Signal Processing Approach for COVID-19 Analysis

Regina Beytrison
Global College, Shanghai Jiao Tong University, China

Ong Dun Yan
Global College, Shanghai Jiao Tong University, China

***Abstract*—This paper presents a signal processing framework for analyzing complex time-series data. We propose methods for periodicity detection, transient event identification, cross-signal correlation analysis and lag analysis. Our approach combines time-frequency analysis with statistical signal processing techniques to extract meaningful patterns from noisy observational data. Experimental results on real-world datasets demonstrate the effectiveness of the proposed methods in identifying underlying structures and temporal relationships. The framework provides valuable insights for applications in epidemiology, environmental monitoring, and financial analysis.**

I. Introduction

TIME-SERIES analysis is fundamental to understanding complex systems and predicting their evolution. The COVID-19 pandemic has generated extensive time-series data that presents unique challenges and opportunities for signal processing methodologies. While time-series analysis has its own limitations such as strict assumptions on non-stationarity, it brings the undeniable advantage of being able to capture the temporal patterns and multiscale dependencies inherent to the data, as well as highlighting the strength and tendency of the successive waves.

Short-Time Fourier Transform (STFT) has proven effective for analyzing non-stationary signals. However, applying these methods to epidemiological data requires careful consideration of the specific characteristics of disease propagation dynamics.

II. Related Work

A. Time-Frequency Analysis in Epidemiology

Several studies have applied signal processing techniques to biomedical analysis and epidemiological data, such as Takefuji's research on Fourier analysis of the number of deaths due to COVID-19 in the US [1], or Bai et al. work on epidemic analysis using Adaptive Fourier Decomposition [2].
In our work, we will focus especially on the algorithm developed by Griffin and Lim [3] to estimate a signal from its Short-Time Fourier Transform.

B. Anomaly Detection in Time-Series

Anomaly detection in time-series data has been approached using various methods. Şahinler et al. [4] employed statistical techniques to compare its effectiveness to unsupervised learning methods, while Choi et al. [5] used different deep learning models. In this project, the Short-Time Fourier Transform (STFT) and Discrete Fourier Transform (DFT) will be used to support the analysis.

C. Cross-Correlation Analysis

Temporal correlation analysis is widely used to identify patterns, time delays, and the strength of linear relationships between two signals or datasets to understand if a change in one signal can influence the other signal.

D. Lag Analysis

In the case of COVID-19, the lag between the time a person was confirmed positive and the time of death can highlight interesting insights on the virulence of some variants, as well as the efficiency of the healthcare system.

III. Methodology

In this subsection, we will introduce the important theoretical steps used to analyze the COVID-19 dataset.

A. Data Preprocessing

First, missing values in the dataset, which can occur as some countries didn't report the cases during the same timeline, need to be cleaned : they can be put to 0 for simplicity. A 7-day rolling average can be used to smooth the data and hence highlight the periodicity components, as well as remove some noise. The formula for a 7-day moving average is as followed:

$$z[n] = \frac{1}{7}\sum_{k=0}^{6} y[n-k] \tag{1}$$

B. Periodicity Analysis using STFT

DFT and DTFT are powerful tools for time series analysis. However, they need to account for all samples ranging from $-\infty$ to ∞: it will then give an average, which can mask some specific behaviors. To solve this problem and better understand how the signal evolves with time, the Short-Time Fourier Transform can be employed to analyze time-varying periodic components:

$$X_w(mS,\omega) = F_t[x_w(mS,l)] = \sum_{n=-\infty}^{\infty} x_w(mS,l)e^{-j\omega l} \tag{2}$$

$$x_w(mS,l) = w(mS-l)x(l) \tag{3}$$

where $w(n)$ is the analysis window; $w(n)$ being assumed to be real and nonzero for $0 \leq n \leq L-1$, equation (1) can be rewritten as:

$$X_w(mS,\omega) = F_t[x_w(mS,l)] = \sum_{n=0}^{N-1} x_w(mS,l)e^{-j\omega l} \tag{4}$$

