Policy Brief
August 2024

Spatial Informatics Group

UNDP

BIG DATA AND ARTIFICIAL INTELLIGENCE FOR MAPPING VULNERABILITY IN CAMBODIA

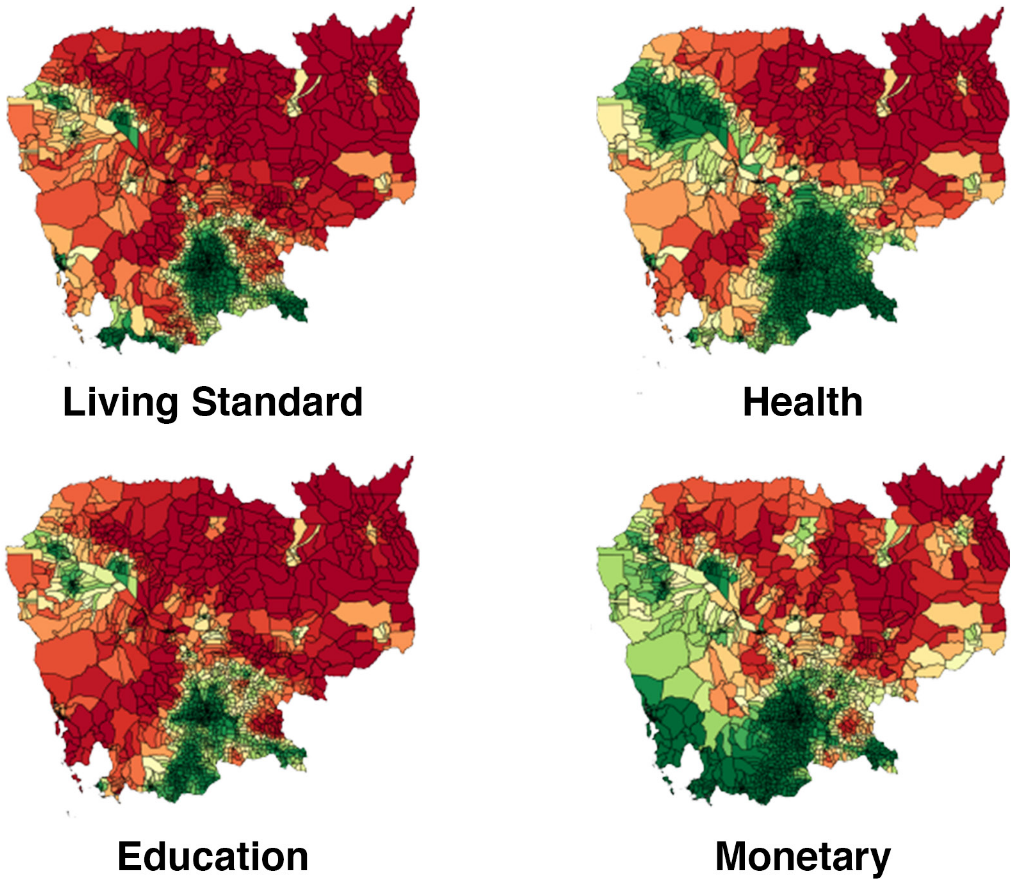
ARTIFICIAL INTELLIGENCE CAN HELP SUPPORT POLICIES ON INCLUSIVE GROWTH, DECENT JOBS AND SECURE LIVELIHOODS

ABSTRACT

Cambodia's economic prosperity in recent years has increased rapidly. Yet despite significant economic growth, it remains one of the most vulnerable countries in Southeast Asia and faces many unresolved challenges. Infrastructure, water, sanitation, roads and solid waste management are, for example, among critical services in need of improvement. Mapping deprived households in different regions of the country is critical to the understanding of vulnerabilities in the dimensions of health, education, living standards, employment and monetary factors.

