Modeling the End Date of COVID-19 Pandemic

COVID-19 Pandemisinin Bitiş Tarihlerinin Modellemesi

Mehmet KOÇAK[a,b]

[a]The University of Tennessee Health Science Center, Department of Preventive Medicine, Division of Biostatistics, Memphis, Tennessee, USA
[b]İstanbul Medipol University, Regenerative and Restorative Research Center, İstanbul, TURKEY

ABSTRACT Objective: As we have been living through COVID-19 pandemic for more than 5 months with its all detrimental impacts on our economic, social, and individual lives, developing models that will accurately inform us about the possible ending date of the pandemic locally or globally becomes ever more critical. In this study, we provide a data-driven model projecting the end-date of a given pandemic, specifically COVID-19. **Material and Methods:** To predict the end date of a given pandemic for early-phase and mature pandemic profiles, we propose a logistic-mixture modelling framework utilizing only the dates and number of infections (i.e., cases), where the level of mixing is determined in a data-driven way with one, two, three or four peaks. We assess the projection accuracy through model convergence and goodness of fit measures for countries that have controlled the pandemic. **Results:** We have shown that our logistic-mixture modelling approach has very favourable convergence and goodness of fit properties, especially when the number of local and global peaks and their timings are provided to the model carefully. Based on the projections of our model, using the available data as of June 01, 2020, the COVID-19 pandemic is ending in early September in Turkey, in early October in the United States of America, and not before December 2020 for the entire world. **Conclusion:** A mixture-logistic modelling framework is a flexible modelling strategy to capture multiple pandemic peaks and, therefore, a reasonable projection approach.

Keywords: COVID-19 pandemics; case projections;
modelling pandemic end-date;
mixture logistic modelling

ÖZET Amaç: COVID-19 pandemisinin yıkıcı ekonomik, sosyal ve bireysel tüm etkilerini son beş aydır yakından yaşarken, pandeminin muhtemel bitiş tarihi hususunda bizleri doğru bir şekilde bilgilendiren modeller geliştirmek, hem yerel hem küresel olarak elzem hale gelmiştir. Bu çalışmada, özelde COVID-19 olmak üzere, herhangi bir pandeminin bitiş tarihini veriye dayalı olarak tahmin eden bir model sunuyoruz. **Gereç ve Yöntemler:** Herhangi bir pandeminin bitiş tarihinin, hem yeni hem de olgunlaşmış pandemi profilleri için, sadece vaka tarihleri ve günlük vaka sayılarını kullanılarak tahminde, karışım seviyesi veriye dayalı olarak bir, iki, üç ve dört zirveli olacak şekilde belirlenen, karma-lojistik modelleme çerçevesi sunuyoruz. Tahminlerin doğruluğunu, model yakınsaması ve pandemiyi kontrol altına aldığı gözlenen ülkelerin verilerine uyum iyiliğiyle değerlendiriyoruz. **Bulgular:** Özellikle yerel ve global zirvelerin modele doğru bir şekilde verildiğinde, karma-lojistik modelleme yaklaşımın, arzu edilen model yakınsaması ve uyum iyiliği özelliklerini taşıdığı gösterilmiştir. 1 Haziran 2020 tarihli verilere dayanan model tahminlerimize göre, COVID-19 pandemisi Türkiye'de Eylül 2020'de, Amerika Birleşik Devletleri'nde Ekim 2020'de sona erecek gözükürken, tüm dünyada Aralık 2020'den önce sona ermeyeceği sonucuna varılmıştır. **Sonuç:** Karma-lojistik modelleme çerçevesi, pandeminin birden çok zirvesini yakalamakta esnek bir yapıya sahip olması hasebiyle, makul bir tahmin yaklaşımıdır.

Anahtar kelimeler: COVID-19 pandemisi; vaka tahmini;
pandemi bitiş tarihi modellemesi;
karma-lojistik modeli

Since its start in Wuhan, China in December 2019, the COVID-19 pandemic impacted our lives in different dimensions. Firstly, the pandemic impacted our psychologies negatively through experiencing panic and anxiety;[1] Wang et al. report from their survey conducted in the earlier phase of the pandemic for the Chinese population that about 54% of the survey participants reported experiencing moderate to severe psychological impact of the outbreak. The negative impacts are especially true for children as children during

the outbreak 'are physically less active, have much longer screen time, irregular sleep patterns, and less favourable diets, resulting in weight gain and a loss of cardiorespiratory fitness.'[2] Another especially vulnerable population experiencing the immediate negative impact of isolation is the elderly population as they are in increased risk of depression and anxiety due to their weak cardiovascular, autoimmune and cognitive state.[3] The pandemic also impacted our social lives and led to increase of unhealthy attitudes towards those who are more out-going as 'unnecessary risk takers and infection spreaders' as well as towards those who are staying indoors and distant.[4] Another issue that becomes immediate with the school closures for pandemic management is the food security for children especially in underdeveloped and developing countries but definitely more prevalent than one thinks in the developed world like the United States and Europe as well; with the closure of the school for a prolonged duration, students living in poverty lose their primary food source through school breakfasts and lunches, which leads to not only malnutrition but decline in academic performance, physical and psychological wellbeing.[5]

Beyond the above dimensions, what is forcing the governments to start a 'new normal' phase is the economic impact of the pandemic and its being not tolerable any longer for all economies small or big. The immediate two fold concerns are that with the long-term social distancing measures, the commerce activities will be much more limited and such restrictions will also come with a heavy load of lost-work time by the workforce, which affects both the businesses and workers and their families.[6] The quantitative measure of the impact of the COVID-19 on the world economy is still unknown due to the uncertainty of when this pandemic will be taken under control. Service-oriented, tourism-based, and foreign trade bound counties will experience the lion share of this negative economic impact where each month of additional pandemic control measures will lead to about 2.5-3% decline in the global GDP.[7]

To help with the data-driven policy making, accurate and timely predictions (i.e., projections) of the ending date of the pandemic becomes more critical than ever. To address this challenge, the versions of SIR and SEIR models have been used.[6,8-10] Another example of pandemic ending projections have been provided by the Singapore University of Technology and Design.[11] In short-term projections to assist pandemic management of the healthcare systems, the utility of time-series models was also evaluated.[12]

In this paper, we propose a logistic-mixture modeling framework to project the ending date of the COVID-19 pandemic for locations that are still in the early phase of a pandemic as well as pandemics with matured profiles with at least one local or global peak.

