DEEP WORKING PAPER 02

How can new technology support better measurement of extreme poverty?

Paul Jasper, Angela Luna Hernandez, Molly Scott, Nikos Tzavidis

May 2021

About DEEP

Our mission is to build evidence, insights, and solutions that help end extreme poverty globally.

We aim to contribute to new global and national data and evidence that governments, decision makers, citizens and researchers can use to improve people's lives and support the world's poorest people in their efforts to escape extreme poverty.

We are a consortium of the Universities of Cornell, Copenhagen, and Southampton led by Oxford Policy Management, in partnership with the World Bank's Development Data Group and funded by the UK Foreign, Commonwealth & Development Office.

Acknowledgements

We are thankful for the discussions with, extensive reviews from, and useful comments received on earlier drafts of this paper by Patrick Ward, Peter Lanjouw, Chris Barrett, John Hoddinott, Jana Harb, Lucy Scott, and others.

Suggested Citation: Jasper, P.; Luna Hernandez, A; Scott, M. and Tzavidis, N. (2021) How can new technology support better measurement of extreme poverty? *DEEP Working Paper 02*. Data and Evidence to end Extreme Poverty Research Programme: Oxford.

See also the accompanying summary: Jasper, P. (2021) How can new technology support better measurement of extreme poverty? *DEEP Research Insight 01*. Data and Evidence to end Extreme Poverty Research Programme: Oxford.
Available at: https://doi.org/10.55158/DEEPWP2

Corresponding Author: Paul Jasper, email paul.jasper@opml.co.uk

Connect with the team:

Abstract

Access to high-quality and timely information about extreme poverty is required to develop and target appropriate policies, strategies, and programmes to tackle it, and to monitor progress towards achieving the Sustainable Development Goals. Despite advances over the last 20 to 30 years, this information remains slow to emerge, is often available only at a high cost, and is frequently only available at high levels of spatial aggregation. Traditional approaches to poverty measurement have relied on household surveys or census data, which are costly and therefore collected infrequently. New sources of data – including remote sensing data, data from the data exhaust, online data, crowd-sourced data, and mobile phone survey data – and novel statistical techniques have the potential to enhance the toolkit of approaches available for measuring and investigating extreme poverty. This paper is one of a series produced by the Data and Evidence to End Extreme Poverty (DEEP) research programme to explore how innovation in data collection, data processing, and data analysis might, with further development, provide solutions to 'pinch points' in policymaking and policy management for poverty reduction. The focus of this paper is on exploring the suite of different data sources that can be used for measuring and investigating poverty. Our focus is on sources that can be used to help measure extreme poverty directly, as well as related attributes, such as determinants, proxies, or correlates of poverty. We adopt a consistent assessment framework to describe and compare different approaches and understand their attributes and limitations. We conclude by discussing to what extent new technologies can enable better insights into the incidence, distribution, and severity of extreme poverty over space and time, and where the most promising avenues for adopting these methods or investing in further research lie.

Table of contents

List of tables, figures, and boxes

List of abbreviations

CAPI	Computer-assisted personal interviewing
CATI	Computer-assisted telephone interviewing
CDR	Call detail record
CNN	Convolutional neural network
DEEP	Data and Evidence to End Extreme Poverty
DHS	Demographic and Health Survey
DMSP-OLS	Defense Meteorological Satellite Program's Operational Linescan System
EBP	Empirical best predictor
EVI	Enhanced Vegetation Index
FSPI	Food Staples Price Index
INLA	Integrated nested Laplace approximation
IPL	International poverty line
IVR	Interactive voice response
LMICs	Low- and middle-income countries
LSMS	Living Standards Measurement Survey
MNO	Mobile network operator
MPIs	Multi-dimensional poverty indices
MSE	Mean squared error
NDVI	Normalised Difference Vegetation Index
NLP	Natural language processing
NTL	Night-time lights
OSM	OpenStreetMap
PPP	Purchasing power parity
SAE	Small area estimation
SDGs	Sustainable Development Goals

Introduction

The post-2015 Sustainable Development Goals (SDGs) enshrine a global commitment to eliminate extreme poverty by 2030. While achieving this goal looks unlikely,[1] particularly in the light of the COVID-19 pandemic, progressing on this ambitious agenda will require the existence of suitable data to inform effective poverty reduction strategies and monitor progress. These evidence needs are substantial and vary depending on who the user of the data is, and their specific context and objectives. Yet, in general, they include describing the incidence and severity of extreme poverty at different levels of geography; how it is distributed across different areas or population sub-groups; the location of people and communities living in extreme poverty; how poverty status and severity varies over time; as well as its drivers and manifestations.

The availability of related data has improved enormously over the past 30 years (Serajuddin *et al*. (2015: 2). Traditionally, extreme poverty has been measured by comparing household *per capita* consumption expenditure to a US$ 1.90 per day international poverty line (IPL), using household survey data, price data, and census data. For example, large-scale surveys, such as the World Bank's Living Standards Measurement Surveys (LSMS), collect detailed information about poverty and wellbeing that can address a range of measurement needs. The volume of these kinds of large-scale surveys has increased markedly since the 1990s, with significant growth in the number conducted across low- and middle-income countries (LMICs).[2] This has been accompanied by a corresponding advancement in data collection methods for ensuring a high quality among such data, as well as systems for using available data to monitor the impact of policies and programmes designed to reduce poverty. Among the most transformational shifts in the past two decades has been the advent of computer-assisted personal interviewing (CAPI), in which interviews are conducted electronically rather than through paper questionnaires. This has led to a radical improvement in the quality, speed, and cost-effectiveness of survey data (Caeyers, Chalmers, and De Weerdt, 2012).

Yet, despite these gains, information still remains frustratingly slow to emerge. Household survey data tend to be expensive to collect, leading to a long lag in time between survey rounds.[3] The minimum gap between nationally representative household surveys is typically around three to five years (Alkire and Samman, 2014), while national censuses are carried out even less frequently. Around one in four LMICs have not conducted a census in the last 10 years (Thomson *et al*., 2020), and there are some countries (including Afghanistan,

---

[1] See, for example, https://datatopics.worldbank.org/sdgatlas/goal-1-no-poverty/ [accessed March 2021].

[2] For example, Serajuddin *et al*. (2015) find that the availability of household consumption survey data grew from a low coverage of 13 countries in the early 1990s to 40 countries in 2001, and to 62 countries by 2011.

[3] For example, the LSMS costs an estimated US$ 1.7 million per round (Sustainable Development Solutions Network, 2015) (based on a sample of International Development Association recipient countries). See Section 3.1 for further discussion of average survey-related costs.

Eritrea, Lebanon, and the Democratic Republic of Congo) that have not conducted a population census in at least 30 years.[4]

Geographically disaggregated information about poverty that may be required to inform intervention targeting or results monitoring at higher granularity is also hard to come by through traditional survey data. Estimates are usually representative to the national or first administrative level. This is due to constraints on survey sample sizes, among other parameters. Lack of access to disaggregated poverty information can frustrate efforts to monitor the distribution of poverty across or within administrative sub-units, which local administrations may require to tackle poverty in the areas under their jurisdiction.

Compounding these issues facing prospective users of poverty information is that the availability and quality of relevant data are often lower in the countries and contexts where the data may be most needed. In poorer settings, a combination of resource constraints, higher logistical challenges associated with data collection (such as lower-quality transport networks or a shortage of required equipment), or a lower concentration of necessary expertise, can all restrain the collection of quality data at frequent intervals. As a result, the gap in information may be the greatest in precisely those contexts where it is needed the most.

Consequently, attention is increasingly shifting to the potential to adopt new technologies outside of the domain of survey data collection methods, to help further improve the frequency, spatial granularity, and availability of poverty data, analysis, and reporting, as well as reducing reporting costs. The term 'data revolution', or 'all data revolution', is sometimes used to describe the implied transformative potential of new forms of data and analytical techniques for the development agenda (Independent Expert Advisory Group on a Data Revolution for Sustainable Development, 2014). This includes a focus on 'big data', which is a generic term used to refer to the huge expansion occurring in the volume and variety of data available through a range of digital or satellite sources. The hope is that, if put to effective and thoughtful use, big data have the potential to support a wide variety of different measurement problems related to the issue of measuring and understanding extreme poverty.

There is a rapidly expanding literature documenting innovative applications of big data and the 'data revolution' to generate new insights into extreme poverty and to shorten the time lag between data generation and actionable findings. Harnessing this potential effectively could contribute to realising the SDGs and informing more finely-tuned and adaptive policy responses. The need for accurate information about extreme poverty has been brought into even sharper relief by the profound challenges presented by COVID-19, as well as increasing awareness of how climate change will impact the poor. The COVID-19 crisis itself threatens to undermine progress in the global fight against poverty, pushing an estimated 71 million to 100 million people into extreme poverty in 2020 (World Bank, 2020). At the same time, as

[4] See https://unstats.un.org/unsd/demographic-social/census/censusdates/ [accessed 29 August 2020].

much of the world has adapted to remote ways of working, attention has also shifted towards alternative means of data collection that deviate from large-scale face-to-face surveys.

Nonetheless, the application of new methods to enhance poverty measurement and advance the SDG agenda remains an emerging field. Some innovations are still in their infancy and efforts to understand how to apply these techniques to best advantage, where the most promising avenues for future exploration lie, and what limitations remain are ongoing.

This paper provides a summary overview of the analytical methods and, in particular, data sources that are in use, or that are being explored, for generating insights into extreme poverty. We start with a brief overview of 'traditional' methods, by which we mean methods that rely on survey and census data. The focus is then on summarising approaches that explore other, more innovative, data sources. The core aim is to lay the foundation for a better understanding of how new technologies and data innovation may help to address different measurement needs relating to extreme poverty, and where gaps still remain. The main messages from this review are also summarised in a DEEP Research Insight.

It is important to emphasise that this review is explicitly not limited to examining methods for direct estimation of the prevalence of extreme poverty as defined by the IPL. Rather, as mentioned above, this review also includes approaches that can generate insights into poverty and extreme poverty more broadly. We include approaches that aim to capture the concept of poverty via different indicators (e.g. consumption, income, asset wealth), that look at components of poverty indicators (e.g. prices, certain types of assets), and that try to answer a variety of questions related to the prevalence, depth, and dynamics of poverty. The objective is to equip the reader with a sense of the breadth of the options that are being explored to suit different measurement needs, and the variety of innovations that are being implemented in this area.

We identify a number of potential key users of this review. The first group includes government officials and policymakers who might be interested in an improved understanding of the poverty situation at national or sub-national levels, in order to inform poverty alleviation strategies and policy design, and help target assistance and evaluate progress. The second group of potential key users of this review are donors and development partners working in support of poverty alleviation efforts. The third group of potential users consists of civil society groups, for whom the evolution of the data landscape and proliferation of new sources may offer expanded opportunities to hold decision makers to account and actively shape the way that poverty is measured and tackled in their contexts.

An additional objective for this paper is to help refine the agenda of the Data and Evidence to End Extreme Poverty (DEEP) Research Programme consortium, which is a UK Foreign, Commonwealth and Development Office-funded research programme designed to produce new, impactful research into global extreme poverty. This paper is being prepared as part of the inception stage of DEEP, to help characterise the current state of knowledge and gaps in extreme poverty measurement, in order to aid in identifying the best entry points for this consortium to help enrich anti-poverty efforts and contribute to the debate.

In Section 1 we provide a description of our approach to this review, which includes a short summary of our review methods. In Section 2 we provide a brief description of the analytical methods that are being used to extract insights from new data sources. In Section 3 we present a review of data sources, starting with one brief section (Section 3.1) on 'traditional' data sources: survey and census data. Section 3 continues with the main focus of the paper, which is a review of approaches that use new, innovative data sources, including big data. Based on our findings there, in Section 4 we discuss how and in what ways new methods could enhance the toolkit of measurement approaches, and where gaps remain. We also reflect on which approaches seem to hold particular promise for enabling new insights into extreme poverty in the future.

1 Our approach to this review

1.1 Key terms and the scope of this review

1.1.1 Defining extreme poverty

There are numerous ways to conceptualise and measure extreme poverty. This paper does not restrict attention to a single definition. This sub-section provides a general overview of different ways of defining extreme poverty and presenting poverty data that may be relevant to policymakers.[5] One of the most widespread methods for defining extreme poverty is to compare household *per capita* consumption expenditure to a reference poverty line, representing a minimum threshold below which it is not possible to achieve a basic standard of living. The current global standard is the World Bank's IPL, set at US$ 1.90 per day, expressed in units of purchasing power parity- (PPP-) adjusted local currency.[6] The IPL is the dominant poverty measure adopted by many international organisations in addition to the World Bank, and is the primary indicator used to track progress against the SDGs.[7]

The IPL has several advantages. Though experiences of poverty are understood to extend well beyond monetary resource availability alone, the IPL is helpful in highlighting a fundamental aspect of absolute need. It is designed to allow for comparisons over time, as well as between and within countries. And while the construction of poverty lines can present technical complexities in practice, the concept of the IPL is compellingly straightforward. It has an appeal as a key uniting concept around which international efforts to eradicate poverty (notably progress towards the SDGs) can be coordinated. Yet the IPL and the degree of influence that it commands has also been the subject of significant critique. A recent report by the United Nations Special Rapporteur on Extreme Poverty and Human Rights singles out reliance on this measure as contributing to a misleadingly positive global narrative on the trajectory of extreme poverty (Center for Human Rights and Global Justice, 2020). The IPL is benchmarked at a much lower level than many national poverty lines, meaning that 'one can "escape" from poverty without an income anywhere near that required to achieve an adequate standard of living, including access to healthcare and education' (OHCHR, 2020).

[5] There is a vast literature on the topic of poverty measurement. For a more authoritative discussion, please see the Atkinson Commission report (World Bank, 2017).

[6] The IPL has been periodically revised upwards since it was first established at US$ 1/day in the 1990 World Bank World Development Report (World Bank, 1990). The line was set at US$ 1.90 in 2011 PPP US dollars (its current value) in 2015. The US$ 1.90 cut-off was defined in relation to the national poverty lines of developing countries, and represents what that amount of money could buy in the United States in 2011. It is typically measured using data on household consumption and an estimate of the consumer price index to translate the IPL into local prices at the time of data collection.

[7] The IPL is an example of an absolute poverty line. This means that it seeks to benchmark a basic, fixed amount of goods and services that is considered the minimum required to attain a basic standard of living. Note that an alternative way of setting a poverty line is to define the threshold according to the relative distribution of income or consumption. This approach represents a different conceptualisation of poverty as a relative concept, where the metric of interest is how the standard of living compares to the distribution within a particular country at a particular time. Relative poverty lines are more often adopted in developed countries where levels of absolute deprivation are low: they are therefore not a focus of this paper.

Poverty lines, both national as well as the IPL, are generally defined using income or consumption expenditure data. This information can be difficult to obtain. Income is not straightforward to estimate in contexts where informal livelihoods dominate. Consumption information is usually used as an alternative to income, and consumption *per capita* is the definition favoured by the World Bank for estimating global poverty (World Bank, 2017: 7). Consumption per adult equivalent measures are also commonly used, to allow for comparison between households of different compositions. Yet consumption can also prove difficult to measure due to the time-consuming nature of standard survey consumption modules, and in contexts where prices (to estimate the value of household consumption) are difficult to obtain. Where income and consumption data are not readily available, alternatives include measurement based on asset or wealth indices. The Demographic and Health Surveys (DHSs) that are collected every four to five years in a range of countries measure a wealth index on the basis of ownership of a range of assets, household dwelling conditions, and access to basic services. The resulting indices are a proxy measure of wealth that correlate with poverty measures developed through full consumption measures, but are simpler and cheaper to collect.

Alongside monetary definitions of extreme poverty, there are also multi-dimensional measures that seek to encapsulate broader experiences of poverty extending beyond a lack of income. A key recommendation of the 2016 World Bank Commission on Global Poverty was that an array of additional indicators should be monitored alongside the IPL, in recognition of the fact that poverty is a multi-dimensional phenomenon (World Bank, 2017). This can be done either through presenting a dashboard of relevant indicators, or combining these indicators into a composite index. Multi-dimensional poverty indices (MPIs) are aggregate indicators that reflect an array of characteristics, spanning both monetary and non-monetary elements. These include aspects such as access to quality health and sanitation services, wellbeing, access to education, ability to claim human rights, and empowerment. Note that while the focus of this paper is on *extreme* poverty, MPIs are concerned with the related, but distinct, concept of *acute* poverty. While the IPL measure of extreme poverty captures the extent to which consumption levels attain a certain threshold, MPI measures of acute poverty are concerned with whether or not multiple minimum standards are simultaneously attained or not (Alkire and Santos, 2014).

Whatever specific indicator definition is chosen, there are a range of ways in which extreme poverty data can be analysed and presented. The simplest of these are headcount measures or proportions, which describe the extent of extreme poverty across some geographic area. Yet more granular information on the distribution and severity of extreme poverty across space and time is often desirable to inform policy questions and resource allocation decisions. For instance, the poverty gap is used to understand distributional characteristics of extreme poverty (Blumenstock, 2019). It measures the average shortfall from the poverty line (where those above the poverty line are counted as having no shortfall), expressed as a percentage of the poverty line (World Bank, 2017). It is helpful in any situation where an understanding of poverty severity below the poverty threshold is required. Other policy questions may require data about distributional aspects of poverty and inequality, requiring information on the percentiles of the income distribution, or inequality measures such as Gini coefficients, as well as data on poverty dynamics – on movements into and out of poverty, or further into poverty – by different groups and for different geographical areas.

1.1.2 New technologies for poverty measurement

The purpose of this paper is to present a review of approaches that explore the use of new technologies to generate insights into extreme poverty, especially those that enable new insights into the incidence, distribution, and severity of poverty over time and space. In terms of new technologies, the main focus is on sources of data for measuring extreme poverty, including big data sources. However, in Section 2, we also include a brief introduction of some of the analytical statistical methods that underpin extreme poverty measurement and the use of new data sources, and that will be referenced in the main review (Section 3).

Note that some of the data sources that we consider in this paper can no longer strictly be considered 'new'. For example, remote sensing data have been incorporated into environmental and social sciences research for over half a century (Landgrebe, 1986). In this sense, the use of these kinds of data is not in itself an innovation. However, we still include these in this review because these are areas where innovation is ongoing, where novel ways of drawing insights continue to be refined, and – in particular – where their application in measuring and investigating extreme poverty is relatively recent.

The review of data sources in Section 3 will follow the broad headings set out below:[8]

- **Household survey and census data:** We will begin with a review of the traditional forms of data collection that have previously been the backbone of poverty measurement approaches. This starts with household survey and census data.
- **Data exhaust**: Data exhaust is a form of big data. This refers to data collected passively as a result of people's interactions with digital services, such as mobile phones, transactions data relating to online purchases, and internet searches.
- **Online information**: This includes content from internet sources, such as news articles and social media interactions.
- **Remote sensing data**: This refers to data collected from satellites, infrared sources, and other kinds of earth observation sensors.
- **Citizen-reported or crowd-sourced data**: This refers to data that are reported directly by citizens that 'opt in' to do so, through specialised platforms.
- **Mobile phone surveys**: These are surveys conducted using mobile phone technology to capture questionnaire data from sampled households or units. They share similarities with traditional household surveys but the use of mobile phones offers different possibilities as well as limitations.

[8] This grouping is adapted from a classification scheme developed by the United Nations Global Pulse initiative (Global Pulse, 2012).

It is important to note that many innovative applications of new technologies involve the use of different forms of data and analytical techniques in combination – as a form of 'data sandwich' where different types of data are layered on top of each other. Such mixed approaches do not fall neatly within the aforementioned schema, but are also discussed in the review.

1.2 Review method

To conduct this review we have drawn on a wide range of literature relating to the analytical methods and data sources that are used to contribute towards poverty measurement. We have included evidence drawn from a variety of different sources, such as academic papers, journal articles, blog posts, and sections in books.

As discussed above, it is important to note that we have not restricted our attention solely to methods that are used to measure the US$ 1.90 IPL, rather we have considered approaches that investigate poverty measurement from different angles. More precisely, we have included approaches that define, measure, and estimate poverty in a variety of different ways, e.g. related to consumption, income, wealth, proxy measures, and asset ownership. We have also included approaches that try to tackle questions related to the estimation of specific intermediate indicators or information needed to estimate poverty, such as prices that individuals face in the markets in which they consume goods. Finally, we have included approaches that look both at static descriptions of the prevalence of poverty (i.e. estimate proportions at a given point in time), as well as poverty dynamics and the depth of poverty (e.g. looking at poverty gaps).

We have implemented this review in a step-wise fashion: our starting point was previous overviews and review papers, both related to poverty measurement (e.g. Alkire and Samman, 2014) and to the issue of new technologies and new data in international development more generally (e.g. Global Pulse, 2012). We then used other publicly available compendia of related literature to identify more recent papers and publications that describe approaches related to the issue of new data and poverty measurement, such as for example a reading list from a 'Big Data and development' course facilitated by Joshua Blumenstock at Berkeley University. In a final step, we submitted initial drafts of this paper for peer review to experts within the DEEP consortium to help identify any gaps or suggest key additional or new references to include. As such, this paper does not represent a systematic literature review. However, given this step-wise approach, we are confident we have captured the main trends and approaches that – at the time of writing – are being investigated with respect to new technologies and poverty measurement.

1.3 Analytical framework

We adopt a simple framework to aid the review of new approaches to poverty measurement covered in this paper. This consists of a set of 10 core criteria selected to capture different attributes of each approach that may be desirable for policymakers. We use this framework to guide our assessment and to help summarise the qualities, limitations, and opportunities represented by each approach. Not all criteria are equally applicable to all approaches; however, using this framework to summarise approaches is intended to help illustrate the differences between them. Our objective in adopting this framework is to highlight attributes and the possibilities afforded by different approaches, and to comment on where promising avenues for adopting these approaches or conducting further research lie, rather than to offer a definitive judgement on which approach performs best. Decisions regarding which approaches are most suitable for the objectives at hand may be guided by referring to this framework, depending on which criteria are considered of relatively higher priority by the decision maker to tackle the problem at hand. We refer back to this framework in Section 4, where we try to summarise the findings of our review.

