SPATIAL MAPPING AND CLUSTER ANALYSIS OF COVID-19: A CASE STUDY OF UTTAR PRADESH, INDIA

S Agrawal[1], S Agrawal[2,*]

[1] GIS Cell, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, Uttar Pradesh 211004, India - shikha2592agrawal@gmail.com
[2] GIS Cell, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, Uttar Pradesh 211004, India - sonam@mnnit.ac.in

Commission IV, WG IV/9

KEY WORDS: COVID-19, Cluster analysis, SatScan, GIS, Mapping.

ABSTRACT:

In public health, the representation and analysis of the incidence of the disease play an important role in assessing the regional disparity and health infrastructure. The geographic information system is one of the best methods to do so. This paper aims to visualise the spread of COVID-19 and perform cluster analysis. The prospective Poisson space-time scan statistic was utilised to detect clusters of COVID-19 at the district level in the Uttar Pradesh state of India. The spatial mapping was performed to assess the situation of COVID-19 and related factors. The log-likelihood ratio and relative risk were calculated monthly from May to December 2020. As per the results, the size and location of clusters kept changing with being more concentrated in Lucknow, Kanpur, Gautam Buddha Nagar districts and NCR regions. The significance of these clusters was less than 0.001. The detection of these clusters helped understand the overall dynamics of the disease spread. The number of confirmed cases declined in most districts, with only a few being at higher risk and needing more attention and resources. It was concluded based on the analysis that the areas of higher economic activities and population density with higher access to hospitals and testing had a larger number of cases and were the regions of hotspots. The findings can help to create health awareness, monitor situations in real-time and evaluate the steps taken to assess their efficacy.

1. INTRODUCTION

In public health, the representation and analysis of the incidence of the disease play an important role in assessing the regional disparity and health infrastructure (Kumar and Agrawal 2020; Rushton 2003). Risk mapping and visualization help the decision-makers to take the necessary action based on the location of the disease incidence. The role of Geographic Information System (GIS) in public health is vast (Tripathi et al.). Every disease has spatial and temporal components which can be dealt with by GIS (Rahman et al. 2020). Virus-borne diseases are closely associated with their environment and hence have a spatial component that can be integrated well with GIS (Kandwal et al. 2009). GIS helps to know who needs the resources and its location (Agrawal and Gupta 2017). GIS can monitor the disease and create risk maps (Lyseen et al. 2014).

In December 2019, a viral pneumonia outbreak was reported in Wuhan City, Hubei Province, China (Murugesan et al. 2020). It was caused by a novel coronavirus–severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It was named coronavirus disease 2019 (COVID-19). As it spread across the globe, the World Health Organization (WHO) declared it a pandemic (Mollalo et al. 2020). Vaccination programs have been started in several countries, but it was hard to contain the virus because of its unpredictable behavior and mutations. Governments were imposing lockdowns, issuing stay at home advisory and wearing masks was compulsory as this virus was hitting intermittently (Cole et al. 2020). Attention has been paid to health and hygiene to stop the spread of the virus. GIS can make the public and professionals aware of this disease through visualizations and dissemination of information (Cordes and Castro 2020; Kang et al. 2020; Kim and Castro 2020).

*Corresponding author

The first such dashboard was developed by John Hopkins University, which has become one of the most followed sources for COVID-19 related data. The case data used in the dashboard has been made publicly available.

In India, the first case of COVID-19 came in January 2020. Gradually, it engulfed all the states of the country. In India, approximately 31% of its population resides in urban areas and contributes to 65% of its GDP as per the Census of India, 2011 (https://www.census2011.co. in/census/). It can be seen in the current times the urban centers are growing at a rapid pace. As per reports and evidence, it is said that there will be a point where approximately 90% of the world's population will shift to these urban centers by the end of this century (Li et al. 2020). These urban agglomerations suffer from numerous problems like high population density, shortage of drinking water, pollution, reduction of green spaces and higher temperatures that lead to poor quality of life and increased disease burden. When the COVID-19 pandemic situation arose, it was necessary to monitor all these places. In India, a complete lockdown was imposed from 25th March 2020 to 31st May 2020. Initially, the cases were very less as all the movement of people, freight and trains stopped. But, slowly, the migrants came back to their home districts. In India's Uttar Pradesh state, a large number of the working population had migrated to different states and because of lockdown, they were forced to return. Despite all the efforts, there was a surge of cases that could be attributed to few factors considering the socio-demographic profile of the state. Some of them were behavioural reluctance to follow the guidelines, poor health infrastructure and high population density. GIS can help to analyze the situation in this pandemic situation (Krzysztofik et al. 2020; Mahdizadeh Gharakhanlou and Hooshangi 2020; Pourghasemi et al. 2020; Requia et al. 2020; Sarwar et al. 2020). It was imperative to study the disease

https://doi.org/10.5194/isprs-annals-X-4-W3-2022-3-2022

spread and map it for better visualization and create awareness amongst the communities to better equip for future health emergencies. This work is an attempt to fill this research gap.
This paper aims to study the COVID-19 spread and the possible cluster formation. The first objective was to perform spatial mapping of COVID-19 and related parameters. The second objective was to construct clusters based on prospective scan statistic method using the Poisson distribution method. The scan statistics method has been used for ages for carrying out disease surveillance, forestry, archaeology, accident studies, chemical attacks and crime rate studies (Kulldorff 1999). In the light of the COVID-19 pandemic, this method can be useful. Based on hotspot development, resource allocation and medical help can be carried out to prevent the disease's further spread.