MATERIAL AND METHODS

Our proposed non-linear modeling approach utilize mixture-logistic probability-density functions in a data-driven manner in that we select different depth of mixture depending on the number of major peaks in the pandemic profiles. We describe the models as follows:

If we have only one peak (i.e., summit), which is the least number of peaks and must be specified even if the pandemic profile is still in an upward direction, and it can be chosen as a future date as the expected potential summit, our model is as in (1) below.

$$Y = \beta_0 . t . \frac{e^{-(t-\mu_1)/s_1}}{s_1\left(1+e^{-(t-\mu_1)/s_1}\right)^2}, s_1 > 0 \tag{1}$$

In Equation-1, $\frac{e^{-(t-\mu_1)/s_1}}{s_1\left(1+e^{-(t-\mu_1)/s_1}\right)^2}$ is the probability density function (pdf) of a logistic random variable with shape and scale parameters μ_1 and s_1, respectively. The role of $\beta_0 . t$ in Equation-1 is to stretch the function to the scale of the data at hand.

If we have two peaks, where one can be a local-summit and the other can be a global summit, our model in (2) below becomes a mixture model of two logistic pdfs.

$$Y = \beta_0.t.\left(m_1 \frac{e^{-(t-\mu_1)/s_1}}{s_1\left(1+e^{-(t-\mu_1)/s_1}\right)^2} + (1-m_1)\frac{e^{-(t-\mu_2)/s_2}}{s_2\left(1+e^{-(t-\mu_2)/s_2}\right)^2}\right), s_1, s_2 > 0 \qquad (2)$$

In Equation-2, m_1 is the mixing parameter and must be between 0 and 1. These mixing parameters, and the ones we will discuss in the upcoming models can be considered as weights of each mixing distribution and their total cannot go beyond 1.0.

Similarly as in Equation-2, if we have three or four peaks, which can be a combination of local and global summits, our models are given in Equation-3 and Equation-4, respective below:

$$Y = \beta_0.t.\left(m_1 \frac{e^{-(t-\mu_1)/s_1}}{s_1\left(1+e^{-(t-\mu_1)/s_1}\right)^2} + m_2\frac{e^{-(t-\mu_2)/s_2}}{s_2\left(1+e^{-(t-\mu_2)/s_2}\right)^2} + (1-m_1-m_2)\frac{e^{-(t-\mu_3)/s_3}}{s_3\left(1+e^{-(t-\mu_3)/s_3}\right)^2}\right), s_1, s_2, s_3 > 0, \qquad (3)$$

In Equation-3, m_1 and m_2 are the mixing parameters and must be between 0 and 1 with a sum not exceeding 1.

$$Y = \beta_0.t.\left(m_1 \frac{e^{-(t-\mu_1)/s_1}}{s_1\left(1+e^{-(t-\mu_1)/s_1}\right)^2} + m_2\frac{e^{-(t-\mu_2)/s_2}}{s_2\left(1+e^{-(t-\mu_2)/s_2}\right)^2} + m_3\frac{e^{-(t-\mu_3)/s_3}}{s_3\left(1+e^{-(t-\mu_3)/s_3}\right)^2} + (1-m_1-m_2-\right.$$
m3 e−(t−μ4)/s4s41+e−(t−μ4)/s42,s1,s2,s3,s4>0
$$\qquad (4)$$

In Equation-4, m_1, m_2 and m_3 are the mixing parameters and must be between 0 and 1 with a sum not exceeding 1.

To implement the above projection models, we have writing a SAS Macro named Pandemic_End_Date whose code is provided in the Appendix along with an example run. This macro uses the NLIN procedure in SAS. The user can specify upto four peaks for the model, indicate whether or not the mean value of the corresponding mixture distribution for each peak is entered manually or estimated by the model, and the upper limit for each mixture proportion. These parameters give the user much needed flexibility for a specific pandemic profile especially when there are some local peaks and the global peak has not been reached yet.

As we are utilizing publicly available COVID-19 summary data for all counties including Turkey, which does not include any human subject data, no Institutional Review Board review is needed for our research. We have conducted this research according to the principles of Helsinki Declaration.

RESULTS

We applied our non-linear logistic-mixture model to the COVID-19 pandemic data as of June 21, 2020, by using the 7-days moving averages of the infection counts.

