

The Nature of Human Settlement: Building an understanding of high performance city design.

Kerry A. Nice^{a,1,*}, Gideon D.P.A. Aschwanden^{a,1}, Jasper S. Wijnands^a, Jason Thompson^{a,c}, Haifeng Zhao^a, Mark Stevenson^{a,b}

^a*Transport, Health, and Urban Design Hub, Faculty of Architecture, Building, and Planning, University of Melbourne, Victoria 3010, Australia*

^b*Melbourne School of Engineering; and Melbourne School of Population and Global Health, University of Melbourne, Victoria, Australia.*

^c*Centre for Human Factors and Sociotechnical Systems, University of the Sunshine Coast, Australia.*

Abstract

In an impending urban age where the majority of the world's population will live in cities, it is critical that we improve our understanding of the strengths and limitations of existing city designs to ensure they are safe, clean, can deliver health co-benefits and importantly, are sustainable into the future. To enable this, a systematic and efficient means of performing inter- and intra-city comparisons based on urban form is required. Until now, methods for comparing cities have been limited by scalability, often reliant upon non-standardised local input data that can be costly and difficult to obtain. To address this, we have developed a unique approach to determine the mix, distribution, and composition of neighbourhood types in cities based on dimensions of block size and regularity, sorted by a self-organising map. We illustrate the utility of the method to provide an understanding of the underlying city morphology by overlaying spatially standardised city metrics such as air pollution and transport activity across a set of 1667 global cities with populations exceeding 300,000. The unique approach reports associations between specific mixes of neighbourhood typologies and quantities of moving vehicles ($r=0.97$), impervious surfaces ($r=0.86$), and air pollution levels (aerosol optical depth $r=0.58$ and NO_2 $r=0.57$). What this illustrates, is that this unique approach can identify the characteristics and neighbourhood mixes of well-performing urban areas while also producing unique 'city fingerprints' that can be used to provide new metrics, insights, and drive improvements in city design for the future.

Keywords: self organizing map, city typologies, neighborhood typologies, urban, city science

1. Introduction

The world's urban areas are rapidly growing in size, density, and in importance with more than half the world's population living in cities; this is projected to increase to two-thirds by 2050 (UNDESA, 2019). Beyond population growth, internal migration from rural areas and the concentration of international arrivals into cities is further driving this urbanisation process (Raymer and Baffour, 2018). These accelerating changes and the growing need to accommodate increasing urban populations presents some key challenges but also opportunities. Poorly managed urbanisation can impact urban residents through increased pollution (Stevenson et al., 2016; Sallis et al., 2016; Landigan et al., 2017), urban heat (Coutts et al., 2012; Bowler et al., 2010), urban sprawl (Frank, 2000; Bettencourt and West, 2010), and social isolation (Vlahov and Galea, 2002). However, urbanisation can also promote economic activities and innovation, increased employment opportunities, and community building through centralised locations, with economies of scale reducing the need for sprawling infrastructure (Kühnert et al., 2006; Bettencourt et al., 2007; Lobo et al., 2013).

*Principal corresponding author

Email address: kerry.nice@unimelb.edu.au (Kerry A. Nice)

¹These authors contributed equally to this work

Managing this process requires data and knowledge about how cities work. However, we lack an objective method to compare cities around the world and especially lack a systematic quantitative method to identify different neighbourhoods and perform inner-city comparisons (Louf and Barthelemy, 2014). This leads to repeated reinvention of solutions for problems that occur globally and limits the possibility to learn from previous examples. Understanding and comparing cities has been done many times where crude aggregates of population density or GDP measures are juxtaposed. Our framework allows us to compare a single element, as small as neighbourhoods of cities, across the globe.

To discover how different cities work, previous research has been conducted through a number of methods to categorise cities and analyse underlying processes. Analysing the basic structural elements of road networks and urban blocks, often the most long lasting part of an urban area, can provide clues as to the processes under which city development occurs (Porta et al., 2006; Strano et al., 2012). In addition, the road structure can point to the dominant modes of transportation and governance systems underlying each urban area, grid structures reflecting a top-down planning system (Crouch and Mundigo, 1977; Courtat et al., 2011) while T-shaped crossings point to more disorderly (Jacobs, 1962) self-organised organic growth (Cardillo et al., 2006). Division of large land blocks (often originally agricultural land) can follow an evolutionary progression, either to medium sized manufacturing or smaller residential plots (Falkowski and Bitner, 2008). In addition, studies show that areas unconstrained by adjoining villages or topography are generally and most efficiently subdivided into smaller grids (i.e. regular rectangular plots) (Strano et al., 2012). While this research based on urban blocks and streets reveal insights into cities and how the resulting built environment influences a wide range of outcomes, including health, transport, and economic opportunity, they are limited by scalability and observer dependence. In addition, validation and applicability of these methods to all types of world-wide cities are limited by the ability to collect urban information that applies to all cities.

To overcome these difficulties, and building on recent advances in computing power, artificial intelligence, and the wide availability of urban imagery, new approaches have been created to discover unique characteristics of cities and how cities function. Large numbers of geo-tagged photos have been used to detect patterns of urban usage and public perception of a number of areas' functional and social attributes (Liu et al., 2016; Zhou et al., 2014). Place Pulse, a database of urban imagery using crowd-sourced classifications (including safety, beauty, and liveliness) has attempted to quantify perceptions of urban areas (Dubey et al., 2016; Naik et al., 2014) and inequality (Salesses et al., 2013). Doersch (Doersch et al., 2012) used a large number of geo-localised street level images to discover common visual features across a number of cities. Enabled by remote sensing data, night-time light data has been used to categorise cities into stages of urbanisation and levels of economic activities (Zhang and Seto, 2013).

