

BMJ Open Questioning the seasonality of SARS-COV-2: a Fourier spectral analysis

Riccardo Cappi,¹ Luca Casini ,² Davide Tosi,¹ Marco Rocchetti ²

To cite: Cappi R, Casini L, Tosi D, *et al.* Questioning the seasonality of SARS-COV-2: a Fourier spectral analysis. *BMJ Open* 2022;**12**:e061602. doi:10.1136/bmjopen-2022-061602

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-061602>).

Received 01 February 2022
Accepted 04 April 2022

ABSTRACT

Objectives To investigate the hypothesis of a seasonal periodicity, driven by climate, in the contagion resurgence of COVID-19 in the period February 2020–December 2021.

Design An observational study of 30 countries from different geographies and climates. For each country, a Fourier spectral analysis was performed with the series of the daily SARS-CoV-2 infections, looking for peaks in the frequency spectrum that could correspond to a recurrent cycle of a given length.

Settings Public data of the daily SARS-CoV-2 infections from 30 different countries and five continents.

Participants Only publicly available data were utilised for this study, patients and/or the public were not involved in any phase of this study.

Results All the 30 investigated countries have seen the recurrence of at least one COVID-19 wave, repeating over a period in the range 3–9 months, with a peak of magnitude at least half as large as that of the highest peak ever experienced since the beginning of the pandemic until December 2021. The distance in days between the two highest peaks in each country was computed and then averaged over the 30 countries, yielding a mean of 190 days (SD 100). This suggests that recurrent outbreaks may repeat with cycles of different lengths, without a precisely predictable seasonality of 1 year.

Conclusion Our findings suggest that COVID-19 outbreaks are likely to occur worldwide, with cycles of repetition of variable lengths. The Fourier analysis of 30 different countries has not found evidence in favour of a seasonality that recurs over 1 year period, solely or with a precisely fixed periodicity.

INTRODUCTION

After almost 2 years from the start of the COVID-19 pandemic, the scientific community is still arguing about many of the characteristics of this virus and its spread, as well as what the best course of action in the fight against it is.^{1 2} While the public may find this lack of consensus disheartening, every scientist knows this was to be expected when dealing with such unprecedented phenomena, especially given the enormous uncertainty around the data that concern it. Every analysis that has tried to leverage information on a global scale, in fact, had to deal with the limitation of having a largely inhomogeneous data set, where differences in the data collection process and, even more importantly, in the

Strengths and limitations of this study

- The use of a Discrete Fourier Transform offers the advantage of a temporal decomposition of the time series of the daily SARS-COV-2 cases allowing to explore the temporal relationships among recurrent outbreaks.
- Applying the Discrete Fourier Transform represents an appropriate method to question the hypothesis of a seasonal structure of COVID-19 assumed to recur on a 1-year long period basis.
- This observational study has avoided the problem to quantify the role attributable to various climatic factors or control measures, like temperatures or vaccination.
- The time series of the daily SARS-CoV-2 cases was not normalised or corrected and did not include the outbreak started in December 2021.
- The Fourier uncertainty principle may render the results we achieved at low frequencies somewhat uncertain given the availability of only 2 years of data.

actions taken by each country contributed to confounding the factors being studied.³

In this complex context, at the end of 2021, as the virus has been around for almost 2 years, a discussion started about the possibility of COVID-19 following a seasonal pattern, similar to many other viral infections, like measles and influenza, for example. This idea gained momentum probably because of how the contagion receded during the summer months in many Western countries, to start climbing back up with the start of autumn and finally reaching a new peak during the winter holidays.

From a scientific perspective, the debate on an infection pattern that repeats over a 1-year period has been driven by several analyses investigating the correlation between SARS-COV-2 and various climatic (and environmental) factors, such as temperature, humidity and UV radiations. The rationale behind this research is that if a negative correlation between SARS-COV-2 and higher temperatures and exposure levels to UV radiation can be demonstrated, then lower COVID-19 infection rates should happen in



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Theoretical and Applied Sciences, University of Insubria, Varese, Lombardia, Italy

²Department of Computer Science and Engineering, University of Bologna, Bologna, Emilia-Romagna, Italy

Correspondence to

Luca Casini;
luca.casini7@unibo.it

some given seasons of the year. Humidity, instead, appears to have a U-shaped relation with SARS-CoV-2 infection rates, as only values that hover 50% seem to shorten the virus life.

Along this line of research, remarkable are the studies of D'Amico *et al* who used a multivariate regression to assess the influence of temperatures and vaccinations on mortality rates in temperate climate countries. They found a negative correlation with temperatures and discovered that the vaccination's effect grew larger as the temperature decreased.⁴ Similarly, Ma *et al* studied the problem in the USA using a generalised additive mixed model. Their findings are that temperatures are negatively correlated with COVID-19, in an almost linear way, in the range of 20°C–40°C.⁵

However, some other research begins to point out the weaknesses of many of the aforementioned studies. For example, Fontal *et al* while studying the negative correlation between the virus and both higher temperature and humidity, found that there are moments in which this correlation can be inverted, typically corresponding to summer outbreaks in certain regions.⁶ The authors suggest that this can be due when various human activities take over, like intensive use of air conditioning, lack of preventive measures and uncontrolled mass gatherings. Also, Sera *et al* have expressed their concerns, concluding that the effect of weather, while present, is negligible when compared with the decisive impact of control interventions.⁷ More interestingly, Baker *et al* have argued that climate factors can play an important factor in the infection when the virus is in the endemic stage. In contrast, during the pandemic stage, it only drives modest changes.⁸ Finally, very inspiring theoretical results were achieved also in a study by Telles *et al* where it has been demonstrated that a combination of factors, including climate, control policies and the use of urban spaces could influence the seasonality of COVID-19.⁹

Beyond these scientific studies, the positive effect of good weather seems to contrast with several COVID-19 contagions that have broken out, with broad impact, even if with unfavourable climates to the spread of the virus. For example, while this is completely anecdotal, one could wonder what mechanisms were behind the resurgence of the contagion in May 2020 in Israel.¹⁰ Similarly, the 2021 Olympic Games in Japan took place during the summer when the weather was optimal, but the virus spread, even in the presence of high security standards.¹¹ In the same period, the European Football Championship took place, and this tournament was connected with an increase of new cases in many involved countries.¹² Not to say about the 4 July 2021, US presidential Speech, when the US President declared the final success in beating the pandemic, but a new peak hit strong just a few weeks later.¹³ Finally, the third wave across Europe started at the end of 2021 summer in many eastern countries, when the temperatures were still relatively high.

