Mapping Rural Accessibility Gaps across Latin America and the Caribbean

Florencia Buccari[1], Leandro Martinez Villa[2], and Tatiana Rosa[3]

[1]CAF – Development Bank of Latin America and the Caribbean

mbuccari@caf.com

[2]Universidad Nacional de La Plata (UNLP)

leandromartinezvilla@gmail.com

[3]CAF – Development Bank of Latin America and the Caribbean

trosa@caf.com

This version: June 19, 2026

Abstract

We present the first open, harmonized Rural Access Index (RAI) for 33 countries in Latin America and the Caribbean, constructed at the subnational-government level, with coverage for 14,840 administrative units. We compute two variants: the standard one, restricted to paved roads, and an alternative that incorporates all roads of tertiary hierarchy or higher, leveraging the deep-learning surface imputation provided by HeiGIT over OpenStreetMap. We document a central finding: the standard measure exhibits a floor effect in the low-accessibility range, losing power to discriminate among municipalities precisely in the segment most relevant for public policy. For Brazilian municipalities with a standard RAI below 40%, agricultural land values show no correlation with the paved-only measure but do correlate significantly with the extended one. The resulting subnational dataset is publicly available.

Keywords: Rural Access Index, rural accessibility, subnational measurement, Latin America and the Caribbean, road infrastructure, geospatial data

---

Small sections of text that are less than two paragraphs may be quoted without explicit permission as long as this document is acknowledged. Findings, interpretations and conclusions expressed in this publication are the sole responsibility of its author(s) and cannot be, in any way, attributed to CAF, its

Executive Directors or the countries they represent. CAF does not guarantee the accuracy of the data included in this publication and is not, in any way, responsible for any consequences resulting from its use.

©2024 Corporación Andina de Fomento

1 Introduction

Rural connectivity is a central driver of economic development. By lowering transport costs, road infrastructure expands market access and fosters the spatial integration of regional economies [Donaldson, 2018, Redding and Turner, 2015], raises agricultural productivity through the adoption of modern inputs and stronger linkages with output markets [Aggarwal, 2018, Sotelo, 2020], and reshapes the structure of local labor markets by enabling rural workers to commute to non-agricultural employment [Asher and Novosad, 2020, Fiorini and Sanfilippo, 2022]. The welfare gains are sizable and concentrated among the rural poor, with documented increases in household consumption, property values, and access to credit following targeted road-paving interventions [González-Navarro and Quintana-Domeque, 2016, Dercon et al., 2009]. In recognition of this evidence, the international community adopted in 2016 a single global metric of rural accessibility as the Indicator 9.1.1 of the Sustainable Development Goal: the Rural Access Index (RAI), defined as the share of the rural population living within two kilometers of an all-season road [World Bank, 2016].

The standard implementation of the RAI proxies the all-season requirement by restricting the calculation to paved roads, on the grounds that paved segments are far more resistant to weather-related disruptions than unpaved ones [World Bank, 2016]. This approximation is reasonable in regions where the unpaved network is residual, but it may produce a distorted picture of accessibility in Latin America and the Caribbean (henceforth LAC), where a large share of the network connecting rural producers to markets is unpaved and yet functionally trafficable for most of the year. The reliance on road-surface information introduces an additional source of bias: in the most widely used global road inventory Humanitarian OpenStreetMap Team [2025], 47% of the road network in South America carry no surface attribute [Randhawa et al., 2024]. Excluding these segments, or implicitly classifying them as unpaved, induces large measurement error in the input to the standard RAI and risks suppressing economically meaningful variation in precisely the segments where it is most relevant for policy.

In this paper, by combining several georeferenced datasets, we construct a harmonized subnational measure of the RAI for 33 countries in LAC, covering 14,840 administrative units. We compute two variants of the indicator: the standard one, restricted to paved roads, and an extended variant that considers all roads of tertiary hierarchy or higher, regardless of surface. Using data from Brazil, we validate both measures against agricultural land values, the outcome most directly expected to capitalize differences in rural accessibility. To the best of our knowledge, this is the first open, harmonized RAI dataset constructed at the subnational government level for Latin America and the Caribbean.

Our results proceed in three steps. First, we find that on average 66% of the rural population in LAC lives 2 kilometers or closer to any type of road, but only 40% from

a paved road. Second, there is high variation in RAI both between and within country. The subnational disaggregation of the measure allows us to show that heterogeneity in connectivity in LAC is not only between country; while some regions have connectivity measures close to OECD countries (Porto Amazonas municipality in Brazil or Mariscal Caceres in Peru) others, as San Pedro in Bolivia or Lloró in Colombia show a RAI lower than 10%, highlighting the need for investment to close these connectivity gaps.

Third, we validate our measure by regressing agricultural land value in Brazil against each of our two RAI measures, building on the causal evidence that links rural connectivity to land values [Donaldson, 2018, González-Navarro and Quintana-Domeque, 2016]. Controlling for state fixed effects, altitude and the degree of urbanization of each municipality, we find a positive correlation between both RAI measures and the outcome. Moreover, to explore potential heterogeneous effects, we discretize the continuous RAI into ten dummy variables, one per decile of its distribution. Results show that for municipalities in the low-accessibility range (i.e. up to values around 40% of the RAI), agricultural land values are uncorrelated with the standard measure but correlate strongly with the extended one, with the two converging only above RAI $\approx$ 50–60 percent. In municipalities with a low share of paved roads, unpaved roads continue to provide effective connectivity and are capitalized into land values, but this variation is invisible to the standard measure because most such municipalities cluster around very low paved-RAI values with little informative dispersion; the extended measure recovers it.

We further show that this divergence tracks the fiscal capacity of local governments: the gap between the all-roads and paved-only semi-elasticities of land value is widest among municipalities in the lowest quintile of own fiscal revenue per capita and narrows monotonically until it vanishes in the highest quintile. Where local governments lack the resources to pave and maintain roads, the unpaved network provides the effective connectivity that is capitalized into land values but that the standard measure fails to capture.

The implication for policy is direct: rankings, investment targeting, and empirical work that rely exclusively on the standard RAI incur systematic mismeasurement in precisely the range where rural-roads policy should be most active.

This paper contributes to the literature on the measurement of rural accessibility, originating in Roberts et al. [2006]'s proposal of the RAI and its subsequent operationalization by World Bank [2016]. We provide the first implementation of the RAI at the subnational-government level for the entire LAC region, and the first to report it jointly under an extended all-roads variant and the standard paved-only definition — a comparison made feasible at the regional scale by the deep-learning surface imputation of HeiGIT [HeiGIT, 2024]. Beyond extending geographic coverage, the paper offers a methodological diagnostic: we document a floor effect by which the standard variant loses discriminating power in the low-accessibility range, precisely the segment in which the indicator is most

policy-relevant.

Finally, to facilitate the use of our results, we release the full subnational dataset and the computation scripts as a public resource.

