Global GDP Prediction With Night-Lights and Transfer Learning

Nathan Price and Peter M. Atkinson

Abstract—Nighttime lights (night-lights) data are useful in predicting gross domestic product (GDP), a key economic indicator used by policymakers and economists. A persistent problem in such prediction is that night-lights under-represent economic activity in rural areas. Attempting to disaggregate night-lights using urban and rural regions is problematic, as the urban–rural dichotomy is increasingly tenuous due to changing economic structures. In response, this article presents a regionalization approach, which is data-driven. Utilizing transfer learning, we trained a model that takes fine spatial resolution daytime satellite sensor imagery and learns an optimal regionalization to disaggregate visible infrared imaging radiometer suite (VIIRS) night-lights for GDP prediction. To make national scale inference feasible, we formulate a novel Monte Carlo importance sampling scheme, and then performed a single-year cross-sectional study across 178 countries using 178 000 images. This achieved an \( R^2 \) between predicted and actual \( \log_{10} \) GDP of 0.86 on the validation set and 0.89 on the whole study area. To benchmark, we perform a subnational study over 396 U.S. counties using 98 500 images in which our method outperformed comparable methods. Interpreting the regionalization, we found that the utility of the urban–rural dichotomy is not supported by the model and argue that seeing the night-lights of some land types as representative of the overall economy is a better way to understand the model.

Index Terms—Convolutional neural networks (CNNs), deep learning (DL), economic indicators, machine learning, Monte Carlo methods, night-lights, regionalization, remote sensing, transfer learning, urban areas, visible infrared imaging radiometer suite (VIIRS).

I. INTRODUCTION

It has been known since the pioneering work of Elvidge et al. [1] (though first noted by Croft [2]) that nighttime illumination, as measured through satellite sensor imagery (also known as night-lights), provides a noisy proxy for economic activity and can be used to predict gross domestic product (GDP) using only traditional remote sensing methods. Compared to National Statistical Office (NSO)-produced GDP figures, which require large costly surveys and typically take months to produce [3], night-light-produced GDP predictions are an attractive alternative for both government and financial institutions.

Manuscript received 27 April 2022; revised 6 July 2022; accepted 10 August 2022. Date of publication 22 August 2022; date of current version 5 September 2022. This work was supported by the Lancaster Environment Centre, Lancaster University. (Corresponding author: Nathan Price.)

Nathan Price is with the Lancaster Environment Center, Lancaster University, LA1 4YQ Lancaster, U.K. (e-mail: n.price1@lancaster.ac.uk).

Peter M. Atkinson is with the Faculty of Science and Technology, Lancaster University, LA1 4YR Lancaster, U.K., and also with the Department of Geography and Environment, University of Southampton, SO17 1BJ Southampton, U.K. (e-mail: pma@lancaster.ac.uk).

Digital Object Identifier 10.1109/JSTARS.2022.3200754

Night-lights have the advantage of being free, public-domain, globally consistent, and ubiquitous in time and space, and have periodic regular releases (e.g., monthly and annually) [4]. The night-lights datasets can be available within a few months of the reference period and can be used to predict GDP within days or even hours of the data’s release. Were the precision of night-light GDP predictions comparable to that of NSO GDP figures, Night-light GDP figures could be used for GDP nowcasting and as an affordable alternative to surveys for developing countries. However, night-lights are a poor proxy for economic activity in rural areas\(^1\) [5], and in particular, they underestimate the productivity of unlit areas [6], which are notably common in many developing nations with large numbers of subsistence farmers [7].

Recognizing that the relationship between night-lights and economic activity is nonlinear and is conditioned upon geographic factors, it is sensible to adopt a nonlinear model, which can account for such heterogeneity. A simple approach is to regionalize whereby the relationship between night-lights and economic activity is treated differently in different “regions.” In a time series setting, it is possible to treat each country separately using a country fixed-effect [8], [9]; but for prediction of GDP based on single-year cross-sectional data, a different sort of regionalization must be used. One approach is to distinguish between urban and rural regions and treat night-lights separately in each, a method that has been used in subnational GDP prediction [10], but also in other settings, such as predicting electrification from night-lights [11]. While useful, the rural–urban distinction is under increasing scrutiny, especially with its lack of regard for rural regions with urban function [12], [13], [14]. Such areas are commonly classified as rural and yet are better treated as urban. Even a consistent rigorous definition of “rural” and “urban” may be difficult to achieve with organizations using differing definitions [15]. Nevertheless, regionalization is a potentially useful tool if appropriate regions can be identified.

One way of addressing the challenge of identifying an appropriate regionalization is a dynamic approach where data and predictive performance inform what constitutes the “best” regionalization. To do this meaningfully, a supplementary data source other than night-lights is needed. Ideally, this data source should preserve the advantages of night-lights (i.e., ubiquity, public-domain, and global consistency). An obvious candidate is land cover data. Sun et al. [16] applied a 1-D convolutional neural network (CNN) on night-light histograms paired with

---

\(^{1}\)For example, Gibson et al. [5] showed that, in Indonesia, the \( R^2 \) of \( \log_{10} \) GDP regressed on \( \log_{10} \) night-lights is around 0.68 in urban areas and 0.01 in rural areas. They also found similar numbers for China.
land cover histograms to predict the GDP of the contiguous USA at the county level for a four-year period. They achieved an $R^2$ averaged across years of 0.83, which was an increase on the traditional remote sensing night-lights method ($R^2$ averaged across years of 0.72). Despite such success, land cover data generally define built-up areas based on the presence of human-made structures [17, Table 3.3], a classification lacking in certain information. For this reason, there is potential to utilize more information-rich data sources.

An attractive alternative is daytime satellite sensor imagery. Information-rich, ubiquitous, public-domain, and globally consistent daytime satellite sensor imagery has the properties we desire, but has the drawbacks of massive data volume, high dimensionality, and complex nonlinear relationships with night-lights and economic activity. This is a textbook use case for the application of machine learning, particularly, deep learning (DL) [18], [19]. The relevance of daytime satellite sensor imagery to the task at hand has been confirmed by Liu et al. [20], who used a CNN trained on such imagery to predict GDP at the county level for China ($R^2 = 0.71$), and Khachiyan et al. [21], who paired a similar CNN with socioeconomic indicators to predict total personal income at fine spatial resolutions in the USA ($R^2 = 0.90$), though neither of these examples took advantage of night-lights data.

In this article, we use transfer learning to adapt an existing CNN to the task of regionalization for the purpose of more precise night-light-produced GDP predictions globally for a single year. Transfer learning, an increasingly popular and effective method to solve modeling problems in remote sensing DL applications [22], is a method, whereby a DL model trained to perform a source task can be modified to perform a target task by preserving the knowledge accumulated in learning the source task [23],[2] We use this to repurpose the feature extractor presented by Jean et al. [29], which was originally used to predict poverty from daytime satellite sensor imagery. Transfer learning is appropriate in this setting for its efficiency, which means that the small number of GDP observations—no more than one for each country for the study year—is not an impediment.

