

SEIR with age compartments for assessing opening of schools and kindergartens in Norway after 20th of April

Geir Evensen

Version compiled April 17, 2020

Abstract The work illustrates how one can use iterative ensemble smoothers to estimate parameters of a SEIR model with age-classes and compartments of sick, hospitalized, and dead. The data conditioned on are the daily numbers of accumulated deaths and the number of hospitalized. Starting from a wide prior for the model parameters, the ensemble conditioning leads to a posterior ensemble of estimated parameters with associated model predictions in close agreement with the observations. The posterior ensemble of model simulations provides predictions in agreement with the observations that have predictive capabilities and include uncertainty estimates.

1 Introduction

2 SEIR model with ageclasses

A simple model for epidemic modeling is the SEIR model (Susceptible, Exposed, Infectious, and Recovered). For simulating somewhat realistically the Corona epidemic the model needs to be extended and a more convenient formulation is the following conceptual model:

$$\left\{ \begin{array}{l} S_1 \rightarrow E_1 \rightarrow I_1 \\ \vdots \\ S_i \rightarrow E_i \rightarrow I_i \rightarrow \left\{ \begin{array}{ll} Q_m & \rightarrow R_m \\ Q_s & \rightarrow H_s \rightarrow R_s \\ Q_f & \rightarrow H_f \rightarrow D \end{array} \right. \\ \vdots \\ S_n \rightarrow E_n \rightarrow I_n \end{array} \right. \quad (1)$$

Here we have added age compartments to the standard SEIR model. First we divide the populations of susceptible S_i , exposed E_i , and infectious I_i , into age groups.

Then it is assumed that after the infectious period when the patients know they are sick they gather in common groups of “Quarantened” patients of mild Q_m , severe Q_s , and fatal symptoms Q_f . The patients with mild symptoms will recover into R_m without going to the hospital. The patients with severe symptoms goes to the hospital H_s , and thereafter recover into R_s . Finally, the fatally ill patients goes to the hospital H_f before they end up in the group of dead D .

The model equations are as follows:

$$\frac{\partial S_i}{\partial t} = -\frac{1}{\tau_{inf}} \left(\sum_{j=1}^n R_{ij}(t) I_j \right) S_i \quad (2)$$

$$\frac{\partial E_i}{\partial t} = \frac{1}{\tau_{inf}} \left(\sum_{j=1}^n R_{ij}(t) I_j \right) S_i - \frac{1}{\tau_{inc}} E_i \quad (3)$$

$$\frac{\partial I_i}{\partial t} = \frac{1}{\tau_{inc}} E_i - \frac{1}{\tau_{inf}} I_i \quad (4)$$

$$\frac{\partial Q_m}{\partial t} = \sum_{i=1}^n \frac{p_m^i}{\tau_{inf}} I_i - (1/\tau_{recm}) Q_m \quad (5)$$

$$\frac{\partial Q_s}{\partial t} = \sum_{i=1}^n \frac{p_s^i}{\tau_{inf}} I_i - (1/\tau_{hosp}) Q_s \quad (6)$$

$$\frac{\partial Q_f}{\partial t} = \sum_{i=1}^n \frac{p_f^i}{\tau_{inf}} I_i - (1/\tau_{hosp}) Q_f \quad (7)$$

$$\frac{\partial H_s}{\partial t} = (1/\tau_{hosp}) Q_s - (1/\tau_{recs}) H_s \quad (8)$$

$$\frac{\partial H_f}{\partial t} = (1/\tau_{hosp}) Q_f - (1/\tau_{death}) H_f \quad (9)$$

$$\frac{\partial R_m}{\partial t} = (1/\tau_{recm}) Q_m \quad (10)$$

$$\frac{\partial R_s}{\partial t} = (1/\tau_{recs}) H_s \quad (11)$$

$$\frac{\partial D}{\partial t} = (1/\tau_{death}) H_f \quad (12)$$

These equations are normalized by the total population number N . For this model system the total population (sum of all model variables) is constant in time and equal to one. There is a flow of people from the S_i groups

$I_0 = 50.0$	Initial infectious
$R_0 = 5.0$	Reproduction Number $t \leq t_1$
$R_{t_1} = 0.63$	Reproduction Number $t_1 < t < t_2$
$\tau_{\text{inf}} = 2.9$	Infections time
$\tau_{\text{inc}} = 5.2$	Incubation time
$\tau_{\text{recm}} = 11.1$	Recovery time mild cases
$\tau_{\text{recs}} = 15.0$	Recovery time severe cases
$\tau_{\text{hosp}} = 5.0$	Time to hospitalization.
$\tau_{\text{death}} = 15.0$	Days to death
$p_f = 0.006$	Case fatality rate
$p_s = 0.012$	Hospitalization rate for severe cases

Table 1 A set of first guess model parameters

towards the \mathbf{R}_m , \mathbf{R}_s , and \mathbf{D} groups. For the population to stay constant the right-hand sides of the equations needs to sum to zero.

The speeds of the transitions from one group to the next are determined by the different time scales shown in Tab. 1. The most critical parameter in the model is the value of the reproduction number $R(t)$. When R is greater than one, one experiences exponential growth of the epidemic, while when R is less than one, the epidemic will die out. In the current model, R is a matrix $\mathbf{R}(t)$ that allows for different reproduction numbers in between different agegroups, and \mathbf{R} can change with time to reflect implementation of social distancing etc. Two first guess values of \mathbf{R} used in the examples below are the constants R_0 used before interventions were implemented, and the R_{t_1} used after the first interventions were implemented.

