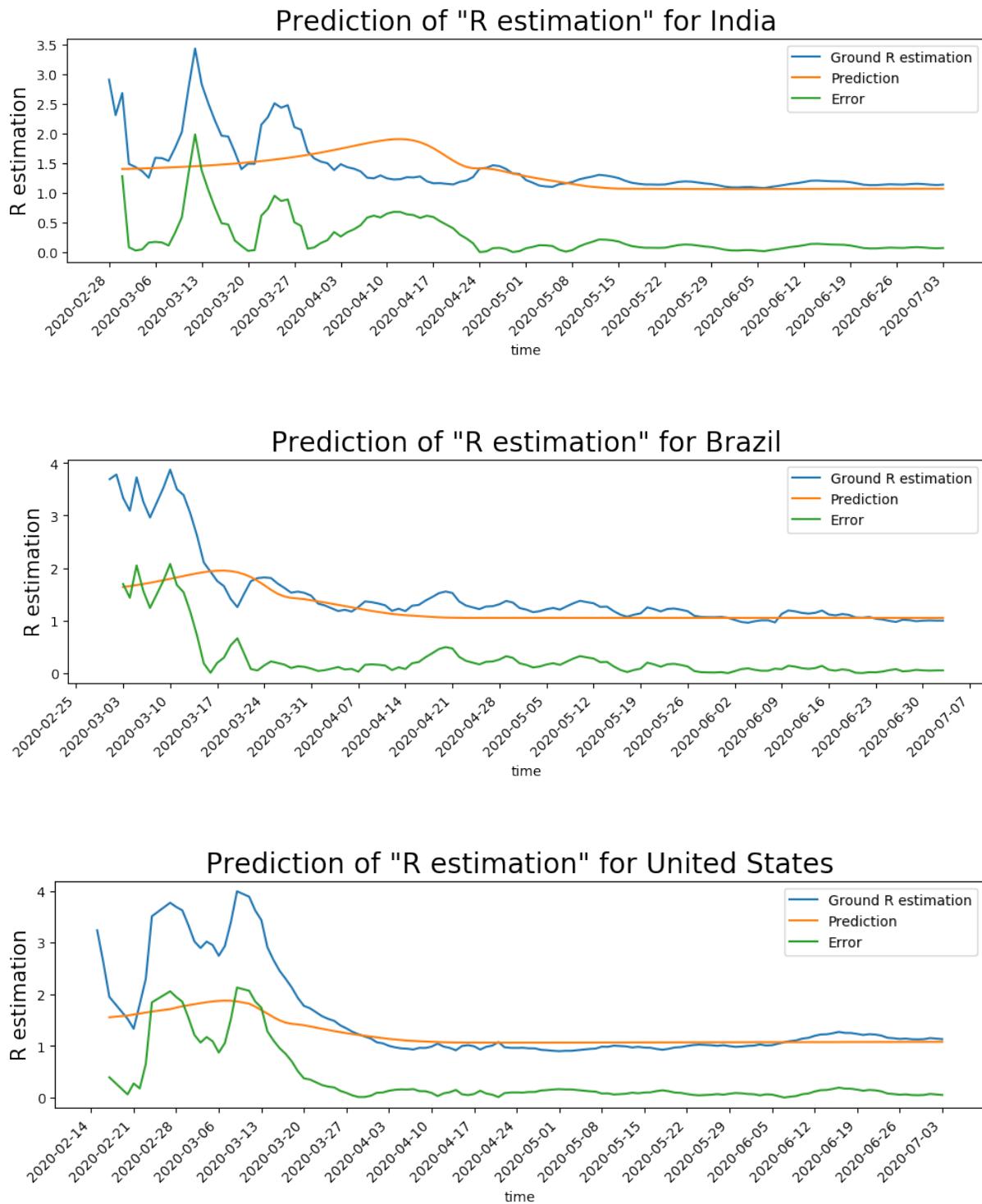


Figure 4.16: Baseline prediction using *outbreak progression* as a unique feature

We see here that the predictions based on the *outbreak progression* feature takes advantage of the fact that a common shape is appearing in many countries, where the R starts at a high

value and decreases after some time. As the R estimation often converges to a value close to 1, this prediction simply considers a peak at the beginning and converges to a value close to 1. It makes it unable to catch any second peaks that do not necessarily match with other countries in terms of progression. We conclude that this prediction will tend to be increasingly inaccurate over the course of the epidemic as the impact of the starting high values is decreasing. For those reasons and with a long term perspective, the *outbreak progression* feature is a feature that we want to avoid in our model.

4.5.2 Feature selection

We show here the successive sequence of added or removed features. The details of the whole feature selection process can be found in Appendix A.4, where each setting is reported with the mean and standard deviation of the MAE over 4 runs.

1. Started with *Google mobility* features group
2. Added *demographic* features group
3. Added *pressure* feature

The obtained list of features after selection is the following:

- **Mobility**
 - Retail and recreation
 - Grocery and pharmacy
 - Parks
 - Transit stations
 - Workplaces
 - Residential
- **Demographic**
 - Population density
 - Median age
 - Aged 65 or older
 - Aged 70 or older
 - GDP per capita
- **Weather**

- Pressure

The data obtained from these 12 features is considered — in each country — only the days for which the reproduction number estimation is valid. It is available in 93 countries (see Appendix A.2) and it corresponds to 120 720 daily data points. We note the presence of all demographic features and the absence of sanitary features. Intuitively, it makes sense as diabetes prevalence and life expectancy is not supposed to influence reproduction numbers, except for some factors that are supposed to be covered by demographic information (age of the population). An even more interesting element is the presence of the *pressure* feature, that reduced the error in prediction by about 4.5%.

4.5.3 LSTM time window selection

The results obtained by comparing various LSTM time windows on our model are displayed in Appendix A.5. We selected a window of **7 days** as it gave the best results for 3 out of 6 countries. For the rest of the countries, the error was still close to the optimal one.

4.5.4 Ultimate reproduction number prediction

We show below the results obtained from the optimized model (Table 4.5.4). The prediction plots of the 6 countries are displayed below the table (Figs. 4.16). The complete list of results from all countries can be seen in Appendix A.6. We also display the learning curve of the model on a randomly selected (with probability 0.2) validation set (Fig. 4.17):

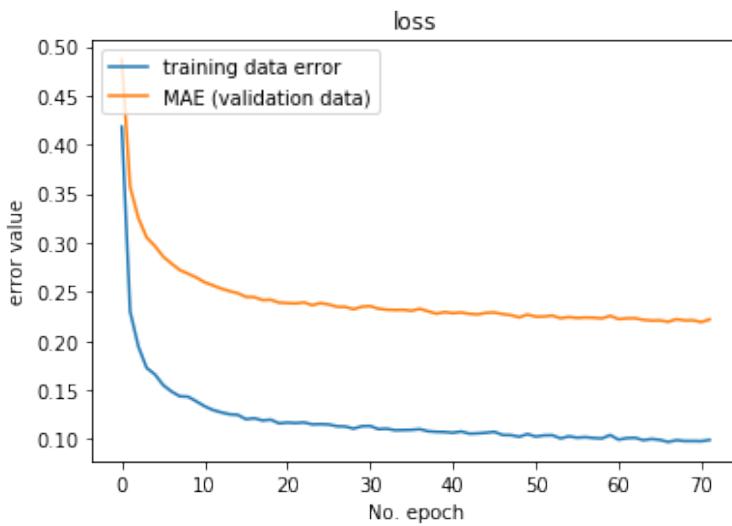


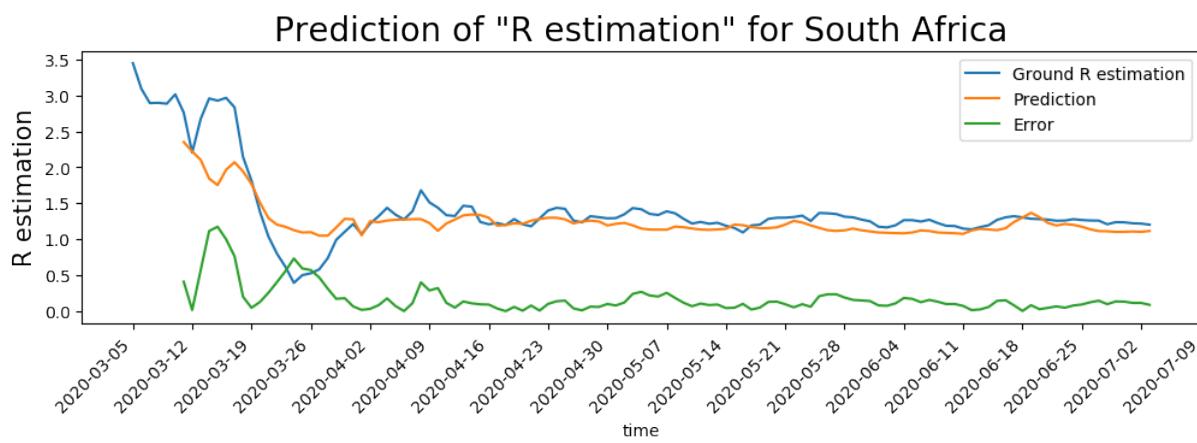
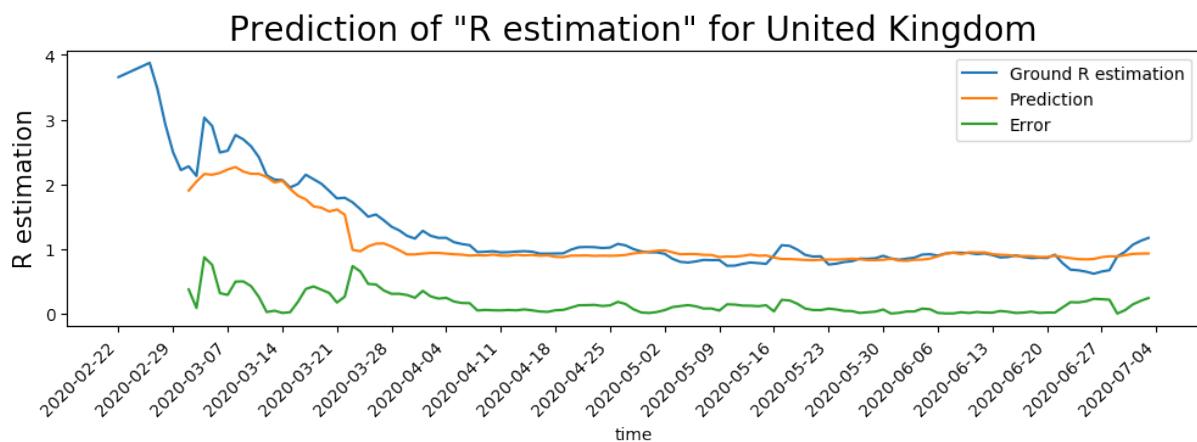
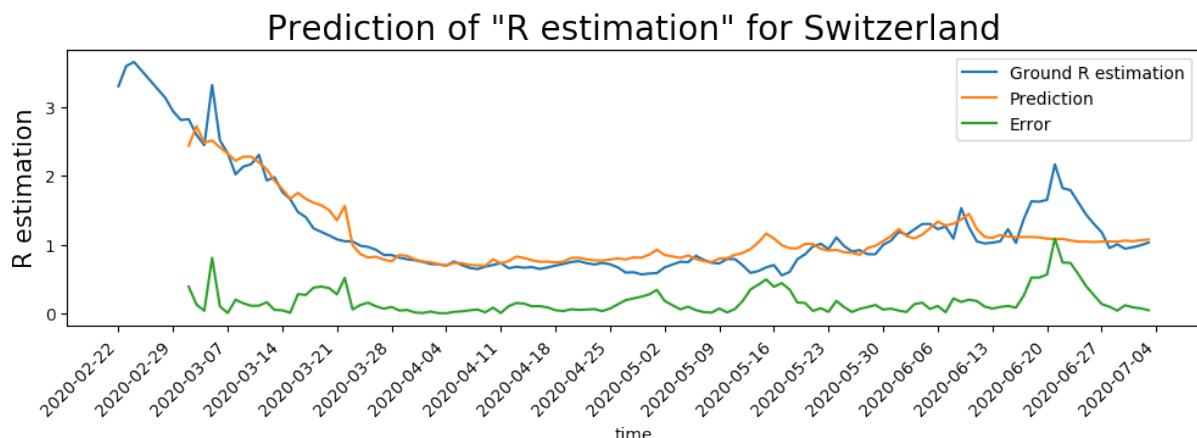
Figure 4.17: Learning curve of the model, stopped by early stopping regularizer

Zone	Mean Absolute Error	% Baseline improvement
World (average of 93 countries)	0.254	13.3
Switzerland	0.169	54.0
United Kingdom	0.163	39.0
South Africa	0.168	49.2
India	0.226	11.4
Brazil	0.201	28.5
United States	0.196	39.0

Figure 4.18: Model performance (mean absolute error) on the 6 selected countries

In comparison with the baseline prediction which takes into account the time evolution of the pandemic only, the improvement (in % of MAE reduction) is the most remarkable in Switzerland (54%) and the least remarkable in India (11.4%). The lower improvement on the global average can be explained by the countries whose reproduction number did not peak since the start of the pandemic, and thus was more predictable by the simple baseline model. On the other hand, the prediction error reduced drastically on the selected countries that are known to have faced high R values with a clear, transient peak (Switzerland) versus those with sustained transmission (India).

Below are the prediction plots for those 6 countries (see Fig. 4.19), as well as the 6 countries for which the model obtained the best predictions (see Fig. 4.5.4 and 4.20).