2. BACKGROUND

Cluster analysis is one of the oldest techniques to study the concentration of events at a particular location—disease hotspots. Various software like ArcGIS, SatScan, R packages can be used to perform cluster analysis. It is an effective method to exhibit areas of high concentration of crime, disease, disasters, etc. The clusters of high and low significance can be mapped using hotspot analysis and the resources can be allocated accordingly. The space scan statistics can be used to study the spatial dynamics of disease spread. This spatio-temporal statistical technique can detect clusters through various time duration and locations. This method can be used to study the significance of any clusters using multiple statistical methods in space and time (Kulldorff et al. 2005; Takahashi et al. 2008). The SatScan software can be used to detect potential geographic clusters (Coleman et al. 2009; Linton et al. 2014; Mulatti et al. 2015; Tonini et al. 2009; Tuia et al. 2008).
In space scan statistics, a scanning window is used. This window can be cylindrical or circular, depending on the type of method used. A circular window is used for purely spatial analysis, while a cylindrical window is used when both space and time dimensions are considered. The cylinder's height represents the temporal dimension, while the size of the base represents the location or space dimension used for making clusters. Then, for the maximum upper bounds for both the parameters, i.e., spatial and temporal, the cylinders are expanded until the maximum is reached.

A scan statistic can be either prospective or retrospective in nature. The retrospective method helps to study the clusters that have occurred in the past, while the prospective method helps to scan clusters based on real-time. The prospective scan statistic method is applied when there is a need to systematically monitor the spread of the disease, predict new regions of an outbreak, analyse clusters and facilitate the future anticipation of the transmission. This method is based on the significance of clusters which is given based on the values of Log-Likelihood Ratio (LLR), Relative Risks (RR) and Monte Carlo test simulation.

The underlying hypothesis in the SatScan software is the null hypothesis (H_0). The null hypothesis says "the events occurring are completely random" i.e., the disease spread is random and follows no spatial clustering. The alternative hypothesis (H_a) says that clusters are formed out of non-randomness, i.e., there is a close relationship between the clustering of disease. For any cluster to be significant, the null hypothesis needs to be rejected. Hence inferential statistics support the evidence that the clusters occurring are non-random and unique in nature. The alternative hypothesis is verified based on the Monte Carlo test simulation values, which give a probability value (p-value). If a p-value is less than 0.001, the cluster is said to be of significance. The risk associated with the significant cluster depends on the values of RR.

3. STUDY AREA

The study area of this research work was the state of Uttar Pradesh in India, as shown in Figure 1. It covers an area of 240,928 km^2. It is divided into 75 districts. As per the latest census, which held in the year 2011, the population of Uttar Pradesh was 199.8 million people making it the most populous state in India. The forecasted population for the year 2020 is estimated to be 237 million, as per the Aadhar India Unique Identification. The population density of Uttar Pradesh is 829 persons per km^2. The climate of this region is defined primarily as humid subtropical and dry winter. The western part of Uttar Pradesh has a semi-arid type of climate.

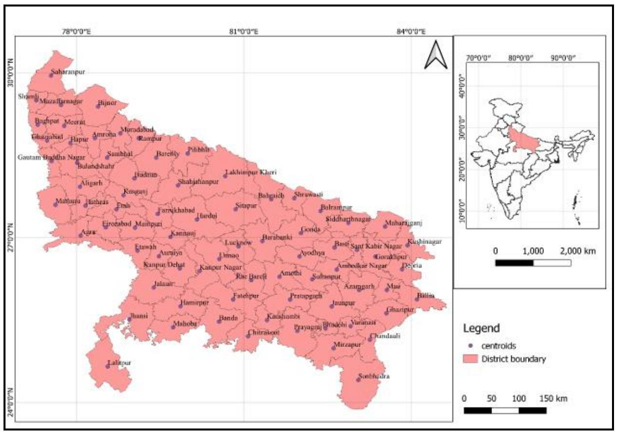

Figure 1. Map showing the study area.

4. DATA USED

Uttar Pradesh has 18 administrative divisions, which are further divided into 75 districts. The daily case data are obtained from the publicly accessible websites covid19india.org from May 2020 to December 2020, which utilizes data from all the government websites and their Twitter handles. The data included case-time series data for the entire country, state and district. It also contained data of confirmed cases, active cases, death and the number of tests done. These data have been converted into a format that was supported by GIS software using MS Excel. The data was then made in a form that was compatible with SatScan software.

The COVID-19 data was district-wise, which have been modified as per the latest shapefiles. So, the data has been corrected to match the shapefile data for the further analysis. To carry out the statistical analysis in SatScan software two types of files are required, namely, the coordinate file and the case file. The Poisson statistic method is used in this study, so a population file is also prepared. In India, the last census was carried out in 2011 and for the accurate analysis of COVID-19 clusters the current population data had to be obtained. For this purpose, the data obtained from the UIDAI (https://uidai.gov.in/) has been used which was updated on 31st May 2020, at the time of this study. District-wise hospital beds data has been obtained from the government website (https://nidm.gov.in/covid19/PDF/covid19/state/Uttar%20Pradesh /159.pdf).

5. METHODOLOGY

The space scan statistics method is used to study the spatial dynamics of disease spread. For this purpose, the SatScan software is used to identify the clusters of COVID-19 cases in the state of Uttar Pradesh. The SatScan software detects the non-random and unique clusters using the prospective Poisson distribution method. Since the disease is viral in nature and spreads from person to person, therefore a densely populated area will be at higher risk than places with less population density. Hence, the Poisson method has been chosen to encapsulate this point. The probability value (p) is calculated by the Monte Carlo simulation method. If the value of p is less than 0.001, then the clusters are considered significant. The software will automatically drop off all the less significance.