The different agegroups of infectious \mathbf{I}_i transition into the different quarantined groups of sick based on the fractions p_m^i , p_s^i , p_f^i . These fractional coefficients sum to one for each agegroup, and allows for specifying how different agegroups are differently affected by the virus. For each age group the fractions should sum to one. Additionally, the weighted average (by age group) of severely and fatally ill should be given by the specified p_s and p_f parameters in Tab. 1. Tab. 2 gives an example of the fractions used in the cases below.

3 Ensemble methods for model calibration

There is a vast litterature on the use of ensemble Kalman filter (EnKF) type methods for sequential data assimilation, model parameter estimation, and solving high-dimensional and nonlinear inverse problems. Ensemble data assimilation is now standard and state of the art in a majority of operational prediction systems in the geosciences (Carrassi et al 2018), including weather prediction (Houtekamer and Zhang 2016) and petroleum applications (Aanonsen et al 2009). The most popular data-assimilation methods build on the Ensemble Kalman Fil-

ter (EnKF) originally developed by Evensen (1994, 2004, 2009b,a).

Ensemble data assimilation in petroleum applications solves a parameter-estimation problem, where the non-linear behavior of a model prediction on the parameters poses a significant problem. Recent developments originating from the petroleum application have led to the use of new iterative ensemble smoothers such as ESM DA by Emerick and Reynolds (2013) and EnRML by Chen and Oliver (2012, 2013), which have become popular methods for solving inverse problems with moderate nonlinearity. Recently Raanes et al (2019); Evensen et al (2019) introduced a new efficient formulation of EnRML, where the solution is searched for in the ensemble subspace, and this method is now operational in petroleum applications Evensen et al (2019); Evensen (2020). In the current application we have used the ESM DA method based on its simplicity of implementation, and its computational efficiency with vast ensemble sizes. For a simple but complete explanation of ESM DA for solving parameter estimation problems see Evensen (2018).

4 A case study

We have used the ensemble methods with the model described above to run a sensitivity simulation to assess the impact of opening children schools and kindergartens after the 20th of April. The age-based model considers 11 age groups as defined in Tab. 2. The \mathbf{R} matrix used in the simulation after letting children back to kinder gardens and schools is presented in Tab. 4. Note that the p factors and the \mathbf{R} matrix are still uncertain and we should gather more data to calibrate these coefficients.

We have run one scenario where we continue today's measures with $R_{t_1} = 0.63$ as shown in Fig. 2, and then three scenaria with different \mathbf{R} matrices giving rise to a stable case Fig. 3, a neutral case Fig. 4, and an unstable case Fig. 5. A summary of results is presented in Fig. 1. More details on the infected age groups for these cases are presented in the two Figs. 6 and 7.

A conclusion based on these simulations is the following: *From these simulations, the opening of children's schools and kindergartens is likely to yield significant growth in new cases for the young age groups. The virus then spreads to their parents, teachers, and the rest of the community. A continuous lockdown in the remainder of society may stabilize the growth, but the measures must be sufficient, or we may experience exponential growth again. With open schools, it is not likely that the rest of the society will manage to keep R below one due to the additional commuting and people going back to work. The extreme transmission rate for the Corona virus makes the opening of parts of the society very risky. We will*

Age group	1	2	3	4	5	6	7	8	9	10	11
Age range	0–5	6–12	13–19	20–29	30–39	40–49	50–59	60–69	70–79	80–89	90–105
Population	351159	451246	446344	711752	730547	723663	703830	582495	435834	185480	45230
p-mild	1.0000	1.0000	1.0000	1.0000	1.0000	0.9640	0.9185	0.9210	0.8900	0.9070	0.9120
p-severe	0.0000	0.0000	0.0000	0.0000	0.0000	0.0360	0.0720	0.0600	0.0720	0.0360	0.0120
p-fatal	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0095	0.0190	0.0380	0.0570	0.0760

Table 2 The population numbers are obtained from SSB and are accurate. The total Norwegian population is 5367580. The p numbers indicate the fraction of sick people in an age group ending up with mild symptoms, severe symptoms (hospitalized), and fatal infection (hospitalized and then dead). With the current numbers, the average case fatality rate is 0.0090 and the average percentage of severe (hospitalized) cases is 0.0280.

Number of hospitalized	$N(\mathbf{H}_s + \mathbf{H}_f)$
Number of recovered	$N(\mathbf{R}_m + \mathbf{R}_s)$
Number of deaths	$N\mathbf{D}$
Number of exposed	$N \sum \mathbf{E}_i$
Number of infectious	$N \sum \mathbf{I}_i$
Number of susceptible	$N \sum \mathbf{S}_i$

Table 3 Diagnostic variables. The total population is N .

likely experience multiple new local exponential blooms of infected people.

In the simulations, we start with a large prior ensemble of sampled parameters and initial conditions and run this ensemble forward in time (see pale curves in the plots). Then we use an EnKF (ESMDA implementation) to update the prior ensembles of parameters and initial conditions by conditioning them to the measured deaths and hospitalization numbers. We then use the posterior ensemble of parameters to generate the posterior prediction (dark lines in the ensemble plots).

