Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Data-driven multiscale modelling and analysis of COVID-19 spatiotemporal evolution using explainable AI

Alvin Wei Ze Chew, Limao Zhang

A. Bentley Systems Research Office, 1 Harbourfront Pl, HarbourFront Tower One, 098633, Singapore
B. School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, 1037 Luoyu Road, Hongshan District, Wuhan, Hubei 430074, China.

ARTICLE INFO
Keywords:
COVID-19
Multiscale modelling
Data-driven prediction
Explainable AI
Deep learning

ABSTRACT
To quantificationally identify the optimal control measures for regulators to best minimize COVID-19’s growth (G-rate) and death (D-rate) rates in today’s context, this paper develops a top-down multiscale engineering approach which encompasses a series of systematic analyses, namely: (global scale) predictive modelling of G-rate and D-rate due to COVID-19 globally, followed by determining the most effective control factors which can best minimize both parameters over time via explainable Artificial Intelligence (AI) with SHAP (SHapley Additive exPlanations) method; (continental scale) same predictive forecasting of G-rate and D-rate in all continents, followed by performing explainable SHAP analysis to determine the most effective control factors for the respective continents; and (country scale) clustering the different countries (>150 in total) into 3 main clusters to identify the universal set of effective control measures. By using the historical period between 2 May 2020 and 1 Oct 2021, the average MAPE scores for forecasting G-rate and D-rate are within 10%, or less on average, at the global and continental scales. Systematically, we have quantificationally demonstrated that the top 3 most effective control measures for regulators to best minimize G-rate universally are COVID-CONTACT-TRACING, PUBLIC-GATHERING-RULES, and COVID-STRINGENCY-INDEX, while the control factors relating to D-rate depend on the modelling scenario.

1. Introduction

The current prevalence of the novel coronavirus disease 2019 (COVID-19) is considered as one of the greatest crises faced by humanity in the 21st century. On 11 Mar 2020, COVID-19 was officially declared as a global pandemic by the World Health Organisation (MBbs et al., 2020), and has continued to take its toll on the global community across all major continents due to the continual emergence of new COVID-19 variants such as Delta - B.1.617.2 and Omicron - B.1.1.529 (Chen et al., 2021a; Rao & Singh, 2021). At present, the number of confirmed COVID-19 cases, inclusive of the variants-related ones, has already crossed 349 million worldwide, while the total number of deaths due to COVID-19 exceeds 5 million globally. Since the inception of COVID-19 in early 2020, regulators around the world have imposed different socioeconomic and restrictive measures to control the virus spread, while attempting to mitigate the adverse impact(s) on their respective local economies to the best possible extent. Generally, many countries in the different continents (Asia, Europe, North America, etc.) have shifted from broad interventions such as widespread lockdowns and closures, to risk-based and targeted strategies (e.g., contact tracing, public gathering rules, etc.) which may best tailor to the socioeconomic characteristics of a specific country (Ministry of Health, 2020b; Pung et al., 2021; Services & Department, 2021), hence resulting in varying government stringency indexes (Ma et al., 2021) being implemented globally. At the same time, COVID-19 vaccinations have now been incorporated into humanity’s defense mechanism despite the continual debate over the effectiveness of the different vaccines and their long-term impacts on human health (WHO, 2020; Pilishvili et al., 2021; Thompson et al., 2021).

Due to the highly infectious nature of COVID-19, regulators in city states faced the difficult challenge of controlling the virus spread, especially in close living environments which include, but not limited to, dormitories, military camps, and prison facilities (Lee et al., 2020). For example, in Singapore, a highly dense and populated urban city, the nation was dealt with large surge of COVID-19 cases across multiple foreign worker dormitories which house a quarter of 1.4 million foreign workers in Singapore (Ministry of Health, 2020c; Ministry of Manpower, 2020), that subsequently drove the local regulators to impose a nationwide lockdown (termed as circuit breaker) between...
April and June 2020. In larger countries having lower population densities such as in the United States and Europe, COVID-19 infectious rate continued to remain high due to the reluctance of portions of their populations to adhere to the recommended control measures such as wearing masks and receiving adequate vaccinations. For example, surveys have quantifiably demonstrated that an approximate 16% of the general sampled population in the United States and Canada are reluctant to wear masks due to their beliefs that masks are ineffective to protect against the virus, and their psychological reactance (Taylor & Asmundson, 2021).

The above-described contextual examples generally highlight that regulators, i.e., governments, across the different nations and continents, having disparate socioeconomic characteristics and cultural behaviours, are confounded in identifying the best optimal control measures which may best control the growth rate in the confirmed number of COVID-19 cases and their associated death rates, as justified by the continual rise in the present reported numbers due to the multiple virus variants spreading globally. On the other hand, the impacts of environmental (climate-related) conditions on the transmission rate of COVID-19 virus continued to attract attention, which further compounds regulators’ difficulties to design optimal control policies that can best control the spread of COVID-19. In March 2021, the United Nations highlighted the possibility of COVID-19 becoming a seasonal disease, i.e., endemic, by pointing out that respiratory-related viral infections are often seasonal, especially during autumn-winter peak period for influenza, and temperate conditions for cold-related coronaviruses, hence underscoring the challenge faced by regulators to control the virus spread in the foreseeable future.

In the present literature, there have already been research studies performed to evaluate the level of correlations between the transmission rates of COVID-19 globally and the complex environmental conditions which include ambient temperature, wind speed, and humidity (Dzialiel et al., 2018; Maiti et al., 2021; Viezzer & Biondi, 2021). As a relational study to COVID-19, Yuan et al. (2006) had earlier reported that the 2003 SARS (coronavirus related) outbreak in Beijing, China peaked at the following conditions: (a) mean temperature of 16.9°C; (b) mean relative humidity of 52.2%; (c) wind speed of 2.8 m/s. For COVID-19 itself, Prata et al. (2020) highlighted that the daily number of confirmed COVID-19 cases in Brazil generally has an inverse relationship with temperatures ranging between 16.8 °C and 27.4 °C from Feb to Mar 2020. Şahin (2020) observed a correlation between the spread of COVID-19 in several cities of Turkey with wind speed values recorded 14 days ago and the temperature of the actual day for the total number of confirmed cases reported. Chew et al., 2021 reported that land surface temperature day has the strongest negative correlation with the growth rate in the number of confirmed COVID-19 cases globally. Other studies for specific countries, such as China, Spain, and the United States (Abdollahi & Rahbaralam, 2020; Liu et al., 2020; Runkle et al., 2020; Shi et al., 2020), have also reported that temperature does affect, to an extent, the transmission rate of COVID-19.

At the same time, there have been recent studies which reported positive or minimum correlations (Ahmadi et al., 2020; Bashir et al., 2020; Toosepu et al., 2020; Xie & Zhu, 2020; Yao et al., 2020) between the transmission rate of COVID-19 and climate conditions, especially for cases where data is generally limited. Analysis for large-scale countries based upon multiple climate zones have also been restricted and inconsistent in several reported studies (Awasthi et al., 2020; Baker et al., 2020; Chiyevsky & Takeuchi, 2020; Islam et al., 2020; Pan et al., 2021a; Sobral et al., 2020). Similar to the global study carried out by Chew et al., 2021, Wu et al. (2020) and Guo et al. (2021) examined the impacts of ambient temperature on the transmission rates of COVID-19 worldwide by too reporting negative correlations between ambient temperature and the transmission rate(s) of COVID-19.

Overall, the above-discussed studies underline the complex relationship between COVID-19 transmission and the dynamic environmental conditions which continues to be an open research topic to be explored in the present pandemic situation. The complexity, however, further underlines the difficulty faced by regulators to implement the most appropriate control measures to best control the spread of COVID-19 due to humanity’s inability to regulate environmental conditions optimally. Therefore, the question remains on how the different communities, at the multiple spatial scales (global, continent, country levels), can best adapt to living with COVID-19 as an endemic by implementing the most suitable control policies (e.g., contact tracing, vaccination rates, face mask policies, etc.) for the key objectives of: (1) stemming the transmission and death rates due to COVID-19 virus and its variants over time for a given spatial context (e.g., global scale, continents, or specific countries) having their inherent environmental (climate-related) conditions and social characteristics; and (2) protecting the stability of the global economy to the best possible extent in the foreseeable future when COVID-19 is expected to become endemic.

To the very best of our knowledge, we note that the present literature lacks a comprehensive model framework which can systematically investigate the impacts of the multitude of control policies on the growth and death rates of COVID-19 from the global to individual country scales, followed by recommending the most critical/effective control measures which can best minimize the growth and death rates due to COVID-19. By far, it is worth highlighting that most research studies are restricted to local countries (Chen et al., 2021b; Zhou et al., 2021; Zhu & Tan, 2021; Pan et al., 2022), or individual continents (Pan et al., 2021b; Sannigrahi et al., 2020) for either investigating the relationship between specific control factors (work resumption policies, home quarantines) and COVID-19, or to determine the optimal policies to control the spread of COVID-19, respectively. For works pertaining to the global scale analysis (Chew et al., 2021; Chew et al., 2021a; Guo et al., 2021; Wu et al., 2020), they are mostly related to investigating the most important environmental factor(s) which have the strongest correlations with the transmission rates of COVID-19 globally. As described in-detail previously, climate-related environmental conditions do contribute complexity to COVID-19’s spatiotemporal dynamics which, however, cannot be optimally regulated by humanity due to (1) lack of complete quantitative understanding of their complex relationship, and (2) inability of humanity to optimize environmental conditions that can best control the spread of COVID-19 and its associated death rates. From a practical aspect, it is instead more realistic and effective for regulators to optimize the selection of control factors that can best minimize the associated growth and death rates of COVID-19, in the near real-time context, for any given spatial context (global, continent and country levels) having inherent socioenvironmental characteristics which cannot be optimally controlled by regulators in general, as highlighted above.