where X_w is the Fourier transform of x_w with respect to l.
The STFT $|Y_w(mS, w)|$ of the output signal $y(n)$ can be decomposed as $|Y_w(mS, \omega)| = |X_w(mS, \omega)||H_w(mS, \omega)|$ only if the analysis window $w(n)$ is long enough to include several impulse responses and short enough so that $x(n)$ is approximately stationary during the entire window length.
To improve the complexity of the algorithm, we will compute the FFT instead of the DFT for each window.
The spectrogram $S(m, k) = |X(m, k)|^2$ visualizes the change of a nonstationary signal's frequency content over time and can be used to analyze the periodicity of a signal.

C. Cross-Correlation Analysis

The normalized cross-correlation between two time-series $x[n]$ and $y[n]$ is computed as:

$$\rho_{xy}[l] = \frac{r_{xy}(l)}{\sqrt{r_{xx}(0) r_{yy}(0)}} \tag{5}$$

where

$$r_{xy}(l) = x(l) * y(l) \tag{6}$$

A preliminary step that needs to be taken before computing the cross-correlation is to ensure that $x[n]$ and $y[n]$ have the same length. In the case where this condition is not met, a zero-padding can be used.
Similarly to the STFT, the FFT can also be used for the convolution.

D. Lag analysis

The cross-correlation can be used to calculate the lag, as it will reach its peak when the time-shift corresponds to the lag between the two signals. This condition can be expressed as followed :

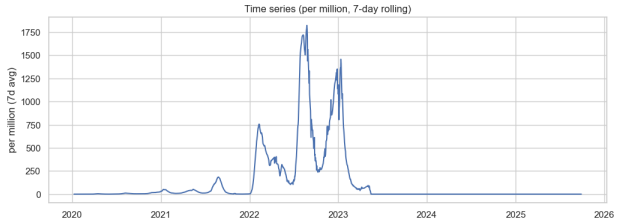
$$\tau = \left(\arg\max_l |r_{xy}(l)| \right) \times T_s \ [seconds] \tag{7}$$

where τ is the lag and $T_s = 24*60*60 = 86400$ (one case per day). This gives the dominant time shift between the signals over the entire period.

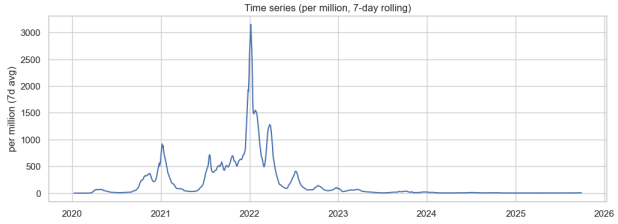
IV. Experimental Results

A. Dataset Description

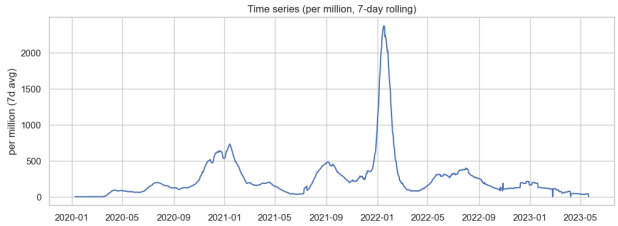
For the dataset is based on the daily new confirmed cases of COVID-19 per million people provided by Our World in Data [6], which is based on the data provided by the World Health Organization. In our project, we decided to take the data concerning the US, the UK and Japan and then compare them on different scales to analyze the different trends and behaviors. Below are the time-series plotting of the data to have a better visualization and understanding of it.