2 The Rural Access Index: Measurement and Validation Strategy

2.1 Definition

The Rural Access Index was originally proposed by Roberts et al. [2006] to capture the proximity of the rural population to quality transport infrastructure. In 2016, it was adopted as Indicator 9.1.1 of the Sustainable Development Goals, monitoring access to safe, affordable, and sustainable transport systems [World Bank, 2016]. The indicator is defined as

$$\text{RAI} = \frac{\text{rural population living within } \leq 2 \text{ km of a road}}{\text{total rural population}} \times 100. \tag{1}$$

The two-kilometer threshold reflects the maximum walking distance to the nearest road [Roberts et al., 2006]. The original methodology additionally restricts the set of admissible roads to those that provide year-round trafficability, defined as roads passable throughout the year by the dominant means of rural transport, allowing for brief and predictable interruptions during adverse weather, particularly on low-volume tracks [World Bank, 2016].

2.2 The two variants and why they matter in LAC

The official RAI is defined over *all-season* roads. In practice, the year-round-trafficability requirement is most commonly proxied by restricting the calculation to paved roads, on the grounds that pavement is more resistant to weather-related disruption than unpaved surfaces [World Bank, 2016]. We report two measures in parallel. The *paved-only benchmark* replicates this implementation: it considers only paved roads of tertiary hierarchy or higher. The *all-roads variant* considers all roads of tertiary hierarchy or higher, regardless of surface type. We treat the paved-only benchmark as the operational standard, and the all-roads variant as a diagnostic complement.

The paved-only benchmark raises two concerns in LAC. *Operationally*, a large share of the regional road network lacks a surface attribute in public road inventories: in OpenStreetMap 47% of the roads' kilometers do not contain the surface attribute [Randhawa et al., 2024]. The HeiGIT database [HeiGIT, 2024] addresses this gap by imputing surface type from OpenStreetMap using deep learning algorithms based on Mapillary images,

slightly lowering the share of unclassified segments to 45%. This imputation is precisely what makes a comparable paved-only measure feasible at the subnational level in the region; the all-roads variant, which does not condition on surface information, is unaffected by missing surface attributes. *Conceptually*, in low-paving environments a non-negligible fraction of unpaved roads is functionally trafficable for most of the year and provides effective connectivity. Restricting the indicator to paved segments therefore not only mechanically lowers the level but may also suppress within-country variation with economic content. The all-roads variant is best read as a measurement diagnostic: it reveals how much of that variation the paved-only benchmark discards, and where.

2.3 Validation of the connectivity measures

To examine the informational content of both measures of connectivity computed in this paper, and to quantify how much variation the paved-only benchmark discards in LAC, we correlate each measure of the RAI with a set of economic outcomes that the literature has linked to rural accessibility.

The empirical literature has documented—in many cases causally—a robust association between rural road accessibility and a set of economic outcomes. We use those documented associations as validation exercises for the two measures of the RAI: a measure that performs as the literature predicts is a measure that carries economic content. We focus on the outcome for which the literature is strongest: agricultural land values, which should capitalise improvements in accessibility [see Donaldson, 2018, González-Navarro and Quintana-Domeque, 2016]. We are not estimating causal effects of road infrastructure on this outcome; we are asking whether each measure of the RAI is informative about the variation in outcomes that the literature has tied to accessibility, and *where* along the connectivity distribution that informativeness lies.

For each outcome we use cross-sectional specifications with fixed effects at the lowest administrative level available (municipality), controls for elevation and degree of urbanization, and report separate results for the two variants of the RAI. At the municipal level m we estimate

$$Y_m = \alpha + \beta\,\mathrm{RAI}_m + \gamma_1\,\mathrm{Elevation}_m + \gamma_2\,\mathrm{UrbanDegree}_m + \mu_e + \varepsilon_m, \tag{2}$$

where Y_m is the outcome of interest (the log of agricultural land value), RAI_m is the index computed either using all roads or using only paved ones, $\mathrm{Elevation}_m$ is the municipality's mean elevation, $\mathrm{UrbanDegree}_m$ is its degree of urbanization according to Schiavina et al. [2023] (urban, semi-urban and rural), and μ_e is a set of state dummies that absorb unobserved heterogeneity at the administrative level immediately above the municipality.

The specification above restricts the effect of the RAI on the economic outcomes to be linear, which is excessively restrictive if, as the discussion of the two variants suggests, the

information content of connectivity varies along the range of the indicator. To measure the non-linear relationship between the RAI and our outcome, we discretize the connectivity variable into deciles and construct ten dummy variables, one for each decile of the RAI distribution. This specification allows us to trace the association between accessibility and the outcome flexibly along the entire distribution of the indicator, without imposing a functional form, and is particularly informative for detecting ranges in which each measure of the RAI loses discriminating power. Formally, we replace the continuous variable RAI_m in Equation 2 with the set of indicators $\{\mathbf{1}[\text{RAI}_m \in D_k]\}_{k=1}^{10}$, where D_k denotes the kth decile of the RAI distribution and $\mathbf{1}[\cdot]$ is the indicator function, omit the first decile as the reference category, and estimate

$$Y_m = \alpha + \sum_{k=2}^{10} \beta_k \, \mathbf{1}[\text{RAI}_m \in D_k] + \gamma_1 \, \text{Elevation}_m + \gamma_2 \, \text{UrbanDegree}_m + \mu_e + \varepsilon_m. \qquad (3)$$

The coefficients β_k recover the average difference in the outcome between municipalities whose RAI falls in decile D_k and those in the reference decile, conditional on the controls. The profile of $\{\beta_k\}$ across deciles characterizes the shape of the relationship between connectivity and the economic outcome.

3 Data and computation

3.1 RAI

The computation of the RAI requires three georeferenced inputs—population distribution, urban settlements, and the road network—together with a layer of administrative boundaries that defines the units of analysis. Population data come from WorldPop [2020], a global gridded population product with a spatial resolution of 100 by 100 meters for the year 2020; the high resolution enables accurate estimates both of the rural population and of the rural population residing within two kilometers of a road. Urban settlements are taken from the Global Rural-Urban Mapping Project (GRUMP) [CIESIN-Columbia University et al., 2011], consistent with the methodology of World Bank [2016]. The road network is drawn from HeiGIT [HeiGIT, 2024], which applies deep-learning techniques to OpenStreetMap [OpenStreetMap Wiki, 2025] to classify the surface condition of road segments lacking that attribute in the original source. We retain segments of tertiary hierarchy or higher (*motorway, trunk, primary, secondary, tertiary*). For Mexico, we use the data from Humanitarian OpenStreetMap Team [2025], since HeiGIT was not available. Our results for Peru were calibrated against the official figures reported in

World Bank [2016], which confirmed the adequacy of the classification.[1] National boundaries are taken from the World Bank's official administrative boundaries [World Bank, 2025], with coverage for all LAC countries considered. Subnational boundaries come from CAF's Atlas of Subnational Governments [CAF, 2025], which covers 29 countries in the region; for the four countries lacking subnational disaggregation in the Atlas (Barbados, Dominica, Grenada, and Saint Vincent and the Grenadines) computation is performed only at the national level.

The calculation proceeds in two steps. First, the total rural population is calculated as the difference between the total population of the administrative unit and the population residing within GRUMP urban settlements. Second, the rural population within the two-kilometer threshold is calculated by constructing a buffer of that distance around the filtered road network, overlaying the buffer with the population layer, and subtracting the urban population that falls within the same buffer. The technical detail is presented in Online Appendix A.1.

