Big Data- and Artificial Intelligence-Based Hot-Spot Analysis of COVID-19: Gauteng, South Africa, as a case study

Benjamin Lieberman[1,*], Roy Gusinow[1,*], Ali Asgary[2], Nicola Luigi Bragazzi[3,4], Joshua Choma[1], Salah-Eddine Dahbi[1], Kentaro Hayasi[5], Deepak Kar[1], Mary Kawonga[6,7], Jude Dzevela Kong[3,8], Mduduzi Mbada[9], Bruce Mellado[1,11], Kgomotso Monnakgotla[1], James Orbinski[10], Xifeng Ruan[1], Finn Stevenson[1], and Jianhong Wu[3,4]

[1]School of Physics and Institute for Collider Particle Physics, University of the Witwatersrand, Johannesburg, Wits 2050, South Africa

[2]Disaster & Emergency Management, School of Administrative Studies and Advanced Disaster, Emergency and Rapid-response Simulation, York University, Toronto, Canada

[3]Department of Mathematics and Statistics, York University, Toronto, ON Canada

[4]Laboratory for Industrial and Applied Mathematics (LIAM), York University, Toronto, Ontario, Canada

[5]School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, Wits 2050, South Africa

[6]School of Public Health, University of the Witwatersrand, Johannesburg, Wits 2050, South Africa

[7]Gauteng Provincial Department of Health, South Africa

[8]Canadian Center for Diseases Modeling (CDM), York University, Toronto, ON Canada

[9]Gauteng Office of the Premier, South Africa
[10]Dahdaleh Institute for Global Health Research, York University, Toronto, Ontario, Canada
[11]iThemba LABS, National Research Foundation, P.O. Box 722, Somerset West 7129, South Africa
*These authors contributed equally to this work.

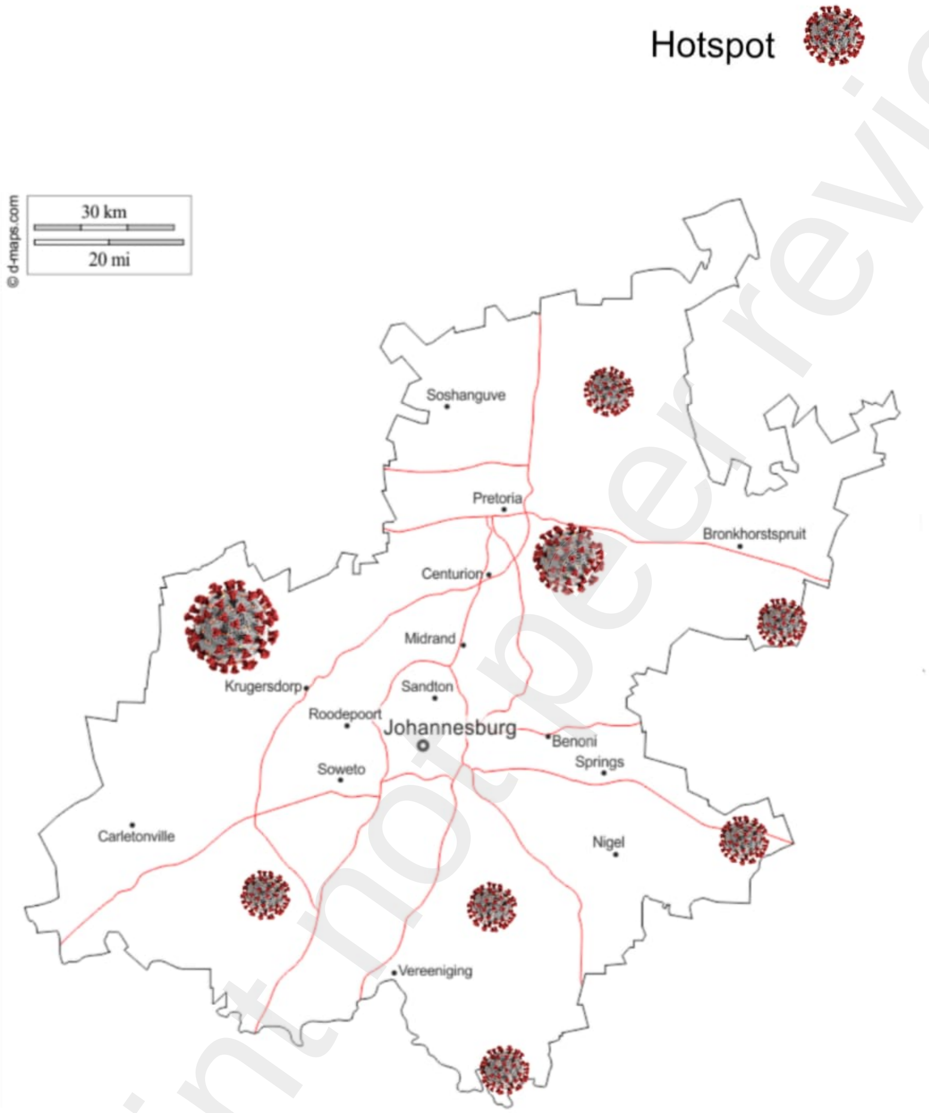
March 11, 2021

Abstract

"Coronavirus Disease 2019" (COVID-19) related data contain many complexities that must be taken into account when extracting information to guide public health decision- and policy-makers. In generalising the spread of a virus over a large area, such as a province, it must be assumed that the transmission occurs as a stochastic process. This statistically random spread of a virus through a population is the core of the majority of Susceptible-Infectious-Recovered-Deceased (SIRD) models and is dependent on factors such as number of infected cases, infection rate, level of social interactions, susceptible population and total population. However, the spread of COVID-19 and, therefore, the data representing the virus progression do not always conform to a stochastic model. In this paper, we have focused on the most influential non-stochastic dynamics of COVID-19, hot-spots, utilizing artificial intelligence (AI) based geo-localization and clustering analyses, taking Gauteng (South Africa) as a case study.

