

Proxying economic activity with daytime satellite imagery: Filling data gaps across time and space

Patrick Lehnert ^{a,*}, Michael Niederberger^{a,b}, Uschi Backes-Gellner ^a and Eric Bettinger ^{c,*}

^aDepartment of Business Administration, University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland

^bDepartment of Geography, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

^cGraduate School of Education, Stanford University, 520 Galvez Mall, Stanford, 94306 CA, USA

*To whom correspondence should be addressed: E-mail: patrick.lehnert@business.uzh.ch (P.L.); ebettinger@stanford.edu (E.B.)

Edited By: Taylor Jaworski

Abstract

This paper develops a novel procedure for proxying economic activity with daytime satellite imagery across time periods and spatial units, for which reliable data on economic activity are otherwise not available. In developing this unique proxy, we apply machine-learning techniques to a historical time series of daytime satellite imagery dating back to 1984. Compared to satellite data on night light intensity, another common economic proxy, our proxy more precisely predicts economic activity at smaller regional levels and over longer time horizons. We demonstrate our measure's usefulness for the example of Germany, where East German data on economic activity are unavailable for detailed regional levels and historical time series. Our procedure is generalizable to any region in the world, and it has great potential for analyzing historical economic developments, evaluating local policy reforms, and controlling for economic activity at highly disaggregated regional levels in econometric applications.

Keywords: daytime satellite imagery, Landsat, machine learning, economic activity, land cover

Significance

Reliably measuring regional economic activity poses a key challenge for social scientists as administrative statistics often have insufficient regional detail, limited time series, or politically motivated biases (e.g. in autocratic regimes). While the use of proxy variables (e.g. night light intensity from satellite data) has resolved some of these issues, these proxies are still insufficient for some settings (e.g. break-off regions from the former Soviet Union bloc states). This paper develops a novel proxy from daytime satellite imagery to measure economic activity in longer time series and much smaller regional units (anywhere in the world) than any previous satellite data.

Introduction

The lack of credible data hampers our understanding of regional economic development, especially in historical contexts. Most countries lack data at the regional or even municipal levels, and the extant data either focus only on recent years or lack consistency across regions and/or time. To fill these data gaps across time and space, researchers have increasingly used satellite data on night light intensity as a proxy for economic activity [e.g. (1–4)].

However, night light intensity data have significant weaknesses. They are available only for a limited time series (from 1992) and, due to their spatial resolution (one kilometer at the equator), they are not reliable for disaggregated regional units such as municipalities or suburbs (5–7). Administrative or survey data on economic activity encounter similar problems. They are typically not available for longer historical time series, not

regionally disaggregated, or otherwise unreliable or unavailable to the research community. In a recent literature review, the lack of long time series and regional scalability have been identified as key weaknesses of former satellite-based metrics (8).

This paper solves these key weaknesses by offering an economic proxy from daytime satellite imagery with worldwide applicability based on a procedure that we developed in 2020 and that applies machine-learning techniques to Landsat imagery (9). We show how this proxy enables economic analyses across time periods and for highly disaggregated spatial units in an example that identifies the innovation effect of higher education institutions in East and West German regions—an analysis that would otherwise be impossible due to unavailable and unreliable data (as in former communist or developing countries). The proxy presents valuable information on economic activity over a uniquely long time series (from 1984) at a level of regional disaggregation that is smaller (30-m resolution) than any alternative.

Competing Interest: The authors declare no competing interest.

Received: September 5, 2022. **Revised:** January 31, 2023. **Accepted:** February 22, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Daytime satellite imagery from the Landsat program has so far received almost no attention in economics applications. The few existing applications rely on visual interpretation for identifying, for example, agricultural land use, (de)forestation, or urbanization [e.g. (10, 11)]. Developing new and more accurate proxies with Landsat data requires novel machine-learning techniques to adapt these data to economic settings. Tools such as the Google Earth Engine facilitate the processing and analysis of Landsat's large geographic datasets (12).

Landsat daytime satellite data have three advantages over other data sources. First, Landsat data have substantially higher disaggregation (30-m resolution) than regional administrative or other satellite data such as night light intensity (1-km resolution) (13). This higher resolution entails more precise information at a much more disaggregated regional level. Our economic proxy can characterize economic development at regional levels and even in much smaller localities such as municipalities or urban districts.

Second, NASA launched the first satellite of the Landsat program (Landsat-1) in 1972, making Landsat the earliest existing source of regionally highly disaggregated satellite data (14, 15). While Landsat did not reach its full potential until 1984, the long time horizon of the data allows researchers to construct longer historical data than other regional economic administrative data or other proxies based on satellite data such as night light intensity (which is available from 1992). In comparison to regional economic data, which might be available for some administrative locations, Landsat daytime data pre-date the break-up of the former Soviet Union, German Reunification, and other significant changes in regional or even local economic development.

Third, Landsat satellites collect multispectral imagery of the earth, that is, they observe the energy that the earth reflects in different spectral bands (e.g. infrared). The geographic remote-sensing literature has been using algorithmic techniques for detecting land cover in Landsat scenes for half a century [e.g. (16–20)]. It provides successful applications of machine-learning techniques that exploit Landsat's multispectral information in the identification of different types of land cover from subsets of Landsat data [e.g. (21–25)]. We extend this literature by creating a procedure that combines all Landsat data available from 1984 to map six different types of land cover, which we refer to as *surface groups*: built-up surfaces, grassy surfaces, forest-covered surfaces, surfaces with crop fields, surfaces without vegetation, and water surfaces.

As some surface groups are more closely related to economic activity than others (26, 27), mapping surface groups yields important information on regional economic activity. For example, increases in built-up surfaces, which include agglomerations of cities or transportation networks, coincide with increases in economic activity (28, 29). Even holding built-up surfaces constant, the other surface groups provide greater predictability of local economic conditions. While previous research finds that the raw spectral values of Landsat-7 imagery can serve as a slightly better proxy than night light intensity in Vietnamese regions (30), we show that identifying the different surface groups through machine-learning techniques results in a substantially improved proxy for economic activity over time and space. In so doing, our approach goes beyond a previous approach for inferring economic data through mapping land cover from single Landsat images in Zhousan City, China (31) by developing an automated procedure for combining multiple Landsat images into annual data composites. Moreover, compared to a previous application that uses Landsat imagery to directly predict village asset wealth in Africa

(32), our surface groups can function both as an indicator of land cover and as a more general proxy for economic activity with the potential for worldwide application. Which proxy to choose for empirical research depends on the concrete research question, with other proxies offering advantages through specialization in, for example, asset wealth (32) and our proxy offering advantages through painting an overall picture of regional (or even more subregional) economic activity.

