



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Estimating spatiotemporal patterns of deaths by COVID-19 outbreak on a global scale

**Citation for published version:**

Valente, F & Laurini, MP 2023, 'Estimating spatiotemporal patterns of deaths by COVID-19 outbreak on a global scale', *BMJ Open*, vol. 11, no. 8, e047002, pp. 1-11. <https://doi.org/10.1136/bmjopen-2020-047002>

**Digital Object Identifier (DOI):**

[10.1136/bmjopen-2020-047002](https://doi.org/10.1136/bmjopen-2020-047002)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

BMJ Open

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# BMJ Open Estimating spatiotemporal patterns of deaths by COVID-19 outbreak on a global scale

Fernanda Valente, Marcio Poletti Laurini 

**To cite:** Valente F, Laurini MP. Estimating spatiotemporal patterns of deaths by COVID-19 outbreak on a global scale. *BMJ Open* 2021;**11**:e047002. doi:10.1136/bmjopen-2020-047002

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online. (<http://dx.doi.org/10.1136/bmjopen-2020-047002>).

Received 18 November 2020  
Accepted 27 July 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

Department of Economics, FEARP, University of Sao Paulo, Ribeirao Preto, Sao Paulo, Brazil

**Correspondence to**  
Marcio Poletti Laurini;  
[laurini@fearp.usp.br](mailto:laurini@fearp.usp.br)

## ABSTRACT

**Objective** Our main objective is to estimate the trend of deaths by COVID-19 on a global scale, considering the six continents.

**Study design** The study design was a retrospective observational study conducted using the secondary data provided by the Our World in Data project on a public domain.

**Setting** This study was conducted based on worldwide deaths by COVID-19 recorded for the Our World in Data project from 29 February 2020 to 17 February 2021.

**Methods** Estimating the trend in COVID-19 deaths is not a trivial task due to the problems associated with the COVID-19 data, such as the spatial and temporal heterogeneity, observed seasonality and the delay between the onset of symptoms and diagnosis, indicating a relevant measurement error problem and changing the series' dependency structure. To bypass the aforementioned problems, we propose a method to estimate the components of trend, seasonality and cycle in COVID-19 data, controlling for the presence of measurement error and considering the spatial heterogeneity. We used the proposed model to estimate the trend component of deaths by COVID-19 on a global scale.

**Results** The model was able to capture the patterns in the occurrence of deaths related to COVID-19, overcoming the problems observed in COVID-19 data. We found compelling evidence that spatiotemporal models are more accurate than univariate models to estimate the patterns of the occurrence of deaths. Based on the measures of dispersion of the models' prediction in relation to observed deaths, it is possible to note that the models with spatial component are significantly superior to the univariate model.

**Conclusion** The findings suggested that the spatial dynamics have an important role in the COVID-19 epidemic process since the results provided evidence that spatiotemporal models are more accurate to estimate the general patterns of the occurrence of deaths related to COVID-19.

## INTRODUCTION

On 11 March 2020, the WHO declared the COVID-19 outbreak as a pandemic, which is currently a public health threat in more than 200 countries. Since the efficacy and safety of pharmaceutical treatments to deal with the COVID-19 require time and effort,<sup>1 2</sup> some

## Strengths and limitations of this study

- Use of novel statistical decomposition model for spatiotemporal analysis to estimate the trend of deaths by COVID-19.
- The proposed model is able to avoid the heterogeneity and error measurement problems associated with the COVID-19 data.
- The model's ability to capture the observed general patterns in the occurrence of deaths related to COVID-19 in a global scale.
- The model's applicability to different fields, such as epidemiology issues as well as climate events.
- Study limitations include the reliability of the data used, which relies on the government actions that differs from country to country.

public health measures have been proposed by countries to cope with the rapid spread of the COVID-19, such as social distancing, self-quarantine and lockdown. However, preventing transmission depends on how well we can assess the number of infected people. Statistical modelling is helpful to obtain the trend component and better understand the evolution of the confirmed cases and deaths of COVID-19 and provide key information about the effects of non-pharmaceutical measures to control the transmission. Regarding the methods for trend estimation, there are several different methods reported in the literature, which differ in their complexity and interpretability. The most prevalent trend estimation methods are model-based trend extraction, non-parametric filtering, singular spectrum analysis and wavelets.<sup>3</sup>

To estimate COVID-19 trends, some of the aforementioned approaches have been applied to the current outbreak.<sup>4-7</sup> However, it should be noted that despite the great variety of papers dedicated to studying the long-movements of COVID-19, there is little understanding about the spread pattern of COVID-19 on a global scale. Also, it is worth noting some problems related to COVID-19 data. First, it is observed that there is a delay

between the onset of symptoms and accurate diagnosis, which may vary from country to country, depending on the local government's strategy.<sup>8</sup> Also, there are unconformities in the report of new deaths which are counted on the day that are included in the system, instead of the actual day of the death.<sup>9</sup> These problems create a relevant measurement error problem, changing the series' dependency structure. In addition, the delays are also related to the lack of personnel available to report cases on weekends, which creates a seasonality structure in the cases and death reports, generating an additional aggregation problem on the time series.<sup>10</sup>

To bypass these problems, we propose to estimate the long-term movements of deaths by COVID-19 on a global scale through a structural decomposition.<sup>11</sup> The main idea is to decompose the temporal variability observed in the data into the trend, seasonal and cycle components, which allows identifying permanent movements, and cyclical and seasonal effects, in the presence of measurement errors. In addition, since we are interested to assess the spread pattern of COVID-19 on a global scale, the importance of a spatial component is undeniable, since the use of this information is important in estimating the trend and cycle components, due to the nature of transmission in an epidemic process, with spatial spread dynamics. In the light of this, we include a component that considers the spatial heterogeneity among different areal unities in the study of COVID-19 evolution. We propose two different spatial formulations which rely on the definition of neighbourhoods. The first is based on regions that present common borders, while the second includes information on the air transport network. Regarding this context, the contribution of this paper is to explore a Bayesian version of the trend-cycle decomposition in combination with count distributions, considering the spatial heterogeneity,<sup>12</sup> in the task of estimating the permanent movements of deaths by COVID-19 outbreak in a global scale.