Following this scientific debate, in this work, we chose another technical perspective in order to investigate

the 1-year seasonality hypothesis, employing techniques from signal processing in order to study the presence of evidence towards periodicity in the COVID-19 recurrent outbreaks, or lack thereof. Relying on a Fourier spectral analysis of the daily SARS-CoV-2 infections data, at a worldwide level, we looked for peaks in the frequency spectrum that could inform us about the length of the recurrent cycles of COVID-19 outbreaks. This has allowed us to question on the sinusoidal seasonality assumed to recur over 1-year period, solely.

METHODS

We focused our study on a wide selection of worldwide countries, chosen following the Köppen climate classification.¹⁴ This classification divides climates into five main groups, where each group is considered based on seasonal precipitation and temperatures. The five main groups are: tropical (A), dry (B), temperate (C), continental (D) and polar (E). We analysed 30 different countries that cover all the five groups, with several of the selected countries belonging to two or more groups, given their vast geography (eg, India, Russia and the USA, to cite a few). The complete list of the 30 countries follow below, each with its prevalent type of climates: Argentina (B, C), Australia (A, B, C), Austria (D, E), Belgium (C), Brazil (A, C), Canada (C, D, E), Chile (B, C, D), Colombia (A, C), Croatia (C), Denmark (D), France (C), Germany (C, D), Hungary (D), India (A, B, C, D), Indonesia (A), Italy (B, C), Japan (A, C, D), Mexico (A, B, C), Morocco (B, C), Norway (D, E), Portugal (C), Russia (D, E), Saudi Arabia (B), South Africa (B, C), South Korea (C, D), Spain (B, C), Sweden (D, E), Turkey (B, C, D), UK (C) and USA (B, C, D, E).

Notice that our selection includes 18 out of the 20 countries of the group of 20. China was excluded just because its SARS-COV-2 data are not made available on a regular basis. Also, the European Union (EU) was not considered as a whole. Yet, in place of EU, we included the following EU members: Austria, Belgium, Croatia, Denmark, France, Germany, Hungary, Italy, Portugal, Spain and Sweden. The period of observation for our study started on 1 February 2020 until 4 December 2021, with the decision not to take into consideration the strong SARS-CoV-2 outbreak that hit Europe in December 2021, as the progression of this wave was still ongoing in many of the investigated countries during our analysis.

The method we adopted for our investigation was a Fourier spectral analysis. In particular, we used a Discrete Fourier Transform (DFT) to examine the periods length in the spectrum of the COVID-19 data, by converting the time series of the number of the new daily SARS-CoV-2 cases to the frequency domain.¹⁵ This Fourier frequency spectrum analysis was performed with the precise intent to obtain a converted peak spectrum, indicating the strength and the recurrence of the pandemic waves. In particular, we looked for peaks in the frequency spectrum that could reasonably indicate a periodicity with a certain

length. Employing a spectral analysis on the time series of the COVID-19 cases has allowed us to understand, with less ambiguity, the period length of the recurrent outbreaks, instead of counting and observing the number of infections, directly.

The 1-dimensional DFT $y[k]$ of length N , of the length- N sequence $x[n]$, is defined as follows:

$$y[k] = \sum_{n=0}^{N-1} e^{-2\pi j \frac{kn}{N}} x[n],$$

where $y[k]$ corresponds to magnitude of the k th frequency and n represents the n th day of the time series, with x being the daily number of SARS-COV-2 cases registered on that n th day of the series. The number of the analysed days (ie, the sampling period) was equal to 730 (2 years). Since our study's real period of observation started on 1 February 2020 (until 24 December 2021), the string x was left padded with zeros to reach 730 samples. This zero padding did not alter the validity of the operation since in all the considered countries no SARS-COV-2 infection was registered before 1 February 2020. To conclude, using a Python library called *SciPy* (<https://scipy.org/>), we performed a DFT of the time series of the SARS-COV-2 data of each country, that returned all the peaks in the frequency spectrum at their corresponding frequency which can be inverted to obtain the repetition period.

It is to notice that all the data used for our DFT investigation are available at the following site <https://github.com/owid/covid-19-data>, with the instructions on how to use made available at: <https://github.com/owid/covid-19-data/blob/master/public/data/README.md>. The results of all the DFT computations are fully reproducible by using the method described above. The DFT code can be downloaded from: https://github.com/mistermagpie/covid_periodicity. It is finally worth mentioning that all that aforementioned data on COVID-19 infections (OWID) are maintained by the Johns Hopkins University Centre for Systems Science and Engineering.¹⁶

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

RESULTS

Figures 1–5 show the FTs obtained for all the countries subject of our study, using two separate plots. For each country, the leftmost plot reports the time series of the new daily SARS-COV-2 infections during the observed period (1 February 2020–24 December 2021); the rightmost plot shows the result of the Fourier transform applied to the time series of the leftmost plot.

In all the leftmost plots, $x[n]$ is the COVID-19 data series of interest, where x is the number of daily new infections per each day n . All the rightmost plots of Figures 1–5 depict, in the y -axis, the Fourier frequency spectrum.

This spectrum comprises a red line with endpoints at the junctions, representing the COVID-19 peaks. Each peak is depicted with its magnitude (y -axis) and with its corresponding frequency k on the x -axis, based, in turn, on a semilogarithmic yearly scale. Two preliminary facts are noteworthy. First, we can observe a peak in the frequency spectrum representing the 7-day cycle associated with the case reporting process, on the rightmost side of all these DFT plots. This was a quite expected fact, since the reporting process causes an oscillation during the week, in almost all the considered countries. Obviously, this peak recurs 52 times in the DFT plot ($k=52$), being 52 the number of the weeks in 1 year. Second, we observe a peak with typically the highest magnitude at the opposite side of the spectrum of all our DFT plots (leftmost side). The meaning of this peak is simply that the COVID-19 phenomenon has occurred each year, during the 2 years of observation, in all the countries of interest. Its repetition cycle is, naturally, equal to 1 ($k=1$).

More interestingly, on the leftmost side of the DFT plots of figures 1–5, we have three different sectors coloured in orange, green and pink (from right to left). Those sectors display temporal intervals, respectively, equal to 3–6 months (orange), 6–9 months (green) and 9–12 months (pink). They should be interpreted as follows: if one observes for a certain country the presence of a peak in a given coloured sector of the plot (say the green one, for example), this means that country has been hit by a COVID-19 outbreak, which has recurred with a period of 6–9 months. More precisely, if that peak lies on the x -axis in correspondence of a value of $k=2$, this implies that we have had two outbreaks of similar magnitude per year in that country. Coming now to our results, our 30 DFT plots of figures 1–5 reveal that, in the observed period, all the 30 investigated countries have seen the recurrence of at least one COVID-19 wave, repeating over a variable period in the range 3–9 months, with a peak of magnitude (roughly equivalent to the number of new infections) at least half as high as that of the highest peak ever experienced since the beginning of the pandemic until December 2021. These findings suggest that strong COVID-19 outbreaks may repeat with cycles of different lengths, without a precisely predictable seasonality of 1 year.

