A New Way to Proxy Levels of Infrastructure Development

Steve Pickering, Brunel University London

December 14, 2016

Abstract

Researchers in many fields have needed to develop a measure of infrastructure, and many proxies have been used toward this end, such as night light data and the Digital Chart of the World (DCW). Yet there are issues in using these methods. This paper presents a new way of proxying infrastructure: analysing the file sizes of map images on the Bing, Google, OpenStreetMap and Sina websites. The paper also demonstrates four ways in which this can be achieved. This approach is by no means perfect and does not solve all of the difficulties presented by other methods. Nevertheless, it does provide a simple and functional alternative proxy for level of infrastructure development.

Background and motivation

Many researchers have tried to include a measure of infrastructure in their work for a wide variety of research goals. Unfortunately, there is no single data set which can give us a geo-coded measure of infrastructure. Accordingly, many proxies have been used, among them, the roads layer from the Digital Chart of the World (DCW), the Gridded Population of the World (GPW) data set (CIESIN & CIAT, 2005), the Gridded Economic (G-Econ) data set (Nordhaus et al., 2006)) and more recently, the Global Roads data set (gRoads: CIESIN & ITOS (2013) and the Global Human Settlement Layer produced by the European Commission's Joint Research Centre (JRC, 2014)).[1] Yet these proxies

[1] A review of research using proxies to look at infrastructure is beyond the scope of this paper, but as a brief sample, researchers have used proxies to model infrastructure in their analyses of biodiversity loss (Alkemade

present numerous issues, in terms of the granularity of their data and their appropriateness to social science research. While projects such as the Global Human Settlement Layer are moving in the right direction, they are not complete yet. We need a simple measure of infrastructure to use in the meantime.

An alternative: using Bing Maps, Google Maps, OpenStreetMaps and Sina Maps

Accordingly, this paper presents an alternative means of proxying the level of infrastructure. By looking at the file sizes of images on internet map servers, such as Bing Maps, Google Maps, OpenStreetMaps and Sina Maps, we can make a reasonable inference into infrastructure development levels. To explain how and why map images are useful to proxy infrastructure, a brief discussion of the formats the map images are stored in is necessary.

Map file format and sizes

Bing, Google, OpenStreetMap and Sina have all chosen to use the Portable Network Graphics (PNG) format to display map images in web browsers and other client software. The reasons for their collective decision are quite clear: the PNG format supports lossless data compression, which means that while file size can be greatly reduced, the map images will not suffer from the compression artefacts typical of other formats, such as those used in Joint Photographic Experts Group (JPEG) files; additionally, PNG files are not subject to the bit-depth limitations and potential patent issues of the earlier Graphics Interchange Format (GIF) standard.

The functional specification document of the PNG format is registered as ISO/IEC 15948 (ISO/IEC JTC 1/ SC 24 Committee, 2004), the abstract of which states '[t]he datastream and associated file format have value outside of the main design goal.' Using map images to proxy infrastructure, we are able to use one of these values. Inherently,

et al., 2009), conflict and malnutrition in Africa (Rowhani et al., 2011; de Sherbinin, 2011), desertification (Okayasu et al., 2010), distribution of HIV and AIDS in Africa (Kalipeni & Zulu, 2012), mobility patterns in the Ivory Coast (Dixon et al., 2014), civil wars in Africa (Buhaug & Rød, 2006), distribution of forest elephants (Yackulic et al., 2011) and the future of China's transport systems (Hou & Li, 2011).

the more complex a PNG image is, the greater its file size. We can use these file sizes as a proxy for the level of infrastructure development.

An illustration of this is made in Figure 1. As well as agreeing on a common image file format, Bing, Google, OpenStreetMap and Sina have also arrived at a common map image size: 256 by 256 pixels. When you visit the website of one of the four map servers, the image you see on screen is a composite of multiple tiles, each of these dimensions. Figure 1 presents individual tile images[2] from the four map servers for four cities: specifically (in descending order of population) Tokyo, Beijing, Littlerock (capital of the US state of Arkansas) and Timbuktu (capital city of the Timbuktu region of Mali). A quick glance at the sixteen tiles makes it clear that for three of the servers, the tiles of Tokyo are clearly the most complex, while those of Timbuktu are the least. This is reflected in the file sizes: for three of the map servers, the file sizes for Tokyo are the largest, and the smallest are for Timbuktu.[3]

It is also worth comparing the same city for the four servers. For Tokyo, the OpenStreetMap image has the largest file size. The reason for this is clear: at this zoom level, the creators of OpenStreetMap have decided to include many roads, whereas Bing have chosen to only include major roads. A note should also be made on place names. As the text of the place names is part of the PNG image, it will therefore contribute to the file size. Bing have used entirely Latin characters; OpenStreetMap have chosen Japanese; Google use both Latin and Japanese characters; Sina, naturally enough, use simplified Chinese. This means that for countries which use non-Latin characters, the file sizes for Google map images will be greater than for those which exclusively use Latin characters.

