MODELLING THE COVID-19 PANDEMIC REQUIRES A MODEL… BUT ALSO DATA!

Marc Lavielle
Inria Saclay & Ecole Polytechnique

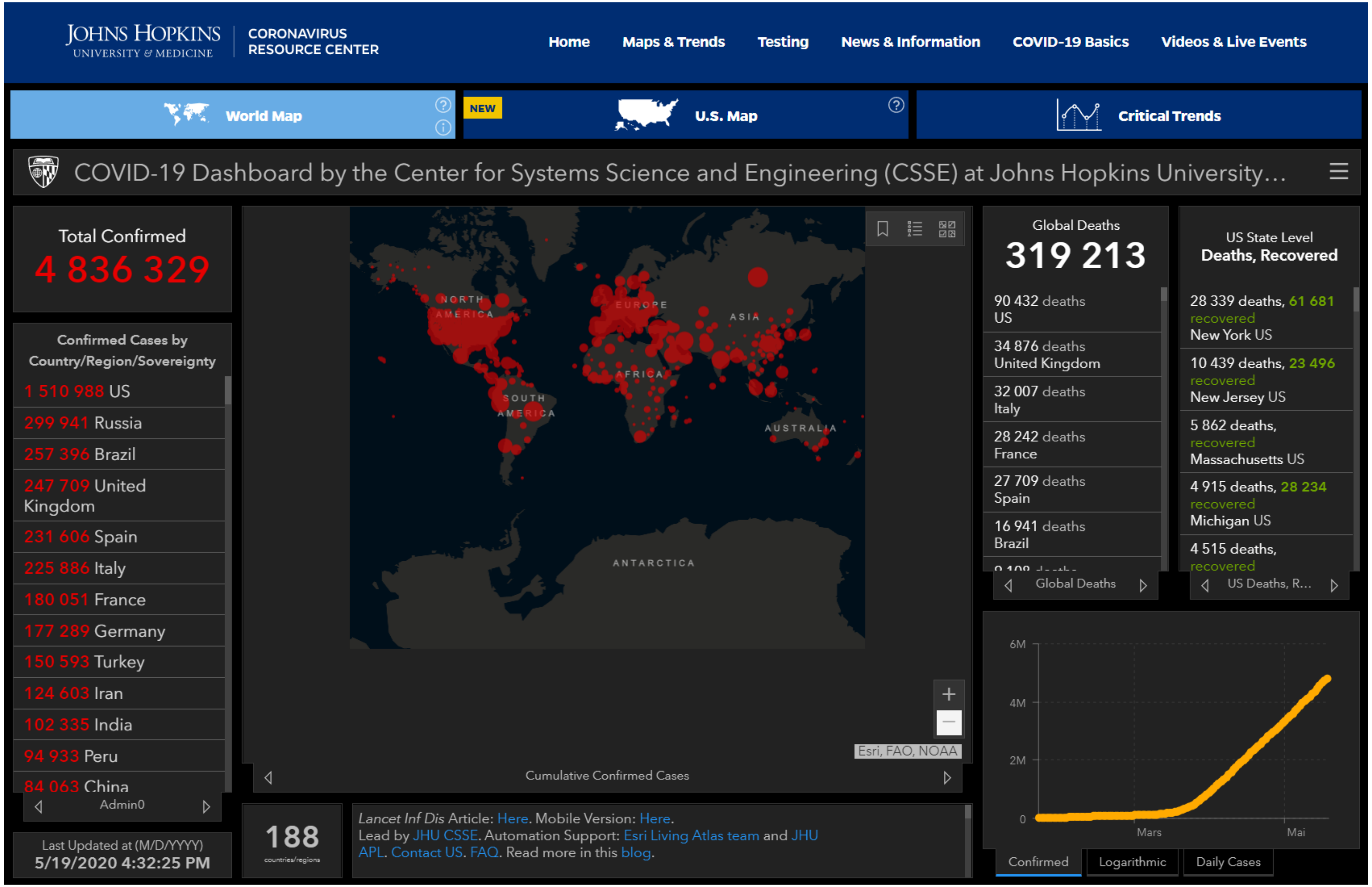
November 18th, 2020

EcoDep

Introduction:

data & objectives

https://coronavirus.jhu.edu/map.html



Why GitHub? Team Enterprise Explore Marketplace Pricing Search Sign in Sign up

CSSEGISandData / **COVID-19**

Watch 932 Star 22.2k Fork 13.2k

Code | Issues 1,318 | Pull requests 254 | Actions | Projects 0 | Security 0 | Insights

Branch: master COVID-19 / csse_covid_19_data / csse_covid_19_time_series / Create new file Find file History

CSSEGISandData automated update Latest commit 8390ba6 13 hours ago

..		
.gitignore	update	3 months ago
Errata.csv	Update Errata.csv	8 days ago
README.md	Update README	18 days ago
time_series_covid19_confirmed_US.csv	automated update	13 hours ago
time_series_covid19_confirmed_global.csv	automated update	13 hours ago
time_series_covid19_deaths_US.csv	automated update	13 hours ago
time_series_covid19_deaths_global.csv	automated update	13 hours ago
time_series_covid19_recovered_global.csv	automated update	13 hours ago

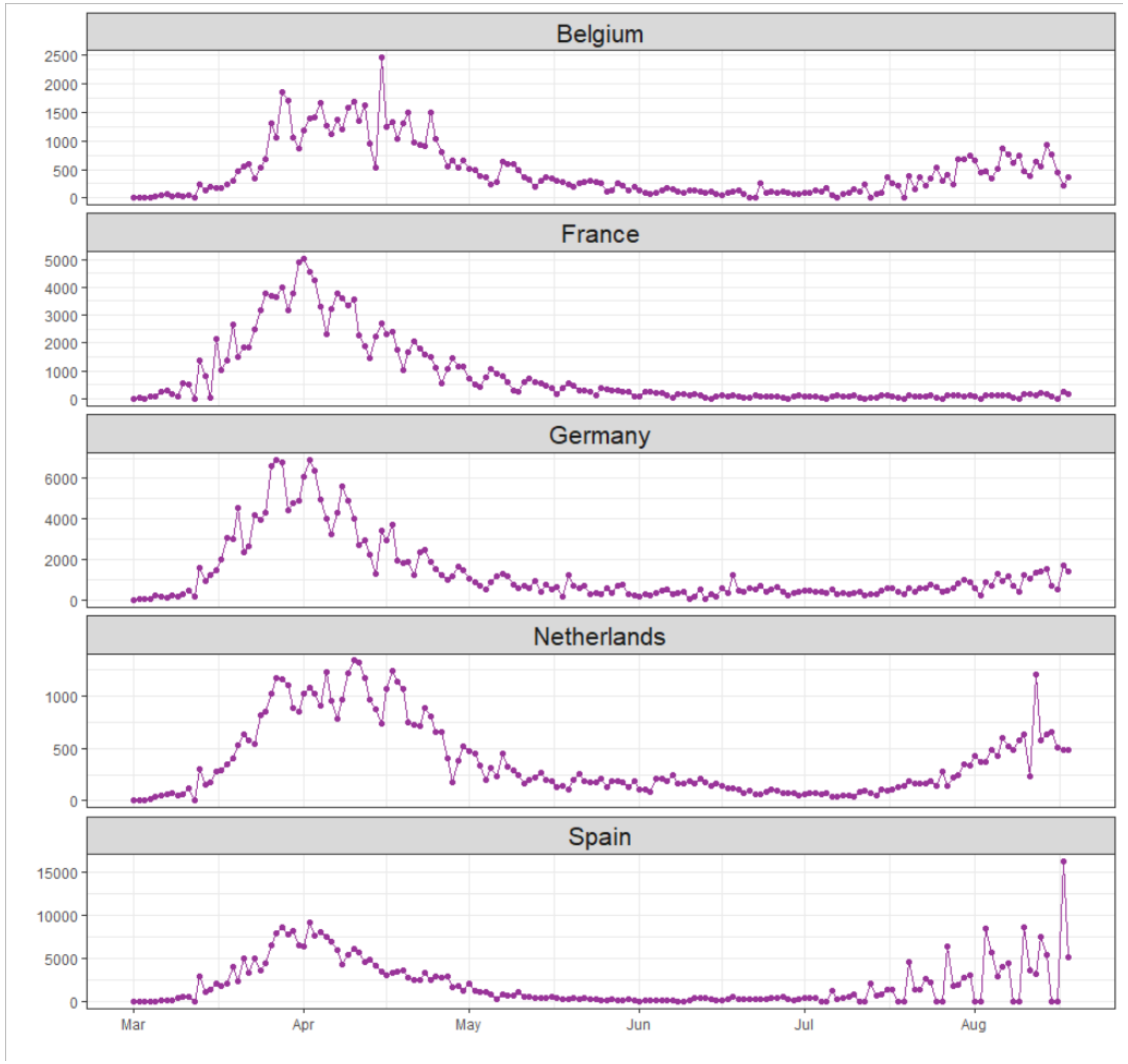
README.md

Time series summary (csse_covid_19_time_series)

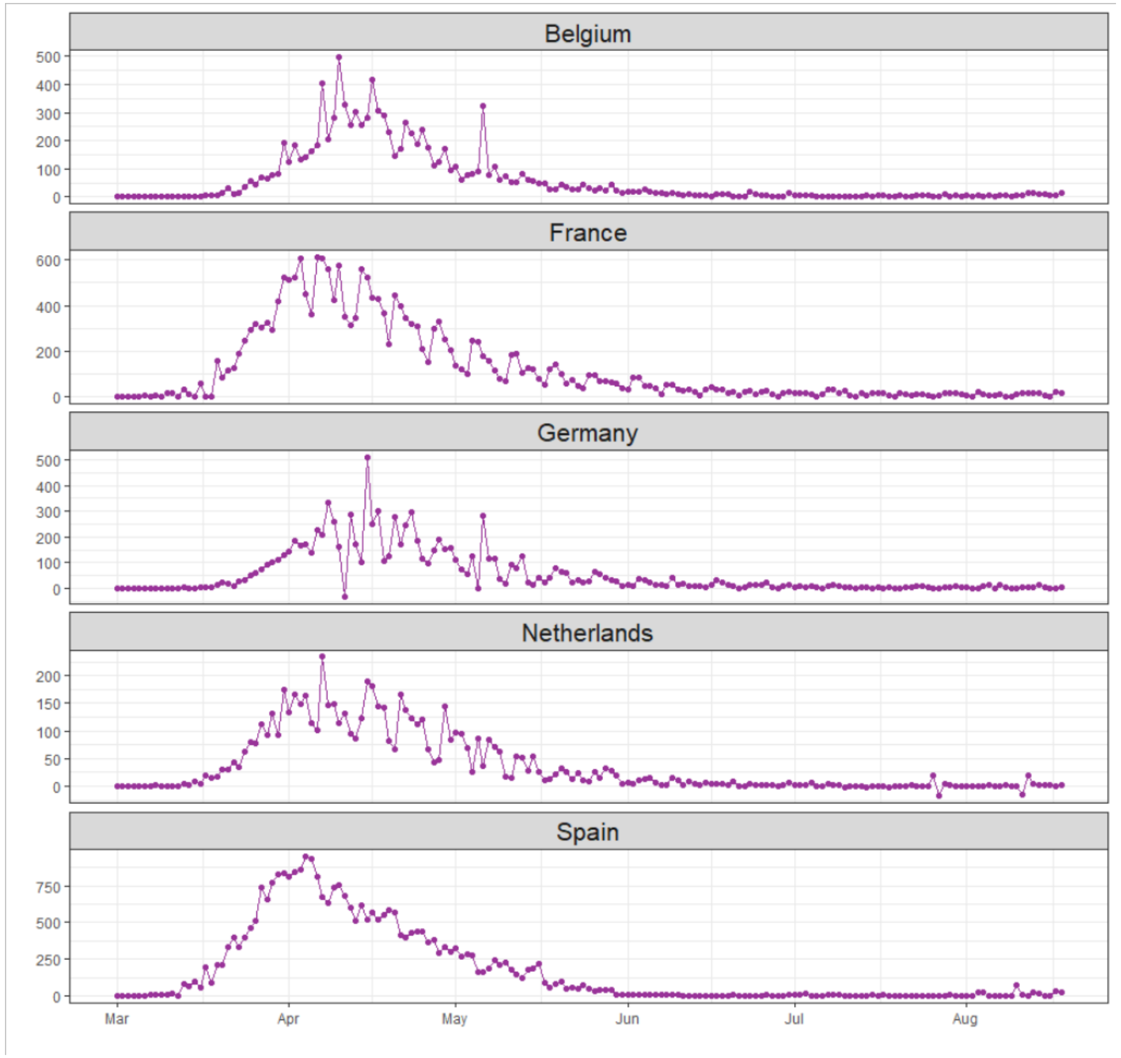
This folder contains daily time series summary tables, including confirmed, deaths and recovered. All data is read in from the daily case report. The time series tables are subject to be updated if inaccuracies are identified in our historical data. The daily reports will not be adjusted in these instances to maintain a record of raw data.

Two time series tables are for the US confirmed cases and deaths, reported at the county level. They are named `time_series_covid19_confirmed_US.csv`, `time_series_covid19_deaths_US.csv`, respectively.

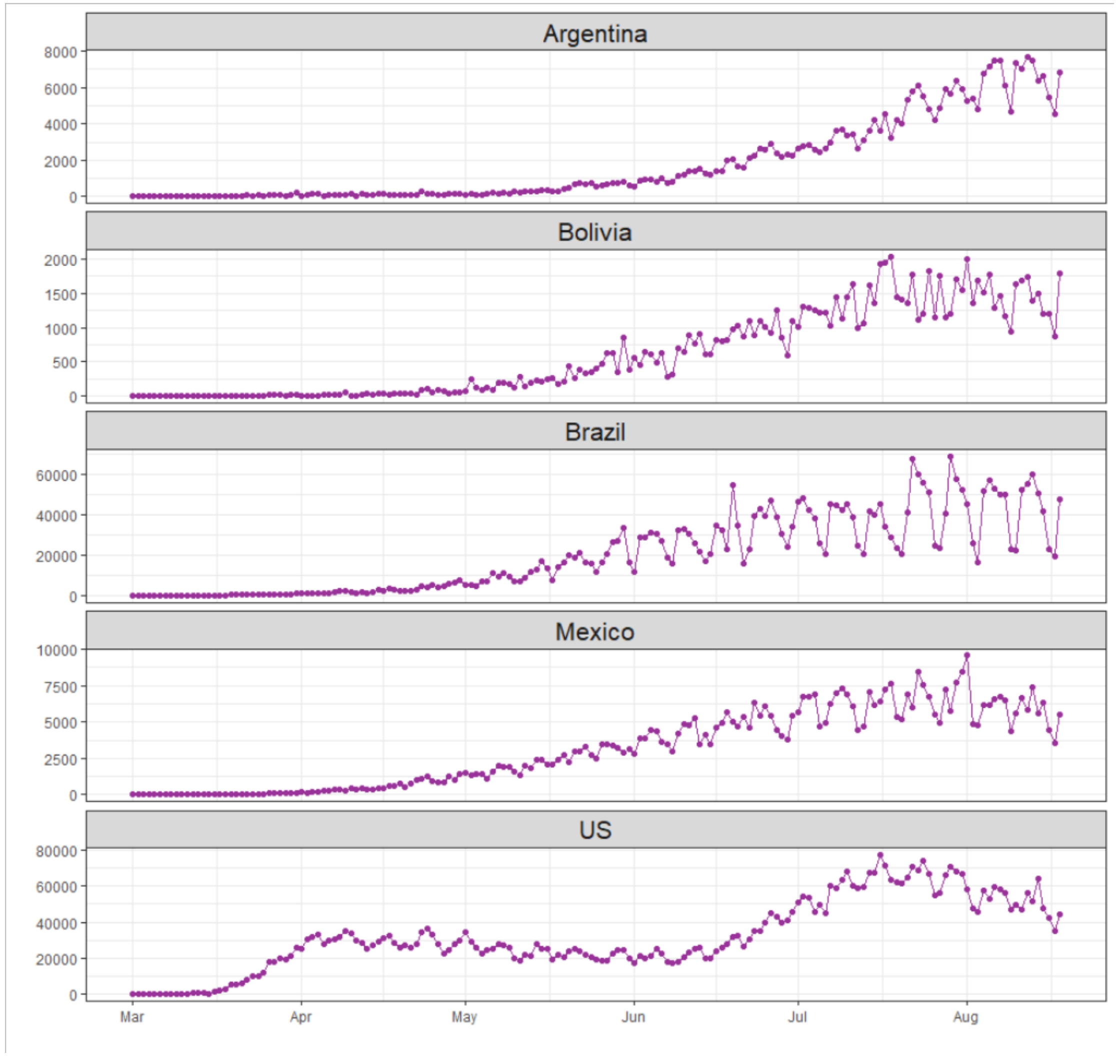
Daily number of confirmed cases



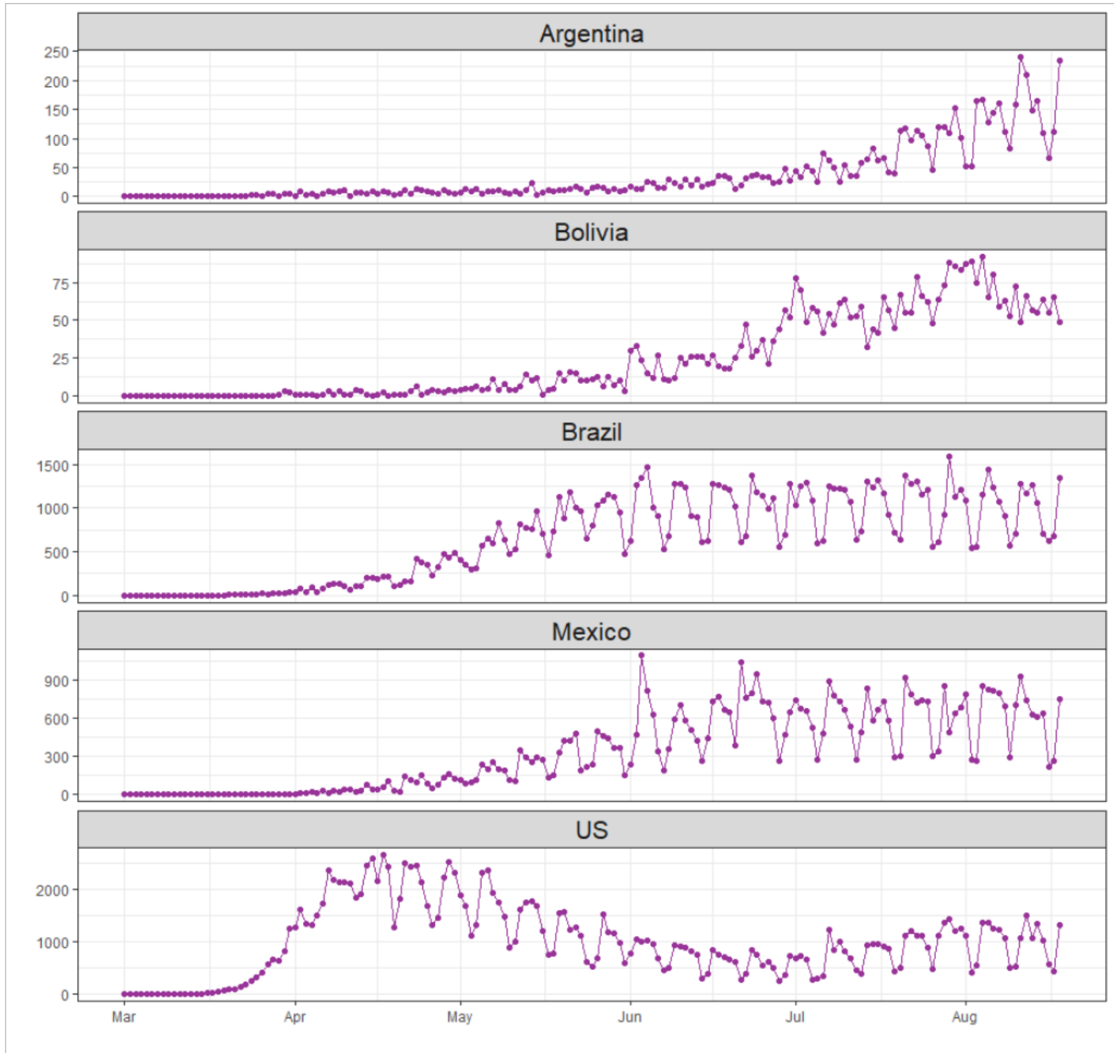
Daily number of deaths



Daily number of confirmed cases



Daily number of deaths



https://www.data.gouv.fr/fr/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19/

Données hospitalières relatives à l'épidémie de COVID-19

Ce jeu de données provient d'un service public certifié

COVID-19

Point d'information : Un établissement hospitalier de l'Essonne (91) a transmis ce jour (18/09/2020) près de 240 dossiers concernant des patients hospitalisés au cours des derniers mois. De ce fait, les indicateurs hospitaliers du 18 septembre 2020, présentés par date de déclaration, présentent une augmentation soudaine dans ce département. Cet impact est également visible à un niveau régional (région Ile-de-France) et national. Cette augmentation du nombre de personnes hospitalisées, déclarées le 18 septembre par cet établissement ne reflète pas des nouvelles hospitalisations mais des nouvelles déclarations.

Les actions de Santé publique France

Santé publique France a pour mission d'améliorer et de protéger la santé des populations. Durant la crise sanitaire liée à l'épidémie du COVID-19, Santé publique France se charge de surveiller et comprendre la dynamique de l'épidémie, d'anticiper les différents scénarii et de mettre en place des actions pour prévenir et limiter la transmission de ce virus sur le territoire national.

Description du jeu de données

Le présent jeu de données renseigne sur la situation hospitalières concernant l'épidémie de COVID-19.

Quatre fichiers sont proposés :

- **Les données hospitalières relatives à l'épidémie du COVID-19 par département et sexe du patient** : nombre de patients hospitalisés, nombre de personnes actuellement en réanimation ou soins intensifs, nombre cumulé de personnes retournées à domicile, nombre cumulé de personnes décédées.
- **Les données hospitalières relatives à l'épidémie du COVID-19 par région et classe d'âge du patient** : nombre de patients hospitalisés, nombre de personnes actuellement en réanimation ou soins intensifs, nombre cumulé de personnes retournées à domicile, nombre cumulé de personnes décédées.
- **Les données hospitalières quotidiennes relatives à l'épidémie du COVID-19 par département du patient** : nombre quotidien de personnes nouvellement hospitalisées, nombre quotidien de nouvelles admissions en réanimation, nombre quotidien de personnes nouvellement décédées, nombre quotidien de nouveaux retours à domicile.
- **Les données relatives aux établissements hospitaliers par département** : nombre cumulé de services ayant déclaré au moins un cas.

Producteur

PRODUCTEUR CERTIFIÉ

Santé publique France

Santé publique France

Santé publique France est l'agence nationale de santé publique. Créée en mai 2016 par ordonnance et décret, c'est un établissement public administratif sous tutelle du ministère...

