**Using big data for evaluating development outcomes:**

**a systematic map**

**Francis Rathinam**, 3ie

**Sayak Khatua**, 3ie

**Zeba Siddiqui**, 3ie

**Manya Malik**, 3ie

**Pallavi Duggal**, 3ie

**Samantha Watson**, Flowminder

**Xavier Vollenweider**, Flowminder

**Draft report**

**30 May, 2020**

****

**Contents**

[Executive summary 3](#_Toc42031660)

[1. Introduction 5](#_Toc42031661)

[1.1. The issue: data challenges 5](#_Toc42031662)

[1.2. Objectives and the research questions 6](#_Toc42031663)

[1.3. Scope of work 7](#_Toc42031664)

[1.4. Overview 8](#_Toc42031665)

[2. Methodology: Definitions, Inclusion and Exclusion 8](#_Toc42031666)

[2.1. Methodological approach 8](#_Toc42031667)

[2.2. Criteria for including the studies 8](#_Toc42031668)

[2.3. Sources of big data 10](#_Toc42031669)

[2.4. Outcomes 11](#_Toc42031670)

[2.5. Evaluation of potential methodological issues 13](#_Toc42031671)

[2.6. Evaluating reporting on privacy and ethical considerations 14](#_Toc42031672)

[2.7. Exclusion criteria 15](#_Toc42031673)

[3. Results 15](#_Toc42031674)

[3.1. Search, screening and coding 15](#_Toc42031675)

[3.1.1. Searching and screening 15](#_Toc42031676)

[3.1.2. Coding of included studies 16](#_Toc42031677)

[3.1.3. Mapping of the studies 16](#_Toc42031678)

[3.2. Characteristics and trends of the evidence base 17](#_Toc42031679)

[3.2.1. Distribution of studies across data sources 21](#_Toc42031680)

[3.2.2. Distribution of studies across development themes 23](#_Toc42031681)

[3.2.3. Units of observation 24](#_Toc42031682)

[3.2.4. Studies using mixed-methods 25](#_Toc42031683)

[3.2.5. Studies with a rural or urban focus 25](#_Toc42031684)

[3.2.6. Studies by fragile context 26](#_Toc42031685)

[3.3. Appraising potential methodological biases, risks and limitations 27](#_Toc42031686)

[3.3.1. Reporting on methodological challenges and data transparency 27](#_Toc42031687)

[3.3.2. Reporting on privacy and ethical challenges 28](#_Toc42031688)

[4. Key findings, discussion and lessons 29](#_Toc42031689)

[4.1. Evidence base and key gaps 29](#_Toc42031690)

[4.2. Analytical, ethical and practical challenges 31](#_Toc42031691)

[5. Using big data in IEs: potentials and challenges 34](#_Toc42031693)

[5.1. Using satellite data in IEs 34](#_Toc42031694)

[5.2. Using CDR analytics to inform disaster management 36](#_Toc42031695)

[6. Limitations of this study and potential next steps 38](#_Toc42031696)

[6.1. Next steps 39](#_Toc42031697)

[7. Conclusions 39](#_Toc42031698)

[References 40](#_Toc42031699)

[Appendix A: OECD definition of fragile states 44](#_Toc42031700)

[Appendix 2: Details of Sub-maps 47](#_Toc42031701)

[Appendix 3: Search strategy and the data bases searched 49](#_Toc42031702)

[Appendix 4: Screening protocol and data extraction 54](#_Toc42031703)

[Appendix 5: Machine learning and manual screening 59](#_Toc42031704)

[Appendix 6: Systematic review appraisal tool and summary 60](#_Toc42031705)

[Systematic review appraisal 62](#_Toc42031706)

[Appendix 7: Additional tables and figures 65](#_Toc42031707)

[Appendix 8: The systematic map of big data sources and outcomes 68](#_Toc42031708)

[Appendix 9: List of included studies 69](#_Toc42031709)

# Executive summary

Data challenges in international development are stark, especially in developing country contexts. Traditional data collection can be costly, target populations may be inaccessible, phenomena cannot always be directly observed and interviewing people could be unethical, dangerous or impossible. Budget constraints can limit the available sample size, information on covariates, the level of aggregation and the frequency of data collection.