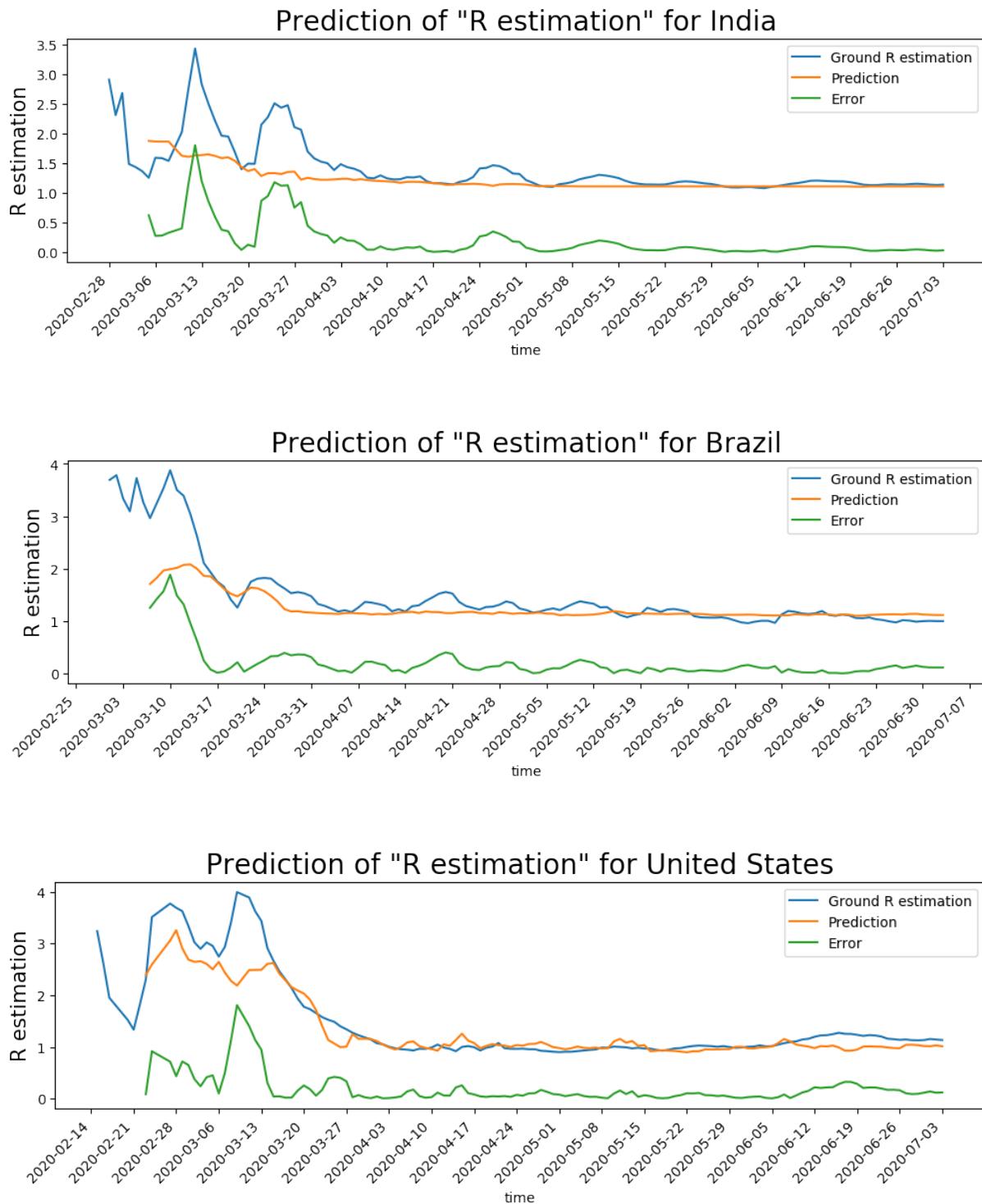


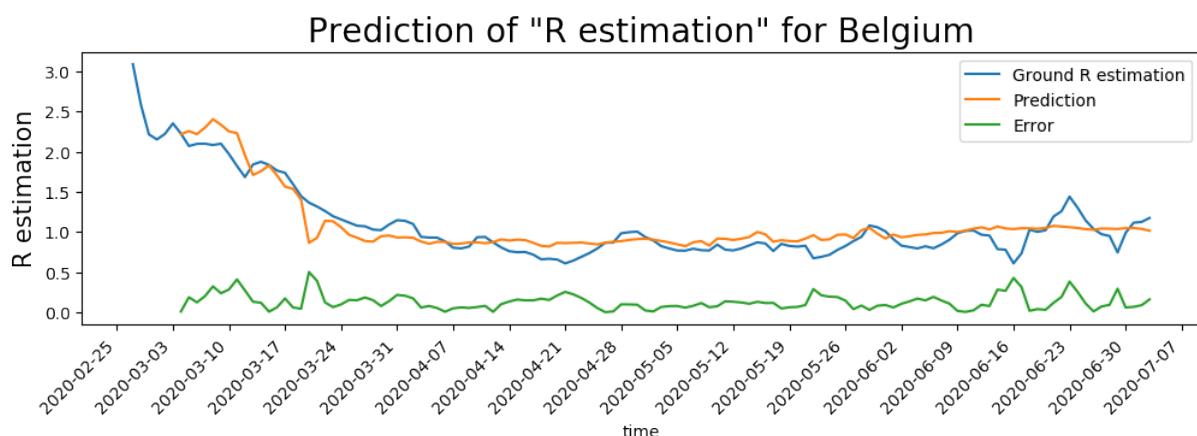
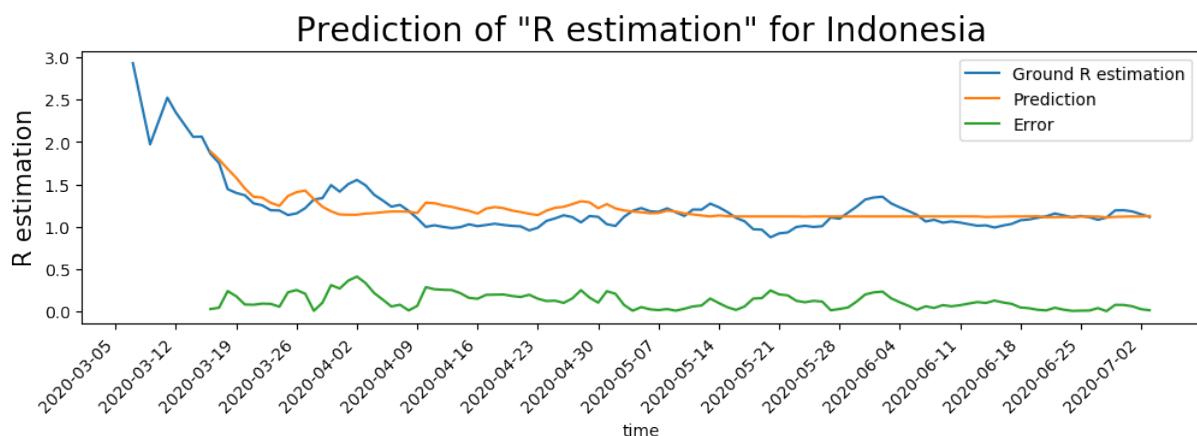
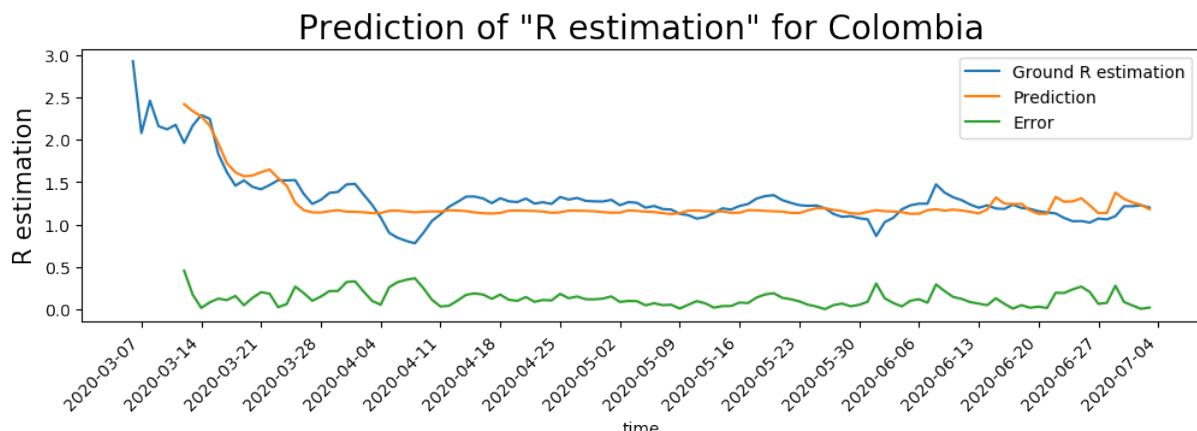
Figure 4.19: Reproduction number predictions using optimized model

We observe from the above graphs that the high values at the beginning of the epidemic are captured quite well by the model, whereas the later variations (e.g. *second waves*) are not as precisely predicted. However, it seems that the model is able to capture at least some late variations, as is the case for Switzerland (Fig. 4.5.4).

Below are the top 6 countries in which the model performed best. The predictions of countries that are the most accurate in terms of error could be due to the low values and the lack of variations in the reproduction number. It is also important to remember that random factors may be responsible for variations in the reproduction number. By random factor, we mean events that cannot be captured by the collected data. For example, a specific event (party, sport event) that would hardly be observable from our data, could be the starting point of a transmission cluster that may produce a rise in the R curve.

Top-6 predictions:

Zone	Mean Absolute Error
Colombia	0.126
Indonesia	0.129
Belgium	0.136
Dominican Republic	0.138
Mexico	0.138
Poland	0.138



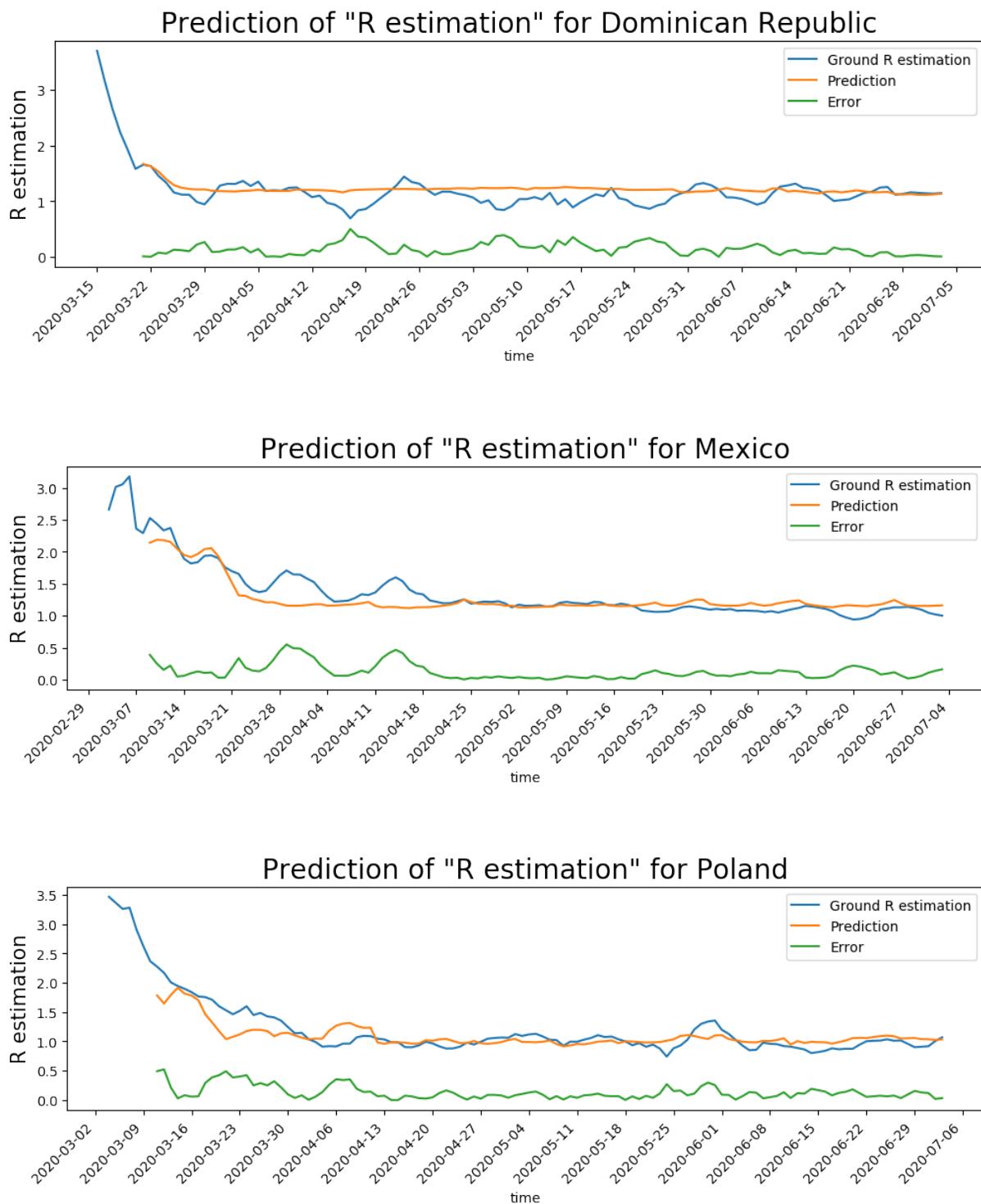


Figure 4.20: Reproduction number predictions using optimized model

4.5.5 Model performance in the face of missing data

One of the applications of our model, can be to compensate for the lack of data. To illustrate this, we chose Tanzania for which the reproduction number prediction is available for 18 days only. Since May 8th, no COVID-19 case or death has been reported, which is largely regarded as a politically motivated deficiency in reporting/testing [9]. In comparison, the reproduction number is available for 121 different days in South Africa. We display below, the epidemic summary plot in Tanzania (Fig. 4.21) as well as the predicted value of the reproduction number (Fig. 4.22).

