Local GDP Estimates Around the World*

Esteban Rossi-Hansberg
University of Chicago

Jialing Zhang
University of Chicago

January 21, 2025

Appendix

Contents

*Esteban Rossi-Hansberg: earossih@uchicago.edu. Jialing Zhang: jialingzhang@uchicago.edu

1

1 Details of Data Source

1.1 GIS data

- Worldwide country level, province-equivalent level and county-equivalent level maps in Geopackage ("`.gpkg`") format are downloaded from GADM version 4.1 "geopackage" action in the sentence "You can also download this version as six separate layers (one for each level of subdivision/aggregation), as a geopackage database" from the following website: `https://gadm.org/download_world.html`. Save the downloaded file with the name "gadm_410−levels.gpkg" in the folder "step1_obtain_gis_data/inputs /CGZA_ADM1/".

- Another version of the province-level geometry file is the CGAZ dataset. Download "ADM1" global map in "geojson" format from the website: `https://www.geoboundaries.org/globalDownloads.html` (Runfola et al., 2020). Save the downloaded file with the name "geoBoundariesCGAZ_ADM1.geojson" in the folder "step1_obtain_gis_data/inputs/CGZA_ADM1/".

- Worldwide large lakes are excluded from the maps. The data used are "Global Lakes and Wetlands Database: Large Lake Polygons (Level 1)" from the website: `https://www.worldwildlife.org/publications/global-lakes-and-wetlands-database-large-lake-polygons-level-1` (Lehner and Döll, 2004). One with higher computer resources can even exclude permanent open water bodies with a surface area greater and equal to 0.1 km^2 and wetlands using other types of data from the website: `https://www.worldwildlife.org/pages/global-lakes-and-wetlands-database`. Save the downloaded files with filenames starting with "glwd_1.xxx" in the folder "step1_obtain_gis_data/inputs/large_inland_waters_geom_GLWD_level1/".

- We also need the spatial data from the paper named "DOSE – Global data set of reported sub-national economic output" (Wenz et al., 2023). Use the link: `https://doi.org/10.5281/zenodo.7659599todo` and download the files from the folder "DOSE replication files/Data/spatial data/". We also need to download files from `https://gadm.org/download_world36.html` by clicking the action "shapefiles" in the sentence "You can also donwload this version as six separate layers (one for each level of subdivision/aggregation), as a geopackage database or as shapefiles). Save the downloaded files in the folder "step1_obtain_gis_data/inputs/DOSE_spatial_data/ ".

- China city level geometry shape files are downloaded from the website: `https://github.com/GaryBikini/ChinaAdminDivisonSHP` (GaryBikini, 2024). Download all files in the folder named "3.City". Save the downloaded files in the folder "step1_obtain_gis_data/inputs/china_city/".

1.2 GDP data

1.2.1 Regional GDP data

Table 1 lists the sources of regional GDP data by country. Table 2 lists countries in the training sample and their average area of regional units used to construct cell-level GDP data for model training. If a country is included in the training sample, the most granular administrative level data collected are used to construct the cell-level GDP. The following details provide information on where to find the regional GDP datasets referenced in Table 1:

- Obtain data from "DOSE – Global data set of reported sub-national economic output"[1] (Wenz et al., 2023):

  - Click link https://doi.org/10.5281/zenodo.7573249, which is their paper's recommended data download link

  - Download the file "DOSE_V2.csv", and save it to folder ``inputs/gdp_data/ regional/DOSE/''

- Obtain data from OECD iLibrary - Regional economy[2] (OECD, 2024b):

  - Click link https://www.oecd-ilibrary.org/urban-rural-and-regional-d evelopment/data/oecd-regional-statistics_region-data-en#archive, which refers to OECD iLibrary - OECD Regional Statistics

  - Click the "Archive 2023" under the "Datasets Archives"

  - Click the "csv" bottom beside "Regional economy (Edition 2023)"

  - Save the file with the name "REGION_ECONOM−2023−1−EN−20240216T100059 2.csv" to folder "inputs/gdp_data/regional/oecd/"

  - Please note that for the following countries: "AUT", "BEL", "CAN", "CHL", "DNK", "ESP", "FIN", "FRA", "GBR", "HUN", "ITA", "LVA", "NLD", "NOR", "PRT", "ROU", and "SWE", the OECD provides some unregionalized data. We have removed this unregionalized data and rescaled the regional data to align with national GDP data from the World Bank or IMF.

- Obtain data from "OECD Data Explorer - Regions"[3] (OECD, 2024a):

  - Use the link https://data-explorer.oecd.org/, which refers to OECD Data Explorer main page.

  - Click "Regions" under "Regions, cities and local areas"

  - Choose the dataset "Gross domestic product - Regions"

---

[1]This dataset will not be updated; therefore, only the available years are utilized for training.

[2]This dataset will also not updated; therefore, we only retrieve data for the years 2012–2020. Data from 2021 should be obtained from "OECD Data Explorer - Regions".

[3]The data was downloaded on June 15, 2024, and may have been updated since. For replication, the latest files should be downloaded, and the corresponding R scripts must be updated to process the new data correctly.

- Click "Download / Unfiltered data in tabular text (CSV)", and save the file with the name "OECD.CFE.EDS,DSD_REG_ECO@DF_GDP,2.0+all.csv" to the folder "inputs/gdp_data/regional/oecd/"

- Also click "OECD Territorial correspondence table (xlsx)" , and save the file with the name "OECD Territorial correspondence − November 2023.xlsx" to the folder "inputs/gdp_data/regional/oecd/". We need this file to understand the heritage relationships between different id names.

- Obtain Russia's regional data after 2019[4] (*Russian Statistical Yearbook* 2022; *Russian Statistical Yearbook* 2023):

  - Click the website link https://eng.rosstat.gov.ru/Publications/document/74811, which refers to Russian Statistical Yearbook

  - Download the "Russian Statistical Yearbook 2023" and "Russian Statistical Yearbook 2022" in "RAR" format. Downloading the "pdf" format is also acceptable because we have to manually collect the data. Save the folders with name "russian_statistical_yearbook_20xx" in the folder " inputs/gdp_data/regional/ RUS/"

  - Match the regions' names from the yearbooks with OECD's definitions of id names. Follow the matches in the file "RUS.xlsx" in the folder "inputs/gdp_data /regional/RUS/". When updating to new years, add the new data to the file "RUS.xlsx".

- Obtain Brazil's regional data in 2021 (Geografia e Estatística (IBGE), 2024):[5]

  - Click the website link https://www.ibge.gov.br/en/statistics/economic/national-accounts/16855-regional-accounts-of-brazil.html, which refers to the dataset "SCR - System of Regional Accounts" on the Instituto Brasileiro de Geografia e Estatística (IBGE)

  - Download the "xls" file for the "GDP under the point of view of Income (2010-2021)". Save the file with the name "PIB_Otica_Renda_UF.xls" to the folder "inputs/gdp_data/regional/BRA/". It is crucial to choose "GDP under the point of view of **income**", instead of "production", as the former corresponds exactly to what OECD uses.

- Obtain China's regional data (Statistics of China, 2024):

  - Click the website link https://data.stats.gov.cn/english/index.htm, which refers to the main page of National Bureau of Statistics of China.

---

[4]Russia's regional GDP data for 2012-2019 are obtained directly from "OECD iLibrary - Regional economy" as described above. Since the OECD no longer updates Russia's data, regional GDP data for future years must be obtained from alternative sources.

[5]As of the OECD data download date (June 15, 2024), Brazil's regional data had not yet been updated. If, during future updates, OECD publishes Brazil's province data, this step can be skipped by directly downloading the updated data using the same procedure as for the "OECD Data Explorer." This step serves as a backup in case Brazil's data remains unpublished.

– Click "Regional / Annual by Province" and choose the following:

  * Year: LATEST20
  * Indicators: Gross

– Download the "xls" format and save it with the name "AnnualbyProvince.xls" to the folder "inputs/gdp_data/regional/CHN/".

- Obtain India's regional data (India, 2023):

  – Use the website link https://www.rbi.org.in/Scripts/AnnualPublication s.aspx?head=Handbook%20of%20Statistics%20on%20Indian%20States, which refers to the "Handbook of Statistics of Indian States" on the Reserve Bank of India.

  – Choose the year 2023, download the excel format of "Table 27: Gross State Domestic Product (Current Prices)" and save the file with the name "27T_15112023E301A02422494F73BFAFD6CDD84EEEAE.XLSX" in the folder "inputs/gdp_data/regional/IND/".

- Obtain Kazakhstan's regional data (National Statistics Agency for Strategic Planning and Republic of Kazakhstan, 2023):

  – Use the link https://stat.gov.kz/en/industries/economy/national-acc ounts/, which refers to the datasets of "Statistics of national accounts" on the Bureau of National Statistics Agency for Strategic Planning and Reforms of the Republic of Kazakhstan.

  – Find the "Gross regional product (January-December 2023)", scroll down, and find the "Gross regional product" under the section of "Dynamic tables". Download the file and save it with the name "1. Gross regional product.xlsx" to the folder "inputs/gdp_data/regional/KAZ/".

- Obtain USA's regional data (Economic Analysis U.S. Department of Commerce, 2023):

  – Use the link https://apps.bea.gov/histdatacore/Regional_Accounts_ne w.html, which refers to the "Regional Economic Accounts - Previously Published Estimates" of the Bureau of Economic Analysis U.S. Department of Commerce.

  – Click "Gross Domestic Product by County and Metro Area / December 8, 2022 / CAGDP2: GDP in Current Dollars by County and MSA". Download the folder and save it with the name "CAGDP2" to the folder "inputs/gdp_data/regional/USA/".

- Obtain Philippine's regional data (Authority, 2023):

  – Use the link https://psa.gov.ph/statistics/grdp/data-series, which refers to the dataset "GRDP Data Series" on the Philippine Statistics Authority

– Find the "2000-2023 Gross Regional Domestic Product" in the "Title" section. Download the file "GRDP by Region" and save it with the name "GRDP_Reg_2018PSNA_2000−2023.xlsx" to the folder "inputs/gdp_data/regional /PHL/".

- Obtain Kyrgyzstan's regional data (Kyrgyz Republic, 2023):

  – Use the link https://www.stat.kg/en/statistics/nacionalnye-scheta/, which refers to the dataset "National accounts" on the National Statistical Committee of the Kyrgyz Republic.

  – Download the file "1.01.00.09 Gross regional product (GRP) at current prices" under the section "Dynamic tables". Save the file with the name "1010009 Валовой региональный продукт (ВРП) в текущих ценах..xlsx" to the folder "inputs/ gdp_data/regional/KGZ/".

Table 1: Regional GDP data source by country

ISO	Available Years	Data Level	Source	Date
ALB	2012-2019	Third	DOSE – Global data set of reported sub-national economic output	June 15th, 2024
AUS	2012-2021	Second	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
AUT	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
BEL	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
BGR	2012-2020	Second, Third	OECD iLibrary - Regional economy (2012-2020);	June 15th, 2024
BIH	2012-2019	Second	DOSE – Global data set of reported sub-national economic output	June 15th, 2024
BLR	2012-2019	Second	DOSE – Global data set of reported sub-national economic output	June 15th, 2024

Table 1: Regional GDP data source by country (Continued)

ISO	Available Years	Data Level	Source	Date
BRA	2012-2021	Second	OECD iLibrary - Regional economy (2012-2020); Instituto Brasileiro de Geografia e Estatística - System of Regional Accounts (2021)	June 15th, 2024
CAN	2012-2021	Second	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
CHE	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
CHL	2012-2021	Second	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
CHN	2012-2021	Second	National Bureau of Statistics of China - Regional - Annual by province	June 15th, 2024
COL	2012-2021	Second	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
CZE	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
DEU	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
DNK	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
ECU	2012-2019	Third	DOSE – Global data set of reported sub-national economic output	June 15th, 2024
ESP	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024

Table 1: Regional GDP data source by country (Continued)

ISO	Available Years	Data Level	Source	Date
EST	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
FIN	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
FRA	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
GBR	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
GRC	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
HRV	2012-2020	Second, Third	OECD iLibrary - Regional economy (2012-2020);	June 15th, 2024
HUN	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
IDN	2012-2020	Second	OECD iLibrary - Regional economy (2012-2020)	June 15th, 2024
IND	2012-2021	Second	Reserve Bank of India - Handbook of Statistics on Indian States (2020-2021)	June 15th, 2024
IRL	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
ISL	2012-2020	Second, Third	OECD iLibrary - Regional economy (2012-2020);	June 15th, 2024

Table 1: Regional GDP data source by country (Continued)

ISO	Available Years	Data Level	Source	Date
ISR	2012-2021	Second	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
ITA	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
JPN	2012-2021	Second	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
KAZ	2012-2021	Third	National Statistics Agency for Strategic Planning and Republic of Kazakhstan	June 15th, 2024
KEN	2013-2017	Third	DOSE – Global data set of reported sub-national economic output	June 15th, 2024
KGZ	2012-2021	Third	National Statistical Committe of the Kyrgyz Republic	June 15th, 2024
KOR	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
LKA	2013-2019	Third	DOSE – Global data set of reported sub-national economic output	June 15th, 2024
LTU	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
LUX	2012-2020	Second, Third	OECD iLibrary - Regional economy (2012-2020);	June 15th, 2024
LVA	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024

Table 1: Regional GDP data source by country (Continued)

ISO	Available Years	Data Level	Source	Date
MEX	2012-2021	Second	OECD iLibrary - Regional economy (2012-2020); Instituto Nacional de Estadística y Geografía - National Accounts System (2021)	June 15th, 2024
MOZ	2012-2019	Third	DOSE – Global data set of reported sub-national economic output	June 15th, 2024
NLD	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
NOR	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
NZL	2012-2020	Second	OECD iLibrary - Regional economy (2012-2020)	June 15th, 2024
PER	2012-2020	Second	OECD iLibrary - Regional economy (2012-2020)	June 15th, 2024
PHL	2012-2021	Second	Philippine Statistics Authority	June 15th, 2024
POL	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
PRT	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
ROU	2012-2020	Second, Third	OECD iLibrary - Regional economy (2012-2020)	June 15th, 2024
RUS	2012-2021	Second	OECD iLibrary - Regional economy (2012-2019); Federal State Statistics Service - Russian Statistical Yearbook (2020-2021)	June 15th, 2024
SRB	2012-2018	Third	DOSE – Global data set of reported sub-national economic output	June 15th, 2024

Table 1: Regional GDP data source by country (Continued)

ISO	Available Years	Data Level	Source	Date
SVK	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
SVN	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
SWE	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
THA	2014-2018	Third	DOSE – Global data set of reported sub-national economic output	June 15th, 2024
TUR	2012-2021	Second, Third	OECD iLibrary - Regional economy (2012-2020); OECD Data Explorer - Regions (2021)	June 15th, 2024
USA	2012-2021	Second, Third	Bureau of Economic Analysis - Regional Economic Accounts	June 15th, 2024
UZB	2012-2019	Second	DOSE – Global data set of reported sub-national economic output	June 15th, 2024
VNM	2012-2018	Third	DOSE – Global data set of reported sub-national economic output	June 15th, 2024

1.2.2 National GDP data

Table 3 lists the sources of national GDP data by country. The following details provide information on where to find the national GDP datasets referenced in Table 3:

- Obtain data from World Bank (Bank, 2024d; Bank, 2024c; Bank, 2024b; Bank, 2024a):

  - Click link https://data.worldbank.org/indicator/SP.POP.TOTL for World Bank Population dataset with file name "API_SP.POP.TOTL_DS2_en_excel_v2_294626 .xls"; link https://data.worldbank.org/indicator/NY.GDP.MKTP.CD for World Bank GDP in current US$ with file name "API_NY.GDP.MKTP. CD_DS2_en_excel_v2_287504.xls"; link https://data.worldbank.org/indicator/NY.GDP.MKTP.PP.CD for World Bank GDP in current PPP adjusted international$ with file name "API_NY.GDP.MKTP.PP.CD_DS2_en_excel_v2_287316

Table 2: Average area of subnational units used for constructing cell-level GDP data by country

Developed Country	Average Subnational Area (km^2)	Developing Country	Average Subnational Area (km^2)
AUT	2383	ALB	2353
BEL	695	BIH	16988
BGR	3961	BLR	34328
CHE	1548	CHL	48669
CZE	5618	COL	34504
DEU	887	ECU	10695
DNK	3894	IDN	55053
ESP	8548	KEN	12276
EST	8641	KGZ	21134
FIN	16717	LKA	7314
FRA	6305	MOZ	70691
GBR	1356	PER	51699
GRC	2524	PHL	17006
HRV	2667	SRB	3122
HUN	4607	THA	6652
ITA	2756	UZB	29715
JPN	7832	VNM	5240
KOR	5635		
LTU	6459		
LVA	10699		
NLD	877		
NOR	29668		
NZL	21908		
POL	4254		
PRT	3668		
ROU	5642		
SWE	20600		
SVK	6107		
SVN	1692		
TUR	9500		
USA	2514		

.xls"; link https://data.worldbank.org/indicator/NY.GDP.MKTP.PP.KD for World Bank GDP in constant 2021 PPP adjusted international\$ with file name "API_NY.GDP.MKTP.PP.KD_DS2_en_excel_v2_288920.xls".

– Download as "Excel" files and save it in the corresponding folders in "inputs/ gdp_data/national/world_bank_data/xxx".

• Obtain the "Annual Index Value" data from US Census Bureau (Bureau, 2024):

- Use the link https://www.census.gov/en.html, which refers to main page of United States Census Bureau.

- Click "Topics / Income and Poverty / Income / Guidance for Data Users / Current versus Constant (or Real) Dollars".

- Download the excel file "Annual Index Value and Annual Percent Change in Price Series: 1947 to 2022" and save it to the folder "inputs/gdp_data/national/ US_census_bureau_data".

- Obtain data from IMF - WEO (Fund, 2024):[6]

  - Use the link https://www.imf.org/en/Publications/SPROLLS/world-economic-outlook-databases#sort=%40imfdate%20descending, which refers to the World Economic Outlook Databases (WEO) in International Monetary Fund (IMF).

  - Choose "World Economic OUtlook Database, April 2024". Click "By Countries / ALL COUNTRIES / Continue".

  - Choose datasets "Gross domestic product, current prices U.S. DOLLARS", "Gross domestic product, current prices PURCHASING POWER PARITY; INTERNATIONAL DOLLARS", "Gross domestic product per capita, constant prices PURCHASING POWER PARITY; 2017 INTERNATIONAL DOLLARS", and "Population PERSONS".

  - Click "Continue" and choose "2012" as "Start Year", "2021" as "End Year". Click "ISO Alpha-3 Code".

  - Download the file and save the file with name "WEO_Data.xlsx" in the folder "inputs/gdp_data/national/IMF_data".

- Obtain data from UNdata (UNdata, 2024):

  - Use the link https://data.un.org/Data.aspx?q=gdp&d=SNAAMA&f=grID%3a101%3bcurrID%3aUSD%3bpcFlag%3a1#SNAAMA, which refers to the dataset "Per capita GDP at current prices - US dollars" in UNdata.