A schematic diagram giving an overview of the method is shown in Fig. 1. Code for this paper is available at https://github.com/NathanJCP/Global GDP prediction_minimal implementation.

II. REGIONALIZATION

A feature extractor is any map $f : \mathcal{X} \rightarrow \mathbb{R}^d$ mapping any input to a $d$-dimensional vector. In the case of this research, $\mathcal{X}$ will be the world’s surface, each point of which has an associated daytime satellite sensor image and $f$ will be a CNN that maps that image to a learned representation. Given a partition of the representation space $\mathbb{R}^d = \bigcup_i A_i$, the preimage of the partition gives a regionalization of the world’s surface $\mathcal{X} = \bigcup_i f^{-1}(A_i)$. Each region in each country will then have a separate night-light total, which will be used in a regression model. If the representation space holds semantic information (which is one of the underlying assumptions of transfer learning), then there should be a meaningful way to partition the representation space.

III. METHODS

A. Study Area

This research utilized a single-year cross-sectional study area. It was decided to work with 2019 as the study year, due to constraints on accessing archived aerial imagery through the Google Static Maps API platform and as the 2019 GDP figures were the most recent GDP figures available when the research commenced. It also had the advantage of being before national lockdowns in response to COVID-19 during 2020, which created a high degree of uncertainty around GDP figures for 2020.

It was decided to predict GDP at the national scale, both because the national scale is the most important scale for policymakers and investors, but also because securing subnational GDP figures for many countries is difficult or impossible, although the method presented should eventually be scalable to subnational scales. Thus, the study area was global with 178 countries selected, and 17 countries were not included, due to the lack of GDP data in 2019. Kiribati was also excluded due to its complex geography, which rendered a majority of sampled satellite sensor images being uninformative pictures of the ocean.

B. Transfer Learning

Here, 178 countries, each with one GDP figure for 2019, amounted to 178 observations. Typically, to train a neural network, many more labeled observations are needed. Transfer learning is applicable in this setting.

Transfer learning is the process whereby, given a neural network $f_0 : \mathcal{X} \rightarrow \mathcal{Y}_0$ trained on a task $T_0$, by factoring $f_0$ into two models $f_0(x) = g_0(h(x))$, then we can use reuse $h : \mathcal{X} \rightarrow \mathbb{R}^n$ in solving a similar task $T_1$. Specifically, if task $T_1$ shares the feature space $\mathcal{X}$, we can learn a new, generally smaller, model $g_1 : \mathbb{R}^n \rightarrow \mathcal{Y}$ so that $f_1(x) = g_1(h(x))$ solves task $T_1$ [23]. Importantly for us, $g_1$ does not have to be a neural network. We call $h$ a feature extractor, as it represents an image as a vector of semantically rich features.

Much as human learning, transfer learning is effective when the tasks are closely related. Transfer learning is advantageous because of the preservation of useful information from one task

---

2To illustrate the growing popularity of DL in remote sensing, some recent examples include [24], [25], [26], [27], and [28].
to the next, which means that less data and less computation are needed for training. Specifically, this is useful when data for $T_0$ are abundant and data for $T_1$ are limited [23].

The task to be solved was regionalization. As discussed earlier, a partitioning of the feature space induces a regionalization onto the Earth’s surface backwards through the feature extractor.

As the starting model, we used the CNN trained by Jean et al. [29], which was desirable for the relevance of the economic task it was trained to perform. Other choices of starting models were available. This CNN takes daytime satellite sensor imagery, a data source rich in information not captured in night-lights but economically informative [29].

To avoid overparameterizing, $g_1$ was desired to have few learnable parameters. We worked with a composition of a principal components analysis (PCA) linear map $P : \mathbb{R}^n \to \mathbb{R}^d$ and a linear classifier $L$ so that $g_1 = LP$. The details of these will be discussed later.

Although the primary motive for using a linear classifier is that it uses few parameters and is a standard “out of the box” method, similar methods are already established in the literature. For example, support vector machines have seen growing popularity in remote sensing and have been utilized effectively in satellite sensor image classification problems [30]. Supposing that the conceptual framework is semantically meaningful and that the feature extractor does contain relevant information to the task, then after the PCA removes colinearity, the regions should be linearly separable, justifying the use of a linear classifier.

C. Data

1) World Bank GDP Figures: For this research, GDP at purchaser’s prices as sourced from the World Bank National Accounts data and OECD National Accounts data files, with all figures in current US$ for 2019, were used. These GDP figures are primarily the product of NSOs.

Rather than predicting GDP, the main focus was on predicting $\log_{10}$GDP as other authors have done [5], [16]. Studying $\log_{10}$GDP has distinct theoretical advantages. First, it is normally distributed; a Shapiro–Wilk test failed to reject the null hypothesis that $\log_{10}$GDP is not normally distributed with $p = 0.930$. This is desirable as trying to predict a log-normally distributed variable using ordinary least squares regression has the effect that a very few large observations are inordinately impactful on the parameters of the model. Second, economists and policymakers take great interest in the percentage change in GDP, so homoscedasticity means an invariant mean percentage error. The distribution of $\log_{10}$GDP is shown in Fig. 2.

NSO GDP figures have data quality problems worth addressing. For developing nations, the quality of NSO GDP figures varies. Missing data, imputations based on outdated surveys, and infrequent rebasings increase the likelihood of underestimation of sectors and for best practices to be unattainable in developing countries [3], [31]. Errors are compounded by data gaps filled using spurious methodologies and political pressures that have de-emphasized the importance of standardized national accounts [32]. The problem comes into stark focus when we consider the problem of GDP series rebasings, which can cause sizable revisions in NSO GDP figures. Two particularly egregious examples are: a) Ghana’s 2010 rebasing, with a revised figure over 60% larger [31]; and b) Nigeria’s 2014 rebasing, with a revised figure 89% larger [33]. These are exceptional, but they illustrate the huge errors that can accumulate even in NSO GDP figures. Then, there is political pressure that can lead to artificially inflated figures, which is more likely to occur in nondemocratic countries [9]. This is not to say that all developing or nondemocratic countries have these issues, but to note that these issues are real and this can limit the minimum achievable error for a regression model.