Keywords: COVID-19, South Africa, Gauteng Department of Health, Risk Adjusted Strategy, Control Interventions, Hot-Spot, Big Data, Artificial Intelligence

1 Graphical Abstract



2 Introduction

In late December 2019, a novel coronavirus, named "Severe Acute Respiratory Syndrome-related Coronavirus type 2" (SARS-CoV-2), emerged in the city of Wuhan, Hubei province of People's Republic of China (Sun et al. 2020). The virus rapidly spread by the 11^{th} of March 2020, resulting in a confirmed global pandemic, known as "Coronavirus Disease 2019" (COVID-19). As of the 5^{th} of March 2021, the virus is affecting more than 218 countries, with the total number of confirmed cases exceeding 116 million and approximately 2.6 million fatalities worldwide being attributed to the effects of the virus. A large, worldwide modeling effort is currently underway to improve public health policy decision-making with regards to the still ongoing COVID-19 pandemic (Mellado et al. 2021). Many research groups and national response teams have looked into country specific intervention strategies and the effects they have on the transmission rate of the virus as well as the impact of pre-existing country characteristics on the transmission rate (Duhon et al. 2020; Kong et al. 2021).

On the 5^{th} of March 2020, South Africa recorded its first COVID-19 case and three weeks later, on the 27^{th} of March, South Africa entered a full, government-enforced lockdown (Lone and Ahmad 2020). This formed part of a five tier risk-adjusted alert levels system (South Africa 2020). The full list of South Africa's moves between lockdown levels can be seen in Table 1 (Ramaphosa 2021). The first wave of COVID-19 continued in South Africa until October 2020 where the number of new cases had settled to a manageable amount. By late November 2020, South Africa's number of cases started to increase and the second wave of the pandemic began. The risk-adjusted system implemented allowed a controlled reopening/closing of the economy influenced by a set of factors, including the virus transmission rate, number of infectious cases, capacity of health facilities, the extent of the effectiveness of the implemented public health interventions and the economic and societal impact of continued restrictions.

Alert Level	Wave	Start Date	Total Cases	Recoveries	Fatalities
5	1	27 March	927	12	0
4	1	1 May	5951	2382	116
3	1	1 June	34357	17291	705
2	1	18 August	592144	485468	12264
1	2	21 September	661936	591208	15992
3	2	29 December	1021451	858456	27568

Table 1: South Africa's 2020 Alert Level Progression

The University of Witwatersrand and iThemba LABS COVID-19 modelling group have formed part of the Gauteng Premier's COVID-19 Advisory Committee, providing an in-depth analysis of the province's progress in the pandemic (Choma et al. 2020). As part of the Gauteng Premier's COVID-19 Advisory Committee, our modeling efforts provide information that government stakeholders use to inform their decisions, thus allowing a statistical grounds for changes in alert levels and distribution of resources.

COVID-19 data contain many complexities that must be taken into account when extracting information to guide public health decision- and policy-makers (Roda et al. 2020). This complexity includes factors such as the large number of misclassified or under-reported infections, inconsistency and limitations in testing as well as fluctuating infection and fatality rates as influenced by social/behavioral dynamics.

As this data is the basis for modeling and therefore, informing decisions around the risk-adjusted policies, understanding and accommodating these complexities in the model is vital. In generalising the spread of a virus over a large area, such as a province, it must be assumed that the transmission occurs as a stochastic process. This statistically random spread of a virus through a population is the core of the majority of Susceptible-Infectious-Recovered-Deceased (SIRD) models and is dependent on factors such as number of infected cases, infection rate, level of social interactions, susceptible population and total population (Choma et al. 2020). However, the spread of COVID-19 and therefore, the data representing the virus progression do not always conform to a stochastic model. In this paper, we will focus on the most influential non-stochastic dynamics of COVID-19 hot-spots.

A virus hot-spot can be defined as a cluster of cases within an area whose spreading dynamics do not conform to the general growth of the pandemic, exhibiting an exponential, short-lived growth. As these collections of cases do not conform to the macro-dynamics of the location, they need to be clearly defined and understood in order to accurately understand and model the virus progression. The geo-localization and clustering analyses of cases for this purpose are therefore, vital and can be done using advanced artificial intelligence (AI) geo-clustering methods. This clustering approach can therefore, be used to define individual clusters as hot-spots and allows the corresponding cases to be removed from the stochastic model - providing stochastic predictions that are not biased by the hot-spot dynamics (Nowzari et al. 2016).

3 Material and methods

3.1 Data

The data required for the hot-spot geo-localization analysis needs to be of a high level of detail. Therefore, for this paper anonymized data provided by the Gauteng Department of Health containing: Case ID, recorded address, test date and geo-localization data (including latitude and longitude coordinates). The data has been prepossessed to remove geo-localization data that has incorrect address recorded or issues interpreting/processing the address.

3.2 Clustering Cases by Geo-Location

In order to analyse the area distribution of COVID-19 cases, AI techniques provide an excellent tool in grouping cases geographically. In this paper we focus on the unsupervised machine learning method, using a Gaussian mixture model. This model allows us to analyse and model the dynamics of the virus within a determined area.

3.2.1 AI and Clustering: Gaussian Mixture Model (GMM)

The given problem is using the location of residence of each COVID-19 case in Gauteng to produce clusters. Once defined, these clusters can be analysed and accurately labelled as hot-spots or non-hot-spots.

There exists various clustering methods within AI through unsupervised machine learning algorithms that can be implemented to solve a 2-dimensional (latitude/longitude co-ordinates) problem, such as the present one. After evaluating various methods including the k-means algorithm, the Gaussian mixture model was chosen.

Gaussian Mixture models provide a probability-based approach to the likelihood of a COVID cases being within a cluster by producing a 2-dimensional Gaussian probability model overlayed onto the Gauteng map area. The clusters produced can overlap with each other, which encapsulates the possibility that hot-spots may very well also overlap with each other. The corresponding weight, ϕ, generated for each cluster, provides a simplistic estimate of the importance of the cluster, as well as another variable for filtering false clusters from actual hot-spots.