Given the well-known Fourier uncertainty principle,¹⁷ we developed a further analysis. We returned to the leftmost plots of figures 1–5, looking for the COVID-19 peaks, recurring in each country, but adopting a more traditional technique. Specifically, using a 7-day rolling average as the raw data of leftmost plots of figures 1–5 present a weekly periodicity, due to the way COVID-19 tests are carried out and registered, we considered a peak has happened in a given day n , if the number of SARS-COV-2 infections registered in that day was larger than the number of daily SARS-COV-2 cases reported in the 28 days both before and after n . Not only, but to be considered a peak, the number of infections registered on that day n had to be larger than a given threshold computed as the 85% of the average of the daily cases reported in

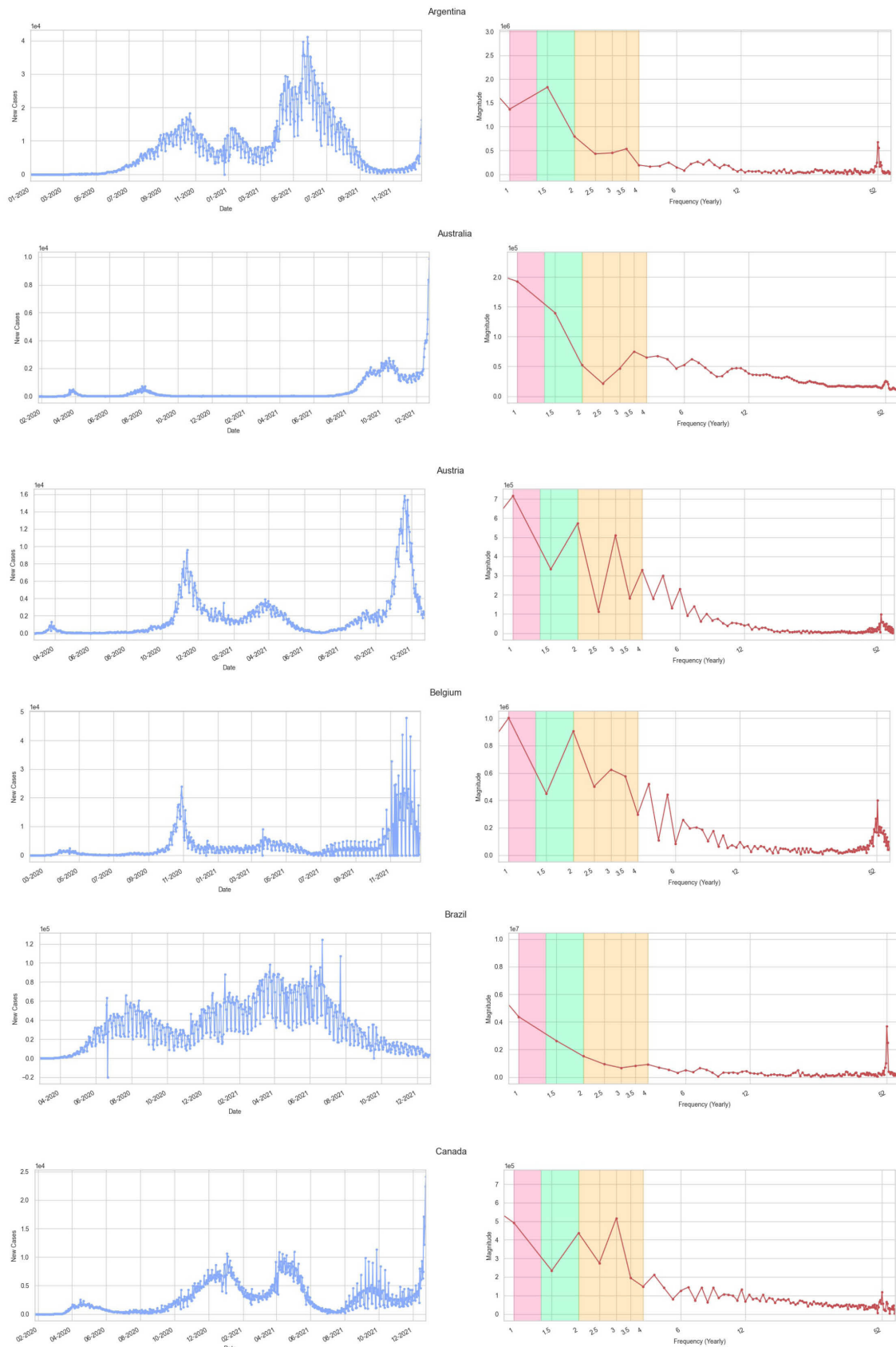


Figure 1 Plots for Argentina, Australia, Austria, Belgium, Brazil and Canada. Leftmost plots: on the y-axis number of new daily SARS-CoV-2 cases in the period 1 February 2020–24 December 2021 (x-axis). Rightmost plots: Fourier frequency spectrum on the y-axis, with dots lying on the red line corresponding to wave peaks. Coloured sectors indicate precise recurring cycles: pink (9–12 months), green (6–9 months) and orange (3–6 months). The k index on the x-axis of the rightmost plots indicated the repetition frequency on a semilogarithmic scale (on a 1-year period).