Some interesting results include:

1. The initial R is likely 4.5–5.0, i.e., much higher than originally thought, but larger numbers are now supported by the Imperial college studies.
2. We estimate the implemented measures to have reduced R to about 0.63 in Norway.
3. The CFR is around 0.6% and the fraction hospitalized is around 3.8% (these number depends of course on the accuracy of the actual number of infected people).

Some remarks which should encourage us to test extensively and without bias in the whole population:

1. We don't know the total number of infected and thus we don't know the basic reproduction number R .
2. We do know the number of deaths and the number of hospitalized, but since we don't know the actual number of infected, the fractions of severely and fatally ill are both unknown.
3. The official numbers are mostly based on a rough estimate of the total infected population (always larger than the number of positive tests).

Thus, in the current model and with the data we have available, we can adjust I_0 and R_0 to increase the popula-

tion of infected, and accordingly reduce the fractions of severely and fatally ill, to still match the observed deaths and hospitalized. Thus, we need one more piece of information to have a fully consistent model system. The fraction of the population testing positive could be this additional information, but for now these data are biased by selective testing and too little testing.

Thus, in the current simulations we started out with 50 initially infectious and 200 exposed at March 1st, assuming these numbers represent the number of imported cases at this point in time.

References

- Aanonsen SI, Naevdal G, Oliver DS, Reynolds A, Valles B (2009) Ensemble Kalman Filter in Reservoir Engineering – A Review. *SPE Journal*, SPE-117274-PA 14(3):393–412, doi: 10.2118/117274-PA.
- Carrassi A, Bocquet M, Bertino L, Evensen G (2018) Data Assimilation in the Geosciences: An overview on methods, issues and perspectives. *Wires Climate Change* 9(5):pp. 50, DOI /10.1002/wcc.535, <https://onlinelibrary.wiley.com/doi/10.1002/wcc.535>
- Chen Y, Oliver DS (2012) Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Math Geosci* 44:1–26
- Chen Y, Oliver DS (2013) Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Computat Geosci* 17:689–703
- Emerick AA, Reynolds AC (2013) Ensemble smoother with multiple data assimilation. *Computers and Geosciences* 55:3–15
- Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J Geophys Res* 99(C5):10,143–10,162
- Evensen G (2004) Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics* 54:539–560
- Evensen G (2009a) Data Assimilation: The Ensemble Kalman Filter, 2nd edn. Springer
- Evensen G (2009b) The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Systems Magazine* 29(3):83–104
- Evensen G (2018) Analysis of iterative ensemble smoothers for solving inverse problems. *Computat Geosci* 22(3):pp. 885–908, DOI 10.1007/s10596-018-9731-y, URL <https://doi.org/10.1007/s10596-018-9731-y>
- Evensen G (2020) Formulating the history matching problem with consistent error statistics. *Computat Geosci* Submitted
- Evensen G, Raanes P, Stordal A, Hove J (2019) Efficient implementation of an iterative ensemble smoother for data assimilation and reservoir history matching. *Frontiers in Applied Mathematics and Statistics* 5:47, DOI 10.3389/fams.2019.00047, URL <https://www.frontiersin.org/article/10.3389/fams.2019.00047>

- Houtekamer PL, Zhang F (2016) Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon Weather Rev* 144:4489–4533
- Raanes PN, Stordal AS, Evensen G (2019) Revising the stochastic iterative ensemble smoother. *Nonlin Processes Geophys* 26:325–338, DOI 10.5194/npg-2019-10, URL <https://doi.org/10.5194/npg-26-325-2019>

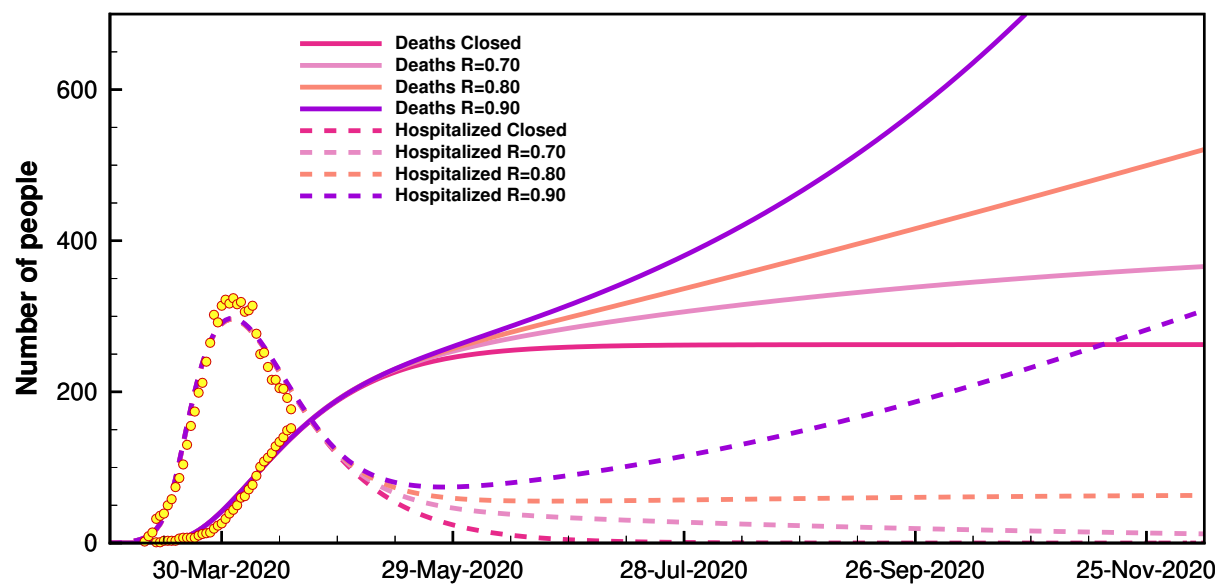


Fig. 1 Summary of cases with the different scenarios we may experience in case we start opening up measures after 20th of April. The plot shows the ensemble means of the number of hospitalized and the accumulated number of deaths for different scenarios. Then the three opening scenarios show that it is vital that one manages to retain R below 0.8 to avoid new exponential growth.