  - Choose "Cuba", "Democratic People's Republic of Korea", "Eritrea", and years 2012 to 2022.

  - Download the files in csv format, value separated in comma. Save it with the name "UNdata_Export_20240613_003754887.csv" to the folder "inputs/gdp_data/national /UN_data".

---

[6]The primary data source is the IMF, as it converts GDP reported on a fiscal year basis into a calendar year format, aligning with predictors that are primarily satellite-based. When updating to future years, care must be taken in selecting the dataset's end year, as some entries may be estimates rather than actual values. Additionally, note that data for certain countries may change during updates.

Table 3: National GDP and population data source by country

ISO	Data	Source	Date Retrieved
BMU, CYM, CUW, GRL, KXK, LIE, MCO, SXM, SYR, TCA, PSE	Population	World Bank - Population	June 15th, 2024
	GDP in current US dollars	World Bank - GDP (current US$)	June 15th, 2024
	GDP in constant 2017 US dollars	Calculated using World Bank - GDP (current US$) and Annual Index Value from US Census Bureau	June 15th, 2024
	GDP in current international dollars PPP adjusted	World Bank - GDP, PPP (current international $)	June 15th, 2024
	GDP in constant 2017 international dollars PPP adjusted	Calculated using World Bank - GDP, PPP (current international $) and World Bank - GDP, PPP (constant 2021 international $)	June 15th, 2024
CUB, ERI, PRK	Population	World Bank - Population	June 15th, 2024
	GDP in current US dollars	Calculated using UNdata - Per capita GDP at current prices - US dollars and World Bank - Population	June 15th, 2024
	GDP in constant 2017 US dollars	Calculated using UNdata - Per capita GDP at current prices - US dollars, World Bank - Population, and Annual Index Value from US Census Bureau	June 15th, 2024
	GDP in current international dollars PPP adjusted	NA	June 15th, 2024
	GDP in constant 2017 international dollars PPP adjusted	NA	June 15th, 2024
Other countries	Population	IMF - WEO - Population	June 15th, 2024
	GDP in current US dollars	IMF - WEO - Gross domestic product, current prices, US dollars	June 15th, 2024
	GDP in constant 2017 US dollars	Calculated using IMF - WEO - Gross domestic product, current prices, US dollars and Annual Index Value from US Census Bureau	June 15th, 2024
	GDP in current international dollars PPP adjusted	IMF - WEO - Gross domestic product, current prices PURCHASING POWER PARITY; INTERNATIONAL DOLLARS	June 15th, 2024
	GDP in constant 2017 international dollars PPP adjusted	Calculated using IMF - WEO - Gross domestic product per capita, constant prices PURCHASING POWER PARITY; 2017 INTERNATIONAL DOLLARS and IMF - WEO - Population	June 15th, 2024

1.2.3 Calculate Different Measures of GDP

There are four measures of GDP in our study: current USD, constant 2017 USD, current PPP-adjusted international dollars, and constant 2017 PPP-adjusted international dollars. Each type of GDP measure serves a specific purpose and requires different calculations to provide accurate economic comparisons. Below, we will explain what these different GDP measures mean and how to calculate them.

- GDP in Current USD:

– This is the Gross Domestic Product (GDP) measured in current U.S. dollars. It represents the total value of all goods and services produced within a country in a given year, converted to USD using the current exchange rates. This measure is influenced by inflation, exchange rate fluctuations, and changes in the price level. GDP in current USD is useful for understanding the nominal size of an economy and comparing it to others using the same currency (USD).

- GDP in Constant 2017 USD:

  – This is the GDP adjusted for inflation and expressed in constant 2017 U.S. dollars. By using a base year (in this case, 2017), it removes the effects of price level changes over time, allowing for comparison of economic output across different years in real terms. This measure helps to isolate the actual growth in economic activity.

  – In our study, this measure is calculated using:

  $$\text{Constant 2017 USD} = \text{Current USD} \times \left(\frac{\text{2017 Price Index}}{\text{Current Year Price Index}} \right) \quad (1)$$

  – The price index is the Chained Consumer Price Index for All Urban Consumers (C-CPI-U) obtained from US Census Bureau.

- GDP in Current PPP-Adjusted International \$:

  – This GDP measure adjusts for purchasing power parity (PPP) and is expressed in current international dollars. Purchasing Power Parity (PPP) is a method of measuring the relative purchasing power of different countries' currencies over the same types of goods and services. International dollars are a hypothetical currency that has the same purchasing power over GDP as the U.S. dollar has in the United States. By using PPP adjustments, this measure accounts for differences in price levels between countries, providing a more accurate comparison of economic output and living standards. Unlike GDP in current USD, which is influenced by exchange rates, the PPP-adjusted measure reflects the value of goods and services in terms of what they can actually buy in each country.

- GDP in Constant 2017 PPP-Adjusted International \$

  – This is the GDP adjusted for both purchasing power parity (PPP) and inflation, expressed in constant 2017 international dollars. By using a base year (2017) and adjusting for PPP, it allows for comparison of economic output across different years and countries, accounting for both inflation and differences in price levels. This measure provides a consistent basis for comparing real economic growth and living standards over time and across countries, isolating real growth from both inflation and price level differences.

  – In our study, this measure is calculated using:

  $$\begin{array}{l} \text{Const 2017} \\ \text{PPP prices} \end{array} = \begin{array}{l} \text{Const 2021} \\ \text{PPP prices} \end{array} \times \left(\frac{\text{2017 GDP in Current PPP Prices}}{\text{2017 GDP in Const 2021 PPP Prices}} \right) \quad (2)$$

1.2.4 China city level GDP data

Table 4: City-Level GDP Data Sources for Seven Major Provinces in China

Province	Source	Dataset Name
Guangdong	http://stats.gd.gov.cn/gdtjnj/	Guangdong Province Statistical Yearbook 2023: 2-14 Gross Domestic Product by City
Henan	https://tjj.henan.gov.cn/tjfw/tjcbw/tjnj/	Statistical Yearbook 2023: 2-9 Gross Domestic Product by City (2021)
		Statistical Yearbook 2022: 2-9 Gross Domestic Product by City (2020)
		Statistical Yearbook 2021: 2-9 Gross Domestic Product by City (2019)
		Statistical Yearbook 2020: 2-8 Gross Domestic Product by City (2018)
		Statistical Yearbook 2018: 2-9 Gross Domestic Product by City (2017)
		Statistical Yearbook 2017: 3-9 Gross Domestic Product by City (2016)
		Statistical Yearbook 2016: 3-9 Gross Domestic Product by City (2015)
		Statistical Yearbook 2015: 3-9 Gross Domestic Product by City (2014)
		Statistical Yearbook 2014: 3-10 Gross Domestic Product by City (2013)
		Statistical Yearbook 2013: 3-10 Gross Domestic Product by City (2012)
Hubei	https://tjj.hubei.gov.cn/tjsj/sjkscx/tjnj/qstjnj/	Hubei Province Statistical Yearbook 2022: 0115-Gross Domestic Product of Cities and Prefectures (2021).xls
		Statistical Yearbook 2021: 0115-Gross Domestic Product of Cities and Prefectures (2020).xls
		Statistical Yearbook 2020: 0115-Gross Domestic Product of Cities and Prefectures (2019).xls
		Statistical Yearbook 2019: 0115-Gross Domestic Product of Cities and Prefectures (2018).xls
		Statistical Yearbook 2018: 0119-Gross Domestic Product of Cities and Prefectures (2017).xls
		Statistical Yearbook 2017: 0119-Gross Domestic Product of Cities and Prefectures (2016).xls
		Statistical Yearbook 2016: 0119-Gross Domestic Product of Cities and Prefectures (2015).xls
		Statistical Yearbook 2015: 0119-Gross Domestic Product of Cities and Prefectures (2014).xls
		Statistical Yearbook 2014: 0119-Gross Domestic Product of Cities and Prefectures (2013).xls
		Statistical Yearbook 2013: 0119-Gross Domestic Product of Cities and Prefectures (2012).xls
Jiangsu	https://www.jiangsu.gov.cn/col/col84736/index.html	Jiangsu Province Statistical Yearbook 2023: 2-12 Gross Domestic Product by Region
		Jiangsu Province Statistical Yearbook 2022: 2-12 Gross Domestic Product by Region
		Jiangsu Province Statistical Yearbook 2021: 2-12 Gross Domestic Product by Region
		Jiangsu Province Statistical Yearbook 2020: 2-12 Gross Domestic Product by Region
		Jiangsu Province Statistical Yearbook 2019: 2-12 Gross Domestic Product by Region

Province	Source	Dataset Name
		Jiangsu Province Statistical Yearbook 2018: 2-12 Gross Domestic Product by Region
Shandong	http://tjj.shandong.gov.cn/col/col6279/index.html	Shandong Province Statistical Yearbook 2023: 2-9 Gross Domestic Product by Region
		Shandong Province Statistical Yearbook 2022: 2-6 Gross Domestic Product by Region (2021)
		Shandong Province Statistical Yearbook 2021: 2-6 Gross Domestic Product by Region (2020)
		Shandong Province Statistical Yearbook 2020: 2-6 Gross Domestic Product by Region
		Shandong Province Statistical Yearbook 2019: 2-9 Gross Domestic Product by Region
		Shandong Province Statistical Yearbook 2018: 2-9 Gross Domestic Product by Region
		Shandong Province Statistical Yearbook 2017: 2-9 Gross Domestic Product by Region
		Shandong Province Statistical Yearbook 2016: 2-8 Gross Domestic Product by Region
		Shandong Province Statistical Yearbook 2015: 2-8 Gross Domestic Product by Region
		Shandong Province Statistical Yearbook 2014: 2-9 Gross Domestic Product by Region
Sichuan	https://oversea.cnki.net/knavi/YearbookDetail?pcode=CYFD&pykm=YSCTN	Sichuan Province Statistical Yearbook 2023: 2-7 Gross Domestic Product by Region
Zhejiang	https://tjj.zj.gov.cn/col/col1525563/index.html	Zhejiang Province Statistical Yearbook 2023: 17-2 Major Indicators of National Economy by City and County (2022)
		Zhejiang Province Statistical Yearbook 2022: 17-2 Major Indicators of National Economy by City and County (2021)
		Zhejiang Province Statistical Yearbook 2021: 17-2 Major Indicators of National Economy by City and County (2020)
		Zhejiang Province Statistical Yearbook 2020: 17-2 Major Indicators of National Economy by City and County (2019)
		Zhejiang Province Statistical Yearbook 2019: 17-2 Major Indicators of National Economy by City and County (2018)
		Zhejiang Province Statistical Yearbook 2018: 17-2 Major Indicators of National Economy by City and County (2017)
		Zhejiang Province Statistical Yearbook 2017: 17-2 Major Indicators of National Economy by City and County (2016)
		Zhejiang Province Statistical Yearbook 2016: 17-2 Major Indicators of National Economy by City and County (2015)
		Zhejiang Province Statistical Yearbook 2015: 17-2 Major Indicators of National Economy by City and County (2014)
		Zhejiang Province Statistical Yearbook 2014: 17-2 Major Indicators of National Economy by City and County (2013)
		Zhejiang Province Statistical Yearbook 2013: 17-2 Major Indicators of National Economy by City and County (2012)

1.3 Predictors data

- Population data:

  - The population data for our study is sourced from the LandScan Global Metadata, developed by the Oak Ridge National Laboratory (Bright et al., n.d.). The link is https://landscan.ornl.gov/metadata.

- Choose "LandScan Global Metadata" and download for years 2000 to 2021. Save the downloaded files in the folder "/inputs/population/".

- Nighttime Light (NTL) data:

  - The Nighttime Light (NTL) data utilized in our study is the VIIRS VNP46A4 product from NASA's Black Marble suite (Román et al., 2018). The link is https://ladsweb.modaps.eosdis.nasa.gov/archive/allData/5000/VNP46A4/.

  - Data for years 2012 to 2022 are all downloaded and processed. Save the corresponding year files in the folder "/inputs/NTL_VNP46A4/20xx/001".

- Net Primary Productivity (NPP) data

  - The Net Primary Productivity (NPP) data is the MOD17A3HGF Version 6.1 product from the MODIS suite (Running and Zhao, 2021). The link is https://search.earthdata.nasa.gov/search?q=C2565791034-LPCLOUD.

  - Data for years 2012 to 2022 are all downloaded and processed. Save the corresponding year files in the folder "/inputs/NPP_V061/20xx/".

- Landcover data:

  - The landcover data that our model uses is the MCD12Q1 version 6.1 product from the MODIS suite (Friedl and Sulla-Menashe, 2022). The link is https://search.earthdata.nasa.gov/search?q=C2484079608-LPCLOUD.

  - Data for years 2012 to 2021 are all downloaded and processed. Save the corresponding year files in the folder "/inputs/landcover_MCD12Q1V061/20xx/".

- Carbon dioxide (CO2) emissions data:

  - The carbon dioxide (CO2) emissions data in our study are sourced from the EDGAR - Emissions Database for Global Atmospheric Research community. Data version "v8.0_GHG 1970-2022 (CO2, CH4, N2O, F-gases)". The link is https://edgar.jrc.ec.europa.eu/dataset_ghg80.

  - They separate the CO2 emissions into "IEA-EDGAR CO2" and "EDGAR_CO2bio". For both of them, we aggregate sector-specific emissions to three main categories: manufacturing combustion, heavy industry, and transportation.

  - For manufacturing combustion category, go to the section "Annual sector-specific gridmaps (1970-2022)", click "IEA-EDGAR CO2" and "EDGAR_CO2bio" under the subsection of "Combustion for manufacturing". Then download the files named "COMPLETE TIMESERIES [2022-1970] ENETCDF (xxx_emi_nc,zip)".

  - For heavy industry category, it includes the subsection "Power industry", "Oil refineries & Transformation industry", "Fuel Exploitation", "Non-metallic minerals production", "Iron and steel production", and "Non-ferrous metals production". Note for "EDGAR_CO2bio", only subsections "Power industry", "Oil refineries & Transformation industry", and "Fuel Exploitation" have the data.

- For transportation category, it includes the subsection "Road transportation", and "Shipping".
- Save the corresponding year files in the folders "/inputs/CO2_bio_specific_sectors /sector name/" and "/inputs/CO2_non_org_specific_sectors/sector name/".

- Gas flare data:

  - The gas flare data in our study are sourced from the Global Gas Flaring Data dataset in the Global Flaring and Methane Reduction Partnership (GFMR) community (Zhizhin et al., 2021; Elvidge, Zhizhin, Hsu, et al., 2013; Elvidge, Zhizhin, Baugh, et al., 2016). The link is https://www.worldbank.org/en/programs/ga sflaringreduction/global-flaring-data.
  - Click the "Dataset: 2012 - 2022 Flare Volume Estimates by Individual Flare Location" to download and save the corresponding file in the folder "/inputs/ gas_flare_data/". The file should be named "GGFR-Flaring-Dashboard-Data-March292023.xlsx".

- Terrain Ruggedness Index data:

  - We use the Terrain Ruggedness Index produced in Nunn and Puga, 2012, calculated as a 30 arc-second grid across the earth. The link is https://diegopuga. org/data/rugged/.
  - The dataset was downloaded on 01/20/2023. Save the corresponding file "/inputs /ruggedness/" and name it "tri.txt".

The figure 1 shows the locations of gas flare spots used in this study.
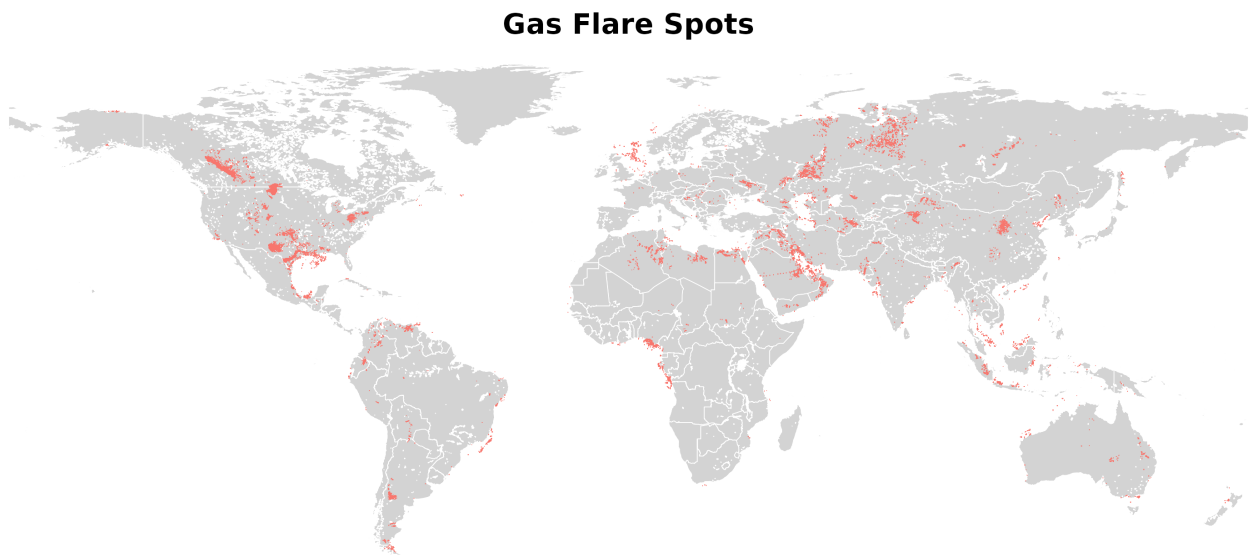
Gas Flare Spots

Figure 1: Locations of gas flare spots used in this study

1.4   Data Processing for Predictors

To prepare the geometry units for aggregating predictor data, the world is divided into grids with resolutions of 1°, 0.5°, and 0.25° in the EPSG:4326/WGS84 coordinate system. These grids are intersected with world country maps (as described in Section 1.1) to create country-cell-level geometry units. For specific countries, the geometry is at the state-equivalent level: Australia (8 units), Brazil (27 units), Canada (13 units), China (31 units), India (33 units), Kazakhstan (16 units), Mexico (32 units), Russia (83 units), and USA (51 units). Large inland water areas are then removed from the maps. This creates the final geometry units used for both GDP and predictor data. For NTL data aggregation only, an additional geometry unit is created by dividing the country/state-cell-level geometry into three land use categories: urban, cropland, and others by intersecting it with corresponding land use data.

Before aggregating predictors' data into the geometry unit created above, the following processing steps are needed:

The original NASA VIIRS VNP46A4 Black Marble NTL dataset separates data by satellite angles and into snow-covered and snow-free periods. For this study, we use the snow-free layer covering all satellite angles. The dataset has already been pre-processed to adjust for atmospheric and lunar reflectance conditions and to remove noise sources such as stray light, twilight, clouds, scan edge artifacts, and temporary lights (e.g., lightning, wildfires, fireworks). However, persistent lights from gas flares and fishing boats still remain. To address this, we further process it by excluding the lights within a 0.2deg radius around gas flaring locations with positive gas flaring volumes. Gas flaring locations, provided by the Global Gas Flaring Data from the GFMR community, include annual point locations with estimated flaring volumes. Positive flaring locations for each year are identified and corresponding exclusion zones are applied. The original NTL data are in a sinusoidal coordinate system and are reprojected to EPSG:4326 using bilinear interpolation. This ensures spatial consistency but can introduces slight changes to original pixel values. The processed data are aggregated to year-country/state-cell-landuse levels.