2) VIIRS-DNB: The visible infrared imaging radiometer suite (VIIRS) day/night band (DNB) nighttime lights data, currently maintained and provided by the Earth Observation Group, Payne Institute for Public Policy, are gaining slow acceptance as the gold standard for economic study through night-lights [5], [34]. For this research, the VIIRS-DNB version 1 February 2020 monthly average product was used, as the 2019 yearly average product was unavailable at the time of acquisition. Furthermore, it was felt that the end of year night-lights would be more representative of 2019 GDP than midyear night-lights, although it was desirable to avoid the brighter night-lights associated with the holiday season [35], which can begin in November and extend into January in many of the world’s countries. This night-lights product is a (tiled) raster of the Earth’s surface where each pixel corresponds to the average nocturnal radiance for a 15 arcsecond (~450 m) grid square. Before averaging, each image is masked to exclude clouds and corrected for stray light, lightning, and lunar illumination. This provides a high-quality globally consistent dataset of average radiance values (the units of which are nW cm$^{-2}$ sr$^{-1}$) [36]. An exception in spatial resolution was made for the Russian Federation where we downsampled the imagery to a spatial resolution of 90 arcseconds (~2700 m) per pixel so that the country raster could fit into memory on a Google Colab session.

It is worth noting that the VIIRS sensor’s approximately 1 A.M. overpass time [37]. A consequence of this is that very little residential light is represented in the VIIRS data, most of the light from urban areas corresponds to streetlights. While this might
be considered less desirable than DSMP-OLS’s overpass time of approximately 7.30 p.m., the benefits of VIIRS far outweigh any drawbacks. For an extended discussion of the benefits of VIIRS over DSMP-OLS, we refer the reader to Elvidge et al.’s [38] comparison of the platforms.

3) Google Static Maps API Daytime Satellite Sensor Imagery: Daytime aerial imagery was sourced from the Google Static Maps API at zoom level 14, which is a fine spatial resolution of \( \sim 2.5 \) m. All images used were 224 \( \times \) 224 px large, 32-b depth full color and were accessed and downloaded on 16 September 2020. Google’s aerial imagery consists of a mosaic of imagery from multiple sources, including aerial surveys and satellite sensor imagery, updated regularly with new imagery [39]. Consequently, an exact date of image acquisition cannot be given, but most structures visible at this spatial resolution, such as buildings, infrastructure, farm structures, and permanent landscape features, change very little across the course of a few months, so it can be expected that Google’s daytime aerial imagery is a good proxy for the late 2019 ground reference data. Vegetation does change across the course of months, but the assumption was made that the variation in the appearance of a patch due to seasonal variation in vegetation is, on average, not so pronounced as to make a patch appear to be of a land type of a dramatically different economic character to that of its actual character.

As imagery has varying sources, images were preprocessed to normalize color channels to ensure comparability. A small fraction of images did contain artifacts from the mosaic process (e.g., a discernible edge where contiguous images were joined), and imagery resolution was occasionally coarse in remote forested areas, but the occurrence of these data issues was judged to be infrequent enough to be negligible.

For each of the 178 countries, 1000 fine spatial resolution images were acquired, which gave 178 000 images for partitioning the representation space. Extra 17 000 images for those countries without GDP figures were included in the PCA to avoid biasing the model toward those countries for which data were obtained.

4) World Country Boundaries: To delineate between countries, the Esri’s [40] World Countries (generalized) dataset was used for its accurate borders at large cartographic scales, which ensured that cities by the sea were not incorrectly clipped when calculating in-country radiances.

D. Feature Extractor

It was decided to use a CNN as the feature extractor due to their suitability for image processing and wide-ranging use in remote sensing [18]. The main competitors were Deepsat V2, a popular satellite sensor image classification framework incorporating hand-crafted features [41], and the model of Jean et al. [29], which utilized a CNN to predict poverty based upon daytime satellite sensor imagery. While the hand-crafted features of Deepsat V2 circumvent the limited capability of CNNs to learn Haralick features [42], this was not enough of a strength to prefer it over the model of Jean et al., which had the advantages of a larger context window (224 \( \times \) 224 px versus Deepsat V2’s 28 \( \times \) 28 px) and a preferable spatial resolution (2.5 m per pixel versus Deepsat V2’s 1 m per pixel) as well as closer task relevance.

We adapted Mathur’s [43] 2020 Pytorch reimplementation of Jean et al.’s [44] CNN, which is based on the VGG-F architecture, an architecture first presented by Chatfield et al. [45], who based it on the VGG architecture of Simonyan and Zisserman [46], an architecture popular for its high prediction accuracy on Imagenet and its open-source weights. This CNN was originally trained to predict discretized nighttime illumination classifications for five African countries: 1) Nigeria; 2) Tanzania; 3) Uganda; 4) Malawi; and 5) Rwanda, though de Lima and Marfurt [22] suggested that it would still generalize well. After removing the last layer, the feature extractor was a seven-layer CNN. This feature extractor takes in full color daytime satellite sensor imagery \( 224 \times 224 \) px large with a 2.5-m spatial resolution as 4096-D feature vectors.

E. Importance Sampling Scheme

It was infeasible to download and run inference on imagery at the chosen spatial resolution for the whole of the world’s land surface, both in terms of cost and computational time and resource.

To overcome this problem, if GDP for a country \( C \) is written as \( \text{GDP}_C = \int_{\mathcal{C}} g(x) \), where \( g \) is a theoretical function that gives the economic activity at each latitude–longitude point \( x \), then Monte Carlo sampling methods can be applied to estimate the value of the integral. Two practical challenges immediately presented themselves as follows.

1) Many countries in the world contain large wilderness expanses. For example, 80% of Algeria’s land surface is covered by the Sahara Desert [47]; so if random uniform sampling was adopted, then 80% of the sample would be pictures of desert sands.

2) The function of \( g(x) \) is spiky with most of the mass concentrated in urban areas. Thus, a uniform sampling scheme would be inefficient.

To address these problems, an importance sampling Monte Carlo scheme was used. The sampling scheme operated as follows.

For a pixel \( i \) of the VIIRS night-light imagery for a country \( C \), let \( r_i \) denote the measured monthly average pixel radiance. Also, for a pair of coordinates, \( x = (\text{lat, lon}) \) denote the pixel that \( x \) lies in by \( i_x \). It was desirable to choose locations with probability proportional to \( r_i \); but to give every pixel a chance of selection, we introduced a parameter \( \epsilon > 0 \)—for which we choose a value of \( \epsilon = 0.01 \)—and set \( r_i' = \max(r_i, \epsilon) \). Locations were selected in the following two stages.

1) A pixel \( i \) of the VIIRS night-light imagery for \( C \) was selected such that each pixel \( i \) in \( C \) had a selection probability \( p_i \propto r_i' \).

2) A pixel defined a rectangle on the Earth’s surface. Within that rectangle, we uniformly sampled a pair of coordinates \( x \).