A Gaussian Mixture model is an algorithm which operates by generating k 2-dimensional Gaussian probability distributions, where k is a hyper-parameter specified. Thus, we are required to generate means, μ_j, covariance Σ_j and weighting, ϕ_j where the index specifies the $j-th$ Gaussian cluster. So, the probability of a new case, $p(x)$, occurring at a given point x is a linear combination of probabilities from all generated clusters:

$$p(x) = \sum_{j=1}^{k} \phi_i \, \mathcal{N}\left(x \mid \mu_j, \Sigma_j\right), \tag{3.1}$$

where $\mathcal{N}$ is the normal distribution. We generate the set of normal distributions (with associated weights, means and covariances) with an algorithm which optimally fits the probability distributions given the set of already known COVID-19 cases and their coordinates.

In order to generate k-Gaussian probability distribution, the **Expectation-Maximisation** algorithm is employed.

At the expectation step, we calculate the probability that a point x_i was generated by the j^{th} Gaussian for all k distributions:

$$\gamma_{ij} = \frac{\phi_j N\left(x_i \middle| \mu_j, \Sigma_j\right)}{\sum_{q=1}^{k} \phi_j N\left(x_i \middle| \mu_q, \Sigma_q\right)}. \tag{3.2}$$

In the maximisation step, the probabilities γ_{ij} are used to generate new cluster parameters. That is, new mean μ_j, co-variance Σ_j and weight ϕ_j are updated as follows:

$$\phi_j = \sum_{i}^{N} \frac{\gamma_{ij}}{N} \quad , \quad \mu_j = \sum_{i}^{N} \frac{\sum_i^N \gamma_{ij} x_i}{\sum_i^N \gamma_{ij}} \quad , \quad \Sigma_j^2 = \frac{\sum_i^N \gamma_{ij}(x_i - \mu_j)(x_i - \mu_j)^T}{\sum_i^N \gamma_{ij}}. \tag{3.3}$$

These steps are iterated until a convergence criteria is met. In our case, the variable $x = \{x, y\}$ is the set of longitudinal, y and latitudinal coordinate, x.

Once the latent variables of the Gaussian probabilities distributions (weights, means, standard deviation) have been found through the processing of COVID-19 cases in Gauteng, it is important to verify which clusters are hot-spots, or highly infectious areas/districts of the province. In order to accomplish this, the time dependent progression of cases is inspected for each cluster independently. That is, the cumulative number of cases was computed as a function of the date the patients were first recorded to have contracted the virus.

An aspect to consider is whether the clusters found follow the *Susceptible-Infection (SI) Curve*, which model the number of susceptible people who get infected, $SI(t)$, over time, within a given area/cluster. The SI equation is shown in Equation 3.4:

$$SI(t) = \frac{SI_0}{1 + e^{-SI_1(t - SI_2)}}, \tag{3.4}$$

where SI_0 is the total number of predicted cases within a cluster once it has saturated the susceptible population, SI_1 represents the rate of infection of the virus, and SI_2 is the number of days before the peak of growth of the cluster. This function is a solution to the logistic differential equation, a simple system which describes the number of infected cases in a given population. The model is applicable as we expect a small increase of infection cases in the early stages of a susceptible population, and then a sharp increase as the disease spreads rapidly throughout the cluster. A plateau is expected once all susceptible people within a cluster are infected.

The SI curve can therefore, be fitted to the time-series of each cluster in order to generate these parameters for the j^{th} cluster, giving more properties to accurately filter clusters into hot-spots. A poorly fit SI curve can indicate that the cluster is not a COVID-19 hot-spot, as it does not follow an accurate description of disease spread.

Once the cases throughout Gauteng province have been clustered and described, each area can be described through the following parameters; Total Cumulative Cases (N_{TC}), 1^{st} and 2^{nd} standard deviation area (A_{1sd} and A_{2sd}, respectively) and the susceptible-infection parameters (SI_0, SI_1 and SI_2).

4 Results and Discussion

4.1 Gauteng Province First Wave Cluster Analysis and Hot-Spot Definition

The density distribution of the clusters, shown in Figure 1, forms a Gaussian like shape at densities from 0 - 350 $\frac{cases}{km^2}$ followed by a sporadic tail of densities of 350 to more than 30000 $\frac{cases}{km^2}$. The uniformity of low density clusters is found to be associated with stochastic growth. Thus by cutting the densities at the two Sigma interval we are able to produce a density threshold of 196.05 $\frac{cases}{km^2}$. Clusters with densities greater than the threshold (denoted $\rho_c(t)$) are found to have rapid, non-stochastic growth. This density threshold therefore, allows us to define hot-spot clusters as any cluster whose density exceeds the determined density threshold.

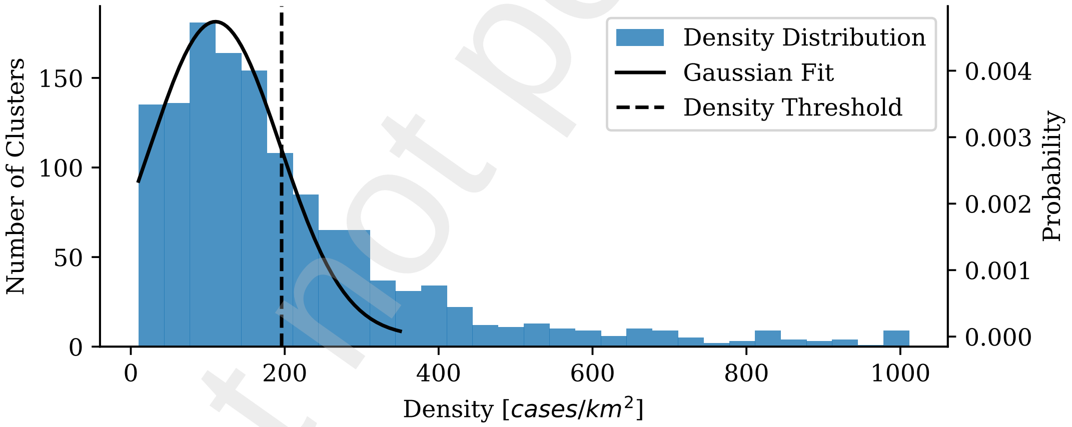


Figure 1: Gauteng Cluster Density Distribution with Gaussian Fit.