Urban characteristics (road geometry, building dimensions and heights, and vegetation heights) have also been used to classify cities into typologies of differing periods of historical design and urban planning (i.e. 19th Century, 1950s, 1970s, etc.) (Hermosilla et al., 2014). The connection between the physical and topological structure of the road network in cities to the structural sociology field, transportation and economics, has been drawn by the 'space syntax' community and Hillier (Hillier, 1996) that established a correlation between configurations of urban forms and variations of human interactions within it.

Most methods described above require some amount of subjective classification of local input data; the quality and availability of which can vary widely across collection or political districts. Existing empirical methods highlight these mechanisms by evaluating the street network typology (Hillier and Hanson, 1989), but neglect their geometrical expanse and their function as places to stay. A method cannot rely solely on topology but needs to incorporate the urban geometry (Louf and Barthelemy, 2014).

Our proposed method uses block size and regularity information within a neighbourhood to comprehensively explore the characteristics of cities across selected domains. The fundamental nature of city blocks, defined as the area bounded by surrounding streets, can be read as a simplified schematic view of the city (Southworth and Ben-Joseph, 2003), highlighting both the structure and organisation of the city, as well as the process of the urban development and morphology. This makes the city block the most used and accessible urban element for urban analysis and basic common elements in city design typologies and theory. Blending theory with globally available datasets at scale can enable new insights into the form and function of the world's cities. This method allows us to find the mix of commonalities and unique elements, a 'fingerprint' of each city, to understand what about each city works and doesn't work, an understanding essential to best manage the world's rapid urbanisation for health and well-being.

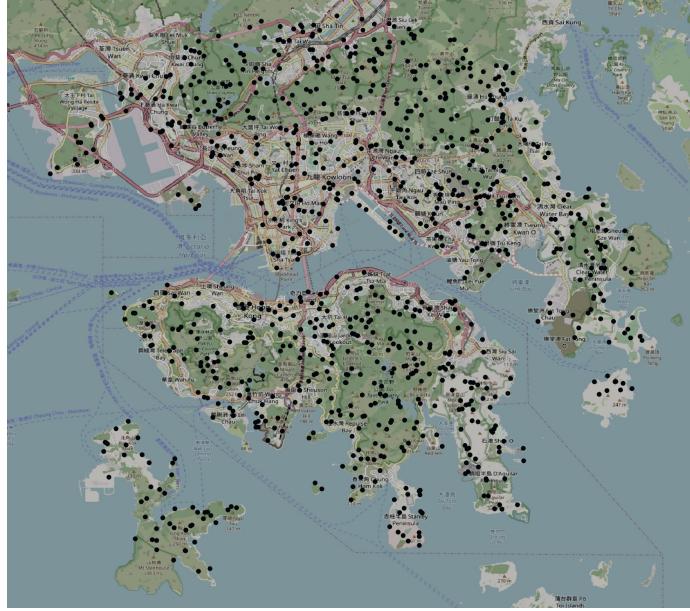


Figure 1: Sampling locations for map imagery (from Hong Kong).

2. Methods

2.1. Map imagery sampling.

The concept employed in this study was to sample maps of individual city sections, calculate block size and regularity of each section, and then use a self organising map (SOM) to organise the images into different urban types. All cities with populations greater than 300,000 people (United Nations, 2014) were selected for analysis. Map imagery from Google Maps (Google Maps, 2017) was used to provide globally consistent data.

A two-stage sampling approach was applied to each city. As no standardised urban boundaries are available for all the cities evaluated in this study, a methodology had to be developed to define these. Firstly, a sampling area extending 1.5 km from the identified city centroid (United Nations, 2014) was set as a baseline. Then the sampling radius r (km) was scaled, increasing by a power of 0.85 to the proportional increase in population size based on Barthelemy (Barthelemy, 2016) in

$$r = \sqrt{\frac{28.27}{\pi} \left(\frac{p}{300,000} \right)^{0.85}} \quad (1)$$

Standardising the sampling area in this manner avoided socio-political discrepancies relating to a city's 'true' (political) boundary and captured differences in population density and shape between small (e.g., Wellington, New Zealand; Izmit, Turkey) and global mega-cities (e.g., Tokyo, Japan; Delhi, India). Location sampling areas were adjusted for the earth's curvature (Sinnott, 1984). Large water-bodies (e.g., oceans but not coastlines) were removed from the sampling area, as they were not indicative of urban form.

These procedures result in a population and water body-adjusted circular area centred on the city's central coordinates, intended to capture the widest extent of each city while minimising the amount of non-urban locations. For example, Figure 1 shows the resulting sampling locations used in collecting imagery for Hong Kong.

2.2. Map imagery source.

320×320 pixel sized map images were sampled for 1692 cities using a zoom level of 16 (covering 750×750m at the equator and down to 335×335m at higher latitudes) using a custom style defined with the Google Static Maps API (Google Maps, 2017) (see Figure 2 for examples of Paris, France). To ensure each map covers the same area,

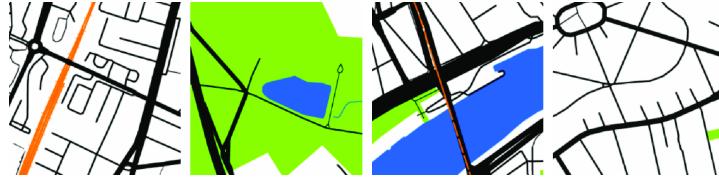


Figure 2: Four sample Google Maps training data images (from Paris, France) (Google Maps, 2017).