We first predicted the Pandemic end date for China, South Korea, Austria as examples of countries that contained the pandemic successfully (Figures 1-3).
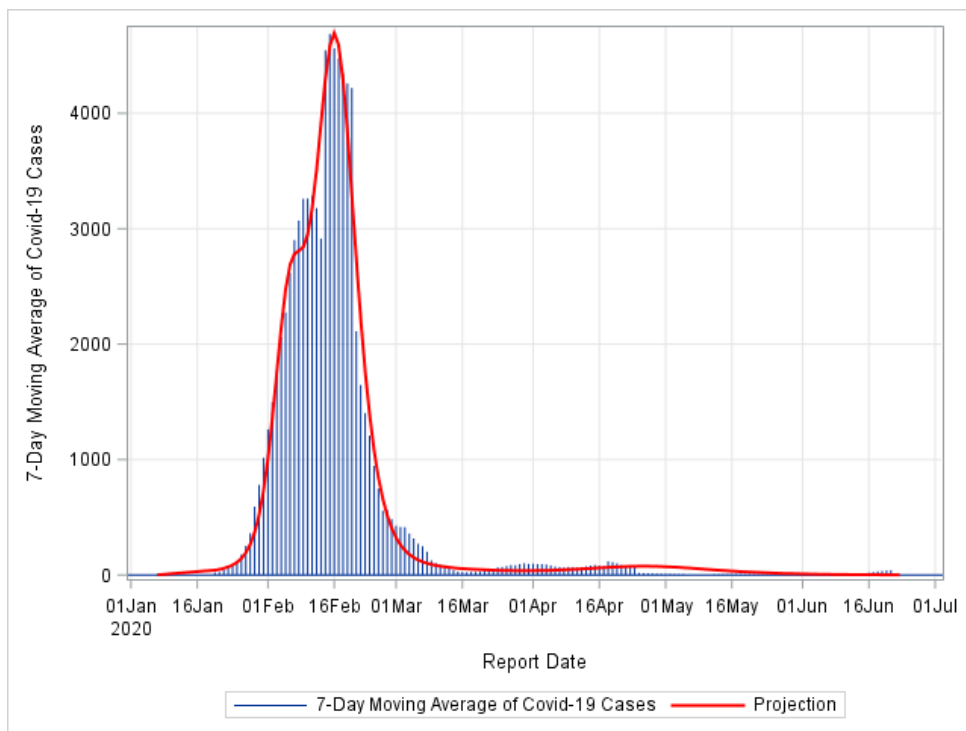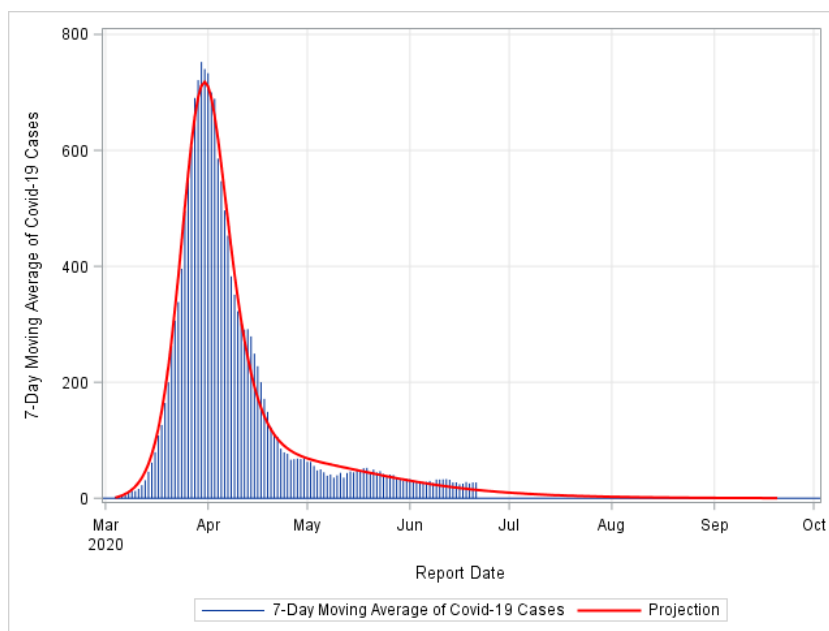
FIGURE 1: Pandemic End Date Projection for China. The model fit was conducted assuming four peaks.

The COVID-19 pandemic is practically contained in China as of early May 2020 and our model indicates that as well.
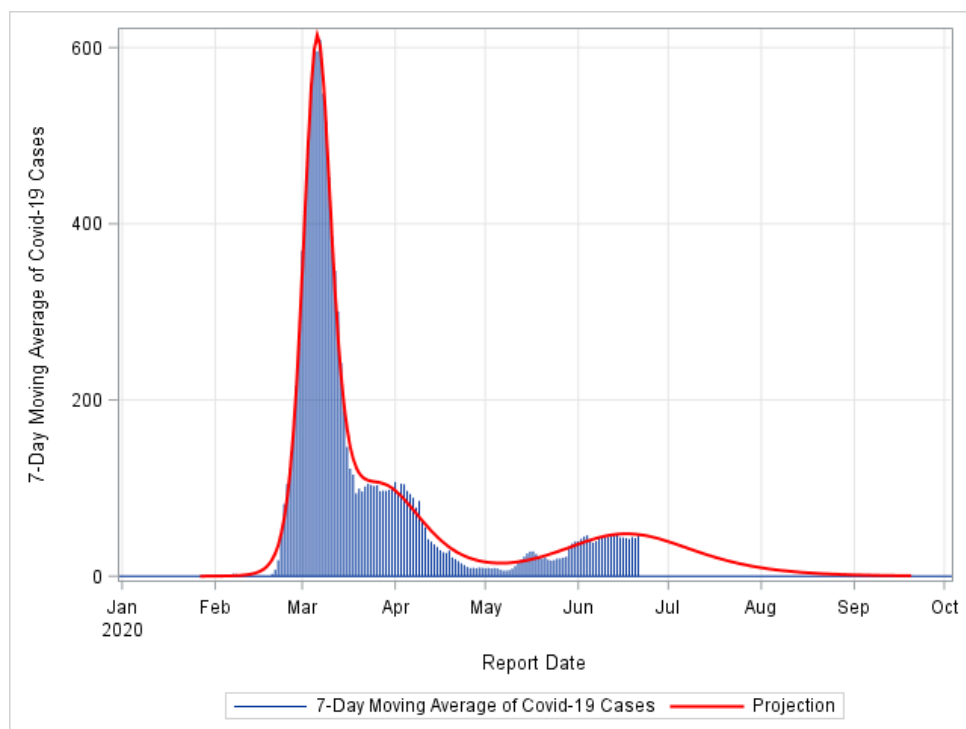
FIGURE 2: Pandemic End Date Projection for South Korea. The model fit was conducted assuming three peaks.

Although South Korea contained the pandemic by early May, recent increases in new cases are closely captured by our model and push the actual ending date towards September 2020.

FIGURE 3: Pandemic End Date Projection for Austria. The model fit was conducted assuming two peaks.

The pandemic ending date for Austria has been projected to be early September 2020.

