Towards a Data-Driven Fuzzy-Geospatial Pandemic Modelling

Amir Pourabdollah and Ahmad Lotfi
School of Science and Technology
Nottingham Trent University
Nottingham, NG11 8NS, UK
amir.pourabdollah|ahmad.lotfi@ntu.ac.uk

***Abstract*—The current Covid-19 worldwide outbreak has many lessons to be learned for the future. One area is the need for more powerful computational models that can support making better decisions in controlling future possible outbreaks, particularly when being made under uncertainties and imperfections. Motivated by the rich data being daily generated during the pandemic, our focus is on developing a data-driven model, not primarily relying on the mathematical epidemiology techniques. By investigating the implications of the current pandemic data, we propose a fuzzy-geospatial modelling approach, in which uncertainties and linguistic descriptions of data, some of which being geo-referenced, are handled by non-singleton fuzzy logic systems. In this paper, we outlining a conceptual model designed to be trained by the available pandemic worldwide data, and to be used to simulate the effect of an enforced controlling measure on the geographical extent of the infection. This can be considered as an uncertain decision support systems (UDSS) in controlling the pandemic in the future outbreaks.**

***Index Terms*—Fuzzy Systems, GIS, Pandemic Models**

I. Introduction

The world has just witnessed the huge impact of the local/national measures introduced in response to the Coronavirus spread, some of which could be made more wisely or at a better time in different countries. This highlights the need for developing advanced modelling techniques to simulate the patterns of the infection spread given the available data, even if not completely certain and trusted.

There are several approaches in modelling the epidemiological information in order to provide close predictions of how infectious diseases spread in communities, as well as supporting the decision-makers. Mathematical epidemiology, backed by about 100 years of history and research, can provide predictive models of disease spread based on specific parameters of a disease such as the reproduction number (R) and population attributes. Examples are the classic SIR model (Susceptible, Infected and Removed) that is widely used in recent outbreaks [1], and recent Covid-19-specific models such as Bats-Hosts-Rseservoir-People (BHRP) and Reservoir-People (RP) transmission network models [2].

The current Covid-19 pandemic, as being one of the greatest disease pandemics in recent years, has been the source of one of the richest data of its kind. Therefore, it is a highly motivating case for developing *data-driven* pandemic models for future similar cases. In the data-driven approach, models are developed and trained by relying on the collected data during the outbreak, rather than on the epidemiological mechanics of the outbreak. In this case, tuning the epidemiological attributes for running a mathematical model is replaced by a reverse-engineering approach, i.e., training the model by the big data collected *after* the pandemic, not by the epidemiological attribute known *before* it.

A challenge in such a data-driven approach lies in the considerable amount of uncertainties naturally or artificially embedded in the released pandemic data. A known example to the world is the inaccuracies in the announced numbers of Covid-19 cases due to the shortage of test kits in many countries. The current data modelling approaches are rarely able to systematically digest the uncertainties into the model. Currently, the uncertainties are usually treated as noise data that deviate the developed model results from close predictions.

Here is where fuzzy systems may act better than the rigid quantised analysis, as they have already shown their capabilities in such scenarios. In fuzzy systems, the uncertainties are naturally embedded in the "computing with words" paradigm. For example, an area can be called "highly infected", or an individual can be classified as being in "low risk" of having the infection, each with different membership grades. There are some research works on fuzzy modelling of epidemiological phenomena [3], [4], however to the best of our knowledge, none of them are trained or tested against a worldwide data collection similar to the Covid-19.

The other challenge is that most of the fuzzy logic systems are based on serialised or single-dimensional numeric data. Pandemic data, on the other hand, are usually multi-dimensional, based on 2-D or 3-D maps, and/or incorporating temporal dimensions. It seems that a fuzzy-spatio-temporal approach is required for developing future models to deal with the full range of pandemic data types.

We aim at developing a data-driven fuzzy-geospatial pandemic models. Rather than relying on the classical epidemiological modelling, our approach does not much care about the mechanics of the infection spread. Instead, by taking advantage of the Covid-19 data availability, it treats the pandemic as a system with multiple inputs and outputs, governed by a knowledge-base trained by the historical data. To fit the model to its specific requirements, it should be able to handle uncertain data, some of which having spatio-temporal nature. That is why the contribution of this paper is proposing a fuzzy-geospatial model of the pandemic.

