Local GDP Estimates Around the World*

Esteban Rossi-Hansberg
University of Chicago

Jialing Zhang
University of Chicago

January 29, 2025

Abstract

We use high-resolution spatial data to build a novel global annual gridded GDP dataset at 1°, 0.5°, and 0.25° resolutions from 2012 onward. Our random forest model trained on local and national GDP achieves an R^2 above 0.92 for GDP levels and above 0.62 for annual changes in regions left out of the training sample. By incorporating diverse indicators beyond population and nighttime lights, our estimates offer more precise subnational GDP measurements for analyzing economic shocks, local policies, and regional disparities. We evaluate the precision of our estimates with a sample case of COVID-19's impact on local GDP in China.

Obtaining worldwide detailed production data over time is particularly challenging because middle- and low-income countries lack reliable subnational statistics. The G-Econ dataset by Nordhaus (2006) has provided essential spatial GDP estimates, primarily using population as a key predictor, but it is limited to a few years (1990, 1995, 2000, 2005). Researchers seek datasets that span more recent and frequent years and move beyond population-based proxies. The work by Henderson, Storeygard and Weil (2012) was pivotal in demonstrating that nighttime lights (NTL) can serve as a reliable proxy for economic activity. The use of NTL data for GDP estimation in data-scarce regions involves various methods, including running regressions with coarser GDP data to generate finer-scale predictions (Vogel et al., 2024), directly using NTL as outcome variable in regressions (Storeygard, 2016; Henderson, Squires and Weil, 2018), and employing light intensity as a weighting factor to distribute national GDP (Chen et al., 2022). It is well known, however, that nighttime lights data have limitations: saturation can lead to underestimation of urban production

*Esteban Rossi-Hansberg: earossih@uchicago.edu. Jialing Zhang: jialingzhang@uchicago.edu. We thank Reigner Kane, Sreyas Mahadevan, and Jordan Rosenthal-Kay for excellent research assistance and contributions during the initial stages of this project.

(Henderson, Storeygard and Weil, 2012), inconsistencies between satellite sensors over time pose challenges for comparing data across years, and sectors such as agriculture and forestry are often underestimated by NTL (Keola, Andersson and Hall, 2015).

There is an increasing demand for more accurate local GDP dataset to enhance quantification of spatial models as well as for reduced form analysis. To meet this need, machine learning techniques combined with high-resolution datasets have been employed (Khachiyan et al., 2022). However, these machine learning approaches are typically restricted to regions with existing reliable subnational GDP data due to the need for training data, and thus have limited applicability for middle- and low-income countries.

Building on these approach, we go a step further to develop a statistical prediction model that generalizes across heterogeneous regions and still accurately predicts local GDP for out of sample regions. Our model can provide reliable local GDP estimates that cover more years and also capture economic shifts and complex patterns, as we described below.

To achieve this goal, we introduce several novel elements. First, alongside gridded population data (Bright et al., 2012–2021) and NASA's new unsaturated nighttime lights dataset (VIIRS VNP46A4 Black Marble product) (Román et al., 2018), we also integrate high-resolution land use data (Friedl and Sulla-Menashe, 2022), emissions data (Emissions Database for Global Atmospheric Research , EDGAR), and net primary productivity (NPP) data (Running and Zhao, 2021).[1] The unsaturated nighttime lights dataset addresses urban GDP underestimation issues and mitigates the challenges of data comparison across years as its sensors remain consistent over time. Integrating land use and NPP data improves GDP estimations in the agricultural sector. Emissions data capture more nuanced GDP changes in response to economic shocks. This diverse set of variables enables more accurate out-of-sample GDP predictions, improves year-over-year GDP growth rate estimations, and captures economic shifts that single proxies may miss. The approach achieves out of sample R^2 values above 0.92 for GDP levels and 0.62 for annual changes, even in the presence of unprecedented shocks, like COVID-19.

Second, we use a random forest algorithm to build our statistical prediction model. This methodology excels at preventing overfitting by aggregating predictions from numerous weak learners (i.e., decision trees) while still capturing complex data patterns. We use it to estimate the function that maps predictor shares to GDP shares using subnational GDP data from North America, South America, Europe, Africa, and Asia. By focusing on GDP share instead of absolute GDP levels, the model disregards country-level fixed effects and can generalize more effectively across regions. To avoid overly optimistic out-of-sample

[1]Net Primary Productivity (NPP) measures the amount of carbon captured by plants through photosynthesis minus the carbon they release via respiration. It serves as an indicator for measuring vegetation.

predictions, we use Group k-Fold Cross-Validation. This method keeps data points from the same country together and thus prevents them from being split into training and testing sets. This ensures that when we select the best model, we do not choose the one that appears to perform well due to shared data dependencies between training and testing or because GDP share values do not vary much.

We further validate the model by applying it to China, a country excluded from the training sample, to assess its ability to detect COVID-19 related economic changes. The model achieves R^2 values above 0.95 for GDP levels and above 0.31 for annual changes across pre-COVID, COVID, and post-COVID periods. Unlike the G-Econ dataset, where population is the primary GDP predictor, our model identifies a convex relationship between population and GDP at all three spatial resolutions, as predicted by the presence of agglomeration economics. Additional analyses presented in the Appendix demonstrate the model's strong in-sample fit, consistency across resolutions, predictive accuracy for future years, and robustness under various checks.

The remainder of this paper is structured as follows. Section I describes the primary data sources, data construction processes, and model training procedures. Section II presents the results and evaluates the model performance. Section III explores potential applications of our datasets and concludes. All data output are accessible for download at `http://bfidatastudio.org/gdp`.