(a) Japan



(b) UK



(c) US

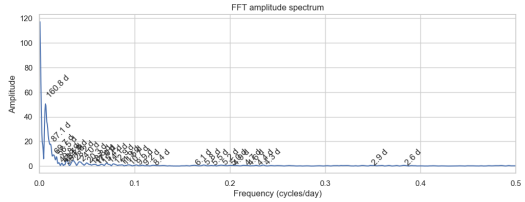
Fig. 1: COVID-19 timelines for different countries

B. Periodicity Analysis

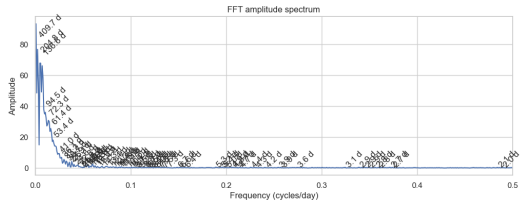
When using the STFT, careful considerations must be taken for the choice of the window used for the filtering, as well as the size of the window. For our analysis, we used a Hanning window. Compared to rectangular window, Hanning window has a lower resolution, but presents better properties for weak signals filtering due to its smaller side lobes. The chosen size of the window is 28 days, which let us have the average behavior for a month. Below are the different spectrograms for the different countries.

Name of window function	Transition width (Hz) (normalized)	Passband ripple (dB)	Main lobe relative to side lobe (dB)	Stopband attenuation (dB) (maximum)	Window function $w_d(k)$, $\lvert k \rvert \le (N-1)/2$
Rectangular	$0.9/N$	**0.7416**	**13**	**21**	**1**
Hanning	$3.1/N$	**0.0546**	**31**	**44**	$0.5 + 0.5\cos\left(\frac{2\pi k}{N}\right)$
Hamming	$3.3/N$	**0.0194**	**41**	**53**	$0.54 + 0.46\cos\left(\frac{2\pi k}{N}\right)$
Blackman	$5.5/N$	**0.0017**	**57**	**75**	$0.42 + 0.5\cos\left(\frac{2\pi k}{N-1}\right) + 0.08\cos\left(\frac{4\pi k}{N-1}\right)$
Kaiser	$2.93/N$ ($\beta = 4.54$)	**0.0274**		**50**	$\frac{I_0\left(\beta\sqrt{1-\left[2k/(N-1)\right]^2}\right)}{I_0(\beta)}$
	$4.32/N$ ($\beta = 6.76$)	**0.00275**		**70**	
	$5.71/N$ ($\beta = 8.96$)	**0.00275**		**90**	

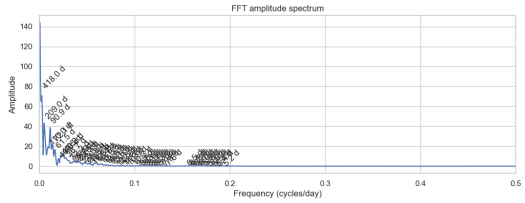
Fig. 2: Properties of different windows [7]



(a) Japan

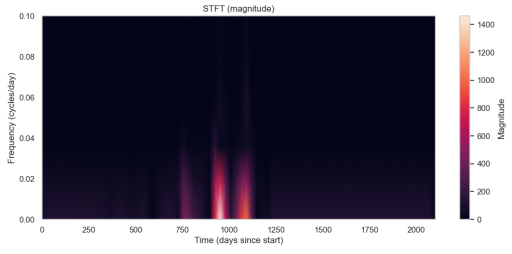


(b) US

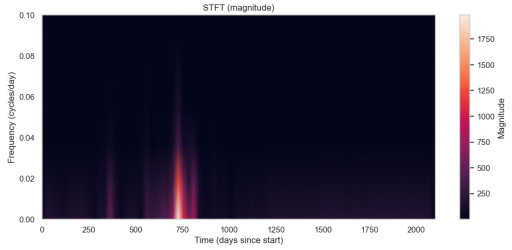


(c) UK

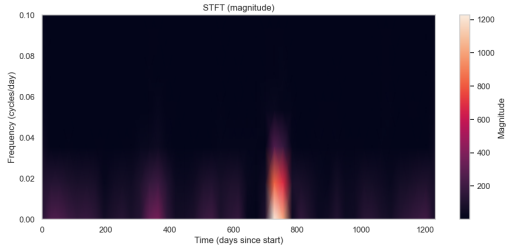
Fig. 3: Time-frequency analysis of COVID-19 cases



(a) Japan



(b) UK



(c) US

Fig. 4: STFT of the number of cases magnitude visualization

The higher the magnitude of the STFT, the stronger will the periodic behavior of the signal be. This periodic trend can be observed in all three countries, even if it doesn't occur at the same time. We can infer that during a certain interval of time, there is indeed a periodic component in the signal. The periodic component is observable around mid-2022 for Japan, and around early 2022 for the US and the UK.

