Landscape and Urban Planning 236 (2023) 104762

Contents lists available at ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

Research Paper

Place-level urban–rural indices for the United States from 1930 to 2018

Johannes H. Uhl [a,b,*], Lori M. Hunter [a], Stefan Leyk [a,c], Dylan S. Connor [d], Jeremiah J. Nieves [a,e], Cyrus Hester [a], Catherine Talbot [a], Myron Gutmann [a,f]

[a] Institute of Behavioral Science, University of Colorado Boulder, Boulder, CO 80309, USA
[b] Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO 80309, USA
[c] Department of Geography, University of Colorado Boulder, Boulder, CO 80309, USA
[d] School of Geographical Sciences & Urban Planning, Arizona State University, Tempe, AZ 85287, USA
[e] Department of Geography and Planning, University of Liverpool, Liverpool L69 7ZT, UK
[f] Department of History, University of Colorado Boulder, Boulder, CO 80309, USA

HIGHLIGHTS

- There are no fine-grained measures of the US urban gradient for the 20th century.
- We created indices of remoteness for all census places in the US from 1930 to 2018.
- These indices enable fine-scale, multi-temporal analysis of urban and rural change.
- We compared our indices to a variety of existing rural–urban classifications.
- These indices provide a significant contribution for rural–urban classifications.

ARTICLE INFO

Keywords:
Rural-urban continuum
Urban gradient
Long-term population dynamics
Human settlements
Spatial demography
Spatial network analysis

ABSTRACT

Rural-urban classifications are essential for analyzing geographic, demographic, environmental, and social processes across the rural–urban continuum. Most existing classifications are, however, only available at relatively aggregated spatial scales, such as at the county scale in the United States. The absence of rurality or urbanness measures at fine spatial resolution poses significant problems when the process of interest is highly localized, as with the incorporation of rural towns and villages into encroaching metropolitan areas. Moreover, existing rural–urban classifications are often inconsistent over time, or require complex, multi-source input data (e.g., remote sensing observations or road network data), thus, impeding the longitudinal analysis of rural–urban dynamics. In order to address this gap, we compare existing rural–urban classifications in the US, and we develop a set of distance- and spatial-network-based methods for consistently estimating the remoteness and rurality of places at fine spatial resolution, over long periods of time, aiming to provide and evaluate temporally consistent rural–urban classifications at fine spatial granularity, but scalable to arbitrary, coarser spatial units. We demonstrate the utility of our approach by constructing indices of urbanness for over 28,000 places in the United States from 1930 to 2018 and further test the plausibility of our results against a variety of evaluation datasets. We call these indices the place-level urban–rural indices (PLURAL) and make the resulting code and datasets publicly available so that other researchers can conduct long-term, fine–grained analyses of urban and rural change. In addition, due to the simplistic nature of the input data, these methods can be generalized to other time periods or regions of the world, particularly to data-scarce environments.

1. Introduction

Over the twentieth century, the growth of towns and cities across the United States profoundly reshaped the geography of the nation's population. The share of the US population living outside of urban areas fell from roughly 60 percent in 1900 to less than 20 percent today (Ratcliffe, 2015), and many communities that were once rural were absorbed into cities through urban expansion. Data constraints have,

* Corresponding author at: Institute of Behavioral Science, University of Colorado Boulder, 483 UCB, Boulder, CO 80309-0483, USA.
E-mail address: johannes.uhl@colorado.edu (J.H. Uhl).

https://doi.org/10.1016/j.landurbplan.2023.104762
Received 23 December 2021; Received in revised form 31 January 2023; Accepted 29 March 2023
Available online 25 April 2023
0169-2046/© 2023 Elsevier B.V. All rights reserved.

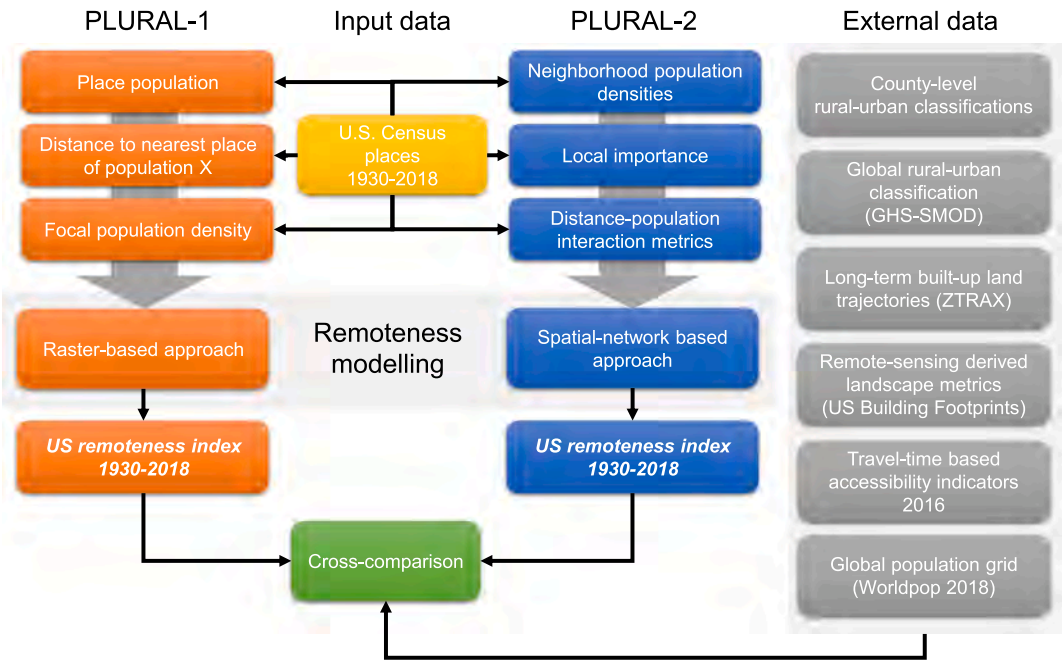

Fig. 1. Flow diagram of the presented methods.

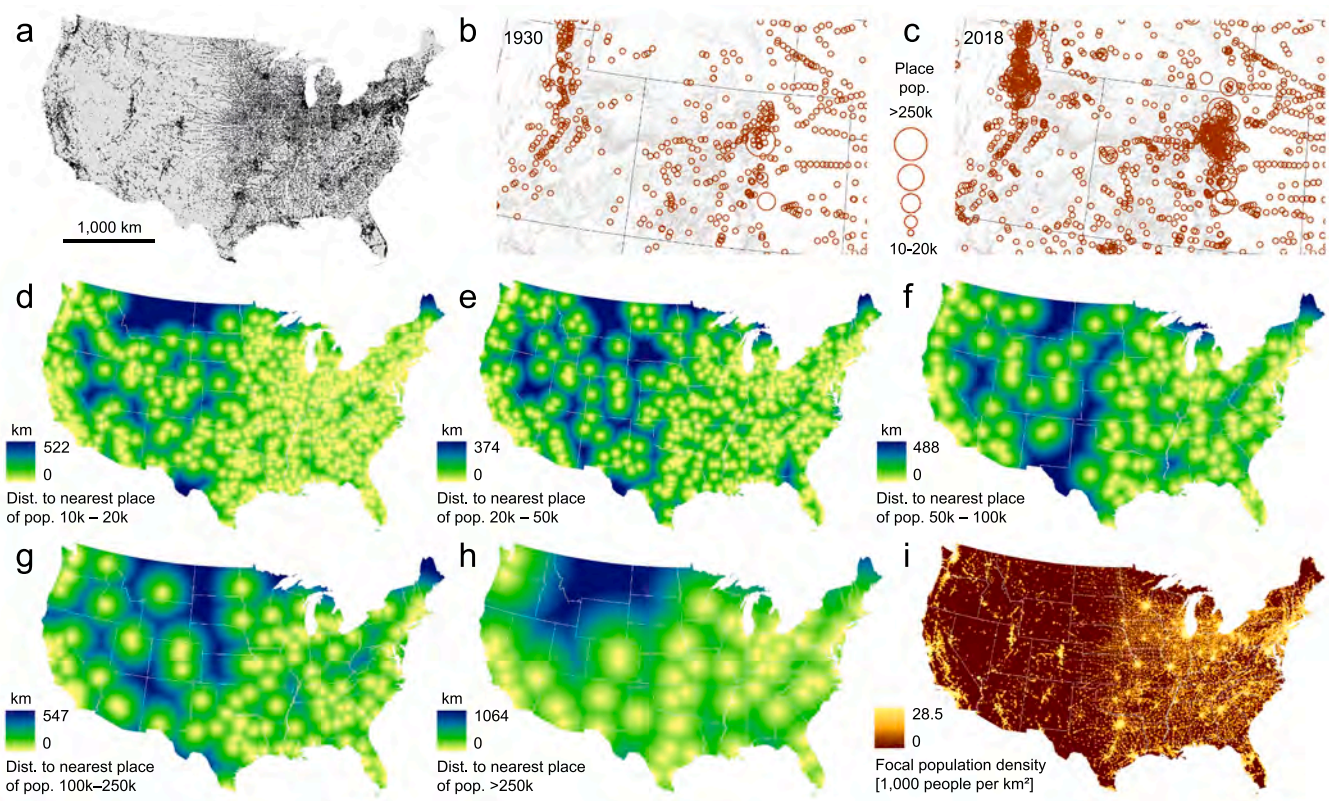

Fig. 2. Source data and created raster datasets used as input for PLURAL-1: (a) US census places in 2018, (b) place-level population for Colorado and Utah in 1930 and (c) in 2018. Panels (d) to (h) show the distance surfaces to the nearest place of population (d) 10,000 to 20,000 (e) 20,000 to 50,000, (f) 50,000 to 100,000, (g) 100,000 to 250,000, and (h) > 250,000 derived from the place population estimates in 2018. Panel (i) shows the focal population density surface derived from the 2018 place populations. All surfaces are generated at a spatial resolution of 1 km.

however, limited our understanding of how this process has unfolded at fine spatial scales and also what is known about the current conditions of rural communities, particularly those with smaller populations (Hunter et al., 2020).

For a wide variety of reasons, researchers often use the county as their analytical unit to study rural communities (e.g., Curtis, DeWaard, Fussell, & Rosenfeld, 2020; Machado, Jayawardana, Mossialos, & Vaduganathan, 2021). The county is often a policy-relevant choice

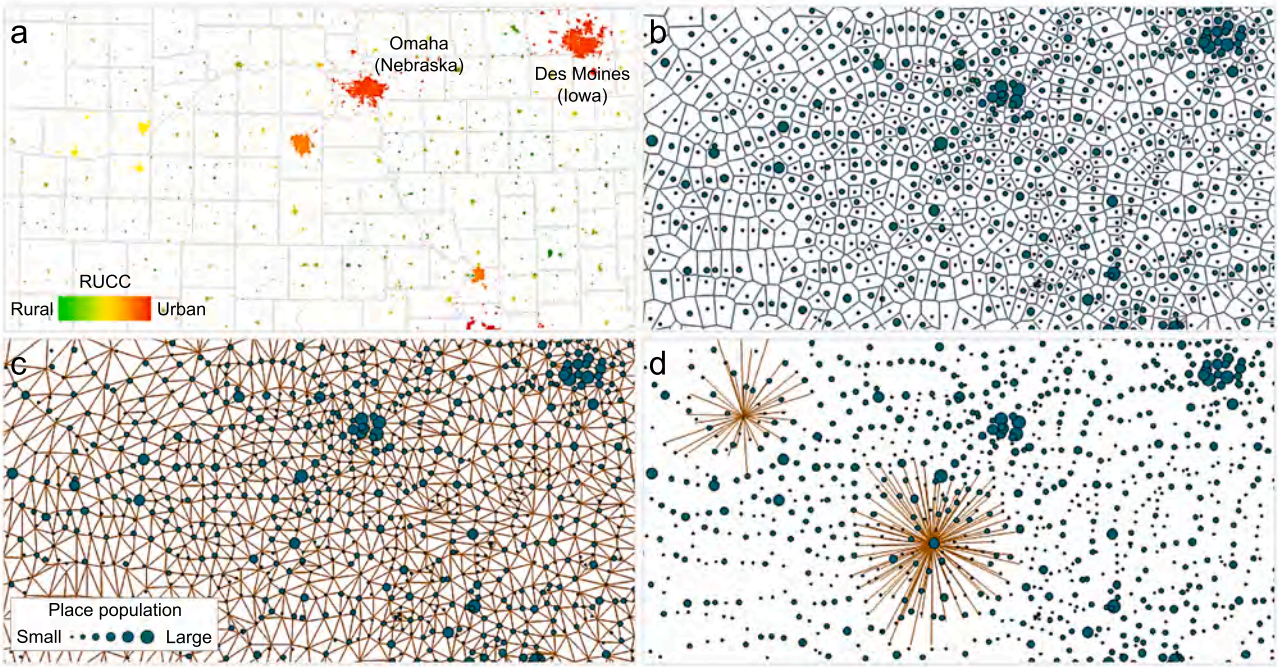

Fig. 3. Illustrating the spatial network generation: (a) US census places in 2010, color-coded by 2013 USDA county-level rural–urban continuum codes, (b) Thiessen polygonization, (c) exhaustive spatial network for neighbors of cardinality 1, (d) exemplary neighborhoods of cardinality 3 (top left) and cardinality 5 (center) shown for two places.

representing local stakeholders within multi-level governance (Homsey, Liu, & Warner, 2019). In addition, a vast amount of information is available at the county scale including sociodemographic characteristics (e.g., US Census Bureau, 2019), health outcomes and behaviors (e.g., CDC, 2021), mortality (e.g., Curtin & Spencer, 2021), and many indices reflecting for example, the rural–urban continuum (e.g., Golding & Winkler, 2020). Counties are, however, coarse descriptors of urban and rural conditions on the ground. In fact, most rural people in the US today – as officially defined – live inside metropolitan counties (Lichter, Brown, & Parisi, 2021). The occlusion of fine-grained variations along the rural–urban continuum by counties is a case of the Modifiable Areal Unit Problem (MAUP), a well-established challenge in the social and spatial sciences that refers to the inferential problems arising from using one unit of spatial aggregation over another (Openshaw, 1979; Flowerdew, Geddes, & Green, 2001, Goodchild, 2022).

The "rural–urban continuum" is a concept that is used to characterize the gradient of urbanization across rural and urban settlements. According to Dewey (1960), the rural–urban continuum captures the intersection of the "population continuum" (as measured by population size and density) and the "cultural continuum" as related to urban and rural ways of life. The rural–urban continuum thus emerges from the interaction of sociocultural and spatial demographic processes (Pahl, 1966) and helps us move beyond artificial binaries between "urban" and "rural" places (Taubenböck et al., 2022). In the literature, there are many neighboring concepts to the rural–urban continuum including the "folk-urban continuum" (Yusuf, 1974), the "rural–urban interface" (López-Goyburu & García-Montero, 2018), or the "rural–urban gradient" (Du Toit & Cilliers, 2011). Although the rural–urban continuum has long been accepted as a useful conceptual and analytic tool, even outside of the US (e.g., Yuan 1964), it has not gone without critique and controversy (see Dewey, 1960, Bell, 1992).

Today, the rural–urban continuum serves as a useful analytic device for examining differences in a wide range of outcomes across contexts and for tracking urbanization patterns over time. This framework has been used to study a variety of sociological and health-related outcomes (e.g., Hillemeier, Weisman, Chase, & Dyer, 2007, Sibley & Weiner, 2011, Lee & Sharp, 2017, Peters, 2020), migration patterns (Golding & Winkler, 2020), land consumption and biodiversity (Murali, Suryawanshi, Redpath, Nagendra, & Mishra, 2019), income inequality (Thiede, Butler, Brown, & Jensen, 2020), mortality (Brooks, Mueller, & Thiede, 2020), political polarization (Scala & Johnson, 2017), and manyother processes (e.g., Pender, Hertz, Cromartie, & Farrigan, 2019, Johnson & Lichter, 2020, Lichter & Johnson, 2020, Lichter et al., 2021). As is evident in recent social mobility research, outcomes often do not shift in a linear fashion across the rural–urban continuum, as transitional, *peri*-urban and "micropolitan" places can seemingly constitute their own unique contexts that require their own forms of policy attention (Weber, Fannin, Cordes, & Johnson, 2017). There are even now efforts to generate future predictions based on the urban–rural continuum (e.g., Abdelkarim, Alogayell, Alkadi, & Youssef, 2022). Confidence in these findings rest, of course, on the assumption that we have appropriately classified the continuum.

One of the most notable features of these classifications in the US context is that they are generally based on county-level units (e.g., Cromartie, Dobis, Krumel, McGranahan, & Pender, 2020). We know, however, that many rural and urban processes and their associated effects play out at the sub-county scale (e.g., places, towns, and villages), and therefore, are masked by county scale analyses. For example, in our recent work on rural social mobility, we document that much of the variation in US social mobility outcomes is between places within the same counties, rather than between places in different counties (Connor, Uhl, Xie, Talbot, Hester, Jaworski, & Hunter, 2022). Furthermore, many other globally-significant socio-environmental processes—such as amenity-driven migration, land-use change, biodiversity loss, displacement and segregation—manifest from predominately local processes (Rockström et al., 2009, Gosnell & Abrams, 2011, Banzhaf & Walsh, 2013). Zoning, income levels and the distribution of public goods (e.g., public spaces, viewscapes, school quality) or nuisances (e.g., crime, pollution) influence local housing markets and drive the differentiation of places (Grineski, Bolin, & Boone, 2007, Glaeser, Resseger, & Tobio, 2009, Banzhaf & Walsh, 2013, York et al., 2014, Banzhaf, Ma, & Timmins, 2019). These same processes have implications for the

Table 1
Overview of the network-based remoteness metrics.

Metric	Reference neighborhood	Description
Population		
POP_{Place}	Place	Population of the place (i.e., node)
NPD_1	Cardinality 1	Population density of places in neighborhood of cardinality 1, referred to the squared maximum distance between places in neighborhood
NPD_2	Cardinality 2	Population of places in neighborhood of cardinality ≤ 2, referred to the squared maximum distance between places in neighborhood
NPD_3	Cardinality 3	Population of places in neighborhood of cardinality ≤ 3, referred to the squared maximum distance between places in neighborhood
Local significance		
MLS_1	Cardinality 1	Median local significance (MLS) of edges connecting each place with its neighbors of cardinality 1
MLS_2	Cardinality 2	MLS of edges connecting each place with its neighbors of cardinality ≤ 2
MLS_3	Cardinality 3	MLS of edges connecting each place with its neighbors of cardinality ≤ 3
Distance-based neighborhood population indices		
$DNPI_{C3}$	Neighborhood of cardinality ≤ 3	DNPI (i.e., AUC of cumulative population of neighbors, sorted by their distance), within neighbohood of $c = 3$.
$DNPI_{250km,500k}$	250 km	DNPI within radius of 250 km, or until reaching a cumulative population of 500,000
$DNPI_{500km,1,000k}$	500 km	DNPI within radius of 500 km, or until reaching a cumulative population of 1,000,000
$DNPI_{MAXPOP}$	d_{MAX}	DNPI until reaching a cumulative population equal to the population of the largest place in the analyzed distribution, or the maximum occuring distance d_{MAX}.

sustainability of landscapes by influencing land uses, resource demands, and development along the rural–urban continuum (Theobald & Romme, 2007, Grimm, Pickett, Hale, & Cadenasso, 2017). Full understanding of these patterns requires that our scale of observation and analysis are well aligned with the true scale at which processes unfold. Failure to achieve this can lead to substantial mischaracterization and bias when using coarse spatial units (e.g., county) to describe finer units (e.., places) within them (Hunter et al., *forthcoming*). Again, these issues are closely related to the MAUP, which may be thought of as the geographical manifestation of the Ecological Fallacy Problem (Piantadosi, Byar, & Green, 1988).

In order to study rural and urban processes, researchers have already generated many indices, classifications, and typologies of rurality, based on a wide range of data (Nelson et al., 2021). Examples of existing rural–urban classifications in the US include the commonly used *rural–urban continuum codes (RUCC)* created by the US Department of Agriculture's (USDA) Economic Research Service (ERS). The RUCC identify nine categories, i.e., three metro and six nonmetropolitan county designations, with metropolitan counties further disaggregated by the encompassing metro area's population size (McGranahan, Hession, Mines, & Jordan, 1986, Butler, 1990). Nonmetropolitan counties are further classified by their degree of urbanization and adjacency to a metropolitan area. Golding and Winkler (2020) refined the RUCC to distinguish explicitly between urban cores and their exurbs and suburbs, resulting in the *rural–urban gradient (RUG)*. The USDA ERS also produces the *rural–urban commuting area (RUCA) codes* (Economic Research Service, 2022; also available by the ZIP code area). The RUCA codes make use of the US Census urbanized areas and urban core designations (Census Bureau, 2020), in combination with census-tract level commuting flow estimates. Combined, the RUCA groups census tracts into 10 classes of commuting levels. Another related measure is the *USDA urban influence codes (UIC)* (Ghelfi & Parker, 1997), which yields nine different classes based on the population of the county's largest city rather than an aggregated urban population as in RUCC.

The National Center for Health Statistics (NCHS) have also released a classification called the *urban–rural classification scheme (URCS)* (Ingram & Franco, 2014), which is based on metropolitan and non-metropolitan county classification in combination with population thresholds, identifying six county designations. Moreover, a continuous classification scheme is provided by the *index of relative rurality (IRR)* (Waldorf, 2006; Waldorf & Kim, 2018), a county-level index based on population size, density, road network distance, and built-up areas. As the IRR method is independent from administrative or census-defined boundaries, the underlying framework can be applied to finer-grained spatial units as well (Waldorf & Kim, 2015). Finally, there are the *Frontier and Remote (FAR) Area Codes* available at the ZIP code level (Cromartie & Nulph, 2015), which provide four classes of remoteness, and are derived from travel times and population estimates. See Waldorf and Kim (2015) and National Academies of Sciences, Engineering, and Medicine (2016) for reviews of these various classification approaches, and Fig. 6 for a visual comparison of these classifications.[1]

However, these existing measures of the rural–urban continuum typically face at least one of three challenges of relevance to rural populations. We describe each of these in turn below.

County-scale data are relatively coarse. Indices at the county-scale face important challenges. The principal problem explicated by the MAUP is that results based on data that have been aggregated to a set of spatial units will change when those units change (Goodchild, 2022). This issue applies implicitly to our measurement of locations along the rural–urban continuum (e.g., Cosby et al., 2019; Johnson & Lichter, 2019; Monnat, 2020) and also when drawing inferences on subcounty processes from coarser county-level data (Homsy, Liu, & Warner, 2019), i.e., the ecological fallacy problem.

County-scale analyses can also face the Uncertain Geographic Context Problem which refers to the challenge whereby relevant conclusions depend on analyses at spatial units corresponding to the true and relevant geographic contexts experienced by individuals (Kwan, 2012). Here we can look to a prominent body of recent rural-related work that examines how the characteristics of places shape individual-level processes including social mobility (Chetty, Hendren, Kline, & Saez, 2014; Connor & Storper, 2020), racial inequality (Manduca & Sampson, 2019), health, and voting (e.g., Shah et al., 2020; Sachdeva, Fotheringham, Li, & Yu, 2021). Indeed, individual lives are typically lived in communities, places that influence life trajectories and places to which meaning can be ascribed. Such meaning furthers a sense of belonging and the development of place-based identities (e.g., Sack, 1997, Manzo & Devine-Wright, 2013, Armstrong & Stedman, 2019). Place identity appears especially strong among rural dwellers (Lewicka, 2005, Anton & Lawrence, 2014) and, while today's rural America is ever-changing, recent work confirms that, in general, rural residents remain deeply tied to place and hold strong commitments to community (Ulrich-Schad & Duncan, 2018). As compared to ZIP codes, places have

[1] Many countries provide individual delineations of urban and rural areas (Workman & McPherson, 2021), often relying on census-based information. At a global scale, researchers typically rely on classifications of the rural–urban continuum derived from remotely sensed earth observation data in combination with population estimates, such as GRUMP (Balk, Pozzi, Yetman, Deichmann, & Nelson, 2005), degree of urbanization (Dijkstra & Poelman, 2014), GHSL-SMOD (Florczyk et al., 2019). However, these methods are confined to recent decades, of relatively coarse spatial resolution, and represent a land perspective more than a population view.

