Data suggest COVID-19 affected numbers greatly exceeded detected numbers, in four European countries, as per a delayed SEIQR model

Sankalp Tiwari
Indian Institute of Technology Kanpur
C. P. Vyasarayani (vcprakash@mae.iith.ac.in)
Indian Institute of Technology Hyderabad https://orcid.org/0000-0002-3396-0484
Anindya Chatterjee
Indian Institute of Technology Kanpur

Research Article

Keywords: COVID 19, undetected population, SEIQR, time delay, Pareto distribution

Posted Date: October 6th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-87296/v1

Version of Record: A version of this preprint was published at Scientific Reports on April 14th, 2021. See the published version at https://doi.org/10.1038/s41598-021-87630-z.

Data suggest COVID-19 affected numbers greatly exceeded detected numbers, in four European countries, as per a delayed SEIQR model

Sankalp Tiwari[1], **C. P. Vyasarayani**[2*], **and Anindya Chatterjee**[1]

[1]Mechanical Engineering, Indian Institute of Technology Kanpur, 208016, India.
[2]Mechanical and Aerospace Engineering, Indian Institute of Technology Hyderabad, 502285, India
[*]vcprakash@mae.iith.ac.in

ABSTRACT

People in many countries are now infected with COVID-19. By now, it is clear that the number of people infected is much more than the number of reported cases. To estimate the infected but undetected/unreported cases using a mathematical model, we can use a parameter called the probability of quarantining an infected individual. This parameter exists in the time-delayed SEIQR model (Scientific Reports, article number: 3505). Two limiting cases of a network of such models are used to estimate the undetected population. The first limit corresponds to the network collapsing onto a single node and is referred to as the mean-β model. In the second case, the number of nodes in the network is infinite and results in a continuum model, treating the infectivity as statistically distributed. We use a shifted Pareto distribution to model the infectivity. This distribution has a long tail and incorporates the presence of super-spreaders that contribute to the disease progression. While both the models capture the *detected* numbers equally well, the predictions of *affected* numbers from the continuum model are more realistic. Results suggest that affected people outnumber detected people by one to two orders of magnitude in Spain, UK, Italy, and Germany.

1 Introduction

For different countries around the world, several researchers [1–7] have concluded that the number of people actually infected, or *affected*, by COVID-19 is far greater than the number of cases actually reported, or *detected* officially. Recent serological surveys for COVID-19 also indicate that the infected people outnumber detected people by about 12 times in Spain[8] and 6 to 24 times in the USA[9]. Other serological surveys suggest that about 18% of people in London[10] and 23% of the people in New Delhi[11] were already infected by mid-April and early July, respectively, far outnumbering the reported cases.

In other words, affected numbers seem to greatly exceed detected numbers. To what extent can this difference be anticipated from purely data fitting of detected people, simple parameter estimation, and simple epidemiological models? That is the question we take up in this paper.

We fit two time-delayed SEIQR (Susceptible, Exposed, Infected, Quarantined or Isolated, Recovered/Removed) models to the numbers of reported cases against time, for four European countries. These countries were chosen because they are not extremely large and diverse (e.g., the USA and India), they have cultural differences amongst them, and yet they are geographically close to each other. In other words, they are different from each other but not vastly different.

These models are obtained by considering two limiting cases of a time-delayed network SEIQR model motivated by the model of Young *et al.*[12–14]. In the network model, the whole population is divided into N sub-populations based on their net infectivity (β) values, and each node represents a sub-population or group. In the first limiting case that we adopt, which is the same as a mean-β model[14], the entire network[13] is collapsed into a single node ($N = 1$). This model was originally proposed by Young *et al.*[12], and for a fast pandemic some simplifications and approximations are possible[14]. In the second limiting case[13], which is a continuum model, we take $N \to \infty$. Here, the infectivity is treated as a continuously distributed parameter in the population.

For Italy, Germany, UK, and Spain, we fit these two models to the data reported under the heading 'total cases' on the Worldometer website[15]. We consider the data from February 15 - June 18 for fitting (125 days). Beyond mid-June, all the countries seemed to be experiencing a second wave of COVID-19 after relaxing social distancing norms, or perhaps due to increased testing rates. Therefore, the constancy of parameters would no longer be a reasonable assumption.

Both these models include a parameter called the probability of detecting an infected individual. This parameter, upon fitting from detected population data, allows us to indirectly estimate the affected but undetected population. We will find that the continuum model fits the data better than the mean-β model. The affected people outnumber the detected people

by 8, 22, 48, and 130 times in Spain, UK, Italy, and Germany, respectively. The order of magnitude of these multiplying factors are consistent with those reported in the serological surveys mentioned earlier; however, it is emphasized that only the officially detected numbers are used for data fitting. The continuum model also suggests the presence of 'super-spreaders' (or 'super-spreading events') in all the countries, in the form of a long tail in the distribution of the infectivity β in the population.

The rest of this paper is organized as follows. In section 2, we briefly discuss the two models (mean-β and continuum) used in this work. In section 3, we present and discuss in detail the results of the optimization calculations (i.e., parameter fitting) for Italy. In section 4, we present the results for the remaining three countries: Germany, UK and Spain. In section 5, we present our conclusions.

2 SEIQR models

A detailed description of the mean-β model[14] ($N = 1$) and the continuum model[13] ($N \to \infty$) can be found in the literature. In this section, we describe them briefly for clarity and completeness.

2.1 Mean-β model

The mean-β model[14] can be derived from the five state SEIQR model of Young *et al.*[12] by assuming no loss of immunity after recovery. This assumption is valid for a fast pandemic like COVID-19. In the mean-β model, exposed (E_m), quarantined (Q_m), and recovered (R_m) states becomes slave variables of susceptible (S_m) and infected (I_m) states whose dynamics are governed by the following DDEs:

$$\dot{S}_m(t) = -\beta_m S_m(t) I_m(t), \tag{1}$$

$$\dot{I}_m(t) = \beta_m S_m(t-\sigma_m) I_m(t-\sigma_m) - p_m e^{-\gamma_m \tau_m} \beta_m S_m(t-\sigma_m-\tau_m) I_m(t-\sigma_m-\tau_m) - \gamma_m I_m(t). \tag{2}$$

$$\tag{3}$$

The parameters p_m, γ_m, τ_m, β_m, and σ_m are described in Table 1. The subscript m in all the quantities serves to distinguish them from those used in the continuum model ($N \to \infty$). By defining

$$V(t) = \int_{-\infty}^{t} I_m(\eta)\, d\eta, \tag{4}$$

and integrating Eq. 1, we get

$$S_m(t) = e^{-\beta_m V(t)}, \tag{5}$$

where we have imposed the initial condition $S_m(-\infty) = 1$. Inserting Eq. 4 and Eq. 5 into Eq. 2, we obtain

$$\ddot{V}(t) = \beta_m e^{-\beta_m V(t-\sigma_m)} \dot{V}(t-\sigma_m) - p_m e^{-\gamma_m \tau_m} \beta_m e^{-\beta_m V(t-\sigma_m-\tau_m)} \dot{V}(t-\sigma_m-\tau_m) - \gamma_m \dot{V}(t).$$

Integrating both sides of the above equation and by defining

$$\bar{p}_m = p_m e^{-\gamma_m \tau_m}, \tag{6}$$

we obtain

$$\dot{V}(t) = \bar{p}_m e^{-\beta_m V(t-\sigma_m-\tau_m)} - e^{-\beta_m V(t-\sigma_m)} - \gamma_m V(t) + 1 - \bar{p}_m. \tag{7}$$

The complete dynamics of the pandemic in the mean-β can be captured by the first-order nonlinear DDE given by Eq. 7. The number of people detected as having contracted the disease is given by

$$h_m(t) = T \bar{p}_m \beta_m \int_{-\infty}^{t} e^{-\beta_m V(t-\sigma_m-\tau_m)} \dot{V}(t-\sigma_m-\tau_m)\, dt, \tag{8}$$

and the total number of people infected (detected plus undetected) till time t is

$$w_m(t) = T(1 - e^{-\beta_m V(t)}), \tag{9}$$

wherein T is the total population. The biological parameters σ_m and γ_m are fixed at values reported in the COVID-19 literature[16–18].