VOIR LE PROFIL

CONTACTER

SUIVRE

Informations

Covid-19

Licence Ouverte / Open Licence version 2.0

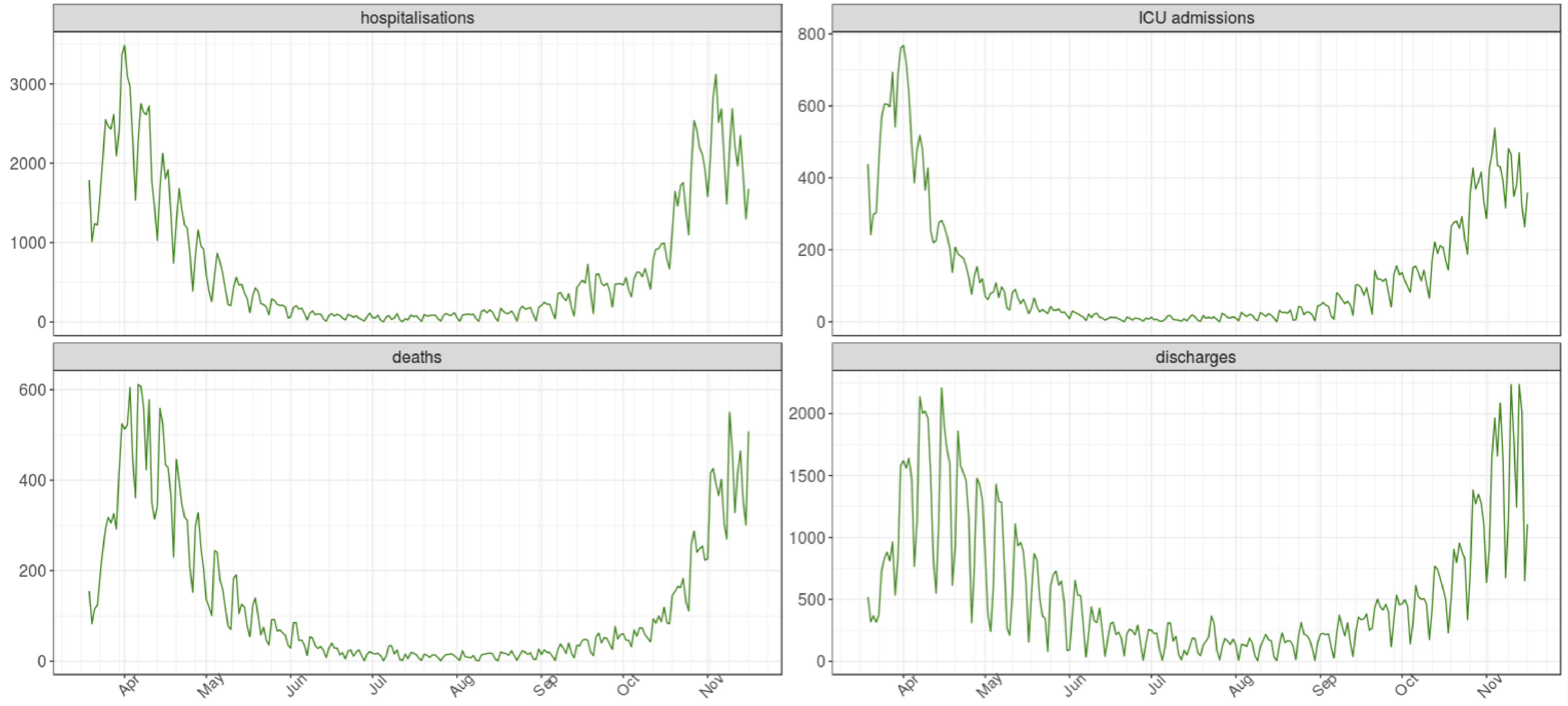
Quotidienne

27 mars 2020

15 novembre 2020

8 avril 2020

coronavius covid-19 covid19 covid19france Suggérer un mot-clé



hospitalisations
ICU admissions
deaths
discharges
0
1000
2000
3000
200
400
600
800
500
1000
1500
2000
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov

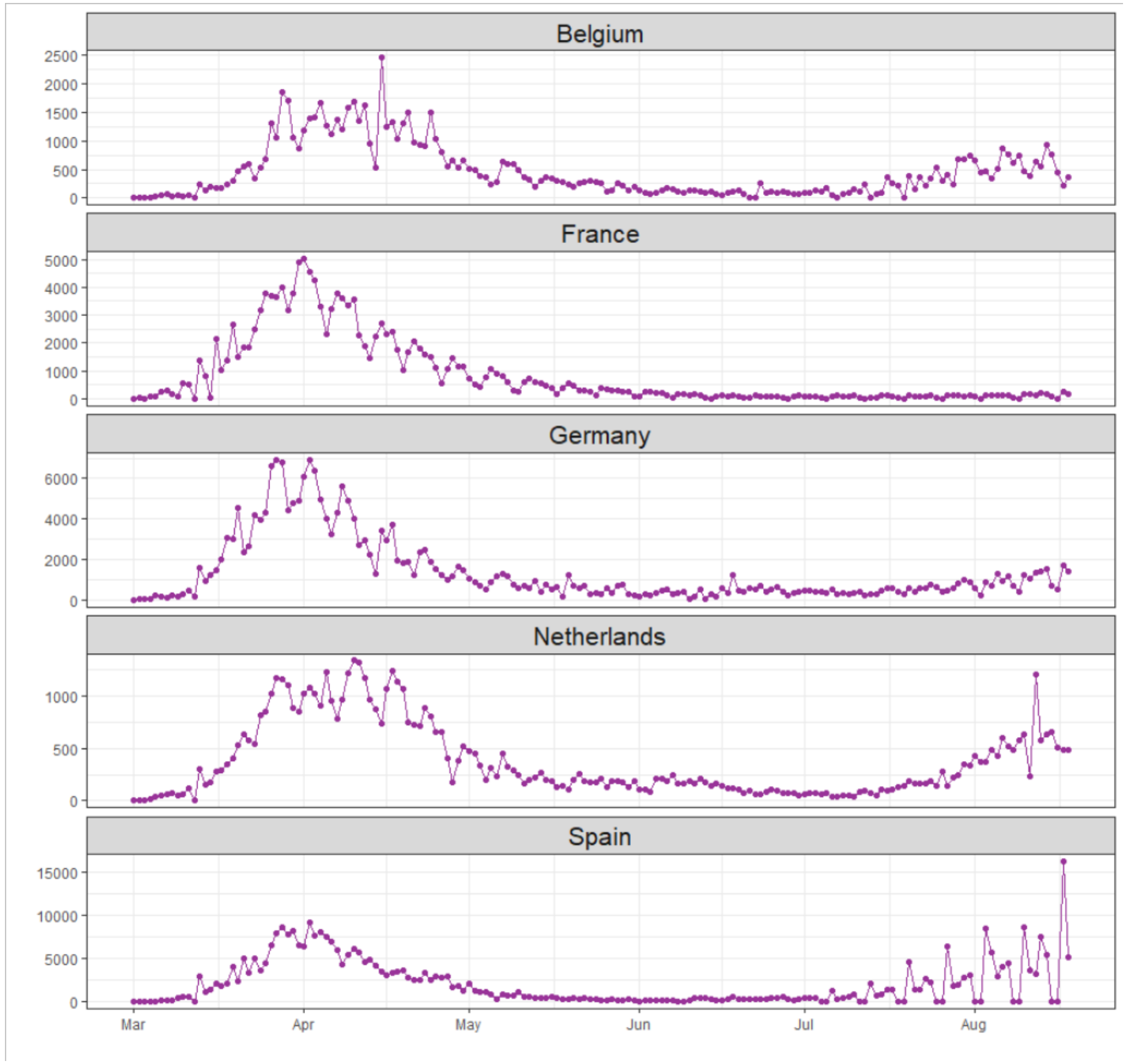
- The objective **is not** to build a model… and try to “calibrate” it in order to fit the data as well as possible

- The objective **is not** to build a model… and try to "calibrate" it in order to fit the data as well as possible

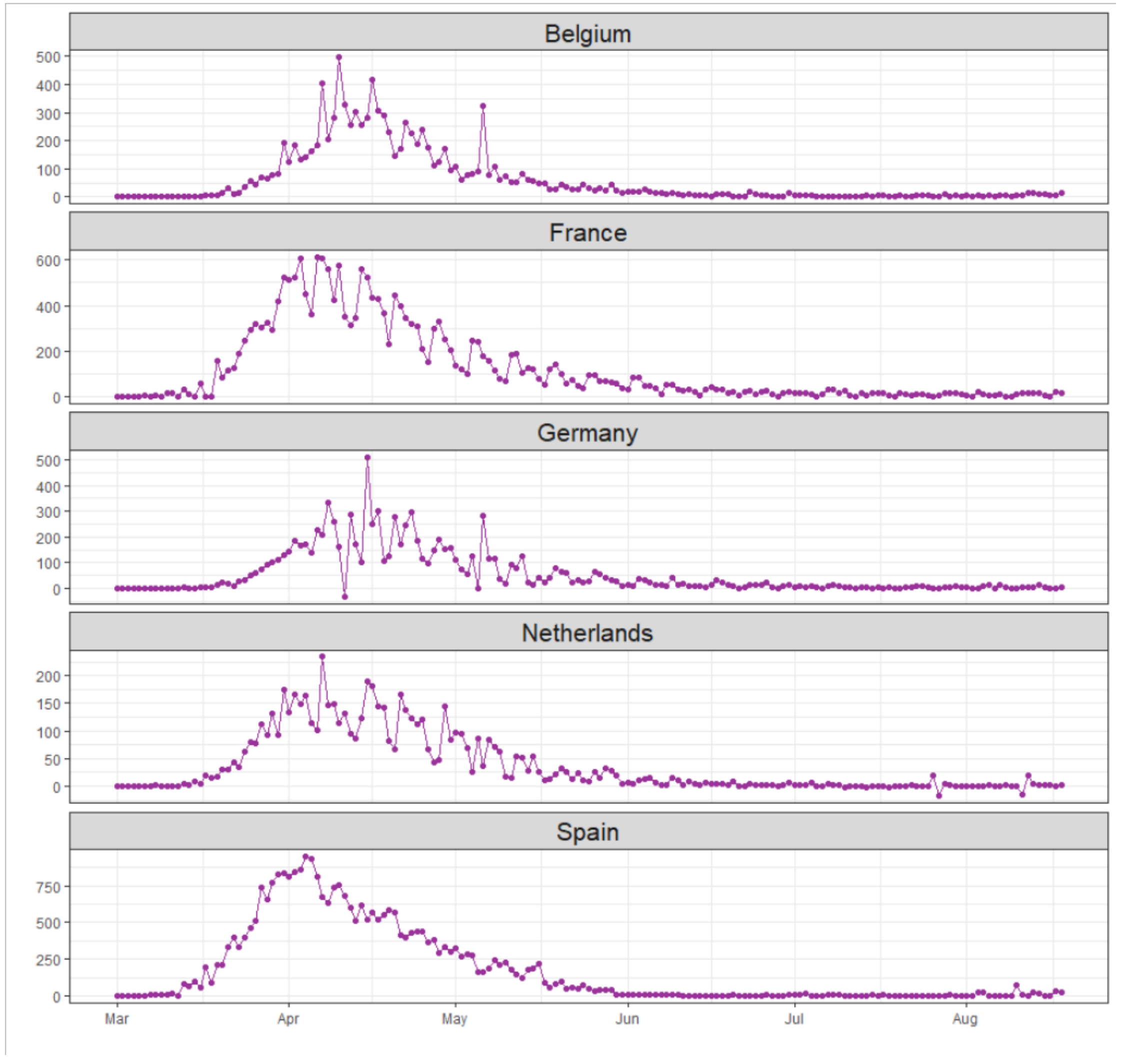
- The objective is to develop a model
  - **for the observed data**, and validated by the data,
  - that provides good short-term predictions,
  - that is implemented as an open-access interactive tool.

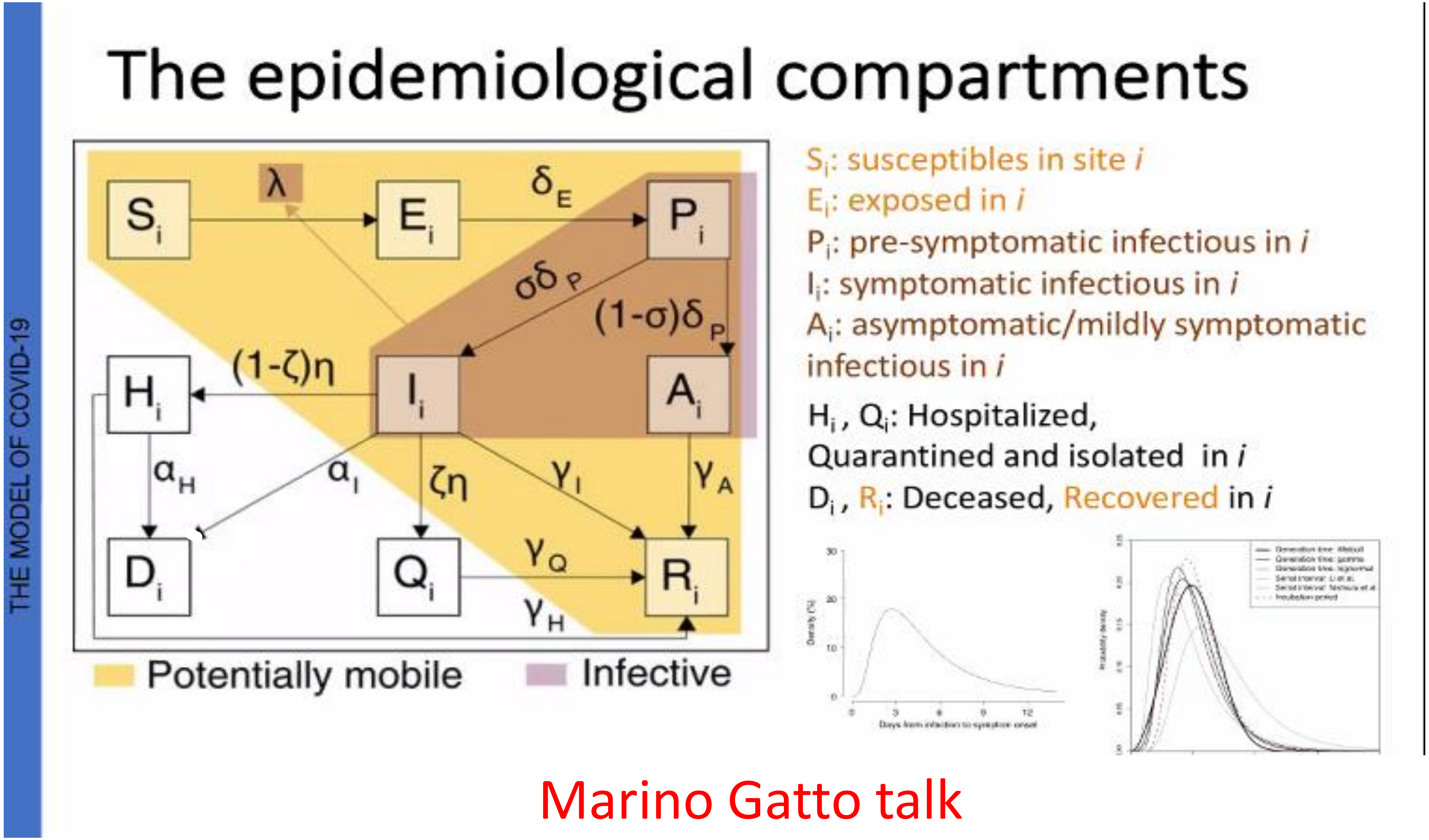
I
Modelling the JHU data

Daily number of confirmed cases



Daily number of deaths



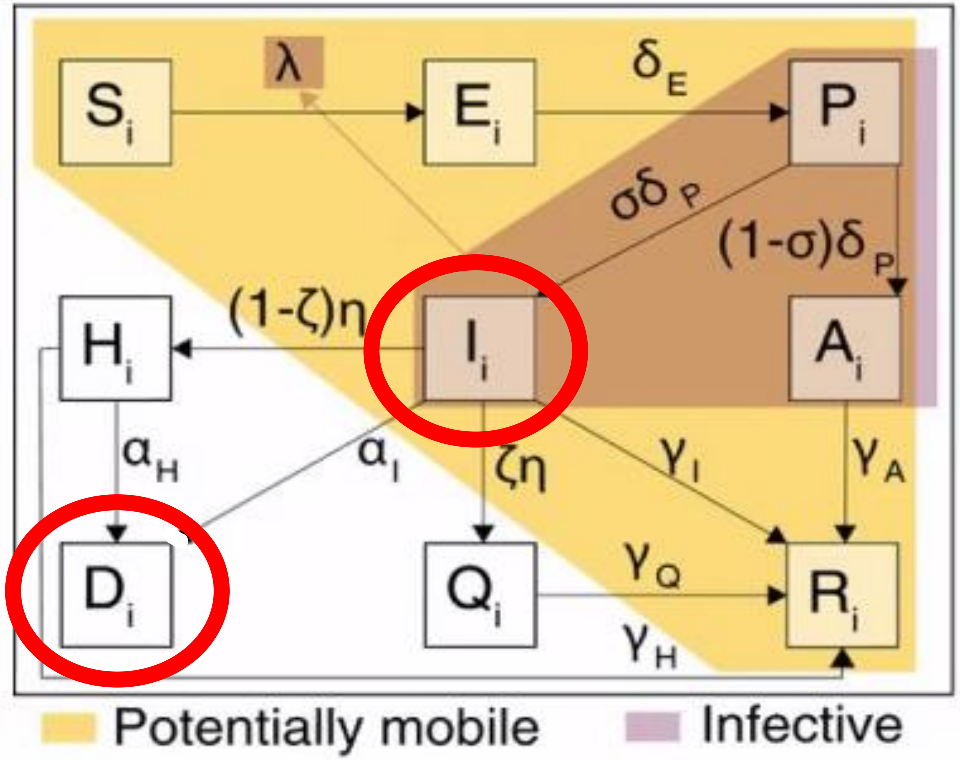


Marino Gatto talk

(Modeling the propagation of Covid-19, May 2020)

THE MODEL OF COVID-19

The epidemiological compartments



S_i: susceptibles in site *i*

E_i: exposed in *i*

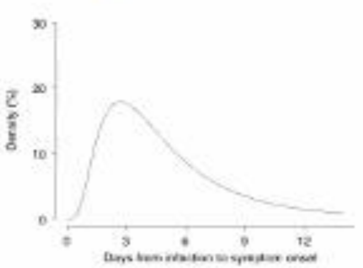
P_i: pre-symptomatic infectious in *i*

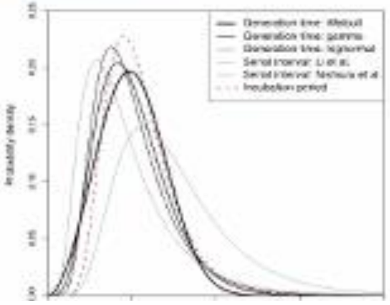
I_i: symptomatic infectious in *i*

A_i: asymptomatic/mildly symptomatic infectious in *i*

H_i, Q_i: Hospitalized, Quarantined and isolated in *i*

D_i, R_i: Deceased, Recovered in *i*

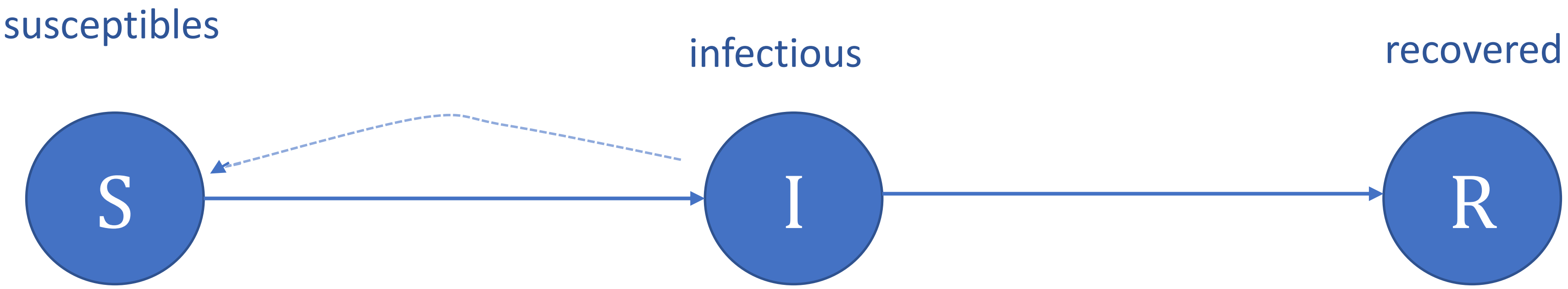




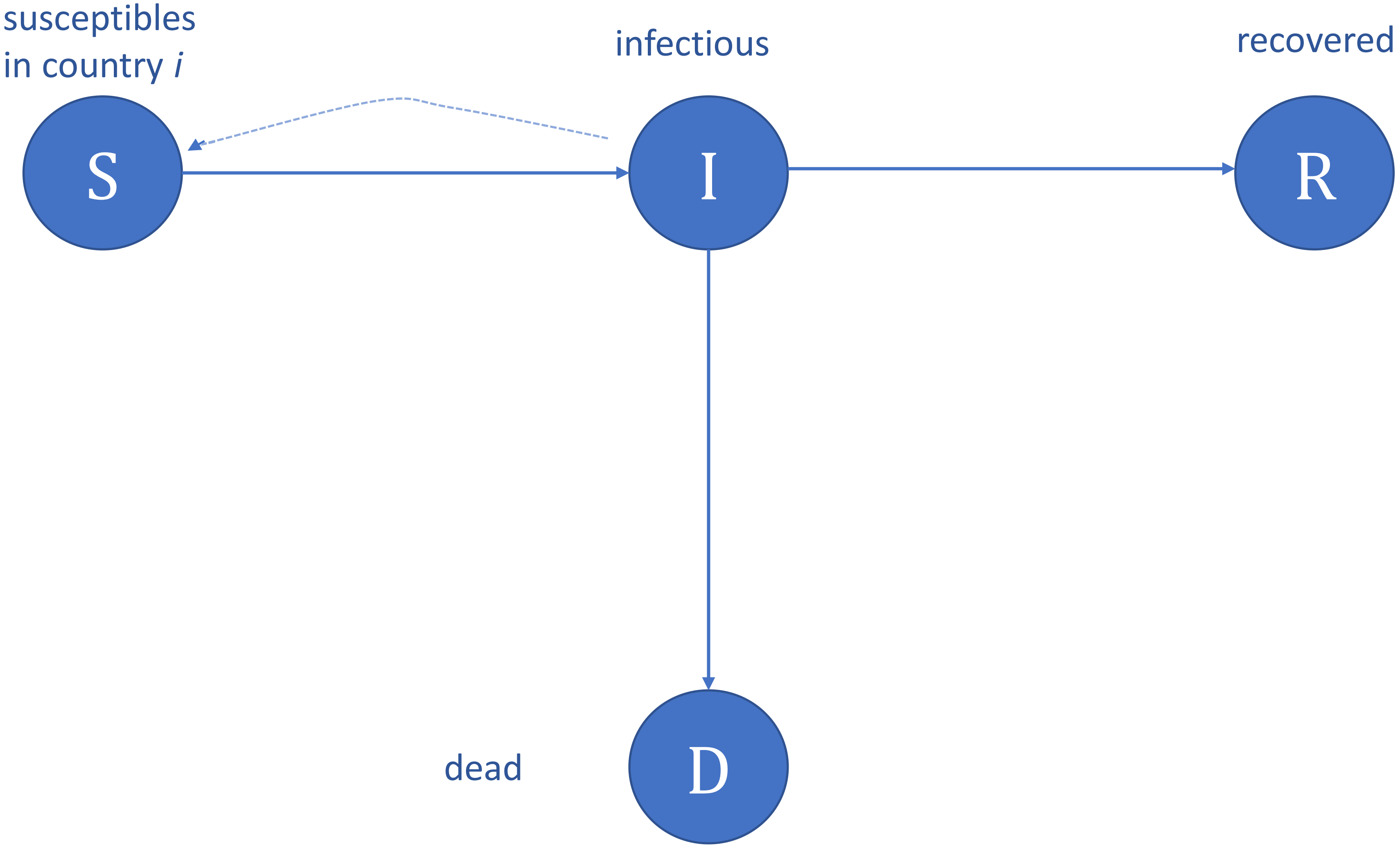
Marino Gatto talk

(Modeling the propagation of Covid-19, May 2020)