The criteria that make up our framework are described in Table 1. These criteria have been developed by the authors based on a literature review, our informed opinion, and the feedback of the peer review experts who have contributed to this review. A key reference for this framework is Alkire and Samman (2014), whose work bears some similarity to our own and who also refer to a set of 10 (differently formulated) criteria to guide their assessment of approaches to monitoring poverty-related SDG indicators.

Table 1: Analytical framework

Criteria	Description and examples	Why is it important?
RELEVANCE	**To what extent does this method capture the information required to measure extreme poverty?**	Policymakers require relevant information in order to achieve their estimation goals, including monitoring progress in relation to the SDGs.
COST CONSIDERATIONS	**How expensive is it to carry out this approach, relative to alternatives?**	Cost is an important consideration in evaluating the uptake of different approaches, and the frequency with which they can be adopted.
FREQUENCY	**How often can poverty estimates be produced?** (Does the approach provide 'snapshot' information at distinct and limited points in time, or can the information be updated continuously to capture dynamic changes?)	Information that is updated infrequently may be out of date and therefore unreliable. Policy development or research objectives may require an understanding of how the rate and severity of poverty are evolving over time.
SPATIAL RESOLUTION	**At what geographical level can poverty estimates be generated through this approach?** (For example, can estimates be generated at the national level only? Or at regional, municipal, or even finer levels?)	Policy development or research objectives may require an understanding of how the rate and severity of poverty vary across space.
ACCESSIBILITY	**What constraints, if any, exist for policymakers in regard to accessing the required data, technology, or skills to adopt this approach?** (For example, are the data proprietary? Are high technical capacity thresholds or computer processing requirements	Accessibility is an important consideration in evaluating the potential for the uptake of different methods by national statistical agencies or other practitioners.

Criteria	Description and examples	Why is it important?
	needed to adopt the method? Are there legal frameworks governing the use of these data? What other barriers are there to uptake of this method, apart from cost?)	
DATA QUALITY	**What are the primary data quality considerations or concerns associated with this method?** (Data quality is a very broad criterion, and assessments need to be implemented for specific data sources;[9] however, of particular interest to us are issues of bias and precision.)	Data quality may be a greater or lesser priority depending on what level of precision is required, and whether or not there is a need to accurately detect small changes over space and time.
ABILITY TO MEASURE UNCERTAINTY	**Considering data quality issues, does the approach enable uncertainty in estimates to be reported?** (For example, can the direction and level of imprecision be described?)	Knowledge of the level of uncertainty is important for data interpretation and understanding the strength of evidence being used to inform decisions.
ABILITY TO MAKE OUT-OF-SAMPLE PREDICTIONS	**To what degree can this method be used to make predictions about units that are not in the sample?** (For example, can findings generated for one population be generalised to other regions, countries, or settings?).	The ability to generalise findings to a wider population may have policy relevance.
REPLICABILITY	**To what extent can the information generated through this approach be reused or replicated by others?**	Replicability may be desirable for purposes of transparency and accountability. Re-usability may also be considered important to enable additional value to be drawn from the data or information generated by other users.
ETHICAL CONSIDERATIONS	**Are there any particular ethical considerations associated with this approach?** (For example, are there issues of data protection associated with this method that need to be navigated? Or potential negative implications for certain groups of adopting this method?)	The advent of new technologies may introduce new forms of ethical risk that need to be well-understood and navigated.

[9] See for example the OECD Data Quality Assessment Framework here: https://stats.oecd.org/glossary/detail.asp?ID=4567.

2 A brief primer on statistical methods

Using, and extracting insights from, many of the new data sources that will be discussed in Section 3 often requires applying a variety of statistical modelling techniques that are derived from the 'statistical learning' toolkit. Defining statistical learning is tricky. Most broadly it can be defined as 'a set of tools for modelling and understanding complex datasets' (James *et al.*, 2013). This can encompass a wide variety of statistical techniques. For the present review, however, we employ the term 'statistical learning' as referring explicitly to methods that derive from more recent advances in the fields of computer science and machine learning.[10]

An assessment of opportunities and caveats related to the use of new data sources for the measurement of extreme poverty requires an understanding of these methods. Many of the approaches reviewed in this paper make use of such methods to investigate questions related to measuring or estimating extreme poverty using new data sources. In particular, they make use of them because of their strength in tackling prediction problems (James *et al.*, 2013: 6). In this section we therefore provide an overview of the main estimation problems that the approaches reviewed in Section 3 are trying to tackle, and we aim to give a sense of how they employ statistical learning methods. By doing so, we also aim to provide a guide to the terminology used in the literature reviewed for the purposes of this paper, which will re-appear in Section 3. Readers who have an understanding of these methods and the problems they try to tackle can skip directly to section 3.

2.1 An overview of statistical learning methods

In order to provide an overview of, and to attempt a categorisation of, statistical learning methods, a good starting point is to look at their target of estimation, i.e. the statistic, parameter, or indicator that they are trying to estimate. This target of estimation varies across the different pieces of empirical work reviewed in Section 3: statistical learning is not only used to predict the proportion of the population living below the IPL, it is also used to tackle a variety of different estimation tasks. As discussed in Section 1.1, this is driven by the nature of the domain area this review is focusing on, as there are a variety of different ways in which poverty and extreme poverty can be defined (e.g. threshold of consumption expenditure vs multi-dimensional poverty), a variety of different indicators and variables that need to be measured to compute poverty estimates (e.g. market prices, household sizes, consumption levels), and a variety of different estimates that policymakers and researchers are interested in with respect to poverty (e.g. changes over time, geographical distribution, poverty gaps, inequality). Which of these definitions, indicators, variables, and estimates an

[10] It is difficult to pin down a clear definition of machine learning, but generally the term refers to methods that train a model (an algorithm) for prediction purposes by some automated, iterative process. The goal is not so much to identify parameters in an assumed data generation process that defines a model, but to let a computer (the machine) decide how to best fit the data for a certain prediction problem (Athey and Imbens, 2019).

approach will focus on when employing statistical learning methods will depend on the application context, the data sources available, and the questions that researchers or policymakers will want to focus on in a particular context.

Importantly, however, depending on the chosen focus, the analytical options and choices from the statistical learning toolkit will differ. For example, analytical approaches are sometimes categorised by whether the target estimate is a categorical variable (e.g. whether a household is poor or not based on an IPL cut-off), in which case they are categorised as ‘classification problems’, or a continuous variable (e.g. the level of per adult equivalent household consumption), in which case they are categorised as ‘regression problems’ (James *et al.*, 2013: 28). Similarly, estimation problems are categorised into ‘supervised’ vs ‘unsupervised’ learning problems. Supervised learning problems relate to situations where, for at least part of the data one has available, both outcomes and covariates for a particular situation are observable. Typical supervised learning problems are prediction problems that involve using the available observed data on the outcome variable and covariates to predict the outcome variable for unobserved objects. Almost all of the use cases of using new data sources that we review in Section 3 are supervised learning problems, e.g. trying to predict poverty or wealth status using a set of geospatial covariates. Unsupervised learning refers to situations where there is no outcome variable. Typical unsupervised learning problems are clustering analyses, e.g. to identify groups of observations that are similar to each other based on a set of descriptors (covariates) only. In our review, we find very few examples of unsupervised learning methods being employed.

The approaches reviewed in Section 3 employ statistical learning methods both for the ultimate purpose of estimating or predicting the desired outcome but also to prepare data from new sources so that they can then – in a second step – be used for analysis. The approach of Sheehan *et al*. (2019) is one example, where a natural language processing (NLP) algorithm (i.e. the quantitative analysis of text data) is used first to deal with unstructured text data from Wikipedia that can then, in a second step, be used to predict the outcome (Sheehan *et al*., 2019) Another example is Jean *et al*. (2016), who first use a neural network algorithm to estimate or construct covariates from satellite imagery, which are then used in a subsequent prediction step *(Jean et al*. 2016) Hence, it is important to emphasise that in many cases where new technologies provide large amounts of unstructured data – e.g. images and text – statistical learning techniques are required to prepare these data for further analysis with supervised learning algorithms.

Another important categorisation of the models that are employed refers to the difference between analytical and descriptive inference. Analytical inference refers to the aim of understanding the relationship between certain covariates and a certain outcome. For example, this could be to understand how certain household characteristics are related to extreme poverty. Descriptive inference problems, on the other hand, relate to situations where one wants to estimate a population parameter (e.g. the proportion of people in extreme poverty in a country or in small geographical areas within a country) but where one is not necessarily interested in understanding the relationship between other variables (covariates) and this outcome (James, 2013: 18ff.) A significant part of the statistical learning and machine learning literature focuses its efforts on developing algorithms and using data to improve our ability to estimate descriptive population parameters. In fact, a large majority of the approaches reviewed in Section 3 explore the possibility of using new data to estimate such parameters relating to poverty, rather than performing analytical inference on, for example, drivers of poverty.

Traditionally, descriptive parameters that policymakers or other stakeholders are interested in at the national level are estimated as part of national statistical institute operations by using survey and administrative data. This typically does not involve the use of statistical learning models. In fact, machine learning and statistical learning algorithms depend on model assumptions and, traditionally, there has been a tendency to avoid relying on such assumptions for producing official statistics that determine important policies. However, with the advent of large sets of data, expanded computational capacity, and the ability to obtain high spatial resolution and high-frequency statistics, machine learning and statistical learning have become a lot better at dealing with estimation or prediction problems, which makes investigations into their use for descriptive inference relevant (Athey and Imbens, 2019).

As our review in Section 3 will show, however, in many use cases the choice of a particular statistical learning method for the final task of predicting poverty indicators seems to have a lesser impact on the quality of predictions than the choice of other factors (e.g. the type of outcome measure predicted). In fact, well-known methods from the statistical learning toolkit that do not rely on complex machine learning algorithms (e.g. linear or logistic regression models), and that are often used even in recent applications and method comparison exercises, suggest that more complex tools are not always needed (see Engstrom, Hersh, and Newhouse (2017); Head *et al.* (2017), and Li *et al.* (2019), among others).[11] Yet it cannot be discounted that new machine learning techniques may produce significant improvements in specific situations. In particular, some literature has highlighted the relevance of taking into consideration a spatial dimension for the case of wealth and health indicators[12] (Okwi *et al*., 2007; De Sherbinin, 2011; Sedda *et al*., 2015; Osgood-Zimmerman *et al*., 2018).

[11] Other machine learning and statistical learning approaches that have been used for the aim of poverty prediction in the literature include Gaussian processes and hierarchical logistic regression, under a Bayesian framework of inference, with or without spatial component, random and rotation forests, support vector machines, and K-nearest neighbour regression.

[12] With respect to uncertainty estimates, except for a few applications, mainly those based on small area estimation models or regression models estimated within the Bayesian framework, no uncertainty estimates for the area-specific point estimates are typically provided in the reviewed literature. The assessment of the predictive capability of a given method is limited to the estimation of an average metric of overall performance obtained through cross-validation. Examples of these metrics are the coefficient of determination (R2), a mean squared error (MSE) of prediction, the area under the curve of a receiver operating characteristic (ROC) curve, or, for feature extraction tasks, a type of Jaccard index or intersection over union coefficient.

2.2 The data gap problem

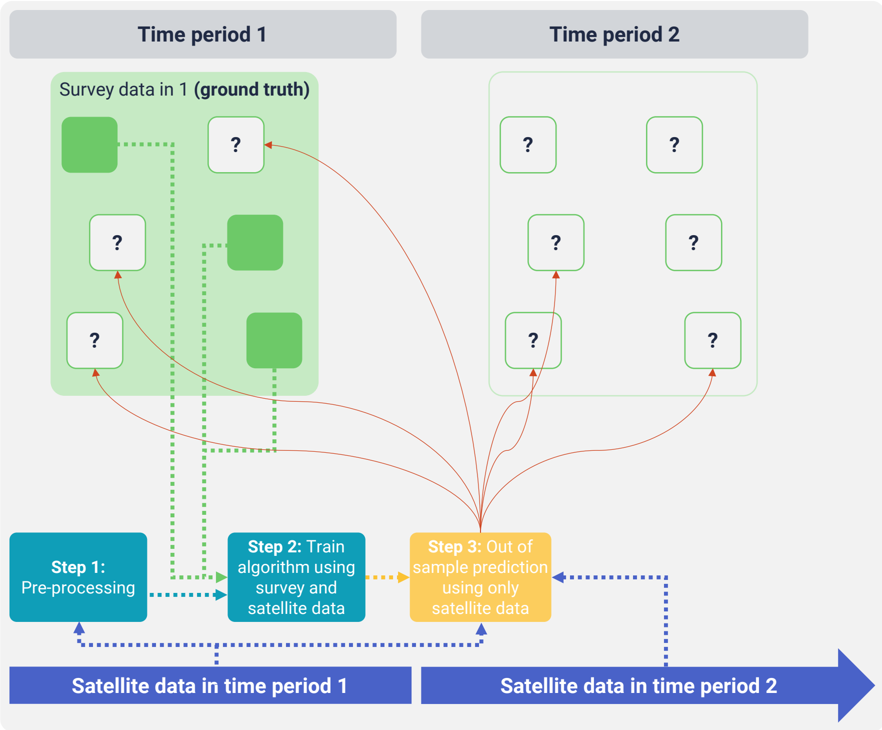
The fact that statistical learning methods together with large sets of new data are seen to perform well at prediction tasks motivates a large part of the use cases of new data sources presented in Section 3. In essence, in many cases the objective is to test whether the predictive capacity of these statistical learning methods, together with new data, can help to address the issue that poverty estimates are only available intermittently and for larger geographies – something we label the 'data gap' here. We present a schematic representation of this idea in Figure 1 using survey and satellite imagery data as examples of the data sources available.

Assume that the large green box in Figure 1 represents a country, for which survey data are available in time period 1 but not in time period 2. Survey data are based on a sample from the three green areas within the country, which means that it is possible to compute an overall descriptive estimate – for example, the average poverty rate – for this country without recourse to statistical learning models. Satellite imagery data are available for the whole country and in both time periods, 1 and 2. The idea is to use statistical learning methods to process satellite imagery data and to prepare those data for analysis (Step 1 – pre-processing). The aim is then to combine those data with survey data from known areas in time period 1 to train an algorithm (Step 2) that predicts the survey outcome using *satellite imagery data alone* (Step 3). The observed survey and satellite imagery data used to train the prediction algorithm in Step 2 are referred to as the 'training data', while the survey data are sometimes called **'ground truth'**, representing the 'truly' measured outcomes.

Because satellite imagery data are available for the entire country and continuously, i.e. in time period 2 as well, they can be combined with this algorithm in order to predict the outcome, either spatially (in time period 1) to previously non-sampled areas or spatio-temporally (in time period 2) to a new time period. Hence, using satellite imagery and statistical learning, the data gap is closed. This is a simple schematic representation of how the use of new data sources is being investigated to tackle the data gap problem in poverty measurement. The success of these approaches depends on the quality of the training data and on how well the satellite data and algorithms perform in predicting the outcome of interest in time periods and areas where survey data are not available or sample sizes are small (see next section).

Another way of framing this problem is to describe it as the challenge of how to produce high-resolution and high-frequency estimates of poverty. 'High-resolution' refers to producing estimates for small areas – such as e.g. the small rectangles within the country in Figure 1 or even smaller geographical areas. 'High-frequency' refers to the objective of producing estimates for time periods in which survey data are not available, e.g. years in between survey rounds, months of a year, or even smaller time periods. As mentioned, a large section of the approaches reviewed in Section 3 aim to produce such high-resolution or high-frequency estimates using new data sources. Given this importance, we provide a bit more detail on the problem of producing high-resolution estimates here.

Figure 1: The data gap problem



2.3 Small area estimation

Surveys, using probability samples, are designed to provide estimates with acceptable precision at national and first-level sub-national administrative levels (e.g. regions or provinces) but usually have insufficient sample sizes to allow for precise estimation at lower levels of aggregation (e.g. municipalities). In Figure 1, this could be seen as surveys being able to provide estimates at the overall country level in time period 1 (the large green box), but not at the level of the different small boxes within the country that might, however, sometimes be of interest to policymakers. This can be referred to as the specification of the geographic target level (spatial resolution) of a survey. Methodological approaches known as small area estimation (SAE) aim at producing estimates at a lower level of aggregation than the usual geographic target level of surveys.

Tackling issues like the one described above is not a new problem. The conventional use of small area methods only requires access to the 'ground truth' data offered by surveys and census data, i.e. what we call traditional sources of data. In contrast, as will be shown in Section 3, a large section of the literature that investigates the use of 'non-traditional' data for poverty estimation explores whether a new feasible data scenario for achieving SEA is one that assumes a combination of new data sources, such as satellite and mobile phone call detail record (CDR) data, with traditional ground truth data, mostly coming from surveys.

SAE is a class of approaches predominantly employing model-based methods to generate high-resolution estimates. Below, we provide an overview of how model-based SAE works. The intuition behind model-based SAE methods comes from the idea that survey and population data have complementary qualities. While census data do not tend to collect information on all outcomes of interest, such as poverty estimates, their core advantage is their ability to disaggregate information over any identified population group and spatial scale without the need for further modelling. By contrast, while survey data only achieve partial population coverage, and their spatial disaggregation is constrained by available sample sizes, they provide a rich source of outcome information. SAE methods seek to exploit these complementarities by combining these data sources, thus enabling estimation of the population-level poverty outcomes at fine levels of spatial granularity.

To understand how this works, consider the following highly stylised example (Figure 2 and Figure 3). Suppose the objective is to develop a high-resolution poverty map comprising areas A, B, and C (for example, these could be districts within a larger region). A recent survey exists which measures the poverty status (outcome Y) or a proxy of sampled households in areas A, B, and C, along with a set of covariates – X_1, X_2, and X_3 – correlated with poverty (for example, household size, the dependency ratio, and education). Now, suppose that a population census is available. The census measured the same covariates – X_1, X_2, and X_3 – that were also collected in the survey, for all households in the population. However, it did not collect the poverty status (as indicated by the dotted lines and green highlights in the right-hand panel of Figure 2 below). A more common scenario, especially when interest is in estimation at very low spatial levels, is where survey data are not available for some of the target areas. This situation is described in Figure 3 with area C being 'out of sample' in the survey, so neither the poverty outcome nor the covariates are available from the survey data for that area.

Figure 2: Stylised example of how SAE methods combine data sources (all areas in survey)

Survey data

Area	*HH*	*Outcome (Y)*	X_1	X_2	X_3
A	1	●	●	●	●
A	2	●	●	●	●
B	3	●	●	●	●
B	4	●	●	●	●
B	5	●	●	●	●
C	6	●	●	●	●
C	7	●	●	●	●

Census data

Area	*HH*	*Outcome (Y)*	X_1	X_2	X_3
A	1		●	●	●
A	2		●	●	●
B	3		●	●	●
B	4		●	●	●
B	5		●	●	●
C	6		●	●	●
C	7		●	●	●

Figure 3: Stylised example of how SAE methods combine data sources (with out-of-sample areas)

Survey data

Area	*HH*	*Outcome (Y)*	X_1	X_2	X_3
A	1	●	●	●	●
A	2	●	●	●	●
B	3	●	●	●	●
B	4	●	●	●	●
B	5	●	●	●	●
C	6				
C	7				

Census data

Area	*HH*	*Outcome (Y)*	X_1	X_2	X_3
A	1		●	●	●
A	2		●	●	●
B	3		●	●	●
B	4		●	●	●
B	5		●	●	●
C	6		●	●	●
C	7		●	●	●

Generally speaking, the first step of the SAE process is to use the survey data to fit a model using the outcome Y and the covariates X. The objective is to estimate a function that relates the observed outcomes and the covariates, and to then use the estimated function parameters to predict outcomes:

$$\hat{Y} = \hat{f}(X) \quad (1)$$

The estimated model parameters or function $(\hat{f})$ can now be applied to the covariate data in the census to predict the outcome for each unit (e.g. household) in the population, including for those that were not included in the survey. Note that (1) can also be used to predict the outcome for households in area C under the scenario described by Figure 3 above, even if the estimation of the model parameters does not involve data (training data) from area C. Hence, in this case we assume that the function we estimate with the data from areas A and B also holds for area C – this is out-of-sample prediction. A final step uses the estimated model parameters to compute an estimate of the target population parameters (e.g. the extreme poverty rate) in areas A, B, and C at an aggregate level.

The function specified in (1) can take a variety of different forms. In its simplest form, this can be a linear model. It can be specified either at the unit – i.e. household – level or at the area level, depending on the availability of auxiliary information. Much of the literature has focused on the use of mixed (random) effects models for SAE. Nevertheless, any statistical model or machine learning algorithm can be used to predict Y, including novel statistical learning approaches that are included in our review in Section 3. Irrespective of the model one uses, it has to allow for an approach that can measure the precision of the estimates at the chosen level of spatial resolution.

This simple example above serves to illustrate how surveys and population data may be combined to generate finely grained poverty estimates. The reliability of estimates generated in this way depends on the quality of the input survey data (the labelled 'ground truth' data), the population data that are used to project the outcomes, and the models chosen to estimate equation (1). The long time lapse between, for example, census rounds means that the covariate information used to make such projections may frequently be out of date, which has two key implications. First, with census data alone 'high-frequency' SAE is not possible. Second, the quality of SAE output might suffer due to the outdated nature of the census data. This implies a need to perform SAE prediction more frequently and – possibly – better, by using information from alternative and newer data sources than traditional censuses.

3 A review of data sources

3.1 Household survey- and census-based methods

Key messages:

- Traditional data sources – chiefly in-person surveys and censuses – currently form the backbone of efforts to measure poverty across the world.
- However, they are expensive and challenging to implement at a high quality, with nationally representative surveys easily costing between 1 and 2 million US dollars.
- This means data collection does not happen very often, with censuses implemented roughly once every 10 years and sample surveys every three to five years in many LMICs. This contributes to the 'poverty data gap', as it means data are often out of date.