Table 1. Parameters used in the mean-β model

S. No.	Parameter	Description	Constraints	Specified/Estimated
1	σ_m	Asymptomatic and non-infectious period	$\sigma_m = 3$	Specified
2	τ_m	Infectious but asymptomatic period	$14 \geq \tau_m \geq 1$	Estimated
3	γ_m	Self-recovery rate	$\gamma_m = 0.07$	Specified
4	p_m	Probability of quarantining symptomatics	$1 \geq p_m \geq 0$	Estimated
5	β_m	Infectivity constant	$\beta_m > 0$	Estimated
6	V_0	Constant history of V	$V_0 > 0$	Estimated

2.2 Continuum model

The other limit of the network model[13] is for the case of $N \to \infty$, which implies that the infectivity (β) is now distributed continuously over the population. The governing differential equations for the states S and I in this case are as follows:

$$\dot{S}(\beta,t) = -\sqrt{\beta}S(\beta,t)\int_0^\infty \sqrt{\xi}I(\xi,t)\,d\xi, \tag{10}$$

$$\dot{I}(\beta,t) = \sqrt{\beta}S(\beta,t-\sigma)\int_0^\infty \sqrt{\xi}I(\xi,t-\sigma)\,d\xi - pe^{-\gamma\tau}\sqrt{\beta}S(\beta,t-\sigma-\tau)\int_0^\infty \sqrt{\xi}I(\xi,t-\sigma-\tau)\,d\xi - \gamma I(\beta,t), \tag{11}$$

where p, γ, τ, and σ are described in Table 2. It was shown[13] that $S(\beta,t)$ admits a solution of the form:

$$S(\beta,t) = \phi(\beta)e^{-f(t)\sqrt{\beta}}. \tag{12}$$

Therefore, if $\phi(\beta)$ is determined using initial conditions, the variation of S over β is simply through $f(t)$. Using algebraic manipulations, it is shown that $f(t)$ satisfies the following non-linear DDE[13] :

$$\dot{f}(t) = -G(f(t-\sigma)) + pe^{-\gamma\tau}G(f(t-\sigma-\tau)) - \gamma f(t) + C_0, \tag{13}$$

where

$$G(f(t)) = \int_0^\infty \sqrt{\beta}\phi(\beta)e^{-f(t)\sqrt{\beta}}\,d\beta, \tag{14}$$

and

$$C_0 = (1-pe^{-\gamma\tau})\int_0^\infty \sqrt{\beta}\phi(\beta)\,d\beta.$$

The quantities of interest are

$$h(t) = Tpe^{-\gamma\tau}\int_{-\infty}^{t} \dot{f}(\bar{t}-\sigma-\tau)G(f(\bar{t}-\sigma-\tau))\,d\bar{t} \tag{15}$$

and

$$w(t) = T\left(1 - \int_0^\infty \phi(\xi)e^{-f(t)\sqrt{\xi}}\,d\xi\right), \tag{16}$$

where $h(t)$, $w(t)$ have the same meaning as in the mean-β model (except we have dropped the subscript m). In Eq. 16, ϕ is the initial distribution of infectivity in the population. In the present work, we assume $\phi(\beta)$ to be of the form

$$\phi(\beta) = \frac{(m-1)a^{m-1}}{(a+\beta)^m}, \quad \beta \geq 0,$$

which is a shifted Pareto distribution[19]. Note that

$$\int_0^\infty \phi(\beta)\,d\beta = 1.$$

A summary of the parameters used in the continuum model is presented in Table 2.

Table 2. Parameters used in the continuum model.

S. No.	Parameter	Description	Constraints	Specified/Estimated
1	σ	Asymptomatic and non-infectious period	$\sigma = 3$	Specified
2	τ	Infectious but asymptomatic period	$14 \geq \tau \geq 1$	Estimated
3	γ	Self-recovery rate	$\gamma = 0.07$	Specified
4	p	Probability of quarantining symptomatics	$1 \geq p \geq 0$	Estimated
5	a	Parameter in initial infectivity distribution	$a > 0$	Estimated
6	m	Denominator exponent in initial infectivity distribution	$m > 2$	Estimated
7	f_0	Constant history of f	$f_0 > 0$	Estimated

2.3 The fitting error

In the mean-β model, for a given set of parameter values, we compute $h_m(t)$ and fit it with the data for the total number of detected cases as reported on the Worldometer website[15]. This is done by minimizing the fitting error

$$E_{0m} = \frac{||h_m - \text{data}||_2}{||\text{data}||_2} \times 100. \tag{17}$$

We see from Table 1 that there are four parameters to be identified in the mean-β model. The fitting error for the continuum model is defined as

$$E_0 = \frac{||h - \text{data}||_2}{||\text{data}||_2} \times 100. \tag{18}$$

We see from Table 2 that there are five parameters to be identified in the continuum model.

3 The case of Italy

The data for the detected cases happened to match extremely well for Italy. Therefore, in this section, we present detailed results for Italy obtained from the two models. The results for other countries will be presented in the next section.

3.1 Results for the mean-β model

We minimize the fitting error E_{0m} (see Eq. 17) using the optimization routine `fminsearch` in MATLAB. Since there are four free parameters in the mean-β model, the input variable for the optimization code is a four-by-one column vector, suitably transformed so that the constraints in Table 1 are automatically satisfied. We have performed several hundred optimization calculations with random initial conditions and have found many converging solutions. Several of these solutions correspond to nearly identical and low values of E_{0m}. Several other local minima yielded significantly higher E_{0m} values, and were discarded.

The parameter set that yields the lowest E_{0m} in all the random trials is reported in the first row of Table 3. The fit generated using these parameters, along with the reported data, is shown in Figure 1a. The reported data, which records the number of detected cases in Italy from February 15 for the following 125 days, is plotted in red circles. For clarity, only the data of alternate days is plotted. To account for the initial uncertainty in the reporting, we ignore the data of the first few days. Specifically, we neglect initial data where the number of cases is less than 1% of the number reported on the 125$^{\text{th}}$ day. The fitted $h_m(t)$ is plotted for a longer duration using a solid line to depict the saturation value clearly. In the figure, the total number of detected cases saturates at 0.238 million (0.4% of Italy's population).

Table 3. Parameter sets from mean-β model yielding the lowest E_{0m}, and subsidiary quantities.

