Revised: 2 March 2021 | Accepted: 2 April 2021

DOI: 10.1002/int.22440

RESEARCH ARTICLE

WILEY

Using self-organising maps to predict and contain natural disasters and pandemics

Raymond Moodley[1] | **Francisco Chiclana**[1,2] | **Fabio Caraffini**[1] | **Mario Gongora**[1]

[1]Institute of Artificial Itelligence, School of Computer Science and Informatics, De Montfort University, Leicester, UK

[2]Andalusian Research Institute on Data Science and Computational Intelligence (DaSCI), University of Granada, Granada, Spain

Correspondence
Raymond Moodley, Institute of Artificial Itelligence, School of Computer Science and Informatics, De Montfort University, The Gateway, Leicester LE1 9BH, UK.
Email: raymond.moodley@dmu.ac.uk

Abstract

The unfolding coronavirus (COVID-19) pandemic has highlighted the global need for robust predictive and containment tools and strategies. COVID-19 continues to cause widespread economic and social turmoil, and while the current focus is on both minimising the spread of the disease and deploying a range of vaccines to save lives, attention will soon turn to future proofing. In line with this, this paper proposes a prediction and containment model that could be used for pandemics and natural disasters. It combines selective lockdowns and protective cordons to rapidly contain the hazard while allowing minimally impacted local communities to conduct "business as usual" and/or offer support to highly impacted areas. A flexible, easy to use data analytics model, based on Self Organising Maps, is developed to facilitate easy decision making by governments and organisations. Comparative tests using publicly available data for Great Britain (GB) show that through the use of the proposed prediction and containment strategy, it is possible to reduce the peak infection rate, while keeping several regions (up to 25% of GB parliamentary constituencies) economically active within protective cordons.

KEYWORDS

epidemiology, pandemic and natural disaster containment, predicting pandemics, protective cordons, self organising maps

1 | INTRODUCTION

Predicting and containing natural disasters, including the spread of diseases through pandemics, are indeed important functions of all tiers of government, global organisations like the United Nations, and relief agencies like the Red Cross. While several artificial intelligence (AI) systems are available to predict natural disasters, these are usually very complex, domain-specific, and expensive, thus limiting its global access, in particular to poor countries that often need these systems the most.[1] In fact, even with these complex systems and wealth of data, the world has been surprised at how ill-prepared most countries were in dealing with a fast spreading pandemic, as was the case with The severe acute respiratory syndrome coronavirus-2 (COVID-19).[2] Globally, governments were criticised for their slow and ineffective response to COVID-19, with many governments implementing strict lockdowns starting in March 2020 as a "knee-jerk" response which still resulted in thousands of fatalities as well as economic and social turmoil.[2-5]

It is well-recognised that recent advances in AI and technology have made significant contributions to predictive modelling, and key to this success has been having access to large volumes of reliable data,[1] highly effective and robust algorithms,[6] and high performance computers.[7] Clearly these require investment, skills and access to data which limits the organisations, and to some extent the countries, that can harness AI for the common good of society. Given this, we take the philosophical approach and ask: Is it possible to create a simple AI model which uses limited computing power and limited historic, publicly available data to predict and contain natural disasters and pandemics? We note that if this is possible, then this approach could provide a host of possibilities in helping organisations better prepare and respond to catastrophic events, thus saving lives and easing suffering, especially among the world's poorest and most vulnerable.

In line with this, we propose a simple AI algorithm that could be useful in the prediction and containment of natural disasters, including future outbreaks of COVID-19 and other pandemics. Our approach centres on simplicity; limited computer processing power requirements; and limited, publicly available historic data to provide accurate modelling that can be used as part of organisations' prediction and containment strategies. Our approach is unique in that it uses AI and data-driven techniques, notably self-organising maps (SOM), to both stratify a country or region and predict the "susceptibility" of each area within that region. This allows decision makers to identify areas that are likely to be most negatively impacted, areas that are most able to provide support, and areas that need to be protected, for example, protecting vulnerable, isolated areas against the spread of disease during a pandemic. The use of SOM in epidemiology is not new, and we leveraged concepts from previous SOM studies as well as from studies involving self-sustaining urban planning to develop our proposed approach.[8-11] We use the UK's experience of the COVID-19 pandemic as a case study. For completeness, we perform a macrocomparison of our approach with that of the UK government's current modelling and public reporting which is based on the well-known "SIR" model (Susceptible, Infectious, and Recovered).[12-14] We also compare our proposed SOM approach with the fuzzy C-means (FCM) clustering approach which has been shown to be highly effective in clustering unstructured data.[15]

The remainder of this paper is organised as follows: Section 2 provides a review of recent literature related to this study, with Section 3 detailing our analytical model and approach. We outline our experimental conditions in Section 4, with results discussed in Section 5. Finally we summarise our findings, discuss the implications of our work and highlight possible future work in Section 6.

2 | LITERATURE REVIEW

2.1 | Modelling natural disasters and pandemics

Performing mathematical modelling and simulations to predict the occurrence and impact of natural disasters and the spread of diseases during pandemics have been well-studied over the years.[4,13,16-18] Indeed, modelling natural disasters and pandemics can be extremely complex depending on the number of variables involved.[4] It is thus not surprising that most models are specific to a specific types of catastrophic events and do not easily lend themselves to the modelling of other events. For example, in Reference [16], Dell'Acqua et al. detailed the development and use of an earthquake monitoring system including the use of geo-spatial and urban planning data, while Currie et al., in Reference [4], detailed four broad categories of pandemic modelling:

- Systems dynamics (SD) models, which use differential equations to model flows;
- Agent-based models (ABM) for interactions between people within a community;
- Discrete event simulation (DES), which is used for resource availability e.g. nurses, hospital beds, and so on;
- Hybrid models, which uses a combination of the other modelling techniques.

Similarly, in their special issue on natural disaster modelling and risk assessment, Teves-Costa, and Mendes[17] published a collection of research papers, each to a specific type of catastrophic event (e.g., earthquakes, hurricanes, flooding, etc.). However, irrespective of the specifics, all studies that have been reviewed emphasised the vital role of data in creating effective models to save lives and reduce risks.

2.1.1 | Modelling epidemics and pandemics

The pandemic modelling categories, highlighted by Currie et al. in Reference [4], have been used for different applications by various governments, notably the UK government, in the ongoing fight against COVID-19. Indeed, one of the key models being used by governments is the "SIR" model, a form of SD model, first developed in 1927 by Kermack and McKendrick.[14] While the "SIR" model for modelling pandemics is mathematically robust, requiring the solution of a set of differential equations with some boundary conditions, it does rely on some debatable basic assumptions. In Reference [18], Epstein noticed that the "SIR" model does not capture complex social interactions and assumes that everyone is identical and nobody adapts their behaviour based on the prevalence of the disease. Epstein concluded that the "SIR" model can be ill-suited to the modern context where society is highly mobile and recommended the use of the ABM model to better capture the interactions between people. In contrast, Araz[19] showed that SD models can be expanded to incorporate several interactions that simulate the "real world" during pandemics including hospitalisation, vaccination programs, reinfections, and deaths. However, these models quickly become very complex, with the number of assumptions also rising rapidly: for example, the model detailed in Reference [19] required 10 differential equations with corresponding boundary conditions being assumed from several other studies and sources.