To contribute to the above-outlined objective, this study develops a top-down multiscale engineering approach, coupled with explainable artificial intelligence (AI), which can comprehensively perform the following systematic analyses: (global scale) predictive modelling of the global growth (G-rate) and death (D-rate) rates due to COVID-19, followed by determining the most critical/effective control factors, via explainable AI, that can best minimize both the G-rate and D-rate target parameters due to COVID-19 on the global scale; (continental scale) predictive modelling of the same parameters for each of the 5 key continents in Asia, Africa, Europe, North America, and South America to quantitatively evaluate any level of similarity with that of the global scale, followed by determining the most critical/effective control factors for each respective continent via explainable AI again; and (country scale) perform time-series clustering of individual G-rate and D-rate parameters pertaining to all available countries globally (> 150 in total) to confirm on universal set(s) of control factors to achieve the same optimisation objectives for both G-rate and D-rate parameters in the different clusters of countries formed. We note that the global scale analysis considers a total of 20 data features, which are commonly known control factors adopted by regulators globally. The selected features encompass socioeconomic factors, government restrictive
policies, and the global community’s aggregated sentiments towards COVID-19, while the continental and country scales consider the same features by only excluding the sentiments component due to the lack of data information at the localised level for multiple countries. In summary, the key research contributions in this paper are as follows:

- Quantificationally identify the optimal modelling scenarios in terms of the number of multiple-time steps, as quantified in days, for assimilating the total amount of historical records for the selected data features to accurately model and forecast the averaged, via smoothing analysis, G-rate, and D-rate time-series profiles at the global and continental scales with a lead-time of 1 day.
- Systematically and quantificationally identify the optimal set of control measures, i.e., data features, at the multiple scales of global, continental, and country level which can best minimize their G-rate and D-rate time-series profiles via explainable AI, that can serve as useful information to regulators to control the spread of COVID-19, and its associated death rates, in the foreseeable future where the virus is expected to become an endemic.

This paper is structured as follows. Section 2 reviews the related studies which quantificationally models the complex relationship between COVID-19 and the different socioeconomic and environmental factors for predictive analyses/modelling of the G-rate and D-rate time-series profiles. Section 3 describes the key details relating to the adopted methodology in this study which encompasses the proposed top-down multiscale engineering approach, data features and management, development of deep learning models for predictive modelling and the associated explainable AI method. Section 4 summarizes the key results obtained from the global, continental, and country scales for the objective(s) of determining the most effective control factors for the respective scales, and the predictive modelling results attained from the global and continental scales analyses. Finally, Section 5 summarizes the key findings obtained from our proposed analysis using the proposed top-down multiscale engineering approach to address the identified problem statement in this study.

2. Related works

In the following, we provide a literature overview on the critically important and relevant studies carried out by far which investigate the quantitative relationship between COVID-19 and the key socioeconomic, restrictive, and environmental factors that have been useful to contribute to the community’s scientific understanding of the current pandemic situation. For better readability, our readers are referred to Table 1 (see end of Section) that summarizes all selected reference studies in relations to the general scope of our present study.

For modelling the impacts of environmental conditions on the dynamics of COVID-19, Wu et al. (2020) developed a novel generalised additive model (GAM) mathematical representation which can forecast the daily global numbers of confirmed COVID-19 cases and the associated death cases with a lead-time of 1 day, as a function of wind speed and other weather variables. Guo et al. (2021) leveraged on multiple meteorological parameters, which are first weighted via population density, for the defined modelling objective at the city- or country-level. Chew et al., 2021 exploited the use of climate-related satellite images to forecast the growth rate in the total number of confirmed COVID-19 cases globally via semantic segmentation analysis. A similar study has recently performed by Zhou et al. (2021) where the authors leveraged on time-series earth observation data for land surface temperature (LST) to quantificationally explore the impacts of the spatiotemporal variations of multiple COVID-19 control measures on industrial production in Wuhan, China. Kim (2021) developed a unique regression analysis approach which further underlines that COVID-19 is likely to be affected by climate change, while also highlighting on the importance to model the complex spatial relationship between weather conditions and the spread of COVID-19 in their future works. Kroumpouzos et al. (2020) reported that low ambient temperatures in China, Japan, South Korea, Northern Italy, and Germany coincided with the inception of COVID-19 towards the end of 2019. The favourable impact of cold temperatures on the survival of CoV (coronavirus strain) has previously been demonstrated in lab experiments, where Casanova et al. (2010) reported that the air temperature range of 5 °C–11 °C, and 47–70% for relative humidity can be the optimal conditions to sustain CoV over time. Despite this finding, countries that generally experience higher temperature across each year which include, but not limited to, Singapore, Thailand, India, and the Middle East, have since reported significant numbers of confirmed COVID-19 cases and the associated death rates due to close

| Refs. | Key notes |
|-------|-----------|
| Casanova et al. (2010) | Reported that air temperature range of 5 °C–11 °C, and 47–70% for relative humidity can be the optimal conditions to sustain CoV over time |
| Huang et al. (2020) | Reported on the need to find an optimal balance between uncontrollable environmental conditions and general control measures to manage COVID-19 |
| Kroumpouzos et al. (2020) | Reported that low ambient temperatures in China, Japan, South Korea, Northern Italy, and Germany coincided with the inception of COVID-19 |
| Li et al. (2020) | Reported on the need to find an optimal balance between uncontrollable environmental conditions and general control measures to manage COVID-19 |
| Saminrahi et al. (2020) | Reported that total population size and income level are most important to regulate the number of death cases due to COVID-19 in the entire European region, and poverty and income can best control COVID-19 spread and death rate in the European region |
| Sun and Zhai (2020) | Reported that social distancing is most effective to minimize the infectious rates of COVID-19 |
| Wu et al. (2020) | Reported that ambient temperature has negative correlation with transmission rates of COVID-19 |
| Chen et al. (2021) | Reported that healthcare infrastructure adequacy and urban governance capacity are the optimal factors which can maximize a nation’s resilience towards the current pandemic situation |
| Chew et al., 2021 | Reported that land surface temperature day (LSTD) has strongest negative correlation with transmission rates of COVID-19 |
| Das et al. (2021) | Correlated living environment deprivation factor with the spatial clustering of COVID-19 hotspots in Kolkata |
| Fu and Zhai (2021) | Reported on the complex relationship(s) between the socioeconomic-environmental nexus and the growth rate of COVID-19 in major urbanised cities |
| Kim (2021) | Reported the influences of climate change on COVID-19’s dynamics |
| Guo et al. (2021) | Reported that ambient temperature has negative correlation with transmission rates of COVID-19 |
| Hu et al. (2021) | Reported on the complex relationship(s) between the socioeconomic-environmental nexus and the growth rate of COVID-19 in major urbanised cities |
| Li et al. (2021) | Reported statistically significant influences of commercial strength and the availability of transportation infrastructures on the number of confirmed cases in any infectious cluster |
| Mati et al. (2021) | Reported on the complex relationship(s) between the socioeconomic-environmental nexus and the growth rate of COVID-19 in major urbanised cities |
| Ugul et al. (2021) | Reported that social distancing is most effective to minimize the infectious rates of COVID-19 |
| Yue et al. (2021) | Reported that government stringency index is the most important social factor to control the transmission rates of COVID-19 in Nepal, South Korea, Japan, and Pakistan. |
| Zhu and Tan (2021) | Reported that home quarantines are useful to control COVID-19 spread, while balancing public protection, individual freedom, and general resources |
| Zhou et al. (2021) | Investigated the general impacts of spatiotemporal variations of multiple COVID-19 control measures on industrial production in Wuhan, China |
human contact in dense urban cities, thus underlining the importance to find an optimal balance between the uncontrollable environmental conditions and the relevant control factors which can encompass population densities, hygiene, space availability, (Huang et al., 2020; Li et al., 2020), socio-economic conditions and governmental restrictive policies.

For the socio-economic considerations, Li et al. (2021) proposed a structural quantitative representation to quantify the relationship(s) between built environment attributes and the cluster size of COVID-19 in Hangzhou, China. Their results reported statistically significant influences of commercial strength and the availability of transportation infrastructures on the number of confirmed cases in any infectious cluster. Similar clustering analysis has also been performed by Das et al. (2021) in the context of Kolkata megacity, India by correlating the level of living environment deprivation on the spatial clustering of COVID-19 hotspots in Kolkata. Multiple data regression models have also been exploited for the same modelling objective, where zero-inflated negative binomial regression (ZINBR) has been determined to effectively correlate the living environment deprivation factor with the spatial clustering of COVID-19 hotspots in Kolkata. Samirgrahi et al. (2020) adopted multiple 1D regression models to report that total population size and income level are most important to regulate the number of death cases due to COVID-19 in the entire European region. The authors also underlined on the importance to quantitatively model the influences of environmental conditions on the growth and death rates of COVID-19 for predictive analysis. Other relevant studies performed by Fu and Zhai (2021), Hu et al. (2021) and Maiti et al. (2021) have further underscored the complex relationship(s) between the socioeconomic-environmental nexus and the growth rate of COVID-19 in different cities that include Washington, New York City, and North America as a whole, which thus constitute to another continental scale analysis as discussed previously. At the same time, the authors also reported the significance of income level, social distancing, and stay-at-home practices to control the growth and death rates of COVID-19 globally. In the context of Asia, Yue et al. (2021) developed a spatiotemporal analysis framework by combining ensemble model (random forest regression) and multi-objective optimisation algorithm to confirm that the government stringency index is the most important social factor to control the transmission rates of COVID-19 in Nepal, South Korea, Japan, and Pakistan. Chen et al. (2021a) exploited city-level data from China using multiple linear regression models to report that healthcare infrastructure adequacy and urban governance capacity are the optimal factors which can maximize the nation’s resilience towards the current pandemic situation. For the case of Hong Kong, Zhu and Tan (2021) combined epidemiological data with the relevant socioeconomic and meteorological data from over 250 cities to examine the effectiveness of home quarantines where the authors reported that the control measure can be as useful as centralised quarantine, while balancing public protection, individual freedom, and general resources. In the European continent, Samirgrahi et al. (2020) extensively investigated the global and local spatial association between key socio-demographic variables and the growth and death rates due to COVID-19 in 31 different European countries by documenting that poverty and income can best minimize both COVID-19 related conditions in the European region. Finally, Sun and Zhai (2020) and Ugail et al. (2021) developed novel mathematical approaches in their respective works to confirm the universal usefulness of social distancing to minimize the infectious rates of COVID-19.