The abovementioned strategy is efficient, because \( r'(x) = r_i \) is closely correlated with \( g(x) \) and so \( p(x) \) was close to the
optimal choice for the sample distribution [48, §5.7]. The Monte Carlo estimator for GDP was

\[
\log_{10}\hat{\text{GDP}}_C = \alpha \log_{10}(\text{NTL}_C) + \delta
\]

(1)

where \(S_C\) is the point sample for the country \(C\), \(\alpha\) and \(\delta\) are the regression coefficients to be learned, and \(\text{NTL}_C\) is the aggregate night-time illumination for the country \(C\). We also write \(w_j = 1/(np(x_j))\) to denote the weights.

To understand \(\epsilon\), consider its properties at the extremes. If \(\epsilon = 0\), then we select according to night-lights, which under-represents rural areas. If \(\epsilon > \sup r_t\), then \(p(x)\) is a uniform distribution, which over-represents rural areas. A well-chosen value of \(\epsilon\) gives greater representation of urban areas without neglecting rural areas.

To sample pixels with the correct selection probabilities, a probability proportional to size systematic sampling scheme was used, as it was computationally efficient and ensured a good spread of sample points.

For each country, \(n = 1000\) pairs of latitude–longitude coordinates from the country’s land surface were sampled randomly. The corresponding Google Static Maps API image centered at that coordinate was then downloaded for each. As an example of the sampling scheme in action, the sampled points for Norway is shown in Fig. 3. As can be seen, most of the sample points congregate around urban centers, but many can still be found scattered throughout the countryside.

\[\text{F. Dimensionality Reduction}\]

Taking the images and running them through the CNN to perform inference, feature vectors with 4096 dimensions were obtained. Given that there were only 178 GDP observations in the study area, the number of dimensions was reduced. Assuming the representation space was semantically rich, PCA was used to reduce the dimensionality while attempting to preserve most of the information. After normalizing the inputs, the representation space was reduced to a number of principal components, similar to the approach taken by Jean et al. [29], who preserved ten principal components. To determine the appropriate number of principal components, the corrected Aikake information criterion (AICC) selection method described in Section III-H was used to determine the optimal number for generalization. We did not weight the PCA procedure as that would bias the procedure so that rural areas might have undue influence.

The reduced dimension representation space will be referred to as simply “the representation space” and the \(N_j(x)\) will denote the \(j\)th principal component of the representation of the satellite sensor image centered at \(x\). Histograms showing the distribution of each \(N_j(x)\) are shown in Fig. 4.

\[\text{G. Validation}\]

While cross-validation was used to prevent overfitting, it remained a risk that the model space still could have been too large relative to the size of the dataset. In such instances, overfitting can still occur [49]. To validate the model, we used a procedure of holding back 35 countries as a validation set \(^3\) to ensure no overfitting, while the rest were used to fit the model.

To account for disparity between random seeds, this procedure was repeated 60 times and results were averaged.

For a model fitted on train set \(T\), we will refer to its \(R^2\) on the associated validation set as its validation \(R^2\). When fitted on the entire study area, we will refer to its \(R^2\) on the entire study area as its final \(R^2\).

\[\text{H. Partitioning the Representation Space}\]

To ensure that the representation space could be partitioned into \(p\) parts (which gives \(p\) regions) without overlap, the partition was specified as follows. Given a vector \(v \in \mathbb{R}^d\) with \(\|v\|_2 = 1\) and real numbers \(t_1, t_2, \ldots, t_{p-1}\) (which we call “splits”), then the part \(i\) was defined to be \(A_i = \{u \in \mathbb{R}^d : t_{i-1} < u \cdot v \leq t_i\}\).

We adopted the conventions that \(t_0 = -\infty\) and \(t_p = \infty\). Each region \(i\) is given by \(R_i = f^{-1}(A_i)\) and region \(i\) in country \(C\) is given by \(R_{iC} = C \cap R_i\).

Each region gives a corresponding value for the night-lights for that region. Precisely, \(\text{NTL}_{iC} = \int_{R_{iC}} r(x)\) for each \(0 < i < p\). The integrals can be approximated using the Monte Carlo estimator from (1). Denoting the sample points in \(R_{iC}\) by \(S_{iC} = S_C \cap R_{iC}\), the final estimator is given by

\[
\log_{10}\hat{\text{GDP}}_{iC} = \sum_{i=1}^{p} \alpha_i \left( \log_{10} \sum_{j \in S_{iC}} w_j r(x_j) \right) + \delta
\]

(3)

\(^3\)Chosen as a simple random sample.
where the $\alpha_i$'s are the regression coefficients and $\delta$ is the intercept, the values of which were to be learned.

To estimate the optimal number of parts $p$, a Monte Carlo simulation was used to estimate the optimal regionalization when $p = 1, 2, 3, \ldots$, denoted $R^{(p)}$, and then the AICc was computed and compared for each resultant model. To avoid arbitrarily small regions arising from noise, the constraint was added that in each regionalization there should be at least 256 training sample points in every region. This minimum region size hyperparameter was determined by repeatedly simulating GDP data and maximizing validation $R^2$. Ultimately, the value of $p$, which minimized AICc, was chosen to be the final number of regions.

AICc is an information theory derived measure of model generalization performance. It can be used to compare competing model families. The model family that scores lowest is selected, as the model family most likely to generalize well.

\[ AICc = AIC + \frac{2k}{n-k+1} \]

Like-likelihood was calculated as

\[ 2L = m \times \log(CVMSE) \]

where $m$ is the number of observations the model was fitted on and CVMSE is the 5-fold cross-validated mean square error (CVMSE) score on the training set. CVMSE was used in place of the estimated mean squared error, as the estimated mean squared error was observed to be noisy and erratic.

After finding and fixing the optimal number of regions $p$, a further Monte Carlo simulation with more iterations found the regionalization $R^{(p)}$, which minimized the (CVMSE) score on the training set.

---

**I. Computing Environment**

All computational work was performed on the following:

1) Dell Optiplex 3010 with an Intel i5-3470 CPU @ 3.20 GHz 4 cores, 4 GB RAM, 229 GB disk space, and no GPU running Python 3.
2) A VMware virtual desktop instance with an Intel Xeon Gold 6248R CPU @ 3.00 GHz 2 cores, 12.0 GB RAM, and no GPU running ArcGIS Pro.
3) A single free-use GPU-enabled Google Colab session with 12.7 GB of RAM and 107.8 GB of disk space running Python 3.

Apart from ArcGIS Pro, all softwares used were open-source and the cloud computing environment was open-access.

---

**IV. Results**

**A. Model**

Without regionalization, that is, regression of $\log_{10}\text{GDP}_C$ on $\log_{10}(\Sigma_{i \in C} t_i)$, the $R^2$ for the study area was 0.729, which gave a night-lights-only baseline to compare performance. As a CNN-only baseline, a regression on $\log\Sigma_{i \in C} w_i \exp(P_f(x))$ with the first 12 principal components was performed, which gave an $R^2$ for the study area of 0.790.