AI techniques, including both supervised and unsupervised learning, to model pandemic-related processes has been previous used, notably by Uhlig et al. in Reference [20]. Uhlig et al. showed some success in using neural networks as part of a top-down approach to track a set of early warning indicators as a means of understanding the progress of COVID-19 and the risk of localised outbreaks. However, the authors did note that further work is required to enhance their model, given that the COVID-19 pandemic was still in progress.[20] Neural networks have also been used in disease management including diagnostics and resource planning. In References [21,22] convolutional neural networks were used as part of algorithms to automate the processing of X-ray images to identify and diagnose COVID-19, while in Reference [23] neural networks were used as part of emergency departments' staffing algorithms to optimise the response of medical facilities during pandemics.

2.1.2 | SOM and its applications

SOM remains a popular unsupervised learning technique due to it being relatively easy to implement while producing powerful, topologically maintained, visual representations of complex data.[24,25] In many ways, SOM can be seen as a constrained clustering process, be it crisp or fuzzy clustering as detailed in Reference [15]. Unlike in traditional clustering, where only the number of clusters (or centroids) is specified, in SOM both the number and location of the centroids are specified, thus the data is "fitted" to the centroids as opposed to the centroids being appropriately chosen to best represent the data around it.[26] The key benefit of SOM over all other unsupervised learning techniques is its topological preservation property, which is achieved through a combination of "competition" and "collaboration" of neighbouring neurons in a neural network. It is this property that has resulted in SOM being widely used across several applications as the technique efficiently combines a series of data features to organise points that are similar in nature, while highlighting dissimilarities between topologically—close points, making feature and outlier detection relatively easy.[24-26] SOM has similar drawbacks to other clustering techniques like FCM, which includes the difficulty in selecting the optimum number of nodes, and increased computational time in higher dimensions. However, as in the case of clustering, these are usually overcome using multiple passes (trial error) and high-performance computing, which is becoming common place in today's society.[15,25]

Given the above, SOM has been successfully used in many applications including fraud detection in banking, detecting cancerous tumours, genetic modelling in medicine, poetry, and literary text analysis in education, and analysing changing weather patterns to quantify the impacts of climate change.[24,25] SOM has also been used successfully in a variety of epidemiological studies, usually in combination with geographic information systems (GIS), to help public health officials identify disease "hotspots," group communities that display similar symptoms, and identify environmental factors that contribute to increased spread of diseases.[8,9,11] In Reference [9], SOM were used to quantify the prevalence of dengue fever in the various regions of Andhra Pradesh, India, while in Reference [11] it was used to map the pollution levels in Atlanta (USA) based on the type of day (e.g., sunny, windy, etc.). Similarly, in Reference [8], 511 communities across New York State were grouped into five clusters based on 92 environmental variables. These studies are indeed quite relevant and methods and techniques have been leveraged as part of this study.

2.2 | Self-sustaining regions

While globalisation has provided several strong, positive benefits for people and cities across the world, it has resulted in communities becoming increasingly reliant on imports, thus making them less self-sustaining.[27] According to References [10,28], a self-sustaining city contains a defined perimeter inside which lies a self-sufficient population, that is, where the economy within the city fully employs its population, and where services and cultural infrastructure are in sufficient supply for its population. This notion can easily be extended to regions. Creating and sustaining self-sustaining regions is quite complex and often involves both a high level of initial investment supported by ongoing commitment from governments, local businesses and the community.[28]

Advocates for self-sustaining cities have typically positioned the benefits from reducing the risks associated with the reliance for food, energy, and other products from external sources, as well as environmental protection due to global financial power imbalances between food producers and food consumers.[27] However, the emergence of lockdowns as a result of pandemics, and its consequential pausing of local economies, has now provided yet another benefit to encourage self-sustaining urban developments. Following the outbreak of COVID-19, there have been several opinion articles on the future of sustainable cities, particularly focusing on local food production, local employment and transportation planning.[29,30] However, it is still early for formal research to be conducted on urban planning for sustainability during pandemics and it is likely that we may see further studies on this topic being published in the near future. Park et al.[10] studied the implications for self-sustaining cities in South Korea and, as part of their study, provided a comprehensive list of over 40 parameters that are considered relevant for the modelling of self-sustaining urban developments. We leveraged a selection of the most relevant parameters for the development of our model.

3 | PROBLEM STATEMENT AND ANALYTICAL MODEL

3.1 | Problem statement

It is realised that some local regions within a country may be more susceptible than others while others may be more self-sustaining. The notion of self-sustaining in this context is as defined in Park et al.[10] Given this, we define the problem being addressed in this study as follows: Let $L = \{L_1, L_2, ..., L_i\}$ be the set of all local regions, each with a set of measurable characteristics, $C = \{c_1, c_2, ..., c_n\}$. Within C there exists $k\,(1 \leq k \leq n)$ "negative" characteristics that increase with increasing susceptibility (e.g., population density) and $q \quad (=n - k)$ "positive" characteristics that may be seen as "contributors" to L, and which may be hampered during a disaster (e.g., economic output). We note that having "positive" characteristics is desired but not essential for susceptibility. Further, let all points L_i be mapped on the curve P, such that each L_i has a corresponding point P_i on the curve P. Thus there may be some points P_i along P that could be considered to be more susceptible during a disaster than other points along P. For example, if L_1 contains higher values for all "negative" characteristics compared with L_2, then L_1 is considered to be at greater risk during a disaster than L_2. Given that most governments are intent on minimising disruption during a disaster, it thus becomes necessary to stratify a country based on points along P, with points of relatively lower risk being protected and allowed to operate relatively normally, for example, implementing protective cordons during a pandemic, or allowed to offer support to more susceptible areas, that is, offering emergency personnel, shelter, and equipment to other areas during a natural disaster. Thus the aim of this study is to provide a framework for stratifying a country, or part of a country, based on P, thus allowing governments to predict a region's susceptibility, and take decisions on which regions could be protected and allowed to operate relatively normally (e.g., during a pandemic) or offer support (e.g., during an earthquake).

3.2 | Analytical model

We commence by noting that an exact solution to P may be quite complex. Further, given that P contains both "positive" and "negative" characteristics which can offset each other mathematically, we note that there may be no great benefit in knowing the exact equation that defines P. However, what is valuable is the identification of all L_is along P with similar characteristics. Thus, by grouping L_is based on their location along P, governments can easily identify (based on predictive modelling) geographical regions that have a high risk (and which will likely require support) and low risk (which may need to be protected and/or leveraged for support).