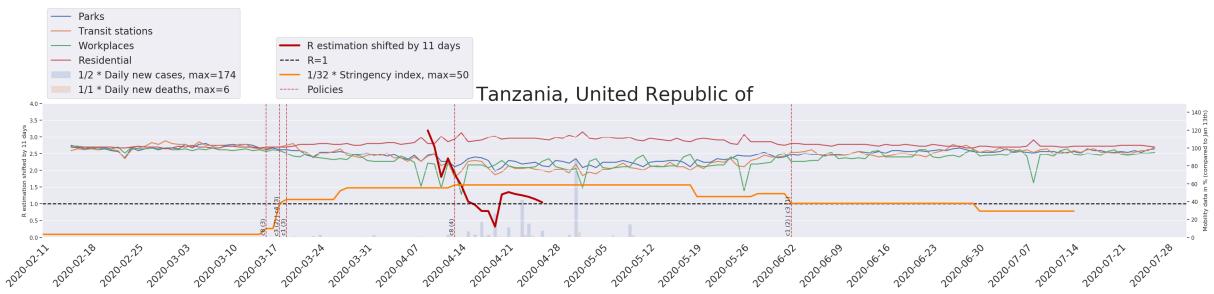


Figure 4.21: Visualization of the COVID-19 outbreak in Tanzania

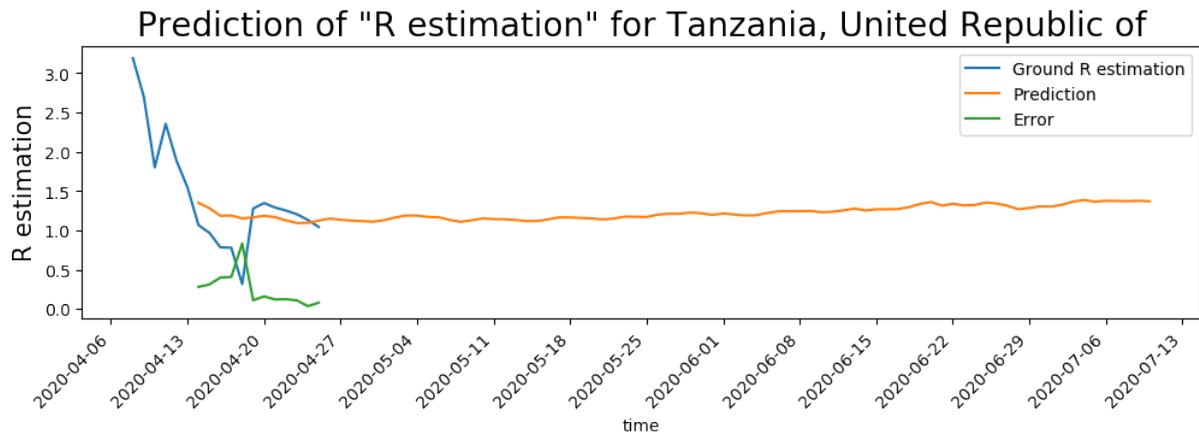


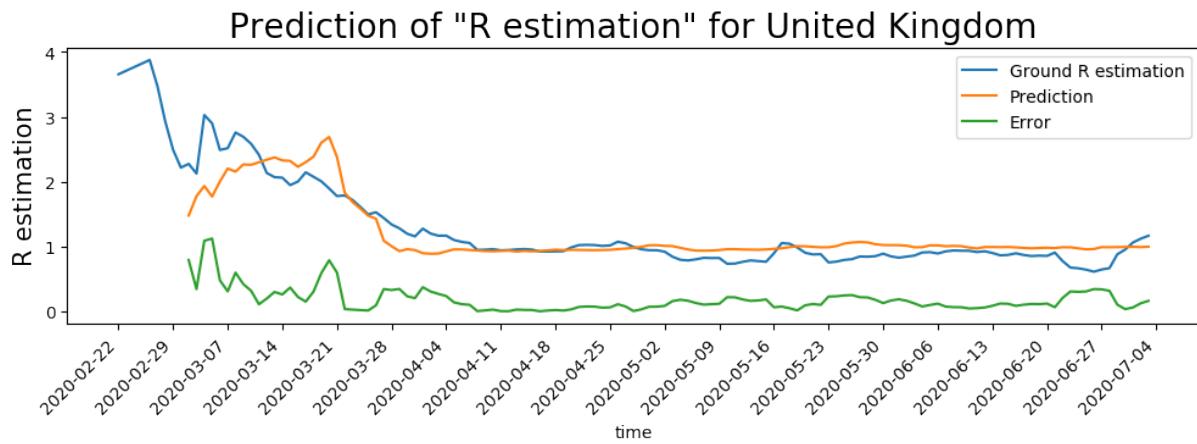
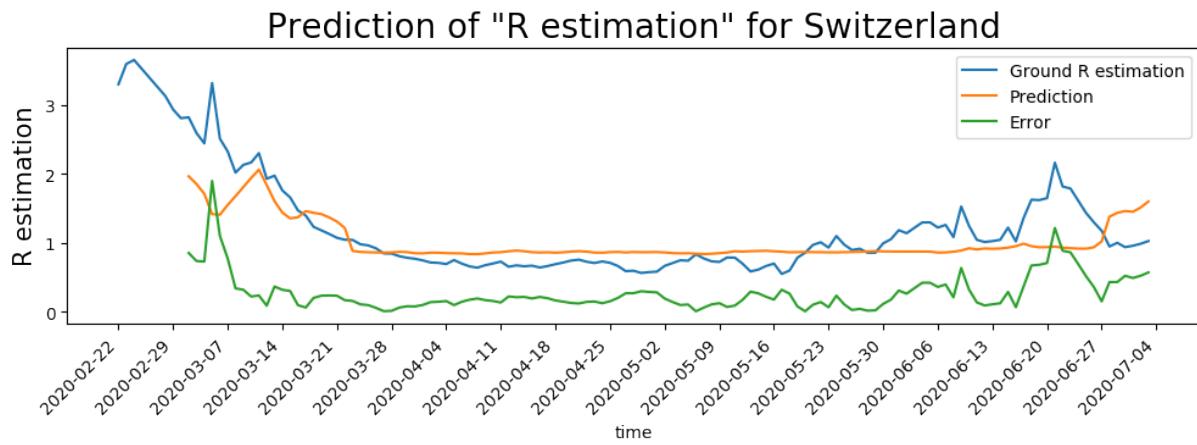
Figure 4.22: Prediction of reproduction number against lack of data

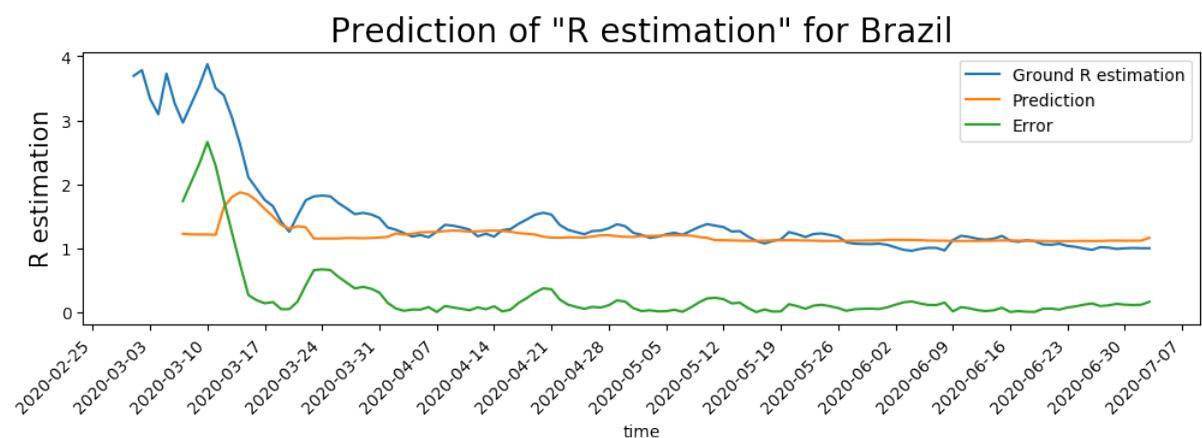
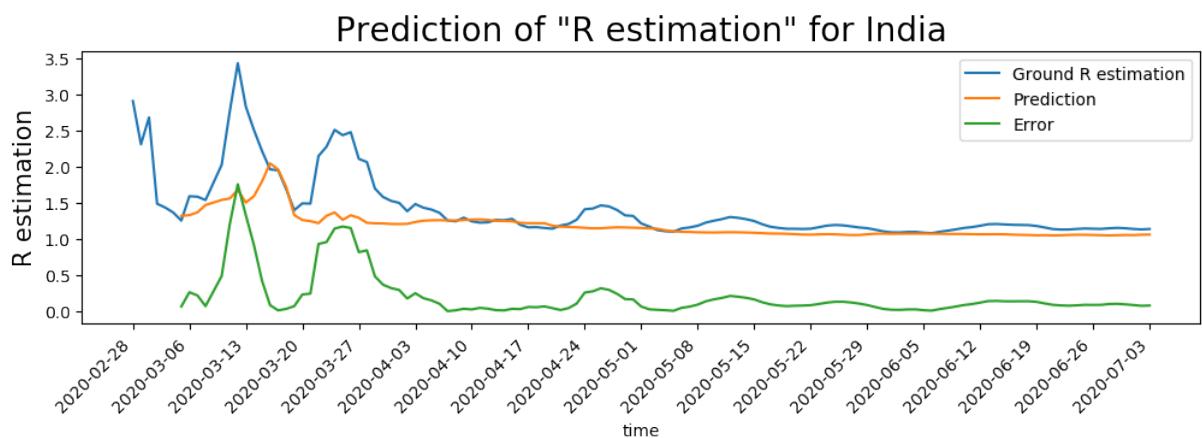
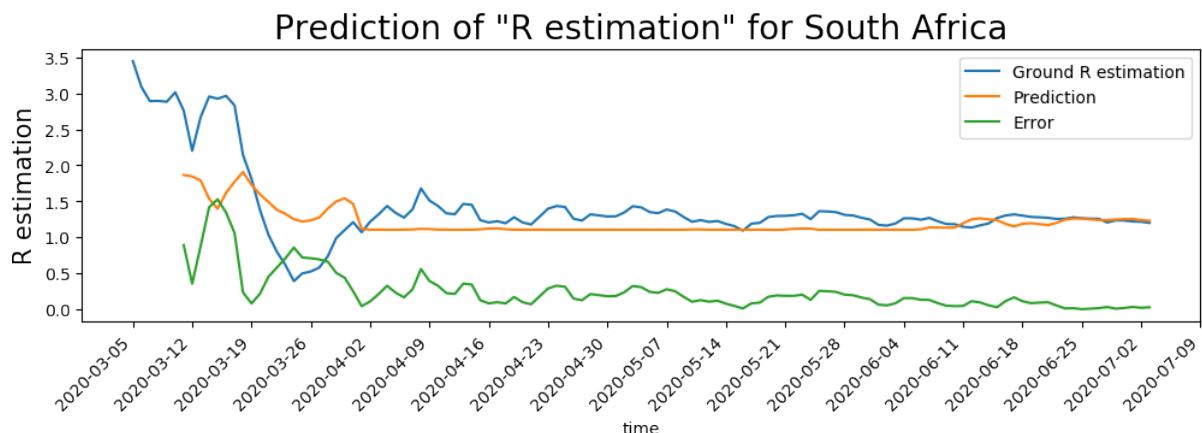
The prediction is slowly increasing in a range between 1 and 1.5 since the last available estimation. It means that unless the epidemic has been totally controlled or stopped — which seems unlikely — the number of real cases is increasing exponentially since the day of the last reported case.

4.5.6 Reproduction number prediction from policies

We show in this section the performance of the model using policies instead of mobility features as inputs. The complete list can be found in Appendix A.7 and the performance and prediction plots concerning the 6 countries that we use as example are shown below:

Zone	Mean Absolute Error
World (average of 115 countries)	0.306
Switzerland	0.232
United Kingdom	0.192
South Africa	0.287
India	0.188
Brazil	0.252
United States	0.391





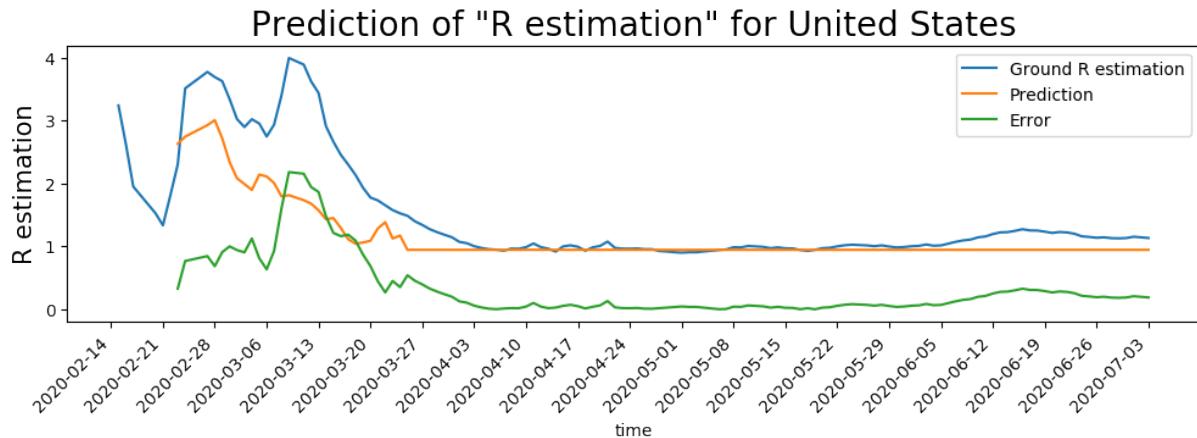


Figure 4.23: Reproduction number predictions from policies using optimized model

As expected, the prediction is not as accurate as the one that takes advantage of mobility data. However, knowing that no time-related feature is part of the input, the model seems to capture enough information from the policies to predict trends even though they are not precise in time. In other words, it poorly detects variations in reproduction number but gives a quite precise prediction when the curve is flat. For example, the flat area in the prediction for India (May-July 2020) lies around 1.2 as the flat area in Switzerland (April 2020) is estimated at about 0.8. In both cases, the prediction is really close to the ground estimation and clearly makes the difference between both situations. It is a peculiarity that we are going to exploit in the next section to evaluate the impact of the policies on the reproduction number.