Overall, the above-outlined research studies have been very useful to provide an extensive overview of the impacts of the different combinations of socio-economic and/or environmental related factors to control the growth and death rates of COVID-19 under multiple spatiotemporal scales. However, to set the basis for our present study, 2 key shortcomings can be identified at this stage, namely:

- Most of the previous studies are mainly confined to specific countries and/or continents without performing any correlation analysis to other countries globally or within the same continents for the defined modelling objective. As such, there lacks a comprehensive framework in the literature which can effectively correlate the results derived from the different countries and/or continents. For example, how we can spatially correlate the findings from one specific continent to countries located in another continent for determining the most effective control factors which can minimize the temporal growth and death rates due to COVID-19.

- Despite the existence of multiple completed works which focus on reporting the optimal control factors for minimising the growth and/or death rates due to COVID-19, the scientific community is still deprived of a universal set of control factors, in an explainable and quantitative manner, which can effectively fulfill the optimisation objective, i.e., minimising the growth and/or death rates due to COVID-19 to the best possible extent, for countries which may belong to the different continents and may comprise of different level of socio-economic living conditions in the current pandemic situation.

To definitively address the above-outlined shortcomings, this study develops a comprehensive top-down multiscale engineering approach which can first model and forecast the growth (G-rate) and death (D-rate) rates target parameters due to COVID-19 on the global and continental (Asia, Africa, Europe, North America, and South America) scales under 15 different modelling scenarios which encompass different combinations of moving average computations and multi-time steps, as measured in days, for data pre-processing and assimilation. The optimal modelling scenario(s) identified for the global and continental scales, respectively, which can achieve the most accurate prediction results for the G-rate and D-rate, are then further exploited for explainable AI (deep learning) analysis to identify the most effective control factors which can best serve the optimisation objective of minimising both COVID-19 related parameters. Upon doing so, country-scale analysis is performed by clustering more than 150 countries globally into multiple individual clusters, based upon their individual growth and death rates due to COVID-19, for the purpose of identifying a universal set of effective control factors for the different clusters of countries, as initially clustered from the diverse continents.

3. Methodology

3.1. Generic multiscale analysis approach

Fig. 1 illustrates the proposed generic top-down multiscale engineering approach to model the spatiotemporal variations of COVID-19 evolution. The approach comprises of 3 key scales, which are outlined as follows:

i) Global scale: By agglomerating the relevant data (20 unique features) for all countries (> 150 in quantity) globally, perform predictive modelling of 2 key parameters of: (1) growth rate in the total number of COVID-19 cases, G-rate; and (2) the death rate, D-rate, due to COVID-19, via deep learning analysis. The predictive modelling step is then proceeded with features importance analysis using SHAP (SHapley Additive exPlanations) method to determine the most effective features, i.e., control factors, which can best minimize G-rate and D-rate parameters at the global scale via quantitative explanations.

ii) Continental scale: Leveraging on the respective data (17 unique features) for 5 key continents in Asia, Africa, Europe, North America, and South America, perform predictive modelling of the same G-rate and D-rate target parameters for each continent via deep learning analysis. Likewise, SHAP analysis is again performed for every continent to determine their most effective features, i.e., control...
Fig. 1. Generic top-down multiscale engineering approach to model spatiotemporal variations of COVID-19 evolution.

Fig. 2. Time-series profiles for G-rate for (i) global and (ii–vi) continental scales for period between 2 May 2020 and 1 Oct 2021 under varying MA computation scenarios: (a) 1-day MA; (b) 3-days MA; (c) 5-days MA; and (d) 7-days MA.
factors, which can best minimize their individual G-rate and D-rate parameters at the continental scale via quantitative explanations.

iii) Country scale: Perform time-series clustering of all available countries (> 150 in quantity) using their individual G-rate and D-rate profiles under multiple moving average computation scenarios. Under each modelled scenario, perform SHAP analysis for every available cluster formed to determine the most common effective features, i.e., control factors, which can best minimize the corresponding G-rate and D-rate parameters pertaining to each cluster of countries via quantitative explanations.

3.2. Data description

For the respective scales, as illustrated in Fig. 1, their G-rate and D-rate time-series profiles are computed by leveraging on global recorded numbers for the confirmed COVID-19 cases and related death cases from an open-source database (https://ourworldindata.org/coronavirus-data). In our analysis, the temporal G-rate and D-rate profiles are computed for the period ranging between 2 May 2020 and 1 Oct 2021, with respect to 1 May 2020 as the time reference, as via the use of Eq. (1).

\[
Y(t) = \frac{X(t) - X(t-1)}{X_0} \times 100\% \tag{1}
\]

where \( Y \) represents either G-rate or D-rate target parameters, \( X \) is the recorded numbers of the confirmed numbers of COVID-19 cases or death cases, \( t \) represents the current state of time (e.g., current day), \( t - 1 \) represents the previous state of time (e.g., previous day), and \( X_0 \) is the reference data (number of cases or death cases) on 1 May 2020.

At the same time, we also consider the moving average (MA) computations of 1-day, 3-days, 5-days, and 7-days, for the computed \( Y(t) \) time-series profiles. Note that the 1-day MA computations essentially represent the original \( Y(t) \) time-series profile for either G-rate or D-rate target parameters. Figs. 2 and 3 illustrate the derived time-series profiles for G-rate and D-rate, respectively, under the different MA scenarios for the global and continental scales, as explained earlier, for the time period ranging between 2 May 2020 and 1 Oct 2021 which constitutes to a total of 518 days.

The respective G-rate and D-rate time-series profiles, under varying MA scenarios, at the global and continental scales, are modelled as a function of available features (total of 17 features) which represent the control factors adopted by countries globally. The historical data for those features, as summarised in Table 2, are extracted, and collated from the same open-source dataset in (https://ourworldindata.org/coronavirus-data). As outlined earlier, the global scale analysis includes an additional 3 data features which represent the positive, negative, and neutral sentiments, in time-series percentages format in Fig. 4, of the global community towards COVID-19. Note that the time-series data for the sentiments components are extracted from another open-source dataset in (https://github.com/lopezbec/COVID19_Tweets_Dataset).

![Fig. 3. Time-series profiles for D-rate for (i) global and (ii–vi) continental scales for period between 2 May 2020 and 1 Oct 2021 under varying MA computation scenarios: (a) 1-day MA; (b) 3-days MA; (c) 5-days MA; and (d) 7-days MA.](image-url)
Table 2

| Data Feature | Description | Type | Range of values | Class 1 | Class 2 | Class 3 | Class 4 | Class 5 | Class 6 | Scale involved |
|--------------|-------------|------|-----------------|---------|---------|---------|---------|---------|---------|----------------|
| X-1          | DEBT-RELIEF | Time-series classification (e.g., Class 1, Class 2, etc.) and X-15 X-16 | Classes 1-3 | No relief | Narrow relief | Broad relief | -nil- | -nil- | -nil- | Global and Continental Scales analysis |
| X-2          | INCOME-SUPPORT | Class 1-3 | No support | Support < 50% salary loss | Support > 50% salary loss | Available to 1 group* | Available to all groups** | -nil- | -nil- | -nil- |
| X-3          | COVID-VACCINATION-POLICY | Classes 1-6 | No vaccinations | Available to 1 group* | Available to all groups** | Universal availability | -nil- | -nil- | -nil- |
| X-4          | COVID-CONTACT-TRACING-POLICY | Classes 1-3 | No tracing | Limited tracing | Comprehensive tracing | -nil- | -nil- | -nil- |
| X-5          | COVID-TESTING-POLICY | Classes 1-4 | No testing | Test symptoms & key groups | Test anyone with symptoms | Open public testing | -nil- | -nil- | -nil- |
| X-6          | INTERNATIONAL-TRAVEL | Classes 1-5 | No measures | Screening | Quarantine from high-risk regions | Ban from high-risk regions | -nil- | -nil- | -nil- |
| X-7          | INTERNAL-MOVEMENT | Classes 1-3 | No measures | Recommend movement restrictions | Impose closing (or strong prohibition) | -nil- | -nil- | -nil- |
| X-8          | PUBLIC-TRANSPORT | Classes 1-3 | No measures | Recommend closing (or reduce volume) | Public officials urging caution | Required in some public spaces | -nil- | -nil- | -nil- |
| X-9          | PUBLIC-CAMPAIGNS | Classes 1-3 | No information sharing | Public officials warning | Coordinated information campaign | Required for some | -nil- | -nil- | -nil- |
| X-10         | FACE-COVERING-POLICIES | Classes 1-5 | No policy | Recommended | Required in all public spaces | Required for key workers | -nil- | -nil- | -nil- |
| X-11         | STAY-AT-HOME | Classes 1-4 | No measures | Recommended | Required (except essentials) | Required for all except key workers | -nil- | -nil- | -nil- |
| X-12         | PUBLIC-GATHERING-RULES | Classes 1-5 | No restrictions | > 1000 people | 100-1000 people | Required at all levels | -nil- | -nil- | -nil- |
| X-13         | PUBLIC-EVENTS | Classes 1-3 | No measures | Recommended cancellations | Required cancellations | Required for some | -nil- | -nil- | -nil- |
| X-14         | WORKPLACE-CLOSURES | Classes 1-4 | No measures | Recommended | Required at some levels | Required at all levels | -nil- | -nil- | -nil- |
| X-15         | COVID-CONTAINMENT-AND-HEALTH-INDEX | 0–100 | | | | | | | | |
| X-16         | COVID-STRINGENCY-INDEX | 0–100 | | | | | | | | |
| X-17         | SCHOOL-CLOSURES | Time-series percentages 0.0–1.0 | No measures | Recommended | Required at some levels | Required at all levels | -nil- | -nil- | -nil- | Only for Global Scale analysis |
| X-18         | POSITIVE SENTIMENTS (%) | | | | | | | | | |
| X-19         | NEGATIVE SENTIMENTS (%) | | | | | | | | | |
| X-20         | NEUTRAL SENTIMENTS (%) | | | | | | | | | |

\[ X'_n = \sum_{k=1}^{m} A_k x_{n,k} \]  