## MATERIAL AND METHODS

In this paper, we present the results of estimating the proposed model for COVID-19-related death data for countries collected by the Our World in Data project (available at <https://ourworldindata.org/covid-deaths>), using daily information on new deaths for the period from 29 February 2020 to 17 February 2021. Therefore, the study design was a retrospective observational study conducted using secondary data on a public domain. Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research. We use the countries on that basis that contained information on the total population, corresponding to a total of 205 countries. Also, we use the continental division of countries as a region definition, and thus  $k=6$ , corresponding to the continents of Asia, Europe, South America, North America, Africa and Oceania.

## Patient and public involvement statement

No patient involved.

## Statistical model

The proposed model is a generalisation of the Bayesian model introduced by Valente and Laurini<sup>12</sup> for estimating the COVID-19 trend for data in Brazil, controlling for problems of measurement error in the construction of death statistics. The model is based on a generalised decomposition of trend, seasonality and cycle components incorporating a time-varying spatial component. The spatial component is based on the conditional autoregressive (CAR) structure,<sup>13</sup> based on the information of occurrences in some definition of the neighbourhood to formulate prior information for the number of occurrences in the region of interest. The structure of the model can be summarised by the following equations:

$$\begin{aligned}
 Y_{(i,t)} &= \text{Poisson}(\exp \lambda_{(i,t)} E), \\
 \lambda_{(i,t)} &= \mu_{(k,t)} + s_{(k,t)} + c_{(k,t)} + \xi_{(i,t)} \\
 \Delta^2 \mu_{(k,t)} &= \eta_{\{\mu_{(k,t)}\}} \\
 s_{(k,t)} &= s_{(k,t-1)} + s_{(k,t-2)} + \dots + s_{(k,t-m)} + \eta_{\{s_{(k,t)}\}} \\
 c_{(k,t)} &= \theta_1 c_{(k,t-1)} + \theta_2 c_{(k,t-2)} + \eta_{\{c_{(k,t)}\}} \\
 \xi_{\{(i,t)\}} \mid \xi_{\{(j,t), i \neq j, \tau\}} &\sim N\left(\frac{1}{n_i} \sum_{\{i \sim j\}} \xi_{\{(j,t)\}}, \frac{1}{\tau n_i}\right) \\
 \xi_{\{(j,t)\}} &= \Phi \xi_{\{(j,t-1)\}}
 \end{aligned}$$

where  $Y_{(i,t)}$  is the total number of occurrences (deaths) in region  $i$  and time  $t$ .  $E$  is the exposure offset, controlling for the total population regarding the cases in  $Y_{(i,t)}$ . This offset allows using common components for regions with different population sizes. It is important to highlight that the used database contains daily information on deaths by COVID-19 of 205 countries, thus the region  $i$  corresponds to the country level. However, to estimate the latent factors, we define the region  $k$ , which contains a certain group of individuals  $i$ . In particular, define  $k$  corresponding to each of the six continents, namely, Asia, Europe, Oceania, Africa, North America and South America, where each continent  $k$  contains a certain group of countries  $i$ . In this sense, the component  $\mu_{(k,t)}$  captures the general trend of the process for a region  $k$  for some definition of region, containing a group of individuals  $i$  and representing the average level of occurrences in period  $t$  for the region  $k$ . In our model,  $k$  represents each continent under analysis. We assume a second-order random walk (RW2) structure for the trend component. The RW2 is a flexible structure suitable for epidemics with rapid spread, which is able to model the evolution of the average number of cases in fast-growing processes since it assumes that both the average level and the rate of growth of the level are non-stationary processes.<sup>12</sup> Also, it is worth noting that this model has a non-parametric nature, since it can be formulated as a spline model, and thus suitable for a new epidemic whose parametric case structure is unknown.

The model also contains region-specific seasonal ( $s_{(k,t)}$ ) and cycle ( $c_{(k,t)}$ ) components. These two components

are used to capture periodic components in the series of deaths related to COVID-19, where we assume that these components are transient, that is, have a sum equal to zero. As discussed by Valente and Laurini,<sup>12</sup> these components are a way of controlling the effects of measurement errors introduced by the mechanism for disclosing the number of cases and deaths related to COVID-19, where the total number of cases reported on a certain day corresponds to the total recorded in the day, instead of the exact date of the diagnosis confirmation or death. Using simulations of an Susceptible-Exposed-Infectious-Removed (SEIR) model, the authors<sup>12</sup> show that this accounting mechanism introduces patterns like seasonality and cycle effects in the series of cases and deaths. The  $s_{(k,t)}$  component represents a sum of stochastic components, with the number of components given by the periodicity of the series and with a sum-to-zero restriction. The  $c_{(k,t)}$  component is represented by a second-order autoregressive process with possibly complex roots, which permits to capture cyclical components, also assuming a restriction of zero-sum. It is worth noting that the cyclic component also allows us to capture other patterns of periodic dependency in the series, such as transmission waves without permanent effects in the series.

The  $\eta_{\{\mu_{(k,t)}\}}$ ,  $\eta_{\{s_{(k,t)}\}}$  and  $\eta_{\{c_{(k,t)}\}}$  components are non-spatial independent innovations with zero mean Gaussian distributions with estimated precision (inverse of variances).  $\xi_{(i,t)}$  is the spatial random effect component, allowing to incorporate spatial variability in the occurrence rate,  $n_i$  is the number of neighbours of region  $i$  and  $i \sim j$  indicates that the two regions  $i$  and  $j$  are neighbours. The CAR structure used defines a prior for the spatial effect in the region  $i$  given by a Gaussian distribution with mean given by the average of the values of the spatial effects for regions  $j$  that are neighbourhood to region  $i$ , and variance controlled by a precision parameter  $\tau$  multiplied by the number of neighbours. As the spatial variability in an epidemic process varies over time, we assume that this component is also time varying. The dynamic formulation for the spatial CAR is parameterised through an autoregressive structure, via a parameter  $\Phi$  controlling for the time dependence for this process.

The spatial formulation adopted in this paper depends on the definition of the neighbourhood structure. We use two definitions of the neighbourhood. The first is the spatial contiguity, that is, two regions are considered neighbours if there is a common border. The second formulation considers both the spatial contiguity and also an air transport network, which allows incorporating into the model the transmission mechanism given by the contact among passengers and air transport crew between two countries. We detail the structure used in the analysis of air transport network in online supplemental material. In online supplemental table S1, the country codes used in determining the air transport network are presented, and in online supplemental table S2 and figure S1, the communities chosen as neighbourhood determinants by the passenger transport criterion are presented.