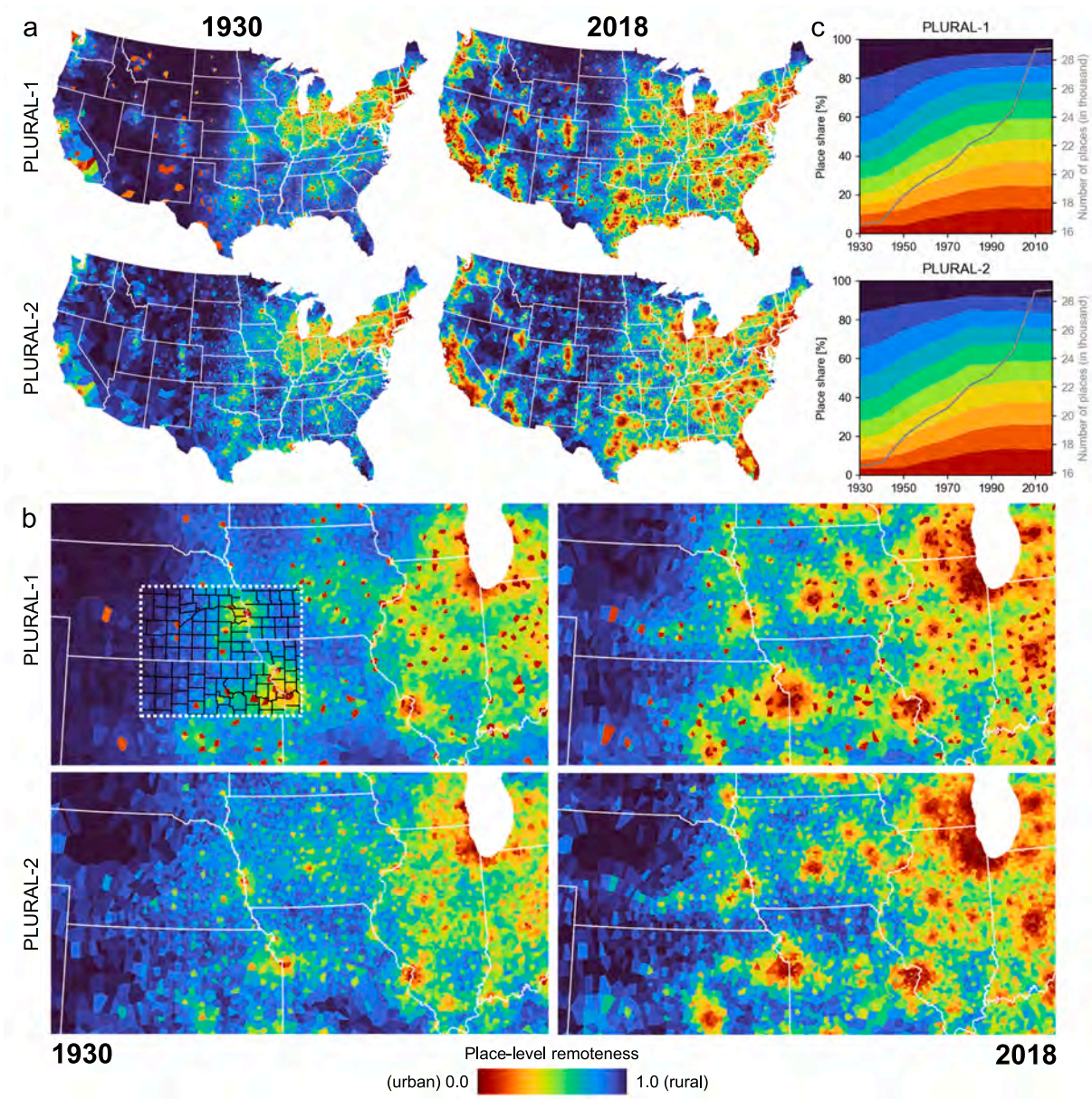

Fig. 4. Equal-weighting schemes for raster-based (PLURAL-1) and network-based (PLURAL-2) remoteness indices in 1930 and 2018, (a) for CONUS and (b) for the US Midwest; indices are scaled jointly across all points in time and thus, comparable over time. In the enlargement, county boundaries are shown in black for a subregion, and state boundaries are shown in white. In these maps, each place is represented by the Thiessen polygon established from the discrete place locations. Panel (c) shows the temporal trends of the PLURAL indices over time, discretized into 10 classes of width = 0.1.

the advantage of being representative of a meaningful social unit. ZIP codes were created to increase the efficiency of mail delivery and can be especially problematic as an analytical unit in rural areas (Grubesic, 2008). "Places", as defined by the US Census provide functions for a concentration of people. They are locally recognized, independent of other places, and can be either incorporated places – defined by criteria within their respective states – or census-designated places, which are not incorporated and lack a municipal government (US Census Bureau, 1994). As such, places are of both practical and social importance in the lives of rural dwellers (Federal Register, 2008). An important constraint of place as a unit of analysis, however, is that it neglects consideration of residents outside of place boundaries. We contend that, as collectives, places serve nearby residents as well as those within specific boundaries.

There is often a lack of temporal consistency. County boundaries change over time, as do the designations of metropolitan and non-metropolitan counties. As a result, classifications such as the RUCC suffer from temporal inconsistencies caused by changes in methodology, and by changing units that cannot be compared between different points in time. These issues may prevent scholars from conducting long-term studies across the rural–urban continuum or may constrain analyses to narrower temporal windows, where classifications are consistent, which may in itself inhibit efforts to produce generalizable findings (see the Modifiable Temporal Unit Problem (MTUP, Çöltekin et al., 2011). Of course, place boundaries also change across time and, as a result, we do not explicitly engage place-based boundaries in the approach articulated below. Instead, we use place population data and incorporate a broader, more general representation of spatial extent based on distance to other places (details below).

Existing approaches are often based on limited and hard-to-acquire data. Some of the more complex indices (e.g., the IRR, FAR) are grounded in data reflecting road networks or built-up areas, information that does not typically offer substantial historical coverage. Such data are more

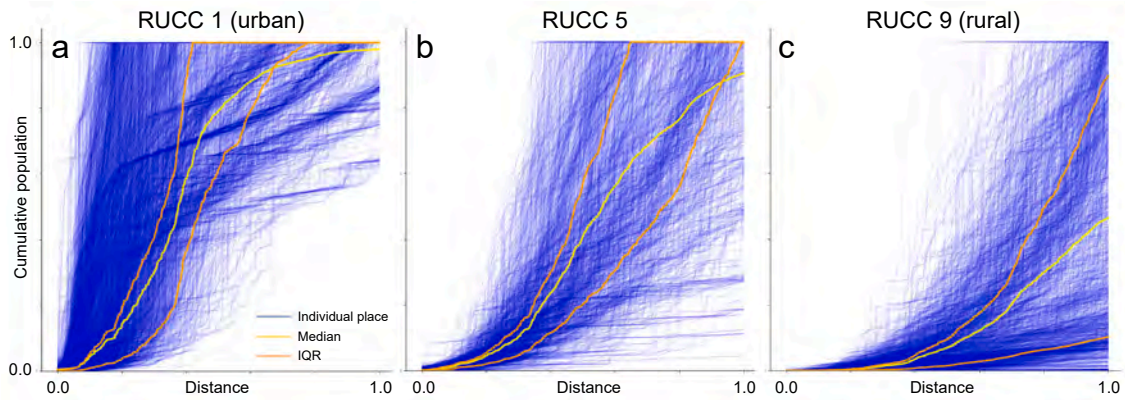

Fig. 5. Illustrating the concept of the distance-based neighborhood population index (DNPI). Cumulative population curves over the neighboring places, sorted by their distance, provide an area under the curve which characterizes the spatial configuration of place populations with respect to a "focal" place. Shown are the curves based on US census places 2010 within counties of (a) RUCC 1 (urban), (b) RUCC 5 (*peri*-urban), and (c) RUCC 9 (most rural) in 2013. The stopping criteria used are a maximum distance of 500 km, and a maximum cumulative population value of 1,000,000.

challenging to acquire compared to Census data and may not be available for early points in time (e.g., gridded population data).

Hence, existing measures of the rural–urban continuum face a combination of these three challenges: a) they are generally derived from county-level data which can be too coarse a scale to describe the population dynamics of rural places; b) they lack spatiotemporal consistency which impedes longitudinal analysis; and/or c) they are constructed based on measures of infrastructure access, commuting patterns, or urbanized land rather than population size, and thus, are constrained to time periods where such data is available. In this article, we evaluate existing characterizations of the rural–urban continuum and propose new classification approaches that helps, in part, address the three issues above: 1) Our classifications are reported at ***fine spatial grain,*** 2) they are ***scalable*** to different spatial units, and 3) their minimalist input data make them ***generalizable*** to data-poor regions.

Such refined indicators enable new possibilities for understanding pressing urban and rural issues, such as disparities in regional development, infrastructure, and social and economic well-being. Until now, analyses of these issues have often been constrained to relatively coarse scales of analysis. Focusing on the spatial distributions of the proposed rural–urban classifications also allows for direct examination of the changing nature of urban and rural places. Analysis of these rural–urban indices over time provides unprecedented insight into the development history and urbanization of the United States and our reliance on publicly available data in the creation of these fine-scale indices provides an accessible and flexible option for scholars and policymakers, particularly those concerned with issues affecting small places and other data scarce environments.

Particularly for rural settings, we contend that a better understanding of sub-county units – especially the spaces where rural dwellers focus their daily, collective activities – is essential for research, planning, and the development of place-relevant policies and programs. The place-level classification approaches presented here advances efforts to address this important gap, without aiming to replace existing, and widely used county-level classifications, but rather providing an additional, scalable and generalizable approach.

2. Data and methods

Overall, efforts to analyze demographic processes across the rural–urban continuum at the place level and over time have been impeded by the lack of spatially fine-grained and temporally consistent indicators of rural and urban places. Because "urbanness" or "rurality" are multivariate processes that evade simple definition, we use the concept of "remoteness" to continuously measure the urban-to-rural spectrum. We consider places as "remote" (i.e., rural) if they have relatively small populations and are surrounded by other small places. Maximum remoteness is achieved if these surrounding, small places are also very distant. We refer to "non-remote" (i.e., urban) places if they are relatively large in population and/or surrounded by other large places, thus implementing the concept of remoteness on a continuous scale.

Our definition of remoteness as a measure of rurality is based around how individuals may or are likely to experience a place, as well as quantitatively describing the average access to other populations, infrastructure, and services. This is motivated by the assumption that quantity, quality, and type of such infrastructure and services are often highly correlated with the population size of any individual place and the spatial arrangement with any neighboring places and these neighbors' own characteristics. We assume remoteness and rurality to be correlated as well. However, by using the term remoteness, we emphasize the underlying, purely geometric modelling strategy, while the term "rurality" also implies identity-related characteristics that we are not able to take into account in our approach.

We propose two methods to derive measures of remoteness of places in the US (and possibly elsewhere) at fine spatial granularity. These methods generate consistent classifications of rural and urban places over long periods of time by implementing simplified characteristics commonly used to define rural–urban classes (e.g., size, distance, local importance and spatial relationships between populated places). Specifically, the first approach is based on population size of places and the weighted (Euclidean) distances to other places of different size categories. While this approach is computationally efficient and can be implemented as a raster-based approach, it may overly generalize local spatial configurations of populated places, and thus, ignore valuable information regarding the local importance of a place. Thus, we propose a second approach, based on a spatial network, that adopts concepts from landscape ecology and network analysis to model remoteness in a more spatially explicit manner. We call the proposed indices the ***place-level urban–rural indices (PLURAL)***. We name the raster-based index PLURAL-1, and the spatial network-based index PLURAL-2. Both approaches rely on the same data input, which are solely derived from public-domain data and allow for the derivation of various (combined) distance and population-based attributes to model remoteness based on different perspectives. These approaches differ in their spatial modelling methodology, the technical skills required for implementation, and their suitability to different research problems. We compare these two methods to ascertain which modeling strategy is more suitable for use, i.e., to generate new methodological knowledge, and in order to provide users with alternative modelling strategies accounting for different levels of modelling skills.

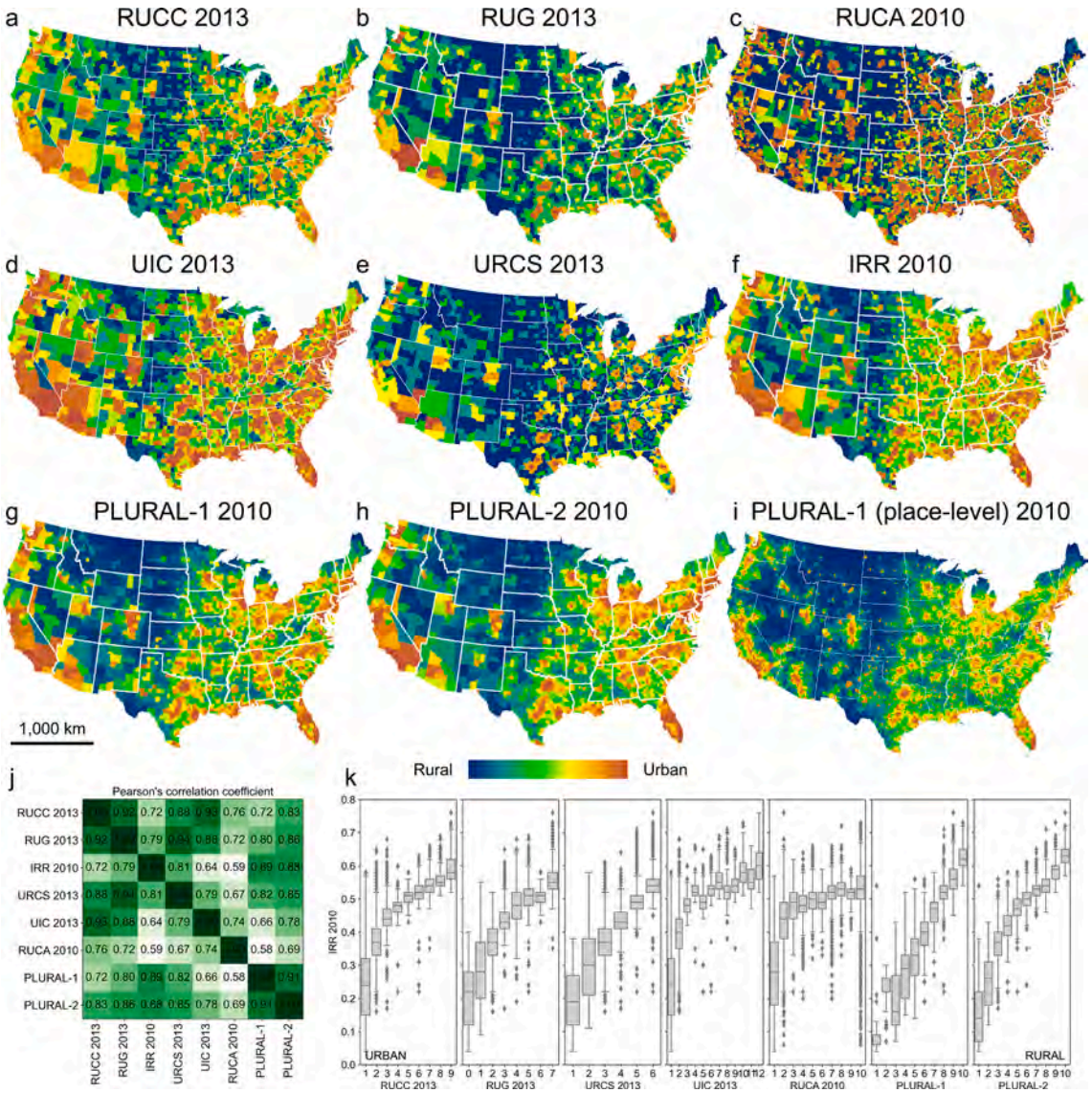

Fig. 6. Cross-comparison of existing county and tract-level rural–urban classifications for the U.S: a) USDA rural–urban continuum codes (RUCC) in 2013 (b) the Rural-Urban Gradient (RUG) 2013, (c) USDA rural–urban commuting areas (RUCA) in 2010 at the tract-level, (d) USDA urban influence codes (UIC) 2013, (e) 2013 NCHS Urban–Rural Classification Scheme (URCS), and (f) the Index of Relative Rurality (IRR) in 2010; (g) and (h) show county-level averages of the place-level PLURAL-1, and PLURAL-2, respectively, and panel (i) shows the place-level PLURAL-1 index to illustrate the difference in spatial granularity. Panel (j) shows cross-correlations between the different rural–urban classifications, computed at the tract-level for RUCA, and at the county-level otherwise. Panel (k) illustrates the variations of the continuous IRR against the other classification schemes. All correlation coefficients in (j) have a p-value < 0.05 and thus, reported correlations are statistically significant (see Supplementary File 1).

Herein, we describe the derivation of the PLURAL indices and their underlying data for the conterminous US (CONUS), as well as a range of cross-comparisons and plausibility analyses (Section 2). We then demonstrate the applicability of the PLURAL for modelling the long-term dynamics of the rural–urban continuum by measuring place-level remoteness in the CONUS for each decade, from 1930 to 2018 (Section 3), and assess the plausibility of the calculated place-level rural–urban classifications by comparing against a range of external data sources (Section 3.4). We conclude with a critical discussion (Section 4) and conclusions (Section 5). The indices for the time period from 1930 to 2018 are publicly available for download as tabular and spatial datasets at https://doi.org/10.6084/m9.figshare.22596946, and the underlying Python code is available at https://github.com/johannesuhl/plural.

Specifically, we describe the input data and the derivation of the PLURAL-1 index based on gridded surfaces (i.e., raster-based approach). We then describe PLURAL-2, a spatial network-based remoteness modelling approach that explicitly accounts for local spatial relationships between populated places by adopting concepts from network analysis and landscape ecology. We also introduce the data sources and strategies used for cross-comparison and plausibility analysis of the results. An overview of the presented approaches is shown in Fig. 1.

2.1. Source data and preprocessing

In this work, we use US census place population counts and locations for each decade from 1930 to 2010, (NHGIS[2]; Manson, Schroeder, Van Riper, Kugler, & Ruggles, 2020, US Census Bureau 1942; 1964). as well as for 2018 (Manson et al., 2020). These place locations are shown in Fig. 2a,b,c. See Appendix 1 for details on the source data. We integrated and harmonized these data and used these integrated datasets as base

[2] https://www.nhgis.org/.

Table 2
A brief summary of the different evaluation analyses of the PLURAL indices between each other and against external datasets.

Assessment	Section	Major finding
Comparing county aggregates of PLURAL to existing county-level classifications	3.3.	PLURAL aggregated to the county is most correlated to the IRR, and least correlated to the RUCA, due to the different input data and modelling techniques.
Comparing PLURAL-1 to PLURAL-2	A6-1	The network-based approach (PLURAL-2) tends to classify places as more urban than the raster-based approach (PLURAL-1).
Comparing place-level indices to county-level classifications	A6-2	PLURAL-2 exhibits far fewer outliers when analyzing distributions within county-level groups; PLURAL-2 appears to agree more with county-level classifications.
Comparison to global gridded population data	A6-3	The remoteness-based concept underlying the PLURAL indices allow for incorporating the relationships between nearby places, by modelling their influence, which raw population counts or densities cannot do directly.
Comparing to FAR codes	A6-4	All indices show high agreement with travel-time based FAR classes, indicating that the Euclidean distance-based approximation does not meaningfully bias the resulting indices.
Comparing to travel-time based accessibility indicators	A6-5	The implemented distance concept using Euclidean distance rather than road network distance introduces low levels of bias overall but may distort the results for a small number of individual places.
Comparing to long-term built-environment trajectories	A6-6	We observe a negative correlation between remoteness and total buildings per place. This negative correlation increases over time, indicating greater plausibility in our indices for more recent points in time.
Comparing to contemporary landscape metrics	A6-7	We observe strong associations between remoteness and morphological characteristics (landscape metrics) for all indices and weighting schemes.
Sensitivity analysis	A7	We observe high levels of robustness, consistent over time, and higher levels of sensitivity to the chosen population thresholds than to focal window size. The metro-centric PLURAL-1 weighting scheme exhibits highest levels of robustness overall.

data for all subsequent data processing and analyses. In total, we obtained 213,827 place locations, across all years (from 15,641 places in 1930, to 28,814 places in 2018), attributed with their population counts.

2.2. Derivation of the raster-based remoteness index (PLURAL-1)

Using place-level population estimates (1930–2018), provided for discrete geospatial locations, we design a method to model the remoteness of places across the region of interest (e.g., the US). The remoteness index of a place is computed based on the size ***s*** (i.e., population) of the place of interest and the distance between that place and the nearest places of varying size categories (10,000–20,000, 20,000–50,000, 50,000–100,000, 100,000–250,000 and > 250,000 people, herein referred to as population categories ***pc***, Fig. 2d-h). These or similar categories are used in numerous studies (e.g., Angel, Parent, Civco, Blei, & Potere, 2011, Cromartie & Nulph, 2015, Nelson et al., 2019). Moreover, we used a population density surface derived from the place-level population counts (Fig. 2i). We implemented this method as a raster-based approach using a GIS-based workflow, which is discussed in detail in Appendix 2.

The place-level urban–rural index ***PLURAL-1*** for a given place ***p***, in a given year ***t*** can be derived as the weighted average of the inverse of the population size $s_{p,t}$, of place ***p*** in year ***t***, and the distance measures $D_{pc,p,t}$, (in meters) to the nearest place of population category ***pc*** (Equation 1). All measures are log-transformed to achieve a uniformly distributed index despite skewed distributions of population and population density (and potentially skewed distributions of distance measures due to the presence of extremely remote places):

$$PL\bar{U}RAL_{p,t} = \left[w_s \cdot \log\left(\frac{s_{p,\max}}{s_{p,t}}\right) + w_{pd} \cdot \log\left(\frac{pd_{\max}}{pd_{p,t}}\right) \sum_{pc=1}^{5} w_{pc} \cdot \log\left(D_{pc,p,t}\right) \right] \tag{1}$$

With

$$w_{p,t} + w_{pd,t} + \sum_{pc=1}^{5} w_{pc} = 1.0 \tag{2}$$

where w_s is the weight for place population, w_{pd} is the focal population density weight, and w_{pc} are weights for the distance measures to different population categories ***pc*** to allow for adjusting the influence of local versus regional population centers. The constants $s_{p,max}$ and pd_{max} are global maximum values of place population and focal population density, respectively, and can either be derived from the data distribution or chosen based on domain knowledge. Herein, we use a maximum place population of $s_{p,max} = 10,000,000$ and a maximum focal population density of $pd_{max} = 15,000$ people / km^2. By log-dividing the upper bounds of population and population density by the place-level values, $\log\left(\frac{s_{p,\max}}{s_{p,t}}\right)$ and $\log\left(\frac{pd_{\max}}{pd_{p,t}}\right)$, respectively, we obtain measures that yield low values for large, densely populated places.