For the case of Japan, it can be related to the COVID outbreak due to several mutations of the Omicron variant [8]. In the case of the US and the UK, these peaks appear approximately at the same time. One possible cause is the lift of the travel ban on 8th November 2021 between the US, the UK and the EU countries [9].

Another way to analyze the periodicity of a signal is to use the auto-correlation. When computing the auto-correlation for the three countries, we obtain these results:

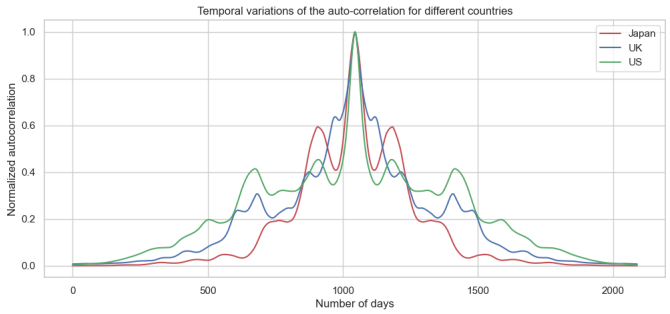


Fig. 5: Auto-correlation curves for each country

The auto-correlation analysis also supports the analysis of periodicity via STFT: there is a clear peak around the 1000 days in the case of Japan. For the US and the UK, there is a delay between the periodicity peak predicted by the auto-correlation and the one obtained with the STFT method.

Thus, the auto-correlation gives a relatively good approximation of the periodicity of a signal, however it cannot grasp the time evolution of it as well as the STFT method.

C. Cross-Country Correlation

For this part, the cross-correlation is applied two by two on the datasets of Japan, the UK and the US.

The Scipy's method $scipy.signal.correlate$ [10] can be used, which calculates the cross-correlation based on the convolution between the two signals. We obtain the following results:

TABLE I: Normalized cross-correlation values between the different datasets

Country 1	**Country 2**	**Max normalized cross-correlation**
Japan	UK	0.7870102887754808
Japan	US	0.7709086563002049
UK	US	0.8796828838329355

Analyzing these maximum values of the cross-correlations, it seems that there is a strong temporal relationship between

these countries. The analysis can be further strengthened by computing the cross-correlation curves.

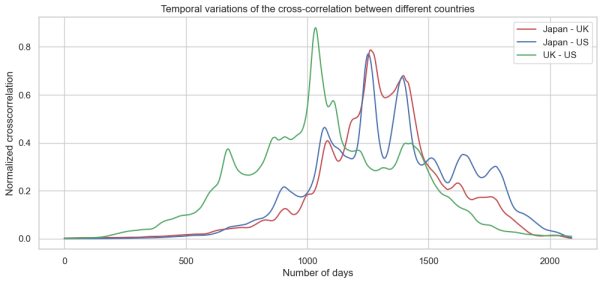


Fig. 6: Cross-correlation curves

The plotting of the cross-correlation curves confirms the presence of a period where the evolution of the number of cases in these different countries seems strongly correlated.

D. Case-Death Lag Analysis

There is a consistent lag between case confirmation and mortality. The delay in each country is presented in Table II.