1 Data and Methods

1.1 Data

We used three primary types of data: remote sensing data, model-processed datasets from other sources, and GDP datasets from international organizations and official agencies. See Section 1 of the Appendix for a detailed description of the data sources.

The remote sensing data include nighttime light (NTL) data, land use, and net primary productivity (NPP) data. The nighttime light data comes from NASA's VIIRS VNP46A4 Black Marble product. It offers uncensored nighttime light intensity at 15 arc-second (approximately 450 to 500 meters) with consistent satellites across years and removes noise and temporary lighting. We further process the data by 1) excluding the lights within a 0.2° cell radius centered around known gas flaring locations with positive gas flaring volumes sourced from the Global Gas Flaring Data dataset in the Global Flaring and Methane Reduction Partnership (GFMR) community (Global Gas Flaring Data, 2012-2023), 2) excluding ocean

and large inland water bodies before extracting the lights emissions in each cell.[2] Different land activities emit varying levels of light and may relate differently to GDP, so we also use NASA MODIS MCD12Q1 version 6.1 land cover dataset to separate lights by urban, cropland, and other land types. This dataset has a spatial resolution of 500 meters. We also include land cover data in our model as separate predictors. To address the limitations of nighttime lights in capturing the agricultural sector, we supplement cropland land use data with net primary productivity (NPP) data from the NASA MODIS MOD17A3HGF version 6.1 product. NPP measures vegetation with a spatial resolution of 500 meters.

Gridded population and CO_2 emissions data are model-processed datasets from other sources. Population data, derived from the LandScan Global dataset, spatially distributes national population counts at a 30 arc-second resolution (approximately $1km$) by applying country-specific machine learning models based on satellite imagery. The CO_2 emissions dataset comes from the Emissions Database for Global Atmospheric Research (EDGARv8.0), specifically the Global Greenhouse Gas Emissions dataset. It has a spatial resolution of 0.1 degrees. We categorize and aggregate sector-specific emissions into six groups: manufacturing combustion, heavy industry, and transportation, each further divided to separately account for fossil CO_2 sources and biofuel CO_2 sources. Alongside the above data, we include the terrain ruggedness index from Nunn and Puga (2012).

National GDP and national population data are sourced from the IMF World Economic Outlook Database, with supplementary data from the World Bank and UNdata for specific countries. Regional GDP data are mainly derived from the OECD Regional Statistics dataset, complemented by additional data for some developing countries obtained from Wenz et al. (2023) and official statistical agencies in Russia, Brazil, China, India, Kazakhstan, USA, Philippines, and Kyrgyzstan. For China, city-level data from the seven provinces with the highest GDP are collected from provincial statistical yearbooks. These city-level data are used exclusively to evaluate the model's ability to detect COVID-19 shocks and are not included in any model training in Section 2.[3] Further details on data sources, processes, and countries included in the training sample can be found in Appendix Section 1.

1.2 Random Forest Model Training

Training datasets are constructed at the cell level for each resolution (1°, 0.5°, and 0.25°) covering the years 2012 to 2021. These datasets consist of cell GDP share as the target values,

[2] 3,067 largest lakes (area greater and equal to 50 km2) and 654 largest reservoirs (storage capacity greater and equal to 0.5 km3) worldwide are excluded from the geometry shapefiles.

[3] China's GDP data is excluded from the training set primarily due to the challenges in data collection, including restricted overseas access to government websites, the need for manual compilation, and discrepancies between the China City Yearbook and individual province statistical yearbooks.

$\mathbf{y}$, and a matrix of input features, $\mathbf{X}$, also at cell level. These inputs include NTL share, population share, CO_2 share, NPP share, land use share, their lagged shares, mean terrain ruggedness, and national GDP per capita. A separate random forest model is developed for each resolution to estimate the function $f : \mathbb{R}^{n\times d} \to \mathbb{R}^n$ such that $\mathbf{y} = f(\mathbf{X}) + \varepsilon$. The random forest algorithm is selected for its ability to capture complex relationships and also mitigate overfitting by averaging the predictions of multiple decision trees (Breiman, 2001).

Before converting predictors to shares, we aggregate the values of pixels k intersecting with cell i to compute the cell's absolute predictor value $\tilde{X}_i$. This is done by multiplying each pixel's density x_k by the intersected area A_{ki},[4]

$$\tilde{X}_i = \sum_{k\in\{k|k\cap i\neq\emptyset\}} x_k \times A_{ki}. \tag{1}$$

The units of x_k vary depending on the predictor. For nighttime light (NTL) data, x_k is measured in nanowatts per square centimeter per steradian ($n\text{Watts}\cdot\text{cm}^{-2}\cdot\text{sr}^{-1}$). For NPP, x_k is expressed in kilograms of carbon per square meter (kgC/m^2). For CO_2 emissions, x_k is in tonnes per square meter (t/m^2). For population, x_k represents the population count per square meter. For land use data, x_k is an indicator variable that identifies specific land use types within the pixel.[5]

After obtaining cell absolute values, we construct predictors as shares relative to country-level or state-level wherever available (Australia, Brazil, Canada, China, India, Kazakhstan, Mexico, Russia, USA). Alongside the above shares, we include lagged shares to help predict growth rates.[6] Besides, terrain ruggedness index is calculated as cell mean, national GDP per capita is also added in the model.

The model is trained using data from countries with county level GDP available for developed nations and a slightly broader administrative level for developing ones. These training countries span North America, South America, Europe, Africa, and Asia. Assuming uniform GDP per capita within counties or comparable areas, we calculate cell-level GDP as $y_i = \sum_{r\in\{r|r\cap i\neq\emptyset\}} \frac{y_r}{p_r} \times p_{ir}$, where $\frac{y_r}{p_r}$ denote GDP per capita of county r, and p_{ir} denotes the population at the intersection of county r and cell i.[7] As when calculating predictors' share,

[4]The area are calculated based on a spherical approximation of Earth.