This results in a total of seven ***remoteness indicators*** (population, population density, and five distance measures). While we tested four exemplary weighting schemes, emphasizing different components of "remoteness" (see Appendix 2), herein we focus on an equal weights scenario for simplicity. The final raster-based index is then calculated by scaling the raw index measures $PL\bar{U}RAL_{p,t}$ into the range [0,1]. This computation yields values close to 0 for large places near other (large and/or small) places, and values close to 1 for small places, remote from other places. By approximating each place by a discrete point location (i.

Table A21
Weighting schemes for raster-based remoteness index.

Weighting scheme	Place population	Focal population density	Distance (10 k-20 k)	Distance (20 k-50 k)	Distance (50 k-100 k)	Distance (100 k-250 k)	Distance (>250 k)
Equal weights	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Place-centric	0.250	0.250	0.100	0.100	0.100	0.100	0.100
Place-centric & metro focus	0.250	0.250	0.033	0.067	0.100	0.133	0.167
Metro focus	0.100	0.100	0.053	0.107	0.160	0.213	0.267

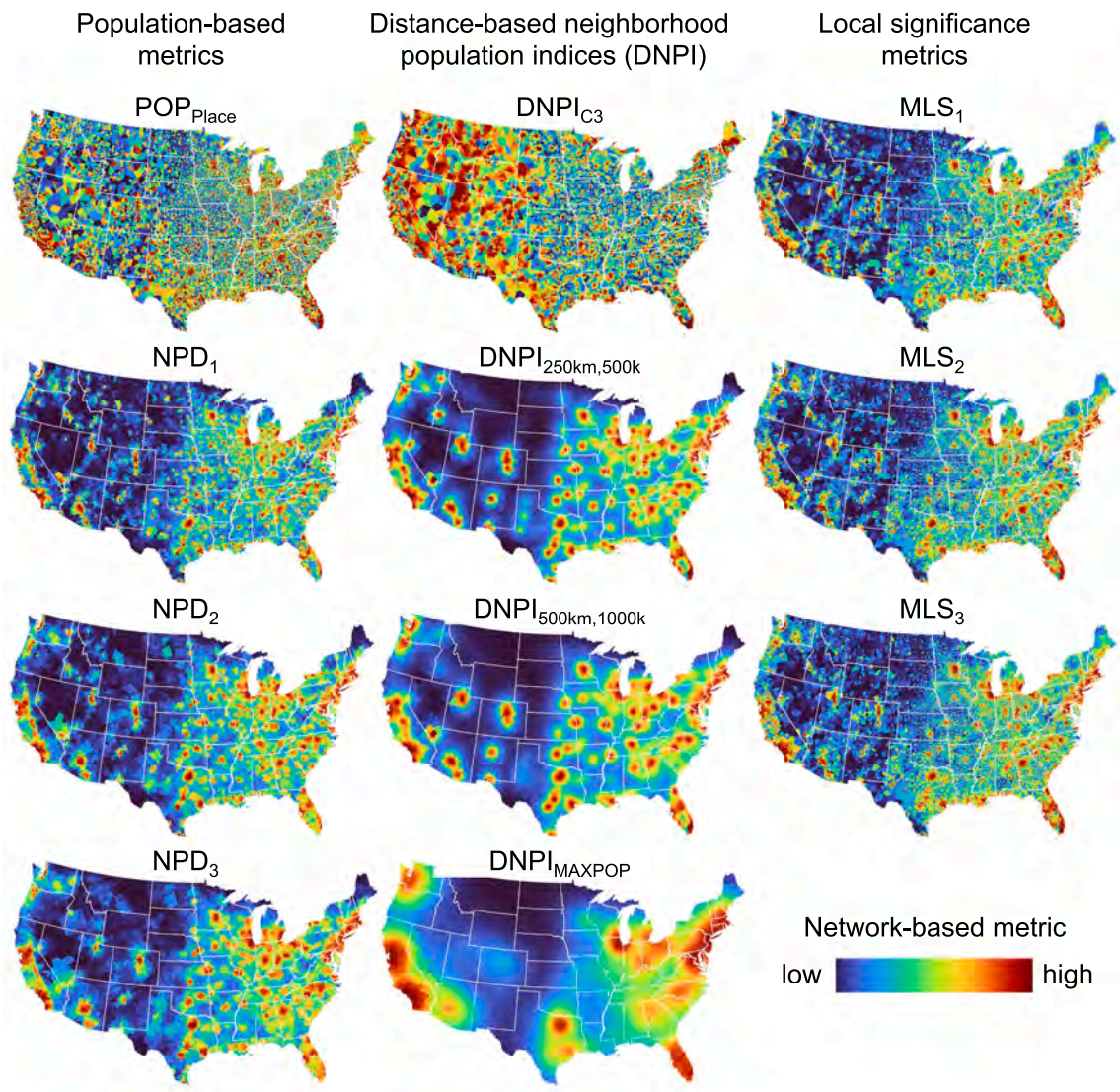

Fig. A3-1. Maps of the 11 network-based metrics in for each census place in 2018, input to the spatial-network-based PLURAL-2 remoteness indices. Abbreviations: NPD = Neighborhood population density, DNPI = Distnnce-based neighborhood population index, MLS = Median local significance. In these maps, each place is represented by the Thiessen polygons established from the discrete place locations.

Table A31
Weighting schemes for network-based remoteness index.

Weighting scheme	POP_{Place}	NPD_1	NPD_2	NPD_3	MLS_1	MLS_2	MLS_3	$DNPI_{C3}$	$DNPI_{250km,500k}$	$DNPI_{500km,1,000k}$	$DNPI_{MAXPOP}$
Equal weights	0.091	0.091	0.091	0.091	0.091	0.091	0.091	0.091	0.091	0.091	0.091
Population focus	0.125	0.125	0.125	0.125	0.071	0.071	0.071	0.071	0.071	0.071	0.071
DNPI focus	0.071	0.071	0.071	0.071	0.125	0.125	0.125	0.125	0.071	0.071	0.071
Significance focus	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.167	0.167	0.167

e., the place polygon centroid), rather than using its areal extent, and by modelling the distances between places using Euclidean rather than road network distances, our approach is highly versatile and generalizable to data-scarce environments and (early) time periods, as retrospective areal place extents and multi-temporal road network data are rarely available for these periods. The chosen parameters such as the thresholds to define the population categories ***pc***, the window size to model focal population density, and the spatial resolution of raster surfaces used (Fig. 2d-i) may potentially affect our PLURAL-1 indices. To assess this, we conducted a sensitivity analysis to these parameters.

2.3. Modelling remoteness based on spatial networks (PLURAL-2)

The previously described raster-based approach is computationally inexpensive. That is, population density and distance-based components can be derived from distance grids easily in commonly used GIS environments. However, this approach may ignore the local, spatial configuration of populated places, which may contain critical information regarding the local importance of a place. Thus, we use concepts from network analysis and landscape ecology to provide a second modelling approach. Such methods and metrics have been applied to human settlement modelling based on remote-sensing derived patches of built-up land (Esch et al., 2014) or for analyzing global land cover

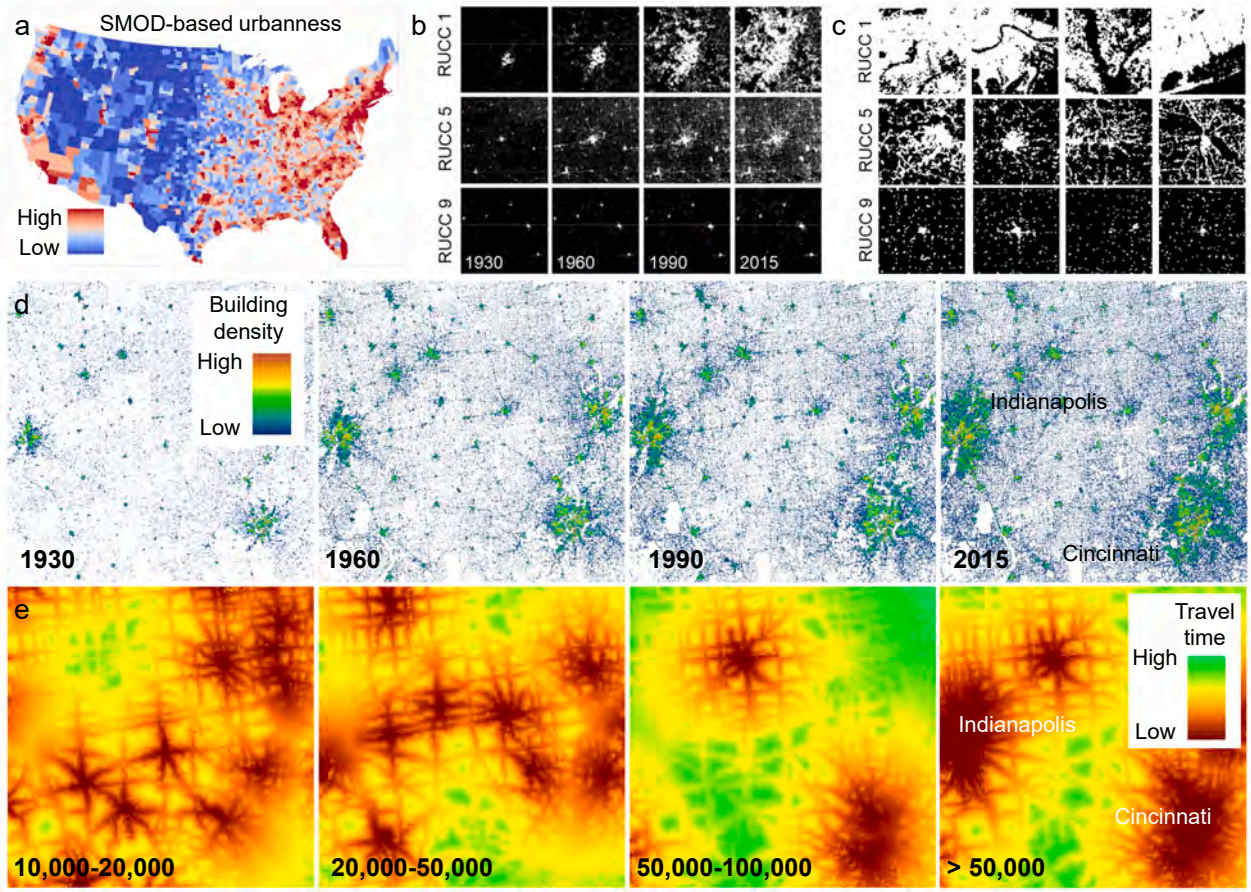

Fig. A4. Data used for the evaluation of the proposed remoteness indices: (a) GHS-SMOD based county-level urbanness score, (b) HISDAC-US place-level built-up areas 1930 – 2015, (c) MSBF built-up areas, both shown for illustrative examples of places within counties of different RUCC, (d) HISDAC-US building density 1930–2015 shown for the Indianapolis-Cincinnati region, and (e) travel-time based accessibility indicators to cities of different population ranges from Nelson et al. (2019) shown for the same region.

patterns (Nowosad & Stepinski, 2018). Using a network to describe the spatial configuration of the point-based places allows for the derivation of topology-based, and thus, density-independent metrics. This is particularly important as the population and settlement density across the United States varies considerably across space and time. Similarly, utilizing local landscape metrics enables the quantification of the localized, place-centric configuration of neighborhood place populations. Moreover, this network-based approach allows for a joint, place-centric assessment of neighboring places, where the raster-based approach only considers the nearest places of each population category only without taking into account the whole spectrum of the spatial context (e.g., the *n*-th nearest place) which may contribute to the rurality of a given place as well.

2.3.1. Establishing place-level spatial networks

Populated places may be given as discrete point locations (see Fig. 2b,c), or, typically for recent points in time, as areal objects (Fig. 3a). In this case, place locations from 1980 onwards are given as polygons, and, prior to that, as discrete point locations attributed with their place population. For consistency, we converted place polygons into discrete locations by using their centroid coordinates, and generated Thiessen polygons (Voronoi, 1908, Thiessen & Alter, 1911) based on these discrete locations (Fig. 3b). Topological relationships between the Thiessen polygons allowed us to construct spatial networks for different levels of neighborhood cardinality which can be understood as varying scales of spatial context (Fig. 3c,d). The concept of neighborhood cardinalities is used to complement the Euclidean distance-based measures and allows us to identify neighborhood relationships between places independent from spatial density variations. See Appendix 3 for more details on the spatial network creation.

2.3.2. Spatial-network based remoteness indicators

Similar to the focal population density metric, based on a fixed focal radius used in the raster-based approach, we modeled ***population density*** in local, topology-based neighborhoods: For each place $\boldsymbol{p}$, we identify the neighboring places of a given cardinality ($\boldsymbol{p_c}$) and calculated the total population $\boldsymbol{s_c}$ in the neighborhood. To calculate the approximate population density, we use a square of size nxn with n being the largest occurring distance $\boldsymbol{dist_{max}}$ between $\boldsymbol{p}$ or any of the neighboring places $\boldsymbol{p_{c1}}$. Thus, the neighborhood population density (***NPD***) for any cardinality $\boldsymbol{c}$ of place $\boldsymbol{p}$ and year $\boldsymbol{t}$ can be calculated as:

$$NPD_{p,t,c} = \frac{\left[s_{p,t} + s_{c,t}\right]}{dist^2_{\max}} \tag{4}$$

To capture local population density at different levels of spatial context around each place, we calculated the ***NPD*** for the cardinalities 1, 2 and 3. Moreover, we adopted a measure of ***local significance***, which was proposed by Esch et al. (2014). The local significance ***LS*** can be calculated for each edge of a spatial network, based on the length of the edge $\boldsymbol{d}$ (i.e., the Euclidean distance between neighboring places) and the size $\boldsymbol{s}$ (i.e., population) of the nodes (places $\boldsymbol{i}$ and $\boldsymbol{j}$) connected by the edge, as:

$$LS_{i,j} = \frac{\left[s_i \times s_j\right]}{\left|d_{i,j}\right|^2} \tag{5}$$

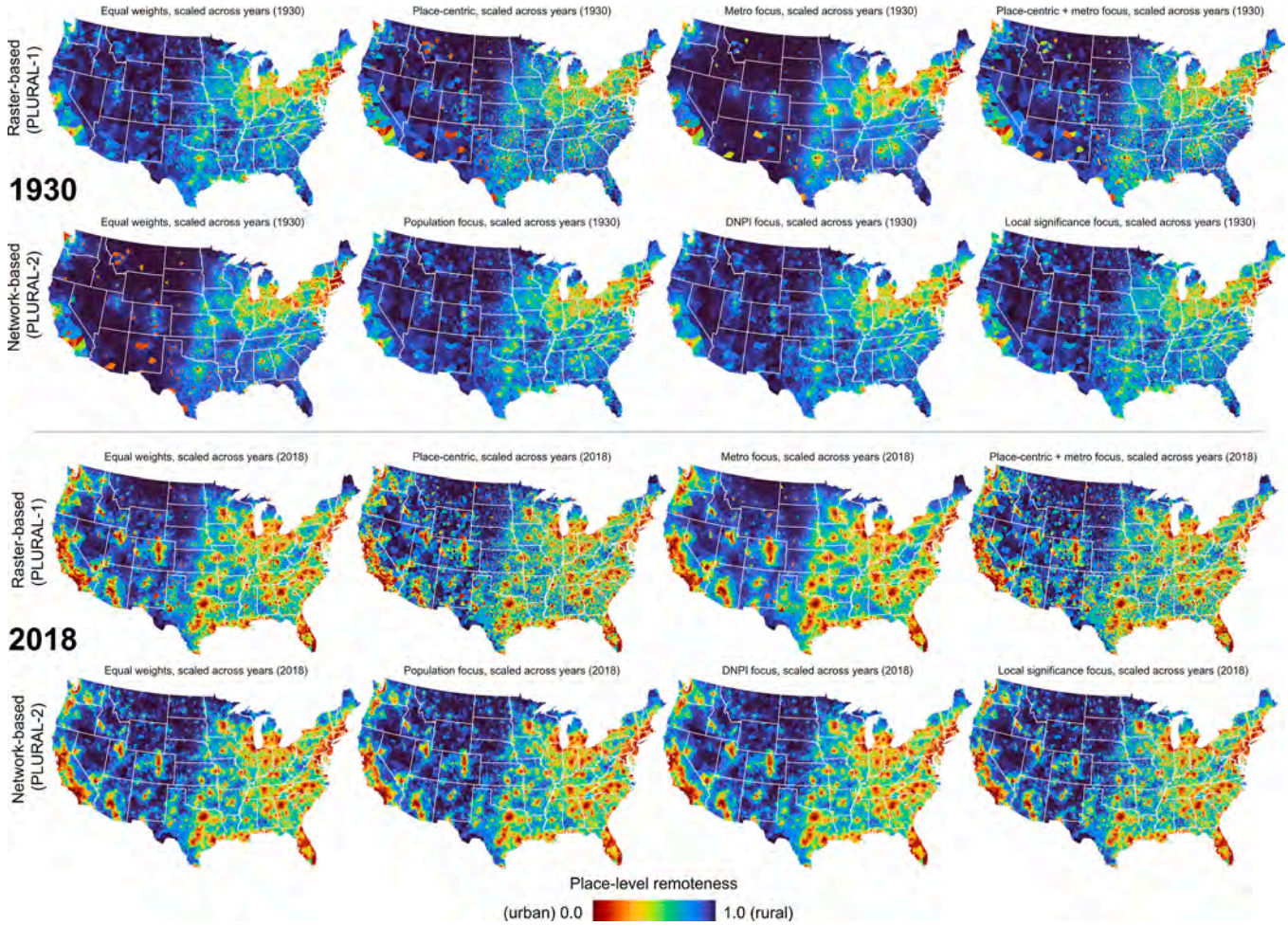

Fig. A5-1. Maps of the remoteness indices in 1930 (top) and 2018 (bottom), each shown for the raster-based approach and network-bases approach, and for each of the four weighting schemes. In these maps, each place is represented by the Thiessen polygons established from the discrete place locations.

Similar to Esch et al. (2014), we calculated the median local significance $\boldsymbol{MLS_C}$ for each place $\boldsymbol{p}$ based on the values obtained for each edge connected to the node that represents the place $\boldsymbol{p}$. Accordingly, we calculated this metric for each place within neighborhoods of cardinality 1, 2 and 3. The MLS yields high values if a place is large and located near other large places, and yields small values if the neighboring places of a given small place are small and distant.

In the raster-based remoteness modelling approach we used the distance to the nearest place of a given population category $\boldsymbol{D_{PC}}$ as a measure of remoteness. However, such a metric ignores the spatial configuration of places below the chosen population threshold that could potentially be located between the place under study and the nearest place of a given population range. For example, a place (A) located 50 km from the nearest city > 50,000 inhabitants (B) would receive the same value for $\boldsymbol{D_{PC}}$, regardless if the area between A and B is completely uninhabited (scenario 1) or if that area contains many small places below the lowest population threshold, and outside the focal window size used for the focal population density calculation (scenario 2). The degree of remoteness of place A should be higher in case of scenario 1 than in scenario 2.

Motivated by this shortcoming of the $\boldsymbol{D_{PC}}$ metrics, we adopted the concept of proximity and isolation metrics, a subgroup of landscape metrics commonly used in landscape ecology and habitat fragmentation analysis (e.g., Bender, Tischendorf, & Fahrig, 2003) to quantify the degree of subdivision of a landscape or of the isolation of specific components (e.g., land cover classes) within a landscape. More specifically, we adopt the concepts of the "degree of landscape division" metric (DIV), proposed by Jaeger (2000) and distance-weighted landscape variables (see Miguet, Fahrig, & Lavigne, 2017). Jaeger (2000) defines the DIV as the area under the curve when sorting patches in a given landscape by their patch area, as a measure of the graininess of a landscape (McGarigal, 1995). Based on this, we designed a place-centric, distance-based metric quantifying the relationships of neighboring place populations and their distances to the "focal" place (i.e., the place under study) in a single metric. Our method identifies the neighbors of a given focal place, either using a Euclidean distance or a topology-based neighborhood criterion, and sorts the neighboring places, including the focal place itself, ascendingly by their distances to the focal place. Then, the cumulative population curve is calculated *over the distance-sorted places*, and finally, the area under the cumulative population curve is obtained. This area under the curve (AUC) represents a metric characterizing the spatial distribution of populated places in dependence of the distance to the focal place. We call this metric the **distance-based neighborhood population index (DNPI).**

In order to make this metric comparable between different focal places, the maximum distance $\boldsymbol{d_{MAX}}$ needs to be specified, as well as a maximum value for the cumulative population $\boldsymbol{CP_{MAX}}$, and the curves need to be scaled by $\boldsymbol{d_{MAX}}$ in x-direction and by $\boldsymbol{CP_{MAX}}$ in y-direction, respectively, so that it is normalized into the range [0,1]. If the curve exceeds $\boldsymbol{CP_{MAX}}$ before $\boldsymbol{d_{MAX}}$ is reached, the cumulative population curve is "trimmed" to 1.0. Thus, the maximum possible AUC is 1.0 for a place of population >= CP_{MAX}. We calculated the DNPI within the neighborhood of cardinality 3, and for a range of distance-population combinations. Table 1 summarizes the network-based metrics of remoteness

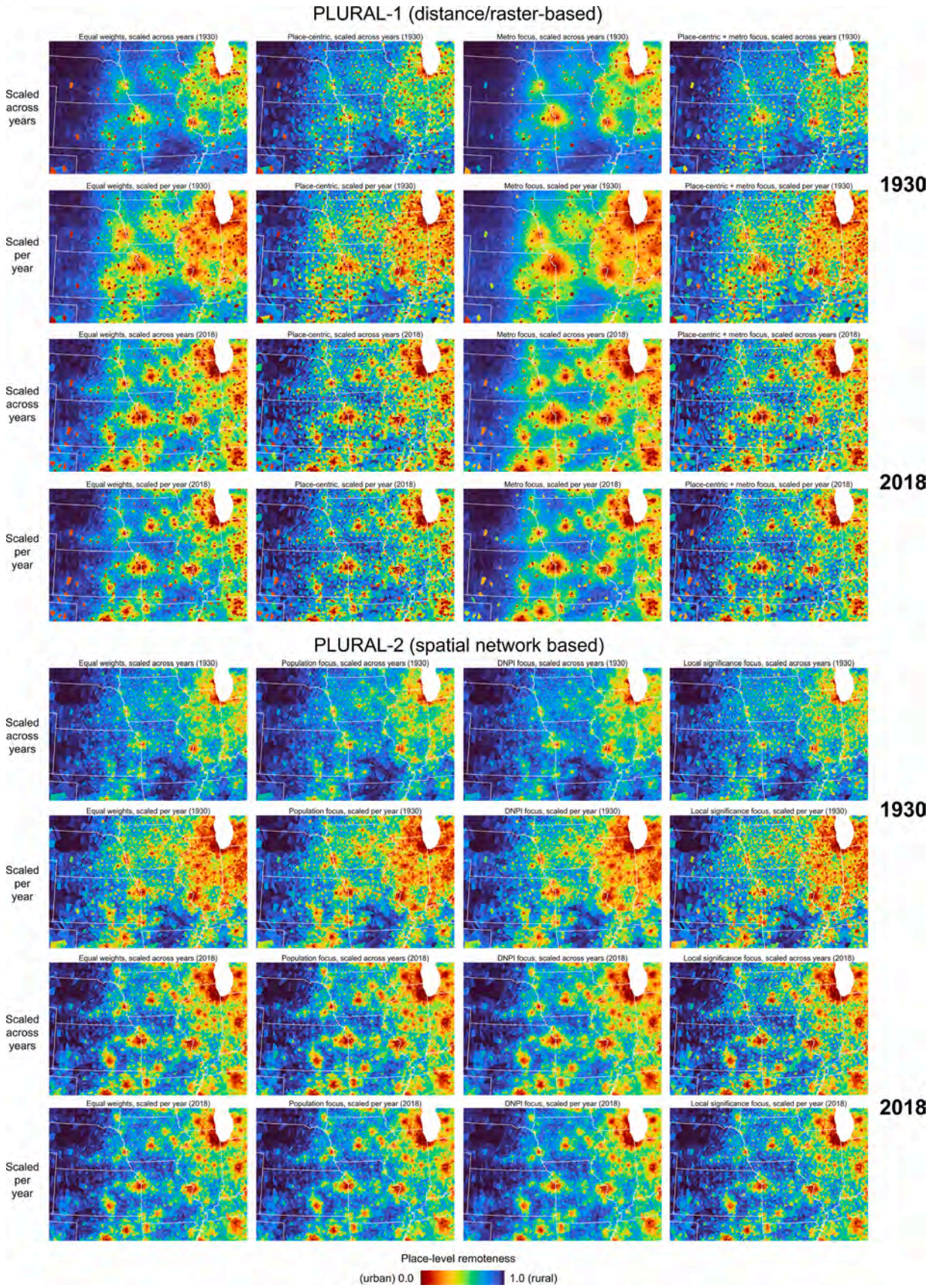

Fig. A5-2. Maps of the remoteness indices in 1930 and 2018, shown for a subset of the data covering parts of the Midwest, shown for the raster-based approach (PLURAL-1) and network-bases approach (PLURAL-2), scaled both across time, and per year, and shown for each of the four weighting schemes. In these maps, each place is represented by the Thiessen polygons established from the discrete place locations.