4.5.7 Policy effect

We display below the evolution of the impact of each *containment and closure* policy in Switzerland. The plots consist of 4 different curves. In blue and orange, as for prediction plots, are the true R estimate and the prediction, respectively. In green is shown the hypothetical prediction of the R value if the said policy would not have been applied. Finally, we show the impact in red which corresponds to the difference between the prediction of both situations. The time grid placed under the plots show the evolving level of application of the policies.

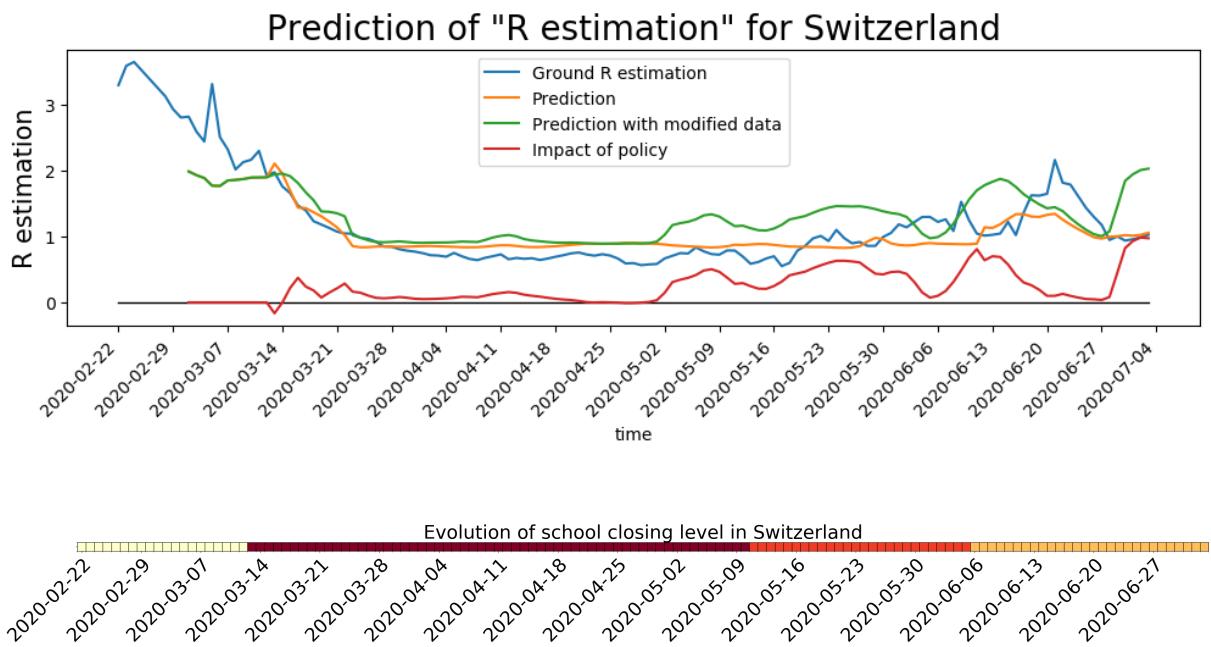


Figure 4.24: **School closing** impact over time in Switzerland

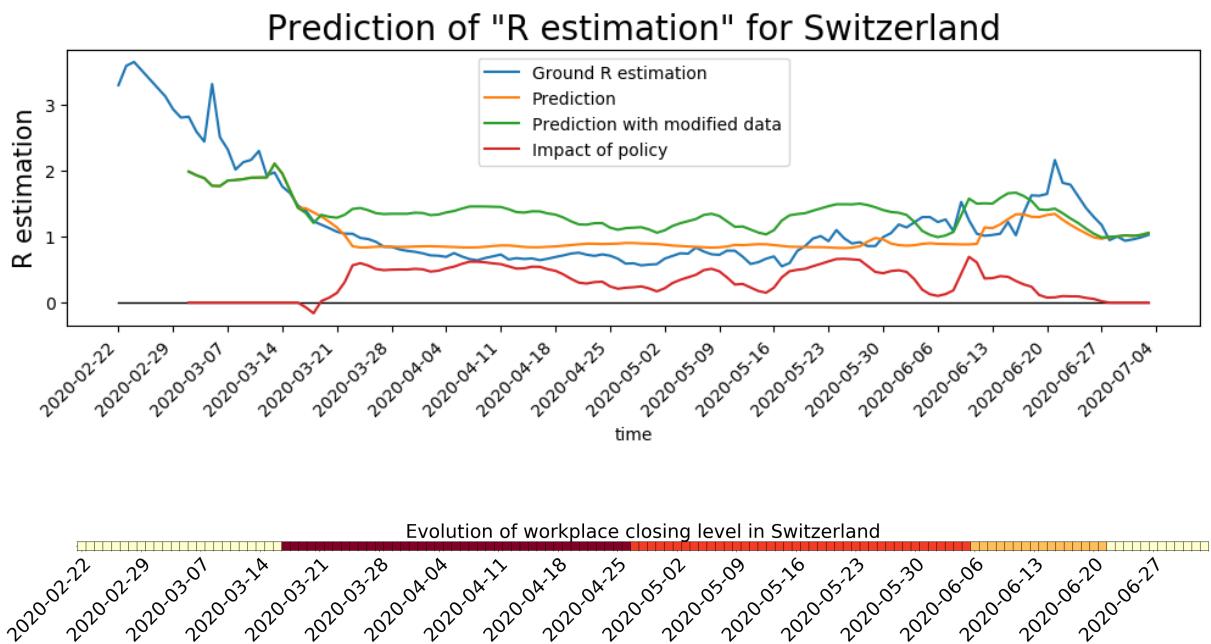


Figure 4.25: **Workspace closing** impact over time in Switzerland

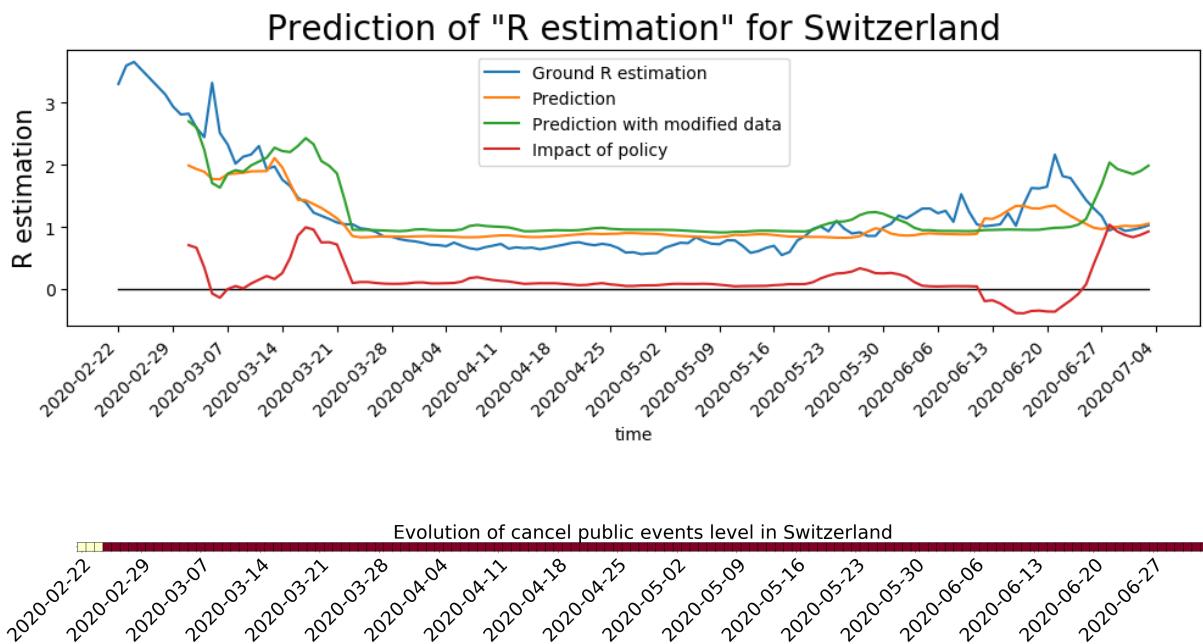


Figure 4.26: **Public events canceling** impact over time in Switzerland

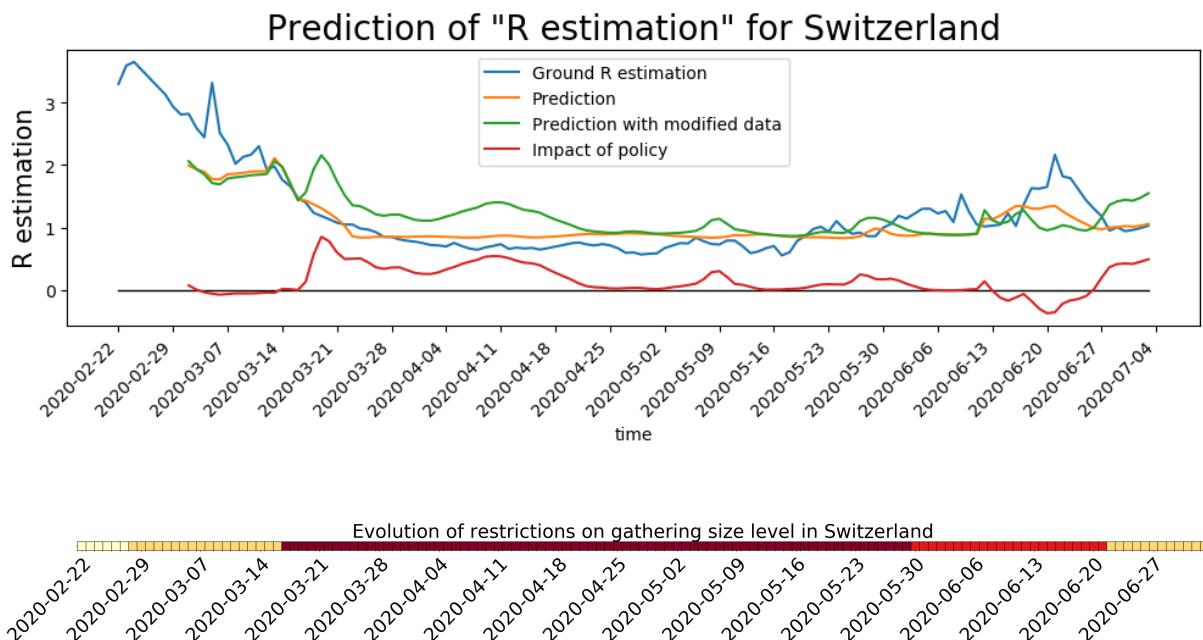


Figure 4.27: **Gathering size restrictions** impact over time in Switzerland

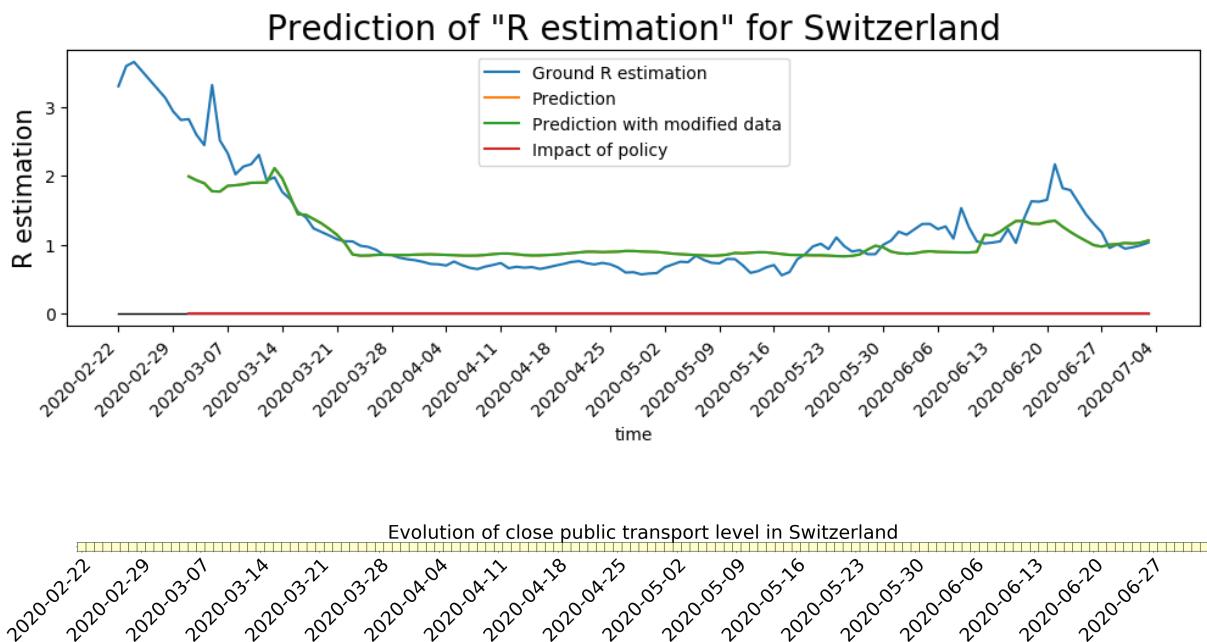


Figure 4.28: **Public transports closure** impact over time in Switzerland

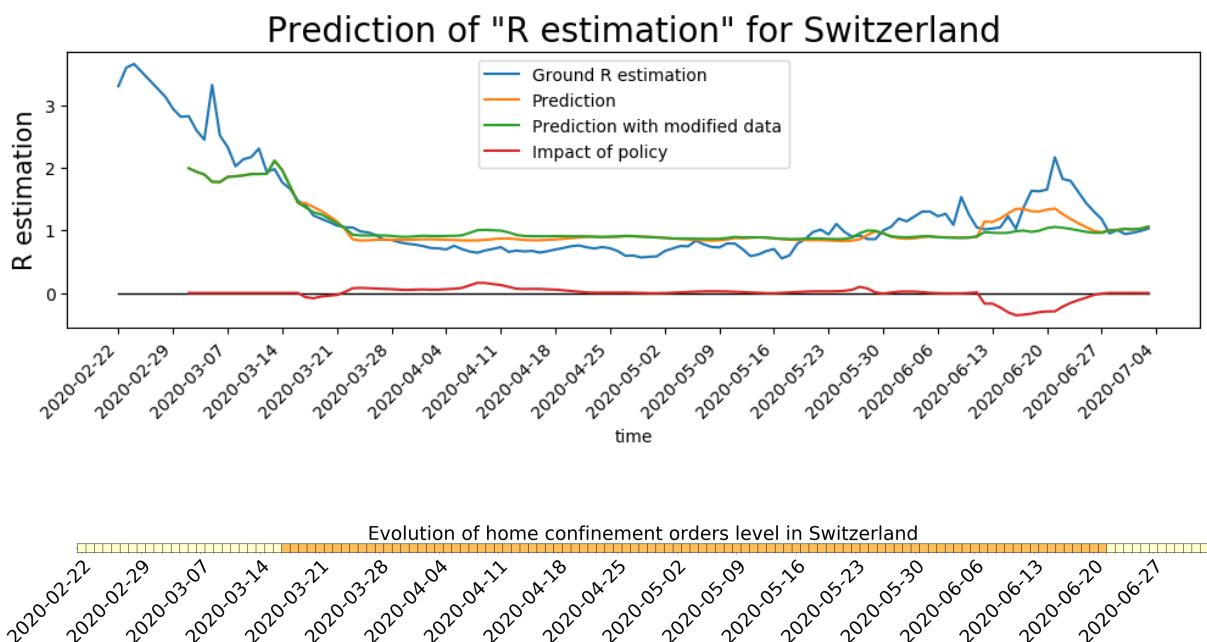
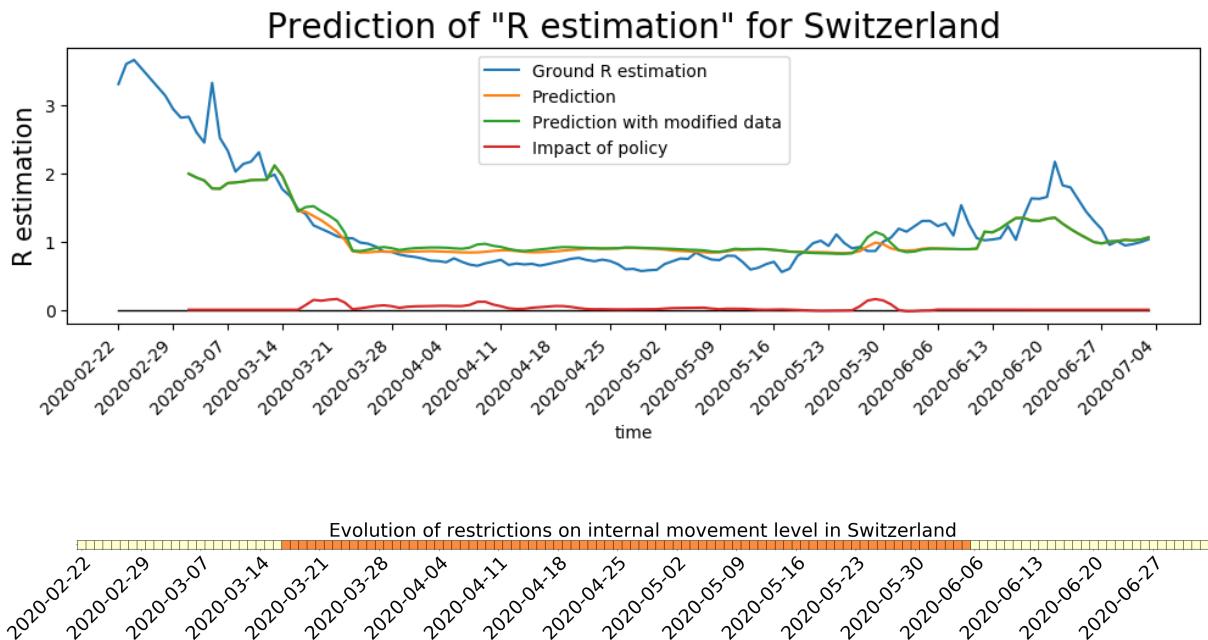
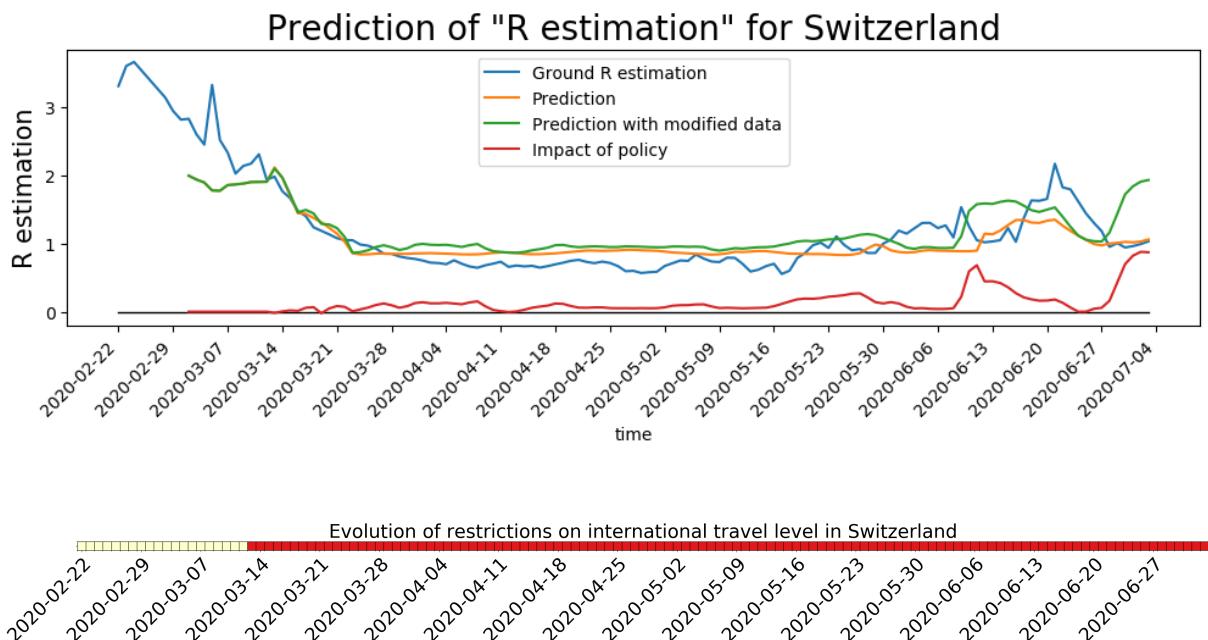


Figure 4.29: **Home confinement order** impact over time in Switzerland

Figure 4.30: **Internal movement restrictions** impact over time in SwitzerlandFigure 4.31: **International travel restrictions** impact over time in Switzerland

These representations of the impact of each policy over time are meaningful and we will try to interpret the most notable ones. Obviously, there is no impact of *public transports closure* because the measure has not been applied in Switzerland. A policy whose impact has an