Spatially and temporally relevant ‘big data’ which does not require data collection in the field, has the potential to provide insights into people’s economic, social, behavioural and political lives, and hence could be used in measuring key development outcomes. Big data consists of human generated data including online searches, social media, citizen reporting or crowdsourced data, process-mediated data like mobile phone call detail records (CRD), commercial transactions data and machine-generated data from satellites, sensors, or drones. The primary value of big data is that it is possible to measure outcomes that could not previously be measured using household surveys at the required temporal and spatial scale. The potential of big data to answer causal attribution, however, is still not widely understood, especially in low and middle income countries (L&MICs).

The report is based on a map of the studies using big data and its objective is to discuss methodological, ethical and practical constraints relating to the use of big data. The systematic map includes impact evaluations (IEs) that use big data to evaluate development outcomes, systematic reviews of big data IEs and other measurement studies that innovatively use big data to measure and validate any development outcomes. This study also explores the sectoral and geographical spread of big data's use in international development.

This map includes studies written in English and published between 2005 and 2019, regardless the target country's income level or population's status. We provide detailed breakdowns in the map for different country income classifications, fragile contexts, and population characteristics. From the initial list of 17,393 studies, we arrived at final list of 437 studies which included 48 impact evaluations, 381 measurement studies and 8 systematic reviews.

**Key findings and lessons**

* **There is considerable potential for measuring various development indicators using big data.** The number of measurement studies across development themes indicates the potential for big data to measure development outcomes. These measurement studies serve as a proof-of-concept for evaluators who look for innovative ways to measure development outcomes.
* **There is potential for more IEs on development interventions**. The map shows that the number of IEs which use big data to measure outcomes or control variables is growing fast. However, there are fewer IEs than measurement studies, and the extent of their thematic and geographical coverage is limited. IEs seem to be concentrated around environmental sustainability, economic development and urban development.
* **Satellite data is used the most.** The use of satellite data for IEs and measurement studies has been facilitated by the availability of pre-processed satellite data, and new machine learning techniques and increased computational capacity to process the satellite images into meaningful measures of development outcomes. However, despite a number of high-profile measurement studies, call record details data has not been not used to rigorously evaluating any development outcome. Similarly, other human-sourced data and process-mediated data have been used only sparingly in IEs. This is a notable gap and a potential area for future exploration.
* **There are potential sectors and themes where systematic reviews will be useful**. Although the number of IEs is small, the map highlights a few potential thematic areas where systematic reviews will be informative. A systematic review of all IEs that have used satellite data across the sectors will help understand the potential and challenges in using satellite data for IEs. Similarly, there is a concentration of IEs on forest management. A systematic review with reference to the data sources used in rigorously evaluating forest cover, and the advantages and challenges thereof, may be useful.
* **There is unrealized potential to conduct studies in fragile contexts**. A number of studies using big data have been conducted in fragile contexts such as conflicts, humanitarian crises, disease outbreaks, natural disasters, and areas situated in difficult terrain. However, the IEs are concentrated around conflict and difficult terrain. The number of measurement studies indicate the potential for more IEs in the fragile contexts.
* **Ethical concerns and transparency issues are substantial**. Ethical issues related to informed consent, data privacy, data security and unintended exclusion are severe for some of the sources of big data. However, this report shows that very few studies report on ethical issues related to using big data. This report calls for IRB review specifically designed for ethical issues related to big data. Similarly, very few studies have reported on data quality and fewer studies have data publically available for replication.
* **Some capacity constraints are acute**. Computational capacity is constrained and technical expertise on large-scale big data analysis is siloed. This report calls for donors to facilitate more interaction among data scientists and development evaluators for collaborations and learning.

This map shows that big data can contribute to the evidence base in development sectors where evaluations are often infeasible due data issues. One of the key ‘absolute gaps’ that the map has identified is that there are fewer IEs than measurement studies. Given the fast-growing availability of big data and improving computation capacity, there is a great potential for the use of big data in future IEs. However, there are several analytical, ethical, and logistical challenges that may hinder the use of big data in evaluations. This report calls for standards to be set for the reporting of data quality issues, data representativeness, and data transparency. More interaction is needed between big data analysts, remote sensing scientists, and evaluators.