Given this, we approach this stratification and prediction problem using SOM as detailed in References [9,24-26]. SOM, being an unsupervised learning technique, is well-suited to problems that involve high dimensionality, parameters that are difficult to quantify, and where the relative position of an outcome is more important than the actual value of the outcome itself. The key principle of SOM is that it performs dimensionality reduction and topological preservation by establishing a one- or two-dimensional map with a set of nodes (e.g., 36 nodes in a 6 by 6 square SOM), where each node is arbitrarily assigned a set of characteristics C and positioned in the space containing P. The nodes are then iteratively adjusted to cover all points along P using an optimisation process usually based on a distance measure (e.g., Euclidean distance). Thus all L_is are then assigned to their nearest node and, as a result, each node of the SOM contains L_is that have similar characteristics. In addition, since topology is maintained in SOMs, nodes located close to each other typically have similar characteristics. Once developed, the resulting SOM serves as a predictive model which may then be visually inspected, and decisions could be taken (e.g., which geographical regions should be protected, and which should offer support, etc.).

4 | EXPERIMENTS

We applied our analytical model to the COVID-19 pandemic within Great Britain (GB).

4.1 | Case study: COVID-19 within GB

The UK has been heavily impacted by COVID-19, having one of the highest death rates globally.[31] In light of this, we use mainland UK, or GB, as a case study to demonstrate the effectiveness of our proposed model in predicting outbreaks and "hotspots," and whether an alternative COVID-19 government response strategy, that is, using protective cordons, would have been more effective in minimising the negative impacts of the pandemic.

Using the notation outlined in Section 3, we used GB as the country under consideration, L, and its 632 parliamentary constituencies as distinct local regions, L_i. Note that we excluded Northern Ireland as it does not form part of GB. We gathered basic, publicly available data for each of the 632 constituencies and used this as the basis of our model. Further details are provided in the sections that follow.

4.2 | Experimental process

Experiments were conducted based on the well-known Knowledge Discovery in Databases (KDD) process, first outlined in Reference [32], and more recently in Reference [26] where it was noticed that the heart of the KDD process is the data mining phase, which leverages models and algorithms to process data into information. In this regard the SOM algorithm forms part of the data mining and postprocessing phases where patterns are identified and interpreted to contribute to overall knowledge.[26]

Our approach's computer code and data is publicly available in Reference [33]; it leverages data and assumptions from several sources as provided in Table 1; and it comprises of the following steps for our case study of the UK's experience during the COVID-19 pandemic:

1. Identify the distinct local regions, L, that are under consideration.
2. Define the set of key characteristics, C, that are relevant for a pandemic. These characteristics must be measurable, or assigned a measure, for each distinct local region L_i. The measure could be a numerical value (e.g., number of people), ordinal (e.g., rank of prosperity relative to the other regions), or categorical (e.g., good, bad).
3. Gather data for each of the characteristics in each local region.
4. Run data using an SOM model, for example, using the Kohonen package in R.[34]
5. Interpret the SOM map and identify nodes on the SOM map that can be cordoned-off, that is, nodes that have lower "negative" characteristics.
6. Overlay all local regions L_i onto the SOM map, and thus identify those L_is that can be cordoned-off or locked-down based on their relative risk profile.

Notice that this process and the SOM map is a predictive, decision-making tool that aids decision making. Decisions are ultimately made by the people responsible for making such decisions. It is likely that some nodes on the SOM map may have higher values for some "negative" characteristics than other nodes and vice versa, hence deciding if either, or both nodes should be locked-down is likely to be a judgement call for the decision makers, who can then potentially support their decisions with other initiatives.

4.2.1 | Selecting and defining characteristics

The set C comprises of six characteristics (defined below) which are based on similar constructs applied in References [4,8-10,27]:

1. Vulnerability: the number of vulnerable people as a ratio of the total population of the region. COVID-19 disproportionately impacted people 60 years and older. Hence

TABLE 1 List of characteristics (a), modelling assumptions (b) and relative sources

(a) Data gathering
Characteristics
Age by population—retrieved from "House of Commons Library" at Reference [35]
Commuters—retrieved from Reference [36]
Primary School Mobility for England; retrieved from Reference [37]
Primary School Mobility for Scotland and Wales—assumed to be 0.01 and 0.013 (lower than all English constituencies)
Population density—retrieved from Reference [38]
Houses per constituency—retrieved from Reference [39]
Gross value added (GVA)—retrieved from Reference [40]
(b) Modelling assumptions
Assumptions
R_0 at Day 1—taken from Reference [13]
Sus_0 at Day 1—assumed to be 0.999998
Inf_0 at Day 1—assumed to be 0.000002
Rec_0 at Day 1—assumed to be 0 (no one recovered as yet)
γ—assumed to be 1/7 (infection is expected to lasts for 7 days[41,42])
Infected fatality rate (IFR)—assumed to be 0.7% of all people infected[13,41]
2.4% critical-care hospitalisation of infected people—value extracted from Reference [41]
R_0 at Day 24 and Day 39—assumed based on findings in Reference [13]

vulnerability in this study was defined as the number of people 60 years and older in the region/total population of the region.
2. Population density: the number of people per square kilometre in the region.
3. Commuter mobility: the absolute difference between the number of jobs available in the region and the number of jobs done by the residents of the region.
4. School mobility: average of the number of primary school children (aged 4 to 11 years) attending school outside of their local authority as a ratio of total number of primary school children in the local authority.
5. People per house: total number of people in the region divided by the total number of houses in the region.
6. Economic output: Gross value added (GVA) of the region, which measures the value of the goods and services produced in a region.

The exact nature of these characteristics are not essential for the objectives of this study, and could be easily adapted or substituted based on expert knowledge. Nevertheless, we realise that there are limitations with our selection of parameters, the largest of which is the ability to obtain the most accurate and up to date data from publicly available sources. In this regard, we believe that this issue can be overcome if this approach is adopted by the various governmental structures within the UK, as they are most likely to have detailed access to the required data.

Other limitations include the assumption that people typically live close to their place of work, and commuter mobility is driven by excess demand or supply and not a mismatch between place of work and residence. We also consider aggregate population density and people per household as reasonable measures to describe the "compactness" of the population. We do acknowledge that the true population density can vary if there are dense pockets in otherwise sparsely populated areas, for example, little villages in farming communities. We also acknowledge that people per house may be understated, particularly in areas where there is a high concentration of houses that are typically not used as primary residences, for example, holiday homes in seaside towns, and investment properties in some high-value areas in London. Here again, several of these variables can be easily adjusted once more accurate data is made available. For completeness, we provide a full list of our assumptions and data sources in Tables 1a and 1b.

4.2.2 | Modelling: Software, algorithms and parameters

We created a computer program, using the collected data as input, to generate a 6 by 6 SOM and assign the SOM groupings to a GB constituency map, using the Kohonen and Parlitools packages in R programming language.[34,43] The data set and the broad layout of the R computer code is provided in Reference [44].

