Data assimilation for tracking and predicting Covid19 epidemic

Javier Amezcua
Alison Fowler
Alberto Carrassi

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

j.amezcuaespinosa@reading.ac.uk

UKRI Natural Environment Research Council

Introduction

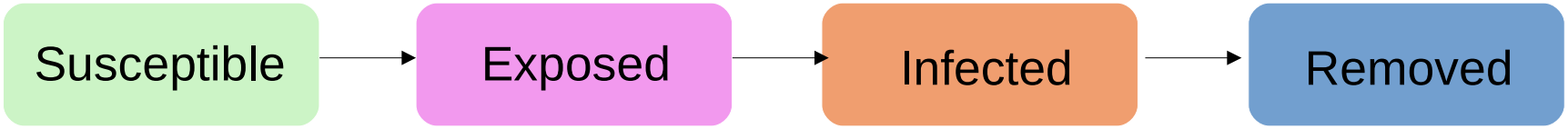
A new coronavirus disease **(Covid-19)** was identified in China in December 2019. By 3 March 2020 it had spread to **every continent** except Antarctica, totalling around 93k confirmed cases and 3k deaths. As of **22 June** there are **close to 9M confirmed cases** and close to **500k deaths**.

We use an **ensemble data assimilation** to perform **state/parameter estimation** in the evolution of the Covid19 pandemic. It belongs to the family of **iterative ensemble Kalman smoothers,** and it is popular in NWP and oil reservoir modelling.

Led by Geir Evensen (Norway), this is **international effort aimed at comparing the course of the epidemic and assessing the effectiveness of the policies across different countries**. These include: Argentina, Brazil, England, France, Norway, The Netherlands, four states of the USA, and the province Quebec of Canada.

The SEIR Model

It is a **compartmental model** often used in epidemiology.



As a **DA problem**, we aim to **estimate state variables and parameters**.

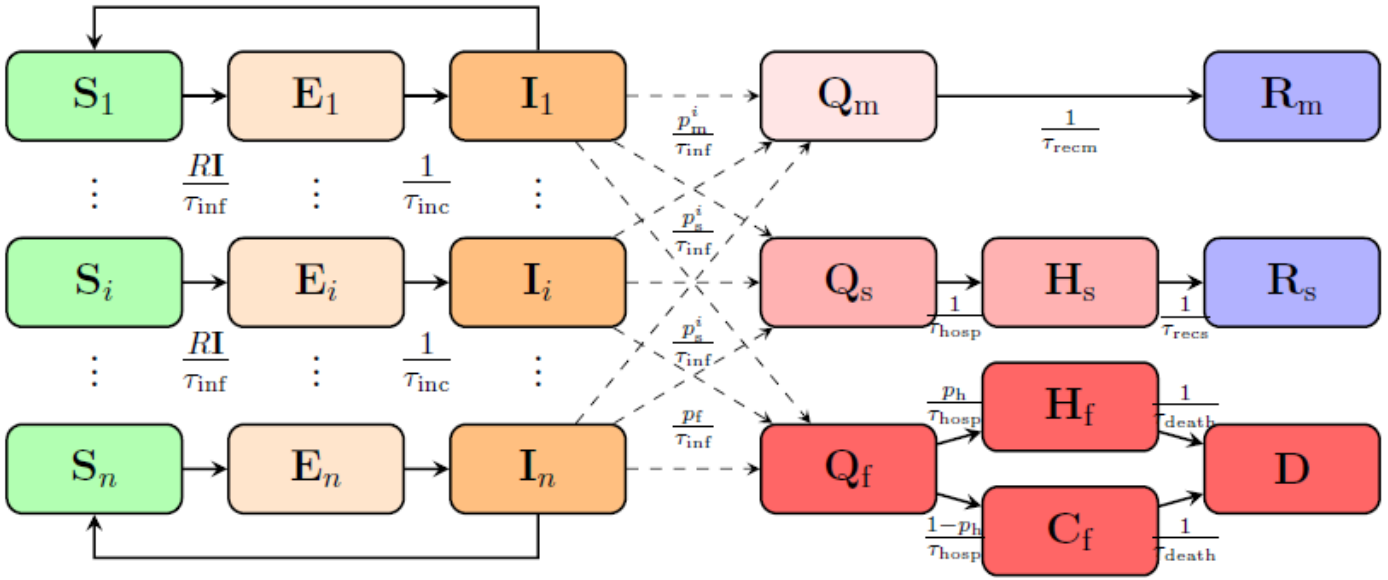
State variables: number of people in each compartment.
Parameters: residence time in each compartment, transition rates.