In an initiative undertaken by the United Nations Development Programme (UNDP) in partnership with SERVIR-Mekong and the Spatial Informatics Group (SIG), big data was combined with information from the Cambodia Socio-Economic Survey (CSES) 2019 to map deprivations across the country using artificial intelligence (AI). A machine learning algorithm was used with big earth data containing information on critical infrastructure, public services such as hospitals, clinics, schools and universities, satellite and satellite-derived datasets to calculate the likelihood of deprivation in living standards, health, education and finance.

Whereas poverty at the household level can be transient which is difficult to map, it was demonstrated that AI accurately predicts spatial patterns for different vulnerability dimensions across Cambodia. AI and big data can thereby support traditional methods of measuring vulnerabilities by taking the spatial context explicitly into account. This enables the study of spatio-temporal dynamics at a high granular level and aggregates the information at different administrative levels. This brief demonstrates how big data and AI can support vulnerability mapping and achievement of the Sustainable Development Goals (SDGs), as it allows for the mapping of different dimensions of vulnerability and deprivation levels at will. For future studies, it is recommended to explicitly attach spatial information for data collection to support the use of big data and AI.



Probability of deprivation across different dimensions of vulnerability.
Red indicates high probability of deprivation, Green indicates low probability.

THE DEVELOPMENT CONTEXT

INTRODUCTION

Historically, it has been difficult to capture up-to-date, detailed information on the factors contributing to poverty. This is particularly true in developing nations. Survey data collection, for instance, is time consuming and tedious. Fortunately, advances in machine learning and big data now allow for mapping at-risk areas within a geographic region.

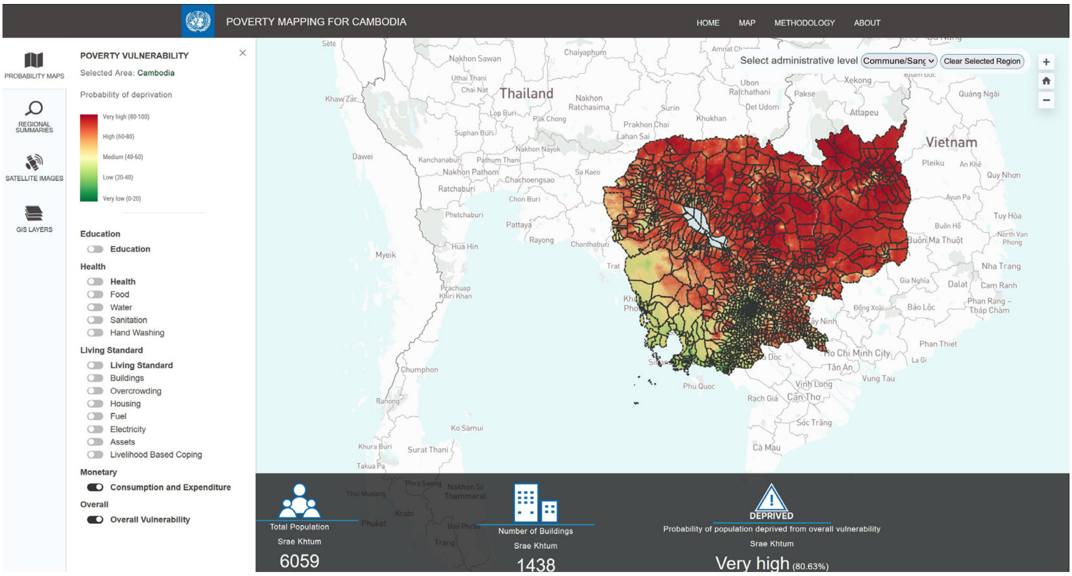
– Big (earth) data brings disruptive innovation that can be leveraged to tackle the biggest global challenges of our lifetime. –

Poverty mapping is a promising tool for developing countries like Cambodia to support targeted aid efforts. More than 17 percent of Cambodia's population lives below the national poverty line, and critical resources like education, water, roads and solid waste management need improvement. Importantly, poverty mapping is now more feasible with the recent arrival of inexpensive and accessible cloud computing platforms and free, publicly available datasets.