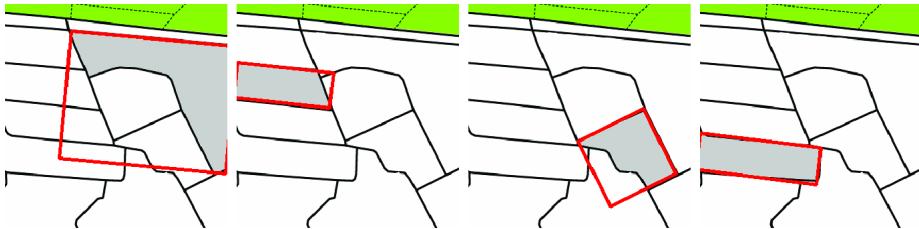


Figure 3: Results of flood filled city blocks showing flood fills of each individual region to determine region size (count of pixels in grey). Differences between region size and pixel counts within bounding boxes (outlined in red) are used as a measure of regularity.

each image was cropped and resized before processing. The sampled city at the highest latitude was at 64 degrees north, so each image was cropped and resized to include a region of 335×335m.

The maps provide a high-level abstraction of road (black) and public transport (orange) networks, green space (green), and water bodies (blue). Any remaining space is coded white. Inconsistent map imagery from 25 South Korean cities (due to South Korean government restrictions on map data (Badalge and Fairchild, 2018)) was removed from the dataset, reducing the number of cities to 1667. 1000 maps were sampled per city. The total dataset consists of nearly 1.7 million images.

2.3. Calculating block size, regularity, and colour counts.

Block size, regularity, and colour counts were calculated for each sampled image with Algorithm 1, using the Java 8 AWT toolkit (Oracle, 2018):

```

Result: Histogram of region sizes and region regularity for a single image
Using latitude of image sample, crop the image to represent 335×335 meters;
Resize image back to 320×320 pixels;
Start at top left point of image;
while White pixels are found do
    Floodfill area using boundaries of all non-white colours (i.e. black, green, blue, orange);
    Count pixel size of region;
    Construct a the smallest bounding box of the cloud of points in the region using the Fast Convex Hull
        algorithm (Javagl, 2017; Google Code Archive, 2011);
    Use the difference of counted pixels between the bounding box size and the region size as measure of
        irregularity;
    Add size and regularity counts to corresponding (pre-specified) histogram bin;
    Locate next white pixel by iterating across rows and columns;
end
Count percentage of blue, orange, green, black, and white pixels in each image.
Combine the two size and regularity histograms along with colour counts into a single histogram vector to be
used in the SOM.

```

Algorithm 1: Calculation of histograms of block sizes and regularity

Samples of size floodfills and regularity floodfills are shown in Figure 3. Sample histograms used in the SOM are shown in Figure 4.



Figure 4: Samples of map regions (top) and resulting histograms (bottom). Region size, regularity, and colour counts are joined into a combined histogram vector, with size frequencies in the first 15 bins, regularity in the second 15 bins and colour pixel counts in the remaining 5 bins.

2.4. City size and regularity histograms.

Using the calculated counts, two vectors were constructed for each image, one each for block size and block regularity. The vectors were sorted into 15 histogram bins (the number of bins determined by Sturges' formula (Sturges, 1926), $\lceil \log_2 n \rceil + 1$, with n being the number of data points). To reduce the clumping of data in the first bin, bins of increasing sizes were used to spread this data across all bins. The first bin starts with a size boundary of 1 and each following bin has a boundary of the current bin boundary times a multiplier. A multiplier of 2.3 was used to fit the maximum count size (320×320 pixels = 102400) into the 15 bins.

The resulting histograms for sample map regions are shown in Figure 4. Histograms input into the SOM were constructed by combining the 15 bins of region size frequencies (on the left side) with the 15 bins of region regularity frequencies (second 15 bins) into a single histogram vector. In addition, the 5 colour pixel percentages are appended to the end of the histogram vector. Finally, the vector values are normalised into a range of 0 to 1.

2.5. Sorting map histograms in the self organising map.

The SOM methodology (Kohonen, 1982) is a data driven technique that transforms a multi-dimensional data source into a lower dimensional space, commonly a two-dimensional map, while keeping the relative proximity of two datapoints intact. The distance in the lower dimensional representation is therefore a similarity index, calculated as the euclidean distance, of the higher dimensional space. Each point in the two-dimensional map has location (x,y) and is associated with a vector of values from the multi-dimensional space.

SOM is a generic, objective and robust methodology that has been deployed in many domains and is used for the visualisation of multi-dimensional data and data exploration (Kolehmainen, 2004). This methodology was chosen for its ability to create two-dimensional maps of smoothly changing patterns from the original high-dimensional space. Additionally, the SOM map spans the extremes observed in the original data and allows for investigation on how the data is distributed, potential paths between two observations.

The 1.7 million map histograms with 35 dimensions were the initial data space used to train the two-dimensional SOM. After the randomised initialisation of the 100x100 nodes of the SOM, a random selection of 5.4 million data points from the initial data space were used to transform the two-dimensional SOM to match it. This iterative process locates nodes that are similar to the training vector and morphs the values of the SOM nodes towards the training values. The number of iterations were determined by the minimum number of iterations to reach the greatest continuity among clusters. Training was stopped when the clusters started becoming discontinuous, checking every 150,000 iterations for a rapid increase in number of clusters. However, experimentally, we found that once a sufficient number of iterations were run, subsequent iterations only had small impacts on the results.

