

Cities, crowding, and the coronavirus: Predicting contagion risk hotspots¹

Gaurav Bhardwaj, World Bank

Thomas Esch, German Aerospace Center

Somik V. Lall, World Bank

Mattia Marconcini, German Aerospace Center

Maria Edisa Soppelsa, World Bank

Sameh Wahba, World Bank

Working paper: This version April 21, 2020

1. Introduction

Today, over 4 billion people around the world—more than half the global population—live in cities. By 2050, with the urban population more than doubling its current size, nearly 7 of 10 people in the world will live in cities. Evidence from today's developed countries and rapidly emerging economies shows that urbanization and the development of cities is a source of dynamism that can lead to enhanced productivity. In fact, no country in the industrial age has ever achieved significant economic growth without urbanization.

The underlying driver of this dynamism is the ability of cities to *bring people together*. Social and economic interactions are the hallmark of city life, making people more productive and often creating a vibrant market for innovations by entrepreneurs and investors. International evidence suggests that the elasticity of income per capita with respect to city population is between 3% and 8% (Rosenthal & Strange 2003).² Each doubling of city size raises its productivity by 5%.

¹ Authors' names in alphabetical order. Correspondence to Somik Lall (slall1@worldbank.org). The authors thank (a) the German Aerospace Center (DLR) for sharing data on building heights that are critical for this approach; (b) the UK DFID for generous financial assistance in deploying this approach and (c) the European Space Agency (ESA) for funding the related research in the project "Artificial Intelligence for Smart Cities" [grant number 4000126100/19/I-EF]. Hogeun Park, Olivia D'Aoust, Swati Sachdeva provided important technical contributions. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

² After controlling for skill level of the labor force

But the coronavirus pandemic is now seriously limiting social interactions. With no vaccine available, prevention through containment and social distancing, along with frequent handwashing, appear to be, for now, the only viable strategies against the virus. The goal is to slow transmission and avoid overwhelming health systems that have finite resources. Hence non-essential businesses have been closed and social distancing measures, including lockdowns, are being applied in many countries.

Will such measures defeat the virus in dense urban areas? In principle, yes. Wealthier people in dense neighborhoods can isolate themselves while having amenities and groceries delivered to them. Many can connect remotely to work, and some can even afford to live without working for a time. But poorer residents of **crowded neighborhoods** cannot afford such luxuries. They are forced to leave their home every day to go to work, buy groceries, and do laundry. This is especially true in low-income neighborhoods of developing countries – many of which are slums and informal settlements. In fact, 60 percent of Africa’s urban population is packed into slums — a far larger share than the average 34 percent seen in other developing countries (United Nations 2015). With people tightly packed together, the resulting crowding increases contagion risk from the coronavirus.

Consider Dar es Salaam where 28 percent of residents are living at least three to a room; in Abidjan, the figure is 50 percent. Residents in these neighborhoods live in substandard housing and overcrowded conditions. They lack open space and suffer from inadequate infrastructure meaning that they share taps and latrines, with as many as 200 people per communal facility. In South Africa, for example, only 44.4% of people have access to water inside their house and only 60.6% have access to a flush toilet.³ Most people living in these communities are daily wage earners in the informal sector, with irregular income and insecure jobs. They cannot realistically practice social distancing, nor can they afford it.

Figure 1: Feasibility of Social Distancing: Density is a symptom; underlying social disparities are the drivers

³ New York Times, “How can you social distance when you share a toilet with your neighbor?”, April 3, 2020.



Sketch by Somya Bajaj

In regions such as Africa, South Asia, and Central America, the pandemic has yet to peak. Their cities, especially the densest ones, will face a great challenge, given their weak infrastructure and limited medical and financial resources. To stave off this crisis, *emerging hotspots* must be anticipated so that medical and civil resources can be targeted to limit diffusion into surrounding areas. *Vulnerable groups* need to be identified in advance, so that they can be supported to weather the storm.

To help city leaders prioritize resources towards places with the highest exposure and contagion risk, we have developed a simple methodology that can be rapidly deployed. This methodology identifies hotspots for contagion and vulnerability, based on:

- The practical inability for keeping people apart, based on a combination of population density and livable floor space that does not allow for 2 meters of physical distancing.
- Conditions where, even under lockdown, people might have little option but to cluster (e.g., to access public toilets and water pumps).

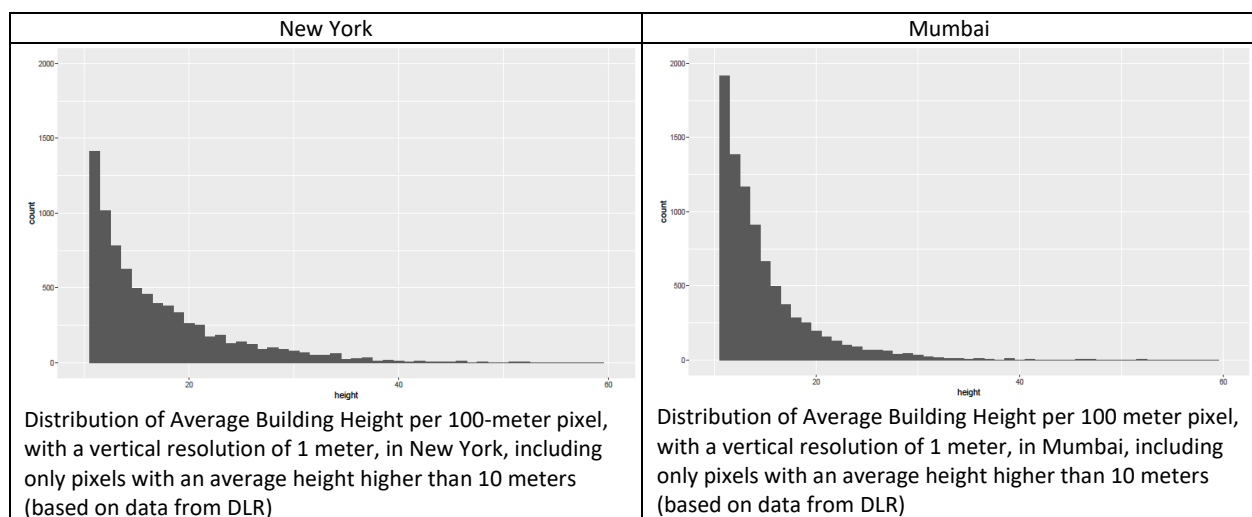
We outline the methodology in section 2. Section 3 describes the data. We present 3 pilot applications in Section 4. Section 5 concludes.

