JDSSV

Journal of Data Science, Statistics, and Visualisation

November 2022, Volume II, Issue VII. *doi: 10.52933/jdssv.v2i7.46*

A Spatial SEIR Model for COVID-19 in South Africa

Inger Fabris-Rotelli
University of Pretoria

Jenny Holloway
Council for Scientific
and Industrial Research

Zaid Kimmie
Foundation for Human Rights

Sally Archibald
University of Witwatersrand

Pravesh Debba
Council for Scientific
and Industrial Research

Raeesa Manjoo-Docrat
University of Witwatersrand

Alize le Roux
Council for Scientific
and Industrial Research

Nontembeko Dudeni-Tlhone
Council for Scientific
and Industrial Research

Charl Janse van Rensburg
South African Medical Research Council

Renate Thiede
University of Pretoria

Nada Abdelatif
South African Medical
Research Council

Arminn Potgieter
University of Pretoria

Sibusisiwe Makhanya
IBM, South Africa

Abstract

The virus SARS-CoV-2 has resulted in numerous modelling approaches arising rapidly to understand the spread of the disease COVID-19 and to plan for future interventions. Herein, we present an SEIR model with a spatial spread component as well as four infectious compartments to account for the variety of symptom levels and transmission rate. The model takes into account the pattern of spatial vulnerability in South Africa through a vulnerability index that is based on socioeconomic and health susceptibility characteristics. Another spatially relevant factor in this context is the level of mobility throughout. The thesis of this study is that without the contextual spatial spread modelling, the heterogeneity in COVID-19 prevalence in the South African setting would not be captured. The model is illustrated on South African COVID-19 case counts and hospitalisations.

Keywords: COVID-19, SEIR model, spatial, excess deaths, South Africa, hospitalisations.

1. Introduction

South Africa is a large, diverse country with marked income inequality and differences in access to adequate housing, basic municipal services, transportation and medical care. Many of those affected by poverty also have increased morbidity risks due to tuberculosis (TB) and human immunodeficiency virus (HIV). These factors contribute to spatially diverse levels of vulnerability to the COVID-19 pandemic, which will result in limited accuracy if not taken into account when modelling.

At the onset of the COVID-19 pandemic, it remained uncertain how the spread of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV-2) would affect the South African healthcare system, particularly the public system. The ability of the government to provide quality health outcomes and to source necessary equipment were in the spotlight and dominated news discussions. The South African government proclaimed COVID-19 a national disaster on the 15^{th} of March 2020 and thereafter, on the 26^{th} of March placed the country into a national lockdown, which halted most economic activity and restricted personal movements by only permitting movement associated with essential activity. This most stringent alert level, that later became known as Level 5, was intended to contain the spread of SARS-CoV-2 and to adequately prepare the state, particularly the healthcare system, for the epidemic.

Modelling initiatives focusing on the key parameters, such as the potential magnitude of infections across the population and possible hospitalisation requirements, became critical as a means to assist government preparations. Hospitalisation requirements were a key component that would assist in evaluating whether the healthcare system would be able to cope with the cases needing hospital care, and identify where the impact would be most severe. Assessing these at different spatial scales was essential, as the country's hospital bed capacities and relevant resources vary across administrative boundaries. Most of the hospital-related care planning for COVID-19 occurred at provincial level. These included quarantine facilities, which involved decisions on building new field hospitals or refurbishing existing infrastructure and identifying quarantine sites. This paper argues for the incorporation of spatial spread into such modelling initiatives.

The Susceptible-Exposed-Infectious-Recovered (SEIR) compartmental model is a well-known epidemiological model, developed by Kermack and McKendrick (1991), for predicting the spread of a disease. As with most compartmental models, the underlying assumption is that each person in the model is equally likely to interact with any other person, referred to as ‘homogeneous mixing’. While this assumption is untrue for almost all cases of disease spread, it was important to account for it when assessing COVID-19 interventions as many of the interventions proposed involved creating spatial constraints to movement. For the case of South Africa, these interactions are spatially constrained and this must be accounted for. Moreover, in South Africa, due to high inequality and a history of spatial segregation there are large differences in the capacity of different segments of society to self-isolate, and therefore in their exposure to the disease. This makes modelling the pandemic at national or provincial scale unsatisfactory, due to small scale variation in a number of societal characteristics.

South Africa has deep-rooted inequalities causing stark differences in the quality and access of communities to basic and critical services such as healthcare, running water, sanitation, housing and social amenities. South Africa’s current GINI coefficient of 0.65 (StatsSA 2017) highlights these inequalities and shows this country to be amongst the most unequal countries globally with regards to income distribution. Wealth inequality is even starker with an estimated 10% of the population sharing close to 95% of all wealth (StatsSA 2017). With the onset of the COVID-19 pandemic in South Africa, this high level of inequality raised concerns amongst global leaders in the health sector about the detrimental effects that the pandemic might have on the society and economy at large, specifically those most vulnerable.

Aside from the differences in income and wealth distribution, the differences in housing conditions and access to basic services such as water and sanitation raised critical concerns about the transmission potential of the virus, and whether this would be skewed to higher densities in informal settlements with limited access to basic services. The health susceptibility of communities was also raised with regards to the stark spatial differences in access to quality health services and the underlying comorbidities present within communities.

Including mobility in the model is an important factor in South Africa even under varying lockdown approaches for the pandemic. The mobility data available for this research shows significant mobility still occurring under lockdown due to very limited access to daily needs such as food and income. A significant proportion of the South African population lives day-to-day and do not have savings or food stores to rely on. Earlier in 2020, the Al Jazeera news network brought to light an issue relevant to all developing countries when they published an article “In Africa, social distancing is a privilege few can afford”[1]. This statement makes it clear that many individuals who work and live in African countries do not possess the economic status to facilitate self-isolation. South Africa is one such country where the vast majority of the population lives under the poverty line (Finn et al. 2014). Certain challenges that impoverished South Africans have to contend with can increase their risk of COVID-19 infection. Given this fact it stands to reason that some areas will have higher transmission rates than others, thus necessitating the inclusion of a spatial element to models that attempt to model COVID-19 in South Africa.

This paper introduces methodology for a spatial SEIR model for COVID-19 in South Africa. The spatially diverse levels of risk at ward level[2] are incorporated with a

---

[1] `https://www.aljazeera.com/opinions/2020/3/22/in-africa-social-distancing-is-a-privilege-few-can-afford` [accessed on 31 March 2021]

[2] A ward is defined in South Africa as an administrative area for which larger municipalities are

South African vulnerability index as well as mobility data between wards obtained from cellular phone-based data on peoples' movement between wards. We add spatial elements into the compartmental epidemiological model. We advocate that any model for the spread of COVID-19 in South Africa should make use of a spatial element due to the societal structure within this developing country. Forecasting cases in South Africa using a single SEIR model at a national level does not allow one to take the spatial diversity present into account or respond effectively from a healthcare point of view. The aim, therefore, of creating the spatially modified SEIR model for South Africa is to investigate whether forecasts can be improved at a local level, by firstly tracking and predicting the spatial spread of the infection using the spatial location of cases, and secondly by bringing in factors specific to each area, such as vulnerability and population mobility.

Other solutions to the homogeneous mixing problem have included running individual-based simulation models. The solution proposed here is less computationally intensive but still, we argue, provides the necessary level of spatial detail for fighting an epidemic. This paper does not claim to provide a model that improves over all other modelling approaches of the COVID-19 pandemic. The aim is rather to illustrate that the proposed model captures the spatial heterogeneity inherent in the nature of the spread of the virus, or in fact any other similar disease, and to demonstrate the importance of incorporating these effects.

The paper proceeds with a literature review in Section 2, and follows with the methodology in Section 3. Section 4 discusses the implementation of the proposed spatial model, Section 5 presents the results, Section 6 provides a critical discussion and Section 7 concludes.

2. Literature Review

The use of SEIR type models in disease modelling is common, but in their simplest form these models assume homogeneous mixing of the population. Alternatives include stratified models. Stratified models with contact matrices between population strata are used to account for different populations in large areas and movement between areas. One of the first models to introduce mobility between spatial units via contact matrices is Sattenspiel and Dietz (1995). Generally the methods using contact matrices are applied when a small number of spatial units are being considered. The homogeneity assumption is relaxed by varying the contact rate between strata, commonly taken as age groups. Rost et al. (2020) follow the contact matrix approach in incorporating spatial dependency into the model for COVID-19 in Hungary. Mobile phone geolocation data has also been used to inform compartmental models, such as Peixoto et al. (2020) who estimate probabilities of movement between cities which is used to adjust the infected equation in a simple SI model. A more elaborate stochastic approach for estimating the mobility terms in the differential equations is followed in Arandiga et al. (2020).

The South African healthcare system consists of both the public and the private sys-

---

subdivided into. http://www.statssa.gov.za/census/census_2001/geo_metadata/geography_metadata.pdf tem. These are characterised by disproportionate spending on healthcare, medical care infrastructure, equipment and supplies, doctor-to-patient ratio as well as the quality of healthcare. The majority (over 80%) of South Africans rely on public healthcare while the healthcare requirements for the remaining (less than 20%) population are covered by private health insurance (Maphumulo and Bhengu 2019; Pillay 2009). The gap between these two healthcare sectors is widened by the disparities in the distribution of medical practitioners, with the public health system being under-resourced and overstretched in comparison to the private health system (Gray and Vawda 2019; Maphumulo and Bhengu 2019). According to the latest edition of the South African Health Review, South Africa has one of the lowest doctor-to-population ratios of 0.9 doctors per 1,000 people (Gray and Vawda 2019), compared to the ratio of 2.5 medical staff per 1,000 people, recommended by the World Health Organization[3]. In the South African private healthcare sector, this ratio is believed to be higher (Gray and Vawda 2019) even though the existing estimates lack consensus due to lack of data, making it difficult to arrive at a widely acceptable estimate. The latest available estimates suggest the ratio of doctor-to-population in the private sector to be around 1.75 doctors per 1 000 people (Competition Commission 2018). An increased doctor to population ratio is important for improved health outcomes (Bloor et al. 2006). Other public health challenges involve poor governance and management as well as the burden of diseases including HIV and TB. Approximately 7.5 million people were estimated to be living with HIV across all ages by the end of 2019 (Fronteira et al. 2021) and according to the National Institute of Communicable Diseases (NICD) surveillance report, around 4.3 million of them are on anti-retroviral treatment. Meanwhile, TB is estimated to be responsible for ill health of around 320 000 people annually and is a leading cause of death for about 80 000 people every year[4]. All these factors have continued to influence the quality of healthcare, particularly that of the public sector.