Country	β_m	p_m	τ_m	$\bar{p}_m = p_m e^{-\gamma_m \tau_m}$	V_0	E_{0m}	A/D	R_0
Italy	0.1825	0.0112	12.1157	0.0048	0.4861	1.8771	228	2.5946
Germany	0.2097	0.0069	13.8161	0.0026	0.5211	2.6831	426	2.9879
UK	0.1636	0.0137	12.2077	0.0058	0.4773	1.6212	185	2.3236
Spain	0.1785	0.0194	12.5517	0.0081	0.7386	2.480	142	2.5293

We also plot the total number of infected people ($w_m(t)$) in Figure 1b with a solid curve. From this model, the total number of people infected in Italy saturates at 54.26 million (89% of Italy's population), which seems too high. In this sense, the mean-β model is unsatisfactory.

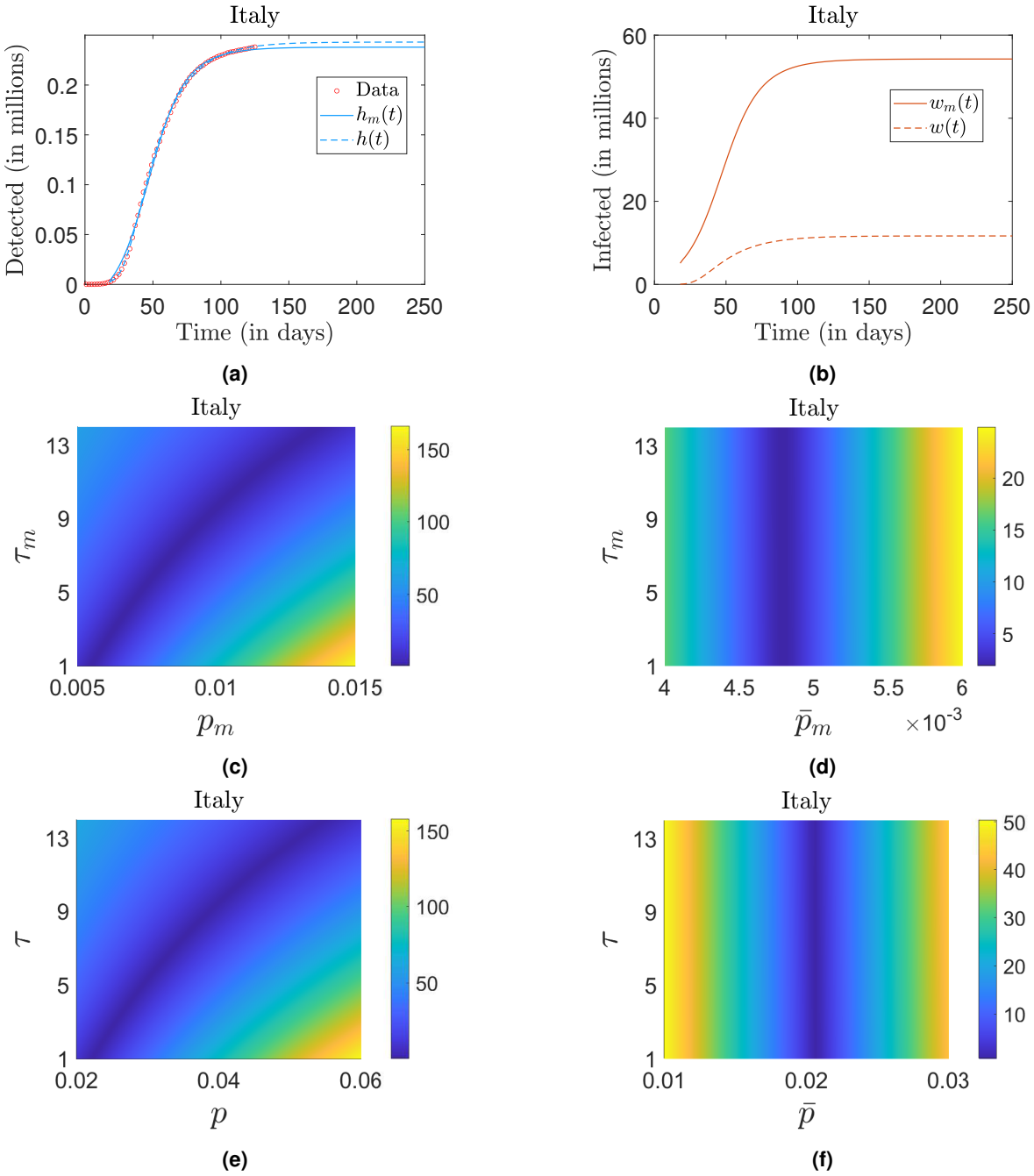

Figure 1. (a) Fitted results for Italy (population = 60 million, $h_m(250) = 0.2380$ and $h(250) = 0.2431$). Data for detected cases, obtained from Worldometer, is plotted using red circles. We have plotted the data of alternate days for clarity. The fit to the detected cases obtained using the mean-β model is shown by a solid line, and that using the continuum model is shown by a dashed line. The parameters used in the mean-β model and continuum model for obtaining the fit are reported in row 1 of Table 3 and Table 4, respectively. (b) Number of infected people obtained from the mean-β model (solid curve) and the continuum model (dashed curve), respectively ($w_m(250) = 54.2598$ and $w(250) = 11.6612$). (c) Variation of E_{0m} in the $p_m - \tau_m$ plane (for low values of p_m) obtained using the mean-β model. The parameters β_m and V_0 are fixed at the values reported in row 1 of Table 3. (d) Variation of E_{0m} in the $\bar{p}_m - \tau_m$ plane. (e) Variation of E_0 in the $p - \tau$ plane (for low values of p) obtained from the continuum model. The parameters a, m, and f_0 are fixed at the values reported in row 1 of Table 4. (f) Variation of E_0 in the $\bar{p} - \tau$ plane.

For this same model, the ratio of the population affected to population detected ($A/D = w_m/h_m$) saturates at 224. The basic reproduction number[12] R_0, for the mean-β model, is found from fitted parameters to be

$$R_0 = \beta_m \left(\frac{1-\bar{p}_m}{\gamma_m} \right) = 2.5946.$$

The mean-β model does offer some further useful insights into data fits, as follows. Upon inspection of the local minima obtained from the `fminsearch` runs, we noted that all the minima corresponding to low values of E_{0m} have nearly identical β_m and V_0 (equal to the values reported in the first row of Table 3), but different values for p_m and τ_m. The fitted values of p_m were consistently low, as well. To investigate further, we fix the values of β_m and V_0, and plot E_{0m} in the $p_m - \tau_m$ plane, in Figure 1c, for low values of p_m. We see that the lowest values for E_{0m} are obtained on a thin band cutting across the $p_m - \tau_m$ plane, which spans the entire range of τ_m and a small range of $0.001 \leq p_m \leq 0.015$. From this plot, we conclude that the mean-β model yields robust estimates for the parameters β_m and V_0; however, τ_m and p_m remain indeterminate. The indeterminacy is high in τ_m while p_m lies in a small range centered around $p_m = 0.01$ (which corresponds to quarantining or isolation of about one in a hundred symptomatic cases). Finally, upon plotting E_{0m} in the $\bar{p}_m - \tau_m$ plane in Figure 1d, we observe that the thin band of Figure 1c corresponds to almost fixed value of $\bar{p}_m \approx 0.0048$, indicating that $\bar{p}_m$ can be robustly identified. We now discuss the superior results obtained using the continuum model.

