

Satellite Data for Nowcasting: Estimating Cambodia's GDP in Real Time Using Satellite Data in a Machine Learning Framework

Iyke Maduako, Dharana Rijal, Alberto Sanchez Rodelgo.

SIP/2026/001

IMF Selected Issues Papers are prepared by IMF staff as background documentation for periodic consultations with member countries. It is based on the information available at the time it was completed on November 5, 2025. This paper is also published separately as IMF Country Report No 25/318.

**2026
JAN**



IMF Selected Issues Paper
Asia Pacific Department

Satellite Data for Nowcasting: Estimating Cambodia's GDP in Real Time Using Satellite Data in a Machine Learning Framework

Prepared by lyke Maduako, Dharana Rijal, Alberto Sanchez Rodelgo.

Authorized for distribution by Kenichiro Kashiwase
January 2026

IMF Selected Issues Papers are prepared by IMF staff as background documentation for periodic consultations with member countries. It is based on the information available at the time it was completed on November 5, 2025. This paper is also published separately as IMF Country Report No 25/318.

ABSTRACT: Cambodia is not alone in facing capacity limitations in the production and timely release of key official statistics needed for data-driven policy decisions. This paper demonstrates that combining satellite-derived indicators (e.g., nighttime lights, NO₂ emissions, vegetation indices) with traditional high-frequency indicators in a machine learning framework significantly improves the accuracy of GDP nowcasts. Moreover, satellite data enables closer examination of subnational patterns, providing granular, near-real-time insights into economic activity. These findings highlight the potential of non-traditional approaches to complement conventional methods and strengthen macroeconomic surveillance in data-scarce environments.

RECOMMENDED CITATION: Maduako, I., Rijal, D. & Sanchez Rodelgo, A. (2025). Satellite Data for Nowcasting: Estimating Cambodia's GDP in Real Time Using Satellite Data in a Machine Learning Framework. IMF Selected Issues Paper No. 2026/001. International Monetary Fund

JEL Classification Numbers:	O53, C53, C52, C44
Keywords:	nowcasting; nowcast; satellite; satellite data; big data; non-traditional data; machine learning; random forest
Author's E-Mail Address:	imaduako@imf.org; drijal@imf.org; asanchez3@imf.org

SELECTED ISSUES PAPERS

Satellite Data for Nowcasting: Estimating Cambodia's GDP in Real Time Using Satellite Data in a Machine Learning Framework

Cambodia

Prepared by lyke Maduako, Dharana Rijal, Alberto Sanchez Rodelgo.



CAMBODIA

SELECTED ISSUES

November 5, 2025

Approved By
Asia and Pacific
Department

Prepared by lyke Maduako, Dharana Rijal, Alberto Sanchez
Rodelgo.

CONTENTS

SATELLITE DATA FOR NOWCASTING	3
A. Motivation – Why Satellite Data	3
B. Data and Methodology	4
C. Results and Interpretation	8
References	10

BOX

1. Satellite Indicators to Gain Timely and Granular Insights on Macroeconomic Developments	5
--	---

SATELLITE DATA FOR NOWCASTING¹

A. Motivation – Why Satellite Data

1. Cambodia faces limited institutional capacity in the production and timely release of quality official statistics, limiting policymakers' ability to make agile and effective policy decisions. While the country has made significant improvements on the availability and quality of national statistics, further strengthening of statistical capacity is needed. GDP data is available only at annual frequency and published with a significant lag, limiting timely analysis of comprehensive economic developments. To address this data gap, various methods can be used to estimate aggregate economic activity using high-frequency indicators that represent key sectors of the economy. However, these input indicators from traditional sources also often come with delays. These data issues are not unique to Cambodia. Many economies confront similar challenges and have been exploring options.