Across the 60 seeds, the average validation $R^2$ was 0.857 and the average final $R^2$ was 0.889. While this demonstrates overfitting, both figures are an improvement over both baselines.

The number of parts $p$ and principal components $d$, which minimized AICc averaged across 60 seeds, was $p = 3$ and $d = 5$, yet only in 13.3% of seeds did the AICc model selection procedure yield these values. Across 60 seeds, the most common values of $p$ and $d$, as selected by the AICc model selection procedure, were $p = 2$ and $d = 5$, optimal in 25% of seeds. Table I gives for how many seeds each $p, d$ combination is optimal.
TABLE I
NUMBER OF SEEDS FOR WHICH EACH DIMENSION–REGIONS (d, p) COMBINATION WAS OPTIMAL

| d | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 4 | 8 | 1 | 1 | 2 |
| 5 | 15 | 8 | 10 | 2 |
| 6 | 7 | 3 | 2 | 1 |
| 7 | 0 | 0 | 0 | 0 |

Fig. 5. Scatterplots of actual GDP versus predicted GDP for seed = 1. Left-hand side is for the validation set when fitted on the train set. Right-hand side is for the model fitted on the entire study area. This seed gives a model better than the average.

Fig. 6. Top: Map showing log_{10}GDP for the final model. Bottom: Map showing log_{10}GDP – log_{10}GDP for the final model with blue indicating underestimation and green overestimation.

For interpretation, the AICc procedure was repeated over the whole dataset to yield a final model. This gave a model with p = 3 regions and d = 5 principal components, v = (−0.29, 0.40, 0.58, 0.08, 0.64), t = (0.19, 0.29), and regression coefficients α = (0.23, −0.64, 1.29) and δ = 5.79. Fig. 5 shows a plot of log_{10}GDP_C against log_{10}GDP_C on the validation set when the model is fitted on the training set and a similar figure for the final model fitted on the entire study area neither of which show heteroscedasticity. Fig. 6 shows maps of log_{10}GDP and of the final model residuals by country. With respect to spatial dependency, the model overestimates GDP for much of Northern Asia and underestimates for much of Western Europe, Southeast Asia, and Oceania.

Though the CNN was trained originally on a dataset of images from only five African countries, the final model appears to generalize across the globe well. This corroborates a wider phenomenon observed by Lima and Marfurt [22] of CNNs trained on other tasks, even nonremote sensing tasks, generalizing well to other remote sensing tasks.

B. Final Regionalization

To understand the composition of the final regionalization, a random subsample of 450 points (150 per region from 100 countries) was drawn and the associated images analyzed and interpreted by visual inspection and then hand-labeled. The categories, to which images were labeled, were specified to capture recurring features observed empirically in the sample. Images were labeled as one of seven categories, which were defined as follows.

1) **Barren:** Images without noticeable presence of farmland, buildings, forest, or mountains, which includes grassland, roads, ocean, desert, and rivers.
2) **Dense Urban:** Images of built-up areas with little greenery.
3) **Farmland:** Images dominated by green fields or with clear evidence of agricultural activity.
4) **Forest:** Images dominated by green forest.
5) **Green Urban:** Images with built-up urban areas and a lot of greenery. Primarily, small settlements, suburbs, and city edges.
6) **Mountain:** Images dominated by mountainous terrain and without presence of farmland, buildings, or forest.
7) **Other Artificial Structure:** Images of human-made structures that do not fit the urban category. For example, dams, mining operations, etc.

The results of this analysis are given in Table II. This table includes estimates of the percentage of total land surface area each region accounts for, the percentage of each region’s land surface area each land category accounts for, and the percentage of each land category’s land surface area each region accounts for. These estimates are representatively weighed and corrected with auxiliary information. Selected images typical of the imagery in each region can be seen in Fig. 7.

Region 3 is the region with the largest associated regression coefficient and the largest mean radiance. Even when accounting for its land surface area being far smaller than that of region 1, a majority of information about a country’s GDP must be contained in the aggregate nighttime illumination of this region. While tempting to attribute the economically productive land to this region alone, there is relatively little dense urban area in this region. As a percentage of the region’s surface area, barren land constitutes a majority (i.e., 67.2%), but this region is constituted to a large degree by green urban land area (i.e., 21.2%), which far exceeds the percentage of green urban land area constituting the other regions. Therefore, even if there is little dense urban area in this region, this region contains many smaller towns, city edges, industrial estates, etc., all of which are highly informative of economic productivity overall.

While most of the land surface area of region 1 is composed of barren, forest, and mountainous land (i.e., 59.4%), over 40%...
Fig. 7. Selected images typical of the imagery in each region with three images per column and each column representing a separate region.

TABLE II
RESULTS OF ANALYZING A RANDOM SAMPLE OF \( n = 450 \) IMAGES TO PROFILE THE COMPOSITION OF EACH REGION

| Region          | 1        | 2        | 3        |
|-----------------|----------|----------|----------|
| Definition      | \( r \cdot v \leq 0.19 \) | \( 0.19 < r \cdot v \leq 0.29 \) | \( 0.29 < r \cdot v \) |
| Final model regression coefficient | 0.23 | -0.64 | 1.29 |
| Mean radiance (nW cm\(^{-2}\) sr\(^{-1}\)) | 0.65 | 0.87 | 1.90 |
| % of land surface area | 71.6% | 10.1% | 18.3% |

| % of region total area |
|------------------------|
| Barren | 44.5% | 66.0% | 67.2% |
| Farmland | 31.2% | 16.4% | 8.1% |
| Forest | 12.8% | 11.4% | 1.4% |
| Green urban | 6.8% | 1.9% | 21.2% |
| Mountain | 2.2% | 4.0% | 1.4% |
| Other human structure | 0.0% | 0.0% | 0.0% |
| Dense urban | 2.6% | 0.2% | 0.7% |

| % of land type total area |
|---------------------------|
| Barren | 62.7% | 13.1% | 24.2% |
| Farmland | 87.7% | 6.5% | 5.8% |
| Forest | 86.6% | 10.9% | 2.4% |
| Green urban | 54.4% | 2.2% | 43.5% |
| Mountain | 70.3% | 18.3% | 11.4% |
| Other human structure | 84.6% | 6.6% | 8.8% |
| Dense urban | 93.0% | 0.9% | 6.2% |

of the land surface area of region 1 is green urban, dense urban, farmland, and other human-made structure area. Moreover, this accounts for 87.7% of the farmland, 93% of the dense urban area, 54.4% of green urban area, and 84.6% of other human-made structure area. Therefore, most productive land is in this region. We would expect the night-lights of this region to contain information about GDP, which is consistent with the positive associated regression coefficient.