3.2 Results for the continuum model

In this case, there are five free parameters to be estimated. Therefore, the input variable to the optimization code is a 5×1 vector, suitably transformed so that the constraints in Table 2 are automatically satisfied. The parameter set that results in the lowest value of E_0 for all the optimization trials is reported in the first row of Table 4. The corresponding fit $h(t)$, again plotted for a longer duration to show saturation, is plotted using a dashed line in Figure 1a. The figure indicates that the fit to the detected data from the continuum model is slightly better than that from the mean-β model (numerically, $E_0 = 0.68$ for the continuum model, while $E_{m0} = 1.88$ for the mean-β model). We also note that the total number of detected cases saturate at 0.2431 million (0.4% of Italy's population), which is only slightly more than that predicted by the mean-β model. The optimizing value of m is found to be 2.6688 and corresponds to a long tail in $\phi(\beta)$'s Pareto distribution as shown in Figure 2.

Table 4. Parameter sets from continuum model yielding the lowest value of E_0 and subsidiary quantities.

Country	a	m	p	τ	$\bar{p} = pe^{-\gamma\tau}$	f_0	E_0	A/D	R_0
Italy	0.1351	2.6688	0.0223	1.0659	0.0207	0.0020	0.6800	48	2.8260
Germany	0.1503	2.5882	0.0199	13.6947	0.0076	0.0008	1.0597	130	3.6223
UK	0.1185	2.7392	0.0537	2.5073	0.0450	0.0048	0.8896	22	2.1856
Spain	0.0527	2.4357	0.1768	5.7101	0.1186	0.0002	1.2844	8	1.5236

A large difference is seen in the estimated number of affected people. We plot $w(t)$ in Figure 1b using a dashed line. The continuum model predicts that the total number of people infected, or affected, in Italy saturates at 11.66 million (19% of Italy's population), with an affected-to-detected ratio (A/D) of 48 at saturation. The basic reproduction number is found from fitted parameters to be

$$R_0 = \beta \left(\frac{1-\bar{p}}{\gamma} \right) = 2.8260.$$

Here, $\beta = \int_0^\infty \xi \, \phi(\xi) \, d\xi$.

Upon inspecting the local minima obtained from the `fminsearch` runs, we found that all the minima corresponding to low values of E have nearly identical values of a, m and f_0 (reported in the first row of Table 4) but different values of p and τ. These observations are similar to those from the mean-β model. Moreover, the values of p are low, while the values of τ vary over its entire range. For more insight, we fix a, m and f_0, and plot E_0 in the $p - \tau$ plane (for low values of p) as shown in Figure 1e. We see that the lowest values of E_0 are obtained on a thin band cutting across the $p - \tau$ plane, spanning the entire range of τ and a small range of $0.02 \leq p \leq 0.06$. Similar to the estimation results of the mean-β model, τ and p remain indeterminate even for the continuum model. The degree of indeterminacy is high in τ, while the corresponding values of p lie in a narrow range centered around $p = 0.04$. This corresponds to quarantining or isolation of about one in twenty five symptomatic cases. Upon plotting E_0 in the $\bar{p} - \tau$ plane in Figure 1f, we observe that the thin band of minimum vaues corresponds to an almost fixed value of $\bar{p} \approx 0.0207$.

In the next section, we report results for Germany, UK, and Spain. We will see that the main features of the results reported for the case of Italy hold for these countries as well.

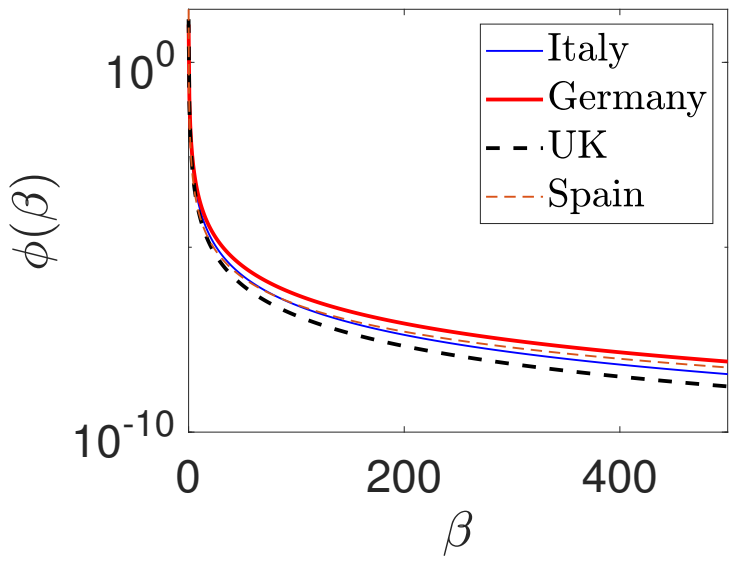

Figure 2. Plots of the long-tail distribution ($\phi(\beta)$) as used in the continuum model for Italy, Germany, UK, and Spain.

4 The cases of Germany, UK, and Spain

The best fits obtained using the two models for these three countries are presented in Figure 3. For Germany (see Figure 3a), the continuum model under-predicts and for Spain (see Figure 3e), the continuum model over-predicts the actual data near the end. However, for the UK (see Figure 3c), the fit is excellent. The variation of E_{0m} obtained from the mean-β model in the $\bar{p}_m - \tau_m$ plane (with β_m and V_0 fixed at values reported in Table 3) is plotted on the left panels of Figure 4 (a, c, e). The variation of E_0 obtained from the continuum model in the $\bar{p} - \tau$ plane (with a, m and f_0 fixed at values reported in Table 4) is plotted on the right panels of Figure 4 (b, d, f). The affected-to-detected (A/D) ratio and the reproduction number (R_0) corresponding to the best fits for the mean-β model and the continuum model are reported in Table 3 and Table 4, respectively.

We see that the fitting results are qualitatively similar to those obtained for Italy. However, there are a few observations that stand out in the results for the continuum model as highlighted below:

- We see in Table 4 that the optimum value of m for Germany, UK, and Spain is around 2.5. This indicates that the infectivity distribution $\phi(\beta)$ for each of these countries has a long tail as can be in Figure 2. We see from the figure that most of the population has small infectivity (low value of β). However, these curves decay to zero very slowly, since the second moment of $\phi(\beta)$ is infinite, i.e.

$$\int_{\beta=0}^{\infty} \beta^2 \phi(\beta)\, d\beta = \infty.$$

  Such a distribution indicates the presence of 'super-spreaders' (people who transmit the virus to a much larger number of people compared to the value of basic reproduction number (R_0)). Alternatively, 'super-spreading events' in the initial stage may also have played an important role in transmitting the virus to a larger population. The existence of super-spreaders or super-spreading events is well-known for SARS[20]. However, it has not been demonstrated yet for COVID-19 using mathematical modelling. Contrarily, super-spreading events have been widely reported for COVID-19 in the medical literature[21,22].

- We see from Table 4 that the affected-to-detected ratio (A/D) is high for all the countries and varies between 8 for Spain and 130 for Germany.

- The ratio of symptomatic cases to detected cases can be approximated from the value of p. Symptomatic cases outnumber the reported cases by about 5 times in Spain, 12 times in UK, 25 times in Italy, and 60 times in Germany.

- It should be noted that the fits obtained from the continuum model are best for Italy and the UK, as compared to Germany and Spain. Therefore, the reported affected-to-detected ratio (A/D) is more reliable for Italy and the UK. Also, due to fitting errors (see Figure 3e), $\bar{p}_m$ cannot be uniquely identified for Spain (see Figure 4f).