The prior structure assumes log-gamma distribution for all precision components, with values (1,  $5e-05$ ) for the trend and seasonal components, and the spatial random effect. For the cycle (AR(2) structure) component, we use a penalised complexity prior<sup>14</sup> for the precision, with values (3, 0.01), and for the first-order and second-order partial autocorrelation parameters, we use penalised complexity priors with values (0.5, 0.5) and (0.5, 0.4). The prior for the coefficient  $\Phi$  is also a penalised complexity prior for the correlation parameter, with values defined as (3, 0.01).

The resulting additive hierarchical structure allows us to perform a Bayesian inference procedure within the Integrated Nested Laplace Approximations (INLA) approach,<sup>15</sup> which provides accurate and efficient approximations on additive hierarchical models that can be represented as Gaussian Markov Random Fields.<sup>16</sup> For reasons of space, we do not detail the INLA method here, which is widely discussed by Rue *et al*<sup>15</sup> and by Valente and Laurini.<sup>12</sup>

## RESULTS

To estimate the trend of deaths by COVID-19 outbreak on a global scale, we performed inference procedures based on the model described in equation (1), considering three different formulations. First, we estimated the parameters without the effects of the spatial component (model M1, hereafter), where the estimated parameters are the precision of the trend ( $1/\eta_{\{\mu_{(k,t)}\}}$ ), seasonal ( $1/\eta_{\{s_{(k,t)}\}}$ ) and cycle ( $1/\eta_{\{c_{(k,t)}\}}$ ) parameters for Asia, Africa, Europe, North America, South America and Oceania, and the parameters of the second-order autoregressive process of the cycle (Partial Autocorrelation Function (PACF)1 and PACF2) for each continent. In the second and third formulations (M2 and M3, respectively), we consider a time-varying spatial component, modelled as a spatial random effect with a CAR structure with time dependency, where the estimated values at any given region are conditional on neighbouring values. The difference between models M2 and M3 is how the neighbourhood structure was defined. While in model M2, the neighbourhood structure was constructed based on contiguity, using the territorial division, in model M3 not only contiguity was considered to define, but also information of an air transport network. In the case of models M2 and M3, the estimated parameters also include the log precision parameter  $\tau$ , and parameter  $\Phi$  of time dependency.

It is worth noting that the precision coefficients are not directly interpretable due to the different scales of the latent components, but a higher precision indicates less variability of the estimated latent component. A higher precision parameter indicates a smoother component or with less temporal variability. In this case, the most direct interpretation is for the estimated path of the latent component. The PACF1 and PACF2 components are the representations of the first-order and second-order autoregressive coefficients in terms of partial

**Table 1** Estimated parameters of deaths reported in worldwide—models M1, M2 and M3

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Model without spatial component (M1)						
Precision for trend (Asia)	47 065.94	1.71e+04	22 255.2	44 171.67	8.86e+04	38 965.84
Precision for trend (Europe)	18 244.38	5.04e+03	10 229.53	17 617.95	2.99e+04	16 431.35
Precision for trend (Africa)	25 340.83	9.38e+03	11 735.89	23 766.84	4.80e+04	20 918.04
Precision for trend (North America)	8287.844	2.39e+03	4505.762	7995.11	1.38e+04	7439.043
Precision for trend (South America)	31 740.07	1.24e+04	13 512.98	29 788.63	6.14e+04	26 043.59
Precision for trend (Oceania)	3631.989	1.67e+03	1285.369	3340.437	7.72e+03	2758.72
Precision for seasonality (Asia)	37 041.63	2.47e+04	8771.08	31 026.05	1.01e+05	21 071.69
Precision for seasonality (Europe)	1480.084	5.50e+02	695.524	1383.102	2.83e+03	1211.254
Precision for seasonality (Africa)	14 441.12	1.08e+04	3037.376	11 564.34	4.29e+04	7337.345
Precision for seasonality (North America)	1904.256	6.96e+02	889.365	1789.016	3.58e+03	1579.576
Precision for seasonality (South America)	2935.316	1.37e+03	1126.925	2651.206	6.40e+03	2174.777
Precision for seasonality (Oceania)	16 849.83	1.94e+04	1802.661	11 049.72	6.73e+04	4780.553
Precision for cycle (Asia)	61.37	5.57e+00	51.092	61.142	7.30e+01	60.716
PACF1 for cycle (Asia)	0.044	7.00e−02	−0.089	0.043	1.84e−01	0.038
PACF2 for cycle (Asia)	−0.002	6.10e−02	−0.121	−0.003	1.20e−01	−0.006
Precision for cycle (Europe)	74.313	8.67e+00	58.553	73.883	9.26e+01	73.107
PACF1 for cycle (Europe)	0.116	7.60e−02	−0.033	0.116	2.64e−01	0.116
PACF2 for cycle (Europe)	−0.011	6.80e−02	−0.147	−0.01	1.20e−01	−0.006
Precision for cycle (Africa)	32.364	3.32e+00	26.235	32.231	3.93e+01	32.01
PACF1 for cycle (Africa)	0.016	7.10e−02	−0.123	0.015	1.56e−01	0.014
PACF2 for cycle (Africa)	0.049	7.20e−02	−0.092	0.048	1.92e−01	0.045
Precision for cycle (North America)	43.924	4.60e+00	35.433	43.744	5.35e+01	43.454
PACF1 for cycle (North America)	0.204	6.80e−02	0.07	0.203	3.37e−01	0.201
PACF2 for cycle (North America)	−0.032	6.70e−02	−0.163	−0.033	1.02e−01	−0.035
Precision for cycle (South America)	20.575	1.89e+00	17.05	20.511	2.45e+01	20.412
PACF1 for cycle (South America)	−0.177	7.20e−02	−0.315	−0.178	−3.30e−02	−0.18
PACF2 for cycle (South America)	0.127	6.50e−02	−0.001	0.127	2.54e−01	0.126
Precision for cycle (Oceania)	6.376	1.75e+00	3.535	6.182	1.04e+01	5.811
PACF1 for cycle (Oceania)	0.103	1.32e−01	−0.161	0.105	3.55e−01	0.11
PACF2 for cycle (Oceania)	−0.363	1.84e−01	−0.68	−0.378	3.50e−02	−0.413
Spatial model with Continent-specific trend, seasonal and cycle components (M2)						
Precision for trend (Asia)	1915.711	188.693	1601.403	1894.92	2337.895	1839.906
Precision for trend (Europe)	2656.69	333.928	2066.338	2633.714	3375.658	2586.415
Precision for trend (Africa)	6258.313	1048.8	4712.686	6076.868	8763.297	5635.089
Precision for trend (North America)	2466.706	332.468	1857.129	2453.747	3160.465	2436.044
Precision for trend (South America)	4711.425	698.155	3365.212	4716.521	6076.105	4777.84
Precision for trend (Oceania)	2331.846	409.061	1748.527	2255.167	3321.72	2066.046
Precision for seasonality (Asia)	1903.494	257.954	1460.44	1880.728	2470.376	1831.028
Precision for seasonality (Europe)	1215.008	184.489	924.436	1189.277	1641.535	1128.188
Precision for seasonality (Africa)	1898.143	265.922	1441.209	1874.799	2482.773	1824.655
Precision for seasonality (North America)	2130.874	308.402	1566.844	2118.774	2773.613	2103.005
Precision for seasonality (South America)	2463.529	466.222	1803.158	2374.758	3596.307	2162.728
Precision for seasonality (Oceania)	2637.863	942.551	1603.34	2379.593	5126.215	1864.649
Precision for cycle (Asia)	96.186	14.71	71.998	94.499	129.396	90.733