We applied this criterion to the first wave of COVID-19 in Gauteng Province where $\rho_{cluster}(t)$ is the case density of a given cluster at a given day and ρ_{th} is the minimum density stipulating hot-spot dynamics. Out of 1,500 clusters, once split on the density threshold 607 of the clusters are defined as hot-spots and the remaining 893 clusters are defined as stochastic.

In order to evaluate this definition further we compare the susceptible-infection parameters of the clusters defined as hot-spots against the stochastic or non-hot-spot clusters. Figure 2 shows that Hot-Spot clusters have on average an increased number of total cases, ±180, than the stochastic clusters, ±90. Hot-Spot clusters also have a slightly increased exponential slope with a period of ±10 days where stochastic clusters period of exponential slope can be seen to be ±11 days.

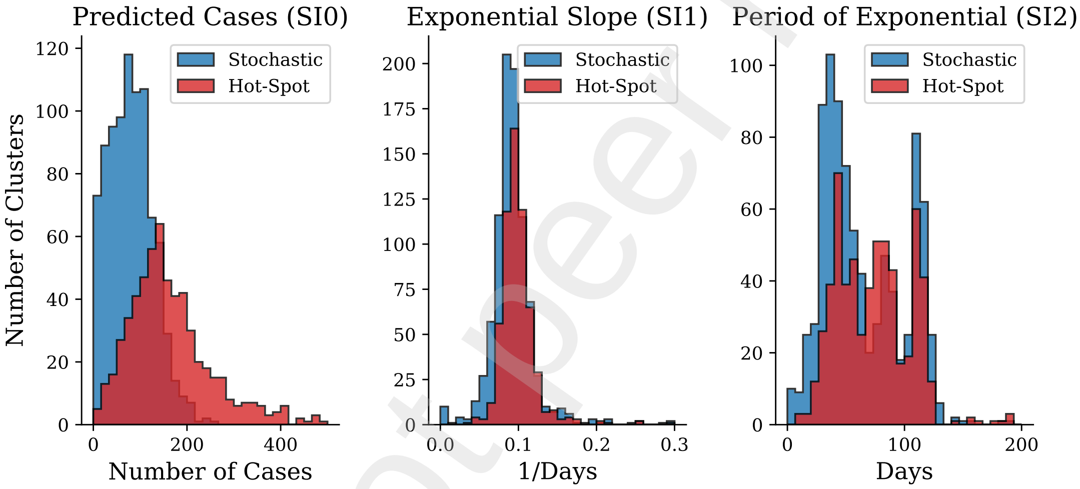


Figure 2: Gauteng Cluster parameter Distributions Comparison.

An evaluation of this hot-spot definition can be seen using a comparison of the total cases in stochastic clusters and hot-spot clusters during the first wave. Figure 3 reflects that during the first wave approximately two thirds of the cases in Gauteng occurred in hot-spot clusters.

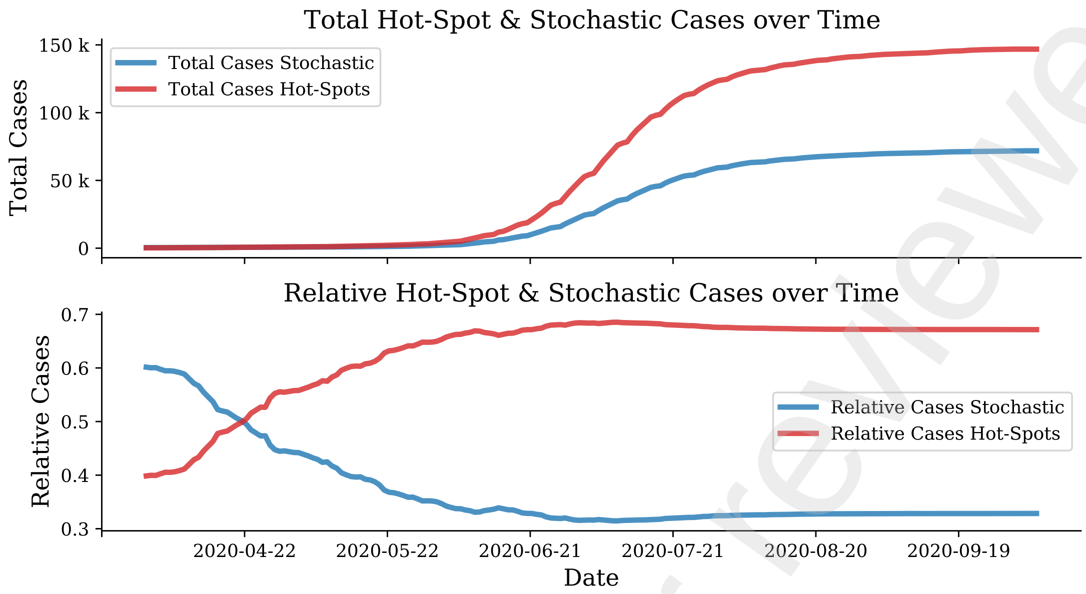


Figure 3: Number of Hot-Spot Cases Over Time During the First Wave.

This case distribution shows excellent coherence with first wave stochastic predictions (Using a Di-SIRD linear control model (Naude et al. 2020)) compared to data, as shown in Figure 4. This example of stochastic prediction demonstrates how the emergence of hot-spots in June 2020 did not follow the expected stochastic progression of the virus.