Urban transport in South Africa exhibits a large degree of spatial heterogeneity. This is largely due to the legacy of apartheid, during which city planning was geared towards restricting rather than facilitating access (Giddy 2019). Formal public transport has not proved sufficient to meet the needs of all lower-income South Africans. This has led to the development of paratransit systems, namely the minibus-taxi transport sector (Jennings and Behrens 2017). Paratransit systems are not formal (de Beer 2019) and arise ad hoc to fulfil local transport needs (Jennings and Behrens 2017). In line with the pre-existing patterns of city development, South African paratransit systems transport people from low-income residential areas such as informal settlements, generally located on the fringes of cities, to city centres providing work opportunities (Czeglédy 2004; Woolf and Joubert 2014), while middle-to-upper-class residents travel by private car (Giddy 2019; Woolf and Joubert 2013). People living in areas serviced by public transport, generally outlying low-income areas (Czeglédy 2004; Woolf and Joubert 2013), therefore have a higher transport-associated risk than those living in higher-income areas, such as suburbs and estates. This increases spatial heterogeneity in the risk of COVID-19 infection across South African cities.

---

[3] World Health Organization's Global Health Workforce Statistics, OECD, supplemented by country data, Physicians (per 1,000 people) - South Africa| Data `https://data.worldbank.org/indicator/SH.MED.PHYS.ZS?locations=ZA`

[4] UNAIDS, 2022. South Africa, `https://www.unaids.org/en/regionscountries/countries/southafrica`

The inherent poverty in communities also affects disease transmission. While some studies conclude that post-apartheid South Africa has seen an improvement with regards to the intensity of poverty (Fransman and Yu 2019), the country still faces a significant challenge regarding poverty and unequal distribution of resources (Finn et al. 2014), and poverty as well as general inequality vary spatially at a municipality level (David et al. 2018). In particular, areas that historically are known to have experienced higher intensities of poverty and inequality are still experiencing such issues (David et al. 2018). These areas are known as informal settlements, the South African term for what could generally be referred to using the umbrella term "slum" internationally. As per this definition, individuals who live in slums do not enjoy much free space and live in small dwellings where it is not necessarily possible to isolate themselves from other family members should they become infected with COVID-19 (Ezeh et al. 2017). Furthermore, the dwellings in these slums are so densely packed that the risk of infecting a neighbour is also very high. We thus expect the rate of transmission in such areas to be higher than more formally established settlements. Estimating the transmission risk in such areas poses a significant challenge since their actual population size may be unknown (Ezeh et al. 2017).

Individuals have been shown to seek medical assistance with varying degrees of urgency, with economic status and location of residency often being key factors (Lynch et al. 2017; Bassett et al. 2017). Some, also, will not seek medical attention due to additional costs associated with certain medical procedures, poor quality of medical services and even the potential stigma that comes from being suspected or confirmed to be infected with COVID-19 (Lynch et al. 2017; Bassett et al. 2017; Choonara and Eyles 2016; Bruns et al. 2020). Once again, these factors tend to affect poverty-stricken individuals more severely than their wealthier counterparts and thus their effect will vary spatially (Lynch et al. 2017; Choonara and Eyles 2016).

3. Materials and Methods

SEIR models are a type of compartmental model that are commonly used to mimic population dynamics, such as the spread of infectious diseases, utilising a series of interrelated mathematical equations. The basic SEIR model divides a relevant population of size N into four components that characterise subsets of the population in terms of their disease status. Individuals potentially start out as susceptible $S(t)$, with some of those becoming exposed $E(t)$ to the disease prior to becoming infectious $I(t)$ and subsequently moving into a state of recovery $R(t)$, with long-lived immunity (Li et al. 1999). In the aftermath of the COVID-19 worldwide pandemic, mathematical models like the SEIR model gained favour in the scientific community and were essential to public and policy discourse (Eker 2020). These models are important in guiding policy and decision making in public healthcare systems, as they are used to forecast, evaluate, and monitor the possible impact of a disease on the community, as well as to design approaches to regulate disease transmission (Chen 2014). Furthermore, these models may be used to examine the impact of non-pharmaceutical treatments, notably those implemented across many nations to combat the spread of SARS-CoV-2, or to explore relevant scenarios in complex disease systems (Lai et al. 2020). In South Africa, for example, various levels of lockdown (restricted mobility of individuals) were implemented

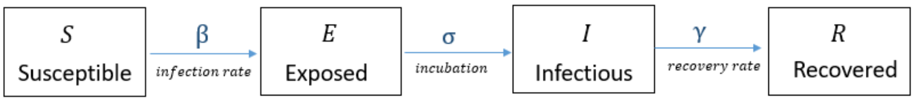


Figure 1: Schematic diagram of a basic SEIR model compartments and parameters

at various times to contain the spread of the disease and to prepare the hospital system for an influx of patients infected with SARS-CoV-2. It was critical in this investigation to measure the degree of SARS-CoV-2 transmission using an SEIR model at various levels of lockdown.

SEIR models are useful for modelling nonlinear patterns of disease development through stages such as time of onset, take-off, peak, and decline, as well as calculating the possible duration of an epidemic disease outbreak Iannelli and Pugliese (2014); Islam et al. (2020). Figure 1 illustrates the compartments in an SEIR model with the transition rates between them. Differential equations are used to estimate and monitor fractions of the original population as they transit from one compartment to the next in each time-step.

Depending on the process by which the disease is propagated, its features, and the assumptions made, several SEIR model extensions can be used. In circumstances where a disease's immunity is only temporary, these models can be expanded to allow the recovered population to revert to a susceptible condition (Li et al. 1999). A basic SEIR model may be specified by quantifying the rates of change in each of the four compartments, which can be represented as follows:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \tag{1}$$

Equation 1 assumes that the change in the susceptible S population from one time-step to the next is precisely proportional to the rate of transmission β (also called the infection rate) when contact is made with an infected individual. This is normalised by the population size, N. Initial contacts can result in secondary infections. These can be measured by the reproductive number, $R_0 = \frac{\beta}{\gamma}$, where γ is the recovery rate and therefore γ^{-1} is the average infectious period. Then R_0 estimates the average number of infections that an infected individual can generate.

Being in contact with an infected person leads to equation 2, where people who have been exposed to the illness will become infectious after a certain amount of time (the average incubation period is σ^{-1}).

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E \tag{2}$$

The transition of exposed persons to a state of infectiousness minus the number who recovered over that time period may be used to calculate the change in the number of infectious individuals (equation 3).

$$\frac{dI}{dt} = \sigma E - \gamma I \tag{3}$$

The rate γ at which the infectious individuals recover, makes it possible to estimate the number of persons recovering at any given moment (equation 4). In the current context, recovered can refer to both individuals who battle the sickness and those who die as a result of it.

$$\frac{dR}{dt} = \gamma I \tag{4}$$

The basic SEIR model has been adapted or extended in many cases to include additional compartments, for example in (Aldila et al. 2020; Zhang et al. 2020). We use an adaption in which we include four infected (I) components for four different types of infected individuals, separated according to the severity of their infection: Asymptomatic (I_1), mild (I_2), severe (I_3), and critical (I_4). For simplicity's sake, individuals only move through one of these four compartments for the infected based on what their final disease stage will be. Each of the infected compartment is allowed to vary in terms of the duration of its infectious period (γ_1^{-1}, γ_2^{-1}, γ_3^{-1}, γ_4^{-1}) and the asymptomatic class is assumed to be ρ times as infectious as all other classes ($\rho < 1$).

SEIR models simulate the spread of a disease for the entire population with an assumption of homogeneous mixing. When implemented at a national level, this does not mimic reality. In order to overcome this underlying assumption and to bring in localised factors, as discussed in Section 1, we add a spatial component to this adapted SEIR model to simulate the spread of COVID-19 through the 4392 different municipal wards of South Africa, using a daily time-step. This means that in each ward we model the course of the infection for that ward at a daily time-step, using the adapted SEIR compartmental model to simulate the dynamics specifically for that ward. In order to simulate the interaction of people between wards, exposed individuals from each ward are distributed amongst the wards, at the start of each time-step, based on movement data from cell phone service providers. This movement of individuals is discussed in more detail further on in this section. Population size of the wards ranges from 293 people to >100 000 people and within the wards people were assumed to mix homogeneously. Figure 2 shows this spatial model in detail with wards $i = 1, 2, ..., n$. The choice of spatial unit is important and in practice involves some compromise between granularity and computational cost. At one extreme, each individual could be located in their own spatial unit, essentially mimicking an agent-based model, while at the other extreme, one could have the standard compartmental model with only one spatial unit, as is the case with the national SEIR model.

The incubation period (σ^{-1}) and the infectious period for each infectious compartment (γ_1^{-1}, γ_2^{-1}, γ_3^{-1}, γ_4^{-1}) were kept the same across geographic wards. Taking $N_i = S_i + E_i + I_{1_i} + I_{2_i} + I_{3_i} + I_{4_i} + R_i$ as the population in ward i, with R_{0_i} as the ward-level R_0 and taking the infection rate, β_i, for that ward as $\beta_i = R_{0_i}\bar{\gamma}$, the flow between the main compartments in the model is implemented within each geographic ward i using the set of differential equations (5)-(11).

$$\frac{dS_i}{dt} = -\frac{\beta_i S_i(\rho I_{1_i} + I_{2_i} + I_{3_i} + I_{4_i})}{N_i} \tag{5}$$

$$\frac{dE_i}{dt} = \frac{\beta_i S_i(\rho I_{1_i} + I_{2_i} + I_{3_i} + I_{4_i})}{N_i} - \sigma E_i \tag{6}$$

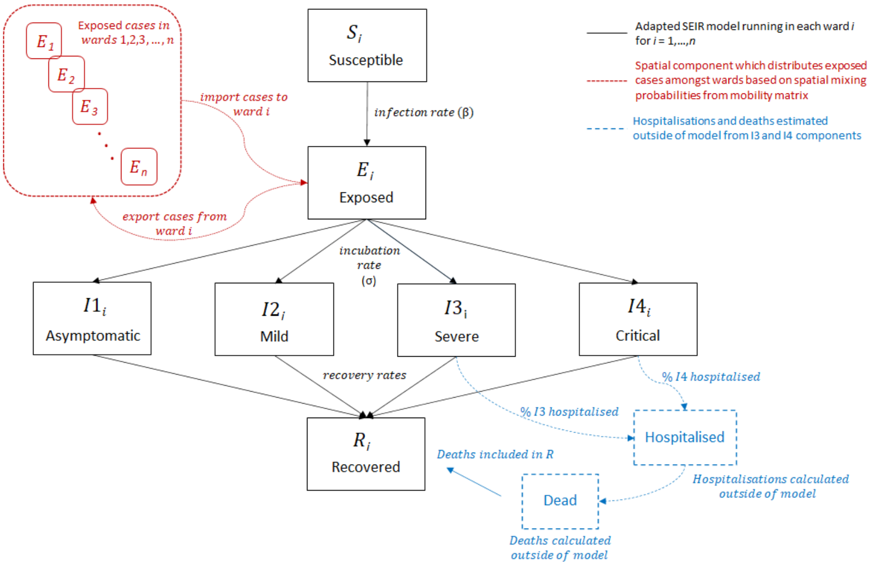


Figure 2: The spatial SEIR model used in this study, indicating the movement of the Exposed class between wards $i = 1, 2, ..., n$.