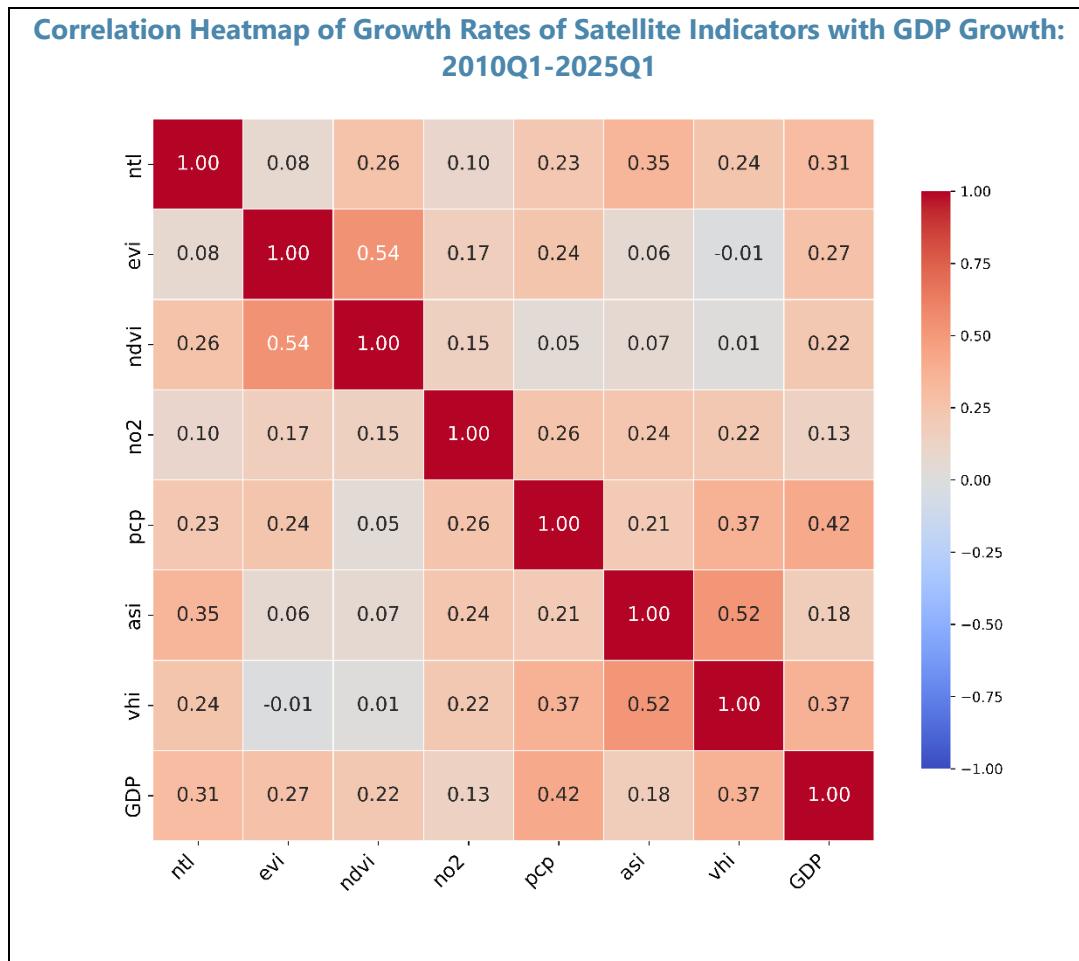
2. Satellite indicators are available for nearly all countries in the world. They exist in near-real time and at granular levels capturing nuances that may otherwise go undetected.

They can serve as proxies for economic activity in various sectors of the economy. In particular, data on nighttime lights, nitrogen dioxide (NO₂) emissions, and vegetation-related indices can help uncover underlying patterns and trends in sectors like manufacturing and agriculture.² In Cambodia, quarterly GDP growth rate (interpolated, see section on Data and Methodology) is positively correlated with changes in nighttime light (NTL), vegetation health index (VHI), and precipitation (PCP) (Figure 1). Satellite indicators can complement and fill the gaps in traditional indicators as they provide near real-time reflection of what is seen and felt on the ground and capture nuances in economic activity at high spatial and temporal granularities.

3. Machine learning models can make the best use of satellite indicators, along with macroeconomic data to analyze their complex interactions for nowcasting GDP. First, the dataset is split into 'train' and 'test' sets. The model learns patterns based on the train set, and its predictions are then evaluated against observed values in the test set—data that was not used during training (i.e. out-of-sample). This approach has the advantage of the model's performance to be assessed based on its ability to generalize to unseen data. In addition, in contrast to linear methods, the predicted value obtained through this approach accounts for complex, non-linear interactions that may exist between various indicators. Lastly, the nowcast can be updated monthly as up-to-date, high-frequency input data become available.