This training is subjugated to a decay function for both magnitude (learning decay, L_i) and distance (radius decay, D_R) in the SOM. Radius decay was calculated using

$$D_R = r_0 e^{-\frac{i}{n} \log_{10}(r_0)} \quad (2)$$

where radius (r_0) is 50 (the width and height of the SOM was 100) and the current training iteration i (of total iterations n of 3.2 million). Learning decay is calculated as

$$L_i = L_0 e^{-\frac{i}{n}} \quad (3)$$

where learn rate (L_0) = 0.05.

After the SOM was trained, each map histogram was classified to find the closest matching node in the SOM. The underlying imagery of the resulting trained SOM was visualised by tiling representative map images from each node (x,y) point in the SOM in Figure 5. Areas with black have no associated map segments with that particular node's (x,y) location (about 5000 nodes). Most nodes are associated with multiple map segments that have similar characteristics. One notable outlier exists that accumulates more than 60,000 map segments of (0,99) with 385,967 maps that are all or nearly all white. Approximately 20 nodes contain 60,000 to 10,000 maps, about 200 contain 10,000 to 1000 maps, 5700 nodes contain 1000 to 1 map.

The node's (x,y) locations were encoded into RGB colour codes using a Java 8 port of Color2D (Jäckle, 2017; Steiger et al., 2015). These colours were used in plotting the (x,y) typologies in QGIS (QGIS Development Team, 2009).

2.6. KernelDensityEstimator2D.

To create the city fingerprints, kernel density plots were made of the SOM x,y locations for each city using the KernelDensityEstimator2D (a bi-variate kernel density smoother for data) module from the Bayesian Evolutionary Analysis by Sampling Trees (BEAST) software package (Suchard et al., 2018) with a smoothed grid size of 100.

2.7. Aerosol Optical Depth (AOD) Dataset

Two near-identical MODIS (Moderate Resolution Imaging Spectroradiometer) instruments are mounted upon the sun-synchronous polar-orbiting NASA satellites Terra and Aqua; these missions have over-pass times of approximately 10:30am and 1:30pm local time and were launched in 1999 and 2004, respectively. The MODIS resolution is $10\text{ km} \times 10\text{ km}$ at nadir. The MOD04 and MYD04 L2 retrievals (V006) were downloaded and gridded to $0.05^\circ \times 0.05^\circ$ (grid-spacing of approximately 5.6km). A range of different retrievable fields are available of which we used "Optical_Depth_Land_And_Ocean" (see Table B1 of Levy et al. (2013) (Levy et al., 2013)), which represents a compromise between quality and coverage. These data were averaged to a monthly temporal resolution on this grid, and the number of MODIS pixels contributing to each averaged value were recorded. At each grid-point, a time-series decomposition into seasonal, trend and irregular (STL) components was applied (Cleveland et al., 1990). A slight modification to the code of the STL algorithm in the R stats package (R Core Team, 2015) was made so that data were weighted by the number of observations involved in creating these observations. The trend component is estimated as a smooth function (via the locally weighted scatterplot smoothing, or LOESS, algorithm of Cleveland et al. (1992) (Cleveland et al., 1992)), however the trend window parameter (defining the smoothing length scale for the trend term) was set sufficiently high that this term was effectively a linear term. From this, we derived the mean concentration at each grid-point accounting for long-term trend and seasonal fluctuations.

2.8. NO₂ Dataset

The tropospheric-column NO₂ data were derived from the TEMIS (Tropospheric Emission Monitoring Internet Service) OMI (Ozone Monitoring Instrument) tropospheric-column NO₂ (tcNO₂) database (Boersma et al., 2007). Monthly gridded averages at $0.125^\circ \times 0.125^\circ$ resolution (a grid-spacing of roughly 13km) were downloaded from the TEMIS website. These are based on the Level-2 tropospheric-column retrievals. While this provides little vertical information, the tcNO₂ product shows a good correlation with surface NO₂ concentrations, when averaged over a sufficient period. The OMI is mounted aboard NASA's sun-synchronous, polar-orbiting satellite Aura (launched July 2004). In normal operation, the OMI pixels are $13\text{ km} \times 24\text{ km}$ at their finest (i.e. at nadir). The combination of the instruments wide swath (spanning about 2600 km) and the 14 orbits daily provides a global coverage every day, however retrievals are not possible everywhere, mainly due to clouds. Airborne aerosols and surface albedo are other significant sources of uncertainty. The monthly tcNO₂ data from 2005-2016 (complete years) were averaged at their native resolution.

2.9. City fractions dataset

Middel et al. (2019,2018) (Middel et al., 2019, 2018) derived fractional breakdowns from Google Street View (GSV) panorama images for 65 million locations across 75 cities of six urban form classes of sky, trees, buildings, impervious surfaces, pervious surfaces, and non-permanent objects (i.e. moving vehicles). Data for each location was indexed by city, latitude, and longitude.

2.10. City correlations

Mean averages for each city were generated using the same centroid area as in the "Map imagery sampling" section above, however locations were sampled at a $400 \times 400\text{m}$ resolution instead of the 1000 randomly selected locations. Using these locations, mean values of AOD and NO_2 were calculated for 1667 cities. A second set of city mean values were generated for 34 cities that matched the cities sampled for this study using the city fractions dataset.

Next, mean averages of pollution (AOD and NO_2) and of urban form fractions were calculated for each $\text{SOM}(x,y)$ location. The latitude and longitude of each point (for the 1.7 million points) was used to look up an individual value from the gridded pollution datasets. For the city fractions dataset, latitude and longitude locations were matched to 0.0001 degree, for a total of approximately 8500 locations.