$$\frac{dI_{1_i}}{dt} = p_1 \sigma E_i - \gamma_1 I_{1_i} \tag{7}$$

$$\frac{dI_{2_i}}{dt} = p_{2_i} \sigma E_i - \gamma_2 I_{2_i} \tag{8}$$

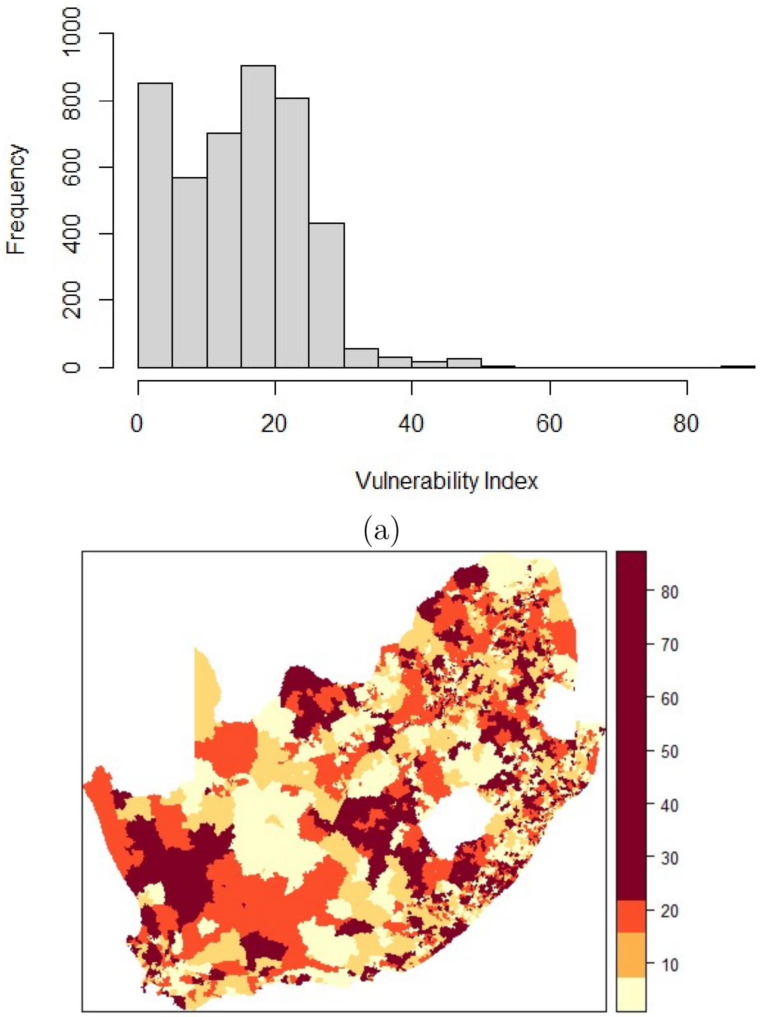
$$\frac{dI_{3_i}}{dt} = p_{3_i} \sigma E_i - \gamma_3 I_{3_i} \tag{9}$$

$$\frac{dI_{4_i}}{dt} = p_{4_i} \sigma E_i - \gamma_4 I_{4_i} \tag{10}$$

$$\frac{dR_i}{dt} = \gamma_1 I_{1_i} + \gamma_2 I_{2_i} + \gamma_3 I_{3_i} + \gamma_4 I_{4_i} \tag{11}$$

3.1. Vulnerability Index

The COVID-19 Vulnerability Index of the population is a composite indicator developed at the early mitigation and prevention phase of the South African government's emergency response to the pandemic (le Roux et al. 2020). In the early disaster management response phase, several sector departments required similar information with regards to the location and characteristics of highly vulnerable communities. The indicator (le Roux et al. 2020) was developed to facilitate a coordinated response by several government sectors with regards to prioritising intervention areas for water provision, sanitation upgrading, social interventions and community risk awareness. The indicator provides a spatial overview of communities that are highly vulnerable to COVID-19 based firstly on how effectively the spread of COVID-19 can be contained (the transmission potential), and secondly on the population's susceptibility to severe disease associated with contracting COVID-19 (the health susceptibility). The transmission potential highlighted areas and communities that would struggle to apply the basic principles of social distancing, hand washing and good basic hygiene by including spatial information on informal settlement areas, communities with a lack of access to



(a)

(b)

Figure 3: The vulnerability index used in the spatial SEIR model (a) Histogram of the vulnerability indices across the wards, (b) Vulnerability index per ward (colour legend representing quantiles)

basic services and areas of high population density. The health susceptibility of individuals was added to account for older populations and populations with a higher disease burden and inadequate access to medical care, by including data on age cohorts, comorbidities present and poverty levels (as a proxy for healthcare and the access thereof). About 93% of South African households have access to improved drinking water sources (piped water inside and outside the dwelling) (Matlala et al. 2017) and nearly 80% of households have access to safely managed sanitation services (StatsSA 2016). Figure 3 provides an illustration of the spatial vulnerability index across South Africa.

In order to use this COVID-19 Vulnerability Index in the model, it is normalised across all the wards and then scaled to have a range of 0.4 with this range centered around 1, resulting in a range of 0.8 to 1.2. It was applied by multiplying the national baseline R_0 by this factor to create an R_{0_i} for each ward $i = 1, 2, ..., n$. For highly vulnerable wards, the R_0 would increase, since their scaled vulnerability factor would be greater than 1, and for the less vulnerable wards the R_0 would reduce. The value of the ward-level modelling is that key parameters could be varied per ward based on the economic and social vulnerability of the people as well as interventions implemented (see Table 3 for the list of interventions and scaling factors). In addition, the proportions p_{2_i}, p_{3_i}, p_{4_i} of exposed people moving into the mild, severe and critical infected classes was adjusted based on the age structure distribution within each ward. For example, 98% of the symptomatic 0-9 year-olds were placed in the mild class while only 52% of the symptomatic individuals in the age 80+ class were placed in the mild class. The severity by age categorisation is given in Table 1 and this breakdown was derived from World Health Organisation data (`https://www.who.int/data`). These data were combined with the number of people in each age class in a ward to get overall percentages per severity class for each ward. Figure 4 also illustrates the age profile spatially across South Africa. There is a clear indication of a young population in South Africa.

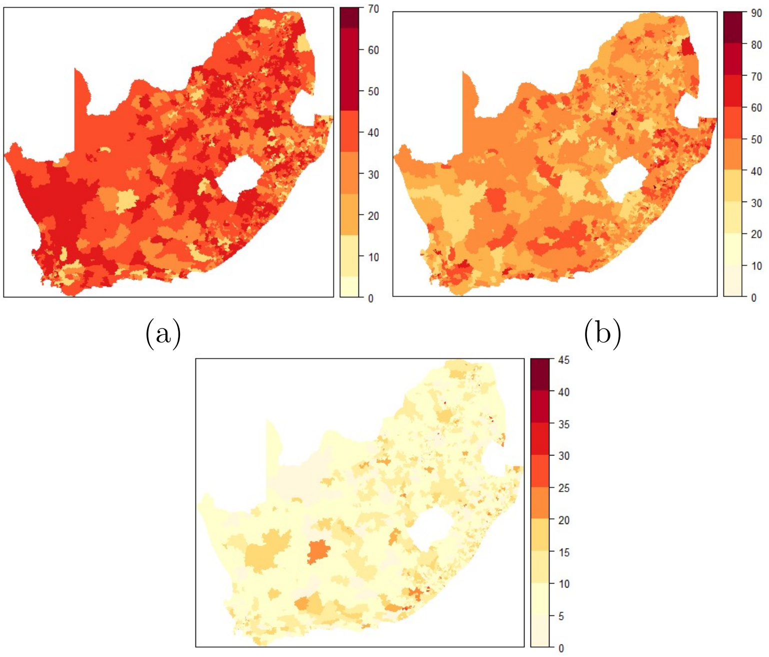
Table 1: Severity percentages per age class applied to symptomatic individuals

Age class	mild	severe	critical
0-9	98.0	2.0	0
10-19	98.0	2.0	0
20-29	90	8.8	1.2
30-39	85	12.6	2.4
40-49	79.0	17.0	4.0
50-59	75.0	19.2	5.8
60-69	69.3	23.0	7.7
70-79	59.7	28.2	12.1
80+	52.3	33.4	14.3

3.2. Mobility

The movement between wards is approximated by daily aggregate movements of the people based in that ward using cell phone location data. An individual is based in a geographic ward if their cell phone is located there between 8pm (of the previous day) and 4am (of the current day). For each ward (ward i with N_i residents say) for each day, we have a vector of length k (where k is the number of wards), where element j of that vector is the number of residents of ward i who appeared in ward j. We use this vector to derive a multinomial distribution (by dividing the vector by its sum) which represents the daily probability that persons from ward i appeared in any other ward.

This movement data was summarised by a k by k matrix representing the average probability of movement between wards for the different interventions imposed by the SA government during this time i.e., for each lockdown scenario one movement matrix was produced representing the average inter-ward movement during these restriction conditions (see Table 2). A movement matrix for a Business As Usual (BAU) scenario prior to government interventions was also created. Since cell phone data was not available for lockdown levels 1 and 2, level 2 was estimated as a slightly constrained version of the BAU mobility matrix and the BAU matrix was used for level 1.



(a) (b)

(c)

Figure 4: Proportions of age groups per ward (a) 0-19 years, (b) 20 - 59 years, (c) 60+ years

Table 2: National interventions used in the model

Intervention	Date	Scaling Factor
Business-as-usual (BAU)	No specific date - used for testing scenarios	1
Hard lockdown-level 5 (LD)	26 March-30 April 2020	0.6
Level 4 lockdown (L4)	1 May-31 May 2020	0.7
Level 3 lockdown (L3)	1 June-17 August 2020	0.75
Level 2 lockdown (L2)	18 August-20 September 2020	0.8
Level 1 lockdown (L1)	21 September-29 December 2020	0.85

3.3. Model Assumptions

Details of the final parameters chosen for this model, and justification based on literature are presented in Table 3. Note that the parameters were not estimated due to the uncertainty in case numbers and testing accuracy.