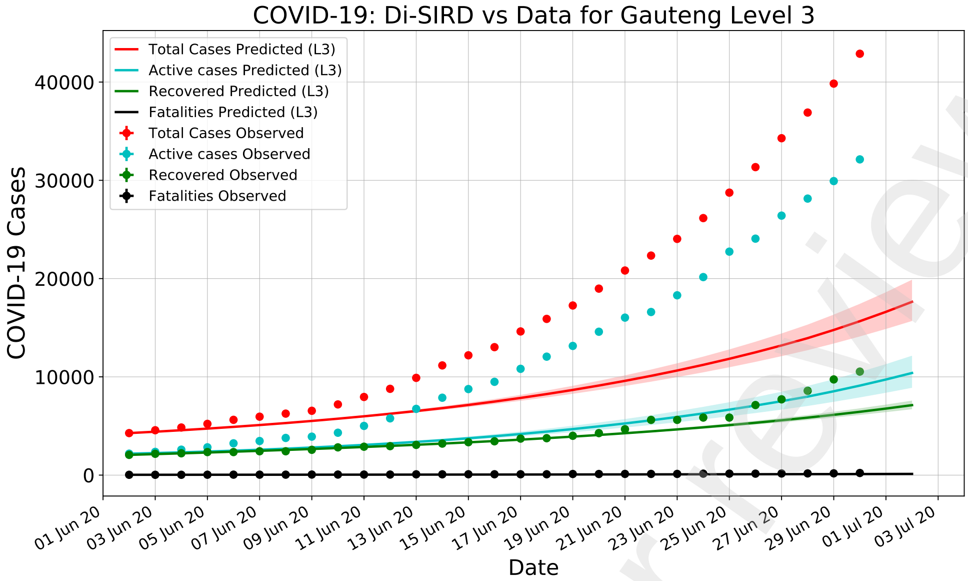


Figure 4: First Wave Stochastic Prediction Vs Data.

Therefore, it can be seen that the density cut-off value defining hot-spot clusters successfully is able to extract the clusters growing more exponentially and sporadically from those with a more uniform, random growth.

4.2 Hot-Spot Activity Analysis

Once a hot-spot cluster's total cases reaches the plateau or passes the peak of a surge, it can be said that the dynamics of the cluster is no longer that of a hot-spot. The activity of a cluster at any point in time can therefore, be quantified as the ratio of the total cases in the cluster at the respective time divided by the total predicted cases of the cluster, SI_0, described in Criteria (4.5):

$$\frac{N_{TC}(t)}{SI_0} < L_{th}, \tag{4.5}$$

where the activity threshold, L_{th}, represents the upper bound on actively growing clusters using the ratio of Total Cases to Total Predicted Cases. As one would expect all the clusters that where defined as hot-spots during the first wave have returned to stochastic dynamics after the first wave completion. More specifically, we are able to determine an activity threshold. The activity threshold assumes that only 1% of clusters remain active in the subsequent period of the first wave with a corresponding proportional error. Therefore, as shown in the activity distribution, Figure 5, the activity threshold is determined to be 0.85.

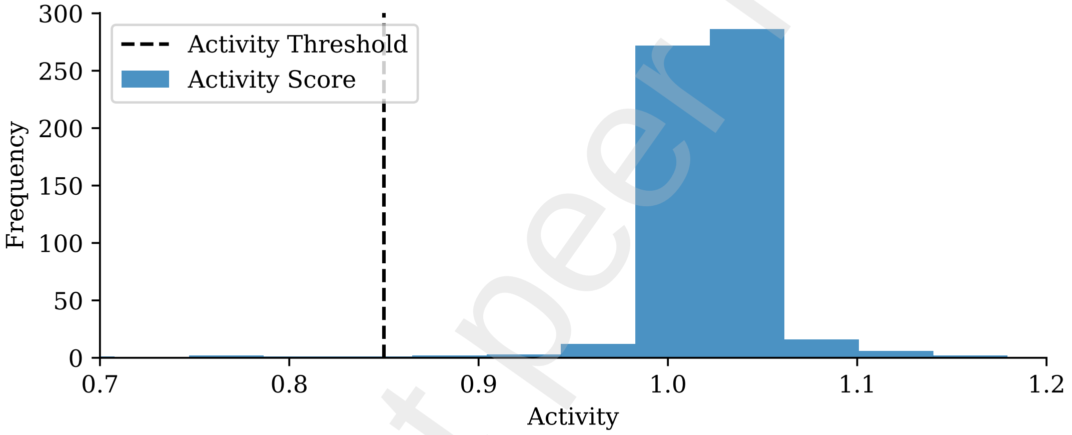


Figure 5: Gauteng Activity Distribution After First Wave Completion.

The time dependent evolution of newly defined hot-spots as well as hot-spots that are returning to stochastic dynamics in Gauteng during the first wave can be analysed using the above defined criteria. These dynamics are visualised in Figures 6 and 7, respectively.

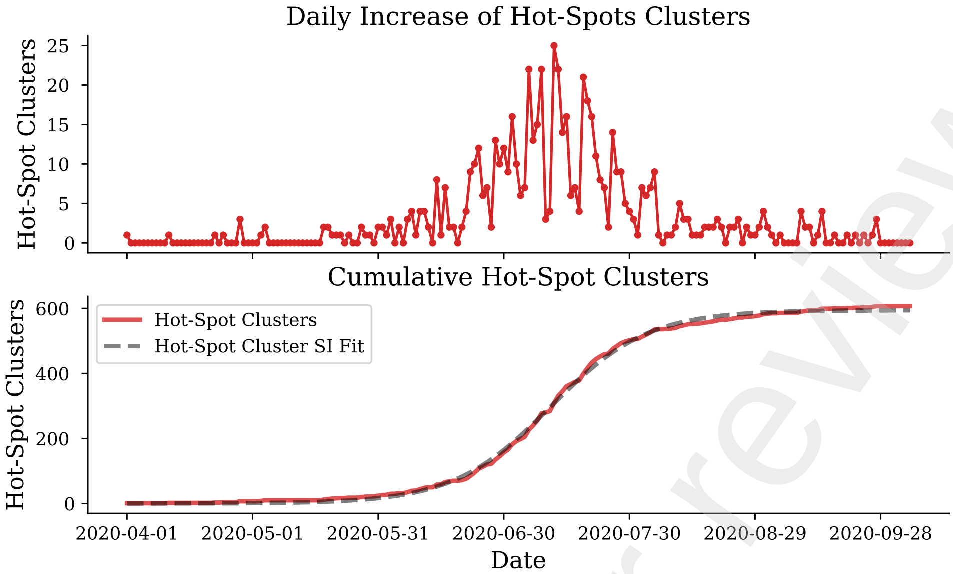


Figure 6: Cumulative and Emerging COVID-19 Hot-Spot Clusters in Gauteng.