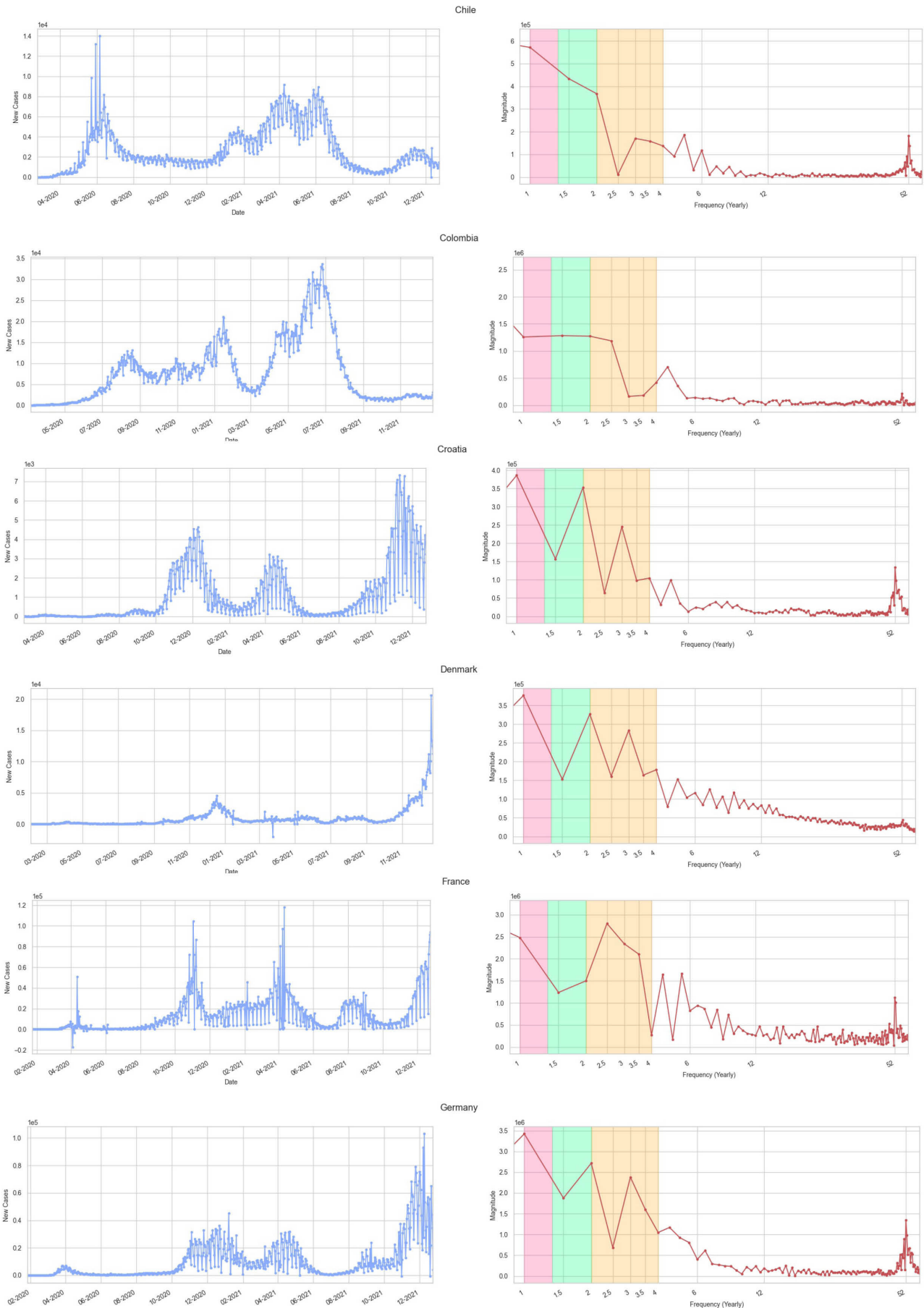


Figure 2 The same plots of figure 1 but for Chile, Colombia, Croatia, Denmark, France and Germany.

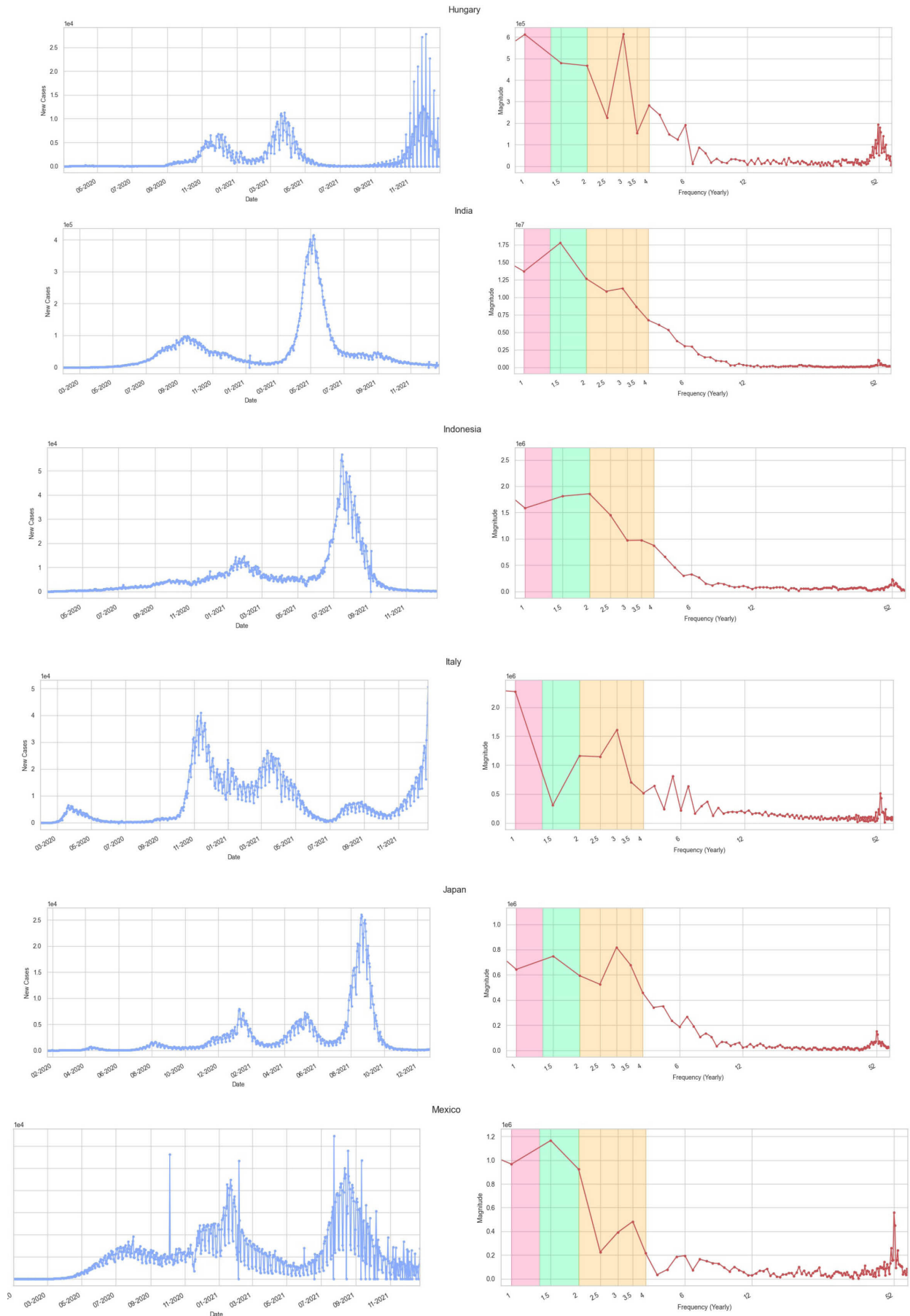


Figure 3 The same plots of figure 1 but for Hungary, India, Indonesia, Italy, Japan and Mexico.