After outlining a conceptual model for future developments, we propose a simplified model as a proof of concept that can deal with limited data taken from the recent Covid-19 spread in an exemplar county in the UK.

II. Background

In this section, some of the background concepts and implications for the outlined pandemic model are reviewed. Particularly, it is useful to explore the implications of two specific aspects of the pandemic data, namely its embedded uncertainties and its geospatial nature.

A. The Uncertain Nature of the Pandemic Data

The embedded uncertainties in pandemic data have already shown the world, how important is incorporating the uncertainties in modelling and to develop uncertain decision support systems (UDSS). A very bold example of such uncertainties is the fact that testing has not been always available for Coronavirus infection. Different countries have struggled to provide enough test kits, leaving the patients with the symptoms, the communities and the decision-makers under doubt about the real number of the disease cases. The daily statistics are normally the confirmed cases but no one knows what has been the real number of the infected cases. The number of tests per day per 1000 patients, if correct and if available, can improve the model and enhance the prediction results by telling the model how wide is the uncertainty.

Another uncertainty source can be the data non-transparencies applied in different countries, at different times and under different policies. Moreover, when a policy measure is applied by governments, the extent of enforcing it and the level in which people in different cultures obey the rules, add uncertainties to the released data. Besides, the disease maps provided by national or international agencies have to be treated as fuzzy, since the viruses do not care about concepts such as human-made rigid country borders.

A fuzzy logic system can translate these uncertainties into the underlying fuzzy sets. Therefore, any model that predicts the future spread based on the current statistics, has to account for the invisible uncertainties. Namely, non-singleton and type-2 fuzzy systems are the more specialised typed of the fuzzy systems that can efficiently capture the uncertainties at their source and digest them in producing the close predictions.

B. The Geospatial Nature of the Pandemic Data

The majority of fuzzy logic systems work with scalar concepts. Recently, fuzzy Geographic Information System (GIS) systems are developed as being rule-based fuzzy logic systems specialised for spatial information in two or three-dimensional environments. If positions can be treated as fuzzy concepts, the uncertainties embedded in human understanding of the positions or the position's associated attributes can be efficiently modelled by fuzzy sets [5].

The idea of fuzzy-GIS is not new and there is an ongoing research on uncovering the range of possibilities and scenarios of applying fuzzy positioning in different intelligent systems. Fuzzy systems has been used for spatial decision support systems, such as in [6]. Some of the current applications of fuzzy-GIS are in agriculture [7], urban design [8], photogrammetry [9] and particularly in healthcare [10], [11]. However, the applications of fuzzy-GIS are still limited, partly due to its more complex underlying mathematics compared to the mainstream fuzzy systems. This is particularly the case for fuzzy topological relations [12], [13], Region Connection Calculus (RCC) [14] and multi-criteria fuzzy search [15].

It is noticeable that fuzzifying the geospatial values does not stop precisely measuring them, nor working with a position as values, however fuzzy-GIS opens a wide new area of applications by including position to another type of variables that can be treated both as values or as fuzzy terms. For example, we may be interested to provide ambulance services for a patient if *the area is highly infected*. This is, in fact, a fuzzy rule that comes from a humanitarian understanding of the environment, rather than a rule that comes from measured numbers. However, to implement such a rule, we would certainly need to measure the position of the individual as precisely as possible since the numbers are ultimately used by computers to make the decisions. In between, we will convert the measured numbers to some membership grades in some fuzzy sets, in which the linguistic terms such as "highly infected" are represented by numbers.

The main challenge towards designing fuzzy logic systems for geospatial concepts is the transition from one-dimensional variables to two or three dimensional geo-referenced variables (such as fuzzy maps), in addition to temporal dimension. Some classical fuzzy operations, such as intersection or negation, need to be redefined, as well as Fuzzy rules and relations. On the other hand, some classical spatial concepts, such as topological relations, may not previously have a fuzzy counterpart, thus need to be conceptually "fuzzified".

Some possible fuzzy expression of spatial terms are:

- Fuzzy boundaries,
- Some uncertain words, such as *close* (as in close to hospotal) or *moderate* (as in moderate infection).
- Combined boundaries and words (as close to moderately infected area)
- Cardinal directions; and,
- spatio-temporal concepts.

Fuzzy boundaries have a major role in developing rule-based fuzzy-GISs, since many rules are based on relative positions of people or assets to some areas, whereas the boundaries of those areas are not precisely defined. Imagine the concept of "high-risk area" in a city, where there is no real boundary between it and the low- or moderate-risk areas of the same city.