The United Nations Development Programme (UNDP) in partnership with SERVIR-Mekong and the Spatial Informatics Group (SIG) developed the web-based Cambodian Poverty Mapping Tool to make vulnerability information widely available. This sophisticated, easy-to-use tool allows users to toggle through maps that assess multiple deficiencies in key sectors contributing to poverty.

With this tool, users can explore risks at the provincial, district or sangkat levels. Users can also explore specific poverty indicators such as education, health, living standards and monetary factors. The tool provides accurate estimates of at-risk households based on when access to education, clean water and electricity is limited. The user interface is also exceptionally easy to navigate. As a result, the tool will benefit all users, including community leaders and government agencies.

The Cambodian Poverty Mapping Tool uses publicly available, area-specific data to highlight different factors contributing to poverty called indicators. These indicators are displayed in colour-coded maps to help users understand different aspects of poverty, such as poor access to education and food.



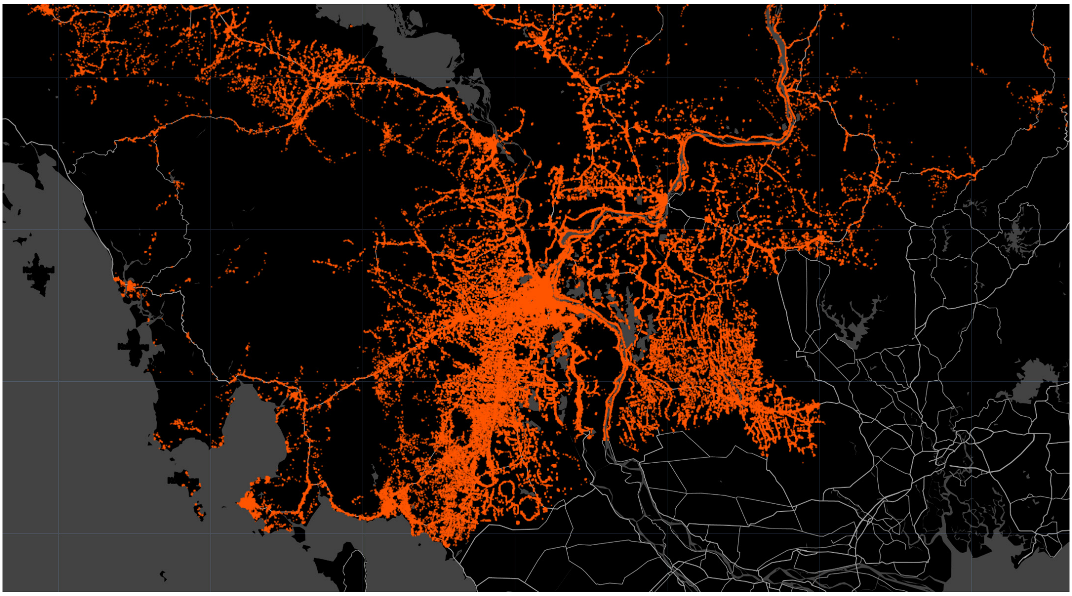
Different vulnerability dimensions can be explored across different administrative areas at https://cambodiapovertymapping.sig-gis.com/.

VITAL INFRASTRUCTURE DATA

Area-specific data is needed to understand poverty in a given region. Some of the many types of important data for understanding poverty include data related to health care, public utilities like water treatment plant locations, and roads and other infrastructure.