Conventional methods for estimating poverty are based on household sample surveys or census data. Household surveys are questionnaires conducted by trained enumerators for a sample of individuals or households. Nowadays, most surveys rely on electronic tablets or mobile devices to record answers.[13] While household surveys involve a sample, census data cover an entire population of interest. This includes national censuses, which cover the entire population of a given country.

The 2017 World Bank Report of the Commission on Global Poverty sets out the traditional approach to measuring extreme poverty in terms of three main elements. The first is the collection of consumption and expenditure information from survey respondents. A leading example of a large-scale household survey including a detailed consumption module is the LSMS, which has been conducted in around 40 countries to date since the early 1980s (World Bank, 2017). The second element is data on domestic prices, which are used to adjust consumption expenditure *per capita* in terms of purchasing power. These price data may come from other sources, such as international efforts like the United Nations International Comparison Programme (ICP). The third element is the poverty line against which *per capita* consumption estimates are compared to identify households falling into extreme poverty. A final, additional, component of the approach to measurement is population estimates from national censuses, which may be used to estimate the absolute number of individuals estimated to experience extreme poverty (World Bank, 2017).

Survey data permit a relatively high degree of flexibility in terms of content. Questionnaires can in theory be tailored to capture a wide variety of information to suit the specific research or policy objectives at hand. This means that survey data can collect information directly on key variables of interest needed to measure various poverty estimates, including consumption and asset-based measures, multi-dimensional indicators, headcount rates, and poverty gaps. This is a core advantage of survey methods over other approaches that can only achieve

[13] Note that this section is solely concerned with surveys that are conducted in-person, for example by enumerators visiting respondents in their homes or meeting at a central location for interview. In Section 3.6 we also review surveys conducted remotely, through mobile phone devices or other remote means.

measurement of poverty proxies. Longitudinal surveys, in which the same participants are surveyed repeatedly at different time points, can provide particularly valuable insights into poverty dynamics over time. This information is vital for understanding patterns of change, whether and to what extent the same households may fluctuate in their poverty status over time, and the factors that determine persistent poverty status or movements in and out of poverty.

Compared to survey questionnaires, censuses are much more restricted in what information they can collect. Questionnaires are usually highly targeted at a limited number of indicators, given the large number of interviews that need to be carried out to cover whole populations. We are not aware of any censuses that have gathered detailed income or consumption information for the purposes of consumption-based poverty estimation. Measurement of non-income poverty or proxy measures based on a limited set of questions may, however, be possible. To enhance the utility of census data, they can be used in conjunction with survey data to help estimate poverty at a fine spatial resolution using SAE techniques (see Section 2.3).

The core advantage of census data, which makes them valuable in regard to SAE techniques, is the fact that information can be disaggregated over any population group that is identified within the data without the need for further modelling. This is of course contingent on the potential disaggregating covariates that exist within the data, which may be limited. In the absence of any recent census, the extent of disaggregation that can be carried out using survey data alone is limited by the sample sizes of available sub-groups within the data. These sample sizes, and the sampling strategy employed by the survey, exert an influence over the precision of estimates derived from survey data.[14] The level of precision associated with survey estimates can be represented by confidence intervals, which describe the margin of error surrounding a given statistic. This ability to model precision through confidence intervals and other parameters is important for helping data users to interpret survey estimates and understand the likely degree of uncertainty.

A disadvantage of traditional household survey and census-based methods for extreme poverty measurement is the high cost of collecting these data. Costs can vary widely between countries and surveys, depending on the sample size, questionnaire design, and associated logistical costs in the country (such as transport and wage costs). The Sustainable Development Solutions Network has estimated the average costs of DHS surveys at US$ 1.6 million, Multiple Indicator Cluster Surveys at US$ 1 million, and LSMS surveys at US$ 1.7 million, for 77 countries receiving the most aid (Sustainable Development Solutions Network, 2015: 19). These high costs can be associated with the low updating frequency, with census data collected roughly once every 10 years (Steele *et al*., 2017) (and sometimes much less: for

[14] Whenever a sample is drawn from a wider population to estimate a population parameter there is a loss of precision. This means that there is some chance that the value of the discrepancy found in the sample does not equal the ‘true’ value found in the population. Larger sample sizes are generally associated with higher precision. In other words, as the sample size increases, the sample mean should become increasingly close to the ‘true’ population mean.

example, in Angola the most recent census before 2014 was conducted in 1970 (Blumenstock, Cadamuro, and On, 2015)), and sample surveys every three to five years. Where a long time lag elapses between survey rounds, the best poverty information available in many countries to make resource allocation decisions can be outdated and inaccurate. The degree of periodicity in survey initiatives measuring poverty indicators varies between regions. Many countries have made substantial investments in survey initiatives, including large-scale panel surveys that collect updated estimates more regularly. For example, the large-scale SUSENAS survey in Indonesia has reported consumption and expenditure data twice per year since 2015.[15]

The volume, depth, and variety of information that can be collected through surveys is not without limit. Interviews need to be kept to within a reasonable time limit so as not to overburden respondents, who are typically not remunerated for their time. Questionnaires are generally highly standardised in order to facilitate aggregation, making them less well-suited to exploring more intangible information that may be relevant to MPI measures (such as subjective wellbeing, self-confidence, and ability to claim rights). The quality of survey data can vary, and will depend on how survey questionnaires are designed, how sampling is carried out, and how surveys are implemented. A lack of reliable and up-to-date sample frames is a key issue for many household surveys.

3.1.1 Summary

Table 2 provides a summary of household survey and census data, according to the analytical framework criteria set out in Section 1.2.

Table 2: Summary of survey and census-based sources

Criteria	Summary comments
RELEVANCE	A key strength of survey-based methods is the flexibility they provide to capture a variety of information related to poverty measurement. This includes the possibility of constructing direct measures of extreme poverty, such as those that are based on consumption or asset information, as well as proxy indices. They can help measure simple prevalence indicators, as well as poverty gap measures. Traditional methods are also well-suited to collecting information on a variety of other factors relevant to understanding multi-dimensional experiences of poverty, and its determinants or correlates (notwithstanding some caveats relating to data quality and limitations associated with the measurement of nuanced or less tangible factors through standardised surveys).
COST CONSIDERATIONS	The cost of survey data collection varies depending on the sample size, questionnaire design, and associated logistical costs in the country (such as transport and wage costs). However, surveys are generally considered an expensive form of data collection. Costs to conduct nationally representative surveys can range from US$ 1 million to US$ 1.7 million.
FREQUENCY	Nationally representative household surveys are normally carried out every three to five years, while censuses are conducted around once per decade.

15 See https://mikrodata.bps.go.id/mikrodata/index.php/catalog/762

Criteria	Summary comments
	There are some countries with very limited relevant survey or census data available at all.
SPATIAL RESOLUTION	Poverty estimates derived from sample surveys are usually available either at national level or at the level of the first administrative unit. Finer levels of spatial granularity are less common due to the demands on sample size. Census data achieve high spatial granularity, and, when combined with survey data, they have the potential to provide poverty estimates at low geographical levels.
ACCESSIBILITY	The majority of countries possess the required capacity to conduct surveys. International best practice standards for the design, execution, and analysis of survey data are well documented. For example, standard modules exist for the collection of poverty statistics that can be easily adapted to suit local conditions, together with detailed guides on how to administer the survey questions and treat the raw data. This helps lower the constraints on rolling out surveys, in terms of skills and capabilities, for countries or institutions that have not conducted them before. Nonetheless, administering surveys to a high standard of quality still requires relatively diverse expertise, which some countries or statistical agencies may not always be able to procure reliably.
DATA QUALITY	The quality of survey and census data is highly variable across different surveys. This is to some extent within the control of survey implementers and a range of best practice standards exist to improve the rigour of survey data collection, and of the statistical methods for reporting poverty estimates from these data. However, a number of known challenges to the quality of raw data exist, including a lack of reliable sample frames, human error, respondent recall, and the influence on survey responses caused by the act of being surveyed (sometimes called Hawthorne effects).
ABILITY TO MEASURE UNCERTAINTY	Quantitative estimates derived from surveys where a probabilistic sampling process has been adopted can be assigned a confidence interval to describe the precision around the estimate. However, non-sampling errors, such as recall bias, social desirability bias, or Hawthorne effects, cannot be readily measured in a quantitative sense.
ABILITY TO MAKE OUT-OF-SAMPLE PREDICTIONS	If surveys use a representative sampling approach, estimates derived from the survey can be generalised to describe the wider reference population (with a given confidence interval). If an up-to-date sample frame does not exist, survey data may cease to be truly representative of the intended population of interest, weakening the ability to draw inferences from the sample statistics to any wider population. Bias may also be introduced through systematic exclusion of some population groups from the survey or census. This includes households living in insecure or remote locations that are hard to reach, or populations that are otherwise difficult to track (for example, pastoralist communities or seasonal migrants).
REPLICABILITY	Online repositories for survey data exist, through which data can be downloaded and used by other parties, subject to some standard conditions governing their use (such as ensuring that data are appropriately cited). However, a significant volume of survey data that could be further analysed are not in the public domain. This can be due to disclosure risk concerns, or a lack of incentives for data producers to make their data publicly available. Some of the data that are publicly available are also difficult to use due to a lack of comprehensive documentation or meta-data.
ETHICAL CONSIDERATIONS	There is an extensive literature describing the ethical considerations relating to primary survey data collection, results reporting, and data storage: for example DFID (2020).

3.2 Remote sensing data

Key messages:

- Remote sensing refers to data that are collected from satellites, planes, and other kinds of earth observation sensors. Attempts are being made to combine remote sensing information with surveys in order to predict poverty, and to close the poverty data gap.
- Features that researchers have explored in this area are night-time lights, vegetation indices, geographic conditions, climate, and characteristics of dwellings.
- To identify features that are relevant to predicting poverty, recent approaches have employed 'transfer learning' analyses, where algorithms are left to their own devices to extract such features. Innovations in this area are happening fast, and methodological advances in statistical learning are feeding into them.
- Overall, given their low cost, high spatial resolution, high coverage, and the frequency with which they are updated, data from remote sensors offer considerable potential for our efforts to measure and address extreme poverty.

The wide availability, wide coverage, low cost, and high frequency of remote sensing data – data obtained either through aircraft or satellite-based sensors – has increased the interest in exploring its use for monitoring the SDGs, particularly in the case of LMICs, where the data gap problem might be most pertinent. For some SDGs, relevant indicators can be directly derived from remote sensing data. Land cover, crop identification, and crop yield, which can contribute to the monitoring of goals such as 'food security' and 'life on land' are area-level indicators that illustrate this situation (Holloway and Mengersen, 2018). In contrast, the use of remote sensing data for the monitoring of poverty-related indicators is more complex, as the latter typically describe individual or household-level characteristics associated with asset ownership, income, or consumption that are not directly visible from space, and hence cannot be derived from satellite imagery alone.

The amount of literature exploring the use of variables derived from remote sensing for poverty monitoring has increased steadily in the last 10 years. As described in Section 2, the majority of current applications see this as a supervised learning problem, where the goal is to obtain a statistical model or to train a machine learning algorithm to predict the value of a given welfare indicator, typically obtained from survey samples and other traditional sources, conditioning on a set of covariates – in the statistical learning literature also called 'features' – extracted from remote sensing data. Estimates of welfare and poverty indicators derived from traditional sources, mostly surveys, also play the role of the 'ground truth' against which the performance of alternative methodologies is assessed. However, given the comparatively limited availability of poverty data from surveys for addressing the learning task in supervised learning approaches in some contexts, a very recent strand of research has also focused on the use of semi-supervised or unsupervised methods for poverty prediction. New developments such as those in Jean, Xie, and Ermon (2018) Perez *et al*. (2019), and Khan and Blumenstock (2019), are promising and will be discussed in this section as well.

3.2.1 Poverty and welfare indicators, features, and geography

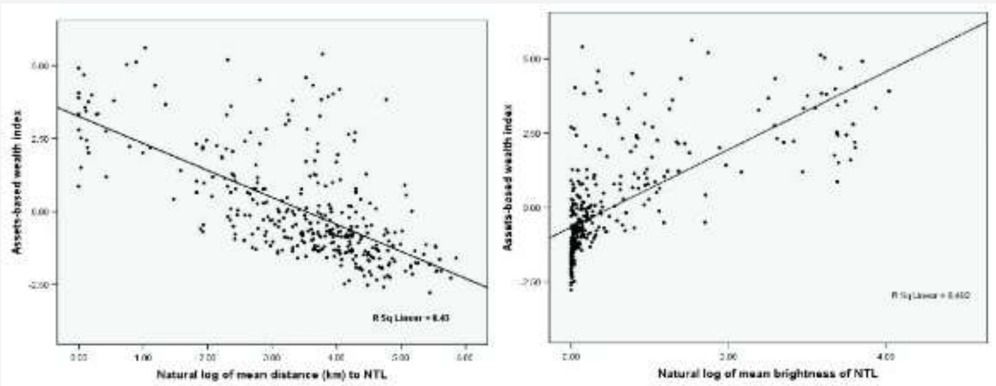
An overview of the literature suggests that success in the use of remote sensing data for poverty monitoring purposes relies heavily on the appropriate choice of the triplet of the poverty or welfare indicator of interest, the set of covariates (e.g. night-time lights (NTL), vegetation coverage, elevation), and the level of geography (see Box 2). For instance, NTL that can be easily obtained from low-resolution satellite imagery have shown high predictive power for several types of welfare indicators at national and rough sub-national levels, enabling the generation of global maps of estimates which allow easier cross-national and temporal comparisons than those based on country-specific poverty thresholds (Elvidge *et al.*, 2009; Noor *et al.*, 2008). However, the capability of NTL on their own to identify poverty at more disaggregated levels of geography, as well as in urban areas and for more developed economic societies, has been considerably lower.

Richer sets of features including NTL as well as landscape features (such as elevation and slope of the terrain), accessibility indicators (e.g. distance to roads or population centres), and climate or environmental variables (such as temperature, precipitation, and vegetation indices), have been shown to explain around 40–60% of the variability in asset-based wealth indices at sub-national levels for LMICs. Similar results have been obtained for income or consumption indicators. On the other hand, attempts at producing estimates of anthropometric and health-related indicators using these same variables, or even including features obtained from daytime satellite imagery, have been much less successful (Head *et al.*, 2017) (see Box 1).

Box 1: Illustrating differences in predictive power depending on the choice of welfare indicator, features, and geography

Figure 4 shows the relevance of NTL for the prediction of a DHS wealth index across 37 countries in Africa at a relatively high level of aggregation (admin level 1, typically the equivalent to state). Notice that finding an appropriate representation of a given covariate matters as both ‘mean distance to NTL’ and ‘mean brightness of NTL’ have similar levels of correlation with the outcome variable, but the former seems much more able to discriminate between levels of wealth or poverty than the latter.

Figure 4: NTL vs DHS wealth index for 37 African countries at admin 1 level



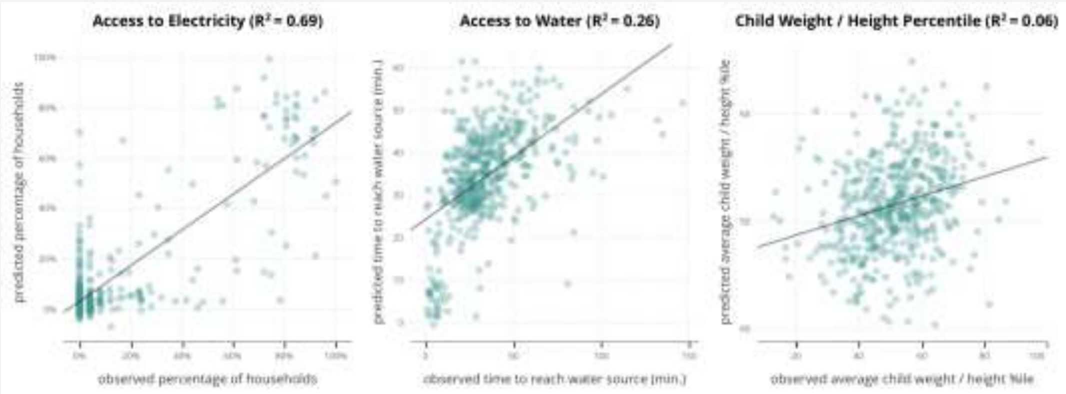
Source: Noor et al., 2008.

Head et al. (2017) assess the use of different approaches to extract relevant features from daylight satellite imagery for the prediction of a welfare indicator in five African countries. Figure 5, taken from this study, illustrates how an off-the-shelf methodology can show varied ranges of success

Box 1: Illustrating differences in predictive power depending on the choice of welfare indicator, features, and geography

depending on the welfare indicator of interest. It shows how predicted values for each welfare indicator compare to the 'ground truth', i.e. the actual values from the corresponding DHS survey, each time estimated using the same methodology. The explained variation (R^2) varies between 0.06 and 0.69 for these individual indicators. So – using the same analytical method and data sources – predictive power varies considerably, depending on the outcome looked at. Also, notice that, as in Figure 4, a relatively high predictive power measured by the R^2 does not necessarily indicate ability as regards identifying the areas in the lower tail of the distribution of the welfare indicator.

Figure 5: Use of the 'transfer learning' approach for several type of welfare indicators at village level in Rwanda



Source: Head et al., 2017. X axis: 'ground truth' estimate obtained from DHS 2010. Y axis: Prediction obtained from remote sensing features using a ridge regression model.

In contrast to the very rich sets of features that can be obtained from remote sensing data with existing methodologies, the relatively scarce availability of poverty data from surveys for algorithm training purposes (see Step 2 in Figure 1) remains a significant hurdle for the prediction of poverty (Burke *et al*., 2020). As mentioned before, training data in the form of poverty estimates (sometimes called 'labelled data') are generally obtained from traditional data sources, such as household surveys, and hence are available just for a modest subset of the areas in a country, and for a few time points in the best of scenarios. Furthermore, very granular survey estimates that would be required for training algorithms at high levels of disaggregation are often highly imprecise. On the one hand, there is a risk of incurring on ecological fallacies and spurious associations if accurate estimates at higher levels of aggregation are used. On the other hand, the additional noise at lower geographical levels seems to translate into reduced predictive power for similarly noisy training data (Yeh *et al*., 2020).

Several alternatives have been proposed to circumvent this problem of scarce poverty data, among them: the 'transfer learning' approach (Xie *et al*. 2016), which uses an intermediate training problem for feature extraction (see Box 3 for more detail); the use of SAE methods to produce a complete starting set of estimates for a given geographic resolution (Engstrom *et al*., 2017); and the pooling of available estimates corresponding to a large set of different countries and time points (Yeh *et al*., 2020). The use of semi-supervised and unsupervised methods to leverage the small sets of available poverty data is a promising area of research, and hence will be discussed in more detail below.

3.2.2 Detail on relevant covariates from remote sensing data

Several variables, covariates, or features that can be extracted from remote sensing data have shown high predictive capacity for welfare indicators, such as DHS wealth indices, as well as indicators of consumption and expenditure. A summary of the most common features extracted from remote sensing data which are relevant for the prediction of welfare indicators is presented below.

NTL

Data on NTL are typically obtained from satellite imagery provided by the US Defense Meteorological Satellite Program's Operational Linescan System (DMSP-OLS) and processed by the US National Oceanic and Atmospheric Administration. With an approximate resolution of 1 km resolution, DMSP-OLS NTL data have been available worldwide since 1994.

The processing of a given satellite image allocates to each pixel an integer between 0 (absence) and 63 (bright), representing the radiance of the light observed. NTL-derived variables include the proportion of pixels illuminated, the average radiance, and the sum of all radiance values in a given geographic area (Addison and Stewart, 2015). Recent work by Li *et al*. (2019) shows that measures that characterise the distribution or spatial homogeneity of NTL in a given area can also be relevant predictors of poverty. Due to the expected absence of lights in uninhabited areas, NTL data are often considered jointly with LandScan population density estimates produced by the US Department of Energy at the Oak Ridge National Laboratory. LandScan estimates spatially disaggregate population census counts using information on administrative boundaries, land cover, elevation, roads, etc., and have been available online since 2000.

Attributes extracted from NTL data have long been recognised to provide useful proxies for monitoring a broad range of socio-economic and environmental indicators, such as GDP, informal economic activity, GDP growth, human development, electricity consumption, and greenhouse gas emissions, among others (Elvidge *et al*., 1997; Elvidge *et al*., 2012; Ebener *et al*., 2005; Sutton *et al*., 2007; Ghosh *et al*., 2009; Ghosh *et al*., 2010; Henderson *et al*., 2012; Doll *et al*., 2000).

As with remote sensing data generally, the main advantages of using NTL data over other traditional sources are their worldwide availability and the consistent measurement methodology, which eases comparability across time and space. For the monitoring of poverty, Noor *et al*. (2008) and Elvidge *et al*. (2009), and others, have illustrated the high predictive capacity of NTL at national levels. For sub-national estimates, the results are mixed, particularly in developing contexts, although a stronger association in rural than urban areas, and for asset-based compared to consumption indicators, (Jean *et al*., 2016), is frequently observed.

Vegetation indices

Vegetation indices are obtained by comparing the reflectance values of different bands of the electromagnetic spectrum that can be obtained from satellite imagery, with the aim of quantifying the level of 'greenness' or vegetation biomass in a given area. More than 100 different types of indices have been proposed in the literature (Xue and Su, 2017) but the most widely used for the purposes of poverty estimation are the Normalised Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI). Both of these indices provide an assessment of vegetation cover in a certain area. The NDVI, measured on a continuous scale varying between -1 and 1, corresponds to the difference in the intensities of reflected light in the red and infrared range divided by the sum of these intensities. The EVI aims to improve the NDVI by correcting distortions in the reflected light that can be induced by variations in atmospheric conditions and the presence of particles in the air, as well as by reducing saturation levels in areas with high levels of greenery, such as rainforests.

Vegetation indices have shown predictive capability for wealth, anthropometric, and health-related indicators, including in LMICs (Lo and Faber, 1997; Sedda *et al*., 2015; Engstrom *et al*., 2017; and Steele *et al*., 2017). According to the literature, a positive association between vegetation and wealth can be observed in well-established urban areas, while a negative one is observed in rural areas.