R(t) is the **reproduction number** and it is very important to determine if the epidemic grows exponentially *(R>1)* or if it dies down *(R<1)*.

The modified SEIR Model

Geir Evensen developed a variation of SEIR model adapted to Covid19.

- The model is **stratified by age groups**.
- The **removed** compartment has several groups:
  {quarantined, hospitalised, recovered and dead}



The modified SEIR Model

Some of the **ODE**'s:

$$\frac{\partial \mathbf{S}_i}{\partial t} = -\left(\sum_{j=1}^{n} \frac{R_{ij}(t)\mathbf{I}_j}{\tau_{\text{inf}}}\right)\mathbf{S}_i$$

$$\frac{\partial \mathbf{E}_i}{\partial t} = \left(\sum_{j=1}^{n} \frac{R_{ij}(t)\mathbf{I}_j}{\tau_{\text{inf}}}\right)\mathbf{S}_i - \frac{1}{\tau_{\text{inc}}}\mathbf{E}_i$$

$$\frac{\partial \mathbf{I}_i}{\partial t} = \frac{1}{\tau_{\text{inc}}}\mathbf{E}_i - \frac{1}{\tau_{\text{inf}}}\mathbf{I}_i$$

$$\frac{\partial \mathbf{Q}_{\text{m}}}{\partial t} = \sum_{i=1}^{n} \frac{p_{\text{m}}^i}{\tau_{\text{inf}}}\mathbf{I}_i - (1/\tau_{\text{recm}})\,\mathbf{Q}_{\text{m}}$$

$$\frac{\partial \mathbf{Q}_{\text{s}}}{\partial t} = \sum_{i=1}^{n} \frac{p_{\text{s}}^i}{\tau_{\text{inf}}}\mathbf{I}_i - (1/\tau_{\text{hosp}})\,\mathbf{Q}_{\text{s}}$$

$$\frac{\partial \mathbf{Q}_{\text{f}}}{\partial t} = \sum_{i=1}^{n} \frac{p_{\text{f}}^i}{\tau_{\text{inf}}}\mathbf{I}_i - (1/\tau_{\text{hosp}})\,\mathbf{Q}_{\text{f}}$$

$$\frac{\partial \mathbf{H}_{\text{s}}}{\partial t} = \frac{1}{\tau_{\text{hosp}}}\mathbf{Q}_{\text{s}} - \frac{1}{\tau_{\text{recs}}}\mathbf{H}_{\text{s}}$$

Parameters in the system:

Parameter first guess (Std. Dev.)	Description
$I_0 = 60$ (6)	Initial infectious
$E_0 = 240$ (24)	Initial exposed
$\tau_{\text{inf}} = 3.8$ days (0.5)	Infection time
$\tau_{\text{inc}} = 5.5$ days (0.5)	Incubation time
$\tau_{\text{recm}} = 14$ days (0.5)	Recovery time mild cases
$\tau_{\text{recs}} = 5$ days (0.4)	Recovery time severe cases
$\tau_{\text{hosp}} = 6$ days (0.5)	Time to hospitalization
$\tau_{\text{death}} = 16$ days (0.5)	Time to death
$p_f = 0.009$ (0.001)	Case fatality rate
$p_s = 0.039$ (0.003)	Hospitalization rate for severe cases
$R_0 = 3.87$ (0.5)	Reproduction rate before intervention
$R_1 = 0.9$ (0.45)	Reproduction rate at the intervention
$p_h = 0.7$	Ratio of population that is hospitalised

The majority of the ODE's are linear in the **state variables** (except for some *IS* products), but not in the **parameters**.

The observations

We have access to **three observed quantities**.

- The number of **positive** Covid19 cases.