Our procedure for detecting surface groups as a proxy for economic activity produces a metric with high internal and external validity. We lay the foundations for computing and validating the proxy using Germany as an example. In the context of the German Reunification, our proxy provides important, previously unavailable, yet reliable information on economic activity in East German regions. As such, the surface groups allow the examination of pre-reunification economic developments at highly disaggregated regional levels and—due to their independence of politically motivated economic statistics produced during the communist era—with very high validity. To demonstrate the necessity of the surface groups proxy for answering important research questions in the social sciences, we apply our proxy in a causal analysis comparing the effect of higher education institutions on regional innovation in East and West German regions. This and similar analyses are otherwise impossible with other data. Our data and their applications easily extend to other settings and geographies throughout the world, for example, those suffering from insufficient regional detail, limited time series, or politically motivated biases such as in autocratic or closed political systems (33, 34).

The value of surface groups as a proxy for economic activity

Features of surface groups

We use a supervised machine-learning algorithm to classify Landsat pixels into one of six surface groups. This classification procedure requires two external data sources. First, the raw imagery of Landsat satellites constitutes the input data to be classified. Before performing the classification, we pre-process this raw imagery to obtain pixel-based annual composites incorporating imagery from multiple Landsat satellites. Second, CORINE Land Cover (CLC) data (which are available only for the five reference years 1990, 2000, 2006, 2012, and 2018) serve as an external source of ground-truth information, that is, they indicate the true surface group for a subset of the input pixels. The training data for the classification algorithm consist of a stratified random sample of Landsat pixels matched to their true surface group from CLC data. The details of the classification procedure are outlined in *Materials and methods* and in the [supplementary material \(Text S1, Fig. S1, and Tables S1 and S2\)](#).

Following prior literature utilizing land cover classifications [e.g. (35–39)], we identify and map six different types of land cover—the surface groups—which are similar to previous work in a Chinese region (25). These groups include the following: (1) built-up surfaces feature buildings of non-natural materials such as concrete, metal, and glass (e.g. residential buildings, industrial plants, roads); (2) grassy surfaces are covered by green plants or groundcover with similar surface reflectance (e.g. natural grassland); (3) surfaces with crop fields include vegetation for agricultural purposes (e.g. grain fields); (4) forest-covered surfaces contain trees or other plants with similar surface reflectance (e.g. mixed forests); (5) surfaces without vegetation have (almost)

no reflective vegetation or buildings (e.g. bare rock); and (6) water surfaces comprise any type of water surface (e.g. lakes). Our algorithm classifies these respective surfaces, which we then combine to form our proxy for economic activity.

The output of our procedure for detecting surface groups is a dataset containing the surface group of every Landsat pixel location in Germany annually from 1984 through 2020. One year comprises more than 630 million Landsat pixels, amounting to more than 23 billion pixel-year observations in the output data. Of these observations, 16.2% are classified as built-up; 20.9% as grass; 29.5% as crops; 25.6% as forest; 3.3% as no vegetation; and 3.8% as water. Only 0.6% of observations contain missing values due to, for example, cloud cover that is uninterrupted within a given year for single pixels in the Landsat data. For applications in research projects, researchers can aggregate this pixel-level information to the geographical unit matching their respective research objective (e.g. administrative regional units or ZIP code areas).

Fig. 1 illustrates the data sources we use and the output data we produce. As examples, the left column of Fig. 1 shows a large-scale area with the metropolitan region of *Nuremberg* (situated in mid-south Germany) in the center of the picture. The right column shows a small-scale area with the village of *Muhr-am-See* (*Muhr-at-the-lake*, situated about 30 miles south-west of *Nuremberg*) in the upper part of the picture and its accompanying lake (*Altmühlsee*) in the lower part of the picture (framed area in the left column). Fig. 1A, which uses Landsat's visible spectral bands to approximate the perception of the human eye, shows the Landsat composite for 2018 (the input data). Fig. 1B illustrates the six different types of land cover we identify from the CLC data (the ground-truth data). Fig. 1C shows the surface group that our classification algorithm produces for every Landsat pixel location in 2018. As a reference, Fig. 1D shows current high-resolution satellite images from Esri World Imagery (40).

Internal validity

To evaluate whether we achieve an accurate classification of Landsat pixels into the six surface groups (i.e. the measure's internal validity), we assess several indicators of prediction accuracy. In so doing, we follow the standard procedure in the remote-sensing literature that uses supervised machine learning to classify land cover [e.g. (21, 22, 41)] and derive these indicators from five-fold cross-validation. This method draws five subsets from the input data and uses these subsets to perform five iterations of pixel classification (see *Materials and methods* and [Text S1.5](#) in the [supplementary material](#) for more details).

Using the classification output from the five-fold cross-validation, we calculate five common indicators of prediction accuracy with respect to each surface group: overall accuracy, true-positive rate, true-negative rate, balanced accuracy, and user's accuracy. Overall accuracy denotes the percentage of pixels correctly classified, true-positive rate the percentage of pixels correctly classified as belonging to the respective surface group, true-negative rate the percentage of pixels correctly classified as not belonging to the respective surface group, balanced accuracy the average of true-positive rate and true-negative rate, and user's accuracy the percentage of pixels correctly classified as belonging to the respective surface group among all pixels belonging to the respective surface group.

Table 1 shows the five-fold cross-validation results with respect to each surface group. With 82.8%, overall accuracy for built-up surface areas is similar to that in other studies detecting built-up land with Landsat data [e.g. (21, 22)]. The other four

indicators are also in line with other studies [e.g. (22, 41)]. Furthermore, we achieve very high overall accuracy for forest (89.5%), areas with no vegetation (87.0%), and water (90.9%).

The five-fold cross-validation results show that our output data constitute an internally valid measure of land cover. All indicators of prediction accuracy reinforce that our classification algorithm accurately identifies the six surface groups, suggesting that we adequately implemented the procedures from the remote-sensing literature. The high internal validity of the surface groups is a prerequisite for their external validity as a proxy for economic activity.

External validity

To evaluate the external validity of surface groups as a proxy for economic activity, we empirically analyze how much they explain of the variation in direct measures of regional economic activity (which are available for parts of our time series). We draw on two such external direct measures: First, from administrative statistics, we extract a regionally disaggregated direct measure of gross domestic product (GDP), the most commonly used economic indicator in the literature evaluating previous satellite-based proxies for economic activity [e.g. (5, 42)]. For Germany, this measure is available at the administrative county (*Kreis*) level from 2000. Second, we use a socioeconomic dataset that provides household income as a further indicator of economic activity with a very high level of regional detail (43). This indicator is available at the level of grid cells sized 1 km² (and thus independent of administrative borders), but annually only from 2009. See *Materials and methods* and the [supplementary material \(Text S2.2\)](#) for more details on the two external validation data sources.