– Poverty is not simply a critical lack of financial resources. Poverty results from limited public services, strained infrastructure and unproductive natural resources. –

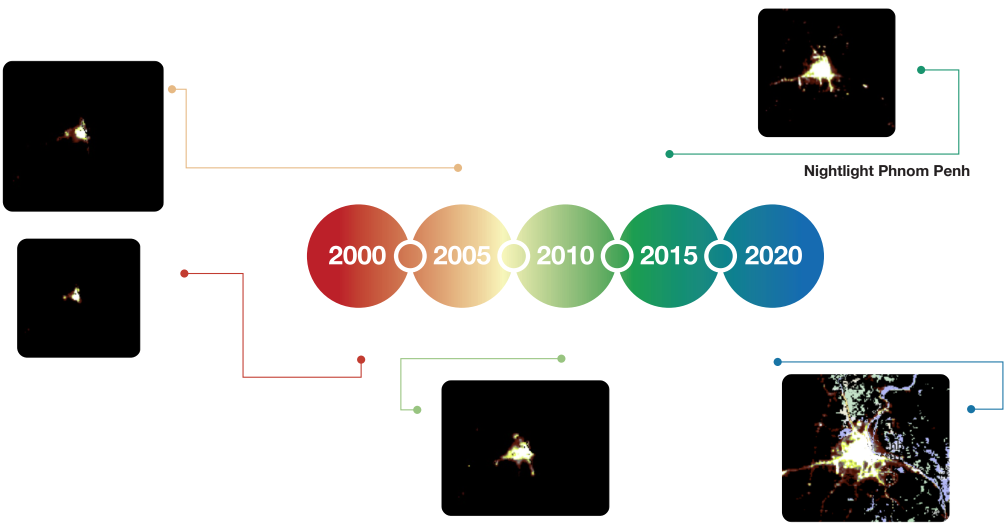
Public datasets on the road network, water supply network, buildings and public utilities were used to calculate accessibility of households to, for example, clean water, sanitation, health care and energy. Combining the spatial information layers can help to identify underserved communities.



The Microsoft Global Building Footprint dataset was generated using AI and provides information on human settlements across Cambodia. This dataset is critical for gaining an understanding of the spatial dependencies that contribute to vulnerability.

BIG EARTH DATA

Earth observation imagery has become an indispensable tool for assessing poverty. Satellite imagery has the distinct advantage of covering large geographic areas regularly and consistently. The Cambodian Poverty Mapping Tool includes data from NASA's Black Marble night light dataset. This data, which shows artificial light visible from orbit during the night, provides a visual display of the country's economic growth over time.



For the period 2000 until 2020, a large increase in night light produced in Phnom Penh can be observed.

A CLOUD-BASED COMPUTING PLATFORM

Cloud-based computing offers strong advantages for teams using big datasets to make complex calculations. SIG used the cloud-based Google Earth Engine (GEE; Gorelick et al., 2017) to develop the Cambodian Poverty Mapping Tool. GEE is a cloud-based geospatial processing platform for large-scale data analysis.

GEE makes available cutting-edge hardware, software and data that would be cost-prohibitive for many institutions to maintain in-house. The low cost of working on the GEE platform democratizes data and computational power, allowing countries like Cambodia to address some of humanity's most difficult issues.

CAMBODIA'S SOCIO-ECONOMIC SURVEY

The Royal Government of Cambodia conducts its national Cambodian Socio-Economic Survey (CSES) every two years. This survey samples ten thousand households throughout the country and provides a comprehensive look at the country's living standards and poverty indicators. The CSES is a 60-page, in-depth questionnaire with a wide range of questions covering areas such as health, education, housing conditions and economic activities, in addition to well-being queries on victimization, violence and vulnerability.

