Description of the data and how they were produced

This text is derived from: Weiss et al (2018) A global map of travel time to cities to assess inequalities in accessibility in 2015. Nature vol 553 p 333-6.
http://dx.doi.org/10.1038/nature25181

Overview

To model the time required for individuals to reach their most accessible city, we first quantified the speed at which humans move through the landscape. The principle underlying this work was that all areas on Earth, represented as pixels within a 2D grid, had a cost (that is, time) associated with moving through them that we quantified as a movement speed within a cost or 'friction' surface. We then applied a least-cost-path algorithm to the friction surface in relation to a set of high-density urban points. The algorithm calculated pixel-level travel times for the optimal path between each pixel and its nearest city (that is, with the shortest journey time). From this work we ultimately produced two products: (a) an accessibility map showing travel time to urban centres, as cities are proxies for access to many goods and services that affect human wellbeing; and (b) a friction surface that underpins the accessibility map and enables the creation of custom accessibility maps from other point datasets of interest. The map products are in GeoTIFF format in EPSG:4326 (WGS84) project with a spatial resolution of 30 arcsecs. The accessibility map pixel values represent travel time in minutes. The friction surface map pixels represent the time, in minutes required to travel one metre. This DANS data record contains these two map products.

Accessibility mapping methodology

The datasets used to construct the friction surface characterize the spatial locations and properties of roads, railroads, rivers, bodies of water, topographical conditions (elevation and slope angle), land cover, and national borders. The datasets were converted into aligning grids with a 30 arcsec resolution, with the pixel values representing speeds of movement. The layers were then combined such that the fastest mode of transport took precedence. The only exception to this logic was national borders, for which a crossing-time penalty was superimposed with priority over all other layers. The borders dataset was created from a UN global administrative units layer (GAUL) such that each border segment had a unique numerical identifier. This approach supports setting border-specific crossing times via a lookup table, however usable data do not presently exist for universally defining this parameter. As such, we used a static, one hour crossing-time penalty for all borders other than those within the Schengen and the UK–Ireland common security zones. Note that for readability the travel speeds for other input layers are provided in km h^{-1}, but the actual units within the friction surface raster are minutes required to travel one metre.

Two road datasets were combined for this research. The first road input layer consisted of vector data extracted from the OSM database, which was created by a user community dedicated to producing open-source, geocoded datasets of infrastructural resources. The OSM dataset was converted into a grid matching the geographic resolution and extent of the eventual friction surface. In cases for which vector features of more than one road type were present within a

single pixel, the road type with the highest associated travel speed took precedence. This rasterization procedure resulted in an integer grid in which pixel values corresponded to a single road type that was subsequently linked to a speed via a lookup table, which we also derived from the OSM database. The lookup table contains the country-specific mean travel speeds associated with each available road type, as derived from attributes linked to individual roads by the OSM user community. We used this lookup table approach rather than direct assignment of road speeds because such speed-of-travel information was infrequently assigned to road vectors within the OSM database. This limitation also necessitated the creation of a global default lookup table, which we created using mean values for each road type from all countries. We applied values from the default table in cases where a country had no speed limit records for one or more road types found within it.

The second, equally important source of road data was the Google distance to roads surface. This Google dataset was also global in extent, although China and the Korean Peninsula were omitted owing to data distribution limitations. To combine the two road datasets the Google distance to roads raster was first restricted to include only pixels with values of 500 m or less, thereby approximating the 1×1 km rasterization of the OSM road vectors. Unlike the OSM data the resulting Google roads raster lacked road-type information. As such the OSM road-type designation took precedence if both layers contained road information for a single pixel. Where only Google road data were available, the pixels were given the default integer value corresponding to the generic 'road' class from OSM. When creating the friction surface, all pixels from the combined roads raster were assigned the road travel speeds from the OSM-based lookup tables. For the lookup procedure, we also used a grid of administrative units to determine each pixel's country association.