# Introduction

## **The issue: data challenges**

Policymakers need access to reliable data to evaluate development outcomes and decide on future resource allocation. Governments, multilateral organisations and other development players in L&MICs use censuses, nationally-representative household surveys, other household surveys and administrative data to evaluate development programmes and policies. With the increasing complexity of development programmes, there is a need to collect a vast array of output, outcomes and contextual variables to robustly assess the impact. However, significant data collection challenges remain. Data challenges for impact evaluations (IEs) include limitations on sample size and power due to budget constraints, inaccessible or difficult to reach sections of target populations, measurement errors due to recall bias, inadequate frequency and level of aggregation, inadequate information on controls and covariates, data collection lag times and difficulties in measuring long-term impact[[1]](#footnote-2) (Wassenich, 2007). Further, in some contexts, like conflicts and humanitarian emergency situations, data collection is often impossible. The data gaps and challenges are particularly significant for the populations and countries where the need for evidence-informed policy decisions are perhaps the greatest (Gaarder and Annan, 2013). Another key shortcoming of survey data is inadequate aggregation at subnational administrative units such as districts, counties or villages, inhibiting evaluation of programmes with spatial attributes.

Big data offers great potential for answering some of these data needs, and more importantly answering the causal questions around which policies or interventions work, including in contexts where traditional methods of data collection are challenging. The UN Global Pulse (2013) defines big data as being digitally generated (as opposed to digitised manually), passively produced (a by-product of digital services, transactions and interactions), automatically collected, and geographically and temporally trackable. Although there is no formal definition for big data, currently the term is characterised by the 3Vs—high volume, velocity, and variety. Satellite images, sensors and drones, mobile phone call detail records, commercial transactions data, online searches, social media, citizen reporting or crowdsourced data are the sources of big data.

Integrating big data with traditional household surveys and administrative data can complement data availability, quality, granularity, accuracy and frequency, and help measure development outcomes temporally and spatially in a number of new ways (York and Bamberger, 2020; BenYishay, et.al., 2018; Salganik, 2017; Lokanathan et.al., 2017; and UN Global Pulse. 2016). For example, satellite images and mobile call records have been used in mapping poverty (Jean, et al, 2016; Blumnenstock et al, 2015), disaster response (Lu et al, 2012; Wilson et al, 2016) and food security (Decuyper et al, 2014). Web searches and social media were used in predicting unemployment and crime instances (Xu et al, 2013; Gerber, 2014).

While big data is increasingly used for tracking indicators and monitoring development progress on Sustainable Development Goals (UN, 2012; Vaitla, 2014; Lokanathan et.al, 2017), available data is less often utilised to address causal questions about the effects of specific policies and programmes. Big data can contribute to answering some of the causal questions around which interventions work. Big data prediction models can generate proxy estimates for key development outcomes such as wealth, human development, infrastructure quality, forest cover, and more, which can be used in experimental (Jayachandran et.al, 2016; Peillegrini, 2019) and quasi-experimental studies (BenYishay et.al, 2018; Jaiswal, 2019). Satellite images such as night light, crop intensity, water availability, land use, proximity to services, and physical attributes such as elevation or slope can be used in IEs as a direct measures of outcomes or as covariates. Furthermore, big data can be used for measuring and evaluating the long-term impacts of policies and programmes, conducting ex-post evaluations, and estimating spatial heterogeneity. For example, satellite data is available at least as far back as 1993 for all places (high resolution pictures are available for the entire globe at a granularity as low as 1×1 meter), allowing measurement of long-term impacts. This can help fill the gaps in evidence that cannot be addressed by traditional data sources.

The potential of big data to answer causal attribution, however, is still not widely understood or used, especially in L&MICs (York and Bamberger, 2020). In this context, a systematic collection of various sources of big data, and ways of measuring and evaluating development outcomes will be a great value addition to the development community’s effort to contributing to evidence informed policy making.

In this paper, we look at impact evaluations, systematic reviews (SRs) and measurement studies[[2]](#footnote-3) that use big data to evaluate development outcomes with a special focus on fragile contexts. The study highlights the new sources of data, how these new data sources can be used for measuring development outcomes innovatively, and how these new measures can be used in IEs. We map different sources of big data onto development outcomes, based on sustainable development goals (SDGs), to identify the current evidence base and its gaps.