There is mathematics behind the statistical analysis. It is assumed based on the past studies that the case counts follow the Poisson distribution. For a disease outbreak, the population is a critical attribute. The null hypothesis assumes that the expected number of cases in each area is proportional to population in the absence of all other covariates. This study takes into consideration only the population and case-time data. μ is the constant risk factor that depends on the population at risk, as given below (Desjardins et al. 2020):

$$\mu = p_i \times \frac{C}{P}$$

where C is the number of cases, P is the total population and p is the population in the particular cylinder i.

According to the alternative hypothesis, it is further assumed that the current number of cases will be exceeded and also that the population in a given geographic location is constant over time. Hence migration is not being considered in the present analysis. Based on the data requirement and the hypothesis, a log-likelihood function is calculated whose higher value denotes the non-randomness of the disease spread. The value inside and outside the cylinder is compared to ascertain the relative risk associated and is given by performing following LLR test given as follows (Hohl et al. 2020):

$$\frac{L(Z)}{L_0} = \left(\frac{\left(\frac{n_z}{\mu(z)}\right)^{n_z} \left(\frac{N - n_z}{N - \mu_{(z)}}\right)^{N-n_z}}{\left(\frac{N}{\mu(T)}\right)^{N}} \right)$$

Where $L(Z)$ is the likelihood function for cylinder Z, L_0 is the likelihood function for H_0, n_z is the number of COVID-19 cases in a cylinder, $\mu(Z)$ *is* the number of expected cases in-cylinder Z, N the total number of observed cases for entire Uttar Pradesh across the time periods and $\mu(T)$ *is* the total number of expected cases in the study area across the time period. When the L(Z) value is greater than one, it means that $\left(\frac{n_z}{\mu(z)}\right) > \left(\frac{N-n_z}{N-\mu(z)}\right)$, and so high chances of this being a cluster. LLR tells about how risk differs inside and outside the cylinder. Monte Carlo simulations are used to understand the impact of risk and uncertainty in prediction and forecasting models. Multiple values are assigned to the uncertain variable to achieve multiple results to estimate the outcome (Cheng and Adepeju 2014) . Using Monte Carlo replications with more than 999 replications under the defined null hypothesis, the clusters' distribution and statistical significance are observed for the multiple values and datasets.

To carry out the disease analysis, the first step is to prepare the data. The data is prepared in MS Excel. It is used to generate a graph showing the trend of confirmed cases and testing in the study area. Along with this cluster heat map is also created to show the month-wise number of cases in each district.

To conduct the disease surveillance, a prospective scan statistic method with Poisson distribution is chosen. The prepared data is made in a GIS-compatible format and is used as the input files in the SatScan software to do the analysis. These input files are the case file, coordinate file and population file. For the analysis, a Poisson statistic method has been chosen and the required parameters are set. The spatio-temporal cluster has been set to 10% of the population at risk, which is the population of a particular district and the shape has been set to circle. In the present study, the temporal maximum cluster size is half of the study period while the minimum is two days. The Monte Carlo simulations are set to 999 and the outputs are generated in the KML (Keyhole Markup Language) and shapefile format. The shapefile data of clusters is imported to QGIS to study the clusters formed. The maps have three visual components. First, the black color dots represents the locations of clusters being considered.. Second, the cluster-circles are color-coded as red for the high rate of risk and blue for low risk based on the values of RR. The RR values talk about the risk within the circle as compared to outside risk. RR ranges from 0 to any value. The clusters with RR values of less than 1 are of low risk, while those greater than 1 is of high risk.

The varying size of clusters show that while analysing how much area is covered, which can vary from being 0 km to user-defined distance. This parameterization is based on the underlying disease characteristic. The reproduction rate of the disease is very high and hence chances of rapid spreading are high. If a certain number of cases are found for more than 2 days, it is considered as a cluster. Therefore, clusters of circular shapes in various sizes are generated. The cluster sizes are based on parameterization only. If, while running analysis, SatScan finds clusters in the specified range, then its size remains unchanged. Otherwise, it keeps on changing the size to achieve the parameters. Based on the RR value, the clusters can have high rates or low rates of disease occurrence. In the results, it could be seen that sometimes various cluster locations are seen lie inside the circles and sometimes a given cluster location itself is significant. The clusters occurring alone or having zero km size have higher RR than circles with varying sizes. So, these locations should be paid more attention.

This cluster analysis has been performed for each month of the study period to see the change in the cluster's size and number. Cluster analysis of each month gives a new set of clusters. Some of these are of higher significance, while others are of lower significance. Based on the significance of clusters, the severity of the disease risk in each of the clusters is identified, which could be helpful for decision making.