4.2.3 | Evaluating the model effectiveness

We demonstrate the effectiveness of our approach by: (1) comparing it with real life events relating to the progress of the pandemic in the UK since March 2020 as well as with events since June 24, 2020, when we first made our model publicly available,[45] and (2) comparing the SOM approach with Fuzzy C-means clustering (FCM).

Comparing our approach with real life events

As part of this comparison we proposed the use of protective cordons and compare this approach against the standard "SIR" model described in Reference [14], and which formed the basis of the UK government's COVID-19 briefings since March 2020.[12,13] In this regard we model the UK government's current approach using the "SIR" model based on publicly available data provided by the government including estimated cumulative total deaths and approximate date of the infection peak. We used the basic "SIR" program code as detailed in Reference [46] to process the data. The following modelling assumptions, some of which based on details provided by recent studies and the government in its daily briefings,[12,13] were made (see Table 1b):

1. *Day 1* is assumed to be March 1, 2020, the approximate date of the first person to person transmission within the UK.
2. The lockdown officially commenced on 24 March 2020, with total freedom of movement within the UK being possible before this date.
3. As per the findings in Reference [13], the initial reproducibility rate is assumed to be $R_0 = 3.6$.
4. The initial infected proportion is assumed to be $Inf_0 = 2 \times 10^{-6}$.
5. The initial susceptibility proportion is assumed to be $Sus_0 = 1 - 2 \times 10^{-6}$.

6. The initial recovery proportion is assumed to be zero, $Rec_0 = 0$.
7. Consistent with the findings in References [13,41,42], the illness lasting for an average of 7 days indicates that the mean recovery rate, γ, can be assumed to be 1/7.

We assume that consistent with the modelling assumptions in Reference [13], R_0 declines dramatically from Day 24 due to the lockdown. We also assume that 14 days later the R_0 value declines slightly as infected households, that have been self-isolating in line with government guidelines, would have recovered and unlikely to infect others.[13,41,47] These three modelling scenarios are compared with a model of the estimated current government plan, the complete lockdown of all local regions, irrespective of their individual set of characteristics, C, or their relative pandemic self-sustainability indicated by their position on P.

Comparing the SOM approach with FCM clustering

We compare our SOM approach with the popular FCM clustering approach that has been widely used for clustering unsupervised data.[15] A detailed theoretical approach and algorithm for FCM is not provided as part of this study, however the approach taken in Reference [15] was leveraged. The "c-means" function was used in R software and incorporated into our computer program which may be accessed in Reference [44]. The results of the FCM approach were compared with the SOM based on two criteria (see Section 5.2.3): (1) ease of interpretation of results, and (2) effectiveness of clustering (i.e., grouping similar items). In the case of interpreting the results, we note that in most cases the end-users (including governments, emergency services, etc.) may not have the time and/ or the technical background to analyse detailed numerical data and formulate an action plan. Hence, having an output that is easy to interpret and intuitive is important as it could accelerate decision making and potentially save lives.

5 | RESULTS AND DISCUSSION

5.1 | Identifying regions

The results of the SOM analysis are provided in Figures 1–3. From Figure 1, it can be seen that the algorithm groups regions with a high population density, and mobility in the lower-left quadrant of the SOM, while regions with high vulnerability are grouped together in the upper-right quadrant. When this data is layered onto the GB constituency map, as shown in Figure 2, it is clear that areas surrounding London and other major metropolitan areas map to the lower-left quadrant of Figure 1 where the population density and mobility is high. It is not surprising, given the close contact of people, that these areas were initially highly impacted by the spread of COVID-19 as noted in March 2020 and were "hotspots" for outbreaks of mutations as noted in December 2020.[47,48] Further, based on the concept of pandemic self-sustainability, these local regions may be considered to be located at points along P where the values of several of the "negative" characteristics are high and thus not ideal candidates for protective cordons.[10,27] At the same time, regions in the upper right quadrant of Figure 1 contain a disproportionately large population of vulnerable people. While these areas may benefit from a protective cordon, the risk of infection, overwhelming the local healthcare system, and consequential death rates may be considered too high, which suggests that a lockdown or self-isolation may be better solutions for these geographic regions. The nodes enclosed within the red perimeter contain several regions that may be ideal candidates for a protective lockdown with free movement

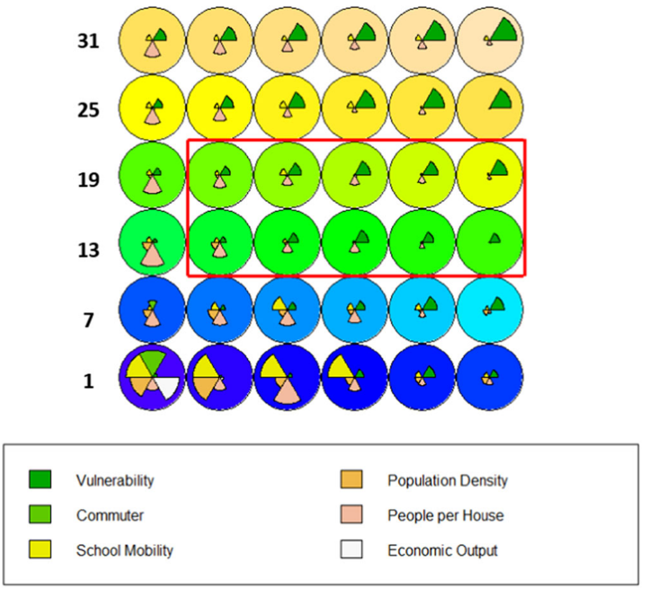

FIGURE 1 Self-organising maps grouping of UK constituencies based on self-sustaining factors [Color figure can be viewed at wileyonlinelibrary.com]

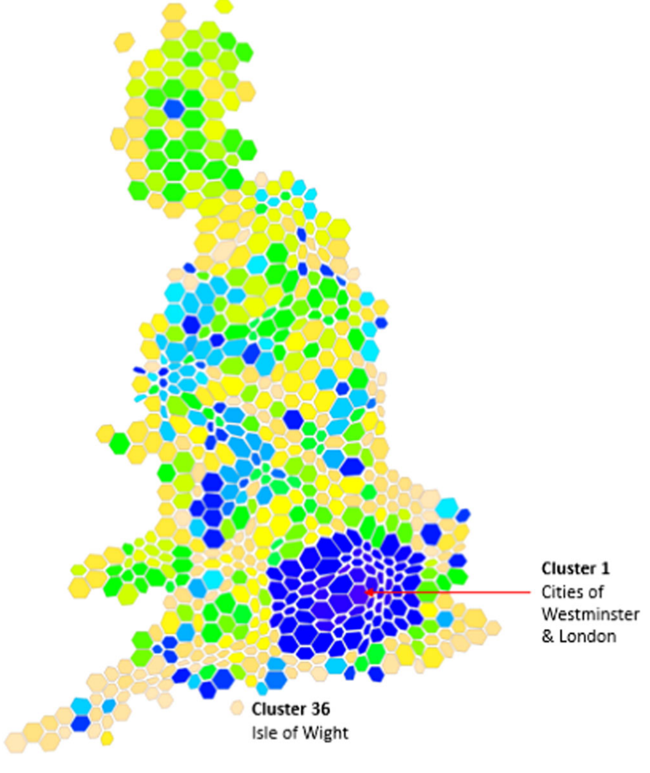

FIGURE 2 Great Britain constituency coding based on self-organising maps analysis [Color figure can be viewed at wileyonlinelibrary.com]