Region 2 is the only region with a negative associated regression coefficient. Barren land, forest, and mountainous area constitute more than 80% of this region’s land surface area. Most of the remaining land surface area is farmland. This region is
largely unproductive, which is consistent with having a small associated regression coefficient. What complicates interpretation is the negative associated regression coefficient; more nighttime illumination in region 2 corresponds to lower GDP. It is worth noting that the sum of the sampling weights is proportional to the land area, so region area is implicitly factored into the model in (3). A larger area of this region in a country downweights the GDP prediction and so might account for dependency on country surface area. It could be hypothesized that developed economies are able to repurpose all available land toward economic functions, but this is inconsistent with other studies, which found weak correlation between total biomass utilization and GDP [52], [53] developed economies instead displacing their land use abroad [54]. It is more likely that a negative $\alpha_2$ acts as a corrective penalty term for barren land in the other regions, the night-lights of which are counted as contributing positively to GDP by the model even if that land does not substantially contribute to GDP in actuality.

C. Sampling

Using bootstrapping to replicate the sampling scheme and keeping everything else constant, including the model, with 50 replications per country, it was found that the sampling error (i.e., coefficient of variation averaged over countries for the model’s $\log_{10}$GDP predictions) was estimated as 0.45%. Furthermore, the Pearson R correlation coefficient between the estimated sampling error for each country and country surface area was 0.11, which was not statistically significant at the 10% significance level.

D. Subnational Experiment

For comparability with other methods, a subnational experiment predicting 2020 real GDP for the USA at the county level was performed using data from the U.S. Bureau of Economic Statistics [55]. To this end, a simple random sample of 394 counties was selected. In each selected county, 250 longitude–latitude points were sampled yielding a dataset of 98,500 images. The baseline $R^2$ regressing $\log_{10}$GDP on $\log_{10}$ aggregate night-lights over this sample was 0.650.

Splitting the sample into a test set of 294 counties and a validation set of 100 counties for 40 different seeds and performing PCA and model selection on the subnational dataset, AICc was minimized by seven principal components in 70% of any country’s economy, there are some land types, such as barren land is more likely that a negative $\alpha_2$ acts as a corrective penalty term for barren land in the other regions, the night-lights of which are counted as contributing positively to GDP by the model even if that land does not substantially contribute to GDP in actuality.

V. DISCUSSION

A. Urban–Rural Dichotomy

In Section IV-B, it was observed that most of the surface area of region 2 is barren, farm, or forest land area. In line with economic intuition, it is tempting to label this region as a rural region, and then try and apply the urban–rural dichotomy to the other regions with some extra nuance. Neither regions 1 nor 2 truly fit the urban label; but, suppose we assume the model fails to disentangle barren land from other land types, an idea we could support by observing that barren land is plentiful in every region, as given in Table II. Assuming this, region 3 could be thought of as green urban. The problem with this is that examining the mean values of $Pf(x) \cdot v$ in region 1 for each land type, the mean for barren land is $-0.22$, for farmland $-0.38$, and for dense urban $-0.12$. The gaps between these values are similar to the widths of the region definitions in Table II, suggesting clear delineation between these land types is possible were that the optimal way to regionalize. A productive–unproductive dichotomy where regions 1 and 3 are considered as the economically productive land and region 2 the economically unproductive land is just as plausible. Rather than apply a reductive dichotomy, the safest interpretation is likely to be that, due to the interrelated nature of any country’s economy, there are some land types, such as green urban, that could be as representative of a country’s entire economy.

B. Model Specification

Taking a moment to consider the model specification, the model used was a linear regression in the logarithm of the aggregate night-lights of the regions, that is,

$$\log_{10}\text{GDP}_C = \sum_{j=1}^{p} (\alpha_j \log_{10}\text{NTL}_{jC}) + \delta + \epsilon_C \quad (4)$$

where $\epsilon_C$ is an error term. This was intended as a weighted average of the logarithm of the GDP of each region, but it is also the logarithm of a Cobb–Douglas production function, which is a production function of the form $f(A_1, \ldots, A_m) = bA_1^{k_1}A_2^{k_2} \cdots A_m^{k_m}$, common in economics [57]. This leads to an interpretation that the NTL of $C$’s are best thought of not as GDP values, but as proxies for some more abstract kind of “inputs” into the economic system, which are the outputs of the economic activity associated with each region. The efficacy of the model in the prediction task suggests that this could be a helpful way of conceptualizing what night-lights represent.
C. Future Research

Extending the method into the time domain is the next natural step though this will require careful thought as to how growth in night-lights in regions relates to growth in GDP. As Northcott [58] made clear, a lack of integration of the economic theory severely limits GDP predictive modeling. Greater integration of economic expertise could be used to strengthen or add depth to the idea of regions representative of GDP as a whole. Domain expertise from remote sensing could also be better integrated, perhaps by integrating hand-crafted Haralick features, which are highly informative, but which CNNs fail to learn [42]. Inclusion of other socioeconomic indicators have increased predictive precision in other studies [21] and could also be integrated.

While progress toward remote-sensed GDP figures of a precision and accuracy comparable to official figures continues, night-light-produced GDP predictions have many other uses. They have been used to show when governments propagate artificial inflations [9], to demonstrate ethnic favoritism in government spending [59], and even used to show when NGO GDP figures underestimate true economic growth [60]. By presenting a framework that makes the production of more precise GDP figures accessible to the research community, we hope to enable more studies that apply these methods to questions of justice and transparency.

VI. Conclusion

While night-lights do provide a noisy proxy for economic activity, alone they lack sufficient context to predict GDP with adequate precision. In this research, we showed that how fine spatial resolution daytime satellite sensor imagery can be used to contextualize night-lights for more precise night-light-produced GDP prediction. The model fitting procedure achieved an average validation $R^2 = 0.857$ and an average final $R^2 = 0.889$. A subnational experiment was also performed where the model fitting procedure outperformed, on average, comparable models.

This predictive accuracy was achieved using a feature extractor trained on a generic night-lights task containing relevant features for predicting GDP, a carefully designed Monte Carlo importance sampling scheme, and fine resolution satellite sensor imagery. This demonstrated that regionalization is a powerful technique for contextualizing night-lights. Despite the temptation to regionalize based on a simple rural–urban dichotomy, it was demonstrated that a data-driven approach agnostic to such a priori convictions can suggest new ways in which to conceptualize night-lights, in this instance, taking the night-lights in green urban areas as representative of overall national GDP. Ultimately, classical notions, such as the urban–rural dichotomy, need not be the default mode by which we interpret night-lights and the economic processes that generate them. The data itself can suggest ways to understand night-lights.

