## POLITECNICO DI TORINO
## Repository ISTITUZIONALE

Analysing the Covid-19 pandemic in Italy with
the SIPRO model

(Article begins on next page)

29 November 2022

# Analysing the Covid-19 pandemic in Italy with the SIPRO model

*Analizzare la pandemia di Covid-19 in Italia con il modello SIPRO*

Martina Amongero, Enrico Bibbona, Gianluca Mastrantonio

**Abstract** We propose an epidemic compartmental model that extends the classical SIR, in order to allow for an estimate of the unobserved infected people that have not been tested. The model is then fitted to the epidemic curves of the 20 Italian regions using Bayesian methods for mixed-effects models. Despite the interpretability of the model, we still face identifiability issues. We explain how we can alleviate them.

**Abstract** *Abstract in Italian* Proponiamo un modello epidemico compartimentale che estende il modello SIR classico, allo scopo di permettere la stima del numero (latente) di persone infette che non sono state testate. Il modello è poi stimato sulla base delle curve epidemiche delle 20 regioni italiane, usando un metodo Bayesiano per modelli ad effetti misti. Nonostante l'interpretabilità del modello, si incontrano comunque alcuni problemi di identificabilità. Li illustriamo e indichiamo come possono essere mitigati.

**Key words:** SIR, Covid-19, MCMC, Compartmental models, Epidemic curves, mixed models

## 1 Introduction

In Italy, as all over the world, the COVID19-pandemic has had a terrible impact on people's lives, and has caused nearly 100 000 deaths by the end of February 2021. The entire scientific community is at work, and statisticians are contributing in several ways. One direction is that of formulating mathematical models that can help the surveillance of the epidemics and can be used to evaluate the effectiveness of public policies.

Martina Amongero, Enrico Bibbona, Gianluca Mastrantonio
DISMA - Politecnico di Torino
e-mail: martina.amongero@polito.it, enrico.bibbona@polito.it, gianluca.mastrantonio@polito.it

Two main approaches have been proposed: mechanistic models and phenomenological ones. Mechanistic models try to interpret the data by the underlying basic mechanisms. Notable work in this direction are [4, 10, 7, 3]. On the other hand, phenomenological models focus on making the best forecast, no matter if the data generating mechanism is interpretable or realistic anyhow. Examples are [2, 5, 1].

This work is about a deterministic compartmental model (SIPRO) that falls in the first category. The main difficulty in this kind of model is that, the more it is detailed and realistic, the more difficult it becomes to fit it to the available data on epidemic curves. Indeed, most of the information is unobserved, and hardly reconstructable from the available data.

The goal of this paper is to propose an extension of the SIR model that includes a compartment of infected people that is not observed directly. When and if these people are tested they enter the positive compartment and, since quarantined, they stop spreading the disease. The motivation is that, as it was shown in [8], the asymptomatic part of the infected population gives a significant contribution to the pandemic's spread and it is often unreported. This model is applied at the level of the 20 Italian regions, with some parameter that is region-specific and some other parameters that are common to all regions. Some of the regional parameters are moreover assumed to be drawn from a common population (random effects).

Despite the model is much simpler than other proposals, the inference problem is still very challenging since the reconstruction of the latent components suffers from poor identifiability. We illustrate how this can be alleviated by adopting suitable expedients. For our analysis, we use the Italian public data collected by the Italian "Protezione Civile", which can be found at the link [6].

## 2 The SIPRO model $(\rho(t), \mu, \alpha, \nu)$

The population is divided into five compartments: Susceptibles ($S$), Infected ($I$), who are infectious, Positives ($P$) who have been tested and quarantined (no longer source of contagion), Recovered ($R$), formerly infected that recovered or died without having been tested, and, finally, Out ($O$), the recovered or dead people who tested positive. The state of the system is identified by the proportion of the population in each compartment, a vector $(i, p, r, o)$. The proportion of susceptibles is $1 - i - p - r - o$. The evolution of the state variables is described by the functions $(I(t), P(t), R(t), O(t))$ that solves the following system of ordinary differential equations (ODE)