5 Conclusions

In this work, we fit the data for the total number of infected people in four western European countries. We use two limiting cases of the time-delayed network SEIQR model: the mean-β model and the continuum model. In earlier works, it was shown that for fast pandemics, each of these two models reduces to one non-linear delay differential equation.

After fixing the values of the biological parameters σ and γ, we need to identify four parameters in the mean-β model and five parameters in the continuum model. In both the cases, we see that there are many parameter sets that minimize the fitting error, yielding almost identical values of the objective function. All these sets have almost identical values of all parameters other than p and τ. Other subsidiary quantities such as the total number of infected people, the affected-to-detected ratio, and the basic reproduction number are also close to each other in value. By plotting the fitting error in the $p-\tau$ plane (with other parameters fixed at their identified values), we see a narrow band yielding minimum error cuts across this plane, spanning the entire range of τ and a small range of p comprising low values.

We see from the results that the continuum model yields superior fits in comparison to the mean-β model. The worst fit obtained from the continuum model is for Spain, with the 2-norm fitting error being only 1.28%. Moreover, it gives reasonable and physically realizable values for all the epidemiological quantities.

The most important conclusion from the models is that both, the total number of affected people (including people who remain asymptomatic) and people who show symptoms, far outnumber the people detected with COVID-19 in all the four countries. The continuum model predicts that the affected-to-detected ratio, in increasing order, is 8, 22, 48, and 130 for Spain, UK, Italy, and Germany, respectively. The symptomatic cases outnumber the reported cases by about 5 times for Spain, 12 times for UK, 25 times for Italy, and 60 times for Germany. These numbers are of the same order as of those reported in recent serological surveys conducted in various countries. The continuum model also indicates the presence of either super-spreaders or super-spreading events in all the four countries during the initial stage of disease progression. This is characterized by the long tails of the corresponding infectivity distribution for each country.

Funding

No funding was received for conducting this research.

Author contributions statement

All authors were equally involved in this research and in writing the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. 2020 Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**(6490) 489–493. (doi:10.1126/science.abb3221).
2. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, Whittaker C, Zhu H, Berah T, Eaton JW and others. 2020 Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584** 257–261. (doi:10.1038/s41586-020-2405-7).
3. Unwin H J T, Mishra S, Bradley VC, Gandy A, Mellan TA, Coupland H, Ish-Horowicz J, Vollmer MAC, Whittaker C, Filippi SL and others. 2020 State-level tracking of COVID-19 in the United States. *medRxiv*. (doi:10.1101/2020.07.13.20152355).
4. Shekatkar S, Pujari B, Arjunwadkar M, Hazra DK, Chaudhuri P, Sinha S, Menon GI, Sharma A, Guttal V. 2020 INDSCI-SIM A state-level epidemiological model for India. Ongoing study at www.indscicov.in/indscisim.
5. Roser M, Ritchie H, Ortiz-Ospina E, Hasell J. 2020 Coronavirus Pandemic (COVID-19). *Our World in Data*. (www.ourworldindata.org/coronavirus).
6. Bohk-Ewald C, Dudel C, Myrskylä M. 2020 A demographic scaling model for estimating the total number of COVID-19 infections. *arXiv preprint*. arXiv:2004.12836 (doi:10.1101/2020.04.23.20077719).
7. Scudellari M. 2020 How the pandemic might play out in 2021 and beyond. *Nature* **584**(7819) 22–25. (https://www.nature.com/articles/d41586-020-02278-5).
8. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, Oteo J, Hernán MA, Pérez-Olmeda M, Sanmartín JL, Fernández-García A, Cruz I, de Larrea NF and others. 2020 Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *The Lancet* **396**(10250) 535–544. (doi:10.1016/S0140-6736(20)31483-5).

9. Havers FP, Reed C, Lim T, Montgomery JM, Klena JD, Hall AJ, Fry AM, Cannon DL, Chiang CF, Gibbons A and others. 2020 Seroprevalence of antibodies to SARS-CoV-2 in 10 sites in the United States, March 23-May 12, 2020. *JAMA Internal Medicine*. (doi:10.1001/jamainternmed.2020.4130).

10. https://www.gov.uk/government/publications/national-covid-19-surveillance-reports/sero-surveillance-of-covid-19.

11. https://indianexpress.com/article/cities/delhi/delhi-serological-survey-covid-19-icmr-6516208/.

12. Young LS, Ruschel S, Yanchuk S, Pereira T. 2019 Consequences of delays and imperfect implementation of isolation in epidemic control. *Scientific reports* **9**(1) 1–9. (doi:10.1038/s41598-019-39714-0).

13. Vyasarayani CP, Chatterjee A. 2020 Complete dimensional collapse in the continuum limit of a delayed SEIQR network model with separable distributed infectivity. *Nonlinear Dynamics*. (doi:10.1007/s11071-020-05785-2).

14. Vyasarayani CP, Chatterjee A. 2020 New approximations, and policy implications, from a delayed dynamic model of a fast pandemic. *Physica D: Nonlinear phenomena* **414**(132701). (doi:10.1016/j.physd.2020.132701).

15. https://www.worldometers.info/coronavirus/.

16. Linka K, Peirlinck M, Sahli CF, Kuhl E. 2020 Outbreak dynamics of COVID-19 in Europe and the effect of travel restrictions. *Computer Methods in Biomechanics and Biomedical Engineering* **23**(11) 710–717. (doi:10.1080/10255842.2020.1759560).

17. Park SW, Bolker BM, Champredon D, Earn DJD, Li M, Weitz JS, Grenfell BT, Dushoff J. 2020 Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: framework and applications to the novel coronavirus (SARS-CoV-2) outbreak. *Journal of the Royal Society Interface* **17**(20200144). (doi:10.1098/rsif.2020.0144).

18. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, Colaneri M. 2020 Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine* **26** 855–860. (doi:10.1038/s41591-020-0883-7).

19. Arnold BC. 2015 Pareto distribution. *Wiley StatsRef: Statistics Reference Online* 1-10. (doi:10.1002/9781118445112.stat01100.pub2).

20. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**(132701) 355–359. (doi:10.1038/nature04153).

21. https://www.washingtonpost.com/graphics/2020/world/coronavirus-south-korea-church/.

22. https://www.scientificamerican.com/article/how-superspreading-events-drive-most-covid-19-spread1/.