interesting shape is the *events canceling*. Its impact is maximal at the beginning and the end of the epidemic although it was applied at the same level since the start of the epidemic. This can be explained by the other policies that have been applied in between. Indeed, if this policy is the only one in operation, we can predict greater effects than if other policies — and particularly the more stringent ones — are in application at the same time. For example and more concretely, if a population is required to stay at home, canceling public events will not change daily habits since people stay at home anyway. This demonstrates the ability of our neural network to capture inter-dependencies between policies.

The last point we highlight is the almost insignificant impact of the home confinement order in Switzerland. While it was applied at the lowest level, we may want to compare the impact of this particular policy with other countries. To this end, we display below the impact plots of confinement orders in Brazil and South Africa.

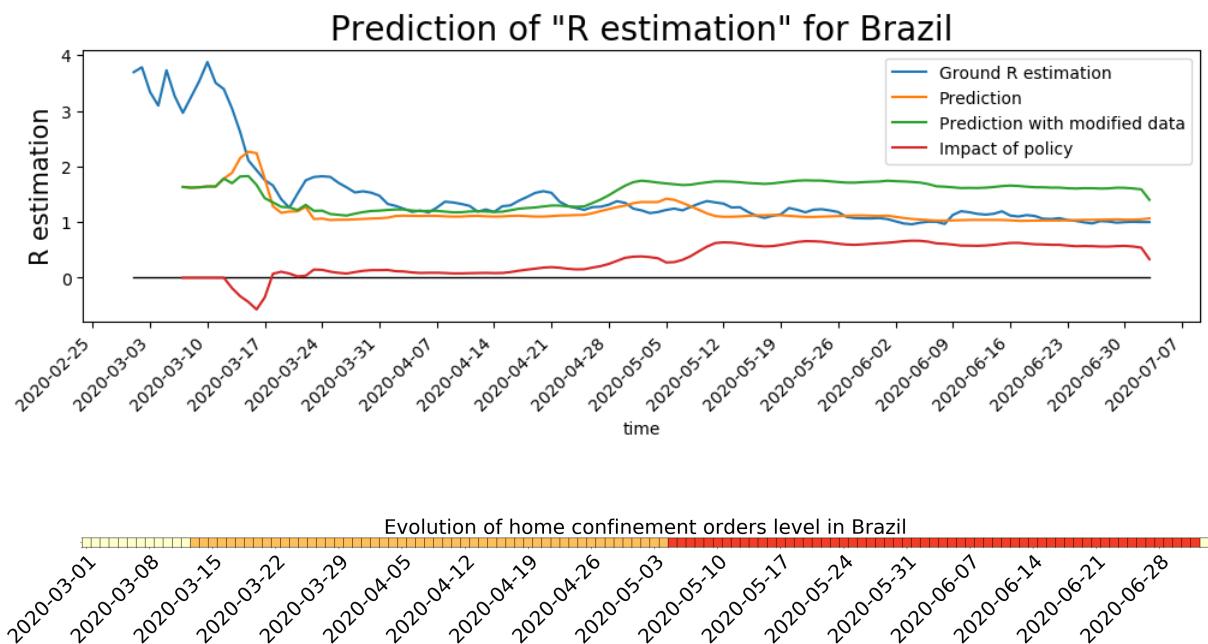


Figure 4.32: **Home confinement order** impact over time in Brazil

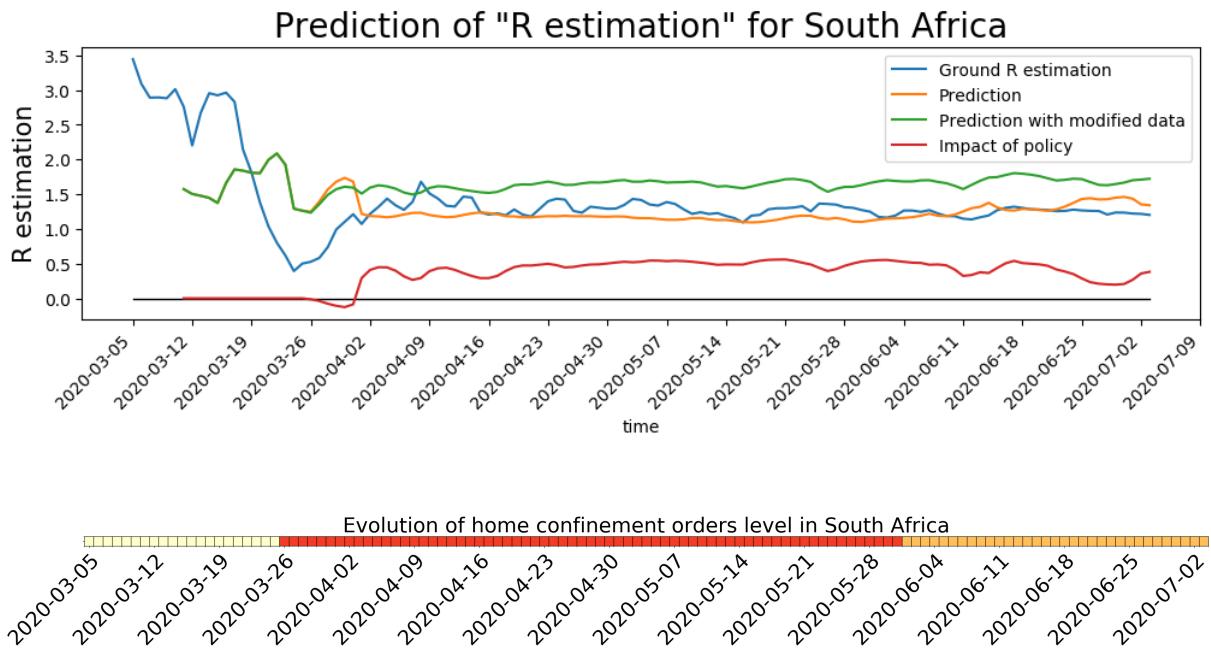


Figure 4.33: **Home confinement order** impact over time in South Africa

In these countries, the impact of the home confinement order is higher than in Switzerland. Although the larger impact seems to be associated with a higher level of stringency, the impact starts to rise before the end of the first level of implementation in Brazil (we note that the model cannot make use of future inputs for a prediction at a certain day). This is a second illustration of the multivariate aspect of this result.

As a complete overview, we provide in the table below the mean impact of each policy in our 6 key countries. Policies with a positive impact are depicted by a blue heatmap overlay, while negative impact policies are shown similarly in red. In each cell, the mean impact of the policy as well as its most stringent level of application are displayed.

