Sustainable Development Report 2023 – Annex

Methodological Summary and Data Tables

SDG TRANSFORMATION CENTER

A.1 Advancing and implementing an updated methodology for the Rural Access Index (SDG 9.1.1)

The Rural Access Index (RAI) is one of the most important global development indicators in the transport sector. It is currently the only indicator for the SDGs that directly measures rural accessibility, and it does so by assessing rural population's access to all-season roads.

The RAI was adopted as Sustainable Development Goal (SDG) indicator 9.1.1 in 2015, and is presented for the first time at the SDR in this current edition. The dataset developed and presented here implements and expands on the most recent official methodology put forward by the World Bank, the custodian of the indicator. This dataset was produced by UN SDSN's SDG Transformation Center and is, to date, the only publicly available application of this particular method at a global scale.

Background

The Rural Access Index (RAI) was developed by the World Bank in 2006, originally as a measure of poverty (Roberts et al., 2006). The original 2006 methodology based itself on pre-existing household surveys, which had several disadvantages including inconsistency across countries, lack of regular updates and cost constraints, which limited the index's sustainability and accuracy (Workman and McPherson, 2021).

A new methodology taking advantage of geospatial techniques was published under the "Measuring rural access using new technologies" report in 2016 (World Bank, 2016), thanks to a partnership established between the World Bank, the Department for International Development (DFID) of the United Kingdom and the Research for Community Access Partnership (ReCAP).

Later, in 2019, the World Bank commissioned a new ReCAP project led by the Transport Research Laboratory (TRL). The RAI Supplemental Guidelines (Workman and McPherson, 2019) provided new and detailed guidance for calculating the RAI, notably with an alternative approach to the all-season aspect of RAI, focusing on the changing accessibility profile of the road networks rather than relying on road surface quality alone or scarce physical measurements for road conditions. Nevertheless, neither the 2016 nor the 2019 methodologies were implemented globally, being restricted to more in-depth studies for selected countries.

The dataset developed and used in this year's SDR is, to the extent of our knowledge, the first global application of the 2019 TRL methodology (Workman and McPherson, 2019), although some changes and refinements have been proposed.

Methodology

The RAI is defined as the "proportion of the rural population who lives within 2 km of an all-season road". There is a common understanding that the 2 km threshold is a reasonable extent for people's normal economic and social purposes, and equates to approximately 20-25 minutes walking time. An all-season road is one that is motorable all year, but may be temporarily unavailable during inclement weather (Roberts, Shyam, & Rastogi, 2006).

The methodology consists of mapping where all-season roads are, applying a 2km buffer to them, and then assessing the proportion of the rural populations that falls within. This generates further questions such as defining what is urban and what is rural and assessing which roads provide all-season access or not, considering that no timely database containing that information is currently available at a global scale.

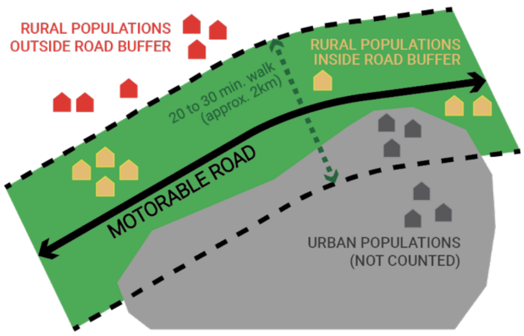

Figure 1 Summary of the RAI methodology

While other methodologies equate road surface to all-season status, the TRL 2019 methodology took into consideration that many rural roads in low-income countries (and even in large high-income countries) are unpaved, and often do provide all-season access. The innovation of this particular method (Workman and McPherson, 2019) lies in how the all-season status of roads is handled: instead of simply removing unpaved roads from the network, factors associated with inaccessibility are superimposed, and the population estimated to have access to a given road is kept in proportion to the **probability** that the road might be all-season.

Data sources

The indicator relies on four major geospatial data sets: those measuring land use (rural or urban), population distribution, road network extent and the all-season status of roads.