The uncertainties about the boundaries of an area can come from different sources [14]:

- **Continuousness** occurs when the boundaries rely on a continuous variable (e.g., the transition of infection rate between country borders).
- **Aggregation** occurs when the boundaries are obtained by aggregating the values of different variables (e.g., population per area unit may indicate a city's boundary).
- **Time Averaging** occurs when the actual boundaries vary in time (e.g., rivers, shorelines).
- **Ambiguity** occurs when linguistic terms are used to define regions (e.g., the highly-infected area of a city).

Based on the above implications of pandemic data, we outline a conceptual fuzzy-geospatial model in the next section.

III. A Conceptual Model

A complete model that can predict the different epidemiological scenarios based on all the various data collected in the past sounds very complicated. Limiting the focus on modelling the geographical and temporal patterns of the spread based on some of their most influencing factors as observed during the Coronavirus pandemic, such as the spatiotemporal patterns, disease epicentres and local/national protection measures can make the model more realistically possible to develop, although still complex. We provide a conceptual outline of such a model, having an outlook that if the developed model showcases its prediction performance, it can be followed up by further model extension incorporating some more influential parameters such as socio-cultural, demographics, transport system, virus types, etc.

A. Model Outline

This model is outlined in Fig. 1. The aim is to simulate and predict the geographical pattern of the disease spread in the presence of enforced quarantine measures. The knowledge-base contains fuzzy sets for different uncertain pandemic concepts such as quarantine level or patient severity (Fig. 2). As far as spatiotemporal data concern, it is necessary to take the uncertainties/imperfections into account when predicting how the disease will spread geographically and temporarily if a certain measure is enforced. The enforced measures, as observed during the pandemic, were in a wide range from limited self-isolation to home stay, from city quarantine to travel ban to a total country lock-down. Not surprisingly, the measures themselves are to be modelled with some levels of uncertainties since they were differentiated with grey borders and/or implemented imperfectly.

B. Data Sources

During (and possibly after the pandemic), different official public and/or research-specific datasets are being developed, some of which are:

- At the global level, WHO (World Health Organisation) publishes detailed reports daily (called Situation Report) including maps, spreadsheets and statistics about the disease spread. WHO-Europe also publishes the daily situation reports at some higher resolutions at European country level [1].
- At the national level, the UK Government releases detailed daily reports (map and statistics) at the UK's cities and county levels [2].
- Research centres around the world, for example, John Hopkins University's Coronavirus Resource Centre that provides world-wide daily maps and statistics [3].
- Another required source of information is the measures that the governments are taking in response to the Coronavirus spread. Besides news and media sources, the data source is provided by the Oxford COVID-19 Government Response Tracker (OxCGRT) project [4].

Moreover, a variety of data sources are expected to be available either publicly or for research communities after the pandemic settlement.

C. Model Training

Training fuzzy systems by example data is has known established methods [16], however, applying them for geospatial data and non-singleton systems is new and therefore needs extra research works. Training leads to fuzzy sets for the system's fuzzifiers (i.e., knowledge-base) and inference engines (i.e., rule-base), however for the geospatial data, the training is partly based on map data, thus the resulted fuzzy sets can be in spatiotemporal forms scattered in 2 or 3 dimensions.

D. Uncertainty Capturing

For achieving a higher level of uncertainty capture than the classical fuzzy systems, the conceptual model is designed as a non-singleton fuzzy logic system (NSFLS) [17]. The NSFLSs are proven to outperform the classical fuzzy systems in working with uncertain inputs since the non-singleton fuzzifiers capture the input uncertainties at their source and directly carry them to the inference engine. This will allow to model the embedded uncertainties not only at the level of linguistic uncertainties, but also at the level of data collection. For example, the system will be capable of making inferences based on fuzzy operations between the knowledge-bases fuzzy maps and the fuzzified input maps (as opposed to the classical rigid-border maps). This would not be possible in singleton fuzzy systems. The difference between singleton and non-singleton fuzzy logic systems is shown in Fig. 3.

Designing the system as NSFLS, makes it essentially similar to noisy time series prediction. The difference is that unlike the time series, the system works with a wide range of non-scalar data. NSFLSs has already shown their relatively high performance in noisy and chaotic time series prediction [17]. In addition, we have already developed some improved methods in employing NSFLSs for noisy time series prediction problems [18], [19], which can be adapted for the pandemic data.