It should be noted that the maps also contain other information, such as rivers. These too will add to the file size. Nevertheless, the more infrastructure in the area should lead to a more complex map, which leads to a greater file size. Also, we have no assurance that the the maps are complete or consistent across states. However, by giving the capacity to use four map servers, we can make some comparisons.

[2]Google, OpenStreetMap and Sina all use the same numbering system for their tiled web maps; Bing uses quadtrees.

[3]The Chinese server, Sina, is the exception to this rule: it presents Beijing as the most complex. However, when using the Sina maps website, it becomes clear that the main focus of Sina's efforts have been on providing maps of China; for the rest of the world, only major roads and features are included. Sina's maps are included in this paper, however, as researchers looking at infrastructure in China may want to consider using this server.

Google Tokyo (41,889 bytes) | Bing Tokyo (23,554 bytes) | Open Tokyo (47,513 bytes) | Sina Tokyo (9779 bytes)

Google Beijing (23,400 bytes) | Bing Beijing (12,358 bytes) | Open Beijing (30,185 bytes) | Sina Beijing (22,239 bytes)

Google Little Rock (22,830 bytes) | Bing Little Rock (13,960 bytes) | Open Little Rock (17,021 bytes) | Sina Little Rock (8514 bytes)

Google Timbuktu (1,537 bytes) | Bing Timbuktu (5,916 bytes) | Open Timbuktu (4,080 bytes) | Sina Timbuktu (4312 bytes)

Figure 1: Comparison of four cities and four map servers at zoom level 9.

Applying the method

Applying this method is quite straightforward. To make things as easy as possible, four methods have been developed and made available on the replication website.[4]

Method 1: Download the data as text files or images To avoid having to make multiple requests to the image servers,[5] the file size data for Sina, OpenStreetMap and Google have been collected at zoom levels up to 9, 10 and 11 respectively and made available on the replication site. This is a snapshot captured in May 2016: further snapshots will be gathered to turn this into a time series.

Method 2: Use the infra package For users of the R language, the **infra** package has been made available to automate the process of creating a grid and downloading the infrastructure data. The package does all of the trigonometry necessary to identify the appropriate maps, then sends requests to the servers to find out the file sizes.

Method 3: Use the SpatialGridBuilder program For researchers who want to avoid programming, the SpatialGridBuilder program is a standalone program which can create grids of any part of the world and download infrastructure data.

Method 4: Write your own, based on the R code The basic R code has been made available and can be used as pseudo-code to build this in other languages. The principle is fairly straightforward and requires relatively little code.

Full details of these methods are available on the replication website.

Choosing a zoom level

One of the first things we need to do when using maps as indicators of infrastructure is to choose a zoom level.

[4] http://168.144.172.143/infra.html

[5] The terms of use of the map websites indicate that they will block users who download too many map images. This is presumably to prevent strain on the servers and to prevent people from re-using their data. In preparing this paper, it was necessary to send several million requests to the map servers; this was done without being blocked. In most instances, it is possible to determine the file size without even downloading it; instead, we send a request to the server to tell us how big the file is.

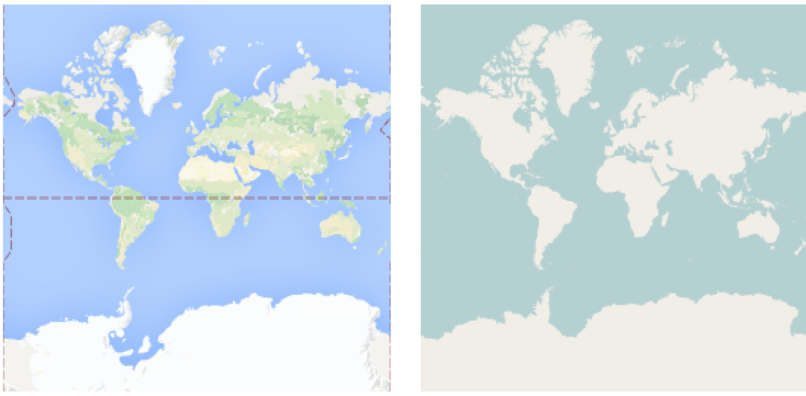


Figure 2: Google and Open, zoom level 0: the world in two 256 by 256 pixel tiles.

Google zoom	Bing zoom	Open zoom	Sina zoom	Tiles (width, height)	Degrees lng per tile
0	-	0	-	1	360
1	1	1	1	2	180
2	2	2	2	4	90
3	3	3	3	8	45
...	...	...	...	...	
18	18	18	18	262,144	0.00137
19	19	19	-	524,288	0.00069
20	-	-	-	1,048,576	0.00034
21	-	-	-	2,097,152	0.00017
22	-	-	-	4,194,304	0.00009

Table 1: Zoom levels, number of tiles, and degrees longitude per tile. Note that while Sina maps continue to zoom level 18 for China, for most of the rest of the world, they only continue to zoom level 9. For Taiwan, interestingly, they continue to zoom level 16, while the disputed Diaoyu/ Senkaku islands are covered at zoom level 18, the same as the Chinese mainland.