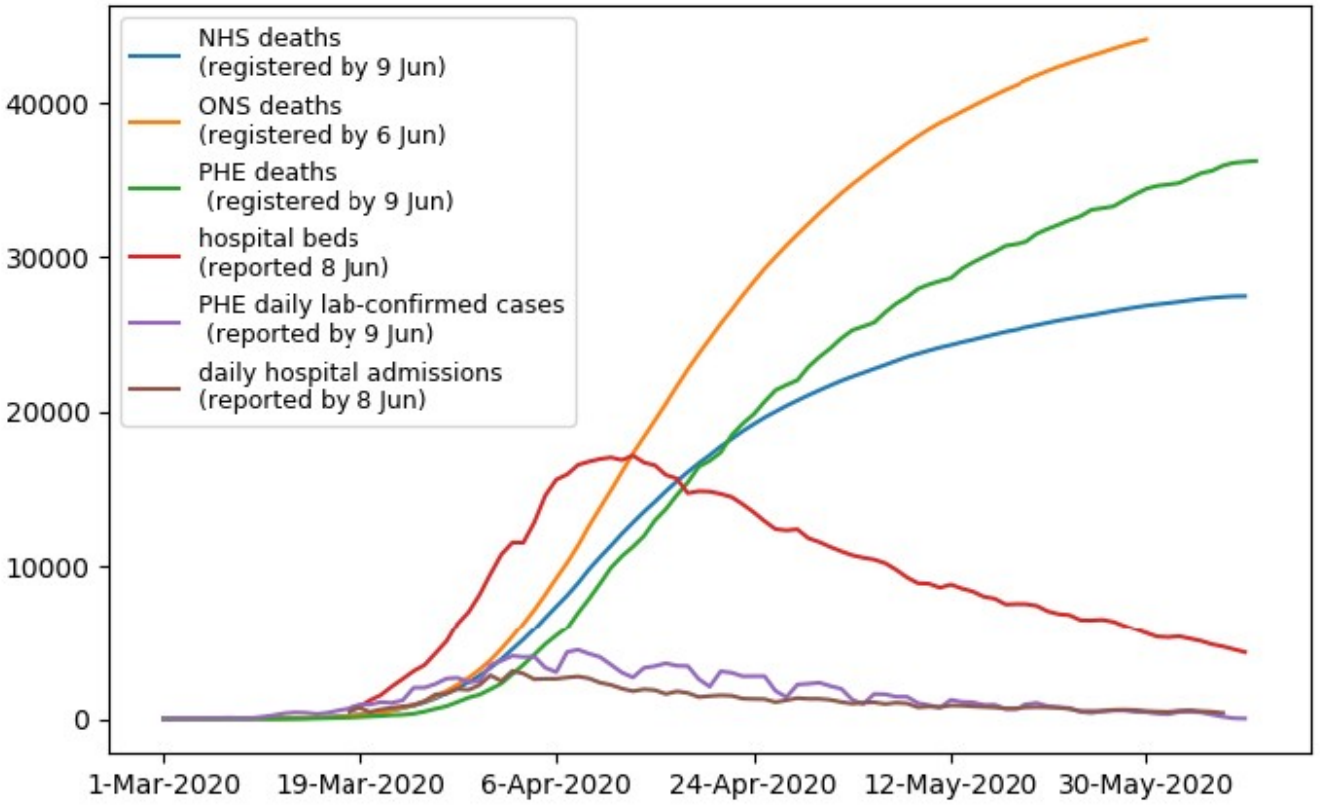
- The number of **hospital beds** with Covid19 patients.

- The number of **deaths** associated to Covid19.

These can be reported: as **new instances per day, or accumulated**.

The type of observation and the way it is reported influences the observational error: **random and/or systematic**, with **time correlation**.

The observations



We use:
- ONS **deaths**.
- **Hospital beds** with Covid19 patient.
- PHE daily **confirmed** cases

A time line of events

Initialisation of the model: 20/02

Observations used for assimilation

Deaths: 05/03-05/06

Hospital beds: 20/03-12/06

Positives*: 05/03-12/06

Lockdown: 23/03-01/06.

Free forecast runs: from 06/06 or 13/06

After the lockdown we consider 3 exit scenarios:

R = 0.5

R = 1.0

R = 1.2

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UKRI Natural Environment Research Council

People in contact

The spread of the virus depends on the **contact rate between people**. In this model the **contact rates** are specified for different age groups. This information comes from medical literature.