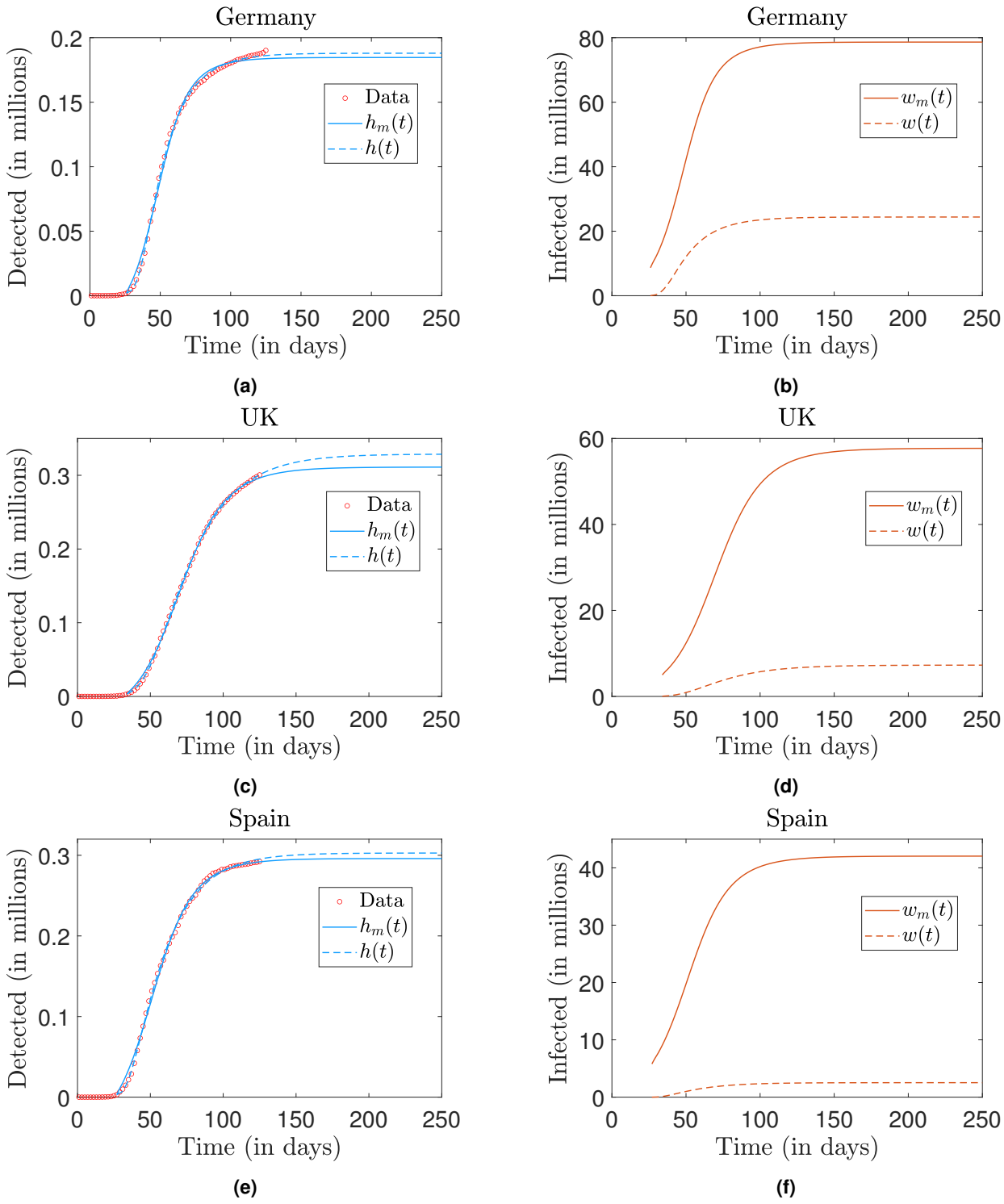

Figure 3. (a) and (b) Fitted results for Germany (population = 84 million, $h_m(250) = 0.1847$, $h(250) = 0.1880$, $w_m(250) = 78.6504$ and $w(250) = 24.3990$); (c) and (d) fitted results for UK (population = 67 million, $h_m(250) = 0.3112$, $h(250) = 0.3286$, $w_m(250) = 57.7096$ and $w(250) = 7.2773$); (e) and (f) fitted results for Spain (population = 47 million, $h_m(250) = 0.2960$, $h(250) = 0.3028$, $w_m(250) = 42.0433$ and $w(250) = 2.5281$). Data for detected cases, obtained from Worldometer, is plotted using red circles. We have plotted the data of alternate days for clarity. The fits obtained from the mean-β model are shown using solid lines, while those from the continuum model are shown using dashed lines. The parameters used in the mean-β model and continuum model for obtaining the fit are shown in Table 3 and Table 4, respectively.

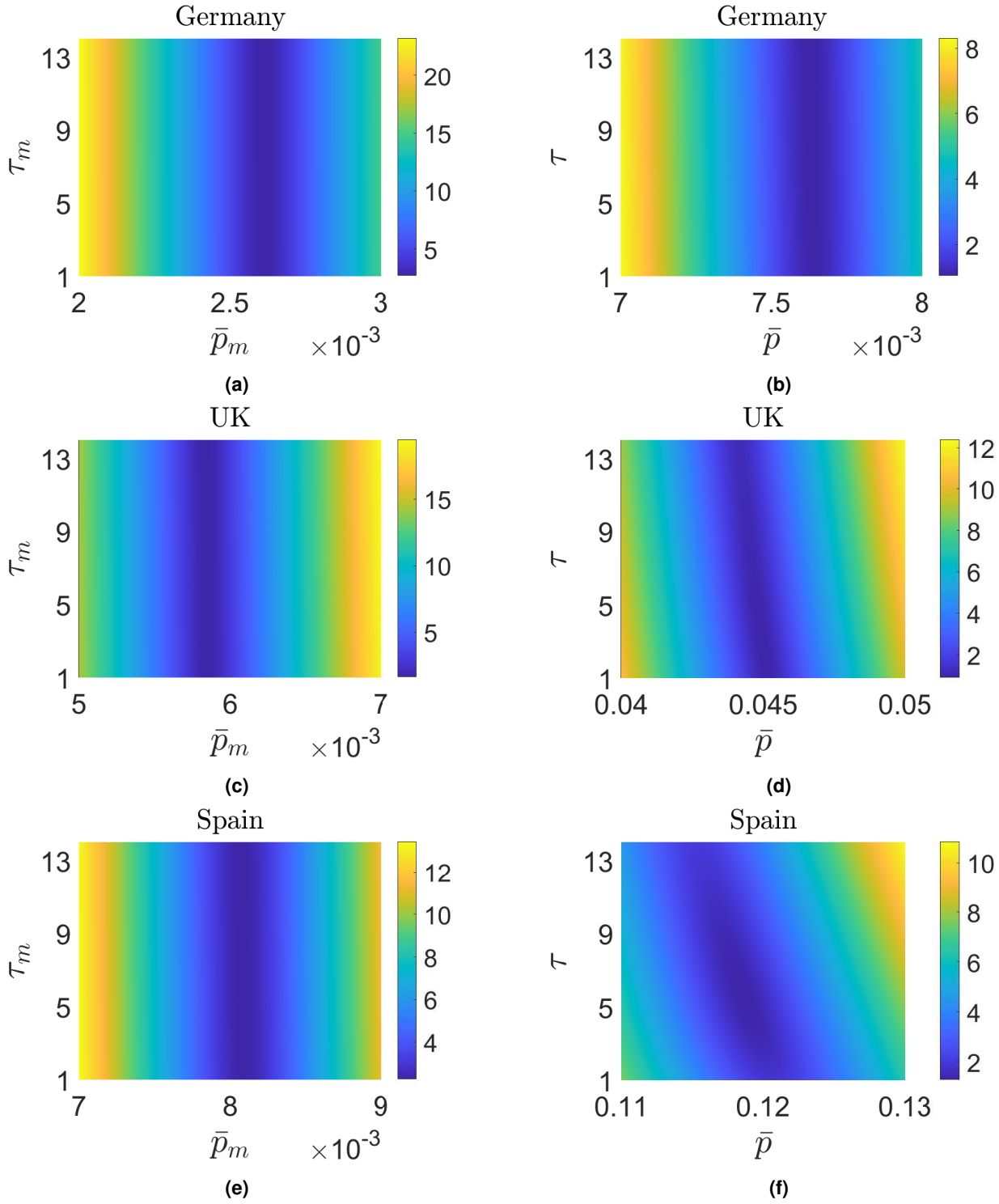

Figure 4. The left side shows the variation of E_{0m} in the $\bar{p}_m - \tau_m$ plane (for low values of $\bar{p}_m$) obtained using the mean-β model. The parameters β_m and V_0 are fixed at the values reported in Table 3. The right side shows the variation of E_0 in the $\bar{p} - \tau$ plane (for low values of $\bar{p}$) obtained from the continuum model. The parameters a, m, and f_0 are fixed at the values reported in Table 4.