Online Appendix Table A.1 summarizes the geographic coverage and the main input statistics for the 33 countries included. CAF's Atlas of Subnational Governments identifies 15,198 subnational units. We compute the index for 14,844 areas in total, of which 4 are national units (Barbados, Dominica, Grenada, and Saint Vincent and the Grenadines) and 14,840 are subnational, covering 97.6% of the units identified.[2] Measurements were taken, whenever feasible, at the smallest available administrative unit: level one in Argentina, Belize, and Guyana; level three in El Salvador, Haiti, Peru, and the Dominican Republic; and level two in the remaining countries. All of these constitute governments that design and target policies related to rural connectivity.

Online Appendix Figure A.1 illustrates the four georeferenced layers used in the computation procedure, taking Bolivia as an example.

3.2 Validation Exercises

The validation exercises draw on data for Brazil, selected for the availability of high-quality georeferenced microdata that can be matched to the subnational RAI. In each case we merge the municipal records to both RAI measures using the official administrative codes, and restrict attention to the units for which the two variants are available.

Our outcome is the value of land without improvements at the municipal level for Brazil in 2022, which is the variable most directly expected to capitalize differences in accessibility. This variable is published by the Ministry of Agricultural Development

[1] The detailed comparison with the alternative GRIP database [Meijer et al., 2018] is available upon request. The HeiGIT database was finally selected for its correspondence with IRF data [International Road Federation, 2025].

[2] Of the 358 units not included, 352 either have no territory outside GRUMP urban agglomerations or have unpopulated territory; the remaining 4 are uninhabited islands: Redonda in Antigua and Barbuda, Isla de Flores and Isla de las Gaviotas in Uruguay, and Anzoátegui in Venezuela.

in Brazil [INCRA, 2023]. We use its natural logarithm as the dependent variable. We also use data on municipalities' fiscal revenue, from Alves et al. [2025] and Secretaria do Tesouro Nacional [2023], to group subnational units into quintiles.

Across all specifications, we include the geographic controls to absorb confounding factors related with terrain and urbanization: the standard deviation of elevation within the unit [Earth Resources Observation and Science (EROS) Center, 2010] and the degree-of-urbanisation from Schiavina et al. [2023].

Table 1 reports summary statistics for the land-value validation exercise, covering the Brazilian municipalities used and their main characteristics.

Table 1: Summary statistics - Validation exercise

	N	Mean	SD	Median	Min	Max
Land value (USD/ha)	5410	6124	6272	3940	80	38387
RAI - all roads (%)	5410	78	19	82	0	100
RAI - paved roads (%)	5410	59	27	63	0	100
Altitude (sd)	5410	76	59	60	1	608
Revenue (USD/population)	5378	171	174	130	2	4918
Degree of urbanisation (% of municipalities)						
Rural areas	5410	56				
Towns and semi-dense areas	5410	37				
Cities	5410	7				

4 Results

This section presents the subnational RAI values for the 33 countries covered, maps the results at the maximum available level of disaggregation, and documents within-country dispersion, which is the new informational content that subnational disaggregation makes it possible to recover.

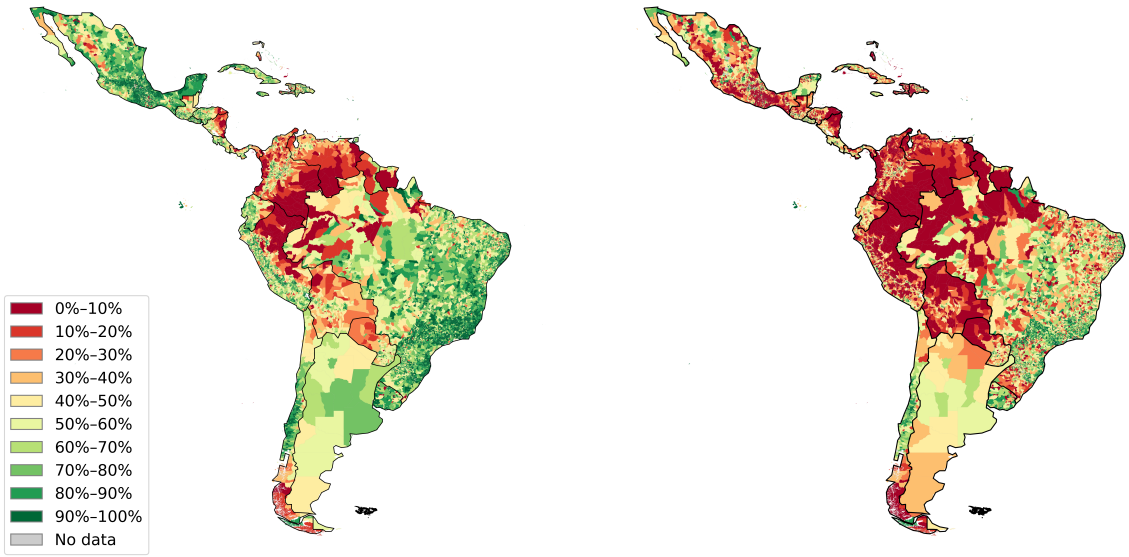
4.1 RAI

Appendix Table A.2 reports the subnational mean of the RAI by country, for both variants, together with the corresponding standard deviations and population-weighted averages. Three stylized facts emerge. *First*, the gap between the extended and the paved variants is markedly larger in countries with less-paved networks (Bolivia, Paraguay,

Suriname, Guyana, and Haiti). *Second*, within-country dispersion is high: in large and geographically diverse countries, the subnational variance of the RAI can be as large as the cross-country variance of national averages. *Third*, subregional averages display a pattern consistent with infrastructure development, with higher accessibility in the more urbanized Southern Cone and in the smaller Caribbean islands, and lower accessibility in Amazonian countries and in the lower-income continental Caribbean.

Figure 1 maps both variants at the maximum available level of disaggregation. The visual contrast between the two panels reinforces the point developed in Section 4.4: the choice of variant substantially reshuffles the geography of accessibility in LAC, especially in countries with little paved network.