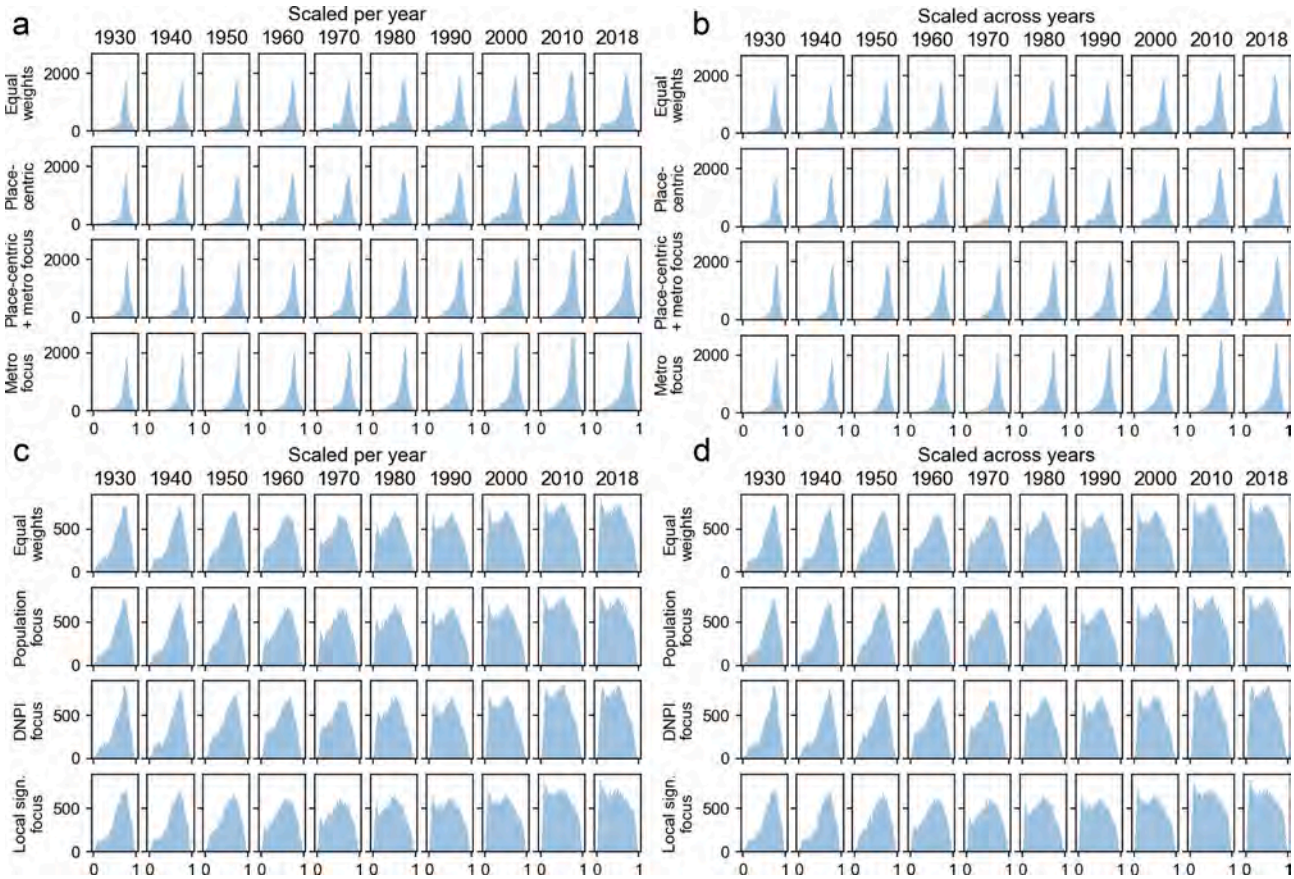

Fig. A5-3. Distributions of remoteness indices for the US 1930–2018 for different weighting schemes and scaling strategies: (a) Raster-based approach (PLURAL-1) scaled per year and (b) scaled across all years, (c) spatial network based approach (PLURAL-2) scaled per year and (d) scaled across all years.

used herein.

The maps of these 11 metrics for the 2018 places are shown in Fig. A3-1. For each of these metrics, we calculated the ranks in a descending order (i.e., the lowest magnitude receives the highest rank), and use these remoteness indicators as input for different weighted averages. We implemented four exemplary weighting schemes (see Appendix 3). Herein we focus on an equal weights scenario to generate the PLURAL-2 for each US census place in the years 1930–2018. Importantly, we calculated both indices using a) annual scaling to the range [0,1], and b) scaling across all years 1930–2018, in order to generate temporally comparable remoteness indices for longitudinal studies. Note that the network-based indices PLURAL-2 are based on a ranking strategy, which makes the indices not directly comparable over time, and thus should not be used for longitudinal analysis.

2.4. Evaluation strategies

We compared our raster-based (PLURAL-1) and spatial network based (PLURAL-2) remoteness indices to a variety of external, independent datasets that are coherent or correlated to the concept of remoteness, or that are assumed to follow systematic patterns across the rural–urban continuum. Note that a rigorous validation is not possible, as the concept of remoteness is not a perfect substitute for rurality or urbanness, and existing data are not generally measured at the scale of places. Further, definitions of remoteness likely have many truths, much as there are many definitions of urban and rural. More specifically, we implemented the following evaluation strategies: We compare the PLURAL-1 and PLURAL-2 indices to existing county-level rural–urban classifications, at both the native place-scale and aggregated to the county, and we compare both indices against each other, to quantify the effects of the different modelling approaches. Additionally, we use the GHS-SMOD data (Florczyk et al., 2019) to test our approaches against a global, remote-sensing derived urban–rural classification, and conduct a visual comparison to global population data (i.e., Worldpop, WorldPop, 2018), in order to assess the differences between the concept of remoteness and population density. Moreover, we compare our created place-level remoteness indices against travel-time based accessibility indicators (Nelson et al., 2019), and conducted a sensitivity analysis of the PLURAL-1 to the chosen population thresholds and focal window size. Finally, we compare our multi-temporal results to historical settlement trends derived from the Historical Settlement Data Compilation for the US (HISDAC-US, Leyk & Uhl, 2018, Uhl et al., 2021) and against landscape metrics derived from Microsoft's building footprint data (Microsoft 2018). The variety of datasets used for comparison is illustrated in Fig. A4.

3. Results

3.1. The rural–urban continuum in the US from 1930 to 2018

Fig. 4a,b shows the spatial distribution of the PLURAL indices for the two modelling approaches, and for the equally weighted scenarios, for 1930 and for 2018 (see Fig. A5-1, A5-2, and A5-3 for maps of all weighting schemes and for data distributions over time). Indices are scaled into the range [0,1] jointly across all years, and thus, the obtained indices are comparable over time. The detailed dynamics of the rural–urban continuum from 1930 to 2018, as modelled by these approaches can be seen in Supplementary Movie 1 and in Supplementary Movie 2. The temporal trends of places per remoteness stratum (Fig. 4c) illustrate the multi-temporal dimension of the PLURAL indices.

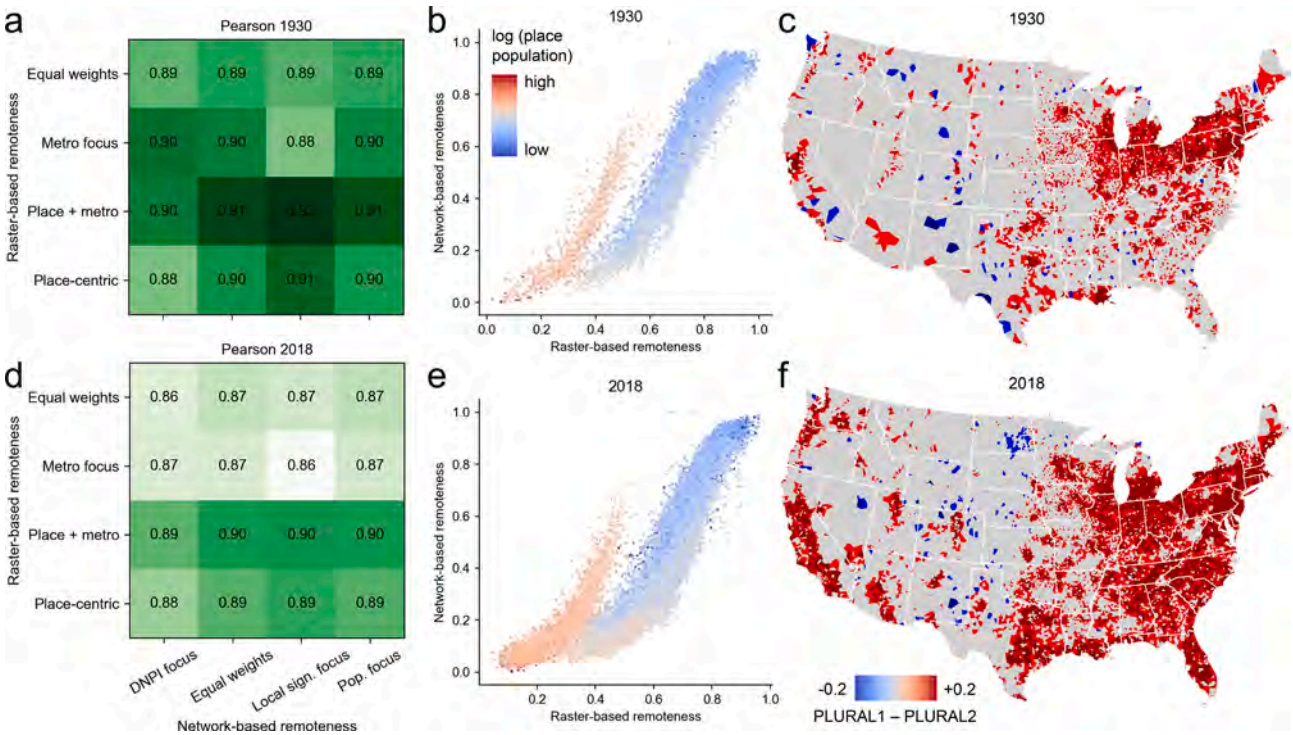

Fig. A-61. Comparison of raster-based and network-based remoteness indices: (a) Pearson's correlation coefficients between the four raster- and network-based weighting schemes, (b) Scatterplot of the equally weighted raster- and network-based remoteness indices in 1930, color-coded by place population, and (c) corresponding map of the differences in remoteness. Panels (d) to (f) show the corresponding results for 2018. In maps (c) and (f), places are represented by the Thiessen polygons established from the discrete place locations. All reported correlation coefficients have a p-value < 0.05 and thus, reported correlations are statistically significant (see Supplementary File 1).

3.2. Distance-based neighborhood population index

One of the components of the network-based remoteness index is the distance-based neighborhood population index (DNPI) proposed herein, which is calculated as the AUC of the cumulative population curve established for the distance-sorted nearest neighboring places. As a proof of concept, we show the cumulative, distance-sorted population curves for 2010 census places in the US, located in urban counties (RUCC 1, Fig. 5a), in *peri*-urban counties (RUCC 5, Fig. 5b) and rural counties (RUCC 9, Fig. 5c). If the focal place is near a large place, the cumulative population curve will increase sharply and yield a large area under the curve (Fig. 5a). Conversely, if a place is small and its neighbors are small too, the cumulative population curve increases only slowly, yielding a small AUC. This visual assessment clearly confirms that the DNPI responds to the degree of urbanness.

3.3. Comparison of existing rural–urban classifications in the conterminous US

We compare the previously discussed, existing county-level classifications to each other and to the PLURAL indices (Fig. 6a-f), and compare them to the PLURAL indices, aggregated to the county level (i.e., average place value per county, Fig. 6g,h). Despite different data sources and methodological approaches, these six classification schemes show generally high levels of correlation (Fig. 6i). The lowest correlation is between IRR and RUCA (Pearson = 0.59) likely driven by the different data sources (i.e., built-up density vs. commuting patterns). Moreover, the continuous IRR exhibits a positive association with the rank-based metrics, which is, on average, strongest and almost linear between IRR and URCS; it shows the least nuanced trend between IRR and the tract-level RUCA (Fig. 6j). The PLURAL indices aggregated to the county show highest correlations to IRR (due to the continuous nature) and, similarly to the IRR, the lowest correlation to RUCA.

3.4. Further evaluation of the PLURAL indices

Our comparisons reveal high levels of agreement and plausibility between the place-level PLURAL indices and other alternate classifications and related data. While we refer to Appendix 6 for the detailed assessments, Table 2 briefly summarizes the major findings from each of our evaluation analyses.

4. Discussion

Our results indicate that the search for a single "correct" approach to measuring the rural–urban continuum will likely remain out of reach. This is evident from the varying levels of agreement (Pearson's correlation coefficients between 0.59 and 0.9) between county-level rural–urban classifications in the United States (Fig. 6). We have presented two modelling approaches to create place-level rural–urban classifications that allow greater flexibility and investigation into rural and urban processes – and the spaces in between – across the United States over time. Our approaches yield similar broad-scale patterns, and reflect the commonly known settlement trends in the US during the 20th century (i.e., early settlements in the Northeast, and late, fast-growing urban areas in the South).

The main differences between our indices appear at finer scales, particularly earlier in the urbanization history of the United States (Fig. 4). While the raster-based approach identifies early settlements such as Tucson (Arizona), Santa Fé (New Mexico), Albuquerque (New Mexico) as highly urban (i.e., low remoteness), such extreme local differences are not present in the 1930 network-based result, which appears to be more sensitive to smaller places of local or regional importance (e.g., lower remoteness levels along historical trade and settlement routes such as the Oregon trail, Fig. 4). When comparing place-level remoteness over time, however, PLURAL-1 and PLURAL-2 reveal consistent trends, both documenting the impressive transition

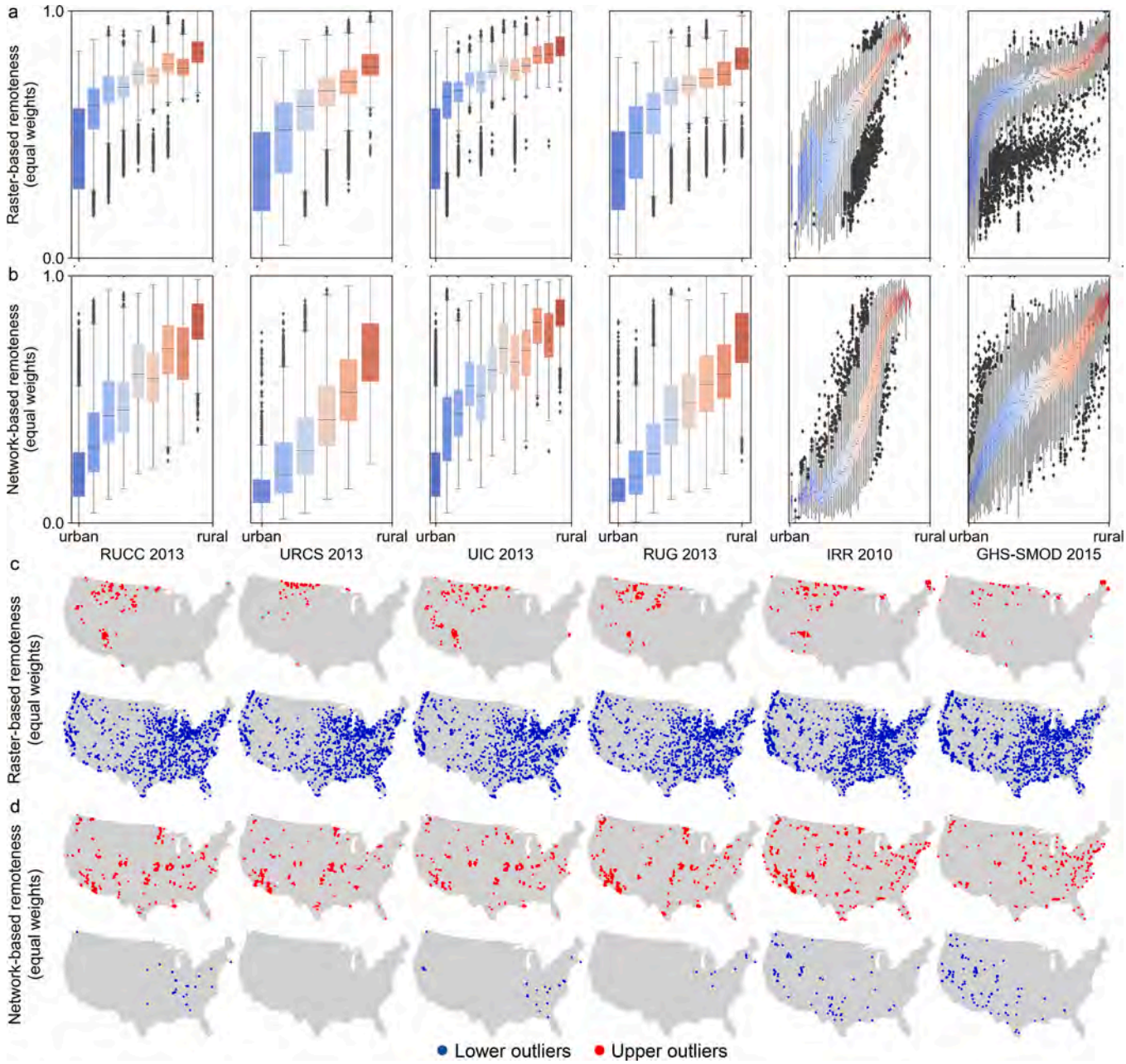

Fig. A6-21. Comparison of the place-level remoteness indices and their county-level rural–urban designations. (a) Raster-based remoteness distributions and (b) network-based remoteness distributions in county-level rural–urban classes, both shown for the equal-weights scenario. Panels (c) and (d) show the spatial distributions of the upper and lower outliers indicated in the box-and-whisker plots, for the raster- and network-based remoteness, respectively. Coloring of the boxes in (a) and (b) corresponds to the classes / values of the county-level classifications (blue = urban, red = rural).

from a large number of rural places in 1930, which recede over time, to a greatly expanded number of urban places by 2018 (Fig. 4c).

Despite qualitatively in a macro sense, the gradients of "urbanness" from small to large places differ between the two approaches (Fig. A6-1), similar to the differences between county-level rural–urban classifications available in the literature (Fig. 6). Neither is necessarily more correct: these differences reflect the definitional and multivariate challenges inherent in measuring and modelling the rural to urban continuum.

The cross-comparison of the created indices against each other, and against several external datasets of different nature, revealed that our indices, despite being based on simple data structures, effectively model the rural–urban gradient patterns. Moreover, our analyses have shown that the network-based remoteness index (PLURAL-2) seems to exhibit higher levels of coherence to the data used for comparison than the raster-based approach (PLURAL-1). However, these differences largely play out at the urban side of the rural–urban continuum, i.e., mostly affecting large places. The indices and weighting schemes have strengths in different areas and we recommend the following: For applications of the PLURAL indices ***identifying rural places***, both methods seem to work well. For studies ***across the whole urban–rural continuum***, we recommend practitioners to use PLURAL-2, as it is more normally distributed across the RUC (see Fig. A6-2, A6-3). For ***longitudinal studies***, the PLURAL-1 (scaled across all years) is more suitable, since the network-based PLURAL-2 contains a rank-based component and thus, is not fully comparable over time. The choice of the weighting scheme depends on the individual application. Importantly, our analysis of place-level estimates of "urbanness" within strata of county-level

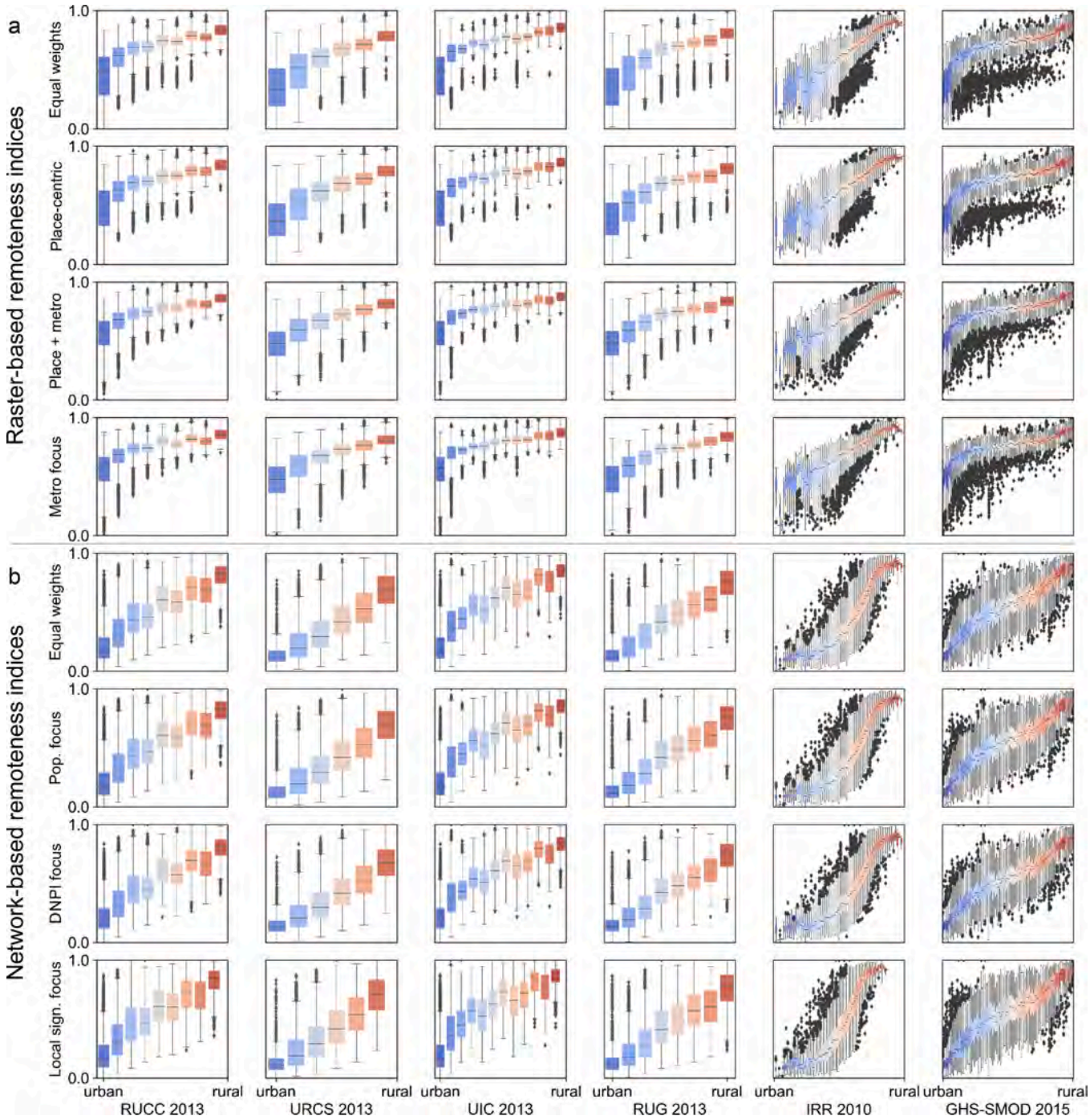

Fig. A6-22. Comparison of county-level rural–urban classifications and the place-level remoteness indices for each weighting scheme of the (a) raster-based approach, and (b) the network-based approach. Color of boxes represent the individual classes / values of the county-level classifications (blue = urban, red = rural).

rural–urban classes revealed considerable dispersion (as measured by the presence of outliers), following strong geographic patterns (Fig. A6-2c,d). In other words, the spatial refinement in using place-level remoteness instead of county-level estimates appears to be particularly effective in understanding the rural–urban patterns at the local scale.