We have also applied the model to Turkey, Germany and the United States (Figures 4-6). We predict that the number of new COVID-19 cases in Turkey will be 1 or less by October 08, 2020, in Germany by October 16, and in the United States by March 11, 2020.
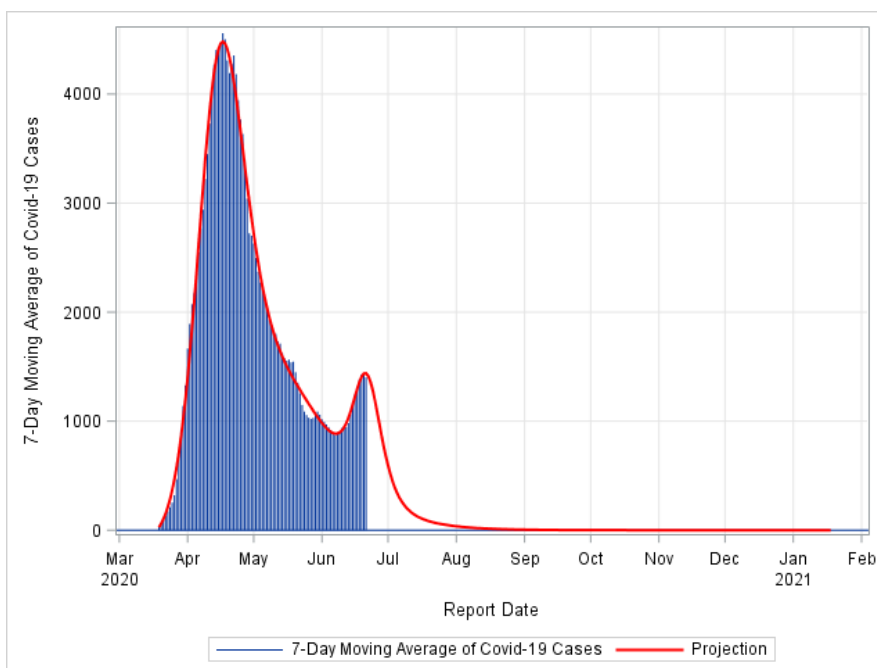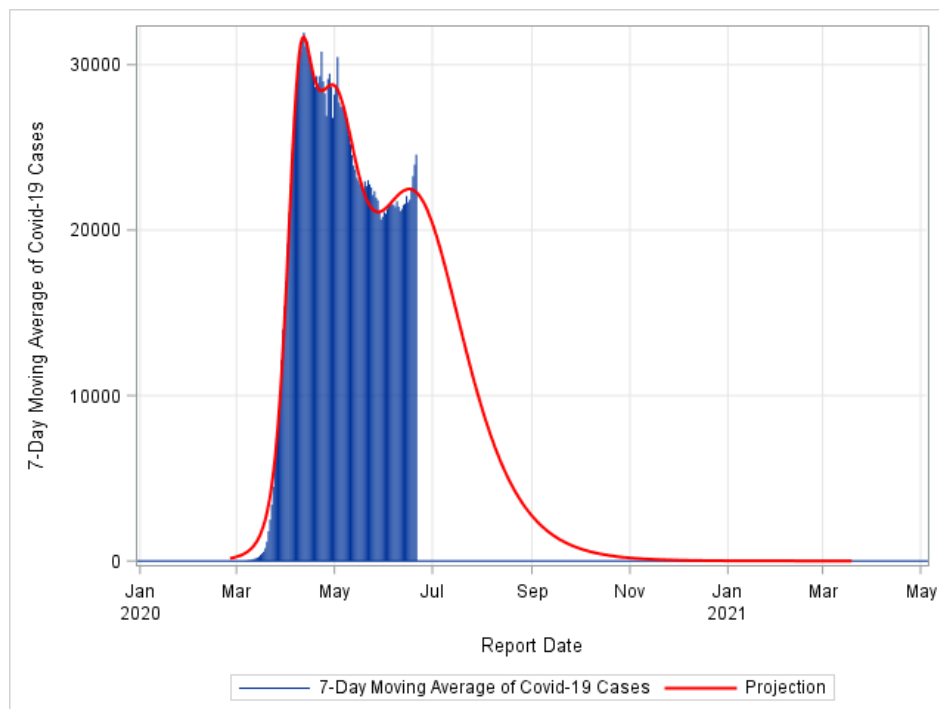

FIGURE 4: Pandemic End Date Projection for Turkey. The model fit was conducted assuming three peaks.
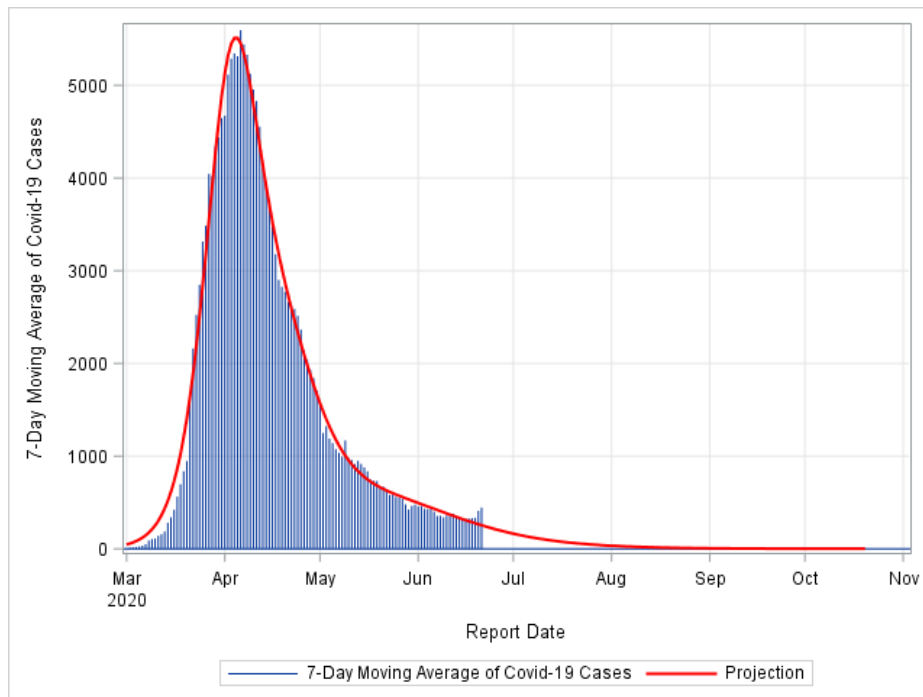
FIGURE 5: Pandemic End Date Projection for Germany. The model fit was conducted assuming three peaks.

FIGURE 6: Pandemic End Date Projection for United States. The model fit was conducted assuming three peaks.