Figure 1: Subnational Rural Access Index in LAC: all roads vs. paved roads only



(a) All roads

(b) Paved roads only

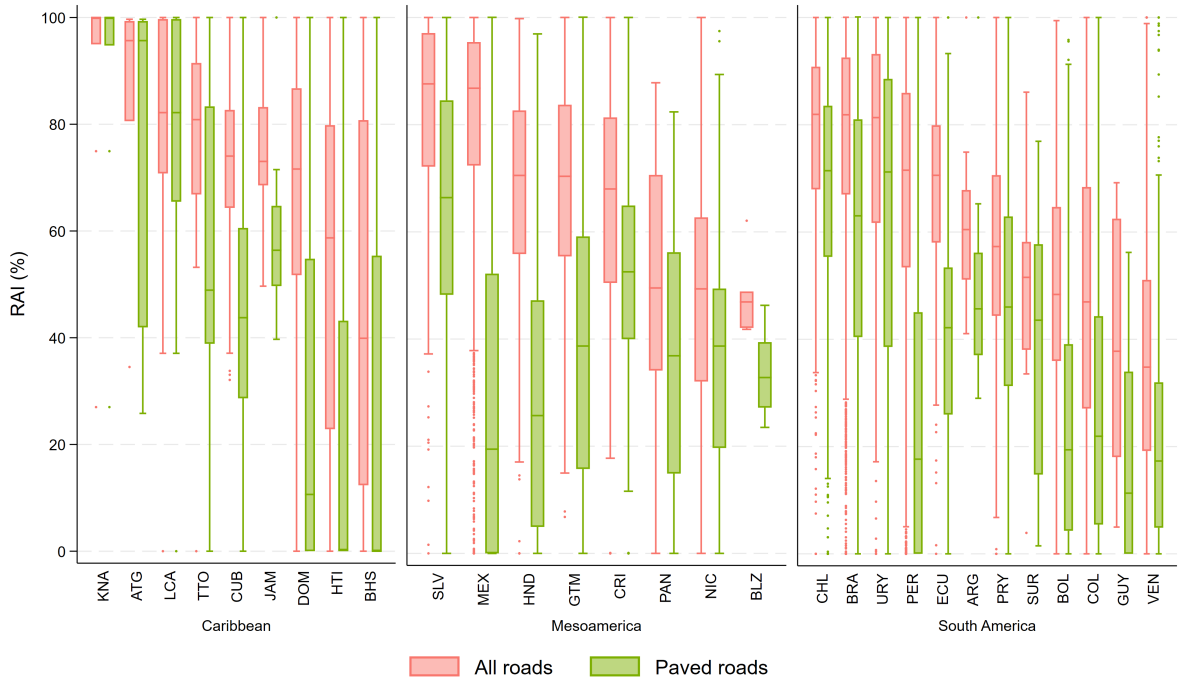
Notes: The figure shows the Rural Access Index measured at the maximum available level of subnational disaggregation. Panel (a) considers all roads of tertiary hierarchy or higher regardless of surface type or trafficability; panel (b) restricts the calculation to paved roads of tertiary hierarchy or higher.
Source: Authors' computations based on World Bank [2025], CAF [2025], CIESIN-Columbia University et al. [2011], WorldPop [2020], HeiGIT [2024], and Humanitarian OpenStreetMap Team [2025].

4.2 Within-country dispersion: the value of subnational disaggregation

Figure 2 reports country-level box plots for both variants of the indicator. In most countries the subnational interquartile range is wide, and for several cases the width of the extended-variant box exceeds the cross-country dispersion of national averages. This shows that the national RAI figures fail to capture the geography that is relevant for public policy: within a given country, administrative units differ in accessibility by

amounts comparable to entire countries at different stages of infrastructure development.

Figure 2: Subnational distribution of the RAI by country, both variants



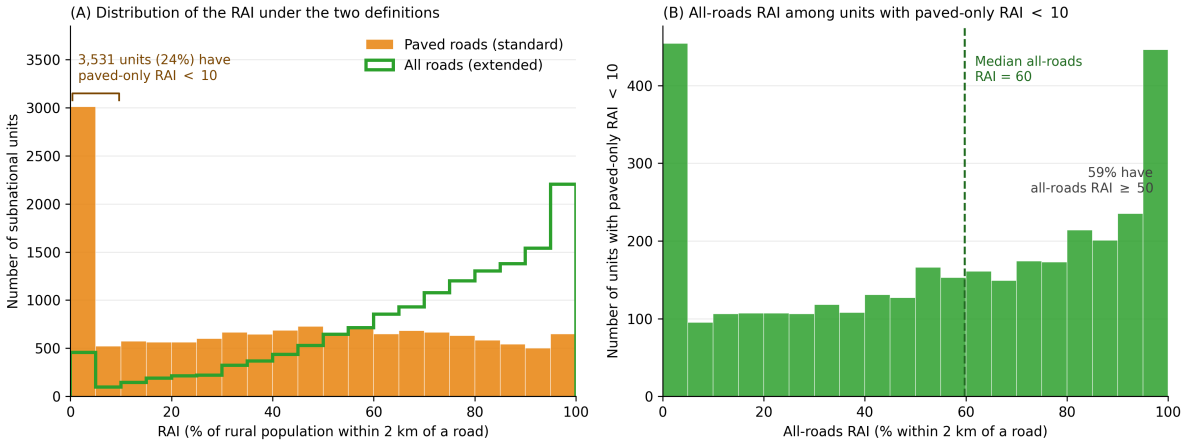
Notes: The figure shows, for each country, the distribution of the subnational RAI under the two variants considered (all roads and paved roads only). Each box summarizes the lower quartile, the median, and the upper quartile; whiskers extend up to 1.5 times the interquartile range. Countries are grouped by subregion (South America, Caribbean, Mesoamerica) and, within each block, sorted in descending order by the median of the all-roads measure. Barbados, Dominica, Grenada, and Saint Vincent and the Grenadines are excluded as they lack subnational disaggregation.

4.3 Comparing the two RAI measures: the floor effect of the paved-only RAI

We compare our two RAI measures by plotting their full subnational distributions, rather than their aggregate means, which mask where the two definitions diverge. Figure 3 reports this comparison. Panel (A) shows the marginal distribution of each variant across the 14,820 subnational units for which both are available: the paved-only RAI concentrates a large mass against the lower bound—3,531 units (24%) record a value below 10—whereas the all-roads RAI is shifted toward the upper range. This compression is the *floor effect*: the standard measure assigns near-identical, near-zero scores to almost a quarter of the units, losing the ability to discriminate among them. Panel (A) alone, however, only establishes that the two marginal distributions differ. The decisive test is conditional, and is reported in Panel (B): restricting attention to the 3,531 units that the paved-only measure places below 10—units it treats as essentially equivalent—we plot the all-roads RAI of those same units. Far from clustering near zero, it is dispersed across the

entire support, with a median of 60 and 59% of units above 50. Even among the 3,531 subnational units with paved-only RAI below 10, the all-roads RAI spans a wide range, revealing substantial variation in effective rural accessibility that the standard measure collapses near zero.

Figure 3: The floor effect of the paved-only RAI



Notes: $N = 14{,}820$ subnational units with both variants available. **Panel (A)** shows the distribution of the subnational RAI under the two definitions: the paved-only variant (orange) piles up near zero—3,531 units (24%) have paved-only RAI below 10—while the all-roads variant (green outline) shifts toward the upper range. **Panel (B)** conditions on those 3,531 units and plots their all-roads RAI: even among units that look indistinguishable under the standard measure, the all-roads RAI spans the full range, with a median of 60 and 59% above 50. The standard measure collapses near zero variation in effective accessibility that the extended measure recovers.

4.4 Validation exercise: RAI and Land Value in Brazil

Table 2 reports the estimates of Equation 2 for Brazil. Consistent with the evidence on paved-road interventions [González-Navarro and Quintana-Domeque, 2016, Donaldson, 2018], rural accessibility is capitalized into agricultural land values: both variants of the RAI enter with positive and highly significant semi-elasticities of comparable magnitude—0.0059 for the all-roads variant and 0.0036 for the paved-only variant. At the linear mean, both measures account for variation in land prices and the choice of variant appears innocuous.