The NPP dataset is also in a sinusoidal coordinate system and undergoes the same reprojection process as the NTL data.

The original land use data classify land into 11 categories. We select and aggregate them into five categories: urban, water, cropland (combining cropland, forest cropland, and herbaceous cropland), forest (combining open and dense forest), and snow ice. Recall that only large inland waters are removed, so smaller water bodies can still appear in the data.

The original CO_2 emissions data categorizes emissions into multiple sectors. We select and aggregate into six categories: fossil CO_2 manufacturing combustion, biofuel CO_2 manufacturing combustion, fossil CO_2 heavy industry (power industry + oil refineries transformation industry + fuel exploitation + non-metallic minerals production + iron and steel production + non-ferrous metals production), biofuel CO_2 heavy industry (power industry + oil refineries transformation industry + fuel exploitation), fossil CO_2 transportation (road transportation + shipping), biofuel CO_2 transporation (road transportation + shipping). EDGAR generates gridded emissions data by downscaling national emissions to each 0.1deg cell using sector-specific spatial proxies. In our study, we will convert all predictors to share terms rather than level terms, so we only care about EDGAR's downscaling

method. EDGAR allocates industry emissions based on the locations and emissions of energy and manufacturing facilities, while transportation emissions are distributed according to the length and intensity of road networks and shipping routes. This distribution process relies mainly on statistical calculations, not machine learning or economic models. And their input data for each sector are primarily sourced from national government agencies, company reports, and international organizations. Excluding these datasets has minimal impact on GDP level predictions but can reduce the accuracy of growth rate predictions.

The LandScan population dataset, when aggregated to the national level, slightly differs from the IMF WEO and World Bank national population datasets used in this study. To ensure consistency, LandScan data are rescaled to match IMF and World Bank population values.

After preprocessing the predictor data, the values for each cell's predictors are then extracted and converted into shares, ready for training the random forest model.

2   Training Countries and Reference for GDP Share Prediction

Our model is trained using data from countries that provide GDP information at the county-equivalent level for developed nations and at a slightly broader administrative level for developing ones. These training countries span Northern America, Southern America, Europe, Africa, and Asia. Figure 2 shows the training countries.

Our model predicts cell GDP share which by definition is the proportion of GDP attributed to a specific cell relative to the aggregate GDP of its higher administrative entity. Correspondingly, our predictors are expressed as shares; for instance, the fraction of population, the proportion of nighttime light intensity, and the percentage of urban areas. The shares are calculated at the national level for most countries, indicating, for example, the share of national nighttime lights that a particular cell occupies. However, to improve the accuracy of our predictions by incorporating more detailed subnational GDP information, we treat province-equivalent units as "parent countries" in nations such as Australia, Brazil, Canada, China, India, Kazakhstan, Mexico, Russia, and the United States, where such province-level GDP data are available. In these cases, both the predictor shares and the predicted cell GDP shares are relative to the provincial level. Figure 2 illustrates the boundaries of "parent countries" in this context.
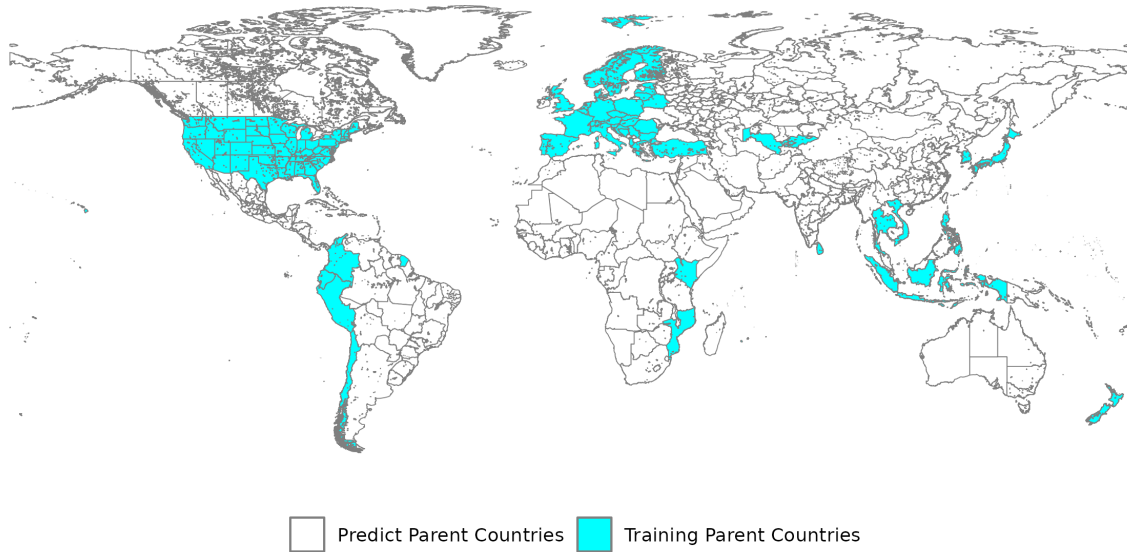
Figure 2: Boundaries of parent countries in the model

3 Post-adjustments

This section presents the number of cells affected by the post-adjustment process for each censoring threshold at each resolution.

Table 5: 1-degree cells affected by post-adjustments: shreshold 0 pop per km2 land areas

Year	Total number of cells	Number of cells affected	Affected cells with zero pop	Percentage of inhabited cells affected	Affected cells with positive GDP
2012	21,194	4,112	3,707	2%	398
2013	21,194	3,888	3,461	2%	420
2014	21,194	3,901	3,453	3%	439
2015	21,194	3,928	3,491	2%	429
2016	21,194	3,812	3,358	3%	446
2017	21,194	3,813	3,339	3%	464
2018	21,194	3,924	3,462	3%	456
2019	21,194	4,210	3,722	3%	484
2020	21,194	4,612	4,079	3%	526
2021	21,194	4,420	3,881	3%	534

Table 6: 1-degree cells affected by post-adjustments: shreshold 0.01 pop per km2 land areas

Year	Total number of cells	Number of cells affected	Affected cells with zero pop	Percentage of inhabited cells affected	Affected cells with positive GDP
2012	21,194	5,283	3,709	9%	415
2013	21,194	5,030	3,462	9%	423
2014	21,194	5,124	3,455	9%	436
2015	21,194	5,119	3,493	9%	432
2016	21,194	5,188	3,361	10%	454
2017	21,194	5,208	3,342	10%	454
2018	21,194	5,253	3,470	10%	436
2019	21,194	5,431	3,729	10%	445
2020	21,194	5,968	4,085	11%	450
2021	21,194	5,848	3,886	11%	461

Table 7: 1-degree cells affected by post-adjustments: shreshold 0.02 pop per km2 land areas

Year	Total number of cells	Number of cells affected	Affected cells with zero pop	Percentage of inhabited cells affected	Affected cells with positive GDP
2012	21,194	5,692	3,709	11%	419
2013	21,194	5,480	3,462	11%	425
2014	21,194	5,550	3,455	12%	430
2015	21,194	5,548	3,493	12%	426
2016	21,194	5,616	3,361	13%	449
2017	21,194	5,633	3,342	13%	445
2018	21,194	5,706	3,470	13%	431
2019	21,194	5,884	3,729	12%	448
2020	21,194	6,318	4,085	13%	444
2021	21,194	6,244	3,886	14%	451

Table 8: 1-degree cells affected by post-adjustments: shreshold 0.05 pop per km2 land areas

Year	Total number of cells	Number of cells affected	Affected cells with zero pop	Percentage of inhabited cells affected	Affected cells with positive GDP
2012	21,194	6,383	3,709	15%	420
2013	21,194	6,212	3,462	16%	422
2014	21,194	6,295	3,455	16%	440
2015	21,194	6,327	3,493	16%	443
2016	21,194	6,376	3,361	17%	454
2017	21,194	6,387	3,342	17%	445
2018	21,194	6,411	3,470	17%	428
2019	21,194	6,555	3,729	16%	448
2020	21,194	6,902	4,085	16%	443
2021	21,194	6,877	3,886	17%	458

Table 9: 0.5-degree cells affected by post-adjustments: shreshold 0 pop per km2 land areas

Year	Total number of cells	Number of cells affected	Affected cells with zero pop	Percentage of inhabited cells affected	Affected cells with positive GDP
2012	73,728	18,327	17,391	2%	919
2013	73,728	17,098	16,071	2%	1,010
2014	73,728	17,211	16,161	2%	1,031
2015	73,728	17,291	16,250	2%	1,020
2016	73,728	17,291	16,205	2%	1,056
2017	73,728	17,370	16,293	2%	1,055
2018	73,728	17,616	16,546	2%	1,042
2019	73,728	18,686	17,601	2%	1,066
2020	73,728	20,570	19,433	2%	1,113
2021	73,728	20,071	18,927	2%	1,128

Table 10: 0.5-degree cells affected by post-adjustments: shreshold 0.01 pop per km2 land areas

Year	Total number of cells	Number of cells affected	Affected cells with zero pop	Percentage of inhabited cells affected	Affected cells with positive GDP
2012	73,728	21,911	17,419	8%	916
2013	73,728	20,655	16,081	8%	972
2014	73,728	20,942	16,179	8%	987
2015	73,728	20,995	16,261	8%	979
2016	73,728	21,221	16,222	9%	1,003
2017	73,728	21,384	16,309	9%	1,000
2018	73,728	21,520	16,583	9%	988
2019	73,728	22,357	17,646	8%	995
2020	73,728	24,452	19,471	9%	962
2021	73,728	24,239	18,969	10%	987

Table 11: 0.5-degree cells affected by post-adjustments: shreshold 0.02 pop per km2 land areas

Year	Total number of cells	Number of cells affected	Affected cells with zero pop	Percentage of inhabited cells affected	Affected cells with positive GDP
2012	73,728	23,365	17,420	11%	935
2013	73,728	22,165	16,081	11%	958
2014	73,728	22,494	16,179	11%	970
2015	73,728	22,582	16,261	11%	965
2016	73,728	22,820	16,223	11%	984
2017	73,728	22,910	16,309	11%	979
2018	73,728	23,075	16,583	11%	964
2019	73,728	23,888	17,646	11%	971
2020	73,728	25,717	19,471	12%	918
2021	73,728	25,648	18,969	12%	937

Table 12: 0.5-degree cells affected by post-adjustments: shreshold 0.05 pop per km2 land areas

Year	Total number of cells	Number of cells affected	Affected cells with zero pop	Percentage of inhabited cells affected	Affected cells with positive GDP
2012	73,728	25,744	17,420	15%	960
2013	73,728	24,710	16,081	15%	931
2014	73,728	25,107	16,181	16%	926
2015	73,728	25,156	16,261	15%	942
2016	73,728	25,331	16,224	16%	953
2017	73,728	25,471	16,311	16%	956
2018	73,728	25,649	16,585	16%	946
2019	73,728	26,301	17,648	15%	931
2020	73,728	27,745	19,473	15%	896
2021	73,728	27,830	18,971	16%	909

Table 13: 0.25-degree cells affected by post-adjustments: shreshold 0 pop per km2 land areas

Year	Total number of cells	Number of cells affected	Affected cells with zero pop	Percentage of inhabited cells affected	Affected cells with positive GDP
2012	271,692	85,793	83,537	1%	2,204
2013	271,692	79,412	76,978	1%	2,365
2014	271,692	80,405	77,966	1%	2,368
2015	271,692	80,784	78,333	1%	2,387
2016	271,692	81,597	79,093	1%	2,429
2017	271,692	82,404	79,860	1%	2,470
2018	271,692	83,015	80,506	1%	2,434
2019	271,692	87,240	84,748	1%	2,433
2020	271,692	94,604	92,138	1%	2,393
2021	271,692	94,284	91,793	1%	2,426

Table 14: 0.25-degree cells affected by post-adjustments: shreshold 0.01 pop per km2 land areas

Year	Total number of cells	Number of cells affected	Affected cells with zero pop	Percentage of inhabited cells affected	Affected cells with positive GDP
2012	271,692	94,485	83,832	6%	2,138
2013	271,692	88,746	77,048	6%	2,213
2014	271,692	89,796	78,119	6%	2,232
2015	271,692	90,074	78,478	6%	2,232
2016	271,692	90,909	79,260	6%	2,283
2017	271,692	91,779	80,032	6%	2,302
2018	271,692	92,196	80,797	6%	2,285
2019	271,692	95,766	85,025	6%	2,284
2020	271,692	103,621	92,397	6%	2,188
2021	271,692	103,731	92,049	7%	2,265

Table 15: 0.25-degree cells affected by post-adjustments: shreshold 0.02 pop per km2 land areas

Year	Total number of cells	Number of cells affected	Affected cells with zero pop	Percentage of inhabited cells affected	Affected cells with positive GDP
2012	271,692	99,391	83,835	8%	2,112
2013	271,692	93,698	77,050	9%	2,144
2014	271,692	94,984	78,121	9%	2,185
2015	271,692	95,406	78,480	9%	2,185
2016	271,692	96,230	79,261	9%	2,207
2017	271,692	97,131	80,032	9%	2,221
2018	271,692	97,528	80,797	9%	2,202
2019	271,692	100,781	85,025	8%	2,180
2020	271,692	108,263	92,398	9%	2,062
2021	271,692	108,850	92,052	9%	2,120

Table 16: 0.25-degree cells affected by post-adjustments: shreshold 0.05 pop per km2 land areas

Year	Total number of cells	Number of cells affected	Affected cells with zero pop	Percentage of inhabited cells affected	Affected cells with positive GDP
2012	271,692	107,285	83,841	12%	2,030
2013	271,692	102,482	77,055	13%	2,014
2014	271,692	104,076	78,125	13%	2,054
2015	271,692	104,417	78,482	13%	2,059
2016	271,692	105,125	79,265	13%	2,066
2017	271,692	106,154	80,034	14%	2,066
2018	271,692	106,446	80,803	13%	2,054
2019	271,692	109,311	85,027	13%	2,010
2020	271,692	115,781	92,402	13%	1,939
2021	271,692	116,742	92,056	14%	1,979

4 Model Within Training Sample Fit

Recall that in the paper Section 2.1 and 2.2, we present the results of models trained using data from 2012 to 2021 for available countries (excluding China) under the optimal hyper-parameters. Here we present the within-sample fit of the models on the training sample. We report the the R^2 of log GDP levels and annual changes for each country in the training sample. Note that the predicted values are out-of-bag (OOB) estimates. The out-of-bag estimate is calculated using predictions from trees that excluded the cell from their training.
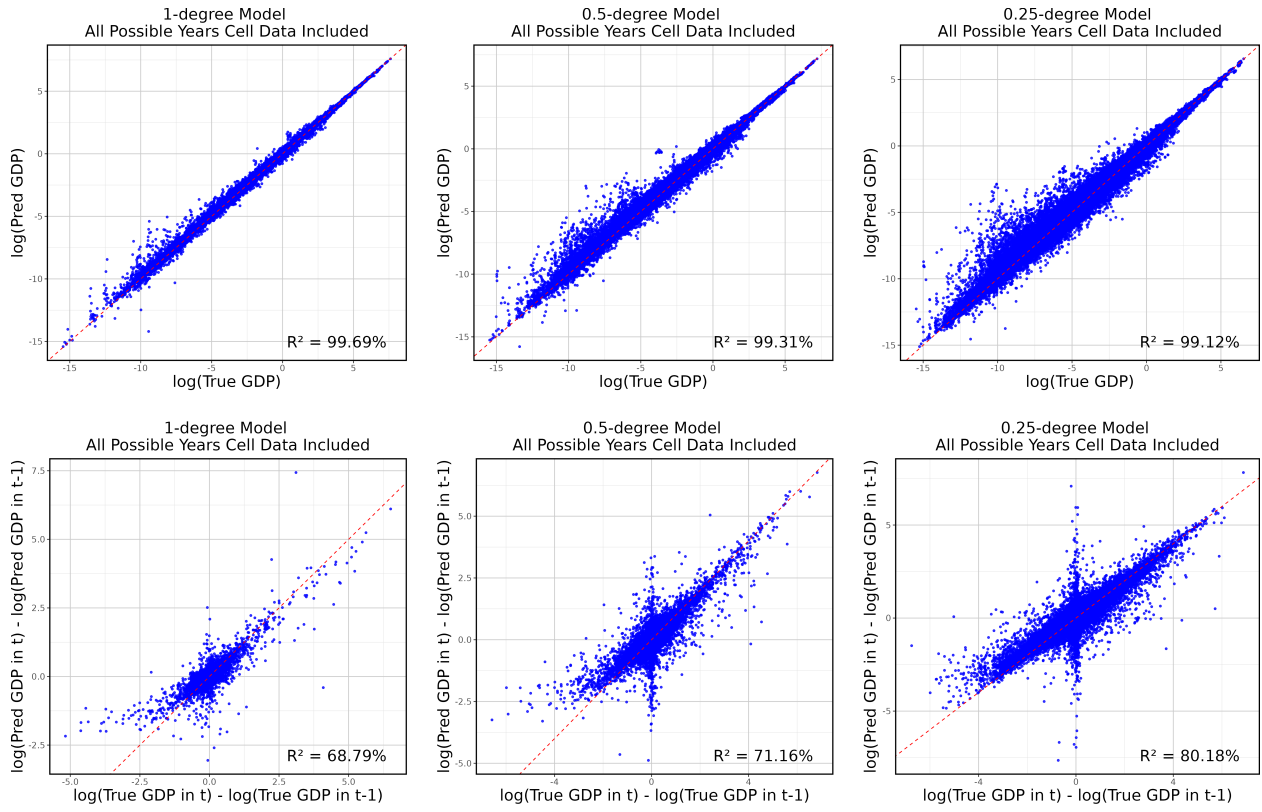
Figure 3: Predicted vs. True GDP Across Different Spatial Resolutions for Training Data

Table 17: R^2 Results for the 1-Degree Model: Comparing Predicted vs. Actual Log Cell GDP

Iso: Developed	R2: Developed	Iso: Developing	R2: Developing
AUT	99.92%	ALB	99.96%
BEL	99.94%	BIH	99.80%
BGR	99.78%	BLR	99.68%
CHE	99.94%	CHL	99.34%
CZE	99.95%	COL	99.58%
DEU	99.90%	ECU	99.76%
DNK	99.93%	IDN	99.73%
ESP	99.86%	KEN	97.49%
EST	99.59%	KGZ	99.63%
FIN	99.86%	LKA	99.82%
FRA	99.55%	MOZ	99.69%
GBR	99.81%	PER	99.49%
GRC	99.79%	PHL	99.72%
HRV	99.61%	SRB	99.83%
HUN	99.80%	THA	99.41%
ITA	99.73%	UZB	99.17%
JPN	99.89%	VNM	99.64%
KOR	99.92%		
LTU	99.57%		
LVA	99.54%		
NLD	99.89%		
NOR	99.75%		
NZL	99.87%		
POL	99.84%		
PRT	99.95%		
ROU	99.73%		
SVK	99.86%		
SVN	99.92%		
SWE	99.85%		
TUR	97.70%		
USA	99.64%		