¹ Prepared by lyke Maduako, Dharana Rijal, and Alberto Sanchez Rodelgo (all STA).

² See analytical examples by Gibson (2020) and McSharry and J. Mawejje (2024) for nighttime lights, Ezran, Morris, Rama, and Riera-Crichton (2023) for nitrogen dioxide (NO₂) emissions, and Puttanapong, Prasertsoong, and Peechapat (2023) and Hu and Xia (2018) for vegetation-related indices.



B. Data and Methodology

4. The machine learning method applies quarterly satellite indicators, along with the traditional variables, for training the nowcasting model. The satellite (“non-traditional”) **indicators** used in this analysis include data on nighttime lights (NTL), NO₂ emissions (NO₂), precipitation (PCP), normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), vegetation health index (VHI), and agricultural stress index (ASI). These indicators are obtained from various sources, including Google Earth Engine and NASA (in the case of NO₂) and defined as follows:

- **Nighttime Lights (NTL)** are satellite-based measurements of the intensity of light emitted at the Earth’s surface, which is shown to be a good proxy for economic activities in many studies.³
- **Nitrogen dioxide (NO₂)** is a pollutant, primarily produced by the combustion of fossil fuels in power plants, industrial facilities, and vehicles. Because NO₂ is emitted in large quantities

³ See, for example, Forbes (2013); Ezran et al. (2023); Gibson et al. (2021)

when economic activity is high, satellite-based observations of NO₂ can approximate the level and distribution of economic activity on the ground.

- **Normalized Difference Vegetation Index (NDVI)** and **Enhanced Vegetation Index (EVI)** are computed using the red (R) and near-infrared (NIR) bands of satellite imagery. These indices measure vegetation health and can be used to proxy agricultural output, and land use changes, as well as expansion of cropland and infrastructure development.
- **Agricultural Stress Index (ASI)** is a satellite-based indicator designed to detect areas of cropland experiencing water stress—such as drought conditions—during the growing season.
- **Vegetation Health Index (VHI)** is computed using NDVI and Land Surface Temperature (LST) as inputs. First, the vegetation condition index (VCI) is derived from NDVI to assess vegetation greenness. Then, the temperature condition index (TCI) is calculated to measure how current surface temperatures deviate from their long-term average, highlighting heat or cold stress. This is also a proxy for agricultural wellness and crop yield.
- **Precipitation indicator (PCP)** obtained from Climate Hazards Center is InfraRed-based precipitation data combined with in-situ station data (CHIRPS). This is a quasi-global rainfall dataset of CHIRPS, which covers a long history (30 plus years) and incorporates 0.05° resolution satellite imagery with in-situ station data, to create gridded rainfall time series for trend analysis and seasonal drought monitoring. This indicator is also related to agriculture and food production.

Non-traditional and Traditional Indicators Used

Non-traditional Indicators

- Nighttime Lights (NTL)
- Nitrogen Dioxide (NO₂)
- Normalized Difference Vegetation Index (NDVI)
- Enhanced Vegetation Index (EVI)
- Vegetation Health Index (VHI)
- Precipitation (PCP)
- Agricultural Stress Index (ASI)

Traditional Indicators

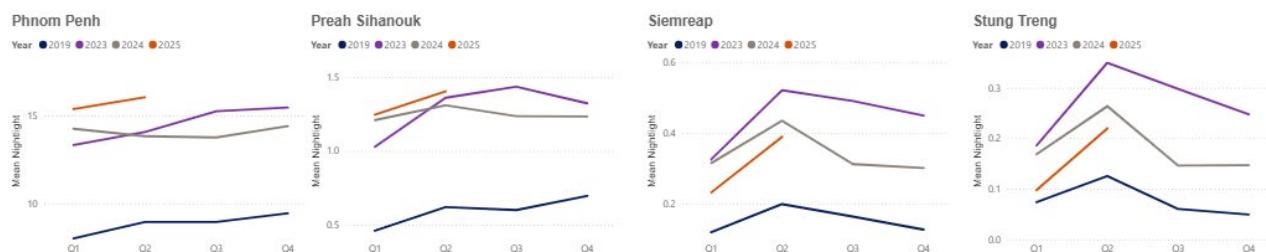
- Exports and Imports
- Broad Money
- Exchange Rates
- Consumer Price Index
- Lending Rate
- Credit
- Tourist Arrivals