We analyze the external validity of our proxy by comparing the amount of variation in GDP that our proxy and night light intensity generate. We obtain this result from comparing Ordinary Least Squares (OLS) regressions of GDP on the surface groups with OLS regressions of GDP on night light intensity (see *Materials and methods* and [supplementary material Text S2.3](#), [Tables S9, S10, S25, and S26](#) for more details on the methodology). Our preferred surface groups specification, which additionally includes year and federal state fixed effects to cancel out any bias due to potential measurement error in the dependent or independent variables, explains 62.3% of the variation in GDP. Using night light intensity instead of surface groups in the same specification explains only 47.1% of this variation, that is, our proxy achieves 32.3% higher precision than previous data at the disaggregated regional level of counties.

The value of surface groups as a proxy for regional economic activity becomes even more obvious at the very small regional level of grid cells. In a similar OLS analysis, our preferred surface-groups specification explains a much larger percentage of the variation in household income than the corresponding night lights specification, with 67.5% vs. 30.7% (i.e. 119.9% higher precision).

The value of surface groups in comparison to night light intensity as a proxy for economic activity thus substantially increases with the degree of regional disaggregation. This finding is supported by an additional analysis on the prediction of county-level GDP by county-size category (see [Text S2.3](#) and [Fig. S5](#) in the [supplementary material](#)). On average, the surface groups explain a larger percentage of the variation in GDP for smaller counties than for larger counties.

Fig. 2 underscores and visualizes these findings. It plots the statistical distribution of the OLS regression residuals, which are

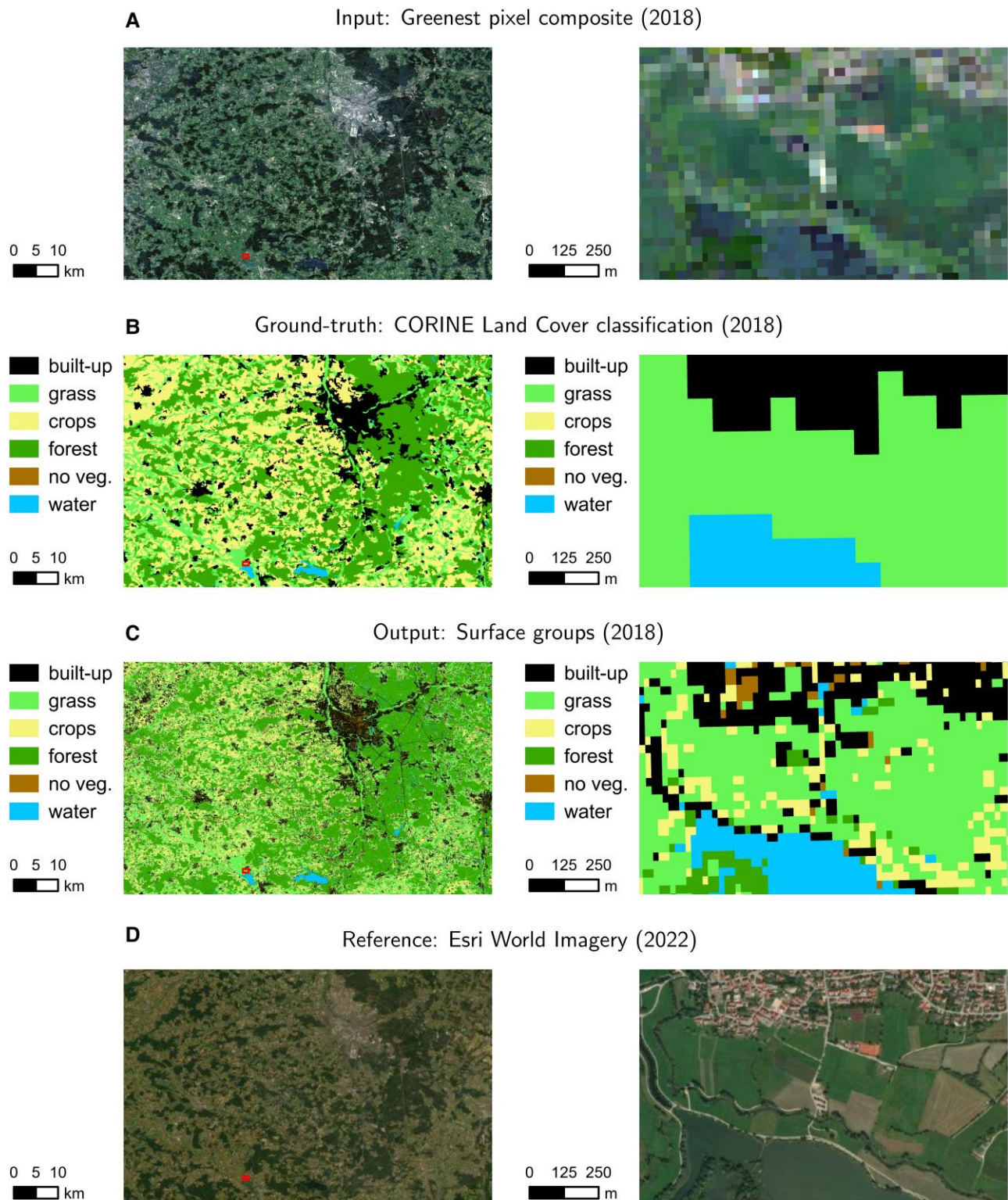


Fig. 1. Visual comparison of data sources. Pictures in the left column show the same approx. 78×49 square kilometers area with the metropolitan region of Nuremberg in the center. Pictures in the right column show the same approx. 1.3×0.8 square kilometers area with the village of Muhr-am-See in the upper part and its accompanying lake (Altmühlsee) in the lower part (framed area in the left column).

smaller when the measure is a better proxy for economic activity. The plots show that, for both GDP and household income, this distribution is smoother and narrower for surface groups (Fig. 2A and C) than for night light intensity (Fig. 2B and D). For household income, the residual distribution of the night lights specification

even exhibits a plateau—instead of a real peak—around the value zero, whereas the surface groups show a very clear peak and a narrow residual distribution.

Furthermore, we conduct four additional validation analyses in the [supplementary material \(Text S2\)](#). First, we find that surface

groups are a temporally and spatially less biased proxy for economic activity than night light intensity (Figs. S2–S4, S6, S7). This feature is important for the surface groups to serve as a valid proxy for comparisons of economic activity over time and between regions. Temporal bias would occur if the OLS residual is constant for a given region throughout all observation years, and spatial bias would occur if this residual is equal for clusters of regions. Surface groups yield a considerably smaller temporal bias that outweighs their somewhat larger spatial bias in comparison to night light intensity. Second, in line with their smaller bias, surface groups offer more information on within-region changes in economic activity than night light intensity through higher within-region heterogeneity (Figs. S8 and S9, Tables S15–S17). That is, our surface groups allow for a more precise determination of which subregional units drive the change in a region's economic activity by isolating the change in each subregional unit. Third, surface groups outperform also newer night light intensity data with higher spatial resolution in proxying economic activity (Tables S11 and S12). Fourth, we validate surface groups as a proxy for economic conditions in developing countries by comparing their predictive power to that of a prior metric of village asset wealth in Africa (32), a similar but more specialized outcome variable (Tables S18 and S19). This analysis shows that the validity of surface groups is not restricted to developed European countries such as Germany, but that surface groups can also provide valuable insights on economic conditions in developing countries across the world.

