"COVID-19 in Italy" Documentation
SIRD Models

April 24, 2020

The SIRD model is a compartmental model used in epidemiology to design the spread of a disease that divides the population into four different groups: susceptible, currently infected, recovered, and deaths. The size of the population is given by the sum of these four variables and it is fixed. The parameters governing the model are the transmission rate, the recovery rate, and the mortality rate.

Another important parameter that wholly describes the spread of an outbreak is the basic reproduction number (R_0), that is computed as the ratio between the transmission rate and the sum of recovery and mortality rate. It is the expected number of individuals that are directly infected by one infected individual, in a population where everyone is susceptible to infection. If it is less than 1, the epidemic will eventually be controlled; if it is larger than 1, the transmission of the disease will increase in the population.

Building on Chen et al. (2020) work, a time-dependent model is proposed in order to let the parameters be free to change over time. By allowing the parameters to vary over time, the effect of containment measures, such as lock-downs, can be somewhat included in the model. Moreover, recovery and mortality rate are likely to depend on the pressure under which hospitals are in, which increases sharply at the beginning of an epidemic (i.e. when a high mortality rate is reported) and then relaxes after the health system capacity is enhanced.

Figure 1 shows the development of these parameters at the national level from the beginning of the outbreak: points indicate real data values while the solid black lines show the model predictions. The dashed line in the R_0 plot represents the threshold where R_0 is equal to 1.

Given the considerable variability of the provincial data, the model is trained on the regional-aggregated data that show smoother trends. Another reason for this choice is that, although the lock-down has been declared almost simultaneously in every Italian province, the virus has spread irregularly in different geographical zones, with the southern areas reporting considerably fewer cases.

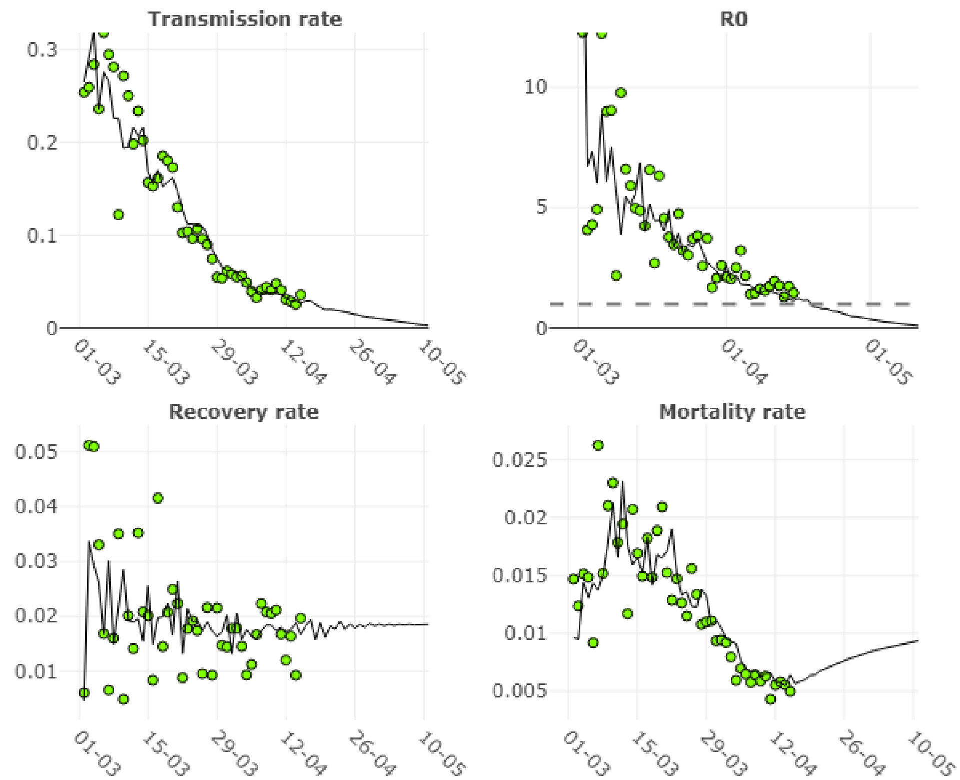

Figure 1: Parameters' Real Values and Model Predictions - Italy

In Figure 2, an example on the Piacenza province is shown. Piacenza is one of the provinces in northern Italy with the highest cumulative rate index (i.e. ratio between reported cases and total population).