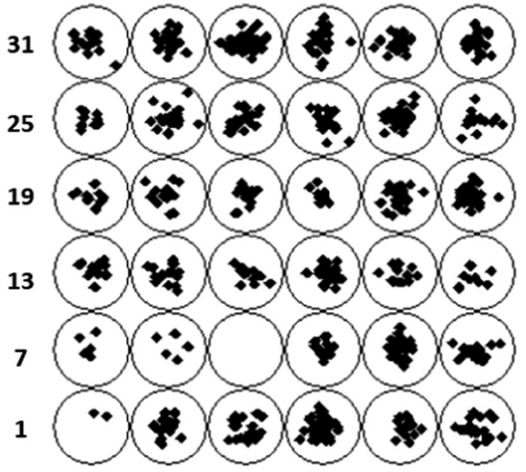

FIGURE 3 Distribution of UK constituencies based on self-organising maps groups

inside their respective boundaries. These regions may be considered to be located at points along *P* where the values of most, if not all, of the "negative" characteristics are low and where the values of the "positive" characteristics may be high. Thus, comparing the data for Figures 2 and 3, it is likely that approximately 25% of GB constituencies fall into the group that could be cordoned-off and allowed to be economically active. Further, based on Figure 2, some of these constituencies may be geographically spread across the country and as a result these protective cordons will likely create pockets of economic activity across the country and not just in a single region. This may also lay the foundations for the country to "ease out of the pandemic" as the perimeter of these protective cordons could gradually be extended to include neighbouring areas when their infection rates decline. Over time, these perimeters may continue to expand outwards and intersect with perimeters of other regions that are being cordoned off, thus creating significantly larger pockets of economic activity. Eventually, these pockets will likely encompass the majority of the country, and may be used in conjunction with the vaccine roll-out, which would return the country to a sense of normality.

5.2 | Evaluating effectiveness

5.2.1 | Comparing our approach with real-life events

The intention of the UK government in March 2020 was to "flatten, and stretch the peak" to save lives and protect the health service from being overwhelmed.[12] Figure 4 shows this *No Intervention* "business as usual" model, with an assumed infected fatality rate (IFR) of 0.7% (consistent with conclusions in References [13,41]) together with the modelling assumption provided in Section 4.2.3, would have resulted in over 460,000 deaths. This is broadly consistent with the UK government's early projections of up to 500,000 deaths and an infection rate of over 80%.[12,47] The peak of the infection would be on *Day 40* with approximately 36.6% of the total population being infected at that point in time. No doubt this would have overwhelmed the UK's National Health Service (NHS), which is free at the point of use for almost all of its approximately 67 million residents.

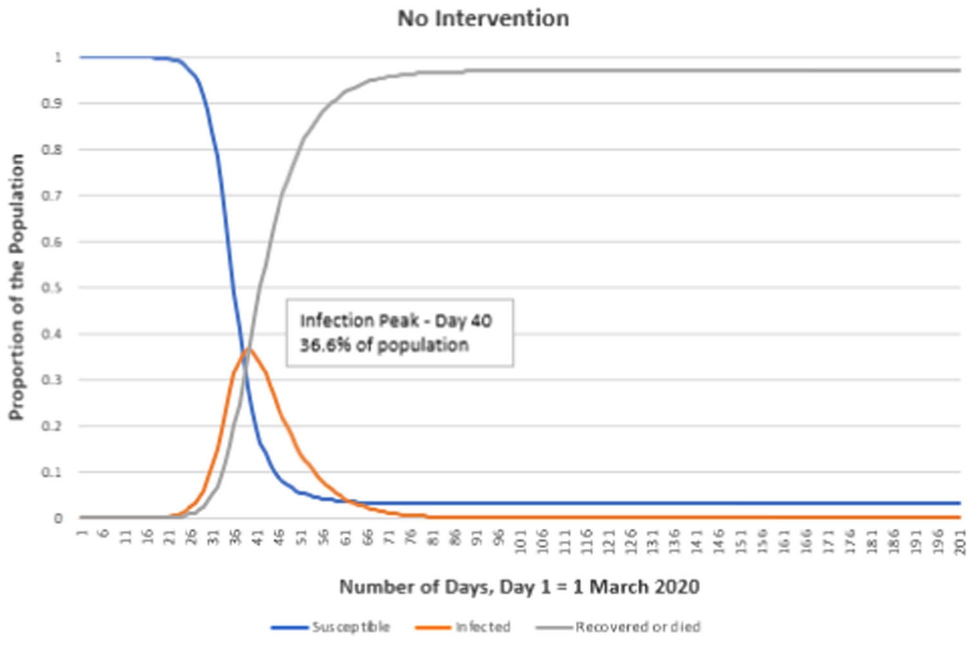

FIGURE 4 No intervention [Color figure can be viewed at wileyonlinelibrary.com]

The actions taken by the UK government to save lives and alleviate the pressure on the NHS, while providing adequate healthcare to its residents was to impose a widespread lockdown from March 24, 2020 with an estimated R_0 value of 1.05. Their messaging was "Stay at Home, Save Lives, Protect the NHS".[5] As a result of the lockdown the economic activity of the majority of UK residents were severely impacted, while the infection curve was both flattened and stretched with the new infection peak around *Day 52*, with a total of 1.14% of the population being infected. Figure 5 shows the estimated *Government Plan*. Note that this model assumes a sustained lockdown. Under this plan, the economy shrank a record 20% in April 2020, in line with the Bank of England and economists' projections.[2,49] In a sustained lockdown, we projected that the estimated UK government model would reach equilibrium after 325 days at which point approximately 19.6% would have been cumulatively infected.