There has been significant progress in the use of remote sensing data and agricultural surveys for the estimation of agricultural productivity indicators, which in turn could act as valuable predictors for local economic conditions in rural areas (Khamala, 2017). However, two main issues need to be addressed before these can be fully used for the purposes of poverty estimation. First, methodologies based on productivity indicators may be very crop-dependent and difficult to generalise both in time and space. Second, the puzzle of measuring productivity for family and subsistence farmers, which may be more relevant for the purposes of poverty and extreme poverty measurement, remains unsolved. It is expected that the increased availability of very high-resolution satellite imagery (for instance through Digital Globe) can substantially improve the situation in this area.

Geographical conditions, climate, and accessibility

Other attributes often used in the literature, with various degrees of success, relate to: the initial conditions of an area (e.g. elevation, slope, and terrain ruggedness); the climate (such as temperature and precipitation, as well as their variation, the potential of evapotranspiration, and aridity indices); different types of land use; as well as the accessibility of areas (for instance, distance to rivers, roads, and population centres). Although some of these attributes are publicly available at various levels of spatial resolution, it is also possible to generate bespoke datasets from public satellite imagery.

In a rough summary, we can say that indicators of slope, elevation, and ruggedness are often positively correlated with poverty. Likewise, lower levels of precipitation or high aridity indices are associated with higher levels of poverty, as are those indicating lower levels of accessibility. The relevance of geographical conditions seems higher for rural than for urban areas, and particularly for developing countries. For variables such as temperature and evapotranspiration, variations seem to be more relevant than average levels. However, it is important to keep in mind that these results are just a generalisation and that the direction of associations, as well as their significance, may differ depending on the specific geographic, economic, and cultural situation of a particular region.

Roof type, informal settlements, and others

Other features, such as the presence of informal settlements, roof type, building height, and the presence of cars, which may be of particular relevance for the prediction of poverty in urban areas, can also be extracted from satellite imagery. However, the scalability of approaches involving these features has been considered limited because commercial high-resolution imagery and large sets of manually labelled data are typically required for the feature extraction. For an illustration of the type of data and processing required for these cases see Engstrom *et al*., 2017).

Some research has shown promise regarding the use of public satellite imagery for the identification of roof type attributes and the mapping of informal settlements. Varshney *et al*. (2015) predict the total number of roofs and the proportion of metallic ones in given areas of Kenya, using satellite imagery from Google Maps and a random forest regression trained on a relatively small set of manually labelled images obtained through volunteer crowd-sourcing. Unfortunately, field checks showed that the predictions from this methodology tended to overestimate the number of roofs, highlighting the difficulties associated with the identification of non-housing or secondary housing units in satellite imagery. Regarding the mapping of informal settlements, a very comprehensive description of the current issues is provided in Kuffer *et al.* (2016). Gram-Hansen *et al*. (2019) propose a low-cost solution in which a Canonical Correlation Forest algorithm is trained to predict the presence of an informal settlement using public Low-Resolution Sentinel-2 images. As for Varshney *et al*. (2015), the algorithm is trained on a relatively small set of manually labelled locations. The method is illustrated in India, Colombia, and several countries in Africa. For training using location-specific labelled data, this approach shows about 80% of pixel accuracy, i.e. predicting 80% of pixels correctly.

Hybrid approaches

Remote sensing data have been used in conjunction with other types of data, such as that obtained passively from CDRs, online, and social media data, as well survey data collected from small area-specific or mobile phone user samples (see the remainder of Section 3 for details on these methods). On the one hand, these types of approaches are useful as they offer the possibility of improving estimates compared with the use of remote sensing data alone, and even in some cases of producing individual or household-level predictions as opposed to only aggregate ones. On the other hand, the collection of additional information involves an additional cost and effort, which may hinder the chances of estimates being updated more frequently compared to when only remote sensing data are used.

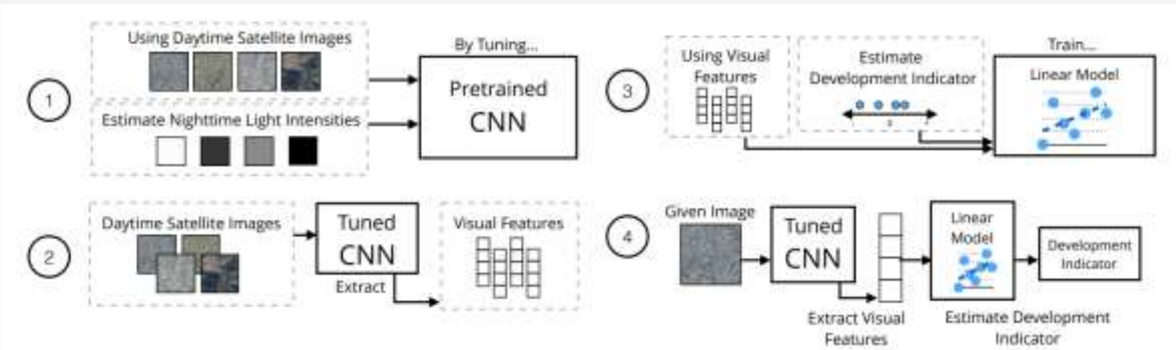
3.2.3 Expanding the set of features with 'transfer learning'

As a response to the limitations shown by NTL for the estimation of poverty indicators in urban areas, as well as the difficulty of extracting some of the potentially most predictive features, much interest has been centred on using other features extracted from daytime remote sensing imagery to improve the predictive capability of remote sensing data for wealth indices, poverty headcount ratio, and consumption indicators. This process may be highly demanding, both in terms of computational capability and because it requires large sets of labelled images, i.e. large sets of training data for which poverty estimates are available. Although convolutional neural networks (CNNs) have been proposed as an alternative for the 'automatic' extraction of poverty-relevant features, the scarcity of 'ground truth' data in the form of poverty estimates makes the use of this methodology challenging. Xie *et al*. (2016) and Jean *et al*. (2016) propose instead to train the CNN using a proxy with abundant labelled data (for instance NTL), as an application of 'transfer learning', as explained in Box 2.

Box 2: The 'transfer learning' process

Starting with a set of high-resolution remote sensing images, 'transfer learning' uses a series of successive classification problems to train a CNN to identify attributes that are useful for the prediction of a welfare indicator. In a first step, labelled images from ImageNet are used to train the CNN to identify low- and mid-level features, such as edges and borders. A second classification problem is then set, which typically consists of the prediction of levels of NTL radiance, either as a continuous measure or summarised in 64 categories, as a proxy for welfare. The last classification problem is to predict poverty estimates at a certain geographical level from the extracted features using a supervised learning tool (e.g. a linear regression model or a classification tree).

Figure 6: The transfer learning process for the prediction of development indicators

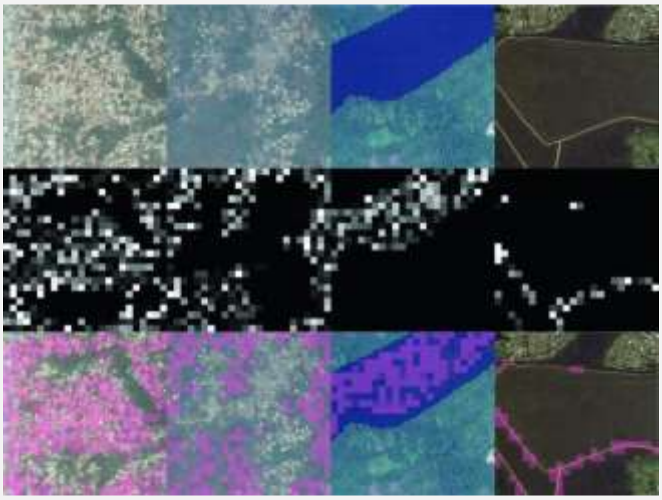


Source: Head et al. (2017)

It is worth highlighting that through the learning process, the CNN learns to extract relevant features, such as type of roof and the presence of roads, without having being explicitly asked to do so and without labelled data on these being provided. In the figure below, each column illustrates a different feature obtained from the corresponding satellite image (top row) via 'transfer learning'. The middle row highlights the characteristic that is relevant for each extracted feature. The bottom row superimposes the characteristic over the corresponding satellite image. Notice that the CNN is able to automatically extract information on water bodies, roads, and urban settlements, which may be highly relevant for poverty estimation, without the use of specifically labelled data.

Box 2: The 'transfer learning' process

Figure 7: Automatic extraction of features



Source: Jean et al. (2016)

The performance of this methodology has been assessed using publicly available images from Google Static Maps for Uganda (Xie *et al*., 2016) and five African countries (Jean *et al*., 2016), as well as for the whole of Africa using public, and more frequently updated Landsat 7 satellite imagery (Perez *et al*., 2017). Tang *et al*. (2018) show that features of similar predictive power can also be extracted from publicly available moderate-resolution maps of vegetation index (NDVI), with additional advantages for the dynamic updating of poverty indicators in rural areas, thanks to the higher sensitivity of NDVI for the measurement of change, as well as in countries with a heavy dependence on agriculture.

More recent efforts have focused on extracting relevant attributes directly from daytime light images, without the intermediate use of NTL. Engstrom *et al*. (2017) train a CNN to identify known relevant attributes, such as agricultural land, presence of cars, shadows (as a proxy for building height in urban areas), and roof type, from very high-resolution Digital Globe daytime images of a sample of areas in Sri Lanka. These attributes, together with spectral and textural characteristics easily obtained from the satellite imagery, show high predictive capability for indicators such as average log consumption and headcount poverty rates using either 10% or 40% of the national average income.

The scarcity of poverty survey data that initially motivated the use of 'transfer learning' is tackled in Yeh *et al*. (2020) by pooling estimates from DHS surveys conducted between 2009 and 2016 across 23 countries in Africa, i.e. creating one large integrated dataset (>500,000 households across approximately 20,000 villages) across several time periods and country contexts, to then train a CNN to predict wealth levels at high resolution. This may be seen as a strength as the CNN learns to identify relevant features from a broad set of contexts which can still be considered reasonably similar.

Instead of using poverty measures obtained directly from a survey for a limited set of areas, Engstrom *et al.* (2017) use small area estimates obtained through the 'ELL' methodology of Elbers, Lanjouw, and Lanjouw (Elbers *et al*., 2003) at a very disaggregated geographical level (an average of 8.4 households per area included in the original survey) as ‘ground truth’ data for an entire country, combining it directly with daytime imagery via a CNN analysis. Although this approach requires census data at the initial calculation of ELL estimates, which may be seen as a limitation, a similar performance is obtained by training the CNN with direct estimates of the welfare indicator obtained from a ‘census extract’, i.e. a sample of similar size to the one of the original survey. Yet the use of SAE in Engstrom *et al.* (2017) does not lend itself to a direct performance comparison with other methods, due to the lower variability in poverty estimates from SAE compared to direct poverty estimates used in other applications. Nevertheless, in this approach, a direct use of daytime imagery appears considerably more efficient than the indirect use of NTL intensity and transfer learning.

Semi-supervised and unsupervised learning

New research has focused on the use of semi-supervised and unsupervised learning methods as an even more flexible alternative that can be used to circumvent the lack of abundant poverty data. In these approaches, multispectral publicly available low-resolution satellite imagery is typically used and some degree of spatial homogeneity or correlation in the space of the features of interest is implicitly assumed. Although this area of research is still relatively new, some of the proposed methodologies seem to perform as well or slightly better than methods such as ‘transfer learning’ for the prediction of poverty (Perez *et al*., 2019; Jean *et al*., 2018; Jean *et al*., 2019).

3.2.4 Summary

A summary of the main characteristics of the approach of using remote sensing data for prediction of poverty indicators is presented in Table 3.

Table 3: Summary of remote sensing data

Criteria	Summary comments
RELEVANCE	Using an appropriate combination of the outcome indicator, set of features, and level of geography, the proposed methodologies have been able to explain on average around 40–60% of the variability observed in ‘ground truth’ estimates, and in some specific applications up to 70%. These results are very encouraging. However, notice that such metrics refer to average measures of poverty, and the discriminant capability for the extreme poor, located at the tail of the distribution of welfare, seems considerably lower in most of the applications reviewed. Furthermore, as poverty measured by the standard indicators (assets, consumption, expenditure) cannot be observed from space, remote sensing data can be used as a proxy but not as ‘ground truth’ for this goal. Additional data sources are required to go beyond the prediction task, to understand the underlying mechanisms and determinants of poverty.
COST CONSIDERATIONS	Estimates obtained from remote sensing are very attractive cost wise as satellite imagery is becoming more public and freely available, even for middle levels of resolution. For bespoke applications, there may be additional costs associated with the extraction/processing of the required features, if high computational power and a very specific set of skills are required. However, there is a culture of data and software sharing supported by governmental organisations in developed countries, NGOs, and academic research groups in this area that can help mitigate these impacts.

Criteria	Summary comments
FREQUENCY	Remote sensing data are very well-suited to the continuous updating of poverty estimates, as long as it can be assumed that the model/algorithm used for prediction remains valid across time. In some cases, the inclusion of variable measures in the set of features can be helpful to make this assumption more reasonable. Although ground truth estimates would be required to formally test this assumption, it is expected that some outcomes and geographical levels are more robust than others to this kind of misspecification. For initial assessments of the use of repeated remote sensing data for continuous monitoring see Bansal *et al*. (2020), Tang *et al*. (2018), and Bennett and Smith (2017).
SPATIAL RESOLUTION	As remote sensing features can be processed at very low levels of resolution, the biggest limiting factors are: i) the availability of accurate training data at low geographic levels; and ii) the predictive capacity of remote sensing features at such low levels. Efforts to expand the set of features are important for this goal, particularly in urban areas.
ACCESSIBILITY	Accessibility is not generally an issue with remote sensing data unless the particular methodology of interest requires high-resolution or very high-resolution satellite imagery, which is usually commercial and expensive. Some of the approaches presented in this section make use of state-of-the-art machine learning tools, which require a high level of skill, but, as said above, a culture of open data and software prevalent in the remote sense community may help in mitigating this issue.
DATA QUALITY	Due to their broad use, quality in the gathering and processing of remote sensing data has been a source of attention of the research community for a number of years and standardised protocols are in place for many types of data. Bespoke processing involving complex machine learning methods may be more sensitive to poor choices of hyperparameters and other method attributes. Regarding the training data, the main quality issues are: i) low accuracy, particularly when survey estimates are generated at lower levels of aggregation than what the survey was designed for; ii) spatial mismatch with the set of features, due to displacement of survey geocoordinates for disclosure control; and iii) the possibility of selection bias due to out-of-date sampling frames, which may systematically miss the poorest subpopulations, such as those residing in informal settlements.
ABILITY TO MEASURE UNCERTAINTY	Except for applications using either SAE or statistical models in a Bayesian framework, typically only summary metrics of overall performance of the method, obtained using cross-validation, are provided under this approach. Notice that this is a quite restrictive way of defining uncertainty and a more comprehensive assessment may involve elements such as the uncertainty in the extraction of features, in the training data, and in regard to the method selection. Further study of the performance of the methods for different portions of the distribution of welfare would substantially increase the potential of these methods for the development of better targeted policies.
ABILITY TO MAKE OUT-OF-SAMPLE PREDICTIONS	Studies of the capability of methods based on remote sensing data for generalising to other spatial contexts have shown mixed success. Machine learning algorithms may be trained using data from several locations, or using semi-supervised or unsupervised algorithms in an attempt to improve spatial generalisability, at the possible expense of predictive power for a specific environment. The bottom line is that considerations about the type of indicator, the level of geography, and the socio-economic conditions of a particular area are necessary.
REPLICABILITY	In general, methods in this category have a high level of replicability, as both the data and the software are often publicly available. For instance, the results of Jean *et al*. (2016) have been replicated successively by several groups of researchers, in order to assess the performance of newly developed methods. However, some of the machine learning algorithms seem to be much more

Criteria	Summary comments
	sensitive to the specification of initial parameters, which may create replicability issues unless the complete set of specifications is made publicly available.
ETHICAL CONSIDERATIONS	Ethical considerations in this area can arise from the more prevalent availability of high-resolution satellite imagery and the processing of survey data at very low levels of aggregation, which could result in disclosure control issues. A higher-level ethical concern arises from the fact that the wide coverage and periodical availability of remote sensing sources can promote spatial and time generalisations without the use of ground truth data, leading to potentially damaging consequences.

3.3 Data exhaust

Key messages:

- The data exhaust refers to sources of data that are inadvertently produced as a by-product of people's interactions with digital services. Mobile phone ownership and internet penetration have increased significantly over the past decade, including in LMICs, with this trend expected to continue. With this has come vast amounts of data from the data exhaust.
- Researchers are experimenting with using these data to estimate or 'predict' poverty estimates for areas, time periods, and at geographical levels that were previously not possible using traditional data. Statistical learning methods are a part of this process.
- A key set of data from the data exhaust are Call Detail Records (CDRs), which contain meta-data on mobile phone activity. Other approaches experiment with data from mobile money, airtime, and credit transactions to map poverty or to identify potential beneficiaries for social assistance.
- There are two important benefits of the data exhaust. First, the data exhaust from mobile phones allows individual-level predictions of phone-owner wealth or poverty. Second, the data exhaust can provide information that is directly relevant to the problem of extreme poverty measurement, since some data is closely connected to poverty and wealth (e.g. mobile money use).
- There are some drawbacks. First, data in the data exhaust are mostly owned by private companies. Second, mobile phone ownership and internet access is not universal, which makes data exhaust coverage non-universal, as well as having the potential to bias results. Third, there are ethical concerns with using these data, relating to issues of privacy and citizen surveillance.

Data exhaust refers to sources of data that are inadvertently produced as a by-product of people's interactions with digital services in their daily lives. Each time someone uses a mobile phone, performs an internet search, sends a monetary transaction through mobile or internet money, or carries out other routine tasks using digital devices, they leave behind a trail of data that can be harnessed for a variety of purposes to study human behaviour and socio-economic characteristics. In this section we discuss the possibilities afforded by mobile phone data, internet search history data, and financial transactions data in relation to estimating poverty.

3.3.1 CDRs

Mobile phone ownership and use across the world has increased rapidly over the past decade. By the end of 2019 there were approximately 5.2 billion mobile phone subscribers worldwide, amounting to roughly 67% of the global population (GSMA, 2020). While some forms of big data are not yet widely produced in developing country settings, mobile phone data records are an exception. Indeed, sub-Saharan Africa is reportedly the fastest growing region in the world in terms of mobile phone subscribers, with an additional 167 million subscribers predicted in the period between 2018 and 2025 (GSMA, 2019a). The rise of mobile phone users has led to the generation of vast amounts of data. A growing class of approaches have developed around the use of these data, especially the information contained in CDRs that are generated by mobile network operators (MNOs). While CDRs do not generally store the actual contents of phone calls or messages, they do contain a variety of other attributes that can be employed in prediction tasks. These include location information for the sender and receiver of phone calls and SMS messages,[16] the time that the call or SMS was made, the duration of phone calls, the size of messages, and network information. CDR data have lent themselves to a wide variety of applications, including analysis of migration patterns and population dynamics (Blumenstock, 2012), understanding social networks, and estimating consumption and expenditure.

Consequently, there is a growing body of work using CDR data – and similar statistical learning techniques as in the case of remote sensing data – to estimate poverty rates and generate poverty maps. For example, a 2017 World Bank study used municipal-level poverty estimates for Guatemala derived from the National Living Conditions survey and most recent census (through SAE techniques) to train a model of poverty prediction using an aggregated CDR dataset (Hernandez *et al*., 2017). The CDR data were found to be significantly predictive of poverty rates, explaining a high degree of variation between areas.

Methods employing CDRs have similar advantages over traditional household survey-based methods as remote sensing data. Firstly, CDR data are generated on an almost continuous basis. A second advantage is that analysis of CDRs can also be performed at significantly lower cost than household survey or census alternatives. The costs of the CDR analysis in the Guatemala study described above were around 20 times cheaper than the cost of a 2014 household survey covering 11,500 households[17] (Hernandez *et al*., 2017).

Finally, CDR data are available at fine levels of spatial resolution. CDR data can be aggregated to the level of the individual phone towers associated with each transaction. The density of phone towers varies across countries and between settings but towers are

---

[16] CDR data are assumed to be linked to the mobile phone tower (or 'base receiver station') nearest to the subscriber at the time when relevant activity is recorded on their device (such as placing or receiving a call or SMS message). In practice there are situations where a phone transaction passes through a more distant tower, due for example to issues relating to network strengths and terrain (Hernandez *et al*., 2017).

[17] The 2014 Encuesta Nacional de Condiciones de Vida survey reportedly cost around US$ 2 million to complete, over two years, compared with around US$ 100,000 for CDR analysis of poverty status on Guatemala presented in this study by Hernandez *et al*. (2017).