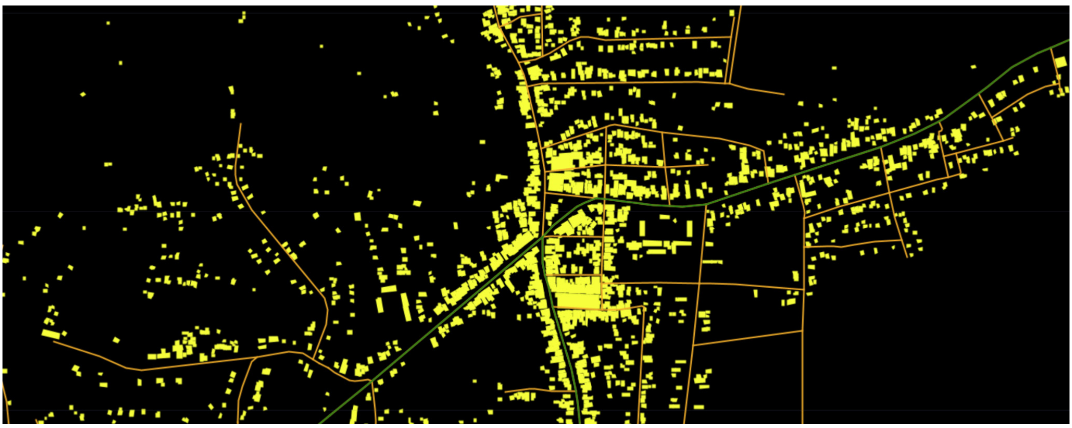
This data is invaluable to fleshing out the factors which contribute to poverty in the region. The CSES also includes information on surveyed participants' locality, which allows the poverty mapping model to add approximate coordinates for each household to further enhance the geospatial value of the data. Whereas CSES data is invaluable to analyzing the socio-economic situation, it does not respond to rapid societal changes such as pandemics, high inflation rates or other types of civil unrest. Moreover, the sample size is generally quite large, but is not exhaustive. Conversely, big data and artificial intelligence can rapidly respond to changing conditions and can also map patterns through space and time which remain unclear in survey data.

AI AND MACHINE LEARNING TO UNCOVER PATTERNS

Artificial intelligence (AI) and machine learning (ML) algorithms were employed throughout the data generating and poverty map model building process. For example, the land cover map derived from Landsat imagery was created using deep neural networks. Neural networks are a ML method that divide images into layers, allowing for the identification of subsects based on distinguishing characteristics within the various layers. Likewise, ML helped produce the WorldPop population density estimations and the Microsoft Global ML Building Footprints dataset.

Combining the datasets to form the Cambodian poverty mapping model also depends upon AI computing. The AI used is called the 'random forest model'. This model translates big data into indicators that help to uncover each area's overall economic vulnerability. An ensemble is a collection of many different models that analyze the data in slightly different ways. The benefit of this strategy is that it is more accurate, as the different models will not have the same errors.

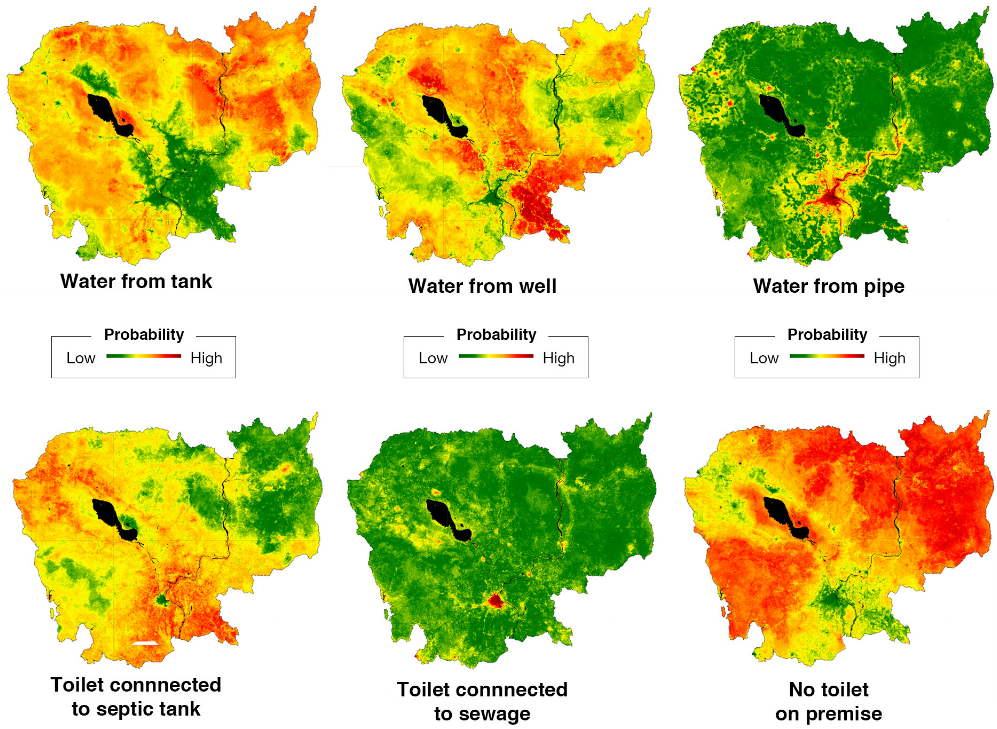
Using this multi-pronged approach, the Cambodian Poverty Mapping Tool can isolate the most impactful variables and translate big earth and socio-economic data into easy-to-understand indicators of poverty. These indicators can help to identify at-risk groups within a larger population.