$$\begin{cases} \dot{I}(t) = \beta(t)I(t)[1 - I(t) - P(t) - R(t) - O(t)] - \mu I(t) - \alpha I(t) \\ \dot{P}(t) = \alpha I(t) - \nu P(t) \\ \dot{R}(t) = \mu I(t) \\ \dot{O}(t) = \nu P(t). \end{cases} \tag{1}$$
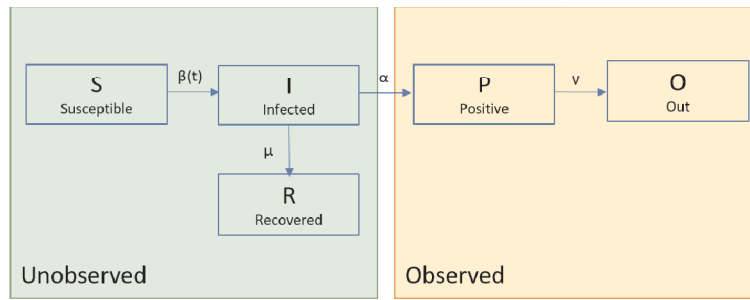
**Fig. 1** SIPRO compartmental model

We can further define $S(t) = 1 - P(t) - R(t) - O(t)$ as the proportion of suscepti-bles. A schematic representation is provided in Figure 1. The rate $\beta(t)$ of infection is allowed to vary over time to model the impact of social distancing measures. The value $1/(\mu + \alpha)$ represents the mean time an individual spends in compartment $I$. The quantities $1/\mu$ and $1/\alpha$ are respectively the time between infection and re-covery and between infection and positivity. Furthermore $1/\nu$ gives the mean time between positivity and recovery. Finally $\rho(t) = \beta(t)/(\mu + \alpha)$ gives the reproduction number at time $t$, and we parameterize the model by $(\rho(t), \mu, \alpha, \nu)$.

## 2.1 The SIPRO model combined with a mixed approach

We build a mixed-model to describe the evolution of the Pandemic in the 20 Italian regions. We assume that within each region, the spread of the pandemic follows the SIPRO model (1), with regional reproduction numbers $\rho_i(t)$, testing policy $\alpha_i$ and recovery rates from positivity $\nu_i$. The recovery rate $\mu$ of the unobserved (asymp-tomatic) infections is common to all regions.

Let $\{(Y_{2,i}(j), Y_{4,i}(j))\}_{\substack{j=1,...,T \\ i=1,...,20}}$ be the daily proportions of Positives and Out indi-viduals derived from the Protezione Civile database [6]. We assume they are noisy observations from the regional SIPRO variables $P_i(j)$ and $O_i(j)$, while the other variables of the model are latent. Since these variables represent proportions, a rea-sonable noise model is given by the Dirichlet distribution. We, therefore, assume

$$\left(Y_{2,i}(j), Y_{4,i}(j), Y_{0,i}(j)\right) \sim \text{Dirichlet}\left(\gamma P_i(j), \gamma O_i(j), \gamma(1 - P_i(j) - O_i(j))\right)$$

where for all $i = 1, ..., 20$ and $j = 1, ..., T$, $Y_{0,i}(j) = 1 - Y_{2,i}(j) - Y_{4,i}(j)$. The pa-rameter $\gamma$ regulates the noise amplitude (in all regions). We estimate directly the regional reproduction number $\rho_i(t)$ from which the infection rate is computed as $\beta_i(t) = \rho_i(t) \cdot (\mu + \alpha_i)$. For simplicity we assume that the function $\rho_i(t)$ is piece-wise linear between a grid of equispaced temporal nodes $\{\tau_k\}_{k=1,\cdots,M}$ with $\tau_1 = 0, \cdots, \tau_M = T$, whose temporal location is specified together with the priors in

the next section. Therefore it is sufficient to estimate the values $\rho_{ik} = \rho_i(\tau_k)$ at the nodes. The logarithms of the regional SIPRO parameters $\rho_{ik}$ and $v_i$ are assumed to be drawn from a Gaussian population (random effects),

$$\log(v_i) \overset{i.i.d.}{\sim} \mathcal{N}(\log(v), \omega_v^2), \qquad \log(\rho_i(k)) \overset{i.i.d.}{\sim} \mathcal{N}(\log(\rho_k), \omega^2)$$

since they are both affected by national and regional public policies, while the $\alpha_i$ are considered as fixed regional effects. The parameter $\mu$ should not be affected by any regulation, and it is therefore kept common to all regions. The initial values $(I(0), P(0), R(0), O(0))$ of the SIPRO equations (1) need also to be estimated.