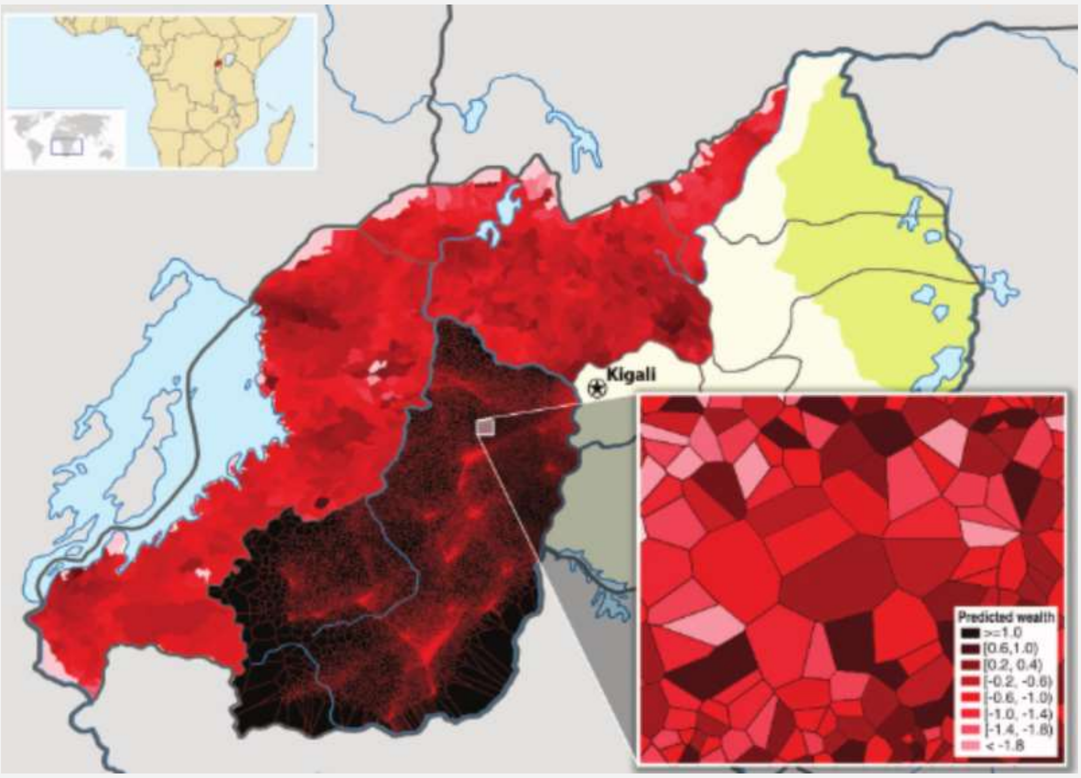
typically located between a few hundred metres and tens of kilometres from each other (Hernandez *et al*., 2017). Coverage is higher in urban areas compared with rural areas (Steele *et al*., 2017).

Box 3: Predicting poverty and wealth from mobile phone meta-data

One example of a novel application of CDR data to predict poverty comes from (Blumenstock *et al.*, 2015), who have used anonymised mobile phone data records to predict individual socio-economic status. Specifically, they used an anonymised dataset containing billions of records from Rwanda's largest mobile phone network to automatically compute a large set of several thousand quantitative metrics for each subscriber. They paired this with a follow-up phone survey of a randomly selected sub-sample of 856 users to train a model for predicting wealth. The elastic net regularisation regression method was used to identify the metrics most strongly predictive of wealth, and to discard the remainder. The resulting model was used to predict characteristics for the remaining mobile subscribers who were not part of the survey sub-sample. When aggregated from the cell level (the finest administrative unit in Rwanda) up to the district level, the estimates were strongly correlated with corresponding wealth estimates from the Rwanda DHS.

These findings show that historical phone transactions data can be used to predict wealth at the individual level. In fact, the authors' modelling approach can be generalised to perform other predictive tasks too (for example, estimating responses to other survey questions), though with variable precision. In combination with the geo-location information in the phone survey, this approach provides the potential to map the wealth distribution at fine levels of spatial resolution (an example of the maps produced is shown in Figure 8). An important contribution of this study is the finding that CDR data can be used to make individual-level wealth predictions, where previous applications had typically been aggregated at higher area levels.

Figure 8: Example of a poverty map created from mobile phone records data



Source: Blumenstock et al. 2015: 1073–1076

Yet while CDR-based approaches have several potential advantages over traditional survey- or census-based methods for poverty measurement, they also have some important limitations. One concern is that CDR data can only generate insights across a population of mobile phone users. While mobile phone uptake has been increasing globally, it is not universal. Mobile phone users are younger and more highly educated than average, more likely to live in developed countries than non-users (Silver *et al.*, 2019), and more likely to live in urban than rural areas (GSMA, 2018). Lack of access to mobile phones, lack of resources to buy phone credit, and lack of electricity or network coverage to use phones can limit uptake among the rural poor. In LMICs, women are also less likely to have access to mobile phones than men (GSMA, 2019b). This means that CDR data can under-represent, or exclude entirely, the very people who are most likely to be living in extreme poverty.

A second potential issue is that CDR data are proprietary, and access must be negotiated with the MNOs that hold the data. MNOs may be protective in disclosing the rights to these data out of concern for their commercial interests, and therefore securing this access can be difficult. In studies where the findings represent CDR data held by one MNO only, the issue of sample selection bias may be further compounded by a restriction to one segment of the market only (even where this represents the largest provider of mobile services).

A final set of concerns around CDR-based approaches relate to subscriber privacy and data protection. Raw CDR data contain a rich set of information about individual subscribers' mobile phone transactions and location (to the nearest cell tower), generated over fine time intervals. Even when data are anonymised there are concerns that complex data processing techniques could be used to uncover the identity of subscribers. Studies have shown that in many cases individuals can be re-identified with an extremely high degree of reliability through a core set of basic demographic characteristics even from datasets that have been through a process of anonymisation (Rocher *et al.* 2019).

Concerns around the privacy and the potential insufficiency of conventional data anonymisation procedures are legitimate. However, they are not necessarily insurmountable. In a 2017 paper, Njuguna and McSharry (2017) show that even an extremely sparse CDR dataset that removes many core characteristics can be used, in conjunction with other forms of data, to generate an MPI index. They used a reduced CDR dataset that did not distinguish between phone calls and SMS messages or record the receiver of the transaction, together with LandScan population data and a NTL dataset, to predict an MPI in Rwanda. The results showed that the model was able to explain around 76% of the variance in MPI across sectors (an administrative area level in Rwanda), indicating that CDRs remain a valuable resource even where privacy concerns require the data to be simplified. Continued work to develop core principles and global standards governing the use of CDR data in a way that adequately protects subscriber privacy rights is crucial.

3.3.2 Other data exhaust sources

In addition to CDR data, there are a range of other forms of data exhaust emanating from the 'data crumbs' that are produced by other use of digital services. The usefulness and coverage of such data across different settings depends to some extent on the level of internet penetration. This concerns not only the level of access to digital services, but also patterns of usage and the extent to which the internet penetrates individuals' lives. For this reason, much of the data exhaust literature we are aware of up to now is associated with OECD countries. There has been comparatively little work concerning poverty measurement in LMICs. However, some examples discussed below provide an indication of the kinds of measurement possibilities that could be feasible through alternative sources of data exhaust.

One example is data mined from internet search histories. The possibilities for using these data for prediction tasks are subject to limits relating to data protection, and it is of course impossible for these kinds of data to be made available to researchers in raw form. However, aggregated data and simplified versions of these that meet appropriate confidentiality standards have the potential to provide a range of insights related to human behaviour. In addition to the coverage issues described above, there are other analytical challenges associated with using these kinds of data for predictive tasks that have occasionally led to misleading results. A leading example of this is Google Flu Trends, an algorithm developed using Google search terms to predict influenza-like illnesses. In 2013 *Nature* reported that Google Flue Trends was routinely overstating the incidence of flu relative to reports from the US Centers for Disease Control and Prevention (Lazer *et al.*, 2014). Part of the explanation seemed to lie in interference in the algorithms of search terms that appeared to be associated with flu-like infections but which were not actually predictors of flu. This underscores the potential pitfalls in seeking to uncover genuine relationships from such a large expanse of data.

A second example of an alternative form of data exhaust that could be used for poverty measurement is data from mobile money, airtime, and credit transactions. While other forms of digital services and online information may be relatively less common in LMICs, mobile money is an exception. In sub-Saharan Africa alone, 50 million registered accounts were reportedly opened in 2019, with GSMA predicting that over half a billion accounts will exist in the region by the end of 2020 (GSMA, 2019c). While this impressive growth may not be indicative of similarly widespread usage among extremely poor populations, the high penetration of mobile money and mobile-based platforms for making transactions presents a key emerging source of data. Researchers have started to make use of these data to study different behavioural traits. For example, Paruthi *et al.* (2016) study data generated from an online peer-to-peer microcredit platform to examine lending behaviour among low-income entrepreneurs.

Other forms of transaction data may also be leveraged for a variety of analytical purposes beyond direct poverty measurement. This includes collecting information related to individual experiences of poverty or that may help inform intervention targeting, although to date the majority of applications of ATM or credit card transactions data are in developed country settings. For example, analysis of purchasing behaviour from credit card transactions has shown potential to help identify groups of individuals with similar characteristics (such as gender, age, and social networks) (Di Clemente *et al*., 2018). The ability to identify clusters of individuals through new analytical tools may lend itself to future applications related to targeting of interventions. Other possibilities include using credit card transactions data to study the effects of social unrest (Dong *et al*., 2018), or using bank payments and ATM data to help understand the effects of natural disasters on populations with different wealth levels (Martinez *et al*., 2016). In a more recent example, such data are used to identify potential beneficiaries for social assistance payments (Blumenstock, 2020).

3.3.3 Summary

A summary of data sources that fit within the 'data exhaust' category is presented in Table 4.

Table 4: Summary of data exhaust

Criteria	Summary comments
RELEVANCE	The majority of applications described in this section use the data to predict estimates of poverty, rather than to measure it directly. The uniting characteristic of the approaches discussed in this section is that the source of data is a by-product of other interactions with digital services. This means that there is little opportunity to influence the kinds of information captured in order to suit specific data requirements. This makes the approaches in this category somewhat less flexible in terms of measurement options than traditional survey-based methods. They may be suitable for measuring proxies of poverty, and understanding other related phenomena, such as migration patterns. However, due to the selective nature of internet access and mobile phone ownership, it is important to reiterate here as well that these measurements might not reflect the situation of the poorest section of the population.
COST CONSIDERATIONS	Data exhaust-based methods can be conducted at significantly lower cost than household surveys or censuses. An example from Guatemala showed that CDR analysis cost around 20 times less than a household survey (Hernandez et al., 2017).
FREQUENCY	Information generated as data exhaust is produced almost continuously, each time a digital interaction takes place. This makes the methods listed under this heading suitable for creating and rapidly updating estimates of poverty or dynamic poverty maps, provided that data can also be accessed frequently.
SPATIAL RESOLUTION	Data from CDRs can be located to within the range of the mobile phone tower closest to the subscriber at the time the phone call or SMS message took place. The coverage of mobile phone towers can vary, but this still represents a very narrow level of spatial resolution compared with traditional survey-based estimates. Other data, like data on mobile money transactions, can even be linked to individual account owners, i.e. they allow individual-level estimates of wealth – a considerably higher level of disaggregation than even remote sensing data allows.
ACCESSIBILITY	The kinds of data described in this section are almost always proprietary, belonging to the owner of the digital service used to record the transaction in question. For example, CDR data are the property of MNOs. This means that access usually needs to be negotiated, and commercial or data protection concerns may act as constraints to access.

Criteria	Summary comments
DATA QUALITY	One of the main data quality considerations with data exhaust-based methods is sample selection. The methods can only capture information about people who use digital services; those in rural areas and living in extreme poverty are more likely to be excluded than those in more urban, developed settings. A second source of selection bias can arise from how the data are obtained. Some studies only seek data from one service provider (for example, one MNO), omitting other providers with a market share that may cover systematically different population sub-groups. This means that the sample covered by the data may not capture the populations of greatest interest for understanding extreme poverty.
ABILITY TO MEASURE UNCERTAINTY	One of the key sources of uncertainty in data exhaust-based methods is the extent of sample bias (discussed above). The extent to which the sample over- or under-represents certain population sub-groups may be difficult to estimate precisely.
ABILITY TO MAKE OUT-OF-SAMPLE PREDICTIONS	Some applications of data exhaust have demonstrated how models trained using ground truth data from one setting can be used to make out-of-sample predictions. For example, Blumenstock et al. (2015) show how their model of wealth derived through a prediction task employing CDR data and a sample survey is also able to predict wealth for users not included in the survey. The out-of-sample predictions, when aggregated at district level, compare well to corresponding DHS estimates (also proxy-based).
REPLICABILITY	The issues of data accessibility described above may frustrate efforts to replicate analysis using data exhaust. An important challenge for advancing work in this area is the development of clear protocols governing data sharing agreements, that can strike a balance between the value that can be obtained from the data, legitimate commercial interests or legal concerns, as well as individual data protection rights (discussed below).
ETHICAL CONSIDERATIONS	A key concern with this group of approaches is data protection rights. Raw data from mobile phone records or internet transactions can hold an array of personal information that individuals have not necessarily consented to being used by third parties. Using these data in a way that does not infringe on individual data protection rights is critical, and work to develop strict codes of practice around responsible handling of these kinds of data is still needed. However, there has been progress in this area; recent work has shown that even a very sparse dataset that strips out significant identifying detail from the raw data can still perform well in prediction tasks (Njuguna and McSharry, 2017).

3.4 Online information

Key messages:

- Online information refers to **data that individuals share online**, such as online articles, social media posts, and Wikipedia entries. In contrast to the data exhaust, these data are not a by-product of online interactions but are explicitly created by.
- Despite the trend of advancing internet penetration, our review identified only a **limited number of applications** of online information for measuring poverty. Some of these employ geo-located online data in the same way as CDR or satellite imagery data (e.g. to produce high-resolution maps of poverty). Others focus on specific components of the poverty measurement challenge: for example, automatically analysing news and social media data to track food price fluctuations.
- As with the data exhaust, online information comes with an **inherent bias towards wealthier populations**. In addition, given its intentional nature, it is biased towards populations with a sufficiently **high level of education** for active interaction with the internet. And in terms of **population coverage it is less relevant than the data exhaust**, as mobile phone ownership is much more prevalent than internet access in LMICs.
- An important feature that differentiates online information from other data sources is that it often comes in an **unstructured text format**, which means researchers may need to employ statistical learning techniques to pre-process, engage with, and make us of it.

Online information refers to data that individuals actively share online, such as online articles, social media posts, and Wikipedia entries. There are a number of promising approaches to harnessing these sources to address questions related to poverty estimation. For example, methods have emerged that use Twitter data to predict food shortages (Global Pulse, 2014), incorporating social media data in estimates of urban poverty (Niu *et al.*, 2020) and using news sources to predict food price fluctuations (Chakraborty *et al.*, 2016). Box 4 outlines a recent example showing how data from Wikipedia can be used to generate poverty maps in developing country settings.

Box 4: Predicting poverty using geo-located Wikipedia articles

An innovative example of how online information could be analysed to measure poverty comes from Sheehan *et al.* (2019), who use geo-located Wikipedia articles to predict community-level asset wealth indicators in Nigeria, Malawi, Tanzania, Uganda, and Ghana. Their paper shows that Wikipedia articles have considerable potential to be used for estimation tasks in a manner that has not previously been exploited. The density of geo-located articles globally is mapped in Figure 9 below, consisting of an estimated 1.02 million articles in the English version of Wikipedia worldwide, and over 50,000 in Africa. These provide a source of detailed text information that can be linked to different locations.

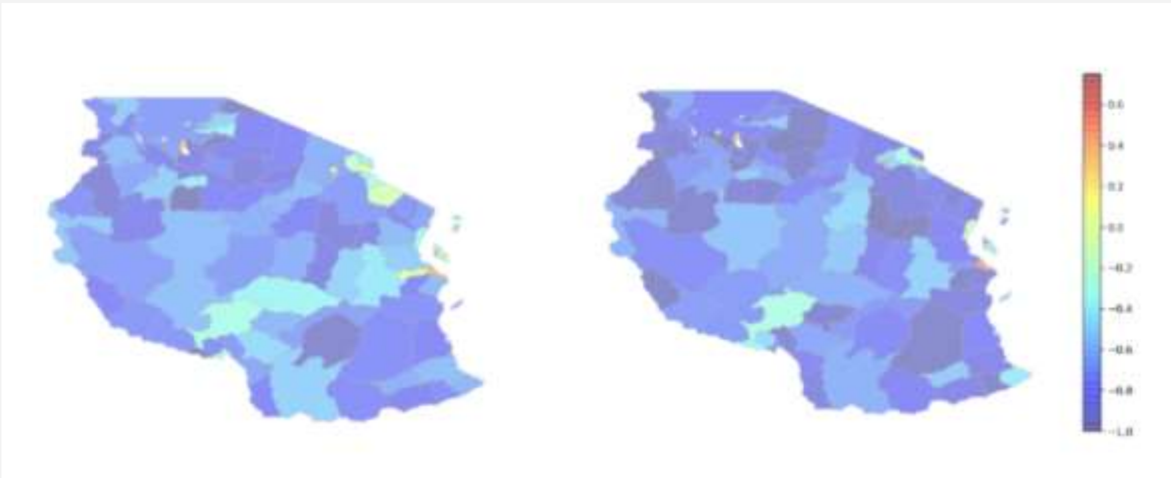
The first step of the authors' approach is to extricate information from Wikipedia articles (which take the form of unstructured text articles) using NLP. They use a method called Doc2Vec to map the articles to a vector representation, and then train a model of wealth using survey data from the DHS. The model is then enhanced by adding NTL data. This results in a model of wealth (trained using the asset wealth index available in the DHS) aggregated at community level. Their results point to a model that effectively predicts poverty at fine levels of spatial granularity, and that is also capable of making out-of-sample predictions. Figure 10 compares ground truth wealth data for Tanzania with predictions made using a model trained in Ghana. The authors find that although the

Box 4: Predicting poverty using geo-located Wikipedia articles

out-of-sample predictions do not always correctly capture the levels of wealth, they do provide good predictions of the relative wealth of different areas. An understanding of relative poverty levels may be valuable for a range of applications, including intervention targeting and monitoring changes in inequality over time.

Figure 9: Distribution of geo-located Wikipedia articles

Figure 10: Comparison between ground truth values (left) and predicted values using Wikipedia data (right), from a model trained on Ghana and applied to Tanzania



Source: Sheehan et al. (2019)

Another key insight from the results of the paper by Sheehan *et al.* (2019) is that the integration of Wikipedia articles together with NTL data improves estimation. The authors posit that these sources are complementary. While prediction based on Wikipedia information alone may be limited by data gaps, NTL (discussed further in Section 3.2) can partially overcome this challenge by providing a more uniform source of data.

Measurement approaches that use online information bear some similarities to those listed under data exhaust in terms of advantages and disadvantages. Online information provides a potentially cheaper alternative to collecting survey data, and contains rich information to support different targets of estimation. On the other hand, information may be relatively scarcer in rural areas, where access to the internet or network coverage is lower than in urban, developed settings. And though the content of online information is eminently varied, there are limited opportunities for researchers or policymakers to tailor what is collated online to fit specific measurement objectives (an exception to this is crowd-sourced data, which we discuss separately in Section 3.5). This means that it is unlikely that online information could be used to support the classical measurement of monetary poverty indicators; proxy measures are more feasible. While some data are open access, other information may be proprietary and unavailable for research purposes. Finally, some of the same concerns about data protection and confidentiality also apply to some of the information held online. Though individuals may freely publish some information online in the expectation that it will be accessed by others (such as public blog posts or social media posts), the use of other information held online (such as search terms) may breach individual data protection rights.

The sheer abundance of data held online also presents some unique challenges. One difficulty is that most online information is stored in the form of unstructured text. The task of sifting through a vast corpus of unstructured information to accurately distinguish useful from irrelevant information poses a considerable computational challenge. A second issue is that user-generated online information (such as blog posts, articles, and social media posts) may include information that is unverified or not easily verifiable, which could weaken the veracity of evidence (Global Pulse, 2012: 27). There are also subtleties within text information that current text-mining and NLP techniques may struggle to recognise, such as use of irony, slang, exaggeration, abbreviated language, and humour.

3.4.1 Summary

A summary of sources that fit within the 'online information' category is presented in Table 5.

Table 5: Summary of online information

Criteria	Summary comments
RELEVANCE	The information that exists online constitutes a vast expanse of varied and unstructured data. This is not likely to include the kinds of information required to generate consumption- or asset-based poverty line measures. However, the existence of relevant information to support the measurement of proxy indicators, or to understand other attributes that are related to extreme poverty, is possible. However, relevant and useful information that is housed online also co-exists alongside an immense volume of irrelevant information.
COST CONSIDERATIONS	The majority of information on the internet is freely accessible, though access to certain domains may be restricted (requiring negotiation and possible costs). The most significant cost constraint may be the cost of accessing the required expertise or software required to process such large volumes of data.
FREQUENCY	Information generated online in theory can be analysed almost continuously.
SPATIAL RESOLUTION	Data from online sources can in theory be analysed at high spatial resolution. In practice, this depends on information being geo-located, as well as on coverage. Note that spatial resolution may be lower in contexts that are more affected by extreme poverty, due to lower access and use of the internet.

Criteria	Summary comments
ACCESSIBILITY	As discussed above (under 'cost considerations'), a vast amount of online information is already in the public domain and available for free.
DATA QUALITY	There are a number of possible challenges to data quality. One risk is the possibility of identifying misleading or spurious relationships from among such a vast array of data. A second risk is the difficulty of interpreting largely unstructured information, distinguishing relevant from irrelevant information, distinguishing fact from conjecture, and making sense of non-straightforward features of these data, such as exaggeration, irony, and humour.
ABILITY TO MEASURE UNCERTAINTY	There are no simple metrics available to represent the likely level and direction of uncertainty or bias associated with estimates from online sources.
ABILITY TO MAKE OUT-OF-SAMPLE PREDICTIONS	Applications of online information have demonstrated a potential to make decent out-of-sample predictions (see Box 4). This may depend on the nature of the application and the form of online information being used.
REPLICABILITY	This is likely to vary. In theory, applications of online information using data that are held in the public domain may document the methods used to enable replicable results. Since the use of online information remains relatively novel, there is not yet a strong precedent in place for replicable research within this field.
ETHICAL CONSIDERATIONS	Issues of individual consent and data protection are pertinent for the analysis of online information. Individuals may post information online without having an understanding or awareness of the potential implications, or the opportunity to explicitly provide consent for their information being used for analytical purposes. The internet is a largely ungoverned space, and methods for data extraction and analysis should acknowledge possible risks for individuals.

3.5 Citizen-reported or crowd-sourced data

Key messages:

- Citizen-reported or crowd-sourced data are data that are **purposefully submitted by citizens** via their mobile phones or the internet. In contrast to surveys, this means that respondents or contributors self-select into sharing data. Depending on the technology available, a wide variety of different data types can be shared.
- The crowd-sourcing of data has been used frequently in environmental sciences, health, geographic mapping, and the monitoring and evaluation of social programmes, but our review found very **few applications of these data for measuring poverty**. One example is crowd-sourcing price data in order to construct price indices. Another is using crowd-sourced geographic data from the OpenStreetMap (OSM) platform and combining them with satellite data in order to improve poverty mapping.
- Proponents of citizen-reported data point to the potentially **empowering nature of contributing data.** It is clear, also, that the variety of data types that can be collected with this approach holds great promise for new ways of trying to estimate poverty. However, this approach can also face issues of **poor data quality and bias**, given the unstructured way in which these data are collected.