Essential improvements in social science research through surface groups data

To demonstrate the usefulness of our surface groups proxy for regional analyses in general and for applications in the social and

economic sciences in particular, we tackle a perennial research question for which an empirical answer requires accurate data on regional economic activity before the German Reunification—a period for which reliable (administrative) East German data do not exist. Studies in education and innovation economics find that higher education institutions improve innovation outcomes in developed regions [e.g. (44–46)]. However, whether similar effects would occur in less developed regions remains open due to the lack of adequate data for causal analyses.

One example for such a less developed region is East Germany, which—like many countries with a history under a communist regime—lagged dramatically behind in economic development compared to West Germany (47, 48). Germany offers an ideal setting for studying whether otherwise identical higher education institutions affect developed and less developed regions differently provided that the necessary data are available. The surface groups proxy provides exactly these necessary and otherwise unavailable data because it constitutes an unconfounded measure of pre-reunification economic activity in all East German regions. As the proxy is also available for all West German regions, we can directly compare regions in both parts of the country with a high degree of disaggregation (e.g. municipalities). With the surface groups proxy, we can estimate causal effects by adjusting for differences in prereunification economic trends. Surface groups are thus crucial for answering the research question on the different effects of higher education institutions. While we use surface groups to exploit the German setting as an example, they can be a similarly crucial source of information on developing countries or other countries without reliable (administrative) data.

To study differences in the effects of higher education institutions in East and West Germany, we compare regional innovation outcomes in East and West German regions with a University of Applied Sciences (UAS) campus. In addition to surface groups, we use a municipality-level dataset [from (49)] that annually indicates whether a municipality lies within the catchment area of a UAS campus. Moreover, these data contain two well-established indicators of regional innovation, patent quantity (the number of priority patent applications per municipality and year) and patent quality (the average number of forward citations three years after a patent's publication per municipality and year). These patent-based indicators are complete for all German municipalities from 1991.

Descriptive analyses show that in 1991 (i.e. immediately after reunification), East German municipalities lag far behind West German ones in both patent quantity and patent quality. Moreover, in both indicators, East German municipalities never reach the same level as West German ones over time. With this descriptive evidence as a starting point, surface groups allow us

Table 1. Five-fold cross-validation results.

Surface group	Overall accuracy	True-positive rate	True-negative rate	Balanced accuracy	User's accuracy
Built-up	0.828	0.606	0.877	0.741	0.514
Grass	0.831	0.451	0.910	0.680	0.511
Crops	0.832	0.381	0.932	0.657	0.563
Forest	0.895	0.685	0.938	0.812	0.708
No veg.	0.870	0.756	0.886	0.821	0.490
Water	0.909	0.672	0.958	0.815	0.765

Notes: Indicators calculated with respect to each surface group. Values indicate the average over all five iterations and all five reference years in the CLC data. See *Materials and methods* and the [supplementary material \(Text S1.5\)](#) for more details (including the results separately for every reference year in Tables S3–S8).

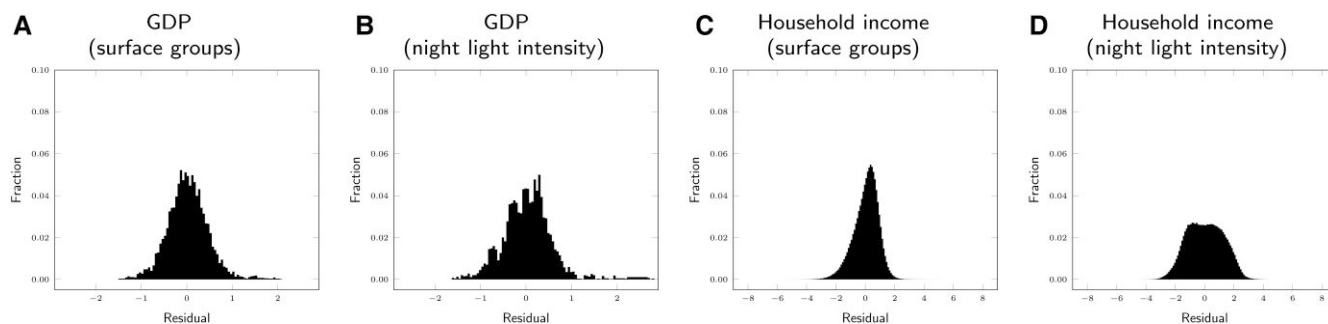


Fig. 2. Statistical distribution of OLS regression residuals. See the [supplementary material \(Text S2.3, Tables S9 and S10\)](#) for details on the regression specifications. Bin width of histograms is 0.05 in panels A and B and 0.1 in panels C and D.

to identify whether UASs have a causal impact on decreasing the innovation gap, thus bringing East German municipalities closer to their West German counterparts.

Our surface groups proxy is the only available reliable measure on pre-reunification economic activity in East German regions, thus enabling a comparison of municipalities with similar pre-reunification economic characteristics. We perform propensity-score matching on average pre-reunification growth in the six surface groups. The estimated propensity score allows us to econometrically adjust for pre-reunification differences.

Our causal propensity-score matching estimations show that the post-reunification increase in patent quantity is significantly smaller in East German UAS regions than in West German ones. Thus, while previous studies on higher education institutions in developed countries in general and on UASs in particular show positive innovation effects [e.g. (50, 51, 46)], the policy instrument of opening UASs for regional development has a much smaller effect in a less developed country than in a developed country. This result highlights the importance of reliable economic data at sufficiently disaggregated regional levels for causal analyses, particularly for less developed countries. Our surface groups proxy provides such data for historical time series and detailed regional levels. For details on the dataset, the methodology, and the results of the UAS analysis, see *Materials and methods* and the [supplementary material](#) (Text S3, Fig. S10, and Table S27).

Surface groups economic proxy

As having one single proxy may be desirable when economic activity is the dependent variable in an analysis, we compute predicted county-level GDP using our OLS model. To assess the external validity of this single-variable proxy, we use one randomly selected quarter of the sample to train the coefficients showing the predictive power of our surface groups proxy, and then for a randomly selected half of our sample compute predicted GDP (see Text S2.5 and Tables S20 and S21 in the [supplementary material](#) for more details). Corroborating the results of the first analysis of external validity, GDP predicted using surface groups explains 63.2% of the variation in actual GDP in the left-out half of the sample, whereas GDP predicted using night light intensity explains only 50.6% of this variation (i.e. 24.9% higher precision). The corresponding values for household income are 67.6% using surface groups vs. 30.9% using night light intensity (i.e. 118.8% higher precision). However, when using the proxy as an independent variable, we recommend using the full set of proxy variables to minimize the noise and measurement error that might come from the predictive process.