2. Methodology

Contagion will increase with social interaction. A person in a crowded slum is likely to have close contact with more people than someone living in a suburban community. Moreover, people in **small homes or shacks** are usually forced to move out more frequently, often for accessing basic services such as water taps and toilets.

People living in such crowded communities are more vulnerable. To get reasonable estimates of crowding, it is important to get information on *livable floor space* available to individuals and families. Raw measures of density (people per unit of land) need to be adjusted to measure livability (people per unit of floor space). Consider two cities: New York City (Manhattan) and Bombay (Mumbai). They both have similar population densities at around 25,000 people per square km. However, the way these densities translate into living space is completely different. Having 25,000 persons on a square km living in 10-story apartment buildings will be very different compared to the same number of people living in single story shacks. Figure 2 shows that distribution of building heights for New York and Mumbai. When we consider building heights, New York's total floor area is almost 4 times that of Mumbai. The additional living space (coupled with amenities) in New York makes it easier for people to maintain the required social distance and to self-isolate. While people in Mumbai will struggle to even maintain social distance at home due to the lack of sufficient floor space, many in slums will be forced to interact outside their homes, for instance, to access communal water and sanitation facilities.

Figure 2: Distribution of Building Heights - New York and Mumbai



We use adjusted densities based on floor areas as the first step in developing our measure of coronavirus contagion. To calculate floor area, we first construct an adjusted measure of population density that incorporates the building structure. We define adjusted population density at the pixel level as follows:

$$D_i = \frac{P_i}{TFA_i}$$

where i identifies a pixel, P_i is the population of pixel i , and TFA_i is the total floor area of pixel i . The total floor area is calculated as:

$$TFA_i = \frac{H_i * L_i}{3}$$

where H_i denotes the average height of pixel i , L_i is the built surface area of pixel i , and 3 is the average floor height (assuming an average floor height of 3 meters).

The pixel level density gives us a first notion of where the most crowded areas are located in a city. Social distancing norms require that any person needs to maintain a distance of at least 2 meters to the nearest person; a person needs a minimum of 3.464 square meters of space around them.⁴ This translates to a density of approximately 0.29 persons per square meter of floor space. Places where density surpasses this threshold are those where people cannot comply with the minimum social distance requirements due to lack of available living space.

However, a static measure of density is not enough. A person will move not only around her “not so big” house but may also need to travel a “relatively small” distance to access certain amenities or shop for basic products (e.g. groceries). She will necessarily interact with the people in her neighborhood, making crowded neighborhoods even more prone to a viral spread. The concentration of people in the neighborhood is then also important to understand how much exposure a person may face.

To calculate the potential exposure of a person, we need to account for the density of her neighborhood. Assuming people will travel short distances during lockdown, we considered the eight neighboring pixels as the limit of mobility (queen contiguity). When using a grid with 100x100 meter resolution, this represents a neighborhood with a radius of approximately 100-150 meters. We take 9 pixels at a time and

⁴ The preferred lattice to accommodate people more efficiently is the hexagon. We draw a hexagon grid and locate people in the center of each hexagon. In order to have people respect the required social distance (2 meters), each hexagon needs to have a side of 1.15 meters, with a surface of 3.464 square meters. Then, one person will need a space of 3.464 square meters. This translates in a threshold of 0.289 (Pop/TFA>0.289).

assign the sum of population per unit of TFA for the nine pixels to the reference (central) pixel. This gives us a better approximation of the exposure that a person living in the central pixel may face.

This measure of density and minimal mobility allows us to identify the first set of potential hotspots. However, people living in informal settlements face an extra challenge: they need to move to access water, toilet facilities, and other basic services, even under a lockdown scenario. We account for mobility and concentration in these service spots and find a set of additional hotspots.

To generate these additional hotspots, we first identify two types of pixels: **service pixels**, where a specific service is located, and **transit** pixels. Before looking into pixel density, we need to consider two facts. First, more people will concentrate on service pixels. It is expected that these pixels will be potential hotspots, since people from the neighborhood will concentrate around the service often (water kiosk, toilet). Second, more people will pass through other pixels on their way to the service pixel. Even though pixels that are not service pixels will not concentrate people per se, they will still experience more traffic of people. The volume of traffic in these transit pixels will depend on how close they are to the service pixel.

We define $d_{i,s}$ as the distance from the centroid of pixel i to the nearest service pixel s . For any given pixel (service or transit), the number of people they receive (concentration of people or through traffic) depends on the density of the neighboring pixels, $\frac{Pop_j}{TFA_j}$, where j denotes all neighboring pixels that will go to the service pixel or pass through a transit pixel. The density of each pixel i is weighted by the distance from pixel i to the neighboring pixel j , denoted as d_j . The assumption is that people in the contiguous pixel will probably pass through pixel i on their way to the service pixel, while people on a pixel that is 500 meters away may take a different route, so density of pixels further away have a smaller impact on pixel i .⁵ Then, the density of pixel i is a function of population, TFA, and two distances: distance from the neighboring pixels j to pixel i and distance to the service pixel, i. e., $D_i = f\left(\frac{Pop}{TFA}, d_j, d_s\right)$.