	Switzer-land	United Kingdom	South Africa	India	Brazil	United States
school closing	0.24 Requ.	0.22 Requ.	0.22 Requ.	0.05 Requ.	0.5 Requ.	0.25 Requ.
workplace closing	0.31 Requ.	0.16 Part. requ.	-0.03 Requ.	-0.03 Requ.	-0.08 Requ.	0.14 Requ.
cancel public events	0.17 Requ.	0.11 Requ.	-0.1 Requ.	-0.18 Requ.	0.12 Requ.	0.04 Requ.
gathering size re-strict.	0.15 <10	-0.07 <10	-0.13 11-100	-0.08 <10	-0.09 <10	-0.0 <10
close public transport	0.0 No meas-ure	0.05 Recomm.	0.01 Recomm.	-0.03 Requ.	0.03 Requ.	-0.02 Recomm.
home confinment orders	-0.01 Recomm.	0.03 Requ. with exceptions	0.36 Requ. with exceptions	-0.01 Requ. (strict)	0.33 Requ. with exceptions	0.06 Requ. with exceptions
internal mov. re-strict.	0.02 Recomm.	0.02 Requ.	0.16 Requ.	-0.18 Requ.	0.19 Requ.	0.07 Requ.
internat. travel restrict.	0.13 Total	-0.0 Quarantine	-0.08 Requ.	-0.04 Requ.	-0.08 Requ.	0.01 Total

Figure 4.34: Table of policies impact in 6 countries (requ. = requirement)

We observe that in India, policies in general did not have a positive association with epidemic control (R-value reduction), as opposed to Switzerland where most of the policies could be associated with positive impact. Indeed, the capacity and willingness of states and populations to follow implemented policies may vary according to numerous and complex socio-economic

and political forces. From a policy point of view, the gathering size restrictions are predicted to have had an impact in Switzerland only, while in South Africa and Brazil, home confinement orders were most valuable.

Chapter 5

Discussion

The results obtained from the hybrid model combining an LSTM layer and a multilayer perceptron gave encouraging results as obtained by the *leave-one-out* validation process. In contrast to other studies presented (see Section 2.1.3), it takes into account inter-dependencies between policies as well as reducing the time and locality bias thanks to the scope of the data. A certain error is theoretically unavoidable because some of the factors that affect the trend of the epidemic cannot be captured by our data. They are therefore considered as noise in our study. To formalize this, we show below an augmented subpart of our causality graph from the beginning (Fig. 3.1):

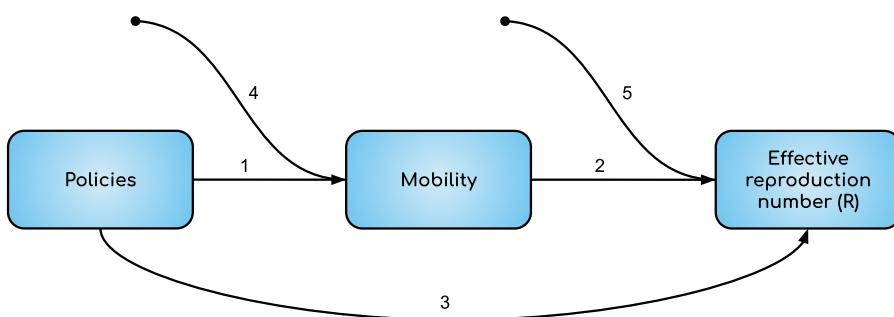


Figure 5.1: Causality subgraph with focus on modeling parts

The factors represented by the arrow no. 5 are independent of our data. We will name those factors *the untracked behavior*, that includes physical distancing habits (handshakes), wearing of masks or global awareness among others. Those factors are very difficult to evaluate and they are becoming more and more important as the epidemic evolves. Another part of the prediction error comes from the reproduction number estimation. We can reasonably assume that in some countries the model prediction is closer to the real reproduction number than the ground estimation, as the reported data may be of bad quality (see Subsection 4.5.5). By

creating our alternative model based on policies data, we lost again some information that is represented by the link no. 4. It mainly consists of the mobility changes that are not affected by government measures as well as the precision in mobility reports that is not achievable from policy data. However, this alternative model can theoretically take advantage of information that is not available in the other model. The link no. 5 represents the effects of the policies that are not visible through mobility. We can think about the restrictions on gathering size, whose part of the effects may be independent of mobility. Based on the model results, we conclude that information brought through link no. 4 has more influence on the reproduction number than information brought by link no. 3.

Using our alternative model based on policies, we put aside the precision in prediction and rather try to take the maximum advantage of the captured information coming from policies and affecting the reproduction number (links no. 1 and 2). Even though the information coming through link 4. may affect the prediction precision, it is not supposed to affect our estimation of the policies impact as it is, by definition, independent of the policies. This part of the results is difficult to assess in terms of precision because the alternative model is mainly based on the performances of the first model and the assumptions on which we based ourselves to switch from one to another. However, the obtained impact measures are intuitively consistent and their evolution over time are interpretable from our knowledge on epidemics.

5.1 Limitations

The first limitations we faced concern the estimation of the reproduction number. As we have seen in our background chapter (see Section 2.2.2), the R_0 (value of the basic reproduction number before the start of the epidemic) is very difficult to evaluate. As it is dependent on many parameters that vary from country to country, this measure would have been determinant to see *how far* each country was from controlling the epidemic ($R_0 \leq 1$) before the application of public health measures. Our work is thus based on the estimations that begin in each country when the epidemic has already started and for some of them the reproduction number has already been reduced. A second limitation related to our reproduction number is the delay from infection to case. We considered an average value based on the Swiss data coming from FOPH (see Fig. 4.1). However, this value may vary across countries and thus resulting in a shifted estimation of the reproduction number. The variance in delays could also be higher in some countries compared to Switzerland, which would produce a less precise estimation. Briefly, the value that is considered by our models as the *ground truth* is already subject to inaccuracies, which limits the performance of the prediction even before starting with the modeling part.

Another source of uncertainty comes from our geographical decomposition by countries. If most of the data features are well represented by this decomposition, the policies may be ill-defined in some large countries such as the United States. Indeed, part of the political response is managed by smaller authorities (states, counties) and are thus not captured by our data from the

Oxford COVID-19 Government Response Tracker (OxCGRT) (see Subsection 2.1.1). It implies that the country-level collected policies do not correspond, on average, to the reality observed by the mobility data. This can affect the results obtained by the alternative model used to measure the impact of policies.

A last limitation that is important to emphasize is the use of the mobility reports provided by *Google Inc.* in our prediction model. While they have released a statement, that they plan to secure public access “so long as public health officials find them useful in their work to stop the spread of COVID-19” [10], the duration of this is uncertain.

Chapter 6

Future work

In this chapter, we provide a list of tasks that could further improve the accuracy, impact and interpretability of this work.

A first scope of improvement would be to improve the model's sensitivity to inherent differences between countries that may confound the behavior of individuals or motivate political willingness to apply the implemented policy. This could be achieved by including new features in the model that represent issues such as *democracy index*, *human development index* or *freedom of press*. Another data-related improvement could be made on weather metrics. Currently, data of the capitals are taken as representative of each country, but a weighted average over all regions of a country would be a more representative projection.

To obtain more information about individual human behavior, social inputs can provide insights that are not captured by mobility data (link no. 5. of the causality subgraph, Fig. 5.1). To this end, we have began exploring the utility of Twitter data (obtained through their API [16]) as a proxy of population level reactions to policies which may differ between various cultural settings and sub-populations. Work is underway focusing on inferring sentiments and topics that may align to deviations from expected mobility trends (e.g. can fear, confusion, boredom correlate to how individual apply the given policies?). Further work on this topic could bring valuable information to our modeling.

As an already planned expansion of the project, we aim to create an interactive web interface, that will allow the public and policy makers to “play” with the trained model: creating hypothetical scenarios of policy/epidemic conditions to answer the question of "what if...". The purpose of this application is to enable interpretable visualizations of policy effectiveness and to help individuals explore the effects of a customized virtual implementation of hypothetical policies.

A new dimension of *cost* may also be added to take into consideration that the implementation of public health measures is not only a matter of reproduction number, but also as a cost for a population (welfare, freedom, economy, ...). Death attributable to the pandemic is

not only due to the infectious agent, but also to indirect effects in societal functioning and can accumulate to a 2.3-fold increase in all-cause mortality [17]. Indeed in the Ebola epidemic, these secondary deaths due to lack of routine care (10,600) almost equaled the lives lost due to Ebola itself (11,300) [3]. Thus, it is important to optimize the trade-off between deaths incurred by COVID-19 to those incurred by the policies to mitigate it.

Chapter 7

Conclusion

In this study, we propose a novel method for assessing the dynamic drivers of effective policy during the COVID-19 pandemic. We explore the impact of a range of heterogeneous parameters from demographics to meteorological information, and describe an innovative way to incorporate and analyze the effect of these changeable inputs on the evolution of epidemics based on Neural Networks. The model predicts an estimation of the reproduction number which is derived from a simple count of infectious individuals over time and is thus an interpretable metric of the efficacy of mitigation strategies.

The approach has many applications such as overcoming unreliable case reporting or detecting inconsistencies in epidemic trends. This is possible thanks to the inclusion of inputs that are sometimes more robust than the target.

Finally, the large geographic scope of our input could also allow us to better assess causality through drawing analogies and finding coherence between populations. Thanks to complex inner structure of Neural Networks, the inter-dependencies between policies and other various features was taken into consideration in the results.

Understanding these complex interactions may allow individuals and policy makers to better adapt the mitigation strategies that optimize the efficacy of the implemented policies.

Bibliography

- [1] Qun Li et al. “Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia”. In: *The New England Journal of Medicine*, 2020; 382:1199-1207 (Mar. 26, 2020). DOI: 10.1056/NEJMoa2001316.
- [2] Francisco Arroyo-Marioli, Francisco Bullano, Simas Kučinskas†, and and Carlos Rondón-Moreno. “Tracking R of COVID-19: A New Real-Time Estimation Using the Kalman Filter”. In: (May 14, 2020). DOI: 10.1101/2020.04.19.20071886.
- [3] Parpia AS, Ndeffo-Mbah ML, Wenzel NS, and Galvani AP. “Effects of Response to 2014-2015 Ebola Outbreak on Deaths from Malaria, HIV/AIDS, and Tuberculosis, West Africa”. In: *Emerg Infect Dis*. 2016;22(3):433-41 (Mar. 2016). DOI: 10.3201/eid2203.150977.
- [4] Nicolas Banholzer, Eva van Weenen, Bernhard Kratzwald, Arne Seeliger, Daniel Tschernutter, Pierluigi Bottrighi, Alberto Cenedese, Joan Puig Salles, Werner Vach, and Stefan Feuerriegel. “Impact of non-pharmaceutical interventions on documented cases of COVID-19”. In: (Apr. 28, 2020). DOI: 10.1101/2020.04.16.20062141.
- [5] Diana Beltekian, Daniel Gavrilov, Charlie Giattino, Joe Hasell, Bobbie Macdonald, Edouard Mathieu, Esteban Ortiz-Ospina, Hannah Ritchie, and Max Roser. *Data on COVID-19 (coronavirus) by Our World in Data*. July 2020. URL: <http://covid.ourworldindata.org>.
- [6] Lemaitre Joseph. C., Perez-Saez Javier, Azman Andrew S., Rinaldo Andrea, and Fellay Jacques. “Assessing the impact of non-pharmaceutical interventions on SARS-CoV-2 transmission in Switzerland”. In: *Swiss Med Wkly*. 2020;150:w20295 (May 3, 2020). DOI: 10.4414/smw.2020.20295.
- [7] Michael J. Coren. *The number of scientific papers on the novel coronavirus is doubling every 14 days*. Apr. 8, 2020. URL: <http://qz.com/1834521/scientific-papers-on-novel-coronavirus-doubling-every-14-days>.
- [8] Benjamin J Cowling, Sheikh Taslim Ali, Tiffany W Y Ng, and Tim K Tsang. “Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study”. In: *Lancet Public Health*. 2020;5(5):e279–88 (Apr. 17, 2020). DOI: 10.1016/S2468-2667(20)30090-6.
- [9] The Economist. *Why Tanzania's statistics look fishy*. July 23, 2020. URL: economist.com/middle-east-and-africa/2020/07/23/why-tanzanias-statistics-look-fishy.