Finally, Fig. 3 demonstrates the usefulness of our economic proxy in the historical context by plotting the three-year moving average of administrative GDP and the surface groups proxy. The curves marked with triangles show the extant data for regional economic activity in four regions of Germany—including areas in both East Germany (Rostock, Börde) and West Germany (Groß-Gerau, Passau). The thicker curves without triangles show our single-variable proxy for economic activity (with OLS coefficients trained on the entire sample) for the years for which reliable administrative data are available. First, Fig. 3 shows that as the surface groups are available from 1984, they almost double the number of available years compared to administrative data (which start in 2000 for Germany). Compared to other proxies such as night light intensity (which starts in 1992), surface groups are the only proxy pre-dating the German Reunification. Second, Fig. 3 shows that all curves exhibit identical trends over time,

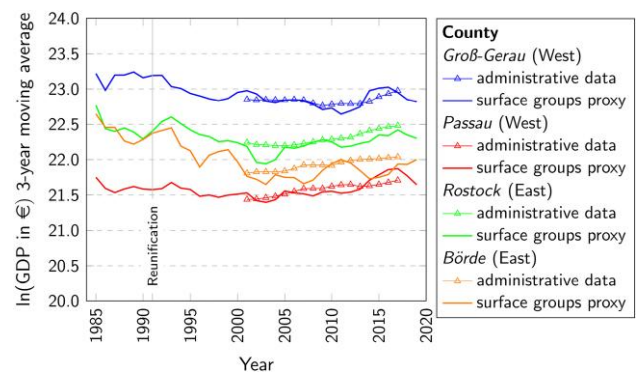


Fig. 3. Time series of GDP measures in four counties. Plots show three-year moving averages. Curves marked by triangles show the natural logarithm of GDP in administrative data for all years for which county-level administrative GDP data are available. Thick curves without triangles show the surface groups proxy for GDP (predicted from OLS estimates, see Text S2.5 and Table S22 in the [supplementary material](#) for details). Groß-Gerau is situated in mid-west Germany, Passau in south Germany (at the border to Austria), Rostock in north Germany (at the Baltic Sea), and Börde in mid-north Germany.

although the variation between years in the surface groups proxy is larger than in the administrative metric. The longer time series and the differences in the developments of the regions over time (e.g. GDP in Börde falls below that in Rostock after reunification) emphasize the proxy's potential for enabling previously impossible analyses.

Furthermore, although reliable, regionally disaggregated data to validate our trends obviously do not exist, we can use other types of historical information to support our claim that the negative trends we find in our economic proxy for the East German regions correspond to real historical developments. Previous literature provides strong quantitative and qualitative support for a declining overall economic trend (in indicators such as employment rate, industrial output, or competitiveness) in East Germany after reunification [e.g. (52–54)]. These findings are thus consistent with the negative trends we find in our disaggregated economic proxy. For later years with sufficiently disaggregated validation data from administrative statistics (see Fig. 3), the trends of our surface groups proxy and the validation data are identical, again supporting the proxy's validity.

Improving precision by combining data sources

Depending on the research purpose, even better proxies can be constructed by combining our surface groups data with additional data sources. As an example, we combine our surface groups with a metric that offers additional information on built-up volume but is available only in five-year intervals (see Texts S2.4 and S2.6 and Tables S13, S14, S23, and S24 in the [supplementary material](#)). As built-up land cover can grow both horizontally and vertically, a metric that uses additional information on built-up volume can increase the precision in proxying economic activity. Our analyses confirm that the combination of surface groups and built-up volume performs well in proxying economic activity in German regions. However, due to the built-up volume metric's availability in five-year intervals only, this combination of datasets does not help in answering research questions that require annual economic data, such as our example of studying immediate economic effects after the fall of the Iron Curtain or similar applications studying local policies with immediate economic consequences. Similar limitations would arise, for example, for a combination

of surface groups and night light intensity, an approach that would not allow studying events before 1992.

Nevertheless, the combination of different metrics with our surface groups proxy to increase precision in proxying economic activity is a powerful tool. Therefore, depending on the regional level and the time series required for a specific research purpose, a proxy that combines our surface groups with other metrics can even outperform the use of single proxies. The methodology for such combinations is provided in the [supplementary material](#) (Text S2.6 and Tables S23 and S24).

Conclusion and discussion

When other data are unreliable, inaccessible, or entirely nonexistent, the proxy we create from daytime satellite imagery is a strong proxy across time periods and across highly disaggregated regional levels (as Fig. 3 demonstrates). Moreover, in this particular example, the proxy provides valuable, previously unavailable information on economic activity for East German regions before the fall of the Iron Curtain.

More generally, our procedure has worldwide relevance. While we apply our procedure to Germany and establish its validity for this country, the procedure is transferable to any region or country in the world (as we demonstrate in [Texts S1.6 and S2.4](#) in the [supplementary material](#)). Our analyses for Germany exemplify that our machine-learning approach using daytime satellite imagery can predict both disaggregated and potentially missing or erroneous economic activity data (e.g. GDP at highly disaggregated levels within a country). However, the methodology and the data it provides for countries across the world can be extended globally to additional contexts where specific economic and developmental markers are needed. Our insight is to demonstrate that our methodology can be helpful for many economic and social science applications where varying degrees of disaggregation are required and where missing or incorrect data are prevalent. Surface groups thus constitute a valuable resource for analyzing historical developments, evaluating local policy reforms, and controlling for economic activity in econometric applications within a country. Although a country's history or industry structure affects the economic importance of different types of land cover (55), the principle that land cover, which the surface groups reflect, relates to economic activity applies to any country in the world. Therefore, surface groups have a potential for economic research that investigates small regions within the same country or within a homogeneous group of countries. Furthermore, surface group measurement could have relevance for such issues as climate change, sustainability, and equity as businesses and policymakers formulate investment and development options for decades to come.

The Landsat daytime satellite data are available for extremely small regional units such as municipalities or urban districts, thus providing new opportunities for urban and regional economic researchers to understand differences in even small regional variation in economic development. The surface groups we derive from these data thus contribute to analyses of the regional impacts of local policy reforms by providing information on economic activity at very detailed regional levels, for which other data sources are entirely unavailable for the necessary observation period, unreliable, less precise, or inaccessible for non-residents of the respective country. With these particular features, the surface groups complement other satellite-based measures for economic activity such as night light intensity.