To illustrate how the model behaves where there is no clear peak, we applied it to the pandemic profile of India and Brazil (Figures 7-8). If we assume that India and Brazil will experience their global peaks within a week or so, then their pandemic will not be over in these two countries before the end of the year.
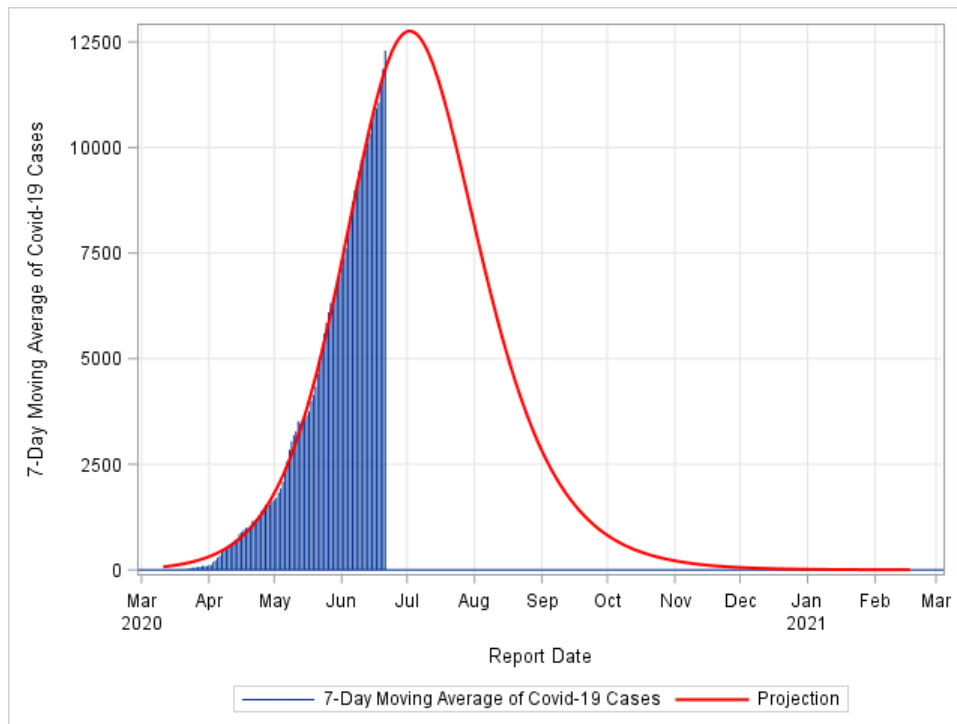
FIGURE 7: Pandemic End Date Projection for India. The model fit was conducted assuming two peaks.
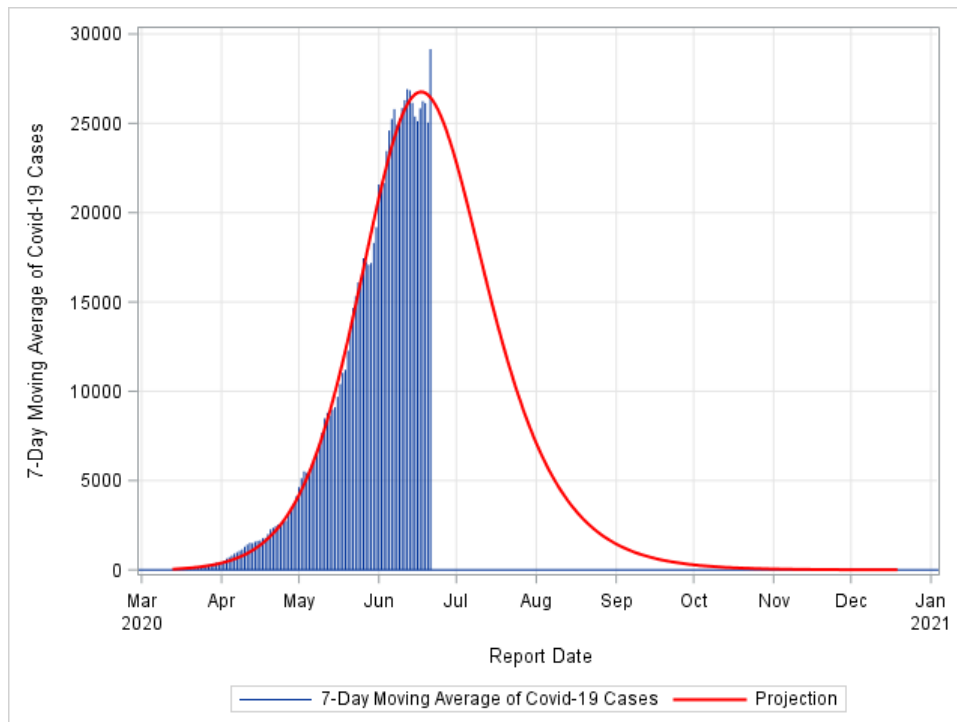

FIGURE 8: Pandemic End Date Projection for Brazil. The model fit was conducted assuming one peak.

Finally, we share the pandemic ending date for the entire World in Figure-9. According to our model, the COVID-19 pandemic will be ending in the entire world not before March 2021.
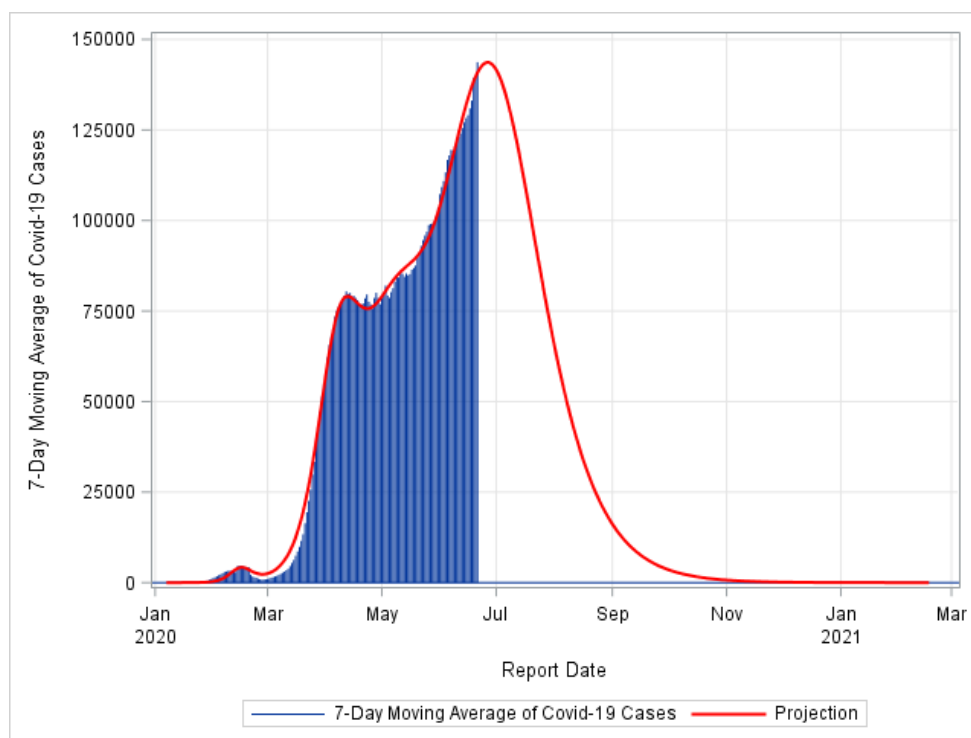
133

FIGURE 9: Pandemic End Date Projection for the entire World. The model fit was conducted assuming four peaks.