Data reported directly by citizens have the potential to contribute towards the measurement of extreme poverty, and in this section we discuss the applications of citizen-reported or crowd-sourced data. These terms can occasionally be used in different ways, but we partly adopt the terminology of crowd-sourced data provided by the United Nations Global Pulse, as information that is 'actively produced or submitted by citizens' via mobile phones, hotlines, online questionnaires, or other ways in which citizens can take the initiative to share information (Global Pulse, 2012: 16).[18]

An important component of this definition is that citizens are active participants in the production of data. This distinguishes this category from data that are passively produced by citizens through the use of digital services (which we discussed under the banner of 'data exhaust' in Section 3.2). This definition of crowd-sourced data also shares some similarities with survey data collection, in which individuals actively contribute to answering interview questions. Yet unlike traditional modes of survey data collection, crowd-sourced data are collected via tools that are deployed remotely – such as mobile phones or apps – without any face-to-face interviewing. In addition, and in contrast to mobile phone sample survey data, we define crowd-sourced data as being collected by individuals who self-select into providing their data, rather than being part of an identified sample (Dabalen *et al*., 2016: 12).

Several different forms of data are included within this category, such as photographs, videos, geographic information system (GIS) data, and other data that can be collected using phones. The aggregation of such data offers powerful opportunities to provide new information on previously unstudied phenomena, or to extend or validate existing data. Crowd-sourced data have been incorporated into a wide variety of applications across a diverse set of fields, including geographic mapping, monitoring and evaluation, and environmental monitoring (Lämmerhirt *et al*., 2018). Another common application is 'crowdmapping' or crowd-sources mapping, which refers to the aggregation of information provided by a large number of contributing individuals to generate a map of key information. A number of initiatives exist to coordinate the generation of crowd-sourced data, including Premise data (discussed in Box 5 below), mClerk (Gupta *et al*., 2012) and txtEagle (Eagle, 2009).

Box 5: Using crowd-sourced data to estimate prices

Information about prices is crucial for the measurement of poverty. For the construction of poverty lines, price and inflation data are required to convert nominal expenditure values into real consumption, as well as to determine the appropriate PPP exchange rate. The ability to monitor prices frequently and at fine levels of spatial granularity can also help policymakers to identify instances of price volatility and to anticipate where risks of rising food insecurity or worsening poverty may be likely. Yet up-to-date and accurate information about prices is often lacking in developing country settings, especially for rural areas, where data collection may be more costly and complicated owing to the potentially informal and mobile nature of markets and vendors where items are typically bought (Gaddis, 2016). Crowd-sourced information on prices has the potential to bridge this gap.

One example of this comes from Monrovia, Liberia, where Blumenstock and Keleher generate a price index based on crowd-sourced data and study how closely this correlates with analogous

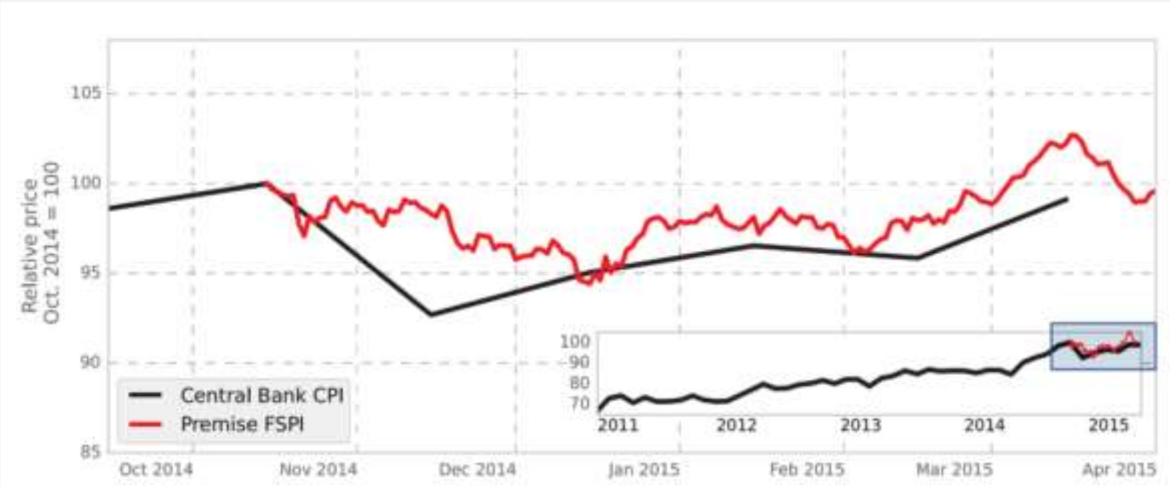
[18] An alternative definition of citizen-generated data given by the DataShift initiative refers to 'data that people or their organisations produce to directly monitor, demand or drive change on issue that affect them. It is actively given by citizens, producing direct representations of their perspectives and an alternative to datasets collected by governments or international institutions' (The DataShift, 2015a).

Box 5: Using crowd-sourced data to estimate prices

indices from official sources (Blumenstock and Keleher, 2015). They use data from Premise, a technology platform that facilitates crowd-sourced information on a variety of economic, social, and demographic information from volunteer contributors through the completion of different structured tasks. Premise currently operates in over 70 countries worldwide.[19] In Monrovia (the focus of this paper), Premise has collected daily observations of prices for a range of food and non-food items since 2014. The authors use these data to compute a monthly aggregated Food Staples Price Index (FSPI). The constructed FSPI is compared with official 'ground truth' consumer price index data from the Central Bank of Liberia (shown in Figure 11 below) and the United Nations World Food Programme. The results show a high level of overall correlation between the crowd-sourced index and official sources, but with some considerable deviations in the estimated prices for individual items.

Blumenstock and Keleher's (2015) paper concludes with a discussion of the relative merits and disadvantages of traditional and crowd-sourced price data. While traditional price data collection benefit from well-established systems for data collection and analysis informed by best practice norms, crowd-sourced data have the potential to provide a lot of information at much finer levels of granularity and frequency. This enables these data to be used to detect volatility that may be missed by traditional sources. However, trend analysis of the crowd-sourced price data in this paper indicated substantial noise, requiring careful work to attempt to distinguish genuine fluctuations from measurement error. Further work is required to understand what may drive discrepancies between the two sources, such as the item-level differences observed in this paper.

Figure 11: Comparing crowd-sourced information on prices with the central bank consumer price index



Source: Blumenstock and Keleher (2015: 117–125)

The diversity in the forms of data that can be collected through crowd-sourcing methods means that there is considerable variation in the characteristics of those data. Nonetheless, there are a number of distinct advantages to crowd-sourced data. In terms of spatial resolution, much of the information collected is geo-located (such as geo-tagged photograph data, or information provided through a mobile device), and therefore permits estimation at potentially fine geographic levels. The same is true of the frequency with which data are collected: depending on the platform used and the frequency of contributor participation, crowd-sourced data can in theory be gathered on an almost continuous basis.

---

[19] See www.premise.com

Of key importance is the extent to which these data can capture the kinds of information that are relevant to the estimation of extreme poverty. The wide variety of applications of these data attest to the diversity of information on economic, socio-economic, and environmental attributes that can be collected through crowd-sourcing approaches. This includes information on local prices (see Box 5), access to quality health and education services, water and air quality, and social attitudes and perceptions. This kind of information does not typically support the direct estimation of extreme poverty through consumption- or income-based measures, but is nonetheless of relevance for understanding broader factors that may be driving poverty rates or experiences of poverty. There is considerable flexibility in how crowd-sourcing initiatives may be designed to suit a range of purposes related to poverty measurement.

The very act of participating in citizen reporting may bring wider benefits too. Crowd-sourcing initiatives have the potential to raise citizen agency in the process of data production and use, which may have some intrinsic benefit. Greater ownership over data and awareness of results can also help to increase transparency and trust in reported results, as well as enabling citizens to hold leaders to account. In contexts where trust in official sources may be low, crowd-sourcing allows people to directly provide some information that may be missing, or corroborate existing data. For example, a crowd-sourcing initiative called FLOAT Beijing has enabled citizens to track air quality directly at a time when reliable official data on air pollution were perceived to be lacking (The DataShift, 2015b).

Box 6: Collecting geo-referenced data that can be used in 'data sandwiches'

While crowd-sourced data are not commonly used in isolation for the measurement of poverty, they are gaining increasing traction as an input to hybrid methods that combine different forms of data and analytical techniques to estimate poverty. A promising example of the integration of crowd-sourced data with other methods is the combination of geo-referenced data from a crowd (similar to satellite imagery data) with other sources.

Tingzon *et al*. (2019) incorporate crowd-sourced data from OpenStreetMap (OSM) in a model of asset-based wealth, together with NTL data, to predict poverty in the Philippines. OSM is a well-known open—source platform for crowd geospatial data sourced from individual volunteers across the world. Anyone can contribute, without any specialised knowledge or training, by adding information or making an edit (such as describing features that are only visible on the ground). In the Tingzon *et al.* study, information about roads, buildings, and other points of interest were extracted from these data, and were integrated within a complex statistical learning approach. Ground truth data from the nationally representative 2017 Philippines DHS survey were used to train a model of asset-based poverty at the 'cluster' level (a grouping of around two to 44 households).

The study findings show that this approach explained up to 63% of the variation in an asset-based wealth indicator. This level of explanatory power compared well with an alternative model built using NTL data and daytime satellite imagery obtained from Google Static Maps, which is proprietary. Yet while satellite data may be proprietary, the OSM and NTL model can be derived entirely from data that are freely and publicly available.

While crowd-sourced data have become relatively well-established in several fields, notably in the environmental sciences, their use for direct measurement of poverty indicators is not widespread. This may be partly due to limitations with the data. One primary issue is related to the fact that contributors typically self-select into participation, rather than being directly approached. In the absence of strict sampling protocols, this means that the resulting data cannot be treated as statistically representative and used to infer statistics over a specified target population.[20] Moreover, the extremely poor and vulnerable may be exactly the kinds of people who are excluded from crowd-sourced datasets through lack of access to means of contributing, or lack of awareness of opportunities to do so.

The uncertain quality of crowd-sourced data has also been the subject of debate. In addition to the issues around sample selection described above, there have also been concerns within the scientific community around a perceived lack of agreed quality standards governing the production of data, and limited opportunities to quality assure individual contributions (Alabri and Hunter, 2010). Nonetheless, quality assurance frameworks (see Meek *et al.*, 2014)) and communities of practice have been established to help assess and improve the quality of crowd-sourced data. The Citizen Science Association Data and MetaData working group, founded in 2015, has been established to help build a set of standards around the production of citizen-reported data and meta-data, and to facilitate greater coordination of initiatives, harmonisation of standards, and improved interoperability of these sources with traditional data sources.

Finally, for all the apparent promise of citizen-reported data as a democratising force in the data ecosystem, there have also been critiques as to whose interests are really served by the continued growth of these platforms. Critics have raised concerns over the potential that the individuals who produce crowd-sourced data may be exploited by the platforms that host their contributions, or the researchers who benefit from access to cheap data (Gleibs, 2017; Fort *et al.*, 2011). The experiences of individual contributors working for crowd-sourcing platforms have been the subject of some negative media attention (Semuels, 2018; Utpal, 2015). A focus of much of this criticism has been Amazon Mechanical Turk (Mturk), a well-established crowd-sourcing platform that enables ‘Requesters’ (researchers) to quickly seek contributions to structured tasks by participating ‘Turkers’. The concern is that the model has created potentially harmful power structures, as many participants have come to rely on the platform as a meaningful source of income, rather than engaging as casual contributors (Gleibs, 2017). With the emergence of ‘professional crowd-workers’, critics have argued that ‘Mturk has developed into an unregulated labour market with very low wages, incomplete contracting, weak access to enforcement and a disciplining role of reputation, in which workers are denied basic workplace rights and the community has no recourse for employer wrong-doing’ (Gleibs, 2017). Thus, while the opportunity to participate in citizen science initiatives may be empowering for some, the ethical practices around the data generation process may demand further scrutiny.

---

[20] In situations where users are providing information about themselves – that is, where individuals are the unit of analysis – the issue of potential selection bias means that the sample captured in the data cannot be treated as representative of the wider population. In situations such as the one described in Box 6, where contributors are providing information about their external environment, the issue of contributor self-selection means that the aspects of the environment described may not align with the representative experience. For example, if crowd-sourced data are being used to collect price information, the data collected may disproportionately reflect prices faced by contributors to crowd-sourcing platforms, which may differ from those faced by non-contributors (if, for example, they shop or acquire goods in different ways).

3.5.1 Summary

A summary of crowd-sourced data is provided in the table below.

Table 6: Summary of crowd-sourced data

Criteria	Summary comments
RELEVANCE	Crowd-sourcing is flexible and initiatives can be designed to suit a range of policymaker or researcher questions. The platforms are not typically well-suited to collecting the kind of information required to directly measure extreme poverty, according to the IPL measure or other consumption- or asset-based measures. However, information about a range of attributes that are related to the experience of poverty, or that may be drivers or correlates of poverty, can be collected through these kinds of data. This can include price data, which are one of the three key components of information needed for poverty measurement.
COST CONSIDERATIONS	Crowd-sourced data are cheap to collect and access.
FREQUENCY	Information can be provided at high frequency (depending on the volume and level of engagement of contributors).
SPATIAL RESOLUTION	Information can be provided at fine levels of spatial resolution (depending on the volume and dispersion of contributors).
ACCESSIBILITY	Crowd-sourced data are either publicly available or cheap to access.
DATA QUALITY	Quality-assuring data produced through crowd-sourced means, and meta-data standards for describing quality, have historically been challenges with crowd-sourced data. New initiatives to bolster data quality may hold promise in regard to improving rigour over time.
ABILITY TO MEASURE UNCERTAINTY	Confidence intervals cannot generally be applied to indicators derived from crowd-sourced data due to the non-probabilistic sample generation process. However, frameworks for describing data quality and meta-data standards continue to be developed to help assess and report on data quality.
ABILITY TO MAKE OUT-OF-SAMPLE PREDICTIONS	Crowd-sourced data are not typically representative of any well-described population, given that participants self-select into contributing.
REPLICABILITY	Crowd-sourced data are usually easily accessible, meaning that results can in theory be replicated.
ETHICAL CONSIDERATIONS	There are concerns about the potential for exploitation of contributors to crowd-sourcing platforms, through lack of appropriate regulatory frameworks or grievance redressal mechanisms. This varies between cases but requires careful consideration.

3.6 Mobile phone surveys to measure poverty proxies

Key messages:

- Remote surveys are **surveys carried out in LMICs using mobile phones**. In contrast to citizen-reported data, these surveys employ a sampling approach and do not rely on self-reporting. In contrast to the traditional face-to-face approach to poverty measurement, data are collected remotely via either phone calls or SMS messages.
- Interest in this data collection approach has significantly increased for three reasons. First, **costs per interview are significantly lower** than in face-to-face surveys. Second, **mobile phone ownership in LMICs has increased** significantly, which means that larger shares of the population can be reached in this way. Third, **the COVID-19 pandemic** has prevented many face-to-face interviews going ahead.
- In remote surveying, **interviews must be short**. This approach can't be used to deliver comprehensive consumption or expenditure modules to measure poverty. Researchers therefore have to investigate **alternative indicators and proxies related to povert**y that can be measured using this approach.
- In our review we found **few uses of remote surveys to measure poverty in LMICs**, which indicates that further research might be needed in this area. On the other hand, there has been a lot of activity since the beginning of the COVID-19 pandemic. The landscape around remote surveys is changing rapidly.

The expansion of the internet, and of mobile phone ownership and use globally, has created the opportunity to collect survey data through mobile phones, smartphones, and online interfaces, as an alternative to face-to-face data collection through traditional surveys or censuses. In the context of COVID-19, such remote data collection has become particularly relevant – for example, in the US with the so-called household pulse survey.[21] In LMICs, remote data collection refers mainly to mobile phone surveys, which can take several forms. This section deals primarily with surveys in which trained enumerators interview respondents over the phone, usually facilitated by computer-assisted telephone interviewing (CATI) software. Other variants of mobile phone surveys are set out in the box below.

Box 7: Forms of mobile phone survey

CATI surveys are surveys administered by trained enumerators who interview respondents over the phone and record their verbal answers to interview questions. Like CAPI software (which is used for in-person surveys) CATI is a tool for enabling electronic recording of survey responses, and supports a range of validation checks and filters to improve data quality. The distinctive feature of CATI software is a call management functionality, which allows interviewers to easily keep track of and schedule phone calls. This is essential when conducting a phone survey at scale, as repeated call-backs may be necessary to reach some respondents or to schedule a call at a time that suits them.

Interactive voice response (IVR) surveys use automated voice recordings to guide respondents through the questionnaire. Rather than using enumerators to administer questions and record responses, IVR surveys rely on respondents answering pre-recorded questions using buttons on the phone's keypad to progress through the interview.

[21] www.census.gov/householdpulse.

Box 7: Forms of mobile phone survey

SMS surveys collect survey responses through SMS messages, sent to respondents one at a time. A benefit of SMS surveys is that respondents can answer at their own convenience, and it is straightforward to provide incentives such as small credit top-ups to respondents in exchange for their participation. However, there are restrictions on the length of questionnaires that can be administered in this way, and SMS surveys are also unsuitable for respondent groups with high illiteracy rates (60 Decibels, 2020).

See 60 Decibels (2020) or Ballivian *et al.* (2015) for a comparison of different mobile-based survey approaches.

Phone surveys are an increasingly popular alternative to traditional survey data collection. Part of their appeal lies in their relatively low costs. While costs per interview do vary depending on the length and complexity of the instrument, the sample size, and the number of follow-ups, phone surveys can be substantially cheaper than in-person surveys. Costs per interview for several recent mobile phone surveys are summarised in an article in the World Bank's Development Impact blog series: the majority are under US$ 10 per interview (Himelein *et al*., 2020). This compares with an estimated cost per household survey of US$ 185 for DHS surveys, US$ 191 for LSMS, and US$ 67 for Multiple Indicator Cluster Surveys.[22] The COVID-19 pandemic has further concentrated attention on phone surveys as a way to overcome restrictions on in-person data collection and to mitigate risk. Phone surveys may also be used in conjunction with in-person surveys: for example, rapid phone surveys can be administered between rounds of a main in-person survey to collect multiple updates of key indicators over time.

The growing literature on mobile phone surveys makes plain that mobile phone surveys are very different to traditional in-person surveys. Among the differences is the average questionnaire length. It is generally accepted that respondents are less inclined to consent to long interviews carried out over the phone than they may be to those carried out in person. Guidance from the World Bank suggests that mobile phone surveys should last for around 15–30 minutes and should cover approximately 20 questions in total (Dabalen *et al*., 2016: 59). This is according to received wisdom that respondent fatigue will set in earlier for phone surveys compared with in-person surveys, leading to deteriorating quality of responses and a weakening of the respondent's willingness to cooperate (although the guidance suggests that, to date, there is a lack of strong supporting evidence for this claim). By contrast, in-person surveys (especially those containing detailed consumption modules used to estimate poverty) can take upwards of 90 minutes to complete. Constraints on survey length mean that the depth of information that can be collected by phone is generally less than that which can be collected through in-person questionnaires. This constraint may be partially overcome by carrying out phone surveys over more than one call, although this introduces other logistical challenges.

[22] https://paris21.org/sites/default/files/monitoringMDG-household-full.pdf.

This need for brevity has implications for the extent to which phone surveys can capture the information needed to measure extreme poverty. Long consumption modules of the kind that underpin traditional survey-based estimates of poverty (see Section 3.1) are not likely to be feasible over the phone. We are not aware of any existing phone surveys that have done this (Boznic *et al*., 2017: 5). Instead, phone surveys are more suitable for collecting poverty proxies that may be used to impute poverty estimates. This points to a promising role for mobile phone surveys in combination with survey-to-survey imputation techniques to estimate poverty (Boznic *et al*., 2017).

A potential disadvantage of mobile phone surveys is sample selection bias, since phone surveys can only reach respondents who are users of phones. As discussed in Section 3.4, phone ownership and use is not universal and is lower than average among communities most affected by extreme poverty. Estimates generated through mobile phone surveys must be carefully caveated in recognition of the sample that is included in the survey, and to what extent this may be considered representative of the population of interest.

In other respects, however, phone surveys may also offer a route to overcoming the issues of sample selection bias faced by traditional survey methods. For example, data collection over the phone provides a means of reaching populations that may be physically inaccessible. While face-to-face interviewing may be impossible for populations living in geographically hard to reach or insecure locations, phone surveys can be administered anywhere where there is a network.

3.6.1 Summary

A summary of mobile phone surveys is provided in the table below.

Table 7: Summary of mobile phone surveys

Criteria	Summary comments
RELEVANCE	Mobile phone survey questionnaires can be tailored to capture a variety of information. However, mobile phone surveys are usually much shorter than in-person household surveys, so there are constraints on the amount of information that can be collected. As a result, detailed consumption modules of the kind that in-person surveys may collect cannot be captured with any reasonable degree of quality through a mobile phone survey.
COST CONSIDERATIONS	Mobile phone surveys are much cheaper to carry out than in-person surveys.
FREQUENCY	Given their lower cost, mobile phone surveys may be conducted more frequently than in-person surveys. However, estimates cannot be produced as rapidly as other ‘big data’ sources that are collected on an almost continuous basis.
SPATIAL RESOLUTION	The level of spatial resolution depends on the sample size and design. Although mobile phone surveys are cheaper than in-person surveys, potentially enabling higher sample sizes, there is still likely to be some trade-off between sample size and logistical feasibility. This places a limit on how finely grained the resulting estimates can be.
ACCESSIBILITY	Administering mobile phone surveys to a high quality does require a certain level of technical expertise and equipment, but in general this is an accessible form of data collection that can be administered fairly easily.