The new pixel density can be defined as follows:

$$D_i = \frac{P_i}{TFA_i} + \sum_{j=1}^8 \frac{P_j}{TFA_j} + S_i \left\{ \sum_{j=9}^N \frac{P_j}{TFA_j} d_j^{-\frac{1}{2}} \right\} + (1 - S_i) \left\{ \frac{1}{8} \sum_{j=1}^8 \frac{P_j}{TFA_j} d_j^{-\frac{1}{2}} \right\} d_{i,s}^{-\frac{1}{2}}$$

⁵ This assumption is motivated by the market access literature that uses distance to discount for the lower effect that some variables (such as demand, wages) have, the further away they are (see Harris, 1954, Davis and Weinstein, 2003; Head and Mayer, 2004)

where S_i is a dummy variable equals to 1 if the pixel is a service pixel, and 0 otherwise. Introducing this dummy variable allows us to separate the fact that service pixels are different from transit pixels. On the one hand, more people will agglomerate in a service pixel. When $S_i = 1$ we add to our original density, the people from the defined neighborhood that will come to the service pixel. We use N to define the number of pixels in the neighborhood. For example, for a neighborhood that includes pixels in a 200-250 meter radius, $N = 25$. On the other hand, people will pass through transit pixels. When $S_i = 0$ we add just a proportion of people that will go through transit pixels on their way to the service pixel.

This equation accounts for both, density considering where people live (first two terms), and density considering concentration at a service facility or some mobility towards it (third and last term, respectively). The last term of the equation is multiplied by a constant, in this case, $\frac{1}{8}$, to consider the fact that not all the people in the defined neighborhood will cross through the pixel to access the service pixel. Since each pixel is surrounded by 8 neighboring pixels, one can think of 8 possible directions from where people can cross. Using $\frac{1}{8}$ simplifies the calculation and considers only one direction. Note here that we are currently using straight line distance (as the crow flies) to measure interactions between where people live and where service points are located. In refining the methodology, we will apply network distances to refine these estimates.

3. Data

The methodology is applied to data using globally sourced datasets on population, average building heights, and location of key services. Because of the importance of deploying this analysis rapidly and since local data coverage may be scarce and difficult to access during the pandemic, this methodology primarily relies on global datasets. Since we are trying to roll out this approach to support decision making over the next 4-6 weeks, we have had to make a trade-off between speed of delivery and granularity of the data. The team is taking a 'good enough' approach for the analysis where the core data and methodology is analytically robust but not 'perfect'. While the team is making all possible attempts to refine the hotspots to take local information on amenities, transport networks, and housing quality into account, relying on global data as the base template makes it possible to immediately support cities during the pandemic.

We use three basic datasets for the analysis.

- First, we use WorldPop population 2019 raster⁶ or Facebook 2019 population raster⁷, depending on the city. These rasters contain the estimated number of people that lives in each pixel, with a resolution of 100x100 meters or 30x30 meters, respectively. While we used global population datasets, the analysis can be easily adjusted if local sources are available. Population data from local sources may be more accurate than a global population raster. Some cities have developed spatial datasets at considerable high levels of disaggregation, although generally not at the pixel level. The challenge relies on creating a population grid based on these datasets, which usually implies modelling on assigning population to pixels, based on land use and built-up areas.⁸
- Second, we use DLR's World Settlement Footprint 3D product (WSF-3D) that is derived from ALOS World 3D (AW3D30) digital terrain model and the WSF 2015 settlement layer. Starting from satellite images at a finer resolution, and using digital terrain models, DLR extracts building heights and creates a 100 x 100 meter raster with the average building height (with a vertical resolution set to 1m).⁹
- The third dataset is a layer with the location of key services, such as water kiosks, public toilets, obtained from Open Street Maps Platform (OSM).¹⁰

Here is an example showing the base data for Mumbai. Figure 3 shows the distribution of population, where denser areas are highlighted in dark red, and the location of public toilets and water points are marked as dots (dots with an H and solid dots, respectively). Figure 4 shows the average heights across for the city, in meters.

⁶ See <https://www.worldpop.org/> for details on WorldPop data and methodology.

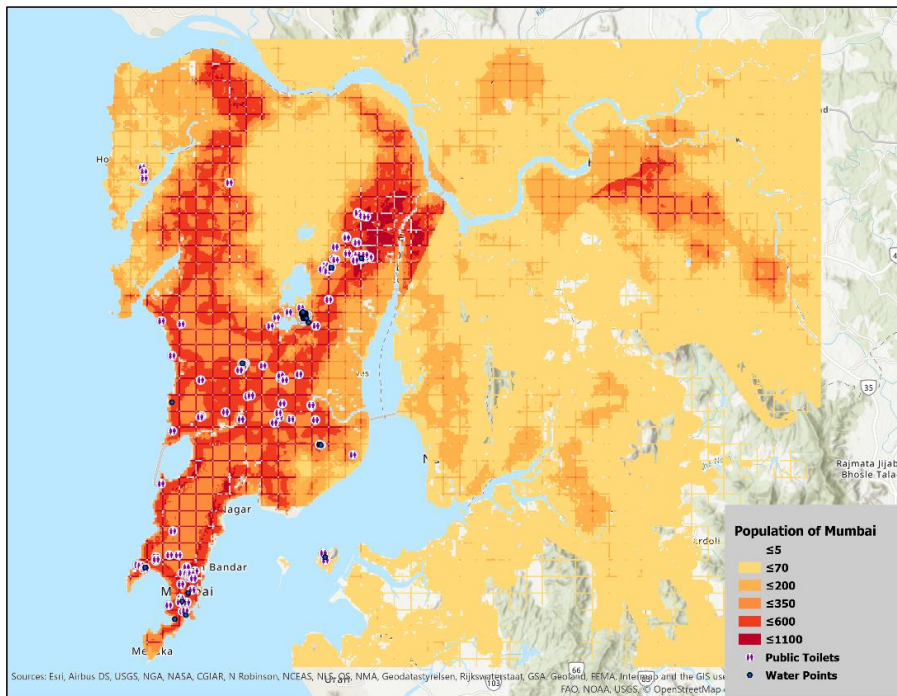
⁷ See <https://dataforgood.fb.com/> for details on Facebook data and <https://dataforgood.fb.com/tools/population-density-maps/> for more details on Facebook population maps.