DISCUSSION

The COVID-19 pandemic has challenged not only global healthcare systems but also all economies and social fabrics all over the world, both in the developed countries as underdeveloped and developing countries. Among the primary actions to manage the spread of the pandemic were to implement social-distancing and hygiene measures, using face masks, and frequent partial and total lock-downs, which proved successful in many countries in slowing the spread of the virus. Despite their success, these pandemic management strategies forced the world societies face trade-offs: staying healthy and COVID-19 infection-free versus being confined in homes, staying away from pandemic centers versus losing business, employment, etc. Such ever-growing challenges from the beginning of the pandemic that become more and more painful in many dimensions of social and economic life forced everyone and every local and central governments to ponder and strive to get an answer for the question of 'When is this ending?' Time series models may give a short-term answer to such a question, and SIR and SEIR models take the issue from the angle of disease population management with tracking susceptible, exposed, infected, recovered populations and help plan for the management of the healthcare burden of the pandemic.

In this study, we propose a data-driven non-linear modeling framework that is established on mixture logistic distributions. We call this approach a data-driven approach as different depth of the mixture is called in depending on the number of observed or anticipated (future) peaks we have in a given COVID-19 case profile. For locations which does not seem to reach a peak yet, our experience suggests that only one antici-pated summit be specified along with earlier local peaks if needed.

We showed through examples that our model is flexible enough to fit to rather complicated data such as the profile of COVID-19 in China, which experienced spiky peaks close to each other. The difficulty of the profile of South Korea is that it has a very sharp climb to the global peak followed by a very sharp decline before a month-long plateau. Our model seems to have a reasonable fit for such a difficult pattern as

well. One of the underlying assumptions behind the choice of logistic probability density function mixture as the backbone of our approach is that the pandemics experience an exponential growth followed by a gradual decline with a plateau at the outskirts of the declining slope of the peak, which is a phenomenon resulting from easing isolation, quarantine and lock-down efforts and thus increased social interactions among infected and susceptible. The typical examples of this phenomenon are South Korea, Austria, and Australia. Logistic distribution function is a great candidate to capture such patterns.

Our proposed model is not a robust approach as the correct choice of the peaks is highly critical to achieve convergence of model parameter estimation. Therefore, when specifying prior values for a peak, it is recommended that couple of alternative within close range be provided; for example, the first local peak in the global pandemic was observed on Day-50; rather than specifying summit1=50, it is better to specify it as summit1=48 50 52 so that a narrow search grid for parameter estimation is provided, which generally helps convergence significantly. The same argument is valid with the search grid specifications for the standard deviations as well as mixing parameters.

Luo at the Singapore University of Technology and Design[11] projected that the theoretical ending date of the pandemic in Italy would be August 12, 2020 based on the data as of April 21, 2020, and similarly on October 10, 2020 (two-months later than the previous projection) based on the data as of April 28. Our model projected that the pandemic in Italy will end on August 04, 2020 and July 29, 2020 based on the data at two cutpoints as above. Based on the data as of June 21, we predict that the pandemic in Italy will end by the end of August 2020. They predicted the ending date for Italy also for data as of May 8 and suggested the end of October while our model suggests July 26, which practically the same as the model prediction using the data as of April 28. Similarly, they projected the pandemic ending date for Singapore at the same date threshold as above and predicted July 19 and July 11, 2020, respectively. We predicted June 6 and June 4, respectively. They suggested the end of October as the pandemic ending date based on data as of May 8 and our model suggested September 03. With the most current data, we predict that the pandemic ending date for Singapore is expected to be in the second half of August 2020 as Singapore experienced two peaks followed by a week-long plateau afterwards and continued her downward move. These examples show the difficulty of pandemic ending predictions especially under the dynamic management of the social and economic lives, which makes what we desire to predict a moving target.

The user of our model should be ready to exercise patience with increased number of peaks specified with increased number of prior values for the search grid as the computation time increases exponentially. Therefore, a close analysis of the profile under review needs to be conducted to identify influential peaks to be specified so that the estimation process does not get overburdened by unnecessary and misleading prior specifications. It is helpful to fix some of the main parameters in the model rather than leaving it completely free to be estimated. The mixing parameters will always be estimated based on the weight needed at different positions of the profile.

Finally, as a limitation of our modeling approach, regardless of how well we project the pandemic ending date, it is a moving target and it will always depend on what goes on in the field with pandemic management such as quarantine restrictions, social distancing enforcement, international travel restrictions and management, etc.; therefore, projections must be updated frequently to guide people and administrators.

Our proposed model was specifically designed targeting the COVID-19 pandemic; however, our model can be employed for other pandemics such as MERS, SARS, Influenza H1N1 and their potential future occurrences, regardless of their initial incline phases, their post-peak behaviors and their number of major local or global peaks.

CONCLUSION

We conclude that logistic-mixture distribution offers a flexible modelling infrastructure which can detect the pandemic signal of both simple and complicated profiles with careful specification of peak locations in time. Based on our model fits, COVID-19 Pandemic is estimated to end in the world by early 2021, in Turkey by mid-September, in Germany by the end of August, and in the United States by October 2020.