Criteria	Summary comments
DATA QUALITY	There are some known risks to data quality. Sample attrition may be higher for phone surveys than it is for in-person surveys, and this attrition may not be random. This means that households that participate in mobile phone surveys may be systematically different from those that opt out, potentially leading to biased estimates. Issues that affect the quality of in-person surveys (such as recall bias, Hawthorne effects, or a pressure that respondents may perceive to provide certain answers to the enumerator) can also affect mobile phone surveys
ABILITY TO MEASURE UNCERTAINTY	Quantitative estimates derived from surveys where a probabilistic sampling process has been adopted can be assigned a confidence interval to describe the precision around the estimate. Non-sampling errors are hard to illustrate with any metric.
ABILITY TO MAKE OUT-OF-SAMPLE PREDICTIONS	If mobile phone surveys use a representative sampling approach, the estimates can be generalised to describe the wider reference population (with a given confidence interval). As with in-person surveys, the claim to representativity depends on the quality of the sample frame used to draw the sample, as well as the extent of sample attrition that is not random. As described above, the risk of non-random attrition may be higher for mobile phone surveys than for in-person surveys. In some respects, mobile phone surveys may be better able to capture certain population groups than in-person surveys. Populations living in conflict-affected or otherwise very remote locations that in-person survey teams struggle to reach may be omitted from traditional household surveys (thereby weakening representativity). The ability of mobile phone surveys to capture these groups is an advantage. In other respects, though, mobile phone surveys may be more likely to miss out some population groups that could be captured by in-person surveys. For example, communities with poor network coverage, or respondents who do not use or own mobile phones, are likely to be excluded. This means that final estimates may not be generalisable across these groups.
REPLICABILITY	As with in-person survey data, online repositories for storing survey data exist, and these enable other groups to download and make use of the data. However, this is not always done.
ETHICAL CONSIDERATIONS	As with in-person surveys, ethical approval is often required to conduct a phone survey, and there are a range of considerations that need to be taken into account. Phone surveys have some differences to in-person surveys. It may be harder to ensure that participant confidentiality is met, since the enumeration team has less control over the setting in which the respondent is conducting the interview (and whether this is quiet and private).

4 Opportunities and caveats of a new data ecosystem

Key messages:

- Our review revealed that there is a **very active and rapidly increasing literature** exploring the benefits and costs of using innovative data approaches to measure poverty in LMICs.
- While most of this literature emphasises the substantial potential of these approaches to close the 'poverty data gap', there are still **many open questions** relating to how to extract the best value out of this new data ecosystem, while dealing with relevant issues around (for example) data quality, ethical concerns, and biases.

We suggest that the following three main areas could be of particular interest for further investigation by DEEP:

- First, our review revealed that there is still comparatively little research on the **potential uses of online data, citizen-reported data, and remote surveys** for measuring poverty in LMICs. Exploring methods to deal with data that are not text-based or questionnaire-based is a particularly exciting – but experimental – frontier.
- Second, we also find that many approaches try to **combine and integrate data from a variety of 'new sources' and traditional data** in order to improve poverty estimates. Further research would is needed on how to best do this integration across data sources, in particular taking into account advanced statistical learning methods and focusing on extreme poverty as the target of estimation.
- Third, we suggest investigating **how such estimation and data collection approaches could be integrated into national statistics ecosystems** in a useful way. This implies tackling questions related to ethical concerns, data quality, trust in highly modelled results, and actual demand for new poverty measures. It could also include testing prototypes and capacity building in specific national statistics offices, who are often the main national institution responsible for producing, analysing, and publishing national statistics on poverty.

4.1 The motivation for a new data ecosystem

In this paper, we have aimed to provide an overview of how different innovative approaches are being used and experimented with in order to measure and research poverty in LMICs. In particular, we have focused our review on the exploration of non-traditional data sources – and innovative analytical techniques – that originate from the use and adoption of new digital technologies. This has involved reviewing approaches that examine data sourced from remote sensing technology, the data exhaust, online activities, crowd-sourcing efforts, and mobile phone surveys for their usefulness in generating insights into poverty in LMICs. We have also touched on statistical learning techniques that allow researchers to extract value from these data sources. In this section we reflect on what the findings of our review mean for the agenda that DEEP is trying to achieve, as well as for national statistics offices and other institutions that require access to high-quality and reliable information about extreme poverty.

First, it is important to emphasise that the application of almost all of the approaches reviewed in the context of this paper is motivated by what in Section 2.2 we labelled the ‘data gap problem’. In a nutshell, this contends that the traditional data ecosystem for measuring poverty in LMICs, comprising official statistics produced by government or international institutions using traditional data sources (surveys, censuses) provides key stakeholders with scarce data that allow them to only draw an incomplete picture of extreme poverty. In essence, there is a gap between the data this ecosystem supplies and the assumed data needs for formulating policies and design interventions that tackle extreme poverty effectively. This relates to geographies (estimates are not available at the right level of disaggregation or for the right locations), time periods (poverty is measured only intermittently), and poverty dynamics (we know relatively little about how the wealth of particular households and individuals changes over time). Given the cost of implementing traditional data collection via face-to-face surveys and censuses, the assumption is that these traditional data sources cannot solve this data gap problem on their own.

Enter the above-mentioned new data sources: the hope is that they, together with new statistical learning methods, can help to close this data gap, given that they are often available at higher frequency, with greater geographical coverage, higher spatial resolution, and lower cost than traditional data. Such data can hence be fed into statistical learning algorithms that are able to produce reliable high-frequency and high-resolution estimates or predictions of poverty for places and in time periods where traditional data cannot. In contrast to traditional poverty estimates, this puts statistical modelling at the core of these new approaches, and heavily bets on the promise of the ‘data revolution’: that statistical learning techniques (e.g. machine learning, neural networks, random forests etc.), together with large sets of new data and computational capacity, are becoming increasingly better at prediction tasks. The resulting poverty estimates can then be used to tackle poverty in a more effective and targeted manner.

It should be noted, then, that this narrative relies on a set of key assumptions. The first is that these statistical learning methods, together with new data sources, perform well for the purpose of measuring and investigating issues around extreme poverty. The second assumption is that key stakeholders, such as governments and international institutions, actually perceive the data gap to be as pertinent as described above, and there is hence a demand for this gap to be filled. The third assumption is that stakeholders have the ability to use, and trust in using, these new poverty measurement approaches to close the information gap, despite any existing ethical concerns and their heavy reliance on statistical modelling. Investigating these assumptions could be a key component of DEEP’s future agenda.

4.2 Opportunities and challenges of the reviewed approaches

Based on the findings of our review, we contend that approaches which integrate a variety of different data sources and analytical techniques are the most promising avenues for future research and practice. In Table 8 we try to visualise those areas where our review identified the opportunities and strengths, as well as the challenges and weaknesses, that the different approaches demonstrate with respect to the criteria included in the summaries at the end of each of the review sub-sections in Section 3. We highlight areas where our summaries indicate particular strengths in bright green and with four stars (****). Similarly, we highlight areas where our summaries indicate particular challenges in light red and with one star (*). In between those two extremes, we indicate assessments of criteria that could be considered rather challenging with light yellow and two stars (**) and those that could be considered as strengths with light green and three stars (***). The objective of this visualisation is not to provide an authoritative judgement on these different approaches – as one could certainly argue over some of the categorisations that we provide – but to provide a visual 'aide-memoire' in reference to the preceding review sections, and to guide our discussion here.

Overall, we find that new technologies offer a potential to improve poverty measurement and provide solutions to previously difficult measurement problems. They show particular promise with respect to the criteria of costs, frequency, and spatial resolution. Yet the review also shows that new technologies are not a silver bullet. Each approach is associated with its own limitations, challenges, and underlying assumptions, and there is no method that outperforms all others against all criteria in our assessment framework. Where prediction is the goal, predictive performance varies, depending on the target of estimation, the data used, and the choice of statistical model. In certain situations, and depending on the measurement objectives and priorities at hand, a single source of data can be all that is required to address the question. However, given the varied attributes of different approaches, the findings of this review suggest that new technologies can be used to their best advantage when they are thoughtfully combined and used as a complement to existing methods. This points towards the emergence of a new multi-faceted data ecosystem that changes the way that extreme poverty is measured, and the way that progress towards eradicating it is monitored.

Table 8: Visual overview of review results: challenges and opportunities

Criteria	Surveys and census	Remote sensing	Data exhaust	Online information	Citizen-generated data	Mobile phone surveys
RELEVANCE	****	*	**	***	***	***
COST CONSIDERATIONS	*	***	****	***	***	***
FREQUENCY	*	****	****	***	****	***
SPATIAL RESOLUTION	**	***	****	***	***	**
ACCESSIBILITY	***	***	*	**	***	**
DATA QUALITY	***	***	**	**	*	**
ABILITY TO MEASURE UNCERTAINTY	****	**	*	*	*	***
ABILITY TO MAKE OUT-OF-SAMPLE PREDICTIONS	***	**	**	**	*	***
REPLICABILITY	***	***	*	**	***	***
ETHICAL CONSIDERATIONS	***	**	*	*	**	***

Face-to-face sample surveys and census data will likely remain the backbone of this ecosystem for several reasons. Firstly, most of the methods reviewed in this paper rely on some form of survey data as 'ground truth' to train prediction models. Household surveys that contain consumption expenditure modules in particular will remain critical as they are the only means currently available of collecting data on household *per capita* consumption, which is fundamental to measuring extreme poverty. Their relevance is therefore unchallenged among the approaches reviewed here. In many of the studies reviewed in this paper, the aim is to build a model that accurately predicts this outcome for areas or time periods where such direct measurement via surveys is not available. To this end, surveys provide the vital 'training' data for such instances. In fact, as our review has shown, researchers point out that the availability of such survey data needs to increase in order to provide better 'poverty-labelled' data for training purposes.

In addition, given the wealth of variables that a survey questionnaire commonly collects data on, there is currently no substitute for the depth of insights that a household survey can provide on different forms of poverty, and their manifestations and determinants. None of the alternative data sources reviewed in this paper can replicate what a well-designed household survey can do in this respect. The ability to understand what drives poverty and how it affects people's lives is as crucial to informing policy and targeting as information about its prevalence and severity. This understanding will continue to be informed by survey data.

Moreover, the fact that such surveys follow established probability sampling rules, often using censuses as sampling frames, means that poverty estimates can be constructed with known and quantifiable uncertainty levels. Censuses can directly provide estimates of certain indicators (if not consumption expenditure-based poverty estimates) for entire sub-groups in the population of interest. This means that surveys and censuses remain crucial for validating results derived from other data sources and estimation techniques – in particular when trying to assess whether these techniques are producing biased results (for example, if the underlying data are not representative of the whole population).

Nonetheless, the review of innovative data sources shows that new sources of data, together with analytical innovations, offer significant potential. As the coding in Table 8 visualises, all the methods reviewed can be employed **more cheaply** than undertaking new survey or census data collection, and therefore can also be **carried out at higher frequency.** This enhances the possibilities for monitoring poverty more dynamically and adapting poverty reduction strategies or interventions to respond to changes over time. Approaches based on forms of data exhaust (such as CDR data) and sensors offer the possibility of **consistently higher spatial resolution** than household surveys. This is also crucial as it potentially enables a much more finely grained understanding of the distribution of poverty, allowing programmes to be better targeted.

While new methods do not offer the same flexibility for measuring outcomes that a survey can provide, they are capable of addressing a variety of relevant measurement needs indirectly. These sources can be utilised in prediction models to construct estimates of poverty through proxy measures. This includes non-big data sources, such as mobile phone surveys, which represent a cheaper alternative to household surveys for producing data on poverty proxies and asset ownership. Other sources (such as satellite data) also lend themselves well to providing **high-quality and frequently measured covariate data** that can be used within SAE methods to make out-of-sample predictions. There are also sources that can provide other relevant information on different dimensions and determinants of poverty. For example, the prominence of digital currency and payment platforms creates the potential to study **financial transactions and household economic behaviour** in a manner that has not previously been possible. Meanwhile, crowd-sourced data have been used to **measure prices and price fluctuations,** and are a potentially flexible form of data that could in theory be leveraged to provide information on a range of attributes. The examples outlined in this paper are just a snapshot of a considerable and expanding literature, where innovation is abundant. Nascent work, such as crowd-sourcing methods to collect more idiosyncratic forms of data (such as photos or voice recordings) through smartphones, may also soon become more mainstream.

All in all, these are exciting frontiers for research and practice and the DEEP programme provides a vital opportunity to harness this potential and to advance the cause of developing better measures of extreme poverty. It is also clear that the enthusiasm that big data sources and new analytical techniques has inspired must be tempered with a good understanding of their limitations, including ethical concerns. This is essential in order for new methods and data to be used thoughtfully and to their best advantage.

4.2.1.1 Possible areas for future research

We now reflect briefly on where this discussion may lead the DEEP consortium next. We identify important areas of investigation that can produce insights into, and learnings on, the key assumptions and challenges identified above. We broadly categorise these areas of investigation as pertaining to research on specific features of individual innovative approaches reviewed in this paper ('New Data'), research on the integration of different data sources ('Integrating Data') in order to build on their respective strengths, and research on the possibilities of inducing and scaling up the use of these approaches among governments and international institutions in LMICs ('Implementation').

New Data

As mentioned above, and as highlighted by the visualisation in Table 8, none of the methods and approaches reviewed in this paper offer a silver bullet in relation to the 'data gap' narrative presented above. Our review has highlighted both opportunities and challenges for them all. A significant section of the literature that we reviewed for this paper focuses only on a subset of approaches, namely integrating traditional survey and census data with remote sensing and the data exhaust in order to build on their benefits regarding coverage, frequency, and low cost. Other approaches appear to have been explored to a much lesser extent. We think that there could be value in further exploring the unique benefits from online information, citizen-reported data, and mobile phone surveys. In particular, our review shows that they share some of the cost and frequency benefits with other approaches, while at the same time allowing for the possibility of collecting data that could be more relevant to measuring extreme poverty than data from remote sensing or the data exhaust. The fact that smartphone ownership and internet penetration in LMICs is increasing significantly adds to the potential benefits of exploring such approaches further.

With respect to **citizen-generated data**, further research into how issues of data quality, bias, and uncertainty could be dealt with might significantly increase the attractiveness of collecting data via such approaches. This is of particular interest when taking into account that in other areas of research (e.g. environmental sciences and health), and for non-LMIC geographies, the usage of these types of data has been increasing. This means that there might be opportunities for learning from those experiences to improve the measurement and investigation of extreme poverty in LMICs. Investigating solutions to issues of bias and representativeness are similarly relevant for high-frequency **mobile phone surveys**. In addition, research here could focus on understanding the use of easy-to-collect proxy measures to track consumption poverty closely in LMICs via these surveys, which would link to the wider research agenda of how to shorten and make surveys in this area more efficient. With respect to **online data**, research could focus on the task of sorting systematically through the vast amount of irrelevant information to identify those data that specifically provide useful information on extreme poverty in different LMICs.

Finally, both for **mobile phone surveys** and **citizen-generated data**, an exciting but highly experimental area of research is investigating using non-questionnaire-based, but human-generated, data for the purposes of investigating and measuring extreme poverty. Smartphones offer the possibility of purposefully collecting data via photographs, videos, voice recordings, and other sensors (e.g. movements), and it can be investigated whether such data could be used to generate insights for DEEP's research agenda. Such research is currently being conducted in other areas (e.g. measuring child malnutrition using photographs (Revell, 2020)), but we did not identify any work in this area with respect to extreme poverty.

Integrating Data

Given our view that integrating methods offers the best means of employing different methods to their advantage, the question remains how best to do this. In fact, many of the approaches reviewed for this paper are instances of integrating different data sources, often combining geo-located survey or census data with some innovative data source (mainly remote sensing or data exhaust). We have labelled these approaches 'data sandwiches',[23] as they layer several different types of data on top of each other and then feed these into an

[23] Disclosure: this term was first coined by Prof. Andrew Tatem from Southampton University.

estimation model that produces a high-resolution output. One important role for DEEP may be to conduct further research and small pilots to deliver better insights into how different data sources and estimation methods in this area interact, how they can be optimally integrated to address specific objectives, what interoperability constraints may exist, and how they can be mitigated. This could imply establishing a learning agenda that enables testing different means of utilising the different sources of data available for each of the different DEEP countries specifically.

It should be noted that a large part of research activities in this area will need to include investigations into the different statistical learning methods that could be employed to integrate and analyse data for high-frequency and high-resolution prediction output. One particularly exciting area of research could be the usage of deep learning or transfer learning approaches to deal with the scarcity of poverty-labelled data – which combine data from different sources as described in Section 3.2 – and which build on some of the most recent advances in statistical learning.

One recurring theme across the approaches reviewed for this paper is that while 'poverty' is often the theme of investigation, the indicator that is being mapped or predicted at high resolution is not directly an estimate of extreme poverty defined with a *per capita* consumption cut-off. Often, the estimation target is income levels, wealth distributions, or other proxy measures (e.g. DHS wealth or asset indices). Investigating applications that more closely focus on extreme poverty, strictly defined with the IPL, and comparing this to the performance of other estimation targets, could also be of relevance to DEEP.

Implementation

Finally, a future area of research could be whether and how the innovative methods reviewed in this paper actually respond to a demand among international institutions and governments. As has been said, the assumption that motivates much of the literature included in our review is that closing the 'data gap' is actually in the interest of stakeholders in LMICs, and that there is a demand for doing so – in order to then act on this improved evidence. Yet little systematic research seems to have been carried out to find out whether and in what contexts this assumption actually holds true, and if not why not. Importantly, given the innovative (and perhaps unusual) and ethically sensitive nature of some approaches included here, issues around how to build institutional capacity, understanding, public trust, and safe usage of these methods will be particularly relevant. In essence, this research will be crucial in order to understand how the vision of an innovative new data ecosystem with a diversity of data sources to track poverty can actually become a reality.

In fact, a core role of DEEP may also be in providing practical support to developing the systems and capacity required among national statistical offices to implement some of the methods outlined in this paper. For example, one possible direction for this work in the future is developing integrated poverty prediction systems that funnel all data that are produced at a particular place (for example, a district), and training an algorithm to produce poverty estimates as the outcome. These models could be periodically fine-tuned using updated survey and census data, but essentially left to run continuously (or as frequently as is required), to dynamically update estimates of poverty as new data are fed in. They might make use of some of the transfer and deep learning approaches reviewed in this paper. DEEP could explore how such a system could be set up, what institutional set-up would be required (taking into account data access issues), and what the methodological challenges are, and it could provide support to the testing of such approaches and help in setting up a team who could be in charge of these.

5 References

60 Decibels (2020) *Remote Survey Toolkit*, Global Innovation Exchange, https://covid19innovationhub.org/resources/60-decibels-remote-survey-toolkit.

Addison, D. and Stewart, B. (2015) ‘Nighttime Lights Revisited: The Use of Nighttime Lights Data as a Proxy for Economic Variables’. *Policy Research Working Papers* 7496, World Bank, http://documents1.worldbank.org/curated/en/613891468195568648/pdf/WPS7496.pdf.

Alabri, A. and Hunter, J. (2010) ‘Enhancing the Quality and Trust of Citizen Science Data’, *Proceedings 6th IEEE International Conference on e-Science*, EScience, https://doi.org/10.1109/eScience.2010.33.

Alkire, S. and Samman, E. (2014) ‘Mobilising the Household Data Required to Progress toward the SDGs’, *Ophi working paper* no. 72, Oxford Poverty & Human Development Initiative, https://ora.ox.ac.uk/objects/uuid:01d904e5-628e-4733-ad4a-10eb59050490.

Alkire, S. and Santos, M.E. (2014) ‘Measuring Acute Poverty in the Developing World: Robustness and Scope of the Multidimensional Poverty Index’, *World Development* 59, pp. 251–74, https://doi.org/10.1016/j.worlddev.2014.01.026.

Athey, S. and Imbens, G.W. (2019) ‘Machine Learning Methods That Economists Should Know About’, *Annual Review of Economics* 11(1), pp. 685–725, https://doi.org/10.1146/annurev-economics-080217-053433.

Ballivian, A., Azevedo, J., and Durbin, W. (2015) ‘Using Mobile Phones for High-Frequency Data Collection’, in *Mobile Research Methods: Opportunities and challenges of mobile research methodologies*, Toninelli et al. (2015). https://doi.org/10.5334/bar.c.

Bansal, C., Jain, A., Barwaria, P., Choudhary, A., Singh, A., Gupta, A. and Seth, A. (2020) ‘Temporal Prediction of Socio-Economic Indicators Using Satellite Imagery’, in *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, pp. 73–81, CoDS COMAD, Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/3371158.3371167.

Bennett, M.M. and Smith, L.C. (2017) ‘Advances in Using Multitemporal Night-Time Lights Satellite Imagery to Detect, Estimate, and Monitor Socioeconomic Dynamics’, *Remote Sensing of Environment* 192, pp. 176–97, https://doi.org/10.1016/j.rse.2017.01.005.

Blumenstock, J.E. (2012) ‘Inferring Patterns of Internal Migration from Mobile Phone Call Records: Evidence from Rwanda’, *Information Technology for Development* 18(2), www.tandfonline.com/doi/abs/10.1080/02681102.2011.643209.

Blumenstock, J. (2019) 'The Knowns and Unknowns of Big Data and Poverty Alleviation', Center for Effective Global Action, https://medium.com/center-for-effective-global-action/joshua-blumenstock-the-knowns-and-unknowns-of-big-data-and-poverty-alleviation-d498c30063da.

Blumenstock, J. (2020) 'Machine learning can help get COVID-19 aid to those who need it most', Nature, 14 May 2020, www.nature.com/articles/d41586-020-01393-7.

Blumenstock, J. and Keleher, N. (2015) 'Statistical Evaluation of a Crowd-Sourced Market Information System in Liberia', Data-Intensive Development Lab, https://didl.berkeley.edu/portfolio/statistical-evaluation-of-a-crowd-sourced-market-information-system-in-liberia/.