The railroad input layer was also created from the rasterized OSM surface. Unlike the OSM roads data, however, the railroads were not differentiated by type within OSM and thus consisted of a single class with a uniform movement speed. The railroad speed used in this project was 24.3 km h^{-1}, which was the mean value assigned to railroad vectors extracted from the OSM database.

Three datasets were used to account for travel time by water within the friction surface. River travel time was added via a global set of navigable rivers rasterized from the CIA World Data Bank II vector rivers dataset. For inland water bodies, we used a newly created global surface-water occurrence dataset, which we first aggregated from its native 30-m resolution to create a layer that enumerated the fraction of each pixel's area that was covered by water at the resolution of the friction surface. In this procedure, all 30-m pixels within the resulting fractional surface-water dataset that were classified as water at least 80% of the time were considered permanent water, as 80% was the lowest occurrence value that we observed within ocean pixels when screening the data. The resulting fractional surface-water layer was then converted into a binary surface in which only pixels that were completely covered by permanent water were coded as a body of water amenable to be crossed by boat. The final dataset relating to water was a land–sea mask, which was used to identify ocean pixels. The movement speeds assigned to the water types within the friction surface were 10 km h^{-1} for rivers and lakes and 19 km h^{-1} for oceans.

The value for rivers was based on inland travel speeds reported in the UK, Ireland, and Australia. The ocean value was the average speed obtained from over 142 million observations of ocean-going passenger ships collected from the Automatic Identification System (AIS) and the Voluntary Observing Ship (VOS) program.

For all pixels not covered by any of the water, road or railroad datasets, we derived a baseline speed of movement overland (that is, on foot) using the MODIS MCD12Q1 land cover product in which we assigned each land cover type a travel speed from a lookup table. The lookup table was created by summarizing results from an online survey designed to crowd-source estimates of how long it takes individuals to traverse each land cover type. The survey consisted of representative photos and global maps of each land cover type. Respondents were asked to estimate the amount of time it would take them to travel one kilometre (or one mile) on foot through each land cover type. The survey received 407 complete responses and, after standardizing the distance units, the median values for the fifteen land cover classes within the survey (in units of km h^{-1}) were as follows: evergreen needleleaf forest = 3.24, evergreen broadleaf forest = 1.62, deciduous needleleaf forest = 3.24, deciduous broadleaf forest = 4.00, mixed forest = 3.24, closed shrublands = 3.00, open shrublands = 4.20, woody savannas = 4.86, savannas = 4.86, grasslands = 4.86, permanent wetlands = 2.00, croplands = 2.50, cropland/natural vegetation = 3.24, snow and ice = 1.62, and barren or sparsely vegetated = 3.00. The two land cover classes we excluded from the survey were (a) urban and built-up, which was given a speed of 5 km h^{-1}, but this value is almost never needed owing to the higher speed (and thus precedence) of roads that dominate urban landscapes at 1×1 km resolution, and (b) open water, which was given a speed of 1 km h^{-1}. The speed for the open water pixels was assigned using the rationale that if these pixels were not considered inland water within the water bodies layer (and would thus have received the inland water speed associated with boat travel) they were probably more akin to permanent wetland pixels that had a very high subpixel fraction of water to circumnavigate on foot. As such, all pixels that were classified as open water in the land cover layer but not as permanent water in the water bodies layer were given a speed half as fast as the crowd-sourced median speed for permanent wetlands of 2 km h^{-1}.