⁸ The German Aerospace Center (DLR) has developed a model to generate population grid datasets based on spatially disaggregated population data. The team is working with them to generate better population grids when local data is available.

⁹ Although this data has been collected between 2006 and 2011, it corresponds to one point in time.

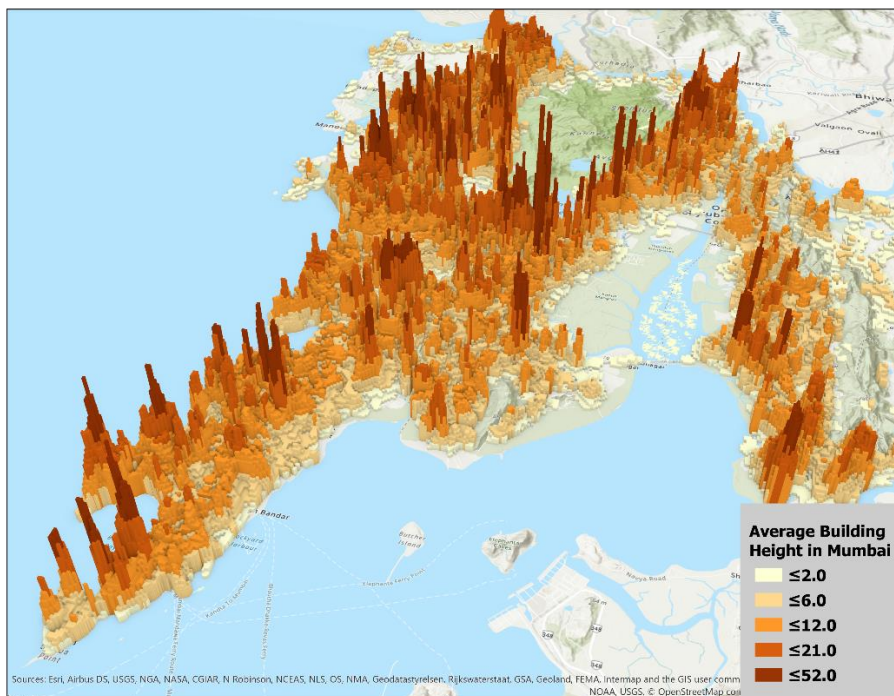
¹⁰ © OpenStreetMap contributors. This database is open source, licensed under opendatacommons.org/licenses/odbl/ (ODbL, opendatacommons.org) <https://www.openstreetmap.org/copyright>.

Figure 3: Mumbai - Population and Location of Key Services



Source: Population data from WorldPop 2019, water points and public toilets from Open Street Maps

Figure 4: Average Heights



Source: Average height in meters, at the pixel level (based on data from the German Aerospace Center)

Other datasets may also be useful in identifying key services. For example, cellular data can be a good source to find specific points of interest towards where people move. By combining this with OSM, one can filter places where people concentrate that could potentially be shuttered under a lockdown scenario (school, places of work, certain amenities). Moreover, if water provision is done through mobile tanks, we can potentially identify open spaces within an informal settlement, as a proxy to where these tanks can be located.

However, these extra data needs compromise the possibility of rapidly scaling up this analysis to several cities across the world. For example, mobile data are currently available only in a small set of developing countries, and coverage sometimes is not as thorough as in US or Europe. Furthermore, the use of cell phones in informal settlements may be limited, which could bias the estimates of hotspots. For these reasons, we rely on global and already available datasets for this analysis¹¹, to show how this could potentially be scaled up globally. Clearly, if there are better data sources available for a particular city, this could be incorporated and will provide more insights on the identification of potential hotspots.

4. Hotspots in Action

We now describe the application of the methodology in three cities – Mumbai, Kinshasa¹², and Cairo. We first combine the population layer with the heights data to obtain our first adjusted density measure. This density measure takes into account the total space available per person. The next step is to account for minimal mobility around the neighboring pixels, using the queen contiguity criteria. After assigning the new density measures to each pixel, we use our pre-defined threshold of 0.29 to identify those pixels with a population density that does not allow people to maintain the required social distance of 2 meters.