Blumenstock, J., Cadamuro, G. and On, R. (2015) 'Predicting Poverty and Wealth from Mobile Phone Metadata', *Science* 350(6264), pp. 1073–76, https://doi.org/10.1126/science.aac4420.

Boznic, V., Katayama, R., Munoz, R., Takamatsu, S. and Yoshida, N. (2017) 'Prospects of Estimating Poverty with Phone Surveys: Experimental Results from Serbia', World Bank.

Burke, M., Driscoll, A., Lobell, D. and Ermon, S. (2020) 'Using Satellite Imagery to Understand and Promote Sustainable Development', *NBER Working Paper Series* 27879, https://web.stanford.edu/~mburke/papers/BurkeDriscollLobellErmon_2020.pdf.

Caeyers, B., Chalmers, N., and De Weerdt, J. (2012) 'Improving Consumption Measurement and Other Survey Data through CAPI: Evidence from a Randomized Experiment', *Journal of Development Economics* 91(1), pp. 19–33.

Center for Human Rights and Global Justice (2020) 'Philip Alston Condemns Failed Global Poverty Eradication Efforts', Center for Human Rights and Global Justice, NYU School of Law, https://chrgj.org/2020/07/05/philip-alston-condemns-failed-global-poverty-eradication-efforts.

Chakraborty, S., Venkataraman, A., Jagabathula, S. and Lakshminarayanan, S. (2016) 'Predicting Socio-Economic Indicators Using News Events', in *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 2016, pp. 1455-1464, https://doi.org/10.1145/2939672.2939817.

Dabalen, A., Etang, A., Hoogeveen, J., Mushi, E., Schipper, Y. and von Engelhardt, J. (2016) 'Mobile Phone Panel Surveys in Developing Countries: A Practical Guide for Microdata Collection', World Bank.

De Sherbinin, A. (2011) 'The Biophysical and Geographical Correlates of Child Malnutrition in Africa', *Population, Space and Place* 17(1), pp. 27–46.

DFID (2020) 'DFID Ethical Guidance for Research, Evaluation and Monitoring Activities', UK Department for International Development, www.gov.uk/government/publications/dfid-ethical-guidance-for-research-evaluation-and-monitoring-activities.

Di Clemente, R., Luengo-Oroz, M., Travizano, M., Xu, S., Vaitla, B. and González, M.C. (2018) 'Sequences of Purchases in Credit Card Data Reveal Lifestyles in Urban Populations', *Nature Communications* 9(1), p. 3330, https://doi.org/10.1038/s41467-018-05690-8.

Doll, C.N.H., Muller, J-P., and Elvidge, C.D. (2000) 'Night-Time Imagery as a Tool for Global Mapping of Socioeconomic Parameters and Greenhouse Gas Emissions', *AMBIO: A Journal of the Human Environment* 29(3), pp. 157–62, https://doi.org/10.1579/0044-7447-29.3.157.

Dong, C., Meyer, J., Shmueli, E., Bozkaya, B. and Pentland, A. (2018) 'Methods for Quantifying Effects of Social Unrest Using Credit Card Transaction Data', *EPJ Data Science* 7(1), pp. 1–25, https://doi.org/10.1140/epjds/s13688-018-0136-x.

Eagle, N. (2009) 'Txteagle: Mobile Crowdsourcing', in *Internationalization, Design and Global Development*, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, pp. 447–56, https://doi.org/10.1007/978-3-642-02767-3_50.

Ebener, S., Murray, C., Tandon, A. and Elvidge, C.C. (2005) 'From Wealth to Health: Modelling the Distribution of Income per Capita at the Sub-National Level Using Night-Time Light Imagery', *International Journal of Health Geographics* 4(1), p. 5, https://doi.org/10.1186/1476-072X-4-5.

Elbers, C., Lanjouw, J.O. and Lanjouw, P. (2003) 'Micro–Level Estimation of Poverty and Inequality', *Econometrica* 71(1), pp. 355–64, https://doi.org/10.1111/1468-0262.00399.

Elvidge, C.D., Baugh, K.E., Kihn, E.A., Kroehl, H.W., Davis, E.R. and Davis, C.W. (1997) 'Relation between Satellite Observed Visible-near Infrared Emissions, Population, Economic Activity and Electric Power Consumption', *International Journal of Remote Sensing* 18(6), pp. 1373–79, https://doi.org/10.1080/014311697218485.

Elvidge, C.D., Baugh, K.E., Anderson, S.J., Sutton, P.C., and Ghosh, T. (2012) 'The Night Light Development Index (NLDI): A Spatially Explicit Measure of Human Development from Satellite Data', *Social Geography* 7(1), pp. 23–35.

Elvidge, C.D., Sutton, P.C., Ghosh, T., Tuttle, B.T., Baugh, K.E., Bhaduri, B. and Bright, E. (2009) 'A Global Poverty Map Derived from Satellite Data', *Computers & Geosciences* 35(8), pp. 1652–60, https://doi.org/10.1016/j.cageo.2009.01.009.

Engstrom, R., Hersh, J. and Newhouse, D. (2017) 'Poverty from Space: Using High-Resolution Satellite Imagery for Estimating Economic Well-Being', *Policy Research Working Papers* 8284, World Bank, https://documents.worldbank.org/en/publication/documents-reports/documentdetail/610771513691888412/poverty-from-space-using-high-resolution-satellite-imagery-for-estimating-economic-well-being

Fort, K., Adda, G. and Cohen, K. (2011) 'Amazon Mechanical Turk: Gold Mine or Coal Mine?' *Computational Linguistics* 37, https://doi.org/10.1162/COLI_a_00057.

Gaddis, I. (2016) 'Prices for Poverty Analysis in Africa', *Policy Research Working Papers*, World Bank, https://doi.org/10.1596/1813-9450-7652.

Ghosh, T., Anderson, S., Powell, R.L., Sutton, P.C. and Elvidge, C.D. (2009) 'Estimation of Mexico's Informal Economy and Remittances Using Nighttime Imagery', *Remote Sensing* 1(3), pp. 418–44, https://doi.org/10.3390/rs1030418.

Ghosh, T., Powell, R.L., Elvidge, C.D., Baugh, K.E., Sutton, P.C. and Anderson, S. (2010) 'Shedding Light on the Global Distribution of Economic Activity', *The Open Geography Journal* 3(1), https://benthamopen.com/ABSTRACT/TOGEOGJ-3-147.

Gleibs, I.H. (2017) 'Are All "Research Fields" Equal? Rethinking Practice for the Use of Data from Crowdsourcing Market Places', *Behavior Research Methods* 49(4), pp. 1333–42, https://doi.org/10.3758/s13428-016-0789-y.

Global Pulse (2012) 'Big Data for Development: Challenges and Opportunities', United Nations Global Pulse, https://beta.unglobalpulse.org/wp-content/uploads/2012/05/BigDataforDevelopment-UNGlobalPulseMay2012.pdf.

Global Pulse (2014) 'Mining Indonesian Tweets to Understand Food Price Crises', United Nations Global Pulse, https://pdfs.semanticscholar.org/39a3/453f5f5697f78c0065da368b187631c196a2.pdf.

Gram-Hansen, B.J., Helber, P., Varatharajan, I., Azam, F., Coca-Castro, A., Kopackova, V. and Bilinski, P. (2019) 'Mapping Informal Settlements in Developing Countries Using Machine Learning and Low Resolution Multi-Spectral Data', in *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 361–368, AIES '19, Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/3306618.3314253.

GSMA (2018) 'Enabling Rural Coverage: Regulatory and Policy Recommendations to Foster Mobile Broadband Coverage in Developing Countries', GSMA, www.gsma.com/mobilefordevelopment/wp-content/uploads/2018/02/Enabling_Rural_Coverage_English_February_2018.pdf.

GSMA (2019a) 'The Mobile Economy Sub-Saharan Africa 2019 – GSMA Sub Saharan Africa', GSMA, www.gsma.com/subsaharanafrica/resources/the-mobile-economy-sub-saharan-africa-2019.

GSMA (2019b) 'The Mobile Gender Gap Report 2019 | Mobile for Development', GSMA, www.gsma.com/mobilefordevelopment/resources/mobile-gender-gap-report-2019/.

GSMA (2019c) 'State of the Industry Report on Mobile Money', GSMA, www.gsma.com/sotir/.

GSMA (2020) 'The Mobile Economy 2020', GSMA, www.gsma.com/mobileeconomy/wp-content/uploads/2020/03/GSMA_MobileEconomy2020_Global.pdf.

Gupta, A., Thies, W., Cutrell, E. and Balakrishnan. R. (2012) 'MClerk: Enabling Mobile Crowdsourcing in Developing Regions'', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1843–1852, CHI '12, Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/2207676.2208320.

Head, A., Tran, N., Manguin, M. and Blumenstock, J. (2017) 'Can Human Development Be Measured with Satellite Imagery?', in *Proceedings of ICTD '17*, Lahore, Pakistan, November 16–19, 2017, 11 pages, https://doi.org/10.1145/3136560.3136576.

Henderson, V.J., Storeygard, A. and Weil, D.N. (2012) 'Measuring Economic Growth from Outer Space', *American Economic Review* 102(2), pp. 994–1028, https://doi.org/10.1257/aer.102.2.994.

Hernandez, M., Hong, L., Frias-Martinez, V., Whitby, A. and Frias-Martinez, E. (2017) 'Estimating Poverty Using Cell Phone Data: Evidence from Guatemala', *SSRN Scholarly Paper* ID 2923520, Social Science Research Network, Rochester, NY, https://papers.ssrn.com/abstract=2923520.

Himelein, K., Eckman, S., Lau, C., McKenzie, D. (2020) 'Mobile Phone Surveys for Understanding COVID-19 Impacts: Part I Sampling and Mode', World Bank blogs, 7 April 2020, World Bank, https://blogs.worldbank.org/impactevaluations/mobile-phone-surveys-understanding-covid-19-impacts-part-i-sampling-and-mode

Holloway, J. and Mengersen, K. (2018) 'Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review', Remote Sensing 10(9), p. 1365, https://doi.org/10.3390/rs10091365.

Independent Expert Advisory Group on a Data Revolution for Sustainable Development (2014) 'A World that Counts: Mobilising the Data Revolution for Sustainable Development', Independent Expert Advisory Group on a Data Revolution for Sustainable Development, United Nations, www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013) 'An Introduction to Statistical Learning: With Applications', in Springer, New York, https://search.library.wisc.edu/catalog/9910207152902121.

Jean, N., Burke, M., Xie, M.W., Davis, M., Lobell, D.B. and Ermon, S. (2016) 'Combining Satellite Imagery and Machine Learning to Predict Poverty', *Science* 353(6301), pp. 790–94, https://doi.org/10.1126/science.aaf7894.

Jean, N., Wang, S., Samar, A., Azzari, G., Lobell, S. and Ermon, S. (2019) 'Tile2Vec: Unsupervised Representation Learning for Spatially Distributed Data', *Proceedings of the AAAI Conference on Artificial Intelligence* 33(01), pp. 3967–74, https://doi.org/10.1609/aaai.v33i01.33013967.

Jean, N., Xie, S.M. and Ermon, S. (2018) 'Semi-Supervised Deep Kernel Learning: Regression with Unlabeled Data by Minimizing Predictive Variance', in *Advances in Neural Information Processing Systems* 31, pp. 5322–5333, Curran Associates, Inc., http://papers.nips.cc/paper/7778-semi-supervised-deep-kernel-learning-regression-with-unlabeled-data-by-minimizing-predictive-variance.pdf.

Khamala, E. (2017) 'Review of the Available Remote Sensing Tools, Products, Methodologies and Data to Improve Crop Production Forecasts', Food and Agriculture Organization of the United Nations.

Khan, M.R., and Blumenstock, J. (2019) 'Multi-GCN: Graph Convolutional Networks for Multi-View Networks, with Applications to Global Poverty', in *Proceedings of the AAAI Conference on Artificial Intelligence*, www.aaai.org/ojs/index.php/AAAI/article/view/3836/3714.

Kuffer, M., Pfeffer, K., and Sliuzas, R. (2016) 'Slums from Space—15 Years of Slum Mapping Using Remote Sensing', *Remote Sensing* 8(6), p. 455, https://doi.org/10.3390/rs8060455.

Lämmerhirt, D., Gray, J., Venturini, T. and Meunier, A. (2018) 'Advancing Sustainability Together? Citizen-Generated Data and the Sustainable Development Goals', Global Partnership for Sustainable Development Data, Open Knowledge International, Public Data Lab.

Landgrebe, D. (1986) 'A Brief History of LARS', www.lars.purdue.edu/home/LARSHistory.html.

Lazer, D., Kennedy, R., King, G. and Vespignani, A. (2014) 'The Parable of Google Flu: Traps in Big Data Analysis', *Science* 343 (6176), pp. 1203–5, https://doi.org/10.1126/science.1248506.

Li, G., Cai, Z., Liu, X., Liu, J. and Su, S. (2019) 'A Comparison of Machine Learning Approaches for Identifying High-Poverty Counties: Robust Features of DMSP/OLS Night-Time Light Imagery', *International Journal of Remote Sensing* 40(15), pp. 5716–36, https://doi.org/10.1080/01431161.2019.1580820.

Lo, C.P. and Faber, B.J. (1997) 'Integration of Landsat Thematic Mapper and Census Data for Quality of Life Assessment', *Remote Sensing of Environment* 62(2), pp. 143–57, https://doi.org/10.1016/S0034-4257(97)00088-6.

Martinez, E., Hernandez Rubio, M., Maestre-Martínez, R., Arias, J.M., Patane, D., Zerbe, A., Kirkpatrick, R. and Luengo-Oroz, M. (2016) 'Measuring Economic Resilience to Natural Disasters with Big Economic Transaction Data', ArXiv.

Meek, S., Jackson, M. and Leibovici, D. (2014) 'A Flexible Framework for Assessing the Quality of Crowdsourced Data', *Connecting a Digital Europe through Location and Place. Proceedings of the AGILE'2014 International Conference on Geographic Information Science, Castellón, June, 3-6, 2014.*

Niu, T., Chen, Y. and Yuan, Y. (2020) 'Measuring Urban Poverty Using Multi-Source Data and a Random Forest Algorithm: A Case Study in Guangzhou', *Sustainable Cities and Society* 54(March), p. 102014, https://doi.org/10.1016/j.scs.2020.102014.

Njuguna, C., and McSharry, P. (2017) 'Constructing Spatiotemporal Poverty Indices from Big Data', *Journal of Business Research* 70(January), pp. 318–27, https://doi.org/10.1016/j.jbusres.2016.08.005.

Noor, A.M., Alegana, V.A., Gething, P.W., Tatem, A.J. and Snow, R.W. (2008) 'Using Remotely Sensed Night-Time Light as a Proxy for Poverty in Africa', *Population Health Metrics* 6(1), p. 5, https://doi.org/10.1186/1478-7954-6-5.

OHCHR (2020) 'Report on the parlous state of poverty eradication', United Nations Office for the High Commissioner for Human Rights, www.ohchr.org/EN/Issues/Poverty/Pages/parlous.aspx

Okwi, P.O., Ndeng'e, G., Kristjanson, P., Arunga, M., Notenbaert, A., Omolo, A., Henninger, N. *et al.* (2007) 'Spatial Determinants of Poverty in Rural Kenya', *Proceedings of the National Academy of Sciences* 104(43), pp. 16769–74, https://doi.org/10.1073/pnas.0611107104.

Osgood-Zimmerman, A., Millear, A.I., Stubbs, R.W., Shields, C., Pickering, B.V., Earl, L., Graetz, N. *et al*. (2018) 'Mapping Child Growth Failure in Africa between 2000 and 2015', *Nature* 555(7694), pp. 41–47, https://doi.org/10.1038/nature25760.

Paruthi, G., Frias-Martinez, E. and Frias-Martinez, V. (2016) 'Peer-to-Peer Microlending Platforms: Characterization of Online Traits', in *2016 IEEE International Conference on Big Data (Big Data)*, pp. 2180–89, https://doi.org/10.1109/BigData.2016.7840848.

Perez, A., Ganguli, S., Ermon, S., Azzari, G., Burke, M. and Lobell, D. (2019) 'Semi-Supervised Multitask Learning on Multispectral Satellite Images Using Wasserstein Generative Adversarial Networks (GANs) for Predicting Poverty', ArXiv, http://arxiv.org/abs/1902.11110.

Perez, A., Yeh, C., Azzari, G., Burke, M., Lobell, D. and Ermon, S. (2017) 'Poverty Prediction with Public Landsat 7 Satellite Imagery and Machine Learning', ArXiv, http://arxiv.org/abs/1711.03654.

Revell, T. (2020) 'An AI can now tell how malnourished a child is just from a photo', New Scientist, 18 May 2018, www.newscientist.com/article/2169382-an-ai-can-now-tell-how-malnourished-a-child-is-just-from-a-photo/#ixzz6saXwjhI4

Rocher, L., Hendrickx, J.M. and de Montjoye, Y-E. (2019) 'Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models', *Nature Communications* 10(1), p. 3069, https://doi.org/10.1038/s41467-019-10933-3.

Sedda, L., Tatem, A.J., Morley, D.W., Atkinson, P.W., Wardrop, N.A., Pezzulo, C., Sorichetta, A. *et al.* (2015) 'Poverty, Health and Satellite-Derived Vegetation Indices: Their Inter-Spatial Relationship in West Africa', *International Health* 7(2), pp. 99–106, https://doi.org/10.1093/inthealth/ihv005.

Semuels, A. (2018) 'The Internet Is Enabling a New Kind of Poorly Paid Hell', *The Atlantic*, 23 January 2018, www.theatlantic.com/business/archive/2018/01/amazon-mechanical-turk/551192/.

Serajuddin, U., Uematsu, H., Wieser, C., Yoshida, N. and Dabalen, A. (2015) 'Data Deprivation: Another Deprivation to End'', *SSRN Scholarly Paper* ID 2600334, Social Science Research Network, Rochester, NY, https://papers.ssrn.com/abstract=2600334.

Sheehan, E., Meng, C., Tan, M., Uzkent, B., Jean, N., Burke, M., Lobell, D. and Ermon, S. (2019) 'Predicting Economic Development Using Geolocated Wikipedia Articles', in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2698–2706, KDD '19, Association for Computing Machinery, Anchorage, AK, USA, https://doi.org/10.1145/3292500.3330784.

Silver, L. (2019) 'Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally', Pew Research Center's Global Attitudes Project (blog), 5 February 2019, www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/.

Steele, J., Sundsoy, P.R., Pezzulo, C., Alegana, V., Bird, T., Blumenstock, J., Bjelland, J. *et al*. (2017) 'Mapping Poverty Using Mobile Phone and Satellite Data', *Journal of the Royal Society – Interface* 14(127), https://doi.org/10.1098/rsif.2016.0690.

Sustainable Development Solutions Network (2015) 'Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development', Sustainable Development Solutions Network, https://opendatawatch.com/news/data-for-development-a-needs-assessment-for-sdg-monitoring-and-statistical-capacity-development/.

Sutton, P.C., Elvidge, C.D. and Ghosh, T. (2007) 'Estimation of Gross Domestic Product at Sub-National Scales Using Nighttime Satellite Imagery', *International Journal of Ecological Economics & Statistics* 8(S07), pp. 5–21.

Tang, B., Sun, Y., Liu, Y. and Matteson, D.S. (2018) 'Dynamic Poverty Prediction with Vegetation Index', OpenReview.net, https://openreview.net/forum?id=Hye0abRRtX.

The DataShift (2015a) 'What Is Citizen-Generated Data and What Is the DataShift Doing to Promote It?', The DataShift, http://civicus.org/images/ER%20cgd_brief.pdf.

The DataShift (2015b) 'Float Beijing: Citizen-Generated Data on Air Quality', The DataShift, http://civicus.org/thedatashift/wp-content/uploads/2015/07/Float-Beijing-case-study.pdf.

Thomson, D.R., Rhoda, D.A., Tatem, A.J. and Castro, M.C. (2020) 'Gridded Population Survey Sampling: A Review of the Field and Strategic Research Agenda', *International Journal of Health Geographics* 19(34), DOI: 10.1186/s12942-020-00230-4

Tingzon, I., Orden, A., Sy, S., Sekara, V., Weber, I., Fatehkia, M., Herranz, M.G. and Kim, D. (2019) 'Mapping Poverty in the Philippines Using Machine Learning, Satellite Imagery, and Crowd-Sourced Geospatial Information', in *AI for Social Good ICML 2019 Workshop*, https://pdfs.semanticscholar.org/9d96/bbd1bab6f66015096336052bd86662e14c6d.pdf.

Utpal, D. (2015) 'My Experience as an Amazon Mechanical Turk (MTurk) Worker', LinkedIn, www.linkedin.com/pulse/my-experience-amazon-mechanical-turk-mturk-worker-utpal-dholakia.

Varshney, K.R., Chen, G.H., Abelson, B., Nowocin, K., Sakhrani, V., Xu, L. and Spatocco, B.L. (2015) 'Targeting Villages for Rural Development Using Satellite Image Analysis', *Big Data* 3(1), pp. 41–53.

World Bank (1990) 'World Development Report 1990 : Poverty', World Bank, https://openknowledge.worldbank.org/handle/10986/5973.

World Bank (2017) 'Monitoring Global Poverty: Report of the Commission on Global Poverty', World Bank, https://openknowledge.worldbank.org/bitstream/handle/10986/25141/9781464809613.pdf.

World Bank (2020) 'Projected Poverty Impacts of COVID-19 (Coronavirus)', World Bank, www.worldbank.org/en/topic/poverty/brief/projected-poverty-impacts-of-COVID-19.

Xie, M., Jean, N., Burke, M., Lobell, D. and Ermon, S. (2016) 'Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping', Arxiv, http://arxiv.org/abs/1510.00098.

Xue, J. and Su, B. (2017) 'Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications', *Journal of Sensors*, https://doi.org/10.1155/2017/1353691.

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S. and Burke, M. (2020) 'Using Publicly Available Satellite Imagery and Deep Learning to Understand Economic Well-Being in Africa', *Nature Communications* 11(1), p. 2583, https://doi.org/10.1038/s41467-020-16185-w.