The land-cover-dependent travel speeds were then adjusted to take into account the effect of topographical properties. Topographical data-sets used in this analysis were produced from the Global Multi-resolution Terrain Elevation Dataset 2010 (GMTED2010), a derivative of the Shuttle Radar Topography Mission data and produced by USGS. The adjustment that we applied to elevation accounts for decreasing atmospheric density (and thus available oxygen) with altitude, which closely parallels the drop in maximal oxygen consumption (that is, VO2 max, a measure of optimal heart and lung function) and thus decreased the predicted travel speed as a function of altitude. On the basis of the standard atmosphere calculation, equation (1) shows the adjustment factor that we associated with elevation (in metres). We treated slope angle (in degrees) similarly, as steep terrain slows humans' ability to traverse it on foot. For the slope adjustment, we used Tobler's Hiking Function as shown in equations (2) and (3), with Tobler's walking speed capped to a maximum of 5 km h^{-1} and then divided by five to convert it into a fraction of maximum travel speed. The elevation and slope adjustment factors were subsequently

multiplied by the land-cover-dependent travel speeds, thus lowering the speed of travel on foot and increasing the time required to traverse each associated pixel within the friction surface.

$$\text{Elevation adjustment factor} = 1.016e^{-0.0001072 \times \text{elevation}} \tag{1}$$

$$\text{Tobler's walking speed} = 6e^{-3.5|\tan(0.01745 \times \text{slope angle})+0.05|} \tag{2}$$

$$\text{Slope adjustment factor} = \text{Tobler's walking speed}/5.0 \tag{3}$$

The final input for the accessibility map was the dataset of urban land cover, which was created using a layer from the Global Human Settlement (GHS) project. This dataset was produced using a combination of satellite imagery and census data to map the spatial distribution of urban areas across the globe. We selected the 'high-density centres' variant of the GHS dataset, which is defined as "contiguous cells with a density of at least 1,500 inhabitants per km^2 or a density of built-up greater than 50% and a minimum of 50,000 inhabitants". The dataset contained a total of 13,840 unique urban areas which consisted of a cluster of pixels, thus effectively representing urban areas as polygons. We used the least-cost-path algorithm to estimate travel time to the borders of these polygons.

The friction surface was created entirely within Google Earth Engine, which was also used to create the majority of the accessibility surface. In contrast to the process used to create the friction surface, deriving the accessibility map was very computationally intensive and required a more complex processing chain. Within Earth Engine, accessibility surfaces were generated using the cumulativeCost function, a least-cost-path function that was an experimental tool implemented specifically for this project but is now freely available within Earth Engine. By harnessing the computational power of the Google cloud-computing system the cumulativeCost function shortened the production time of the global accessibility surface from several months (when relying on local computing resources alone) to approximately two weeks. Despite reducing the production time substantially, the cumulativeCost function was still an evolving tool that was not yet capable of producing the global accessibility map in a single run or reliably producing output for latitudes above 60° if the friction surface was in geographical coordinates (that is, units of degrees latitude and longitude). As such, we created the global accessibility map by mosaicking a set of 31 tiles, 24 of which encapsulated the most computationally demanding areas and were generated within Earth Engine, and seven of which we created outside Earth Engine. The limitations of the least-cost-path function within Earth Engine at high latitudes were due to the nature of processing raster data stored in geographical coordinates because distances at high latitudes span far more degrees of longitude (and thus more pixels) than comparable distances at low latitudes. In order to parallelize computations efficiently, the Earth Engine cumulativeCost function required specification of a maximum search distance from the source points (that is, high-density urban land cover pixel centres), which we set to 1,500 km for most of the globe but reduced to 1,000 km in areas from 50° to 55° latitude owing to the afore-mentioned processing limitations at high latitudes. For latitudes above 50° we calculated accessibility tiles using the gdistance package in R36, thus ensuring an overlapping area of five-degrees latitude and providing data with which to compare the output maps from the differing sources (pixel

values in these areas proved to be almost identical). We also calculated accessibility times locally for very remote islands at lower latitudes that were beyond the 1,500 km search distance threshold from their closest cities. Lastly, the cumulativeCost function in Earth Engine could not account for wrapping at ±180° longitude, so we created an alternative version of the friction surface centred at this longitude and reprocessed approximately one-fifth of the globe outside of Earth Engine to ensure that any pixels that had their closest cities on the opposite side of this 'edge' were ascribed accurate travel times. We then mosaicked all of the tiles together by selecting the minimum travel times for all pixels that fell within overlapping portions of multiple tiles.