Continued

Table 1 Continued

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
PACF1 for cycle (Asia)	0.177	0.074	0.02	0.182	0.309	0.199
PACF2 for cycle (Asia)	0.027	0.07	-0.114	0.029	0.159	0.036
Precision for cycle (Europe)	78.772	11.986	57.889	77.871	104.812	76.1
PACF1 for cycle (Europe)	0.052	0.094	-0.136	0.054	0.23	0.061
PACF2 for cycle (Europe)	-0.23	0.102	-0.432	-0.227	-0.034	-0.215
Precision for cycle (Africa)	65.884	9.914	48.994	64.971	87.786	63.037
PACF1 for cycle (Africa)	0.024	0.082	-0.142	0.025	0.18	0.032
PACF2 for cycle (Africa)	-0.03	0.074	-0.178	-0.029	0.112	-0.025
Precision for cycle (North America)	57.175	9.509	41.701	56.03	78.822	53.53
PACF1 for cycle (North America)	0.085	0.082	-0.092	0.092	0.227	0.119
PACF2 for cycle (North America)	0.145	0.088	-0.006	0.137	0.335	0.103
Precision for cycle (South America)	47.607	9.048	32.823	46.561	68.159	44.419
PACF1 for cycle (South America)	-0.091	0.076	-0.248	-0.087	0.05	-0.073
PACF2 for cycle (South America)	0.344	0.069	0.2	0.347	0.469	0.357
Precision for cycle (Oceania)	45.551	6.951	32.586	45.403	59.724	45.408
PACF1 for cycle (Oceania)	0.221	0.086	0.074	0.213	0.405	0.178
PACF2 for cycle (Oceania)	0.005	0.08	-0.133	-0.003	0.177	-0.032
log $\tau$	0.067	0.003	0.062	0.067	0.072	0.066
Group $\Phi$	0.96	0.002	0.957	0.96	0.963	0.96
<b>Spatial model with Continent-specific trend, seasonal and cycle components—air transport network (M3)</b>						
Precision for trend (Asia)	4225.738	229.884	3747.123	4240.222	4640.095	4301.895
Precision for trend (Europe)	2464.551	112.490	2249.928	2462.482	2691.357	2459.068
Precision for trend (Africa)	5411.900	318.028	4733.933	5439.884	5957.384	5559.366
Precision for trend (North America)	2287.834	110.097	2067.379	2290.439	2498.605	2303.331
Precision for trend (South America)	4521.493	309.205	3851.897	4554.247	5031.378	4697.282
Precision for trend (Oceania)	2484.957	119.962	2271.991	2476.214	2741.473	2450.545
Precision for seasonality (Asia)	1963.974	93.720	1769.153	1969.625	2133.869	1993.574
Precision for seasonality (Europe)	1356.319	63.771	1235.083	1355.026	1485.311	1352.741
Precision for seasonality (Africa)	2401.577	124.529	2191.461	2388.591	2676.463	2348.164
Precision for seasonality (North America)	2462.174	135.863	2173.940	2473.277	2698.285	2520.300
Precision for seasonality (South America)	2238.563	110.550	2043.071	2230.278	2475.860	2205.895
Precision for seasonality (Oceania)	1781.246	118.396	1526.034	1793.136	1978.952	1844.691
Precision for cycle (Asia)	136.511	5.603	126.185	136.254	148.163	135.540
PACF1 for cycle (Asia)	0.209	0.022	0.167	0.209	0.252	0.208
PACF2 for cycle (Asia)	0.076	0.024	0.031	0.075	0.125	0.072
Precision for cycle (Europe)	124.948	5.534	114.676	124.719	136.403	124.114
PACF1 for cycle (Europe)	0.013	0.023	-0.034	0.014	0.056	0.017
PACF2 for cycle (Europe)	-0.064	0.024	-0.112	-0.064	-0.020	-0.060
Precision for cycle (Africa)	104.274	4.748	94.815	104.360	113.409	104.815
PACF1 for cycle (Africa)	0.091	0.024	0.048	0.090	0.140	0.084
PACF2 for cycle (Africa)	0.031	0.025	-0.014	0.030	0.083	0.025
Precision for cycle (North America)	95.808	4.498	86.843	95.899	104.456	96.369
PACF1 for cycle (North America)	0.304	0.026	0.248	0.307	0.347	0.317
PACF2 for cycle (North America)	0.023	0.024	-0.029	0.024	0.066	0.030
Precision for cycle (South America)	73.152	3.728	65.438	73.359	79.950	74.245

Continued

Table 1 Continued

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
PACF1 for cycle (South America)	0.129	0.023	0.085	0.129	0.176	0.127
PACF2 for cycle (South America)	0.188	0.025	0.144	0.187	0.241	0.180
Precision for cycle (Oceania)	48.144	2.449	43.900	47.928	53.471	47.276
PACF1 for cycle (Oceania)	0.192	0.023	0.148	0.191	0.238	0.189
PACF2 for cycle (Oceania)	-0.066	0.023	-0.113	-0.066	-0.022	-0.064
$\log \tau$	0.014	0.000	0.014	0.014	0.015	0.014
Group $\Phi$	0.961	0.001	0.959	0.961	0.963	0.961

PACF, Partial Autocorrelation Function.

correlation coefficients, which have some advantages in terms of computational representation, especially for the imposition of the stationarity of the component that is important in the definition of the cyclic component. The cycle components can be interpreted as the correlation between component in the period  $t$  and  $t_1$  (PACF1) and  $t_2$  (PACF2), controlling for the other correlation placed in this specification. We emphasise that the cycle component serves to capture the structures of transient dependence in the series of deaths, which are generated by the component of measurement error or by the composition

of other non-permanent patterns in the series of observed deaths.