Table 18: R^2 Results for the 0.5-Degree Model: Comparing Predicted vs. Actual Log Cell GDP

Iso: Developed	R2: Developed	Iso: Developing	R2: Developing
AUT	99.67%	ALB	99.88%
BEL	99.80%	BIH	98.52%
BGR	99.09%	BLR	99.45%
CHE	99.74%	CHL	99.17%
CZE	99.89%	COL	99.17%
DEU	99.69%	ECU	98.99%
DNK	99.83%	IDN	99.37%
ESP	99.46%	KEN	96.30%
EST	99.11%	KGZ	98.96%
FIN	99.80%	LKA	99.76%
FRA	99.51%	MOZ	98.72%
GBR	99.60%	PER	99.12%
GRC	99.51%	PHL	99.56%
HRV	98.32%	SRB	99.44%
HUN	99.53%	THA	98.70%
ITA	98.70%	UZB	98.83%
JPN	99.76%	VNM	99.61%
KOR	99.71%		
LTU	99.12%		
LVA	99.28%		
NLD	99.86%		
NOR	99.75%		
NZL	99.39%		
POL	99.58%		
PRT	99.71%		
ROU	99.32%		
SVK	99.69%		
SVN	99.76%		
SWE	99.82%		
TUR	97.44%		
USA	98.87%		

Table 19: R^2 Results for the 0.25-Degree Model: Comparing Predicted vs. Actual Log Cell GDP

Iso: Developed	R2: Developed	Iso: Developing	R2: Developing
AUT	99.54%	ALB	99.75%
BEL	99.50%	BIH	95.72%
BGR	98.33%	BLR	98.62%
CHE	99.57%	CHL	99.27%
CZE	99.77%	COL	99.17%
DEU	99.18%	ECU	98.74%
DNK	99.67%	IDN	98.71%
ESP	99.35%	KEN	97.05%
EST	97.65%	KGZ	98.63%
FIN	99.74%	LKA	99.70%
FRA	98.97%	MOZ	97.11%
GBR	99.21%	PER	99.29%
GRC	99.42%	PHL	98.89%
HRV	99.16%	SRB	97.62%
HUN	98.96%	THA	98.33%
ITA	98.59%	UZB	98.70%
JPN	99.04%	VNM	98.73%
KOR	99.42%		
LTU	98.05%		
LVA	98.62%		
NLD	99.71%		
NOR	99.50%		
NZL	99.51%		
POL	98.68%		
PRT	99.50%		
ROU	98.26%		
SVK	99.20%		
SVN	99.70%		
SWE	99.77%		
TUR	96.41%		
USA	98.63%		

Table 20: R^2 Results for the 1-Degree Model: Comparing Predicted vs. Actual Log Differences in Cell GDP (log(cell GDP in t) - log(cell GDP in t-1))

Iso: Developed	R2: Developed	Iso: Developing	R2: Developing
AUT	74.01%	ALB	72.07%
BEL	93.34%	BIH	51.13%
BGR	62.86%	BLR	73.50%
CHE	-6.46%	CHL	81.19%
CZE	90.37%	COL	62.36%
DEU	81.85%	ECU	15.80%
DNK	86.68%	IDN	45.25%
ESP	65.69%	KEN	11.17%
EST	86.40%	KGZ	41.75%
FIN	93.46%	LKA	-5.69%
FRA	73.63%	MOZ	49.29%
GBR	74.89%	PER	75.69%
GRC	72.00%	PHL	75.95%
HRV	61.65%	SRB	70.33%
HUN	66.10%	THA	-230.63%
ITA	88.55%	UZB	41.28%
JPN	77.90%	VNM	-173.09%
KOR	44.79%		
LTU	85.46%		
LVA	43.90%		
NLD	86.26%		
NOR	69.27%		
NZL	84.53%		
POL	83.28%		
PRT	72.87%		
ROU	74.08%		
SVK	87.83%		
SVN	92.46%		
SWE	92.18%		
TUR	-121.43%		
USA	61.08%		

Table 21: R^2 Results for the 0.5-Degree Model: Comparing Predicted vs. Actual Log Differences in Cell GDP (log(cell GDP in t) - log(cell GDP in t-1))

Iso: Developed	R2: Developed	Iso: Developing	R2: Developing
AUT	69.06%	ALB	80.34%
BEL	-38.74%	BIH	14.34%
BGR	35.22%	BLR	71.10%
CHE	74.55%	CHL	82.73%
CZE	86.45%	COL	70.65%
DEU	63.62%	ECU	-129.28%
DNK	93.93%	IDN	32.60%
ESP	24.29%	KEN	31.61%
EST	78.96%	KGZ	19.06%
FIN	93.15%	LKA	41.59%
FRA	80.62%	MOZ	27.60%
GBR	58.03%	PER	76.19%
GRC	69.50%	PHL	73.31%
HRV	-19.16%	SRB	48.81%
HUN	65.32%	THA	-23.98%
ITA	82.86%	UZB	52.55%
JPN	70.19%	VNM	-36.71%
KOR	55.23%		
LTU	77.82%		
LVA	53.67%		
NLD	70.10%		
NOR	90.34%		
NZL	86.16%		
POL	90.24%		
PRT	-6.94%		
ROU	67.21%		
SVK	86.04%		
SVN	74.32%		
SWE	94.33%		
TUR	20.76%		
USA	66.22%		

Table 22: R^2 Results for the 0.25-Degree Model: Comparing Predicted vs. Actual Log Differences in Cell GDP (log(cell GDP in t) - log(cell GDP in t-1))

Iso: Developed	R2: Developed	Iso: Developing	R2: Developing
AUT	76.41%	ALB	72.34%
BEL	49.77%	BIH	43.71%
BGR	49.93%	BLR	74.30%
CHE	71.46%	CHL	85.92%
CZE	77.71%	COL	81.25%
DEU	31.91%	ECU	-64.50%
DNK	90.76%	IDN	28.69%
ESP	82.40%	KEN	69.11%
EST	79.47%	KGZ	13.96%
FIN	95.56%	LKA	30.66%
FRA	78.90%	MOZ	21.96%
GBR	69.53%	PER	91.88%
GRC	86.89%	PHL	29.41%
HRV	67.03%	SRB	41.62%
HUN	-33.87%	THA	-52.89%
ITA	72.43%	UZB	76.24%
JPN	25.17%	VNM	-186.18%
KOR	68.67%		
LTU	84.60%		
LVA	59.82%		
NLD	65.97%		
NOR	93.88%		
NZL	90.89%		
POL	65.76%		
PRT	76.27%		
ROU	35.40%		
SVK	67.80%		
SVN	80.76%		
SWE	95.18%		
TUR	-1.16%		
USA	82.00%		

5 Model Performance Under COVID Shock

In the paper, we evaluate the model's performance under the COVID-19 shock using China's data, which is excluded from the training sample. Here, we further demonstrate that the model's strong performance on COVID-affected data is not due to prior exposure to COVID years in the training sample. To verify this, we train the models using data from 2012 to 2019 for all available countries (excluding China) and assess whether the model's performance remains consistent on out-of-sample data. Note that all years of data for China are treated as out-of-sample and only the years 2020 and 2021 are considered out-of-sample for the training countries.

The results of the China test, as shown in Figure 4, align closely with the findings presented in the paper. Similarly, the tests conducted on the training countries, presented in Figures 5, 6, and 7, for the years 2020 and 2021 also demonstrate strong performance. These results highlight the model's ability to generalize effectively, even in the absence of specific shocks in the training data.
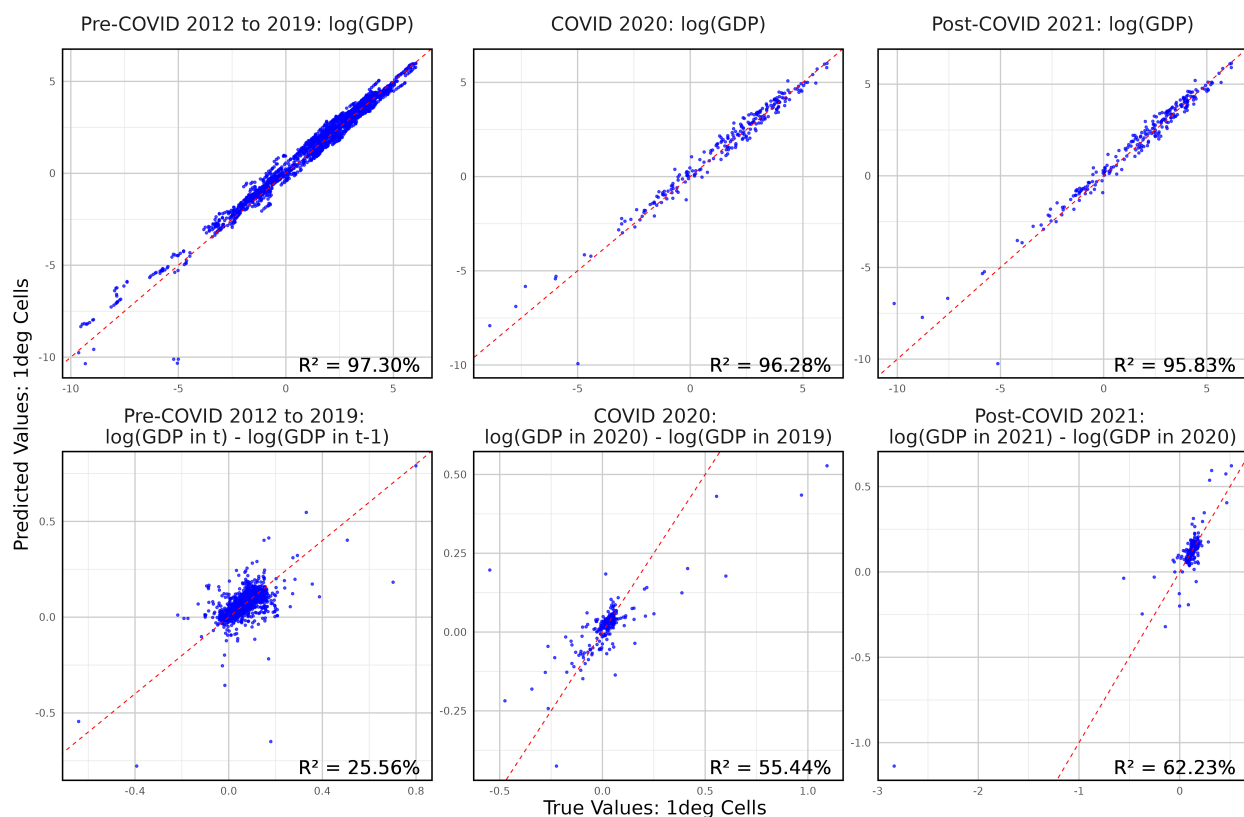
Figure 4: Model Predictions Against Actual Values in Billion Constant 2017 USD for Seven Leading Provinces in China

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.
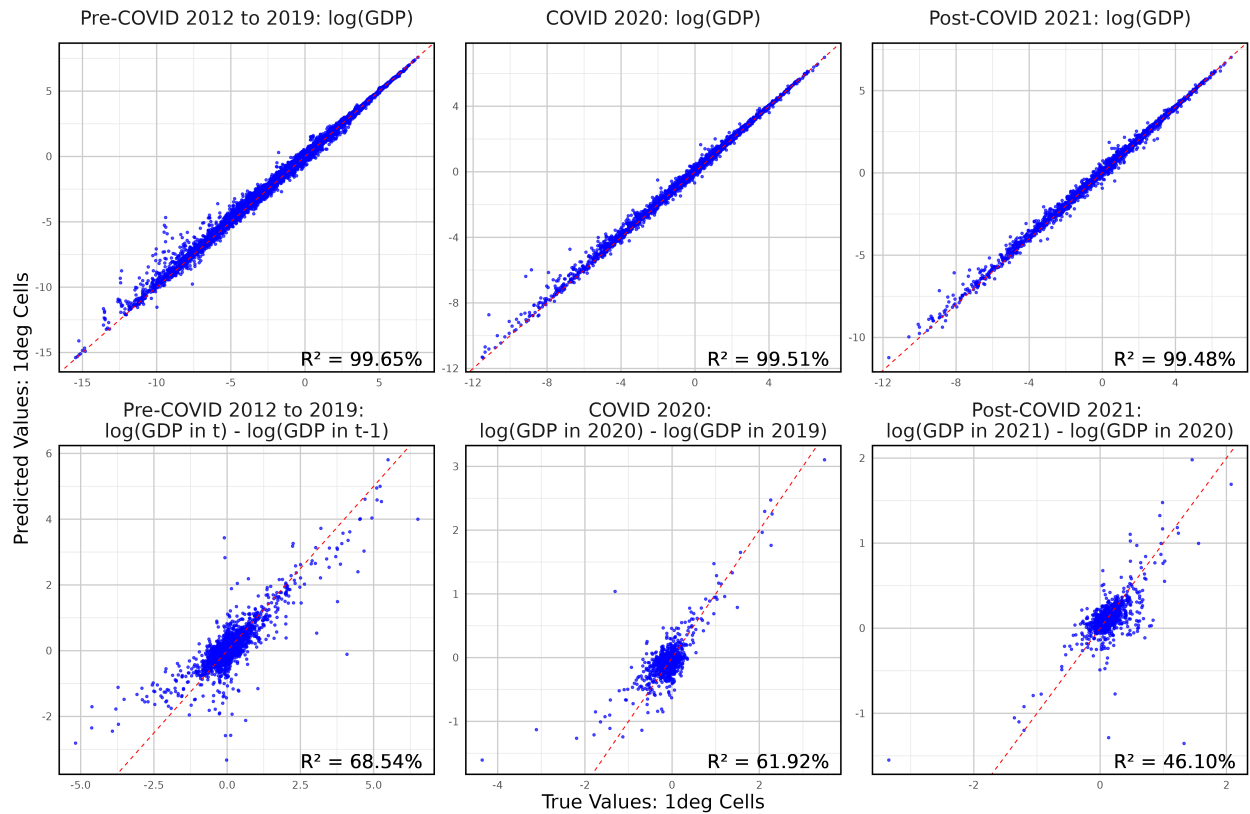
Figure 5: 1deg Model Predictions Against Actual Values in Billion Constant 2017 USD for All Training Countries

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. Data for 2012 to 2019 are within training sample, so use out-of-bag predictions. Data for year 2020 and 2021 are predictions from the model.
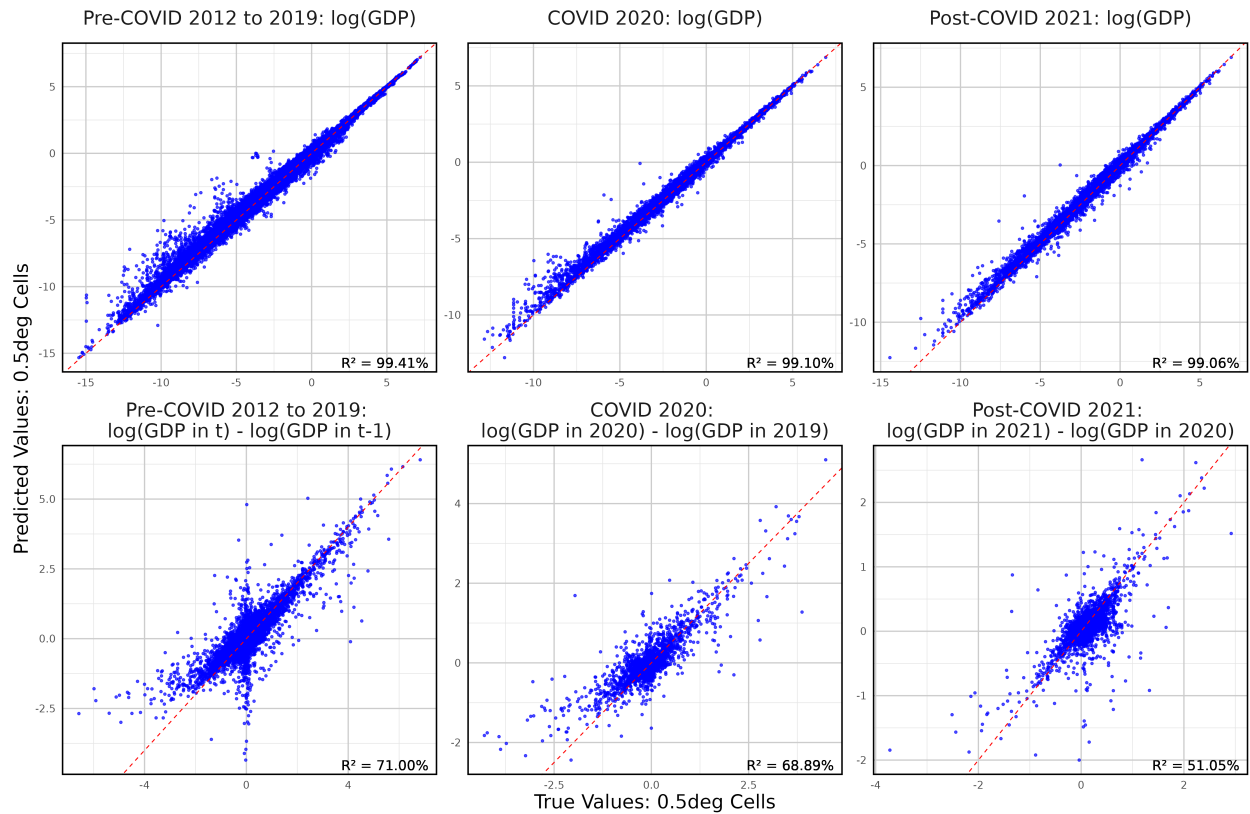
Figure 6: 0.5deg Model Predictions Against Actual Values in Billion Constant 2017 USD for All Training Countries

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. Data for 2012 to 2019 are within training sample, so use out-of-bag predictions. Data for year 2020 and 2021 are predictions from the model.
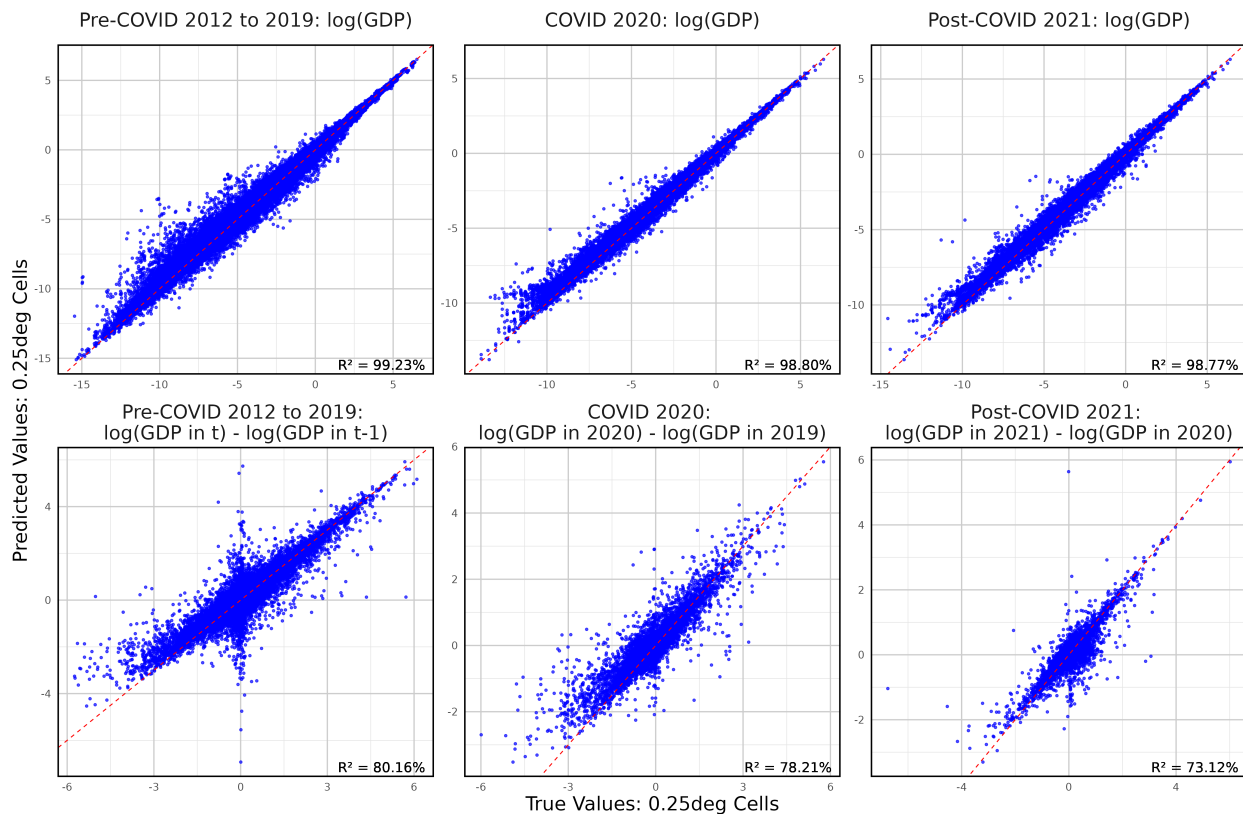
Figure 7: 0.25deg Model Predictions Against Actual Values in Billion Constant 2017 USD for All Training Countries

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. Data for 2012 to 2019 are within training sample, so use out-of-bag predictions. Data for year 2020 and 2021 are predictions from the model.