TABLE II: Delay between cases and number of deaths

Country	Delay (in number of days)
Japan	28
UK	-343
US	21

In the case of Japan and the US, the predicted delay between the number of cases and the number of deaths corresponds to what was observed during the pandemic. However, the delay of -343 days for the UK implies that the case curve is delayed compared to the death curve. The number of cases curve and the number of deaths curve for UK can be plotted to better understand this negative delay:

Fig. 7: Number of cases curve (in red) and number of deaths curve (in blue)

The peak of the number of deaths curve is ahead of the number of cases curve. This shows that the number of deaths was higher during the early stage of the pandemic in the UK and, while the number of cases reached a peak later, the number of deaths didn't rise significantly.

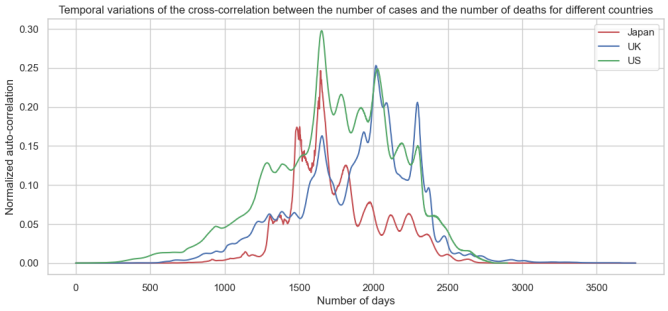


Fig. 8: Cross-correlation curves between the number of cases and the number of deaths

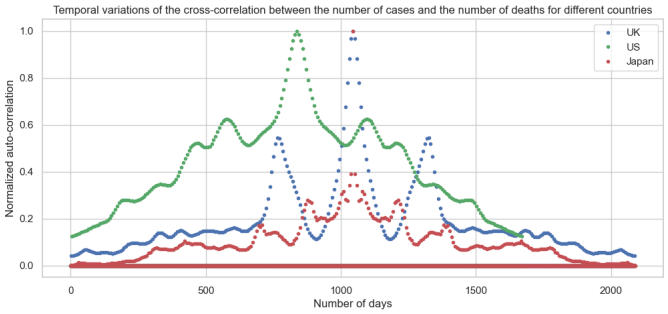
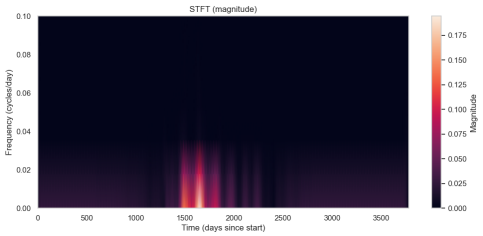
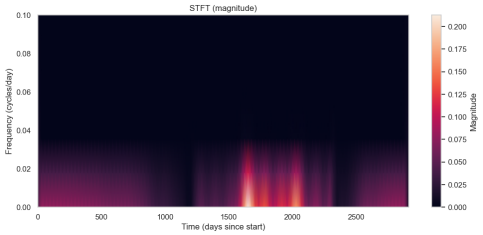


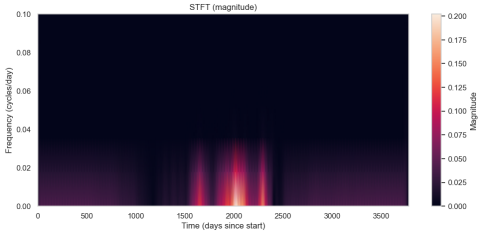
Fig. 9: Auto-correlation curves of the number of deaths



(a) Japan



(b) UK



(c) US

Fig. 10: STFT of the number of deaths magnitude visualization

In Fig. 9, the peak of cross-correlation between the number of cases and the number of deaths occur almost at the same time in the cases of Japan and the UK, while there is a shift in time with the case of the US. This observation can be confirmed with the visualization of the STFT (Fig. 10) applied to the cross-correlation: the peaks in the magnitude of the STFT do correspond to peaks in the cross-correlation curves.

V. Discussion

The periodic components identified via STFT and autocorrelation (Figs. 3-5) correspond to distinct pandemic waves. For instance, the $\sim$1000-day peak in Japan's auto-correlation aligns with the interval between major resurgences, such as the Delta wave in mid-2021 and the Omicron wave in early 2022.