Land cover data (urban/rural distinction)

Since the indicator measures the access of rural populations, it's important to define what is and what isn't rural. This implementation uses primarily the DegUrba Methodology, proposed by the UN Expert Group on Statistical Methodology for Delineating Cities and Rural Areas (United Nations Expert Group, 2019). This approach has been developed by the European Commission into the Global Human Settlement Layer (GHS-SMOD) dataset, which is designed to confer consistency for definitions based on population density and built-up area (European Commission et al., 2021). While GHS-SMOD offers the best available temporal resolution, being updated at least every 5 years, its spatial definition (1km pixel) isn't ideal, and in some cases urban areas can't be well delineated. For this reason, data from NASA SEDAC CIESIN's GRUMP (CIESIN et al., 2018), or Global Rural Urban Mapping Project, is also used. The Urban Extent Polygons provided by GRUMP are limited to the year of 1995, but have a better spatial definition due to generalization of pixels into concave hull vectors. The GHS-SMOD and GRUMP datasets were put together and the overlap of urban areas from both datasets is used as the final urban land cover extent to be excluded from the analysis for RAI.

Population distribution

The source for population distribution data is WorldPop (WorldPop, 2023). It uses national census data, projections and other ancillary data from countries to produce aggregated, 100m² population data, making it the most spatially disaggregated population data currently available at global scale.

Road extent

One of the main issues identified in previous attempts to calculate RAI at global scale is ensuring that all roads are being taken into account. To respond to that, a redundancy strategy is used by simultaneously adopting three widely-recognized road datasets: the real-time updated, crowd-sourced OpenStreetMap (OSM, 2023), GLOBIO's 2018 GRIP database (Meijer et al., 2018), which draws data from official national sources, and Microsoft BING's Road Detection Project (Microsoft Bing, 2023), which identifies roads through Neural Network models. Each of these sources represents at least one advantage compared to the others:

- The GRIP database is the only global road network database containing information about the all-season status of roads, but to the detriment of its temporal resolution – which is restricted to the year of 2018 – and its coverage, restricted to what national authorities could provide at the time.

- OpenStreetMap (2022 reference year) provides excellent temporal resolution and at least two attributes from which the all-season status can be derived: surface and hierarchy. Although it provides good coverage, the network is limited by the volunteers' interest in certain regions, which might skew the coverage towards urban centres to the detriment of rural areas.

- Microsoft Bing's recent Road Detection (2022 reference year) project is used to ensure completeness. This dataset is completely derived from machine learning methods applied over satellite imagery, and detected 1,165 km of roads missing from

OSM, though there are currently no attributes associated to any of the roads.

The three datasets are put together in order to generate two final road subsets, all-season (paved) and exposed (unpaved). The distinction is important, as unpaved roads deteriorate rapidly and in a different way to paved roads (Workman and McPherson, 2019): unpaved roads are more exposed to water ingress to the surface, softening materials and making them vulnerable to traffic.

The first subset contains roads classified as all-season by GRIP and roads tagged as paved and/or as a hierarchy often (≥60%) associated with paved surfaces by OpenStreetMap (see Table 1). The population living within 2 km of these roads will be considered to have full access to an all-season road. The second subset contains all roads identified by Microsoft Bing's Road Detection project (as those aren't qualified in any way) and roads tagged as unpaved and/or as any of the remaining hierarchy tags, given that they're not also tagged as paved by OpenStreetMap.

The roads in the second subset are considered to be exposed to factors associated with difficulty of access. Their probability of being all-season is calculated by the superimposition of passability criteria, which are described in the following section. The population living within 2km of these roads will be considered in equal proportion to the probability that the road provides all-season access (i.e. if it's established that there's a 10% chance that a road is all-season, only 10% of the population living within 2km of it will be considered to have access to it).

Roads' all-season status

The World Bank's original 2006 methodology defines the term all-season as "… a road that is motorable all year round by the prevailing means of rural transport, allowing for occasional interruptions of short duration" (Roberts, Shyam, & Rastogi, 2006). TRL's 2019 (Workman and McPherson, 2019) methodology proposed that passability should equate to the all-season status of a road, along with the assumption that typically the wet season is when roads become impassable, especially so in steep roads that are more exposed to landslides.

This dataset implements a passability index, where each component is used as a multiplying factor ranging from near 0 to 1 over the population distribution layer whenever they're located exclusively inside a buffer generated by an exposed (unpaved) road. The proposed use of passability factors relies on the following aspects:

- Climate. Precipitation has a significant effect on the condition of unpaved roads, being a significant factor in its deterioration. We use the Copernicus Programme's (C3S, 2017) yearly accumulated precipitation data, which is made available freely at ~30km pixel resolution for reference year 2022.