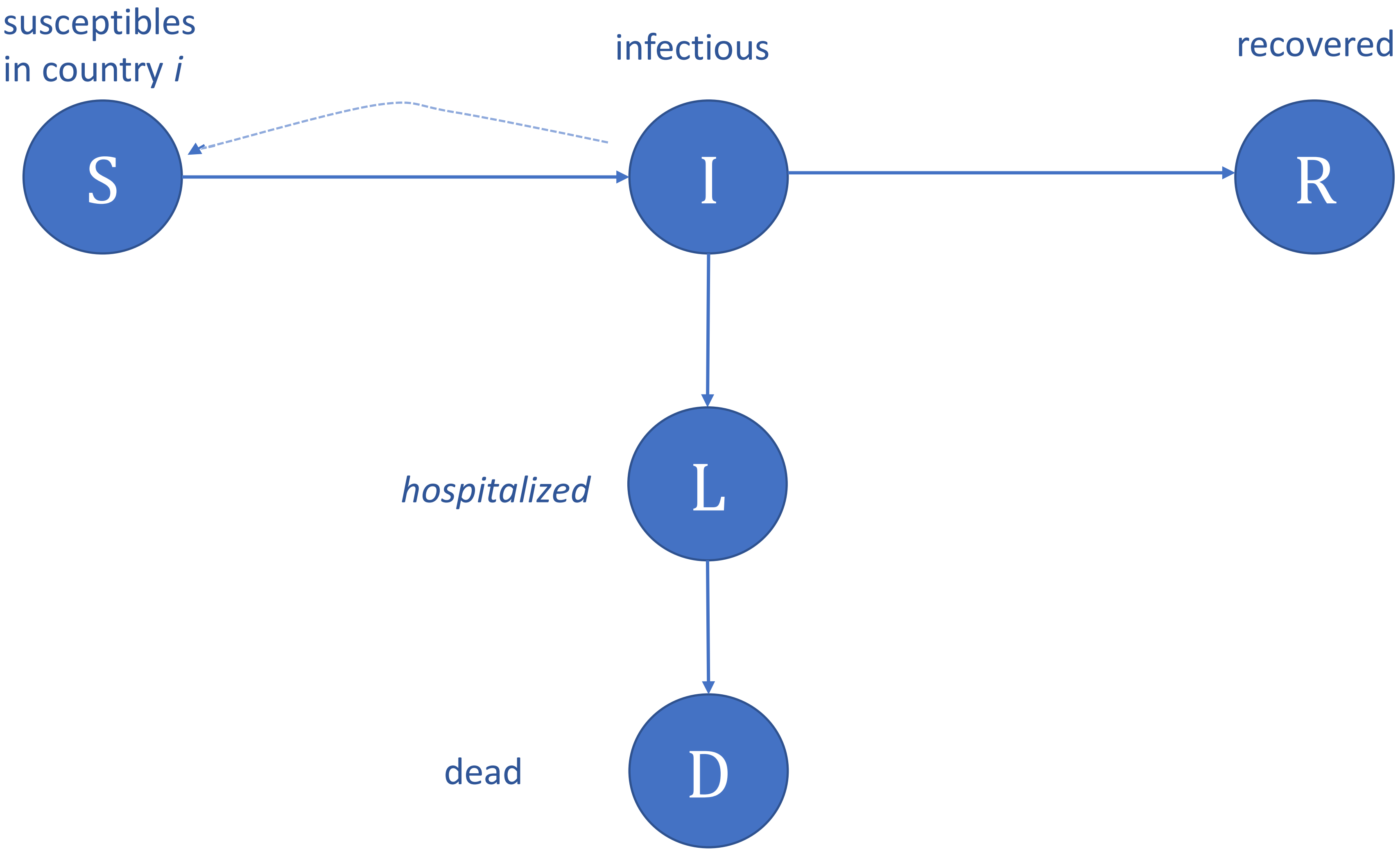
The epidemiological compartments



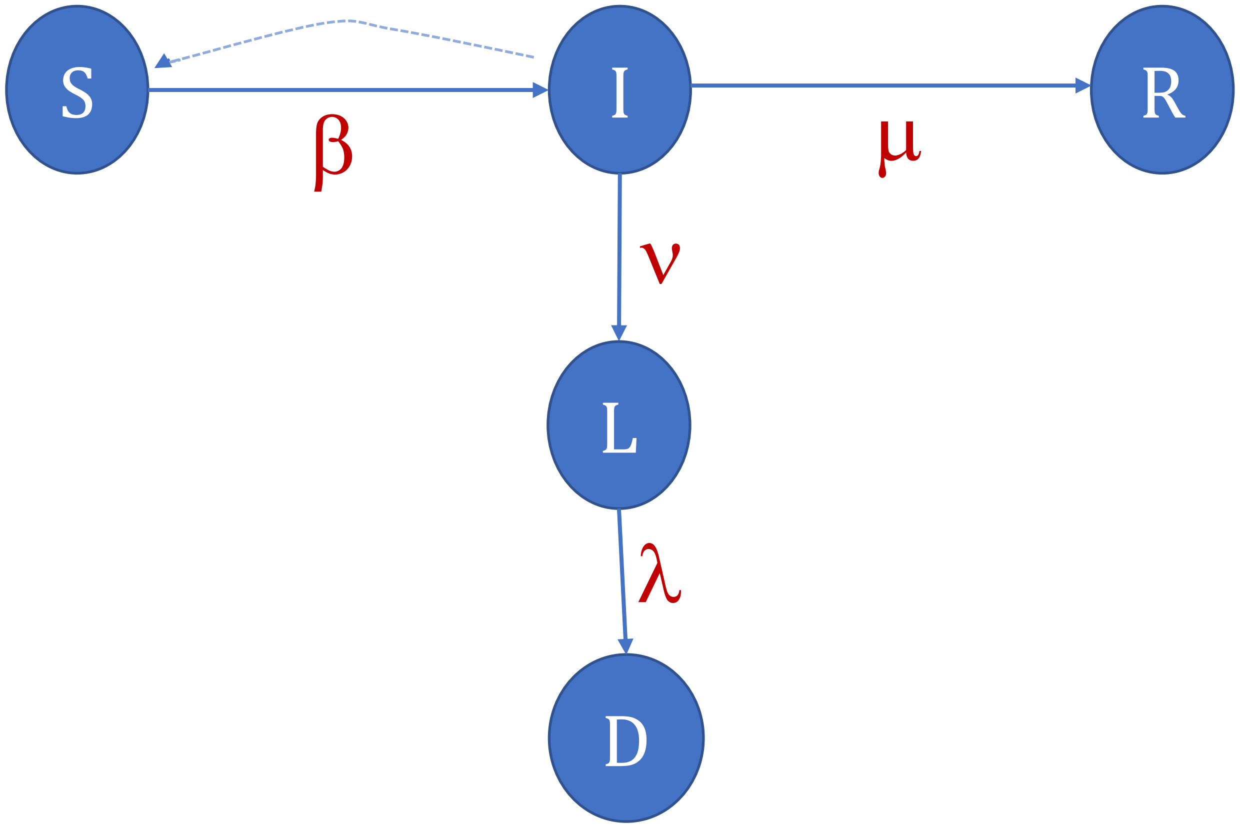
The epidemiological compartments



The epidemiological compartments



The epidemiological compartments



$$\dot{S}(t) = -\beta \frac{S(t)}{N} I(t)$$

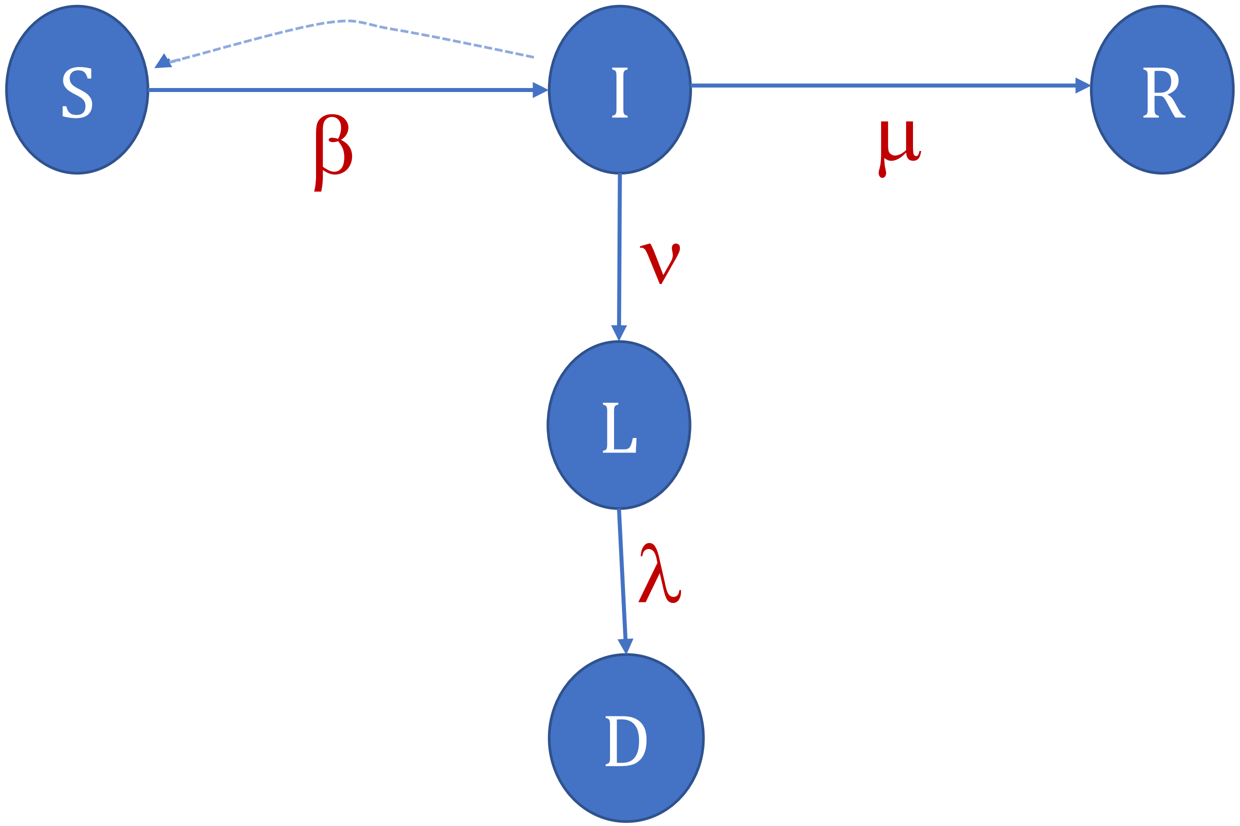
$$\dot{I}(t) = \beta \frac{S(t)}{N} I(t) - \mu I(t) - \nu I(t)$$

$$\dot{R}(t) = \mu I(t)$$

$$\dot{L}(t) = \nu I(t) - \lambda L(t)$$

$$\dot{D}(t) = \lambda L(t)$$

The epidemiological compartments



$$\dot{S}(t) = -\beta \frac{S(t)}{N} I(t)$$

$$\dot{I}(t) = \beta \frac{S(t)}{N} I(t) - \mu I(t) - \nu I(t)$$

$$\dot{R}(t) = \mu I(t)$$

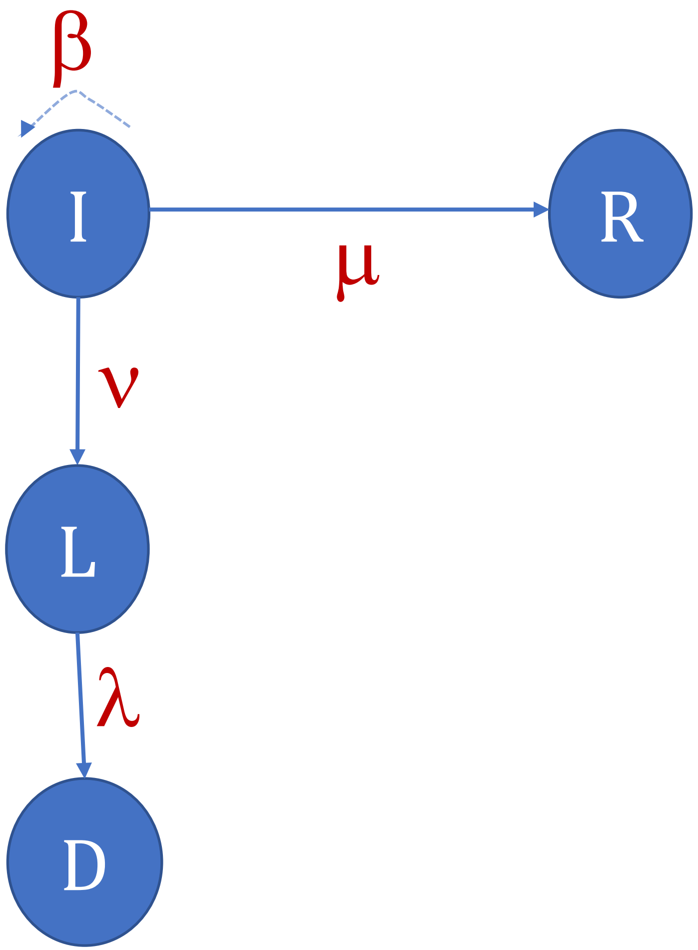
$$\dot{L}(t) = \nu I(t) - \lambda L(t)$$

$$\dot{D}(t) = \lambda L(t)$$

Approximation : $S(t) = N$

$(I(t)/S(t) < 0.01)$

The epidemiological compartments



$$\dot{I}(t) = \beta I(t) - \mu I(t) - \nu I(t)$$

$$\dot{R}(t) = \mu I(t)$$

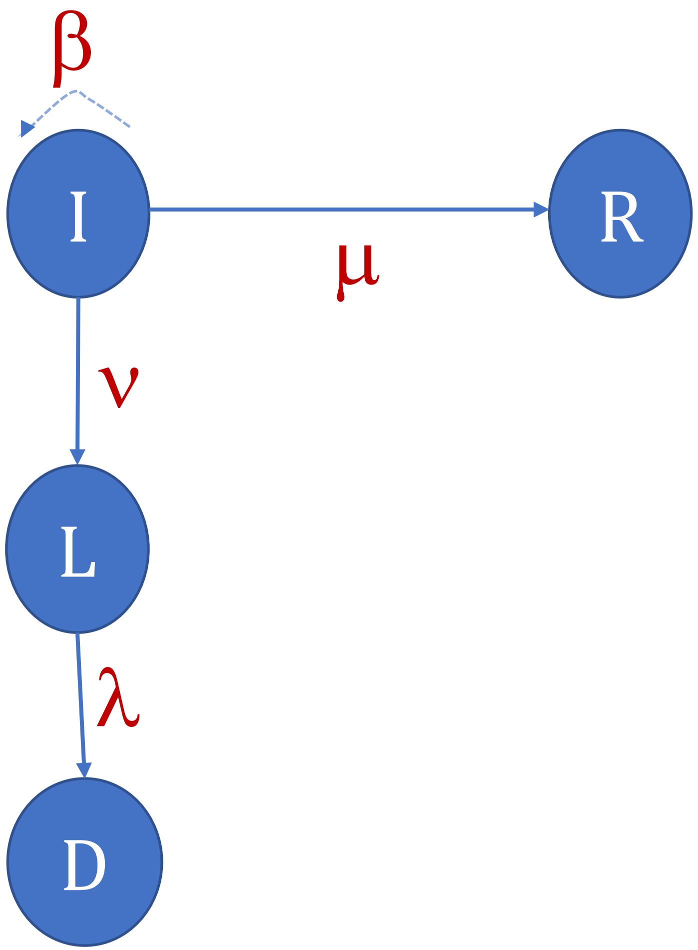
$$\dot{L}(t) = \nu I(t) - \lambda L(t)$$

$$\dot{D}(t) = \lambda L(t)$$

Approximation : $S(t) = N$

$(I(t)/S(t) < 0.01)$

The epidemiological compartments



$$\dot{I}(t) = \beta I(t) - \mu I(t) - \nu I(t)$$

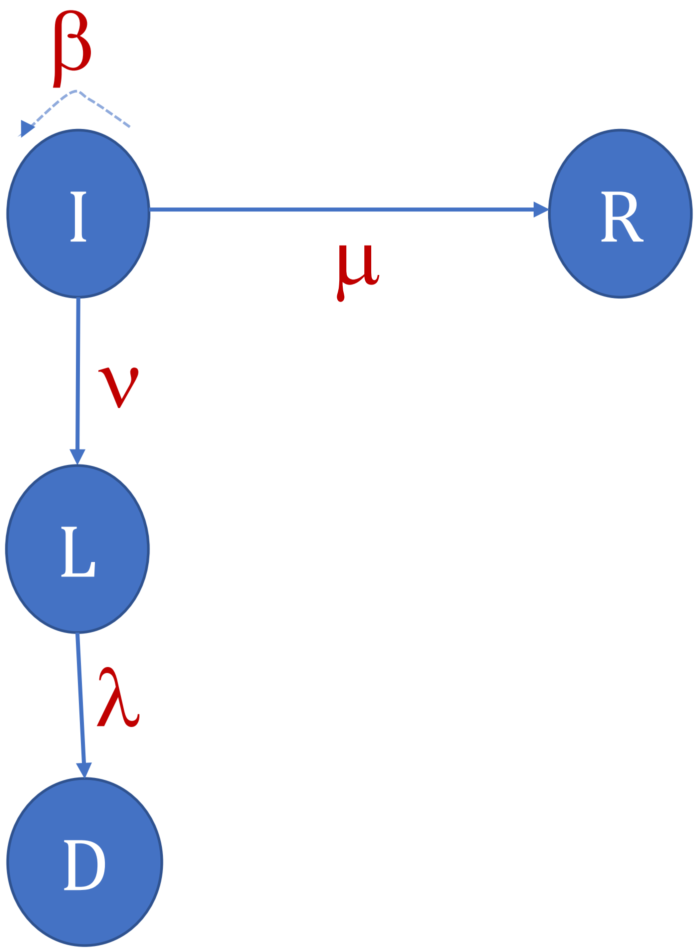
$$\dot{R}(t) = \mu I(t)$$

$$\dot{L}(t) = \nu I(t) - \lambda L(t)$$

$$\dot{D}(t) = \lambda L(t)$$

R not needed for fitting the data

The epidemiological compartments



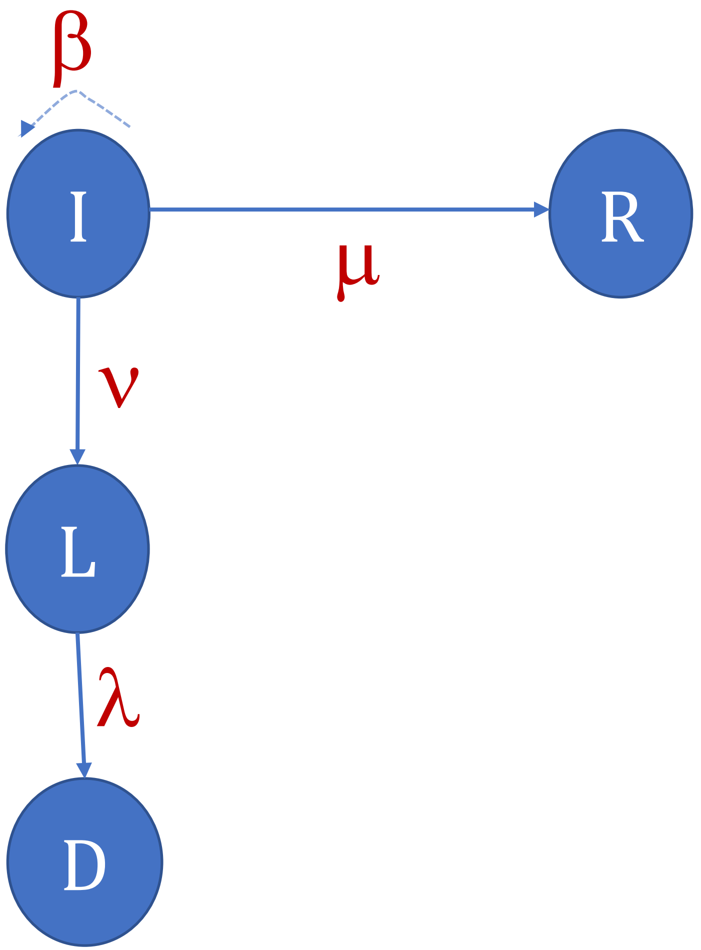
$$\dot{I}(t) = \beta I(t) - \mu I(t) - \nu I(t)$$

$$\dot{L}(t) = \nu I(t) - \lambda L(t)$$

$$\dot{D}(t) = \lambda L(t)$$

R not needed for fitting the data

The epidemiological compartments



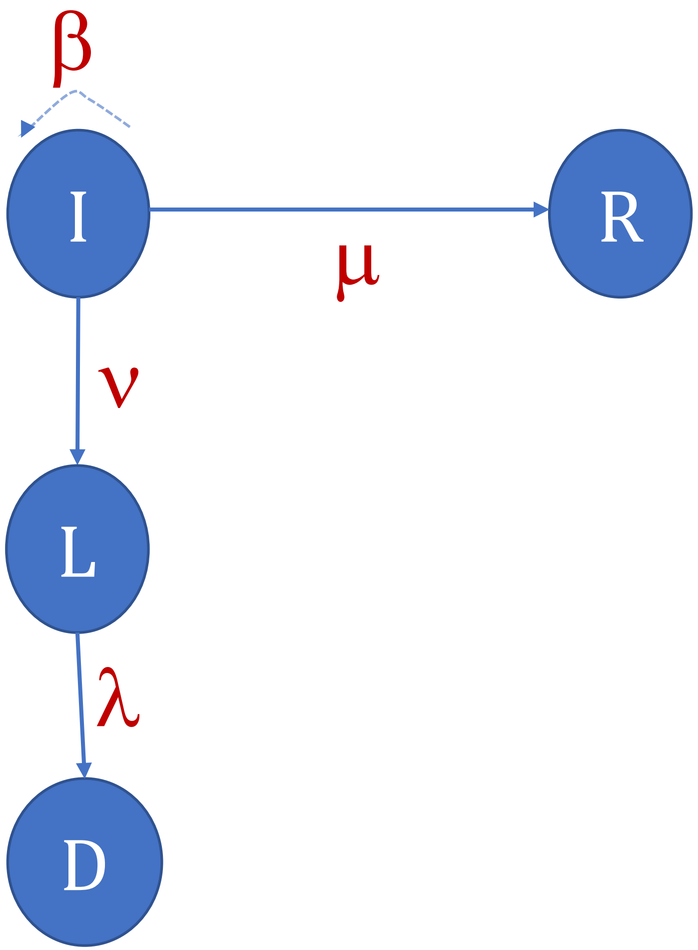
$$\dot{I}(t) = \beta I(t) - \mu I(t) - \nu I(t)$$

$$\dot{L}(t) = \nu I(t) - \lambda L(t)$$

$$\dot{D}(t) = \lambda L(t)$$

The transmission rate β changes over time

The epidemiological compartments



$$\dot{I}(t) = \beta(t)I(t) - \mu I(t) - \nu I(t)$$

$$\dot{L}(t) = \nu I(t) - \lambda L(t)$$

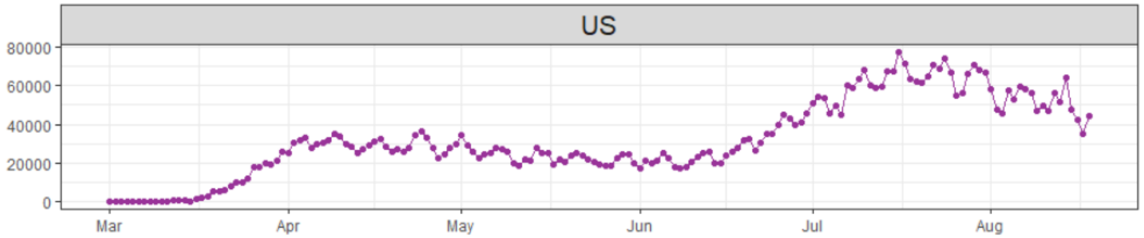
$$\dot{D}(t) = \lambda L(t)$$

The transmission rate β changes over time

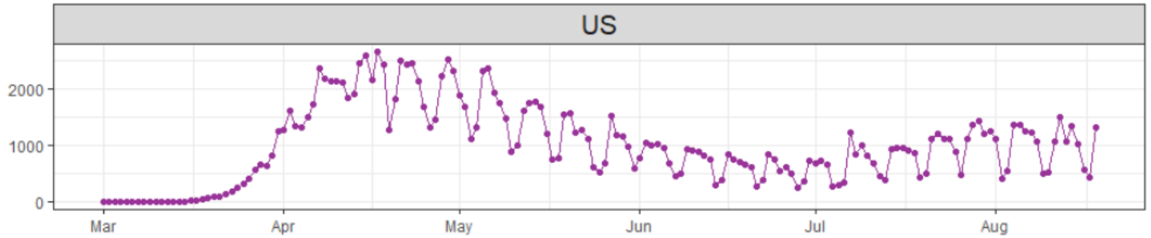
The observation model

Available data:

- $w = (w_j\ , j = 1, 2, \ldots)$ where w_j is the number of **new** confirmed cases on day j



- $d = (d_j\ , j = 1, 2, \ldots)$ where d_j is the number of **new** deaths on day j



The observation model

The data:

- daily number of *new confirmed cases* (w_j)
- daily number of *new deaths* (d_j)

The epidemiological model:

$$\dot{I}(t) = \beta(t)I(t) - \mu I(t) - \nu I(t)$$

$$\dot{L}(t) = \nu I(t) - \lambda L(t)$$

$$\dot{D}(t) = \lambda L(t)$$

The observation model

The data:

- daily number of *new confirmed cases* (w_j)
- daily number of *new deaths* (d_j)

We assume that only a fraction $\alpha(t)$ of the infected people are *confirmed* at time *t*