UNITED NATIONS SUSTAINABLE DEVELOPMENT GOALS

AI can play a significant and novel role in achieving the United Nations Sustainable Development Goals (SDGs) – particularly the targets of zero hunger, health and well-being, quality education, and clean water and sanitation. Many of the SDGs have explicit spatial dimensions. Whereas the individual risk of poverty is transient, the risk of poverty at the regional level can be persistent. If persistence of regional poverty is the case, then local factors are contributing to poor outcomes in the area. These factors could include a lack of services like water treatment and education, or a lack of personal safety. Identifying the need for critical services using big earth data, spatial informatics and AI can go a long way in meeting the targets set out in the SDGs.

Development Goal 6 on clean water and sanitation, for example, depends on access to a reliable and safe source of water and decent sanitary facilities. The three main sources of (drinking) water are storage tanks (often filled with rainfall), wells and water from a pipe. A clear spatial pattern can be observed with people in the main cities reporting that they obtain water from a pipe, and people in remote areas reporting that they obtain water from tanks and wells. It is notable that households east of Phnom Penh are more likely to obtain water from a well than from a tank. Water from unprotected sources is a leading cause of child morbidity in Cambodia (Poirot et al., 2020)



Probability of access to water from different sources (top) and probabilityof access to different sanitation facilities (bottom).

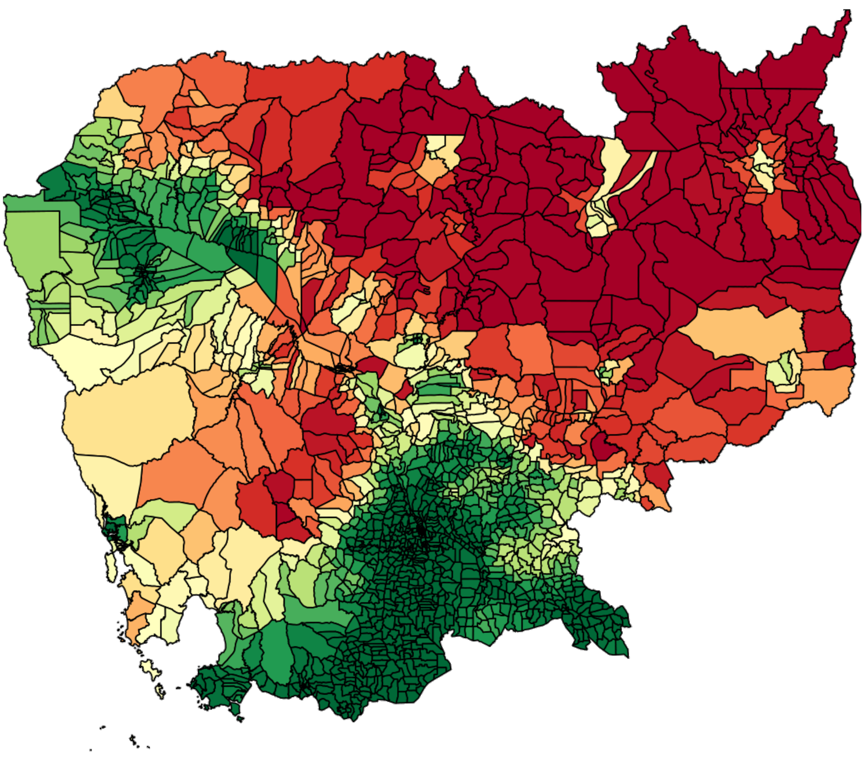
There are also clear spatial patterns for access to sanitary facilities. Households in the capital and other dense urban centres are more likely to have access to a sewage system, while households in the remote areas often reported having septic tanks, which can be seen in spatial representation. Very remote areas are classified with a high probability of there being no access to a toilet on the premises. Poor sanitation has important implications for the health and nutritional status of children (Vyas et al., 2016).