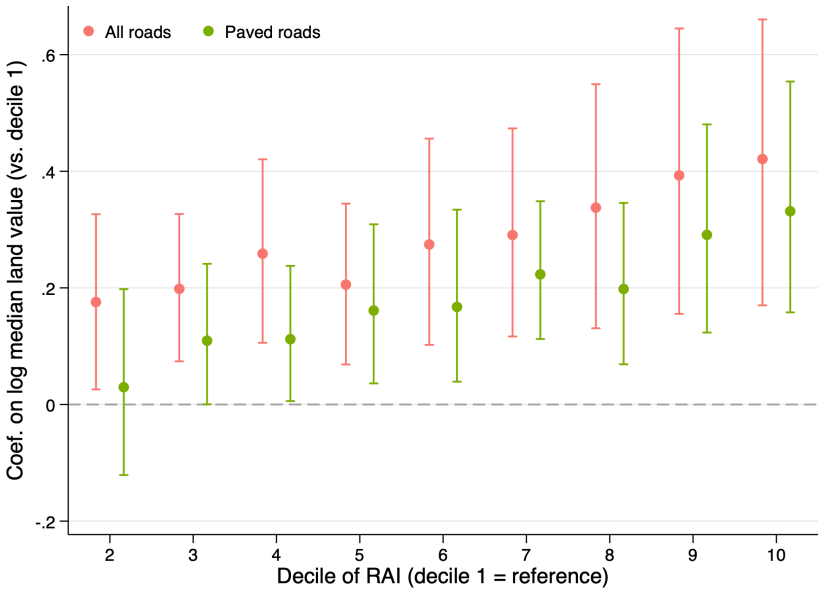
Table 2: RAI and Agricultural Land Values in Brazil

	(1)	(2)	(3)
RAI – paved roads	0.00363***		0.00179***
	(0.00074)		(0.00062)
RAI – all roads		0.00583***	0.00416**
		(0.00146)	(0.00169)
Constant	8.69799***	8.56463***	8.56722***
	(0.05836)	(0.09292)	(0.09411)
State fixed effects	Yes	Yes	Yes
Altitude control	Yes	Yes	Yes
DEGURBA control	Yes	Yes	Yes
Observations	5,410	5,410	5,410
R-squared	0.797	0.798	0.798

Notes: Dependent variable: log of unimproved agricultural land value at the municipal level in Brazil. Column (1) reports the estimate for the paved-only RAI; column (2) for the all-roads RAI; column (3) includes both measures jointly. All columns control for the standard deviation of elevation, the degree of urbanisation (DEGURBA) and state fixed effects. In column (3), when the two measures compete to explain land value, the all-roads coefficient is the larger of the two, so the extended measure accounts for more of the variation in land value; it is, however, estimated less precisely (noisier), as expected for a measure that also incorporates unpaved segments whose trafficability is imputed. Standard errors clustered at the state level in parentheses; statistical significance confirmed by wild cluster bootstrap (9,999 replications) given the small number of clusters (24 states). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

This apparent equivalence breaks down once we relax linearity. Letting each variant enter through a set of dummy variables—one per decile of the RAI distribution, with the first decile as the omitted reference—uncovers a sharp asymmetry between the two measures in the low-accessibility range. Figure 4 and Appendix Table A.3 report the coefficients. The all-roads variant is positive and rises monotonically across the distribution, whereas the paved-only variant is statistically indistinguishable from zero for municipalities in the bottom deciles of the RAI distribution and turns significant only higher up, with point estimates that remain systematically below those of the all-roads variant. The variation that the linear specification averages over is precisely what disappears when the indicator is restricted to paved roads—in the segment of the distribution where rural-roads policy is most active.

Figure 4: RAI and Land Value in Brazil, Non-Linear Specification



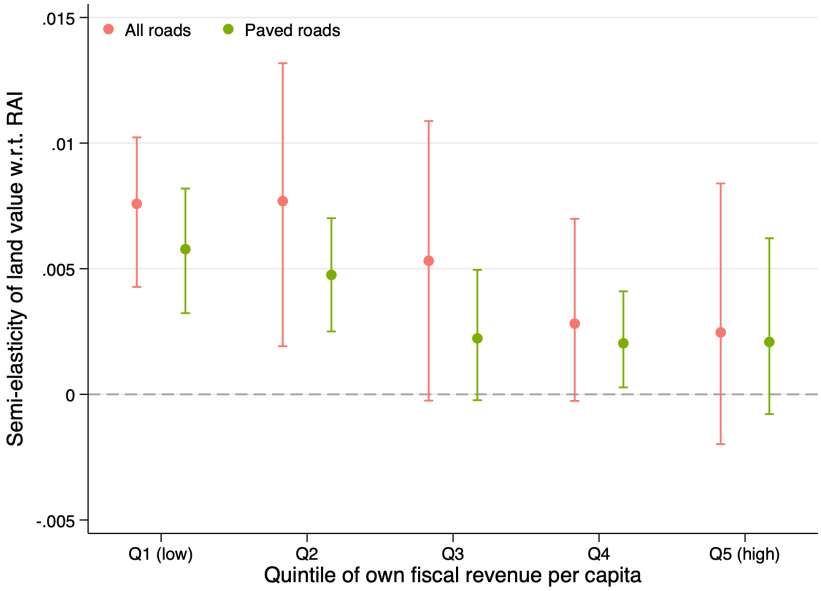
Notes: The figure reports the non-linear relationship between the RAI and agricultural land value in Brazil, obtained by relaxing the linear specification of Equation 3 and letting the indicator enter as a set of dummy variables, one per decile of the RAI distribution. The first decile is the omitted reference category, so each marker reports the average difference in the outcome between municipalities in the indicated decile and those in the first decile. The dependent variable is the log of unimproved agricultural land value at the municipal level. All regressions include state fixed effects, a control for the standard deviation of elevation, and fixed effects for the first-level DEGURBA degree-of-urbanization category. Green markers correspond to the all-roads RAI and orange markers to the paved-only RAI. Vertical lines are 95% confidence intervals; standard errors are clustered at the state level, and the confidence intervals are obtained by wild cluster bootstrap given the small number of clusters (24 states).

Moreover, we re-estimate Equation 2 allowing the semi-elasticity of land value to the RAI to differ across municipalities grouped by their own fiscal revenue. Figure 5 and Appendix Table A.4 report the marginal effect of each variant by quintile of municipal fiscal revenue per capita. The pattern mirrors the one found across the RAI distribution: the gap between the all-roads and paved-only semi-elasticities is widest among low-revenue municipalities and narrows monotonically as fiscal capacity rises, until the two become statistically indistinguishable in the top quintile. The two variants convey similar information among municipalities with higher fiscal revenue per capita. In lower-revenue municipalities, by contrast, the association between land values and the paved-only RAI weakens relative to the all-roads measure, suggesting that the standard indicator misses part of the accessibility variation captured by the broader road network.