Since the model was initially run in a real-time context rather than a post-pandemic context, only sources that were available up to June 2020 were used to calibrate the model. South African sources of parameter values were predominantly used since these had either been calculated on South African data or decided on by experts from the South African COVID-19 Modelling Consortium (SACMC)[5]. There is a large variation of possible values for R_0 (the number of secondary infections produced by one primary infection in a fully susceptible population) in the literature, and since the model is also highly sensitive to the R_0, choosing an appropriate value is difficult. Herein R_0 was chosen as the mean value from a systematic review of 81 papers from January to July 2020 which estimated R_0 values for the beginning of the pandemic, i.e., prior to the implementation of non-pharmaceutical interventions (Thiede et al. 2020).

The South African COVID-19 Modelling Consortium (SACMC) have also done regular updates of the predicted non-ICU and ICU beds required per province (see SACMC (2020) for example) although not all updates were released to the public. Using the spatial SEIR model, our aim was to see whether we could create reliable projections at a high spatial resolution so as to better understand the spatial spread of hospital requirements and expected peaks across the country. Various projections were done from several starting dates. It was found that using a starting date from too early on in South Africa's fight against the pandemic did not provide sufficient case data across the spatial units, with cases being limited to a few epicentres. For the purposes of this paper, the starting date of 1st June 2020 is used to show the model results since this represents the date at which South Africa moved to level 3 of lockdown, thus allowing for far more freedom of movement of people between spatial units (represented by the L3 mobility matrix). This was also at a point in the pandemic where cases were being detected in all areas across the country.

The main concern in South Africa has been hospital capacity, thus the projection of potential hospitalisation cases of COVID-19 for each spatial unit was investigated. An assumption had to be made regarding the proportion of severe (I_3) and critical (I_4) individuals that would be admitted to hospital. Since critical cases are defined to be the proportion requiring either ICU, ventilation or oxygen, 100% of these individuals are assumed to be hospitalised. Given the difficulty in accessing healthcare for a large portion of the South African population, as discussed in the background, it was assumed that not all individuals with severe cases of COVID-19 would report to hospital. It was therefore necessary to determine what proportion of our projected severe cases were likely to go to hospital. Studies from Italy (Reno et al. 2020) and the US (Jehi et al. 2020) have reported statistics of 20% and 21.1%, respectively, of symptomatic cases requiring hospitalisation, while a South African study, focusing on the asymptomatic spread, Anguelov et al. (2020), calculated an average of 4.02% requiring hospitalisation. The latter figure was based on the percentages of severe symptomatic cases per age group taken from Ferguson et al. (2020) and the age structure of South Africa's

---

[5] Details and reports at www.sacmcepidemicexplorer.co.za

Table 3: Base parameters for spatial SEIR model with supporting literature

Description	**Notation**	**Value Used**	**Literature**
National baseline reproductive number	R_0	$\text{gamma}(57.2, 0.05)$ with mean 2.86	
Proportion of asymptomatic cases	p_1	0.75	Anguelov et al. (2020); SACMC (2020)
Percentage of mild cases amongst symptomatic	$p_{2_i}/(1-p_1)$	ward-based with average 87.7%	
Percentage of severe cases amongst symptomatic	$p_{3_i}/(1-p_1)$	ward-based (avg: 10.0%)	
Percentage of critical cases amongst symptomatic	$p_{4_i}/(1-p_1)$	ward-based (avg: 2.3%)	
Relative infectiousness of asymptomatic cases	ρ	0.75	Ngonghala et al. (2020); Li et al. (2020); SACMC (2020)
Incubation period	$1/\sigma$	$\text{gamma}(2, 1)$ days	NICD (2020a)
Infectious period for asymptomatic cases	$1/\gamma_1$	$\text{gamma}(2, 7/2)$ days	SACMC (2020); Ferguson et al. (2020); Tang et al. (2020)
Infectious period for mild cases	$1/\gamma_2$	$\text{gamma}(2, 7/2)$ days	SACMC (2020); Ferguson et al. (2020)
Infectious period for severe cases	$1/\gamma_3$	7 days	SACMC (2020); Ferguson et al. (2020)
Infectious period for critical cases	$1/\gamma_4$	7 days	SACMC (2020); Ferguson et al. (2020)
Infectious period until hospitalisation		7 days	SACMC (2020); Palmieri et al. (2020); Ferguson et al. (2020)
Duration of hospital stay		12 days	SACMC (2020); CHSUFT (2020); Ferguson et al. (2020)
Relative reduction in R_0 due to national interventions		see Table 2	SACMC (2020)
Relative change in R_0 due to vulnerability		ward-based (range$[0.8, 1.2]$)	

population. A statistic close to this 4% was also used by SACMC in their May 2020 presentation (SACMC 2020). Although the proportions of symptomatic cases that fall in the severe and critical compartments in our spatial SEIR model vary per spatial unit based on the age structure in each area, on average across the country the critical cases constitute 2.5% of the symptomatic cases while the severe cases represent 10.7% of all symptomatic cases. These are clearly well below the 20% recorded in other countries, but well above the 4% estimated in the two South African studies mentioned. Using hospitalisation data that was available for the Gauteng[6] and Western Cape[7] provinces on 1st June 2020 (using a 7-day moving average) and comparing this to the confirmed cases (assuming for simplicity sake that confirmed cases were similar in number to symptomatic cases), we calculated the proportion of our severe (I_3) cases reporting to hospitals in Gauteng to be roughly 30%, while for the Western Cape this number was found to be only 10%. Together with the critical (I_4) cases, this represents 5.7% of all symptomatic cases for Gauteng and 3.6% for Western Cape. Although the population in Gauteng is on average a much wealthier population and has the highest access to medical care out of all the provinces in South Africa (StatsSA 2011), it is unclear as to why the proportion in the Western Cape was so much lower. One might surmise that it was due to their rigorous testing regime (NICD 2020c) in the early stages of the pandemic which actually resulted in a much higher detection ratio than elsewhere in the country.

Since the delay from initial infection to time of getting a confirmed test result was estimated to be about 7 days in the early stages of the pandemic (although this could be longer in some cases), the model was initialised with the number of confirmed active cases in each ward, taken from 7 days after the start date for the simulation run. It is well known that the confirmed cases are only a portion of the actual cases in the population, but the exact ratio is unknown and would depend on the testing strategy per country. An earlier study from Italy (Reno et al. 2020) used a ratio of 10 undetected cases to every 1 confirmed positive case while a more recent study (Böhning et al. 2020) done on several European countries, including Italy, calculated a ratio of 2.3. A USA study (Wu et al. 2020) using data up to the 18th April 2020 calculated that the number of infections was nationally 9 times higher than the reported cases, although this factor varied per state and region. We chose to use a ratio of 5 to 1, thus assuming that 20% of the actual infected people were tested and confirmed as cases.

4. Implementation

The spatial SEIR model is simulated over a number of runs and number of time-steps. The model is initialised with a number of cases in each compartment of the model. Then the model is simulated for subsequent time-steps, simulating the movement of cases between compartments and spatial units.

We assume, for simplicity, that the population of each ward is fixed, i.e., no births,

---

[6] Gauteng Province, Department of Health, 2020, COVID-19 Daily Media Reports on Gauteng COVID-19 Confirmed Cases District Breakdown. [Online] Available at: `https://www.gauteng.gov.za/Publications`. [Accessed June 2020 to November 2020]

[7] Western Cape Government. Update on the coronavirus by Premier Alan Winde. [Online] Available at: `https://coronavirus.westerncape.gov.za/news/`. [Accessed June 2020 to November 2020]

non-COVID-19 deaths or permanent movement between wards. At each time-step the number of new exposed people is produced by running the SEIR model for each ward. The new exposed are then randomly distributed over the current and other wards based on the mobility data, essentially using one draw from the appropriate multinomial distribution. This assumes that the spread of the virus matches mobility patterns, in particular that infections are equally likely to occur at any point of an individual's movements throughout the day. At each subsequent time-step the spatial allocation of exposed is taken into account by adding up the new exposed that have been allocated to that ward and subtracting the exposed that have been allocated elsewhere to provide the initial condition for the SEIR at that time-step.

Explicit in this formulation is the assumption that for a given ward, the ward level population parameters (number of susceptible, exposed, infected and recovered) drive the number of new exposed cases. This is clearly a simplification, since as individuals move out of their ward they interact with a wider group and other individuals may move into the ward for various time periods. This is more likely to be problematic at the end of the epidemic when some wards have low numbers of susceptible individuals while others are still highly susceptible. Moreover, the mobility data does not allow one to distinguish between transitory movements through a ward and extended stays in a ward. For example, wards which contain transport hubs and train stations may be allocated more infections than is realistic.

The model was run on facilities made available by the Centre for High Performance Computing (CHPC)[8] which consisted of 10 x 24 core machines. Four simulations were run per core per machine and hence the model was run a total of 960 times, each time with different parameter combinations as presented in Table 3 and random draws from the multinomial movement matrix. The final model outputs were summarised into data structures containing selected percentiles (10th, 25th, 50th, 75th, 90th) calculated across all simulations, as well as the mean and standard deviation, with values for each model compartment (S_i, E_i, I_{1i}, I_{2i}, I_{3i}, I_{4i}, R_i) given per ward and per day. We selected the mean for display purposes in the results section.