(2)

where \( X'_n \) represents the summated value for data feature \( y \) for each continent, \( A_k \) is the population density for country \( n \), \( m \) the total number of countries within each continent, and \( x_{n,k} \) the value of the data feature \( y \) for country \( n \).
3.3. Predictive deep learning models

For analyses at both the global and continental scales (see Fig. 1), predictive modelling of the respective D-rate and G-rate, as represented by \( Y(t) \), time-series profiles under the multiple MA computation scenarios are performed by leveraging on the processed data features, as

**Fig. 4.** Time-series representations of the percentages of positive, negative, and neutral sentiments globally for period between 1 May 2020 and 1 Oct 2021.

\[
Y(t) = f(X_1(t-2-M), X_1(t-1-M), ..., X_1(t-2), ..., X_N(t-2-M), ..., X_N(t-2), Y(t-2-M), ..., Y(t-2))
\]

**Fig. 5.** Modelling and forecasting \( Y(t) \) of either G-rate or D-rate in its current state of time: (a) example of data assimilation using \( M \) days of MS; (b) design of DNN model for predictive modelling.

A.W.Z. Chew and L. Zhang
For each respective data feature and the target parameter (G-rate or D-rate) itself, while maintaining a fixed lead-time of 1 day to forecast the same target value in its current state of time, as generically represented in Eq. (3). To illustrate, Fig. 5a exemplifies the use of M days of MS, i.e., multi-time steps, for leveraging on the historical records for the required data features and target parameter, while maintaining a fixed lead-time of 1 day, to model and forecast Y(t) in its current state of time.

\[ Y(t) = f(X_1(t-2-M), X_1(t-1-M), \ldots, X_1(t-2), Y(t-2)), Y(t-1-M), \ldots, Y(t-2) \]  

(3)

Multiple MS scenarios are adopted to perform the predictive modelling step, where we consider the MS days of 1 day, 3 days, 5 days, 7 days, 9 days, and 11 days. Each MS scenario is then coupled with the above MA (moving scenario) computation scenario of 3 days, 5 days, 7 days, and 7 days, hence generating 15 combinations of modelling scenarios to model and forecast Y(t), of either G-rate or D-rate, with lead-time of 1 day. Training and validation of the predictive model for each unique combination of MS and MA parameters is performed under varying batch sizes of 4, 8, and 16 with a fixed number of epochs of 500, as summarised in Table 3 for the design of the DNN (see Fig. 5b) and selected hyperparameters for the computational experiments in this study. Note that the number of neurons in the input layer of the DNN model is represented by the multiplication value between the total number of data features plus the unique target parameter (G-rate or D-rate) and the number of MS days. For example, if the total number of processed features (Eq. (2)) is 20 to model and forecast G-rate with MS of 3 days, then the total number of neurons in the input layer equates to 63 (21 × 3). Finally, we note that 80% of the total available data instances (in daily time-series format ranging between 2 May 2020 and 1 Oct 2021) in each modelling scenario is leveraged for the model training and validation steps, while the remaining 20% is used to test the trained predictive model for modelling and forecasting the target parameter in G-rate and D-rate. No random shuffling of the time-series data is performed prior to the model training, validation, and testing steps.

3.4. Explainable AI via SHAP for feature importance analysis

In our predictive modelling analysis in this study, it is assumed that each processed data feature, as derived from the use of Eq. (2), contributes to the prediction of the specific target value in either G-rate or D-rate in each of the 15 modelling scenarios, as explained previously. The quantitative contribution from each data feature is measured via explainable AI (XAI), which generally refer to quantitative processes and methods that enable modellers, and other stakeholders to, quantitatively understand and trust the predictions generated from machine learning and/or AI modelling algorithms. In short, XAI is useful to describe the selected algorithm for the modelled problem, and its subsequent impact in addressing the issue to achieve expected benefits.

To quantify the contributions from each processed data feature via XAI in our analysis, we compute the SHAP (SHapley Additive exPlanations) values (Lundberg & Lee, 2017) to explain individual predictions by the processed data features. Readers are referred to the added reference for the exact mathematical details involved. In short, the adopted SHAP explanation method computes Shapley values from coalescent game theory, where the raw feature values pertaining to every data instance serve as the defined "players" in the coalition. For each raw feature value, the respective Shapley value can be generally computed by following the computations in pseudo-Algorithm 1.

Note that each player can be an individual processed feature value or can also be a group/cluster of processed feature values. Our analysis thus focuses on computing the importance of each processed data feature to the respective prediction values, followed by determining the top few most critical features, via the computed Shapley values, which can best minimize the modelled G-rate and D-rate parameters over time. For this purpose, the most critical processed features, i.e., control factors, for the associated modelling scenarios (different combinations of MS and S parameters) are determined by finding the associated data features having the largest average negative SHAP values computed from the model’s testing phase for the predictive modelling analysis. Note that the SHAP values for every modelling scenario are computed by using the corresponding trained DNN model for the model’s testing phase.

3.5. Model performance evaluation

During the model training and validation steps for each of the 15 modelling scenarios, the traditional mean squared error (MSE) in Eq. (4), as indicated in Table 2, is leveraged to minimize the error difference between the measured and predicted target values in G-rate or D-rate. At the end of all epoch runs, the trained model is then subsequently evaluated during its testing phase via the use of the mean absolute percentage error (MAPE) as defined in Eq. (5).

\[ \text{MSE} = \frac{1}{Z} \sum_{i=1}^{Z} (Y_{pi} - Y_{mi})^2 \]  

(4)

\[ \text{MAPE} = \frac{1}{Z} \times \frac{100\%}{N} \sum_{i=1}^{Z} \left| \frac{Y_{pi} - Y_{mi}}{Y_{mi}} \right| \]  

(5)

where Z is the number of data quantity being analysed, \( Y_{pi} \), the predicted target value on a specific day (\( t = i \)), and \( Y_{mi} \), the measured target value on a specific day (\( t = i \)).

4. Results and discussions

4.1. Global scale analysis of COVID-19 temporal evolution

On the global scale, Fig. 6a (i,ii) compares the average predicted and measured values for G-rate and D-rate, respectively, by using their corresponding optimal model configurations, i.e., MS and S parameters. At this stage, it is found that the respective combination of 3 days MS and 5 days MS, coupled a common 5 days smoothing for performing the moving average computations, can best model and forecast the temporal variations in the G-rate and D-rate target parameters during the model’s testing phase. Note that the average predicted profiles for the respective parameters are based upon the average of the predicted profiles from the corresponding batch sizes of 4, 8, and 16 as shown in Fig. 6a (i,ii). At the same time, Fig. 6b (i,ii) summarizes the mean absolute percentage error

| Hyper-parameters | Values |
|------------------|--------|
| No. of neurons in input layer | \((N + 1) \times \text{MS}, N \rightarrow \text{total no. features}\) |
| No. of neurons in hidden layer 1 | \(\text{int}((N + 1) \times \text{MS})/2), N \rightarrow \text{total no. features}\) |
| No. of neurons in hidden layer 2 | \(\text{int}((N + 1) \times \text{MS})/4), N \rightarrow \text{total no. features}\) |
| No. of neurons in hidden layer 3 | \(\text{int}((N + 1) \times \text{MS})/8), N \rightarrow \text{total no. features}\) |
| No. of neurons in hidden layer 4 | 3 |
| No. of neurons in output layer | 1 |
| Batch Size | 4, 8, 16 |
| Number of Epochs | 500 |
| Learning rate | 0.0001 |
| Number of cross validations | 10 |
| Activation function | Rectified Exponential Linear Unit (ReLU) |
| Optimisation function | Adam |
| Key cost function | Mean Squared Error (MSE) |
For the features importance analysis via SHAP, Fig. 7a (i,ii) illustrates the ranking of the different features (X-1 to X-20) in terms of their average impacts (either positive or negative) on the G-rate and D-rate parameters, respectively, from the testing phase of the corresponding trained predictive model. Note that negative SHAP value indicates that (MAPE) scores for all 15 model configurations of varying combinations of MS and S parameters, where the current MAPE values for the optimal model configuration for modelling and forecasting G-rate (3 days MS, 5 days S) and D-rate (5 days MS, 5 days S) are around 7.7% and 7.8%, respectively.