[1] https://who.sprinklr.com/
[2] https://coronavirus.data.gov.uk/
[3] https://coronavirus.jhu.edu/
[4] https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker

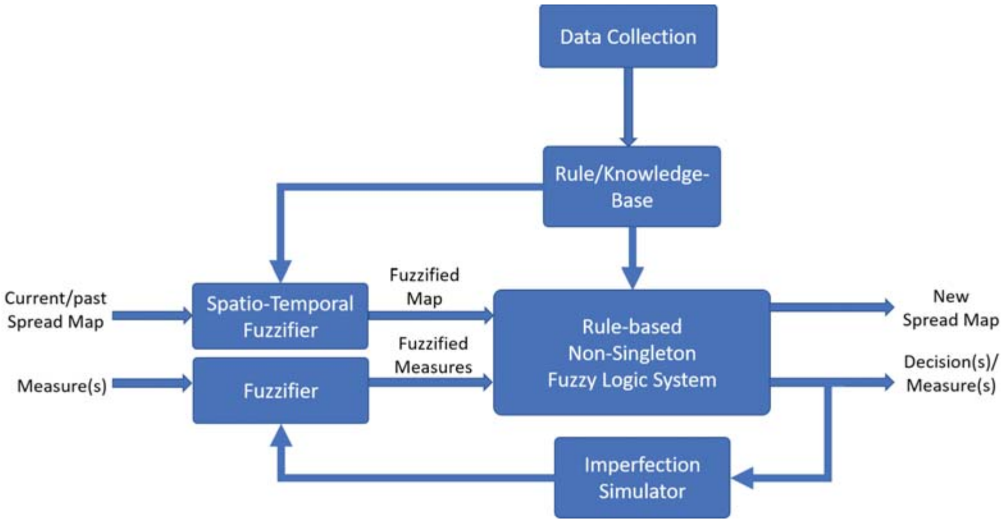


Fig. 1: The Fuzzy rule-based system outline incorporating spatio-temporal concepts.

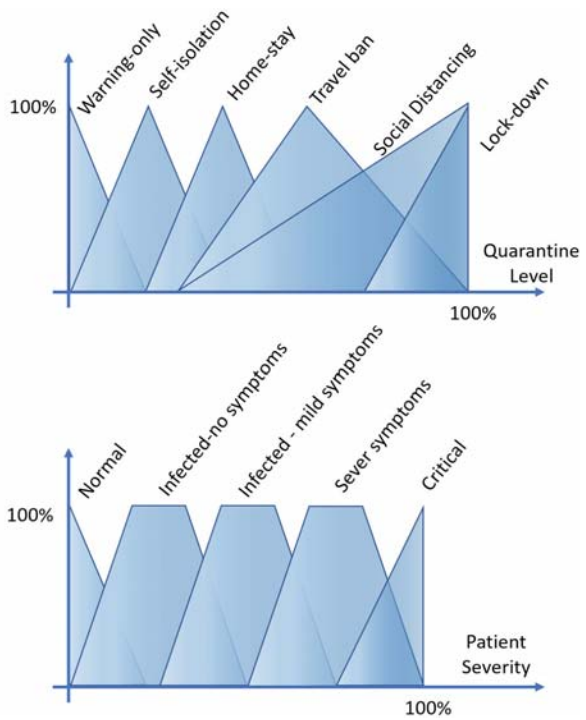


Fig. 2: Exemplar fuzzy sets in the system's knowledge-base.

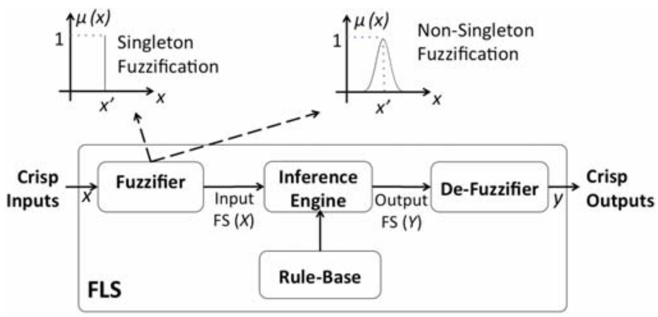


Fig. 3: The difference between singleton and non-singleton fuzzy logic systems.

IV. A Simplified Model

In order to showcase how the described conceptual model can be realised in a limited local extent, we consider a simplified modelling scenario based on the available public datasets in the UK and generate a non-singleton fuzzy rule-based system to model the scenario.