[5]For cells located along country borders, the cells are divided into segments based on the boundaries. The aggregated value for each segment is calculated separately to account for the portion of the cell that falls within each country.

[6]Since the Black Marble nighttime lights data constrains our data to start in 2012, no lag variable is available for that year; instead, we use the 2012 data as its own lagged value.

[7]Again, for cells located along country borders or state borders (in cases where state-level GDP is used), the cells are divided into segments based on the boundaries and the aggregated value for each segment is calculated separately.

we calculate cell GDP share by country, and for countries with state-level GDP available (Australia, Brazil, Canada, China, India, Kazakhstan, Mexico, Russia, USA), we calculate state-level shares.

To prevent overfitting, we use cross-validation to tune three random forest hyperparameters: terminal node size, number of candidate variables at each split, and number of trees. To predict GDP shares for cells in entirely out-of-sample countries, we cannot use the traditional approach of creating folds through splitting data points, as this could result in each fold containing the same cell across different years. If a cell's GDP shares remain stable, this approach may yield overly optimistic results on test folds while still failing for unseen cells in new countries. Instead, we randomly divide the training countries into five folds, using cells from four folds for training and cells from the remaining fold for testing to measure "out-of-training-sample" R^2 in log levels and R^2 in year-over-year log differences for both developed and developing countries' cells (see Section 2.3 for results). For each hyperparameter combination, this five-fold approach provides five "out-of-training-sample" R^2 values. We select the optimal hyperparameters by maximizing the weighted R^2 of year-over-year log differences. This involves calculating the mean R^2 across the five folds for each group (developed and developing), then taking a weighted mean of these two values based on the proportion of cells belonging to each group in the world.[8]

After selecting the optimal hyperparameters for each resolution, we train the standard random forest model on the entire training dataset. To address the imbalance between cells from developed and developing countries in the training data, models are trained with weights that rescale the shares of these cells to reflect their actual proportions in the real world. The weights affect the probability of each cell being selected in the bootstrap sample to build the decision tree and allow a balanced representation of both developing and developed countries during training.[9]

1.3 Predictions and Post-adjustments

After building the random forest model, we proceed with predictions. For training cells, we use out-of-bag predictions to provide an unbiased estimate. The out-of-bag estimate is calculated using predictions from trees that excluded the cell from their training. For

[8]Random forest models typically tune hyperparameters by minimizing mean squared error (MSE) rather than focusing on year-over-year log differences. However, tuning based on year-over-year log differences, while resulting in a slight increase in MSE, can significantly improve the accuracy of log difference predictions. But this adjustment has minimal impact on the accuracy of GDP level predictions. Refer to the Appendix Section 6.1 for results from models tuned using MSE and their comparison with the benchmark models presented in the paper.

[9]Refer to the Appendix Section 6.2 for a comparison showing that weights have minimal impact on model performance.

out-of-training-sample cells, predictions are generated directly by the model. To avoid overestimating GDP per capita in sparsely populated cells, we censor the predicted cell GDP shares to zero for those with population densities equal to 0.[10] Then GDP shares for the new set of cells are rescaled within parent areas to sum to 1, and GDP predictions are obtained by multiplying these rescaled shares by the GDP of the parent area.

2 Local GDP Predictions

2.1 Global GDP Distribution and Growth Patterns

This section presents the results of the 1-degree model trained using data from 2012 to 2021 for all available countries (excluding China) under the optimal hyperparameters, and then generate predictions for global GDP at a 1-degree resolution. Figure 1 presents the results of GDP and GDP per capita in both levels and changes. Grey cells are those with zero population.

The 2019 GDP map reveals the well-known large regional disparities. Dense clusters of high GDP are observed around major cities, such as São Paulo and Santiago in South America, Moscow in Russia, Beijing, Shanghai and Seoul in Asia, and Cairo and Johannesburg in Africa. Coastal areas, especially those along major maritime shipping routes, display higher GDP values compared to inland regions. For instance, the Asia-Europe route, connecting China's coastal regions through critical chokepoints like the Strait of Malacca and the Suez Canal, terminates in Europe via Gibraltar or the Cape of Good Hope in Africa, with some segments bypassing the Gulf of Guinea in West Africa. Key nodes along this route consistently exhibit high GDP levels. Similarly, the Atlantic coastal regions of Brazil along the South America-Europe route are economically prominent. Conversely, areas with challenging geographical features, such as the Himalayan Mountains, the Sahara and Arabian Deserts, central Australia's arid zones, and the Amazon rainforest, show low economic activity and the only GDP clusters are confined to urban centers and water-abundant locations.

The 2019 GDP per capita map also reveals significant regional disparities. Major cities exhibit higher GDP per capita compared to surrounding rural areas. In China, eastern coastal cities demonstrate substantially higher GDP per capita, and values decline progressively inland. Similarly, in Brazil, the northern regions have lower GDP per capita than

[10] Cells located along country or state borders may form irregular polygons instead of standard grid cells. For these polygons, censorship is applied at the polygon level. Thus, after aggregation into standard cells, some cells may still display positive GDP values. We also provide datasets where censorship is applied to cells with population densities below thresholds of 0.01, 0.02, or 0.05 individuals per km^2 of cell land area. All results in Section 2 use the dataset adjusted for cells with population densities equal to 0. See Appendix Section 3 for a count of cells affected by the adjustments.

the south, likely due to environmental and geographical challenges, limited infrastructure, or lower population density. Interestingly, some sparsely populated areas, for example some parts of Sub-Saharan Africa and Russia, exhibit high GDP per capita. This is often attributed to natural resource extraction, particularly oil and gas, which disproportionately contribute to GDP in these regions.