The use of satellite data is a significant advancement in measuring regional economic activity and over time will generate new

opportunities to strengthen our understanding of local economic conditions. While our paper is one of the first to proxy for economic activity at scale, further improvements are possible. For example, analyzing daytime satellite data with image segmentation procedures based on convolutional neural networks (CNNs) such as U-Net (56) or ResNet (57) could provide an even more accurate classification of land cover and thus a better economic proxy by allowing consideration of contextual information from neighboring pixels in the classification process. Moreover, CNNs could allow researchers to be more discerning about built-up surfaces (e.g. differentiating between building types such as stores or industrial buildings, evaluating housing quality) at a global scale. While CNNs have been successfully applied in classifying land cover for specific geographic study areas [e.g. (58, 59)], their extension to economics could provide new insights. Although the application of CNNs to land cover classification could have higher computing-power demands and may require additional region-specific calibration of ground-truth data [e.g. mentioned in (60, 61)], these additional challenges could be solved by introducing additional instruments such as hyperparameter tuning and model pre-training. Therefore, using CNNs to classify land cover has large potential for future research to investigate regional or even subregional economic activity at a global scale.

Moreover, retrieving more sophisticated metrics on economic activity requires satellite data with an even finer spatial resolution than Landsat data, such as the Advanced Spaceborne Thermal Emissions and Reflection Radiometer (ASTER) or the Sentinel mission. These or other satellite data also offer promising venues for future research, for which this paper lays first methodological foundations. While the ASTER and Sentinel data are less valuable for historical analyses because they cover only substantially shorter time series than Landsat, they are potentially very valuable for research studying more recent events, particularly in areas for which reliable data are otherwise unavailable.

Materials and methods

Computation of surface groups

In developing our procedure for detecting surface groups, we follow the remote-sensing literature that has successfully applied machine-learning techniques to identifying, for example, built-up land cover from subsets of Landsat data [e.g. (23, 62)]. Our procedure adds to this literature by combining data from four Landsat satellites to produce a time series of data on different types of land cover starting in 1984. We produce these data in Google Earth Engine and apply supervised machine-learning techniques with the objective of classifying the annual type of land cover of every Landsat pixel location. We proceed in three steps that we shortly outline here and describe in detail in the [supplementary material](#) (Text S1).

First, we prepare the Landsat data to retrieve the input data for the classification algorithm. We combine the data of Landsat-4, Landsat-5, Landsat-7, and Landsat-8 to produce composite data containing the qualitatively best observation per pixel location and year. In so doing, we choose those observations that best differentiate between vegetated and unvegetated areas, because we expect economic activity to concentrate in urban or industrial areas. The composite data constitute the input data that we pass on to the classification algorithm.

Second, to be able to classify observations in the input data, we add CLC data as an external source of ground-truth information. This ground-truth dataset comes from a pan-European project commissioned by the European Environment Agency and maps

land cover in 44 categories. To obtain a classification of land cover types that we can use to train our algorithm, we survey the literature that uses CLC data or Landsat data for classifying land cover [e.g. (35–37)] and aggregate the 44 categories to larger groups with similar surface characteristics—the six surface groups. The classification algorithm requires this ground-truth information on surface groups for a subset of the input pixels to be able to recognize patterns in the input data and link these patterns to the different surface groups. By using external ground-truth data, we overcome the resource-intensive necessity of visually interpreting (i.e. manually classifying) input pixels to retrieve ground-truth information.

Third, we produce the training data for the classification algorithm. To obtain these training data, we draw a stratified random sample of pixels from the input data and match the ground-truth information on surface groups to the pixels in this sample. We then use the training data to train a Random Forest algorithm, which classifies every observation in the input data into one of the six surface groups.

Although we apply various filters for excluding invalid Landsat pixels (e.g. cloud shadow) from the composite input data, potentially erroneous pixel classifications might occur in few regions in years with scarce Landsat imagery (particularly in the 1980s). When applying our surface groups proxy in empirical analyses, we recommend removing outlier observations for these particular regions and years from these analyses. We do so in the comparison of county-level GDP developments in Fig. 3 and in the municipality-level analysis of higher education institutions. From 1984 through 2020, we identify 6.2% of all county-year observations and 8.4% of all municipality-year observations as outliers, which are independent of the analyses (e.g. independent of the locations of higher education institutions). For more details on this outlier removal, see the [supplementary material \(Texts S2.5 and S3\)](#).

External validity analyses

We obtain two indicators of regional economic activity for the external validity analyses, which we shortly outline here and describe in more detail in the [supplementary material \(Text S2\)](#). First, we use administrative GDP, which the German Federal Statistical Office provides at the county-level from 2000 and which we deflate for our analyses. Second, we use RWI-GEO-GRID (43), a dataset containing socioeconomic indicators collected from a variety of public and private sources but annually only available from 2009. This dataset indicates household income at the level of grid cells sized 1 km², an extremely high level of regional detail. Again, we use deflated household income for our analyses.

In addition, to compare the quality of the surface groups as a proxy for economic activity to that of night light intensity, we use night lights data from the U.S. Air Force Defense Meteorological Satellite Program Operational Linescan System (DMSP OLS), available from 1992 through 2013. Similar to previous research (5, 42), we use stable night lights (which are corrected for unusual lighting). To achieve regional correspondence with the administrative GDP data and RWI-GEO-GRID, we calculate average night light intensity at the county and at the grid level. In the [supplementary material \(Text S2.4\)](#), we proceed similarly for comparing the surface groups proxy to Visible Infrared Imaging Radiometer Suite (VIIRS) night light intensity.

Application to analysis of higher education institutions

We use data from prior work (49) that combines patent data from the European Patent Office's Worldwide Patent Statistical

Database (October 2019 version) with self-collected data on campus openings. These data contain two established patent-based indicators for regional innovation (patent quantity and patent quality) and each municipality's annual treatment status. We describe the details on these data and on the methodology used in our analysis in the [supplementary material \(Text S3\)](#).

Combination of surface groups and built-up volume

In our combination of data sources, we use data from the Global Human Settlement Layer (GHSL). Among other things, these data include information on regional built-up surfaces and built-up volume in five-year intervals. We describe the details on these data and how we use them for our analyses in the [supplementary material \(Text S2\)](#).

Acknowledgments

We thank participants of the World Congress of the Regional Science Association International in Marrakech, the Meeting of the Economics of Education Association in Zaragoza, the European Regional Science Association Congress in Bolzano, the Annual Conference of the Verein für Socialpolitik in Cologne, and participants of seminars at the University of Zurich, the Julius-Maximilians-Universität Würzburg, the Max Planck Institute for Innovation and Competition in Munich, and the ZEW Leibniz Center for European Economic Research in Mannheim. We thank Simone Balestra, Thomas Dohmen, Tor Eriksson, David Figlio, Dietmar Harhoff, Simon Janssen, Philip Jörg, Daniel Kükenbrink, Edward Lazear, Mark Long, Jens Mohrenweiser, Guido Neidhöfer, Harald Pfeifer, Natalie Reid, Michael E. Rose, Monika Schnitzer, Dinand Webbink, and Niels Westergård-Nielsen for helpful comments. Furthermore, this paper benefited greatly from very helpful comments by three anonymous reviewers.

Supplementary material

[Supplementary material](#) is available at PNAS Nexus online.