The high cross-correlation values in case trends, notably between the UK and US (0.88), reveal strong temporal synchronization in pandemic waves. This synchronicity likely stems from shared structural and policy factors: both nations function as major international travel hubs, facilitating the simultaneous importation of new variants; they often aligned in the timing of restriction easements, leading to concurrent surges; and their surveillance systems exhibited similar reporting artifacts, such as weekend drops. It is crucial to note, however, that this correlation does not imply direct epidemiological causation but may instead reflect common external drivers, such as the global emergence of dominant variants like Omicron.

The case-to-death lag provides a critical window into underlying healthcare system and demographic factors. While the US and Japan display shorter delays (21 days and 28 days) which could be attributed to a rapidly vulnerable elderly demographic and potentially more efficient, timely death registration. Conversely, the negative and long delay for the UK may be attributed to a better handling of the pandemic after the initial wave, or may point to more complex systemic dynamics, which could reflect factors such as extended hospitalization durations, regional heterogeneity in reporting, or distinct national protocols, including potential periodic batch updates of death records and mortality occurring weeks after hospital discharge.

VI. Conclusion

This study demonstrates that signal processing techniques can transform raw epidemiological data into actionable insights. By interpreting periodicities as pandemic waves, correlations as global synchronization, and lags as healthcare system indicators, we bridge the gap between mathematical abstraction and public health practice. Future work should integrate multivariate signals (vaccination rates, mobility data) within this framework to build more comprehensive predictive models.

References

[1] T. Yoshiyasu, "Fourier analysis using the number of covid-19 daily deaths in the us," *Epidemiology and infection*, vol. 149, no. e64, 2021. DOI: 10.1017/S0950268821000522.

[2] X. Bai, X. Li, Y. Li, K. Guo, K. Gu, and C. Hon, "Adaptive fourier decomposition in epidemic analysis: Precision evaluation of vietnam's policies," *Contemporary Mathematics*, pp. 5015–5037, Aug. 2025. DOI: 10.37256/cm.6420257234.

[3] D. Griffin and J. Lim, "Signal estimation from modified short-time fourier transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 2, pp. 236–243, 1984. DOI: 10.1109/TASSP.1984.1164317.

[4] R. Şahinler, C. Öztürk, B. Karazeybek, and B. Kahraman, "Anomaly detection in time series data using unsupervised machine learning and statistical methods," in *2024 15th National Conference on Electrical and Electronics Engineering (ELECO)*, 2024, pp. 1–5. DOI: 10.1109/ELECO64362.2024.10847072.

[5] K. Choi, J. Yi, C. Park, and S. Yoon, "Deep learning for anomaly detection in time-series data: Review, analysis, and guidelines," *IEEE Access*, vol. 9, pp. 120 043–120 065, 2021. DOI: 10.1109/ACCESS.2021.3107975.

[6] E. Mathieu et al., "Coronavirus (covid-19) cases," *Our World in Data*, 2020, https://archive.ourworldindata.org/20251129-063338/covid-cases.html.

[7] B. Panomruttanarug and R. Longman, "Frequency based optimal design of fir zero-phase filters and compensators for robust repetitive control," vol. 123, Jan. 2006, pp. 219–238.

[8] T. Shiino et al., "Molecular epidemiology of sars-cov-2 genome sentinel surveillance in commercial covid-19 testing sites targeting asymptomatic individuals during japan's seventh epidemic wave," en, *Scientific Reports*, vol. 14, no. 1, p. 20 950, 2024, ISSN: 2045-2322. DOI: 10.1038/s41598-024-71953-8. [Online]. Available: https://www.nature.com/articles/s41598-024-71953-8.

[9] en-GB, Nov. 2021. [Online]. Available: https://www.bbc.com/news/uk-59197366.

[10] P. Virtanen et al., "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020. DOI: 10.1038/s41592-019-0686-2. [Online]. Available: https://doi.org/10.1038/s41592-019-0686-2.