ASKED AND ANSWERED

The Cambodian poverty mapping project provides the means to answer pressing questions regarding the root causes of economic vulnerability within communities. Perhaps the most important question that this tool allows a user to answer is: what is the extent and severity of poverty within a given area?

This information then allows users to follow up with more in-depth inquiries, such as: what resources are lacking in these areas that have limited the economic outcomes of the inhabitants? The answer to such a question could be, for example, a lack of infrastructure like paved roads, or inadequate health services.

Poverty mapping tools like this one also help to answer questions about where to direct new services to have the most impact. These services include the location of basic infrastructure developments, the allocation of grant resources and the piloting of conditional cash transfers to replace subsidies. Thus, more efficient programmes may be designed and implemented for the most at-risk members of a population.



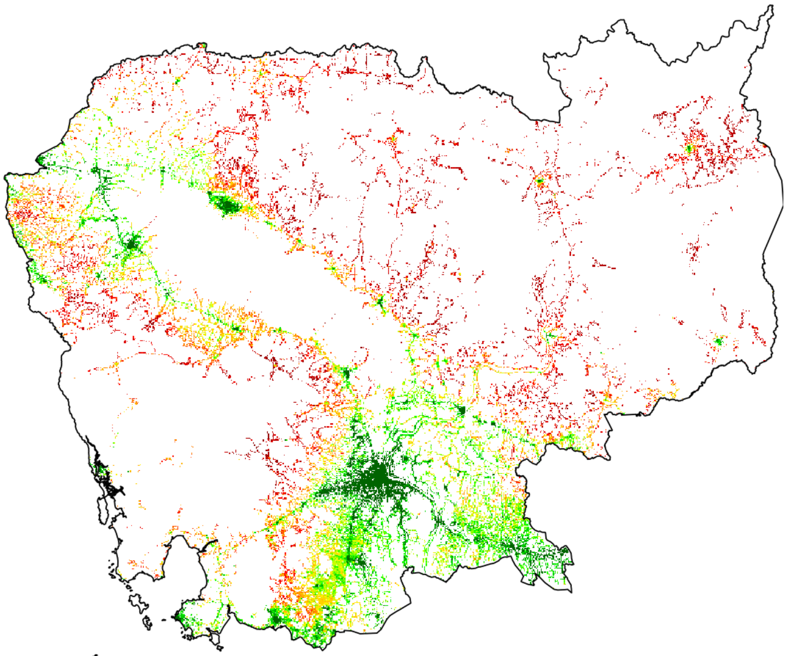
Overall vulnerability at the commune level in Cambodia.
Red colours indicate high levels of vulnerability, Green colours indicate lower levels.