Age group	1	2	3	4	5	6	7	8	9	10	11
1	3.80	2.00	2.00	1.50	1.50	1.10	0.80	0.80	0.80	0.80	0.80
2	2.00	3.80	2.00	1.50	1.50	1.50	0.80	0.80	0.80	0.80	0.80
3	2.00	2.00	1.00	1.00	0.90	0.80	0.80	0.80	0.80	0.80	0.80
4	1.50	1.50	1.00	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
5	1.50	1.50	0.90	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
6	1.10	1.50	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
7	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
8	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
9	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
10	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
11	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80

Table 4 The R matrix allows for using different transmission factors in between different age groups. This matrix was used after opening up children schools and kinder gardens. On the diagonal the value gives the transmission of disease within the same age group. The off-diagonal terms are the transmissions between age groups. Here it is assumed that open kinder gardens and schools leads to “normal” transmission within these groups $R = 3.8$. We also assume that there are increased transmission between parent groups and children.

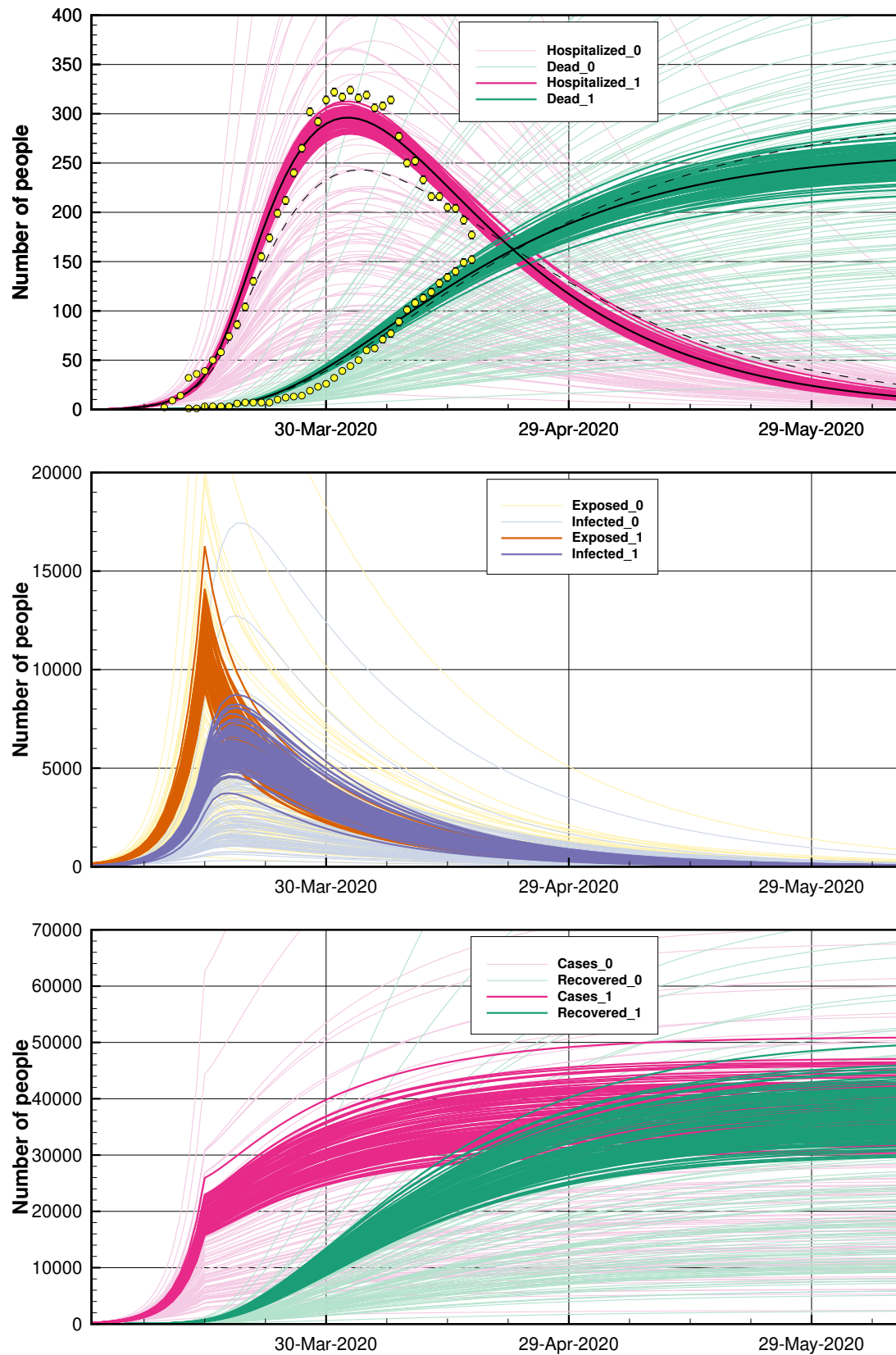


Fig. 2 Continue as today case: Stable case where the current interventions are continued with $R = 0.63$. The upper plot is an ensemble prediction of Hospitalized and accumulated Deaths. The pale colored lines are the prior realizations while the darker lines are the posterior realizations. The black dashed and solid lines are the ensemble means for the prior and the posterior. The middle plot shows the ensemble of exposed and infectious persons, while the lower plot gives the number of cases and recovered.