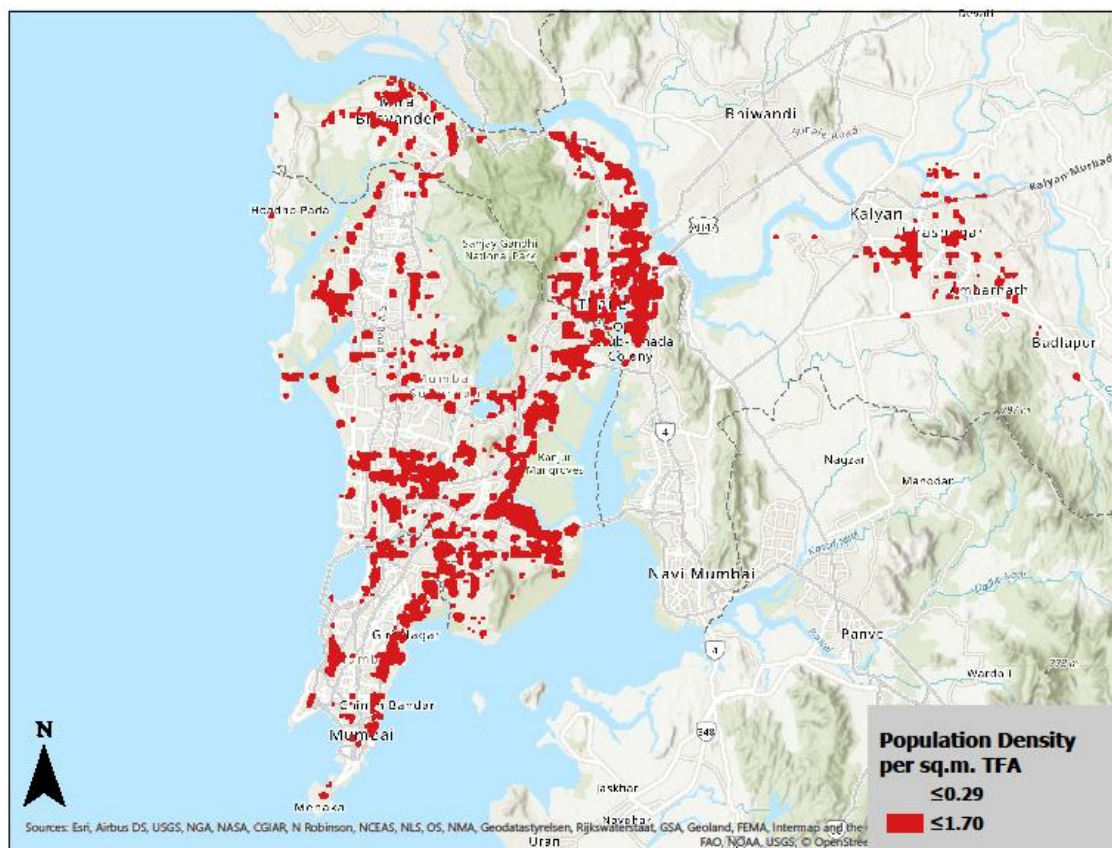
We identify these sets of hotspots in figures 5 to 7. The red areas correspond to potential hotspots, that is, areas where the population density (accounting for minimum mobility around the neighboring pixels) surpasses the minimum threshold required to maintain social distance. In Mumbai, these hotspots represent approximately 104.5 km^2 that account for approximately 4.5 million people. In a city with an estimated total population of 20 million, we find that approximately 20% of the population will not be

¹¹The simplified WSF3D dataset developed by DLR is currently available for a sample of 397 cities. We provide a list of countries for which we have heights data (for a subset of cities, depending on the country) in Annex I.

¹² Facebook population raster for Kinshasa shows a total of 5.7 people. However, current population in Kinshasa is approximately 14 million. We adjusted Facebook population estimates to reach this number, assuming the extra population follows the same distribution shown in the Facebook population raster.

able to maintain social distance and be at risk of rapid contagion. In Kinshasa, the hotspots cover approximately 105 km^2 , with a total of 6.11 million people at risk (approximately 43% of the total population).¹³ In Cairo hotspots cover almost 84 km^2 , accounting for 5.5 million people (approximately 25% of the total population).

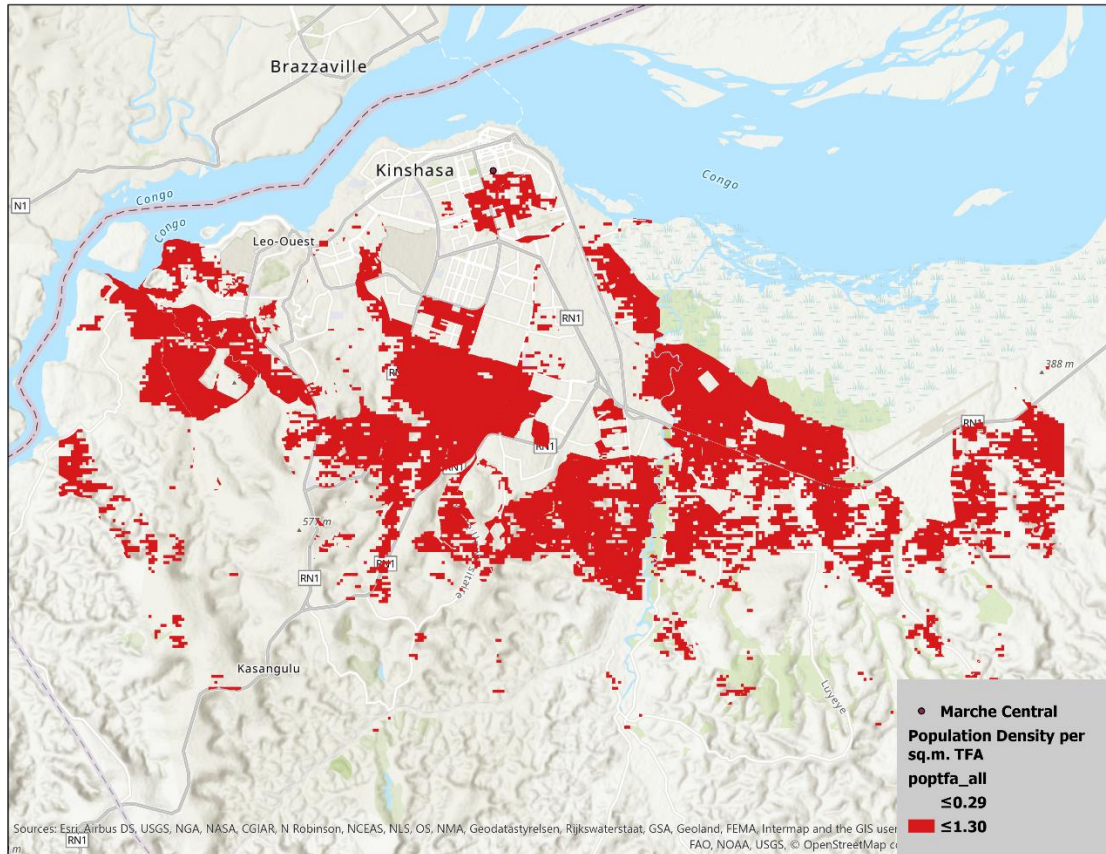
Figure 5: Potential Hotspots in Mumbai - Population Density and Minimal Mobility



Source: This map combines two pixel-level datasets: population data from WorldPop and heights data from DLR. The red areas represent the hotspots.