6 Robustness Check

6.1 Compare Benchmark Models with Models Tuned Based on Mean Square Error

In this section, we compare the results of two models: the model trained using data from 2012 to 2021 for all available countries (excluding China) with hyperparameters chosen to maximize the R^2 of annual log changes in GDP, and a model trained on the same data but with hyperparameters chosen to minimize the mean squared error (MSE) of GDP share. This comparison is essential because traditional hyperparameter tuning in machine learning is often based on minimizing MSE for the predicted variable - in our case, GDP share.

Table 23 shows the cross-validation performance of the model tuned to minimize MSE. Compare this table with the corresponding table in the paper, we show that while GDP level predictions are largely similar between the two models, tuning for maximizing annual log change R^2 improves the accuracy of year-over-year GDP changes, especially for 1deg model.

We then use the two models to predict all global cells and compare their predictions. Figure 8 demonstrates that the level predictions are highly consistent, while the annual changes exhibit slight differences especially at 1deg resolution.

We also performed the same tests on the MSE-tuned model for comparison with the benchmark model, as shown in Figures 9, 10, and 11. The results indicate that the benchmark model outperforms the MSE-tuned model in capturing annual changes in GDP.

Table 23: Cross-Validated Performance Metrics Across Spatial Resolutions for MSE Tuned Models

	1-degree Model	0.5-degree Model	0.25-degree Model
Panel A: Mean Square Error (MSE)			
Mean Square Error (Developed)	0.00019	0.00005	0.00001
Mean Square Error (Developing)	0.00011	0.00002	0.00001
Mean Square Error (All)	0.00017	0.00004	0.00001
Weighted Mean Square Error	0.00014	0.00003	0.00001
Panel B: R^2 of Log GDP Level			
R^2 (Developed)	96.66%	97.86%	97.82%
R^2 (Developing)	93.73%	94.41%	92.99%
R^2 (All)	96.01%	97.09%	96.84%
Weighted R^2	94.7%	95.56%	94.59%
Panel C: R^2 of log(GDP in t) - log(GDP in t-1)			
R^2 (Developed)	42.16%	75.16%	84.56%
R^2 (Developing)	49.55%	64.79%	59.48%
R^2 (All)	39.93%	71.41%	78.77%
Weighted R^2	47.09%	68.24%	67.82%
Panel D: Variables and Importance Scores			
NTL from urban	28.76	6.82	4.44
Lag NTL from urban	22.28	4.55	4.43
Population	21.27	39.38	15.7
Lag population	22.98	22.46	18.12
Urban areas	6.05	1.06	1.18
Lag urban areas	5.87	0.75	0.51
Fossil CO2 from transportation	2.04	0.14	0.19
Lag fossil CO2 from transportation	1.88	0.1	0.25
Biofuel CO2 from manufacturing	0.76	0.21	0.24
Lag biofuel CO2 from manufacturing	0.92	0.23	0.24
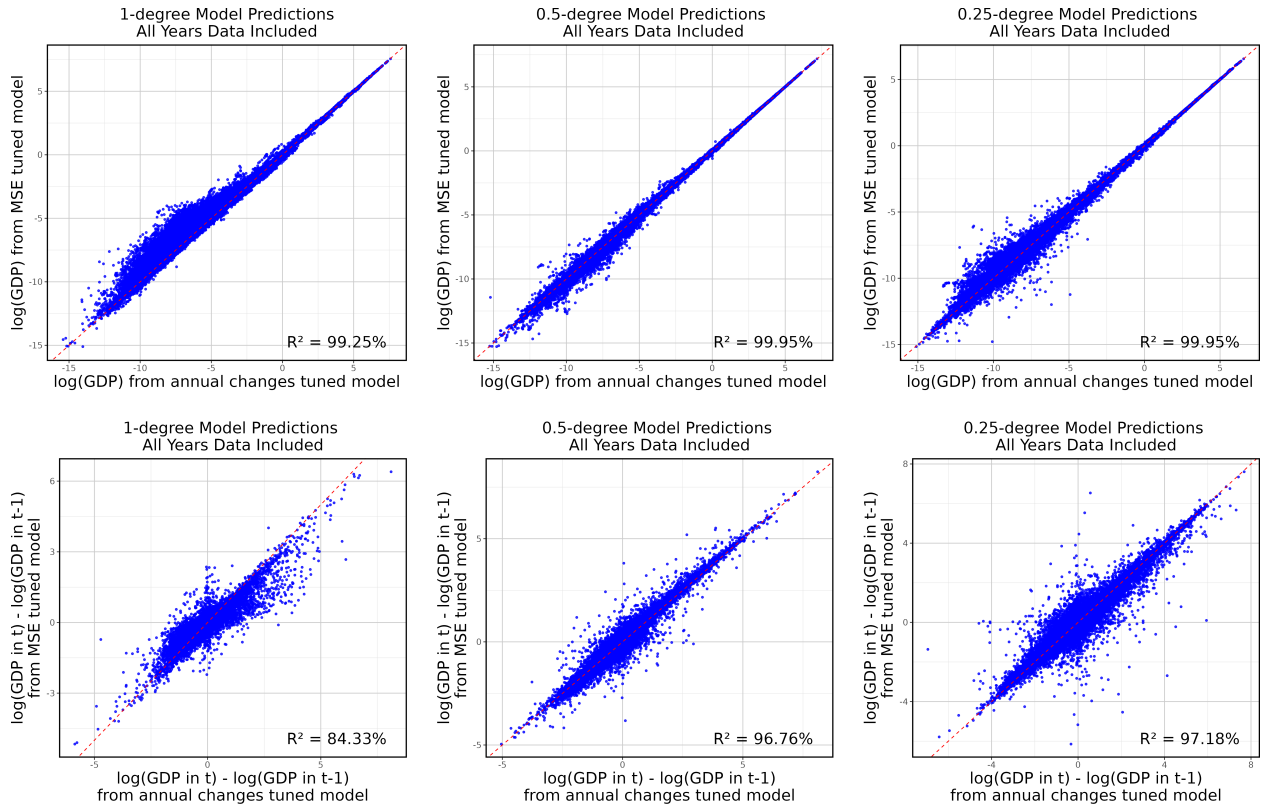
Figure 8: Comparison of Model Predictions: MSE-Tuned vs. Annual Change-Tuned Models
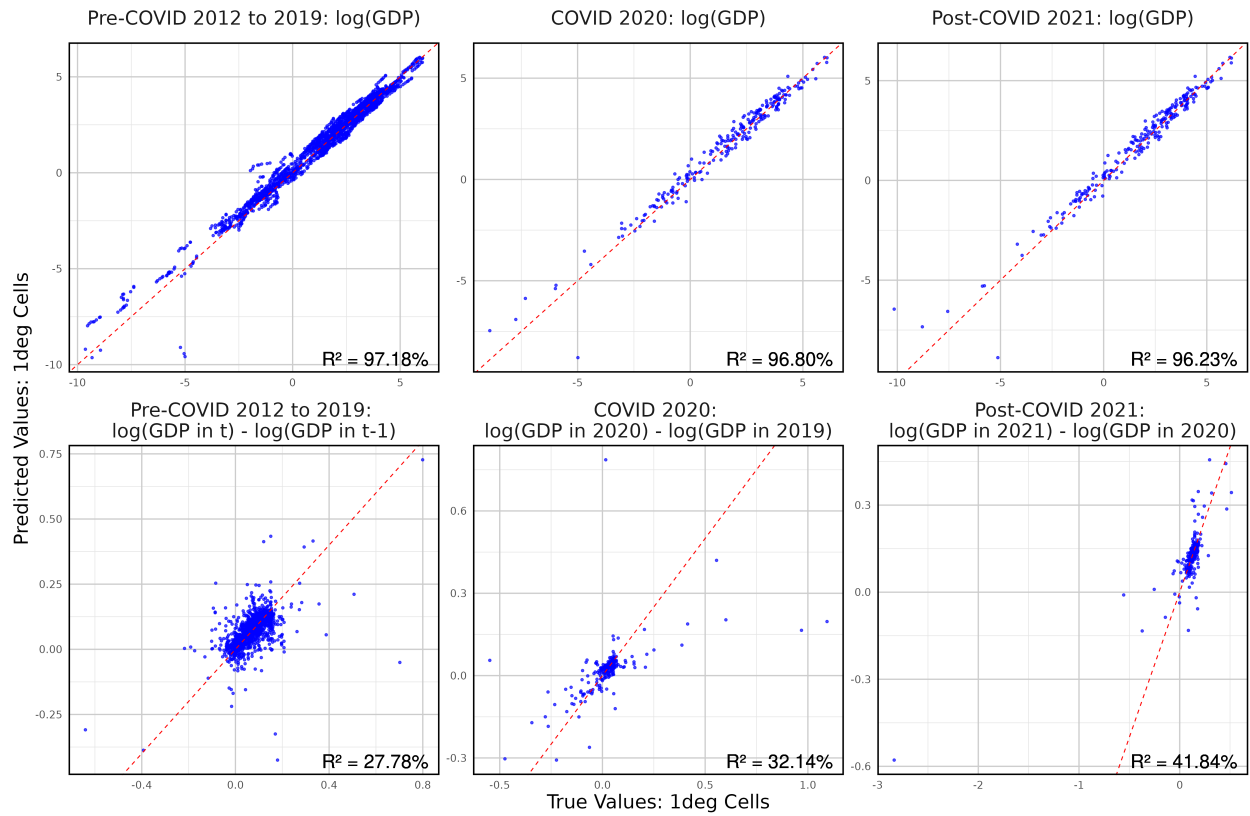
Figure 9: Model Predictions Against Actual GDP Values for China's Seven Leading Provinces, Using the Model Trained on year 2012 to 2021 Data and Tuned for Minimizing MSE

Note: This plot is comparable with the same plot in the paper. The only difference is the way hyperparameters are tuned.
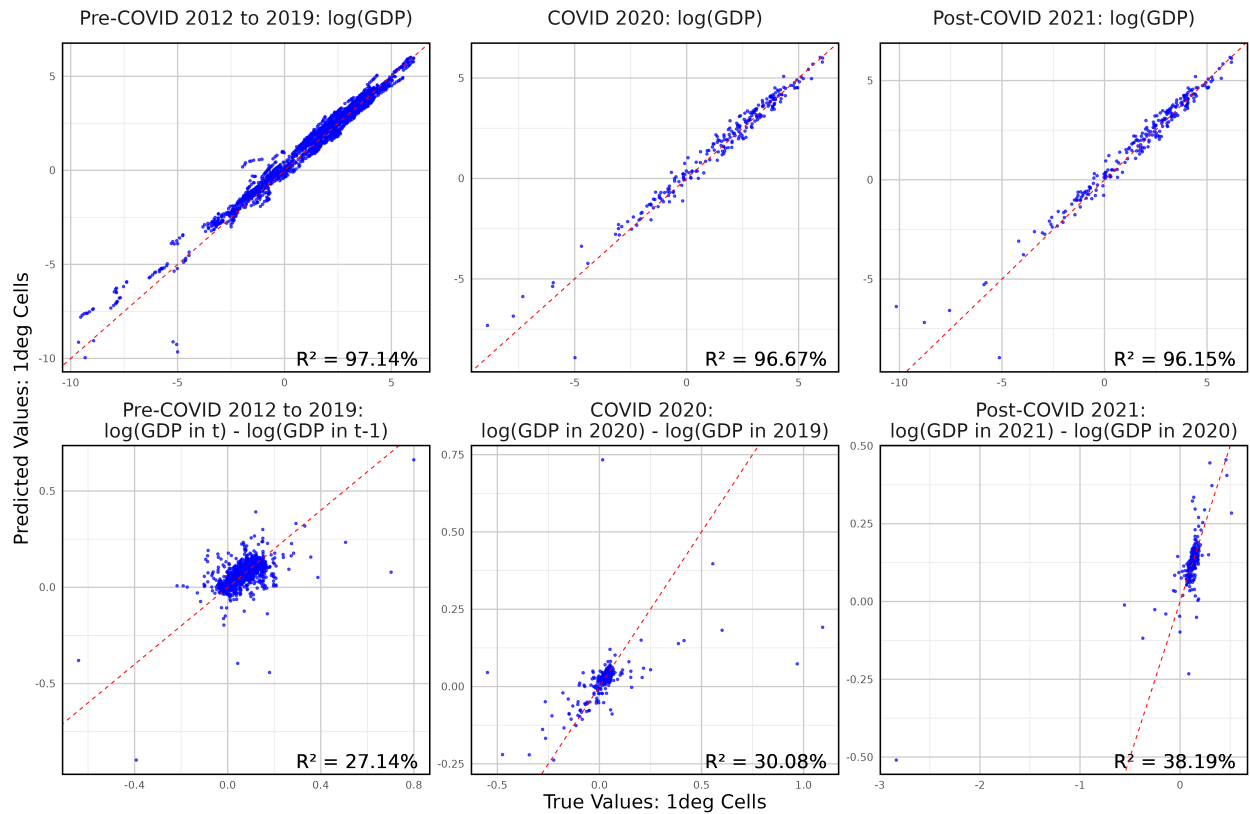
Figure 10: Model Predictions Against Actual GDP Values for China's Seven Leading Provinces, Using the Model Trained on year 2012 to 2019 Data and Tuned for Minimizing MSE

Note: This plot is comparable with 4. The only difference is the way hyperparameters are tuned.
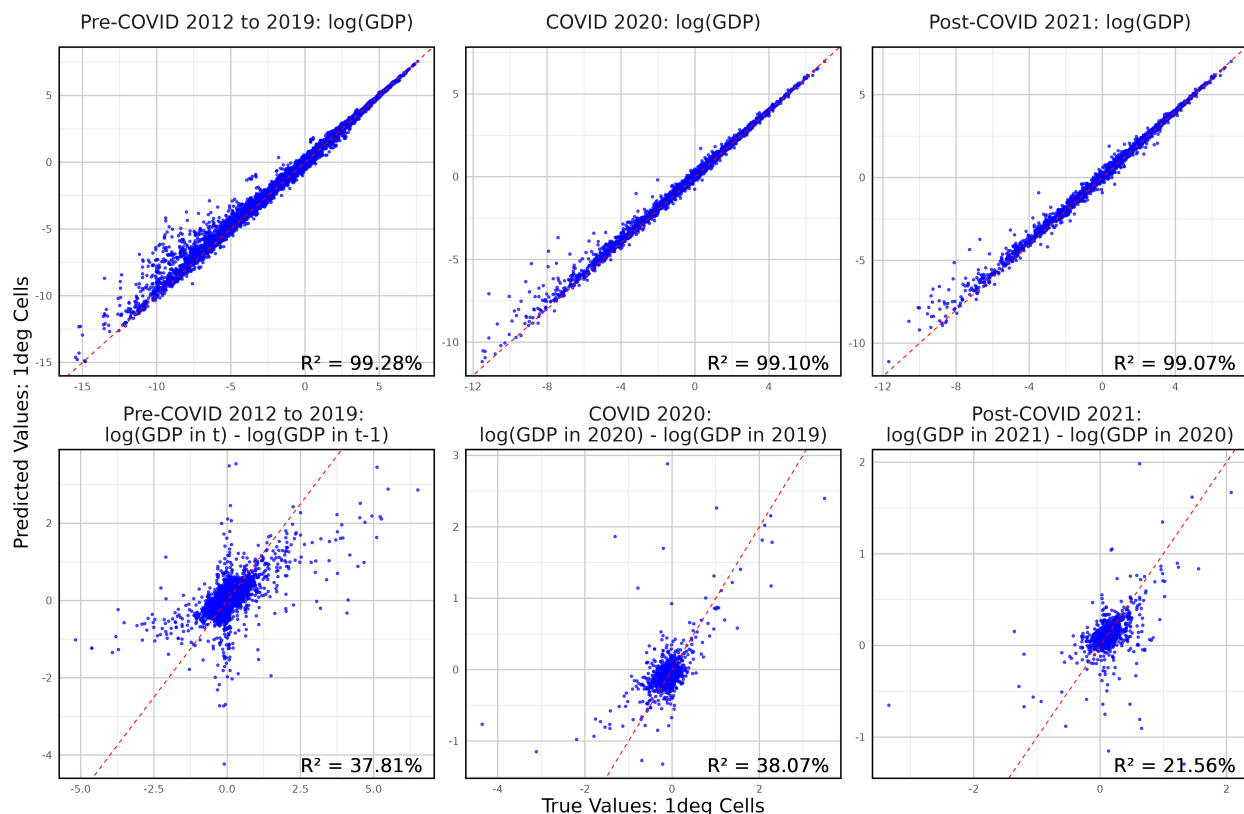
Figure 11: Model Predictions Against Actual GDP Values for All Training Countries, Using a Model Trained on year 2012 to 2019 Data and Tuned for Minimizing MSE

Note: Data for 2012 to 2019 are within training sample, so use out-of-bag predictions. Data for year 2020 and 2021 are predictions from the model. This plot is comparable with 5. The only difference is the way hyperparameters are tuned.