The geographical extent of this scenario is Nottingham city in the UK, and its surroundings. The county of Nottinghamshire consists of 8 local authorities, 4 of which surround the Nottingham city local authority, namely Ashfield, Broxtowe, Rushcliffe and Gedling. The daily maps and statistics of the spread in each local authority is publicly released by the UK Government, together with maps at local authorities level. The sample of these two daily maps are shown in Fig. 4 at the upper tier and Fig. 5 at the lower tier respectively.

The aim of this model is to predict the number of new cases in Nottingham based on the available data as follows.

A. Inputs and Outputs

- The number of confirmed cases in the previous day in Nottingham city (Fig. 7.a);
- The average number of the confirmed cases in the 4 neighbour local authorities (Fig. 7.a). The idea is to consider the effect of the infection rates in the vicinity of the city, thus for having a better estimation, we calculate a weighted average of the reported cases in the 4 neighbouring areas, in which the weights are the ratios of the border lengths between Nottingham city area and each

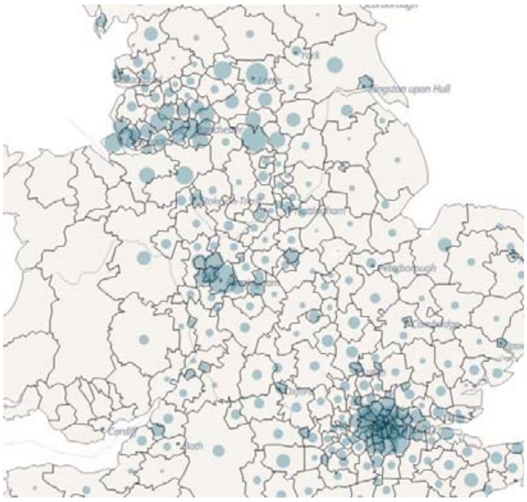


Fig. 4: A sample outbreak map of the UK at the upper tier level local authority (UTLA) level (from https://coronavirus.data.gov.uk/).

Fig. 5: A sample outbreak map of the UK at the lower tier level local authority (LTLA) level, showing Nottingham and its surrounding local authorities. Darker shades have higher rates. (from https://coronavirus.data.gov.uk/).

of the surrounding local authorities, so that the longer the border, the higher the spread possibility.

- The cumulative number of cases per 1000 residents (Fig. 7.b). This is to account for epidemiology or immunology factors since the possibility of the virus spread slows down by increasing the number of already infected cases.
- The number of days passed since the start of the national lock-down in the UK (23 March 2020) (Fig. 7.c). The idea is that the longer the lock-down is enforced, the lower the possibility of the spread.
- The number of tests per 1000 residents (Fig. 7.d). This factor is not a direct input to the system, rather it changes the non-singleton fuzzification attribute related to the first two inputs. The idea is that if more tests are done, the number of daily cases becomes more accurate, thus a narrower membership function can be used to model the uncertainty in the non-singleton fuzzification block.

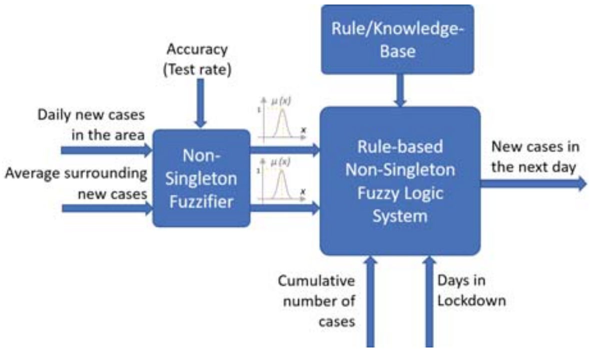


Fig. 6: The outline of the simplified model. The non-singleton fuzzifier is governed by the daily test rate.

Fig. 6 shows the architecture of this simplistic model. For simplicity, we considered a single output variable for the model. This is the prediction of the new daily cases in the next day.