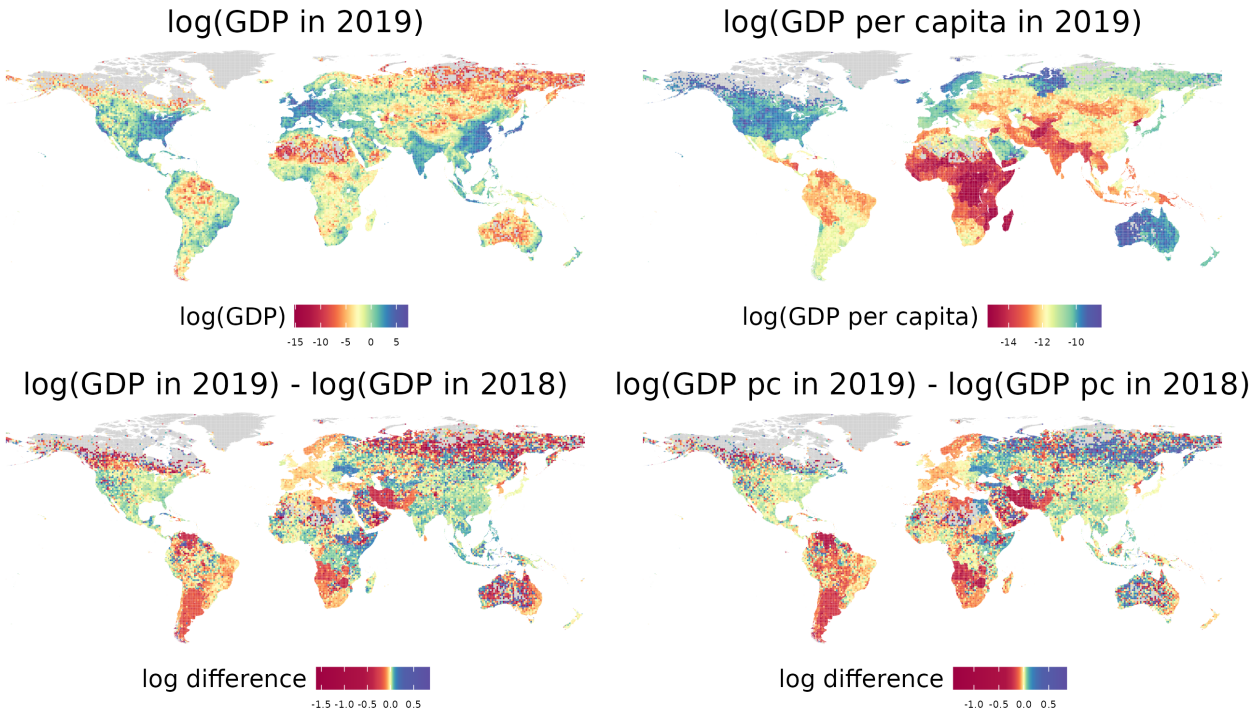

Figure 1: Map of Global 1deg Cell log(GDP), log(GDP Per Capita) and Annual Growth Rates

Note: GDP unit is constant 2017 billion USD. Grey cells are those with a population density of 0 people per km^2, which have been bottom-coded to a value of 0.

The second-row depicts annual changes in GDP and GDP per capita. In sparsely populated areas, cells can show sharp changes and differ from adjacent cells, as seen in deep red and deep blue cells in Russia, Canada, Sub-Sahara Africa, and Australia. In contrast, densely populated areas exhibit more consistent changes, with neighboring cells showing similar levels of growth or decline. In 2019, positive growth in GDP and GDP per capita is concentrated in densely populated regions of North America, South Asia, Eastern Europe, and parts of Africa, while large negative growth is observed in South America, Europe, and Southern Africa. Notably, sharp cross-border contrasts are evident in Africa, whereas Asia, Europe, and North America display much less pronounced cross-border differences and more uniform regional growth.

2.2 The Relationship Between Population and GDP

A key aspect of assessing the usefulness of our estimate is the extent to which they capture factors beyond population. Figure 2 plots cell-level GDP against population count, in log-log scale, across three resolutions. It shows that our model reveals a concave relationship for very low population density areas and largely a convex relationship in most other regions. The concavity observed in sparsely populated areas may stem from congestion in production factors, like land, while the convexity aligns with standard agglomeration forces, where higher density yields productivity gains. This plot also shows a significant disparity, an average variance of approximately seven log points of constant 2017 billion USD in cell GDP values with the same cell population. This substantial variation underscores the influence of factors beyond population in determining GDP.