Box 1. Satellite Indicators to Gain Timely and Granular Insights on Macroeconomic Developments ¹

Satellite indicators can serve as proxies for economic activity in various sectors of the economy (Annex I). For example, data on nighttime lights (NTL) and vegetation-related indices can help uncover underlying patterns and trends in local economic activity in manufacturing and agriculture. They can complement traditional high-frequency indicators by providing real-time reflection of what is seen and felt on the ground. Satellite indicators can also reveal granular, regional variations in economic activity and guide policy formulation in a targeted manner.

(A) Nighttime lights (NTL):

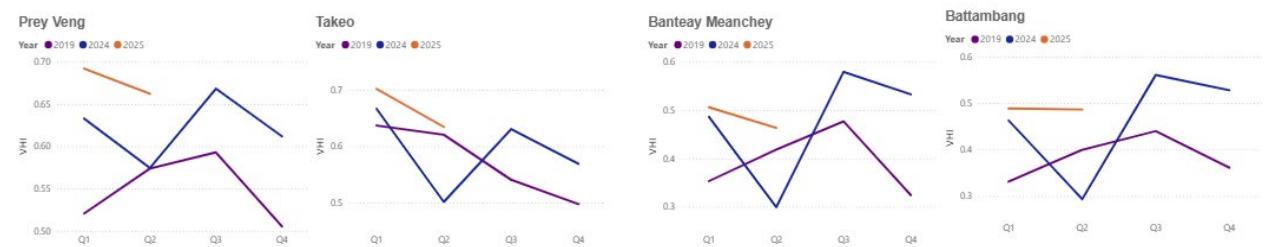
Nighttime lights, which capture the radiance or brightness of observed light can shed light on local economic activity. Compared to 2019, nighttime lights have increased across Cambodia in 2025. Urban regions, such as Phnom Penh and Preah Sihanouk, show higher nighttime lights in the first two quarters of 2025 compared to the same periods in 2023 and 2024. However, Siem Reap, a major destination for tourism, shows lower levels of nighttime lights in 2025 compared to recent years, which indicates a possible slump in the tourism sector. Other provinces in the country also show lower levels of nighttime lights in 2025, reflecting a slowdown in economic activity in the northeastern provinces such as Stung Treng, Ratanak Kiri, Mondul Kiri, and Kratie.



Sources: National Aeronautics and Space Administration (NASA/VIIRS/002/VNP46A2) and IMF Staff Calculations

(B) Vegetation-related indices:

Vegetation-related indices are calculated based on the amount of light reflected by plants and serve as proxies for vegetation health. Among these indices, the Vegetation Health Index (VHI) takes into account both vegetation greenness and data on surface temperatures, thereby serving as a proxy for agricultural wellness. In Cambodia, average VHI in the provinces with higher shares of cropland indicate healthier vegetation in the first two quarters of 2025 as compared to the same periods in 2019 and 2024.



Sources: FAO - Agricultural Stress Index System (ASIS), <http://www.fao.org/gIEWS/earthobservation/>, [Date accessed: 09-23-2025] and IMF Staff Calculations.

¹ This box was prepared by Dharana Rijal (STA).

5. The machine learning model applies interpolated series when some data points are missing in the traditional high-frequency macroeconomic indicators. The model takes key macro variables (Table 1), most of which are available starting 2010q1. In case some observations are missing, we impute data based on some historical patterns as needed. For GDP, Cambodia has annual data only. We have applied the quarterly GDP series (year-on-year growth rates) of Cambodia's major trading partners,⁴ aggregated with respective export weights, for producing Cambodia's quarterly GDP series. This interpolation methodology is applied since export growth in Cambodia drives its business cycles and navigates economic growth over time.