6. RESULTS

6.1 Spatial Mapping of COVID-19 cases

The choropleth map of the population density of each district of Uttar Pradesh is shown in Figure 2. To accurately understand the disease trend a graph is plotted to see how the testing has been improved over the months and to understand the intensity of the disease spread. It could be seen in Figure 3 that as the

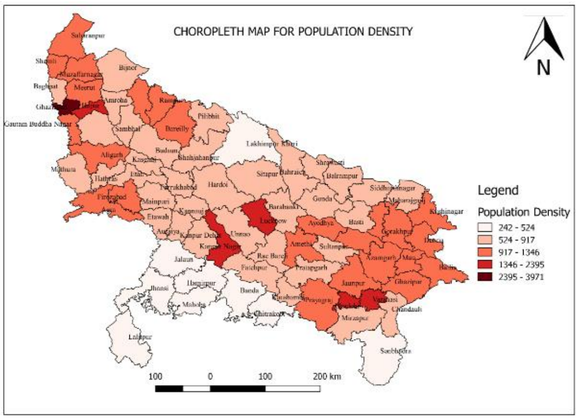

Figure 2. Map showing population density in each district of Uttar Pradesh

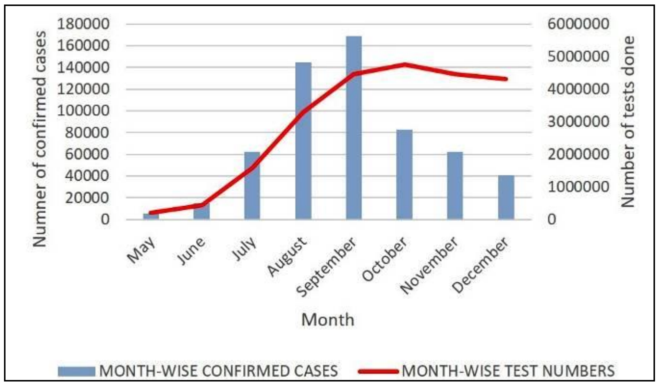

Figure 3. Month-wise testing and number of confirmed cases trend in the study area in the year 2020.

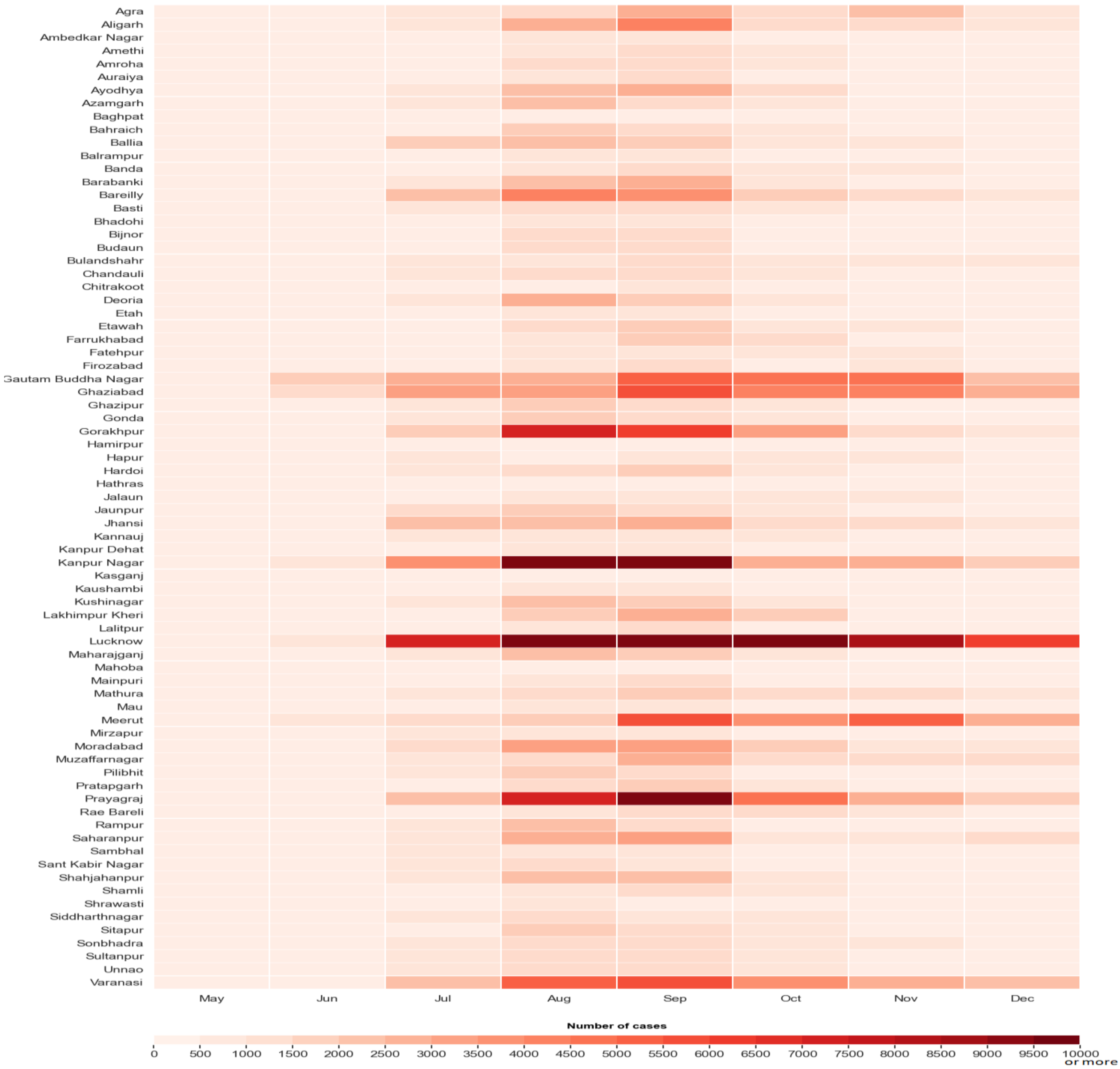

Figure 4. Cluster heat map showing the month-wise number of cases in each district.

months passed by, the number of testing increased. By the end of the study period, the number of cases kept on reducing, showing a decline in the number of cases. It could also be seen that in September the highest number of cases were found.