REFERENCES

[1] C. D. Elvidge, K. E. Baugh, E. A. Kihn, H. W. Kroehl, E. R. Davis, and C. W. Davis, “Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption,” Int. J. Remote Sens., vol. 18, no. 6, pp. 1373–1379, 1997.
[2] T. A. Croft, “Nighttime images of the earth from space,” Sci. Amer., vol. 239, no. 1, pp. 86–101, 1978.
[3] A. Heston, “A brief review of some problems in using national accounts data in level of output comparisons and growth studies,” J. Develop. Econ., vol. 44, no. 1, pp. 29–52, 1994.
[4] N. Levin et al., “Remote sensing of night lights: A review and an outlook for the future,” Remote Sens. Environ., vol. 237, 2020, Art. no. 111443.
[5] J. Gibson, S. Olivia, G. Boe-Gibson, and C. Li, “Which night lights data should we use in economics, and where?”, J. Develop. Econ., vol. 149, 2021, Art. no. 102602.
[6] S. Keola, M. Andersson, and O. Hall, “Monitoring economic development from space: Using nighttime light and land cover data to measure economic growth,” World Develop., vol. 66, pp. 322–343, 2015.
[7] T. Bundervoet, L. Mairoy, and A. Sanghi, Bright Lights, Big Cities: Measuring National and Subnational Economic Growth in Africa From Space, With an Application to Kenya and Rwanda. Washington, DC, USA: World Bank, 2015.
[8] J. V. Henderson, A. Storeygard, and D. N. Weil, “Measuring economic growth from outer space,” Amer. Econ. Rev., vol. 102, no. 2, pp. 994–1028, 2012.
[9] L. Martinez, “How much should we trust the dictator’s GDP growth estimates?,” 2021. [Online]. Available: https://ssrn.com/abstract=3932966 or http://dx.doi.org/10.2139/ssrn.3932966.
[10] X. Wang, M. Rafa, J. D. Moyer, J. Li, J. Scheer, and P. Sutton, “Estimation and mapping of sub-national GDP in Uganda using NPP-VIIRS imagery,” Remote Sens., vol. 11, no. 2, 2019, Art. no. 163.
[11] G. Falchetta, S. Pachauri, S. Parkinson, and E. Byers, “A high-resolution gridded dataset to assess electrification in sub-Saharan Africa,” Sci. Data, vol. 6, no. 1, 2019, Art. no. 110.
[12] E. G. Irwin, K. P. Bell, N. E. Bockstael, D. A. Newburn, M. D. Partridge, and J. Wu, “The economics of urban-rural space,” Annu. Rev. Resour. Econ., vol. 1, no. 1, pp. 435–459, 2009.
[13] K. C. Seto et al., “Urban land teleconnections and sustainability,” Proc. Nat. Acad. Sci. USA, vol. 109, no. 20, pp. 7687–7692, 2012.
[14] G. Bosworth and V. Venhorst, “Economic linkages between urban and rural regions—What’s in it for the rural?,” Regional Stud., vol. 52, no. 8, pp. 1075–1085, 2018.
[15] G. W. Jones, “Urbanization trends in Asia: The conceptual and definitional challenges,” in New Forms of Urbanization, London, U.K.: Routledge, 2017, pp. 113–131.
[16] J. Sun, L. Di, Z. Sun, J. Wang, and Y. Wu, “Estimation of GDP using deep learning with NPP-VIIRS imagery and land cover data at the county level in CONUS,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 13, pp. 1400–1415, Mar. 2020.
[17] A. Di Gregorio, Land Cover Classification System: Classification Concepts and User Manual: LCCS, vol. 2, Rome, Italy: Food & Agriculture Org., 2005.
[18] J. Sun, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, “Deep learning in remote sensing applications: A meta-analysis and review,” ISPRS J. Photogrammetry Remote Sens., vol. 152, pp. 166–177, 2019.
[19] H. Storm, K. Baylis, and T. Heckelei, “Machine learning in agricultural and applied economics,” Eur. Rev. Agric. Econ., vol. 47, no. 3, pp. 849–892, 2020.
[20] H. Liu et al., “Nightlight as a proxy of economic indicators: Fine-grained GDP inference around Chinese mainland via attention-enhanced CNN from daytime satellite imagery,” Remote Sens., vol. 13, no. 11, 2021, Art. no. 2067.
[21] A. Khachiyan et al., “Using neural networks to predict micro-spatial economic growth,” Amer. Econ. Rev.: Insights, to be published, doi:10.3386/w20560.
[22] R. de Lima and K. Marfurt, “Convolutional neural network for remote-sensing scene classification: Transfer learning analysis,” Remote Sens., vol. 12, no. 1, 2020, Art. no. 86.
[23] K. Weiss, T. M. Khoshgoftaar, and D. D. Wang, “A survey of transfer learning,” J. Big Data, vol. 3, 2016, Art. no. 9.
[24] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, “Deep learning earth observation classification using ImageNet pretrained networks,” IEEE Geosci. Remote Sens. Lett., vol. 13, no. 1, pp. 105–109, Jan. 2016.
[25] Z. Chen, T. Zhang, and C. Ouyang, “End-to-end airplane detection using transfer learning in remote sensing images,” Remote Sens., vol. 10, no. 1, 2018, Art. no. 139.
[26] M. Rostami, S. Kolouri, E. Eaton, and K. Kim, “Deep transfer learning for few-shot SAR image classification,” Remote Sens., vol. 11, no. 11, 2019, Art. no. 1374.
[27] H. Lu et al., “Landslides information extraction using object-oriented image analysis paradigm based on deep learning and transfer learning,” Remote Sens., vol. 12, no. 5, 2020, Art. no. 752.

[28] Y. Koga, H. Miyazaki, and R. Shibasaki, “A method for vehicle detection in high-resolution satellite images that uses a region-based object detector and unsupervised domain adaptation,” Remote Sens., vol. 12, no. 3, 2020, Art. no. 575.

[29] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, “Combining satellite imagery and machine learning to predict crop yield,” Science, vol. 353, no. 6301, pp. 790–794, 2016.

[30] U. Maulik and D. Chakraborty, “Object-oriented image classification: A survey of supportvector-machine-based advanced techniques,” IEEE Geosci. Remote Sens. Mag., vol. 5, no. 1, pp. 33–52, Mar. 2017.

[31] African Development Bank Group, “Situational analysis of the reliability of economic statistics in Africa: Special focus on GDP measurement,” 2013. [Online]. Available: https://www.afdb.org/en/documents/document/economic-brief-situational-analysis-of-the-reliability-of-economic-statistics-in-africa-special-focus-on-gdp-measurement-32429

[32] M. Lipton, “Africa’s national-accounts mess,” J. Develop. Stud., vol. 49, no. 12, pp. 1765–1771, 2013.