The epidemiological model:

$$\dot{I}(t) = \beta(t)I(t) - \mu I(t) - \nu I(t)$$

$$\dot{L}(t) = \nu I(t) - \lambda L(t)$$

$$\dot{D}(t) = \lambda L(t)$$

The observation model

The data:

- daily number of *new confirmed cases* (w_j)
- daily number of *new deaths* (d_j)

We assume that only a fraction $\alpha(t)$ of the infected people are *confirmed* at time t

$\alpha(t)$ depends on the number of tests performed at time t

The epidemiological model:

$$\dot{I}(t) = \beta(t)I(t) - \mu I(t) - \nu I(t)$$

$$\dot{W}_c(t) = \alpha(t)\beta(t)I(t)$$

$$\dot{L}(t) = \nu I(t) - \lambda L(t)$$

$$\dot{D}(t) = \lambda L(t)$$

The observation model

The data:

- daily number of *new confirmed cases* (w_j)

w_j predicted by $W_c(t_j) - W_c(t_{j-1})$

- daily number of *new deaths* (d_j)

d_j predicted by $D(t_j) - D(t_{j-1})$

The epidemiological model:

$$\dot{I}(t) = \beta(t)I(t) - \mu I(t) - \nu I(t)$$

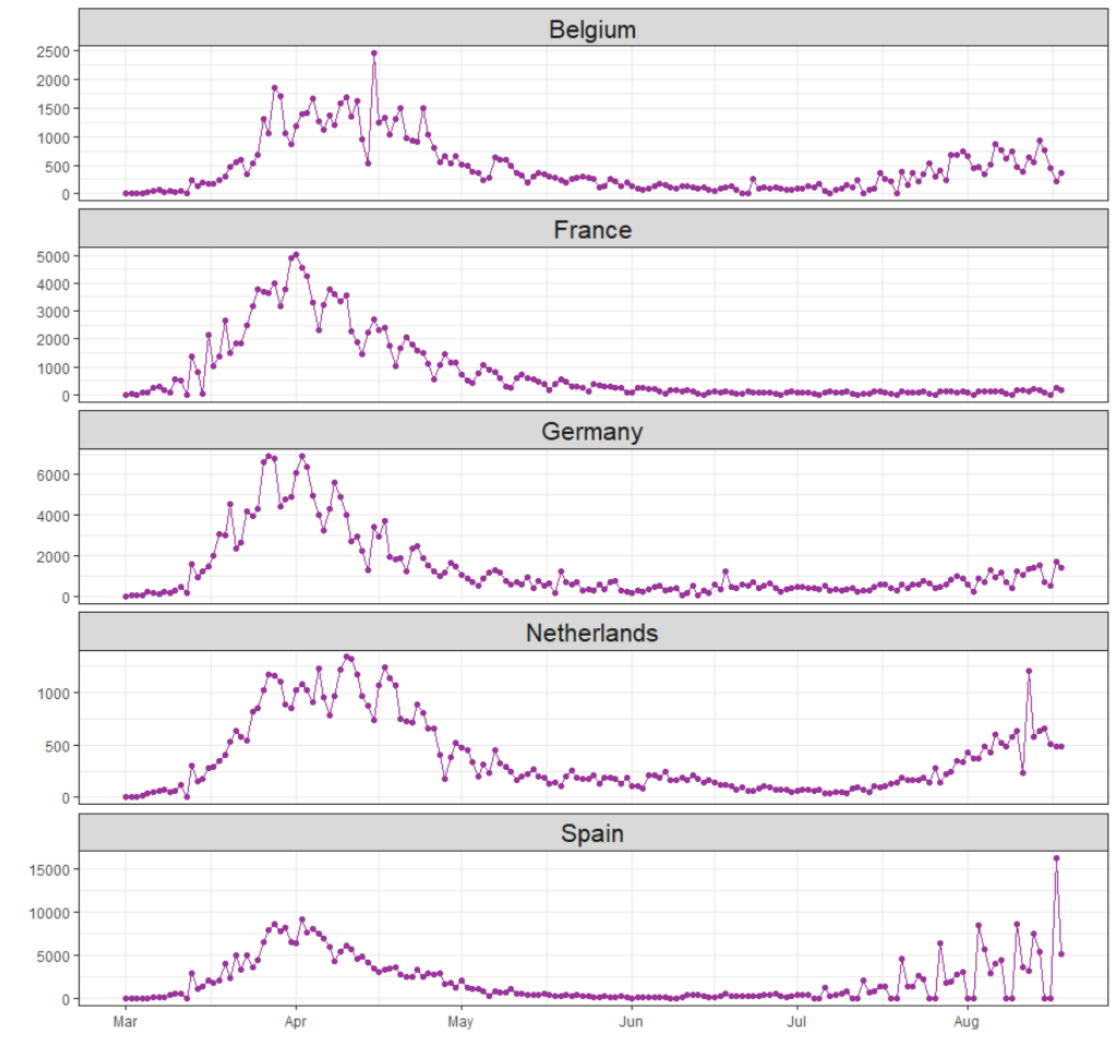
$$\dot{W}_c(t) = \alpha(t)\beta(t)I(t)$$

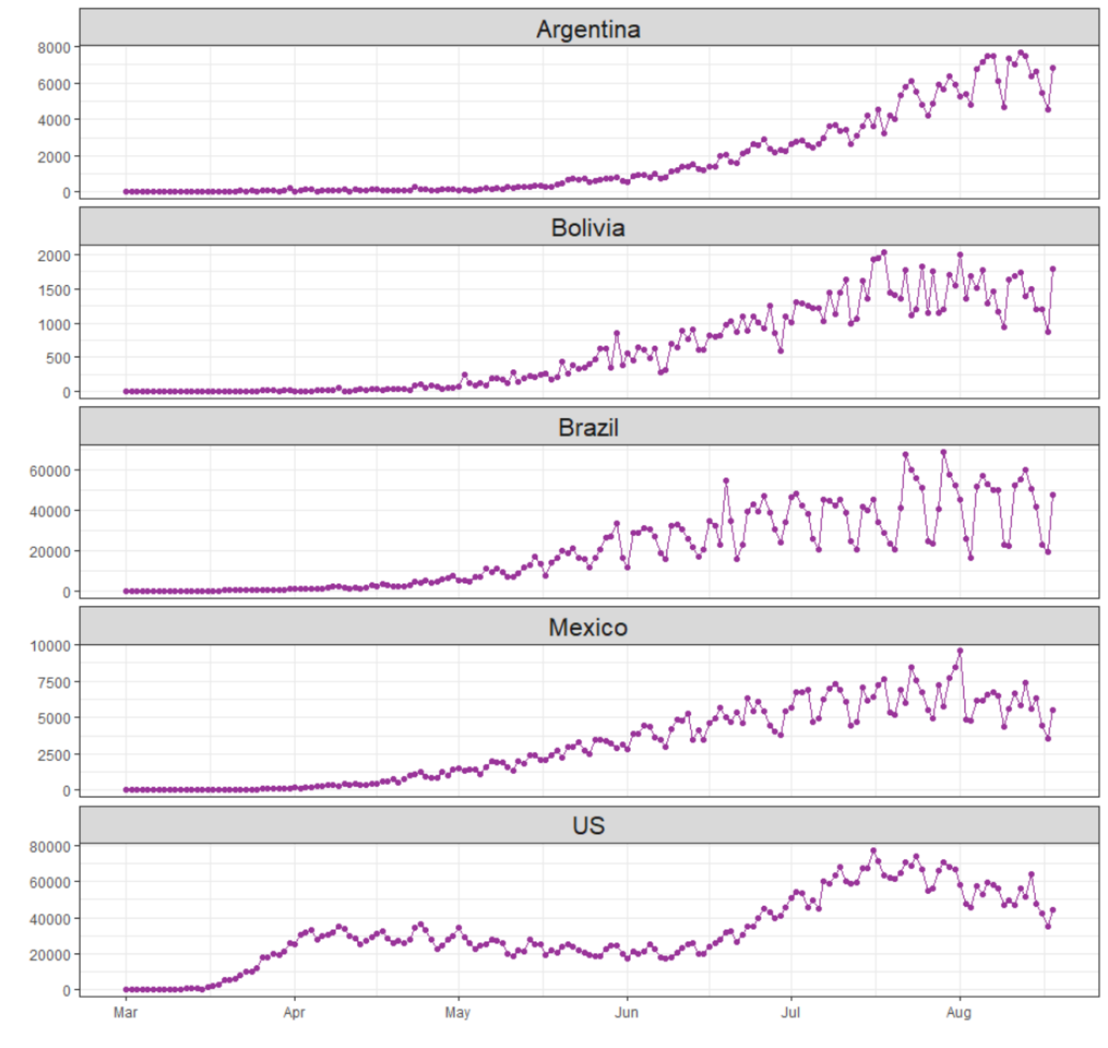
$$\dot{L}(t) = \nu I(t) - \lambda L(t)$$

$$\dot{D}(t) = \lambda L(t)$$

Daily data clearly exhibit a weekly periodic component

Daily number of confirmed cases





The observation model

A statistical model for the daily counts:

$$w_j = (W(t_j) - W(t_{j-1}))\left(1 + A\cos(\frac{2\pi}{7}t_j + \phi)\right)(1 + e_j)$$
$$e_j \sim \mathcal{N}(0, \sigma_e^2)$$

$$d_j = (D(t_j) - D(t_{j-1}))\left(1 + B\cos(\frac{2\pi}{7}t_j + \phi)\right)(1 + u_j)$$
$$u_j \sim \mathcal{N}(0, \sigma_u^2)$$

The epidemiological model:

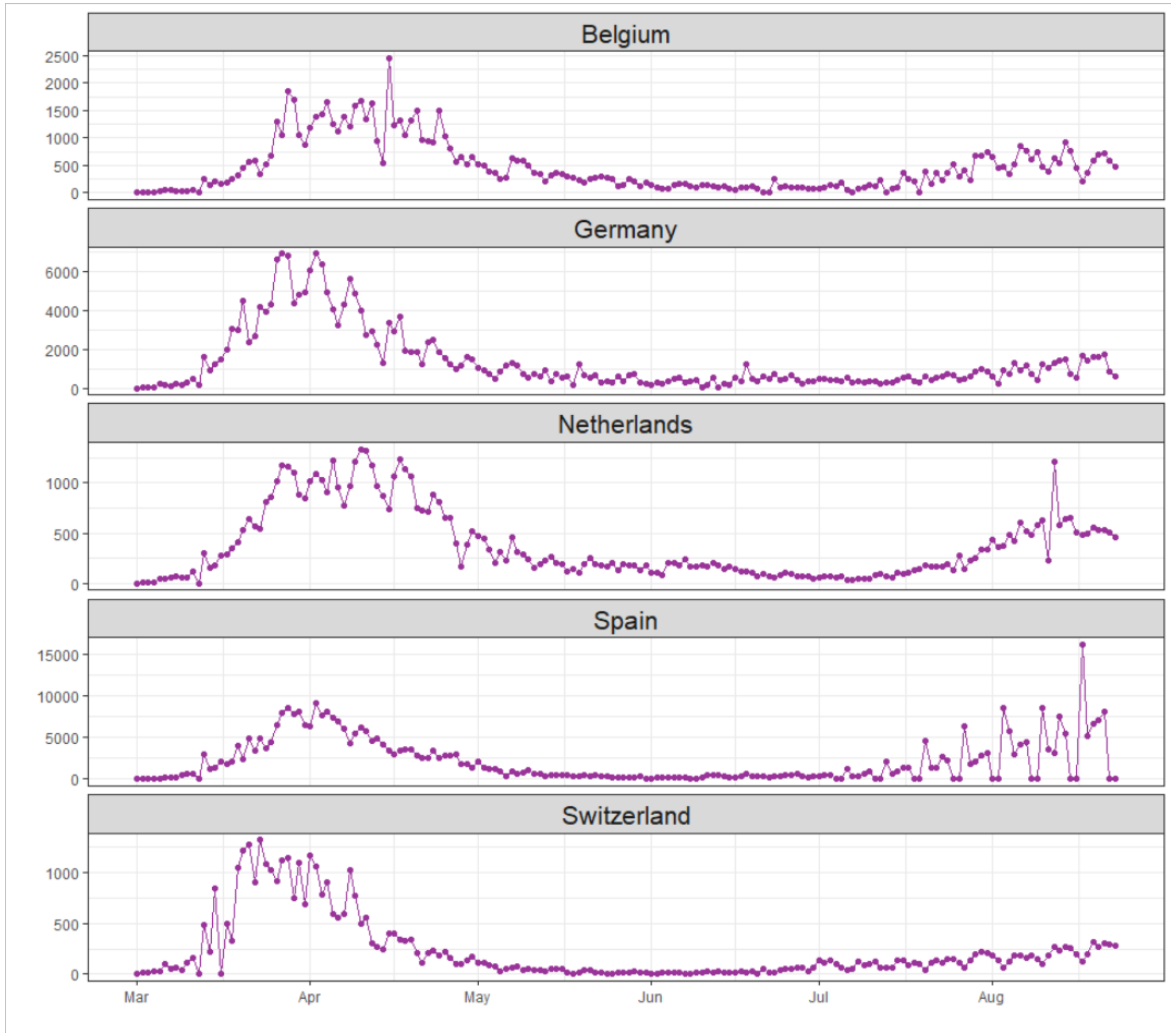
$$\dot{I}(t) = \beta(t)I(t) - \mu I(t) - \nu I(t)$$

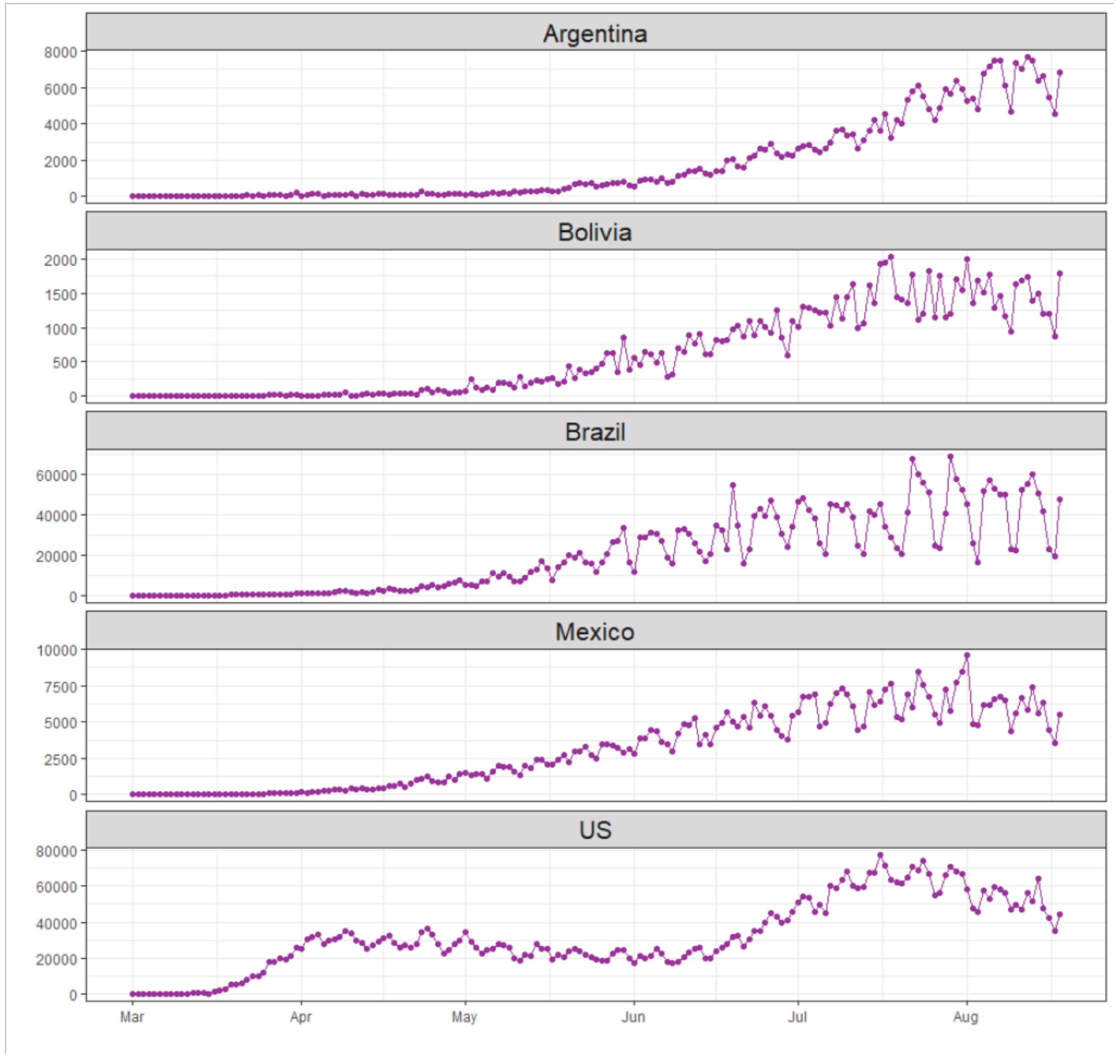
$$\dot{W}_c(t) = \alpha(t)\beta(t)I(t)$$

$$\dot{L}(t) = \nu I(t) - \lambda L(t)$$

$$\dot{D}(t) = \lambda L(t)$$

The dynamics of the infection process seems to have changed from July onwards





Daily number of confirmed cases

A model for the “first wave”
(before July)

The epidemiological model:

$$\dot{I}(t) = \beta(t)I(t) - \mu I(t) - \nu I(t)$$

$$\dot{W}_c(t) = \alpha(t)\beta(t)I(t)$$

$$\dot{L}(t) = \nu I(t) - \lambda L(t)$$

$$\dot{D}(t) = \lambda L(t)$$

- The transmission rate is a piecewise linear function

$$\beta(t) = \beta_0 + a\,t + \sum_{k=1}^{K} h_k\,(t - \tau_k) \times \mathbb{1}\{t \geq \tau_k\}$$

- The fraction of confirmed cases is constant over time

$$\alpha(t) = \alpha$$

A model for the "first wave"
(before July)

The epidemiological model:

$$\dot{I}(t) = \beta(t)I(t) - \mu I(t) - \nu I(t)$$

$$\dot{W}_c(t) = \alpha(t)\beta(t)I(t)$$

$$\dot{L}(t) = \nu I(t) - \lambda L(t)$$

$$\dot{D}(t) = \lambda L(t)$$

- The transmission rate is a piecewise linear function

$$\beta(t) = \beta_0 + a\,t + \sum_{k=1}^{K} h_k\,(t - \tau_k) \times \mathbb{1}\{t \geq \tau_k\}$$

- The fraction of confirmed cases is constant over time

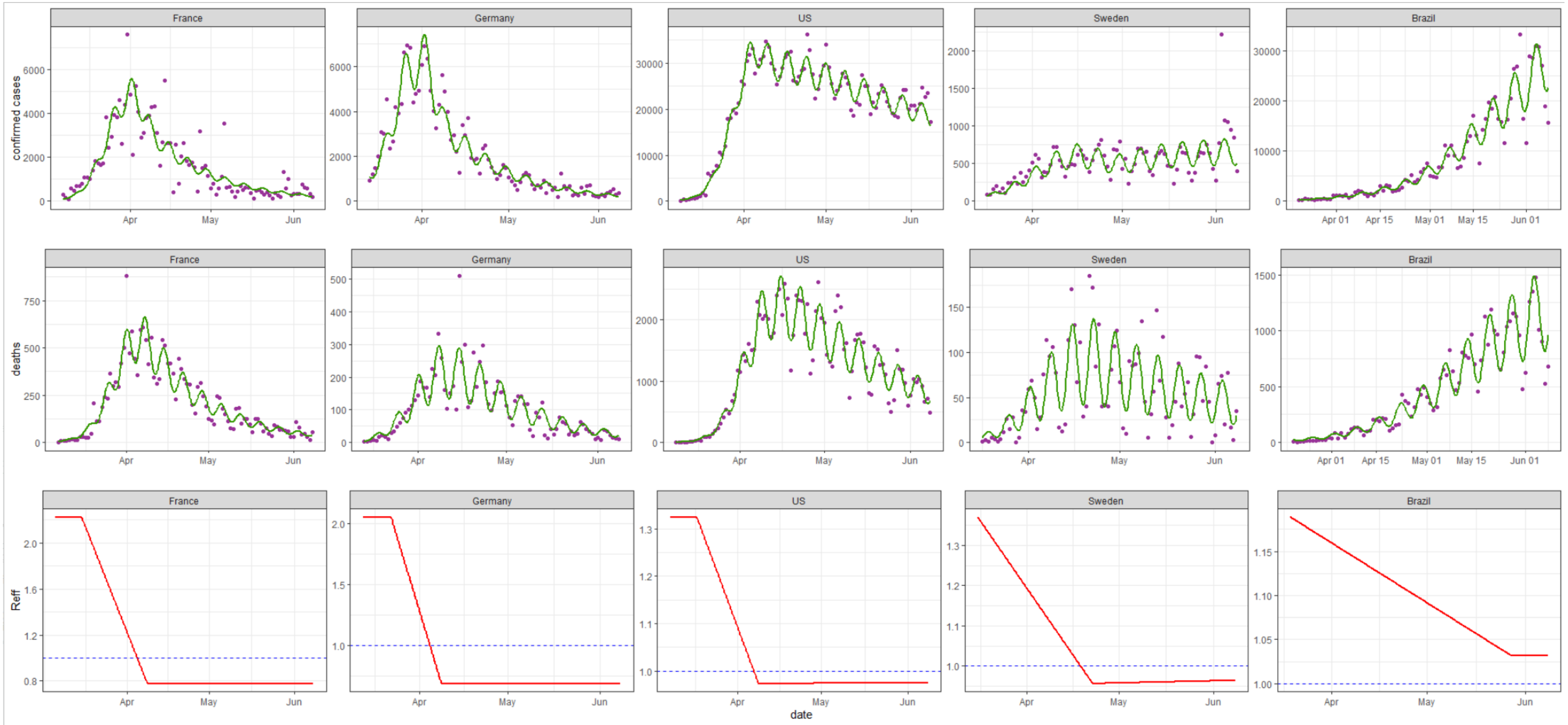
$$\alpha(t) = \alpha$$

Parameters of the model: $\theta = (\alpha, \beta_0, a, h_1, \ldots, h_K, \tau_1, \ldots, \tau_K, \mu, \nu, \lambda, I_0, L_0, D_0, \sigma_e^2, \sigma_u^2)$

θ obtained by Maximum Likelihood (ML) Estimation

K obtained by minimizing the Bayesian Information Criteria (BIC)

Some fits (with the periodic component)

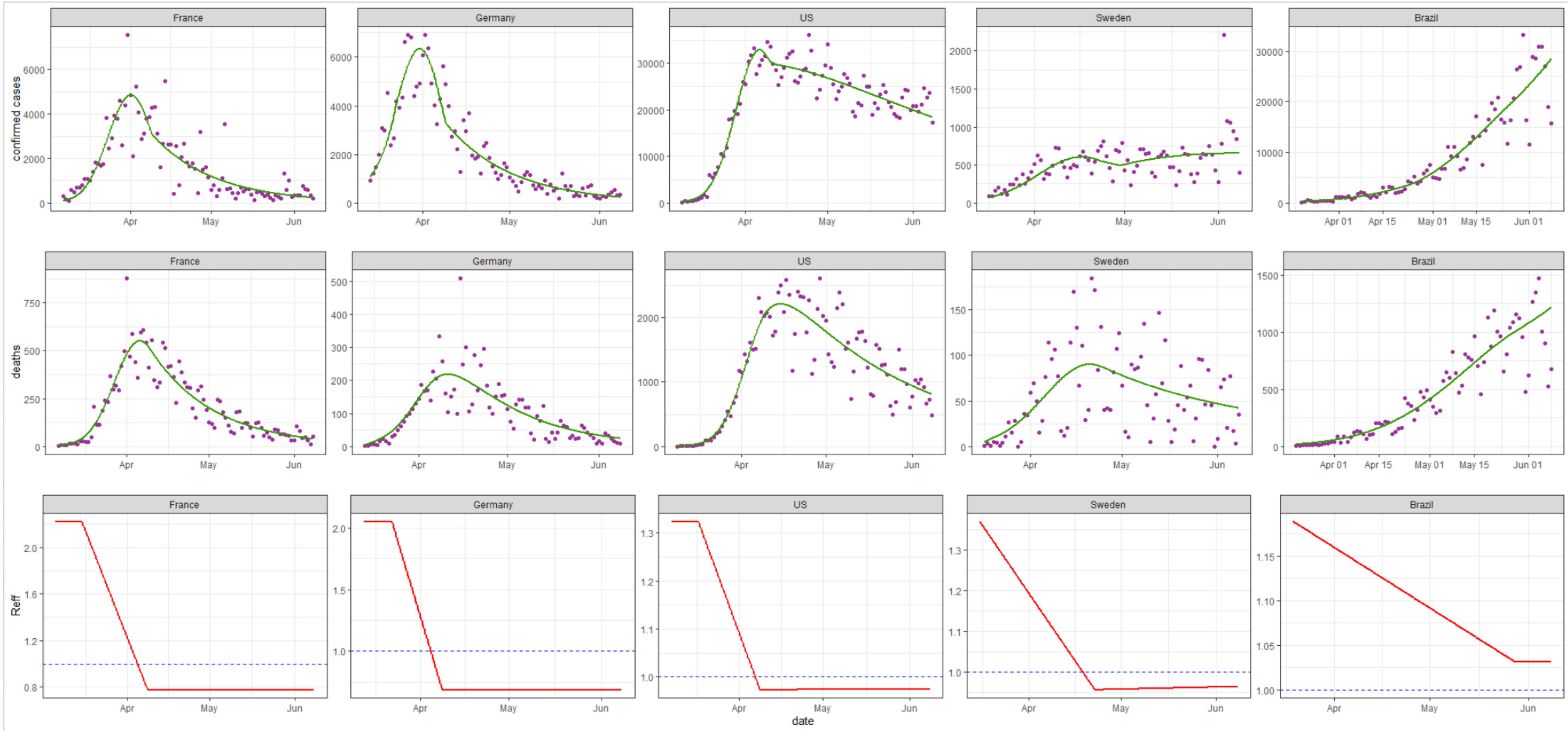


Basic reproduction number:

(expected number of cases generated by one case in a population where all individuals are susceptible to infection)

$$R_0 = \frac{\beta(t)}{\mu + \nu}$$

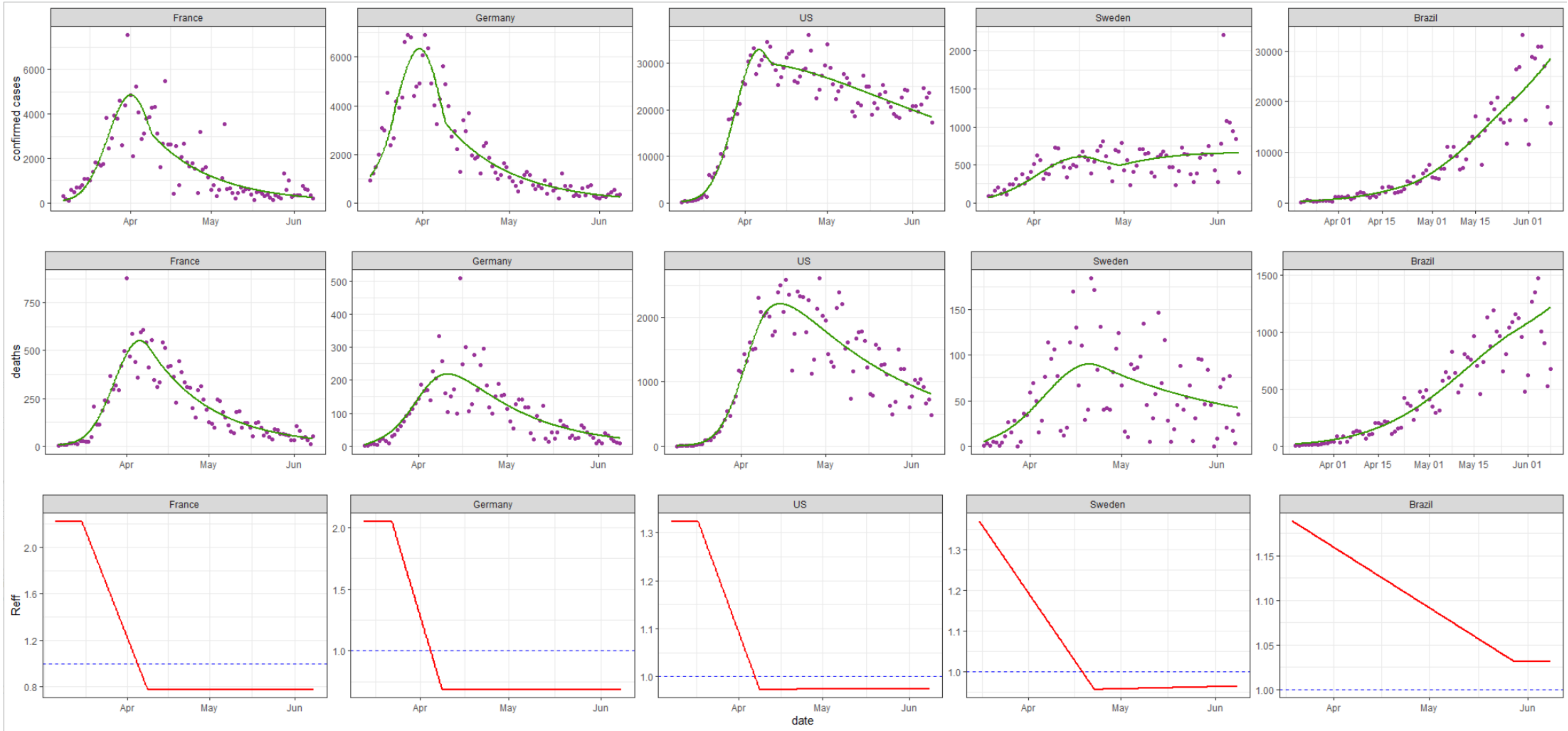
Some fits (without the periodic component)



Basic reproduction number:

(expected number of cases generated by one case in a population where all individuals are susceptible to infection)

$$R_0 = \frac{\beta(t)}{\mu + \nu}$$



t2	2.9	4.7	2.6	4.4	6.9
t1/2	16.1	15.7	31.4	44.1	-

A monitoring tool

This tool can be useful for analyzing "unexpected" changes in the dynamics of the epidemics.

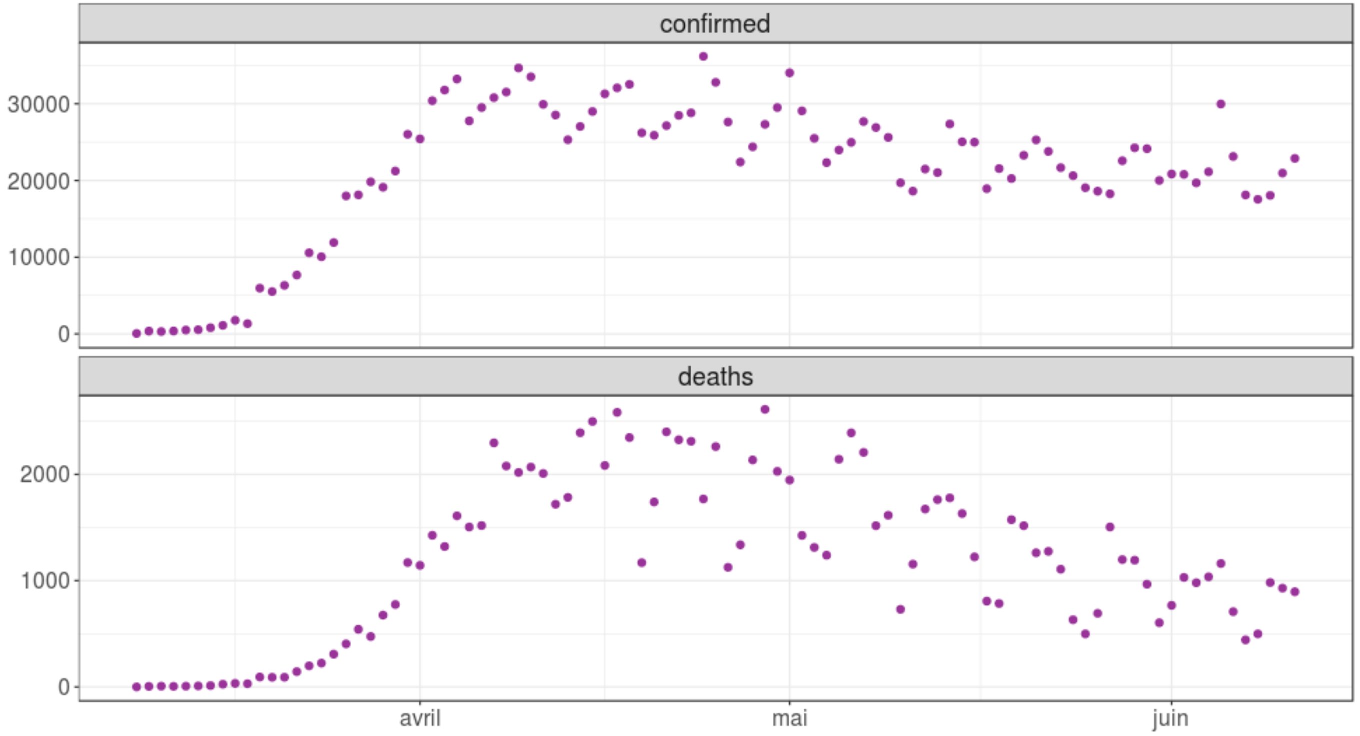
› Actualité › Fiches › Guide Vie quotidienne › Coronavirus

Coronavirus dans le monde : hausse inquiétante des décès aux USA, les chiffres

La Rédaction, Mis à jour le 11/06/20 17:35

Partager sur Facebook | Twitter | Email

CORONAVIRUS. Le nombre de contaminations et de morts liés au Covid-19 semble repartir à la hausse aux Etats-Unis, selon le dernier bilan en date. De nombreux autres pays craignent ou constatent une résurgence du nombre de cas. Le point sur la pandémie dans le monde.



A monitoring tool

This tool can be useful for analyzing “unexpected” changes in the dynamics of the epidemics.

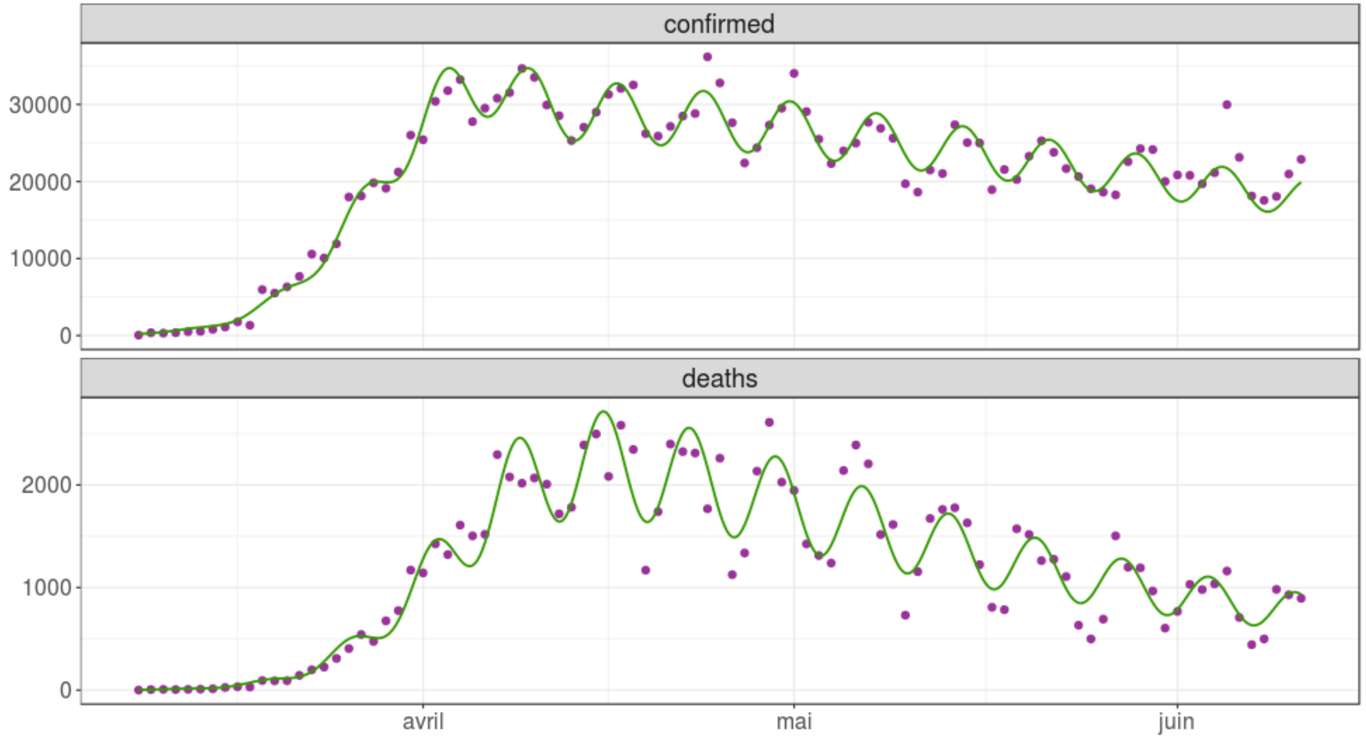
Actualité › Fiches › Guide Vie quotidienne › Coronavirus

Coronavirus dans le monde : hausse inquiétante des décès aux USA, les chiffres

La Rédaction, Mis à jour le 11/06/20 17:35

Partager sur Facebook | Twitter | Email

CORONAVIRUS. Le nombre de contaminations et de morts liés au Covid-19 semble repartir à la hausse aux Etats-Unis, selon le dernier bilan en date. De nombreux autres pays craignent ou constatent une résurgence du nombre de cas. Le point sur la pandémie dans le monde.



A monitoring tool

This tool can be useful for analyzing “unexpected” changes in the dynamics of the epidemics.

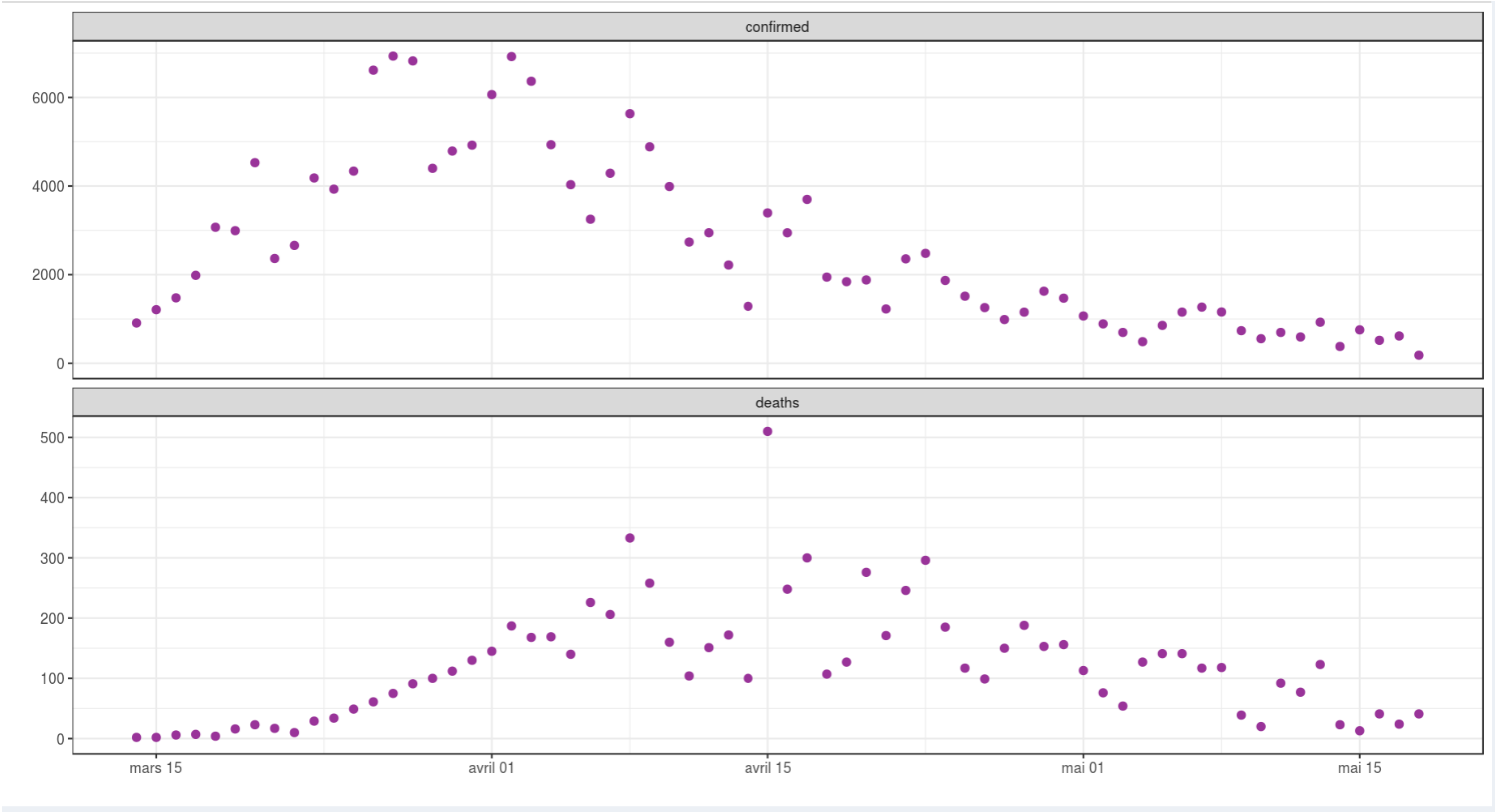
HOME BREAKING NEWS BUSINESS SPORTS HEALTH MIX

Home › Breaking News › Notable rebound in Germany of new cases and deaths from coronavirus

Breaking News

Notable rebound in Germany of new cases and deaths from coronavirus

May 12, 2020



A monitoring tool

This tool can be useful for analyzing “unexpected” changes in the dynamics of the epidemics.


Web 24 NEWS

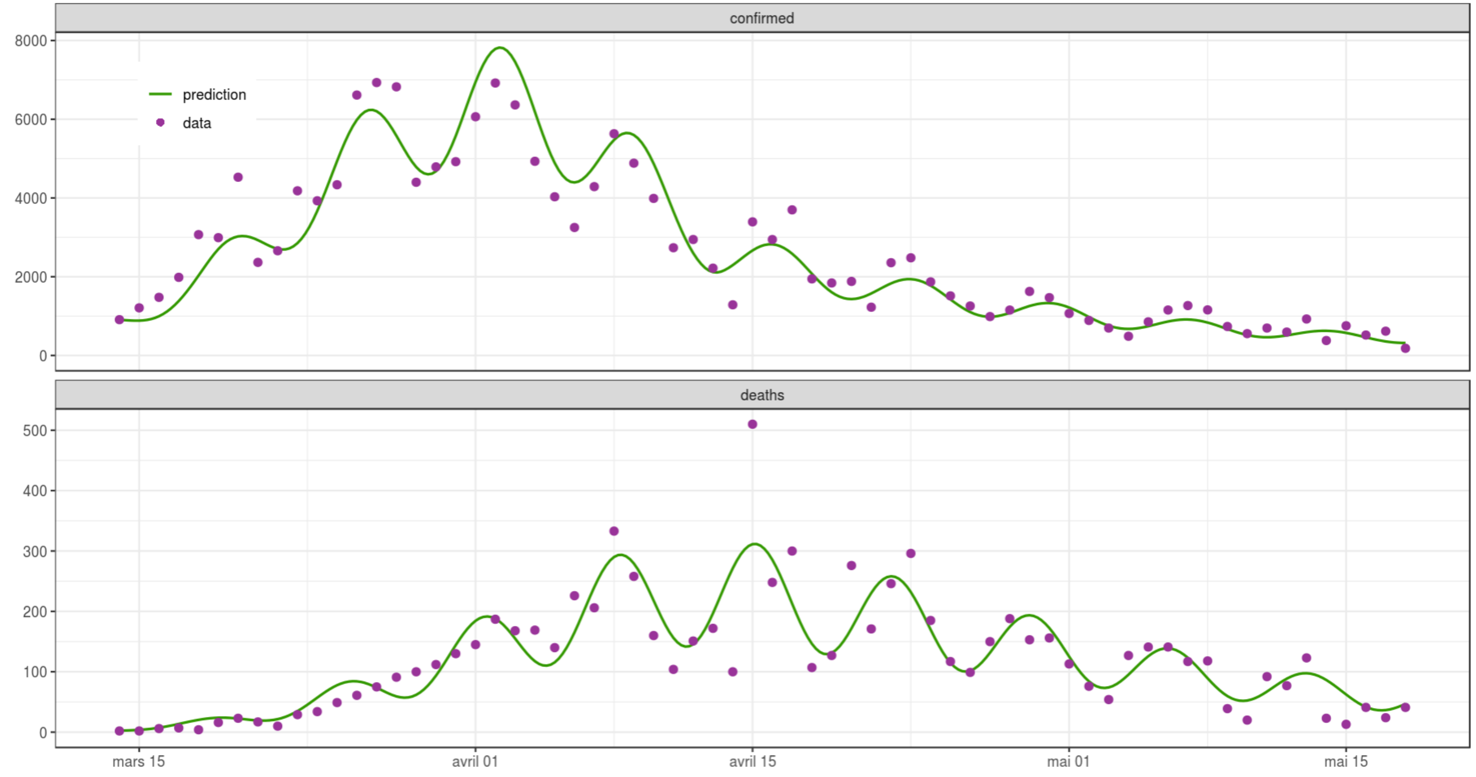
HOME BREAKING NEWS BUSINESS SPORTS HEALTH MIX

Home › Breaking News › Notable rebound in Germany of new cases and deaths from coronavirus

Breaking News

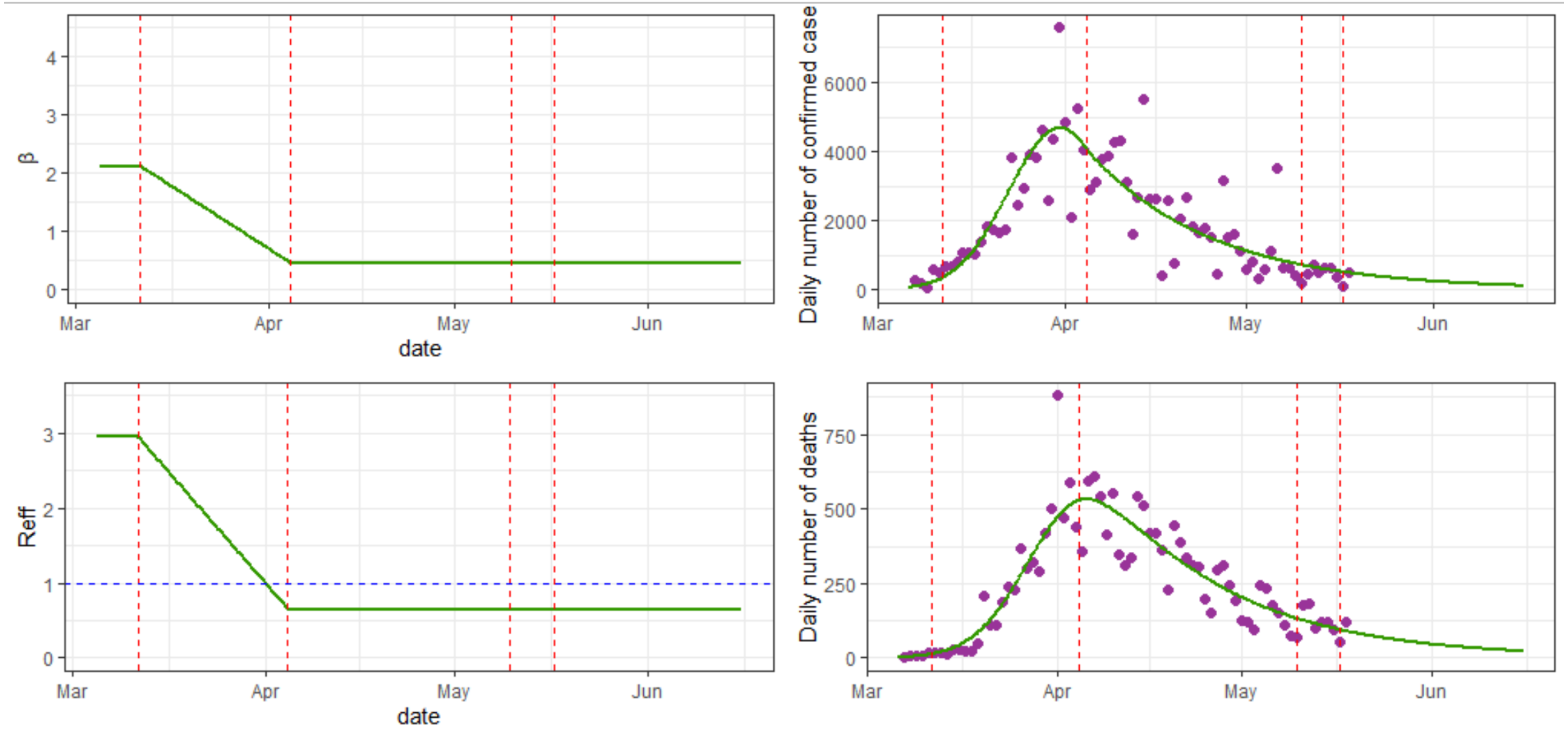
Notable rebound in Germany of new cases and deaths from coronavirus