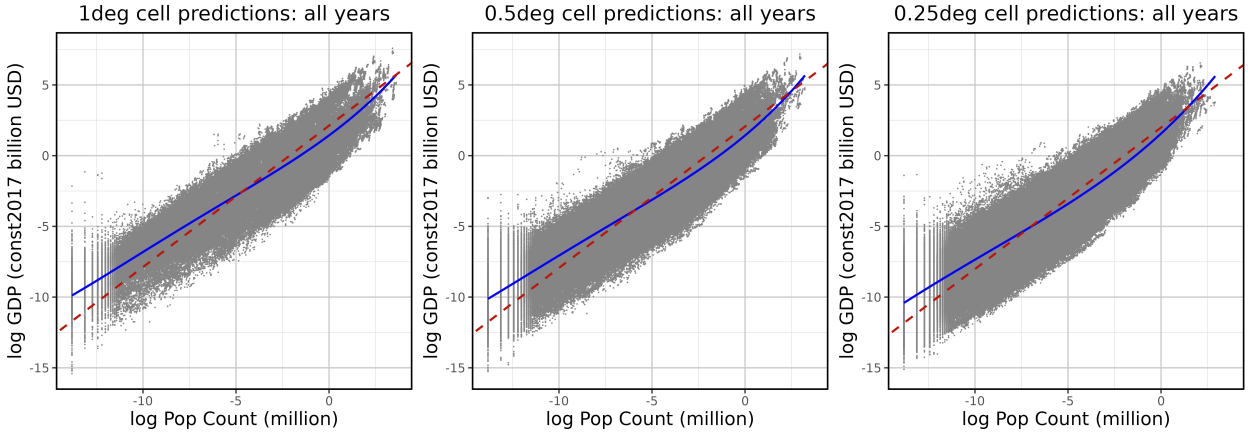

Figure 2: Cell Population Against Cell GDP

Note: For each resolution, the blue solid line represents the estimate from a fourth-degree polynomial regression, while the red dashed line represents the estimate from a linear regression. Both models are fitted using data from all years. For the 1-degree resolution, the polynomial regression is $y = 1.43 + 0.98x + 0.04x^2 + 0.004x^3 + 0.0001x^4$ and the linear regression is $y = x + 2.12$. For the 0.5-degree resolution, the polynomial regression is $y = 1.43 + 1.1x + 0.05x^2 + 0.004x^3 + 0.0001x^4$ and the linear regression is $y = x + 2.06$. For the 0.25-degree resolution, the polynomial regression is $y = 1.56 + 1.21x + 0.06x^2 + 0.003x^3 + 0x^4$ and the linear regression is $y = x + 1.99$.

This characteristic of our predictions is key for researchers seeking to capture nuanced regional output differences and their drivers. For instance, if one aims to estimate local total factor productivity (TFP) across regions, analyze the impact of advanced technology adoption, variations in labor skill levels, measure the impacts of infrastructure improvements on regional welfare, or perform counterfactual analyses to identify optimal policies, a dataset that allocates GDP solely based on population may yield misleading conclusions. Similarly, if researchers seek to assess the effects of trade openness on regional economic re-

silience or evaluate how proximity to ports or major markets influences economic outcomes, GDP estimates solely dependent on population might obscure critical spatial patterns and underestimate the role of trade infrastructure.

2.3 Method Performance

We evaluate the predictive performance of our methodology on out-of-sample countries using Group 5-Fold Cross-Validation. Table 1 presents the average results across these five folds, reporting metrics separately for developed, developing countries, and as a weighted mean of both groups. For each fold, the model is trained on a subset of the data (excluding one fold) using the optimal set of hyperparameters,[11] and then tested on the held-out fold. The R^2 values are calculated individually for developed, developing, and all countries, with the final R^2 values in the table representing the average across all five folds. The weighted value is computed as a weighted average of the developed and developing group values, with weights of 0.332 and 0.668, respectively, reflecting each group's share of total cells globally.

The R^2 of at least 0.92 shown in panel A of the Table 1 indicate strong predictive power for GDP levels. The model at the 1-degree resolution shows slightly better performance compared to the other two resolutions, likely due to the lower potential for error with fewer cells. The model also demonstrates slightly better performance for developed countries because of developed countries' availability of more reliable and detailed subnational data. Despite the coarser resolution of subnational data in developing countries (often exceeding 1-degree cells), we still include these regions.[12]

For year-over-year predictions of cell-level GDP changes, the models attain R^2 values of at least 62%. Again, more reliable and detailed regional data allow predictions for developed countries to outperform those for developing regions. When comparing across resolutions, the model at the 0.25-degree resolution demonstrates the best performance, likely due to the method used to construct the true GDP data at this scale rather than an inherent advantage of the model at finer resolutions. Recall that true GDP values are calculated by assuming constant GDP per capita within each county and distributing GDP based on population, which is also utilized as a predictor in the model. At the 0.25-degree resolution, many counties' areas are larger than the corresponding grid cells, potentially aligning the training process with the predictor structure. This alignment may contribute to the higher performance metrics observed. However, this does not imply that population alone determines the model's predictions. As demonstrated in Section 2.2, the 0.25-degree model's predictions

[11] The optimal set of hyperparameters is the same as Figure 1 and 2.

[12] In the Appendix Section 6.3, we compare models trained with and without data from developing countries.

also reveal a convex relationship between GDP and population and local GDP still varies substantially at equal population levels.

Panel C in the Table 1 presents some of the key variables shared across the three resolutions and their corresponding importance scores. These scores are calculated by summing the contributed variance reductions at each split across all trees in the forest. The values are meaningful for comparing across variables within a single model. The results highlight the critical role in capturing economic activity of nighttime lights (NTL) emitted from urban areas, population, and their lagged counterparts. Other important variables include urban areas and CO_2 emissions from transportation and manufacturing. Despite having lower importance scores, these variables still offer valuable complementary information that enhances the model's predictive performance across varying spatial resolutions.