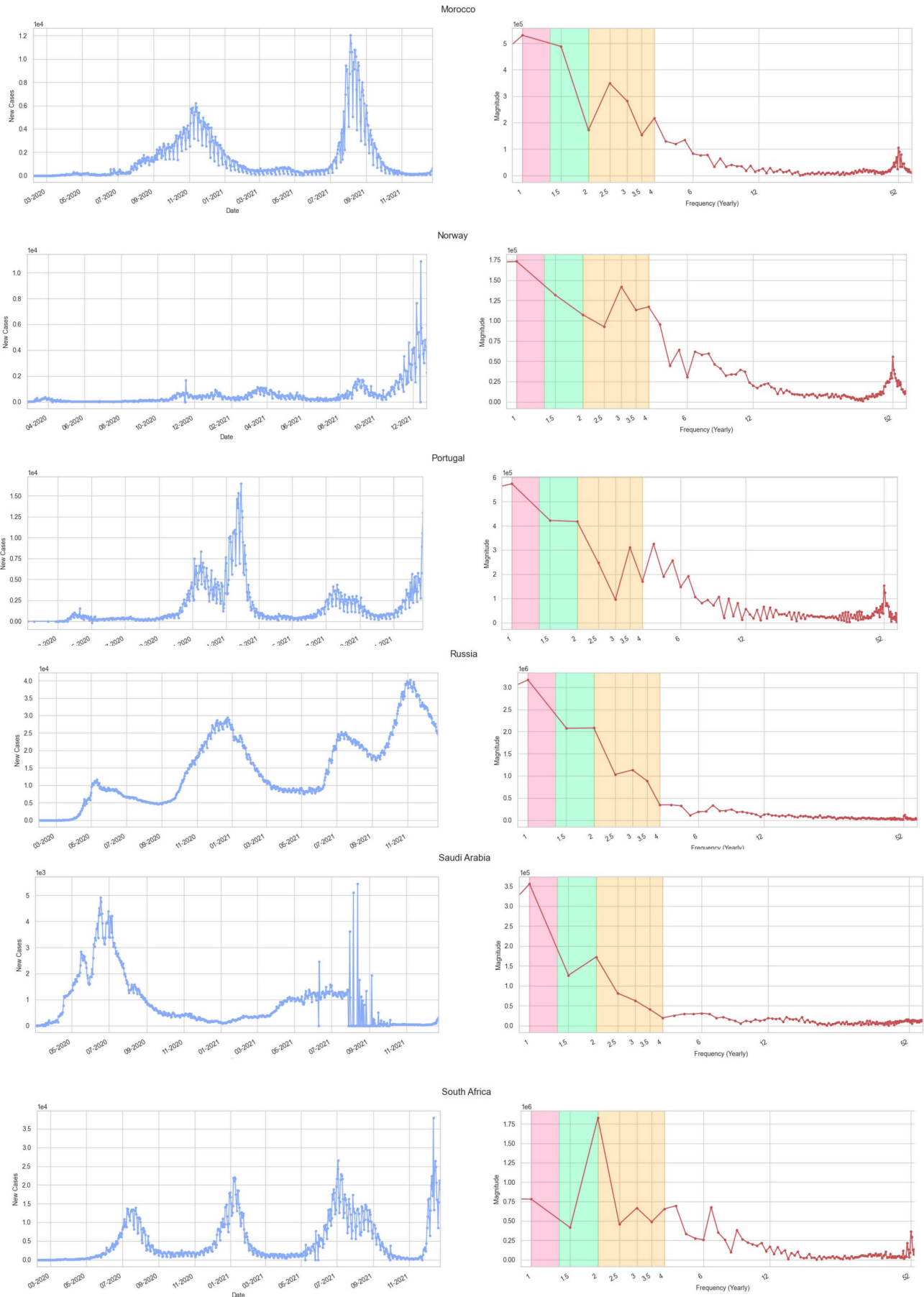


Figure 4 The same plots of figure 1 but for Morocco, Norway, Portugal, Russia, Saudi Arabia and South Africa.