The results of the estimation of models M1, M2 and M3 are shown in table 1. To provide a better interpretation of the results, we plotted the estimated trend (see figure 1), seasonality (see figure 2) and cycle (see figure 3) components for all six considered continents, individually, considering the model M2. For reason of space only the estimated components obtained with model M2 are shown in this paper, and the other results are available on request from the authors.

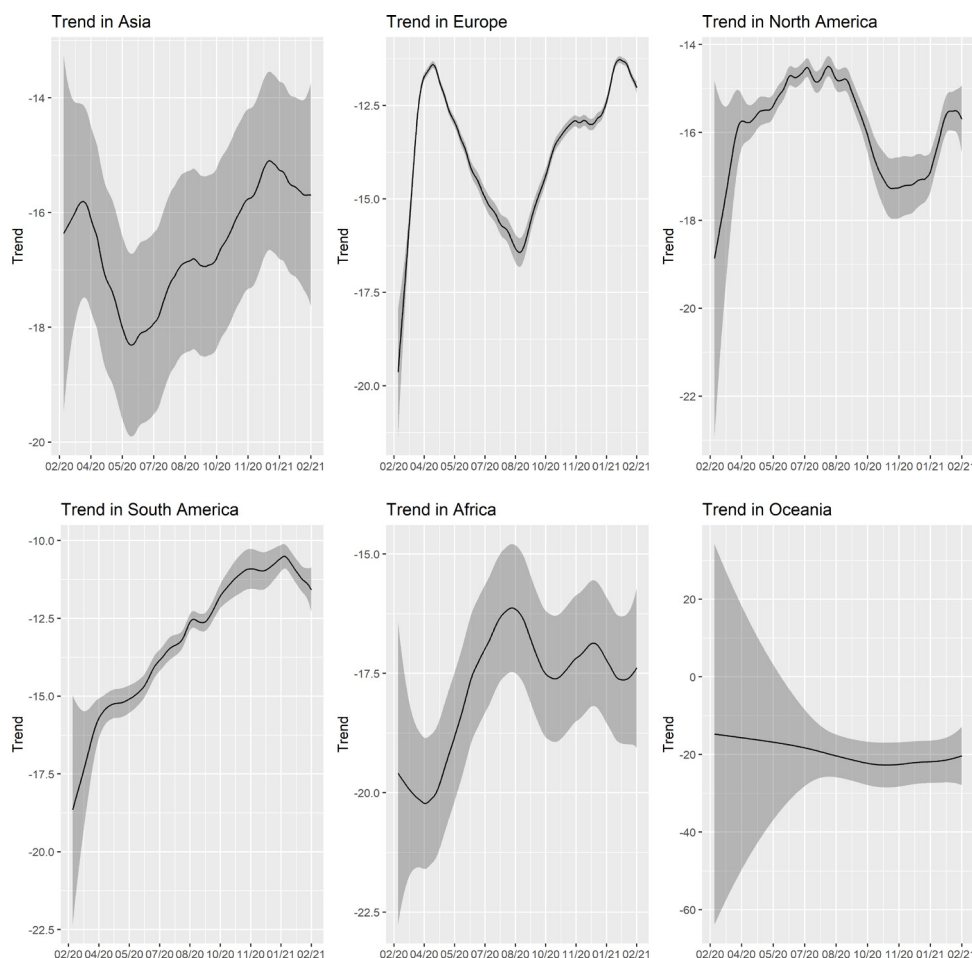
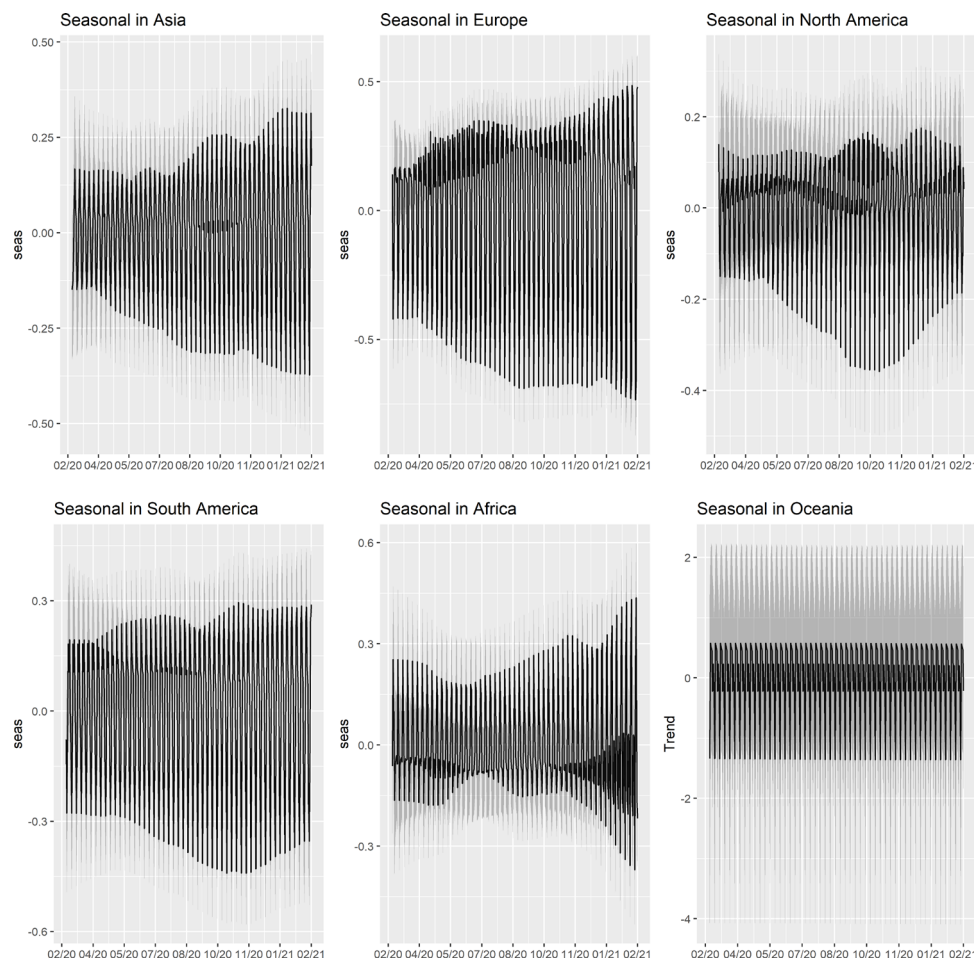


Figure 1 Estimated trends—spatiotemporal model with region specific trends, seasonal and cycle components (model M2).