Figures

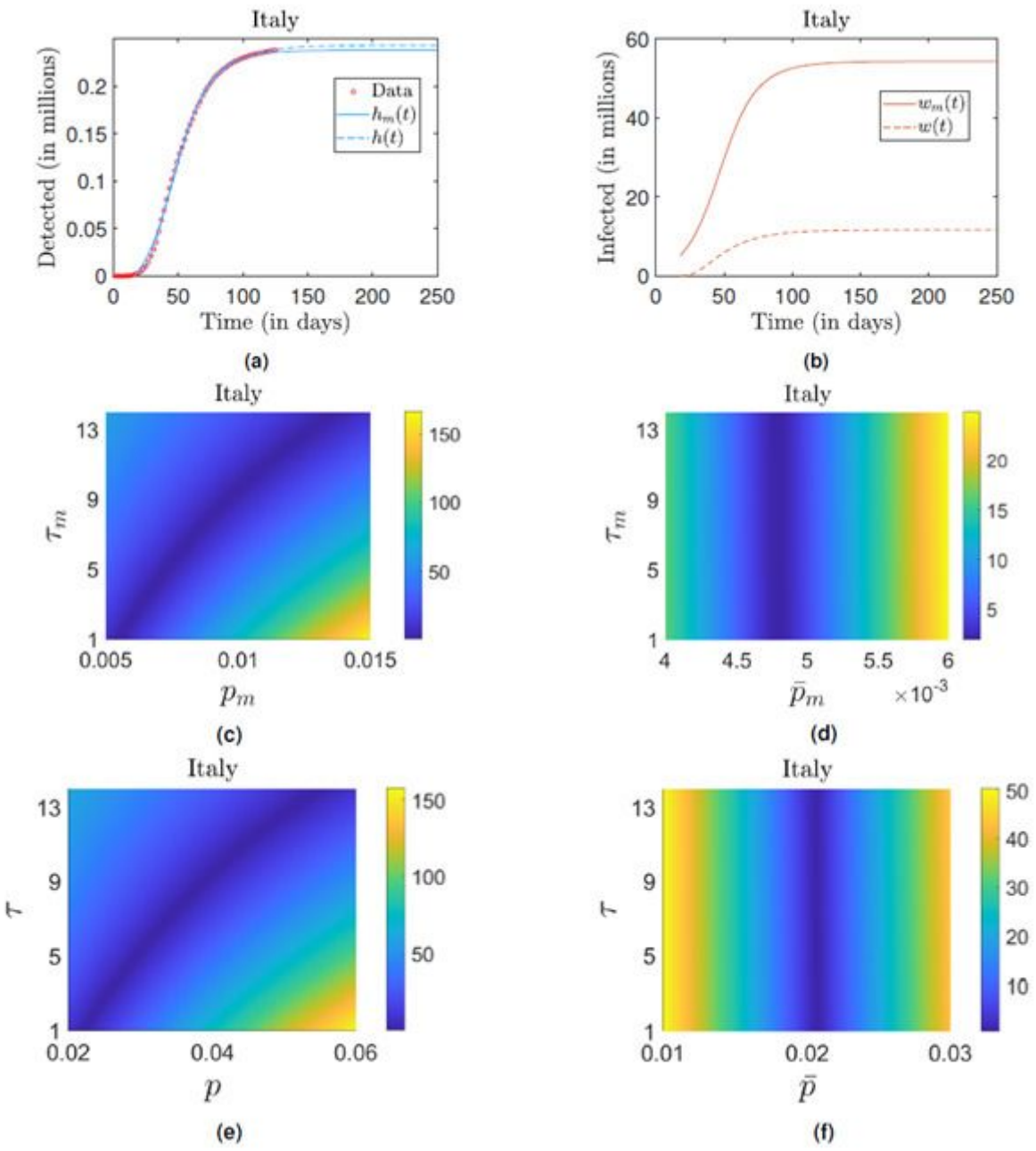

Figure 1

Please see PDF for Figure 1 legend.

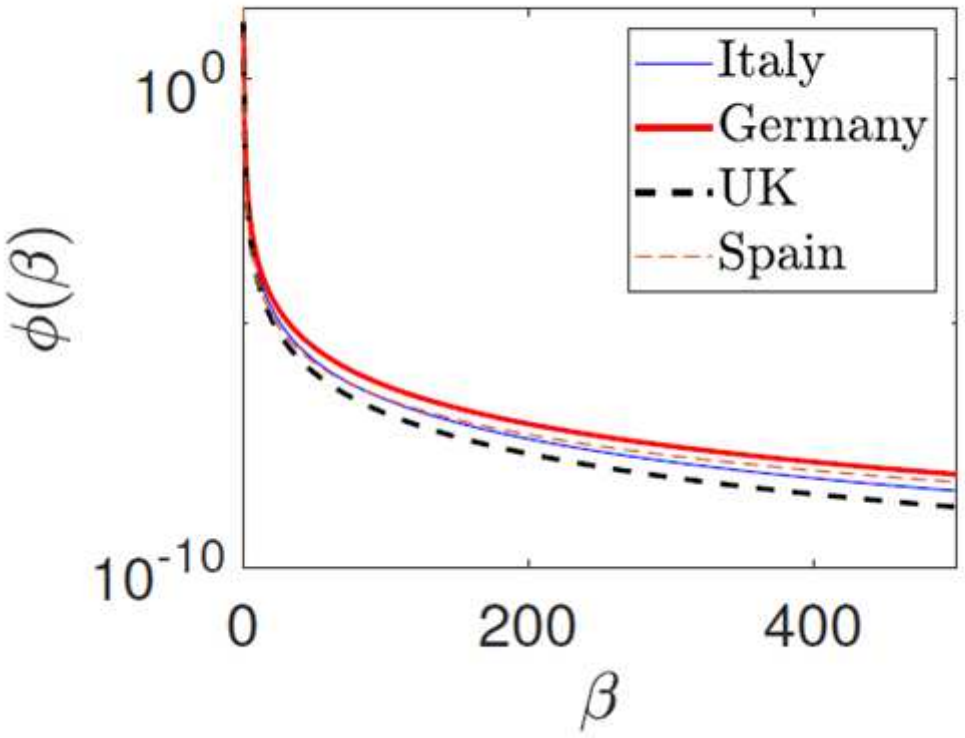

Figure 2

Please see PDF for Figure 2 legend.