In this work, remoteness is modelled based on Euclidean distance, rather than travel time or road network distances. While this simplistic approach may locally bias our results, we found high levels of agreement of our PLURAL indices with road network distance based FAR remoteness measures (Fig. A6-4). We also found plausible results when testing our indices against travel time-based accessibility indicators (Fig. A6-5), indicating that for most parts of the CONUS, the bias introduced by using Euclidean distance instead of road network distance is within acceptable margins. As mentioned before, the observed disagreement of our indices as compared to road network-based distance measures is, in part, likely due to the differences in the underlying analytical units and their spatial granularities (i.e., census places vs. cities, vs. ZIP code areas). Comparing our multi-temporal place-level urban–rural classifications to historical land development trends derived from HISDAC-US, we find high correlations in recent decades which decrease for earlier points in time (Fig. A6-6). While the discussion of the drivers for these changes over time is out of the scope of this paper, a possible reason could be lower levels of completeness in the HISDAC-US and underlying ZTRAX data in early points in time (Uhl et al., 2021).

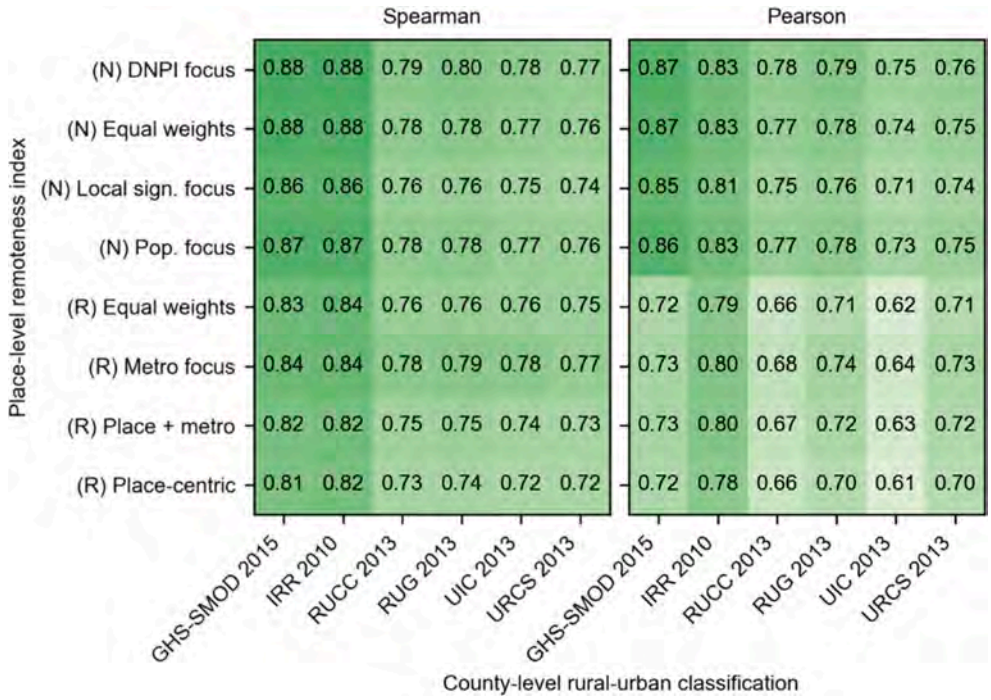

Fig. A6-23. Spearman and Pearson correlation coefficients between the place-level remoteness indices (R = raster-based, N = network-based) and the corresponding county-level rural–urban designations. All correlation coefficients have p-values < 0.05 and thus, the reported correlation coefficients are statistically significant (see Supplementary File 1).

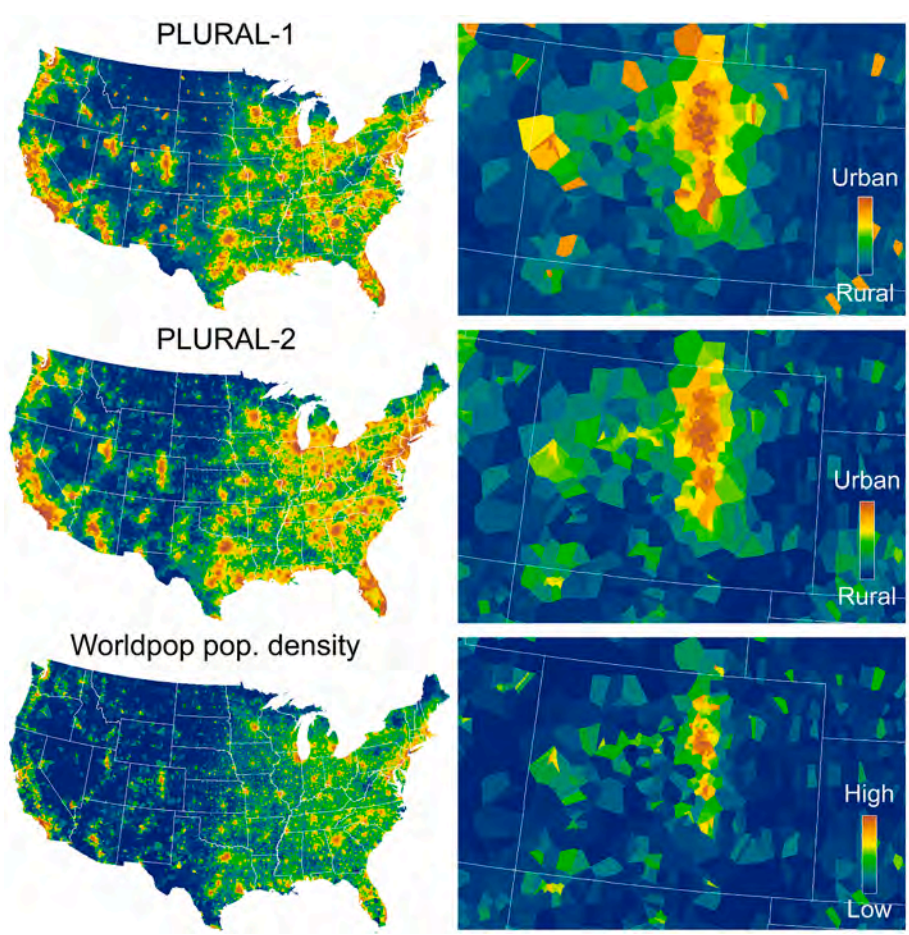

Fig. A6-3. Visual comparison of place-level remoteness indices PLURAL-1 and PLURAL-2 (equal weighting schemes) to population density from Worldpop, shown for the year 2018. Gridded population estimates from Worldpop were aggregated to the Thiessen polygons representing each place.

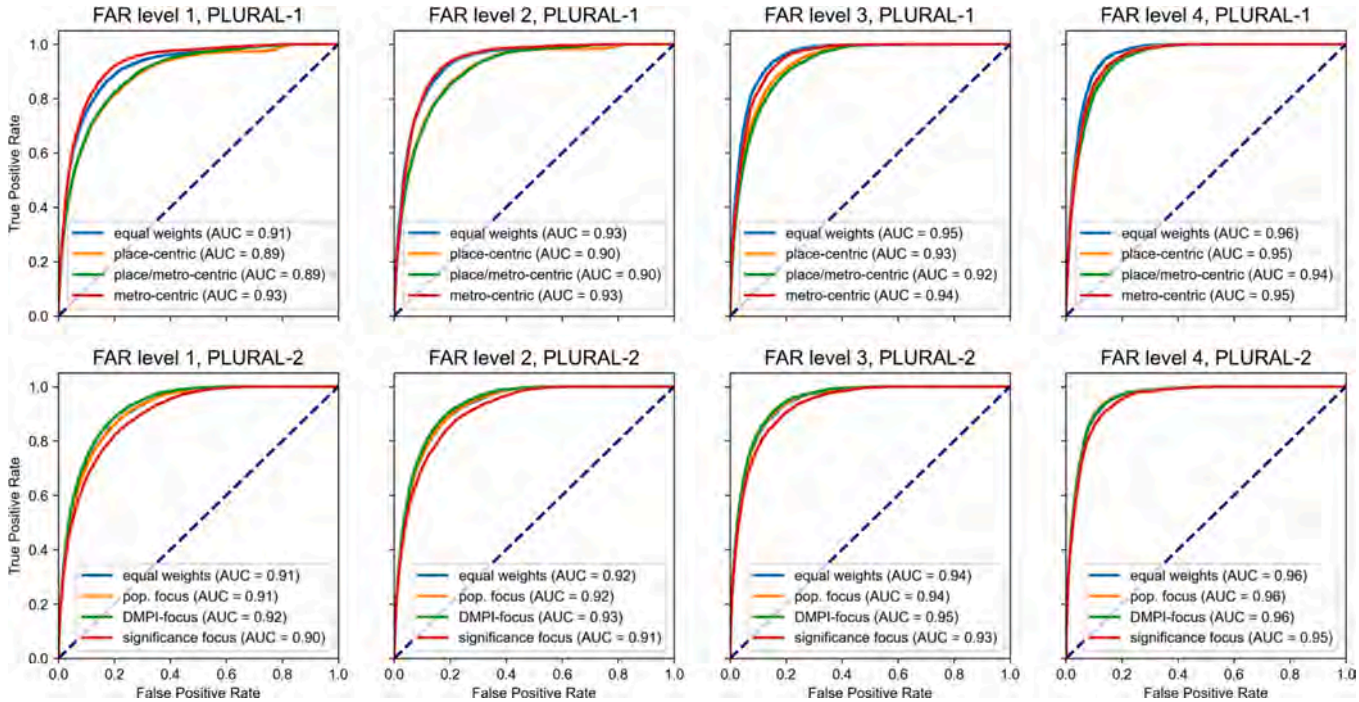

Fig. A6-4. Receiver-operator-characteristic (ROC) analysis to test how well the place-level remoteness indices can mimic fine-grained, travel time based remoteness classes from the Frontier and Remote area (FAR) codes dataset, shown for FAR classes 1–4, and for each weighting scheme of the raster-based (top row) and network-based approach (bottom row). The dashed line represents the ROC curve of a random relationship between the remoteness indices and FAR class memberships for comparison.

Moreover, the PLURAL indices, despite their simplistic input data, are strongly correlated with size and structure-related place characteristics, as measured by means of landscape metrics of fine-grained developed / built-up areas (Fig. A6-7). These findings confirm once more that our remoteness models are consistent with the literature and exhibit an expected behavior when compared to landscape metrics. Finally, our sensitivity analysis (Fig. A7-1) revealed high levels of robustness of the PLURAL indices to the choice of user-defined parameters such as population thresholds or focal window size. These low sensitivity levels are largely consistent over time, and the Q-Q plots in Fig. A7-1 indicate that the choice of these parameters does not affect our results in a qualitative way (i.e., the order of places according to their level of "urbanness" is stable) but may slightly affect the numerical results. This appears acceptable, however, given the variability of quantitative representations of the existing rural–urban continuum models at the county level (Fig. 6). Furthermore, our indices are largely invariant to the spatial resolution of the underlying raster data (Appendix 8), and the weighting schemes implemented in both PLURAL indices appear to be effective (Appendix 9).

Based on these findings, we argue that scholars, planners, or policy makers dealing with social or natural processes that unfold locally could benefit from the examination of the fine-grained spatio-temporal variations of urbanness, as modelled by our PLURAL indices. The many possible use cases are revealed in work on social mobility, rural and child poverty, economic development, and climate resilience (Partridge, Rickman, Olfert, & Tan, 2015, Meerow & Newell, 2019, Chetty, 2021, Bardin & Kedron, 2022), where place-based policies are a rising priority for policy-minded scholars and planners. Federal funding, action and coordination around these problems often occur at coarse political scales like states, counties and cities. It is prescient, however, to consider the unique challenges faced by urban and rural communities, even among those within the same locales. The PLURAL indices provide a flexible approach from which to identify where smaller communities sit on the rural to urban continuum, to assess how that situation is likely to change in the future, and to calibrate place-based polices appropriately to these conditions. The PLURAL indices can thus be a tool for identifying granular community distinctions that are of immediate relevance to place-based policy and planning, within larger political jurisdictions.

Related benefits to the PLURAL approach extend to the temporal and spatial identification of rural–urban transformations (Dudwick, 2011). These transformations tend to occur at the interface of agricultural and industrial land uses, with major impacts for energy consumption, CO_2 emissions, as well social issues around economic inequality (Zhou et al. 2015, Henderson & Wang, 2005). In particular, the conversion of an agricultural economy to a rural non-farm economy is key for sustainable *peri*-urban development (Zhou et al. 2015), but these transition zones also require effective policies for harm- and poverty-mitigation, particularly in developing countries (Christiaensen & Todo, 2014). Such changes are also often accompanied by spatial demographic changes with respect to aging, migration, and population growth (OECD/PSI, 2020), as well as local land-use and land cover transitions (McGee, 2008). Each of these issues could be better addressed and planned for through an expanded knowledge of transformations taking place at the intersections of rural and urban contexts.

The PLURAL indices will enable researchers to work at spatial units that are potentially more meaningful to certain rural processes such as the role of place characteristics in social mobility (Chetty et al., 2014; Connor & Storper, 2020), health (Manduca & Sampson, 2019; Shah et al., 2020), and voting (Sachdeva et al., 2021). Such refined scales can be also used to gain novel insight into the spatial distributions of social vulnerability (Spielman et al., 2020), public health issues (Andersen, Harden, Sugg, Runkle, & Lundquist, 2021) , or the exposure to natural hazards risks (Braswell, Leyk, Connor, & Uhl, 2021). Moreover, the presented indices cover a long time period, 1930 to 2018, and are fully consistent over time (PLURAL-1), enabling longitudinal analyses of long-term, dynamic processes along the rural–urban continuum. Some exemplary research questions where the PLURAL indices could directly be employed, include: How has the rural–urban continuum in the US evolved over time? How does it co-vary with other long-term spatial processes (e.g., historical land use / land cover change? Can local, long-

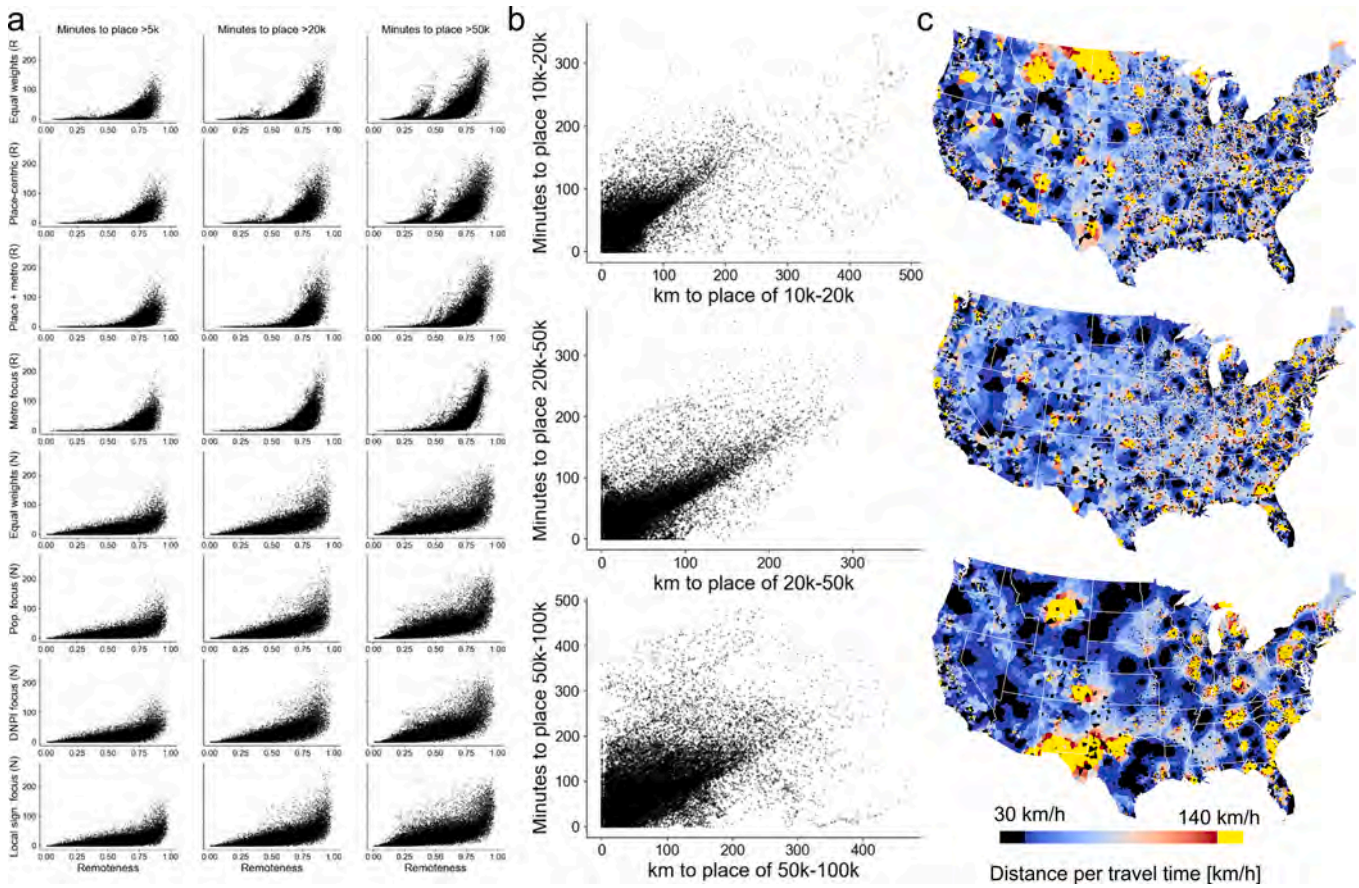

Fig. A6-5. Comparing the remoteness indices and their components to travel-time based accessibility indicators (Nelson et al. 2019). (a) Relationships between the different weighting schemes of the proposed raster- and network-based remoteness indices and travel times to the nearest city of population > 5 k, > 20 k, and > 50 k, respectively, (b) relationships between Euclidean distance based components of the raster-based remoteness approach and travel times, and (c) place-level plausibility analysis of Euclidean distance and travel time by visualizing theoretical travel speed. In (c), places are represented by the Thiessen polygons established from the discrete place locations.

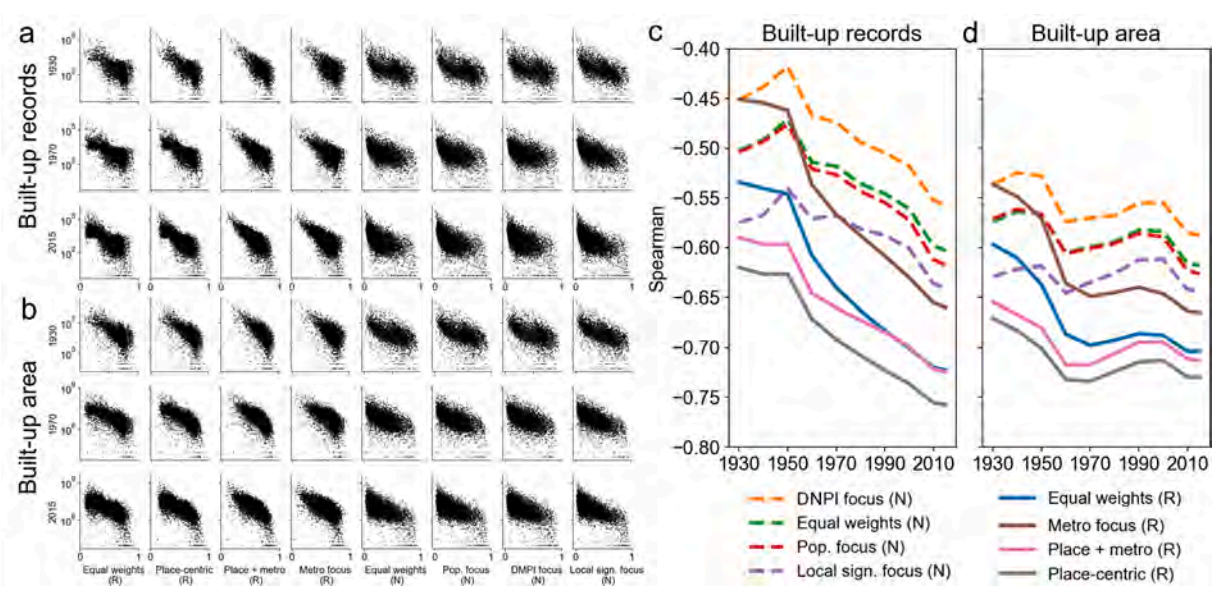

Fig. A6-6. Relationships of ZTRAX-based built environment characteristics and the proposed remoteness indices based on the raster approach (R) and the network approach (N) over time. (a) Scatterplots of place-level remoteness and building counts, and (b) built-up area in 1930, 1970, and 2015, respectively; Panel (c) and (d) show corresponding time series of Spearman's correlation coefficient. Note that 2015 built environment characteristics are compared to the 2018 remoteness indices. All correlation coefficients have a p-value < 0.05 and thus, reported correlations are statistically significant (see Supplementary File 1).