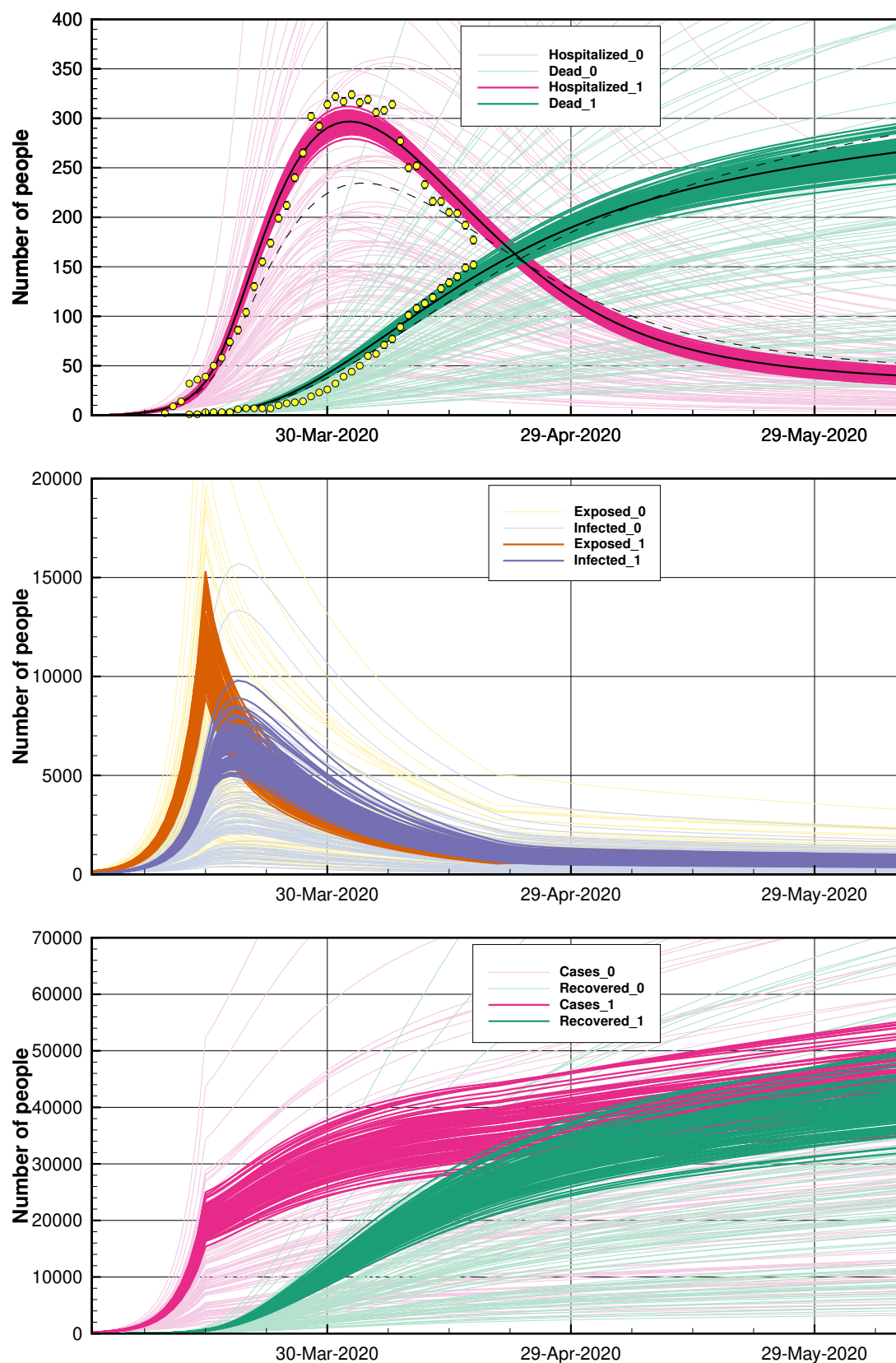


Fig. 3 Open up case 0.7: Stable case where we open kinder gardens and schools for children aged 6–12 years old after 20th of April. The \mathbf{R} matrix from Tab. 4 is used after 20th of April with all 0.8 numbers reduced to 0.7, to resemble today's effect on the measures.