## **Objectives and the research questions**

The overarching aim of this report is to inform policymakers and evaluators of existing evaluations based on big data and to provide a database of big data-based IEs and studies that could inform future IEs. Specifically, the objectives of the research are to:

* Identify rigorous impact evaluations, systematic reviews and the studies that have innovatively used big data to measure any development outcomes, with special reference to fragile contexts.
* Summarise current understanding of potential uses, pros and cons, reliability, biases, risks and ethical issues in using big data for measurement and evaluation of development outcomes.
* Generate interest and awareness among key stakeholders (evaluators, researchers, donors, practitioners, implementers and policymakers) of the potential as well as challenges of using big data.

This systematic map addresses the following questions:

* How have different types of big data and methods been used for measuring and evaluating development outcomes?
* How dispersed or concentrated is the use of big data across development goals and geographies?
* What are the potential biases, measurement reliability issues, pros and cons, risks and ethical issues in using big data for measuring and evaluating development outcomes?
* What are some of the unexplored but promising applications of big data for impact evaluations?

## **Scope of work**

For the purpose of this research, we define big data sources as digitally generated, passively produced and automatically collected data, as defined in UN Global Pulse (2013). The sources of big data include satellite images, sensors and drones, mobile phone call detail records, commercial transactions data, online searches, social media, citizen reporting or crowdsourced data. See table 2 for more details on various sources of big data adapted from UN Global Pulse (2012 and 2013), United Nations Economic Commission for Europe (2014) and Blazquez and Domenech (2018).

3ie evidence gap maps compile IEs and SRs. However, in this study, we include IEs and SRs as well as measurement studies—the studies that have innovatively used big data to measure and validate any development outcome. These are multi-disciplinary studies that use state-of-the-art methods from computer science and statistics to collect, clean and analyse big data for measuring development outcomes. For example, Jean et.al (2016) use transfer learning techniques, and daytime and night light data from satellite images to estimate consumption expenditure at the cluster (village) level to map poverty in five African countries—Nigeria, Tanzania, Uganda, Malawi, and Rwanda. While there are a number of such studies that have used big data for measuring various development outcomes, there are not many IEs that have used these innovative big data-based outcome measures. These measurement studies, we hope, would serve as proofs of concept for innovative use of machine learning and big data that can be used in future evaluations.

3ie defines impact evaluations as a “study of the attribution of changes in the outcome to the intervention[[3]](#footnote-4)”. For the purpose of this systematic map, we define big data based IEs as any experimental or quasi-experimental study that uses any form of big data to measure the outcomes of interest, and/or the confounding variables.

We use OECD definition of fragile contexts, which includes conflicts, institutional fragility, social fragility, environmental risks, health risks and climatic risks. This list is more inclusive than the list used by DFID and the World bank, which includes conflict, institutional and social fragility (DFID, 2016). Please see Appendix 1 for more details on the classifications and country list. We use OECD definition for classifying fragile contexts based on the following:

* Difficult terrain
* Natural Disaster
* Conflict or humanitarian crisis
* Chemical or radio-nuclear issues
* Disease outbreaks or epidemics

Using big data in evaluation poses a number of analytical challenges on issues including data quality, transparency, generalisability, and privacy and ethical challenges such as consent for using data and anonymization of the data. This report explores how the included studies dealt with these challenges as well.

## **Overview**

Section 2 of the report provides an overview of methodology, inclusion and exclusion criteria, and defines big data sources and development goals used in the report. Section 3 provides a detailed description of the trends and distribution of the included studies, and how the included studies dealt with analytical and ethical challenges. Section 4 summarises the key findings and pointers for future studies. Section 5 provides an overview of how effectively big data can be used in IEs. Section 6 discusses the limitations of the study, and the possible next steps for the evaluators, practitioners and the donors. Section 7 presents concluding remarks.

# Methodology: Definitions, Inclusion and Exclusion

## **Methodological approach**

We follow 3ie's methodology and process for evidence gap maps (Snilstveit et al., 2017). To create this map, we used systematic methods to identify any completed and ongoing IEs, SRs and big data measurement studies relevant to our research objectives. We conducted systematic searches and data extraction as described in Appendices 3 and 4. The studies identified are mapped on to the framework of big data sources and SDG outcomes to provide a visual display of the volume of and the trends in the evidence base. We also coded how the included studies have dealt with ethical and transparency related challenges. The systematic map is available through an online interactive platform on the 3ie website and allow users to explore the available evidence through different filtering options[[4]](#footnote-5). There are links to study summaries in the 3ie repositories wherever applicable, and confidence ratings for the SRs.