Finally, mean averages were calculated using the $\text{SOM}(x,y)$ averages for the 1000 locations for 1667 and 34 cities for pollution and urban form respectively. Then the `cor()` function, using "pairwise.complete.obs" and "pearson", from the R stats package (R Core Team, 2015) was used to find correlations between the two sets of computed city averages for the pollution and city fractions datasets.

3. Results

This project provides a means to examine all built environment designs from the largest cities around the world with greater than 300,000 population using an unsupervised sorting algorithm on maps segments of $300 \times 300\text{m}$ sampled from these cities. This provides insights at an unprecedented granularity, uncovering the scale of the neighbourhood, and the nature of cities. We have found unique individual neighbourhood types that can have twins in different cities across the globe. We also show the extent to which individual cities are heterogeneous within their bounds and the variation in distribution of neighbourhood types they display.

The methodology we present uses data on land use and block size extracted from maps and deploys a Self Organising Map (SOM) algorithm to sort them into a tractable array. This method is both universally applicable and extendible to incorporate additional datasets available while scaling linearly. The output of the SOM (Figure 5) contains an organised representation of all currently existing neighbourhood typologies in the urban areas where a large proportion (2.2 billion) of humans live in the world. The SOM sorts these neighbourhoods according to their typological elements, finding similarities in a high-dimensional space and projecting them into two-dimensions. Due to the stability of the physical structure of cities, undergoing long term small incremental changes (Wegener et al., 1986), this collection represents a collective representation of the built history of human settlements.

A look at the range of neighbourhood typologies from individual cities shows that many cities are an eclectic mixture of different neighbourhood typologies (Figure 6) but all are built using common elements. An individual city's uniqueness (in this case, the city's structure through its blocks and streets), its 'fingerprint', is reflected in an individual mix of neighbourhood typologies (Figure 7). These individual fingerprints can be used to compare cities around the world at a granularity of individual $300 \times 300\text{m}$ neighbourhoods. The uniqueness of cities are thus derived from a subtle mix and distribution of common elements (Figure 8).

As an illustration of the utility of this technique, we calculated correlations between mean average values of pollution and elements of urban form calculated for each city compared to averages calculated from a weighted average of the mix of $\text{SOM}(x,y)$ locations for each city. The datasets used are a global gridded pollution datasets of Aerosol Optical Depth (AOD) and NO_2 and dataset of fractions of six urban form classes of sky, trees, buildings, impervious surfaces, pervious surfaces, and non-permanent objects (i.e. moving vehicles) at 65 million locations in 70 cities from Middel et al. (2019,2018) (Middel et al., 2019, 2018). These values are reported in Table 1.

4. Broader Implications

Our contribution is a method that allows global inter-comparisons of neighbourhoods while highlighting that cities are constructed using fundamental particles; neighbourhood typologies. The implications are both fundamental and practical. The results provide for the first time a comprehensive overview of the nature of human settlements, their distribution and prevalence. This allows improvements to other systems such as health, transportation, and employment that are built on these fundamental components. The system can provide guidance for designers, engineers, stakeholders and policy makers by harnessing insights from a comparison across the globe.

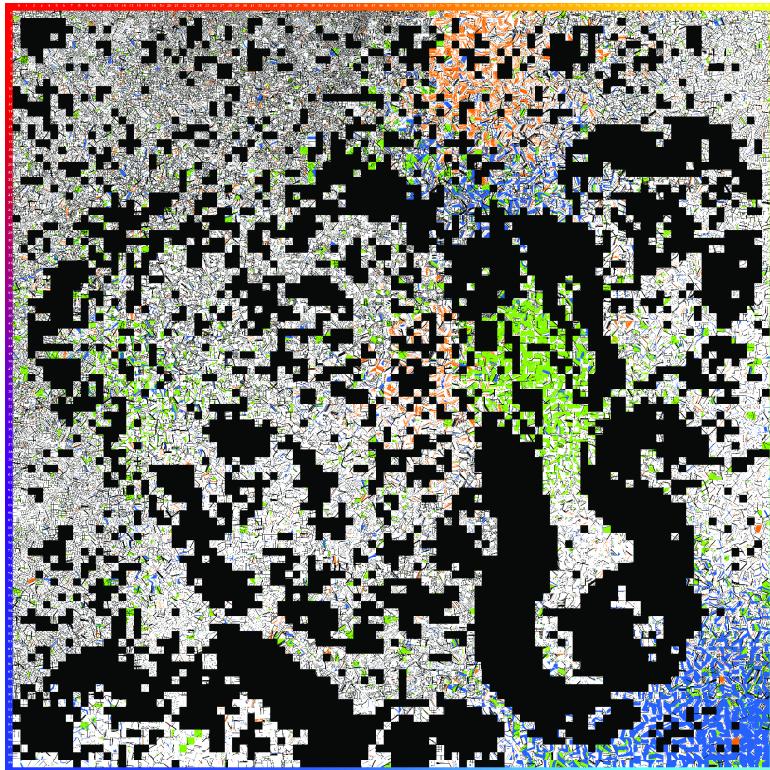


Figure 5: A visualisation of the 2-dimensional 100×100 SOM trained with 1.7 million map images from 1667 cities. Each (x,y) point shows a representative image associated with each node while nodes without associated images are shown in black. Border shows colour coding scheme for SOM (x,y) locations used in Figure 8.

Table 1: Correlations between mean average values by city and by (x,y) location within the SOM.

Parameter	Correlation value
Movable objects fraction	0.97
Impervious surfaces fraction	0.86
Sky fraction	0.75
Building fraction	0.56
Mean AOD	0.58
Mean NO_2	0.57



Figure 6: Sample representative maps for top SOM (x,y) locations for cities a) Jakarta, b) Tokyo, c) New York, d) Paris, e) Nairobi, f) Beijing, g) Barcelona, h) Melbourne, i) Sydney, and j) Brasília.