Funding

This study was partly funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) through its “Leading House VPET-ECON: A Research Center on the Economics of Education, Firm Behavior and Training Policies” (P.L. and M.N., grants 1315002234 and 1315000868).

Authors' contributions

P.L.: conceptualization, data curation, methodology and formal analysis, validation and visualization, writing (draft and editing); M.N.: data curation, methodology and software, visualization; U.B.-G.: conceptualization, funding acquisition, supervision, writing (draft, review, and editing); E.B.: methodology and validation, writing (review and editing). All authors have contributed fair shares, with P.L. having the largest contribution and M.N. the second largest. The order of authors listed in the manuscript has been approved by all authors.

Previous presentations

These results were previously presented at the Annual Conference of the Verein für Socialpolitik (VfS) in Cologne

(September 2020), at the World Congress of the Regional Science Association International (RSAI) in Marrakech (May 2021), at the Meeting of the Economics of Education Association (AEDE) in Zaragoza (July 2021), at the European Regional Science Association (ERSA) Congress in Bolzano (August 2021), at seminars at the University of Zurich (January 2020, February 2020), at a seminar the Max Planck Institute for Innovation and Competition in Munich (May 2022), and at a seminar at the Leibniz Center for European Economic Research (ZEW) in Mannheim (May 2022).

Preprints

Working paper versions of this article are available in the Swiss *Leading House "Economics of Education" Working Paper* series and in the *IZA Discussion Paper* series. A version of this article is part of a chapter in P.L.'s dissertation.

Data availability

The code developed in this paper and the surface groups used for the analyses are available in a GitHub repository (<https://github.com/Neduzen/lhvpetecon-surface>). Surface groups data for other countries are made available as georeferenced Tagged Image Format (TIF) files to the scientific community via SWISSUbase (see the GitHub repository for download links), with the goal of worldwide coverage. The underlying Landsat satellite data are publicly available through Google Earth Engine (free of charge for research, education, and nonprofit use), so that interested researchers can independently investigate their regions of interest. Night light intensity data are publicly available from the National Oceanic and Atmospheric Administration (DMSP OLS) and from the Colorado School of Mines (VIIRS). The GDP data used in the external validity analyses are publicly available from the German Federal Statistical Office, whereas the household income data are available for a fee upon request only from the Research Data Center Ruhr at the Leibniz Institute for Economic Research (RWI). An extract of the data on higher education institutions and regional patenting from prior work (49) is available in the GitHub repository. The data from prior work in Africa used for additional analyses in the supplementary material are available as a supplement to this work (32). The GHSL data are publicly available from the European Commission. All data sources and weblinks for access are documented in the supplementary material.

References

- Dingel JI, Miscio A, Davis DR. 2021. Cities, lights, and skills in developing economies. *J Urban Econ.* 125:103174.
- Hodler R, Raschky PA. 2014. Regional favoritism. *Q J Econ.* 129(2): 995–1033.
- Michalopoulos S, Papaioannou E. 2013. Pre-colonial ethnic institutions and contemporary African development. *Econometrica.* 81(1):113–152.
- Pinkovskiy M, Sala-i-Martin X. 2016. Lights, camera ...income! Illuminating the national accounts–household surveys debate. *Q J Econ.* 131(2):579–631.
- Chen X, Nordhaus WD. 2011. Using luminosity data as a proxy for economic statistics. *Proc Natl Acad Sci USA.* 108(21):8589–8594.
- Kulkarni R, Haynes K, Stough R, Riggle J. 2011. Light based growth indicator (LGBI): exploratory analysis of developing a proxy for local economic growth based on night lights. *Reg Sci Policy Pract.* 3(2):101–113.
- Mellander C, Lobo J, Stolarick K, Matheson Z. 2015. Night-time light data: a good proxy measure for economic activity? *PLoS ONE.* 10(10):e0139779.
- Burke M, Driscoll A, Lobell DB, Ermon S. 2021. Using satellite imagery to understand and promote sustainable development. *Science.* 371(1219):eabe8628.
- Lehnert P. 2020. Higher education institutions and their impact on employment and innovation: regional identification and empirical analyses [dissertation]. Zurich: University of Zurich.
- Burchfield M, Overman HG, Puga D, Turner MA. 2006. Causes of sprawl: a portrait from space. *Q J Econ.* 121(2):587–633.
- Foster AD, Rosenzweig MR. 2003. Economic growth and the rise of forests. *Q J Econ.* 118(2):601–637.
- Gorelick N, et al. 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. *Remote Sens Environ.* 202:18–27.
- Donaldson D, Storeygard A. 2016. The view from above: applications of satellite data in economics. *J Econ Perspect.* 30(4):171–198.
- Morain SA. 1998. A brief history of remote sensing applications, with emphasis on Landsat. In: Liverman D, Moran EF, Rindfuss RR, Stern PC, editors. *People and pixels: linking remote sensing and social science.* Washington, DC: The National Academies Press. p. 28–50.
- Williams DL, Goward S, Arvidson T. 2006. Landsat: yesterday, today, and tomorrow. *Photogramm Eng Remote Sens.* 72(10): 1171–1178.
- Bruzzzone L, Serpico SB. 1997. An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images. *IEEE Trans Geosci Remote Sens.* 35(4):858–867.
- Fung T. 1992. Land use and land cover change detection with Landsat MSS and SPOT HRV data in Hong Kong. *Geocarto Int.* 7(3):33–40.
- Gautam NC, Chennaiah GC. 1985. Land-use and land-cover mapping and change detection in Tripura using satellite LANDSAT data. *Int J Remote Sens.* 6(3–4):517–528.
- Stauffer ML, McKinney RL. 1978. LANDSAT image differencing as an automated land cover change detection technique. Silver Spring: Computer Sciences Corp. Technical Report Contract NAS 5-243500, Task Assignment 206.
- Ton J, Sticklen J, Jain AK. 1991. Knowledge-based segmentation of Landsat images. *IEEE Trans Geosci Remote Sens.* 29(2):222–232.
- Dewan AM, Yamaguchi Y. 2009. Land use and land cover in Greater Dhaka, Bangladesh: using remote sensing to promote sustainable urbanization. *Appl Geogr.* 29(3):390–401.
- Goldblatt R, You W, Hanson G, Khandelwal AK. 2016. Detecting the boundaries of urban areas in India: a dataset for pixel-based image classification in Google Earth Engine. *Remote Sens (Basel).* 8: 634.
- Liu X, et al. 2018. High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine platform. *Remote Sens Environ.* 209:227–239.
- Pekkarinen A, Reithmaier L, Strobl P. 2009. Pan-European forest/non-forest mapping with Landsat ETM+ and CORINE Land Cover 2000 data. *ISPRS J Photogramm Remote Sens.* 64(2):171–183.
- Yu W, Zang S, Wu C, Liu W, Na X. 2011. Analyzing and modeling land use land cover change (LUCC) in the Daqing City, China. *Appl Geogr.* 31(2):600–608.
- Keola S, Andersson M, Hall O. 2015. Monitoring economic development from space: using nighttime light and land cover data to measure economic growth. *World Dev.* 66:322–334.
- Sutton PC, Costanza R. 2002. Global estimates of market and non-market values derived from nighttime satellite imagery,