First, a brief explanation of how zoom levels work for the four map servers. Once again, Google, Bing, OpenStreetMap and Sina have adopted a common approach, although there are some differences in the level to which they will zoom. The zoom level is an exponent used to determine how many tiles the world will be divided into, both latitudinally and longitudinally.[6] Hence at zoom level 0 (only used by Google and Open), the number of map tiles needed to cover the globe is $2^0 = 1$ (or 1 tile wide $\times 1$ tile high; see Figure 2). Longitudinally, this means that one tile covers 360 degrees. The number of degrees longitude covered by each tile decreases exponentially as the zoom level increases, as is shown in Table 1.

The zoom level we use depends on our unit of analysis. Let us assume for the moment that we are using gridded data, and that our resolution is 0.5 degrees per grid cell

[6]Maps are Mercator projected, with polar cut-offs at around 85 degrees.

(such as with the PRIO-GRID used by conflict researchers - see Tollefsen et al. (2012)). The zoom level we choose should be at least as high as our grid resolution, otherwise we will be repeatedly downloading the same map image. As the zoom level 10 divides the world into $2^{10} = 1024$ tiles longitudinally, which are therefore 0.35 degrees longitude per tile, we can use level 10 as our starting point; it is, of course, possible to experiment.

Visualisations

Now that the basics of using map size as a proxy of infrastructure have been outlined, it is possible to present some renderings of the measure. Figure 3 shows the infrastructure levels of the world at three different zoom levels (9, 10 and 11) for Sina, Open and Google. As was mentioned earlier, the priority of Sina is to provide maps of China: this is evident in Figure 3. For Open, Europe has by far the highest values, with Japan, South Korea and the east coast of the US also showing high numbers. These places also achieve high values in the Google maps, but at this zoom level, other regions also show up: mountainous regions such as the Himalayas and the Andes are clearly visible.

Comparison with night light data

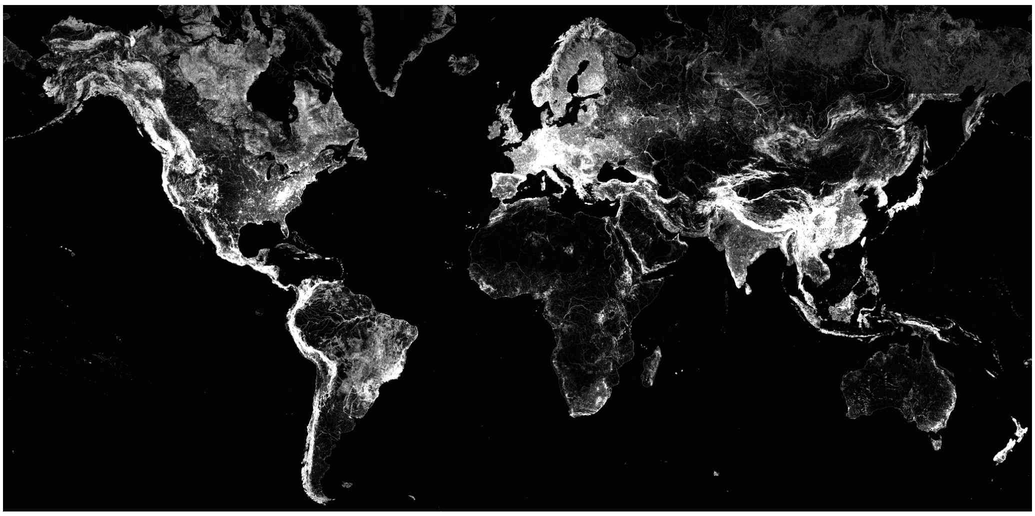
Night light data have also been used as a proxy for infrastructure. The data come from the National Geophysical Data Center of National Oceanic and Atmospheric Administration (NGDC-NOAA). The NGDC Earth Observation Group (EOG) provides nighttime observations of lights and combustion sources worldwide. The group started working with the Defense Meteorological Satellite Program (DMSP) data in 1994 and has produced a time series of annual cloud-free composites of DMSP nighttime lights.