- [10] Jen Fitzpatrick and Karen DeSalvo. *Helping public health officials combat COVID-19*. URL: <http://blog.google/technology/health/covid-19-community-mobility-reports?hl=en>.
- [11] Christophe Fraser. "Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic". In: *PLoS ONE* 2(8): e758 (Aug. 22, 2007). DOI: 10.1371/journal.pone.0000758.
- [12] Andres Hernandez, Esteban Correa-Agudelo, Hana Kim, Adam J Branscum, F DeWolfe Miller, Neil MacKinnon, and Diego F Cuadros. "On the impact of early non-pharmaceutical interventions as containment strategies against the COVID-19 pandemic". In: (May 9, 2020). DOI: 10.1101/2020.05.05.20092304.
- [13] Thomas Hotz, Matthias Glock, Stefan Heyder, Sebastian Semper, Anne Böhle, and Alexander Krämer. "Monitoring the spread of COVID-19 by estimating reproduction numbers over time". In: (Apr. 18, 2020). URL: <http://stochastik-tu-ilmenau.github.io/COVID-19/>.
- [14] Apple Inc. *Mobility Trends Reports*. July 2020. URL: <http://apple.com/covid19/mobility>.
- [15] Google Inc. *Community Mobility Reports*. July 2020. URL: <http://google.com/covid19/mobility>.
- [16] Twitter Inc. *Filter realtime Tweets*. URL: <http://developer.twitter.com/en/docs/tweets/filter-realtime/overview>.
- [17] Simonsen L, Spreeuwenberg P, Lustig R, Taylor RJ, Fleming DM, and Kroneman M et al. "Global mortality estimates for the 2009 Influenza Pandemic from the GLaMOR project: a modeling study". In: *PLoS Med.* 2013;10(11):e1001558 (Nov. 26, 2013). DOI: 10.1371/journal.pmed.1001558.
- [18] Kevin D. Lafferty. "The ecology of climate change and infectious diseases". In: *Ecological Society of America* (Apr. 1, 2009). DOI: 10.1890/08-0079.1.
- [19] Marc Lipsitch, Ted Cohen, Ben Cooper, James M Robins, Stefan Ma, Lyn James, Gowri Gopalakrishna, Suok Kai Chew, Chorh Chuan Tan, Matthew H Samore, David Fisman, and Megan Murray. "Transmission dynamics and control of severe acute respiratory syndrome". In: *Science* (June 20, 2003). DOI: 10.1126/science.1086616.
- [20] Ying Liu, Albert A. Gayle, Annelies Wilder-Smith, and Joacim Rocklöv. "The reproductive number of COVID-19 is higher compared to SARS coronavirus". In: *Journal of Travel Medicine* (Feb. 13, 2020). DOI: 10.1093/jtm/taaa021.
- [21] Joanna Nowakowskaa, Joanna Sobocińska, Mateusz Lewicki, Żaneta Lemańska, and Piotr Rzymski. "When science goes viral: The research response during three months of the COVID-19 outbreak". In: *Biomedicine & Pharmacotherapy, Volume 129, September 2020, 110451* (June 21, 2020). DOI: 10.1016/j.bioph.2020.110451.
- [22] Christopher Olah. *Understanding LSTM Networks*. Aug. 27, 2015. URL: <http://colah.github.io/posts/2015-08-Understanding-LSTMs>.

- [23] World Health Organization. *Coronavirus disease 2019 (COVID-19) Situation Report 73*. Apr. 2, 2020. URL: <http://who.int/docs/default-source/coronavirus/situation-reports/20200402-sitrep-73-covid-19.pdf>.
- [24] World Health Organization. *Virtual press conference on COVID-19–11 March 2020*. Mar. 11, 2020. URL: <http://who.int/docs/default-source/coronavirus/transcripts/who-audio-emergencies-coronavirus-press-conference-full-and-final-11mar2020.pdf>.
- [25] World Health Organization. *WHO Timeline - COVID-19*. Apr. 27, 2020. URL: <http://who.int/news-room/detail/27-04-2020-who-timeline---covid-19>.
- [26] Anna Petherick, Beatriz Kira, Noam Angrist, Thomas Hale, Toby Phillips, Samuel Webster, et al. *Coronavirus Government Response Tracker*. 2020. URL: <http://bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker>.
- [27] Anna Petherick, Beatriz Kira, Noam Angrist, Thomas Hale, Toby Phillips, Samuel Webster, et al. *Oxford Covid-19 Government Response Tracker (OxCGRT)*. 2020. URL: <https://github.com/OxCGRT/covid-policy-tracker/>.
- [28] Anna Petherick, Beatriz Kira, Noam Angrist, Thomas Hale, Toby Phillips, and Samuel Webster. “Variation in government responses to COVID-19”. In: *BSG Working Paper Series, BSG-WP-2020/032* (May 2020).
- [29] Alasdair Sandford. *Coronavirus: Half of humanity now on lockdown as 90 countries call for confinement*. Apr. 3, 2020. URL: <http://euronews.com/2020/04/02/coronavirus-in-europe-spain-s-death-toll-hits-10-000-after-record-950-new-deaths-in-24-hou>.
- [30] Giulio Viceconte and Nicola Petrosillo. “COVID-19 R₀: Magic number or conundrum?” In: *PAGEPress* (Feb. 24, 2020). DOI: 10.4081/idr.2020.8516.
- [31] James Vincent. *France is using AI to check whether people are wearing masks on public transport*. May 7, 2020. URL: <http://theverge.com/2020/5/7/21250357/france-masks-public-transport-mandatory-ai-surveillance-camera-software>.
- [32] *World Weather Online API*. WorldWeatherOnline. URL: <http://worldweatheronline.com/developer/api>.

Appendix A

Modeling

A.1 Considered countries in cross correlations

(120 countries)

- Afghanistan
- Brazil
- Ecuador
- India
- Angola
- Bulgaria
- Egypt
- Indonesia
- Argentina
- Burkina Faso
- El Salvador
- Iraq
- Aruba
- Cambodia
- Estonia
- Ireland
- Australia
- Cameroon
- Finland
- Israel
- Austria
- Canada
- France
- Italy
- Bahrain
- Cabo Verde
- Gabon
- Jamaica
- Bangladesh
- Chile
- Georgia
- Japan
- Belarus
- Colombia
- Germany
- Jordan
- Belgium
- Costa Rica
- Ghana
- Kazakhstan
- Benin
- Côte d'Ivoire
- Greece
- Kenya
- Bolivia, Plurinational State of
- Croatia
- Guatemala
- Kuwait
- Bosnia and Herzegovina
- Czechia
- Haiti
- Kyrgyzstan
- Botswana
- Denmark
- Honduras
- Latvia
- Dominican Republic
- Hungary
- Lebanon
- Libya

- Lithuania
- Luxembourg
- Malaysia
- Mali
- Mauritius
- Mexico
- Moldova, Republic of
- Mongolia
- Morocco
- Mozambique
- Myanmar
- Nepal
- Netherlands
- New Zealand
- Nicaragua
- Niger
- Nigeria
- Norway
- Oman
- Pakistan
- Panama
- Paraguay
- Peru
- Philippines
- Poland
- Portugal
- Puerto Rico
- Qatar
- Romania
- Russian Federation
- Rwanda
- Saudi Arabia
- Senegal
- Serbia
- Singapore
- Slovakia
- Slovenia
- South Africa
- Korea, Republic of
- Spain
- Sri Lanka
- Sweden
- Switzerland
- Tajikistan
- Tanzania, United Republic of
- Thailand
- Togo
- Trinidad and Tobago
- Turkey
- Uganda
- Ukraine
- United Arab Emirates
- United Kingdom
- United States
- Uruguay
- Venezuela, Bolivarian Republic of
- Viet Nam
- Yemen
- Zambia
- Zimbabwe

A.2 Considered countries in feature selection

(93 countries)

- Afghanistan
- Angola
- Argentina
- Australia
- Austria
- Bahrain
- Bangladesh
- Belarus
- Belgium
- Bolivia, Plurinational State of
- Bosnia and Herzegovina
- Brazil
- Bulgaria
- Cameroon
- Canada
- Cabo Verde
- Chile
- Colombia
- Costa Rica
- Côte d'Ivoire
- Czechia
- Denmark
- Dominican Republic
- Ecuador
- Egypt

- El Salvador
- Japan
- Norway
- Korea, Republic of
- Estonia
- Kazakhstan
- Oman
- Spain
- Finland
- Kenya
- Pakistan
- Sri Lanka
- France
- Kuwait
- Panama
- Sweden
- Gabon
- Kyrgyzstan
- Paraguay
- Switzerland
- Germany
- Libya
- Peru
- Tajikistan
- Ghana
- North Macedonia
- Philippines
- Tanzania, United Republic of
- Greece
- Malaysia
- Poland
- Turkey
- Guatemala
- Mali
- Portugal
- Uganda
- Haiti
- Mexico
- Puerto Rico
- Ukraine
- Honduras
- Moldova, Republic of
- Qatar
- Romania
- United Arab Emirates
- Hungary
- Morocco
- Russian Federation
- United Kingdom
- India
- Nepal
- Saudi Arabia
- United States
- Indonesia
- Netherlands
- Senegal
- Yemen
- Iraq
- New Zealand
- Singapore
- Zambia
- Ireland
- Nicaragua
- South Africa
- Zimbabwe
- Israel
- Nigeria

A.3 Errors of mobility predictions with linear model

Mobility target	Prediction zone	Std. dev.	Prediction error (MAE)
retail and recreation	full data	27.571	13.392
	Switzerland	29.546	15.494
	United States	17.354	22.855
	United Kingdom	28.986	17.5
	India	30.385	15.521
	South Africa	26.986	9.497
	Brazil	22.292	10.346
grocery and pharmacy	full data	22.162	12.226
	Switzerland	16.45	10.463
	United States	9.387	17.665
	United Kingdom	14.755	7.443
	India	23.818	13.283
	South Africa	20.009	9.815
	Brazil	12.037	17.306
parks	full data	44.075	26.838
	Switzerland	40.518	32.728
	United States	27.679	36.415
	United Kingdom	33.979	32.251
	India	24.193	11.909
	South Africa	15.809	18.9
	Brazil	21.66	22.262
transit stations	full data	26.684	12.975
	Switzerland	20.024	6.342
	United States	18.127	16.298
	United Kingdom	24.385	13.805
	India	24.97	7.879
	South Africa	28.971	10.778
	Brazil	22.299	12.472
workplaces	full data	24.347	13.36
	Switzerland	19.105	8.624
	United States	18.317	8.944
	United Kingdom	24.514	17.213
	India	23.89	10.094
	South Africa	26.952	12.129
	Brazil	19.784	17.406

	full data	10.574	5.248
residential	Switzerland	8.15	3.252
	United States	7.181	6.563
	United Kingdom	9.479	4.646
	India	10.276	3.883
	South Africa	12.478	5.675
	Brazil	6.894	5.342
	full data	39.462	21.924
driving	Switzerland	29.046	18.054
	United States	28.673	31.381
	United Kingdom	29.209	14.789
	India	39.761	14.105
	South Africa	35.961	15.836
	Brazil	23.78	18.232
	full data	36.873	13.776
transit	Switzerland	29.644	10.516
	United States	32.19	11.832
	United Kingdom	39.908	12.048
	India	No data	No data
	South Africa	No data	No data
	Brazil	36.712	13.737
	full data	39.62	19.858
walking	Switzerland	27.153	19.018
	United States	30.275	28.904
	United Kingdom	36.213	14.294
	India	35.783	9.445
	South Africa	34.309	12.682
	Brazil	31.614	16.855

A.4 LSTM feature selection process

1. Forward grouped feature selection ([Google mobility] as base):

Added group	Country	MAE mean	MAE std. dev.	Improvement
-	Switzerland	0.286	0.013	
-	United Kingdom	0.207	0.005	
-	South Africa	0.232	0.007	
-	India	0.253	0.021	
-	Brazil	0.23	0.009	
-	United States	0.33	0.006	
Demography	Switzerland	0.253	0.02	0.033
Demography	United Kingdom	0.178	0.01	0.029
Demography	South Africa	0.237	0.017	-0.005
Demography	India	0.219	0.009	0.034
Demography	Brazil	0.243	0.013	-0.013
Demography	United States	0.252	0.011	0.078
Sanitary	Switzerland	0.25	0.013	0.036
Sanitary	United Kingdom	0.191	0.012	0.016
Sanitary	South Africa	0.218	0.011	0.014
Sanitary	India	0.22	0.007	0.033
Sanitary	Brazil	0.242	0.013	-0.012
Sanitary	United States	0.281	0.01	0.049
Policies	Switzerland	0.276	0.019	0.01
Policies	United Kingdom	0.18	0.012	0.027
Policies	South Africa	0.282	0.009	-0.05
Policies	India	0.257	0.028	-0.004
Policies	Brazil	0.223	0.021	0.007
Policies	United States	0.309	0.006	0.021
Continent	Switzerland	0.262	0.017	0.024
Continent	United Kingdom	0.179	0.01	0.028
Continent	South Africa	0.255	0.023	-0.023
Continent	India	0.223	0.015	0.03
Continent	Brazil	0.25	0.012	-0.02
Continent	United States	0.285	0.01	0.045
Weather	Switzerland	0.241	0.01	0.045
Weather	United Kingdom	0.183	0.007	0.024
Weather	South Africa	0.279	0.031	-0.047
Weather	India	0.252	0.026	0.001
Weather	Brazil	0.277	0.014	-0.047
Weather	United States	0.291	0.007	0.039

Added group: **Demographic**

2. Backward feature selection on individual demographic features ([Google mobility, demography] as base):

Removed feature	Country	MAE mean	MAE std. dev.	Improvement
-	Switzerland	0.243	0.017	
-	United Kingdom	0.172	0.003	
-	South Africa	0.221	0.004	
-	India	0.217	0.007	
-	Brazil	0.25	0.009	
-	United States	0.252	0.016	
pop. density	Switzerland	0.248	0.008	-0.005
pop. density	United Kingdom	0.171	0.007	0.001
pop. density	South Africa	0.23	0.017	-0.009
pop. density	India	0.223	0.009	-0.006
pop. density	Brazil	0.229	0.021	0.021
pop. density	United States	0.266	0.02	-0.014
median age	Switzerland	0.248	0.012	-0.005
median age	United Kingdom	0.185	0.009	-0.013
median age	South Africa	0.233	0.015	-0.012
median age	India	0.219	0.016	-0.002
median age	Brazil	0.25	0.004	0.0
median age	United States	0.255	0.013	-0.003
65 older	Switzerland	0.255	0.012	-0.012
65 older	United Kingdom	0.175	0.014	-0.003
65 older	South Africa	0.223	0.009	-0.002
65 older	India	0.227	0.016	-0.01
65 older	Brazil	0.238	0.009	0.012
65 older	United States	0.274	0.006	-0.022
70 older	Switzerland	0.266	0.011	-0.023
70 older	United Kingdom	0.177	0.011	-0.005
70 older	South Africa	0.227	0.009	-0.006
70 older	India	0.205	0.005	0.012
70 older	Brazil	0.247	0.01	0.003
70 older	United States	0.253	0.012	-0.001
gdp per capita	Switzerland	0.262	0.011	-0.019
gdp per capita	United Kingdom	0.188	0.007	-0.016
gdp per capita	South Africa	0.237	0.02	-0.016
gdp per capita	India	0.22	0.009	-0.003
gdp per capita	Brazil	0.231	0.008	0.019
gdp per capita	United States	0.277	0.016	-0.025