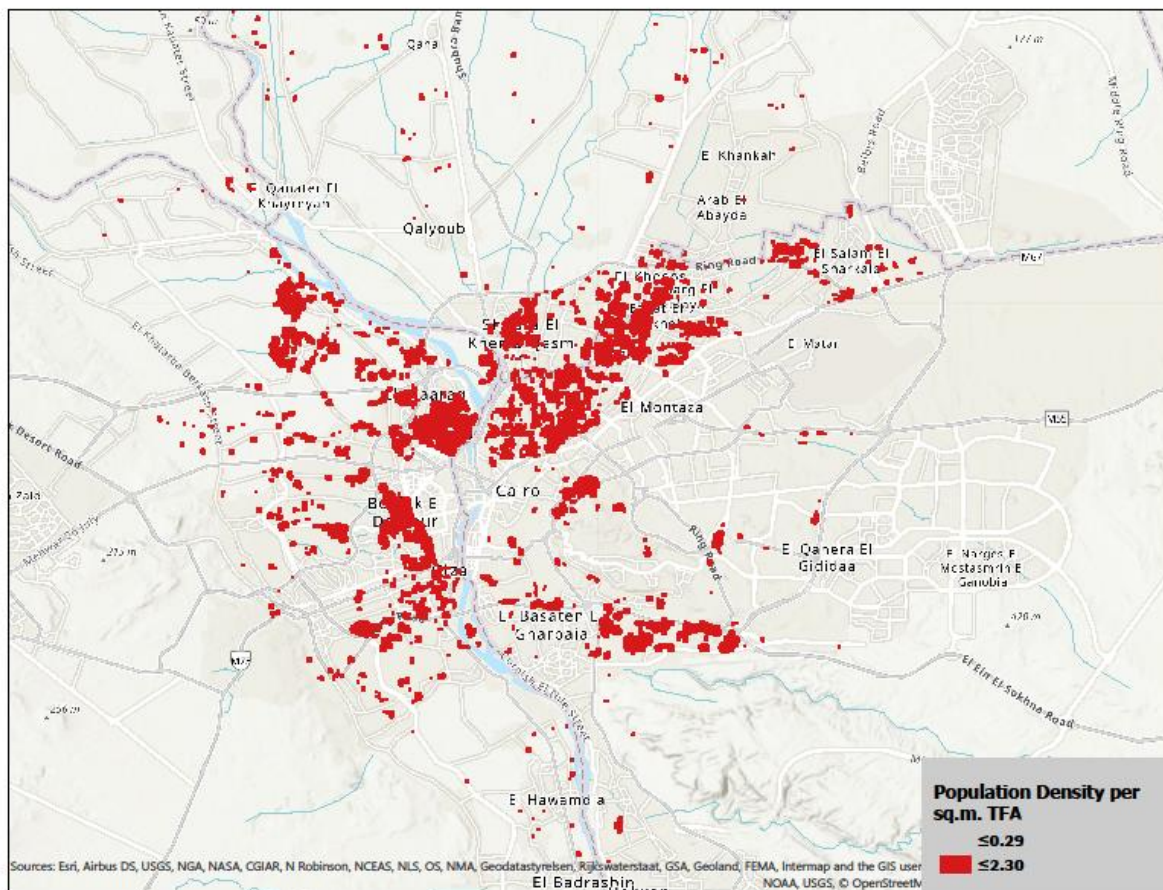
¹³ For Kinshasa, we focus on hotspots in areas where construction is precarious, using data from a World Bank survey on Access to Housing and Services in Kinshasa Province.

Figure 6: Potential Hotspots in Kinshasa - Population Density and Minimal Mobility



Source: This map combines two pixel-level datasets: population data from Facebook and heights data from the DLR, and data from the World Bank survey on Access to Housing and Services in Kinshasa Province showing areas where construction is precarious. The red areas represent the hotspots.

