Use of VIIRS Nighttime Light Data (NTL) for prediction of GDP and Energy Consumption

Author: Ayesha Siddiqa, Amna Ahmed

Abstract:

This study focuses on the use of VIIRS Nighttime Light to predict the GDP and energy consumption. It analyzes the satellite light intensity values from 2013 to 2023 that include median, median masked, minimum and maximum. We apply machine learning models to find the correlation between them. Our results highlights the potential of NTL for finding the energy consumption and GDP values.

Introduction:

GDP and Energy Consumption are key indicator for any country economic development and life quality. Over the period of time their values have played a critical role for policymakers, organizations, government sector, researchers as it directly correlate to planning and execution of development and sustainability goals of a region. In traditional way, these indicators are measured thorough surveys and other traditional government methods, which seem to be inaccurate and often delayed.

In recent years, specially after 2012 , satellite have been used to gather night time light data which prove as a valuable asset for measuring human activity, economic conditions and energy usage. Studies done by researcher over the years have found a correlation between NTL and other important economic indicators.

This paper aims to use regression and time series models to estimate GDP and energy consumption of countries using band values of NTL data. This shows that NTL correlated positively with these indicators and can be used for forecasting. The main goal is to make use of available remote sensing data to monitor and predict these important factors in specially data scarce environment.

Literature Review:

Estimating GDP Growth Using VIIRS Night-Time Light Data [1] is a research paper that studies how viirs night time light data can be used in estimation of GDP especially in countries with limited statistics. The new thing done here was addition of synthetic data(noise) into official GDP data , and then combining it with NTL data, to study how satellite imagery can complement the economic record. The approach used in this paper is STL decomposition for the cleaning VIIRS light data.Creation of GDP data similar to low-quality official statistics finally regression models combining noisy GDP and Total Night-Time Light to evaluate. The dataset used is VIIRS VNP46A3 monthly composites (2012–2019) and World Bank’s World Development Indicators (22 countries) and the techniques used wre STL ,Linear regression with variable noise levels. The main finding of this paper were TNL improves GDP forecasting under high data uncertainty and smoothing techniques improved signal quality and model stability.VIIRS lights can help estimate economic activity in absence of statistics.

A Global Annual Simulated VIIRS Nighttime Light Dataset from 1992 to 2023 [2] study introduces the SVNL dataset, a global continuous night-time light dataset spanning 1992 to 2023. It aims to create VIIRS-like nightlight imagery for the time before VIIRS with help of machine learning, for long-term economic analysis across decades.Newly created annual VIIRS-equivalent NTL data for 1992–2023 by combining DMSP-OLS, VIIRS-DNB, and Landsat NDVI data and focused on generating high-resolution NTL imagery, not direct GDP estimation.Datasets used is DMSP-OLS (1992–2013): Historical nightlight data,VIIRS (2012–2023): Modern high-resolution imagery,Landsat NDVI: Vegetation index used as a helping input to improve cross-temporal consistency and global GDP data (used only for validation).Techniques mentioned are modified U-Net CNN for simulating high-resolution VIIRS-like NTL imagery,Cross-sensor calibration and normalization to align DMSP and VIIRS data and Regression and correlation analysis for validation of simulated outputs. The SVNL dataset provides visually and statistically consistent NTL data from 1992–2023.Cross-sensor calibration improves alignment which enables long-term, global economic analysis using light data where GDP data is sparse or unreliable.

Shedding Light on Development – Leveraging the New Nightlights Data to Measure Economic Progress [3] is about how harmonized VIIRS-DMSP night time light data can estimate subregional differences in wealth, GDP, and HDI in Sub-Saharan Africa. The main point is the additional support this article brings in estimation of various development factors, particularly where there are no official economic statistics or they are unreliable.Approach used is to map NTL values to regional GDP, HDI and wealth indices.Controlled for population density and included country-year fixed effects in regression models and finally compared model performance across urban and rural areas.Datasets used were VIIRS-DMSP night-time light dataset (2004–2019) and Wealth Index, GDP and HDI.Techniques was Ordinary Least Squares (OLS), 10-fold cross-validation for robustness. NTL data showed variation in wealth and development metrics and urban areas showed stronger correlations than rural regions.