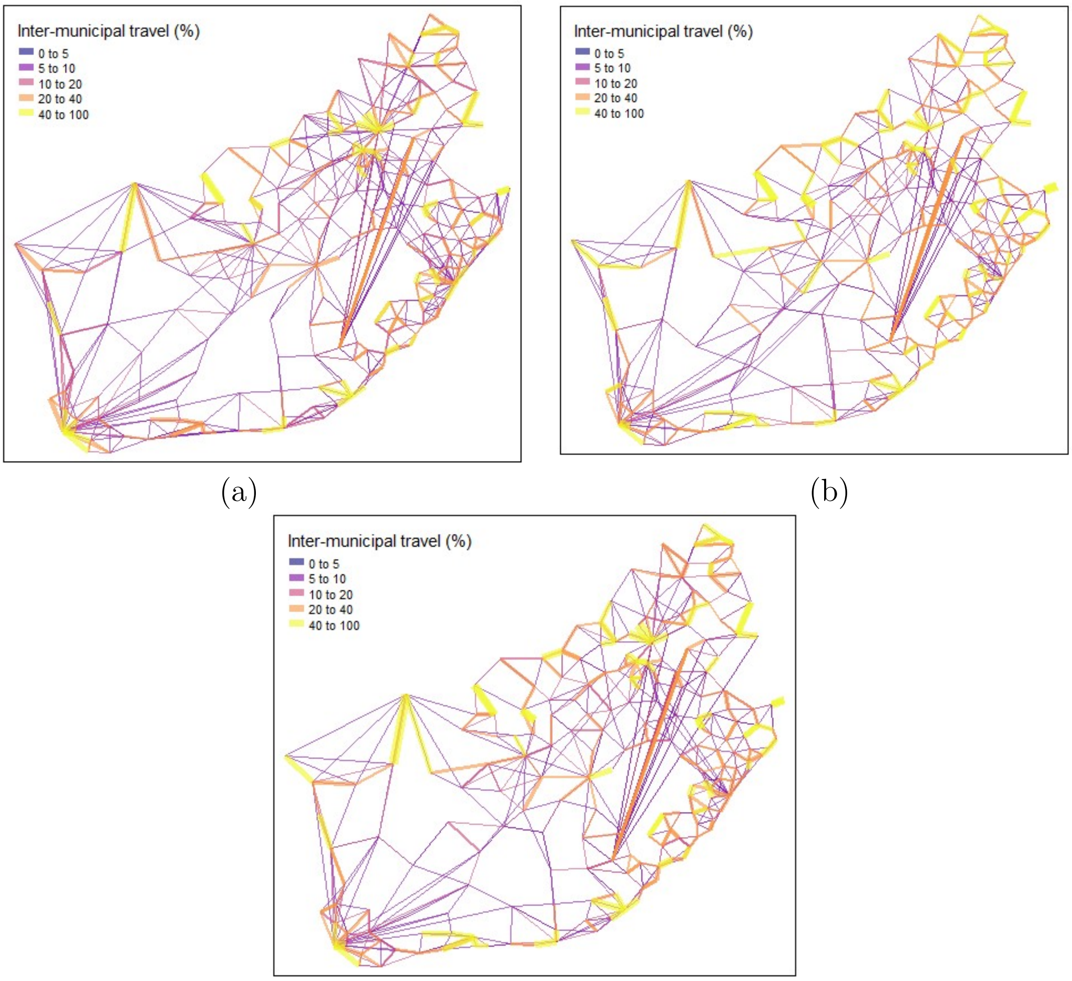
The data made use of in this paper consists of daily case data at ward level in South Africa from 6 March 2020 to 22 July 2020. In addition, hospitalisation data is freely available for two provinces, Gauteng and Western Cape for the period June 2020 up to mid-October 2020. In this paper we refer to hospitalisations as the number of cases occupying hospital beds on a daily basis i.e., capacity requirements, and not the number of daily hospital admissions. The mobility data is cellular data from a local cellular provider from before the pandemic (BAU) up to Lockdown Level 3. Figure 5 provides a connectivity visualisation of some of the cellular data available.

5. Results

The value of including a spatial component to an SEIR model is demonstrated in Figure 6, where the ward level spatial SEIR model is compared with an SEIR model run at a national level. Since both the national and spatial SEIR models were initialised with cases inflated by the undetected factor (see Section 4), for the purposes of the

---

[8] www.chpc.ac.za



(a) (b)

(c)

Figure 5: Connectivity strength relationship for movements between local municipalities, (a) business-as-usual (BAU) movement, (b) lockdown level 5, (c) lockdown level 4

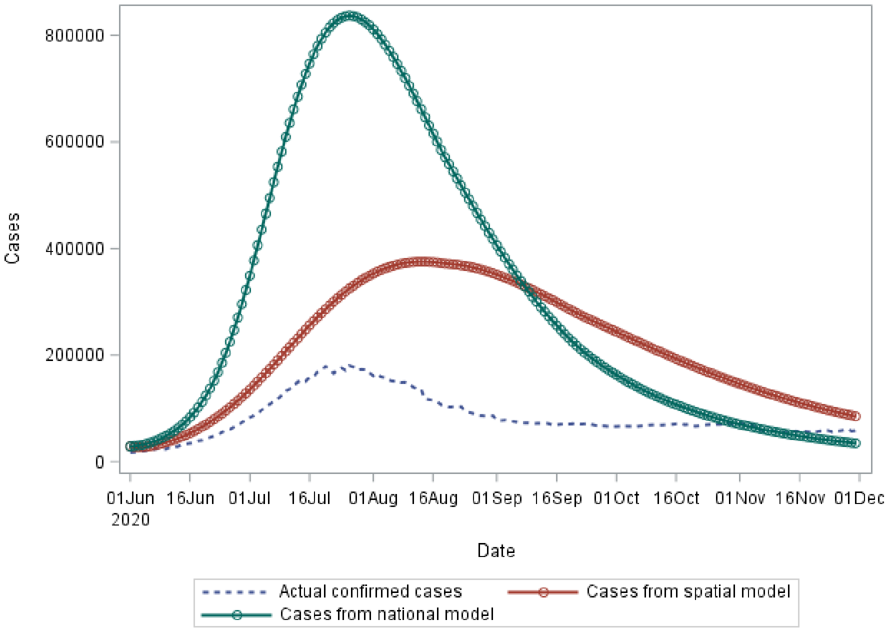


Figure 6: A national non-spatial model (green) vs. national spatial model (red). Actual active case data shown in dashed blue.

comparison to actual confirmed cases in Figure 5, the outputs from these models were divided by this same factor to show estimated recorded cases rather than estimated total cases.

Figure 6 shows the overestimation of the non-spatial approach. Our proposed spatial model produces more conservative estimates that appear to be a closer match to the reported case numbers. The time-course of the spatial model more accurately follows the observed case number, while the national model predicts far too rapid an increase in infections at the beginning, as it assumes a mixed population. The discrepancy with actual cases can be attributed to a number of complexities in the testing and reporting processes, as well as cases not picked up such as asymptomatic cases as well as symptomatic cases avoiding testing due to stigma, for example.

Although the NICD reports hospital admission cases for all provinces (NICD 2020b) these include only a small proportion of public hospitals and therefore no complete hospital data was available for the remaining seven provinces i.e., for all provinces excluding Gauteng and the Western Cape. Due to lack of other available data and assuming that the remaining seven provinces, being more rural in nature, would have a much lower hospitalisation proportion than Gauteng (given in Section 3, a value of 20% was taken as the proportion of severe cases that would actually report to hospital). This equates to 4.6% of the symptomatic population in these provinces. Based on the estimates for all nine provinces together with a starting date of 1st June 2020 and hospital-related parameters as given in Table 3, the projected hospitalisation cases i.e., hospital bed requirements, aggregated per province, are shown in Figure 7. This graph shows how the provinces peak at different times which provides an understanding of how the virus spread across South Africa.

Figure 8 provides a closer look at the provinces Gauteng and Western Cape for which

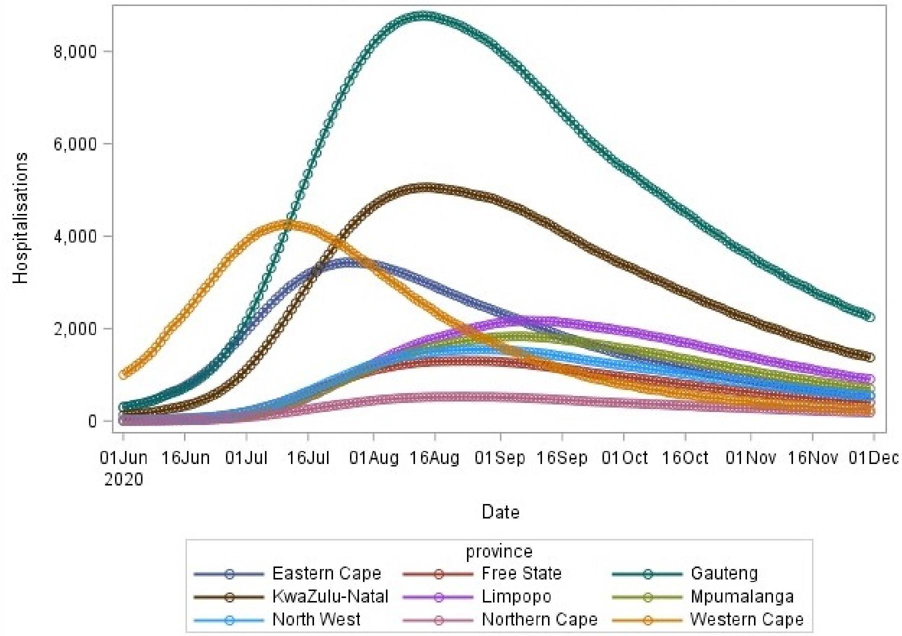
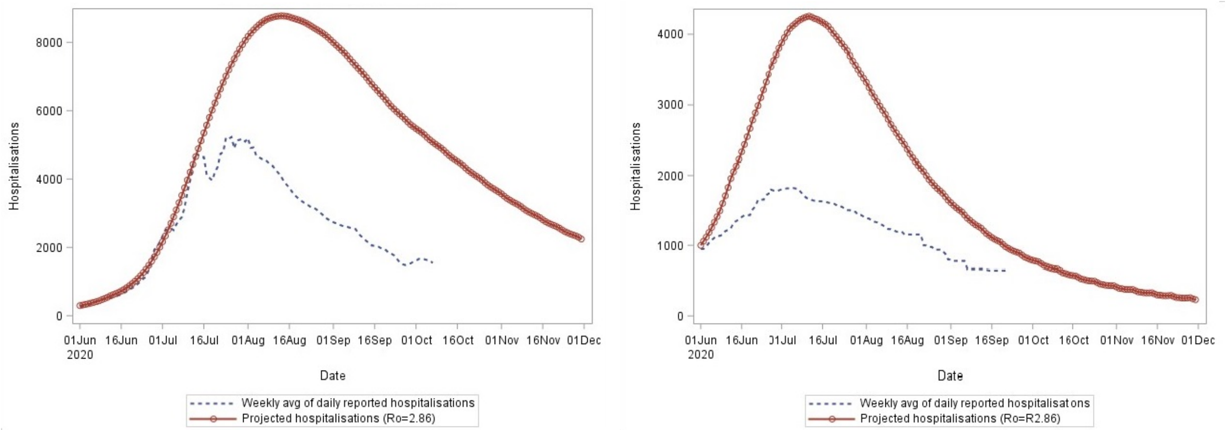


Figure 7: Projected hospitalisations per province in South Africa using $R_0 = 2.86$.



(a) Gauteng Province

(b) Western Cape Province

Figure 8: Actual hospitalisations (dashed blue) compared to predicted in two main provinces Gauteng and Western Cape (red).

hospitalisation data was available. The predictions also overestimate here, but capture the data decently well. Complexities of the hospitalisation data include COVID-19 deaths not captured as well as excess non-COVID-19 deaths. In addition, considerable stigma exists in the South African population regarding COVID-19 (and other diseases) and many severe cases do not end up in hospitals (Turner-Musa et al. 2020; Roelen et al. 2020).