## 3 Tuning parameters and results

We focus on the data from all Italian regions from the 24th of February 2020 to the 30th of June 2020, corresponding to the so-called first wave. The parameters are estimated using a Bayesian algorithm, but since the model is poorly identifiable, any attempt to estimate the parameter solely based on uninformative priors fails. A lot of information is hidden in the unobserved components, and we need to guide the algorithm by the use of prior information. The parameters that are linked to transition between two unobserved compartments are the most difficult to estimate, namely $\rho_i(t)$, $\alpha_i$, and $\mu$. We emphasize that the value of $\rho_i(t)$ at the first and last node is particularly hard since only half of the information on it is available. This will be reflected in the amplitude of the credible intervals. Finally, the inference is more difficult for regions that had few cases.

To obtain sensible values it is necessary to

- Use informative prior on $\alpha_i$. In particular we impose $\frac{1}{\alpha_i} \in [0.001, 30]$.
- Fix different values of $\mu \in \{\frac{1}{5k} \; s.t. \; k = 1, ..., 9\}$ and then choose the best values using the DIC criterion.
- Use a quite informative prior on the initial state of the system.

The prior on $1/\alpha_i$ is set by constructing an auxiliary variable $\varepsilon_i$ with a Gaussian prior distribution

$$\log(\varepsilon_i) \overset{i.i.d.}{\sim} \mathcal{N}(-2, 0.7^2),$$

and then by remapping it to the interval $[0.001, 30]$ (in days) by the transformation

$$\frac{1}{\alpha_i} = 0.001 \cdot \mathbf{1}_{(-\infty, 0.001)}\left(\frac{1}{\varepsilon_i}\right) + \frac{1}{\varepsilon_i}\mathbf{1}_{[0.001, 30]}\left(\frac{1}{\varepsilon_i}\right) + 30 \cdot \mathbf{1}_{(30, +\infty)}\left(\frac{1}{\varepsilon_i}\right)$$

The prior for the initial conditions is specified by the vector $d = 100 \cdot (0.96, 0.01, 0.01, 0.01, 0.01)$, setting for all $i$

$$(S_i(0), I_i(0), P_i(0), R_i(0), O_i(0)) \overset{i.i.d.}{\sim} \mathscr{D}(d)$$

The other priors are

$$\log(\nu) \sim \mathcal{N}(0,100) \qquad \text{and} \qquad \log(\omega) \sim \mathcal{N}(0,100)$$
$$\log(\rho_k) \sim \mathcal{N}(0,100), \ \forall k \in K \quad \text{and} \qquad \omega_v \sim \mathcal{U}[0,100]$$
$$\log(\gamma) \sim \mathcal{N}(0,100)$$

**Table 1** SIPRO model: Global Parameters Estimated Values

| Parameter | Mean | 95% CI | Parameter | Mean | 95% |
|-----------|------|--------|-----------|------|-----|
| $\rho(t_1)$ | 2.60 | [2.12, 3.17] | $\rho(t_6)$ | 0.59 | [0.13, 1.10] |
| $\rho(t_2)$ | 1.41 | [1.15, 1.71] | $\rho(t_7)$ | 0.21 | [0.02, 0.82] |
| $\rho(t_3)$ | 0.87 | [0.70, 1.05] | $1/\nu$ | 31.17 | [27.06, 35.84] |
| $\rho(t_4)$ | 0.33 | [0.23, 0.45] | $1/\mu$ | 10 | - |
| $\rho(t_5)$ | 0.06 | [0.02, 0.20] | $\log\gamma$ | 10.04 | [10.00, 10.08] |