Estimating Economic Activity in Urban South Sudan Using Satellite Data [4]
In this research paper remote sensing is used for estimation of economic indicators in region South Sudan. This paper shows how ML can be used in forcasting of these indicators by using night time light data and satellite data.Night time light data combined with economic indicator to form complete datastet and creation of engineered features eg rainfall thresholds, for better result finally applied regression and machine learning for prediction.Datasets used were VIIRS-DNB, NASA climate data, NDVI, CO_2 emissions and economic indicators, CPI, population growth, oil prices,etc.Techniques used were Linear regression with backward stepwise selection,LASSO and MRMR and Leave-one-out cross-validation.Acurate prediction of urban economic trends and credit and seasonal effects were key predictors.Feature engineering improved model performance.

A Global Annual Simulated VIIRS Nighttime Light Dataset from 1992 to 2023 [5]This paper presents a new dataset named SVNL which is formed by combining DMSP, VIIRS, and Landsat NDVI data using deep learning resulting in nightlight time series dataset from 1992 to 2023. The goal is to provide a longterm night light dataset for countries lacking historical satellite records, and evaluate it for GDP estimation in Egypt.Approach used was to train a modified U-Net model to generate simulated VIIRS lights across three decades.Merged SVNL with weather, population to predict GDP.Compared the accuracy of multiple ML models in GDP prediction.Datasets used were SVNL dataset from DMSP, VIIRS, and NDVI sources and Egypt provincial GDP.Techniques were Ridge Regression, Random Forest, SVM, ANN, XGBoost, KNN and Log transformation of inputs and 5-fold cross-validation.Ridge Regression achieved highest accuracy ($R^2 = 0.955$).SVNL provided consistent and reliable NTL estimates over time.Good for cross-decade economic prediction.

Exploring the Relation between NPP-VIIRS Nighttime Lights and Carbon Footprint, Population Growth, and Energy Consumption in the UAE [6]

This research elaborates the correlation between the Nighttime Lights and the three environmental indicators in UAE that include population Growth, electricity consumption and CO2 emissions. The main idea of this article is that the brighter the night light which means the higher human activity that indicates higher population, energy use and emissions.The dataset used in this research is imagery from 2012-2021 taken through the NASA VIIRS-DNB sensor.Further it includes the UAE population statistics and electricity records.The techniques used are to align with the ground-truth datasets' timescale, daily NTL data was filtered and aggregated.To investigate the relation between the three factors the linear regression was used and R^2 was calculated.City wise analysis was done.The study is specific to UAE.This lacks the atmospheric or lunar interference in the VIIRS data thus this could affect the accuracy.

Estimation Model and Spatio-Temporal Analysis of Carbon Emissions from Energy Consumption with NPP-VIIRS-like Nighttime Light Images [7]

The research focuses on the carbon emission model that correlates night light data with the actual energy consumption statistics to estimate the carbon emissions. This develops a regression model. The dataset used in this form the **NPP-VIIRS** satellite (Visible Infrared Imaging Radiometer Suite).This covers the China data form 2012 to 2019 of all its major provinces. Official data form the Chinese provincial yearbooks. Methodology includes the filtering the NTL images and then uses the Linear regression model and then the model calculates the emission based on NTL alone. Limitation includes the linear link between the NTL and emissions ,which may not be captured fully.

Research on Energy Consumption Carbon Emissions at Grid Scale Based on NPP-VIIRS Nighttime Light Data" [8]

This research focuses on the estimating carbon emissions from energy consumption at a fine spatial resolution using NTL. The dataset used id NPP-VIIRS light data at national/regional consumption and population stats. The techniques uses the regression analysis to link light intensity with emissions.This also uses the GIS-bases spatial analysis to map emissions.Limitations in this includes the Sensor saturation in bright urban areas can reduce accuracy and limited ground truth data for model validation.

Multi-Scale Dynamics and Spatial Consistency of Economy and Population Based on NPP/VIIRS Nighttime Light Data and Population Imagery: A Case Study of the Yangtze River Delta [9]

This research investigates the spatial and temporal relationship between economic activity and population distribution in the Yangtze River Delta (YRD) region of China. The Approach used in this indices to measure the geographic concentration and mismatch between economy and population. Dataset used in this is NPP/VIIRS nighttime light, LandScan Population and stats for GDP and resident population. Techniques used **are Geographic Concentration Index (L),** Inconsistency Index (I) and **Correlation analysis** for validation. Limitations includes Nighttime lights may miss economic activity from non-illuminated sectors.