The economic reading is direct. In municipalities with low paving or low income, unpaved roads continue to provide effective connectivity and are capitalized into land values, but this variation is invisible to the standard measure because most such municipalities cluster around very low paved-RAI values with little informative dispersion. The

all-roads measure, by incorporating functionally trafficable unpaved roads, recovers that variation. In measurement terms, the paved-only RAI carries *non-classical* error: the gap between the paved-only measure and effective accessibility does not average out but is correlated with true accessibility, widening systematically toward the low-accessibility range. It therefore does not merely attenuate estimates toward zero, as classical measurement error would, but biases them by an amount that shifts along the distribution, deflating the estimated return to accessibility precisely where rural-roads policy is most active. Rankings, investment targeting, and empirical work that rely exclusively on the standard RAI inherit this bias in the range that matters most.

Figure 5: RAI and land value in Brazil: heterogeneity by municipal fiscal revenue



Notes: Each marker reports the semi-elasticity of agricultural land value with respect to the RAI—the average marginal effect of the RAI on the log of unimproved agricultural land value—estimated separately for each quintile of municipal own fiscal revenue per capita. Estimates are obtained from Equation 2 augmented with interactions between the continuous RAI and the quintile-of-fiscal-revenue dummies, so that each marker is the semi-elasticity recovered directly from the corresponding interaction coefficient. All regressions include state fixed effects, a control for the standard deviation of elevation, and fixed effects for the first-level DEGURBA degree-of-urbanization category. Green markers correspond to the all-roads RAI and orange markers to the paved-only RAI. Vertical lines are 95% confidence intervals; standard errors are clustered at the state level and the intervals are obtained by wild cluster bootstrap given the small number of clusters (24 states).

The evidence tells a consistent story. The literature has established—in many cases causally—that rural road connectivity is associated with higher agricultural land values [Donaldson, 2018, González-Navarro and Quintana-Domeque, 2016]. Measuring connectivity through paved roads alone, as the standard implementation of the RAI does, recovers part of that signal at the linear mean but discards a substantively large fraction of it in the segment of the distribution that matters most for public policy. In poorer

or less-connected municipalities—the natural target of rural-roads programmes—the unpaved network does the heavy lifting, and the all-roads variant is what picks it up. The evidence we report is, on this point, far from cosmetic.

5 Conclusions

This paper makes a measurement contribution to the study of rural accessibility in Latin America and the Caribbean. We construct and release the first harmonised subnational dataset of the Rural Access Index for 33 countries in the region, covering 14,840 administrative units. We report two measures in parallel: a paved-only benchmark, which replicates the standard implementation and which we make feasible at the subnational level in LAC by exploiting HeiGIT's deep-learning surface imputation over OpenStreetMap, and an all-roads variant that does not condition on surface attributes. The two measures agree at the population-weighted aggregate but diverge sharply within countries: the paved-only benchmark loses discriminatory power among low-accessibility municipalities, while the all-roads variant continues to track the variation in land values that the literature has tied to rural connectivity.

The implications are operational. *For public policy*, the targeting of rural-connectivity investment in LAC cannot rely on the paved-only RAI alone in low-accessibility subsamples; the all-roads variant is needed as a diagnostic complement to identify priority municipalities. *For empirical research*, studies that use the RAI as an explanatory or control variable in LAC should report results under both measures and explicitly check whether the variation they exploit lies in the segment where the paved-only benchmark is informative. *For statistical production*, national reports of SDG Indicator 9.1.1 in low-paving environments should be accompanied by subnational measures and by the all-roads diagnostic, in line with the practice we propose here.

References

Shilpa Aggarwal. Do rural roads create pathways out of poverty? Evidence from India. *Journal of Development Economics*, 133:375–395, 2018. doi: 10.1016/j.jdeveco.2018.01.004.

Guillermo Alves, Pablo Brassiolo, Florencia Buccari, Carolina Camacho, Rodrigo Cifuentes, Ricardo Estrada, and Gustavo Fajardo. Nearby solutions: The role of regional and local governments in latin america and the caribbean. Report on economic development, CAF—Development Bank of Latin America and the Caribbean, Caracas, 2025. URL https://scioteca.caf.com/handle/123456789/2432.

Sam Asher and Paul Novosad. Rural roads and local economic development. *American Economic Review*, 110(3):797–823, 2020. doi: 10.1257/aer.20180268.

CAF. Base de Datos de Gobiernos Locales y Regionales en América Latina y el Caribe, 2025. URL https://atlasgobiernoslocales.caf.com/es/acerca-de.

CIESIN-Columbia University, IFPRI, World Bank, and CIAT. Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extents Grid, 2011. URL https://earthdata.nasa.gov/data/catalog/sedac-ciesin-sedac-grumpv1-ext-1.00.

Stefan Dercon, Daniel O. Gilligan, John Hoddinott, and Tassew Woldehanna. The impact of agricultural extension and roads on poverty and consumption growth in Fifteen Ethiopian Villages. *American Journal of Agricultural Economics*, 91(4):1007–1021, 2009. doi: 10.1111/j.1467-8276.2009.01325.x.

Dave Donaldson. Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4–5):899–934, 2018. doi: 10.1257/aer.20101199.

Earth Resources Observation and Science (EROS) Center. Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010), 2010. URL https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-multi-resolution-terrain-elevation.

Matteo Fiorini and Marco Sanfilippo. Roads, jobs and structural change: Quasi-experimental evidence from Ethiopia. *The World Bank Economic Review*, 36(4):999–1020, 2022. doi: 10.1093/wber/lhac003.

Marco González-Navarro and Climent Quintana-Domeque. Paving streets for the poor: Experimental analysis of infrastructure effects. *Review of Economics and Statistics*, 98 (2):254–267, 2016. doi: 10.1162/REST_a_00553.

HeiGIT. Road Surface Data, November 2024. URL https://data.humdata.org/organization/heidelberg-institute-for-geoinformation-technology.

Humanitarian OpenStreetMap Team. Humanitarian OpenStreetMap Team, 2025. URL https://www.hotosm.org/en/.

INCRA. SIMET: Land Market System: Report of Land Values by Municipality. [online database], 2023. URL https://simet.incra.gov.br/#/relatorio-valores-terra-municipio. Accessed: 2026-03-30.

International Road Federation. World road statistics data warehouse. https://worldroadstatistics.org/data-warehouse/, 2025. Accessed: 2026-06-19.

Johan R Meijer, Mark A J Huijbregts, Kees C G J Schotten, and Aafke M Schipper. Global patterns of current and future road infrastructure. *Environmental Research Letters*, 13(6):064006, June 2018. doi: 10.1088/1748-9326/aabd42. URL https://iopscience.iop.org/article/10.1088/1748-9326/aabd42.

OpenStreetMap Wiki. Main page, 2025. URL https://wiki.openstreetmap.org/wiki/Main_Page.

Perry, Matthew T. rasterstats: Summary statistics of geospatial raster datasets based on vector geometries, 2015. URL https://pythonhosted.org/rasterstats/index.html.

Sukanya Randhawa, Eren Aygun, Guntaj Randhawa, Benjamin Herfort, Sven Lautenbach, and Alexander Zipf. Paved or unpaved? a deep learning derived road surface global dataset from mapillary street-view imagery, 2024. URL https://arxiv.org/abs/2410.19874.