No feature removed

3. Forward group feature selection ([Google mobility, demography] as base):

Added group	Country	MAE mean	MAE std. dev.	Improvement
-	Switzerland	0.242	0.018	
-	United Kingdom	0.183	0.013	
-	South Africa	0.226	0.009	
-	India	0.218	0.006	
-	Brazil	0.24	0.011	
-	United States	0.265	0.017	
Sanitary	Switzerland	0.238	0.024	0.004
Sanitary	United Kingdom	0.186	0.009	-0.003
Sanitary	South Africa	0.231	0.017	-0.005
Sanitary	India	0.239	0.016	-0.021
Sanitary	Brazil	0.24	0.014	-0.0
Sanitary	United States	0.258	0.024	0.007
Policies	Switzerland	0.25	0.009	-0.008
Policies	United Kingdom	0.172	0.007	0.011
Policies	South Africa	0.257	0.012	-0.031
Policies	India	0.197	0.006	0.021
Policies	Brazil	0.247	0.022	-0.007
Policies	United States	0.293	0.009	-0.028
Continent	Switzerland	0.249	0.012	-0.007
Continent	United Kingdom	0.188	0.011	-0.005
Continent	South Africa	0.243	0.02	-0.017
Continent	India	0.206	0.01	0.012
Continent	Brazil	0.224	0.009	0.016
Continent	United States	0.231	0.014	0.034
Weather	Switzerland	0.227	0.011	0.015
Weather	United Kingdom	0.158	0.009	0.025
Weather	South Africa	0.257	0.004	-0.031
Weather	India	0.217	0.022	0.001
Weather	Brazil	0.256	0.012	-0.016
Weather	United States	0.272	0.014	-0.007

No group added

4. Forward selection of individual features ([Google mobility, demography] as base):

Added feature	Country	MAE mean	MAE std. dev.	Improvement
-	Switzerland	0.235	0.011	
-	United Kingdom	0.179	0.005	
-	South Africa	0.222	0.005	
-	India	0.223	0.013	
-	Brazil	0.237	0.017	
-	United States	0.259	0.004	
diabetes prev.	Switzerland	0.242	0.013	-0.007
diabetes prev.	United Kingdom	0.181	0.016	-0.002
diabetes prev.	South Africa	0.226	0.015	-0.004
diabetes prev.	India	0.239	0.011	-0.016
diabetes prev.	Brazil	0.238	0.005	-0.001
diabetes prev.	United States	0.271	0.026	-0.012
life expectancy	Switzerland	0.238	0.019	-0.003
life expectancy	United Kingdom	0.182	0.011	-0.003
life expectancy	South Africa	0.228	0.014	-0.006
life expectancy	India	0.221	0.01	0.002
life expectancy	Brazil	0.237	0.008	-0.0
life expectancy	United States	0.268	0.016	-0.009

Added feature	Country	MAE mean	MAE std. dev.	Improvement
-	Switzerland	0.255	0.01	
-	United Kingdom	0.167	0.004	
-	South Africa	0.221	0.009	
-	India	0.215	0.004	
-	Brazil	0.234	0.007	
-	United States	0.264	0.014	
max temp.	Switzerland	0.208	0.013	0.047
max temp.	United Kingdom	0.174	0.007	-0.007
max temp.	South Africa	0.213	0.004	0.008
max temp.	India	0.26	0.022	-0.045
max temp.	Brazil	0.229	0.008	0.005
max temp.	United States	0.269	0.014	-0.005
min temp.	Switzerland	0.228	0.009	0.027
min temp.	United Kingdom	0.174	0.008	-0.007
min temp.	South Africa	0.206	0.004	0.015
min temp.	India	0.22	0.02	-0.005
min temp.	Brazil	0.225	0.004	0.009
min temp.	United States	0.278	0.01	-0.014
perceived temp.	Switzerland	0.218	0.024	0.037
perceived temp.	United Kingdom	0.175	0.013	-0.008
perceived temp.	South Africa	0.215	0.022	0.006
perceived temp.	India	0.216	0.008	-0.001
perceived temp.	Brazil	0.244	0.004	-0.01
perceived temp.	United States	0.282	0.011	-0.018
humidity	Switzerland	0.266	0.012	-0.011
humidity	United Kingdom	0.173	0.01	-0.006
humidity	South Africa	0.236	0.03	-0.015
humidity	India	0.255	0.006	-0.04
humidity	Brazil	0.272	0.021	-0.038
humidity	United States	0.267	0.015	-0.003
pressure	Switzerland	0.223	0.01	0.032
pressure	United Kingdom	0.165	0.011	0.002
pressure	South Africa	0.208	0.009	0.013
pressure	India	0.211	0.01	0.004
pressure	Brazil	0.245	0.019	-0.011
pressure	United States	0.237	0.022	0.027

Added feature: **Pressure**

5. Forward selection of individual features ([**Google mobility group, demography group, pressure**] as base):

Added feature	Country	MAE mean	MAE std. dev.	Improvement
max temp.	Switzerland	0.222	0.019	0.001
max temp.	United Kingdom	0.152	0.004	0.013
max temp.	South Africa	0.224	0.013	-0.016
max temp.	India	0.218	0.01	-0.007
max temp.	Brazil	0.242	0.01	0.003
max temp.	United States	0.269	0.004	-0.032
min temp.	Switzerland	0.203	0.012	0.02
min temp.	United Kingdom	0.174	0.013	-0.009
min temp.	South Africa	0.219	0.022	-0.011
min temp.	India	0.221	0.012	-0.01
min temp.	Brazil	0.249	0.006	-0.004
min temp.	United States	0.271	0.007	-0.034
perceived temp.	Switzerland	0.225	0.009	-0.002
perceived temp.	United Kingdom	0.165	0.014	0.0
perceived temp.	South Africa	0.225	0.005	-0.017
perceived temp.	India	0.218	0.011	-0.007
perceived temp.	Brazil	0.241	0.01	0.004
perceived temp.	United States	0.27	0.009	-0.033
humidity	Switzerland	0.25	0.007	-0.027
humidity	United Kingdom	0.162	0.008	0.003
humidity	South Africa	0.207	0.014	0.001
humidity	India	0.216	0.012	-0.005
humidity	Brazil	0.275	0.019	-0.03
humidity	United States	0.259	0.005	-0.022
diabetes prev.	Switzerland	0.245	0.021	-0.022
diabetes prev.	United Kingdom	0.18	0.017	-0.015
diabetes prev.	South Africa	0.209	0.025	-0.001
diabetes prev.	India	0.215	0.009	-0.004
diabetes prev.	Brazil	0.287	0.03	-0.042
diabetes prev.	United States	0.253	0.005	-0.016
life expectancy	Switzerland	0.24	0.016	-0.017
life expectancy	United Kingdom	0.172	0.009	-0.007
life expectancy	South Africa	0.212	0.009	-0.004
life expectancy	India	0.218	0.009	-0.007
life expectancy	Brazil	0.234	0.006	0.011
life expectancy	United States	0.252	0.02	-0.015

No feature added

A.5 LSTM time window selection

Model performance by using time windows from 0 to 8 days:

Country	Window size	MAE std. dev.	MAE mean
1 day	Switzerland	0.016	0.244
2 days		0.007	0.209
3 days		0.014	0.212
4 days		0.01	0.198
5 days		0.015	0.195
6 days		0.007	0.188
7 days		0.012	0.179
8 days		0.015	0.179
1 day	United Kingdom	0.01	0.194
2 days		0.016	0.188
3 days		0.019	0.172
4 days		0.013	0.154
5 days		0.016	0.149
6 days		0.011	0.172
7 days		0.018	0.142
8 days		0.006	0.149
1 day	South Africa	0.004	0.193
2 days		0.005	0.199
3 days		0.012	0.191
4 days		0.015	0.181
5 days		0.019	0.219
6 days		0.013	0.208
7 days		0.006	0.201
8 days		0.007	0.235
1 day	India	0.021	0.23
2 days		0.007	0.223
3 days		0.015	0.226
4 days		0.018	0.219
5 days		0.004	0.206
6 days		0.02	0.214
7 days		0.01	0.21
8 days		0.025	0.226

1 day	Brazil	0.007	0.2
2 days		0.008	0.195
3 days		0.004	0.204
4 days		0.003	0.182
5 days		0.01	0.185
6 days		0.01	0.185
7 days		0.016	0.176
8 days		0.007	0.19
1 day	United States	0.017	0.259
2 days		0.019	0.236
3 days		0.035	0.229
4 days		0.008	0.213
5 days		0.014	0.23
6 days		0.019	0.216
7 days		0.015	0.226
8 days		0.044	0.229

A.6 Ultimate reproduction number prediction

The reproduction number prediction by optimized model:

Zone	Mean Absolute Error
World (average of 93 countries)	0.254
Afghanistan	0.192
Angola	0.379
Argentina	0.218
Australia	0.367
Austria	0.267
Bahrain	0.229
Bangladesh	0.163
Belarus	0.327
Belgium	0.136
Bolivia, Plurinational State of	0.155
Bosnia and Herzegovina	0.219
Brazil	0.201
Bulgaria	0.257
Cameroon	0.488
Canada	0.168
Cabo Verde	0.275

Chile	0.188
Colombia	0.126
Costa Rica	0.279
Côte d'Ivoire	0.206
Czechia	0.344
Denmark	0.243
Dominican Republic	0.138
Ecuador	0.309
Egypt	0.185
El Salvador	0.166
Estonia	0.283
Finland	0.189
France	0.244
Gabon	0.31
Germany	0.14
Ghana	0.257
Greece	0.318
Guatemala	0.212
Haiti	0.347
Honduras	0.304
Hungary	0.216
India	0.226
Indonesia	0.129
Iraq	0.25
Ireland	0.262
Israel	0.426
Italy	0.148
Japan	0.377
Kazakhstan	0.3
Kenya	0.229
Kuwait	0.342
Kyrgyzstan	0.332
Libya	0.316
North Macedonia	0.221
Malaysia	0.379
Mali	0.256
Mexico	0.138
Moldova, Republic of	0.195
Morocco	0.284
Nepal	0.392

Netherlands	0.138
New Zealand	0.423
Nicaragua	0.377
Nigeria	0.2
Norway	0.241
Oman	0.228
Pakistan	0.181
Panama	0.181
Paraguay	0.517
Peru	0.271
Philippines	0.27
Poland	0.138
Portugal	0.222
Puerto Rico	0.263
Qatar	0.491
Romania	0.229
Russian Federation	0.194
Saudi Arabia	0.243
Senegal	0.211
Singapore	0.262
South Africa	0.168
Korea, Republic of	0.511
Spain	0.164
Sri Lanka	0.48
Sweden	0.165
Switzerland	0.169
Tajikistan	0.246
Tanzania, United Republic of	0.251
Turkey	0.197
Uganda	0.306
Ukraine	0.174
United Arab Emirates	0.24
United Kingdom	0.163
United States	0.196
Yemen	0.472
Zambia	0.602
Zimbabwe	0.389

A.7 R prediction from policy features

The reproduction number prediction by optimized model trained on policies features instead of mobility:

Zone	Mean Absolute Error
World (average of 93 countries)	0.254
Afghanistan	0.244
Albania	0.254
Algeria	0.393
Angola	0.379
Argentina	0.188
Armenia	0.153
Australia	0.45
Austria	0.236
Azerbaijan	0.241
Bahrain	0.233
Bangladesh	0.164
Belarus	1.216
Belgium	0.197
Bolivia, Plurinational State of	0.17
Bosnia and Herzegovina	0.244
Brazil	0.252
Bulgaria	0.254
Cameroon	0.489
Canada	0.209
Cabo Verde	0.262
Central African Republic	0.269
Chad	0.348
Chile	0.317
China	0.472
Colombia	0.156
Costa Rica	0.3
Côte d'Ivoire	0.216
Cyprus	0.281
Czechia	0.216
Congo, The Democratic Republic of the	0.254
Denmark	0.227
Djibouti	0.591

Dominican Republic	0.175
Ecuador	0.311
Egypt	0.203
El Salvador	0.17
Estonia	0.286
Finland	0.24
France	0.264
Gabon	0.34
Germany	0.306
Ghana	0.269
Greece	0.436
Guatemala	0.209
Guinea	0.224
Guinea-Bissau	0.358
Haiti	0.345
Honduras	0.302
Hungary	0.229
Iceland	0.464
India	0.188
Indonesia	0.167
Iran, Islamic Republic of	0.193
Iraq	0.268
Ireland	0.265
Israel	0.492
Italy	0.279
Japan	0.851
Kazakhstan	0.244
Kenya	0.2
Kuwait	0.4
Kyrgyzstan	0.271
Libya	0.268
North Macedonia	0.214
Malawi	0.319
Malaysia	0.551
Maldives	0.299
Mali	0.223
Mauritania	0.318
Mexico	0.21
Moldova, Republic of	0.207
Morocco	0.304

Namibia	0.509
Nepal	0.396
Netherlands	0.143
New Zealand	0.463
Nicaragua	0.732
Nigeria	0.2
Norway	0.394
Oman	0.185
Pakistan	0.179
Panama	0.188
Paraguay	0.516
Peru	0.253
Philippines	0.297
Poland	0.144
Portugal	0.191
Puerto Rico	0.263
Qatar	0.327
Romania	0.152
Russian Federation	0.284
Saudi Arabia	0.207
Senegal	1.138
Sierra Leone	0.22
Singapore	0.641
South Africa	0.287
Korea, Republic of	0.411
Spain	0.212
Sri Lanka	0.455
Sudan	0.304
Suriname	0.271
Sweden	0.27
Switzerland	0.232
Tajikistan	0.286
Tanzania, United Republic of	0.233
Turkey	0.157
Uganda	0.528
Ukraine	0.222
United Arab Emirates	0.233
United Kingdom	0.192
United States	0.391
Uzbekistan	0.198

Yemen	0.277
Zambia	0.559
Zimbabwe	0.45