Figure 9 provides a spatial representation of the predicted hospitalisation cases aggregated to Local Municipal (LM) level[9] on the day of the projected provincial peak. The maps clearly highlight the spatial variability of cases and the concentration within the Metropolitan Cities. Figure 9(a) shows a concentration within the City of Johannesburg, City of Tshwane and Ekhuruleni Metropolian Cities. Figure 9(b) shows the highest concentration within the City of Cape Town Metropolitan City followed by the Drakenstein and Breede Valley Local Municipalities.

Figure 10 indicates the model prediction simulated from the case numbers on 1 May 2020. The effect of R_0 is obvious and only the higher values of R_0 capture the timing of the peak well enough. The parameter R_0 is not the only parameter in our spatial SEIR model. Optimisation of the choice for it is therefore not enough to get a perfect model fit. The nuances of testing rates, accuracy of tests as well as stigma of getting tested, should be considered.

Further, in order to validate the model, we conducted a local sensitivity analysis using a spatial SEIR model we developed for the COVID-19 cases in South Africa. We consider the sensitivity of the proportion of individuals that had been infected by the end of 201 days to the spatial R_{0i}. The base R_0 was sampled from a $\mathrm{U}(1, 5)$ distribution over 960 simulations, and the correlation between the R_0 and proportion of infected individuals was calculated. Figure 11 illustrates the results of this spatial sensitivity analysis visualised as correlations.

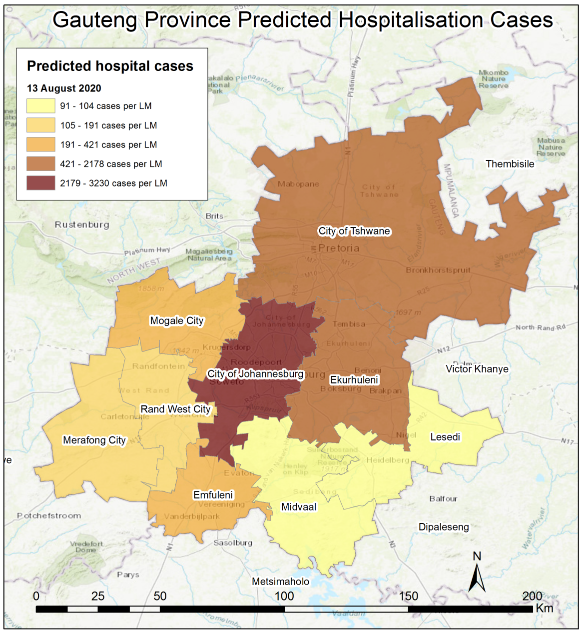
Additionally, we considered the sensitivity of the proportion of individuals that had been infected by the end of 201 days to the spatial R_0. Figure 11 shows that many correlations are between 0.8 and 0.9, indicating a strong relationship between the spatial R_0 and the proportion of infected individuals. It further demonstrates that the strength of this relationship varies across geographical space. This necessitates the use of a spatially varying R_0 to model COVID-19, as is proposed here.

6. Discussion

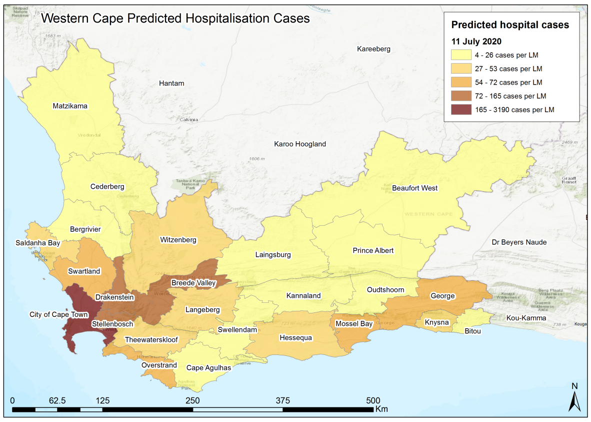
The spatial SEIR model herein can be used to capture the variation in peaks across the country and identify wards that could potentially become high risk by incorporating a vulnerability measure at ward level across the country. Age-structure per ward was also allowed for in the study on hospitalisations and identifies areas where hospitalisations are expected to be higher. This takes into account more vulnerable risk age groups in the hospitalisations. The model thereby accounts for the heterogeneous mixing occurring at a societal level.

---

[9] Local municipality level is defined in South Africa as an administrative area which are divided into wards `http://www.statssa.gov.za/census/census_2001/geo_metadata/geography_metadata.pdf`



(a) Gauteng Province predicted hospitalisation cases per local municipality. 13 August 2020



(b) Western Cape Province predicted hospitalisation cases per local municipality. 11 July 2020

Figure 9: Hospitalisation cases at local municipality level for the provinces Gauteng and the Western Cape.

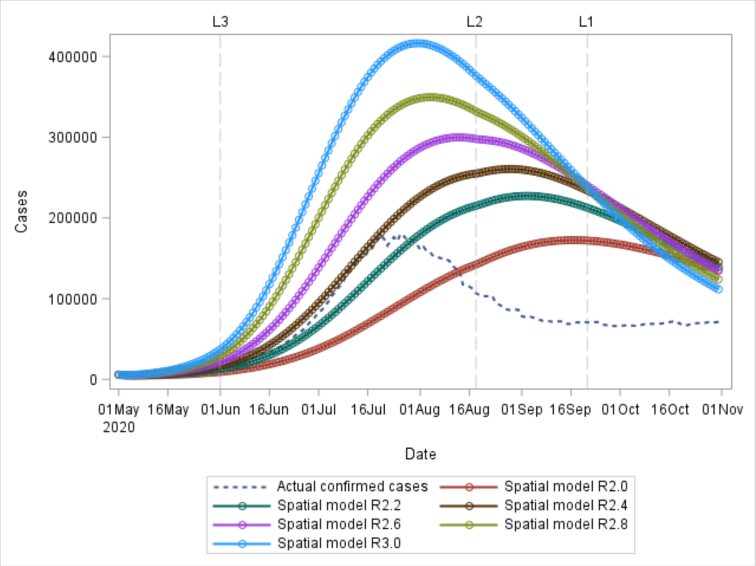


Figure 10: Illustration of the effect of varying R_0 on the spatial model run at a national level. The vertical lines indicate the lockdown dates as detailed in Table 2.

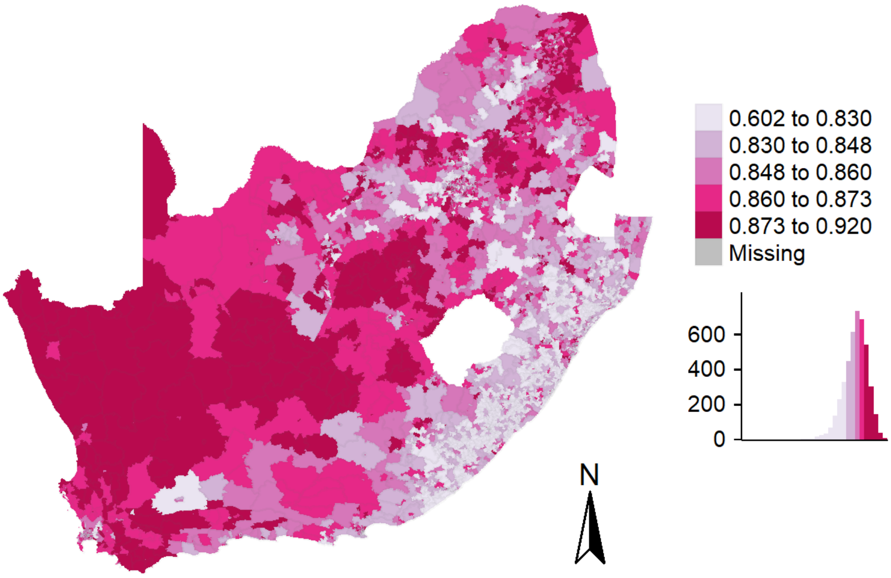


Figure 11: Graph of the correlations between R_0 and the proportion of infected individuals, at South African ward level as per the 2016 census.

The advantages of using mobility matrices at a fine local level is that mixing between wards is identified, and thus helping to reduce over-predictions relative to the national model. Using different movement matrices at the different lockdown interventions also allowed this mixing to vary at different stages of the disease spread. Alternative mobility data could also be considered, see for example, Potgieter et al. (2021).

The parameters in Table 3 were obtained using expert knowledge and literature. These parameters can easily be updated in the model and re-fitted. In addition, if suitable mobility data and a proxy for the vulnerability data are in place for another country, the model can also quite easily be extended to other spatial areas.

The model does not perfectly capture or predict COVID-19 due to the disease complexities still being researched as the pandemic continues. In addition, the data available have been collected centrally from a number of local sources, such as district clinics and testing centres, and may not be accurate in terms of location. It has become apparent that address data is not always the patient's address but perhaps that of the referring doctor. Due to stigma the patient's address may not be captured truthfully at collection; and in some cases an algorithm attempts to automate addresses during peaks. The mobility data is also difficult to prove as accurate as the service provider only covers 46% of the market in South Africa. It is likely that certain mobility is not represented with the mobility matrices (Tizzoni et al. 2014). The advantage, however, of still running the model at ward level is that it allows an aggregation upwards, thereby reducing the noise effect at fine levels, as shown in Figure 9.

Over-prediction in hospitalisations is similar to other projections in South Africa (SACMC 2020; NICD 2020) which also expected hospital capacity to be breached in provinces like the Western Cape and Gauteng and which also projected a later peak. Although spatial differences have been captured in the model in terms of vulnerability, age structure and mobility, there appear to be other factors affecting the nature of the disease spread in the different provinces since, in particular, the Western Cape was over-predicted to a greater degree than other provinces in the spatial model and in other projections seen (NICD 2020). It is unknown why the Western Cape was so much lower than expectations compared to other provinces but one possible reason could be behavioural differences in the population in terms of adhering to restrictions and quarantine rules. Potentially the scaling factors due to government interventions (indicated in Table 2) should differ between provinces due to behavioural differences even though restrictions are uniform across the country.

The prediction of peaks during the pandemic is important in order to understand the hospital capacity preparedness. With good focused healthcare a severe COVID-19 patient has a higher chance of surviving. Over-stretching the capacities of hospitals during this pandemic has been a major concern across the world, and many studies have been done to estimate the impact of COVID-19 on hospital beds, ventilator beds and ICU beds (CHSUFT 2020; Reno et al. 2020; Barrett et al. 2020).