Defense Meteorological Satellite Program (DMSP) night light data

As the name suggests, the fleet of satellites was developed in order to provide the US defense community with meteorological data. It was found, though, that a very nice additional benefit of the satellites was that they could be used to produce nightlight images. However, as producing nightlights was not their primary purpose, the satellites were not

Sina infrastructure

Open infrastructure



Google infrastructure

Figure 3: Renderings of infrastructure at three different resolutions and zoom levels: Sina at zoom level 9, 0.7 degrees per grid cell; Open at zoom level 10, 0.35 degrees per grid cell; Google at zoom level 11, 0.18 degrees per grid cell. All three images have been brightened and cropped for publication. For convenience, the data are available from the author's website at 168.144.172.143/infra.html

Zoom level	Google	Open	Sina
3	0.640061	0.6473644	0.5272382
4	0.5517033	0.6199686	0.5424614
5	0.5653091	0.7200928	0.5459509
6	0.5267032	0.6547015	0.6436776
7	0.5061606	0.6409491	0.6230644
8	0.5094356	0.6408868	0.6036957
9	0.5149355	0.634378	0.591096
10	0.4597864	0.633084	-

Table 2: Correlations between DMSP night light data (2013) and the infrastructure proxy at seven zoom levels

inter-calibrated; put another way, making year-on-year comparisons between nightlight data is possible, but presents issues.[7]

Table 2 presents a comparison between the infrastructure measure, derived from three map servers, and DMSP night light data. As can be seen, there is a positive correlation between the night light data and the three infrastructure measures. For the infrastructure measure based on Google maps, the correlation would appear to diminish as the zoom level increases; the reverse is the case for Sina, while correlations remain fairly constant for Open. More research is needed into why this might be the case.

We can explore this a little more, though, by looking at the parts of the world which have highest light levels, but comparatively low infrastructure levels. Interestingly, several of these points appear in Siberia. While attempts have been made in the night light data to remove surface light reflection, for instance from snow, this has not been entirely successful. The infrastructure measure is not affected by this problem.

Visibile Infrared Imaging Radiometer Suite (VIIRS) night light data

In 2012, the Earth Observation Group introduced new nightlight data from the Visible Infrared Imaging Radiometer Suite (VIIRS). The VIIRS was an improvement on the DMSP in several important ways: i) the satellites capture night light data at a higher resolution (15 arc second grids, compared with the earlier 30 arc seconds); ii) they offer more levels of variance in the light data (65,535 levels compared with the older 63 levels); and iii) the

[7] See Elvidge et al. (1997); Elvidge et al. (2009); Elvidge et al. (2014); Li et al. (2013).

	Night light (DMSP 2000)	Night light (VIIRS Oct 2012)	Google
DMSP 2000	-	0.622282	0.6804555
VIIRS Oct 2012		-	0.613364
Google			-

Table 3: Correlations between night light data and Google infrastructure zoom level 17: Tokyo

instruments on the VIIRS were designed with inter-calibration in mind, in order that time series comparisons could be made without additional mathematics.

Figure 4 shows three renderings of Tokyo. The image on the left is DMSP night-light data for the year 2000 at its native resolution. It is very easy to determine which parts of the image are city and which parts are water, but within the city itself, it is hard to determine much variance: much of the city is registering at the highest value for the DMSP (63). The higher resolution VIIRS data in the centre image provide a considerable improvement. However, even though the goal of the nightlight time series is to create cloud-free averages, this is not always possible, as is shown by the slight blurring of detail in parts of this image. This is not an issue for the infrastructure image on the right. Indeed, this image captures infrastructure data which the other two cannot: looking very closely, we can see that it has captured the 14km long Tokyo Bay Aqua Line. Night light satellites cannot capture this infrastructure because i) it is too narrow; ii) it is too dark; iii) over 9 kilometres of it is underwater. Table 3 shows that the two night light measures, plus the infrastructure proxy are all positively correlated, at a very similar level.

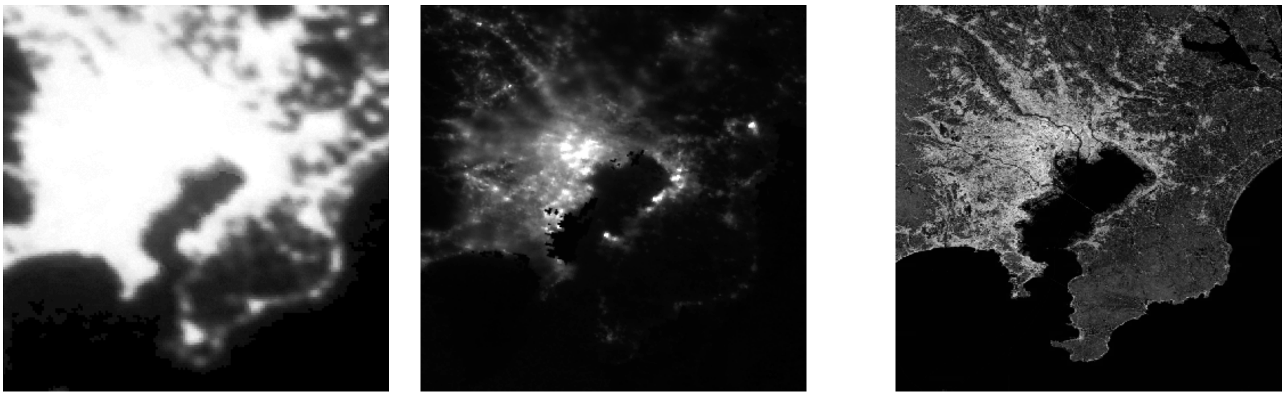


Figure 4: Tokyo comparison: DMSP night lights (2000) at native 0.008 degrees per grid cell; VIIRS night light (Oct 2012), at native 0.004 degrees per grid cell; Google infrastructure zoom level 17, at 0.004 degrees per grid cell