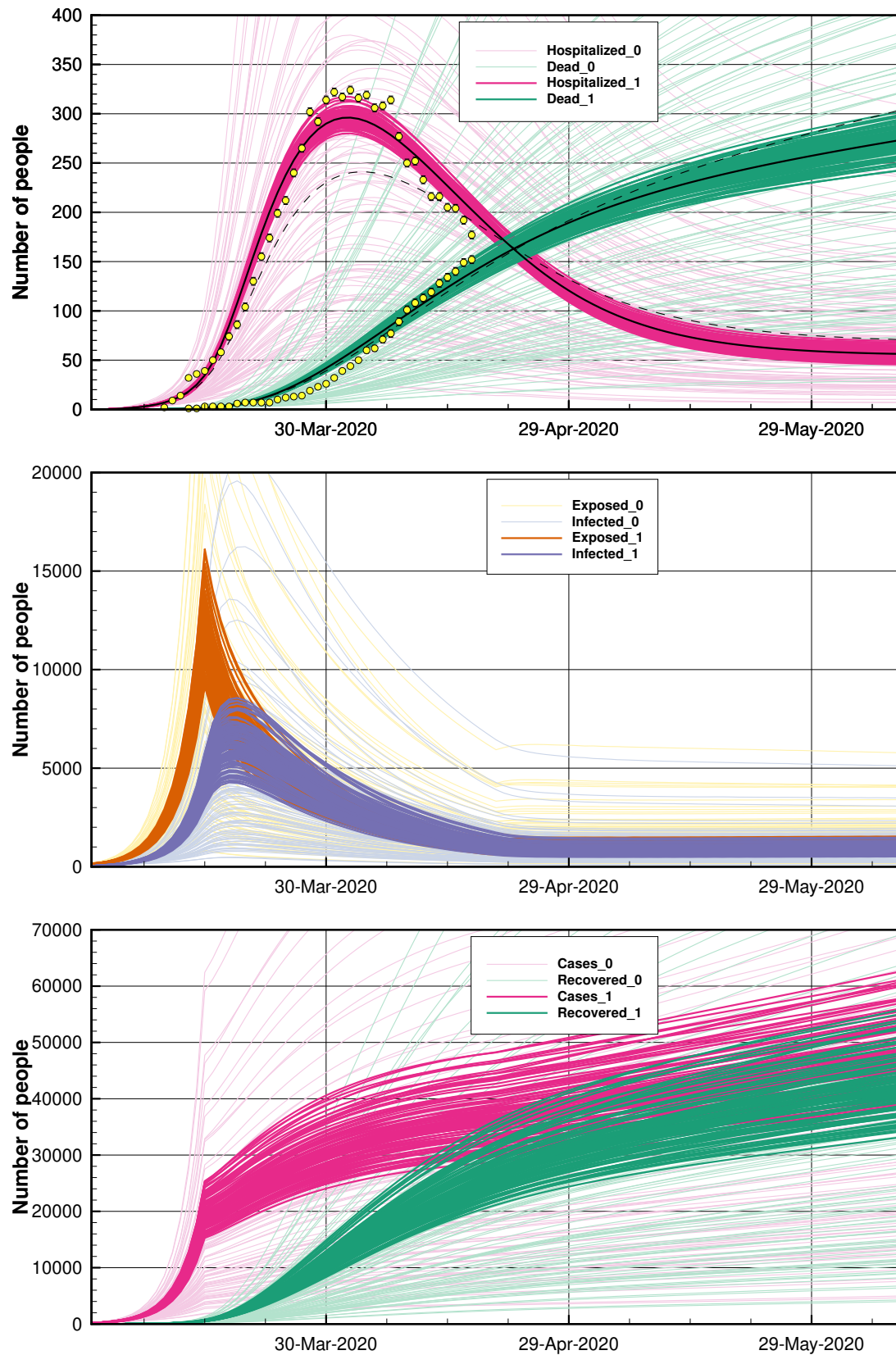


Fig. 4 Open up case 0.8: Neutral case where we open kinder gardens and schools for children aged 6–12 years old after 20th of April. The \mathbf{R} matrix from Tab. 4 is used after 20th of April with all 0.8 numbers are retained at 0.8.