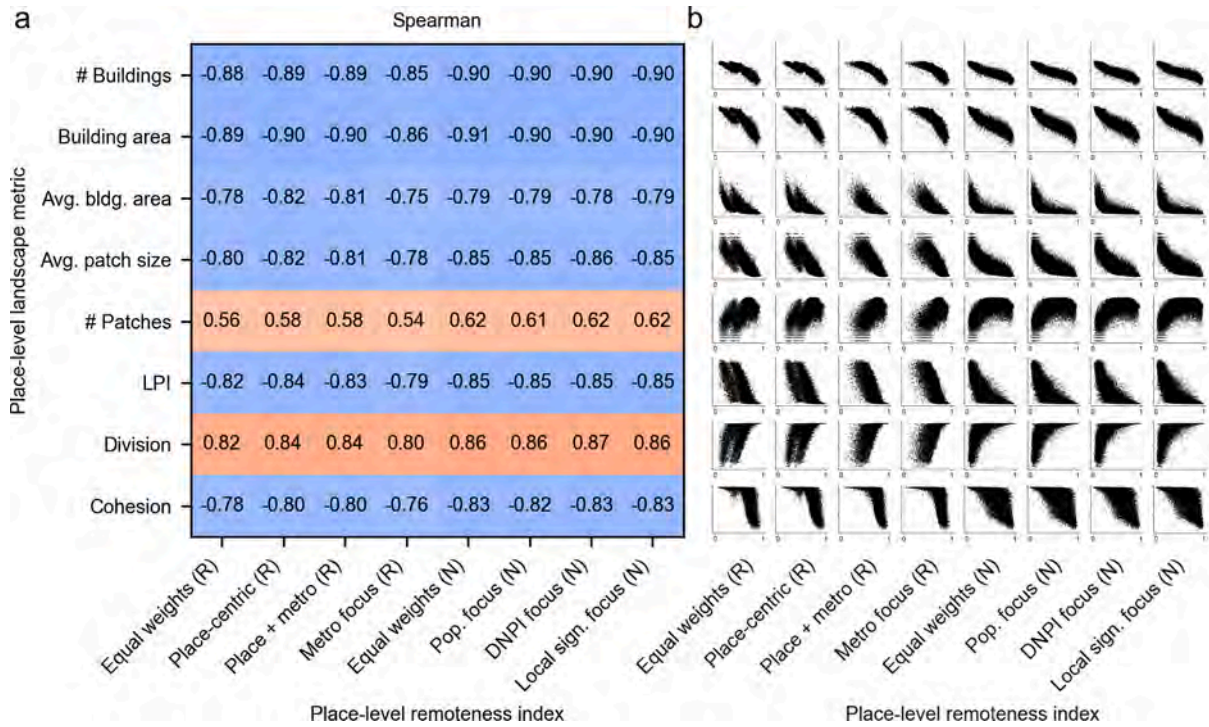

Fig. A6-7. Assessing the relationships between the remoteness indices and contemporary building and landscape metrics derived from Microsoft building footprint data. (a) Spearman's rank correlation, and (b) corresponding scatterplots. All correlation coefficients have a p-value < 0.05 and thus, reported correlations are statistically significant (see Supplementary File 1).

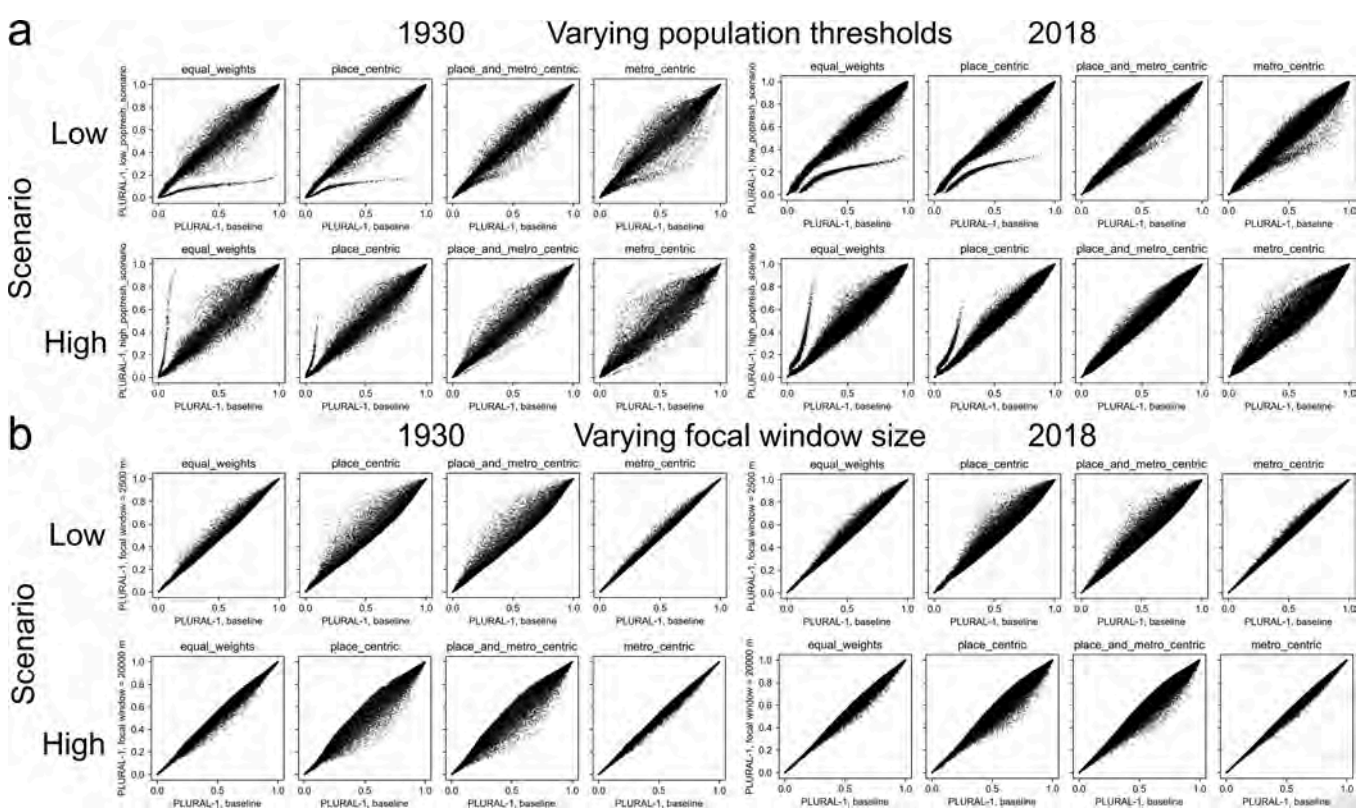

Fig. A7-1. Sensitivity analysis of PLURAL-1 to user-set parameters. Shown are QQ-plots of the PLURAL-1, comparing baseline parameters against a low and high value scenario, for (a) populations thresholds and (b) focal window size for population density estimation, each in 1930 and 2018. Baseline parameters for population thresholds and focal window radii are shown on the x-axis, and scenarios with parameter values lower and higher than the baseline, for both 1930 and 2018, on the y-axis. The degree of disorder (i.e., deviation from the main diagonal) indicates the level of sensitivity of the PLURAL-1 classification to these parameters.

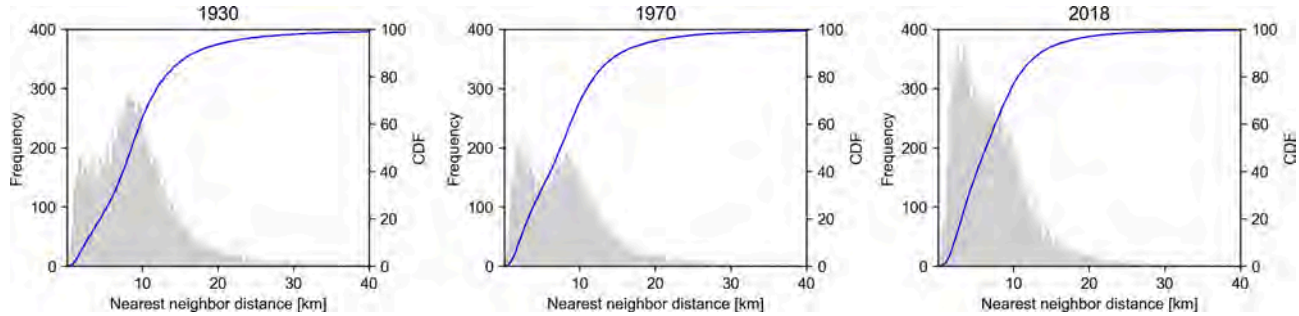

Fig. A8-1. Distributions of nearest-neighbor distances between US places in 1930, 1970, and 2018. Shown are the histograms (grey) and the cumulative distribution function (CDF).

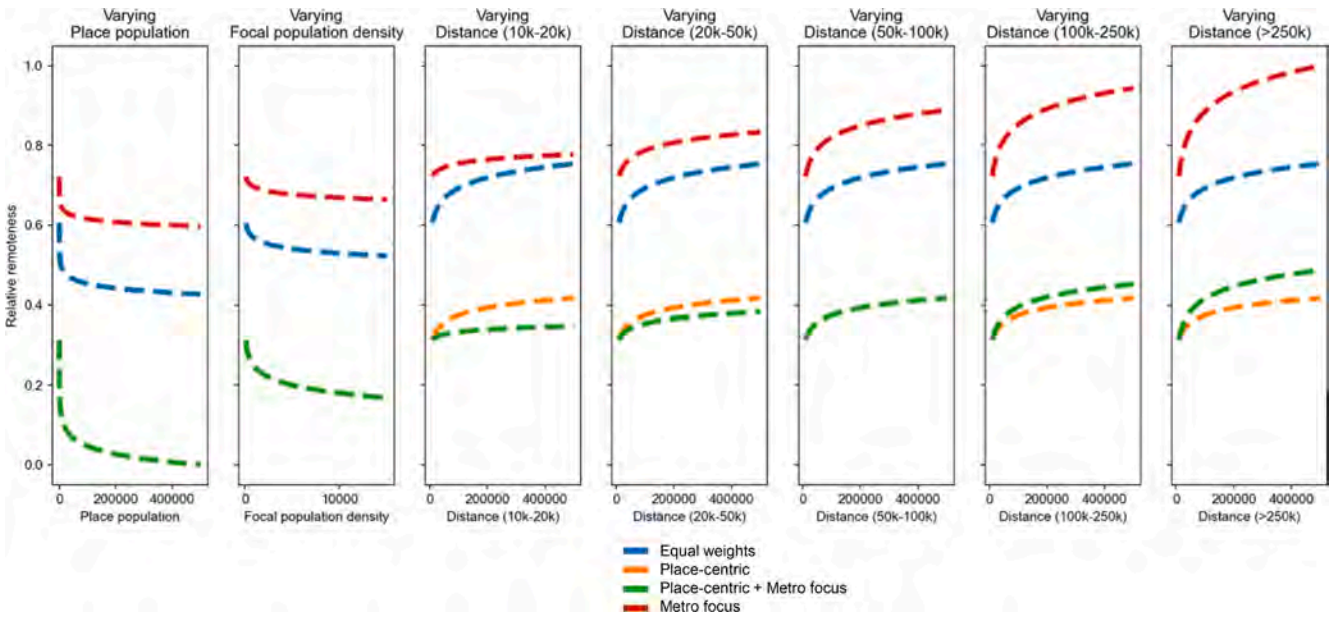

Fig. A9-1. Sensitivity of different weighting schemes to changes in individual remoteness indicators for the raster-based approach. A synthetic baseline place with randomly initialized remoteness indicator was used, and each of the seven indicators was systematically increased. For each increment, the resulting remoteness index was calculated, for each of the four weighting schemes. The curves show the different levels of sensitivity of the weighting schemes.

term trajectories of "rurality" tell us something about contemporary, physical or social characteristics of places?

Users should be aware, however, that any description of the rural–urban continuum, irrespective of the unit of analysis (e.g., county, place, grid cell), rest on analytical and theoretical assumptions that get operationalized in subjective fashion. While there is therefore no entirely correct way to model the rural to urban continuum, we recommend that: 1) analysts consider the scale of the process that they are studying and choose rural and urban definitions that are well aligned with that scale, and 2) all analyses include sensitivity checks based on different definitions of the rural to urban continuum, many of which we have reviewed here.

Despite our creation of a continuous fine-scale measure of the rural–urban continuum, there is often conceptual and analytic value in categorical, and even binary, distinctions between rural and urban contexts. From a practical perspective, it is often necessary to conduct statistical tests with a limited sample size or given certain model specifications (e.g., fixed effects), which may necessitate a binary or categorical approach to rural and urban classification. From the perspective of theory, it has been well noted that even though it is impossible to draw a definite line between cities and their rural hinterlands, rural and urban contexts tend to exhibit strongly contrasting qualities (Scott & Storper, 2015). While changes in the lived experience of people along the continuum can be gradual in some contexts, they can be quite sharp in others. Our transparent methodology and publication of the PLURAL indices provide strong future opportunities for assessing the nature of these gradients and possible "break points" between rural and urban places.

Despite being a useful addition for many applications, the PLURAL indices face some noteworthy limitations. For example, the raster-based approaches are based on population density calculated within focal windows of a fixed radius and based on commonly used population categories. If these indices were to applied to other geographic regions, sensitivity is needed with respect to the size of the focal windows and population thresholds. Moreover, the PLURAL indices do not account for territory and populations outside of bounded census places, and thus, cannot be used for assessments of spatial processes occurring on unincorporated land outside of official place boundaries. As noted above, however, we contend that proximate places are of relevance to non-place-based populations as a hub of local social and economic activity. From this, we infer that the level of rurality assigned to a place is usually a strong reflection of the rurality of surrounding areas. Furthermore, in order to apply these methods to other countries, border effects need to be taken into account, given the potentially high levels of cross-country mobility (e.g., in Europe). The role of *peri*-urban open spaces (e.g., parks, natural reserves; Žlender & Thompson 2017, Wandl & Magoni, 2017) used as recreational areas needs to be investigated, as such uninhabited areas may affect our indices, while their impact on the level of

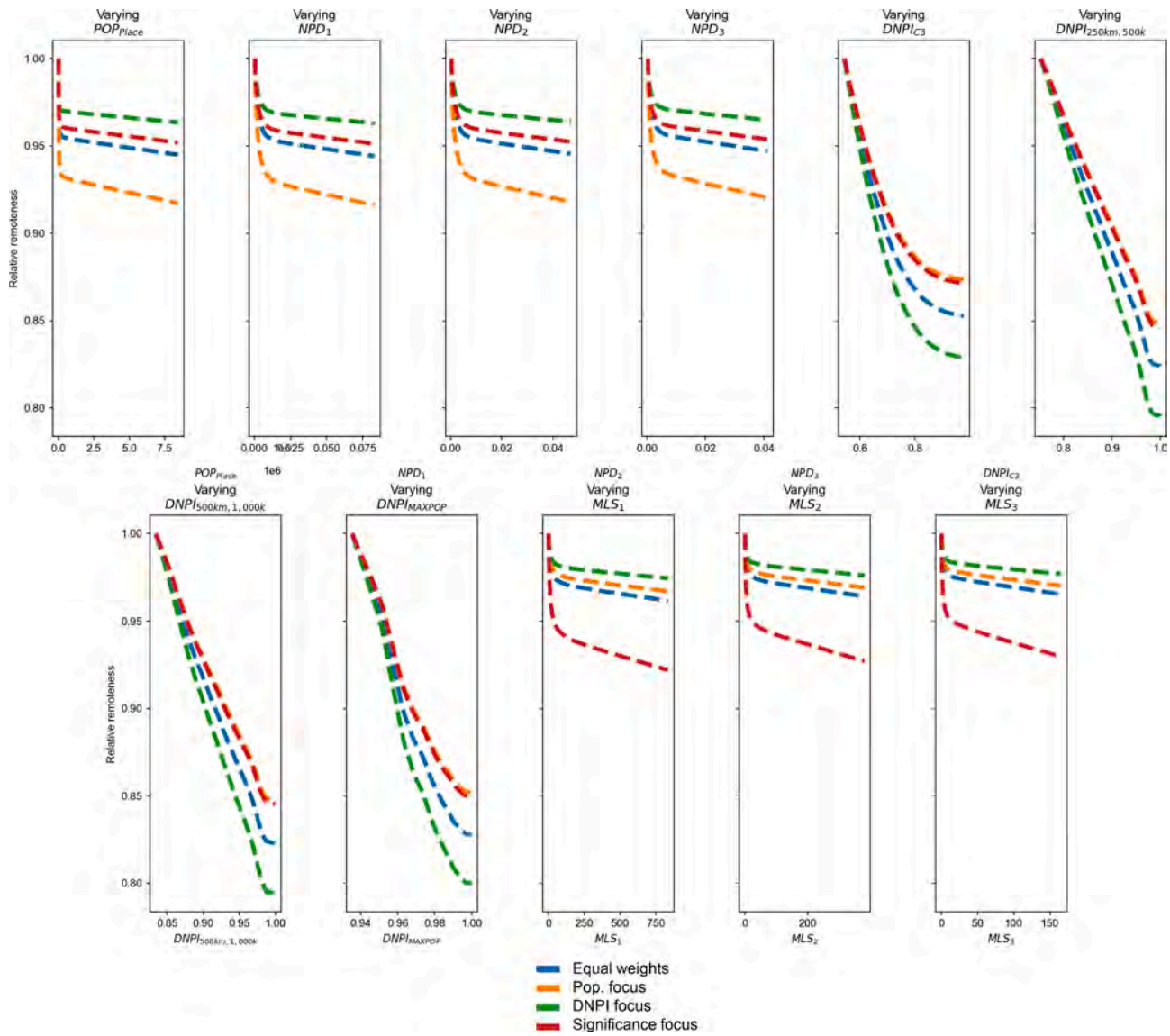

Fig. A9-2. Sensitivity of different weighting schemes to changes in individual remoteness indicators for the network-based approach. A synthetic baseline place with randomly initialized remoteness indicator was used, and each of the seven indicators was systematically increased. For each increment, the resulting remoteness index was calculated, for each of the four weighting schemes. The curves show the different levels of sensitivity of the weighting schemes.

"urbanness" of nearby places may be low. Lastly, it should be noted that some measurements of demographic and social processes are currently reported with high levels of error at the place level, and thus, may limit the usability of the PLURAL indices at place-level granularity. In such cases, users are encouraged to aggregate the place-level PLURAL indices to meaningful spatial units, or to apply the proposed methodology at the granularity of interest.

Future work will include the generation of continuous, local remoteness measures, e.g., at the grid cell level, to create a fine-grained, gapless model of the rural–urban continuum. The Thiessen polygons that we used to visualize places (Fig. 4) and to generate spatial networks (Fig. 3b) could be further combined to model the areal influence of places, potentially enabling us to allocate non-place populations to proximate places. Future work will also include the application of the presented approaches to other countries where comparable data is available, or even to expand such efforts to continental scales.

5. Conclusions

Herein, we described and evaluated two approaches to model the rural–urban continuum at fine spatial granularity (i.e., the place level) and over long time periods (i.e., 1930–2018) for the conterminous US. This work makes several contributions: (a) We fill a gap in the US data landscape by providing temporally consistent, place-level rural–urban classifications, refining existing county-level classifications, and complementing finer-grained classifications at the ZIP code level by providing a demographically meaningful analytical unit (i.e., the census place), applicable over very long time periods, (b) We generate methodological knowledge by developing and comparing a raster-based approach and a spatial network approach. (c) Specifically, we adopt elements of a spatial network approach developed from a physical perspective on human settlements to a population perspective of human settlements; Moreover, we introduce a novel and effective metric for measuring remoteness based on population distributions discretized to point data, which we call the ***distance-based neighborhood population index*** (***DNPI***). (d) We make all of our place-level remoteness indices publicly available, calling them the ***place-level urban–rural index*** (***PLURAL***), enabling researchers of various disciplines to conduct fine-grained, cross-sectional and (in case of PLURAL-1) longitudinal analyses across the rural–urban continuum, and over a time period of almost

90 years.

While there is no single way to measure or model rurality and urbanness, different existing rural–urban classifications may have been developed for specific purposes (e.g., population-centric vs. landscape or infrastructure focused approaches to study demographic processes, or land change processes, respectively). Herein, we propose and advance a set of population-centric indices, that will enable users to study processes at the place scale, using minimalist data sources as inputs. Despite these minimalist (population-based) inputs, our cross-comparisons reveal that the PLURAL indices are also correlated with landscape- and infrastructure-based indices. As these indices are continuous (rather than discrete) measures of urbanness, reported at fine granularity, at a meaningful, policy-level scale, they are both *generalizable* to other regions / time periods, and they are *scalable* to arbitrary, coarser spatial units, both temporally consistent or of dynamic nature. Moreover, these indices are available at fine spatial grain and can be used in situations where the Uncertain Geographic Context Problem would critically bias analytical results, e.g., in the case of highly segregated settings. Moreover, many demographic (e.g., population sorting) and land use (e.g., zoning) processes are heavily influenced by local conditions and therefore better understood at the place-level compared to a coarser scale. Thus, for studying nationally extensive but locally driven processes, place-level metrics are expected to have particular merit.

Finally, because the PLURAL indices are minimalist with respect to their input data, they can be used to study social as well as physical spatio-temporal processes, as circular inference (i.e., using the studied geographic process itself for rural–urban stratification purposes) will be kept to a minimum. Finally, we would like to emphasize that the PLURAL indices do not aim to represent a replacement of existing rural–urban classifications at coarser analytical unit, but rather a versatile, fine-grained dataset that can be used to produce rural–urban classifications at arbitrary spatial units and consistent over time. Lastly, we aim to advance methodological knowledge, as our approach to generating the PLURAL indices can be applied to non-US and historical contexts, which may be particularly advantageous for data-poor contexts. In conclusion, it is our hope that the PLURAL remoteness indices will enable researchers to add a long-term temporal dimension to rural and rural–urban studies, at a refined spatial granularity, and ultimately, contribute to more informed planning and decision-making.

6. Data and code availability

The PLURAL indices and the underlying historical place-level population counts, as well as the derived remoteness indicators used to establish the PLURAL-1 and PLURAL-2 indices are available as tabular data (CSV format). Moreover, separate spatial vector data files containing the place locations for each point in time (1930–2018) attributed with the PLURAL indices are available at https://doi.org/10.6084/m9.figshare.22596946. Python code purely based on free and open source technologies to apply the PLURAL framework to other place-level datasets is available at https://github.com/johannesuhl/plural.

Funding

Funding for this research was provided by the University of Colorado Population Center (CUPC; Project 2P2CHD066613-06), funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development, United States. Moreover, this research was partially funded by Projects R21HD098717-01A1 and 5R21HD098717-02 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, United States. The content is solely the responsibility of the author and does not necessarily represent the official views of the CUPC, National Insitutes of Health (NIH) or University of Colorado Boulder.

CRediT authorship contribution statement

Johannes H. Uhl: Conceptualization, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft. **Lori M. Hunter:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing. **Stefan Leyk:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing. **Dylan S. Connor:** Conceptualization, Writing – review & editing. **Jeremiah J. Nieves:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Cyrus Hester:** Data curation, Writing – review & editing. **Catherine Talbot:** Data curation, Writing – review & editing. **Myron Gutmann:** Conceptualization, Data curation, Funding acquisition, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The PLURAL indices for the US places from 1930 to 2018 are publicly available as tabular data and geospatial vector data at https://doi.org/10.6084/m9.figshare.22596946. A Python script to apply the PLURAL method to other place-level population datasets is available at https://github.com/johannesuhl/plural.