6.2 Compare Benchmark Models with Models Trained Without Weights

The models used in the paper are trained with weights to address the imbalance between cells from developed and developing countries in the training sample. The weights are the rescaling factors of cell shares from developed and developing countries in the training sample, adjusted to match their real-world proportions. Cells can only be assigned one of the two weights: one assigned to cells from developed countries and the other to cells from developing countries, depending on their classification. These weights influence the probability of each cell being selected in the bootstrap sample used to build each decision tree. By assigning higher probabilities to underrepresented groups, the model ensures that both developed and developing regions are adequately represented during training.

Here we compare the predictions from models trained with and without weights. The table 24 and figure 12 illustrate that the differences between the two approaches are minimal and the weights have a limited impact on the overall predictive performance of the methodology.

We also performed the same tests on the model trained without weights for comparison with the benchmark model (trained with weights), as shown in Figures 13, 14, and 15. The results continue to show the limited impact of weights on predictions.

Table 24: Cross-Validated Performance Metrics Across Spatial Resolutions for Models Trained Without Weights

	1-degree Model	0.5-degree Model	0.25-degree Model
Panel A: R^2 of Log GDP Level			
R^2 (Developed)	98.05%	97.8%	97.58%
R^2 (Developing)	96.1%	94.27%	92.97%
R^2 (All)	97.63%	97.16%	96.89%
Weighted R^2	96.75%	95.44%	94.51%
Panel B: R^2 of log(GDP in t) - log(GDP in t-1)			
R^2 (Developed)	67.79%	77.76%	84.49%
R^2 (Developing)	61.76%	62.38%	66.58%
R^2 (All)	64.13%	73.05%	79.81%
Weighted R^2	63.76%	67.5%	72.53%
Panel C: Variables and Importance Scores			
NTL from urban	20.66	12.17	6.96
Lag NTL from urban	13.25	7.35	6.51
Population	83.39	52.82	27.6
Lag population	49.97	35.6	18.52
Urban areas	2.55	2.83	2.39
Lag urban areas	2.52	1.87	1.68
Fossil CO2 from transportation	0.3	0.4	0.25
Lag fossil CO2 from transportation	0.44	0.19	0.54
Biofuel CO2 from manufacturing	0.31	0.3	0.29
Lag biofuel CO2 from manufacturing	0.36	0.3	0.34
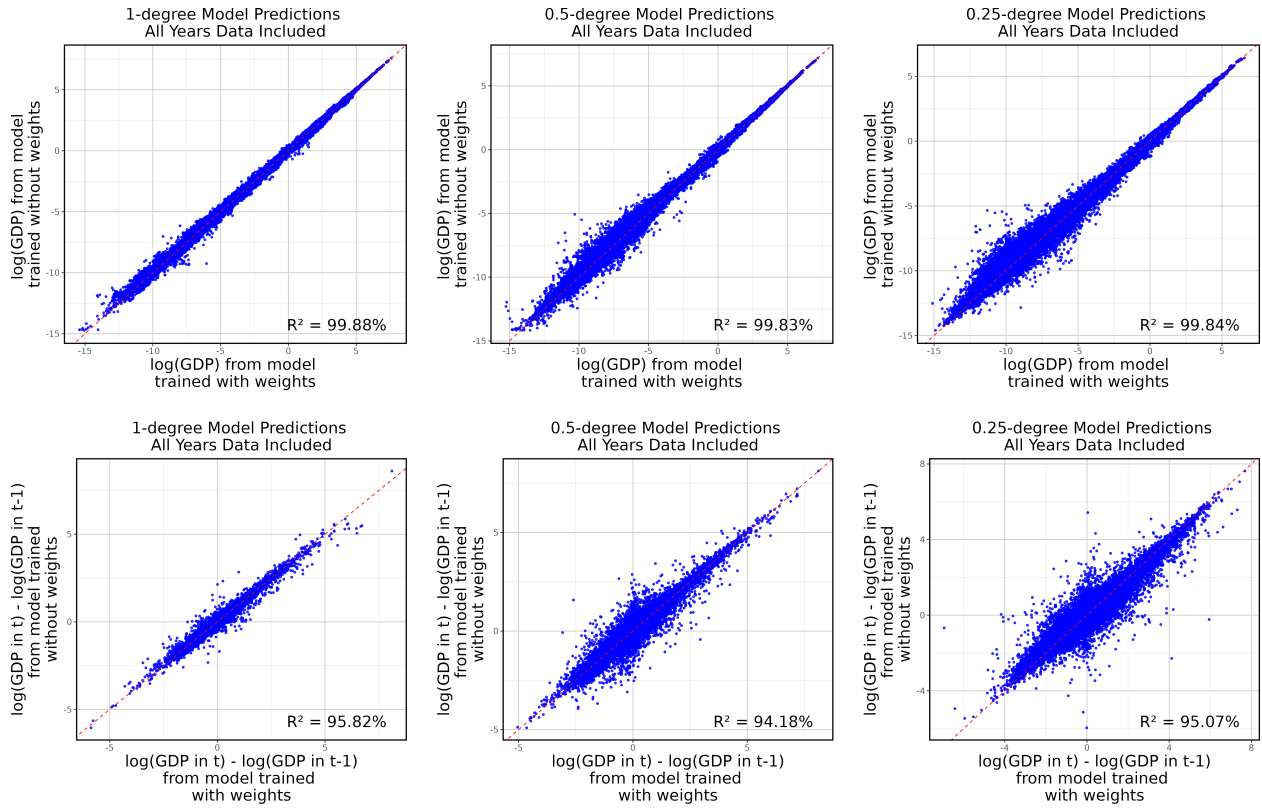
Figure 12: Comparison of Model Predictions: Models Trained With and Without Weights
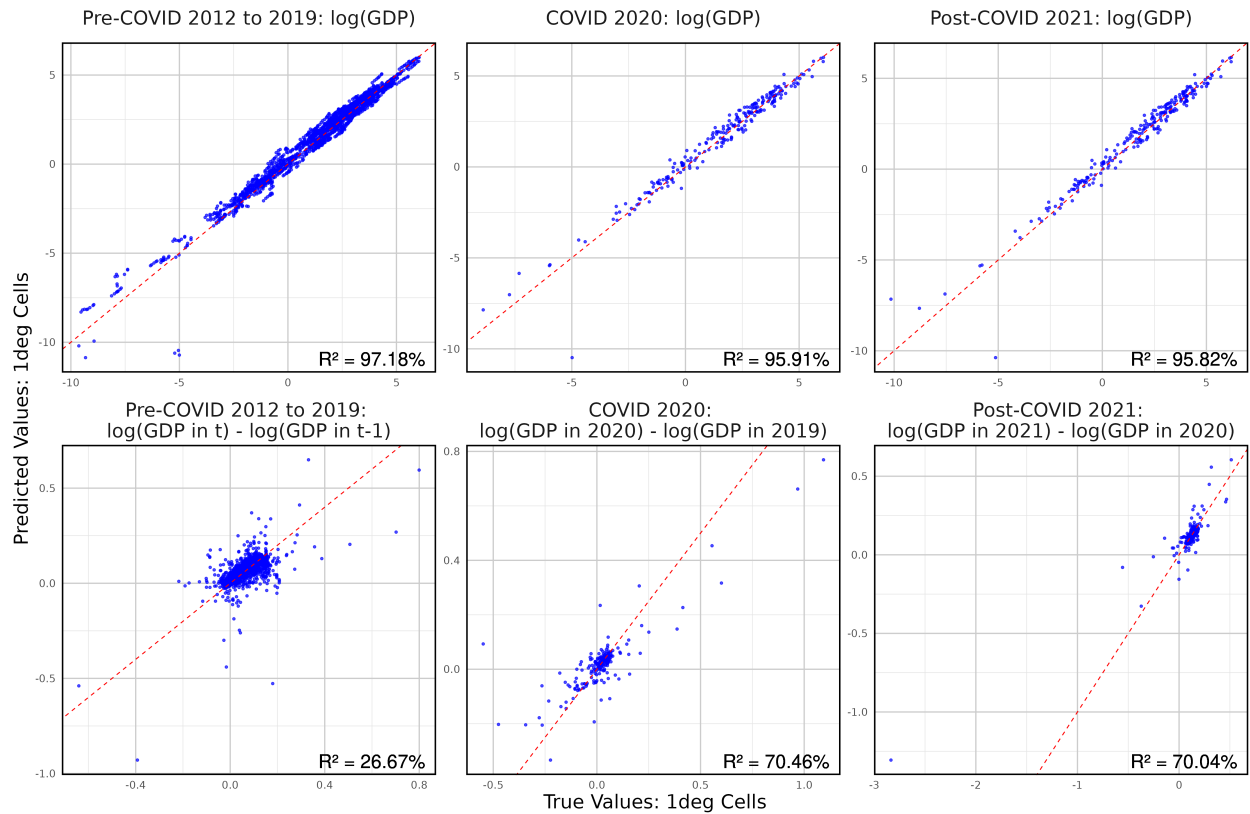
Figure 13: Model Predictions Against Actual GDP Values for China's Seven Leading Provinces, Using the Model Trained on year 2012 to 2021 Data and Without Weights

Note: This plot is comparable to the corresponding plot in the paper, with the only difference being whether the models were trained with or without weights.
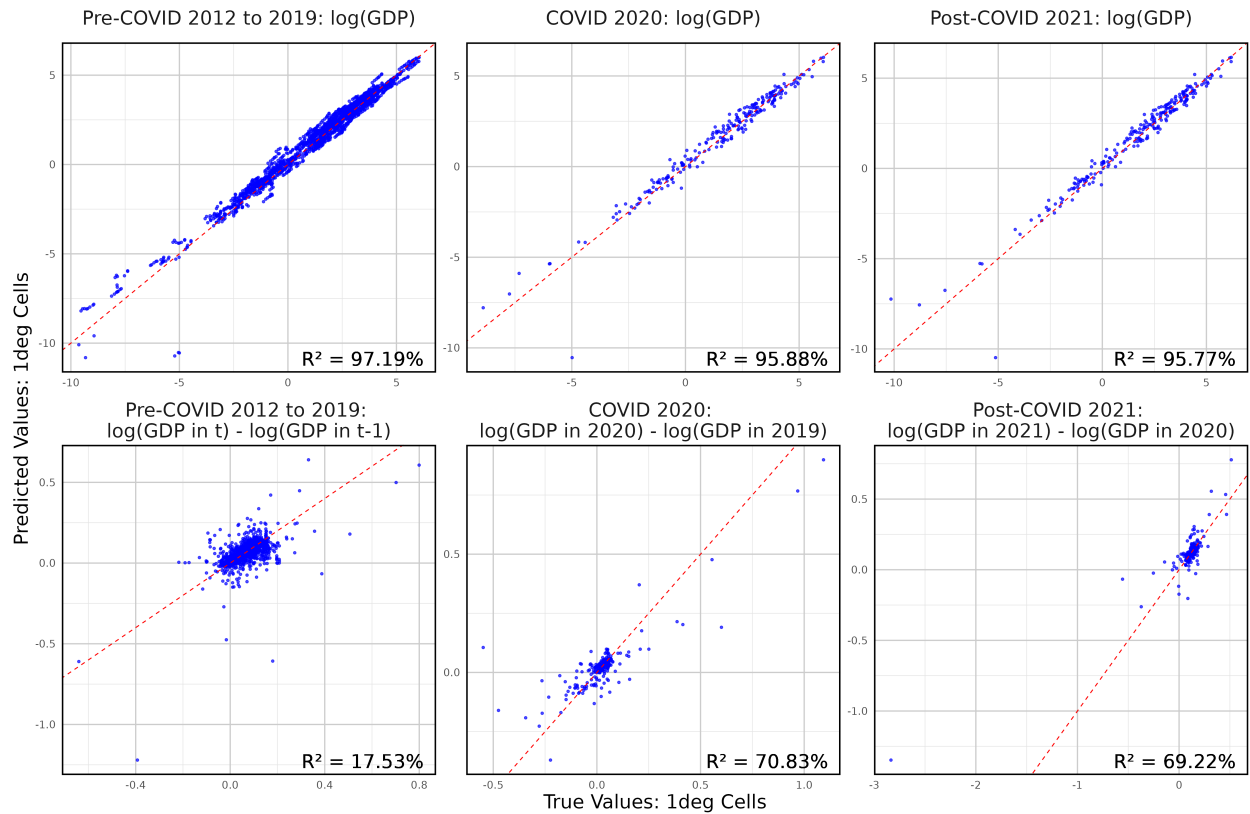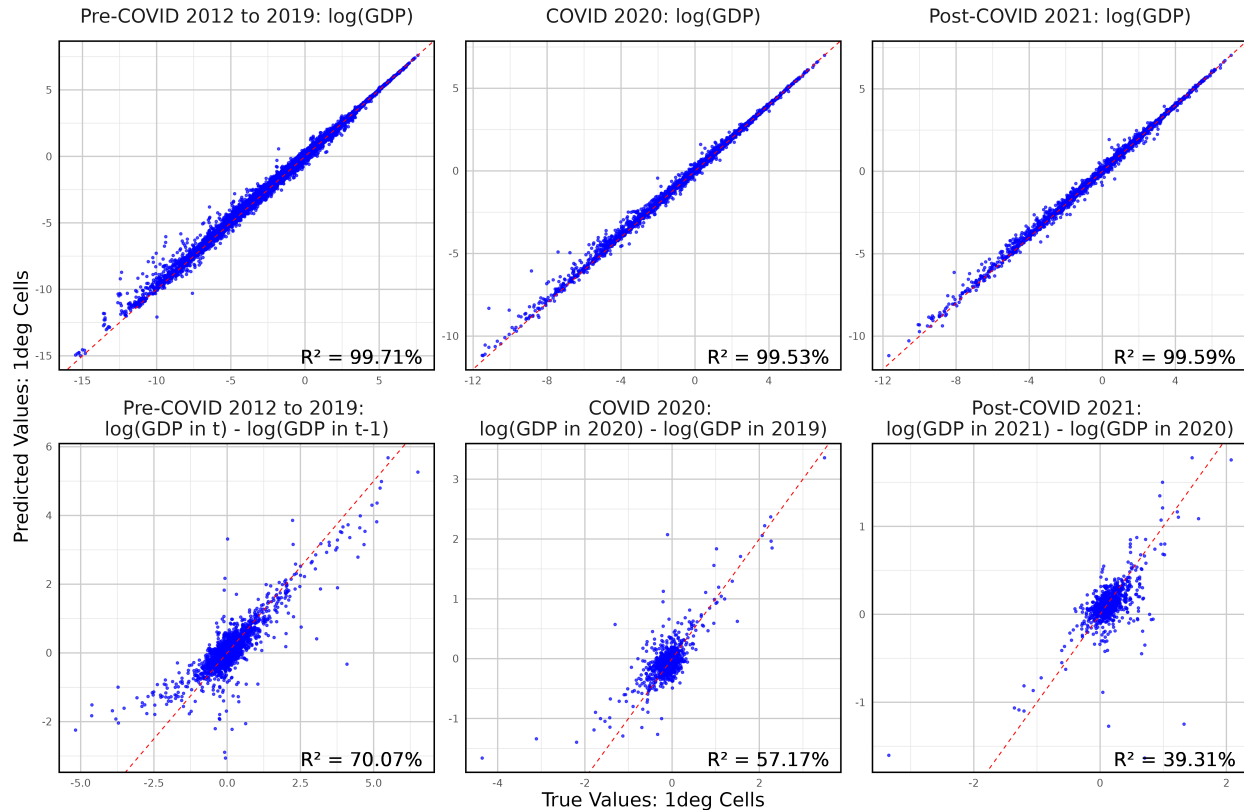
Figure 14: Model Predictions Against Actual GDP Values for China's Seven Leading Provinces, Using the Model Trained on year 2012 to 2019 Data and Without Weights

Note: This plot is comparable with 4. The only difference lies in whether the models were trained with or without weights.

Figure 15: Model Predictions Against Actual GDP Values for All Training Countries, Using the Model Trained on year 2012 to 2019 Data and Without Weights

Note: Data for 2012 to 2019 are within training sample, so use out-of-bag predictions. Data for year 2020 and 2021 are predictions from the model. This plot is comparable with 5. The only difference lies in whether the models were trained with or without weights.

6.3 Compare Benchmark Models with Models Trained Without Developing Countries Data

This section presents the results of models trained without data from developing countries. Note that weights are also excluded when developing country data is not included. Table 25 summarizes the cross-validated performance across all three resolutions. No large differences are observed. The models were then used to predict data for cells in developing countries, and the results were compared with the truth. Tables 26, 27, and 28 demonstrate that the models perform well in predicting developing country data, even without having been trained on it. When applied to predict all world cells, figure 16 reveals that level predictions are highly consistent, while slight differences are observed in annual changes.

We also performed the same tests on the model trained without developing countries for comparison with the benchmark model (trained with developing countries), as shown in Figures 17, 18, 19 and 20. The results indicate that excluding developing countries leads to slightly better performance in these tests. However, the improvement is not substantial enough to justify switching models.