Analysis of Spatiotemporal Changes in Energy Consumption Carbon Emissions at District and County Levels Based on Nighttime Light Data—A Case Study of Jiangsu Province in China [10]

This research developed a regression model using NPP-VIIRS Nighttime Light to estimate the carbon emission using the energy consumption data. It's analysis is conducted at district and county levels in Jiangsu Province, China from 2012–2021.It explores the temporal trends and spatial differences. Dataset used NPP-VIIRS Nighttime Light data (2012–2021) ,carbon emission and Socio-economic stats. Limitation includes that in Urban areas with strong lighting,sensor reaches a saturation level that underestimates real energy use or emissions**.**

Data and Preprocessing:

Data Source:

The data needed for this paper were GDP, energy consumption and NTL.

Nighttime Light Data: VIIRS (Visible Infrared Imaging Radiometer Suite) NTL band DND (day/night band) annual composite data was gathered through Google Earth Engine (GEE). This data obtained is highly accurate considering we are not dealing with images but band vales , also the representation is cloud-free which makes it optimal for large scale modelling.

GDP data for all state member countries of the world were obtained from open source websites World Bank Open Data platform, which ensure the credibility and consistency of data.

Energy consumption data is extracted form International Energy Agency (IEA) , the world bank. These figures are in unit of kilowatt-hours (KWh) of energy.

Spatial and Temporal Scope:

A global shapefile containing all country boundaries was downloaded form an open source website and loaded into GEE, which helped us in extracting data region wise. To ensure consistency and easy mapping on GDP and energy consumption data, we filtered this data to only include Member State countries. The data collected spans from years 2013 to 2023 as this range is covered by NTL data.

Nighttime Light Bands:

For each country and each year , we ended up choosing following 4 band values that conatin most relevant data:

- Median (masked)
- Average (masked)
- Minimum
- Maximum

Data Transformation:

In order to facilitate the use of NTL data , the data over the countries and years was transformed into both wide and long format.

Data Cleaning and Preprocessing:

Several preprocessing steps were done to ensure the data was ready to fed into model:

Country name matching: countries names across different data available like GDP, energy consumption and NTL was inconsistent and in different format. A mapping script was generated that helped us in aligning and standardizing the naming convention.

Null values handling: some of the GDP and energy consumption data was missing values , we ended up excluding those regions from our final data. NTL data was also checked for missing values to ensure no data inconsistency.

Methodology:

This section outline the main modeling techniques applied to our data for prediction of GDP and energy consumption using NTL.

Model Selection:

We ended up using different regression models in order to capture the complexity of band values of NTL to predict the GDP and energy consumption.

Linear Regression: A model for linear dependencies between NTL data and economic indicators.**Ridge Regression**: A regularized linear model for reducing overfitting.**Lasso Regression**: Similar to Ridge but with L1 regularization.**Random Forest Regressor**: An ensemble of decision trees that can model non linear relationships and interactions between variables.**XGBoost Regresso**r and **MLP Regressor (Neural Network)** .**SARIMAX**: A Seasonal ARIMA mostly suitable for capturing temporal dependencies in GDP and energy consumption series with NTL data as exogenous variables and finally **LSTM (Long Short-Term Memory)** and TCN recurrent neural network capable of learning long-term dependencies.

Feature Engineering:

Several feature engineering and transformation techniques were used in order to improve the model performance.

From NTL data for each country , year wise band values were extracted. In models like LSTM and SARIMAX, past values of our target variable GDP and energy consumption were also used as predictors to make use of temporal structure. Also Log Transform was used as target variable to address skewness in data and stabilize it. Input features were scaled using normalization for regression models.

Training and Testing Strategy:

The data was split into training and testing using built in train test split (80-20) rule. For time series models , the input was a single country and its target variable values over the range of under observation years to mimic the real world forecasting.