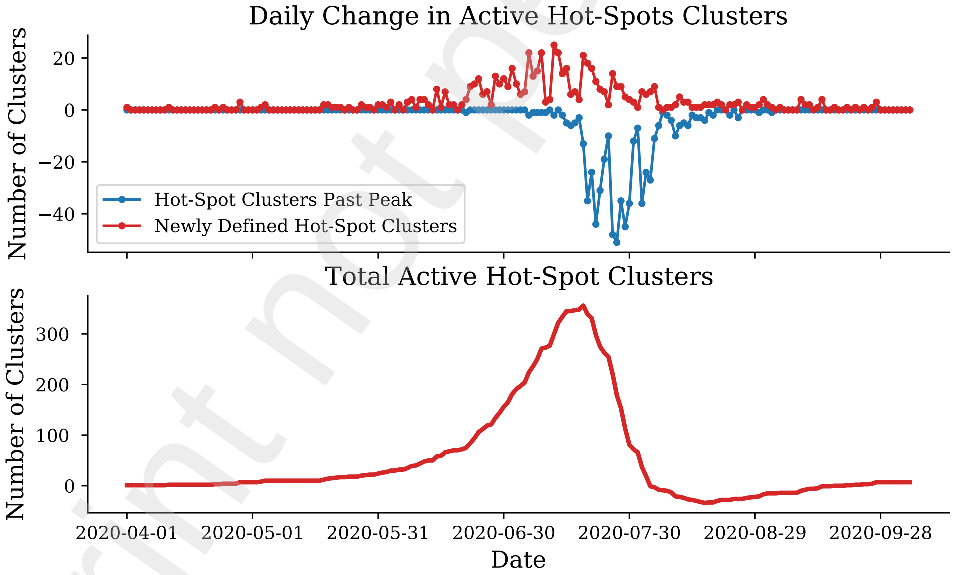


Figure 7: Number of Active Hot-Spot Clusters.

246 To understand the growth of the hot-spot clusters an SI curve is fit to the

cumulative number of hot-spot clusters shown in Figure 6. The daily increase of hot-spot clusters peaks in mid July, which is confirmed by the SI_2 parameter which determines the inflection of exponential growth to occur on the 10th of July, 101 days after the 1st of April. The cumulative hot-spot clusters reaches its plateau in mid August coinciding with South Africa's move from level 3 to level 2, with 594 of the total 1,500 clusters having already developed into hot-spots. The SI fit to the cumulative number of hot-spot clusters describes the period of the exponential growth to be approximately 12 days ($1/SI_1$).

Figure 7 shows not only the emergence of hot-spot clusters but also when hot-spots progress back to a stochastic dynamics, described by Equation 4.5. Clusters experiencing hot-spot dynamics start to reach their peak, and therefore, progress back to stochastic clusters, from mid July. By the end of August a maximum of 39 hot-spots have reached their peak and by the end of September all but 21 cluster have progressed back to stochastic dynamics.

4.3 Second Wave Risk Index Definition

Figure 8 shows the risk index at which a cluster can be defined as at risk in Gauteng Province.

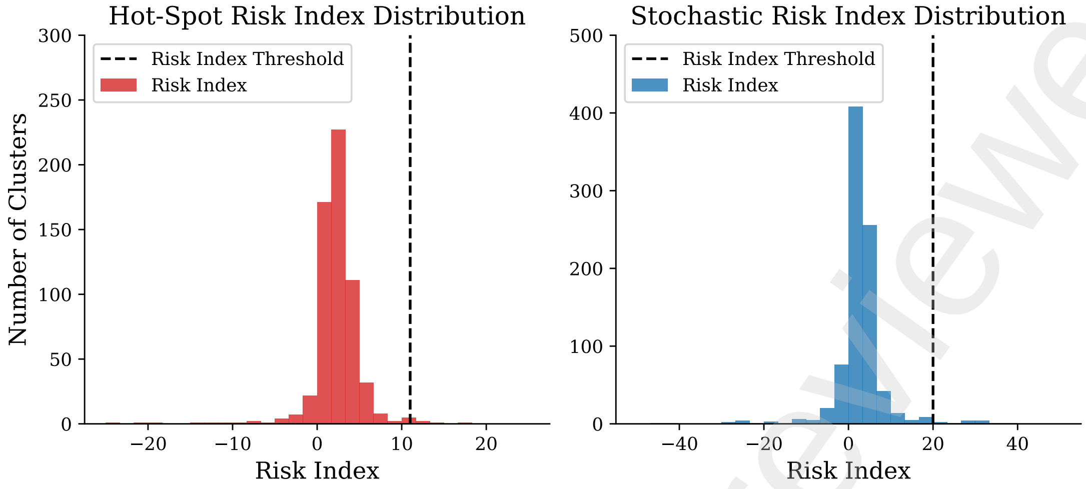


Figure 8: Gauteng First Wave Distributions of Second Wave Risk Index split into Hot-Spots and Stochastic Clusters.

Therefore, in the second wave analysis, a hot-spot cluster with a RI greater than 11 can be classified as a high risk hot-spot. Similarly a stochastic cluster with a RI greater than 20 can be classified as a developing high risk cluster.

4.4 Applications of Hot-Spot Definition for Second Wave

The definition and parameterization of clustered cases provides various applications in informing stakeholders in their decisions related to COVID-19 interventions and preventative measures. Section 4.4 discusses two of these applications. The first application allows for the hot-spot dynamics to be integrated into epidemiological models, while the second and more vital role is to expose potentially problematic areas in order to inform intervention strategies and advance social awareness and adoption of proper behaviors.