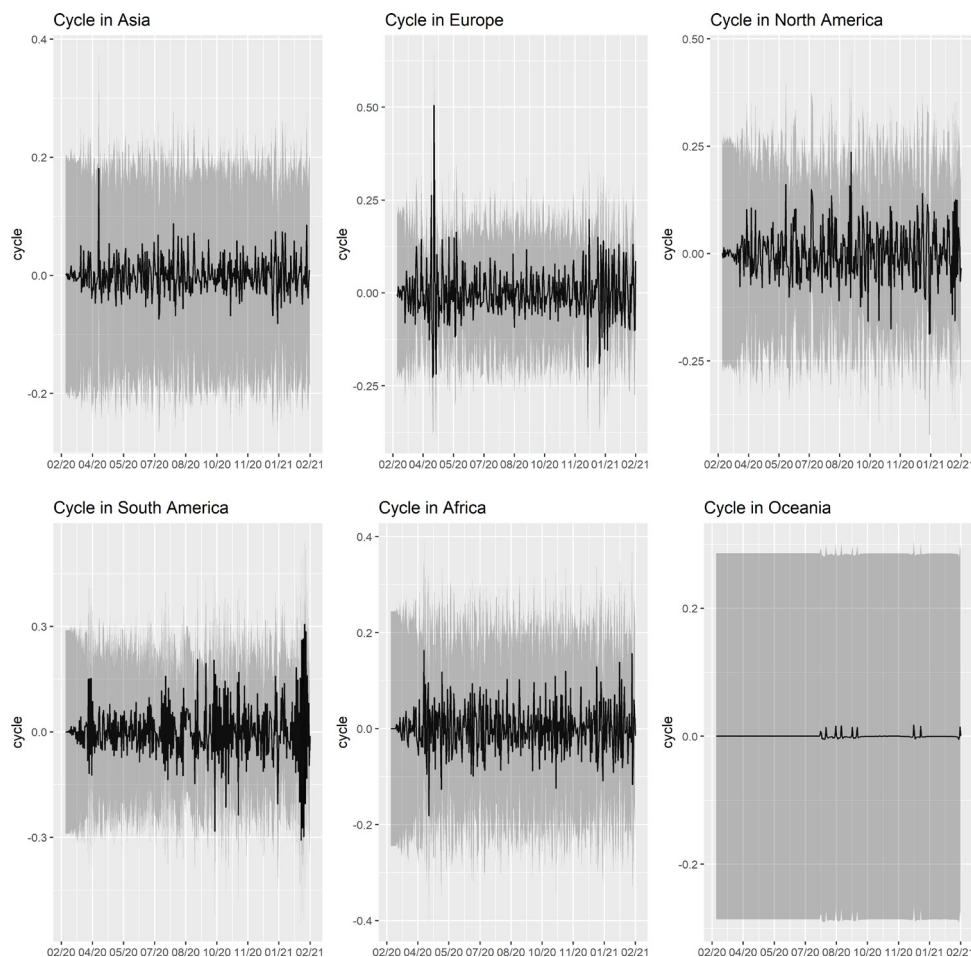
**Figure 2** Estimated seasonality—spatiotemporal model with region specific trend, seasonal and cycle components (model M2).

To provide goodness-of-fit measures to compare how well the models fit the data, we calculate the fit statistics of all the models, such as the mean error (ME), root mean squared error (RMSE) and mean absolute error (MAE), reported in table 2. These measures serve as general parameters of model adequacy. The ME is important to verify the presence of relevant biases in the model's predictions. In the presence of high MEs, the model may be underestimating or overestimating the number of deaths observed. The RMSEs and MAEs are measures of dispersion of the model's predictions in relation to the observed cases, and serve as a measure of the model's accuracy, using different weighting metrics (squared error for the RMSE and absolute error in the MAE). Quadratic loss is a traditional measure of predictive accuracy, linked, for example, to estimation by ordinary least squares, penalising squared error and avoiding the cancellation of errors with distinct signs. The MAE has a similar interpretation, since absolute errors also avoid cancellation of positive and negative errors but has the advantage of being less sensitive to some deviant observation. In general, the results indicate an adequate adjustment for the models estimated in the article, such as almost zero MEs indicating the absence of predictive

bias, and very low RMSE and MAE measures indicating great precision in fitting the observed deaths.

In summary, we present the results of analysing data for COVID-19 in 205 countries from 29 February 2020 to 17 February 2021, divided according to the following six continents: Asia, Europe, North America, South America, Africa and Oceania. It is worth noting that we did not consider the Antarctica continent since the continent was only affected by the coronavirus pandemic in December 2020. The results show that our model was able to capture the long-term movements in cases that resulted in death cases for the six continents, showing an accelerated growth pattern until mid-April in Asia, and a second wave driven by the spread of the disease in the Middle East, South Asia and Southeast Asia. The estimated trend component for Europe shows that the continent has experienced the peak of the first wave at the end of April, and a new accelerated growth phase at the end of the analysed sample, mostly due to the relaxation of the isolation measures taken by the governments. For North and South America, the trend shows that the former presents a sustainable high peak plateau since mid-May, whereas the latter still shows a pattern of rising. For Africa, the results show that the acceleration phase is over after a relatively low





**Figure 3** Estimated cycles—spatiotemporal model with region specific trends, seasonal and cycle components (model M2).

number of reported deaths, despite the experts' prediction of millions of COVID-19 deaths in the continent. For Oceania, the results show that the number of deaths by COVID-19 was small and well controlled by the measures taken by the local governments. Additionally, our results were able to show the importance of the use of spatial information in the task of estimating the trend and mean-reverting components, which are related to the spatial transmission dynamics in an epidemic process.

## DISCUSSION

The first epicentre of the disease was in East Asia, mostly in China and South Korea, which reached the epidemic peak around mid-April. The observed second peak of the COVID-19 in Asia was mostly due to the spread of the disease in the Middle East, South Asia and South East Asia. As of December 2020, deaths by COVID-19 have risen in Iraq, Indonesia and India, which the latter has the world's third-most deaths per 100 000 people and are driving the numbers in Asia. After reaching the peak of the second wave in December, the number of deaths by COVID-19 in the continent has slowly decreased.