Table 1 Field criteria used for each road subset

Source	Field criteria for inclusion	Final subset
OpenStreetMap	highway = *primary, primary_link, secondary, secondary_link, trunk, trunk_link, motorway, motorway_link* AND/OR surface = *paved, asphalt, chipseal, concrete, plates, paving_stones, sett, unhewn_cobblestone, cobblestone, metal, wood*	(1) All-season
GLOBIO'S GRIP	IsSeasonal = 2 (No, all year access)	
OpenStreetMap	highway = *unclassified, track, service, road, footway, bridleway, steps, path, tertiary, tertiary_link* AND/OR surface = *unpaved, compacted, fine_gravel, gravel, rock, pebblestone, ground, dirt, earth, grass, grass_paver, mud, sand, woodchips, snow, ice, salt*	(2) Exposed
Microsoft BING Road Detection	None (all roads included)	

- Terrain. The gradient and altitude of roads also has an effect on their passability. Steep roads become impassable more easily due to the potential for scouring during heavy rainfall, and also due to slipperiness as a result of the road surface materials used. Here this is drawn from slope calculated from SRTM Digital Terrain data (Jarvis et al., 2007), provided at ~30m pixel resolution.

- Road maintenance. The ability of local authorities to repair damage caused by precipitation and scouring is proposed as a correcting factor to the previous ones. Ideally, this would be measured by the % of GDP invested in road construction and maintenance, but this isn't available for all countries. For this reason, GDP per capita for reference year 2022 is adopted as a proxy, as provided by the World Bank (World Bank, 2023).

It's important to note that, differently from the suggestions of datasets made by TRL (Workman and McPherson, 2019), we exclusively chose datasets with at least medium spatial resolution, in raster format, and with temporal resolution of at least 1-year. This ensures that the results won't be the exactly the same when RAI is calculated every year.

In order for RAI to account for the probability that the roads people are using in are all-season or not, the disaggregated factors for accessibility are applied to the spatialized disaggregated population data at pixel level through raster algebra. The final passability index is measured on a scale of 0 to 1, with 1 being 100% probability that the roads are all-season. For example, a road in a flat area with low rainfall and high investment in infrastructure maintenance would have an accessibility factor of 1.0, as this road is designed to be accessible all year round and the environmental effects on its impassability are minimal. The lower and upper thresholds for the each one of the factors ranges are close but never reach 0 and 1, ensuring that when multiplied, the final passability gets incrementally closer to 0 in the lower end and 1 in the higher end.

The multiplication of the climate and terrain factors (each ranging from 0.25 to 0.95) generates the first iteration of the passability criteria, which ranges from 6% to 90% (0.0625 – 0.9025). This first iteration does not take road maintenance into account.

The GDP per capita data is then normalized in such a way that a road maxed out in terms of precipitation and slope (accessibility score of 0.0625) in a country at the top of the GDP per capita range is brought to the higher end of the accessibility score (1), while the accessibility score of a road meeting the same passability conditions in a country where GDP per capita is towards the lower end is further lowered. A mathematical threshold is applied in order to ensure that values higher than 1 are replaced by the final range's maximum (1, or 100%).

Table 3 Showcase of multiplication results for the first iteration of passability criteria

		Terrain: Minimum (0.25)	Terrain: Median (0.85)	Terrain: Maximum (0.95)
Climate	Minimum (0.25)	**0.06**	0.21	0.23
Climate	Median (0.9)	0.22	**0.76**	0.85
Climate	Maximum (0.95)	0.23	0.85	**0.95**

The multiplication of the three factors take place in a GIS environment, through raster algebra,

Table 2 Passability criteria and range transformations

Data source	Original range	Normalized range
Yearly accumulated precipitation (Copernicus)	0 – 10 m	0.25 – 0.95 (reversed)
Slope (SRTM Digital Terrain)	0 – 90 degrees	0.25 – 0.95 (reversed)
GDP per capita	221 – 234,315 USD	0.6 - 17

with the smaller pixel size being the final resolution. The final index ranges from virtually 0 to 1.

Table 4 Showcase of multiplication results for the second iteration of passability criteria

		Terrain x Climate		
		Minimum (0.06)	Median (0.76)	Maximum (0.95)
GDP per capita	Minimum (0.6)	**0.03**	0.45	0.57
	Median (0.93)	0.05	**0.7**	0.88
	Maximum (17)	1	1	**1**

Data processing

The data processing begins with filtering out all the pixels overlapping areas classified as urban from the population layer. The result is a rural **population raster layer** at 100m pixel resolution.

The two subsets of roads (all-season and exposed) have a 2km buffer applied to them. As this operation is quite resource-intensive at a global scale, the roads are rasterized, a Euclidean distance calculation is performed, and all pixels with values higher than 2km are filtered out.