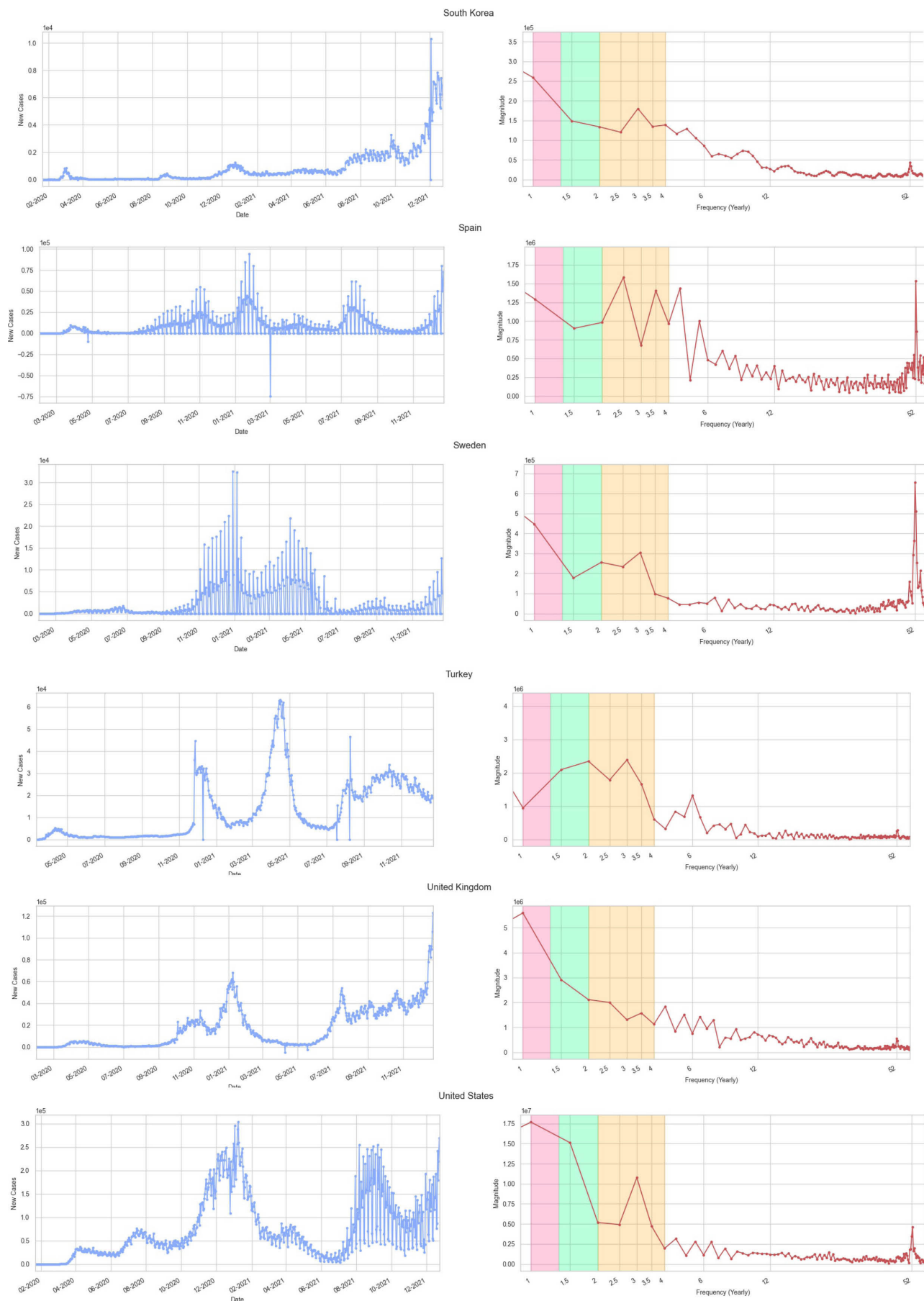


Figure 5 The same plots of figure 1 but for South Korea, Spain, Sweden, Turkey, UK and USA.

all the days since the beginning of the pandemic until n . Choosing 28 days comes from the working definition of wave as provided in reference,¹⁸ where the three quarters of the upward periods of many studied COVID-19 waves lasted less than a month. Similarly, for the downward periods. The rationale behind the concept of having a threshold came, instead, from the need to filter out all the micropeaks. Finally, on computation of all the peaks for each country during the period of interest, we chose the two highest ones. Then, we computed for each such pair the distance in days between them (the Python code for this simple algorithm can be downloaded from: <https://github.com/riccardocappi/Seasonality-SARS-CoV-2>). **Table 1** reports the corresponding results.

Precisely, the number of peaks, the distance in days between the two highest ones and their corresponding dates are given for each country. It is interesting to notice that if we average, all over the 30 countries, the values of the temporal distances between the two highest peaks, we obtain a mean of 190 days (SD 100). In other words, we have obtained a confirmation for all our 30 countries of the recurrence of peaks, with an average period of almost 6 months and a SD of nearly 3 months. Moreover, the 80% of the examined countries has that (maximal) temporal distance which falls below the value of 1 year. Even if we restrict this analysis to only the 13 European countries of our data set: Austria, Belgium, Croatia, Denmark, France, Germany, Hungary, Italy, Norway, Portugal, Spain, Sweden and the UK, we achieve an average of almost five peaks in a 2 years period, with a mean distance between the two highest ones equal to 171 days (SD 85), once again confirming the hypothesis that strong COVID-19 waves may repeat with cycles whose duration break the seasonality pattern of 1 year.

Table 1 also reports, for each registered peak, the variant of the virus that could be considered prevalent at the time of the corresponding outbreak. In particular, to individuate the variant to be associated to each peak, we utilised, for each period and for each country, both the proportion of the total number of sequences collected over time, that fall into some given variant groups and the corresponding phylogenetic tree. These data are extrapolated, respectively, by the two following initiatives: Covariants.org (specifically, <https://covariants.org/per-country>)¹⁹ and Nextstrain.org (specifically, <https://nextstrain.org/ncov/gisaid/global>).²⁰ It is worth noticing that both these initiatives are enabled by data shared by the GISAID.org project²¹ that collects all the genome sequences of COVID-19, worldwide. While it is surely of interest the relation between a peak and the frequencies of the sequences collected during a given outbreak, it should be considered that the information about the clades, portrayed in **table 1**, cannot be always assumed as necessarily representative. The motivation is that genome sampling may not be equal across different countries and periods, with some countries with low sequencing numbers or even with some samples more likely to be sequenced than others. It is worth concluding by pointing

out that each mentioned variant in **table 1** has been identified based on the conventional names proposed by the genome sequencing initiatives mentioned above (ie, Covariants.org and Nextstrain.org.). Essentially, each variant's name is comprised of a two-digit number that represents the year, a progressive alphabetical letter, plus a letter from the Greek alphabet as provided by the WHO organisation (eg, 21J Delta).

DISCUSSION

Seasonality typically refers to a single, recurring pattern with a fixed frequency. The results we achieved highlight how, with COVID-19, strong evidence of a seasonal pattern that repeats over a 1-year fixed period cannot be found. Instead, several repeating outbreaks, not necessarily occurring with a fixed frequency, can be observed. In particular, the Fourier spectral analysis of the time series of the SARS-CoV-2 cases of all the 30 countries we have studied has revealed the recurrence of at least one COVID-19 wave (often two or more), repeating over a variable period, in the range 3–9 months, with peaks of magnitude comparable to that of the highest peak ever experienced since the beginning of the pandemic until December 2021. Indeed, the situation is more complicated than previous studies have revealed. In fact, while some of them consider COVID-19 as a seasonal low-temperature infection, the precise role of temperature, humidity and exposition to UV radiation remains poorly understood. From this perspective, our study has tried to follow an alternative path with spectral analysis of the series of the daily SARS-CoV-2 cases, while looking for peaks in the frequency spectrum to understand the presence of repeating cycles and quantify their lengths.

We recognise that, with our analysis, we have avoided quantifying precisely the role attributed to various climatic factors or control measures, like temperatures or vaccination. This can be seen as a limitation of our study as the magnitude of the effects of those factors should be investigated thoroughly. Yet, there are precise motivations behind our choice. On one side, we have decided to avoid taking part in the scientific discussion about the dominant role of climate versus control measures, including vaccination, as the best solutions that can drive substantial changes to the pandemic trajectory. On the other side, we have tried to observe a natural phenomenon, just resorting to a mathematical technique able to detect the presence of evidence towards periodicity/non-periodicity in the spread of COVID-19, with neutrality and regardless of the underlying factors. Another factor that could have an influence on the seasonality is how urban spaces are lived and the corresponding impact on the spreading patterns of the virus.⁹ While we recognise that this factor is important and comparable with those of climate and policies, we admit that we have not addressed this issue in this study.