REFERENCES

1. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate psychological responses and associated factors during the initial stage of the 2019 Coronavirus Disease (COVID-19) epidemic among the general population in China. Int J Environ Res Public Health. 2020;17(5):1729. PMID: 32155789 [Crossref] [PubMed] [PMC]
2. Wang G, Zhang Y, Zhao J, Zhang J, Jiang F. Mitigate the effects of home confinement on children during the COVID-19 outbreak. Lancet. 2020;395(10228):945-7. [Crossref] [PubMed]
3. Armitage R, Nellums LB. COVID-19 and the consequences of isolating the elderly. Lancet Public Health. 2020;5(5):e256. [Crossref]
4. Van Bavel JJ, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav. 2020;4(5):460-71. [Crossref] [PubMed]
5. Van Lancker W, Parolin Z. COVID-19, school closures, and child poverty: a social crisis in the making. Lancet Public Health. 2020;5(5):e243-4. [Crossref] [PubMed]
6. Atkeson A. What will be the economic impact of COVID-19 in the us? Rough estimates of disease scenarios. Working Paper 26867. National Bureau of Economic Research; 2020 Mar 19. [Crossref] [PubMed]
7. Fernandes N. Economic effects of coronavirus outbreak (COVID-19) on the world economy. 2020 Mar 22. https://foroparalapazenelmediterraneo.es/wp-content/uploads/2020/03/SSRN-id3557504.pdf.pdf [Link]
8. Simha A, Prasad RV, Narayana S. A simple stochastic SIR model for covid 19 infection dynamics for karnataka: learning from Europe. arXiv preprint arXiv:2003.11920. 2020 Mar 26.
9. Calafiore GC, Novara C, Possieri C. A modified sir model for the covid-19 contagion in Italy. arXiv preprint arXiv:2003.14391. 2020 Mar 31.
10. Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis. 2020;12(3):165-74. [Crossref] [PubMed] [PMC]
11. Luo J. When will COVID-19 end? Data-driven prediction. Singapore University of Technology and Design (http://www. sutd. edu. sg). 2020 Apr. [Link]
12. Koçak M. A comparison of time-series models in predicting COVID-19 cases. Turkiye Klinikleri J Biostat. 2020;12(1):89-96. [Crossref]

APPENDIX

The SAS code of the macro program, Pandemic_End_Date, is provided below. The data must be prepared as the date and infection counts (cases) or its version such as 5-d moving average, 7-d moving average, etc.:

```
*************************************************************************
*** AUTHOR: Mehmet Kocak, Ph.D., Associated Professor of Biostatistics         ***
*** DATE: 06/22/2020                                                            ***
*** MACRO NAME: Pandemic_end_date                                              ***
*** PURPOSE: To project the ending date of a given pandemic                    ***
*** INPUT PARAMETERS:                                                          ***
***           data: Name of the input dataset.                                ***
***           var: This is the variable capturing the daily pandemic cases.   ***
***           datevar: This is the time variable in date format.              ***
***           add_days: No. of Days to be projected into the future.          ***
***           casecut: Minimum number of cases to be considered for the model.***
***           nofpeaks: No. of local or global peaks in the pandemic profile to ***
***                           be modelled.                                     ***
***           summit1: Days at which Peak-1 was observed (must be specified)   ***
***           summit2: Days at which Peak-2, if any, was observed (optional)   ***
***           summit3: Days at which Peak-3, if any, was observed (optional)   ***
***           summit4: Days at which Peak-4, if any, was observed (optional)   ***
***           peak1manual: Logical parameter where YES indicates that we       ***
***                           manually enter the mean value of the mixture     ***
***                           distribution for Peak-1. Otherwise, it estimated.***
***           peak2manual: Logical parameter where YES indicates that we       ***
***                           manually enter the mean value of the mixture     ***
***                           distribution for Peak-1. Otherwise, it estimated.***
***           peak3manual: Logical parameter where YES indicates that we       ***
***                           manually enter the mean value of the mixture     ***
***                           distribution for Peak-1. Otherwise, it estimated.***
***           peak4manual: Logical parameter where YES indicates that we       ***
***                           manually enter the mean value of the mixture     ***
***                           distribution for Peak-1. Otherwise, it estimated.***
***           minmix1: Mixture proportion Lower Bound for the Peak-1 Distr.    ***
***           minmix2: Mixture proportion Lower Bound for the Peak-2 Distr.    ***
***           minmix3: Mixture proportion Lower Bound for the Peak-3 Distr.    ***
***           maxmix1: Mixture proportion Upper Bound for the Peak-1 Distr.    ***
***           maxmix2: Mixture proportion Upper Bound for the Peak-2 Distr.    ***
***           maxmix3: Mixture proportion Upper Bound for the Peak-3 Distr.    ***
***           priormix1: Prior values of Mixture Proportion for the Peak-1 Distr.***
***           priormix2: Prior values of Mixture Proportion for the Peak-2 Distr.***
***           priormix3: Prior values of Mixture Proportion for the Peak-3 Distr.***
***           sdprior1: Search grid for the standard deviations for the Peak-1 Distr.***
***           sdprior2: Search grid for the standard deviations for the Peak-2 Distr.***
***           sdprior3: Search grid for the standard deviations for the Peak-3 Distr.***
***           sdprior4: Search grid for the standard deviations for the Peak-4 Distr.***
***           outpred: Name of the dataset containing the projections.        ***
***********************************************************************************;

%macro Pandemic_end_date(data=allcorona_cases, var=cases, datevar=daterep, add_days=90, casecut=0,
nofpeaks=1, summit1=40, summit2=40, summit3=40, summit4=40, peak1manual=no, peak2manual=no,
peak3manual=no, peak4manual=no,
minmix1=0, minmix2=0, minmix3=0, maxmix1=0.5, maxmix2=0.5, maxmix3=0.5, priormix1=0.5, priormix2=0.5,
priormix3=0.5,
sdprior1=10, sdprior2=10, sdprior3=10, sdprior4=10, outpred=_predicted); title; footnote;
proc sql; create table middata as select distinct * from &data where &var^=. order by &datevar; quit;
proc sql noprint; select distinct max(&var), min(daterep), max(daterep) into :maxval, :mindate, :maxdatadate from
middata; quit;
```

```
data middata; set middata; days=&datevar-&mindate+1; run;
proc sql noprint; select distinct max(days) into :maxday from middata; quit;

proc sql noprint; select distinct min(days) into :minday from middata where &var>=&casecut; quit;
data priors1; mlt=0.25; output; mlt=1; output; mlt=20; output; mlt=50; output; mlt=100; output; run;
data priors1; set priors1; multiples=&maxval*mlt; run;
proc sql; select distinct multiples into :scalevals separated by ' ' from priors1; quit;

data dayadd; do days=&maxday+1 to &maxday+&add_days+1; output; end; run; data middata; set middata dayadd;
run;
proc sgplot data=middata; where &var^=.; scatter y=&var x=days; run;

*** Modeling Starts here ***;
proc nlin data=middata plots=all noitprint maxiter=15000; where &var>&casecut or &var=.;
parms a1=&scalevals;
%if &nofpeaks=1 %then %do;
%if %upcase(&peak1manual)=NO %then %do; parms m1=&summit1; parms s1=&sdprior1;  %end; %else %do;
m1=&summit1; s1=&sdprior1; %end;
model &var=a1*days*(pdf('logistic', days, m1, s1));
%end;

%else %if &nofpeaks=2 %then %do;
parms mcut1=&priormix1; bounds &minmix1<mcut1<&maxmix1;
%if %upcase(&peak1manual)=NO %then %do; parms m1=&summit1; parms s1=&sdprior1; %end; %else %do;
m1=&summit1; s1=&sdprior1; %end;
%if %upcase(&peak2manual)=NO %then %do; parms m2=&summit2; parms s2=&sdprior2; %end; %else %do;
m2=&summit2; s2=&sdprior2; %end;
model &var=a1*days*(mcut1*pdf('logistic', days, m1, s1)+(1-mcut1)*pdf('logistic', days, m2, s2));
%end;

%else %if &nofpeaks=3 %then %do;
parms mcut1=&priormix1 mcut2=&priormix2; bounds &minmix1<mcut1<&maxmix1; bounds
&minmix2<mcut2<&maxmix2;
%if %upcase(&peak1manual)=NO %then %do; parms m1=&summit1; parms s1=&sdprior1; %end; %else %do;
m1=&summit1; s1=&sdprior1; %end;
%if %upcase(&peak2manual)=NO %then %do; parms m2=&summit2; parms s2=&sdprior2; %end; %else %do;
m2=&summit2; s2=&sdprior2; %end;
%if %upcase(&peak3manual)=NO %then %do; parms m3=&summit3; parms s3=&sdprior3; %end; %else %do;
m3=&summit3; s3=&sdprior3; %end;
model &var=a1*days*(mcut1*pdf('logistic', days, m1, s1)+mcut2*pdf('logistic', days, m2, s2)+(1-mcut1-
mcut2)*pdf('logistic', days, m3, s3));
%end;

%else %if &nofpeaks=4 %then %do;

parms mcut1=&priormix1; parms mcut2=&priormix2; parms mcut3=&priormix3;
bounds &minmix1<mcut1<&maxmix1; bounds &minmix2<mcut2<&maxmix2; bounds
&minmix3<mcut3<&maxmix3;
%if %upcase(&peak1manual)=NO %then %do; parms m1=&summit1; parms s1=&sdprior1; %end; %else %do;
m1=&summit1; s1=&sdprior1; %end;
%if %upcase(&peak2manual)=NO %then %do; parms m2=&summit2; parms s2=&sdprior2; %end; %else %do;
m2=&summit2; s2=&sdprior2; %end;
%if %upcase(&peak3manual)=NO %then %do; parms m3=&summit3; parms s3=&sdprior3; %end; %else %do;
m3=&summit3; s3=&sdprior3; %end;
%if %upcase(&peak4manual)=NO %then %do; parms m4=&summit4; parms s4=&sdprior4; %end; %else %do;
m4=&summit4; s4=&sdprior4; %end;
model &var=a1*days*(mcut1*pdf('logistic', days, m1, s1)+mcut2*pdf('logistic', days, m2, s2)+mcut3*pdf('logistic',
days, m3, s3)
+(1-mcut1-mcut2-mcut3)*pdf('logistic', days, m4, s4));
```