6. The nowcasting model uses the random forest machine-learning algorithm to predict year-on-year quarterly GDP growth rate. Random forest (Breiman, 2001) is a collection of decision trees, with each built on various subsamples of data drawn with replacement (i.e., bootstrapping). For each tree, a random subset of predictors is selected at each split. At each node of the decision tree, the algorithm chooses the feature and split point that minimizes the root mean squared error (RMSE). This process continues recursively until a stopping criterion is met, such as minimum node size, or if additional splits no longer reduce the RMSE. The final prediction is obtained by averaging the predictions from all trees, a process known as bootstrap aggregation or bagging. The final prediction can be represented as:

$$y(x) = \frac{1}{M} \sum_{m=1}^M T^m(x)$$

where $y(x)$ is the predicted value; x is the vector of input variables we use to make a prediction; m is the index of each individual Decision Tree in the Random Forest, ranging from 1 to M , where M is the total number of Trees; and $T^m(x)$ is the prediction made by the m^{th} decision tree for input x . The Random Forest prediction is obtained by averaging the outputs of all individual trees in the ensemble.

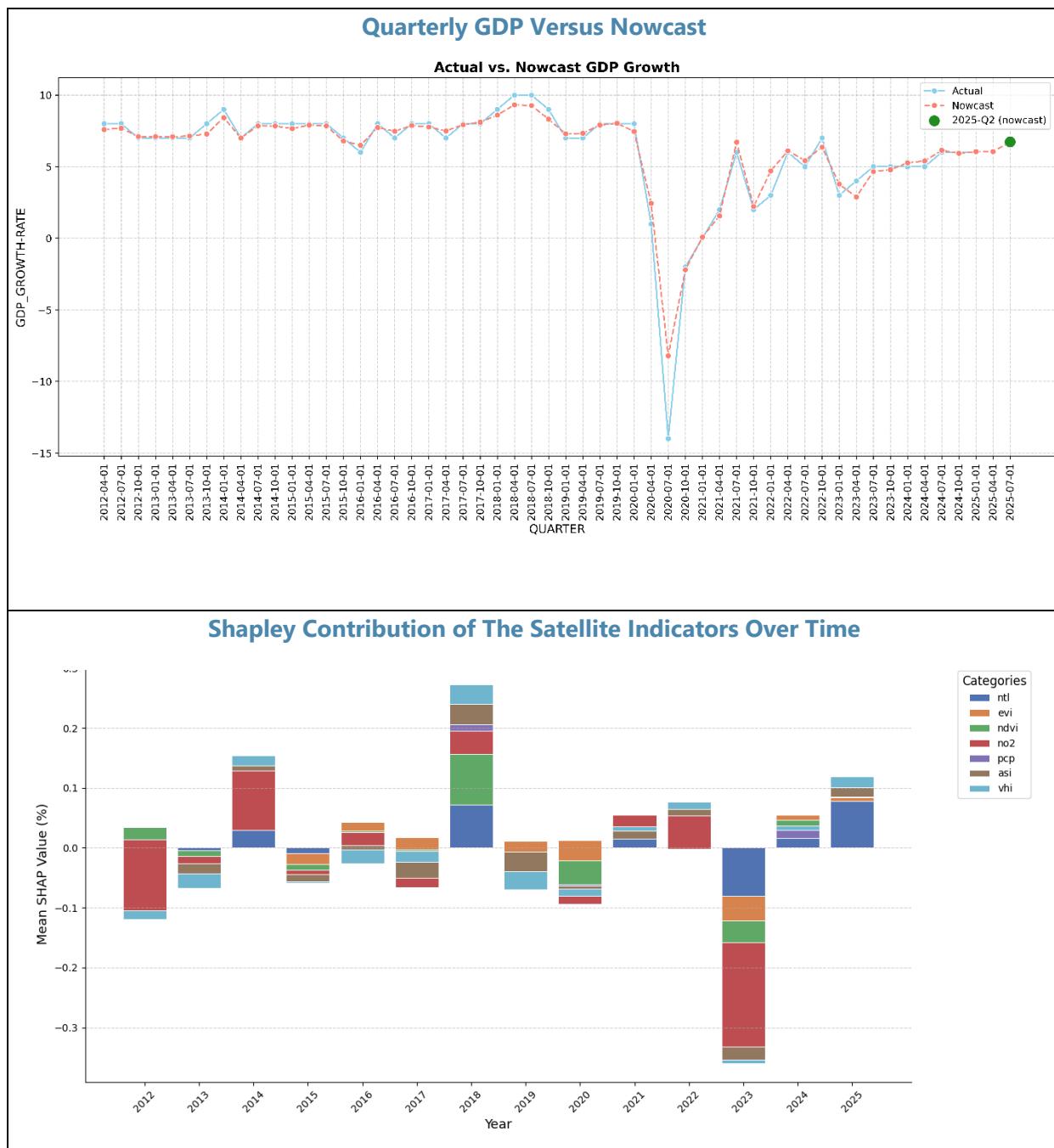
7. The machine learning algorithm analyzes the underlying relationships among key variables by splitting the dataset into two groups (“training” and “test” datasets). This is the key feature of the algorithm which exploits the “training” dataset to learn the relationships in the past, uses the statistical relationships to predict values based on the “test” dataset, and evaluates goodness of the fit based on the difference between the model-based predicted values and the actual values in the “test” dataset. We implement cross-validation in a chronological order where 85 percent of the historical data is treated as the “training” dataset representing the “past”, and the remaining 15 percent of the historical data is treated as the “test” dataset (or “holdout sets”) representing the “future”.