For the alternative approach, we propose a strong perimeter around cordoned areas allowing only vital movement of people and goods in and out, testing facilities for all that need it including using temperature sensors in public places (e.g., malls, large stores, etc.), enforced isolation of the infected, mandatory face coverings in public, and social distancing and regular hand sanitation being encouraged. We propose cordoning-off 25% of selected GB parliamentary constituencies and locking down the rest from Day 24 in line with the model output shown in Figure 1. The assumed R_0 values for the cordoned-off constituencies at *Day 1* is 1.05 as these areas start off with low infection rates. As shown in Figure 6, the peak infection rate occurs on Day 39, with 1.13% of the population being infected at that time, and the demand for critical-care hospitalisation beds being approximately 17,800. Under this alternative, a total of 16.0% of the population would likely be infected by Day 325. The computation of critical-care hospitalisation beds was obtained using the approach in Reference [42] which found that the median age of hospitalisations in the UK was 72, with mean hospitalisation duration of 7 days. This 17,800 demand is well within the capacity of the NHS especially given the NHS Nightingale

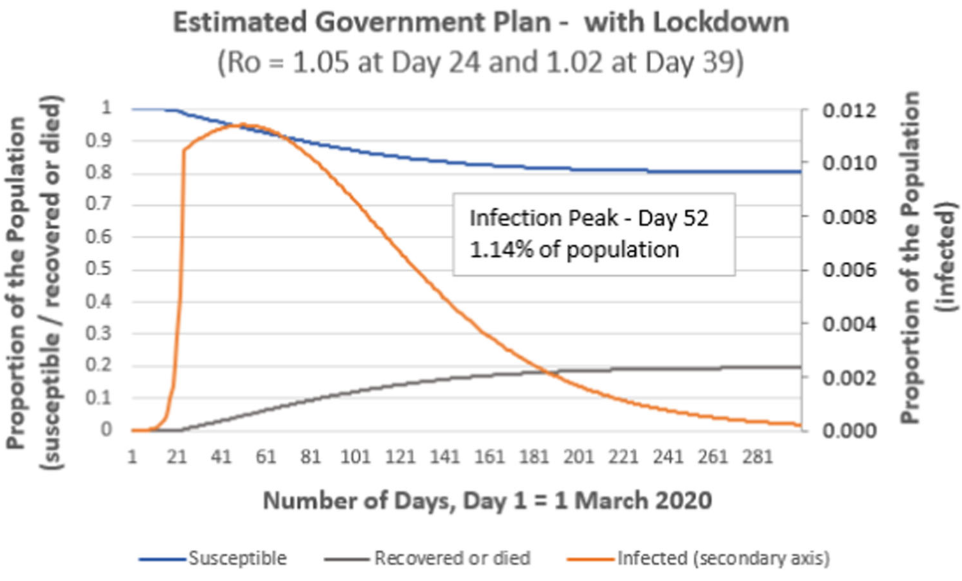

FIGURE 5 Estimated Government's current model [Color figure can be viewed at wileyonlinelibrary.com]

operation in which over 11,000 critical-care beds were added within 4 weeks of the lockdown being announced.[50]

5.2.2 | The UK government's "tiered" system and protective cordons

The UK government introduced the "tiered" system, and later strengthened it, to contain localised outbreaks.[51] Every part of England was placed into a tier based on the prevalence of

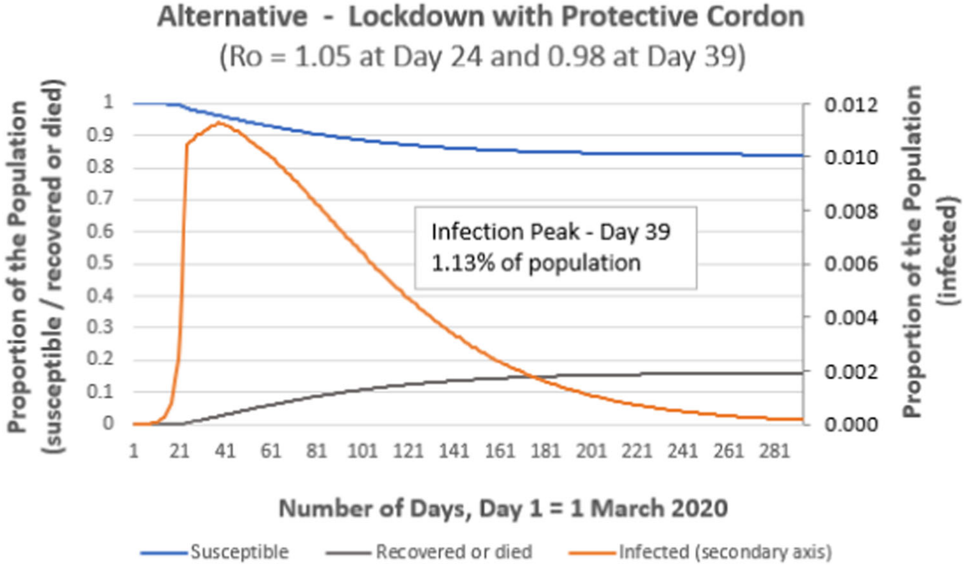

FIGURE 6 Proposed alternative model [Color figure can be viewed at wileyonlinelibrary.com]

COVID-19, with Tier 4 being the most stringent. Differing rules applied to the various tiers, and as part of this approach, mobility across tiers was strongly discouraged, with police intervening and issuing fines to those that broke the rules.[51] There are indeed several similarities to our proposed alternative approach of protective cordons. To a large extent, the "tiered" system was considered to be a failure, driven largely by the lack of proper enforcement, notably around preventing inter-tier mobility. Another area that was considered to be a failure was the effectiveness of the "track and trace" system which proved to be ineffective in supporting the enforcement of quarantine and isolation measures of those infected, or presumed to be infected.[51,52] We believe that these inefficiencies would not occur in our alternative proposal which is predicated on proper enforcement notably around mobility and isolation.

Given this, we conclude that a strong protective cordon, with lockdown in highly impacted areas during a pandemic, may be better than a full lockdown in terms of controlling the spread of the disease and protecting the economic and social integrity of a country. The combined protective cordon and lockdown approach enables some of the regional economies to continue with minimal disruption, while allowing for a lower-risk exit strategy where the perimeters of the protective cordons can gradually be expanded to encompass neighbouring regions when it becomes safe to do so.

5.2.3 | Comparing the SOM approach with FCM clustering

The FCM model was run using the same data set and the process outlined in Section 4.2.3 with the detailed results available in Reference [44]. We used the clustering of the London parliamentary constituencies (73 out of the 632) and the 10 most vulnerable parliamentary constituencies as a measure of clustering effectiveness. This measure of efficacy was considered relevant as London was highly impacted by the pandemic given its size and population density while people in vulnerable constituencies needed shielding from visitors and tourists as outlined in Reference [48]. Hence, it was important that a model was suitably able to distinguish between these groups. The results of the efficiency comparison are presented in Table 2. From Table 2, it is clear that both the SOM and FCM approaches are equally as effective in clustering without overlap both highly impacted areas as well as highly vulnerable areas.