4.4.1 Implementation of Hot-Spot Analysis into susceptible -infected-recovered-death (SIRD) Model

A problem encountered in modeling the COVID-19 pandemic is that SIRD models generally function stochastically (random β dependent spread through susceptible population). However, pockets of cases developing usually in high density areas undergo independent, rapid infection that does not fit into larger model. This micro-system cluster is referred to as a hot-spot and undergoes independent non-stochastic hot-spot dynamics. In order to classify a specific group of cases in an area as a hot-spot the cases must first be grouped and their characteristics modeled, using each groupings characteristics to define a hot-spot cluster.

It therefore, follows that in order to produce informative predictions for governmental policy- and decision-makers, such as estimate numbers of hospital beds, use of intensive care units (ICUs) wards and when the peak will occur, the hot-spot cluster cases must be extracted from the data the stochastic SIRD model is calibrated on. The model then is able to interpret the progression of COVID-19 without the inconsistencies incurred by the non-conforming hot-spot cases.

This is done by extracting the daily ratio of stochastic cumulative cases from the total cases in all clusters and applying this ratio to the recorded data before it is used to inform the model:

$$I_{stoch} = \frac{I_s}{I_s + I_{hs}} * I_d, \tag{4.6}$$

where I_{stoch} is the stochastic active cases, I_s is the active cases in stochastic clusters, I_h is the active cases in hot-spot clusters and I_d is the active cases recorded.

Hot-Spot Classification	Cluster Activity	Risk Index
If cluster can be described as a hot-spot	At what point through its progression a cluster is	The severity of infection rate and scale of a cluster

Table 2: Summary of specifications of classified clusters

4.4.2 Exposing Hot-Spot and High Risk Clusters

The primary need for COVID-19 Hot-Spot classification is to target clusters/areas where non-conforming, exponential growth is occurring. Using the definition of hot-spot clusters developed in the previous sections, clusters can effectively be classified and their progression and dynamics described. Table 2 summarizes descriptive parameters of a classified cluster.

These three parameters describing each cluster are able to inform stakeholders not only on what areas are considered COVID-19 high growth areas but also the period of time the cluster will last and how severe the dynamics of the cluster is. This can then be visualised in an interactive map for stakeholders as shown in Figure 9. The colour code of the clusters visually displays the severity using the RI.

Emerging spatio-temporal hot-spot analysis is of crucial importance for public health policy- and decision-makers and can provide valuable information that would not possible to achieve with other techniques, enabling to capture specific clustering patterns in terms of particular districts and areas that would be otherwise classified as being at low risk for spreading COVID-19. Hot-spot analysis can complement classical epidemiological and surveillance approaches, shedding light on COVID-19 spatio-temporal trends and the possible evolution of its trajectories. Furthermore, the hot-spot analysis enables to easily visualize data in a way that

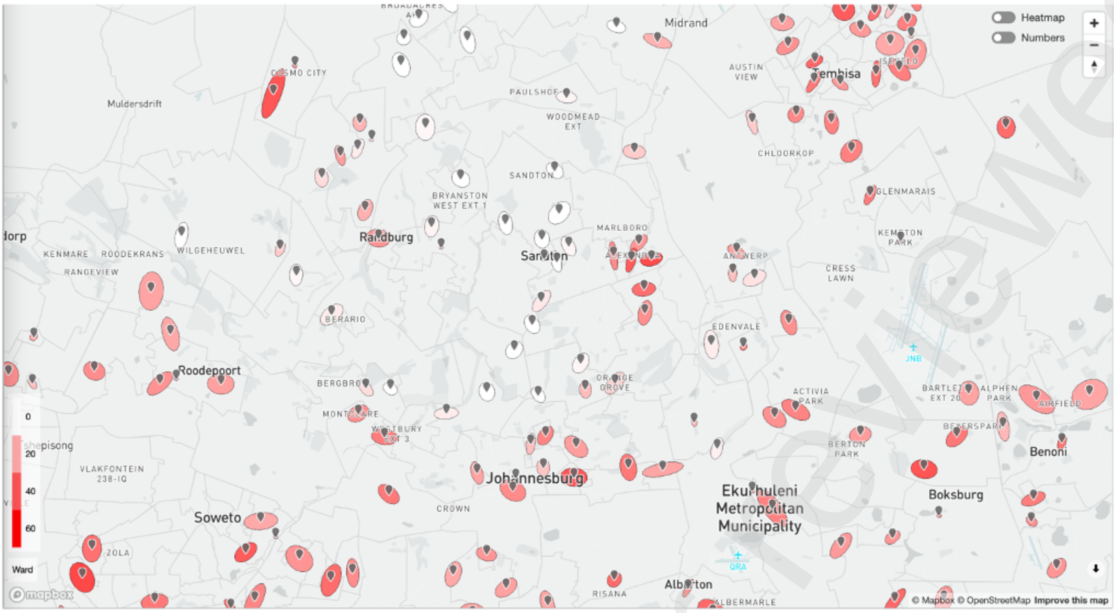


Figure 9: Hot-Spot visualisation on gpcoronavirus.co.za. Courtesy of IBM South Africa.

is accessible for stakeholders and helps them in the decision-making process.

In the existing scholarly literature, some studies have performed a hot-spot analysis of COVID-19. For instance, Shariati and colleagues (Shariati et al. 2020) have computed Anselin Local Moran's I indices to identify high- and low-risk clusters of COVID-19 worldwide. Authors were able to locate San Marino and Italy as territories characterized by a dramatically high toll of deaths, with infectious hot-spots widespread in northern Africa as well as southern, northern and western Europe. Noteworthy, infectious cases occurring in these hot-spots represent about 70 percent of all global infectious cases.