We have briefly mentioned reasons for the overestimation still present in the proposed spatial model. We expand on that now. Firstly, there is a difference between modelling on a data set post-occurrence and with a disease that is mostly understood, compared to modelling in a still evolving pandemic with a disease that is only now slowly being understood. There is specifically difficulty in selecting parameters in a real-time environment when much is still unknown. Far more information is now known in the post-first wave environment internationally. Secondly, the issue of using confirmed cases to initialise the model may be temperamental since screening and testing vary due to factors such as the cost of tests, access to testing sites (closely correlated with poverty and lack of transport), varying local testing strategies, and stigma of being tested. Some studies[10] (Russell et al. 2020) use data on deaths as a more reliable indicator of the severity of a disease while others, (Woolf et al. 2020; Faust and Del Rio 2020; Weinberger et al. 2020), indicate COVID-19 related death reporting as unreliable. In South Africa, this also seems to be the case since the study by South African Medical Research Council (SAMRC) and University of Cape Town Centre for Actuarial Research (Moultrie et al. 2020; Bradshaw et al. 2020) reveal that between 6 May and 8 December 2020, excess deaths for persons aged one year and above were around 56 607, while the total COVID-19 deaths reported by the NICD in the same period was 22432. A September 2020 update of projected cases, NICD (2020), also estimated that about 80% of the excess deaths are a result of COVID-19.

Furthermore, not all cases that should be hospitalised were actually admitted, due to dying on arrival or prior to being admitted and tested, as seen from the excess deaths reported. The September 2020 report, NICD (2020), estimates that the probability of seeking hospital-level care for severely and critically ill cases ranges from 50% to 97%. Reasons for individuals in South Africa failing to report to hospital include lack of transport, no access to a medical facilities and also the stigma attached to having COVID-19 in certain communities, resulting in some individuals being too afraid to be tested or to go to hospital. The poor quality of care in public hospitals also serves as a deterrent to some individuals seeking care even when they need it. In addition, individuals may not have been admitted to medical facilities during the peak period due to capacities at certain facilities being overstretched.

It has more recently become known that some individuals carry immunity (Ni et al. 2020) and therefore the size of the susceptible population in each ward could be lower than what was initially assumed. Our assumption that all previously-infected individuals have permanent immunity also remains to be demonstrated but adds complexity to the model as the absence of immunity would then require a return from the R_i to S_i compartment.

The model naturally has limitations. These include the data accuracy discussed above, and furthermore the exposed case distribution at each time-step. Since the mobility data is not a personal trajectory, the location at the next time point of that individual is not certain. The model simply assumes an infected individual can be exposed but will return/remain in the same ward. Longer-term movement may still play a role. The model still assumes homogeneous mixing within a ward (but provides spatial heterogeneity nationally), which may not be accurate for a diverse country such as South Africa.

---

[10] Ritchie, H., Ortiz-Ospina, E., Beltekian, D., Mathieu, E., Hasell, J., Macdonald, B., Giattino, C. and Roser, M., 2020. Mortality Risk of COVID-19. [online] Our World in Data. Available at: `https://ourworldindata.org/mortality-risk-COVID` [Accessed 15 February 2021].

7. Conclusion

We have successfully demonstrated that a spatially explicit version of a classic SEIR model can effectively improve planning and preparation for COVID, and provide a better estimation of both the timing and the peak of the epidemic. This study has shown the benefit of accounting for the spatial dimension by considering local-level spatial units when using a SEIR-type model in modelling the spread of COVID-19. By adjusting the model for social vulnerability and distributing cases according to mobility data at ward level we allowed for important spatial influences in predicting the spread of the disease. Setting up a spatial compartmental model and appropriately calibrating it at a low aggregation of spatial units is useful to improve decision making once disease characteristics are understood better after the initial outbreak, especially in contexts where social factors are strongly at play.

Future work includes improving the accuracy of the mobility data through triangulation of multiple data sources, stratifying the model to account for co-morbidity in sub-populations as well as modelling multiple waves of infections and/or strains. The model can also be expanded to include a death compartment, as well as allow for levels of population immunity as data becomes available and as vaccines are rolled out.

Computational Details

The data used in this study is not directly available without approval, so cannot be shared directly with the paper. Thank you to the NICD, South Africa for providing access to the ward level infection data and additional accompanying data. In addition, thanks to GeoTerraImage (Pty) Ltd for the age data provided to the CSIR for the COVID-19 response. The code is available at `https://github.com/TheCoolRob/sp-covid19` however requires the use of a high performance computer to be implemented. The user may access the code at the Github link.

The results in this paper were obtained using R 3.6.0, C++, SAS 9 and ArcGIS 10.8.1.

Acknowledgments

We acknowledge the contributions of Rob Bennetto and Ronald Richman, who assisted with the coding and implementation of the model on the CHPC. We also thank Gerbrand Mans from the CSIR for data aggregation assistance.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

This research is also funded by Canada's International Development Research Centre (IDRC) (Grant No. 109559-001).

The use of the Centre for High Performance Computing (www.chpc.ac.za) also made this work possible.

References

Aldila, D., Khoshnaw, S., Safitri, E., Anwar, Y., Bakry, A., Samiadji, B., Anugerah, D., Gh, M., Ayulani, I., and Salim, S. (2020). A mathematical study on the spread of COVID-19 considering social distancing and rapid assessment: The case of Jakarta, Indonesia. *Chaos, Solitons and Fractals*, 139(110042).

Anguelov, R., Banasiak, J., Bright, C., Lubuma, J., and Ouifki, R. (2020). The big unknown: The asymptomatic spread of COVID-19. *Biomath*, 9(1):1–11.

Arandiga, F., Baeza, A., Cordero-Carrion, I., Donat, R., Marti, M., Mulet, P., and Yanez, D. (2020). A spatial-temporal model for the evolution of the COVID-19 pandemic in Spain including mobility. *Mathematics*, 8(1677).

Barrett, K., Khan, Y. A., Mac, S., Ximenes, R., Naimark, D. M., and Sander, B. (2020). Estimation of COVID-19–induced depletion of hospital resources in Ontario, Canada. *Canadian Medical Association Journal*, 192(24).

Bassett, I. V., Coleman, S. M., Giddy, J., Bogart, L. M., Chaisson, C. E., Ross, D., Flash, M. J., Govender, T., Walensky, R. P., Freedberg, K. A., et al. (2017). Barriers to care and 1-year mortality among newly diagnosed HIV-infected people in Durban, South Africa. *Journal of Acquired Immune Deficiency Syndromes*, 74(4):432.

Bloor, K., Hendry, V., and Maynard, A. (2006). Do we need more doctors? *Journal of the Royal Society of Medicine*, 99(6):281–287.

Böhning, D., Rocchetti, I., Maruotti, A., and Holling, H. (2020). Estimating the undetected infections in the COVID-19 outbreak by harnessing capture–recapture methods. *International Journal of Infectious Diseases*, 97:197–201.

Bradshaw, D., Laubscher, R., Dorrington, R., Groenewald, P., and Moultrie, T. (2020). Report on weekly deaths in South Africa: 1 January-8 December 2020 (week 49). Technical report, Burden of Disease Research Unit, South African Medical Research Council.

Bruns, D. P., Kraguljac, N. V., and Bruns, T. R. (2020). COVID-19: Facts, cultural considerations, and risk of stigmatization. *Journal of Transcultural Nursing*, 31(4):326–332.

Chen, D. (2014). *Analyzing and modeling spatial and temporal dynamics of infectious diseases*, chapter Modeling the spread of infectious diseases: A review, pages 19–42.

Choonara, S. and Eyles, J. (2016). Out of control: profit-seeking behaviour, unnecessary medical procedures and rising costs of private medical care in South Africa. *BMJ Global Health*, 1(1).

CHSUFT (2020). Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator days and deaths by US state in the next 4 months. Technical report, COVID-19 health service utilization forecasting team.

Competition Commission (2018). Health market inquiry: Final findings and recommendations report. Technical report, Pretoria: Competition Commission.

Czeglédy, A. P. (2004). Getting around town: transportation and the built environment in post-apartheid South Africa. *City & Society*, 16(2):63–92.

David, A., Guilbert, N., Hamaguchi, N., Higashi, Y., Hino, H., Leibbrandt, M., and Shifa, M. (2018). Spatial poverty and inequality in South Africa: A municipality level analysis. Technical report, University of Cape Town, South Africa.

de Beer, L. (2019). Strategic assessment of the impact of the priority infrastructure on the performance of minibus-taxi services in the context of South Africa. In *Proceedings of the Southern African Transport Conference*.

Eker, S. (2020). Validity and usefulness of COVID-19 models. *Humanities and Social Sciences Communications*, (1):1–5.

Ezeh, A., Oyebode, O., Satterthwaite, D., Chen, Y.-F., Ndugwa, R., Sartori, J., Mberu, B., Melendez-Torres, G. J., Haregu, T., Watson, S. I., et al. (2017). The history, geography, and sociology of slums and the health problems of people who live in slums. *The Lancet*, 389(10068):547–558.

Faust, J. S. and Del Rio, C. (2020). Assessment of deaths from COVID-19 and from seasonal influenza. *Journal of the American Medical Association Internal Medicine*, 180(8):1045–1046.

Ferguson, N., Laydon, D., Nedjati Gilani, G., Imai, N., Ainslie, K., and Baguelin, M. (2020). Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Technical report, Imperial College COVID-19 Response Team.

Finn, A., Leibbrandt, M., and Oosthuizen, M. (2014). Poverty, inequality, and prices in post-apartheid South Africa. Technical report, WIDER Working Paper.

Fransman, T. and Yu, D. (2019). Multidimensional poverty in South Africa in 2001–16. *Development Southern Africa*, 36(1):50–79.

Fronteira, I., Sidat, M., Magalhães, J. P., de Barros, F. P. C., Delgado, A. P., Correia, T., Daniel-Ribeiro, C. T., and Ferrinho, P. (2021). The SARS-CoV-2 pandemic: A syndemic perspective. *One Health*, page 100228.

Giddy, J. K. (2019). The influence of e-hailing apps on urban mobilities in south africa. *African Geographical Review*, 38(3):227–239.

Gray, A. and Vawda, Y. (2019). Health legislation and policy. *South African Health Review*, 1:3–16.

Iannelli, M. and Pugliese, A. (2014). *An Introduction to Mathematical Population Dynamics*, chapter Mathematical Modeling of Epidemics, pages 209–264. Springer.

Islam, M., Mamun, K., Kim, B., and Kim, S. (2020). COVID-19 transmission: Bangladesh perspective. *Mathematics*, 8(1793).