Figure 7: Potential Hotspots in Cairo - Population Density and Minimal Mobility



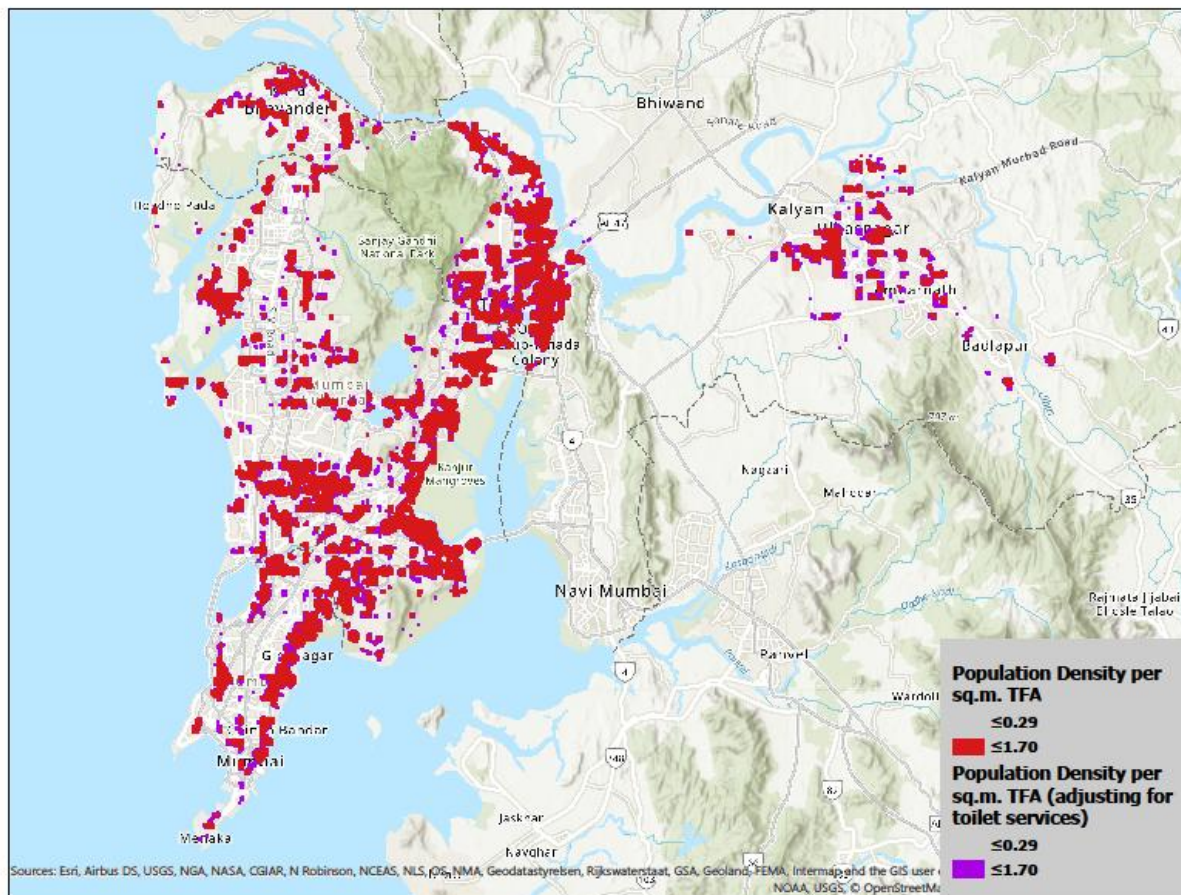
Source: This map combines two pixel-level datasets: population data from WorldPop and heights data from the DLR. The red areas represent the hotspots.

The fact that most of the affected population may live in informal settlements poses an extra concern. For example, people in these hotspots may not have access to water, which limits their ability to practice a proper hygiene routine, increasing the chances of contagion and viral spread. Some areas may have high levels of air pollution due to the type of cooking fuel used by households or illegal dumping and burning in the area. Since air pollution can cause lung and heart disease, these people may have a higher risk of complications from COVID-19.

After incorporating location of key services and mobility towards these services, we identify a second set of hotspots (Figures 8 to 10, in purple). For Mumbai, we use the location of public toilets as an example to a key service that people will still use even under lockdown. We find that additional hotspots cover an extra 15 km^2 , located often nearby the previous set of hotspots, with a few new areas affected as well. These extra hotspots account for more than 600 thousand people, bringing the total affected population

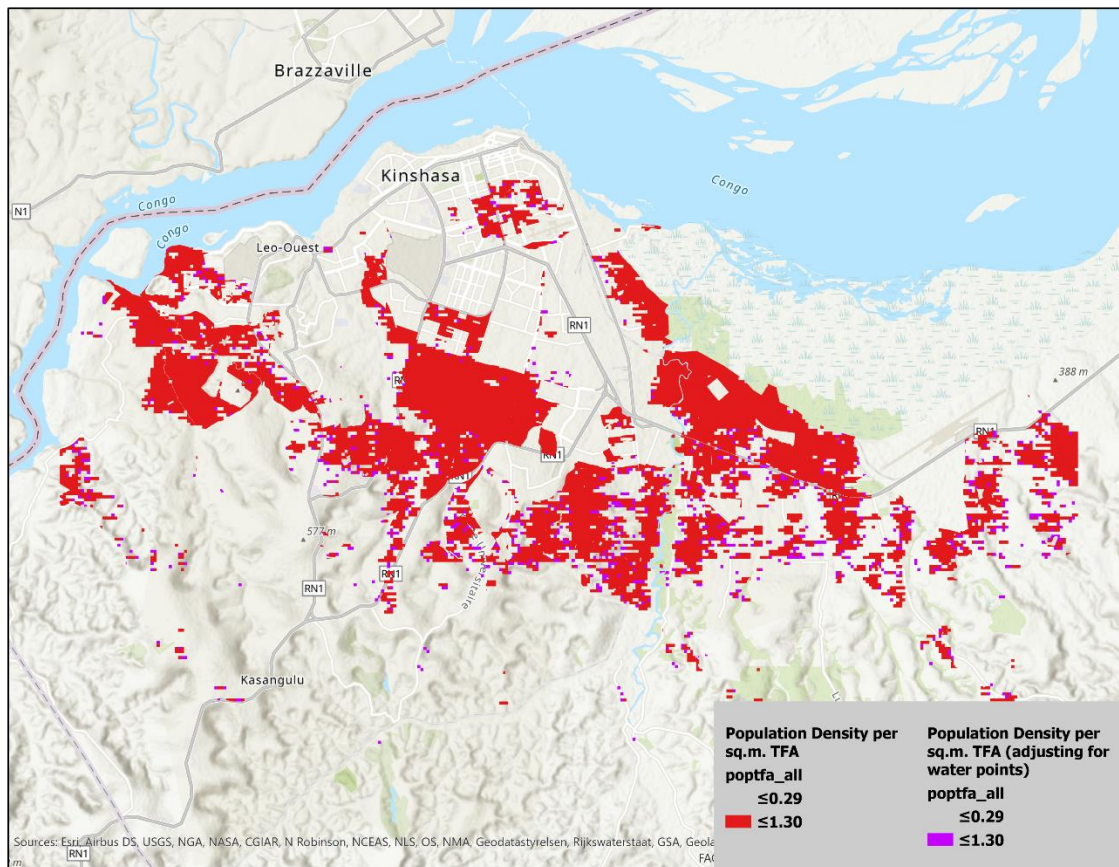
to 5.2 million (representing an increase of 15 percentage points of people at risk). In Kinshasa, we use the location of water kiosks as the key service of interest. We find that additional hotspots cover approximately 10 km^2 more, bringing the total number of people affected to 6.42 million (an increase of 5 percentage points of people at risk). In Cairo, we use the location of public toilets to find additional hotspots. These additional hotspots add 15 km^2 (to a total of almost 100 km^2) and the total number of affected people increases almost 11 percentage points (reaching 28% of the total population).