The cluster heatmap is created to depict the number of cases district-wise, shown in Figure 4. In this, the colour of each cell gives information about the number of cases. It can be seen that the highest number of cases were found in the district of Lucknow which is followed by Kanpur and Prayagraj districts.
The spatial heat map for the confirmed maps is shown in Figure 5. It is showing that the higher clustering and hotspot are formed around the capital city Lucknow, National capital region (NCR)region, Varanasi, Prayagraj, Kanpur and Gorakhpur. It can also be seen that the districts where higher population density is present are also the district where the highest number of cases have been found like the highest population density is in the Ghaziabad district followed by other districts like Varanasi, Hapur and Lucknow. The higher population density suggests higher chances of disease spread and infection.

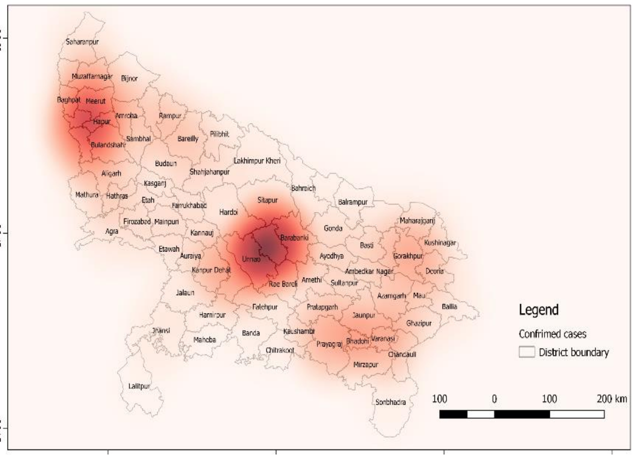

Figure 5. Spatial heat map showing the density of number of confirmed cases in the study period.

6.2 Cluster Analysis

This study takes the case of Uttar Pradesh state, where, as of May 1 2020, there were 2328 positive confirmed cases. The analysis is carried out each month from May 2020 till December 2020, as shown in Figure 6. When the analysis is done for May, eleven clusters are reported, out of which nine are statistically significant. Of the 75 locations, 29 are cluster locations, and their LLR and RR are of high value. Immediate attention is needed in these locations so the disease does not spread to new locations. Isolated locations are rarely seen, and most cluster locations are close to each other. Lucknow, the capital of UP and the regions of UP surrounded by the NCR are seen continuously in the high relative risk area. From June 1 2020, the lockdown was further relaxed, with shops outside markets being allowed to remain open. As a result, it could be seen that areas like Chandauli, Mahoba and Maharajganj started to become regions of risk towards the end of their study period. However, LLR in all these is relatively low. More and more relaxations were granted in "Unlock 1", and the surge in the cases can be seen. More and more people started to get out of their homes for various reasons, and soon the disease spread more, which can be seen in the maps as June and August have more clusters and new risk locations. The LLR of Ghaziabad and Unnao has increased to 4105.754 and 250.729 from 718.765 and 15.16, respectively, while it declined for Amethi. Mahoba, though less significant in terms of P-value and LLR(p-value>0.0001 and LLR below than critical value of 18), show a high RR of 1.45. Since it is a new cluster, authorities should consider reducing its risk by adopting a mitigation policy of wide lockdown and closing of borders and areas from where more cases are coming. Kannauj and Rampur are now not seen as a cluster of risk.

In July and August, the disease spread to new locations, and new clusters were also seen with high LLR and RR. Until July, Lucknow was seen as part of cluster Unnao. However, from August onwards, it became a location of high risk suggesting the risk of widespread disease infection. Lucknow tops the chart with the highest LLR of 6420.533 and 18916.289, respectively. Rampur is again seen as a high-risk location with an LLR of 794.255 and RR of 2.91. LLR and RR of the NCR region have declined to 763.176 and 3.43 in August from 3100.638 and 2.35, respectively. In August 2020, 14 clusters were formed, with 18 cluster locations. Rampur again is not seen in August and September.

In September, Lucknow had the highest LLR of 27772.204 with an RR of 7.58. In September 2020, 13 clusters were formed with 19 locations; out of these 12 clusters, 18 locations have high RR. It could be seen that the clusters have reduced in size and are now clustered more closely. Hence, there was a higher risk of transmission at these locations. The centres of clusters are still the same districts like Lucknow, Amethi, Shajanpur, NCR, Kanpur, Prayagraj, and Varanasi. The higher number of clusters with less risk suggests that the containment strategy is effective, and now there is a downward trend in the spread of disease with normalcy coming back. The state was under third unlock during this time, and more and more relaxation was granted. This was also the time when Lucknow had the maximum number of cases. The state has the peak of cases again attributed to more testing done here and more relaxations provided.