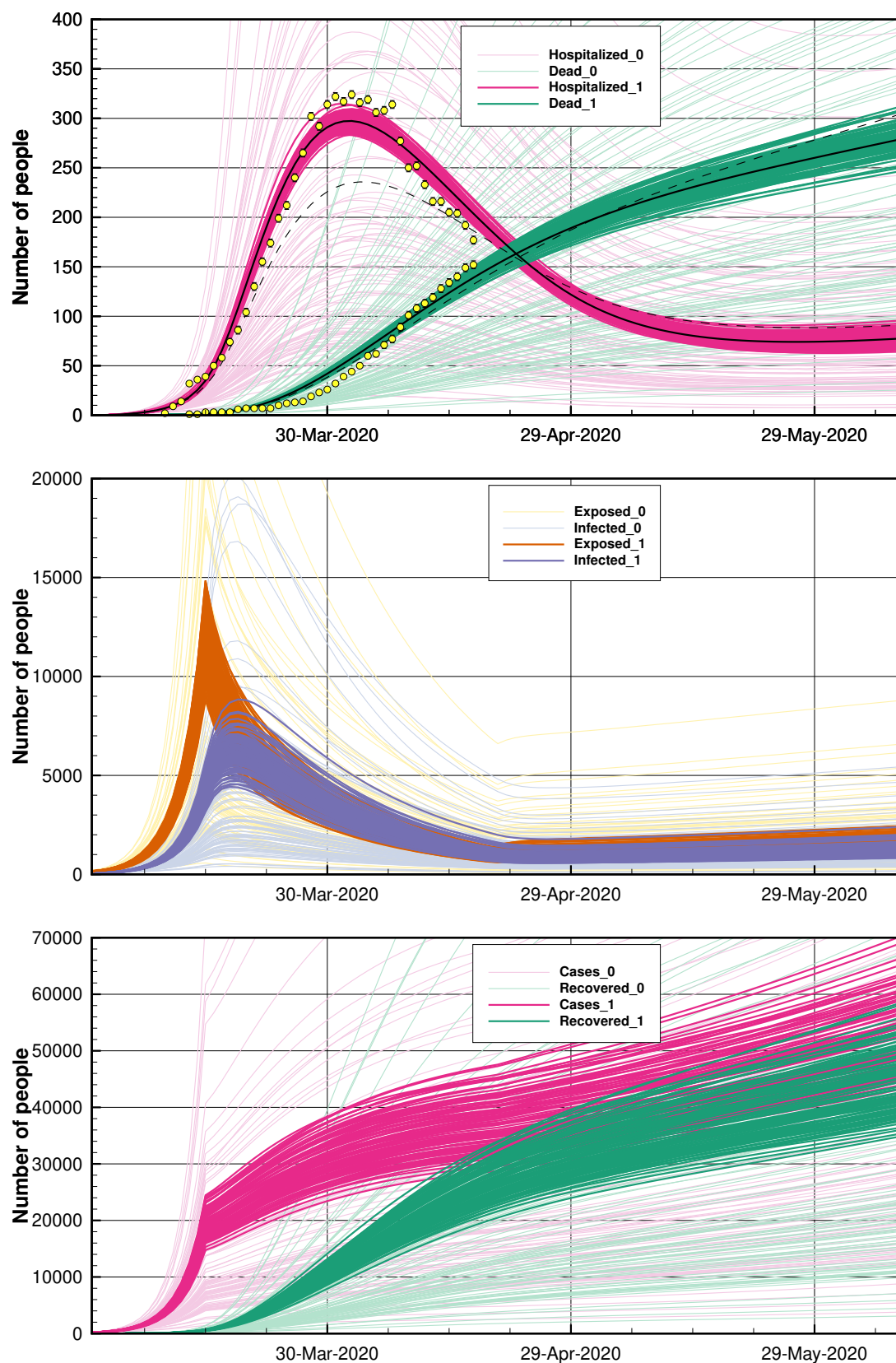


Fig. 5 Open up case 0.9: Unstable case where we open kinder gardens and schools for children aged 6–12 years old after 20th of April. The \mathbf{R} matrix from Tab. 4 is used after 20th of April but with all 0.8 numbers set to 0.9.