The number of lag days represents the number of previous values included in the autoregressive models used to estimate each parameter's evolution. This hyper-parameter is set by default to 7 but is free to be changed within a range: as the number of employed lags grows, the predictions show a more optimistic scenario. Figure 3 shows the model predictions using a number of lags equal to 10, applied to the Piacenza province.

Note that this is a preliminary model whose accuracy tends to decrease sharply as the predictions are done further ahead in the future. Adjustments and improvements are currently being studied and results will be updated regularly.

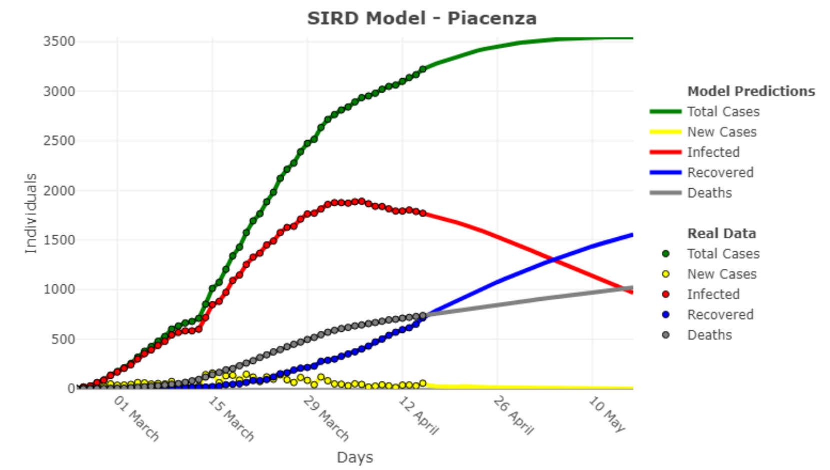

Figure 2: SIRD Model on Piacenza Province (Number of Lags: 7)

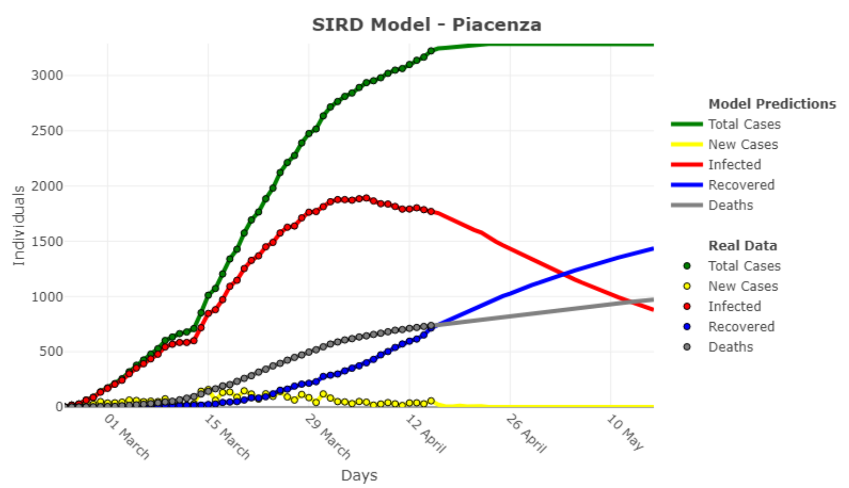

Figure 3: SIRD Model on Piacenza Province (Number of Lags: 10)