Evaluation Metric:

Model performance was evaluated using metrics like R^2 which indicates the proportion of variance explained by model and MAE measures the average magnitude of prediction error.

Results and Analysis:

1- **GDP:**

The performance of all above mentioned models was evaluated using R squared and MAE, some models were trained with log transform on GDP to study the impact of distribution normalization on prediction of models.

Without log transform:

Model	**R^2 Score**	**MAE (in actual GDP units)**
Ridge Regression	0.7180	3.93×10^{11} (393.36B)
Lasso Regression	0.5984	4.34×10^{11} (434.39B)
Random Forest	0.2397	5.08×10^{11} (508.22B)
XGBoost	0.2262	5.20×10^{11} (520.39B)
Linear Regression	0.1335	7.47×10^{11} (747.21B)
MLP Regressor	–0.0571	6.96×10^{11} (695.94B)

With log transform:

Model	**R^2 Score**	**MAE (in actual GDP units)**
Ridge Regression	0.5275	5.09×10^{11} (508.68B)
XGBoost	0.3146	4.60×10^{11} (460.12B)
Random Forest	0.2110	5.34×10^{11} (533.79B)
Lasso Regression	0.1849	6.16×10^{11} (615.53B)
Support Vector Regressor	0.1093	6.46×10^{11} (646.40B)
MLP Regressor	–0.0571	6.96×10^{11} (695.93B)

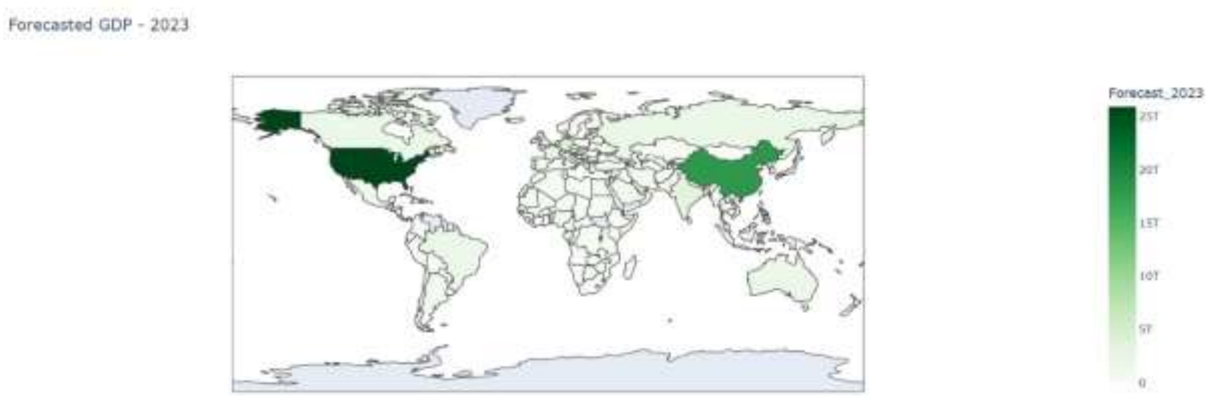
Ridge with r2 score of 0.71 without log transformed captured the variance in data better than with log transform after which it dropped to 0.52.

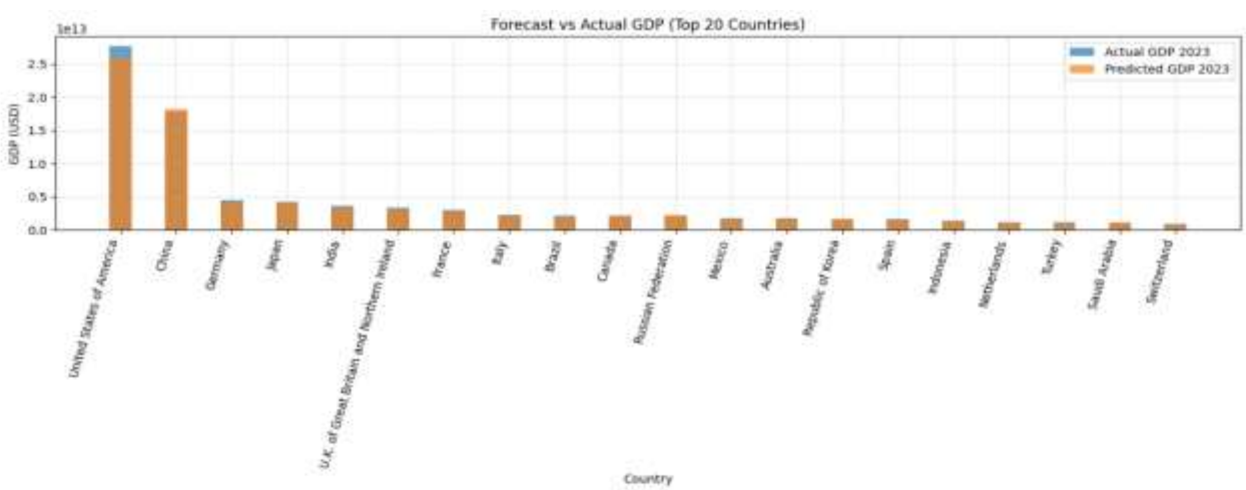
The time series models and their performance is summarized as follow in table.