The layers for precipitation, slope and road maintenance are rescaled and realigned to match the pixel grid in the rural population layer, allowing for raster algebra operations that do not require resampling. The three layers are multiplied by one another (limiting the upper threshold to 1), creating the passability index layer.

Pixels from the rural population layer falling over exposed road buffers have their values multiplied by the passability index. The resulting probability corrected population layer is combined with the population falling over all-season road buffers through raster algebra by making use of a maximum rule. This ensures that whenever the same population pixel is intersected by buffers of the two road subsets, the largest value (not corrected by the passability index) is kept. The resulting layer represents the rural population with access to an all-season road.

The total rural population and the rural population with access to an all-season road raster layers are each used as input for zonal statistics operations to determine the total sum by country. The population with access is divided by the total rural population in order to obtain the proportion, which is the final Rural Access Index (RAI).

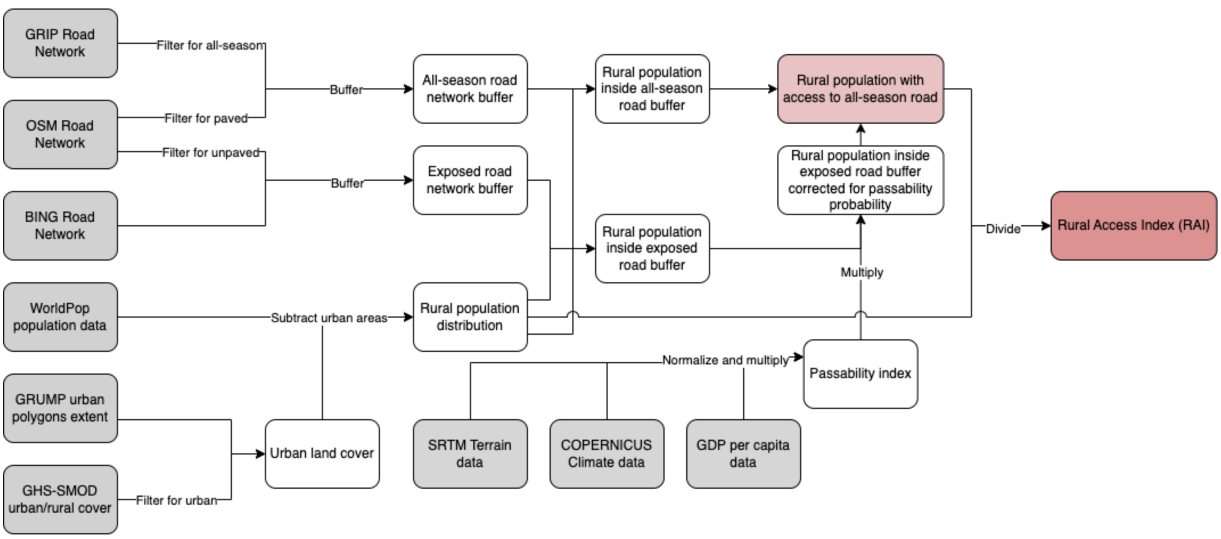

Figure 2 Overview of data processing

Data validation

Several checks were performed in order to assess the validity of the data produced.

Construct validity is assessed by calculating the correlation coefficient with other previous attempts at calculating RAI. The two other pre-existing datasets covering RAI at national scale globally are distributed by NASA SEDAC's CIESIN (CIESIN, 2022) and by Azavea (Azavea, 2019). Both implemented simpler methodologies, either by using exclusively the GRIP database and removing roads not classified as all-season or by removing roads not tagged as very high hierarchy level from OpenStreetMap. Table 5 presents the Pearson coefficient found for each of the datasets. Though the coefficients are high (>80), it's adequate that they aren't extremely high (>95), showing that implementing the present method does yield different results, arguably better ones.

Table 5 Construct validity results

	CIESIN (NASA SEDAC methodology)	Azavea (simplified ReCAP methodology)
Pearson coefficient	0.88	0.81

Convergent validity is assessed through correlation coefficients regarding variables that are expected to correlate with rural accessibility. Here GDP per capita and the Human Development Index (HDI) for the same reference year (2022) are used. Table 6 presents the Pearson coefficients found for each variable. It's telling that RAI has a high correlation (0.76) with HDI, as both can be used as evidence to shield or validate claims about the state of social justice or injustice.