For the features importance analysis via SHAP, Fig. 7a (i,ii) illustrates the ranking of the different features (X-1 to X-20) in terms of their average impacts (either positive or negative) on the G-rate and D-rate parameters, respectively, from the testing phase of the corresponding trained predictive model. Note that negative SHAP value indicates that (MAPE) scores for all 15 model configurations of varying combinations of MS and S parameters, where the current MAPE values for the optimal model configuration for modelling and forecasting G-rate (3 days MS, 5 days S) and D-rate (5 days MS, 5 days S) are around 7.7% and 7.8%, respectively.
the specific feature has a general negative correlation with the target parameter, and vice versa for positive SHAP values. In our analysis, the focus is to determine the most influential features which can best control the G-rate and D-rate, i.e., minimising both target parameters, via the largest negative SHAP values. Readers are referred to Figs. A1 and A2 (Supplementary Figures) which plot the variations between the temporal normalised (with respect to the maximum value) raw values and SHAP values for every data feature to forecast the G-rate and D-rate target parameters from the model’s testing phase. In summary, the following key observations can be derived from Figure Ugail et al. (2021), namely:

- For modelling and forecasting the global G-rate target with the optimal model configuration of 3 days MS and 5 days S parameters, X-11 and X-10 are determined to be the 2 most effective control factors which can best minimize the G-rate values over time due to their largest negative average SHAP values as shown in Fig. 8(i). The combined effect of both X-11 (STAY-AT-HOME) and X-10 (FACE-COVERING-POLICIES) features solely in affecting the G-rate target is illustrated in Fig. 9(a) where the combination of the respective normalised values of 0.995 and 0.52 can result in the lowest possible G-rate of 6.14 during the model’s testing phase.
- Similarly, from Fig. 8(ii), it can be found that X-3 (COVID-VACCINATION) and X-9 (COVID-PUBLIC-CAMPAIGNS) are the 2 most effective control factors which can best minimize the D-rate values over time due to their largest negative average SHAP values. In Fig. 9(b), it can then be interpolated that the combined normalised values of 0.4 and 0.6 will result in average positive SHAP values, especially for Fig. A1 (Supplementary Figures) where there are significantly higher positive SHAP values which indicate the ineffectiveness of school closures in controlling the global G-rate over time.

4.2. Continental scale analysis of COVID-19 spatiotemporal evolution

Moving into the next scale (continental), Fig 10a–f (i,ii) compares the average predicted and measured values for G-rate and D-rate, respectively, by using their corresponding optimal model configurations, i.e., MS and S parameters, for the different continents (Asia, Africa, Europe, North America, South America) as shown. For brevity, Table 4 summarizes the respective optimal model configurations and their resulting MAPE scores for each of the 5 continents, in relation to the predictive modelling step for G-rate and D-rate targets. At this stage, the average
MAPE score for forecasting G-rate across all continents during the model’s testing phase is around 7.5% which is very close to that of the lowest possible MAPE score (~7.7%) for the global analysis. On the contrary, the average MAPE score for forecasting D-rate across all continents for the same testing phase is around 3.0% higher than that of the lowest possible MAPE score (~7.8%) for the global analysis.

• Asia (Figs. 11a, 12a, 13a, 14a, A3, A4): X-4 (COVID-CONTACT-TRACING) and X-12 (PUBLIC-GATHERING-RULES-COVID) are considered as the 2 most effective control factors (data features) in minimising G-rate in Asia continent over time due to their respective largest negative average SHAP values as shown in Fig. 13a(i). Fig. 14a(i) indicates that the sole combination of the normalised range of values of 0.49–0.60 and 0.57–1.00 for X-12 and X-4, respectively, can best minimize the G-rate in Asia to between 12.0 and 13.0. For forecasting the D-rate during the model’s testing phase, X-4 (COVID-CONTACT-TRACING) and X-2 (INCOME-SUPPORT-COVID) are determined to be the effective control factors in minimising D-rate in Asia over time as shown in Fig. 13a(i). Note that X-9 and X-1 are not considered, despite being ranked as the top 2 negative average SHAP values in Fig. A4 (Supplementary Figures), as the bulk of their SHAP values are close to the zero line in the figure as illustrated. Fig. 14a(ii) shows that the sole combination of the normalised range of values of 0.13–0.46 and, 0.57–0.82 or 0.97–0.99 or 0.99–1.0, for X-2 and X-4, respectively can result in the lowest possible D-rate values of 10.8–10.9 approximately. The combinations of (a) X-11 (STAY-AT-HOME) and X-9 (PUBLIC-CAMPAIGNS); and (b) X-12 (PUBLIC-GATHERING-RULES) and X-6 (INTERNATIONAL-TRAVEL) are deemed to the least effective control factors to minimize G-rate and D-rate values in Asia due to their largest positive average SHAP values in Fig. 13a(i) and Fig. 13a(ii), respectively.

• Africa (Figs. 11b, 12b, 13b, 14b, A5, A6): In the Africa continent, X-16 (COVID-STRINGENCY-INDEX) and X-13 (PUBLIC-EVENTS-COVID) are considered as the 2 most effective control factors in minimising G-rate in Africa over time due to their respective largest negative average SHAP values as shown in Fig. 13b(i). Fig. 14b(i) indicates that the sole combination of the normalised range of values of 0.24–0.83 and 0.97–1.00 for X-16 and X-13, respectively, can best minimize the G-rate in Africa to between 17.5 and 17.6. For forecasting the D-rate during the model’s testing phase, X-4 (COVID-CONTACT-TRACING) and X-3 (COVID-VACCINATION-POLICY) are determined to be the effective control factors in minimising D-rate in Africa over time as shown in Fig. 13b(ii). Fig. 14b(ii) shows that the sole combination of the normalised range of values of 0.24–0.81 and 0.29–0.80 for X-4 and X-3, respectively can result in the lowest possible D-rate values of 10.5–11.0 approximately. The combinations of (a) X-15 (COVID-CONTAINMENT-AND-HEALTH-INDEX) and X-4 (COVID-CONTACT-TRACING); and (b) X-16 (COVID-STRINGENCY-INDEX) and X-9 (PUBLIC-CAMPAIGNS) are deemed to the least effective control factors to minimize G-rate and D-rate values in Africa due to their largest positive average SHAP values in Fig. 13b(i) and Fig. 13b(ii), respectively.

• Europe (Figs. 11c, 12c, 13c, 14c, A7, A8): In the large European continent, X-10 (FACE-COVERING-POLICIES) and X-17 (SCHOOL-CLOSURES) are considered as the 2 most effective control factors in minimising G-rate in Europe over time due to their respective largest negative average SHAP values as shown in Fig. 13c(i). Fig. 14c(i) indicates that the sole combination of the normalised range of values of 0.19–0.23 and 0.87–0.93 for X-10 and X-17, respectively, can best minimize the G-rate in Europe to between 5.00 and 5.03. For forecasting the D-rate during the model’s testing phase, X-16 (COVID-STRINGENCY-INDEX) and X-10 (FACE-COVERING-POLICIES) are determined to be the effective control factors in minimising D-rate in Europe over time as shown in Fig. 13c(ii). Fig. 14c(ii) shows that the sole combination of the normalised range of values of 0.76–0.77 and 0.94–1.00 for X-16 and X-10, respectively can result in the lowest possible D-rate values of 0.70–0.71 approximately. The combinations of (a) X-15 (COVID-CONTAINMENT-AND-HEALTH-INDEX) and X-16 (COVID-STRINGENCY-INDEX); and (b) X-7 (INTERNAL-MOVEMENT) and X-8 (PUBLIC-TRANSPORT) are deemed to the least effective control factors to minimize G-rate and D-rate values in Europe due to their largest positive average SHAP values in Fig. 13c(i) and Fig. 13c(ii), respectively.

• North America (Figs. 11d, 12d, 13d, 14d, A9, A10): In North America having 50 cities/states, X-9 (PUBLIC-CAMPAIGNS) and X-16 (COVID-STRINGENCY-INDEX) are considered as the 2 most effective control factors in minimising G-rate in North America over time due to their respective largest negative average SHAP values as shown in Fig. 15d(i). Fig. 14d(i) indicates that the sole combination of the normalised range of values of 0.24–1.00 and 0.23–0.89 for X-9 and X-16, respectively, can best minimize the G-rate in Europe to between 3.18 and 3.20. For forecasting the D-rate during the model’s testing phase, X-3 (COVID-VACCINATION-POLICY) and X-9 (PUBLIC-CAMPAIGNS) are determined to be the effective control factors
in minimising D-rate in North America over time as shown in Fig. 13d(ii). Fig. 14d(ii) shows that the sole combination of the normalised range of values of 0.52–0.98 and 0.99–1.00 for X-3 and X-9, respectively can result in the lowest possible D-rate values of 0.68–0.69 approximately. The combinations of (a) X-6 (INTERNATIONAL-TRAVEL) and X-17 (SCHOOL-CLOSURES); and (b) X-7 (INTERNAL-MOVEMENT) and X-4 (COVID-CONTACT-TRACING) are deemed to the least effective control factors to minimize G-rate.

Fig. 10. Predictive modelling of (i) G-rate and (ii) D-rate using 1 day lead-time under continental scale analysis: (a–e) best predictive results using optimal modelling scenario for Asia, Africa, Europe, North America, and South America, respectively, and (f) summary of MAPE (%) scores by averaging across all batch sizes for all modelling scenarios in each continent.
and D-rate values in North America due to their largest positive average SHAP values in Fig. 13d(i) and Fig. 13d(ii), respectively.