Acknowledgments

This research has been supported by Project Number R21HD098717-01A1, "Health, Social, and Demographic Trends in Rural Communities" funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), United States. The project has also benefited from research, administrative, and computing support, also provided by NICHD, to the University of Colorado Population Center, United States (CUPC; Projects 2P2CHD066613-06 and 5R21HD098717-02). The content is solely the responsibility of the author and does not necessarily represent the official view of the CUPC, National Institutes of Health (NIH) or CU Boulder. Furthermore, the authors would like to thank the student workers Brenda Blanco Soto and Ciara Coughlan for their help with digitizing historical census records used in this study.

Appendix 1. Source data acquisition

We used US census place population counts from 1970, 1980, 1990, 2000, and 2010, obtained from NHGIS (Manson et al., 2020), containing both, incorporated places and census-designated places (US Census Bureau, 1994). From the same data source, we obtained census place geometries for each decade between 1930 and 2010, as well as for 2018. Census places are represented as polygonal features (1980–2018) and as discrete locations (i.e., point features) for the years 1930–1970. For consistency, we converted the place polygons (1980–2018) into discrete point data, by using the polygon centroid, in order to keep the input data as simplistic as possible). Moreover, we use census place populations (i.e., 5-year population estimates 2014–2018) from the American Community Survey (ACS). Importantly, we digitized decadal place populations for 1930 to 1970, obtained from counts published in the 1940 and 1960 decennial reports (Tables 5 and 8, respectively; US Census Bureau 1942, 1964). We integrated and harmonized the place population estimates from these different sources and joined them to the spatial data (i.e., point features representing the census places), based on the place identifier, resulting in a set of place locations and their population, for each year.

Appendix 2. Raster-based remoteness index

Appendix 2.1: GIS workflow for creating the raster-based PLURAL indices

1) In a given year, for each population category ***pc***, we identify the places in ***pc*** based on their population.
2) For these places, we generate a Euclidean Distance surface, indicating for each grid cell (in a grid of 1 km × 1 km) the distance to the nearest place of the category ***pc***, resulting in a total of 5 Euclidean Distance surfaces per year, one for each ***pc***.
3) For all places existing in a given year, we extracted the raster values from each of these five distance surfaces at each place location.
4) Moreover, a focal population density estimate ***pd*** within a radius of 10 km is used in order to characterize the population distributions in the place neighborhood, and to reduce the sensitivity of the index to arbitrary partitioning of places (e.g., neighborhoods in large cities are typically recorded as individual places). We chose a radius of 10 km since based on some initial experiments, a circle with a 20 km diameter is likely to meaningfully aggregate the individual neighborhoods of a large city, without overly aggregating dispersed rural places. The focal population density is obtained by applying a Point Density tool to the geospatial place point data, for each year, calculating the total population within a circular focal window, resulting in a continuous surface of focal population density in the same spatial grid used for the Euclidean Distance surfaces.
5) Likewise, we extract the raster value of this focal population density surface at each place location.
6) This process is repeated for all years in the observation period, attributing each place, in each year, with its location, population, focal population density, and the distances to the nearest places of each population category.

Table A21.

Appendix 2.2: Weighting schemes for raster-based remoteness index (see also Appendix 9)

These weighting schemes are exemplary, and do not represent any attempt to cover different systematic choices or a broad spectrum of weight combinations. Our main goal is to illustrate the flexibility of resulting rural–urban classifications due to different weight combinations, aiming to emphasize different components of the indices.

- **Equal weights:** All remoteness indicators are weighted equally.
- **Place-centric:** Place population and focal population density share 50% of the weight, and each distance-based component receives a weight of 0.1. This weighting scheme emphasizes the local context of a place, and gives less importance to the distance to large places.
- **Metro focus:** Place population and population density receive low weights, and the distance-based components receive higher weights, with the distance to places > 250,000 receiving the highest weight. This weighting scheme emphasizes the influence of large cities in proximity to a place, and gives less importance to the local context of a place.
- **Place-centric + metro focus:** Place population and focal population density share 50% of the weight, and the weights for the distance components are constructed such that distance to large (metro) areas has the highest weight. This weighting scheme combines both concepts.

Appendix 3. Spatial network based remoteness index

We built spatial networks based on the neighborhood relationships of Voronoi diagrams constructed around census place locations. These relationships are measured by means of cardinalities: Neighbors of cardinality 1 are direct neighbors of a node (i.e., connected by an edge), and a cardinality of 2 includes the neighbors of cardinality 1 of the cardinality 1 neighbors of a given place, etc. This is relevant when having a consistent but locally flexible method that can be applied to the densely populated Northeast as well as the sparsely populated regions in the Southwest of the US These networks consist of nodes (i.e., place locations) and edges (i.e., connections between neighboring places), shown in Fig. 3c for the neighborhood of cardinality 1 (i.e., connecting places whose Thiessen polygons share a common boundary). These topological relationships enable the efficient identification of different neighborhood levels for each place. Note that for implementing the spatial networks efficiently, we use the centroids of the Thiessen polygons rather than the actual place locations as network nodes. While the original distances between place locations are assigned to each edge, the use of Thiessen polygon centroids as network nodes decreases processing time, as the spatial networks can be built from the Thiessen polygons without being re-joined to the original place data. This causes the slight offsets between network nodes and place locations in Fig. 3c (Table A31).

Weighting schemes for the network-based remoteness index (see also Appendix 9).

These weighting schemes are exemplary, and do not represent any attempt to cover different systematic choices or a broad spectrum of weight combinations. Our main goal is to illustrate the flexibility of resulting rural–urban classifications due to different weight combinations, aiming to emphasize different components of the indices.

- **Equal weights:** All remoteness indicators are weighted equally.
- **Population focus:** Place population and NPD metrics share 50% of the weight, all other metrics receive equal weights. This weighting scheme emphasizes the local context of a place, and gives less importance to the distance to large places.
- **DNPI focus:** DNPI based metrics receive 50% of the weight, all other metrics receive equal weights. This scheme has a heavy focus on the spatial embedding of a place, and gives less importance to the size of the place itself.
- **Significance focus:** MLS-based metrics receive 50% of the weight, all other metrics receive equal weights. This weighting scheme focuses on whether a place is significant to its neighboring places or not, and thus, models the concept of central places.

Appendix 4. Evaluation data

Appendix 5. Results

Appendix 6. Evaluation results

Appendix 6.1: Comparing the two methods

The quantitative comparison of the two modelling approaches yields high levels of correlation (Pearson > 0.8) between any modelling approach and weighting scheme, and these correlations slightly decrease over time (Fig. A6-1a,d). The scatterplots of the equal-weights schemes (Fig. A6-1-b,e) indicate that for the large majority of places, the network-based approach yields a more conservative remoteness estimate, i.e., most places are below the main diagonal, thus, PLURAL-2 indicates higher levels of urbanness. Notably, the relationship between the two approaches differs between smaller and larger places (as defined by the place population), exhibiting two clusters, shifted along the x-axis: While for large places, PLURAL-1 and PLURAL-2 yield similar values, for small places, PLURAL-1 yields higher levels of remoteness (approximately +0.2) than PLURAL-2. Moreover, the spatial patterns of the differences between PLURAL-1 and PLURAL-2 estimates per place (Fig. A6-1c,f) exhibit a strong spatial pattern in East-West direction,

indicating that the previously observed systematic offset of +0.2 is particularly prevalent in the East, whereas places where network-based remoteness exceeds raster-based remoteness, are mostly smaller places in the West. This is likely due to the higher sensitivity of the network-based approach to the regional or local importance of places, and a result of the east–west population density gradient in the CONUS.

Appendix 6.2: Comparing to county-level, and global rural–urban classifications (GHS-SMOD)

Since the county-level urban–rural classifications discussed in Section 1 were designed exclusively for the US and exhibit a relatively large temporal gap to the most recent set of US census places (i.e., ACS population estimates from 2018), we decided to compute a county-level urbanness index for the US based on the recently released, globally available Settlement Model (SMOD), compiled by the Global Human Settlement (GHS) project (Pesaresi et al., 2013). The GHS-SMOD is based on population data and built-up areas derived from Landsat observations (Pesaresi et al., 2016) and implements the REGIO/EUROSTAT taxonomy (Dijkstra & Poelman 2014) for defining classes of urbanness (Florczyk et al. 2019). GHS-SMOD is available globally at a spatial resolution of 1 km. The GHS-SMOD classifies the Earth into 1 km grid cells of seven density-based levels of urbanness, ranging from "very low density" (class 11) to "urban center" (class 30) (Pesaresi & Freire, 2016). We calculated the area proportions of each class per county, and computed a weighted average per county, based on these area proportions, giving weight 1 to the "very low density" class, and weight 7 to the "urban center". We then scaled the resulting county-level scores into the range of [0,1] (Fig. A4a) and assessed the distributions of our place-level indices within strata of county-level, SMOD-based urbanness estimates for consistency and plausibility.

The county-level rural–urban classifications shown in Fig. 6 provide valuable baseline models on the rural–urban continuum in the US. We employed these to evaluate whether the created remoteness indices exhibit similar trends, measured by the correlation between our place-level indices and the rural–urban designation assigned to the county containing each place; and assessed the outliers in the place-level distributions of the PLURAL indices within strata of county-level rural–urban designations. Moreover, we compared our place-level results to the FAR area codes at the ZIP code area level.

Comparing the place-level remoteness index distribution within strata defined by the county-level urban–rural designations, we generally observe a coherence between the PLURAL indices and county-level classes, manifesting in increasing place-level remoteness with increasing county-level rurality, for all county-level classification schemes and for both modelling approaches (Fig. 10a,b). Notably, we observe frequent "lower" outliers (i.e., below the lower whisker defined as 1.5xIQR) for the raster-based approach (PLURAL-1) (Fig. A6-21a), located evenly across the CONUS (Fig. A6-21c). This effect is not present when comparing the network-based remoteness index (PLURAL-2) against the county-level designations (Fig. A6-21b,d), thus indicating that the spatial-network approach better approximates the rural–urban gradient models underlying the county-level classifications. Such an effect is also observed for the other weighting schemes (Fig. A6-22). This effect is particularly strong when comparing against the GHS-SMOD based strata (Fig. A6-21a,b), indicating that the spatial network approach, which is based on population data only, is capable of capturing the spatial configuration of places, and thus able to approximate a population-density and built-up land-based RUC modeling approach based on population data only. The strong spatial patterns of "upper" outliers, which are similar for the various county-level classifications (Fig. A6-21c,d) indicate counties of presumably high levels of within-county variability of remoteness, for example in Arizona and North Dakota (Fig. A6-21c).

Visually, the trends observed in Fig 10a,b indicate varying relationships between the place-level remoteness estimates and county-level classes. We formally tested linearity- and rank-based correlation between the county-level RUC classifications and our place-level indices. While the rank-based Spearman's correlation coefficient is high (>0.7) for all indices and county-level classes (Fig. A6-23), we observe highest levels of linearity for the network-based approaches and GHS-SMOD, confirming the previous observations.

Appendix 6.3: Comparison to gridded population data

We visually assessed how the concepts of remoteness and population density differ. To do so, we used gridded population counts in 2018 from Worldpop (https://hub.worldpop.org/doi/10.5258/SOTON/WP00660) and re-aggregated them to the Thiessen polygons used to depict each place. When comparing the PLURAL indices and population density, we observe a general positive association, which is expected, as population density is a component in both the distance-based PLURAL-1 and the network-based PLURAL-2. However, we also observe larger bands around metropolitan areas and medium-size cities (Fig. A6-3), resulting from the incorporation of the relationships between neighboring or nearby places in our PLURAL indices. These relationships model the influence of larger places on smaller places in their proximity, which is the core of our remoteness concept. Hence, the use of population density alone to measure the rural–urban continuum is not advised, as the spatial context will not be taken into account.

Appendix 6.4: Comparison to 2010 ZIP-code level FAR remoteness measures

Moreover, we compared our place-level results to the Frontier And Remote area codes (FAR), at the ZIP code area level. There are four FAR area codes, temporally referenced to 2010, identifying ZIP code areas with populations living more than specific travel time thresholds from urban areas of specific population thresholds (see Cromartie & Nulph, 2015 for details). As opposed to the county-level classifications, the FAR classes are not mutually exclusive. Thus, we use receiver-operator-characteristic (ROC) analysis (Green & Swets, 1966) to test whether there are thresholds that can be applied to our remoteness indices that yield high levels of agreement when comparing to each of four FAR classes. To join ZIP code area FAR designations to the places, we applied a spatial join based on 2010 place polygon centroids to the 2010 ZIP code areas (US Census Bureau, 2019).

While most of the discussed rural–urban classifications for the US are based on proximity to metropolitan areas, or commuting patterns, the FAR codes available at the ZIP code level, is the only index explicitly implementing the concept of remoteness by means of road-network derived travel times, and population sizes. The ROC plots shown in Fig. A6-4 show generally high Area-under-the-Curve values, indicating that for each of our remoteness indices, there is a threshold that allows for mimicking the FAR classes at low levels of type I and type II errors (i.e., False positive rate < 0.2, True positive rate > 0.9). The remaining disagreement may be due to the ambiguous spatial relationship between ZIP code areas and census places, their difference in spatial granularity and different modelling strategies (i.e., use of Euclidean distance versus road network distance). Here, it is worth noting that ZIP code areas outside of place boundaries are not taken into account in this assessment.

Appendix 6.5: Comparison to travel-time accessibility indicators

The accessibility indicators used herein are derived from globally available travel time estimates (Nelson, 2019). These travel time estimates are based on road network data from OpenStreetMap and ancillary data on travel times, on land cover and terrain characteristics, as well as urban areas derived from the Global Human Settlement Layer (see Nelson et al., 2019; Weiss et al., 2018 for detailed method descriptions). These indicators are available as global, gridded surfaces at a

spatial resolution of 30 arc-seconds, indicating the travel time from each grid cell to the nearest city of a specific population range, approximately in 2015 (see Fig. A4e for some examples). We used the surfaces associated with population ranges comparable to the population ranges used for the distance-based PLURAL index, i.e., travel times to the nearest place of 10,000–20,000, 20,000–50,000, and 50,000–100,000 inhabitants, and extracted the travel time for each place centroid of our 2018 places dataset. We then qualitatively assessed the relationship between these travel times and our remoteness indices, and computed the theoretical travel speed based on the (road-network based) travel time given from the accessibility indicators, and the Euclidean distances used to construct the distance-based PLURAL index in 2018. We assessed the plausibility of these theoretical travel speed estimates in order to quantify the bias of using Euclidean distances instead of road network distances in our remoteness models.

We assess the impact of using Euclidean distance instead of road network distance to model remoteness, using the travel times from Nelson et al. (2019) observed at each 2018 place centroid. We generally observe increasing travel times with increasing remoteness, while this trend seems to be slightly more linear for the network-based indices than for the raster-based indices, as the scatterplots in Fig. A6-5a suggest. For the raster-based equal-weights and place-centric weighting schemes, we observe an additional peak in travel times for remoteness values around 0.4, in particular for the travel times to places > 50,000 inhabitants, as an effect of the local focus of these weighting schemes. Comparing travel times to the corresponding Euclidean distances $\boldsymbol{D_{PC}}$ (see Eq. 1) used for the raster-based remoteness models, we observe a noisy, but roughly linear relationship (Fig. A6-5b). Places that deviate heavily from the main diagonal are expected to be located nearby geographic obstacles such as mountain ranges, rivers, borders, or lakes, causing deviations of road network routes from the shortest (Euclidean) route (Fig. A6-5b). The visualization of the theoretical travel speed estimates obtained from the Euclidean distances and travel time estimates (Fig. A6-5c) shows that for most parts of the CONUS, these travel speed measures range within plausible values, i.e., between 30 km/h and 140 km/h. We derived these thresholds from the typical speed limits (25 miles/hour in cities, 80 miles/hour on interstates) plus approximately 10 km/h of tolerance to account for differences in the definition of places in our modelling approaches and the Nelson et al. (2019) accessibility indicators. For example, Nelson et al. seem to model Chicago as a single place, whereas US census places distinguish between different neighborhoods of the city. Only in a few regions (in black, and yellow, respectively) this range of plausible travel speed is exceeded.

Appendix 6.6: Long-term place-level trends of the built environment across the rural–urban continuum

Although the temporal patterns in our remoteness indices directly reflect spatial population change over time, we would also expect these changes to be correlated with other processes related to urbanization, such as urban and rural building patterns. Using spatially explicit data on built-up areas for most of the US, at fine spatial granularity (i.e., a grid of 250 m × 250 m resolution) from the HISDAC-US, we constructed place-level trajectories of total built-up area and the number of buildings over time for each place where HISDAC-US data is available, for each decade from 1930 to 2010, as well as for 2015. We then calculated the correlations of these metrics with the PLURAL indices over time, with the expectation that the PLURAL indices (indicating remoteness) would be negatively correlated with new building (indicating urban expansion). Examples of the place-level built-up areas, over time and for three approximate levels of rurality, as indicated by the county-level RUCC, are shown in Fig. A4, illustrating the different growth trajectories across the rural–urban gradient. Similarly, Fig. A4d illustrates how building density increases in *peri*-urban areas, over the long term, while staying relatively stable in scattered, rural settlements, strengthening our hypothesis of different built-up land trajectories along the rural–urban gradient.

When comparing the number of built-up records (i.e., approximate number of buildings) and the built-up area against our remoteness indices, per census place and over time, we observe a highly nuanced relationship between built-environment characteristics and the rural–urban continuum as modelled by our remoteness indices. As expected, high levels of remoteness are associated with low building counts and small built-up areas. These patterns are highly stable over time, and across weighting schemes when visually assessing the scatterplots in Fig. A6-6a,b. These point patters differ slightly between the raster-based approach and the network-based approach, indicating that the ability to model a built environment perspective of rurality based on population data varies between approaches. Moreover, the negative correlations between building counts / built-up area and remoteness increase over time (Fig. A6-6c,d), in particular for the built-up records variable (Fig. A6-6c).

Appendix 6.7: Relationship of remoteness to contemporary landscape and settlement characteristics

While the built-up areas as provided by the HISDAC-US are derived from cadastral property data and may suffer from low levels of spatial accuracy in rural areas, we also used Microsoft's building footprint data (Microsoft 2018), reflecting the state of built-up areas in approximately 2016, derived from high-resolution remote sensing imagery, at high levels of accuracy (Uhl et al., 2021). We rasterized these building footprint data to fine, CONUS-wide spatial grids of 250 × 250 m. From these gridded surfaces (i.e., indicating the number of buildings, and the total building footprint area per grid cell), we computed commonly used settlement metrics, for each place, such as the number of buildings, total built-up area, average building area (measuring built-up intensity), and landscape metrics, such as the average area of contiguous patches of built-up land, number of patches, largest patch index (measuring spatial segregation), landscape division index, and patch cohesion index (measuring the segregation and connectedness of built-up land) (McGarigal, 2015). We then assessed the correlations of these settlement and landscape metrics with our remoteness indices in 2018, motivated by previous work reporting strong relationships between the size and structure of built-up land and the rural–urban gradient (Luck and Wu, 2002, Vizzari, 2011, Vizzari & Sigura, 2013). More specifically, we used the bounding box of each 2018 place polygon, buffered by 1 km in all directions, as a focal window in which the landscape metrics were computed, and assessed the correlation between these landscape metrics and the PLURAL indices. By doing so, we were able to not only characterize the size, shape, and structure of the built-up areas representing each place, but also the unincorporated land in proximity to the places. Some examples of the extracted built-up areas per place are shown in Fig. 5c, illustrating the difference in size and structure of built-up areas across the rural–urban continuum.

The assessment of the relationship between size and structure related characteristics of the contemporary built-up land (derived from Microsoft's US building footprint data) at the place level reveals strong association of these characteristics with the place-level remoteness, which is in line with related work using landscape metrics as a proxy for the rural–urban gradient (Luck and Wu, 2002, Vizzari, 2011, Vizzari & Sigura, 2013). We generally observe strong negative correlations between remoteness and the settlement and landscape characteristics, indicating that remote places are characterized by few, small buildings, organized in small, highly disconnected patches of built-up area of similar sizes (Fig. A6-7a). These negative correlations are slightly higher for the network-based modelling approach than for the raster-based remoteness indices (Fig. A6-7a), and the patterns of these measures across the RUC are highly similar for the different weighting schemes (Fig. A6-7b).

Appendix 7. Sensitivity analysis of PLURAL-1 to user-set parameters

We assessed the sensitivity of the PLURAL-1 indices to two user-defined parameters: (a) the population thresholds for place distance calculations, and (b) the focal window size used for focal population density estimation. As can be seen in the QQ-plots in Fig. A7-1, in all tested cases, sensitivity is relatively low (i.e., data points are near the main diagonal). Moreover, sensitivity is independent of the year, and sensitivity of the PLURAL-1 is higher to the chosen population thresholds used for place-to-place distance calculation, and low to the choice of the focal window radius used for focal population density estimation. Importantly, the metro-centric weighting scheme exhibits highest levels of robustness to both parameters, consistent in 1930 and 2018.

Appendix 8. Distributions of distance to nearest neighboring places over time

For PLURAL-1, the distances to places of different population categories are derived from Euclidean Distance raster surfaces generated at a spatial resolution of 1 km, which represents a trade-off between granularity and computational cost. While we did not formally test the sensitivity of the results to the spatial resolution of these gridded surfaces, we analyzed the distributions of the nearest neighbor distances for the places in each year, shown for 1930, 1980, and 2018 in Fig. A8-1, illustrating that the large majority (i.e., around 95%) of places is farther than 1.4 km (grid cell diagonal) away from the nearest neighboring place, and thus, only 5% of the places may share a 1 km $\times$ 1 km grid cell with another place, which may result in distance values of 0. Such rare cases are expected to occur in urban areas, and hence, limits PLURAL-1 in a way that distances $<$ 1.4 km between places cannot be differentiated. As a side note, the shift in the frequency peaks from 1930 (10 km) to 2018 (3 km), illustrating the densification process that occurred over the intervening period.

Appendix 9

Sensitivity Analysis of weighting schemes to changes in remoteness index components.

Fig. A91. Fig. A92.

Appendix 10. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.landurbplan.2023.104762.

References

Abdelkarim, A., Alogayell, H. M., Alkadi, I. I., & Youssef, I. (2022). Spatial–temporal prediction model for land cover of the rural–urban continuum axis between Ar-Riyadh and Al-Kharj cities in KSA in the year of 2030 using the integration of CA–Markov model, GIS-MCA, and AHP. *Applied Geomatics*, 1–25.

Andersen, L. M., Harden, S. R., Sugg, M. M., Runkle, J. D., & Lundquist, T. E. (2021). Analyzing the spatial determinants of local Covid-19 transmission in the United States. *Science of the Total Environment, 754*, Article 142396.

Angel, S., Parent, J., Civco, D. L., Blei, A. M., & Potere, D. (2011). The dimensions of global urban expansion: estimates and projections for all countries, 2000–2050. *Progress in Planning, 75*, 53–107.

Anton, C. E., & Lawrence, C. (2014). Home is where the heart is: The effect of place of residence on place attachment and community participation. *Journal of Environmental Psychology, 40*, 451–461.

Armstrong, A., & Stedman, R. C. (2019). Understanding local environmental concern: The importance of place. *Rural Sociology, 84*(1), 93–122.

Balk, D., Pozzi, F., Yetman, G., Deichmann, U., & Nelson, A. (2005). The distribution of people and the dimension of place: Methodologies to improve the global estimation of urban extents. In *In International Society for Photogrammetry and Remote Sensing, proceedings of the urban remote sensing conference* (pp. 14–16).