47

Table 25: Cross-Validated Performance Metrics Across Spatial Resolutions for Models Trained Without Developing Countries Data

	1-degree Model	0.5-degree Model	0.25-degree Model
Panel A: R^2 of Log GDP Level			
R^2 (All)	97.7%	97.75%	97.74%
Panel B: R^2 of log(GDP in t) - log(GDP in t-1)			
R^2 (All)	56.86%	78.4%	86.39%
Panel C: Variables and Importance Scores			
NTL from urban	17.21	9.81	4.24
Lag NTL from urban	12.16	7.18	5.07
Population	63.97	48.7	27.63
Lag population	52.19	30.22	17.55
Urban areas	2.63	2.53	1.21
Lag urban areas	3.03	1.69	1.22
Fossil CO2 from transportation	0.3	0.29	0.2
Lag fossil CO2 from transportation	0.31	0.1	0.12
Biofuel CO2 from manufacturing	0.22	0.25	0.15
Lag biofuel CO2 from manufacturing	0.25	0.26	0.19

Table 26: 1-degree Model Performance Metrics for Developing Group

ISO	R^2 for log(GDP)	R^2 for log(GDP in t) - log(GDP in t-1)
ALB	99.02%	97.99%
BIH	97.82%	89.27%
BLR	98.48%	96.67%
CHL	96.37%	96.04%
COL	95.53%	97.44%
ECU	94.63%	91.88%
IDN	95.60%	93.87%
KEN	87.87%	89.95%
KGZ	96.41%	96.85%
LKA	97.97%	87.81%
MOZ	94.34%	94.00%
PER	97.08%	97.96%
PHL	97.95%	99.03%
SRB	97.21%	97.00%
THA	87.60%	92.13%
UZB	98.56%	97.68%
VNM	90.92%	91.83%

Table 27: 0.5-degree Model Performance Metrics for Developing Group

ISO	R^2 for log(GDP)	R^2 for log(GDP in t) - log(GDP in t-1)
ALB	98.55%	98.53%
BIH	95.09%	94.39%
BLR	97.98%	97.83%
CHL	94.14%	94.06%
COL	96.61%	97.16%
ECU	87.95%	87.19%
IDN	95.10%	94.82%
KEN	87.01%	88.35%
KGZ	95.62%	96.40%
LKA	98.84%	98.09%
MOZ	91.09%	91.22%
PER	97.12%	97.34%
PHL	97.20%	98.20%
SRB	80.98%	80.15%
THA	88.23%	90.41%
UZB	97.26%	96.85%
VNM	93.30%	93.77%

Table 28: 0.5-degree Model Performance Metrics for Developing Group

ISO	R^2 for log(GDP)	R^2 for log(GDP in t) - log(GDP in t-1)
ALB	97.50%	97.98%
BIH	91.02%	90.98%
BLR	96.27%	95.64%
CHL	94.89%	90.77%
COL	96.23%	94.14%
ECU	89.69%	88.62%
IDN	94.60%	94.34%
KEN	91.57%	92.45%
KGZ	95.90%	93.50%
LKA	98.62%	98.26%
MOZ	88.84%	89.02%
PER	96.72%	96.98%
PHL	97.09%	93.82%
SRB	83.45%	83.46%
THA	86.52%	88.37%
UZB	96.78%	96.57%
VNM	91.86%	92.47%
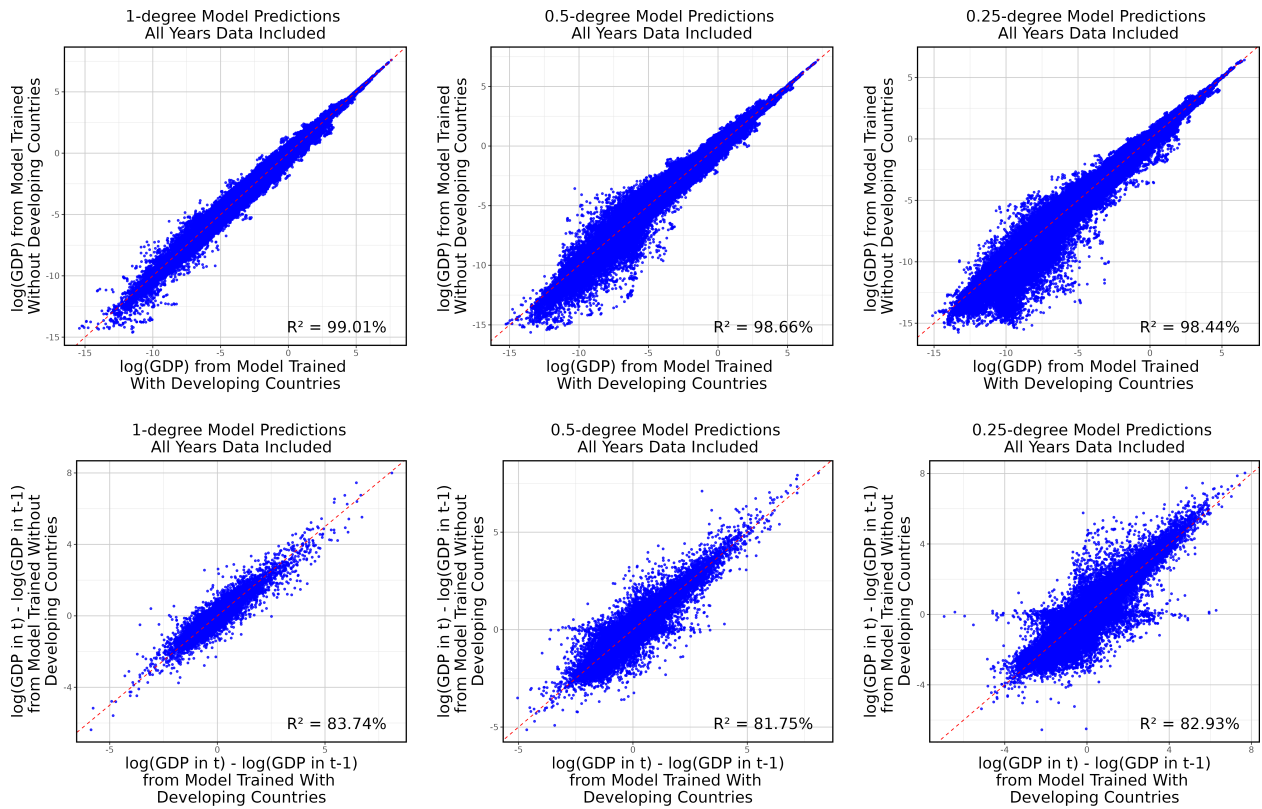
Figure 16: Comparison of Model Predictions: Models Trained With and Without Developing Countries Data
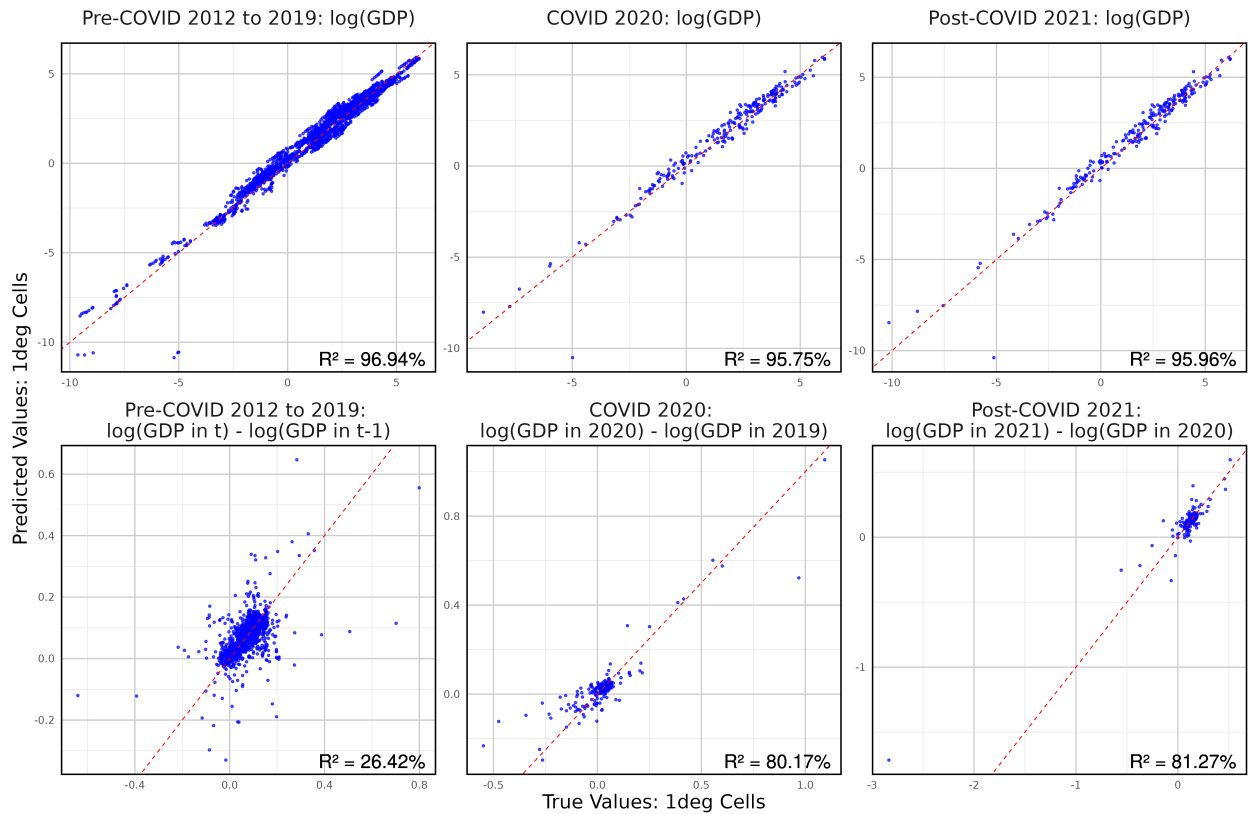
Figure 17: Model Predictions Against Actual GDP Values for China's Seven Leading Provinces, Using a Model Trained on year 2012 to 2021 Data and Without developing countries data

Note: This plot is comparable to the corresponding plot in the paper, with the only difference being whether the models were trained with or without developing countries data.
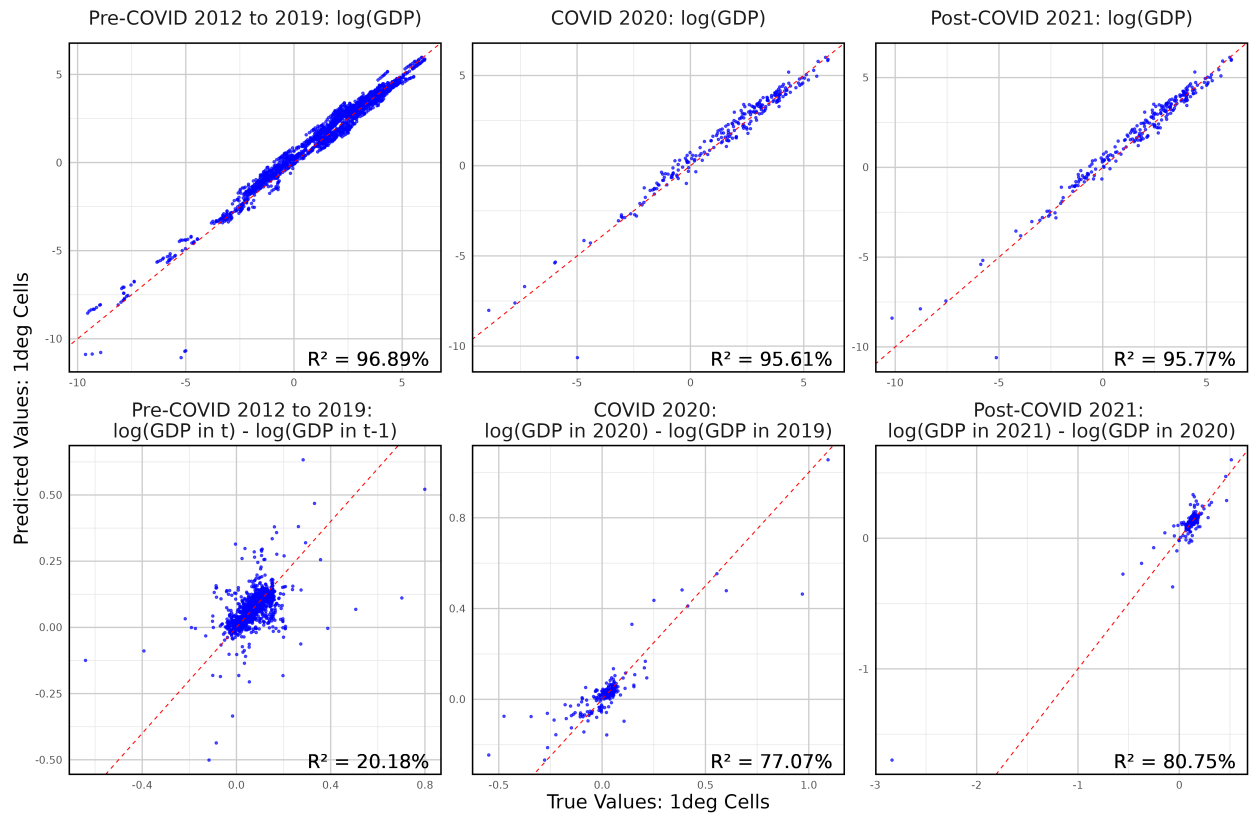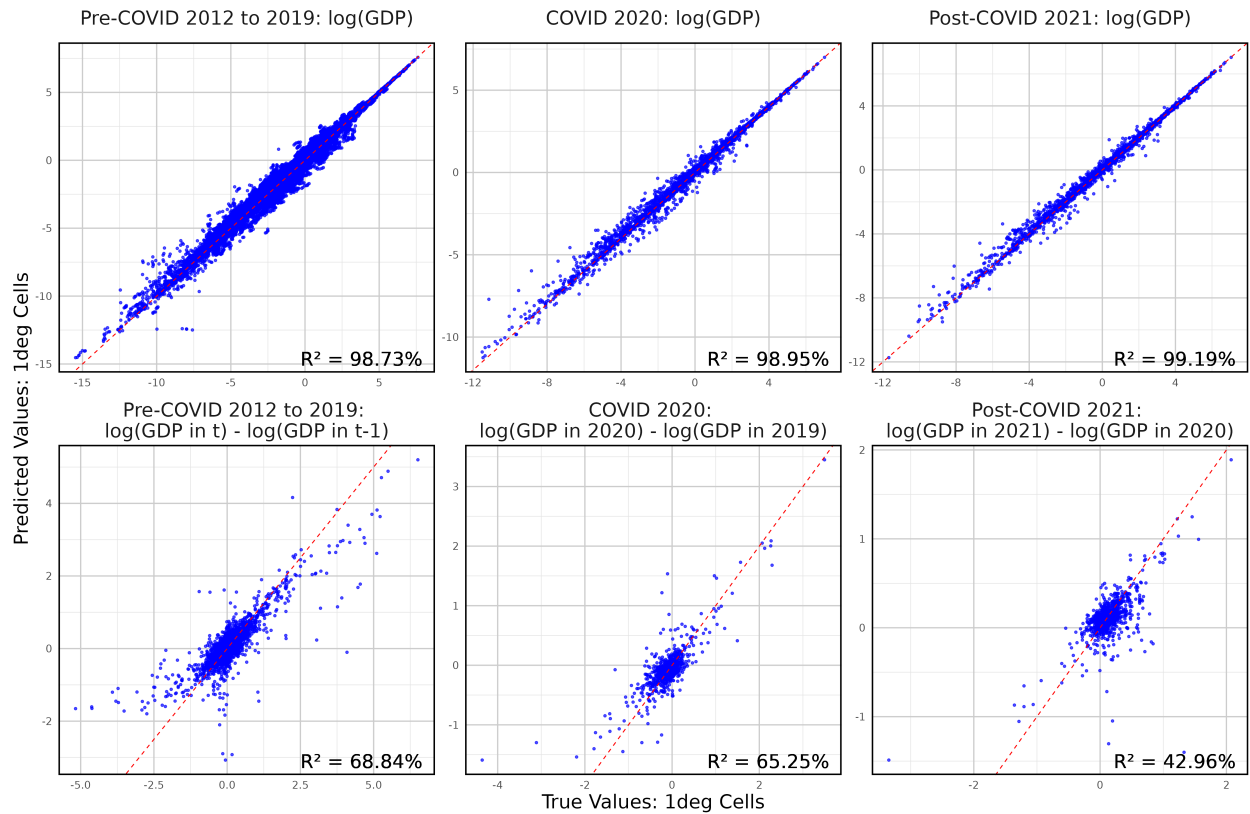
Figure 18: Model Predictions Against Actual GDP Values for China's Seven Leading Provinces, Using a Model Trained on year 2012 to 2019 Data and Without developing countries data

Note: This plot is comparable with 4. The only difference lies in whether the models were trained with or without developing countries data.

Figure 19: Model Predictions Against Actual GDP Values for All Training Countries, Using a Model Trained on year 2012 to 2019 Data and Without developing countries data

Note: Data for 2012 to 2019 are within training sample, so use out-of-bag predictions. Data for year 2020 and 2021 are predictions from the model. This plot is comparable with 5. The only difference lies in whether the models were trained with or without developing countries data.
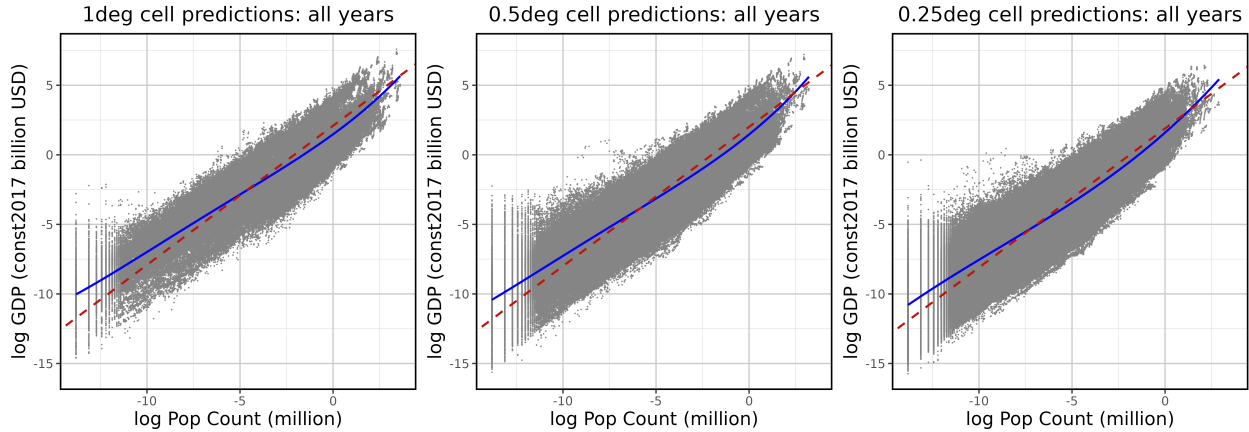
Figure 20: Cell Population Against Cell GDP, Using Models Trained Without developing countries data

Note: For the 1-degree resolution, the blue solid line is $y = 1.47 + 0.97x + 0.04x^2 + 0.004x^3 + 0.0001x^4$ and the red dashed line is $y = x + 2.09$. For the 0.5-degree resolution, the blue solid line is $y = 1.46 + 1.08x + 0.05x^2 + 0.004x^3 + 0.0001x^4$ and the red dashed line is $y = x + 2$. For the 0.25-degree resolution, the blue solid line is $y = 1.59 + 1.18x + 0.04x^2 + 0.002x^3 + 0x^4$ and the red dashed line is $y = x + 1.89$.