Another technical limitation of our approach was the decision not to put a special focus on those countries where

Table 1 Country, type of climates for that country, number of COVID-19 outbreak peaks, distance in days between the two highest peaks, dates of the two highest peaks, dates of the remaining peaks and more frequent clades per peak

Country	Climate types	Number of peaks	Distance between the two highest peaks (days)	Dates of the two highest peaks (more frequent clades)	Dates of the remaining peaks (more frequent clades)
Argentina	B/C	3	214	21 October 2020 (20B/C/D) 23 May 2021 (20J Gam., 21G Lambda)	11 January 2021 (20B, 20I Alpha, 20J Gam.)
Australia	A/B/C	3	432	4 August 2020 (20B/C/F) 10 October 2021 (21J Delta)	30 March 2020 (19A/B, 20A/B/C)
Austria	D/E	5	376	13 November 2020 (20A) 24 November 2021 (21J Delta)	28 March 2020 (20A/B/C) 1 April 2021 (20I Alpha) 15 September 2021 (21J Delta)
Belgium	C	4	148	30 October 2020 (20A/B) 27 March 2021 (20I Alpha)	15 April 2020 (20A/C) 12 August 2020 (20A/C)
Brazil	A/C	4	87	27 March 2021 (20J Gamma) 22 June 2021 (20J Gamma)	28 July 2020 (20B) 12 January 2021 (20B, 20J Gamma)
Canada	C/D/E	4	93	9 January 2021 (20B/C/G) 12 April 2021 (20I Alpha, 20J Gamma)	22 April 2020 (19A, 20B/C) 13 September 2021 (21I/J Delta)
Chile	B/C/D	6	55	14 April 2021 (20J Gam., 21G Lambda) 8 June 2021 (20J Gam., 21G Lambda)	12 June 2020 (19A, 20B/D) 1 October 2020 (20B/D) 25 January 2021 (20B/D/G, 20I Alpha) 15 November 2021 (21J Delta)
Colombia	A/C	4	159	20 January 2021 (19A, 20B/C) 28 June 2021 (21H Mu)	16 August 2020 (20A/B) 2 November 2020 (19A, 20B)
Croatia	C	6	333	13 December 2020 (20B) 11 November 2021 (21J Delta)	1 April 2020 (20A) 15 July 2020 (20B) 29 August 2020 (20B) 21 April 2021 (20I Alpha)
Denmark	D	6	145	18 December 2020 (20B/E) 12 May 2021 (20I Alpha)	8 April 2020 (20A/C) 23 September 2020 (20A/B/E) 16 March 2021 (20I Alpha) 16 August 2021 (21J Delta)
France	C	5	161	7 November 2020 (20A) 17 April 2021 (20I Alpha)	18 April 2020 (19B, 20A) 2 November 2021 (20I Alpha) 16 August 2021 (21J Delta)
Germany	C/D	4	123	23 December 2020 (20A/E) 25 April 2021 (20I Alpha)	2 April 2020 (19B, 20A/C) 10 September 2021 (21J Delta)
Hungary	D	3	113	3 December 2020 (20A) 26 March 2021 (20A, 20I Alpha)	13 April 2020 (20A)
India	A/B/C/D	2	234	16 September 2020 (20A/B) 8 May 2021 (21A/J Delta)	–
Indonesia	A	3	167	1 February 2021 (20A/B) 18 July 2021 (21I/J Delta)	26 September 2020 (20A/B)
Italy	B/C	5	126	16 November 2020 (19A, 20A) 22 March 2021 (20E)	26 March 2020 (20E, 20I Alpha) 11 January 2021 (20I Alpha) 27 August 2021 (20J Alpha)
Japan	A/C/D	5	226	11 January 2021 (20B) 25 August 2021 (21J Delta)	15 April 2020 (19B, 20A/B) 9 August 2020 (20B) 14 May 2021 (20I Alpha)
Mexico	A/B/C	4	213	21 January 2021 (20A/B/C) 22 August 2021 (21I/J Delta)	1 August 2020 (20A/B) 9 October 2020 (20A/B/C)
Morocco	B/C	4	266	17 November 2020 (20A/B) 10 August 2021 (21J Delta)	22 April 2020 (20A) 25 June 2020 (20A)
Norway	D/E	6	167	22 March 2021 (20I Alpha) 5 September 2021 (21I/J Delta)	29 March 2020 (19A, 20A/B) 23 November 2020 (20A/B/C/E) 10 January 2021 (20A/B/E, 20I Alpha) 26 May 2021 (20I Alpha)

Continued

Table 1 Continued

Country	Climate types	Number of peaks	Distance between the two highest peaks (days)	Dates of the two highest peaks (more frequent clades)	Dates of the remaining peaks (more frequent clades)
Portugal	C	5	70	19 November 2020 (20B/E) 28 January 2021 (20E, 20I Alpha)	3 April 2020 (20B) 13 July 2020 (20B) 23 July 2021 (21J Delta)
Russia	D/E	4	315	26 December 2020 (20B/C) 6 November 2021 (21J Delta)	12 May 2020 (20/B) 15 July 2021 (21J Delta)
Saudi Arabia	B	3	412	20 June 2020 (20A) 6 August 2021 (21I Delta)	2 July 2021 (21I Delta)
South Africa	B/C	4	178	11 January 2021 (20A, 20hours Beta) 8 July 2021 (21I/J Delta)	19 July 2020 (20B/D) 22 August 2021 (21J Delta)
South Korea	C/D	7	46	15 August 2021 (21I Delta) 30 September 2021 (21I/J Delta)	4 March 2020 (19B, 20C) 27 August 2020 (20A/C) 25 December 2020 (20C) 20 February 2021 (20C, 20I Alpha, 21D Eta) 23 April 2021 (20A, 20I Alpha)
Spain	B/C	5	174	26 January 2021 (20E, 20I Alpha) 19 July 2021 (21J Delta)	31 March 2020 (19B, 20A/B) 4 November 2020 (20E) 27 April 2021 (20I Alpha)
Sweden	D/E	4	91	11 January 2021 (20A/E) 12 April 2021 (20I Alpha)	29 April 2020 (19A, 20B/C) 18 June 2020/ (20B)
Turkey	B/C/D	5	139	2 December 2020 (20A/B, 20I Alpha) 20 April 2021 (20I Alpha, 20hours Beta)	16 April 2020 (20A) 15 August 2021 (21A/J Delta) 15 October 2021 (21J Delta)
UK	C	6	193	9 January 2021 (20I Alpha) 21 July 2021 (21J Delta)	22 April 2020 (19A, 20D) 16 November 2020 (20B/E) 8 September 2021 (21J Delta) 23 October 2021 (21J Delta)
USA	A/B/C/D/E	5	245	11 January 2021 (20B/C/G, 21C Epsilon) 13 September 2021 (21J Delta)	10 April 2020 (19A/B, 20B/C) 22 July 2020 (20B/C/G) 14 April 2021 (20I Alpha)