## **Criteria for including the studies**

Table 1 summarises the criteria we used for searching, screening and including the studies for the map.

**Table 1: Selection criteria for studies**

|  |  |
| --- | --- |
| **Category** | **Description** |
| **Population** | This map includes all population from all countries, but we provide break-downs for rural areas, urban areas, conflicted-affected persons and ethnic minorities. We also provide break-downs for L&MICs and fragile contexts separately. |
| **Sources of big data** | Big data may originate from any of the following sources:   * Human-sourced information   + Social Networks   + Internet searches   + Mobile data content   + Citizen Reporting or Crowdsourced data * Process-mediated data (traditional business systems and websites)   + Data produced by public agencies   + Data produced by businesses   + Cell phone call record details (CRD) * Machine-generated data (automated systems)   + Data from fixed sensors   + Data from mobile sensors (tracking)   + Data from Satellites |
| **Outcomes** | 1. Economic development and livelihoods 2. Agriculture and food security 3. Health and well-being 4. Quality of education 5. Governance and human rights 6. Water and Sanitation 7. Energy, industry and infrastructure provision 8. Urban development 9. Environmental sustainability 10. Partnerships for Goals |
| **Study Design** | **Impact evaluations:**   * Randomised controlled trial; * Regression discontinuity design; * Controlled before-and-after study using appropriate methods to control for selection bias and confounding, such as: propensity score matching or other matching methods; * instrumental variable estimation or other methods using an instrumental variable such as the Heckman Two step approach; * difference-in-differences; * or a fixed- or random-effects model with an interaction term between time and intervention for baseline and follow-up observations; * Natural experiments; * Other quasi-experimental studies inducing synthetic control studies; * Survey, laboratory or lab-in-the-field type experiments; * Cross-sectional or panel studies with an intervention and comparison group using methods to control for selection bias and confounding as described above.   **Measurement studies**   * We included the studies that innovatively used big data to measure and validate any development outcomes. These studies use big data to measure components that would have been difficult to measure using survey data.   **Systematic reviews**  We include only the reviews that specifically looked at studies which used big data to measure development outcomes, and explicitly described the search, data collection and synthesis methods according to a standard systematic review protocol such as the 3ie systematic review protocol. |

## **Sources of big data**

Innovations in the type of devices available for measurement (satellites and sensors), daily personal use (mobiles, wearables, Internet of Things, etc.) social interaction (blogs, Facebook, Twitter, WhatsApp, etc.) and recording business transactions digitally (CRD, e-transactions, mobile money, credit card payment, etc.) have led to an explosion of data that is automatically collected. However, there is no official definition of big data. McKinsey report defined it broadly as data “whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze” (Manyika et al, 2011).

UN Global Pulse (2013) defines big data for the purposes of development as being digitally generated (as opposed to digitised manually), passively produced (a by-product of digital services, transactions and interactions), automatically collected, and geographically or temporally trackable. While the size, velocity and veracity are all defining characteristics of big data, the definition relevant for impact evaluation is that these are non–sampled data, passively left behind by humans using digital devices and services or automatically collected by the services providers for the purpose other than statistical inference (Letouzé, 2015; UN Global Pulse, 2016). Hence, unlike the conventional survey data where the respondents say what they do or feel, big data captures what people actually do. The implication of this is that big data is non-reactive i.e. there is less likelihood of social desirability bias (Salganik, 2017). The other key characteristic of big data that matters for evaluation is that it is near real-time and is can be available in multiple frequencies (e.g. hourly, daily) over a long period of time. Table 2 provides the types of big data, sub-classifications, definitions and the sources.

Using these definitions, we include the following broad classification of big data[[5]](#footnote-6):

* Human-sourced information from social networks is provided voluntarily by users.
* Process-mediated data from traditional business systems and websites includes digitally-recorded business activities
* Machine-generated data from automated systems includes information from sensors and machines that measure and record events and situations in the physical world.