**South America** (Figs. 11e, 12e, 13e, 14e, A11, A12): In South America, X-5 (COVID-TESTING-POLICY) and X-9 (PUBLIC—CAMPAIGNS) are considered as the 2 most effective control factors in minimising G-rate in South America over time due to their respective largest negative average SHAP values as shown in Fig. 13e(i). Fig. 14e(i) indicates that the sole combination of the normalised range of values of 0.45–0.91 and 0.38–0.96 for X-5 and X-9, respectively, can best minimize the G-rate in South America to between 5.03 and 5.05. For forecasting the D-rate during the model’s testing phase, X-13 (PUBLIC-EVENTS) and X-12 (PUBLIC-GATHERING-RULES) are determined to be the effective control factors in minimising D-rate in South America over time as shown in Fig. 13e(ii). Fig. 14e(ii) shows that the sole combination of the normalised range of values of 0.32–0.87 and 0.32–0.98 for X-13 and X-12, respectively can result in the lowest possible D-rate values of 1.56–1.60 approximately. The combinations of (a) X-11 (STAY-AT-HOME) and X-6 (INTERNATIONAL-TRAVEL); and (b) X-6 (INTERNATIONAL-TRAVEL) and X-3 (COVID-VACCINATION-POLICY) are deemed to the least effective control factors to minimize G-rate and D-rate values in South America due to their largest positive average SHAP values in Fig. 13e(i) and Fig. 13e(ii), respectively.

To provide clarity to our readers, Table 5 summarizes the top 2 most effective features/factors in minimising the G-rate and D-rate targets in the respective continents, together with their optimal range of normalised values for the corresponding control factors identified from our continental scale analysis by far. By grouping the effective control factors, as summarised above, amongst all 5 continents, the 3 most common control factors which may conservatively minimize G-rate and D-rate across all continents are X-4 (COVID-CONTACT-TRACING), X-9 (PUBLIC-CAMPAIGNS), and X-16 (COVID-STRINGENCY-INDEX). In finer considerations, the 3 common control factors are generally more effective for the following continents, namely: (a) X-4 for Asia and Africa; (b) X-9 for North America and South America; and (c) X-16 for Africa, Europe, and North America. On the other hand, X-6 (INTERNATIONAL-TRAVEL) and X-7 (INTERNAL-MOVEMENT) can generally be identified as the least effective control factors in minimising both target parameters, while X-11 (STAY-AT-HOME) may also be considered as less effective which can be indicative of individuals violating stay-home notices. Finally, X-6 and X-7 control factors can also be considered as less effective for Asia, Europe, North America, and South America.

### 4.3. Country scale analysis of COVID-19 spatiotemporal evolution

To definitively determine the most effective, i.e., optimal, control factors which can best minimize the G-rate and D-rate universally across all countries (with available data records), the previous continental scale analysis is further built upon by performing country scale clustering via the multiple 1-day, 3-days, 5-days, and 7-days moving-average (MA) profiles derived for each of the target parameter where each profile directly constitute to a unique scenario (total of 8 for G-rate and D-rate combined) for the clustering analysis. Under each unique scenario for model’s testing phase, clustering of the corresponding time-series profiles pertaining to more than 150 countries is carried out via k-means algorithm where dynamic time warping (DTW) is traditionally used to measure similarity between two temporal sequences. For example, considering 2 independent time-series profiles in \( A(n) \) and \( B(m) \), where \( n \) and \( m \) represent the total number of data instances for the corresponding profiles. By applying the DTW technique to examine both \( A(n) \) and \( B(m) \) profiles, the method will compute the shortest distance difference of each data point of \( A(n) \) with every single data point of \( B(m) \) hence resulting in \((n \times m)\) comparisons. Hence, a warped distance path is formed between \( A(n) \) and \( B(m) \) which minimizes the Euclidean distance between both time-series profiles. We note that this is particularly useful for clustering time-series profiles which may not align exactly along the time-axis. In our case, however, the time-series profiles for all 150 countries align temporally to perform the clustering analysis with DTW under each unique MA computation scenario.

Figs. 15(a–d) and 16(a–d) illustrate the following: (i) the distribution of the different countries (>150 in quantity) globally into 3 main clusters; and (ii) the subsequent impact on the clustered SHAP values for each cluster under every MA computation scenario, with respect to G-rate and D-rate target parameters, respectively. Within each formed cluster under every MA scenario, we compile the top 3 most effective control factors for the respective countries which directly relate to their own continent, as extensively discussed in the preceding section. For example, if Cluster 1 has \( N_1 \) countries where \( n_i \) countries correspond to each of the 5 major continents, then the average SHAP value for the respective feature, i.e., control factor, can be computed as follows via simple averaging:

\[
S_{X_{y,i},a} = \frac{\sum_{i=1}^{N_1} n_i S_{X_{y,i}}}{N_1}
\]

where \( n_i \) is the total number of countries belonging to a specific continent, \( N_1 \) total number of countries from all continents within the defined cluster, \( m \) total number of possible continents within the defined cluster, \( i \) the continent index, and \( S_{X_{y,i}} \) the SHAP value for feature \( X_{y} \) (\( y = 1, 2, 3, \ldots \)).

With respect to the multiple MA computation scenario using G-rate, the map representations in (i) of Fig. 15(a–d) illustrate that the bulk of the >150 countries globally can be clustered into Cluster 1 where the order of magnitude of the average G-rate values is within \( O(10^3) \), while the respective orders of magnitude for Clusters 2 and 3 are within \( O(10^4) \) and \( O(10^5) \) as shown in the average G-rate time-series profiles in plot (ii) of Fig. 15(a–d). Within the same plot (ii) of Fig. 15(a–d), the respective stacked bar plots illustrate the distributions of the top control factors as collated from all possible continents within each formed cluster. Consistently, it can be easily observed from the collated stacked bar plots in (ii) of Fig. 15(a–d) for all MA computation scenarios that the top 3 conservative features (control factors) which can best minimize G-rate universally across all countries in the 3 main clusters are: (1) X-4 (COVID-CONTACT-TRACING); (2) X-12 (PUBLIC-GATHERING-RULES); and (3) X-16 (COVID-STRINGENCY-INDEX).

On the contrary, when using D-rate parameter under the different MA computation scenarios for the time-series clustering analysis, the map representations in (i) of Fig. 16(a–d) generally show less consistency in the clustering of the different countries globally where the 1-day and 5-days MA scenario can result in the bulk of the countries to reside within Cluster 2, as shown. At the same time, the orders of magnitude of the average D-rate values generally vary across the different MA scenarios for the same cluster (e.g., Cluster 1) formed. For example, comparing the average D-rate time-series profiles in the plot (ii) of Fig. 16b (3-days MA) and Fig. 16d (7-days MA) for Cluster 2, the respective order of magnitude is \( O(10^1) \) and \( O(10^3) \). The variations are likely ascribed to the complexity in D-rate profiles for the different

| Continents          | G-rate | D-rate |
|---------------------|--------|--------|
|                     | Optimal model configuration | MAPE (%) | Optimal model configuration | MAPE (%) |
| Asia                | 1 day MS, 7 days S | 3.3% | 1 day MS, 7 days S | 9.5% |
| Africa              | 1 day MS, 7 days S | 7.6% | 1 day MS, 7 days S | 9.5% |
| Europe              | 3 days MS, 5 days S | 5.2% | 3 days MS, 7 days S | 11.6% |
| North America       | 1 day MS, 7 days S | 11.0% | 7 days MS, 7 days S | 14.7% |
| South America       | 3 days MS, 7 days S | 10.0% | 1 day MS, 7 days S | 13.6% |

**Table 4** Summary of optimal model configuration and resulting MAPE scores for modelling G-rate and D-rate across different continents.
clusters formed under the multiple MA computation scenarios (Fig. 16 (a–d)). This is further highlighted by the case that the stacked bar plots in (ii) of Fig. 16(a–d) reveal different sets of important features (control factors) which can best minimize the D-rate across the 3 main clusters of the different countries globally under the respective MA scenarios, as follows:

- **1-day MA** (Fig. 16a): X-10 (FACE-COVERING-POLICIES), X-12 (PUBLIC-GATHERING-RULES), and X-16 (COVID-STRINGENCY-INDEX)
- **3-days MA** (Fig. 16b): X-1 (DEBT-RELIEF), X-4 (COVID-CONTACT-TRACING), X-9 (PUBLIC-CAMPAIGNS), X-10 (FACE-COVERING-
4.4. Discussions

Building upon the optimal control measures, i.e., data features, determined from the country-scale analysis, we first associate the present results with several reported findings from the literature to check on their overall level of consistency, followed by discussing in-detail on how the recommended measures can be useful to control G-rate and D-rate at the multiple spatial scales (global, continental, country-level).
Based on the key summarised notes in Table 1 for the most relevant literature studies at this stage, it is first worth noting that there has been a lack of previous research studies which investigated the impacts of a range of control measures (at least more than 10) on COVID-19, in terms of the resulting G-rate and D-rate parameters at the multiple scales. By far, the closest related studies have been performed by Yue et al. (2021) and Chen et al. (2021a) where the authors reported that government influences, in the forms of stringency index and governance capacity, respectively, are most effective to minimize the transmission rates of COVID-19, particularly in Asia. Hence, their findings are generally consistent with our reported results that X-16 (COVID-STRINGENCY-INDEX) (1 of our top 3 control measures) is expected to be best minimize the reported G-rates across all clusters of countries from the previous country-scale analysis. The effectiveness of X-12 (PUBLIC-GATHERING-RULES) to control the G-rate parameter also aligned with the recent reported studies by Das et al. (2021) and Zhu and Tan (2021) where the authors highlighted that the level of spatial interactions amongst individuals, in the forms of home quarantines and level of living conditions, significantly affect the spread of COVID-19, especially in urbanised cities. Overall, there is a good level of confidence that the recommended control measures of (1) X-4 (COVID-CONTACT-TRACING), (2) X-12 (PUBLIC-GATHERING-RULES), and (3) X-16 (COVID-STRINGENCY-INDEX), can be both conservative and effective to control the spread of COVID-19 in the foreseeable future, and its subsequent death rates (D-rates), while not severely compromising one nation’s economic stability. For example, the
implementation of work-from-home (WFH) status can still enable the working population to contribute to the nation’s economic activities, while reducing social interactions.