138

```
%end;
output out=&outpred p=pred r=resid lcl=lcl ucl=ucl; run;
data &outpred; set &outpred; repdate=&mindate+days-1; format repdate mmddyy10.; run;
proc sgplot data=&outpred; needle y=&var x=repdate; series y=pred x=repdate/lineattrs=(color='red' thickness=2);
label repdate='Report Date' pred='Projection'; xaxis grid; yaxis grid; run;
options notes;
data finaltable; set &outpred; if &var=. and repdate>&maxdatadate; run;
proc sql; create table summarytable as select distinct * from (
select distinct 1 as rowno label=", "below 5,000" as col1 label='If the pandemic management continues as is, the date by
which the daily cases will be:',
min(repdate) as col2 label='Date' format=mmddyy10., round(max(pred),1) as col3 label='Projected Value' from
finaltable where pred<5000 union
select distinct 2 as rowno, "below 1,000", min(repdate) format=mmddyy10., round(max(pred),1) from finaltable where
pred<999.5 union
select distinct 3 as rowno, "below 500", min(repdate) format=mmddyy10., round(max(pred),1) from finaltable where
pred<499.5 union
select distinct 4 as rowno, "below 100", min(repdate) format=mmddyy10., round(max(pred),1) from finaltable where
pred<99.5 union
select distinct 5 as rowno, "below 10", min(repdate) format=mmddyy10., round(max(pred),1) from finaltable where
pred<9.5 union
select distinct 6 as rowno, "below 5", min(repdate) format=mmddyy10., round(max(pred),1) from finaltable where
pred<4.5 union
select distinct 7 as rowno, "below 1", min(repdate) format=mmddyy10., round(max(pred),1) from finaltable where
pred<0.5)
order by rowno; quit;
proc sql; select distinct col1, col2, col3 from summarytable order by rowno; quit;
proc sql; drop table middata, finaltable, summarytable; quit;
%mend;
*** Sample Run ***;
proc sql; create table covid19data_tr as select distinct daterep label="Report Date", days,
sum(case_7dma) as case_7dma label="7-Day Moving Average of Covid-19 Cases",
sum(case_5dma) as case_5dma label="5-Day Moving Average of Covid-19 Cases",
sum(cases) as cases label="Covid-19 Cases" from corona.covid19_data_21jun2020
where fullname='Turkey' group by daterep; quit;
proc sgplot data=covid19data_tr; needle x=days y=cases; run;
%Pandemic_end_date(data=covid19data_tr, datevar=daterep, var=case_7dma, casecut=0, nofpeaks=3,
summit1=31, sdprior1=10, peak1manual=no, minmix1=0.65, maxmix1=0.8, priormix1=0.85,
summit2=60, sdprior2=10, peak2manual=no, minmix2=0.01, maxmix2=0.2, priormix2=0.02,
summit3=97, sdprior3=10, peak3manual=no, add_days=210, outpred=_pandemic_tr);
```