7   Consistency of Predictions Across Resolutions

In this section, we demonstrate that the predictions from models at different spatial resolutions are consistent. Figure 21 compares the predictions generated by our benchmark models in the paper: trained using data from all available countries (excluding China) from 2012 to 2021 with optimized hyperparameters for each resolution. To evaluate consistency, we aggregate the predictions from finer resolutions to match the coarser resolution and compare the results. Figure 21 shows that the predictions remain relatively consistent and confirms the robustness of our models across different spatial scales.
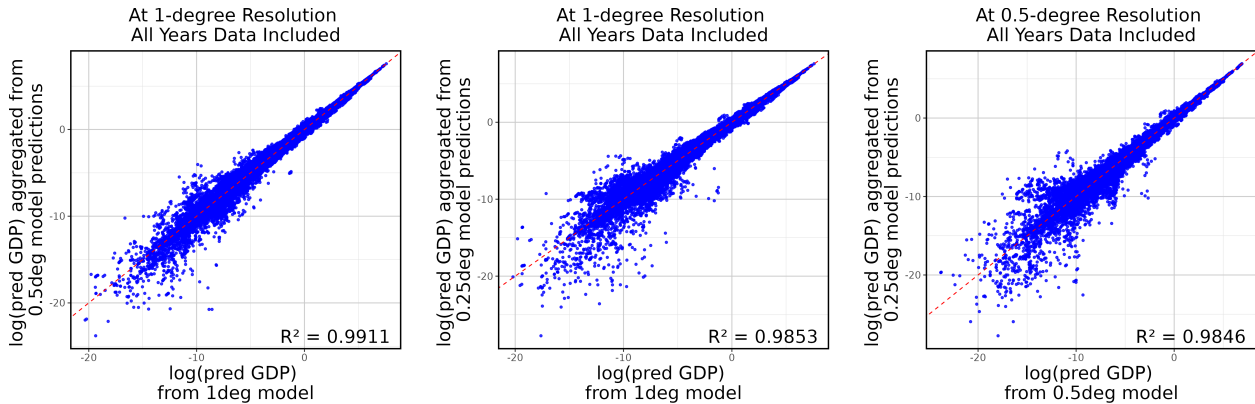

Figure 21: Comparison of Model Predictions Across Resolutions: Aggregated vs. Direct Predictions

8   Consistency of Predictions Across Dataset Versions

Our datasets are updated annually, and this process involves incorporating new training samples. As a result, predictions may change, not only for the new years but also for the previously published years. This highlights the importance of testing whether our models produce consistent predictions across different dataset versions. To conduct this test, we trained two versions of the model: 1) uses data from all available countries (excluding China) for the years 2012 to 2019 2) uses data from the same countries but extends the coverage to include years 2012 to 2021. We compared the predictions for both the newly updated years (2020 and 2021) and selected previous years (2018 and 2019). Figures 22 and 23 illustrate that, despite minor discrepancies, the models produce highly similar results across versions for both new and old years.
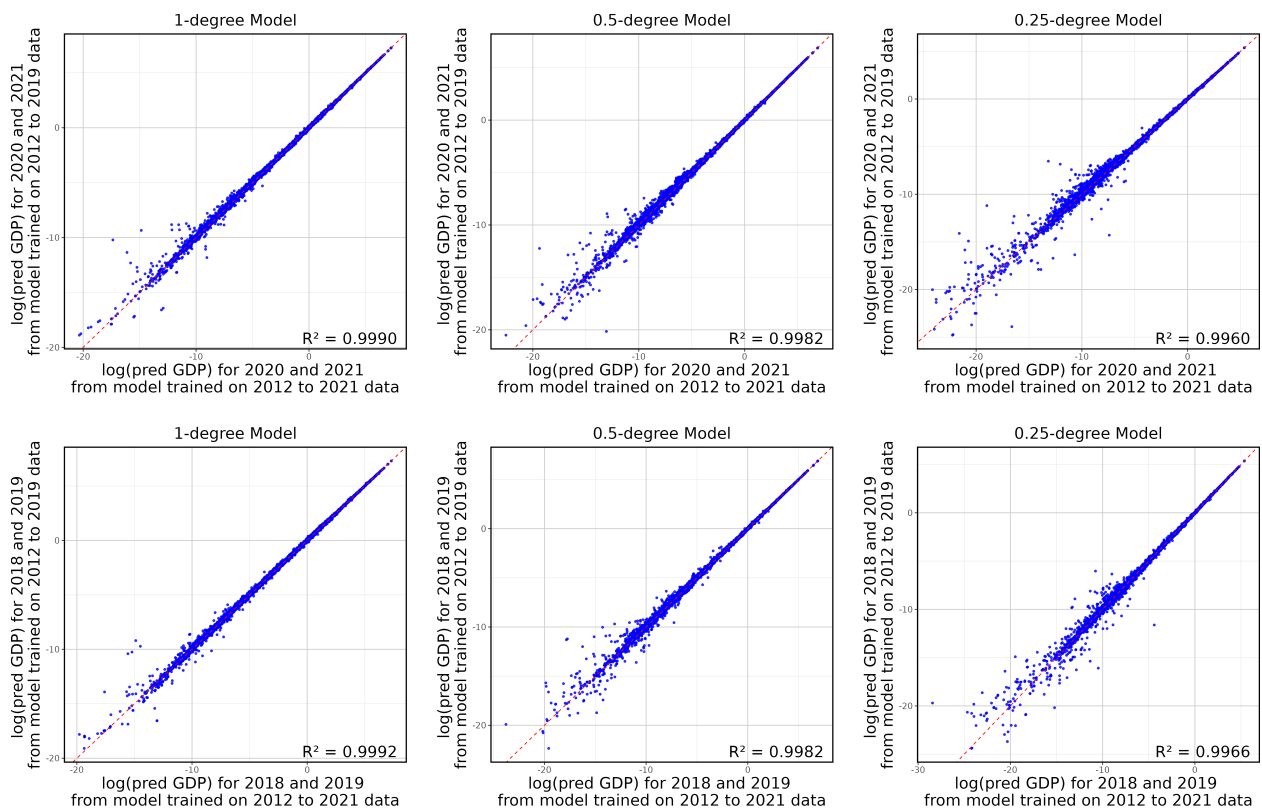

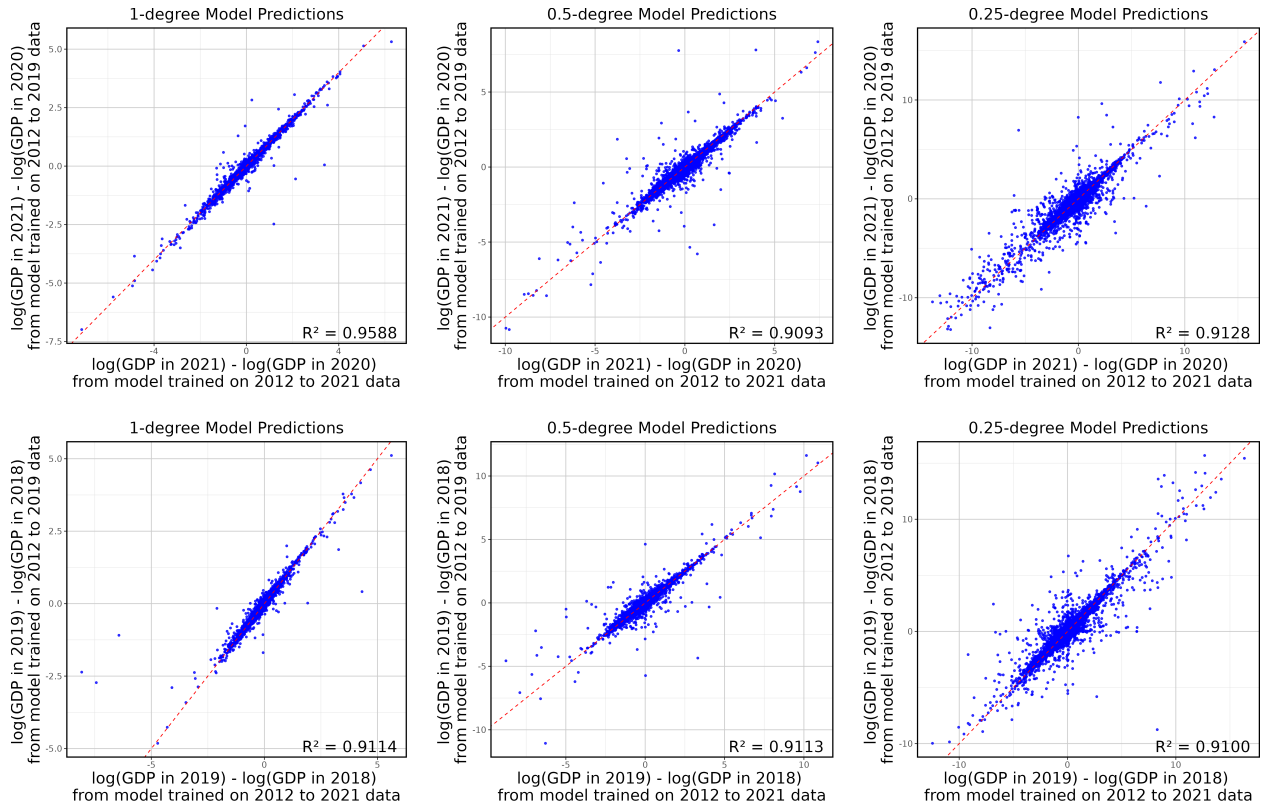Figure 22: Comparison of Predicted Cell GDP Levels Across Model Versions

Figure 23: Comparison of Predicted Cell GDP Annual Changes Across Model Versions

References

Authority, Philippine Statistics (2023). *2000-2023 Gross Regional Domestic Product - GRDP by Region*. GRDP Data Series (Database). Accessed on 15th June 2024. URL: https://psa.gov.ph/statistics/grdp/data-series.

Bank, World (2024a). *GDP in Constant 2021 PPP Adjusted International$ (Dataset)*. The World Bank Group. Accessed on 15th June 2024. URL: https://data.worldbank.org/indicator/NY.GDP.MKTP.PP.KD.

— (2024b). *GDP in Current PPP Adjusted International$ (Dataset)*. The World Bank Group. Accessed on 15th June 2024. URL: https://data.worldbank.org/indicator/NY.GDP.MKTP.PP.CD.

— (2024c). *GDP in Current US$ (Dataset)*. The World Bank Group. Accessed on 15th June 2024. URL: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD.

— (2024d). *Population (Dataset)*. The World Bank Group. Accessed on 15th June 2024. URL: https://data.worldbank.org/indicator/SP.POP.TOTL.

Bright, Edward, Patricia Coleman, Andrew King, Amy Rose, Michael Urban, Jarrod McKee, Kelsey Sims, and Andrew Reith (n.d.). *LandScan Global*. Oak Ridge National Laboratory. Accessed: 2023-12-20. URL: https://doi.org/10.48690/1524196,%20https://doi.org/10.48690/1524197,%20https://doi.org/10.48690/1524198,%20https://doi.org/10.48690/1524199,%20https://doi.org/10.48690/1524200,%20https://doi.org/10.48690/1524201,%20https://doi.org/10.48690/1524202,%20https://doi.org/10.48690/1524203,%20https://doi.org/10.48690/1524204,%20https://doi.org/10.48690/1524205,%20https://doi.org/10.48690/1524206,%20https://doi.org/10.48690/1524207,%20https://doi.org/10.48690/1524215,%20https://doi.org/10.48690/1524208,%20https://doi.org/10.48690/1524209,%20https://doi.org/10.48690/1524210,%20https://doi.org/10.48690/1524211,%20https://doi.org/10.48690/1524212,%20https://doi.org/10.48690/1524213,%20https://doi.org/10.48690/1524214,%20https://doi.org/10.48690/1523378,%20https://doi.org/10.48690/1527702.

Bureau, U.S. Census (2024). *Annual Index Value and Annual Percent Change in Price Series: 1947 to 2022*. U.S. Census Bureau Database. Accessed on 15th June 2024. URL: https://www.census.gov/en.html.

Economic Analysis U.S. Department of Commerce, Bureau of (2023). *CAGDP2: GDP in Current Dollars by County and MSA*. Regional Economic Accounts - Previously Published Estimates: Gross Domestic Product by County and Metro Area. Accessed on 15th June 2024. URL: https://apps.bea.gov/histdatacore/Regional_Accounts_new.html.

Elvidge, Christopher D, Maxim Zhizhin, Kimberly Baugh, Feng-Chi Hsu, and Tilottama Ghosh (2016). "Methods for Global Survey of Natural Gas Flaring from Visible Infrared Imaging Radiometer Suite Data". In: *Energies* 9.1, p. 14. DOI: 10.3390/en9010014.

Elvidge, Christopher D, Maxim Zhizhin, Feng-Chi Hsu, and Kimberly E Baugh (2013). "VIIRS Nightfire: Satellite Pyrometry at Night". In: *Remote Sensing* 5, pp. 4423–4449. DOI: 10.3390/rs5094423.

Friedl, Mark and Damien Sulla-Menashe (2022). *MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V061*. Distributed by NASA EOSDIS Land Processes

Distributed Active Archive Center. Accessed 2023-12-20. URL: https://doi.org/10.5067/MODIS/MCD12Q1.061.

Fund, International Monetary (2024). *World Economic Outlook Database, April 2024*. IMF World Economic Outlook Database. Accessed on 15th June 2024. URL: https://www.imf.org/en/Publications/SPROLLS/world-economic-outlook-databases#sort=%40imfdate%20descending.

GaryBikini (2024). *China Administrative Division Shapefiles*. Version 24.02.06. DOI: 10.5281/zenodo.10624971. URL: https://doi.org/10.5281/zenodo.10624971.

Geografia e Estatística (IBGE), Instituto Brasileiro de (2024). *GDP under the point of view of Income (2010-2021)*. SCR - System of Regional Accounts (Database). Accessed on 15th June 2024. URL: https://www.ibge.gov.br/en/statistics/economic/national-accounts/16855-regional-accounts-of-brazil.html.

India, Reserve Bank of (2023). *Table 27: Gross State Domestic Product (Current Prices)*. Handbook of Statistics of Indian States. Accessed on 15th June 2024. URL: https://www.rbi.org.in/Scripts/AnnualPublications.aspx?head=Handbook%20of%20Statistics%20on%20Indian%20States.

Kyrgyz Republic, National Statistical Committee of the (2023). *1.01.00.09 Gross regional product (GRP) at current prices*. National accounts (Database). Accessed on 15th June 2024. URL: https://www.stat.kg/en/statistics/nacionalnye-scheta/.

Lehner, Bernhard and Petra Döll (2004). "Development and validation of a global database of lakes, reservoirs and wetlands". In: *Journal of Hydrology* 296.1-4, pp. 1–22.

National Statistics Agency for Strategic Planning, Bureau of and Reforms of the Republic of Kazakhstan (2023). *Gross regional product*. Statistics of national accounts (Database). Accessed on 15th June 2024. URL: https://stat.gov.kz/en/industries/economy/national-accounts/.

Nunn, Nathan and Diego Puga (2012). "Ruggedness: The Blessing of Bad Geography in Africa". In: *Review of Economics and Statistics* 94.1, pp. 20–36.

OECD (2024a). *Gross domestic product - Regions. OECD Regional Economy (Database)*. Accessed on 15th June 2024. URL: https://data-explorer.oecd.org/vis?fs%5B0%5D=Topic%2C1%7CRegional%2C%20rural%20and%20urban%20development%23GEO%23%7CRegions%23GEO_REG%23&fs%5B1%5D=Topic%2C2%7CRegional%2C%20rural%20and%20urban%20development%23GEO%23%7CRegions%23GEO_REG%23%7CRegional%20economy%23GEO_REG_ECO%23&pg=0&fc=Topic&snb=39&df%5Bds%5D=dsDisseminateFinalDMZ&df%5Bid%5D=DSD_REG_ECO%40DF_GDP&df%5Bag%5D=OECD.CFE.EDS&df%5Bvs%5D=2.0&lo=5&lom=LASTNPERIODS&dq=A..AUS%2BAU1%2BAU2%2BAU3%2BAU4%2BAU5%2BAU6%2BAU7%2BAU8%2BAUT%2BAT11%2BAT111%2BAT112%2BAT113%2BAT12%2BAT121%2BAT122%2BAT123%2BAT124%2BAT125%2BAT126%2BAT127%2BAT13%2BAT130%2BAT21%2BAT211%2BAT212%2BAT213%2BAT22%2BAT221%2BAT222%2BAT223%2BAT224%2BAT225%2BAT226%2BAT31%2BAT311%2BAT312%2BAT313%2BAT314%2BAT315%2BAT32%2BAT321%2BAT322%2BAT323%2BAT33%2BAT331%2BAT332%2BAT333%2BAT334%2BAT335%2BAT34%2BAT341%2BAT342..GDP..Q.USD_PPP_PS&ly%5Brw%5D=REF_AREA%2CTERRITORIAL_LEVEL&ly%5Bcl%5D=TIME_PERIOD&to%5BTIME_PERIOD%5D=false.

— (2024b). *Regional Economy. OECD iLibrary - OECD Regional Statistics (Database)*. Accessed on 15th June 2024. URL: https://www.oecd-ilibrary.org/urban-rural-and-regional-development/data/oecd-regional-statistics_region-data-en.

Román, Miguel O, Zhuosen Wang, Qingsong Sun, Virginia Kalb, Steven D Miller, Andrew Molthan, Lyle Schultz, Jordan Bell, Erika C Stokes, Bhaskar Pandey, and Karen C Seto (2018). "NASA's Black Marble nighttime lights product suite". In: *Remote Sensing of Environment* 210, pp. 113–143. DOI: 10.1016/j.rse.2018.03.017.

Runfola, Daniel, Dylan Anderson, Daniel Mahr, Will R. Turner, and Martin Wegmann (2020). "geoBoundaries: A global database of political administrative boundaries". In: *PLoS ONE* 15.4, e0231866. DOI: 10.1371/journal.pone.0231866. URL: https://doi.org/10.1371/journal.pone.0231866.

Running, Steven and Maosheng Zhao (2021). *MODIS/Terra Net Primary Production Gap-Filled Yearly L4 Global 500m SIN Grid V061.* Distributed by NASA EOSDIS Land Processes Distributed Active Archive Center. Accessed 2023-12-20. URL: https://doi.org/10.5067/MODIS/MOD17A3HGF.061.

Russian Statistical Yearbook (2022). Federal State Statistics Service (Rosstat). Accessed on 15th June 2024.

Russian Statistical Yearbook (2023). Federal State Statistics Service (Rosstat). Accessed on 15th June 2024.

Statistics of China, National Bureau of (2024). *Annual by Province - Gross.* Regional (Database). Accessed on 15th June 2024. URL: https://data.stats.gov.cn/english/index.htm.

UNdata (2024). *Per Capita GDP at Current Prices - US Dollars.* UNdata Database. Accessed on 15th June 2024. URL: https://data.un.org/Data.aspx?q=gdp&d=SNAAMA&f=grID%3a101%3bcurrID%3aUSD%3bpcFlag%3a1#SNAAMA.

Wenz, Leonie, Robert Devon Carr, Noah Kögel, Maximilian Kotz, and Matthias Kalkuhl (2023). "DOSE–Global data set of reported sub-national economic output". In: *Scientific Data* 10.1, p. 425.

Zhizhin, Maxim, Andrey Matveev, Tilottama Ghosh, Feng-Chi Hsu, Mark Howells, and Christopher Elvidge (2021). "Measuring Gas Flaring in Russia with Multispectral VIIRS Nightfire". In: *Remote Sensing* 13.16, p. 3078. DOI: 10.3390/rs13163078.