On the other hand, X-1 (DEBT-RELIEF) control measure appears to be dominant in controlling D-rate for the modelling scenarios of 3-days MA, 5-days MA, and 7-days MA as quantificationally demonstrated in our country-level analysis. At this stage, there is some level of similarities with the previous studies performed by Sannigrahi et al. (2020) and Li et al. (2021) where the authors also reported that income level and the nation’s level of commercial strength can control the number of confirmed COVID-19 cases and related deaths in the European region and China. However, as noted earlier, there is a need to re-focus on the predictive modelling analysis for D-rate to definitively determine a universal set of optimal control measures which can best minimize the reported D-rates at the multiple spatial scales. Finally, it is worth highlighting that several of the recommended control measures, from the present study, do make intuitive sense from a practical aspect where they have been demonstrated to be reasonably effective in some

---

**Fig. 14.** 2D dependence plot of (i) G-rate and (ii) D-rate with top 2 features in accordance to their largest negative average SHAP value under continental scale analysis: (a) Asia; (b) Africa; (c) Europe; (d) North America; and (e) South America.
control the growth-rate (G-rate) and death-rate (D-rate) behaviours due to COVID-19 at the multiple spatial scales (global, continent, and country scales). Importantly, the present literature lacks a generic model framework which can systematically and effectively investigate the impacts of the multitude of control measures on the modelled G-rate and D-rate target parameters from the global to individual country scales, followed by recommending the most critical/effective control measures which can best minimize both COVID-19 related parameters over time. To contribute to the global community’s growing pool of knowledge pertaining to COVID-19 control and management, this paper develops a top-down multiscale engineering approach which combines predictive modelling and explainable AI, via SHAP (Shapley Additive Explanations) method, to perform COVID-19 predictive analyses at the 3 multiple systematic scales of global, continental, and country levels. The key results for the respective spatial scales are delineated as follows:

- **Global scale**: Optimal model configuration of: (1) 3 days MS and 5 days S parameters; and (2) 5 days MS and 5 days S parameters, can achieve lowest possible MAPE scores of 7.7% and 7.8% for forecasting G-rate and D-rate, respectively, during model’s testing phase. Using the respective model configuration for explainable AI via SHAP analysis, the combined effect of: (1) X-11 (STAY-AT-HOME) and X-10 (FACE-COVERING-POLICIES); and (2) X-3 (COVID-VACCINATION) and X-9 (COVID-PUBLIC-CAMPAIGNS), can best minimize the G-rate and D-rate globally over time.

- **Asia scale**: Optimal model configuration of: (1) 1 day MS and 7 days S parameters; and (2) 1 day MS and 7 days S parameters, can achieve lowest possible MAPE scores of 3.3% and 9.5% for forecasting G-rate and D-rate, respectively, during model’s testing phase. Using the respective model configuration for the required SHAP analysis, the combined effect of: (1) X-4 (COVID-CONTACT-TRACING) and X-12 (PUBLIC-GATHERING-RULES-COVID); and (2) X-4 (COVID-CONTACT-TRACING) and X-2 (INCOME-SUPPORT-COVID), can best minimize the G-rate and D-rate in Asia over time.

- **Africa scale**: Optimal model configuration of: (1) 1 days MS and 7 days S parameters; and (2) 1 day MS and 7 days S parameters, can achieve lowest possible MAPE scores of 7.6% and 9.5% for forecasting G-rate and D-rate, respectively, during model’s testing phase. Using the respective model configuration for the required SHAP analysis, the combined effect of: (1) X-16 (COVID-STRINGENCY-INDEX) and X-13 (PUBLIC-EVENTS-COVID); and (2) X-4 (COVID-CONTACT-TRACING) and X-3 (COVID-VACCINATION-POLICY), can best minimize the G-rate and D-rate in Africa over time.

- **Europe scale**: Optimal model configuration of: (1) 3 days MS and 5 days S parameters; and (2) 5 days MS and 5 days S parameters, can achieve lowest possible MAPE scores of 5.2% and 11.6% for forecasting G-rate and D-rate, respectively, during model’s testing phase. Using the respective model configuration for the required SHAP analysis, the combined effect of: (1) X-16 (COVID-STRINGENCY-INDEX) and X-13 (PUBLIC-EVENTS-COVID); and (2) X-4 (COVID-STRINGENCY-INDEX) and X-3 (COVID-VACCINATION-POLICY), can best minimize the G-rate and D-rate in Europe over time.

- **North America scale**: Optimal model configuration of: (1) 1 day MS and 7 days S parameters; and (2) 7 days MS and 7 days S parameters, can achieve lowest possible MAPE scores of 5.2% and 11.6% for forecasting G-rate and D-rate, respectively, during model’s testing phase. Using the respective model configuration for the required SHAP analysis, the combined effect of: (1) X-16 (FACE-COVERING-POLICIES) and X-17 (SCHOOL-CLOSURES); and (2) X-16 (COVID-STRINGENCY-INDEX) and X-10 (FACE-COVERING-POLICIES), can best minimize the G-rate and D-rate in Europe over time.

- **South America scale**: Optimal model configuration of: (1) 3 days MS and 7 days S parameters; and (2) 1 day MS and 7 days S parameters, can achieve lowest possible MAPE scores of 10.0% and 13.6% for forecasting G-rate and D-rate, respectively, during model’s testing phase. Using the respective model configuration for required SHAP

---

**Table 5**

Summary of optimal model configuration and resulting MAPE scores for modelling G-rate and S-rate across different continents.

| Continents | G-rate | D-rate |
|------------|--------|--------|
|            | Top 2 most effective control factors | Normalised range of values | Top 2 most effective control factors | Normalised range of values |
| Asia       | X-4    | 0.57–1.00 | X-4 | 0.57–0.82 |
|            | X-12   | 0.49–0.60 | X-2 | 0.13–0.46 |
| Africa     | X-16   | 0.24–0.83 | X-4 | 0.24–0.81 |
|            | X-13   | 0.97–1.00 | X-3 | 0.29–0.80 |
| Europe     | X-10   | 0.19–0.23 | X-16 | 0.76–0.77 |
|            | X-17   | 0.87–0.93 | X-10 | 0.94–1.00 |
| North      | X-9    | 0.24–1.00 | X-3 | 0.52–0.98 |
| America    | X-16   | 0.23–0.89 | X-9 | 0.99–1.00 |
| South      | X-5    | 0.45–0.91 | X-13 | 0.32–0.87 |
|            | X-9    | 0.38–0.96 | X-12 | 0.32–0.98 |

* Refer to Table 2 for the exact descriptions of the different features (X).

---

**Algorithm 1**

Shapley approximation for each feature value.  

1. **Input:** Number of iterations (M), data instance (x), feature index (j), data matrix (X), and trained predictive model (machine learning or deep learning) (f)  
2. **Output:** Approximated Shapley value for the value of the feature (xj)  
3. For all m = 1, . . . , M:  
   a. Draw random data instance z from the data matrix X  
   b. Select random permutation of the original feature values  
   c. Order instance x: x0 = (x1, . . . , xk)  
   d. Order instance z: z0 = (z1, . . . , zk)  
   e. Construct two new data instances  
      - With respect to j: xj = (x1, . . . , xj, . . . , xk)  
      - Without j: x: x0 = (x1, . . . , xj, . . . , xk)  
   f. Compute marginal contribution: φj m = f(xj) − f(xj)  
4. Compute Shapley value as the average: φj (∑Mm=1 φj m)
analysis, the combined effect of: (1) X-5 (COVID-TESTING-POLICY) and X-9 (PUBLIC-CAMPAIGNS); and (2) X-13 (PUBLIC-EVENTS) and X-12 (PUBLIC-GATHERING-RULES), can best minimize the G-rate and D-rate in South America over time.

- **Country scale:** In every modelling scenario (varying MS and S parameters) for clustering more than 150 countries globally, (1) X-4 (COVID-CONTACT-TRACING); (2) X-12 (PUBLIC-GATHERING-RULES); and (3) X-16 (COVID-STRINGENCY-INDEX) are determined to best minimize G-rate universally across all countries in the 3 main clusters formed. However, identifying the optimal control factors for minimising D-rate across all modelling scenarios appear difficult at

---

**Fig. 15. Time-series clustering of local countries based upon varying moving average computation scenarios for G-rate (a–d):** (i) global spatial distributions of local countries into 3 main clusters under respective scenario; and (ii) agglomeration of largest negative average SHAP values for 3 main clusters under respective scenario.
this stage due to variations in the identified factors amongst the different scenarios.