The second element of the comparison between the SOM and FCM approaches was the ease of interpretation of results. As noted in Section 2.1.2, there several benefits of SOM over other methods including FCM, such as preserving the topology of the data and powerful visual representations. This has been clearly demonstrated in this study as outlined in Figures 1–3 as

TABLE 2 Comparison of SOM and FCM approaches

Characteristics	SOM	FCM
(a) London		
Cluster number(s) containing London constituencies	1, 2, 3, 4	4, 7, 12, 30
Do these clusters contain vulnerable constituencies?	No	No
(b) Top 10 Vulnerable Constituencies		
Cluster number(s) containing Top 10 vulnerable constituencies	36	16
Do these clusters contain higher-risk constituencies?	No	No

well as in Table 2. In Figure 1 the output of the SOM was easily overlaid on the map of GB thus allowing nontechnical end-users to easily identify the locations of highly impacted geographical regions. This cannot be easily achieved with FCM as it does not preserve data topology. This is evident in Table 2 where the cluster numbers in SOM increase chronologically with cluster 1 being highly impacted while cluster 36 being highly vulnerable. In contrast, the cluster numbers for FCM appear to be random, and will thus require postprocessing to obtain a structure similar to that of the SOM. Hence, it can be concluded that SOM is easier to use and more intuitive than FCM.

6 | CONCLUSIONS

A natural disaster and pandemic prediction and containment model based on SOM was developed. In terms of predicting disasters, the model could be used to identify areas that are likely to be susceptible as well as those areas the could be self-sustaining and thus could be allowed to continue to operate under relatively normal conditions or be used to support impacted regions. SOM were used to identify regions that could be ring-fenced, either to be protected, or locked-down, or serve as a support centre. This was based on a set of publicly available variables and sustainability theory. A case study based on the UK's response to the COVID-19 pandemic was used as a basis to test and demonstrate the effectiveness of our proposal. Tests conducted using publicly available data for GB showed that it is possible to ring-fence approximately 25% of GB's parliamentary constituencies to operate relatively normally within a defined protective cordon in line with our proposed alternative strategy. Our proposed alternative showed that it is possible to have a peak infection rate of 1.13% at Day 39 versus an estimate of the government's current plan which peaked at Day 51 at 1.14% during the first lockdown. Further, the required additional hospital capacity of 11,000 critical-care beds through its NHS Nightingale operation, was sufficient to cater for the peak under proposed approach.[50] Comparative tests between SOM and FCM approaches showed that while both approaches performed equally well in clustering efficiency, the SOM was more easier to use in that it offered the nontechnical end-user an easy to use and intuitive output without the need for additional postprocessing.

6.1 | Summary of theoretical and practical implications

There are several implications of this study, notably its applicability in future natural disaster and pandemic prediction and containment strategies. Governments across the world continue to evaluate their COVID-19 measures notably the efficiency of their current lockdown approaches, particularly within the context of the economy, and are looking for approaches that predict "flare-ups," optimise disease containment, and accelerate the return to "business as usual." We believe that this approach provides a viable alternative for governments to explore for this and future natural disasters and pandemics.

From a theoretical perspective, this study highlights the universal applicability of SOM. The approach used in this study focused on developing a solution that was both easily adaptable and simple to use. As a result, this study provides a strategy that can be adapted for other scenarios, beyond epidemiology, with a flexible data structure that allows for data parameters to be easily adjusted as required.

6.2 | Limitations

We have highlighted some limitations with regard to the selection of parameters in Section 4.2.1. We believe that the primary limitation with this approach is the adequate collection of the required data. Our approach and model is flexible enough to cater for most applications, and could prove to be a useful prediction and/or decision making tool given adequate, reliable data.

6.3 | Future work

Future work for potential end-users of this study, for example, governments, will be to adequately define the parameters of greatest interest to their application and proceed to collect timely and accurate data. Through this study, we have realised the tremendous potential for simple, easy to use SOM models that improve the outcomes for society. We have also realised the benefits of having and using these models as early as possible, notably the potential to save lives and preserve economic activity. As a result, we will continue to extend this approach and study to other applications that, like pandemics, have the potential to be hugely disruptive. Some of these applications include disaster and emergency planning, crime prevention, and cybersecurity breaches.

ACKNOWLEDGEMENTS

All authors made significant contributions throughout this piece of research and agreed to submit the manuscript in the current form. The first author made major contributions in terms of writing and implementing software. All the authors contributed in terms of conceptualisation, writing, and revising the manuscript. All authors have read and agreed to the published version of the manuscript.

ORCID

Raymond Moodley https://orcid.org/0000-0003-4471-2272
Francisco Chiclana https://orcid.org/0000-0002-3952-4210
Fabio Caraffini https://orcid.org/0000-0001-9199-7368
Mario Gongora https://orcid.org/0000-0002-7135-2092

REFERENCES

1. Forbes, ed. How AI can and will predict disasters; 2019. https://www.forbes.com/sites/cognitiveworld/2019/03/15/how-ai-can-and-will-predict-disasters/?sh=4fc4d1cf5be2/. Accessed December 28, 2020.
2. Time, ed. Bank of England predicts british economy to shrink by most since 1706; 2020. https://time.com/5833533/uk-economy-coronavirus/. Accessed December 28, 2020.
3. BBC, ed. Coronavirus: pandemic sends US jobless rate to 14.7%; 2020. https://www.bbc.co.uk/news/business-52591262/. Accessed December 28, 2020.
4. Currie CS, Fowler JW, Kotiadis K, et al. How simulation modelling can help reduce the impact of COVID-19. *J Simul*. 2020;14(2):1-15.
5. ITV, ed. Government launches new coronavirus advert with stay at home or 'people will die' message; 2020. https://www.itv.com/news/2020-04-02/stay-at-home-or-people-will-die-government-launches-new-coronavirus-ad-blitz/. Accessed December 28, 2020.
6. Florez-Lozano JM, Caraffini F, Parra C, Gongora M. A robust decision-making framework based on collaborative agents. *IEEE Access*. 2020;8:150974-150988.