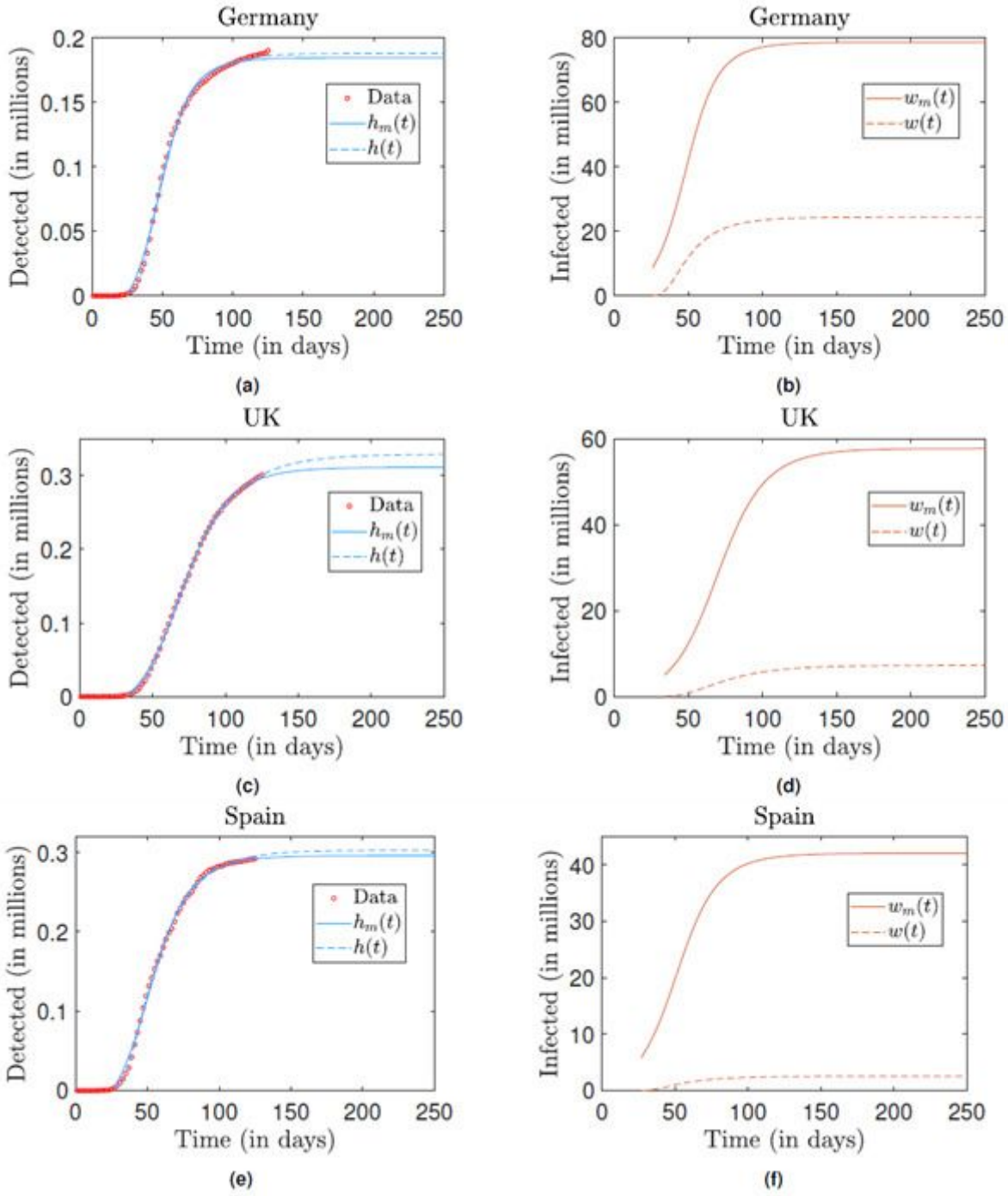

Figure 3

Please see PDF for Figure 3 legend.

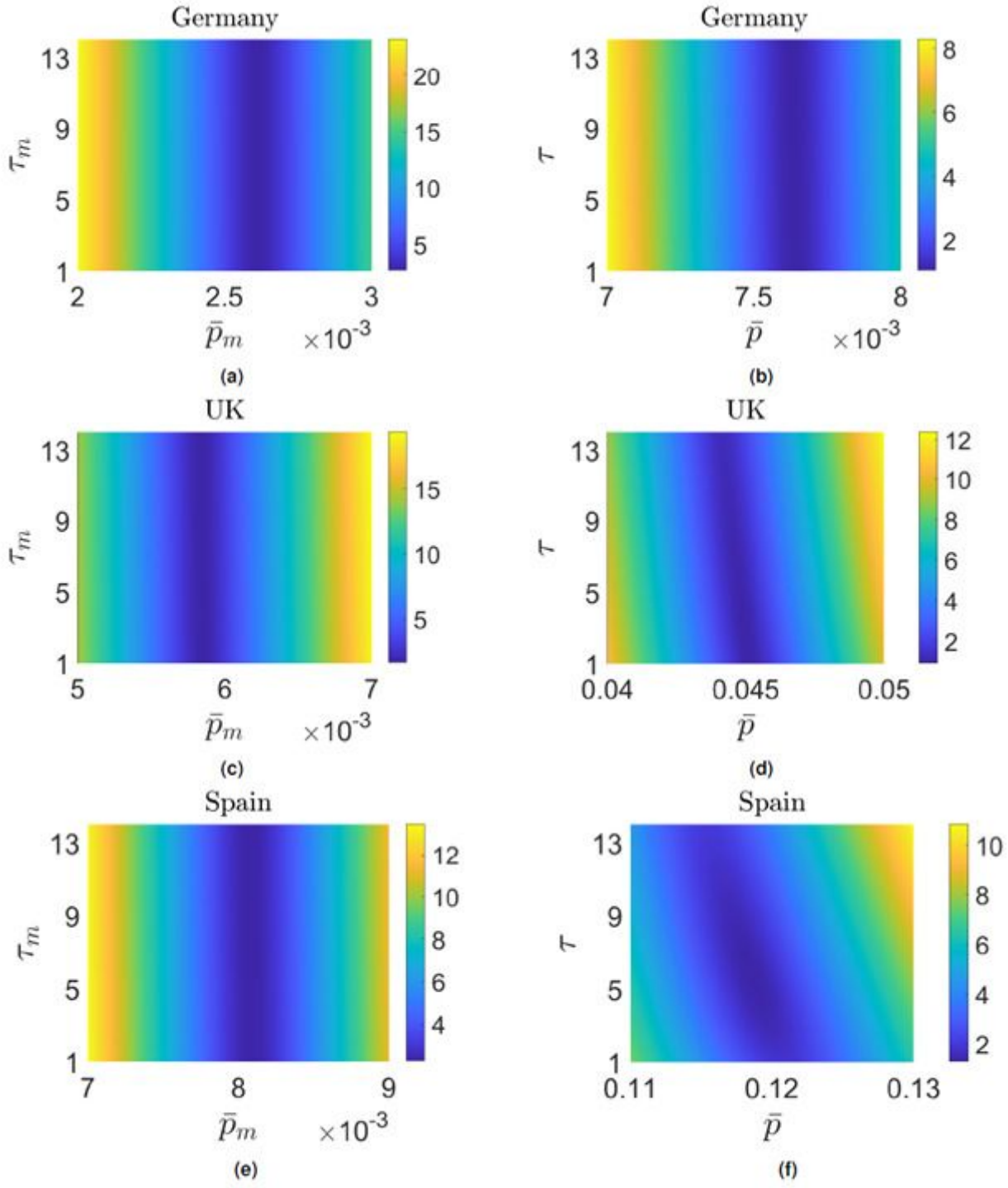

Figure 4

Please see PDF for Figure 4 legend.