May 12, 2020



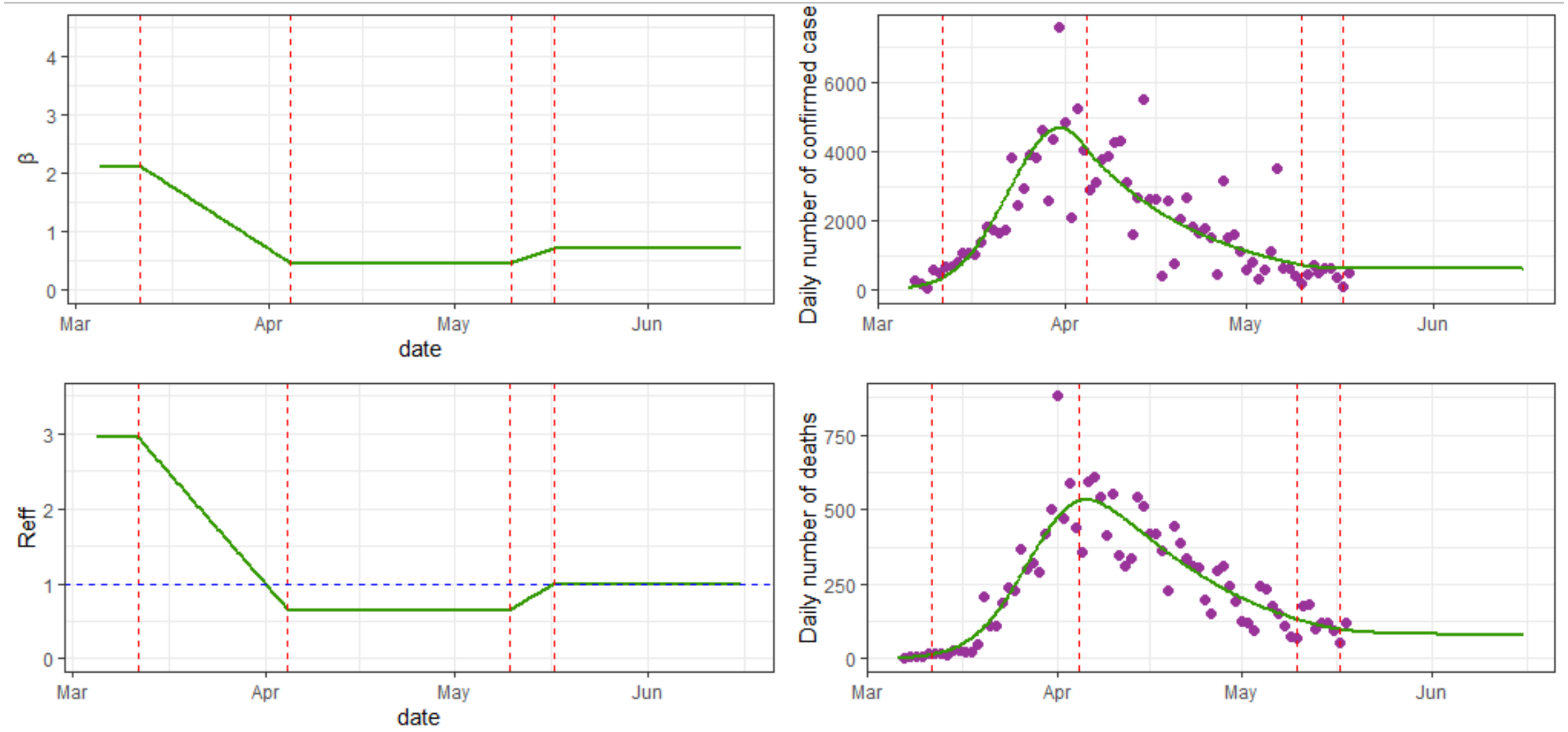
Possible scenarios after the end of the lockdown

1) The transmission rate remains the same



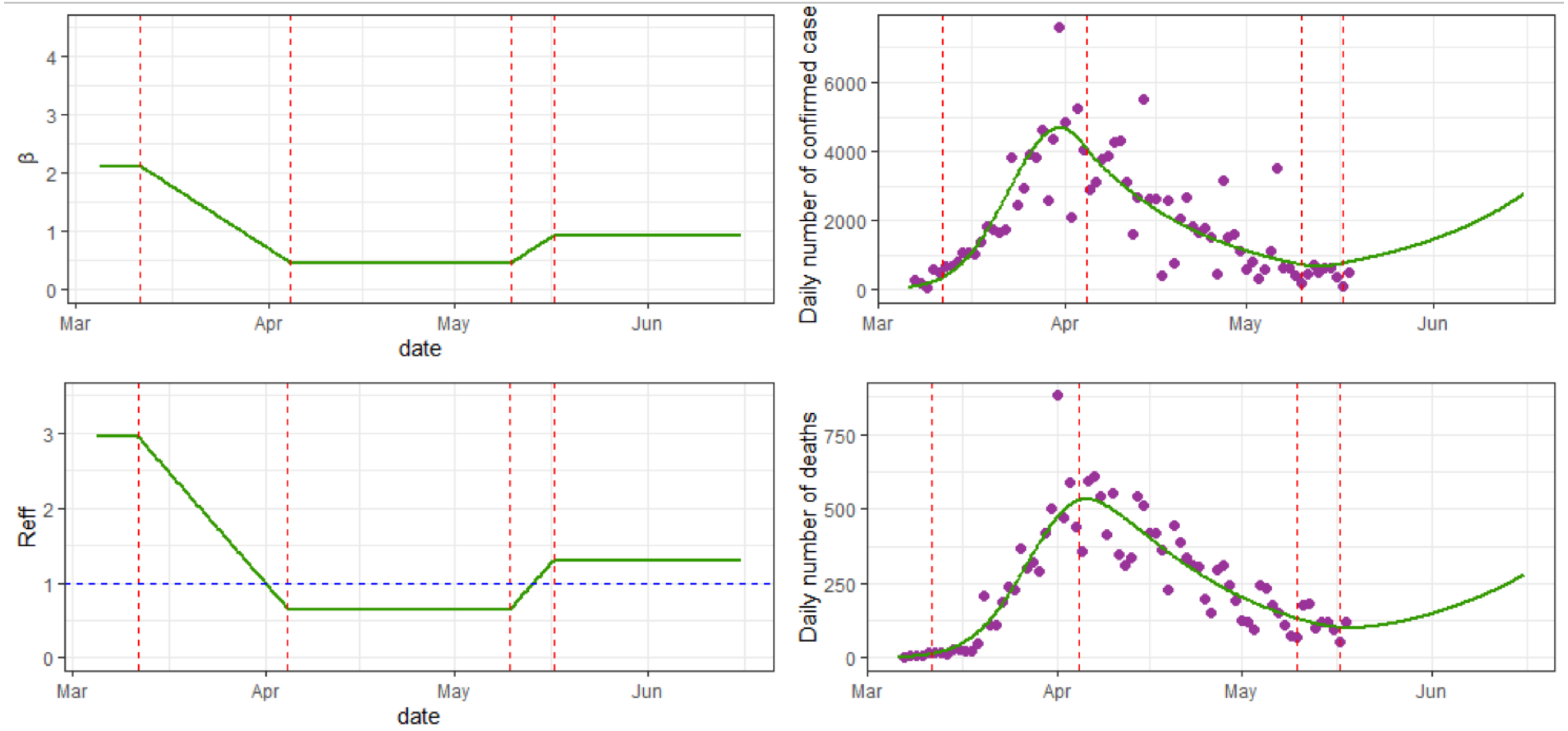
Possible scenarios after the end of the lockdown

2) The transmission rate is multiplied by 1.5



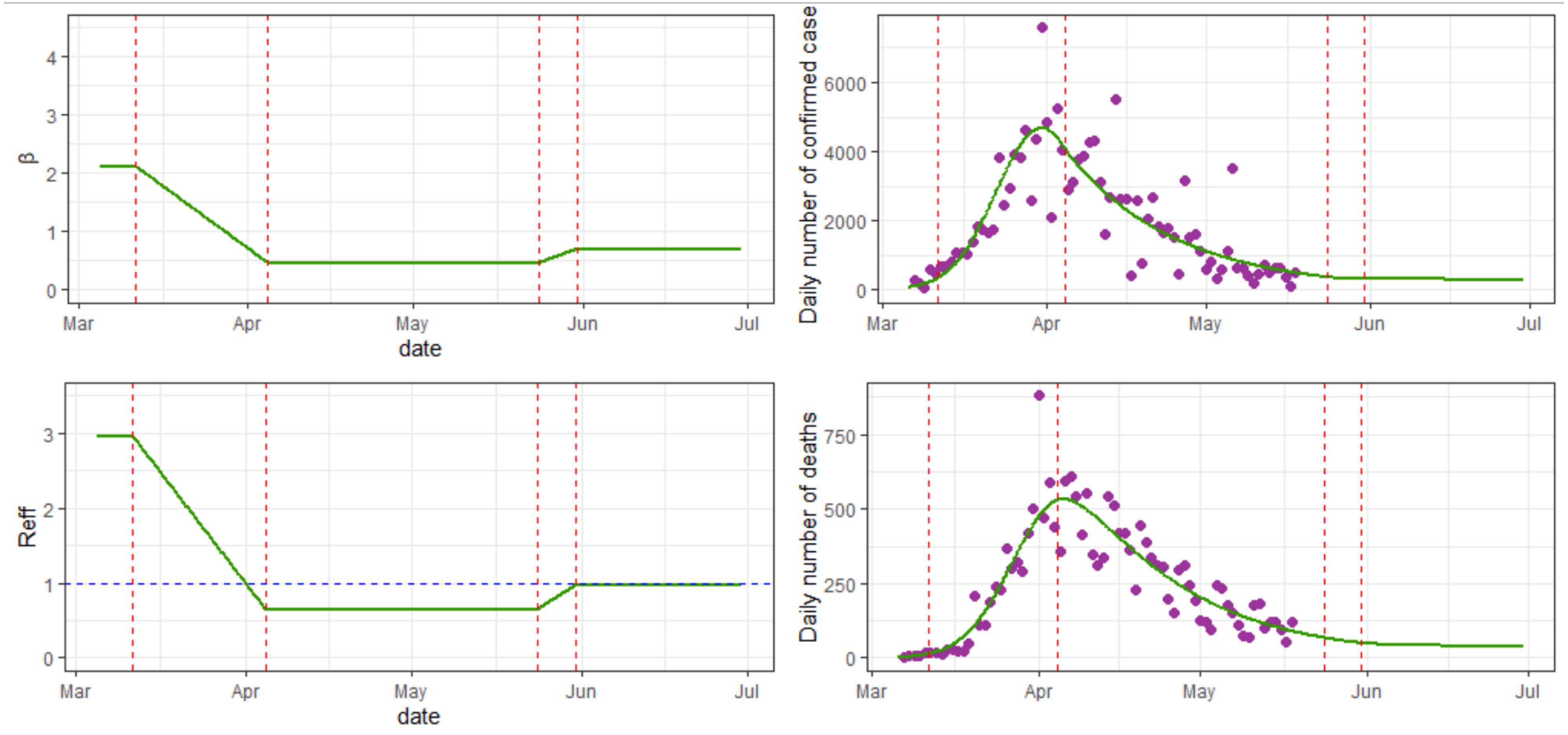
Possible scenarios after the end of the lockdown

3) The transmission rate is multiplied by 2



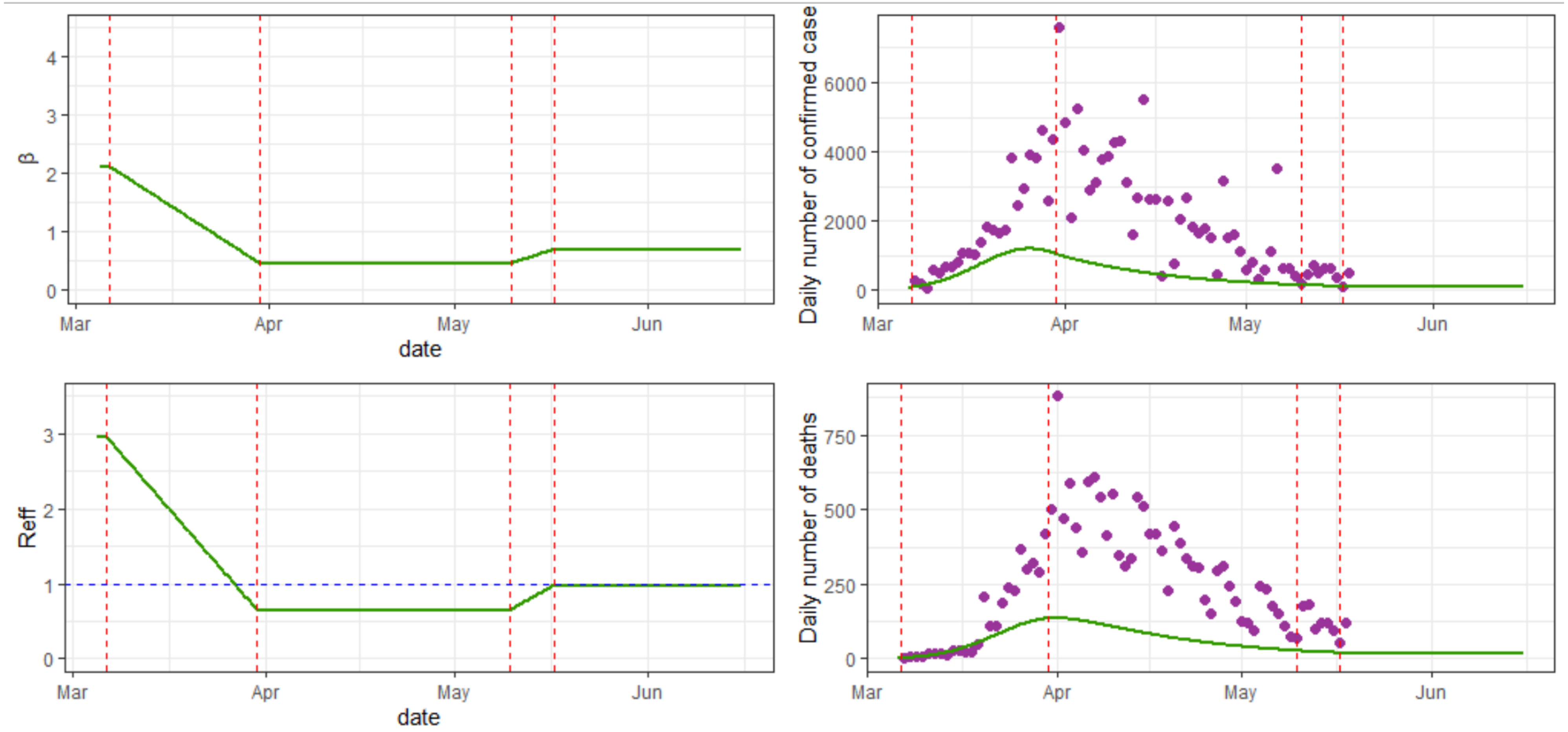
Possible scenarios after the end of the lockdown

4) The lockdown ends 2 weeks later (May 25)



Possible scenarios before/after the lockdown

5) The lockdown starts one week before (March 10)

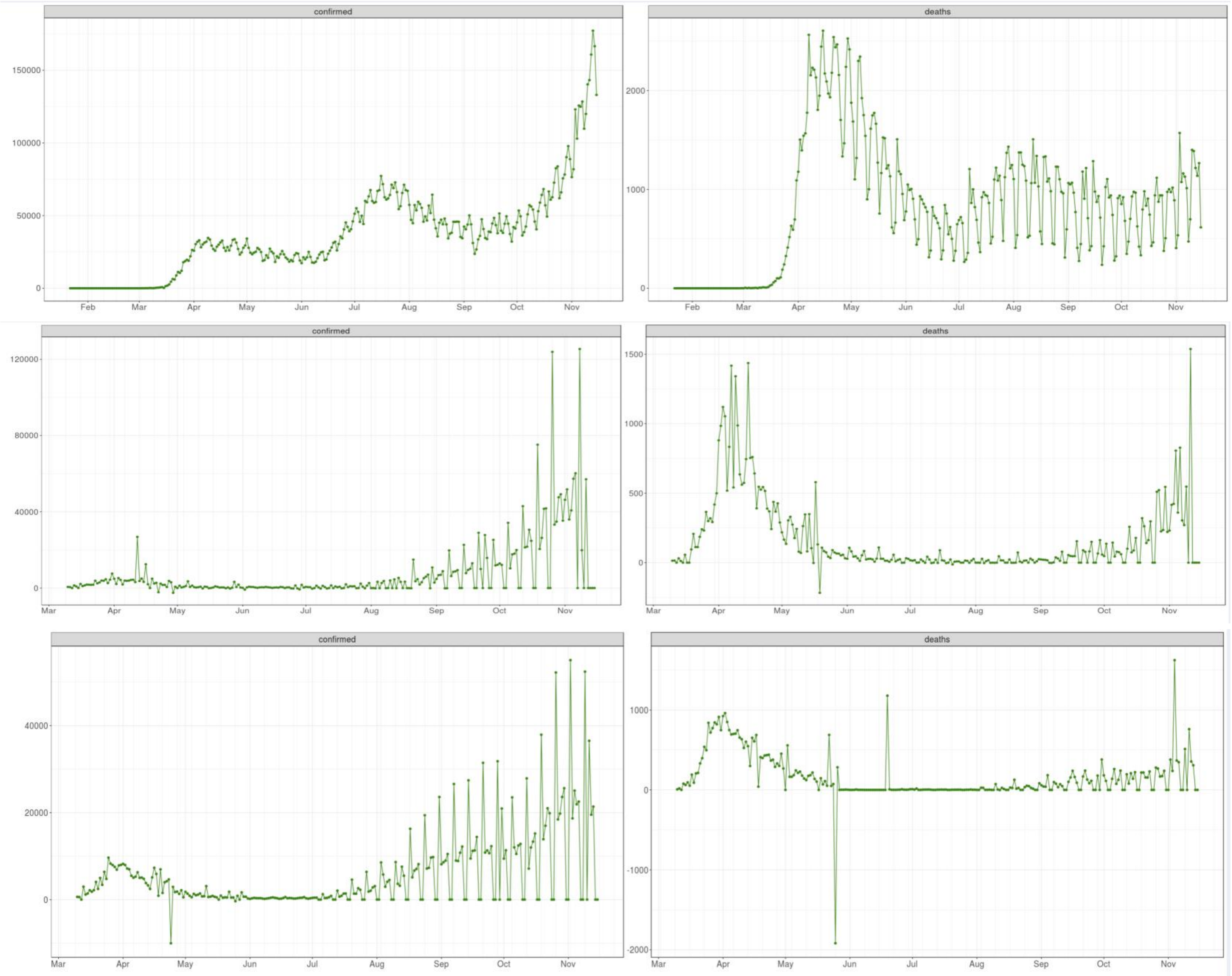


Things get more complicated from July on...