- land cover, and ecosystem service valuation. *Ecol Econ.* 41(3): 509–527.
- 28 Davis MA, Fisher JDM, Whited TM. 2014. Macroeconomic implications of agglomeration. *Econometrica.* 82(2):731–764.
- 29 Holl A. 2004. Manufacturing location and impacts of road transport infrastructure: empirical evidence from Spain. *Reg Sci Urban Econ.* 34(3):341–363.
- 30 Goldblatt R, Heilmann K, Vaizman Y. 2020. Can medium-resolution satellite imagery measure economic activity at small geographies? Evidence from Landsat in Vietnam. *World Bank Econ Rev.* 34(3):635–653.
- 31 Chen C, et al. 2020. Analysis of regional economic development based on land use and land cover change information derived from Landsat imagery. *Sci Rep.* 10:12721.
- 32 Yeh C, et al. 2020. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nat Commun.* 11:2583.
- 33 Frey BS, Moser L, Bieri S. 2022. When do governments manipulate official statistics? An empirical analysis. <http://dx.doi.org/10.2139/ssrn.4244682>
- 34 Martínez LR. 2022. How much should we trust the dictator's GDP estimates? *J Political Econ.* 130(10):2731–2769.
- 35 Balzter H, Cole B, Thiel C, Schmullius C. 2015. Mapping CORINE land cover from Sentinel-1A SAR and SRTM digital elevation model data using random forests. *Remote Sens (Basel).* 7(11): 14876–14898.
- 36 Han K-S, Champeaux J-L, Roujean J-L. 2004. A land cover classification product over France at 1 km resolution using SPOT4/VEGETATION data. *Remote Sens Environ.* 92(1):52–66.
- 37 Neumann K, Herold M, Hartley A, Schmullius C. 2007. Comparative assessment of CORINE2000 and GLC2000: spatial analysis of land cover data for Europe. *Int J Appl Earth Obs Geoinf.* 9(4):425–437.
- 38 Pérez-Hoyos A, García-Haro FJ, San-Miguel-Ayanz J. 2012. A methodology to generate a synergetic land-cover map by fusion of different land-cover products. *Int J Appl Earth Obs Geoinf.* 19: 72–87.
- 39 Waser LT, Schwarz M. 2006. Comparison of large-area land cover products with national forest inventories and CORINE land cover in the European Alps. *Int J Appl Earth Obs Geoinf.* 8(3):196–207.
- 40 Esri, Maxar, Earthstar Geographics, USDA FSA, USGS, AeroGrid, IGN, IGP, GIS User Community. 2009. World imagery (updated January 14, 2022). Redlands (CA): Dataset, Esri.
- 41 Goldblatt R, et al. 2018. Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover. *Remote Sens Environ.* 205:253–275.
- 42 Henderson JV, Storeygard A, Weil DN. 2012. Measuring economic growth from outer space. *Am Econ Rev.* 102(2):994–1028.
- 43 Leibniz Institute for Economic Research (RWI) and Micromarketing-Systeme and Consult GmbH (microm). 2019. RWI-GEO-GRID: socio-economic data on grid level—scientific use file (wave 8). version: 1. Essen: Dataset, RWI.
- 44 Andrews M. 2020. How do institutions of higher education affect local invention? Evidence from the establishment UF U.S. Colleges. <http://dx.doi.org/10.2139/ssrn.3072565>
- 45 Cowan R, Zinovyeva N. 2013. University effects on regional innovation. *Res Policy.* 42(3):788–800.
- 46 Toivanen O, Väänänen L. 2016. Education and invention. *Rev Econ Stat.* 98(2):382–396.
- 47 Dickey H, Widmaier AM. 2021. The persistent pay gap between Easterners and Westerners in Germany: a quarter-century after reunification. *Pap Reg Sci.* 100(3):604–631.
- 48 Schnabel C. 2016. United, yet apart? A note on persistent labour market differences between western and eastern Germany. *J Econ Stat.* 236(2):157–179.
- 49 Lehnert P, Pfister C, Harhoff D, Backes-Gellner U. 2022. Innovation effects and knowledge complementarities in a diverse research landscape. Zurich: Swiss Leading House VPET-ECON. Leading House “Economics of Education” Working Paper No. 164.
- 50 Fritsch M, Aamoucke R. 2017. Fields of knowledge in higher education institutions, and innovative start-ups: an empirical investigation. *Pap Reg Sci.* 96(S1):S1–S27.
- 51 Pfister C, Koomen M, Harhoff D, Backes-Gellner U. 2021. Regional innovation effects of applied research institutions. *Res Policy.* 50(4):104197.
- 52 Akerlof GA, et al. 1991. East Germany in from the cold: the economic aftermath of currency union. *Brookings Pap Econ Act.* 1991(1):1–105.
- 53 Burda M, Hunt J. 2001. From reunification to economic integration: productivity and the labor market in Eastern Germany. *Brookings Pap Econ Act.* 2001(2):1–92.
- 54 Franz W, Steiner V. 2000. Wages in the East German transition process: facts and explanations. *Ger Econ Rev.* 1(3):241–269.
- 55 Henderson JV, Squires T, Storeygard A, Weil D. 2018. The global distribution of economic activity: nature, history, and the role of trade. *Q J Econ.* 133(1):357–406.
- 56 Ronneberger O, Fischer P, Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. Medical image computing and computer-assisted intervention—MICCAI 2015. Lecture Notes in Computer Science, Vol. 9351. Munich: Springer. p. 234–241.
- 57 He K, Zhang X, Ren S, Sun J. 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas. p. 770–778.
- 58 Wagner FH, et al. 2019. Using the U-Net convolutional neural network to map forest types and disturbance in the Atlantic rainforest with very high resolution images. *Remote Sens Ecol Conserv.* 5(4):360–375.
- 59 Wang M, Zhang X, Niu X, Wang F, Zhang X. 2019. Scene classification of high-resolution remotely sensed image based on ResNet. *J Geovis Spat Anal.* 3:16.
- 60 Boston D, Van Dijk A, Rozas Larraondo P, Thackway R. 2022. Comparing CNNs and random forests for Landsat image segmentation trained on a large proxy land cover dataset. *Remote Sens (Basel).* 14(14):3396.
- 61 Latifovic R, Pouliot D, Campbell J. 2018. Assessment of convolutional neural networks for surficial geology mapping in the South Rae Geological Region, Northwest Territories, Canada. *Remote Sens (Basel).* 10(2):307.
- 62 Schneider A. 2012. Monitoring land cover change in urban and peri-urban areas using dense time stacks of Landsat satellite data and a data mining approach. *Remote Sens Environ.* 124:689–704.