Normal life

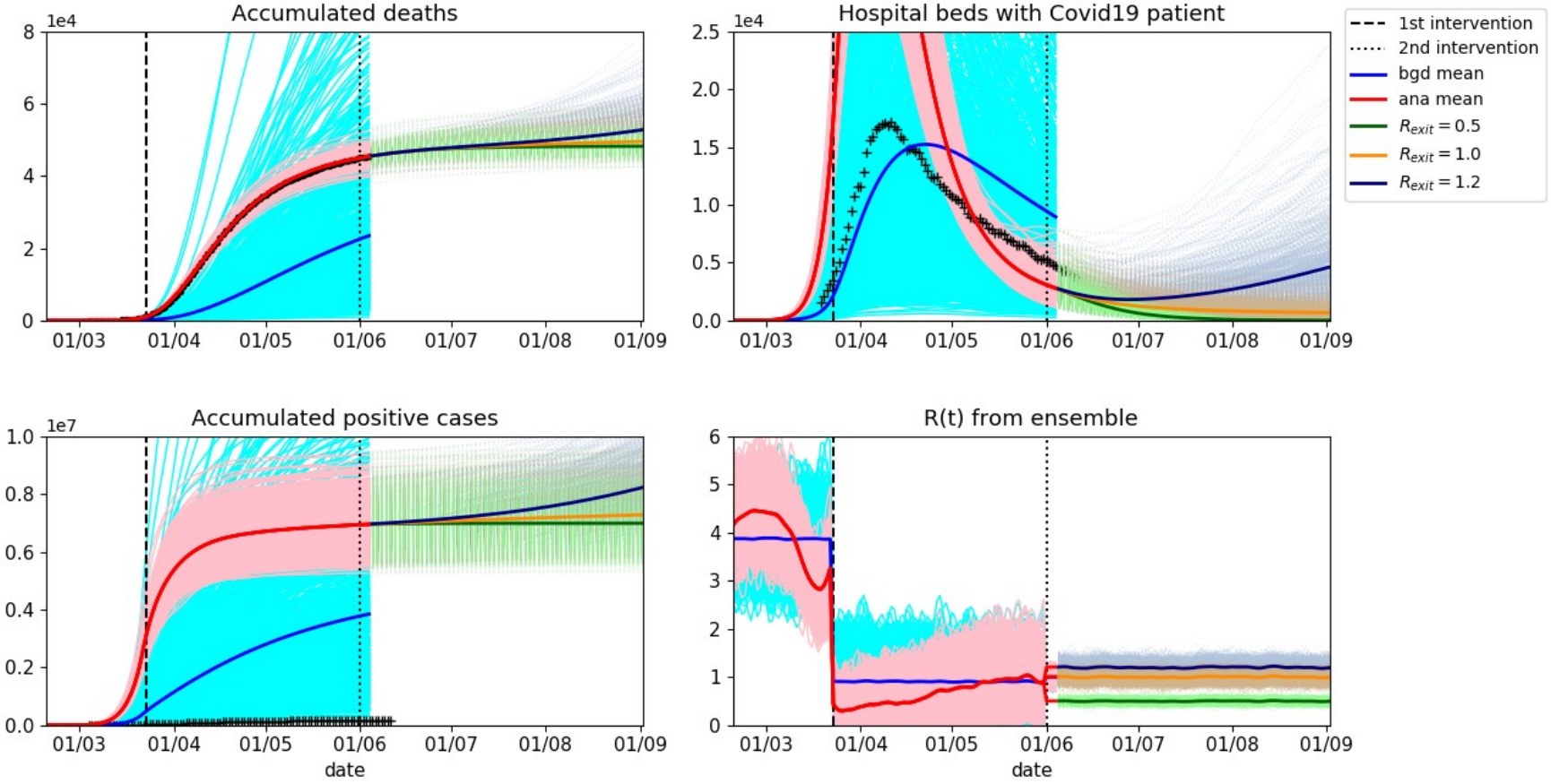
Age groups	1	2	3	4	5	6	7	8	9	10	11
1	**2.0**	1.5	1.5	1.0	1.5	0.5	0.5	0.5	0.4	0.4	0.4
2	0.5	**8.0**	6.0	2.0	2.5	2.5	1.5	1.4	0.9	0.9	0.9
3	0.5	6.0	**8.0**	2.0	2.5	2.5	1.5	1.4	0.9	0.9	0.9
4	0.5	2.5	2.5	**6.0**	2.0	2.0	1.9	1.5	0.9	0.9	0.9
5	1.2	2.5	2.5	2.0	**3.0**	2.0	1.9	1.8	0.5	0.5	0.5
6	0.5	2.3	2.3	2.0	2.0	**3.0**	1.9	1.5	1.4	1.4	1.4
7	0.5	2.0	2.0	1.5	1.5	1.5	**2.0**	1.5	0.9	0.9	0.9
8	0.5	1.9	1.9	1.0	1.2	1.2	1.9	**1.5**	0.9	0.9	0.9
9	0.5	1.5	1.5	0.9	0.9	1.2	1.0	1.5	**1.5**	1.5	1.5
10	0.4	1.0	1.0	0.9	0.7	1.2	1.0	1.0	1.5	**1.5**	1.5
11	0.4	0.9	0.9	0.9	0.7	1.2	1.0	1.0	1.5	1.5	**1.5**