Challenges and limitations of big data and artificial intelligence for mapping vulnerabilities include the availability of good and unbiased training data. Without a substantial amount of training data, the machine learning algorithms cannot find meaningful and reliable results. Moreover, the methods provide generalized results, thereby predicting an averaged level of vulnerability for a given area, not accounting for specific inequalities within a given area. Furthermore, there are many important ethical dilemmas including, but not limited to, ownership, privacy, compliance, transparency, consent and trust.

WHAT DOES THE FUTURE HOLD FOR POVERTY MAPS?

The poverty mapping tools of tomorrow promise further democratization of big data and supercomputing. In the future, poverty mapping will rely less on expensive, infrequent and inconsistent survey-derived data and more on remote sensing and big earth data that is collected frequently and at a regular rate. Numerous promising studies use AI and ML to estimate wealth using image-based data with great accuracy. These techniques have the advantage of capturing data often and consistently. They are also inexpensive for developing countries to use as the data is free to access.

This data further allows for auto-updating. This feature will let poverty maps display the most recent changes to areas of vulnerability. Imagine a scenario where people become displaced due to a conflict or natural disaster. In such a scenario, it would be extremely useful to know the location of new settlements and the impact they have on the surrounding area over time. This scenario hints at another advantage provided by remotely sensed poverty maps, which is the ability to track changes in vulnerability over time. For example, following a targeted aid campaign, the changes in vulnerability distributions could be tracked and the impact of the investment quantified.



Number of deprivations per household. Green colours indicate a low number of deprivations, while red colours indicate a high number.

IN CONCLUSION

Recent innovations in cloud-based computing, in combination with big earth data and survey-generated datasets, have made the creation of the Cambodian Poverty Mapping Tool possible. This innovative tool provides small-scale, geographically-specific information that promotes focused humanitarian efforts and government projects that meet many of the aims set out in the Sustainable Development Goals.

This tool can be utilized by policymakers and community members to combat poverty, regardless of technical know-how. The tool is critical for addressing Cambodia's rapid economic and climate changes as it efficiently addresses the needs of the most at-risk populations.

ACKNOWLEDGEMENTS

This policy brief was developed by the United Nations Development Programme (UNDP) and Spatial Informatics Group in close collaboration with SERVIR-Mekong. The policy brief was prepared by Dr. Ate Poortinga (SIG Service Lead Southeast Asia), Dr. Iván González de Alba (UNDP Country Economist), Theara Khoun (UNDP Policy Specialist, Social and Economic Affairs) and Nyein Soe Thwal (ADPC Data Scientist). We thank Anika Funk, Nicholas Miley and Dr. Karen Dyson for their editorial work and constructive feedback.

Cover and Layout Design by Mr. Panha Som

August 2024

DISCLAIMER

The views expressed in this working paper do not represent those of the member countries of the United Nations, the UNDP Executive Board or of those institutions of the United Nations system that are mentioned herein. Working papers serve the purpose of quickly disseminating ongoing research, data and new findings amongst development practitioners. Their content is the full responsibility of the individual author. The designations and terminology employed, and the presentation of material do not imply any expression or opinion whatsoever on the part of the United Nations concerning the legal status of any country, territory, city or area, or of its authority, or of its frontiers or boundaries.

REFERENCES

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. and Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202,* p.18-27.

Poirot, E., Som, S.V., Wieringa, F.T., Treglown, S., Berger, J. and Laillou, A., 2020. Water quality for young children in Cambodia—High contamination at collection and consumption level. *Maternal & Child Nutrition, 16,* p.e12744.

Saah, D., Tenneson, K., Poortinga, A., Nguyen, Q., Chishtie, F., San Aung, K., Markert, K.N., Clinton, N., Anderson, E.R., Cutter, P. and Goldstein, J., 2020. Primitives as building blocks for constructing land cover maps. *International Journal of Applied Earth Observation and Geoinformation, 85,* p.101979.

Vyas, S., Kov, P., Smets, S. and Spears, D., 2016. Disease externalities and net nutrition: Evidence from changes in sanitation and child height in Cambodia, 2005–2010. *Economics and Human Biology, 23,* p.235-245.