Figure 8: Potential Hotspots in Mumbai - Population Density, Service Location and Mobility



Source: This map combines two pixel-level datasets: population data from WorldPop and heights data from the DLR, as well as location of public toilets (OSM). The purple areas represent additional hotspots.

Figure 9: Potential Hotspots in Kinshasa - Population Density, Service Location and Mobility



Source: This map combines two pixel-level datasets: population data from Facebook and heights data from DLR, location of water kiosks (OSM), and data from a World Bank survey on Access to Housing and Services in Kinshasa Province showing areas where construction is precarious. The purple areas represent additional hotspots.

The map shows the city of Cairo with its major roads and districts. Red areas represent high population density per square meter of Total Floor Area (TFA), while purple areas represent high population density per square meter of TFA after adjusting for toilet facilities. The Nile River flows through the center of the city.

Population Density per sq.m. TFA

- ≤0.29
- >0.29

Population Density per sq.m. TFA (adjusting for toilets)

- ≤0.29
- >0.29

Sources: Esri, Airbus DS, USGS, NGA, NASA, CGIAR, N Robinson, NCEAS, NLS, OS, NMA, Geodatastyrelsen, Rijkswaterstaat, GSA, Geoland, FEMA, Intermap and the GIS user community, NOAA, USGS, © OpenStreetMap contributors

5. Conclusion

The methodology is simple and can be easily replicated using global data. The main data innovation has been to adjust population densities by introducing measures of floor space. This distinction is important

in considerations of the extent to which people can maintain spatial separation of 2 meters to decrease the risk of contagion.

More granular data on location of informal settlements and mobility towards specific spots can enrich the detection of potential hotspots. Future work can focus on refining the methodology and incorporating more granular data sources. Mobile data is a potential good data source to identify specific spots that are the most frequently visited. Location of slums and informal settlements in a city can help validate results and gain a better understanding of the magnitude of a viral spread and the number of people at risk.

References

Facebook Population Data, available at <https://dataforgood.fb.com/>

Lall, Somik Vinay, J. Vernon Henderson, and Anthony J. Venables. Africa's cities: Opening doors to the world. The World Bank, 2017.

Marconcini, M., Metz-Marconcini, A., Üreyen, S., Palacios-Lopez, D., Hanke, W., Bachofer, F., Zeidler, J., Esch, T., Gorelick, N., Kakarla, A., Strano, E. (2019): Outlining Where Humans Live – The World Settlements Footprint 2015. Scientific Data. Submitted. <https://arxiv.org/ftp/arxiv/papers/1910/1910.12707.pdf>

New York Times, “How can you social distance when you share a toilet with your neighbor?”, April 3, 2020, <https://www.nytimes.com/2020/04/03/opinion/coronavirus-south-africa.html>

Open Street Map dataset, © OpenStreetMap contributors. This database is open source, licensed under opendatacommons.org/licenses/odbl/ (opendatacommons.org/licenses/odbl/, opendatacommons.org/licenses/odbl/)
<https://www.openstreetmap.org/copyright>.

Palacios-Lopez, D., Bachofer, F., Esch, T., Heldens, W., Hirner, A., Marconcini, M., Sorichetta, A., Zeidler, J., Kuenzer, C., Dech, S., Tatem, A.J., Reinartz, P. (2019): New Perspectives for Improved Global Population Mapping arising from the World Settlement Footprint. Sustainability 2019, 11, 6056; <https://doi.org/10.3390/su11216056>

Sclar, E.D., Garau, P. and Carolini, G. (2005) The 21st Century Health Challenge of Slums and Cities. Lancet, 365, 901-903. [http://dx.doi.org/10.1016/S0140-6736\(05\)71049-7](http://dx.doi.org/10.1016/S0140-6736(05)71049-7)

United Nations, 2015, Millenium Development Goals Indicators. Indicator 7.10 Proportion of Urban Population Living in Slums. <http://mdgs.un.org/unsd/mdg/seriesdetail.aspx?srid=710>.

WorldPop population dataset, <https://www.worldpop.org/>

World Bank (2020). Profiling Living Conditions of DRC Urban Population: Access to Housing and Services in Kinshasa Province. Washington, DC: World Bank

Annex I

List of Countries		
Afghanistan	United Kingdom	Philippines
Angola	Ghana	Poland
United Arab Emirates	Guinea	North Korea
Argentina	Greece	Palestine
Australia	Guatemala	Romania
Austria	Hungary	Russia
Azerbaijan	Indonesia	Rwanda
Belgium	India	Saudi Arabia
Bangladesh	Iran	Sudan
Belarus	Iraq	Senegal
Bolivia	Israel	Singapore
Brazil	Italy	El Salvador
Canada	Japan	Somalia
Switzerland	Kazakhstan	Serbia
Chile	Kenya	Syria
China	South Korea	Thailand
Côte d'Ivoire	Libya	Tajikistan
Cameroon	Lithuania	Tunisia
Cameroon	Morocco	Turkey
Democratic Republic of the Congo	Mexico	Taiwan
Colombia	Mali	Tanzania
Cuba	Myanmar	Uganda
Germany	Mongolia	Ukraine
Denmark	Mozambique	United States
Algeria	Malaysia	Uzbekistan
Ecuador	Nigeria	Venezuela
Egypt	Nicaragua	Vietnam
Spain	Netherlands	Yemen
Ethiopia	Nepal	South Africa
Fiji	New Zealand	Zambia
France	Pakistan	