Other hot-spot analyses have been carried out at the nation level. Mo and coworkers (Mo et al. 2020) coupled local outlier analysis with hot-spot analysis based on space-time cube metrics in mainland China. Authors were able to demonstrate a rather quick, uneven spreading of the outbreak from the cities of Wuhan and Shiyan to the neighbouring areas and provinces. In Italy, combining a variety of geospatial analytical methods (spatial autococorrelation, spatio-temporal clustering and kernel density techniques), infodemiology (Google Trends and web searches analysis) and AI methods (machine learning and Adaboost algorithm for single-factor modelling), Niu and collaborators (Niu et al. 2020) were able to provide an in-depth assessment of the COVID-19 outbreak, in terms of its distribution and spreading characteristics. Hot-spots could be identified mainly in northern Italy.

Purwanto and colleagues (Purwanto et al. 2021) explored COVID-19 distribution patterns in East Java (Indonesia). Authors were able to identify Surabaya as major hot-spot, from which the outbreak reached cities characterized by high density of roads, food venues, and commercial and financial facilities.

In the present investigation, we have provided a robust statistical method for distinguishing between hot-spots and areas characterized by stochastic spreading of COVID-19 cases. We applied this analytical framework to the first and second waves, taking Gauteng province, South Africa, as a case study. These methods are general-purpose and can be, as such, applied to other countries as well.

Hot-spot analysis represents an advanced statistical approach that can be effectively utilized for outbreak analytics and visualization. It can equip public health policy- and decision-makers with updated, real-time assessment of the pandemic trends and its future projected trajectories. Furthermore, it can complement classical epidemiological surveys, leading to the identification of patterns that would be otherwise classified as low-risk ones. In conclusion, hot-spot analysis has been highly helpful in promptly recognizing high-risk clusters, and to adopt/adjust proper public health measures. Since the COVID-19 pandemic is a highly changeable and constantly under flux situation, we can anticipate that hot-spot analysis can aid stakeholders in making informed, evidence-based and data-driven decisions, while several countries are currently facing the third wave of the outbreak and are making efforts in vaccine roll-out.

Bibliography

Choma, J., Correa, F., Dahbi, S.-E., Dwolatzky, B., Dwolatzky, L., Hayasi, K., Lieberman, B., Maslo, C., Mellado, B., Monnakgotla, K. et al. (2020), 'Worldwide effectiveness of various non-pharmaceutical intervention control strategies on the global covid-19 pandemic: A linearised control model', *medRxiv* .

Duhon, J., Bragazzi, N. and Kong, J. D. (2020), 'The impact of non-pharmaceutical interventions, demographic, social, and climatic factors on the initial growth rate of covid-19: A cross-country study', *Science of The Total Environment* **760**, 144325.

Kong, J. D., Tekwa, E. and Gignoux-Wolfsohn, S. (2021), 'Social, economic, and environmental factors influencing the basic reproduction number of covid-19 across countries', *medRxiv* .

Lone, S. A. and Ahmad, A. (2020), 'Covid-19 pandemic–an african perspective', *Emerging microbes & infections* **9**(1), 1300–1308.

Mellado, B., Wu, J., Kong, J. D., Bragazzi, N. L., Asgary, A., Kawonga, M., Choma, N., Hayasi, K., Lieberman, B., Mathaha, T. et al. (2021), 'Leveraging artificial intelligence and big data to optimize covid-19 clinical public health and vaccination roll-out strategies in africa', *Available at SSRN 3787748* .

Mo, C., Tan, D., Mai, T., Bei, C., Qin, J., Pang, W. and Zhang, Z. (2020), 'An analysis of spatiotemporal pattern for coivd-19 in china based on space-time cube', *Journal of medical virology* **92**(9), 1587–1595.

Naude, J., Mellado, B., Choma, J., Correa, F., Dahbi, S., Dwolatzky, B., Dwolatzky, L., Hayasi, K., Lieberman, B., Maslo, C. et al. (2020), 'Worldwide effectiveness of various non-pharmaceutical intervention control strategies on the global covid-19 pandemic: A linearised control model', *medRxiv* .

Niu, B., Liang, R., Zhang, S., Zhang, H., Qu, X., Su, Q., Zheng, L. and Chen, Q. (2020), 'Epidemic analysis of covid-19 in italy based on spatiotemporal geographic information and google trends', *Transboundary and emerging diseases* .

Nowzari, C., Preciado, V. M. and Pappas, G. J. (2016), 'Analysis and control of epidemics: A survey of spreading processes on complex networks', *IEEE Control Systems Magazine* **36**(1), 26–46.

Purwanto, P., Utaya, S., Handoyo, B., Bachri, S., Astuti, I. S., Utomo, K. S. B. and Aldianto, Y. E. (2021), 'Spatiotemporal analysis of covid-19 spread with emerging hotspot analysis and space–time cube models in east java, indonesia', *ISPRS International Journal of Geo-Information* **10**(3), 133.

Ramaphosa, P. C. (2021), 'South africa's response to coronavirus covid-19 pandemic', `https://tinyurl.com/2hbrby83`. Accessed: 2021-03-08.

Roda, W. C., Varughese, M. B., Han, D. and Li, M. Y. (2020), 'Why is it difficult to accurately predict the covid-19 epidemic?', *Infectious Disease Modelling* **5**, 271–281.

Shariati, M., Mesgari, T., Kasraee, M. and Jahangiri-Rad, M. (2020), ‘Spatiotemporal analysis and hotspots detection of covid-19 using geographic information system (march and april, 2020)’, *Journal of Environmental Health Science and Engineering* **18**(2), 1499–1507.

South Africa, G. (2020), ‘South africa corona virus online portal 2020’, `https://sacoronavirus.co.za/covid-19-risk-adjusted-strategy/`. Accessed: 2021-03-08.

Sun, J., He, W.-T., Wang, L., Lai, A., Ji, X., Zhai, X., Li, G., Suchard, M. A., Tian, J., Zhou, J. et al. (2020), ‘Covid-19: epidemiology, evolution, and cross-disciplinary perspectives’, *Trends in molecular medicine* **26**(5), 483–495.