B. *Fuzzy Sets*

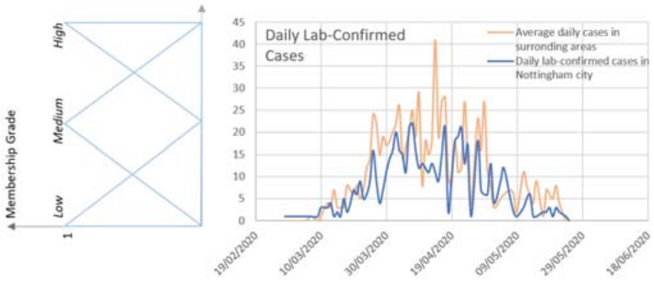
For each of the system inputs, we define three linguistic labels namely *Low, Medium and High*. Three points including the minimum, middle and the maximum values for the range of each input/output variable is selected. Then for each variable, three triangular membership functions are defined as illustrated Fig. 7.a-c. The three defined membership functions represent the three linguistic labels *Low, Medium and High*.

C. *Non-Singleton Fuzzification*

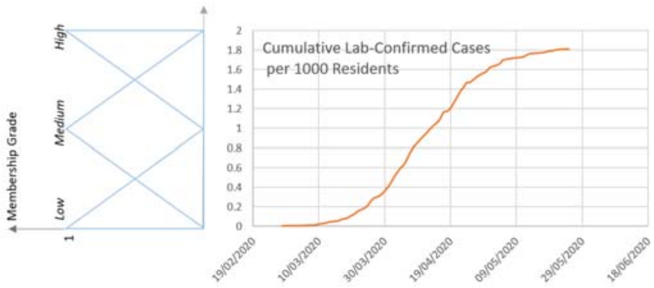
We consider a symmetrical normalised triangular membership function for the fuzzification block, in which the triangle base length is reversely proportional to the number of tests per 1000 residents. As illustrated in Fig. 7.d., when the daily tests are low (high), the daily cases have high (low) uncertainty, thus the input data is fuzzified with a wider (narrower) triangular membership function.

V. Conclusion and the Future Works

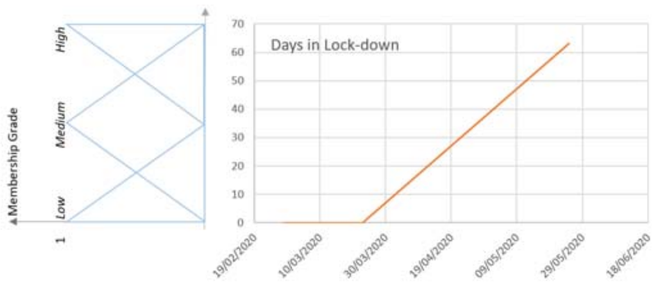
In this paper we outlined the research work towards a novel pandemic model, motivated by the rich data sets being created during the Covid-19 pandemic. A complete predictive model that accounts for all the features found in the current pandemic data, and for a large geographical extent could be far too complex. We proposed the outline of a novel modelling approach by limiting the number of features to some of the most important ones, and by considering the following three methodologies, collectively called data-driven fuzzy-geospatial modelling:



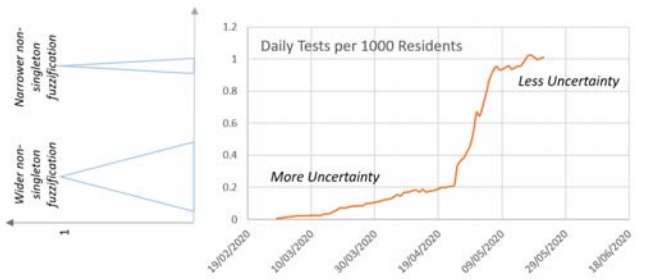
(a)



(b)



(c)



(d)

Fig. 7: Illustrations of the available pandemic data for Nottingham city and its surrounding areas at the time of writing this paper (from https://coronavirus.data.gov.uk/), and their associated fuzzy sets; (a) The daily lab-confirmed cases in Nottingham city and its surrounding areas (the average of 4 local authorities around Nottingham city). The fluctuations are due to reporting delays, however the graph can be statistically smoothed before being fed into the model. Three fuzzy sets (low, medium and high) are associated to different levels of the reported cases, (b) The cumulative lab-confirmed cases rate in Nottingham city together with the associated three fuzzy sets, (c) The number of days passed after the UK's national lockdown (23/03/2020) and its associated three fuzzy sets, (d) The daily test rates and two illustrative fuzzy sets for non-singleton fuzzification of the data (here, the daily cases). The less the tests, the more the uncertainty, the wider the fuzzy set.