**Table 2: Sources of big data**

|  |  |
| --- | --- |
| Data type | Source of Data |
| Human-sourced information | |
| Social Networks | Text, meta-data and location data from social networking sites, opinion platforms, blogs, pictures, videos, etc. such as Twitter, Facebook, LinkedIn, YouTube, Wiki pages. |
| Internet searches | Internet text and internet search queries i.e. Google Trends; web logs (open data). |
| Mobile data content | Text messages |
| Citizen Reporting or Crowdsourced data | OpenStreetMap (OSM), Data from Humanitarian Data Exchange (HDX) platform, etc.  Collected from a large group of people who voluntarily provide information. This is a useful source for recording events and receiving people’s opinion and feedback on issues. Collected through hotlines, user-generated maps, marking of instances, events, etc. |
| Process-mediated data | |
| Data produced by public agencies | Medical records  Postal data  Tax data, etc. |
| Data produced by businesses | E-commerce transaction records and credit card transaction records, ATM transactions, mobile money transfers, savings and loan repayments (collected by the service provider as a part of regular business operation and monitoring. Proprietary and commercially sensitive data.)  Tolls and public transport card use data. |
| Cell phone call record details | Mobile call detail records (CDR) that provide metadata on when the call took place, cost, time and the recipient of the call; location of the caller and recipient; and the users’ mobility, social interaction and airtime transaction details; top up data. |
| Machine-generated data | |
| Data from fixed sensors | Home automation, Weather / pollution sensors, Traffic sensors / webcam, Security / surveillance videos/images, and activity records such as electricity meters (mostly administrative data collected by the authorities as a part of regular monitoring). |
| Data from mobile sensors (tracking) | Community or privately owned drones (common property or privately held data).  Mobile phone GPS (open data available from google).  Vehicle GPS and Self-tracking Apps (proprietary data) |
| Data from Satellites | Open data available from a number of sources. For example,   * DMSP-OLS * VIIRS * Landsat * ESA Land Cover, etc. |

Source: Adapted from UN Global Pulse (2012 and 2013), United Nations Economic Commission for Europe (2014) and Blazquez and Domenech (2018).

## **Outcomes**

We use the SDGs as the basis for identifying the outcome categories, similar to Phillips et.al. (2017)[[6]](#footnote-7). We have regrouped them into smaller thematic groups based on complementarities between the sectors (White, et.al forthcoming). Table 3 below provides the definition of each of the outcomes as defined in UN (2019). Appendix 2 provide more details on the sub-maps where further breakdowns have been provided for key sub-classifications where relevant.

**Table 3: Outcome categories and definitions**

|  |  |
| --- | --- |
| **Category** | **Definition** |
| Economic development and livelihoods  (SDG 1 & 8) | End poverty in all its forms everywhere;  Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all |
| Agriculture and food security  (SDG 2) | End hunger, achieve food security and improved nutrition and promote sustainable agriculture |
| Health and well-being  (SDG 3) | Ensure healthy lives and promote well-being for all ages |
| Quality of education  (SDG 4) | Ensure inclusive and equitable quality education and promote life-long learning opportunities for all |
| Governance and human rights  (SDG 5, 10 & 16) | Achieve gender equality and empower all women and girls;  Reduce inequality within and among countries;  Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels |
| Water and Sanitation  (SDG 6) | Ensure availability and sustainable management of water and sanitation for all |
| Energy, industry and infrastructure provision  (SDG 7 & 9) | Ensure access to affordable reliable, sustainable, and modern energy for all;  Build resilient infrastructure promote inclusive and sustainable industrialization and foster innovation |
| Urban development  (SDG 11) | Make cities and human settlements inclusive, safe, resilient and sustainable |
| Environmental sustainability  (SDG 12, 13, 14 & 15) | Ensure sustainable consumption and production patterns;  Take urgent action to combat climate change and its impacts;  Conserve and sustainable use the oceans, seas and marine resources for sustainable development;  Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss |
| Partnerships for Goals  (SDG 17) | Strengthen the means of implementation and revitalize the global partnership for sustainable development |

Source: Adopted from UNDP SDG Official list (2016)

## **Evaluation of potential methodological issues**

While big data can help resolve many data related challenges, there are considerable methodological, analytical, logistical and ethical challenges in using it in measuring development outcomes (Lokanathan et.al, 2017; Salganik, 2017; Olteanu et.al, 2019; Letouze). This section briefly discusses some of the prominent challenges for which we evaluated studies. As big data is varied in type, quality and composition, we also discuss if the challenges are specific to any particular type of big data.