Table 1: Cross-Validated Performance Metrics Across Spatial Resolutions

Metric	1-degree Model	0.5-degree Model	0.25-degree Model
Panel A: R^2 of Log GDP Level			
R^2 (Developed)	98.01%	97.81%	97.65%
R^2 (Developing)	96.08%	94.15%	92.94%
R^2 (All)	97.58%	97.10%	96.93%
Weighted R^2	96.72%	95.37%	94.51%
Panel B: R^2 of log(GDP in t) - log(GDP in t-1)			
R^2 (Developed)	66.99%	77.12%	86.01%
R^2 (Developing)	62.00%	62.05%	64.62%
R^2 (All)	63.31%	71.82%	80.60%
Weighted R^2	63.66%	67.07%	71.73%
Panel C: Variables and Importance Scores			
NTL from urban	42.34	8.24	4.08
Lag NTL from urban	25.44	5.47	4.11
Population	24.23	35.46	16.08
Lag population	20.75	22.51	18.88
Urban areas	1.76	1.54	0.90
Lag urban areas	1.64	1.57	0.46
Fossil CO2 from transportation	0.38	0.17	0.17
Lag fossil CO2 from transportation	0.60	0.22	0.20
Biofuel CO2 from manufacturing	0.35	0.23	0.24
Lag biofuel CO2 from manufacturing	0.42	0.23	0.24

2.4 Model Performance During The COVID-19 Pandemic

To test the predictions during extraordinary unprecedented episodes, we use one of the most significant recent disruptions: the COVID-19 pandemic. China, which experienced

considerable economic effects during the COVID-19 pandemic, has city-level GDP data that can be collected individually and was not included in our training set, making it an ideal out-of-sample test case. We collected city-level GDP data for seven major provinces in China: Guangdong, Henan, Hubei, Jiangsu, Shandong, Sichuan, and Zhejiang, and aggregated it into 1-degree cells. Figure 3 compares true and predicted values for pre-COVID years (2012 to 2019), the COVID year (2020), and the post-COVID year (2021). The predicted GDP values are obtained directly from our 1-degree model.

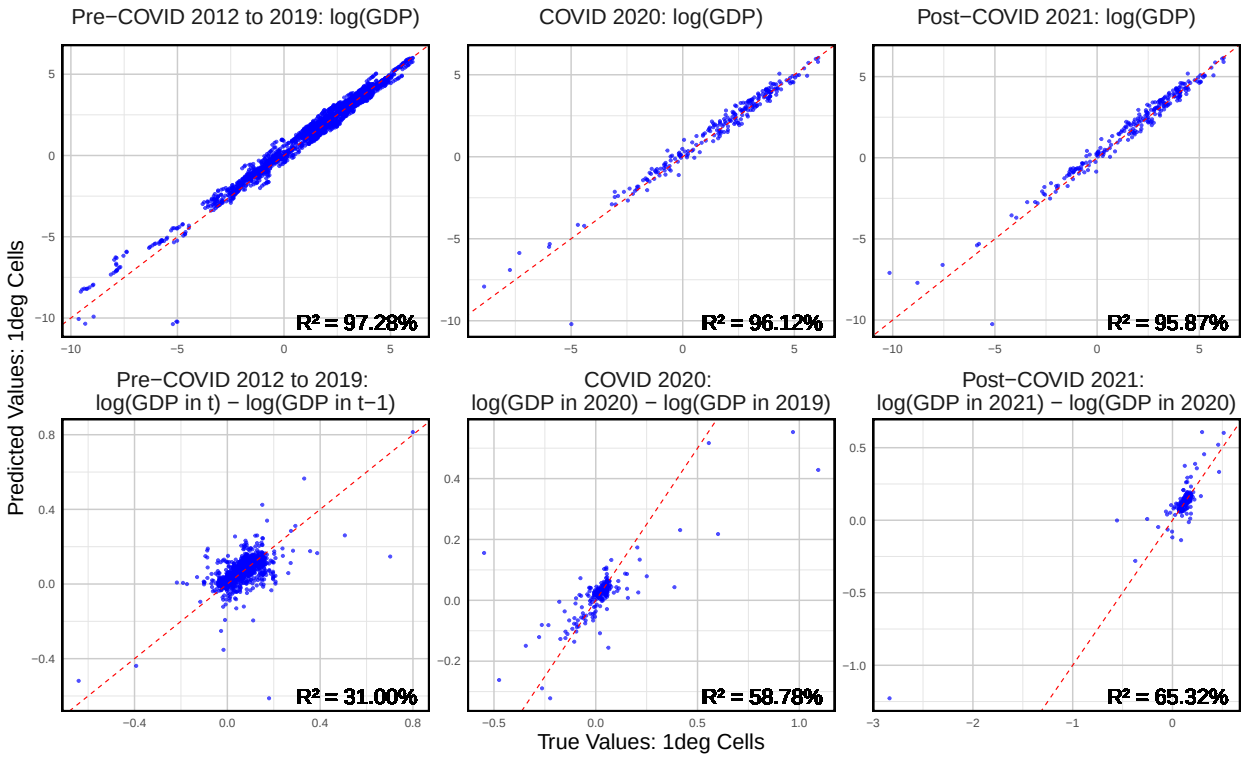

Figure 3: Model Predictions Against Actual Values in Billion Constant 2017 USD for Seven Leading Provinces in China

Note: The red dashed line represents the 45-degree line. Cells with a GDP value of zero are omitted to enable the calculation of logarithmic values. Only two out of 2380 data points are excluded due to zero GDP value.

The model shows strong performance in predicting out-of-sample GDP levels and year-over-year growth rates during both the COVID-19 shock and the subsequent recovery. The R^2 are greater than 0.95 for levels and at least 0.31 for annual changes. For the COVID-19 and post-COVID years, the model's performance on annual changes exceeds 0.58, consistent with the cross-validated performance shown in Panel C of Table 1. The R^2 value of 0.31 for log changes in GDP prior to 2019 could be attributed to issues with data reliability. Only after 2019 did the Chinese central government implement a standardized approach for

local governments to calculate regional GDP. This reform improved the comparability and accuracy of GDP data across regions.