The mean distance between the two highest peaks is 190 days (SD 100).

the number of cases has had high variability, the reason being most likely that the number of tests done each day can have varied as much. We could have normalised those cases with the number of tests, before subjecting them to the DFT, but this datum is often unreliable and may lead, in turn to unrealistic normalised values, so we decided to avoid this. An additional technical limitation of this study is that the Fourier transform may return results, especially at the lowest frequencies, with a variable degree of uncertainty. Hence, to confirm our results, we have developed a parallel analysis directly performed on the number of the new daily SARS-COV-2 cases of interest. Alternatively, we could have performed a spectral analysis of our epidemiological time-series with wavelets. These techniques appear somewhat attractive because they are more appropriate to treat non-stationary signals, but still have to deal with the natural limitation represented by the fact that the pandemic has been in progress so far, and we only have a 2-year data set. Not to mention the exclusion from our research of the intense SARS-CoV-2 outbreak started in December 2021 in many of the 30 considered countries.

The motivation for this exclusion is that the progression of this wave was still ongoing at the moment of our study.

CONCLUSION

We have applied a mathematical technique from signal processing to investigate in 30 different countries the hypothesis whether COVID-19 outbreaks either repeat with a fixed periodicity or occur following unpredictable patterns. The Fourier spectral analysis of the time series of the SARS-CoV-2 cases we have examined has suggested that strong COVID-19 waves may repeat with cycles of different lengths, without a precisely predictable periodicity (1 year, or similar). With the scientific community that appears divided into two factions, which alternatively maintain the importance of the role of meteorological factors versus control measures (including vaccination), we argue we have provided an improved understanding of how the virus may spread, regardless of the presence of several factors that can further confound the scenario.



Contributors RC, DT, LC and MR equally contributed to conceive, design, write, manage and revise the manuscript. MR acts as a guarantor for this study.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data of this study are available in a public, open access repository (<https://github.com/owid/covid-19-data>). Instructions on how to use those data are available at: <https://github.com/owid/covid-19-data/blob/master/public/data/README.md>. The DFT code is available at: https://github.com/mister-magpie/covid_periodicity. The Python code for the peak searching algorithm is available at: <https://github.com/riccardocappi/Seasonality-SARS-CoV-2>.

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 iDs

Luca Casini <http://orcid.org/0000-0002-3468-6974>

Marco Roccetti <http://orcid.org/0000-0003-1264-8595>

REFERENCES

- Baum F, Freeman T, Musolino C, *et al*. Explaining COVID-19 performance: what factors might predict national responses? *BMJ* 2021;372:n91.
- Casini L, Roccetti M. Reopening Italy's schools in September 2020: a Bayesian estimation of the change in the growth rate of new SARS-CoV-2 cases. *BMJ Open* 2021;11:e051458.
- Tosi D, Campi A. How data analytics and big data can help scientists in managing COVID-19 diffusion: modeling study to predict the COVID-19 diffusion in Italy and the Lombardy region. *J Med Internet Res* 2020;22:e21081.
- D'Amico F, Marmiere M, Righetti B, *et al*. COVID-19 seasonality in temperate countries. *Environ Res* 2022;206:112614.
- Ma Y, Pei S, Shaman J, *et al*. Role of meteorological factors in the transmission of SARS-CoV-2 in the United States. *Nat Commun* 2021;12:3602.
- Fontal A, Bouma MJ, San-José A, *et al*. Climatic signatures in the different COVID-19 pandemic waves across both hemispheres. *Nat Comput Sci* 2021;1:655–65.
- Sera F, Armstrong B, Abbott S, *et al*. A cross-sectional analysis of meteorological factors and SARS-CoV-2 transmission in 409 cities across 26 countries. *Nat Commun* 2021;12:5968.
- Baker RE, Yang W, Vecchi GA, *et al*. Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. *Science* 2020;369:315–9.
- Roberto Telles C, Lopes H, Franco D. SARS-COV-2: SIR model limitations and predictive constraints. *Symmetry* 2021;13:676.
- Bar-On YM, Goldberg Y, Mandel M, *et al*. Protection of BNT162b2 vaccine booster against Covid-19 in Israel. *N Engl J Med* 2021;385:1393–400.
- Kato T. Opposition in Japan to the Olympics during the COVID-19 pandemic. *Humanit Soc Sci Commun* 2021;8:327.
- Casini L, Roccetti M. A Bayesian analysis of the inversion of the SARS-COV-2 case rate in the countries of the 2020 European football championship. *Future Internet* 2021;13:212.
- Jacobs J, Yang W, Cai S. Biden declares success in beating pandemic in July 4 speech, Bloomberg, Politics, July 4, 2021.
- Köppen W, Köppen W, Geiger R. Das geographische system der Klimat. In: (Hrsg): *Handbuch Der Klimatologie Bd. 1. Teil C. Bornträger*, Berlin, 1936.
- Takefuji Y. Fourier analysis using the number of COVID-19 daily deaths in the US. *Epidemiol Infect* 2021;149:e64.
- Miller M. Novel coronavirus COVID-19 (2019-nCoV) data repository. *Bulletin of Association of Canadian Map Libraries and Archives* 20192020;164.
- Cazelles B, Chavez M, Magny GCde, *et al*. Time-dependent spectral analysis of epidemiological time-series with wavelets. *J R Soc Interface* 2007;4:625–36.
- Zhang SX, Arroyo Marioli F, Gao R, *et al*. A second wave? What do people mean by COVID waves? - a working definition of epidemic waves. *Risk Manag Healthc Policy* 2021;14:3775–82.
- Hodcroft EB, Zuber M, Nadeau S, *et al*. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature* 2021;595:707–12.
- Hadfield J, Megill C, Bell SM, *et al*. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018;34:4121–3.
- Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* 2017;1:33–46.