Model	R^2 Score	MAE (in actual GDP units)
LSTM	0.9992	3.04×10^{10} (30.39B)
SARIMAX	0.9965	3.69×10^{10} (36.94B)
TCN	0.9910	8.73×10^{10} (87.27B)

Discussion:

For GDP, seeing the results gives us useful insight. We tried both temporal and non temporal models on our data. Among the non temporal models, without log transform Ridge Regression seems to capture the relationship between GDP and NTL band values nicely. Ensemble models like XGBoost and RFR also performed moderately well. But in the end the time series models like SARIMAX and LSTM outperformed these by r2 score of 0.99 and 0.996 showing the temporal dependencies effect.





2- Energy Consumption

The performance of various models was evaluated using **R^2 Score** and **RMSE**. To explore the impact of target normalization, some models were also trained using log-transformed energy consumption values.

Without Log Transform:

Model	**R^2 Score**	**RMSE**
Ridge Regression	1.0000	12.11
Linear Regression	1.0000	12.11
Lasso Regression	0.9998	51.11
XGBoost	1.0000	19.87
Random Forest	0.8738	1222.91
SVR	0.0691	3321.10

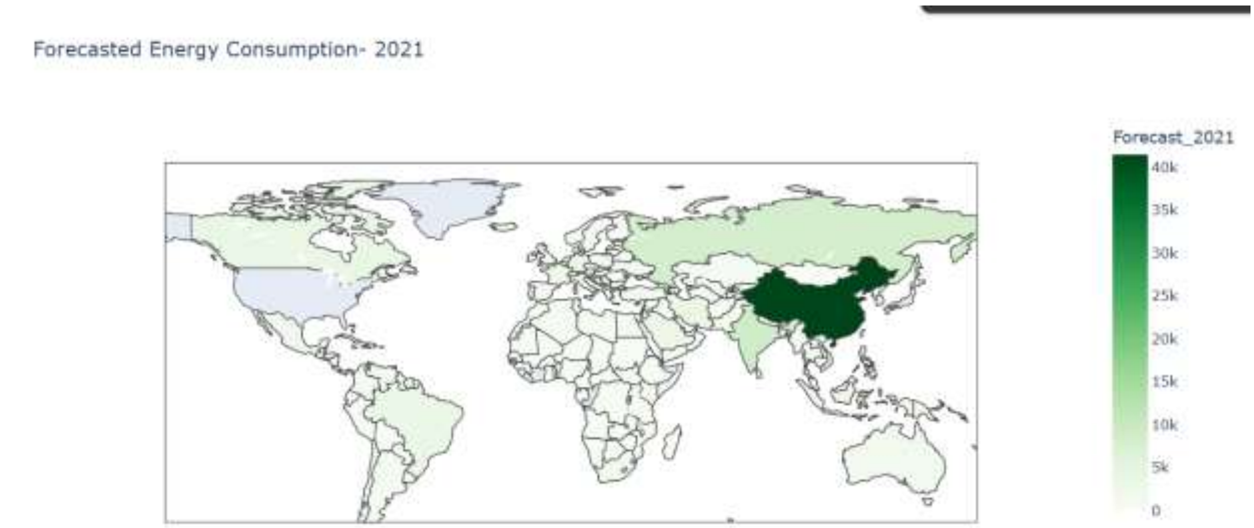
With Log Transform:

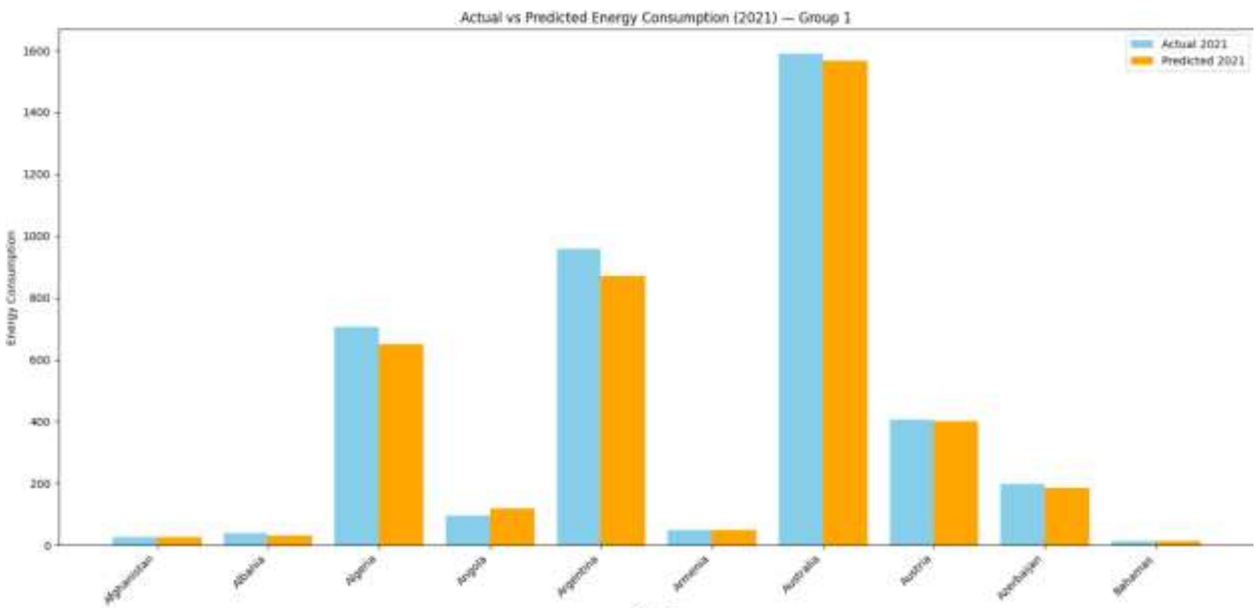
Model	**R^2 Score**	**RMSE**
SVR	0.9873	387.29
Random Forest	0.7586	1691.24
Linear Regression	0.6974	1893.65
Ridge Regression	0.6974	1893.64
Lasso Regression	-0.1226	3647.06
XGBoost	0.9988	120.38

Time Series Models:

Model	**R^2 Score**	**RMSE (in actual energy units)**
SARIMAX	0.9969	191.63

LSTM	0.9763	91.45
TCN	0.8276	246.81





Discussion:

Without log transform, **Ridge** and **Linear Regression** achieved near-perfect R^2 scores (1.0000) with minimal RMSE, suggesting overfitting. **XGBoost** and **Random Forest** also performed well, while **SVR** showed poor performance ($R^2 = 0.0691$).

With log transform, **SVR** and **XGBoost** significantly improved, with SVR reaching **R^2 = 0.9873**. However, **linear models** like Ridge and Lasso saw reduced accuracy, and Lasso performed poorly (R^2 = –0.1226).

Time series models outperformed most traditional models. **SARIMAX** and **LSTM** delivered excellent results (R^2 = 0.9969 and 0.9763 respectively), capturing temporal dependencies well. **TCN** also performed robustly (R^2 = 0.8276).

NTL data extracted form GEE proves to be a good predictor of GDP and Energy Consumption , this study highlights the importance and potential of using satellite data in prediction of economic factors worldwide.

We did face some challenges like mismatch of country names across datasets and issue in temporal alignment of our data and NTL and sensitivity of some models to feature scaling.