7. Erhard A, Wallman JJ, Postler L, et al. Characterizing large-scale quantum computers via cycle benchmarking. *Nat Commun*. 2019;10(1):1-7.
8. Basara HG, Yuan M. Community health assessment using self-organizing maps and geographic information systems. *Int J Health Geographics*. 2008;7(1):67.
9. Mutheneni SR, Mopuri R, Naish S, Gunti D, Upadhyayula SM. Spatial distribution and cluster analysis of dengue using self organizing maps in Andhra Pradesh, India, 2011-2013. *Parasite Epidemiol Control*. 2018; 3(1):52-61.
10. Park M, Kim Y, Lee Hs, Han S, Hwang S, Choi MJ. Modeling the dynamics of urban development project: focusing on self-sufficient city development. *Math Comput Modell*. 2013;57(9-10):2082-2093.
11. Pearce JL, Waller LA, Mulholland JA, Sarnat SE, Strickland MJ, Chang HH, et al. Exploring associations between multipollutant day types and asthma morbidity: epidemiologic applications of self-organizing map ambient air quality classifications. *Environ Health*. 2015;14(1):55.
12. BBC, ed. Coronavirus: what is the R number and how is it calculated?; 2020. https://www.bbc.co.uk/news/health-52473523. Accessed December 28, 2020.
13. Flaxman S, Mishra S, Gandy A, et al. Report 13: estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries; 2020. https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-13-europe-npi-impact/
14. van den Driessche P. Reproduction numbers of infectious disease models. *Infectious Disease Modelling*. 2017;2(3):288-303.
15. Moodley R, Chiclana F, Caraffini F, Carter J. A product-centric data mining algorithm for targeted promotions. *J Retail Consumer Serv*. 2019.
16. Dell'Acqua F, Gamba P, Jaiswal K. Spatial aspects of building and population exposure data and their implications for global earthquake exposure modeling. *Natural hazards*. 2013;68(3):1291-1309.
17. Teves-Costa P, Mendes JM. *Natural Hazards, Modelling, Risk Assessment and the Role of Scale*. Springer; 2018.
18. Epstein JM. Modelling to contain pandemics. *Nature*. 2009;460(7256):687.
19. Araz OM. Integrating complex system dynamics of pandemic influenza with a multi-criteria decision making model for evaluating public health strategies. *J Syst Sci Syst Eng*. 2013;22(3):319-339.
20. Uhlig S, Nichani K, Uhlig C, Simon K. Modeling projections for COVID-19 pandemic by combining epidemiological, statistical, and neural network approaches. medRxiv; 2020.
21. Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. arXiv preprint arXiv:200310849; 2020.
22. Wang L, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. arXiv; 2020: arXiv-2003.
23. Rico FA. Emergency department capacity planning for a pandemic scenario: nurse allocation; 2009. http://scholarcommons.usf.edu/etd/2163
24. Johnsson M. Applications of self-organizing maps. BoD-Books on Demand; 2012.
25. Yin H. The self-organizing maps: background, theories, extensions and applications. In: *Computational Intelligence: A Compendium*. Springer; 2008: 715-762.
26. Tan PN. *Introduction to Data Mining*. Pearson Education India; 2018.
27. Grewal SS, Grewal PS. Can cities become self-reliant in food? *Cities*. 2012;29(1):1-11.
28. Barbosa JA, Bragança L, Mateus R. New approach addressing sustainability in urban areas using sustainable city models. *Int J Sustain Building Technol Urban Devel*. 2014;5(4):297-305.
29. The-Conversation, ed. Coronavirus: we're in a real-time laboratory of a more sustainable urban future; 2020. https://theconversation.com/coronavirus-were-in-a-real-time-laboratory-of-a-more-sustainable-urban-future-135712. Accessed December 28, 2020.
30. Thomson-Reuters-Foundation, ed. OPINION: post-pandemic cities—how can we use the lessons of lockdown?; 2020. https://news.trust.org/item/20200521114134-6wx0e. Accessed December 28, 2020.
31. Sky-News, ed. How Coronavirus is spreading around the world; 2021. https://news.sky.com/story/coronavirus-how-covid-19-is-spreading-around-the-world-12061281. Accessed: January 13, 2021.
32. Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. *AI Magazine*. 1996;17(3):37.
33. Moodley R, Caraffini F, Chiclana F, Gongora M. Data & code for "living through pandemics—using protective cordons to enhance continuity." De Montfort University; 2020. https://figshare.dmu.ac.uk/

articles/dataset/Data_Code_for_Living_through_Pandemics_-_Using_Protective_Cordons_to_Enhance_Continuity_/12477515/1
34. Wehrens R, Buydens LM. Self-and super-organizing maps in R: the Kohonen package. *J Stat Software*. 2007;21(5):1-19.
35. UK-Parliament, Constituency data: population, by age; 2020. https://commonslibrary.parliament.uk/local-data/constituency-statistics-population-by-age/. Accessed December 28, 2020.
36. ONS. Earnings and hours worked, place of work and place of residence by Parliamentary constituency: ASHE Tables 9 and 10; 2019. https://www.ons.gov.uk/datasets/ashe-tables-9-and-10/editions/time-series/versions/1. Accessed December 28, 2020.
37. UK-Government. Secondary and primary school application and offers: 2019; 2019. https://www.gov.uk/government/statistics/secondary-and-primary-school-application-and-offers-2019. Accessed December 28, 2020.
38. ONS. Population density by constituency for 2015; 2019. https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/populationdensitybyconstituencyfor2015. Accessed December 28, 2020.
39. UK-GOV. Live tables on dwelling stock (including vacants); 2020. https://www.gov.uk/government/statistical-data-sets/live-tables-on-dwelling-stock-including-vacants. Accessed December 28, 2020.
40. ONS. Gross Value Added (GVA); 2020. https://www.ons.gov.uk/economy/grossvalueaddedgva. Accessed December 28, 2020.
41. EU, ed, Coronavirus disease 2019 (COVID-19) in the EU/EEA and the UK—eighth update; 2020. https://www.ecdc.europa.eu/sites/default/files/documents/covid-19-rapid-risk-assessment-coronavirus-disease-2019-eighth-update-8-april-2020.pdf. Accessed December 28, 2020.
42. Docherty AB, Harrison EM, Green CA, et al. Features of 16,749 hospitalised UK patients with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol. medRxiv; 2020.
43. Evan-Odell. Parlitools: Tools for analysing UK politics in R. R package version 0.4.1; 2017. https://cran.r-project.org/package=parlitools
44. Raymond-Moodley, Github, ed. Data and code repository—living with pandemics; 2020. https://github.com/Raymondmood/Living-through-Pandemics. Accessed June 02, 2020.
45. ITV, ed. Researchers at De Montfort University say lockdown could have been avoided for around a quarter of the UK; 2020. https://www.itv.com/news/central/2020-06-23/researchers-at-de-montfort-university-say-lockdown-could-have-been-avoided-for-around-a-quarter-of-the-uk/. Accessed December 28, 2020.
46. Keeling MJ, Rohani P. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press. 2011.
47. UK-Government, ed. Coronavirus (COVID-19) in the UK; 2020. https://coronavirus.data.gov.uk/#category=utlas&map=rate. Accessed December 28, 2020.
48. BBC, ed. New coronavirus variant: What do we know?; 2020. https://www.bbc.co.uk/news/health-55388846/. Accessed December 28, 2020.
49. BBC, ed. Bank of England'ready to act' as economy shrinks record 20%; 2020. https://www.bbc.co.uk/news/business-53019360/. Accessed December 28, 2020.
50. BBC, ed. Coronavirus: Birmingham and Manchester temporary hospitals announced?; 2020. Accessed December 28, 2020.
51. New-Statesman, ed. How Boris Johnson's government lost control of Covid-19 again; 2020. https://www.newstatesman.com/politics/uk/2020/12/how-boris-johnson-s-government-lost-control-covid-19-again. Accessed January 13, 2021.
52. Frowde R, Dove ES, Laurie GT. Fail to prepare and you prepare to fail: the human rights consequences of the UK government's inaction during the COVID-19 pandemic. *Asian Bioethics Rev*. 2020;12(4):459-480.

How to cite this article: Moodley R, Chiclana F, Caraffini F, Gongora M. Using self-organising maps to predict and contain natural disasters and pandemics. *Int J Intell Syst*. 2022;37:2739-2757. https://doi.org/10.1002/int.22440