The second outbreak epicentre was in Western Europe, especially in Spain, Italy, France and the UK. The continent has experienced the peak of the first wave in the

end of April. However, after consistent drops in the number of daily recorded deaths, the trend in Europe has presented a pattern of rising since the beginning of September, mostly due to the loosening of lockdown and social distance measures. Due to the increased number of deaths, most European countries induced new restriction measures to control the new peak in deaths. As of the end of February, daily deaths by COVID-19 has been failing in most European countries, but in some others, like France, Italy and the Czech Republic, highest numbers of deaths have been recorded in February, helping to maintain the European trend in a high level.

The last epicentre occurred in South and North America, mostly driven by Brazil and the USA, which have, up to 5 October, the second and first highest numbers of recorded deaths by 100 000 people in the world, respectively. While North America presents a sustainable high peak plateau since around mid-May, the South America still shows a pattern of rising. Between October and December 2020, the trend for North America showed a slight decrease, however, returning to the previous level in January 2021, when the USA has recorded the highest levels of daily deaths, and Canada has experienced a winter surge. In the of the analysed sample, the trend component for North America gives signals that the

**Table 2** Model fit statistics

	ME	RMSE	MAE
<b>Model without spatial component (model M1)</b>			
World	-6.7495e-07	90.4915	28.1972
Asia	-3.1729e-08	86.3027	29.1699
Europe	-6.8675e-07	97.3253	34.6779
South America	-1.3301e-06	156.1249	55.9166
North America	-3.2541e-06	126.8047	47.4272
Africa	-9.1082e-08	32.2776	8.7439
Oceania	-6.9796e-08	1.6949	0.6356
<b>Spatiotemporal model (model M2)</b>			
World	-2.0870e-08	1.9503	0.8978
Asia	-2.1111e-07	1.3780	0.6438
Europe	-6.1200e-07	2.3485	1.2610
South America	-3.5786e-06	4.0655	2.0413
North America	4.8557e-06	1.2455	0.6629
Africa	-2.9075e-07	1.3234	0.6380
Oceania	-9.7523e-08	0.07023	0.01209
<b>Spatiotemporal model—air transport network (model M3)</b>			
World	-1.8349e-06	1.8429	0.8790
Asia	-1.2525e-06	1.5854	0.7988
Europe	-1.8629e-06	2.7506	1.4923
South America	-3.8558e-06	2.4099	1.2811
North America	-6.1345e-06	0.9105	0.4961
Africa	-2.5206e-07	0.9226	0.4686
Oceania	-7.8468e-08	0.4489	0.1707

Measures are calculated in relation to the observed number of deaths. The model fit is constructed using the median of the posterior distribution for the predicted deaths. MAE, mean absolute error; ME, mean error; RMSE, root mean squared error.

component has fallen, which may be associated with the advancement of vaccination campaigns. The estimated trend component for South America reached the peak in early January, and gives signals of decreasing at low rates. In this continent, there is a particular concern with Brazil, which has recorded more than 265 000 deaths, the world's second highest number of death by COVID-19, and due to the circulation of the new Brazilian variant of the virus (known as P.1 or VOC202101/02).

It is possible to note that the initial acceleration phases of the Africa and Oceania COVID-19 outbreak grew at a lower rate than those observed in Asia, Europe and Americas. Africa and the most affected by COVID-19 countries have significant differences in reliable reporting, demography and sociocultural aspects. However, despite the experts' prediction of millions of COVID-19 deaths in Africa,<sup>17</sup> the prevalence and mortality are low after two peaks of the disease, in mid-August 2020 and December 2020. The reason for this outcome is still unclear, however, it may be related to the limited number of tests available

and the lack of high-quality data.<sup>18</sup> However, it is worth noting that in the end of the analysed sample, the trend component shows a slight growth pattern. The Oceania has a natural geographical advantage, associated with effective measures, making the cumulative number of cases and deaths small and controllable.<sup>19</sup> In particular, Australia and New Zealand have been praised for the effective lockdown measures taken in both countries, and their relative low number of deaths by COVID-19.

Additionally, the estimated seasonal components (see figure 2) are significantly different among continents, reflecting different patterns of the spreading of the COVID-19, and the government's strategy to reduce the transmission of the disease, which varies from country to country. In particular, it is possible to see the estimated seasonality components for Asia, Europe, South America and Africa with an increasing amplitude over time, while in North America, the increasing amplitude in the seasonality component was followed by a decrease in the end of the analysed sample. For the Oceania, the estimated seasonality component is roughly constant over time. Regarding the cycle components, for the Oceania, it was significantly less noisy than the estimated components for the other continents.

It is worth noting that the latent components of our proposed model are formulated in terms of the exposure offset for the number of populations observed in each region, allowing the counts of COVID-19 death to be comparable. However, it is not an easy task to directly interpret the estimated trend component. Therefore, to provide better understandable results, we carried out a transformation of the estimated trend component towards the number of daily deaths per million inhabitants per region (see online supplemental figure S2), where it is possible to note that Europe reached the peak faster than the other continents until the end of April 2020, where the death trend reveals a turning point to a decreasing pattern, followed by a new acceleration phase in September. Also, it is possible to see that the trend in South America reached the relative highest peak in early January 2021, followed by a decreasing pattern. Despite the Asia has been the first epicentre of the COVID-19 outbreaks in early 2020, Europe and South America have had more deadly COVID-19 outbreaks. For the North America continent, the trend remains stable with a slight decrease in the end of the analysed sample, whereas for the Asia, Oceania and Africa not too much details can be discerned, given the scale of the graph.

Regarding on the fit of the three different formulations, it is possible to observe that M2 and M3 models presented a similar adjustment result, with few gains in the use of air transport information in defining the model's neighbourhood structure. Also, it is important to note that the similarity in model fit of models M2 and M3 can be related to massive travel restrictions and border control to mitigate the outbreak of the COVID-19 pandemic. As discussed in the literature, a larger reduction in the passenger air flow led to a gradual decrease the spread of COVID-19,<sup>20</sup>



which makes this transmission channel less important. In addition, it is worth noting that both models with spatial component (M2 and M3) are clearly superior to the univariate model (M1) in terms of RMSE and MAE, showing the importance of spatial transmission patterns in the COVID-19 dynamics.