Dealing with mountains and rugged terrain

Looking back to Figure 3, the fact that mountainous regions and other regions of rugged terrain are identifiable in the Google map is potentially a problem: how can we separate mountains from infrastructure? There are two workarounds. The first is to use SRTM topography data and apply the King's move method to determine ruggedness (Anon: 2015): by doing this, we can essentially subtract the value of the mountains from the affected grid cells. Figure 5 gives an example. Looking at the top left image, we have a grid built at 0.35 degrees per grid cell, using Google zoom level 10. This shows high values in some of the places we would expect: again, Europe shows high values, with some cities identifiable. But mountainous regions are also identifiable. The top right image shows the hundred highsest values: some places in Europe, Japan, South Korea, Taiwan, but also places such as the Himalayas. The bottom left image shows the hundred most rugged terrains in the world, at this resolution, with places such as the Himalayas and the Andes showing up clearly. By subtracting the value of the rugged terrain from the map file sizes, we can find the hundred highest values of infrastructure in the bottom right image, which shows Europe, the east coast of the US and some key Asian cities. The method is not perfect: there are still some points in the Himalayas. Nevertheless, if research includes a mountainous region, this method can be applied.

The second method is much more simple: run the analysis at a higher resolution and zoom level. Mountain features take up less of the map image on higher zoom levels, but buildings and roads will take up more detail. Figure 6 demonstrates this. Looking at the Bing and Google maps, we can see a black swath running down the island. The ruggedness map, again based on a King's Move analysis of SRTM data, shows that this black swath is actually the most rugged part of Taiwan (the central mountain chain). Yet as it is shown in black on the Google and Bing images, at this higher resolution, the mountains are not being falsely interpreted as infrastructure.

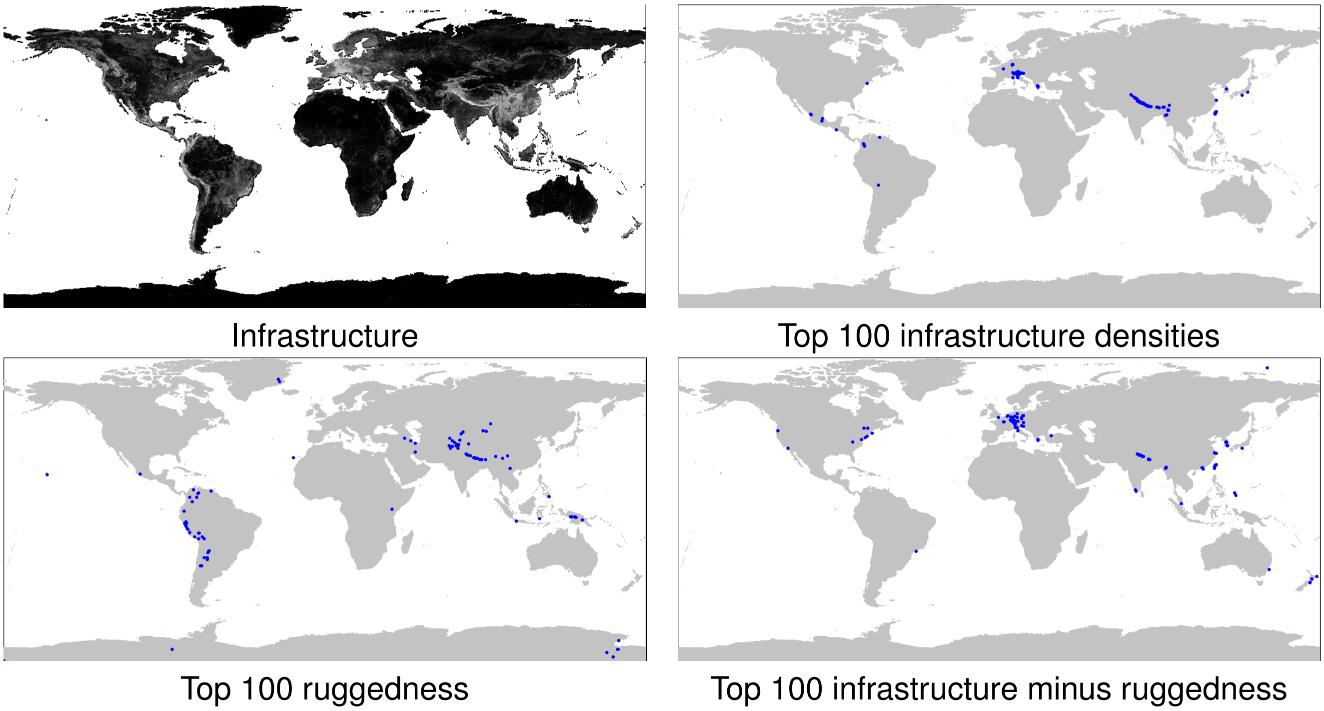


Figure 5: Infrastructure based on Google zoom 10, 0.35 degrees per cell; top 100 highest file sizes; top 100 most rugged regions, based on King's move analysis of SRTM data; top 100 infrastructures, based on filesize minus ruggedness of terrain.