In October 2020, 13 clusters with 18 locations were analysed. There is a continuous downward trend in the size and number of clusters, with the locations being the same. Lucknow has the highest LLR, and its RR is 6.21. The RR of the NCR region is decreased to 3.32, but its LLR is still high, close to Lucknow's, i.e. 6282.754. Jhansi is seen as a new area to consider, with RR being 1.40 and LLR 56.94

In November 2020, 9 clusters with 19 locations were formed. Eight clusters have high rates of disease spread. The NCR region has the highest LLR value of 12440.177, with RR being 5.57. It can be seen that now the relative risk is declining, but the location is the same. The relative risk in Lucknow has increased to 6.78 compared to previous values. However, its LLR has declined since October, suggesting the situation is improving. Varanasi again became a high-risk centre, with RR being 2.27 showing a sudden spike in cases. This might occur due to the festive season. The risks in Kanpur and nearby districts were also reduced to 1.65

In December 2020, the software identified 9 clusters with 14 locations. Out of these, 8 clusters with 13 locations are of high risk. Lucknow district remained at the top with an RR of 8.15 and LLR of 7442.048. The second high-risk zone is still in the NCR region. Amethi has shown the highest RR value of 107.59 with an LLR of 1068.380. Higher RR is due to an unexpected surge in cases, while LLR is lower than the Lucknow and NCR region. The other districts like Varanasi and Prayagraj are

https://doi.org/10.5194/isprs-annals-X-4-W3-2022-3-2022

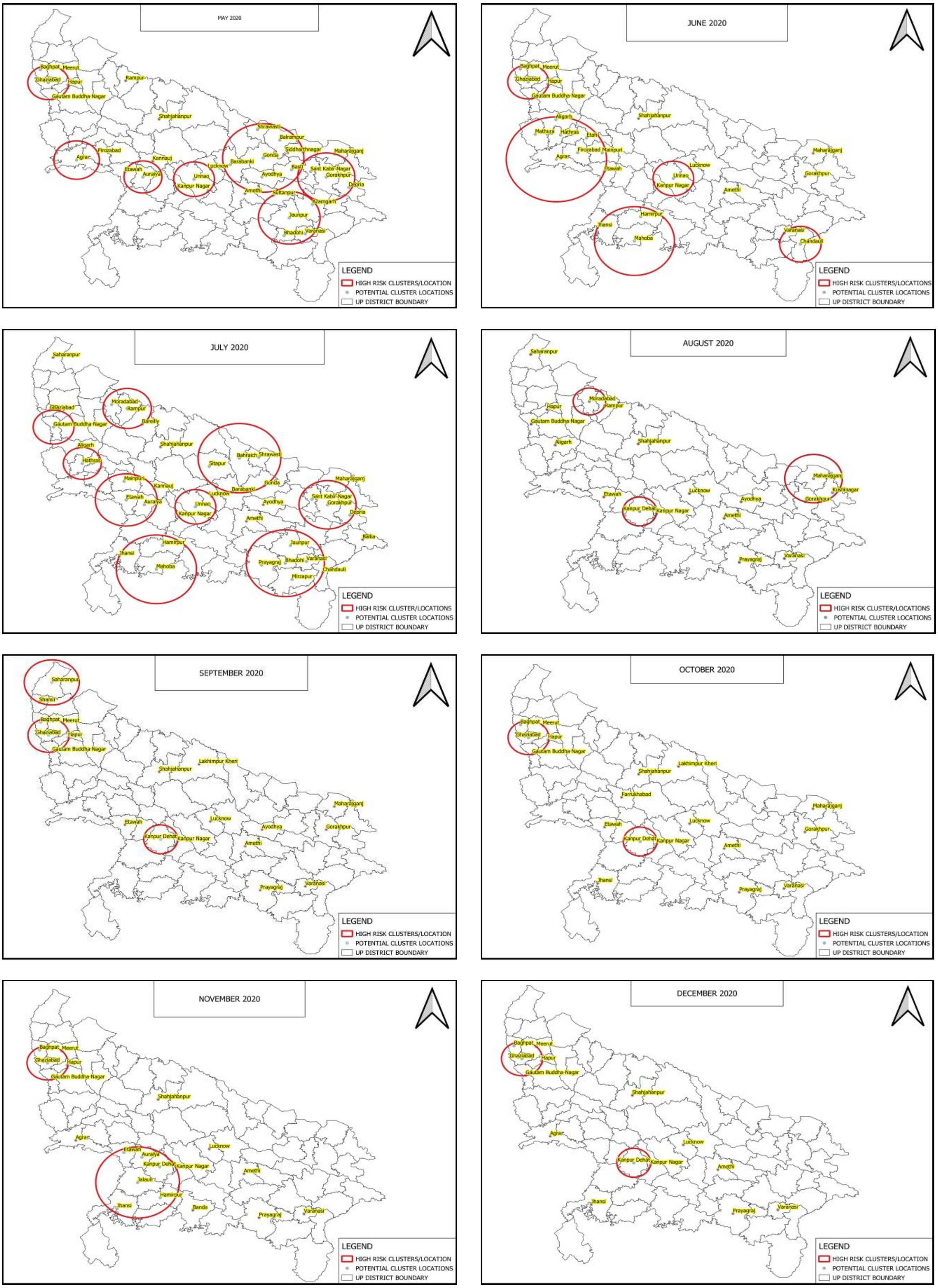

Figure 6. COVID-19 clusters of the study area from May 2020 to December 2020

important centres of disease spread. Only two clusters are seen, and the rest are isolated locations suggesting borders should be kept close to contain the disease in specific locations only.