[33] The Economist, “How Nigeria’s economy grew by 89%,” Accessed: Oct. 30, 2021. [Online]. Available: https://www.economist.com/the-economist-explains/2014/04/07/how-nigerias-economy-grew-by-89-overnight

[34] M. M. Bennett and L. C. Smith, “Advances in using multitemporal night-time lights satellite imagery to detect, estimate, and monitor socioeconomic dynamics,” Remote Sens. Environ., vol. 192, pp. 176–197, 2017.

[35] M. O. Román and E. C. Stokes, “Holidays in lights: Tracking cultural patterns in demand for energy services,” Earth’s Future, vol. 3, no. 6, pp. 182–205, 2015.

[36] C. D. Elvidge, K. Baugh, M. Zhizhin, F. C. Hsu, and T. Ghosh, “VIIRS night-time lights,” Int. J. Remote Sens., vol. 38, no. 21, pp. 5860–5879, 2017.

[37] M. Buchhorn et al., “Copernicus global land service: Land cover 100m: Collection 3: Epoch 2018: Globe,” Sep. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3518038

[38] C. D. Elvidge, K. E. Baugh, M. Zhizhin, and F.-C. Hsu, “Why VIIRS data are superior to DMSP for mapping nighttime lights,” Proc. Asia-Pacific Adv. Netw., vol. 35, 2013, Art. no. 62.

[39] P. Schottenfels, “Ask a techspert: How do satellite images work?,” 2020. [Online]. Available: https://blog.google/products/maps/how-do-satellite-images-work/

[40] Esri, “World countries (generalized),” 2019. [Online]. Available: https://www.arcgis.com/home/item.html?id=2b93b06dc0dc4e809d3c8db5cb96ba69

[41] Q. Liu et al., “DeepSat V2: Feature augmented convolutional neural nets for satellite image classification,” Remote Sens. Lett., vol. 11, no. 2, pp. 156–165, 2020.

[42] S. Basu et al., “Deep neural networks for texture classification—A theoretical analysis,” Neural Netw., vol. 97, pp. 173–182, 2018.

[43] J. Mathur, “Predicting-poverty-replication,” 2020. [Online]. Available: https://github.com/jmathert625/predicting-poverty-replication

[44] N. Jean et al., “Combining satellite imagery and machine learning to predict poverty: Supplementary material,” 2016. [Online]. Available: http://sustain.stanford.edu/predicting-poverty

[45] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, “Return of the devil in the details: Delving deep into convolutional networks,” 2014, arXiv:1405.3531.

[46] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, arXiv:1409.1556.

[47] Geography, Layout, Travel, Jan. 2019. [Online]. Available: https://www.ala.org/travel/geography/

[48] R. Y. Rubinstein and D. P. Kroese, Simulation and the Monte Carlo Method. Hoboken, NJ, USA: Wiley, 2016.

[49] A. Y. Ng et al., “Preventing ‘overfitting’ of cross-validation data,” in Proc. 14th Int. Conf. Mach. Learn., 1997, pp. 245–252.

[50] J. E. Cavanaugh, “Unifying the derivations for the Akaike and corrected Akaike information criteria,” Statist. Probability Lett., vol. 33, no. 2, pp. 201–208, 1997.

[51] R. A. Gordon, Regression Analysis for the Social Sciences. London, U.K.: Routledge, 2015.

[52] J. K. Steinberger and F. Krausmann, “Material and energy productivity,” Environ. Sci. Technol., vol. 45, no. 4, pp. 1169–1176, Feb. 2011, doi: 10.1021/es1028537.

[53] M. Bildirici and Ö. Ersin, “An investigation of the relationship between the biomass energy consumption, economic growth and oil prices,” Procedia—Social Behav. Sci., vol. 210, pp. 203–212, 2015.

[54] J. Weinzettel, E. G. Hertwich, G. P. Peters, K. Steen-Olsen, and A. Galli, “Affluence drives the global displacement of land use,” Glob. Environ. Change, vol. 23, no. 2, pp. 433–438, 2013.

[55] Bureau of Economic Analysis, “Gross domestic product by county,” 2020. [Online]. Available: https://www.bea.gov/data/gdp/gdp-county-metro-and-other-areas

[56] N. Zhao, F.-C. Hsu, G. Cao, and E. L. Samson, “Improving accuracy of economic estimations with VIIRS DN/imag products,” Int. J. Remote Sens., vol. 38, no. 21, pp. 5899–5918, 2017.

[57] P. H. Douglas, “The Cobb–Douglas production function once again: Its history, its testing, and some new empirical values,” J. Political Economy, vol. 84, no. 5, pp. 903–915, 1976.

[58] R. Northcott, “Big data and prediction: Four case studies,” Stud. Hist. Philosophy Sci. Part A, vol. 81, pp. 96–104, 2020.

[59] G. De Luca, R. Hodler, P. A. Raschky, and M. Valsecchi, “Ethnic favoritism: An axiom of politics?,” J. Develop. Econ., vol. 132, pp. 115–129, 2018.

[60] H. Clark, M. Pinkovskiy, and X. Sala-i-Martin, “China’s GDP growth may be understated,” China Econ. Rev., vol. 62, 2020, Art. no. 101243.

Nathan Price received the B.Sc. degree in mathematics from the University of Warwick, Coventry, U.K., in 2015, and the M.Sc. degree in mathematics from the University of St. Andrews, St Andrews, Scotland, in 2016. He is currently working toward the Ph.D. degree in environmental science with Lancaster University, Lancaster, U.K.

He was an Assistant Methodologist with the Office for National Statistics, Newport, U.K., and a Techni cal Expert with the Office of the Chief Government Statistician, Stonetown, Zanzibar, as part of the ODI Fellowship Scheme. His research interests include sample design and estimation as well as data science, machine learning, and remote sensing.

Peter M. Atkinson received the M.B.A. degree from the University of Southampton, Southampton, U.K., in 2012, and the Ph.D. degree from the University of Sheffield, Sheffield, U.K., in 1990, both in geography.

He was a Professor of Geography with the University of Southampton, where he is currently a Visiting Professor. He was the Belle van Zuylen Chair with Utrecht University, Utrecht, The Netherlands. He is currently a Distinguished Professor of Spatial Data Science with Lancaster University, Lancaster, U.K.

He is also a Visiting Professor with the Chinese Academy of Sciences, Beijing, China. He has authored or coauthored more than 350 peer-reviewed articles in international scientific journals and around 50 refereed book chapters, and also edited more than ten journal special issues and eight books. His research interests include remote sensing, geographical information science, and spatial and (space-time) statistics applied to a range of environmental science and socioeconomic problems.

Prof. Atkinson was the recipient of the Peter Burrough Award of the International Spatial Accuracy Research Association and the NERC CASE Award from Rothamsted Experimental Station. He is an Editor-in-Chief for Science of Remote Sensing, a sister journal of Remote Sensing of Environment, and an Associate Editor for Computers and Geosciences. He sits on various international scientific committees.