To further demonstrate that the model's robust performance on COVID-19 affected data does not arise from prior exposure to COVID-19 years in the training dataset, we repeated the analysis by training the model on all available countries excluding China (as in Figure 3), but also exclude data in 2020 and 2021. We then tested the model predictions on the years 2020 and 2021. The results, which are shown in Appendix Section 5, maintain comparable R^2 values and high predictive accuracy. This demonstrates that the model's good performance in these tests does not rely on training on the COVID-19 experience in other countries.

Wuhan, one of the COVID-19 affected cities, can provide a useful case study for assessing the model's accuracy. Figure 4 compares the actual GDP and our estimates for Wuhan's main economic activity hub from 2012 to 2021. Our model effectively captures the city's overall GDP trend, successfully identifying periods of slower growth in 2016 and years of rapid expansion in 2017 and 2018. It also identifies the large economic disruptions in 2020 caused by the COVID-19 pandemic and the subsequent recovery in 2021. These findings demonstrate that the statistical prediction model can forecast both short-term economic shocks and long-term trends.

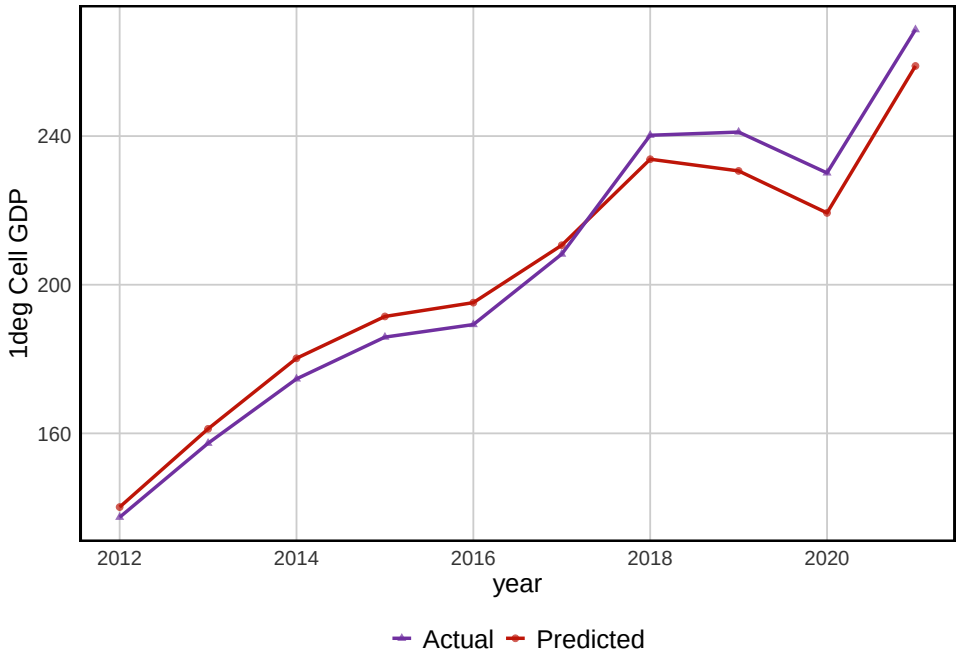

Figure 4: Model Predictions Against Actual Values in Billion Constant 2017 USD for Cell 21535: Main Economic Activity Hub of City Wuhan

3 Discussion

In this paper, we present a new method to predict out of sample local GDP based on predictor variables—such as remote sensing data, population data, and emissions data. By training the model on countries with detailed and accurate subnational data, we extend local GDP estimates to regions lacking granular economic data. The predictor data used are updated annually, allowing our global local GDP estimates to also be updated yearly and extend beyond the current 2012–2021 range. All data are accessible for download at http://bfidatastudio.org/gdp.

One immediate application of this dataset is its direct use as an outcome variable in various analyses. This is particularly relevant in development economics, where researchers often rely on nighttime light intensity as a proxy for economic development or income indicators. Our dataset, however, offers more precise subnational data to assess regional disparities and the development of low-income areas. For example, it could refine insights in studies like Storeygard (2016) analysis of economic growth in Sub-Saharan African cities, investigations into the effects of sanctions on North Korea (Lee, 2018), or analyses of the impact of agroclimatic similarity on migrant productivity in Indonesia (Bazzi et al., 2016).

This dataset also has significant implications for research using structural spatial models. These models often need observed local GDP per capita to estimate local productivity parameters through inversion procedures. Or, when applying exact-hat algebra, structural models may also require reliable income data to assess shock impacts. Compared to G-Econ, which relies heavily on population-based estimates, or GDP data derived only from nighttime lights, our dataset can enhance the accuracy of counterfactual predictions and provide additional data points and years for model estimation. This improvement benefits studies analyzing the heterogeneous effects of global warming across regions (Cruz and Rossi-Hansberg, 2024), assessing natural disaster impacts (Desmet et al., 2021), evaluating the effects of easing migration restrictions (Desmet, Nagy and Rossi-Hansberg, 2018), identifying optimal transport networks (Fajgelbaum and Schaal, 2020), and analyzing the welfare effects of transportation infrastructure improvements (Allen and Arkolakis, 2022) beyond the United States.

Finally, while our new dataset effectively captures broad patterns of economic growth, it has limitations in capturing highly localized economic growth. Forecasting economic growth, especially for out-of-sample countries, is inherently challenging, and users should approach specific regional growth analyses with caution, as our generalized models may lack precision at such granular levels. Nonetheless, the dataset remains a valuable and robust resource for economic research.