Overall, we have quantificationally demonstrated that the combined control measures of X-4 (COVID-CONTACT-TRACING), X-12 (PUBLIC-GATHERING-RULES), and X-16 (COVID-STRINGENCY-INDEX) can be most useful to regulators and policy makers globally to best control the spread of COVID-19, in terms of its global and localised growth rates (G-rates). The general effectiveness of the identified measures has been demonstrated to be useful, to a significant extent, for real-world examples such as Singapore and France, where both X-4 (COVID-CONTACT-TRACING), X-12 (PUBLIC-GATHERING-RULES) are dominant control

---

**Fig. 16. Time-series clustering of local countries based upon varying moving average computation scenarios for D-rate (a–d):** (i) global spatial distributions of local countries into 3 main clusters under respective scenario; and (ii) agglomeration of largest negative average SHAP values for 3 main clusters under respective scenario.
measures implemented by the local regulators. In addition, several previous research studies have been shown to align closely with our reported findings, as explained in this paper. To continue contributing to the global community’s knowledge pool on COVID-19 control and management, our future works involve 2 main research directions, namely:

- To leverage on the most important control measures, i.e., data features, identified for the respective scales to perform multi-objective optimisation analysis to minimize the corresponding G-rate and D-rate profiles at the multiple spatial scales (global, continent, and country level). For example, combining suitable evolutionary optimisation algorithm such as non-dominated sorting genetic algorithm II (NSGA-II) to minimize selected test problems pertaining to time-series predictive modelling. A multi-objective optimization approach (Zhang & Lin, 2021) can be further proposed to identify the optimal conditions for the most effective control measures, as determined from the current study, that can best minimize G-rate and D-rate target parameters over time for a given spatial context.

- To quantificationally explain the ineffectiveness of various control measures, i.e., those having large positive average SHAP values, though they are expected to control the spread of COVID-19 globally, e.g., vaccination rates. The quantitative results derived will thus be useful to assist regulators around the world to better manage their control policies and avoid resources diversion into measures which may not be effective in controlling the spread of COVID-19, and its associated death rates.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors declare no conflict of interest. This study is explored solely based upon the authors’ research interests.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.scs.2022.103772.

References

Abdollahi, A., & Rahbaralam, M. (2020). Effect of temperature on the transmission of COVID-19: A machine learning case study in Spain. MedRxiv, 2020.05.01.20087759. 10.1101/2020.05.01.20087759.

Ahmadi, M., Sharifi, A., Dorosti, S., Fazleabas Ghouchshi, S., & Ghanbari, N. (2020). Investigation of effective climatology parameters on COVID-19 outbreak in Iran. Science of the Total Environment, 729, Article 138705. 10.1016/j.scitotenv.2020.138705.

Awasthi, R., Nagori, A., Singh, P., Pal, R., Joshi, V., & Sethi, T. (2020). Temperature and humidity do not influence global COVID-19 incidence as inferred from causal modelling. Total Environment, 7579. https://doi.org/10.1126/science.aat6030 (New York, N.Y.).

Baharudin, H. (2021, January 20). Nearly all in Singapore are on Covid-19 contact tracing programme TraceTogether. https://www.straitstimes.com/singapore/political-nearly-all-in-singapore-are-on-covid-19-contact-tracing-programme-tracetherough. 10.1101/2020.04.10.20060459.

Baker, R. E., Yang, W., Vecchi, G. A., Metcalfe, C. J. E., & grenfell, B. T. (2020). Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. Science, 369(6501), 315LP-31319. https://doi.org/10.1126/science.abc2553 (New York, N.Y.).

Bashir, M. F., Ma, B., Bilal, Komal, B., Bashir, M. A., Tan, D., et al. (2020). Correlation between climate indicators and COVID-19 pandemic in New York, USA. Science of the Total Environment, 728, Article 138835. 10.1016/j.scitotenv.2020.138835.

Chen, J., Wang, R., Gilby, N.B., & Wei, G.W. (2021). Omicron (B.1.1.529): Infecitivity, vaccine breakthrough, and antibody resistance. arxiv.org/abs/2112.03138.

Chen, J., Guo, X., Pan, H., & Zhong, S. (2021b). What determines cities’ resilience against epidemic outbreak: Evidence from China’s COVID-19 experience. Sustainable Cities and Society, 70, Article 102892. https://doi.org/10.1016/j.scs.2021.102892. March.
Pung, R., Lin, B., Maurer-Stroh, S., Sirota, F. L., Mak, T. M., Octavia, S., et al. (2021). The efficacy of social distance and ventilation effectiveness in preventing COVID-19 transmission. Sustainable Cities and Society, 75, Article 103254. https://doi.org/10.1016/j.scs.2021.103254, March.

Prata, D. N., Rodrigues, W., & Bermejo, P. H. (2020). Temperature significantly changes SAR-CoV-2 transmission in subtropical cities of Brazil. Science of the Total Environment, 729, Article 138862. https://doi.org/10.1016/j.scitotenv.2020.138862

Rao, S., & Singh, M. (2021). The newly detected B.1.1.529 (Omicron) variant of SARS-CoV-2 with multiple mutations. 1, 7–10.

Runkle, J. D., Sugg, M. M., Leeper, R. D., Rao, Y., Matthews, J. L., & Rennie, J. J. (2020). Short-term effects of specific humidity and temperature on COVID-19 morbidity in select US cities. The Science of the Total Environment, 740, Article 140093. https://doi.org/10.1016/j.scitotenv.2020.140093

Sahin, M. (2020). Impact of weather on COVID-19 pandemic in Turkey. Science of the Total Environment, 728, Article 138810. https://doi.org/10.1016/j.scitotenv.2020.138810

Sangmagali, S., Pilla, F., Basu, B., Basu, A. S., & Molter, A. (2020). Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. Sustainable Cities and Society, 62, Article 102418. https://doi.org/10.1016/j.scs.2020.102418

Services & Department, H. K. I., (2021). No title government of the Hong Kong Special Administrative Region. COVID-19 testing to intensify. http://www.news.gov.hk/eng/2021/02/20210201/20210201_172508_389.html

Shi, P., Dong, Y., Yan, H., Li, X., Zhao, C., Liu, W. et al. (2020). The impact of temperature and absolute humidity on the Coronavirus Disease 2019 (COVID-19) outbreak: evidence from China. MedRxiv, 2020.03.22.20038919. 10.1101/2020.03.22.20038919.

Sobral, M. F. F., Duarte, G. B., da Penha Sobral, A. I. G., Marinho, M. L. M., & de Souza Melo, A. (2020). Association between climate variables and global transmission of SARS-CoV-2. Science of the Total Environment, 729, Article 138997. https://doi.org/10.1016/j.scitotenv.2020.138997

Sun, C., & Zhai, Z. (2020). The efficacy of social distance and ventilation effectiveness in preventing COVID-19 transmission. Sustainable Cities and Society, 62, Article 102390. https://doi.org/10.1016/j.scs.2020.102390

Taylor, S., & Asmundson, G. J. G. (2021). Negative attitudes about facemasks during the COVID-19 pandemic: The dual importance of perceived ineffectiveness and psychological reactance. PloS One, 16(2), Article e0246317. https://doi.org/10.1371/journal.pone.0246317

Thompson, M. G., Stenehjem, E., Grannis, S., Ball, S. W., Naleway, A. L., Ong, T. C., et al. (2021). Effectiveness of Covid-19 vaccines in ambulatory and inpatient care settings. New England Journal of Medicine, 384(15), 1355–1371. https://doi.org/10.1056/nejmoa2110362

Tosepu, R., Gunawan, J., Effendy, D. S., Ahmad, L. O. A. I., Lestari, H., Bahar, H., et al. (2020). Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. Science of the Total Environment, 725, Article 138436. https://doi.org/10.1016/j.scitotenv.2020.138436

Ugail, H., Aggarwal, R., Iglesias, A., Howard, N., Campuzano, A., Suarez, P., et al. (2021). Social distancing enhanced automated optimal design of physical spaces in the wake of the COVID-19 pandemic. Sustainable Cities and Society, 68, Article 102791. https://doi.org/10.1016/j.scs.2021.102791

Vieierz, J., & Bienodi, D. (2021). The influence of urban, socio-economic, and eco-environmental aspects on COVID-19 cases, deaths and mortality: A multi-city case in the Atlantic Forest, Brazil. Sustainable Cities and Society, 69, Article 102859. https://doi.org/10.1016/j.scs.2021.102859

Wu, Y., Jing, W., Liu, J., Ma, Q., Yuan, J., Wang, Y., et al. (2020). Effects of temperature and humidity on the daily new cases and new deaths of COVID-19 in 166 countries. Science of the Total Environment, 729, Article 139051. https://doi.org/10.1016/j.scitotenv.2020.139051

Xie, J., & Zhu, Y. (2020). Association between ambient temperature and COVID-19 infection in 122 cities from China. Science of the Total Environment, 724, Article 138201. https://doi.org/10.1016/j.scitotenv.2020.138201

Yao, Y., Pan, J., Liu, Z., Meng, X., Wang, W., Kan, H., et al. (2020). No Association of COVID-19 transmission with temperature or UV radiation in Chinese cities. European Respiratory Journal, Article 2000517. https://doi.org/10.1183/13993003.00517-2020

Yuan, J., Yun, H., Lan, W., Wang, W., Sullivan, S. G., Jia, S., et al. (2006). A climatologic investigation of the SARS-CoV outbreak in Beijing, China. American Journal of Infection Control, 34(4), 234–236. https://doi.org/10.1016/j.ajic.2005.12.006

Zhang, L., & Lin, P. (2021). Multi-objective optimization for limiting tunnel-induced damages considering uncertainties. Reliab. Eng. Syst. Saf., 216, 107945. https://doi.org/10.1016/j.ress.2021.107945

Zhou, Y., Feng, L., Zhang, X., Wang, S., & Wu, T. (2021). Spatiotemporal patterns of the COVID-19 control measures impact on industrial production in Wuhan using time-series earth observation data. Sustainable Cities and Society, 75, Article 103388. https://doi.org/10.1016/j.scs.2021.103388, September.

Zhu, P., & Tan, X. (2021). Is compulsory home quarantine less effective than centralized quarantine in controlling the COVID-19 outbreak? Evidence from Hong Kong. Sustainable Cities and Society, 74, Article 103222. https://doi.org/10.1016/j.scs.2021.103222, March.