Banzhaf, S., Ma, L., & Timmins, C. (2019). Environmental justice: The economics of race, place, and pollution. *Journal of Economic Perspectives, 33*(1), 185–208.

Banzhaf, S., & Walsh, R. P. (2013). Segregation and Tiebout sorting: The link between place-based investments and neighborhood tipping. *Journal of Urban Economics, 74*, 83–98.

Bardin, S., & Kedron, P. (2022). *A Geographic Perspective on Place-Based Policies.*

Bell, M. M. (1992). The Fruit of difference: The rural-urban continuum as a system of identity 1. *Rural Sociology, 57*(1), 65–82.

Bender, D. J., Tischendorf, L., & Fahrig, L. (2003). Using patch isolation metrics to predict animal movement in binary landscapes. *Landscape Ecology, 18*(1), 17–39.

Braswell, A., Leyk, S., Connor, D., & Uhl, J. (2021). Creeping disaster along the US coastline: Understanding exposure to sea level rise and hurri-canes through historical development (No. j4k3e). Center for Open Science.

Brooks, M. M., Mueller, J. T., & Thiede, B. C. (2020). County reclassifications and rural-urban mortality disparities in the United States (1970–2018). *American Journal of Public Health, 110*(12), 1814–1816.

Butler, M. A. (1990). Rural-urban continuum codes for metro and nonmetro counties (No. 9028). US Department of Agriculture, Economic Research Service, Agriculture and Rural Economy Division.

CDC (2021). Diabetes Data and Statistics. Available online: https://www.cdc.gov/diabetes/data/index.html. Accessed on December 22, 2021.

Chetty, R. (2021). Improving equality of opportunity: New insights from big data. *Contemporary Economic Policy, 39*(1), 7–41.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics, 129*(4), 1553–1623.

Christiaensen, L., & Todo, Y. (2014). Poverty reduction during the rural–urban transformation–the role of the missing middle. *World Development, 63*, 43–58.

Çöltekin, A., De Sabbata, S., Willi, C., Vontobel, I., Pfister, S., Kuhn, M., & Lacayo, M. (2011). Modifiable temporal unit problem. https://doi.org/10.5167/uzh-54263.

Connor, D. S., Uhl, J. H., Xie, S., Talbot, C., Hester, C., Jaworski, T., ... & Hunter, L. M. (2022). Rising community poverty reduces social mobility for rural children. *Available at SSRN 4127500*.

Connor, D. S., & Storper, M. (2020). The changing geography of social mobility in the United States. *Proceedings of the National Academy of Sciences, 117*(48), 30309–30317.

Cosby, A. G., McDoom-Echebiri, M. M., James, W., Khandekar, H., Brown, W., & Hanna, H. L. (2019). Growth and persistence of place-based mortality in the United States: The rural mortality penalty. *American Journal of Public Health, 109*(1), 155–162.

Cromartie, J., and Nulph, D. (2015). Frontier and Remote Area Codes. Available: http://www.ers.usda.gov/data-products/frontier-and-remote-area-codes.aspx [March 2015].

Cromartie, J., Dobis, E., Krumel, T., McGranahan, D., & Pender, J. (2020). Rural America at a glance: 2020 edition. *Amber Waves: The Economics of Food, Farming, Natural Resources, and Rural America, 2020*(1490-2020-1865).

Curtin S.C. & Spencer M.R. (2021). Trends in death rates in urban and rural areas: United States, 1999–2019. NCHS Data Brief, no 417. Hyattsville, MD: National Center for Health Statistics. DOI: https://dx.doi.org/10.15620/cdc:109049.

Curtis, K. J., DeWaard, J., Fussell, E., & Rosenfeld, R. A. (2020). Differential Recovery migration across the rural-urban gradient: Minimal and short-term population gains for rural disaster-affected gulf coast counties. *Rural Sociology, 85*(4), 856–898.

Dewey, R. (1960). The rural-urban continuum: Real but relatively unimportant. *American Journal of Sociology, 66*(1), 60–66.

Dijkstra, L., & Poelman, H. (2014). Regional Working Paper 2014. A harmonised definition of cities and rural areas: the new degree of urbanisation. *European Commission Directorate-General for Regional and Urban Policy: Working Paper 01/2014.*

Du Toit, M. J., & Cilliers, S. S. (2011). Aspects influencing the selection of representative urbanization measures to quantify urban–rural gradients. *Landscape Ecology*, *26*(2), 169-181.Economic Research Service (ERS). Rural-urban commuting area codes. Washington (DC): US Department of Agriculture; 2013. http://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx. Accessed September 14, 2020.

Dudwick, N. (2011). *From farm to firm: Rural-urban transition in developing countries*. World Bank publications.

Economic Research Service (2022). Available online at https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/. Last accessed on 20 April 2023.

Esch, T., Marconcini, M., Marmanis, D., Zeidler, J., Elsayed, S., Metz, A., … Dech, S. (2014). Dimensioning urbanization–An advanced procedure for characterizing human settlement properties and patterns using spatial network analysis. *Applied Geography, 55*, 212–228.

Federal Register (2008). Census Designated Place (CDP) Program for the 2010 Census—Final Criteria. 2008;73(30). https://www.gpo.gov/fdsys/pkg/FR-2008-02-13/pdf/E8-2667.pdf.

Florczyk A.J., Corbane C., Ehrlich D., Freire S., Kemper T., Maffenini L., Melchiorri M., Pesaresi M., Politis P., Schiavina, M., Sabo F. & Zanchetta L. (2019). GHSL Data Package 2019, EUR 29788 EN, Publications Office of the European Union, Luxembourg, 2019, doi:10.2760/290498, JRC 117104.

Flowerdew, R., Geddes, A., & Green, M. (2001). Behaviour of regression models under random aggregation. *Modelling Scale in Geographical Information Science*, 89–104.

Ghelfi, L. M., & Parker, T. S. (1997). A county-level measure of urban influence. Rural America/Rural Development. *Perspectives, 12*(2221–2019-2617), 32–41.

Glaeser, E. L., Resseger, M., & Tobio, K. (2009). Inequality in cities. *Journal of Regional Science, 49*(4), 617–646.

Golding, S. A., & Winkler, R. L. (2020). Tracking urbanization and exurbs: Migration across the rural–urban continuum, 1990–2016. *Population Research and Policy Review, 39*(5), 835–859.

Goodchild, M. F. (2022). The Openshaw effect. *International Journal of Geographical Information Science, 36*(9), 1697–1698.

Gosnell, H., & Abrams, J. (2011). Amenity migration: Diverse conceptualizations of drivers, socioeconomic dimensions, and emerging challenges. *GeoJournal, 76*(4), 303–322.

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics (Vol. 1, pp. 1969-12). New York: Wiley.

Grimm, N. B., Pickett, S. T. A., Hale, R. L., & Cadenasso, M. L. (2017). Does the ecological concept of disturbance have utility in urban social–ecological–technological systems? *Ecosystem Health and Sustainability, 3*(1), e01255.

Grineski, S., Bolin, B., & Boone, C. (2007). Criteria air pollution and marginalized populations: Environmental inequity in metropolitan Phoenix, Arizona. *Social Science Quarterly, 88*(2), 535–554.

Grubesic, T. H. (2008). Zip codes and spatial analysis: Problems and prospects. *Socio-economic Planning Sciences, 42*(2), 129–149.

Henderson, J. V., & Wang, H. G. (2005). Aspects of the rural-urban transformation of countries. *Journal of Economic Geography, 5*(1), 23–42.

Hillemeier, M. M., Weisman, C. S., Chase, G. A., & Dyer, A. M. (2007). Individual and community predictors of preterm birth and low birthweight along the rural-urban continuum in central Pennsylvania. *The Journal of Rural Health, 23*(1), 42–48.

Homsy, G. C., Liu, Z., & Warner, M. E. (2019). Multilevel governance: Framing the integration of top-down and bottom-up policymaking. *International Journal of Public Administration, 42*(7), 572–582.

Hunter, L. M., Talbot, C. B., Connor, D. S., Counterman, M., Uhl, J. H., Gutmann, M. P., & Leyk, S. (2020). Change in US Small Town Community Capitals, 1980–2010. *Population Research and Policy Review, 39*(5), 913–940.

Ingram, D. D., & Franco, S. J. (2014). 2013 NCHS urban-rural classification scheme for counties (No. 2014). US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Jaeger, J. A. (2000). Landscape division, splitting index, and effective mesh size: new measures of landscape fragmentation. *Landscape Ecology, 15*, 115–130.

Johnson, K. M., & Lichter, D. T. (2019). Rural depopulation: Growth and decline processes over the past century. *Rural Sociology, 84*(1), 3–27.

Johnson, K. M., & Lichter, D. T. (2020). Metropolitan reclassification and the urbanization of rural America. *Demography, 57*(5), 1929–1950.

Kwan, M. P. (2012). How GIS can help address the uncertain geographic context problem in social science research. *Annals of GIS, 18*(4), 245–255.

Lee, B. A., & Sharp, G. (2017). Ethnoracial diversity across the rural-urban continuum. *The ANNALS of the American Academy of Political and Social Science, 672*(1), 26–45.

Lewicka, M. (2005). Ways to make people active: The role of place attachment, cultural capital, and neighborhood ties. *Journal of environmental psychology, 25*(4), 381–395.

Leyk, S., & Uhl, J. H. (2018). HISDAC-US, historical settlement data compilation for the conterminous United States over 200 years. *Scientific data, 5*(1), 1–14.

Lichter, D. T., Brown, D. L., & Parisi, D. (2021). The rural–urban interface: Rural and small town growth at the metropolitan fringe. *Population, Space and Place, 27*(3), e2415.

Lichter, D. T., & Johnson, K. M. (2020). A demographic lifeline? Immigration and Hispanic population growth in rural America. *Population Research and Policy Review, 39*(5), 785–803.

López-Goyburu, P., & García-Montero, L. G. (2018). The urban-rural interface as an area with characteristics of its own in urban planning: A review. *Sustainable cities and society, 43*, 157-165.

Luck, M., & Wu, J. (2002). A gradient analysis of urban landscape pattern: a case study from the Phoenix metropolitan region, Arizona, USA. *Landscape Ecology, 17*(4), 327–339.

Machado, S. R., Jayawardana, S., Mossialos, E., & Vaduganathan, M. (2021). Physician density by specialty type in urban and rural counties in the US, 2010 to 2017. *JAMA network open, 4*(1). e2033994–e2033994.

Manduca, R., & Sampson, R. J. (2019). Punishing and toxic neighborhood environments independently predict the intergenerational social mobility of black and white children. *Proceedings of the national academy of sciences, 116*(16), 7772–7777.

Manson, S., Schroeder, J., Van Riper, D., Kugler, T., & Ruggles, S. (2020). *IPUMS National Historical Geographic Information System: Version 15.0*. Minneapolis, MN: IPUMS.

Manzo, L., & Devine-Wright, P. (2013). *Place attachment: Advances in theory, methods and applications*. Routledge.

McGarigal, K. (1995). *FRAGSTATS: spatial pattern analysis program for quantifying landscape structure* (Vol. 351). US Department of Agriculture, Forest Service, Pacific Northwest Research Station.

McGarigal, K. (2015). FRAGSTATS Help. Accessible online at: https://www.umass.edu/landeco/research/fragstats/documents/fragstats.help.4.2.pdf. Last accessed on 04 October 2020.

McGee, T. G. (2008). Managing the rural–urban transformation in East Asia in the 21st century. *Sustainability Science, 3*(1), 155–167.

McGranahan, D. A., Hession, J. C., Mines, F. K., & Jordan, M. F. (1986). Social and economic characteristics of the population in metro and nonmetro counties, 1970–80. *Economic Research, 800*, 336–4700.

Meerow, S., & Newell, J. P. (2019). Urban resilience for whom, what, when, where, and why? *Urban Geography, 40*(3), 309–329.

Microsoft. (2018). USBuildingFootprints. GitHub. https://github.com/Microsoft/USBuildingFootprints.

Miguet, P., Fahrig, L., & Lavigne, C. (2017). How to quantify a distance-dependent landscape effect on a biological response. *Methods in Ecology and Evolution, 8*(12), 1717–1724.

Monnat, S. M. (2020). Trends in US working-age Non-Hispanic white mortality: Rural-urban and within-rural differences. *Population Research and Policy Review, 39*(5), 805–834.

Murali, R., Suryawanshi, K., Redpath, S., Nagendra, H., & Mishra, C. (2019). Changing use of ecosystem services along a rural-urban continuum in the Indian Trans-Himalayas. *Ecosystem Services, 40*, 101030.

National Academies of Sciences, Engineering, and Medicine. (2016). *Rationalizing rural area classifications for the Economic Research Service: A workshop summary*. National Academies Press.

Nelson, K. S., Nguyen, T. D., Brownstein, N. A., Garcia, D., Walker, H. C., Watson, J. T., & Xin, A. (2021). Definitions, measures, and uses of rurality: A systematic review of the empirical and quantitative literature. *Journal of Rural Studies, 82*, 351–365.

Nelson, A., Weiss, D. J., van Etten, J., Cattaneo, A., McMenomy, T. S., & Koo, J. (2019). A suite of global accessibility indicators. *Scientific Data, 6*(1), 1–9.

Nelson, A. (2019): Travel time to cities and ports in the year 2015. figshare. Dataset. https://doi.org/10.6084/m9.figshare.7638134.v3.

Nowosad, J., & Stepinski, T. F. (2018). Global inventory of landscape patterns and latent variables of landscape spatial configuration. *Ecological Indicators, 89*, 159–167.

OECD/PSI (2020), Rural Development Strategy Review of Ethiopia: Reaping the Benefits of Urbanisation, *OECD Development Pathways,* OECD Publishing, Paris, https://doi.org/10.1787/a325a658-en.

Openshaw, S. (1979). A million or so correlation coefficients, three experiments on the modifiable areal unit problem. *Statistical Applications in the Spatial Science*, 127–144.

Pahl, R. E. (1966). The rural-urban continuum1. *Sociologia ruralis, 6*(3), 299–329.

Partridge, M. D., Rickman, D. S., Olfert, M. R., & Tan, Y. (2015). When spatial equilibrium fails: Is place-based policy second best? *Regional Studies, 49*(8), 1303–1325.

Pender, J., Hertz, T., Cromartie, J., & Farrigan, T. (2019). *Rural America at a glance* (No. 1476-2020-044).

Pesaresi, M., & Freire, S. (2016). GHS settlement grid, following the REGIO model 2014 in application to GHSL Landsat and CIESIN GPW v4-multitemporal (1975-1990-2000-2015). European Commission, Joint Research Centre [Dataset]. http://data.jrc.ec.europa.eu/dataset/jrc-ghsl-ghs_smod_pop_globe_r2016a.

Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A. J., Freire, S., Halkia, S., Julea, A. M., Kemper, T., Soille, P., & Syrris, V. (2016). Operating Procedure for the Production of the Global Human Settlement Layer from Landsat Data of the Epochs 1975, 1990, 2000, and 2014. *Publications Office of the European Union.*

Pesaresi, M., Guo, H., Blaes, X., Ehrlich, D., Ferri, S., Gueguen, L., … Zanchetta, L. (2013). A global human settlement layer from optical HR/VHR remote sensing data: Concept and first results. *IEEE Journal of Selected Topics in Applied Earth Observation & Remote Sensing, 6*(5), 2102–2131.

Peters, D. J. (2020). Community susceptibility and resiliency to COVID-19 across the rural-urban continuum in the United States. *The Journal of Rural Health, 36*(3), 446–456.

Piantadosi, S., Byar, D. P., & Green, S. B. (1988). The ecological fallacy. *American Journal of Epidemiology, 127*(5), 893–904.

Ratcliffe, M. (2015). A century of delineating a changing landscape: The Census Bureau's urban and rural classification, 1910 to 2010. *In Annual Meeting of the Social Science History Association*.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., III, Lambin, E., … Foley, J. (2009). Planetary boundaries: Exploring the safe operating space for humanity. *Ecology and Society, 14*(2).

Sachdeva, M., Fotheringham, A. S., Li, Z., & Yu, H. (2021). Are we modelling spatially varying processes or non-linear relationships? *Geographical Analysis*.

Sack, R. D. (1997). *Homo geographicus: A framework for action, awareness, and moral concern*. Baltimore, MD: Johns Hopkins University Press.

Scala, D. J., & Johnson, K. M. (2017). Political polarization along the rural-urban continuum? The geography of the presidential vote, 2000–2016. *The ANNALS of the American Academy of Political and Social Science, 672*(1), 162–184.

Scott, A. J., & Storper, M. (2015). The nature of cities: The scope and limits of urban theory. *International Journal of Urban and Regional Research, 39*(1), 1–15.

Shah, P., Owens, J., Franklin, J., Mehta, A., Heymann, W., Sewell, W., … Doshi, R. (2020). Demographics, comorbidities and outcomes in hospitalized Covid-19 patients in rural southwest Georgia. *Annals of Medicine, 52*(7), 354–360.

Sibley, L. M., & Weiner, J. P. (2011). An evaluation of access to health care services along the rural-urban continuum in Canada. *BMC Health Services Research, 11*(1), 1–11.

Spielman, S. E., Tuccillo, J., Folch, D. C., Schweikert, A., Davies, R., Wood, N., & Tate, E. (2020). Evaluating social vulnerability indicators: Criteria and their application to the Social Vulnerability Index. *Natural Hazards, 100*(1), 417–436.

Taubenböck, H., Droin, A., Standfuß, I., Dosch, F., Sander, N., Milbert, A., … Wurm, M. (2022). To be, or not to be 'urban'? A multi-modal method for the differentiated measurement of the degree of urbanization. *Computers, Environment and Urban Systems, 95*, Article 101830.

Theobald, D. M., & Romme, W. H. (2007). Expansion of the US wildland–urban interface. *Landscape and Urban Planning, 83*(4), 340–354.

Thiede, B. C., Butler, J. L., Brown, D. L., & Jensen, L. (2020). Income inequality across the rural-urban continuum in the United States, 1970–2016. *Rural Sociology, 85*(4), 899–937.

Thiessen, A. J., & Alter, J. C. (1911). Precipitation averages for large areas. *Monthly Weather Review, 39*, 1082–8437.

Uhl, J. H., Leyk, S., McShane, C. M., Braswell, A. E., Connor, D. S., & Balk, D. (2021). Fine-grained, spatiotemporal datasets measuring 200 years of land development in the United States. *Earth System Science Data, 13*(1), 119–153.

Ulrich-Schad, J. D., & Duncan, C. M. (2018). People and places left behind: Work, culture and politics in the rural United States. *The Journal of Peasant Studies, 45*(1), 59–79.

US Census Bureau (1942). US Census of Population: 1940. Vol. I, Number of Inhabitants. US Government Printing Office, Washington, D.C.

US Census Bureau (1964). US Census of Population: 1960. Vol. I, Characteristics of the Population. Part I, United States Summary. US Government Printing Office, Washington, D.C.

US Census Bureau (1994). Geographic areas reference manual. Chapter 9 – Places. US Dept. of Commerce, Economics and Statistics Administration, Bureau of the Census. Available at https://www2.census.gov/geo/pdfs/reference/GARM/Ch9GARM.pdf. Accessed on 25 September 2020.

US Census Bureau (2019a). County Population by Characteristics: 2010-2019. Available at https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html. Accessed on 21 December 2021.

US Census Bureau (2019b). 5-Digit ZIP Code Tabularion Areas (ZCTA5). Available online: https://catalog.data.gov/harvest/2019-zcta510. Accessed on 22 September 2021.

US Census Bureau (2020). Urban and Rural. Available at https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html. Accessed on 25 September 2020.

Vizzari, M., & Sigura, M. (2013). Urban-rural gradient detection using multivariate spatial analysis and landscape metrics. *Journal of Agricultural Engineering, 44*(s2).

Vizzari, M. (2011, June). Spatio-temporal analysis using urban-rural gradient modelling and landscape metrics. In *International Conference on Computational Science and Its Applications* (pp. 103-118). Springer, Berlin, Heidelberg.

Voronoi, G. (1908). Nouvelles applications des parametres continus a la theorie des formes quadratiques, Deuxieme Memoire, Recherches sur les paralleloedres primitifs. *Journal Reine Angew. Math., 134*, 198–287.

Waldorf, B., & Kim, A. (2015, April). Defining and measuring rurality in the US: From typologies to continuous indices. In Commissioned paper presented at the Workshop on Rationalizing Rural Area Classifications, Washington, DC.

Waldorf, B., & Kim, A. (2018). The Index of Relative Rurality (IRR): US County Data for 2000 and 2010. *Purdue University Research Repository*. https://doi.org/10.4231/R7959FS8

Waldorf, B. S. (2006). A continuous multi-dimensional measure of rurality: Moving beyond threshold measures (No. 379-2016-21891).

Wandl, A., & Magoni, M. (2017). Sustainable planning of peri-urban areas: Introduction to the special issue. *Planning Practice & Research, 32*(1), 1–3.

Weber, B. A., Fannin, J. M., Cordes, S. M., & Johnson, T. G. (2017). Upward mobility of low-income youth in metropolitan, micropolitan, and rural America. *The ANNALS of the American Academy of Political and Social Science, 672*(1), 103–122.

Weiss, D. J., Nelson, A., Gibson, H. S., Temperley, W., Peedell, S., Lieber, A., … Mappin, B. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature, 553*(7688), 333–336.

Workman, R., & McPherson, K. (2021). Measuring rural access for SDG 9.1. 1. *Transactions in GIS, 25*(2), 721–734.

WorldPop (www.worldpop.org - School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University (2018). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076). https://dx.doi.org/10.5258/SOTON/WP00660.

York, A., Tuccillo, J., Boone, C., Bolin, B., Gentile, L., Schoon, B., & Kane, K. (2014). Zoning and land use: A tale of incompatibility and environmental injustice in early Phoenix. *Journal of Urban Affairs, 36*(5), 833–853.

Yuan, D. Y. (1964). The rural-urban continuum: A case study of Taiwan. *Rural Sociology, 29*(3), 247.

Yusuf, A. B. (1974). A reconsideration of urban conceptions: Hausa urbanization and the Hausa rural-urban continuum. *Urban Anthropology*, 200–221.

Zhou, Y., Liu, Y., Wu, W., & Li, Y. (2015). Effects of rural–urban development transformation on energy consumption and CO_2 emissions: A regional analysis in China. *Renewable and Sustainable Energy Reviews, 52*, 863–875.

Žlender, V., & Thompson, C. W. (2017). Accessibility and use of peri-urban green space for inner-city dwellers: A comparative study. *Landscape and Urban Planning, 165*, 193–205.

Further reading

Cyriac, S. (2022). Dichotomous classification and implications in spatial planning: A case of the Rural-Urban Continuum settlements of Kerala, India. *Land Use Policy, 114*, Article 105992.

Hunter, L. M., Talbot, C.B., Connor, D.S., Gutmann, M.P, Leyk, S. & Uhl, J. (2020b). Understanding Small Town America Requires Small Town Data: The mismatch associated with county-scale analyses. 2020 Annual Meeting of the Population Association of America.

Porter, Jeremy R. and Frank M. Howell. 2016. "A Spatial Decomposition of County Population Growth in the United States: Population Redistribution in the Rural-to-Urban Continuum, 1980–2010." Pp. 175–98 in *Recapturing Space: New Middle-Range Theory in Spatial Demography*, edited by F. M. Howell, J. R. Porter, and S. A. Matthews. Springer.