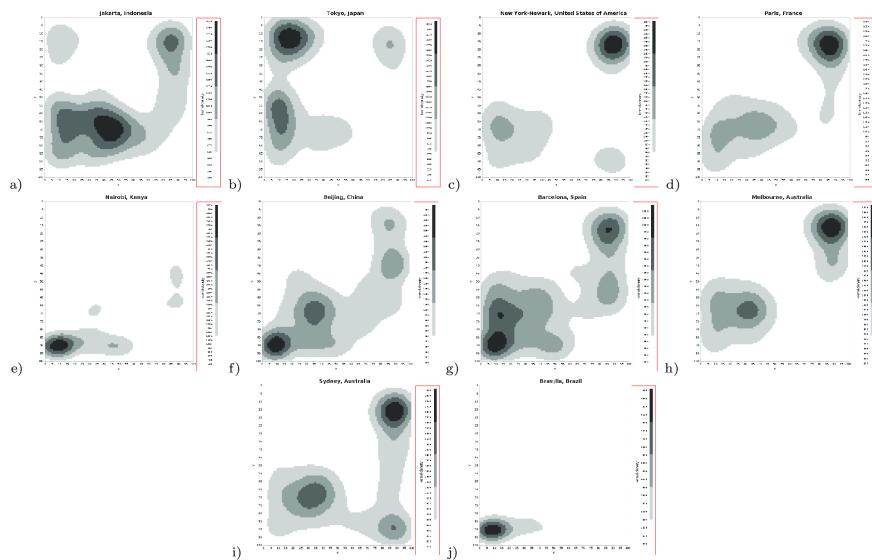


Figure 7: City fingerprints generated by kernel density maps of SOM (x,y) locations for cities a) Jakarta, b) Tokyo, c) New York, d) Paris, e) Nairobi, f) Beijing, g) Barcelona, h) Melbourne, i) Sydney, and j) Brasília.

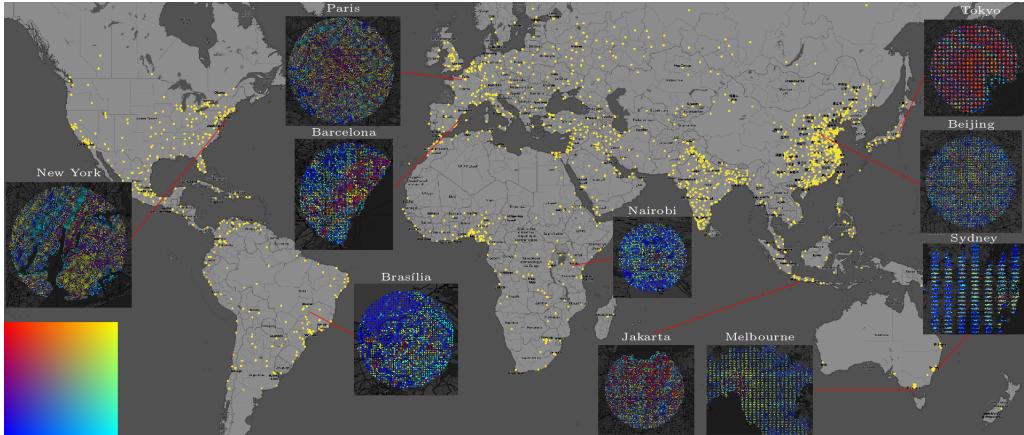


Figure 8: **Sampled world cities with inserts showing detail of New York, Paris, Barcelona, Brasília, Nairobi, Jakarta, Melbourne, Tokyo, Beijing, and Sydney.** City detail maps use the same SOM (x,y) location colour scheme as the border of Figure 5 and of colour map insert image (lower left).

As shown in the results, we find strong correlations between different mixes of neighbourhood typologies derived through block size and regularity and the morphology and composition (through fractions of movable vehicles, sky view, buildings and impervious surfaces) of these areas. In addition, we find correlations between the mix of neighbourhood typologies and the performance of the city in terms of pollutants (AOD and NO₂).

Our method has highlighted that cities are not unique and that individual neighbourhoods can be compared across continents. The findings also indicate that there is a common structure and that city typologies should be built across cities rather than within, that city centres are more comparable to other centres than to other elements of the same city. It also allows us now to compare cities on a global scale and investigate the nature of human settlements. In contrast to previous methods which were limited by the available data from a handful of cities, this method is capable of spanning the globe and can accommodate new datasets as they arise. For example, the link between urban form and vehicle emissions has been shown on an aggregate level (Frank, 2000). This can now be investigated at a neighbourhood scale. Similarly, transportation mode choices are associated with block size and accessibility that are primary enabler / inhibitor for residence using active modes of transportation. Finally, additional variables can be added to the vectors, such as demographic characteristics associated with each location (Kropp, 1998), to undercover another dimension of associations within the resulting neighbourhood typologies.

The implications of this work are two fold, it enhances our understanding of the environments that impact our individual decisions, the systems we create and the way we are producing them as well as has concrete application in several domains impacted by the built environment.