Table 1 reports the estimates of the global parameters with 95% credible intervals. As an example, we present the estimates for two different regions: Lombardy, the Italian region that counted more cases, and Abruzzo. Figure 2 graphically represents the results obtained. The functions $\rho_i(t)$ clearly reflects the effect of social
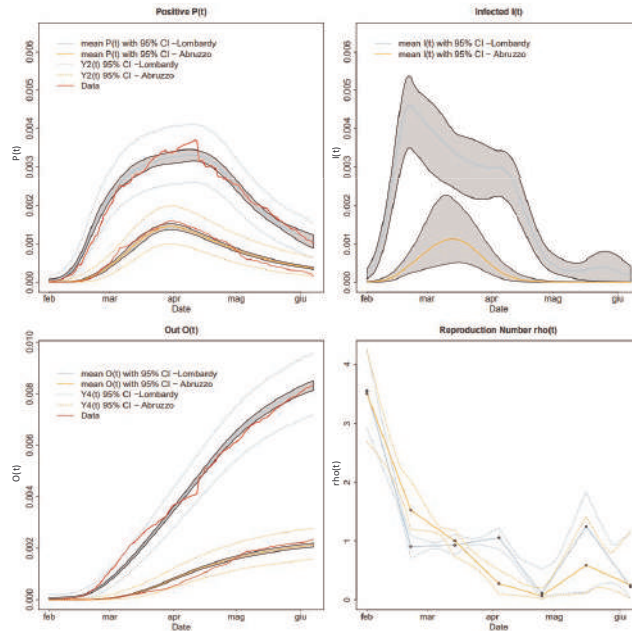


**Fig. 2** $P(t)$, $O(t)$ and $\rho(t)$ mean curves, with 95% CI, for Lombardy and Abruzzo. Red lines represent data. The black dots in the last panes represents the grid of time nodes of the linear interpolation. They are set every three weeks on Mondays

distancing measures. As expected, the function decreased after the lockdown. The estimates of $\frac{1}{\alpha_i}$ for Lombardy and Abruzzo, are $\frac{1}{\alpha_L} \in [20.55, 30]$ with posterior mean 27.67 and $\frac{1}{\alpha_A} \in [7.25, 30]$ with posterior mean 16.19.

## 4 Conclusion

We analyzed and modeled the data collected from February to June (the so-called first phase of the pandemic) using a simple but realistic compartmental model. We highlighted the difficulties in the inference and proposed some solutions. Finally, we presented our estimates that are in agreement with what has been reported in other works. Future work will be needed to compare our model and our results to the ones obtained using other approaches, and to apply the SIPRO-mixed-model to the data collected in September-January, the so-called second phase, to evaluate the effectiveness of the new containment measures that are activated on a regional basis.

## References

1. Agosti, A. and Giudici, P. A Poisson Autoregressive Model to Understand COVID-19 Contagion Dynamics. *Risks*, 8 (3), 2020
2. Farcomeni, A., Maruotti, A., Divino, F., Jona-Lasinio, G., Lovison, G., An ensemble approach to short-term forecast of COVID-19 intensive care occupancy in Italian regions. *Biometrical Journal*, 2020
3. Gatto, M., Bertuzzo, E., Mari, L., Miccoli, S., Carraro, L., Casagrandi, R., and Rinaldo, A. (2020). Spread and dynamics of the covid-19 epidemic in italy: Effects of emergency containment measures. *Proceedings of the National Academy of Sciences* **117,** 10484–10491.
4. Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., and Colaneri, M. (2020). Modelling the covid-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine* pages 1–6.
5. Griette, Q. and Demongeot, J. and Magal, P. A robust phenomenological approach to investigate COVID-19 data for France. *medRxiv*, https://doi.org/10.1101/2021.02.10.21251500
6. https://github.com/pcm-dpc/COVID-19.
7. Jia, W., Han, K., Song, Y., Cao, W., Wang, S., Yang, S., Wang, J., Kou, F., Tai, P., Li, J., et al. (2020). Extended SIR prediction of the epidemics trend of COVID-19 in italy and compared with Hunan, China. *Frontiers in Medicine,* 7
8. Lavezzo, E., Franchin, E., Ciavarella, C., Cuomo-Dannenburg, G., Barzon, L., Del Vecchio, C., Rossi, L., Manganelli, R., Loregian, A., Navarin, N., et al. (2020). Suppression of a SARS-Cov-2 outbreak in the Italian municipality of Vo'. *Nature* 584, 425–429 .
9. Prague, M., Wittkop, L., Clairon, Q., Dutartre, D., Thiébaut, R., and Hejblum, B. P. (2020). Population modeling of early COVID-19 epidemic dynamics in French regions and estimation of the lockdown impact on infection rate. *medRxiv* .
10. Russo, L., Anastassopoulou, C., Tsakris, A., Bifulco, G., Campana, E., Toraldo, G., and Siettos, C. (2020). Tracing day-zero and forecasting the fade out of the COVID-19 outbreak in Lombardy, Italy: a compartmental modelling and numerical optimization approach. *Plos one*, 2020.