Stephen J. Redding and Matthew A. Turner. Chapter 20 - Transportation Costs and the Spatial Organization of Economic Activity. In Gilles Duranton, J. Vernon Henderson, and William C. Strange, editors, *Handbook of Regional and Urban Economics*, volume 5 of *Handbook of Regional and Urban Economics*, pages 1339–1398. Elsevier, January 2015. doi: 10.1016/B978-0-444-59531-7.00020-X. URL https://www.sciencedirect.com/science/article/pii/B978044459531700020X.

Peter Roberts, Shyam Kc, and Cordula Rastogi. Rural access index: a key development indicator. Technical report, World Bank, 2006.

M. Schiavina, M. Melchiorri, and S. Freire. GHS-DUC R2023A - GHS Degree of Urbanisation Classification, application of the Degree of Urbanisation methodology (stage II) to GADM 4.1 layer, multitemporal (1975-2030), 2023. URL https://human-settlement.emergency.copernicus.eu/ghs_duc2023.php.

Secretaria do Tesouro Nacional. Receitas orçamentárias (anexo i-c) [database]. Siconfi—National Treasury Secretariat. Version 2.0. Annual Accounts. Brazil, 2023. URL https://siconfi.tesouro.gov.br/siconfi/pages/public/historico_contas_anuais/historico_contas_anuais_list.jsf. Retrieved June 6, 2024.

Sebastián Sotelo. Domestic trade frictions and agriculture. *Journal of Political Economy*, 128(7):2690–2738, 2020. doi: 10.1086/706859.

World Bank. Measuring rural access : using new technologies, 2016. URL https://documents1.worldbank.org/curated/en/367391472117815229/pdf/107996-REVISED-PUBLIC-MeasuringRuralAccessweb.pdf.

World Bank. World bank official boundaries. Dataset, 2025. URL https://datacatalog.worldbank.org/search/dataset/0038272/world-bank-official-boundaries. Accessed 2025-03.

WorldPop. Global 100m population total adjusted to match the corresponding UNPD estimate, 2020. URL https://www.worldpop.org/doi/10.5258/SOTON/WP00660.

A Online Appendix

A.1 Technical appendix: computation procedure

The methodology consists of: (i) identifying the geographic location of the population; (ii) identifying the population residing outside urban settlements, which is taken as rural; (iii) identifying the geographic location of roads, retaining those relevant to the calculation; and (iv) computing the share of the rural population living within two kilometers of any of the roads identified in (iii). This last step requires dividing the result of (iv) by the total rural population, defined as the total population outside urban centers.

Alternatively, the indicator can be computed indirectly along the following path: (i) identify the geographic location of the population and the total population; (ii) identify the geographic location of the selected roads; (iii) compute the total number of people living within two kilometers of any of the selected roads; (iv) identify the geographic location of urban settlements; (v) compute the number of people living within urban settlements; (vi) compute the number of people living within urban settlements and within two kilometers of a selected road; and (vii) subtract the result of (vi) from that of (iii), yielding the rural population living within two kilometers of a selected road. That figure is then divided by the total rural population, obtained by subtracting (v) from the total population in (i).

Although the first path appears more direct, the available technology is more straightforward to implement along the alternative path. Specifically, the computation of the population within the two-kilometer buffer uses the `zonal_stats` function of the `rasterstats` package [Perry, Matthew T., 2015]. This function takes as inputs a file of geographic polygons contained in a vector layer and a file of raster band attributes, and returns summary statistics of the attribute in question for each vector polygon. In our case the raster attribute is the population count in each cell, and our polygons are, on the one hand, the territorial boundaries of each country or subnational area and, on the other, the road network expanded into a two-kilometer buffer together with the urban settlements within the territorial boundaries considered.

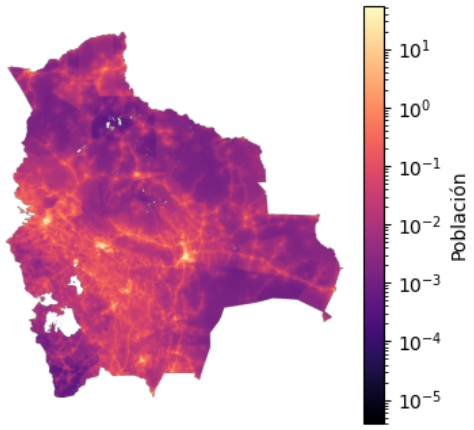
Applying `zonal_stats` with the population raster (WorldPop) and the country's official border polygons as inputs yields the total national population. Applying the same function with GRUMP urban-settlement polygons instead of national boundaries yields the urban population. The total rural population is obtained by subtracting the urban from the total population.

The next step is to identify the total urban population living within two kilometers of any of the roads considered. To do so, we first overlay the urban-settlement polygons with the multi-polygon layer of buffered roads. Figure A.1 illustrates the population layer in panel a, the urban centers in panel b, and the roads buffered to two kilometers in panel

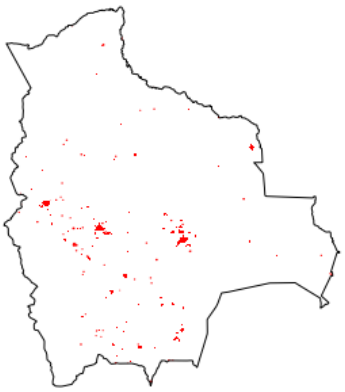
c. The intersection of panels c and d is then used as an additional input to `zonal_stats` together with the population raster, yielding the urban population living within two kilometers of a road. This figure is subtracted from the total population satisfying the road-proximity condition, yielding the rural population within two kilometers of any of the roads considered. This number, divided by the total rural population, is our accessibility indicator.

Note that this is the step at which the road filters—by hierarchy, trafficability, or surface—become relevant. In this case the roads appear as in the original source, as individual lines with independent geometries. To account for the curvature of the Earth's surface and mitigate its effects on the constructed measures, we reproject the road layers to a coordinate reference system compatible with the centroid of the boundary polygon. Applying `zonal_stats` with the population raster and the multi-polygon of roads as inputs, we compute the total number of people in the territory of Bolivia who live within two kilometers of any of the roads considered in the analysis.

Figure A.1: Population distribution, urban settlements, and road network (Bolivia)



(a) Population



(b) Urban settlements



(c) Roads



(d) Subnational boundaries

Source: Authors' computations based on CIESIN-Columbia University et al. [2011], WorldPop [2020], HeiGIT [2024], and World Bank [2025].
Notes: The figure illustrates the geographic layers used to identify the spatial distribution of the population (heatmap in panel a) and the geographic location of urban settlements (red marks in panel b). Applying the `rasterstats` package to the population layer and the area polygon yields total population. Applying the same function to the population layer and the urban-settlement polygons yields the total urban population. The rural population is obtained by subtraction. Panel c shows the network of tertiary-or-higher roads, expanded by a two-kilometer buffer and aggregated into a single multi-polygon geometry; panel d shows the subnational area boundaries. The overlay of panels a and c yields the total population with road access, while the overlay of panels a, b, and c yields the urban population with road access. The rural population within the threshold is then computed by subtraction.