7. DISCUSSION

This paper uses a prospective space-time scan statistic method to detect emerging clusters of COVID-19 in the Uttar Pradesh state at the district level, providing results during eight months of study. To our knowledge, this study is the first to utilise space-time scan statistics to determine clusters of COVID-19 in UP. The method is very useful for monitoring the real-time situation to make informed and scientifically backed-up decisions. The prospective approach utilised in this study can be useful for state and local health departments to monitor the outbreaks in a timely fashion. The main strength of the prospective approach is the ability to add updated COVID-19 counts and rerun the statistic to identify new clusters and see the dynamics of disease spread. Despite the strengths of our study, there are limitations worth mentioning. First, due to the unavailability and uncertainty of the disease, probable and vulnerable cases were not taken, and only daily case data was used. Therefore, the true magnitude of the COVID-19 pandemic will not be known for some time. Second, localised or town/village/colony-wise surveillance could not be taken up due to the availability of aggregated data. As SaTScan is an exploratory statistic, this would have been a true representation of this analysis to understand the transmission dynamics of the current and future emerging clusters. Third, COVID- 19 is more severe for the elderly, kids, and those with pre-existing medical conditions. Future studies can implement other methodologies like case/control cluster techniques with death and case counts (e.g. space-time Bernoulli models). Few other covariates like age, sex, race, and social indicators can also be infused to assess the spread of the disease for a more detailed analysis. Although lockdown was in place, the intra-district movement was still happening for health facilities or other needs, due to which disease transmission could not be prevented. A more detailed study needs to be done to see the effect of steps taken by the authorities to understand and better equip for future spread. This analysis was carried out monthly, but it can be extended to assess daily cases where updated, and retiring clusters can be carefully studied to contain the spread.

8. CONCLUSIONS

The use of space scan statistics for the study of COVID-19 has been demonstrated in this work. This analysis is performed for the Uttar Pradesh state of India from May 2020 to December 2020. It is seen that from May 2020 to August 2020 that the LLR value varied from 6.727 to 18916.289, and RR varied from 691.24 to 1.27. In September, the LLR and RR values are reported to be highest, reaching 27772.205 and 104.92, respectively. These results are in sync with the reported cases of COVID-19, shown in the cluster maps. Post-September 2020, there is a decline in the values of the LLR and RR. More attention should be paid to the areas where the highest LLR values occurred to stop further infection. The virus can only be controlled when the chain of transmission is stopped; hence, the regions can be considered for lockdowns based on scientific evidence as suggested by this outcome. The regions with higher COVID-19 cases are vulnerable to rapid disease spread due to congested spacing, high population density and more economic activities. As seen in the outcomes of the monthly cluster analysis, the clusters were concentrated in the same geographical regions suggesting that better resources need to be centred in these areas on a priority basis. Since the district authorities were given authority to devise their containment strategy, the maps prepared in this work can be very helpful in analysing the ground situation and doing the necessary follow-up. The prospective scan statistic method gave a clear picture of the high-risk clusters. This method is also ideal to see the changes in the formation of clusters along with the evaluation of mitigation strategies followed in any district. The mitigation strategy can be changed based on the outcome of space scan statistical analysis. Along with the cluster analysis, various graphs and maps are also generated to show the detail of COVID-19 cases and related issues. This will help to understand the dynamics of the disease spread. This research can be extended further to assess various other contextual factors associated with the disease spread and based on further availability of data on a location basis.

REFERENCES

Agrawal, S. & Gupta, R. D. (2017). Web GIS and its architecture: a review. *Arabian Journal of Geosciences*, 10(23), 1–13. doi:10.1007/s12517-017-3296-2

Cheng, T., & Adepeju, M. (2014). Modifiable Temporal Unit Problem (MTUP) and Its Effect on Space-Time Cluster Detection. *PLoS ONE*, *9*(6), e100465. doi:10.1371/journal.pone.0100465

Cole, M. A., Elliott, R. J. R., & Liu, B. (2020). The Impact of the Wuhan Covid-19 Lockdown on Air Pollution and Health: A Machine Learning and Augmented Synthetic Control Approach. *Environmental and Resource Economics*, *76*(4), 553–580. doi:10.1007/s10640-020-00483-4

Coleman, M., Coleman, M., Mabuza, A. M., Kok, G., Coetzee, M., & Durrheim, D. N. (2009). Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes. *Malaria Journal*, *8*(1), 68. doi:10.1186/1475-2875-8-68

Cordes, J., & Castro, M. C. (2020). Spatial analysis of COVID-19 clusters and contextual factors in New York City. *Spatial and Spatio-temporal Epidemiology*, *34*, 100355. doi:10.1016/j.sste.2020.100355

Desjardins, M. R., Hohl, A., & Delmelle, E. M. (2020). Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Applied Geography*, *118*, 102202. doi:10.1016/j.apgeog.2020.102202

Hohl, A., Delmelle, E. M., Desjardins, M. R., & Lan, Y. (2020). Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States. *Spatial and Spatio-temporal Epidemiology*, *34*, 100354. doi:10.1016/j.sste.2020.100354

Kandwal, R., Garg, P. K., & Garg, R. D. (2009). Health GIS and HIV/AIDS studies: Perspective and retrospective. *Journal of Biomedical Informatics*, *42*(4), 748–755. doi:10.1016/j.jbi.2009.04.008

Kang, D., Choi, H., Kim, J.-H., & Choi, J. (2020). Spatial epidemic dynamics of the COVID-19 outbreak in China. *International Journal of Infectious Diseases*, *94*(January), 96–102. doi:10.1016/j.ijid.2020.03.076

Kim, S., & Castro, M. C. (2020). Spatiotemporal pattern of COVID-19 and government response in South Korea (as of May 31, 2020). *International Journal of Infectious Diseases*, *98*, 328–333. doi:10.1016/j.ijid.2020.07.004

Krzysztofik, R., Kantor-Pietraga, I., & Spórna, T. (2020). Spatial and functional dimensions of the COVID-19 epidemic in Poland. *Eurasian Geography and Economics*, *61*(4–5), 573–586. doi:10.1080/15387216.2020.1783337

Kulldorff, M. (1999). Spatial Scan Statistics: Models, Calculations, and Applications. In J. Glaz & N. Balakrishnan (Eds.), *Scan Statistics and Applications*. Boston, MA: Birkhäuser, Boston, MA. doi:10.1007/978-1-4612-1578-3