⁴ Major trading partners include Canada, China, Germany, Japan, Korea, Malaysia, Singapore, Thailand, United States, and Vietnam.

C. Results and Interpretation

8. **The random forest machine-learning technique demonstrates a strong fit, pointing to a nowcasting result of 5.7 and 6.7 percent GDP growth year-on-year in 2025q1 and 2025q2, respectively with the underlying stories.** The alignment between the actual and nowcast regression lines can be measured by root mean square error (RMSE) of 0.9 (Figure 2). Shapley decomposition shows contributions of the variables to the predictive power of the model. (Figure 3). It is important to note that shapely contributions are not linked to causality, but contributions of the variables to the ability of the model to accurately predict the GDP growth rate. The addition of satellite variables to the list of indicators used to train the model improved the model accuracy by over 20 percent. This percentage might not seem substantial because of the weight of the traditional indicators, which is significantly higher. However, in situations where these traditional indicators are scarce or not collected on time, satellite indicators can fill the gap and contribute more to the accuracy of models.

- Amongst satellite indicators, Nightlight and NO₂ seem to be the most influential variable across time. Looking over the period of 2012-2025, we find that NO₂ emission seems to have the largest influence on model predictions among satellite indicators. This suggests NO₂ emission levels (as a proxy for industrial activity) are highly predictive of GDP dynamics in Cambodia.
- In the recent period of 2021-2025, the nighttime light (NTL) shows stronger influence on the model, indicating its growing alignment with economic activities visible from space at night. Factors, such as urbanization, tourism, household electricity access and consumption, might explain the growing influence.
- The vegetation indices show modest but consistent contributions to economic activities over the years, with spikes in 2017 and the first quarter of 2025. This reflects important roles of agriculture in Cambodia's economy when it faces volatility in production potentially under the influence of climate change.
- The nowcast for the first quarter of 2025 indicates a year-over-year growth rate of 5.68 percent. Our analysis shows that non-traditional indicators complement traditional ones and serve as good alternatives when traditional indicators are scarce. Including non-traditional indicators in the nowcasting model improved RMSE and MAE metrics by over 20 percent, reducing RMSE from 1.2 to 0.9 and MAE from 1.0 to 0.8.



References

Ezran, I., Morris, S. D., Rama, M. and Riera-Crichton, D. (2023), *Measuring global economic activity using air pollution*, World Bank.

Gibson, J., Olivia, S., Boe-Gibson, G. and Li, C. (2021), 'Which night lights data should we use in economics, and where?', *Journal of Development Economics* **149**, 102602.

Hu, M. and Xia, B. (2019), 'A significant increase in the normalized difference vegetation index during the rapid economic development in the Pearl River Delta of China', *Land degradation & development* **30**(4), 359–370.

McSharry, P. and Mawejje, J. (2024), '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* **203**, 123399.

Puttanapong, N., Prasertsoong, N. and Peechapat, W. (2023), 'Predicting provincial gross domestic product using satellite data and machine learning methods: A case study of Thailand', *Asian Development Review* **40**(02), 39–85.

Breiman, L. (2001), 'Random forests', *Machine learning* **45**, 5–32.