Within the limitations of this study, there are some concerns on the reliability of the data used. It is well known that the reporting number of the COVID-19 cases and deaths do not reflect the reality, varying these under-reporting rates from country to country, depending on the number of available tests and the measures taken by the authorities. In particular, the number of available tests has changed over time and space. Despite the limitations, this study provides evidence of the effectiveness of public health measures taken and vaccines and may be used as a tool to guide the authorities to propose new restrictions to control the transmission and to assess the effects of the restrictions loosening. Previous methods to estimate the trend of COVID-19 have been useful in this context, for example, Yang *et al*<sup>21</sup> who found evidence of the effectiveness of the control measures on January 2020 to reduce the COVID-19 epidemic size in China, and Post *et al*<sup>22</sup> who developed dynamic metrics to inform public health surveillance and to identify where and when corrective measures are necessary to cope with the spread of the COVID-19. Based on our outlined spatiotemporal analysis methodology, it is possible to estimate the trends related to COVID-19 in a more reliable way since the proposed model can control for the measurement errors induced by the case accounting processes, making it a useful tool to design strategy for preventing the spread of the disease, targeting vaccines and for the planning of measures of economic reopening.

## CONCLUSION

Based on hierarchical formulations using Bayesian inference methods for spatiotemporal processes, we estimated the trend component of deaths by COVID-19 in 205 considered countries, divided by six continents, namely Asia, Europe, North America, South America, Africa and Oceania. The model was able to capture the observed general patterns in the occurrence of deaths related to COVID-19, overcoming the common problems observed in COVID-19 data, such as spatial heterogeneity, and a delay between onset of symptoms and accurate diagnosis or death, which creates measurement error problem. The proposed model also provided evidence that spatiotemporal models are more accurate than univariate models to estimate the general patterns of the occurrence of deaths related to COVID-19, showing the importance of the spatial transmission dynamics in an epidemic process.

The article's contribution to the formulation of public health policies lies in the development of a spatiotemporal estimation methodology of the global trend of deaths related to COVID-19, controlling for the measurement errors induced by the case accounting processes,

and a more general definition of spatial neighbourhood using the air passenger transport network. Solid measures of these trends are fundamental for the construction of measures of social distancing, targeting of vaccines and other preventive control measures for the reduction of new cases and for the planning of measures of economic reopening in the affected regions.

**Contributors** FV and MPL designed research, performed research and wrote the paper.

**Funding** The authors acknowledge funding from CNPq (grant number: 306023/2018-0), FAPESP (grant number: 2018/04654-9), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (grant number: Finance Code 001) and Instituto Escolhas (no grant number available, information at URL <https://www.escolhas.org/en/catedra/editais-de-bolsas/>).

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available in a public, open access repository. Data of deaths by COVID-19 may be obtained from Our World in Data Project, which is available at <https://ourworldindata.org/covid-deaths>.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iD

Marcio Poletti Laurini <http://orcid.org/0000-0001-5733-2231>

## REFERENCES

- Dong L, Hu S, Gao J. Discovering drugs to treat coronavirus disease 2019 (COVID-19). *Drug Discov Ther* 2020;14:58–60.
- Sanders JM, Monogue ML, Jodlowski TZ, *et al*. Pharmacologic treatments for coronavirus disease 2019 (COVID-19): a review. *JAMA* 2020;323:1824–36.
- Alexandrov T, Bianconcini S, Dagum EB, *et al*. A review of some modern approaches to the problem of trend extraction. *Econom Rev* 2012;31:593–624.
- Li Q, Feng W, Quan Y-H. Trend and forecasting of the COVID-19 outbreak in China. *J Infect* 2020;80:469–96.
- Gupta R, Pal SK. Trend analysis and forecasting of COVID-19 outbreak in India. *medRxiv* 2020.
- Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci Total Environ* 2020;729:138817.
- Perone G. An ARIMA model to forecast the spread and the final size of COVID-2019 epidemic in Italy. *medRxiv* 2020.
- Contreras S, Biron-Lattes JP, Andrés Villavicencio H, Nyna Llanovarcud-Kawles, and Álvaro Olivera-Nappa. Statistically-based methodology for revealing real contagion trends and correcting delay-induced errors in the assessment of COVID-19 pandemic. *Chaos Solitons Fractals* 2020;12376.
- Russell TW, Hellewell J, Abbott S, *et al*. Using a delay-adjusted case fatality ratio to estimate under-reporting. *CMMID Repository* 2020.
- Skiera B, Jürgensmeier L, Stowe K. How to best predict the daily number of new infections of Covid-19. *arXiv* 2020;03937.
- Harvey AC. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, 1990.

- 12 Valente F, Laurini MP. Robust trend estimation for covid-19 data, 2020. Available: <https://www.dropbox.com/s/6dqhsy6lehwar45/COVIDpreprint.pdf?dl=0>
- 13 Besag J. Spatial interaction and the statistical analysis of lattice systems. *J Royal Stat Soc* 1974;36:192–225.
- 14 Simpson D, Rue H, Riebler A, *et al.* Penalising model component complexity: a principled, practical approach to constructing priors. *Stat Sci* 2017;32:1–28.
- 15 Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J Royal Stat Soc* 2009;71:319–92.
- 16 Rue H, Held L. *Gaussian Markov random fields: theory and applications*. CRC Press, 2005.
- 17 Martinez-Alvarez M, Jarde A, Usuf E, *et al.* COVID-19 pandemic in West Africa. *Lancet Glob Health* 2020;8:e631–2.
- 18 Mbow M, Lell B, Jochems SP, *et al.* COVID-19 in Africa: dampening the storm? *Science* 2020;369:624–6.
- 19 Beck MJ, Hensher DA, Wei E. Slowly coming out of COVID-19 restrictions in Australia: implications for working from home and commuting trips by CAR and public transport. *J Transp Geogr* 2020;88:102846.
- 20 Linka K, Peirlinck M, Sahli Costabal F, *et al.* Outbreak dynamics of COVID-19 in Europe and the effect of travel restrictions. *Comput Methods Biomech Biomed Engin* 2020;23:710–7.
- 21 Yang Z, Zeng Z, Wang K, *et al.* Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis* 2020;12:165–74.
- 22 Post LA, Issa TZ, Boctor MJ, *et al.* Dynamic public health surveillance to track and mitigate the US COVID-19 epidemic: longitudinal trend analysis study. *J Med Internet Res* 2020;22:e24286.