References

Allen, Treb, and Costas Arkolakis. 2022. "The Welfare Effects of Transportation Infrastructure Improvements." *The Review of Economic Studies*, 89(6): 2911–2957.

Bazzi, Samuel, Arya Gaduh, Alexander D Rothenberg, and Maisy Wong. 2016. "Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia." *American Economic Review*, 106(9): 2658–2698.

Breiman, Leo. 2001. "Random Forests." *Machine Learning*, 45: 5–32.

Bright, E., A. Rose, M. Urban, K. Sims, J. McKee, A. Reith, et al. 2012–2021. "LandScan Global Population Dataset." Oak Ridge National Laboratory. https://landscan.ornl.gov/ (accessed Jan 1, 2024).

Chen, Jiandong, Ming Gao, Shulei Cheng, Wenxuan Hou, Malin Song, Xin Liu, and Yu Liu. 2022. "Global 1 km× 1 km Gridded Revised Real Gross Domestic Product and Electricity Consumption During 1992-2019 Based on Calibrated Nighttime Light Data." *Scientific Data*, 9(1): 202.

Cruz, José-Luis, and Esteban Rossi-Hansberg. 2024. "The Economic Geography of Global Warming." *Review of Economic Studies*, 91(2): 899–939.

Desmet, Klaus, Dávid Krisztián Nagy, and Esteban Rossi-Hansberg. 2018. "The Geography of Development." *Journal of Political Economy*, 126(3): 903–983.

Desmet, Klaus, Robert E. Kopp, Scott A. Kulp, Dávid Krisztián Nagy, Michael Oppenheimer, Esteban Rossi-Hansberg, and Benjamin H. Strauss. 2021. "Evaluating the Economic Cost of Coastal Flooding." *American Economic Journal: Macroeconomics*, 13(2): 444–86.

Emissions Database for Global Atmospheric Research (EDGAR). 2012-2021. "Global Greenhouse Gas Emissions, Version 8.0, 2012-2021." European Commission, Joint Research Centre (JRC), and International Energy Agency (IEA). https://edgar.jrc.ec.europa.eu/ (accessed Jan 1, 2024).

Fajgelbaum, Pablo D, and Edouard Schaal. 2020. "Optimal Transport Networks in Spatial Equilibrium." *Econometrica*, 88(4): 1411–1452.

Friedl, M., and D. Sulla-Menashe. 2022. "MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V061 [Data Set]." NASA EOSDIS Land Processes Distributed

Active Archive Center. https://doi.org/10.5067/MODIS/MCD12Q1.061 (accessed Jan 1, 2024).

Global Gas Flaring Data. 2012-2023. "Flare Volume Estimates by Individual Flare Location." National Oceanic and Atmospheric Administration, Payne Institute at the Colorado School of Mines, World Bank, Global Flaring and Methane Reduction Partnership (GFMR). https://www.worldbank.org/en/programs/gasflaringreduction/global-flaring-data (accessed Jan 1, 2024).

Henderson, J Vernon, Adam Storeygard, and David N Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review*, 102(2): 994–1028.

Henderson, Vernon, Tim Squires, and David N. Weil. 2018. "The Global Spatial Distribution of Economic Activity: Nature, History, and the Role of Trade." *Quarterly Journal of Economics*, 133(1): 357–406.

Keola, Souknilanh, Magnus Andersson, and Ola Hall. 2015. "Monitoring Economic Development from Space: Using Nighttime Light and Land Cover Data to Measure Economic Growth." *World Development*, 66: 322–334.

Khachiyan, Arman, Anthony Thomas, Huye Zhou, Gordon Hanson, Alex Cloninger, Tajana Rosing, and Amit K Khandelwal. 2022. "Using Neural Networks to Predict Microspatial Economic Growth." *American Economic Review: Insights*, 4(4): 491–506.

Lee, Yong Suk. 2018. "International Isolation and Regional Inequality: Evidence from Sanctions on North Korea." *Journal of Urban Economics*, 103: 34–51.

Nordhaus, William D. 2006. "Geography and Macroeconomics: New Data and New Findings." *Proceedings of the National Academy of Sciences*, 103(10): 3510–3517.

Nunn, Nathan, and Diego Puga. 2012. "Ruggedness: The Blessing of Bad Geography in Africa." *Review of Economics and Statistics*, 94(1): 20–36.

Román, M. O., Z. Wang, Q. Sun, V. Kalb, S. D. Miller, A. Molthan, L. Schultz, J. Bell, E. C. Stokes, B. Pandey, K. C. Seto, et al. 2018. "NASA's Black Marble Nighttime Lights Product Suite." *Remote Sensing of Environment*, 210: 113–143.

Running, S., and M. Zhao. 2021. "MODIS/Terra Net Primary Production Gap-Filled Yearly L4 Global 500m SIN Grid V061 [Data Set]." NASA EOSDIS Land Processes

Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MOD17A3HGF.061 (accessed Jan 1, 2024).

Storeygard, Adam. 2016. "Farther on Down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa." *The Review of Economic Studies*, 83(3): 1263–1295.

Vogel, Kathryn Baragwanath, Gordon H Hanson, Amit Khandelwal, Chen Liu, and Hogeun Park. 2024. "Using Satellite Imagery to Detect the Impacts of New Highways: An Application to India." *NBER Working Paper No. 32047*.

Wenz, Leonie, Robert Devon Carr, Noah Kögel, Maximilian Kotz, and Matthias Kalkuhl. 2023. "DOSE–Global Data Set of Reported Sub-national Economic Output." *Scientific Data*, 10(1): 425.