Jehi, L., Ji, X., Milinovich, A., Erzurum, S., Merlino, A., Gordon, S., Young, J. B., and Kattan, M. W. (2020). Development and validation of a model for individualized prediction of hospitalization risk in 4,536 patients with COVID-19. *PloS One*, 15(8):e0237419.

Jennings, G. and Behrens, R. (2017). The case for investing in paratransit: strategies for regulation and reform. Technical Report 01720253, Swedish National Road and Transport Research Institute (VTI).

Kermack, W. and McKendrick, A. (1991). Contributions to the mathematical theory of epidemics-i. *Bulletin of Mathematical Biology*, 53(1-2):33–35.

Lai, S., Ruktanonchai, N., Zhou, L., Prosper, O., Luo, W., Floyd, J., Wesolowski, A., Santillana, M., Zhang, C., Du, X., and Yu, H. (2020). Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature*, 585(7825):410–413.

le Roux, A., Cooper, A., Ludick, C., Arnold, K., and Mans, G. (2020). *COVID-19: Geospatial Information and Community Resilience*, chapter Creating a set of high-resolution vulnerability indicators to support the disaster management response to the COVID-19 pandemic in South Africa.

Li, M., Graef, J., Wang, L., and Karsai, K. (1999). Global dynamics of a SEIR model with varying total population size. *Mathematical Biosciences*, 160(2):191–213.

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., and W Yang, J. S. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490):489–493.

Lynch, E. A., Cadilhac, D. A., Luker, J. A., and Hillier, S. L. (2017). Inequities in access to inpatient rehabilitation after stroke: an international scoping review. *Topics in Stroke Rehabilitation*, 24(8):619–626.

Maphumulo, W. T. and Bhengu, B. R. (2019). Challenges of quality improvement in the healthcare of South Africa post-apartheid: A critical review. *Curationis*, 42(1):1–9.

Matlala, M., Bannister, M., and Mochotlhi, D. (2017). Sustainable development goal 6 –water and sanitation integrated monitoring. Technical report, Department of Water and Sanitation, South Africa, Available from www.sdg6monitoring.org/country-experience/south-africa/. [Accessed: 30 March 2021].

Moultrie, T. et al. (2020). Excess deaths: additional measures and approaches to understanding COVID-19 related mortality in South Africa. Technical report, NICD.

Ngonghala, C., Iboi, E., Eikenberry, S., Scotch, M., MacIntyre, C., Bonds, M., and Gumel, A. (2020). Mathematical assessment of the impact of non-pharmaceutical interventions on curtailing the 2019 novel Coronavirus. *Mathematical Biosciences*, 325.

Ni, L., Ye, F., Cheng, M.-L., Feng, Y., Deng, Y.-Q., Zhao, H., Wei, P., Ge, J., Gou, M., Li, X., et al. (2020). Detection of SARS-CoV-2-specific humoral and cellular immunity in COVID-19 convalescent individuals. *Immunity*, 52(6):971–977.

NICD (2020b). COVID-19 daily sentinel hospital surveillance report. Technical report, National Institute for Communicable Diseases.

NICD (2020c). COVID-19 testing summary for week 45. Technical report, National Institute for Communicable Diseases.

NICD (2020). Estimating cases for COVID-19 in South Africa. assessment of alternative scenarios. Technical report, National Institute for Communicable Diseases (NICD).

NICD (May 2020a). Clinical management of suspected or confirmed COVID-19 disease version 4. Technical report, National Institute of Communicable Diseases.

Palmieri, L., Vanacore, N., Donfrancesco, C., Lo Noce, C., Canevelli, M., Punzo, O., Raparelli, V., Pezzotti, P., Riccardo, F., Bella, A., et al. (2020). Clinical characteristics of hospitalized individuals dying with COVID-19 by age group in Italy. *The Journals of Gerontology: Series A*, 75(9):1796–1800.

Peixoto, P., Marcondes, D., Peixoto, C., and Oliva, S. (2020). Modeling future spread of infections via mobile geolocation data and population dynamics. an application to COVID-19 in Brazil. *PLOS ONE*, 15(7: e0235732).

Pillay, R. (2009). Work satisfaction of professional nurses in South Africa: a comparative analysis of the public and private sectors. *Human resources for Health*, 7(1):1–10.

Potgieter, A., Fabris-Rotelli, I. N., Kimmie, Z., Dudeni-Tlhone, N., Holloway, J. P., Janse van Rensburg, C., Thiede, R. N., Debba, P., Manjoo-Docrat, R., Abdelatif, N., and Khuluse-Makhanya, S. (2021). Modelling representative population mobility for COVID-19 spatial transmission in South Africa. *Frontiers in Big Data*, 4:82.

Reno, C., Lenzi, J., Navarra, A., Barelli, E., Gori, D., Lanza, A., Valentini, R., Tang, B., and Fantini, M. P. (2020). Forecasting COVID-19-associated hospitalizations under different levels of social distancing in Lombardy and Emilia-Romagna, Northern Italy: results from an extended SEIR compartmental model. *Journal of Clinical Medicine*, 9(5):1492.

Roelen, K., Ackley, C., Boyce, P., Farina, N., and Ripoll, S. (2020). COVID-19 in LMICS: the need to place stigma front and centre to its response. *The European Journal of Development Research*, 32(5):1592–1612.

Rost, G., Bartha, F., Bogya, N., Boldog, P., Dénes, A., Ferenci, T., Horváth, K., Juhasz, A., Nagy, C., Tekeli, T., Vizi, Z., and Oroszi, B. (2020). Early phase of the COVID-19 outbreak in Hungary and post-lockdown scenarios. *Viruses*, 12(708).

Russell, T. W., Hellewell, J., Jarvis, C. I., Van Zandvoort, K., Abbott, S., Ratnayake, R., Flasche, S., Eggo, R. M., Edmunds, W. J., Kucharski, A. J., et al. (2020). Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. *Eurosurveillance*, 25(12):2000256.

SACMC (2020). Estimating cases for COVID-19 in South Africa update: 19 may 2020. Technical report, South African Covid-19 Modelling Consortium.

Sattenspiel, L. and Dietz, K. (1995). A structured epidemic model incorporating geographic mobility among regions. *Mathematical Biosciences*, 128(1):71–91, ISSN: `0025-5564`.

StatsSA (2011). Use of health facilities and levels of selected health conditions in South Africa: Findings from the general household survey. Technical report, Statistics South Africa.

StatsSA (2016). Water and sanitation: In-depth analysis of the general household survey 2002–2015 and community survey 2016 data. Technical report, Statistics South Africa.

StatsSA (2017). Inequality trends in South Africa: A multidimensional diagnostic of inequality. Technical report, Statistics South Africa.

Tang, B., Bragazzi, N. L., Li, Q., Tang, S., Xiao, Y., and Wu, J. (2020). An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). *Infectious Disease Modelling*, 5:248–255.

Thiede, R., Abdelatif, N., Fabris-Rotelli, I., Manjoo-Docrat, R., Holloway, J., van Rensburg, C. J., Debba, P., Dudeni-Tlhone, N., Kimmie, Z., and Roux, A. l. (2020). Spatial variation in the basic reproduction number of COVID-19: A systematic review. *arXiv preprint arXiv:2012.06301*.

Tizzoni, M., Bajardi, P., Decuyper, A., King, G. K. K., Schneider, C. M., Blondel, V., Smoreda, Z., González, M. C., and Colizza, V. (2014). On the use of human mobility proxies for modeling epidemics. *PLoS Computational Biology*, 10(7):e1003716.

Turner-Musa, J., Ajayi, O., and Kemp, L. (2020). Examining social determinants of health, stigma, and COVID-19 disparities. In *Healthcare*, volume 8, page 168. Multidisciplinary Digital Publishing Institute.

Weinberger, D. M., Chen, J., Cohen, T., Crawford, F. W., Mostashari, F., Olson, D., Pitzer, V. E., Reich, N. G., Russi, M., Simonsen, L., et al. (2020). Estimation of excess deaths associated with the COVID-19 pandemic in the United States, March to May 2020. *Journal of the American Medical Association Internal Medicine*, 180(10):1336–1344.

Woolf, S. and Joubert, J. (2014). A look at paratransit in South Africa. Technical report, University of Pretoria, South Africa.

Woolf, S. and Joubert, J. W. (2013). A people-centred view on paratransit in South Africa. *Cities*, 35:284–293.

Woolf, S. H., Chapman, D. A., Sabo, R. T., Weinberger, D. M., and Hill, L. (2020). Excess deaths from COVID-19 and other causes, March-April 2020. *Journal of the American Medical Association*, 324(5):510–513.

Wu, S., Mertens, A., Crider, Y., Nguyen, A., Pokpongkiat, N., Djajadi, S., Seth, A., Hsiang, M., Colford, J., and Reingold, A. e. a. (2020). Substantial underestimation of SAR-CoV-2 infection in the United States. *Nature Communications*, 11:1–10.

Zhang, J., Dong, L., Zhang, Y., Chen, X., Yao, G., and Han, Z. (2020). Investigating time, strength, and duration of measures in controlling the spread of COVID-19 using a networked meta-population model. *Nonlinear Dynamics*, 101(3):1789–1800.

Affiliation:

Inger Fabris-Rotelli, Renate Thiede, Arminn Potgieter
Department of Statistics
University of Pretoria
E-mail: inger.fabris-rotelli@up.ac.za, renate.thiede@gmail.com, arminnpotgieter@gmail.com

Jenny Holloway, Pravesh Debba, Alize le Roux, Nontembeko Dudeni-Tlhone
Council for Scientific and Industrial Research
South Africa
E-mail: jhollowa@csir.co.za, pdebba@csir.co.za, aleroux1@csir.co.za, ndudenitlhone@csir.co.za

Zaid Kimmie
Foundation for Human Rights
South Africa
E-mail: zaid@kimmie.co.za

Sally Archibald
School of Animal, Plant and Environmental Sciences
University of Witwatersrand
E-mail: sally.archibald@wits.ac.za

Raeesa Manjoo-Docrat
Department of Statistics and Actuarial Science
University of Witwatersrand
E-mail: raeesa.docrat@wits.ac.za

Charl Janse van Rensburg, Nada Abdelatif
Biostatistics Research Unit
South African Medical Research Council
E-mail: charl.jansevanrensburg@mrc.ac.za, nada.abdelatif@mrc.ac.za

Sibusisiwe Makhanya
IBM, South Africa
E-mail: sibusisiwe.makhanya@ibm.com

Journal of Data Science, Statistics, and Visualisation https://jdssv.org/
published by the International Association for Statistical Computing
http://iasc-isi.org/

ISSN 2773-0689
November 2022, Volume II, Issue VII *Submitted:* 2021-05-17
doi:10.52933/jdssv.v2i7.46 *Accepted:* 2022-01-10