During lockdown

Age groups	1	2	3	4	5	6	7	8	9	10	11
1	**1.0**	0.9	0.9	0.8	1.0	0.5	0.5	0.4	0.3	0.3	0.3
2	0.5	**2.0**	1.5	0.9	1.0	1.0	0.5	0.4	0.3	0.3	0.3
3	0.5	1.5	**2.0**	0.9	1.0	1.0	0.5	0.4	0.3	0.3	0.3
4	0.5	1.0	1.0	**1.2**	1.0	1.0	0.9	0.5	0.4	0.3	0.3
5	0.8	1.0	1.0	0.9	**1.1**	0.9	0.9	0.5	0.4	0.3	0.3
6	0.5	1.0	1.0	1.0	1.0	**1.1**	0.9	0.5	0.4	0.3	0.3
7	0.5	0.6	0.6	0.9	0.9	0.9	**1.0**	0.7	0.5	0.5	0.5
8	0.5	0.6	0.6	0.8	0.9	1.0	1.0	**1.0**	0.5	0.5	0.5
9	0.5	0.6	0.6	0.6	0.5	1.0	0.9	0.9	**1.1**	1.1	1.1
10	0.5	0.6	0.6	0.6	0.5	1.0	0.9	0.9	1.1	**1.1**	1.1
11	0.5	0.6	0.6	0.6	0.5	1.0	0.9	0.9	1.1	1.1	**1.1**

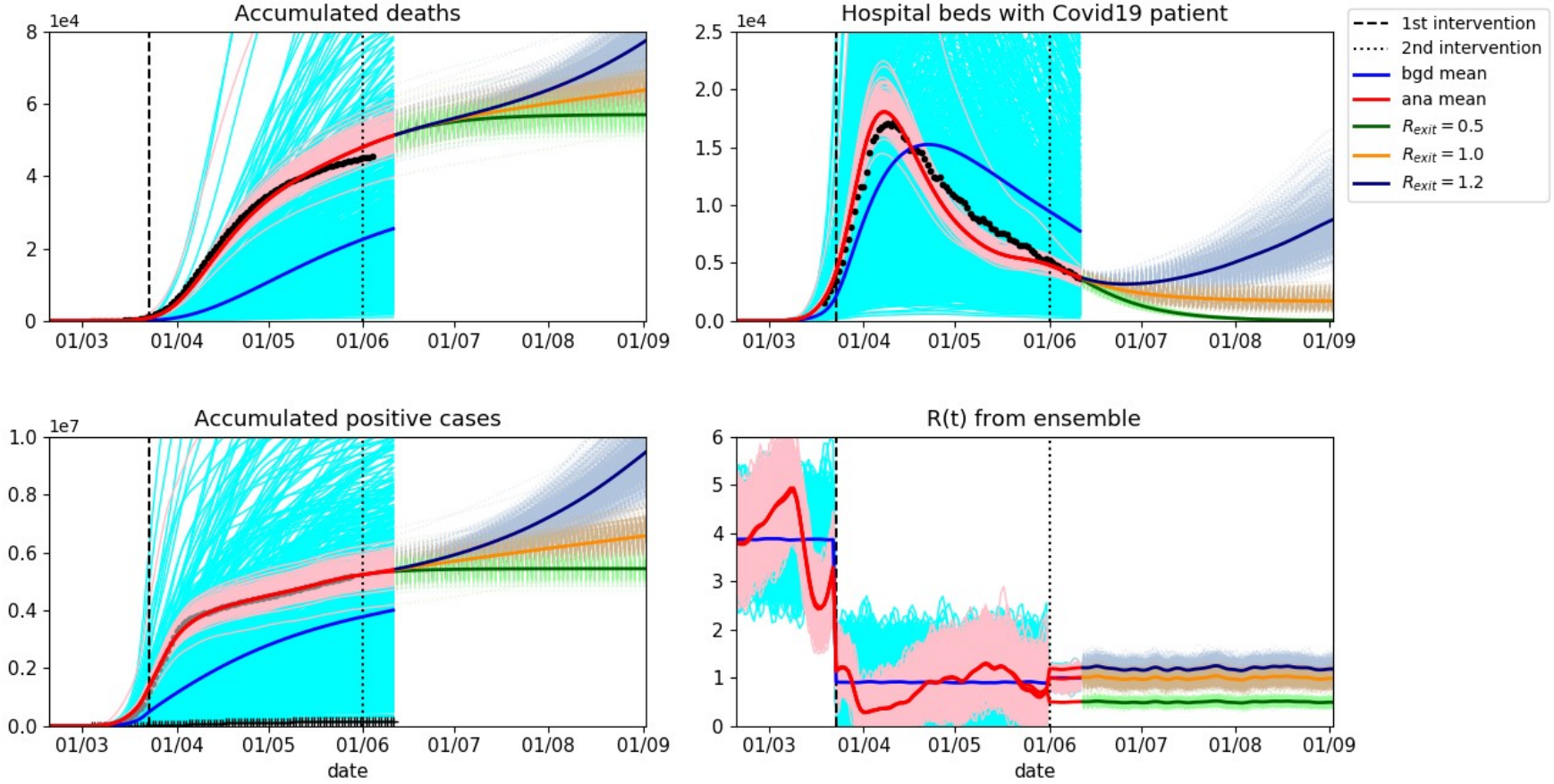
The diagonal elements indicate interactions within groups, while the off-diagonal elements between groups (not symmetric).