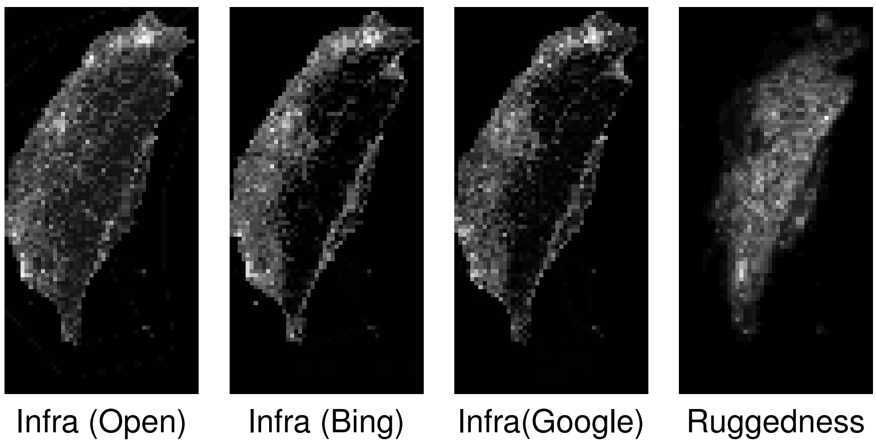


Figure 6: Infrastructure and terrain ruggedness in Taiwan. Zoom level 15

Comparison with road-based data sets: DCW and gRoads

Many researchers have tried to include a measure of infrastructure in their work for a wide variety of research goals. To do this, they have often used the roads layer from the Digital Chart of the World (DCW, later VMap0), because it was regarded as 'the first comprehensive global dataset' (Potere et al., 2009, 6536),[8] or as Nelson et al. (2006, 12) point out, it 'provides global coverage and it is one of the best and most widely-used publicly available road network data sets.'[9]

Yet while recognising the importance of the DCW, many have expressed their dissatisfaction with regard to its completeness and consistency (see Smith & Langaas (1995); Nelson et al. (2006, 12); Uchida & Nelson (2010, 10); Storeygard (2013); de Sherbinin (2011, 41)).[10]

Rather like the DCW before it, data from the gRoads project have also been considered as an infrastructure proxy. This is an attempt to create a comprehensive road data set based on modern data standards. Yet as Figure 7 illustrates, when operating at a very high resolution, data from gRoads do not have sufficient granularity to be used as an infrastructure proxy. Indeed, as the image demonstrates, using internet mapping data allows us to work at an even higher resolution than that used by the VIIRS project.[11]

Summary

This paper has pointed out that many researchers have needed to create a measure of infrastructure, and have used several proxies as a means to accomplish this. Some old proxies presented issues for researchers, while some new ones are moving in the

[8] A review of research using the DCW to look at infrastructure is beyond the scope of this paper, but as a brief sample, researchers have used the DCW to model infrastructure in their analyses of biodiversity loss (Alkemade et al., 2009), conflict and malnutrition in Africa (Rowhani et al., 2011; de Sherbinin, 2011), desertification (Okayasu et al., 2010), distribution of HIV and AIDS in Africa (Kalipeni & Zulu, 2012), mobility patterns in the Ivory Coast (Dixon et al., 2014), civil wars in Africa (Buhaug & Rød, 2006), distribution of forest elephants (Yackulic et al., 2011) and the future of China's infrastructure (Hou & Li, 2011).

[9] Nelson et al. (2006) write in reference to the Vector Smart Map Level 0 (VMap0) , which is the successor to the DCW. The differences between the DCW and VMap0 are essentially data formatting issues; in terms of actual road data content, they can be regarded as functionally equivalent. The specifications for VMap0 can be found at Defense Mapping Agency (1995b) and Defense Mapping Agency (1995c).

[10] For more on the history of the DCW, see National Imagery and Mapping Agency (1995); (Defense Mapping Agency, 1995a, 2); Langaas (1995, 10); anon 2016.

[11] Correlations are not presented here, as comparing raster and vector data is methodologically problematic.