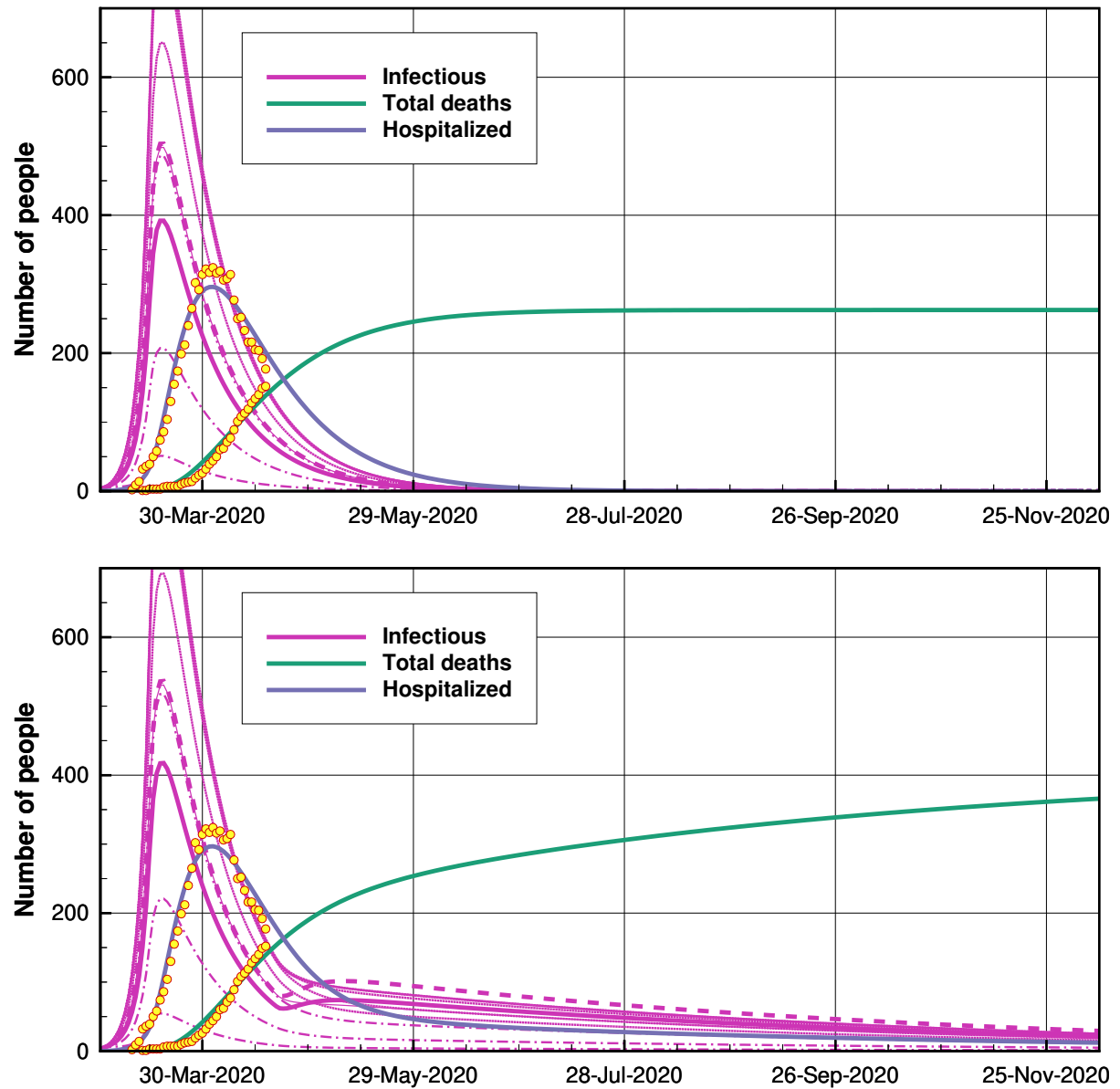


Fig. 6 These plots shows the number of infecteous in the different agegroups. The tick violet line is the youngest age group, the dashed line is the second youges, the dash-dot lines show the three oldes agegroups. Upper plot is the stable case where we stay closed, the lower plot is the case with opening but $R = 0.7$ in the adult population.

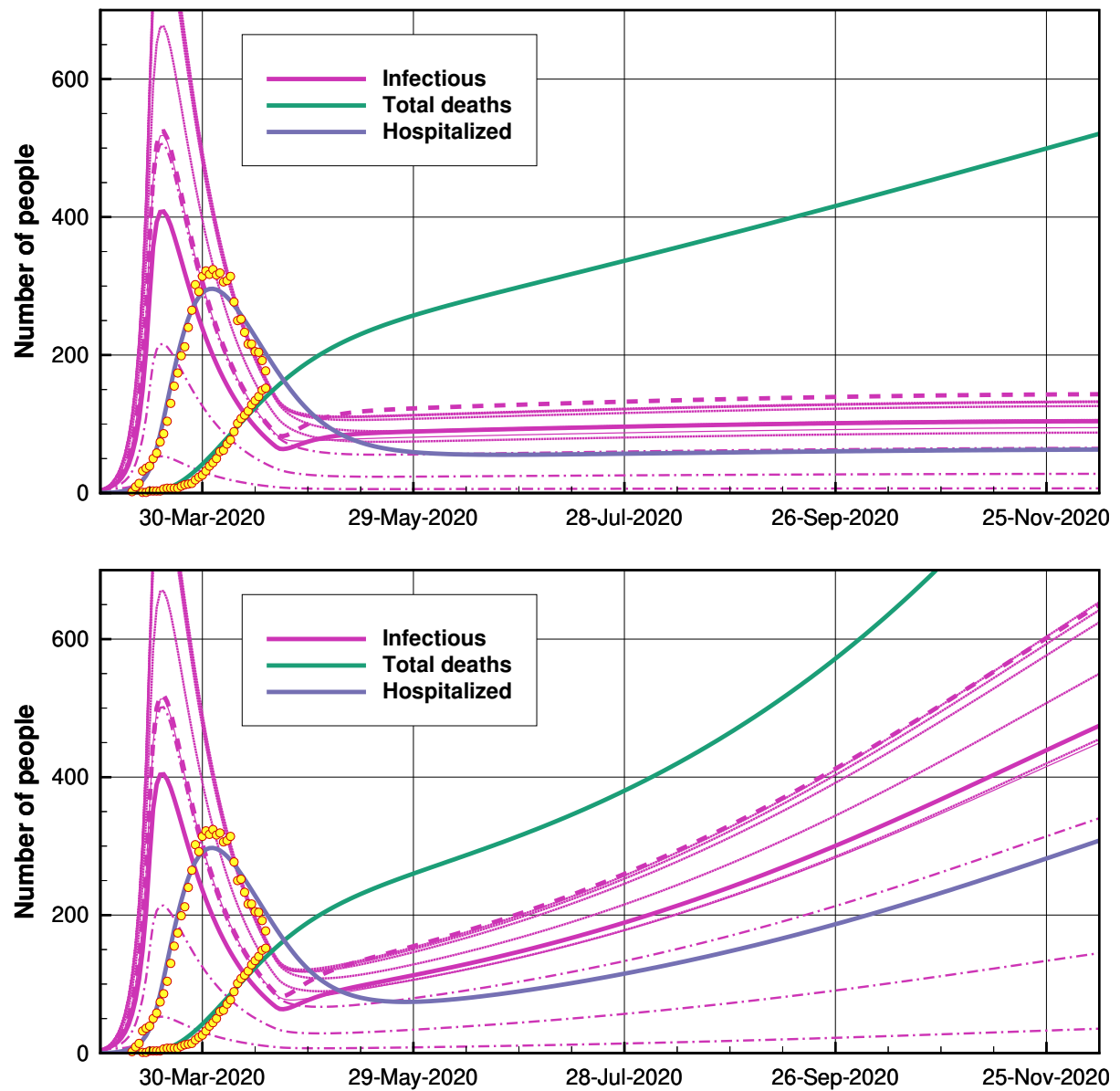


Fig. 7 Same as Fig. 6 Upper plot is the case for opening but with $R = 0.80$ in the adult population, the lower plot is the case with opening but $R = 0.9$ in the adult population.