Kulldorff, M., Heffernan, R., Hartman, J., Assunção, R., & Mostashari, F. (2005). A Space–Time Permutation Scan Statistic for Disease Outbreak Detection. *PLoS Medicine*, *2*(3), e59. doi:10.1371/journal.pmed.0020059

Kumar, S., & Agrawal, S. (2020). Prevention of vector-borne disease by the identification and risk assessment of mosquito vector habitats using GIS and remote sensing: a case study of Gorakhpur, India. *Nanotechnology for Environmental Engineering*, 5(2), 1-15, doi: 10.1007/s41204-020-00084-y

Li, W., Batty, M., & Goodchild, M. F. (2020). Real-time GIS for smart cities. *International Journal of Geographical Information Science*, *34*(2), 311–324. doi:10.1080/13658816.2019.1673397

Linton, S. L., Jennings, J. M., Latkin, C. A., Gomez, M. B., & Mehta, S. H. (2014). Application of Space-Time Scan Statistics to Describe Geographic and Temporal Clustering of Visible Drug Activity. *Journal of Urban Health*, *91*(5), 940–956. doi:10.1007/s11524-014-9890-7

Lyseen, A. K., Nøhr, C., Sørensen, E. M., Gudes, O., Geraghty, E. M., Shaw, N. T., & Bivona-Tellez, C. (2014). A Review and Framework for Categorizing Current Research and Development in Health Related Geographical Information Systems (GIS) Studies. *Yearbook of Medical Informatics*, *23*(01), 110–124. doi:10.15265/IY-2014-0008

Mahdizadeh Gharakhanlou, N., & Hooshangi, N. (2020). Spatio-temporal simulation of the novel coronavirus (COVID-19) outbreak using the agent-based modeling approach (case study: Urmia, Iran). *Informatics in Medicine Unlocked*, *20*, 100403. doi:10.1016/j.imu.2020.100403

Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of The Total Environment*, *728*, 138884. doi:10.1016/j.scitotenv.2020.138884

Mulatti, P., Mazzucato, M., Montarsi, F., Ciocchetta, S., Capelli, G., Bonfanti, L., & Marangon, S. (2015). Retrospective space–time analysis methods to support West Nile virus surveillance activities. *Epidemiology and Infection*, *143*(1), 202–213. doi:10.1017/S0950268814000442

Murugesan, B., Karuppannan, S., Mengistie, A. T., Ranganathan, M., & Gopalakrishnan, G. (2020). Distribution and Trend Analysis of COVID-19 in India: Geospatial Approach. *Journal of Geographical Studies*, *4*(1), 1–9. doi:10.21523/gcj5.20040101

Pourghasemi, H. R., Pouyan, S., Heidari, B., Farajzadeh, Z., Fallah Shamsi, S. R., Babaei, S., et al. (2020). Spatial modeling, risk mapping, change detection, and outbreak trend analysis of coronavirus (COVID-19) in Iran (days between February 19 and June 14, 2020). *International Journal of Infectious Diseases*, *98*, 90–108. doi:10.1016/j.ijid.2020.06.058

Puri, S. K. (2006). Technological frames of stakeholders shaping the SDI implementation: A case study from India. *Information Technology for Development*, *12*(4), 311–331. doi:10.1002/itdj.20050

Rahman, M. R., Islam, A. H. M. H., & Islam, M. N. (2020). Geospatial modelling on the spread and dynamics of 154 day outbreak of the novel coronavirus (COVID-19) pandemic in Bangladesh towards vulnerability zoning and management approaches. *Modeling Earth Systems and Environment*, (0123456789). doi:10.1007/s40808-020-00962-z

Requia, W. J., Kondo, E. K., Adams, M. D., Gold, D. R., & Struchiner, C. J. (2020). Risk of the Brazilian health care system over 5572 municipalities to exceed health care capacity due to the 2019 novel coronavirus (COVID-19). *Science of The Total Environment*, *730*, 139144. doi:10.1016/j.scitotenv.2020.139144

Rushton, G. (2003). Public Health, GIS, and Spatial Analytic Tools. *Annual Review of Public Health*, *24*(1), 43–56. doi:10.1146/annurev.publhealth.24.012902.140843

Sarwar, S., Waheed, R., Sarwar, S., & Khan, A. (2020). COVID-19 challenges to Pakistan: Is GIS analysis useful to draw solutions? *Science of The Total Environment*, *730*, 139089. doi:10.1016/j.scitotenv.2020.139089

Takahashi, K., Kulldorff, M., Tango, T., & Yih, K. (2008). A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. *International Journal of Health Geographics*, *7*(1), 14. doi:10.1186/1476-072X-7-14

Tonini, M., Tuia, D., & Ratle, F. (2009). Detection of clusters using space - time scan statistics. *International Journal of Wildland Fire*, *18*(7), 830. doi:10.1071/WF07167

Tripathi, A.K., Agrawal, S. & Gupta, R.D. (2021). Comparison of GIS-based AHP and fuzzy AHP methods for hospital site selection: a case study for Prayagraj City, India. *GeoJournal* doi: 10.1007/s10708-021-10445-y

Tuia, D., Ratle, F., Lasaponara, R., Telesca, L., & Kanevski, M. (2008). Scan statistics analysis of forest fire clusters. *Communications in Nonlinear Science and Numerical Simulation*, *13*(8), 1689–1694. doi:10.1016/j.cnsns.2007.03.004