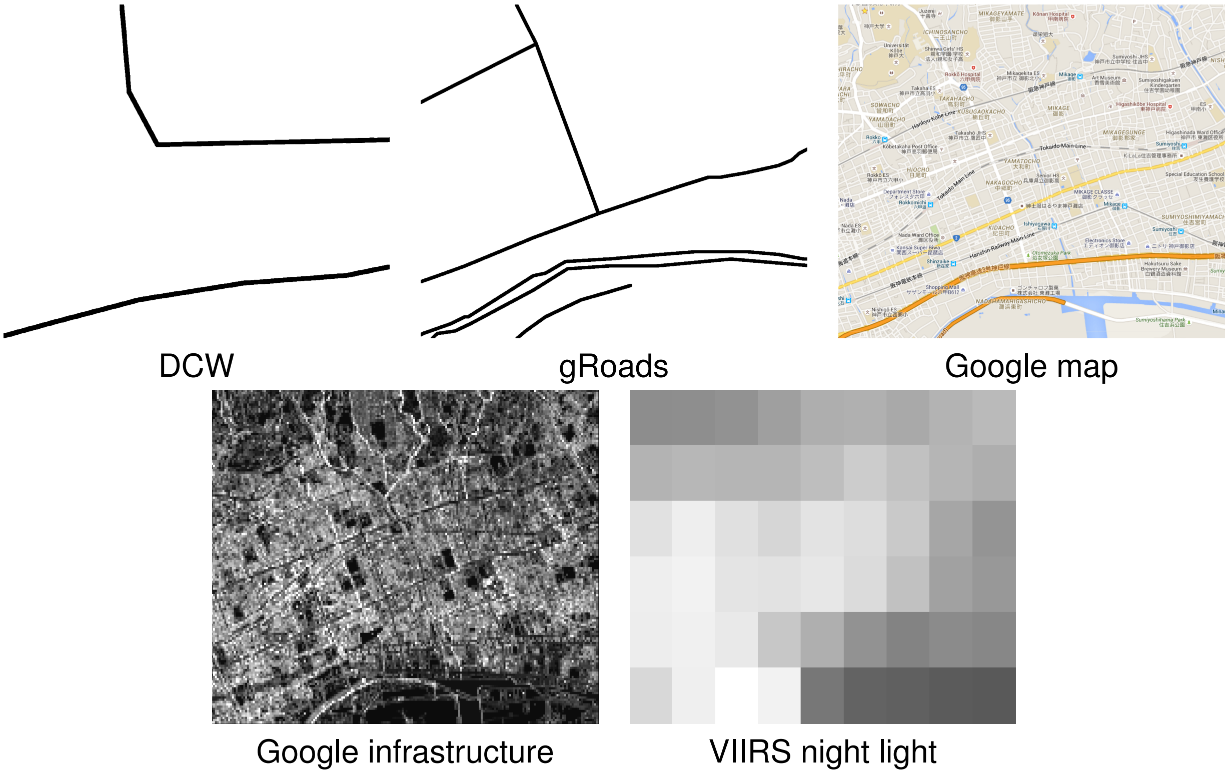


Figure 7: Kobe comparison: Digital Chart of the World (DCW, based on VMap0); gRoads; Google map image, zoom level 15; Google infrastructure, zoom level 21, 0.00017 degrees per grid cell (brightened for publication); VIIRS (Oct 2012) at native 0.004 degrees per grid cell

right direction, but are not yet ready. An easy-to-use alternative has been presented, which allows the researcher to capture four separate sources to create a data set. This alternative is by no means perfect: it too is subject to missing or inconsistent data, and introduces new issues, such as mountains and text affecting the results. Nevertheless, this paper has argued that map file sizes from Bing, Google, OpenStreetMap and Sina can act as a useful, if imperfect, proxy for the level of infrastructure. Future research will test whether this new measure can predict local wealth, in a more granular way than night light data (see Weidmann & Schutte (2016)).

References

Alkemade, R., van Oorschot, M., Miles, L., Nellemann, C., Bakkenes, M., & ten Brink, B. (2009). Globio3: a framework to investigate options for reducing global terrestrial biodiversity loss. *Ecosystems*, *12*(3), 374–390.

Buhaug, H., & Rød, J. K. (2006). Local determinants of African civil wars, 1970–

2001. *Political Geography*, *25*(3), 315–335.

CIESIN, & CIAT. (2005). *Gridded Population of the World (GPW), Version 3.* Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agrigultura Tropical (CIAT). Available at: http://earth-info.nga.mil/publications/specs/printed/VMAP0/vmap0_main.doc. Palisades, NY.