Conclusion and Future Work:

This study highlights the effectiveness of VIIRS Nighttime Light (NTL) data in predicting GDP and energy consumption using machine learning models.There is a strong correlation between the night time light in predicting the energy consumption and GDP by applying different machine learning models.In futre,our aim is to explore more deep learning models to improve prediction accuracy and support finer-scale analysis.

COLAB LINK:

Dataset link:

https://drive.google.com/drive/folders/1cYYwowYAiy5tQywdaW73m9TtOtGJlvbd?usp=sharing

Visualization of energy Consumption:

https://colab.research.google.com/drive/1F0IfIgWA0mbwB2t-bUGe7T3bCHbXM-LR#scrollTo=kFpS-RelKz6B

Energy Consumption:
https://colab.research.google.com/drive/19OekOEzoNk_9Mg5i7xHgeMnvfugsbLS7#scrollTo=5hRavVRE8Uv5

GDP: https://colab.research.google.com/drive/1e5i5HNgXqQWs-e7are9RMGZWJwNSBWCm?usp=sharing

References:

[1] Y. Yu, X. Li, and J. Zhao, “Estimating GDP growth using VIIRS night-time light data,” *ResearchGate*, 2023. OnlineOnlineOnline. Available: https://www.researchgate.net/publication/370480584

[2] X. Chen, Z. Wang, F. Zhang, G. Shen, and Q. Chen, “A global annual simulated VIIRS nighttime light dataset from 1992 to 2023,” *Scientific Data*, vol. 11, no. 1380, Dec. 2024. doi: 10.1038/s41597-024-04228-6

[3] P. Jhamb, S. Ferreira, and S. Ferreir, “Shedding light on development: Leveraging the new nightlights data to measure economic progress,” *PLOS ONE*, vol. 20, no. 2, Feb. 2025. doi: 10.1371/journal.pone.0318482

[4] P. McSharry and J. Mawejje, “Estimating urban GDP growth using nighttime lights and machine learning techniques in data poor environments: The case of South Sudan,” *Technological Forecasting and Social Change*, vol. 203, p. 123399, 2024. doi: 10.1016/j.techfore.2024.123399

[5] H. Suleiman, “Illuminating the Nile: Estimating subnational GDP in Egypt using nighttime lights and machine learning,” *GeoJournal*, vol. 89, p. 117, May 2024. doi: 10.1007/s10708-024-11106-6

[6] arXiv preprint arXiv:2308.06366, Aug. 2023. [Online]. Available: https://arxiv.org/abs/2308.06366

[7] Y. Zhang, X. Zeng, and Y. Zhao, “Estimation Model and Spatio-Temporal Analysis of Carbon Emissions from Energy Consumption with NPP-VIIRS-like Nighttime Light Images: A Case Study in the Pearl River Delta Urban Agglomeration of China,” Remote Sensing, vol. 14, no. 22, p. 5776,. [Online]. Available: https://www.mdpi.com/2072-4292/14/22/5776

[8] H. Wang, L. Yu, Y. Wang, Y. Li, and X. Wang, “Research on Energy Consumption Carbon Emissions at Grid Scale Based on NPP-VIIRS Nighttime Light Data,” Sustainability, vol. 14, no. 16, p. 10212,. [Online]. Available: https://www.mdpi.com/2071-1050/14/16/10212

[9] W. Hu, Y. Liu, and B. Li, “Multi-Scale Dynamics and Spatial Consistency of Economy and Population Based on NPP/VIIRS Nighttime Light Data and Population Imagery: A Case Study of the Yangtze River Delta,” ISPRS International Journal of Geo-Information, vol. 12, no. 1, p. 16, Jan. 2023. [Online]. Available: https://www.mdpi.com/2220-9964/12/1/16

[10] X. Liu, J. Zhang, H. Yang, and L. Zhou, “Analysis of Spatiotemporal Changes in Energy Consumption Carbon Emissions at District and County Levels Based on Nighttime Light Data—A Case Study of Jiangsu Province in China,” Remote Sensing, vol. 16, no. 18, p. 3514, 2024. [Online]. Available: https://www.mdpi.com/2072-4292/16/18/3514