- Taking the advantage of the rich data generated during the pandemic: Unlike most of the epidemiological approaches, this model is data-driven not mathematical.
- Incorporating the range of uncertainties embedded in the pandemic data into the model, by means of designing non-singleton fuzzy logic systems.
- Extending the classical rule-based fuzzy systems to include geo-referenced data required by the pandemic data.

Based on the above ideas, a conceptual model is outlined and a simplified model is designed in order to showcase how to locally realise the model. The simplified model is driven by the official data of a limited geographical area in Nottinghamshire, UK. Further model development in the future has the potential to act as an uncertain decision support systems (UDSS) in controlling the pandemic in future outbreaks.

References

[1] J. C. Blackwood and L. M. Childs, "An introduction to compartmental modeling for the budding infectious disease modeler," *Letters in Biomathematics*, vol. 5, no. 1, pp. 195–221, 2018.

[2] T.-M. Chen, J. Rui, Q.-P. Wang, Z.-Y. Zhao, J.-A. Cui, and L. Yin, "A mathematical model for simulating the phase-based transmissibility of a novel coronavirus," *Infectious diseases of poverty*, vol. 9, no. 1, pp. 1–8, 2020.

[3] E. Massad, N. R. S. Ortega, L. C. de Barros, and C. J. Struchiner, *Fuzzy logic in action: applications in epidemiology and beyond*. Springer Science and Business Media, 2009, vol. 232.

[4] N. R. S. Ortega, P. C. Sallum, and E. Massad, "Fuzzy dynamical systems in epidemic modelling," *Kybernetes*, 2000.

[5] W. Kainz, "Fuzzy logic and GIS," *Vienna: University of Vienna*, 2007.

[6] D. Zheng, "A neural-fuzzy approach to linguistic knowledge acquisition and assessment in spatial decision making," 2001.

[7] M. Faridi, S. Verma, and S. Mukherjee, "Integration of GIS, spatial data mining and fuzzy logic for agricultural intelligence," in *Soft Computing: Theories and Applications*. Springer, 2018, pp. 171–183.

[8] N. Arefiev, V. Terleev, and V. Badenko, "GIS-based fuzzy method for urban planning," *Procedia Engineering*, vol. 117, no. 1, pp. 39–44, 2015.

[9] X. Zhao, X. Chen, and A. Stein, "Integrating multi-source information via fuzzy classification method for wetland grass mapping," *The inter. archives of the photogrammetry, remote sensing and spatial information sciences*, vol. 37, pp. 1463–1470, 2008.

[10] A. Mardani, R. E. Hooker, S. Ozkul, S. Yifan, M. Nilashi, H. Z. Sabzi, and G. C. Fei, "Application of decision making and fuzzy sets theory to evaluate the healthcare and medical problems: a review of three decades of research with recent developments," *Expert Systems with Applications*, vol. 137, pp. 202–231, 2019.

[11] I. Pasha, *Ambulance management system using GIS*. Universitetsbibliotek, 2006.

[12] G. K. Palshikar, "Fuzzy region connection calculus in finite discrete space domains," *Applied Soft Computing*, vol. 4, no. 1, pp. 13–23, 2004.

[13] A. Cohn and M. Gotts, "The 'egg—yolk'representation of regions with indeterminate," *Geographic objects with indeterminate boundaries*, vol. 2, p. 171, 1996.

[14] S. Schockaert, M. De Cock, and E. E. Kerre, *Reasoning about fuzzy temporal and spatial information*. World Scientific, 2011, vol. 3.

[15] C. Kumar, W. Heuten, and S. Boll, "Visual overlay on openstreetmap data to support spatial exploration of urban environments," *ISPRS Int. J. of Geo-Information*, vol. 4, no. 1, pp. 87–104, 2015.

[16] L.-X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Trans. on systems, man and cybernetics*, vol. 22, no. 6, pp. 1414–1427, 1992.

[17] J. M. Mendel, "Uncertain rule-based fuzzy systems," in *Introduction and new directions*. Springer, 2017, p. 684.

[18] A. Pourabdollah, C. Wagner, J. H. Aladi, and J. M. Garibaldi, "Improved uncertainty capture for nonsingleton fuzzy systems," *IEEE Trans. on Fuzzy Systems*, vol. 24, no. 6, pp. 1513–1524, 2016.

[19] A. Pourabdollah, R. John, and J. M. Garibaldi, "A new dynamic approach for non-singleton fuzzification in noisy time-series prediction," in *2017 IEEE Int. Conf. on Fuzzy Systems*. IEEE, 2017, pp. 1–6.