CIESIN, & ITOS. (2013). *Global Roads Open Access Data Set, Version 1 (gROADSv1).* Center for International Earth Science Information Network (CIESIN), Columbia University; Information Technology Outreach Service (ITOS), University of Georgia. Available at: http://dx.doi.org/10.7927/H4VD6WCT. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Defense Mapping Agency. (1995a). *Military Specification – Operational Navigation Charts (ONC), MIL-O-89102.* http://earth-info.nga.mil/publications/specs/printed/89102/89102.pdf. ([Online; accessed 12 February 2015])

Defense Mapping Agency. (1995b). *Military Specification – Vector Smart Map (VMap) Level 0, MIL-V-89039.* http://earth-info.nga.mil/publications/specs/printed/VMAP0/vmap0_main.doc. ([Online; accessed 12 February 2015])

Defense Mapping Agency. (1995c). *Military Specification – Vector Smart Map (VMap) Level 0, MIL-V-89039 Appendix.* http://earth-info.nga.mil/publications/specs/printed/VMAP0/vmap0_app.doc. ([Online; accessed 12 February 2015])

de Sherbinin, A. (2011). The biophysical and geographical correlates of child malnutrition in Africa. *Population, Space and Place*, *17*(1), 27–46.

Dixon, M. F., Aiello, S. P., Fapohunda, F., & Goldstein, W. (2014). Detecting mobility patterns in mobile phone data from the Ivory Coast. *Mobile Phone*

Data for Development, Analysis of mobile phone datasets for the development of Ivory Coast, *9*, 656–665.

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., Davis, E. R., & Davis, C. W. (1997). Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing*, *18*(6), 1373-1379.

Elvidge, C. D., Hsu, F.-C., Baugh, K. E., & Ghosh, T. (2014). National trends in satellite observed lighting: 1992–2012. In Q. Weng (Ed.), *Global urban monitoring and assessment through earth observation.* CRC Press.

Elvidge, C. D., Ziskin, D., Baugh, K. E., Tuttle, B. T., Ghosh, T., Pack, D. W., ... Zhizhin, M. (2009). A fifteen year record of global natural gas flaring derived from satellite data. *Energies*, *2*(3), 595-622.

Hou, Q., & Li, S.-M. (2011). Transport infrastructure development and changing spatial accessibility in the Greater Pearl River Delta, China, 1990–2020. *Journal of Transport Geography*, *19*(6), 1350–1360.

ISO/IEC JTC 1/ SC 24 Committee. (2004). *Iso/iec 15948:2004*. ISO.

JRC. (2014). *Global human settlement layer.* `http://ghslsys.jrc.ec.europa.eu/`. European Commission. (Joint Research Center)

Kalipeni, E., & Zulu, L. C. (2012). HIV and AIDS in Africa: a geographic analysis at multiple spatial scales. *GeoJournal*, *77*(4), 505–523.

Langaas, S. (1995). *Completeness of the digital chart of the world (dcw) database*. UNEP/GRID-Arendal.

Li, X., Chen, X., Zhao, Y., Xu, J., Chen, F., & Li, H. (2013). Automatic intercalibration of night-time light imagery using robust regression. *Remote sensing letters*, *4*(1), 45–54.

National Imagery and Mapping Agency. (1995). *Vector Map Level 0 (VMap).* http://earth-info.nga.mil/publications/vmap0.html. ([Online; accessed 12 February 2015])

Nelson, A., de Sherbinin, A., & Pozzi, F. (2006). Towards development of a high quality public domain global roads database. *Data Science Journal*, *5*, 223–265.

Nordhaus, W., Azam, Q., Corderi, D., Hood, K., Victor, N. M., Mohammed, M., ... Weiss, J. (2006). The g-econ database on gridded output: methods and data. *Yale University, New Haven*.

Okayasu, T., Okuro, T., Jamsran, U., & Takeuchi, K. (2010). Desertification emerges through cross-scale interaction. *Global Environ Res*, *14*, 71–77.

Potere, D., Schneider, A., Angel, S., & Civco, D. L. (2009). Mapping urban areas on a global scale: which of the eight maps now available is more accurate? *International Journal of Remote Sensing*, *30*(24), 6531–6558.

Rowhani, P., Degomme, O., Guha-Sapir, D., & Lambin, E. F. (2011). Malnutrition and conflict in east Africa: the impacts of resource variability on human security. *Climatic change*, *105*(1-2), 207–222.

Smith, C. G., & Langaas, S. (1995). *A survey of digital chart of the world (dcw) use and data quality*. UNEP/GRID-Arendal.

Storeygard, A. (2013). Farther on down the road: transport costs, trade and urban growth in sub-saharan Africa. *World Bank Policy Research Working Paper*(6444).

Tollefsen, A. F., Strand, H., & Buhaug, H. (2012). PRIO-GRID: A unified spatial data structure. *Journal of Peace Research*, *49*(2), 363–374.

Uchida, H., & Nelson, A. (2010). *Agglomeration index: towards a new measure of urban concentration* (Tech. Rep.). Working paper/World Institute for Development Economics Research.

Weidmann, N. B., & Schutte, S. (2016). Using night light emissions for the prediction of local wealth. *Journal of Peace Research*, 0022343316630359.

Yackulic, C. B., Strindberg, S., Maisels, F., & Blake, S. (2011). The spatial structure of hunter access determines the local abundance of forest elephants (Loxodonta africana cyclotis). *Ecological Applications*, *21*(4), 1296–1307.