5. References

References

- Badalge, K., Fairchild, C., 2018. One Thing North Korea Has That the South Doesn't: Google Maps.
URL <https://nationalinterest.org/blog/the-buzz/one-thing-north-korea-has-the-south-doesnt-google-maps-24650>
- Barthelemy, M., 2016. The Structure and Dynamics of Cities: Urban Data Analysis and Theoretical Modeling. Cambridge University Press.
- Bettencourt, L. M. A., Lobo, J., Helbing, D., Kuhnert, C., West, G. B., 2007. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences* 104 (17), 7301–7306.
URL <http://www.pnas.org/cgi/doi/10.1073/pnas.0610172104>
- Bettencourt, L. M. A., West, G., 2010. A unified theory of urban living. *Nature* 467 (7318), 912–913.
URL <https://www.nature.com/articles/467912a>
- Boersma, K. F., Eskes, H. J., Veefkind, J. P., Brinksma, E. J., van der A, R. J., Sneep, M., van der Oord, G. H. J., Levelt, P. F., Stammes, P., Gleason, J. F., Bucsela, E. J., 2007. Near-real time retrieval of tropospheric NO₂ from OMI. *Atmos. Chem. Phys.* 7, 2103–2118.
- Bowler, D. E., Buyung-Ali, L., Knight, T. M., Pullin, A. S., 2010. Urban greening to cool towns and cities: A systematic review of the empirical evidence. *Landscape and Urban Planning* 97 (3), 147–155.

- Cardillo, A., Scellato, S., Latora, V., Porta, S., 2006. Structural properties of planar graphs of urban street patterns. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 73 (6), 1–8.
- Cleveland, R. B., Cleveland, W. S., McRae, J., Terpenning, I., 1990. STL: A Seasonal-Trend Decomposition Procedure Based on LOESS. *Journal of Official Statistics* 6 (1), 3–73.
- Cleveland, W. S., Grosse, E., Shyu, W. M., 1992. Local regression models. In: Chambers, J., Hastie, T. (Eds.), *Statistical Models*. Wadsworth & Brooks/Cole, Ch. 8.
- Courtat, T., Gloaguen, C., Douady, S., 2011. Mathematics and morphogenesis of cities: A geometrical approach. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 83 (3), 38–40.
- Coutts, A. M., Tapper, N. J., Beringer, J., Loughnan, M., Demuzere, M., 2012. Watering our Cities: The capacity for Water Sensitive Urban Design to support urban cooling and improve human thermal comfort in the Australian context. *Progress in Physical Geography* 37 (1), 2–28.
URL <http://ppg.sagepub.com/cgi/doi/10.1177/0309133312461032>
- Crouch, D. P., Mundigo, A. I., 1977. The City Planning Ordinances of the Laws of the Indies Revisited. Part II: Three American Cities. *The Town Planning Review* 48 (4), 397–418.
URL <https://www.jstor.org/stable/40103295>
- Doersch, C., Singh, S., Gupta, A., Sivic, J., Efros, A., 2012. What Makes Paris Look like Paris? *ACM Transactions on Graphics*, Association for Computing Machinery 31 (4).
- Dubey, A., Naik, N., Parikh, D., Raskar, R., Hidalgo, C. A., 2016. Deep learning the city: Quantifying urban perception at a global scale. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9905 LNCS, 196–212.
- Fialkowski, M., Bitner, A., 2008. Universal rules for fragmentation of land by humans. *Landscape Ecology* 23 (9), 1013–1022.
- Frank, L. D., 2000. Land Use and Transportation Interaction: Implications on Public Health and Quality of Life. *Journal of Planning Education and Research* 20 (1), 6–22.
- Google Code Archive, 2011. convex-hull, available at <https://code.google.com/archive/p/convex-hull/>.
- Google Maps, 2017. Google Static Maps API, Available from <https://developers.google.com/maps/documentation/static-maps>. (accessed 15 June 2017).
- Hermosilla, T., Palomar-Vázquez, J., Balaguer-Besér, Á., Balsa-Barreiro, J., Ruiz, L. A., 2014. Using street based metrics to characterize urban typologies. *Computers, Environment and Urban Systems* 44, 68–79.
- Hillier, B., 1996. Space is the machine. Cambridge University Press.
- Hillier, B., Hanson, J., 1989. *The Social Logic of Space*. Cambridge University Press, Cambridge.
- Jäckle, D., 2017. Color2D, available at <https://github.com/dominikjaeckle/Color2D>.
- Jacobs, J., 1962. *The Death and Life of Great American Cities*. Vintage Books, New York.
- Javagl, 2017. Geom: Utility classes for geometry computations, available at <https://github.com/javagl/Geom>.
- Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 43 (1), 59–69.
- Kolehmainen, M., 2004. Data exploration with self-organizing maps in environmental informatics and bioinformatics. Phd, Helsinki University of Technology.
- Kropp, J., 1998. A neural network approach to the analysis of city systems. *Applied Geography* 18 (1), 83–96.
URL <https://www.sciencedirect.com/science/article/abs/pii/S0143622897000489>
- Kühnert, C., Helbing, D., West, G. B., 2006. Scaling laws in urban supply networks. *Physica A: Statistical Mechanics and its Applications* 363 (1), 96–103.
URL <https://www.sciencedirect.com/science/article/abs/pii/S037843710600094X>
- Landigan, P. J., Fuller, R., Acosta, N. J., Adeyi, O., Arnold, R., Basu, N., Baldé, A. B., Bertollini, R., Bose-O'Reilly, S., Boufford, J. I., Breysse, P. N., Chiles, T., Mahidol, C., Coll-Seck, A. M., Cropper, M. L., Fobil, J., Fuster, V., Greenstone, M., Haines, A., Hanrahan, D., Hunter, D., Khare, M., Krupnick, A., Lamphier, B., Lohani, B., Martin, K., Mathiasen, K. V., McTeer, M. A., Murray, C. J., Ndahimananjara, J. D., Perera, F., Potočnik, J., Preker, A. S., Ramesh, J., Rockström, J., Salinas, C., Samson, L. D., Sandilya, K., Sly, P. D., Smith, K. R., Steiner, A., Stewart, R. B., Suk, W. A., van Schayck, O. C., Yadama, G. N., Yumkella, K., Zhong, M., 2017. The Lancet Commission on pollution and health. *The Lancet* 6736 (17).
- Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., Hsu, N. C., 2013. The Collection 6 MODIS aerosol products over land and ocean. *Atmos. Meas. Tech.* 6, 2989–3034.
- Liu, L., Zhou, B., Zhao, J., Ryan, B. D., 2016. C-IMAGE: city cognitive mapping through geo-tagged photos. *GeoJournal* 81 (6), 817–861.
- Lobo, J., Bettencourt, L. M., Strumsky, D., West, G. B., 2013. Urban Scaling and the Production Function for Cities. *PLoS ONE* 8 (3).
URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0058407>
- Louf, R., Barthélémy, M., 2014. A typology of street patterns. *J. R. Soc. Interface* 11, 1–7.
- Middel, A., Lukasczyk, J., Maciejewski, R., Demuzere, M., Roth, M., 2018. Sky View Factor Footprints for Urban Climate Modeling. *Urban Climate* 25, 120–134.
URL <https://www.sciencedirect.com/science/article/abs/pii/S2212095518301883>
- Middel, A., Lukasczyk, J., Zakrzewski, S., Arnold, M., Maciejewski, R., 2019. Urban form and composition of street canyons: A human-centric big data and deep learning approach. *Landscape and Urban Planning* 183, 122–132.
URL <https://doi.org/10.1016/j.landurbplan.2018.12.001>
- Naik, N., Philipoom, J., Raskar, R., Hidalgo, C., 2014. Streetscore – Predicting the Perceived Safety of One Million Streetscapes. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 793–799.
- Oracle, 2018. Java Platform Standard Edition 8 Documentation. <https://docs.oracle.com/javase/8/docs/index.html>.
- Porta, S., Crucitti, P., Latora, V., 2006. The network analysis of urban streets: A primal approach. *Environment and Planning B: Planning and Design* 33 (5), 705–725.
- QGIS Development Team, 2009. QGIS Geographic Information System. Open Source Geospatial Foundation. URL <http://qgis.osgeo.org>.
- R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