Assimilating deaths



Note that the peak is missed both in time and number.

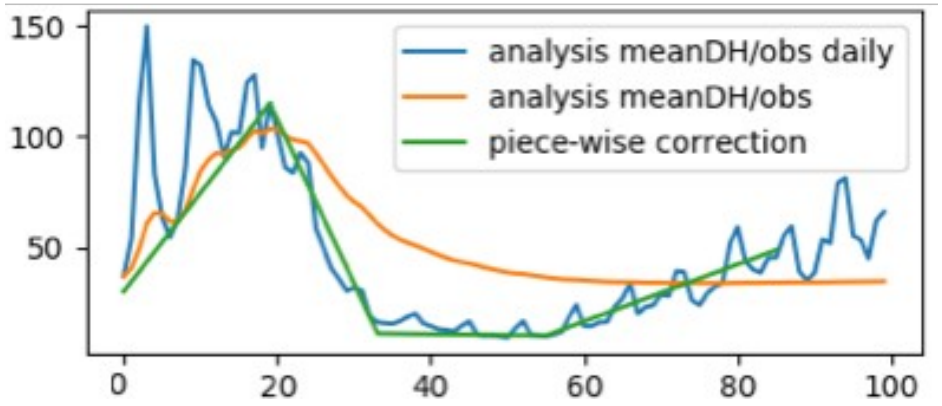
Assimilating deaths and hospital beds



The results get both the number of deaths and hospitalisations. We use these results to calibrate the observation of positive cases.