Table A.1: Descriptive statistics

Country	Sub-national name	Areas		Population (thousands)			Roads (thousands km)		
		RED (2025)	Analyzed	Total	Rural	Rural (%)	All	Paved	Paved (%)
Antigua and Barbuda	Parish	8	7	96	2	2	0	0	79
Argentina	Province	24	24	45137	6182	14	2998	1333	44
Bahamas	District	33	33	381	64	17	2	1	62
Barbados	Country	0	1	261	6	2	1	1	89
Belize	District	6	6	393	266	68	23	13	57
Bolivia	Municipality	344	343	11607	5849	50	66	18	27
Brazil	Municipality	5572	5540	206289	43183	21	8445	4176	50
Chile	Municipality	345	319	15676	3492	22	92	56	60
Colombia	Municipality	1122	1114	49927	15044	30	75	37	49
Costa Rica	Canton	82	71	4249	1342	32	9	5	60
Cuba	Municipality	168	162	10369	5042	49	23	9	40
Dominica	Country	0	1	65	18	28	0	0	74
Dominican Republic	Municipality	393	370	9984	2600	26	122	60	49
Ecuador	Canton	222	222	17512	6002	34	43	21	48
El Salvador	District	266	252	5199	2226	43	91	50	55
Grenada	Country	0	1	99	28	28	0	0	67
Guatemala	Municipality	342	330	16813	11395	68	23	10	43
Guyana	Region	10	10	778	353	45	33	7	20
Haiti	Communal section	570	565	10507	7532	72	57	17	29
Honduras	Municipality	298	298	9878	5512	56	19	6	30
Jamaica	Parish	14	14	2931	745	25	5	4	72
Mexico	Municipality	2479	2362	113140	23169	20	301	114	38
Nicaragua	Municipality	153	151	6586	3878	59	10	6	57
Panama	District	82	81	3881	1407	36	8	6	68
Paraguay	District	263	257	6243	2709	43	31	16	50
Peru	District	1874	1802	23317	10637	46	1450	316	22
Saint Kitts and Nevis	Parish	10	10	52	3	6	0	0	85
Saint Lucia	District	14	14	180	22	12	0	0	88
Saint Vincent and the Grenadines	Country	0	1	82	21	26	0	0	69
Suriname	District	10	10	587	214	36	2	1	51
Trinidad and Tobago	Municipality	15	14	1353	176	13	3	2	65
Uruguay	Department	141	129	2529	246	10	15	11	73
Venezuela	Municipality	338	330	27891	7140	26	51	29	57
Caribbean		1225	1192	36278	16237	45	215	95	44
Mesoamerica		3708	3551	160138	49195	31	484	210	43
South America		10265	10100	407492	101050	25	13302	6019	45
Total		15198	14844	603990	166504	28	14000	6324	45

Notes: The table reports descriptive statistics for the main inputs used to compute the Rural Access Index. It lists, for each country, the administrative category considered, the number of subnational areas identified, the number of areas for which the index was computed, total and rural population, and total road length under both the all-roads and the paved-only definitions. Countries without subnational units (a value of zero in the third column) were computed using their total national area, which produces a value of one in the column reporting analyzed areas; this is the case of Barbados, Dominica, Grenada, and Saint Vincent and the Grenadines. The figures do not include areas for which the indicator was not computed.
Source: Authors' computations based on World Bank [2025], CAF [2025], CIESIN-Columbia University et al. [2011], WorldPop [2020], HeiGIT [2024], and Humanitarian OpenStreetMap Team [2025].

A.2 Geographic layers used to construct the RAI

A.3 Descriptive statistics

Table A.2: Subnational Rural Access Index: mean and standard deviation by country

Country	All roads			Paved roads		
	Mean	Sd	Weighted mean	Mean	Sd	Weighted mean
Antigua and Barbuda	85.6	23.4	64.4	76.1	30.5	57.6
Argentina	60.8	12.5	64.8	48.2	15.4	51.1
Bahamas	45.9	36.6	65.9	26.3	37.2	49.6
Barbados			100.0			100.0
Belize	48.1	7.5	47.0	33.7	8.5	35.2
Bolivia	50.5	21.6	54.1	24.6	22.4	29.2

Table A.3: RAI Semielasticity: Non Linear Specification

	(1)	(2)
	All roads	Paved roads
Decile 2	0.17577**	0.02974
	(0.06850)	(0.07134)
Decile 3	0.19834***	0.10946*
	(0.05888)	(0.05475)
Decile 4	0.25870***	0.11201**
	(0.07387)	(0.05228)
Decile 5	0.20562***	0.16110**
	(0.06218)	(0.06287)
Decile 6	0.27449***	0.16729**
	(0.07973)	(0.06747)
Decile 7	0.29068***	0.22327***
	(0.08013)	(0.05476)
Decile 8	0.33769***	0.19813***
	(0.09456)	(0.06115)
Decile 9	0.39290***	0.29099***
	(0.10736)	(0.07934)
Decile 10	0.42100***	0.33146***
	(0.11241)	(0.08619)
Observations	5,410	5,410
R-squared	0.798	0.797

Notes: Non-linear specification: the continuous RAI is replaced by decile dummies and the first decile is the omitted reference category, so each coefficient is the average difference in the log of unimproved agricultural land value between municipalities in the indicated decile and those in the first decile. Column (1) reports the all-roads RAI; column (2) the paved-only RAI. Reading down a column traces the profile of the relationship across the accessibility distribution, and the contrast between columns is the floor effect: the paved-only coefficients stay flat and insignificant in the bottom deciles while the all-roads coefficients rise monotonically. All regressions control for the standard deviation of elevation, the degree of urbanisation (DEGURBA) and state fixed effects. Standard errors clustered at the state level in parentheses; the confidence intervals in the corresponding figure are obtained by wild cluster bootstrap (9,999 replications, 24 state clusters). *** p<0.01, ** p<0.05, * p<0.1.

Table A.4: RAI Semielasticity and Fiscal revenues

	(1)	(2)
	All roads	Paved roads
RAI × Q1 (low revenue)	0.00758***	0.00578***
	(0.00135)	(0.00109)
RAI × Q2	0.00769***	0.00476***
	(0.00235)	(0.00096)
RAI × Q3	0.00531**	0.00223**
	(0.00238)	(0.00105)
RAI × Q4	0.00282*	0.00203**
	(0.00147)	(0.00084)
RAI × Q5 (high revenue)	0.00246	0.00208
	(0.00201)	(0.00146)
Observations	5,378	5,378
R-squared	0.801	0.800

Notes: Each coefficient is the semi-elasticity of the log of unimproved agricultural land value with respect to the RAI, estimated within the indicated quintile of municipal own fiscal revenue per capita. Column (1) reports the semi-elasticity for the all-roads RAI; column (2) for the paved-only RAI. Reading down each column shows the gradient across fiscal-capacity groups, while comparing columns within a quintile shows the gap between the two measures, which is largest in the bottom quintiles and vanishes in the top one. All regressions control for the standard deviation of elevation, the degree of urbanisation (DEGURBA), quintile-of-fiscal-revenue fixed effects and state fixed effects. Standard errors clustered at the state level in parentheses; confidence intervals in the corresponding figure obtained by wild cluster bootstrap (9,999 replications, 24 state clusters). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.