Table 6 Convergent validity results

	GDP per capita	Human Development Index
Pearson coefficient	0.43	0.76

Known limitations

Scale considerations

Some very small countries, such as Small Island Developing States (SIDS), are excluded from the final result if considered to be entirely urban by the land cover layers used (GRUMP and GHS-SMOD). The remaining small island states with rural populations will tend to achieve very high scores, as the road infrastructure distribution will tend to be much more homogeneous where the rural-urban divide is less clear.

Mobility infrastructure not included

While access to all-season roads offers a fair representation of a population's overall accessibility and mobility, it might provide and under-assessment in places where transportation by other means, such as motorcycle trails and navigable waterways, are relevant. Communities living in the Amazon rainforest, for example, are highly dependent on fluvial transportation, which represented as much as 13% of the total modal share in Brazil as a whole in 2012 according to the Brazilian Agency for National Aquatic Transportation (ANTAQ). To respond to this limitation, the 2019 TRL methodology (Workman and McPherson, 2019) recommended that a secondary, supplementary indicator be developed to allow countries to take into account local infrastructure that might not be included in the standard RAI measurement.

Ground-truthing and construct validity

No ground-truth is assessed at any point in this implementation. The SDG Transformation Center is interested in designing and executing a new project specifically to this end, with the final objective of refining passability factors and the overall methodology. The project would assess road conditions through remote or on-site methods such as visiting and interviewing communities to ascertain how long roads might be closed due to climate or terrain issues. The results of the ground truthing would then be compared to the desktop assessment, and used to refine the accessibility factors as necessary, enhancing the indicator's robustness.

GDP as a proxy for road maintenance

While data on infrastructure maintenance related to preserving the existing transport network exists, it's collected by ITF only for OECD countries. This proxy is to be revisited in the future should any better options become available.

References

Azavea. (2019). **Rural Access Index Measurement Tool.** Retrieved May 9, 2023, from https://rai.azavea.com/

Center for International Earth Science Information Network (CIESIN). (2022). **Rural Access Index**. Palisades: NASA Socioeconomic Data and Applications Center (SEDAC). Retrieved from https://services2.arcgis.com/IsDCghZ73NgoYoz5/arcgis/rest/services/CIESIN_SDG_911_v1/FeatureServer

Center for International Earth Science Information Network Columbia University; CUNY Institute for Demographic Research, International Food Policy Research Institute, The World Bank, and Centro Internacional de Agricultura Tropical. (2018). **Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extent Polygons**. Palisades: NASA Socioeconomic Data and Applications Center (SEDAC).

European Commission, and Statistical Office of the European Union. (2021). **Applying the Degree of Urbanisation — A methodological manual to define cities, towns and rural areas for international comparisons** — 2021 edition. Luxembourg: Publications Office of the European Union. https://doi.org/10.2785/706535

GHS-SMOD R2023A - **GHS settlement layers, application of the Degree of Urbanisation methodology (stage I) to GHS-POP R2023A and GHS-BUILT-S R2023A, multitemporal (1975-2030)**. (2023). European Commission, Joint Research Centre (JRC). PID: http://data.europa.eu/89h/a0df7a6f-49de-46ea-9bde-563437a6e2ba, doi:10.2905/A0DF7A6F-49DE-46EA-9BDE-563437A6E2BA

Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). **Hole-filled SRTM for the globe Version 4**, available from the CGIAR-CSI SRTM 90m Database. https://srtm.csi.cgiar.org/

Meijer, J. R., Huijbregts, M. A. J., Schotten, C. G. J., & Schipper, A. M. (2018). **Global patterns of current and future road infrastructure**. Environmental Research Letters, 13(6), 064006. https://doi.org/10.1088/1748-9326/aabf45. Data is available at www.globio.info

Microsoft Bing. (2023). **Road detections from Microsoft Maps aerial imagery** [Data set]. GitHub repository. https://github.com/microsoft/RoadDetections

OpenStreetMap contributors. (2023). **OpenStreetMap database**. OpenStreetMap Foundation: Cambridge, UK. © OpenStreetMap contributors. Available under the Open Database Licence from: openstreetmap.org.

Roberts, P., K.C., Shyam, & Rastogi, C. (2006). **Rural Access Index: A Key Development Indicator**. Transport Paper TP-10. World Bank, Washington DC.

Schiavina, M., Melchiorri, M., & Pesaresi, M. (2023). **GHS-SMOD R2023A - GHS settlement layers, application of the Degree of Urbanisation methodology (stage I) to GHS-POP R2023A and GHS-BUILT-S R2023A, multitemporal (1975-2030).** European Commission, Joint Research Centre (JRC). PID: http://data.europa