US

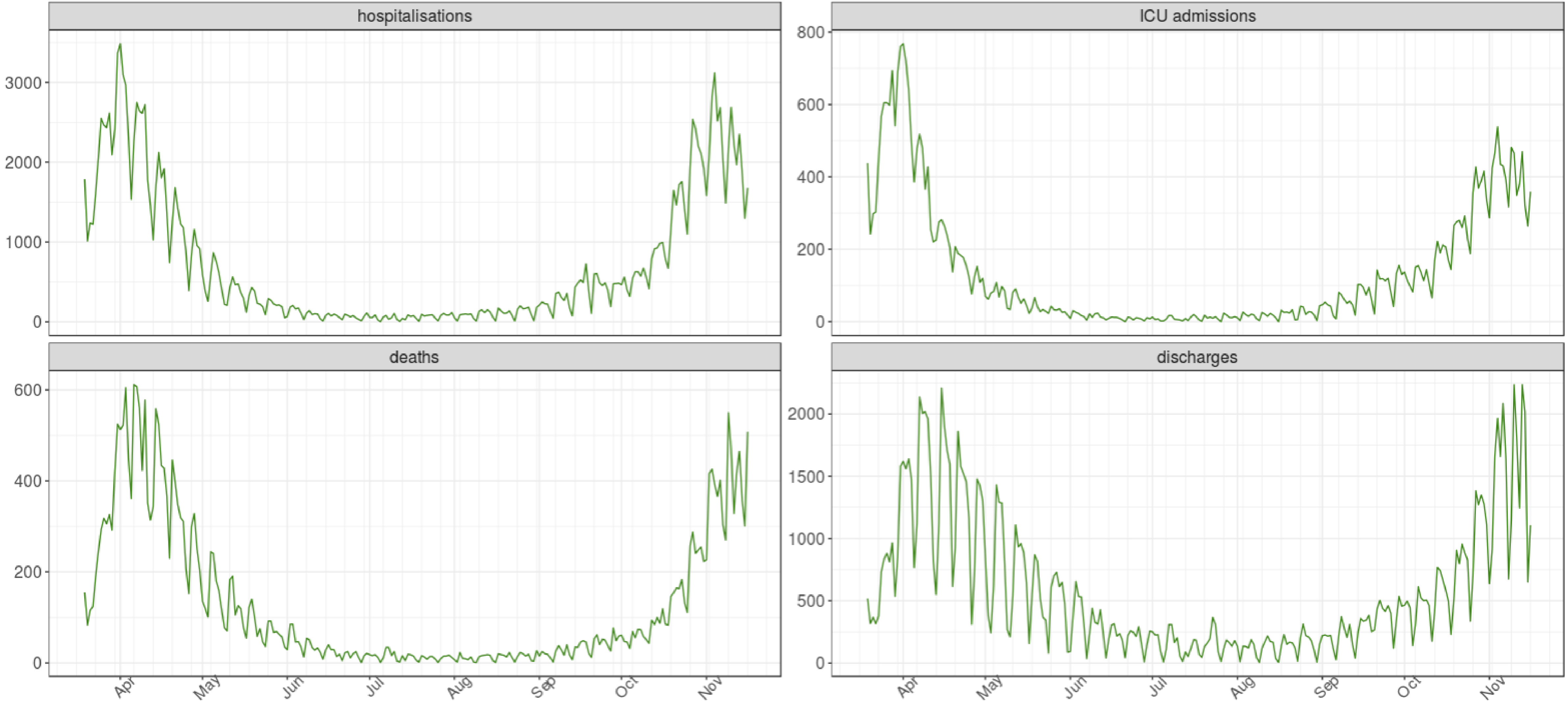
France

Spain

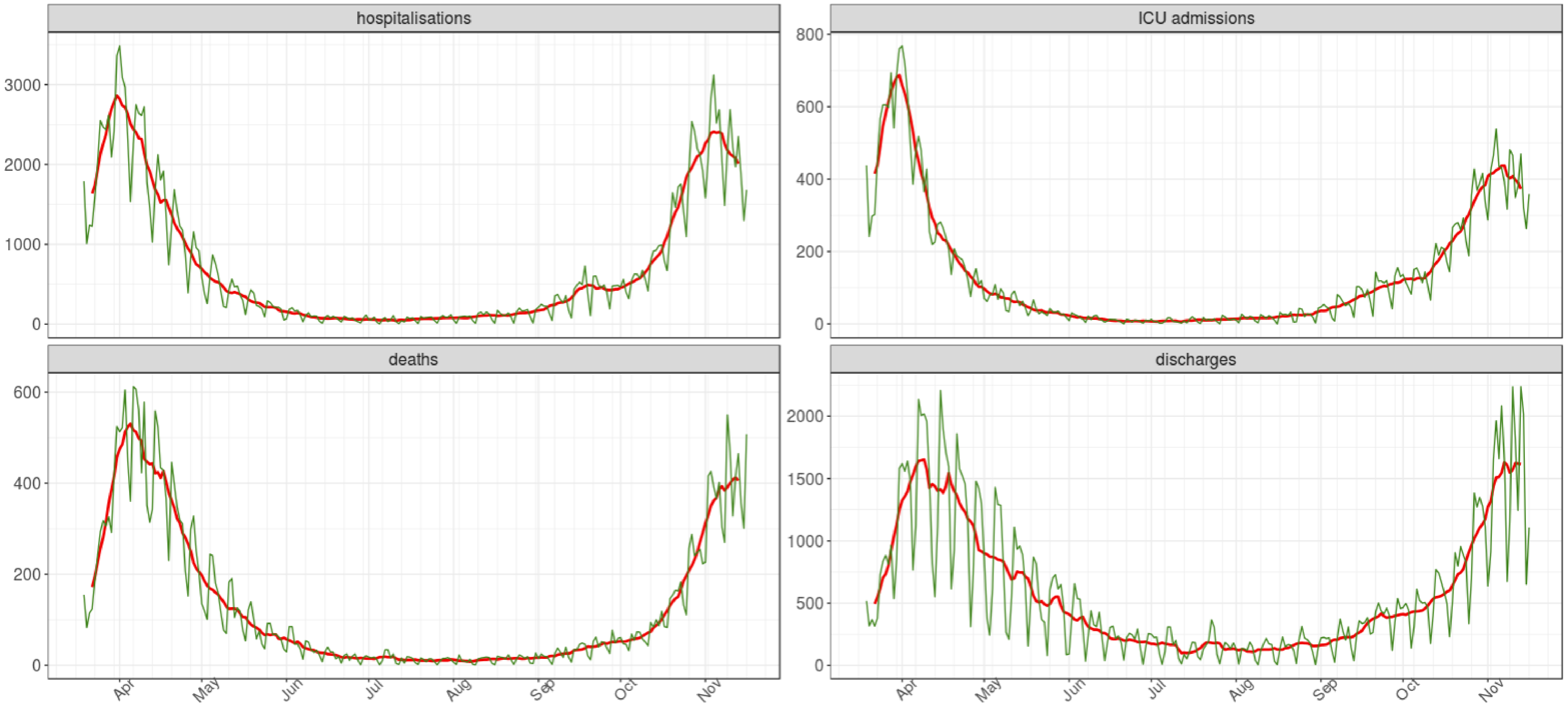


II
Modelling the SPF data

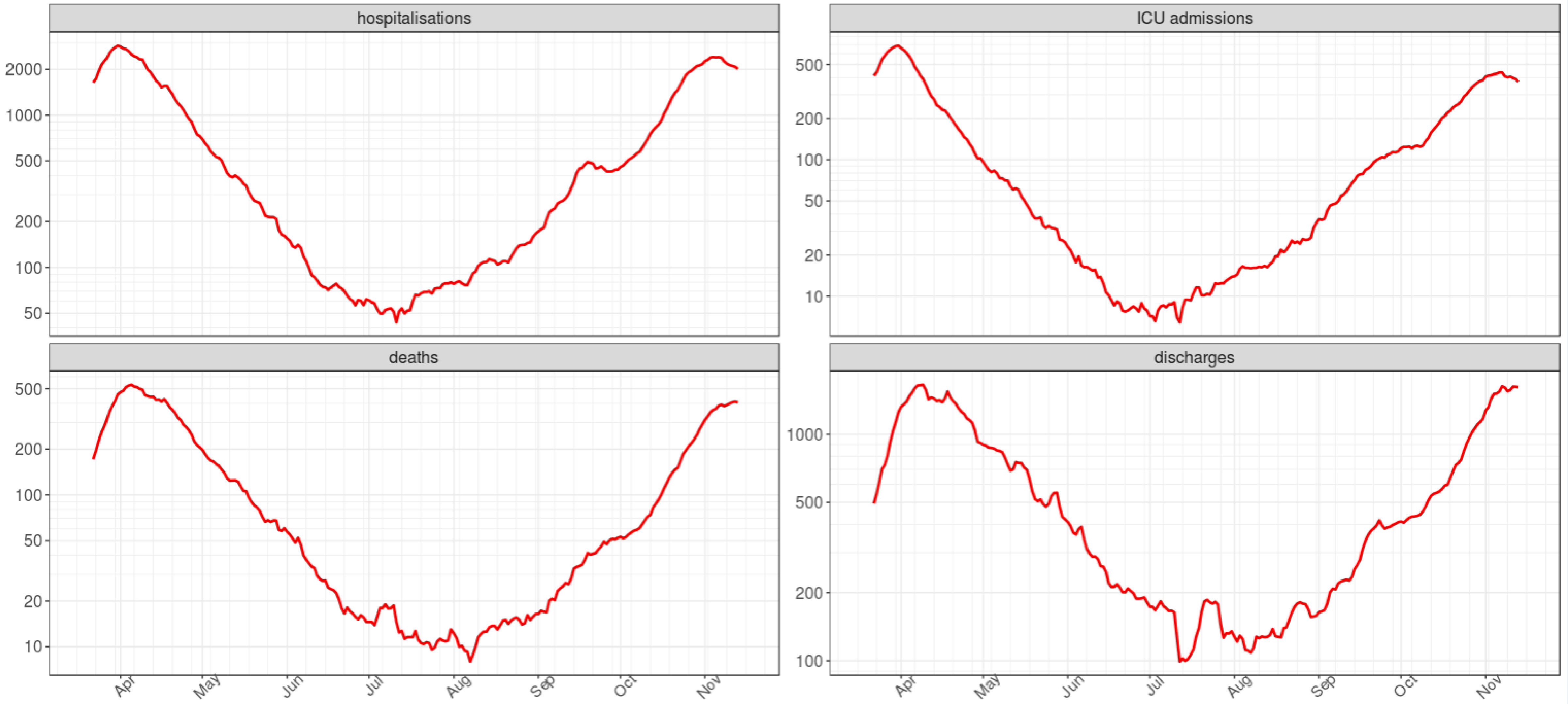
Original (French) daily data



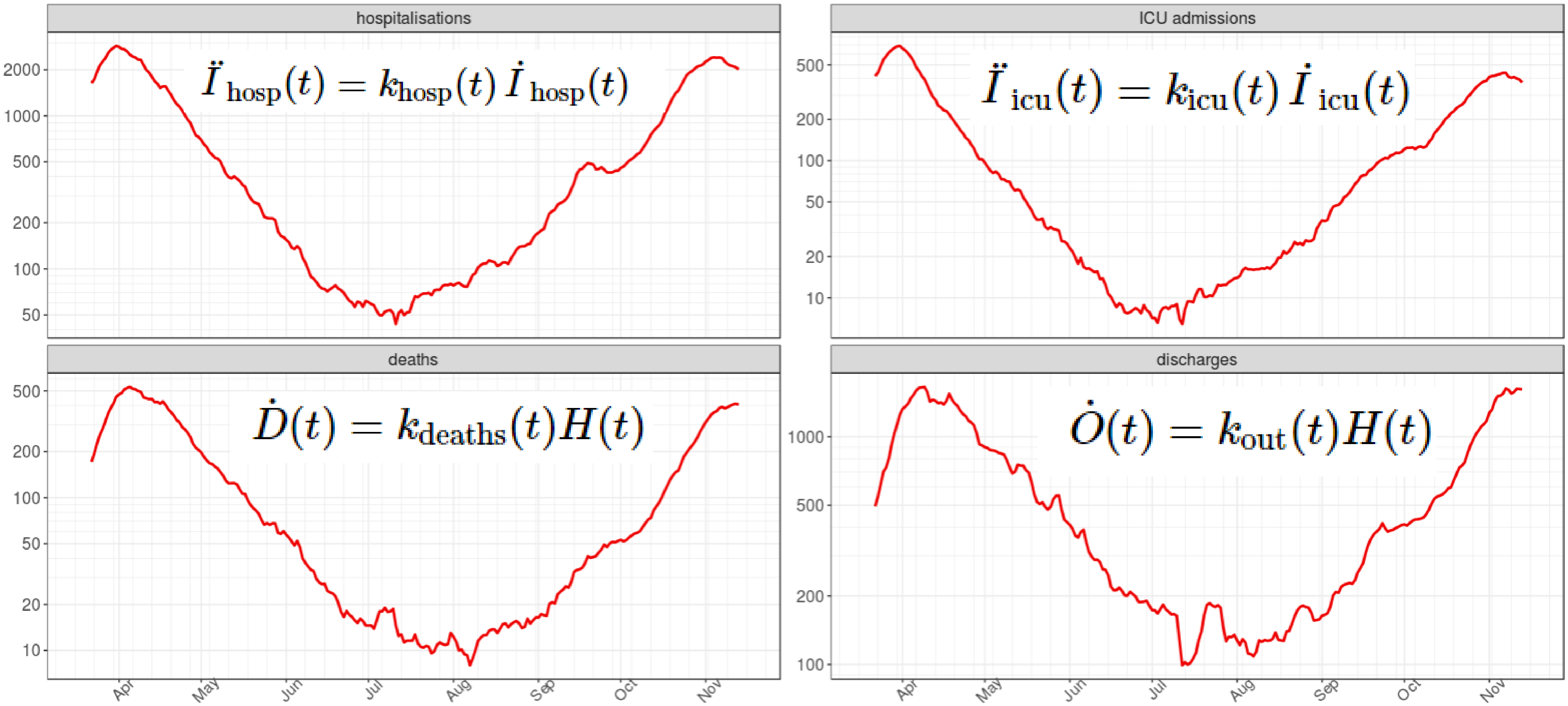
Original (French) daily data + 7-days moving average



7-days moving average & semi-log scale

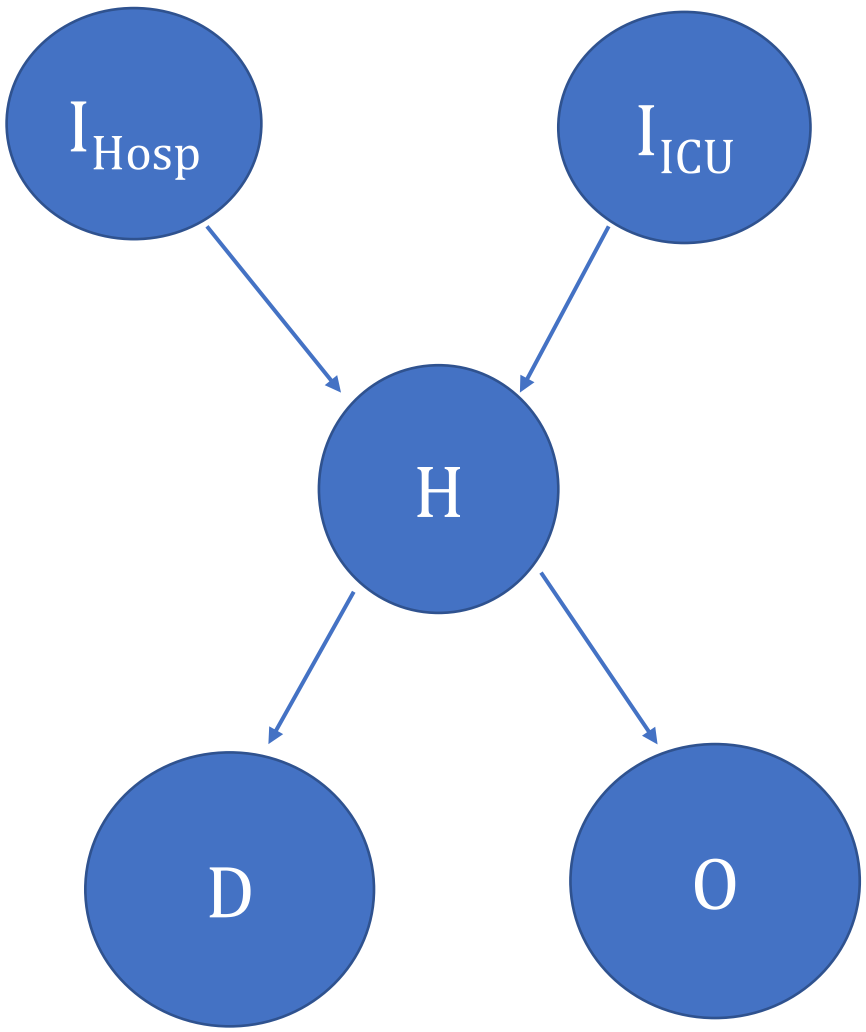


The model



H: number of individuals in hospital or in intensive care unit

The model



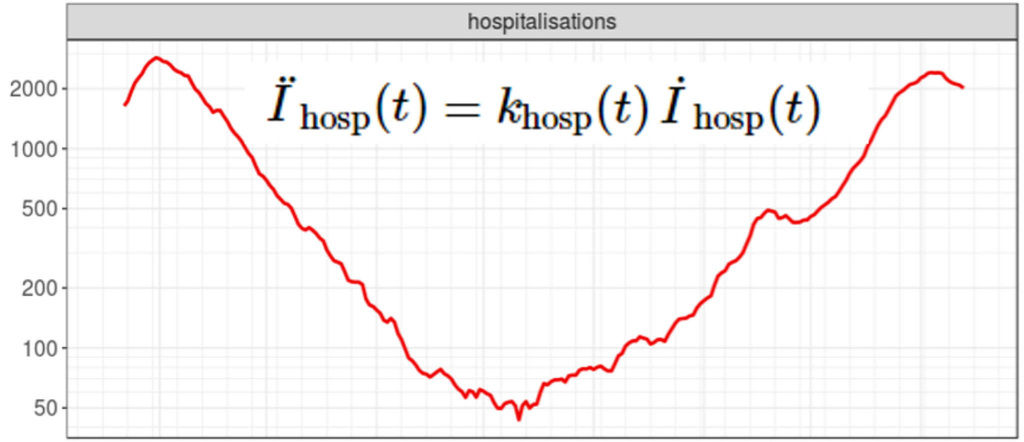
$$\dot{H}(t) = \dot{I}_{\text{hosp}}(t) + \dot{I}_{\text{icu}}(t) - \dot{D}(t) - \dot{O}(t)$$

$$\ddot{I}_{\text{hosp}}(t) = k_{\text{hosp}}(t)\,\dot{I}_{\text{hosp}}(t)$$

$$\ddot{I}_{\text{icu}}(t) = k_{\text{icu}}(t)\,\dot{I}_{\text{icu}}(t)$$

$$\dot{D}(t) = k_{\text{deaths}}(t) H(t)$$

$$\dot{O}(t) = k_{\text{out}}(t) H(t)$$



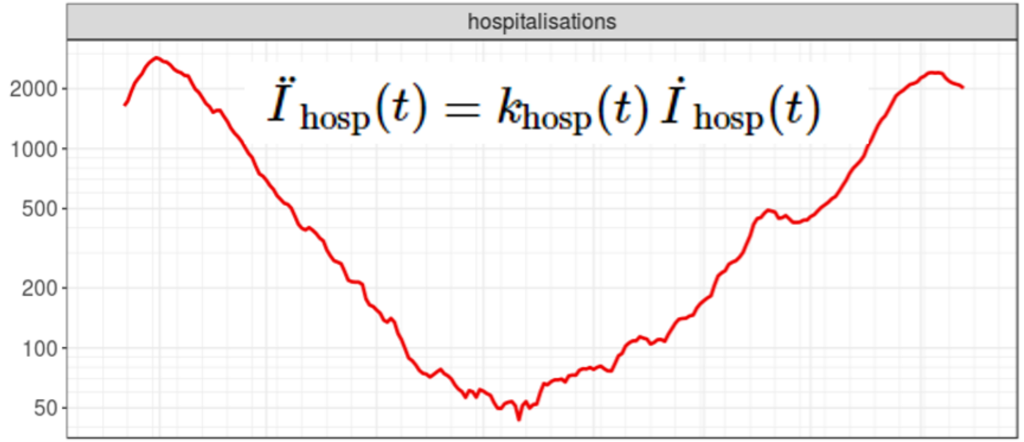
k_{hosp}: continuous piecewise linear function

Let K be the number of segments. Then, for $k = 1, 2, \ldots, K$ and $t \in (\tau_{k-1}, \tau_k)$,

$$\ddot{I}_{\text{hosp}}(t) = (b_k + c_k\, t)\,\dot{I}_{\text{hosp}}(t)$$

$$\dot{I}_{\text{hosp}}(t) = \exp\{a_k + b_k\, t + \frac{c_k}{2}\, t^2\}$$

$$\log(\dot{I}_{\text{hosp}}(t)) = a_k + b_k\, t + \frac{c_k}{2}\, t^2$$



k_{hosp}: continuous piecewise linear function

Let K be the number of segments. Then, for $k = 1, 2, \ldots, K$ and $t \in (\tau_{k-1}, \tau_k)$,

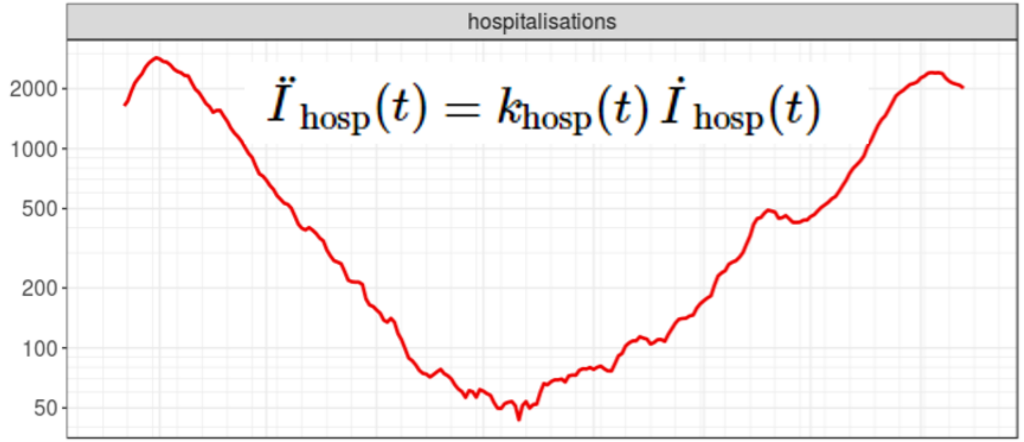
$$\ddot{I}_{\text{hosp}}(t) = (b_k + c_k\, t)\,\dot{I}_{\text{hosp}}(t)$$

$$\dot{I}_{\text{hosp}}(t) = \exp\{a_k + b_k\, t + \frac{c_k}{2}\, t^2\}$$

$$\log(\dot{I}_{\text{hosp}}(t)) = a_k + b_k\, t + \frac{c_k}{2}\, t^2$$

Continuity constraint:

$$a_k + b_k\, \tau_k + 0.5 c_k\, \tau_k^2 = a_{k+1} + b_{k+1}\, \tau_k + 0.5 c_{k+1}\, \tau_k^2$$
$$b_k + c_k\, \tau_k = b_{k+1} + c_{k+1}\, \tau_k$$



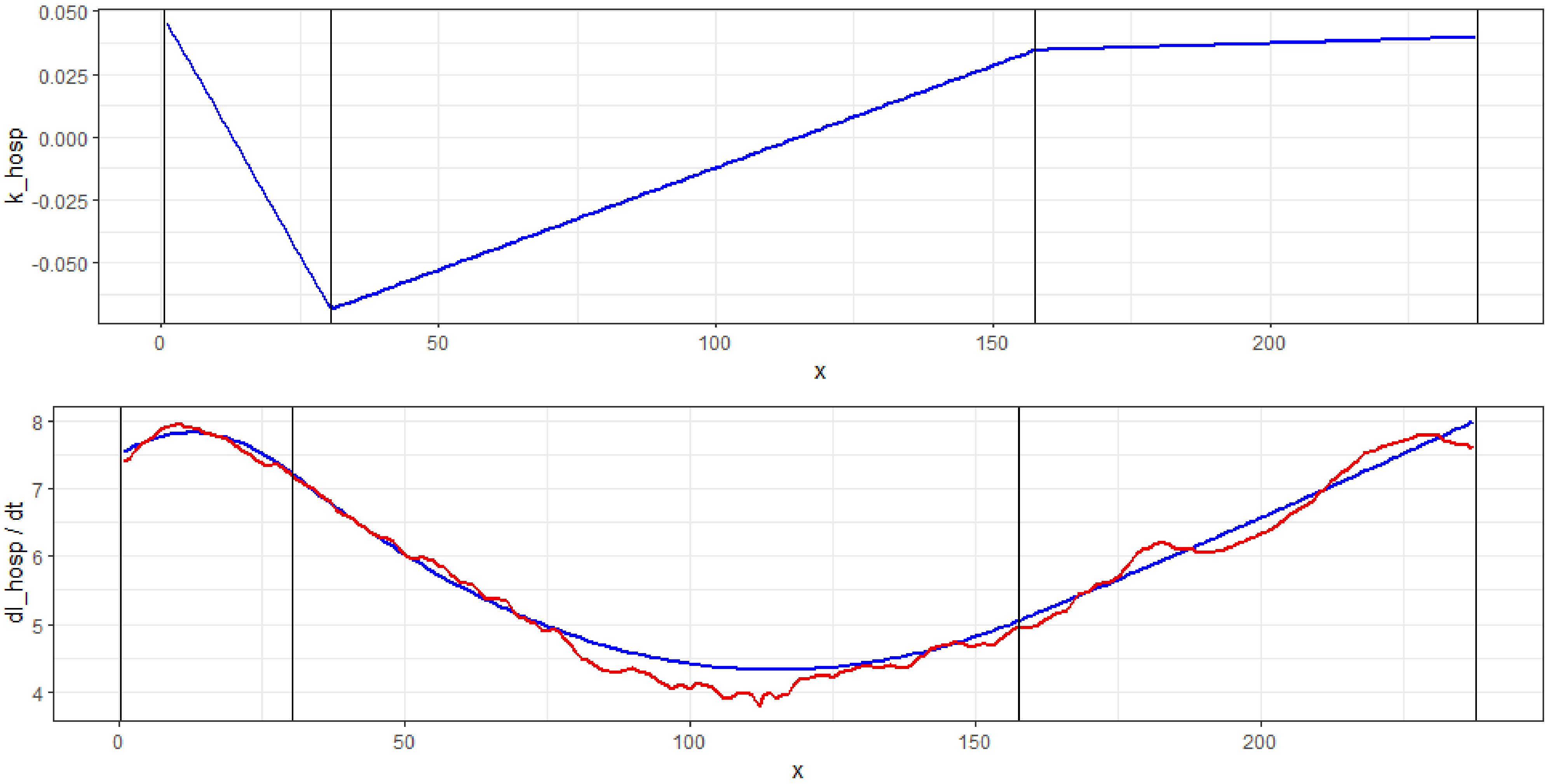
Other parameterization:

$$\log(\dot{I}_{\text{hosp}}(t)) = a_1 + b_1\, t + \frac{c_1}{2}\, t^2 + \sum_{k=1}^{K-1} h_k (t-\tau_k)^2 \times 1\{t \geq \tau_{\text{hosp},k}\}$$

The problem then becomes a problem of change-points detection:

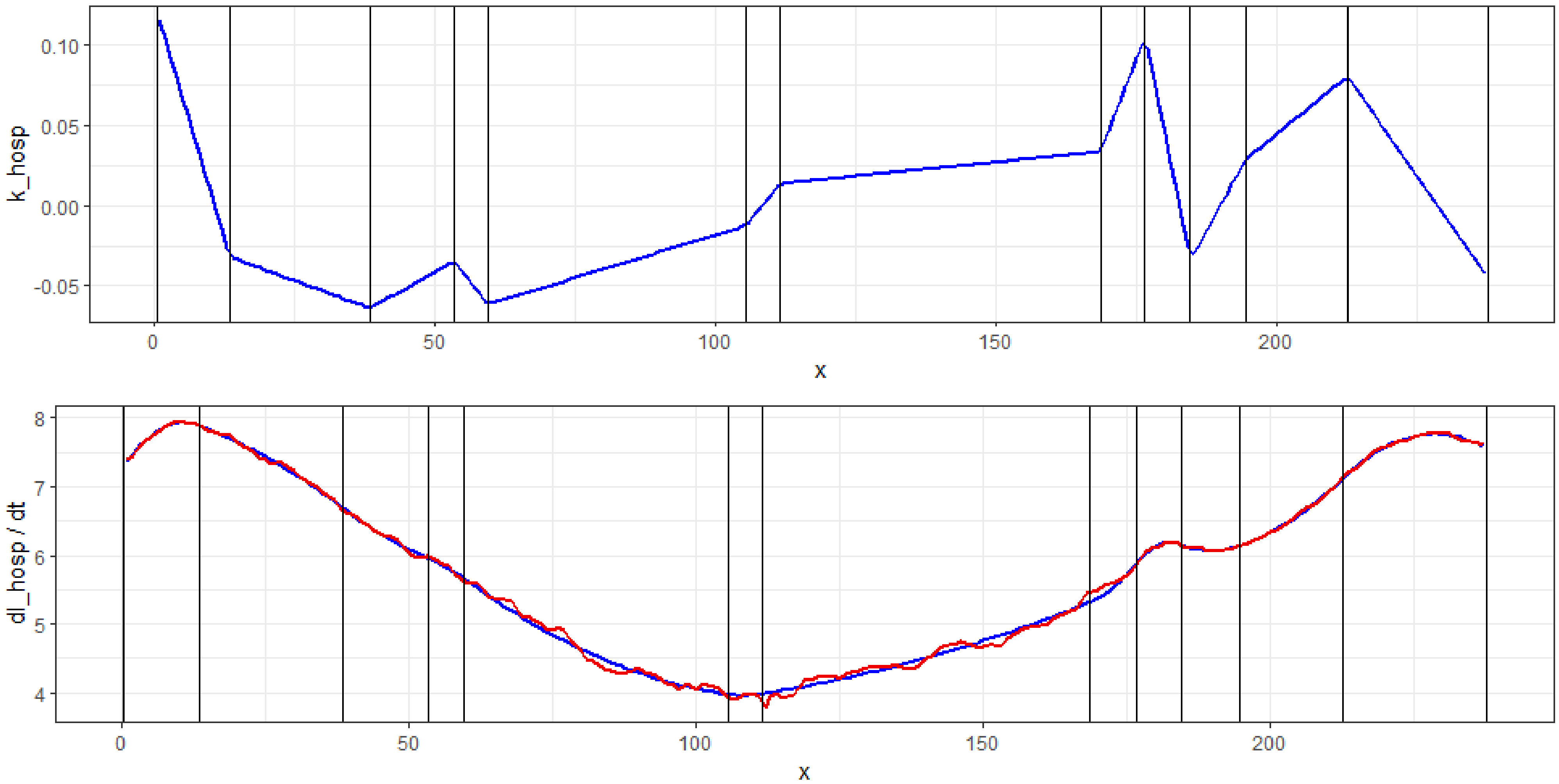
- For a given number of segments K,
  - Find the locations of the $K-1$ change points $\tau_1, \ldots, \tau_{K-1}$,
  - Estimate the parameters of the model $a_1, b_1, c_1, h_1, h_2, \ldots, h_{K-1}$
- Select the ``best'' model, i.e. select the number of segments K

K=3



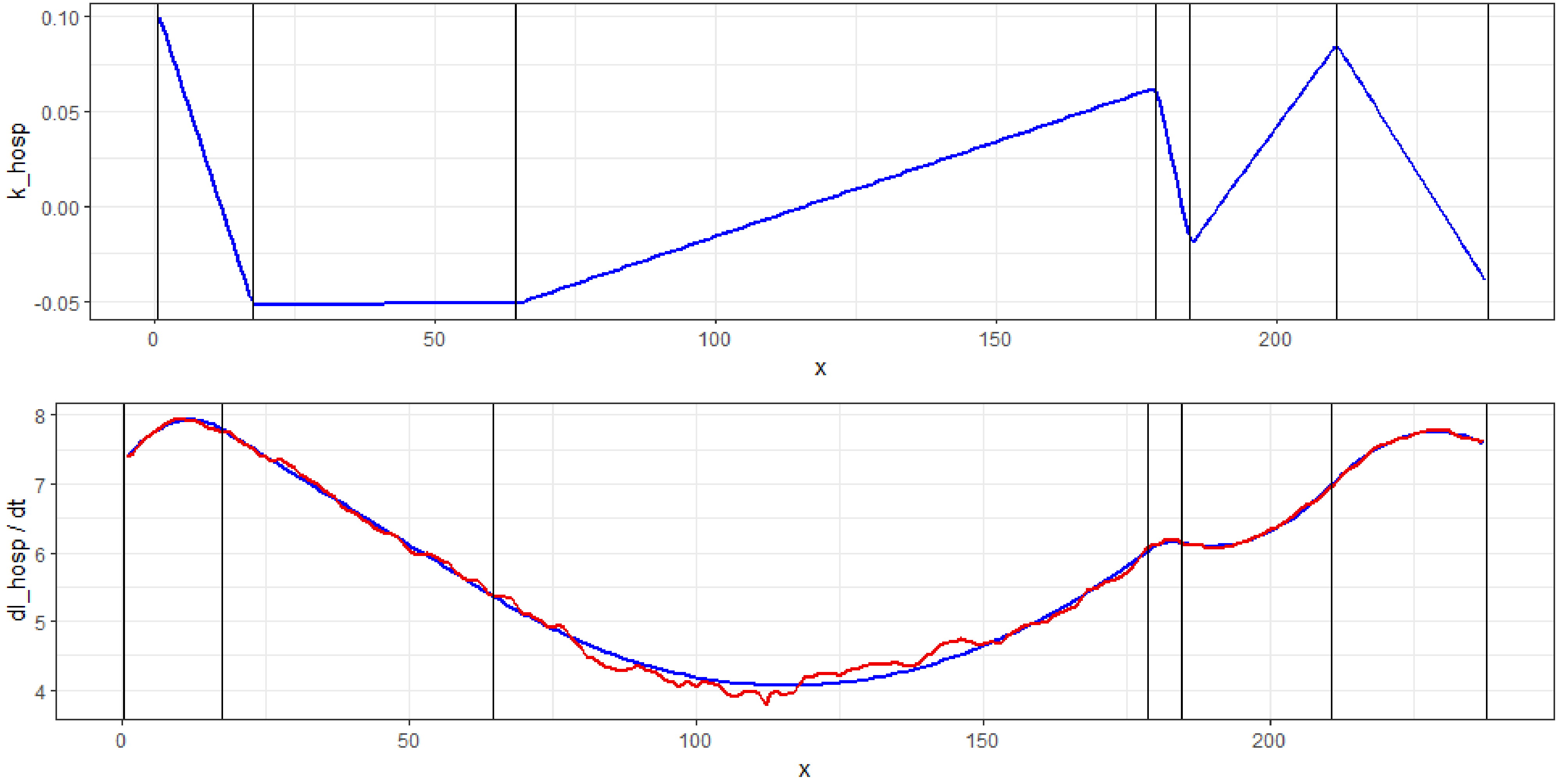
0.050
0.025
0.000
-0.025
-0.050
k_hosp
0
50
100
150
200
x
8
7
6
5
4
dl_hosp / dt
0
50
100
150
200
x

K=12



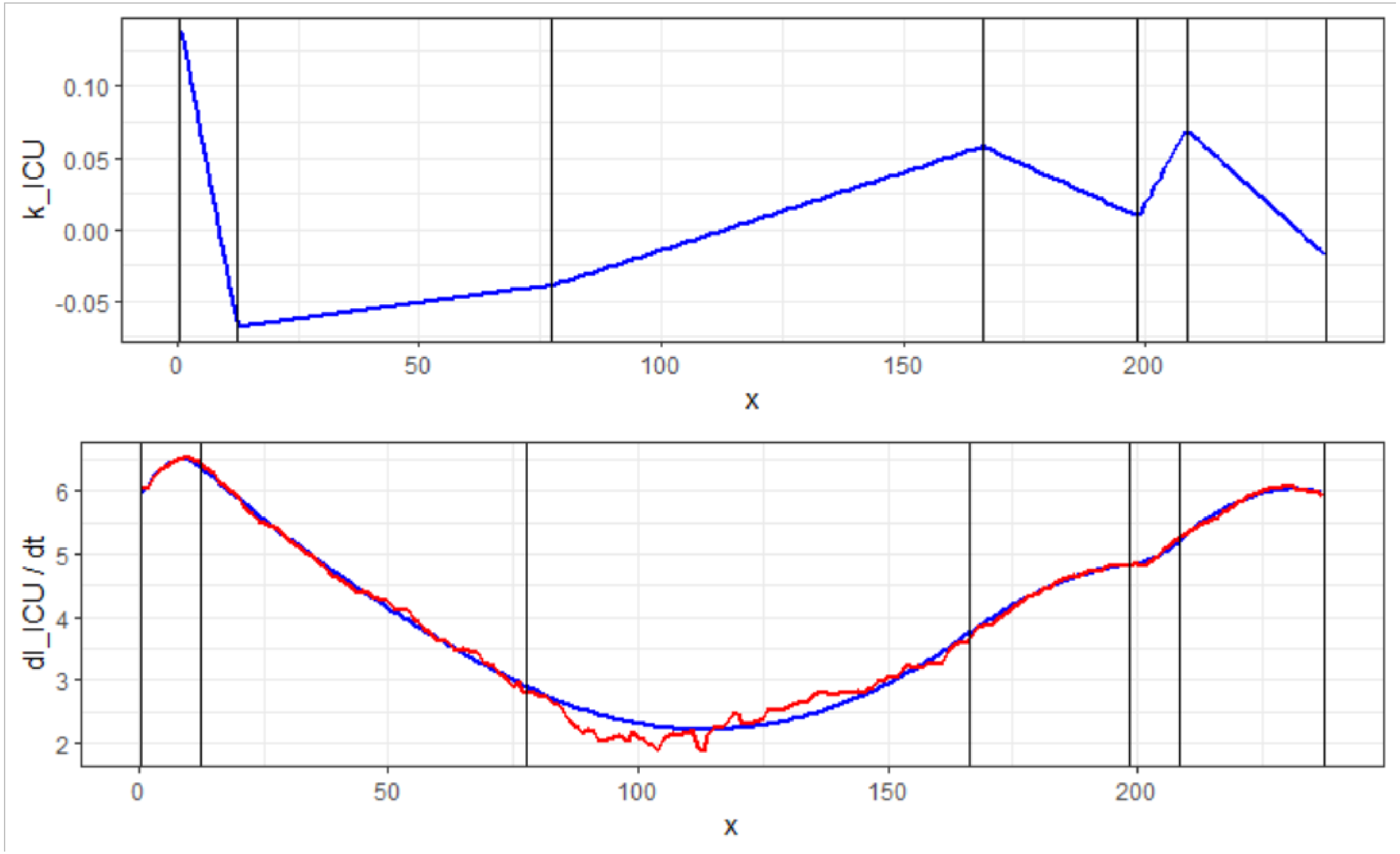
k_hosp
0.10
0.05
0.00
-0.05
0
50
100
150
200
x
dl_hosp / dt
8
7
6
5
4
0
50
100
150
200
x

K=6

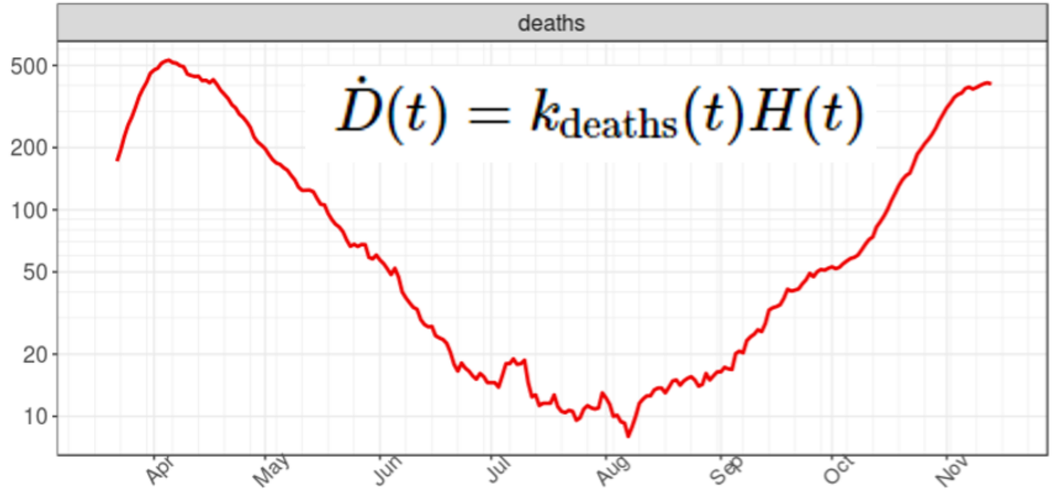


0.10
0.05
0.00
-0.05
k_hosp
0
50
100
150
200
x
8
7
6
5
4
dl_hosp / dt
0
50
100
150
200
x

K=6



k_ICU
0.10
0.05
0.00
-0.05
0
50
100
150
200
x
dI_ICU / dt
6
5
4
3
2
0
50
100
150
200
x



$$\dot{H}(t) = \dot{I}_{\text{hosp}}(t) + \dot{I}_{\text{icu}}(t) - \dot{D}(t) - \dot{O}(t)$$

$$\ddot{I}_{\text{hosp}}(t) = k_{\text{hosp}}(t)\,\dot{I}_{\text{hosp}}(t)$$

$$\ddot{I}_{\text{icu}}(t) = k_{\text{icu}}(t)\,\dot{I}_{\text{icu}}(t)$$

$$\dot{D}(t) = k_{\text{deaths}}(t)H(t)$$

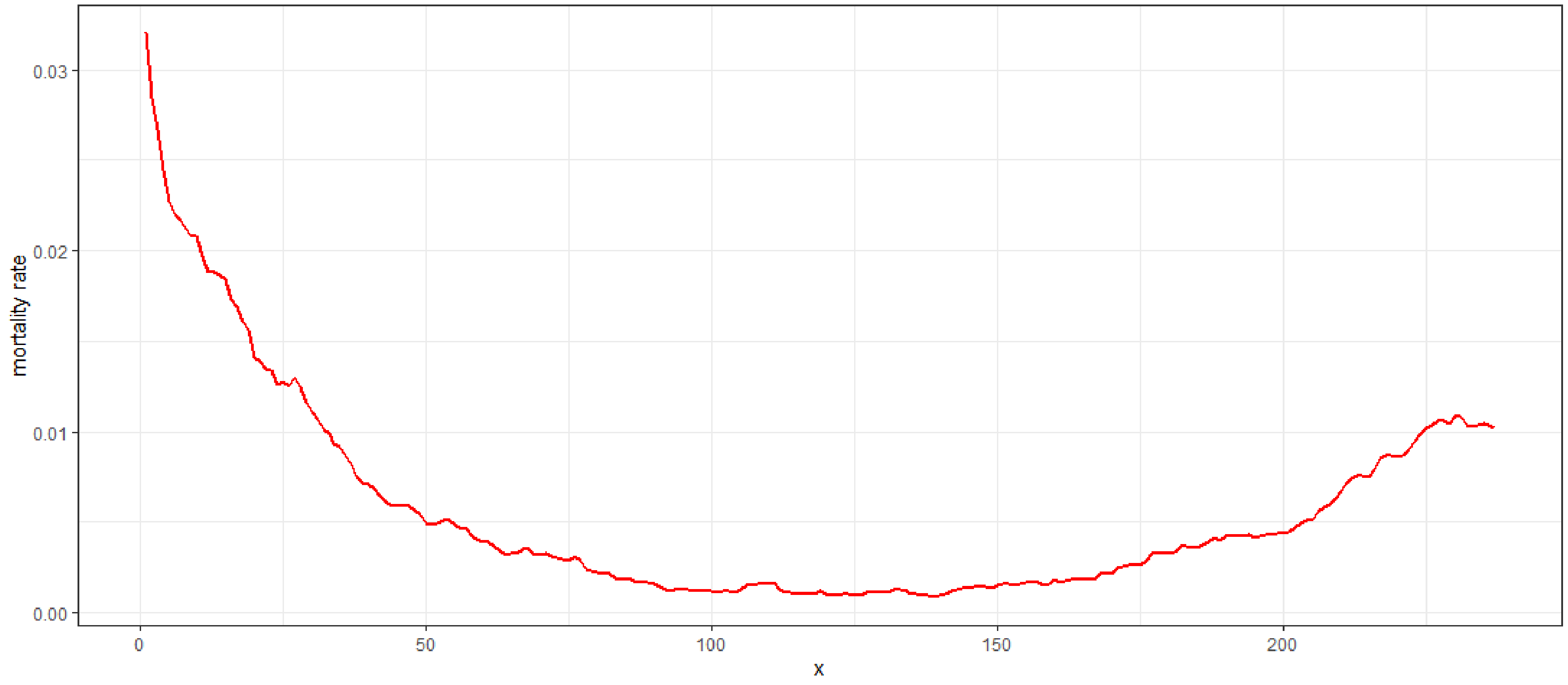
$$\dot{O}(t) = k_{\text{out}}(t)H(t)$$

Which model for the in-hospital mortality rate k_{deaths} ?

$$\dot{H}(t) = \dot{I}_{\text{hosp}}(t) + \dot{I}_{\text{icu}}(t) - k_{\text{deaths}}(t)H(t) - k_{\text{out}}(t)H(t)$$

Empirical in-hospital mortality rate:

$\kappa_{deaths}(j)$ = (number of deaths on day j) / (number of hospitalized patients on day j)

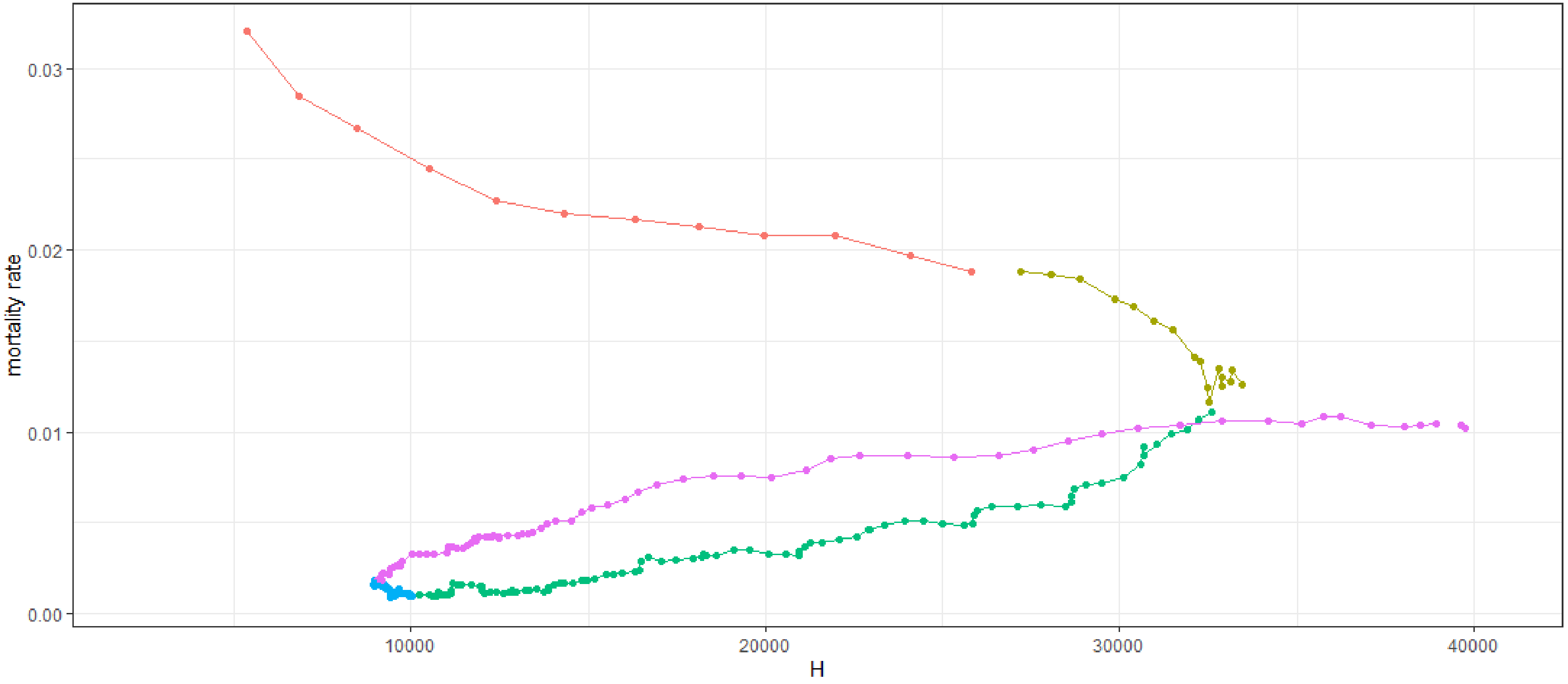


$$\dot{H}(t) = \dot{I}_{\text{hosp}}(t) + \dot{I}_{\text{icu}}(t) - k_{\text{deaths}}(t)H(t) - k_{\text{out}}(t)H(t)$$

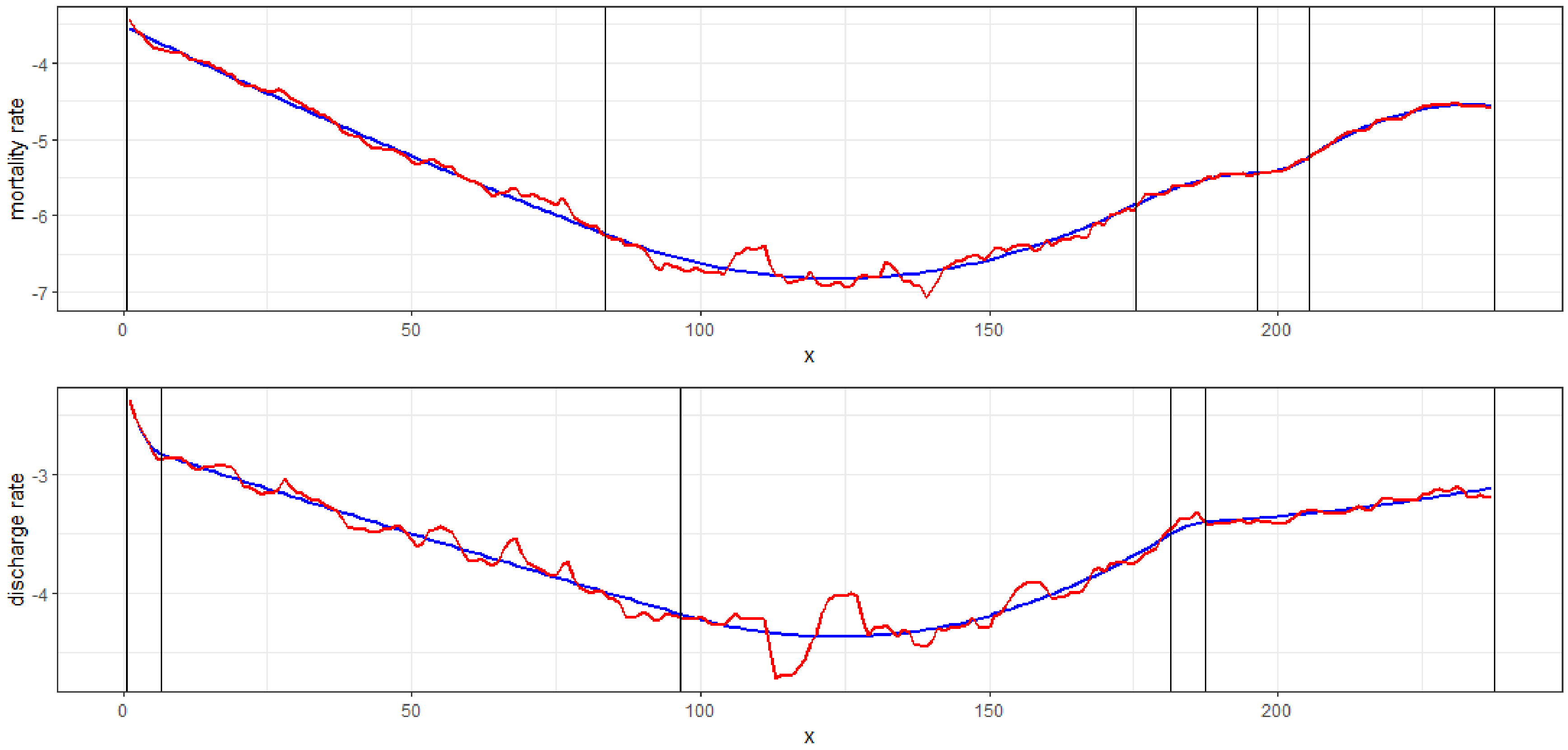
Empirical in-hospital mortality rate:

$\kappa_{deaths}(j)$ = (number of deaths on day *j*) / (number of hospitalized patients on day *j*)

$H(j)$ = number of hospitalized patients on day *j*



Piecewise quadratic models:



Final fits + confidence & prediction intervals