- URL www.R-project.org
- Raymer, J., Baffour, B., 2018. Subsequent Migration of Immigrants Within Australia, 19812016. Population Research and Policy Review.
 URL <https://doi.org/10.1007/s11113-018-9482-4>
- Saleses, P., Schechtner, K., Hidalgo, C. A., 2013. The Collaborative Image of The City: Mapping the Inequality of Urban Perception. PLoS ONE 8 (7).
- Sallis, J. F., Bull, F., Burdett, R., Frank, L. D., Griffi, P., Giles-Corti, B., Stevenson, M., 2016. Use of science to guide city planning policy and practice: how to achieve healthy and sustainable future cities. *The Lancet* 6736 (16), 1–11.
 URL [http://dx.doi.org/10.1016/S0140-6736\(16\)30068-X](http://dx.doi.org/10.1016/S0140-6736(16)30068-X)
- Sinnott, R., 1984. Virtues of the Haversine. *Sky & Telescope* 68, 159.
- Southworth, M., Ben-Joseph, E., 2003. Streets and the shaping of towns and cities. Island Press.
- Steiger, M., Bernard, J., Thum, S., Mittelstädt, S., Hutter, M., Keim, D., Kohlhammer, J., 2015. Explorative Analysis of 2D Color Maps. In: WSCG 2015 Conference on Computer Graphics, Visualization and Computer Vision. pp. 151–160.
- Stevenson, M., Thompson, J., Sá, T. H. D., Ewing, R., Mohan, D., Mcclure, R., Roberts, I., Tiwari, G., Giles-Corti, B., Sun, X., Wallace, M., Woodcock, J., 2016. Land use, transport, and population health: estimating the health benefits of compact cities. *The Lancet* 6736 (16), 1–11.
 URL [http://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(16\)30067-8/abstract](http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(16)30067-8/abstract)
- Strano, E., Nicosia, V., Latora, V., Porta, S., Barthélémy, M., 2012. Elementary processes governing the evolution of road networks. *Scientific Reports* 2.
- Sturges, H. A., 1926. The Choice of a Class Interval. *Journal of the American Statistical Association* 21 (153), 65–66.
- Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., Rambaut, A., 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution* 4 (1), 1–5.
- UNDESA, 2019. World Urbanization Prospects 2018. Tech. rep.
 URL <https://population.un.org/wup/Publications/>
- United Nations, 2014. Department of Economic and Social Affairs, Population Division, World Urbanization Prospects: The 2014 Revision, CD-ROM Edition.
- Vlahov, D., Galea, S., 2002. Urbanization, Urbanicity, and Health. *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 79 (4), 1–12.
- Wegener, M., Gnad, F., Vannahme, M., 1986. The Time Scale of Urban Change. In: Hutchinson, B., Batty, M. (Eds.), *Advances in Urban Systems Modelling*. Spiekermann-Wegener, Amsterdam: North Holland, pp. 175–197.
- Zhang, Q., Seto, K. C., 2013. Can night-time light data identify typologies of urbanization? A global assessment of successes and failures. *Remote Sensing* 5 (7), 3476–3494.
- Zhou, B., Liu, L., Oliva, A., Torralba, A., 2014. Recognizing city identity via attribute analysis of geo-tagged images. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), *Computer Vision ECCV 2014. ECCV 2014. Lecture Notes in Computer Science*. Vol. 8691. Springer, Cham, pp. 519–534.