Calibrating the positive cases

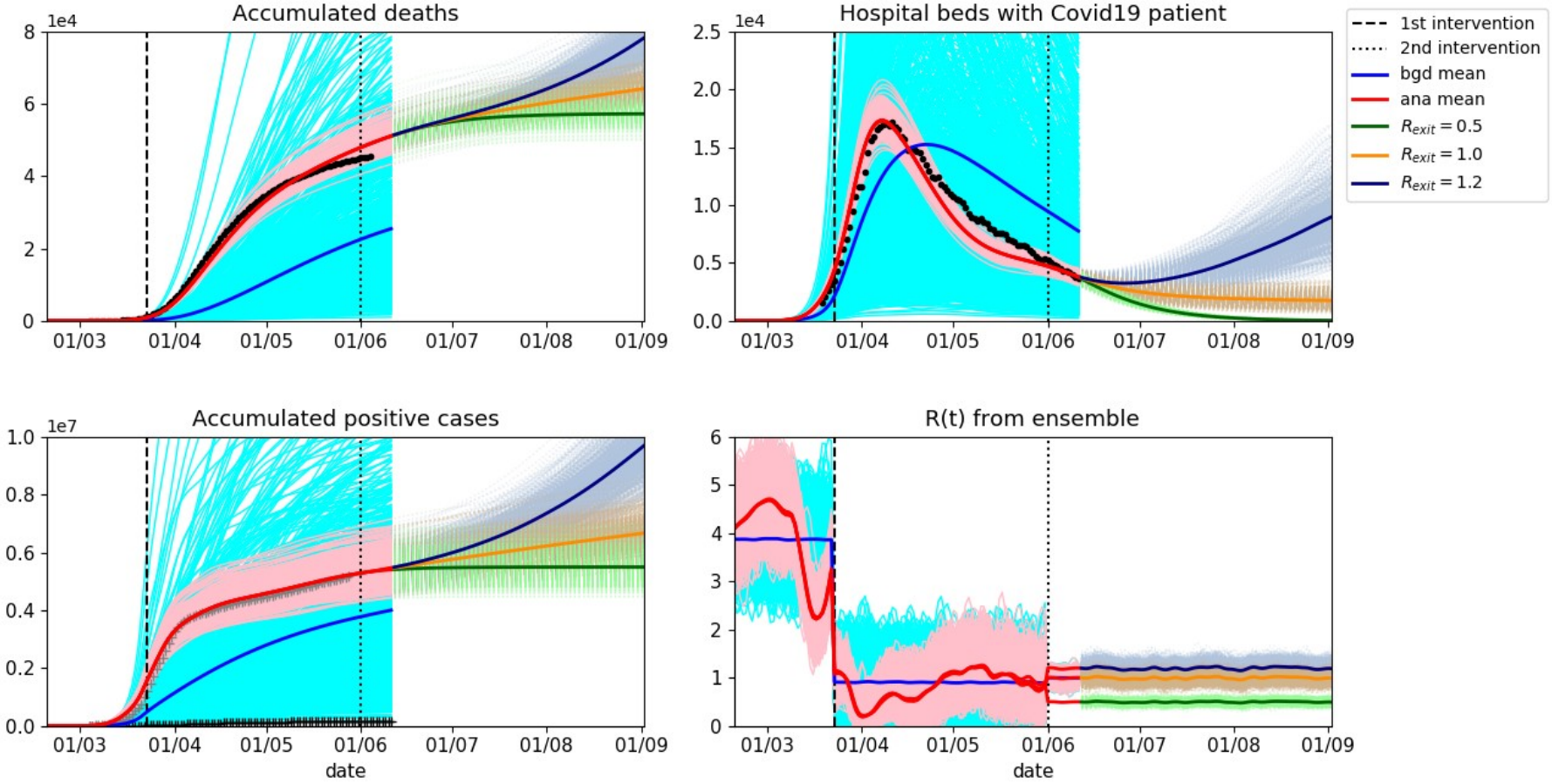
To correct the observations of positive cases, we use the analysis mean value of the DH experiment smoothed in time. This allows us to compute a **piece-wise correction factor** which is then applied to the raw observations. This is a **correction for systematic error**, not random.



By **late March**, only around **1%** of positve cases were being reported. From **early May**, about **2%** of the accumulated cases were reported.

The UK government achieved its target of 100,000 tests per day on **May 1st**. We conclude that approximately **10%** of the accumulated symptomatic cases are being reported from the beginning of May.

Assimilating the 3 observations



The analysis spread reduces. The behaviour in R improves.

Summary

We have **successfully implemented an ensemble data assimilation method for tracking and predicting the Covid19 pandemic in England**. We demostrated the strength of DA to provide a common grounded protocol to study the epidemic in completely different contexts.

The method allowed a dynamical estimation of the **reproduction number R(t)**. The effects of the **lockdown** are clear in **reducing it below the critical unitary value**. As of 1st June, approximately 45,000 deaths were attributed to Covid-19 in England in all settings (source, ONS). Our projections by 1st September for total deaths are: **57,000±1,900 (R=0.5)**, **63,600±2,700 (R=1)** and **76,400±4,900 (R=1.2)**.

Improvements of the method include considering **regional variations** of the state variables and parameters, as well as **stochastic elements** to represent model error.

Geir Evensen, Javier Amezcua, Marc Bocquet, Alberto Carrassi, Alban Farchi, Alison Fowler, Peter Houtekamer, Christopher K. R. T. Jones, Rafael de Moraes, Manuel Pulido, Christian Sampson, Femke Vossepoel: An international assessment of the COVID-19 pandemic using ensemble data assimilation. Submitted to Foundations of Data Science. Preprint on medRxiv. doi: https://doi.org/10.1101/2020.06.11.20128777