We have grouped all the big data challenges that may affect measuring development outcomes credibly and lead to questionable internal and external validity of the studies.

*Non-representativeness of data and selection bias*: Big data may unintentionally exclude certain sections of the population or marginalised communities, thereby making the sample unrepresentative of the population being analysed. Large sample does not solve this systematic bias. This, however, is not a challenge when using satellite data that has universal coverage, but with human-generated and CRD data, non-representativeness is a serious challenge. Human-generated data such as Twitter, Facebook, or web searches, and cell phone use that generate CRD data are not representative as the usage is limited by income, education, infrastructure, etc. However, clarity on what is the sample frame i.e. who is included and who is excluded will help interpret big data results appropriately. Non-representative sample is still useful for within-sample comparisons, but may lead to erroneous out-of-sample generalizations (Salganik, 2017; Olteanu et.al, 2019).

*Construct validity*: Construct validity is whether the proposed measure actually measures what it claims to be measuring. This becomes important when the construct is unobservable and has to be operationalised via some observed attributes (Olteanu et al., 2019). For example, whether night light data truly reflects local GDP and other development outcomes such as health and education? Or what in the CRD data reflect people’s income, employment status? In many cases, the big data based measures may not be straight forward, and it is good practice to clearly state construct validity and provide necessary support to back the claim in the papers.

*Data quality issues:*

* *Comparability of data over time*: Since most of these data are collected routinely as a part of business, the nature and quality of data may change with the technology and business requirements. This may happen due to the underlying technology having changed or the people who use have changed, etc. For example, satellite data are not readily comparable across the years as there is a vast quality difference (Jain, 2020); and good flu trends based on online searches peaked comparing to officially reported data when the underlying google algorithm started prompting people to query more and broke the relationship between google searches and flu prevalence (Archie et.al. 2018).
* *Lack of completeness*: Most big data are a by-product of peoples’ everyday action and/or result of system logs of the government and businesses. It may not contain all the necessary information such as demographic characteristics. However, combining multiple sources of data, especially the big data and the administrative data can help resolve this problem (Salganik, 2017).

*Generalisability*: Generalisability or external validity refers to the applicability of the findings of a study to population or context other than it was produced. In the context of big data, generalisability would mean applicability of the model to a setting different from the setting of the data that the model was trained on. For example, a model trained on satellite data from a specific geographical region may not be generalised (Jean et.al, 2016; Head et.al. 2017). It is a good practice that the studies report on the representativeness of training data.

*Data transparency*: Transparency in this context refers to publishing all relevant materials, including data and code, used in a study in the public domain for independent verification. Sharing of raw data in the public domain is often crucial to establish confidence and reliability in the results. There are two challenges here: some of the major sources of big data such as CRD are proprietary and may not be allowed to be shared beyond the closed group of researchers; second the data has to be de-identified before it can be shared and it is crucial to check whether there are variables or a combination of variables that can be used to re-identify research subjects.

We assessed whether the studies included in the systematic map examined the following questions:

* Is the data representative of the population of interest?
* Is the construct validity explained? i.e. is there a discussion on how the big data based indicator measures what the study claims to measure?
* Are there data quality issues in the dataset used and how are they addressed?
* Are the results generalizable? For example, are the research findings generalizable to other situations such as other platforms (data source), communities, or over time?
* Are data and codes publically available for replication?

## **Evaluating reporting on privacy and ethical considerations**

There are concerns over data access, privacy, consent and ethics in using big data. Although, these are foundational issues for both small and big data studies, the challenges posed by big data have greater repercussions.

When using big data sources such as mobile data, most mobile operators have “inform and consent” policies that mandate disclosure of all relevant information to potential participants, who can then evaluate this information and give explicit permission. However, these policies often contain legal language which is generally not discernible, and it is not clear if explicit consent is obtained to repurpose the data. This kind of informed consent may be completely absent in research leveraging social media data due to the impracticality of obtaining consent from millions of users.

Mobile phone user data and social media data are some of the most used sources of big data that can inform researchers about individuals’ behaviour. Even if the data is de-identified there still remains concerns over the consent and ethics of sharing such data with researchers. Hence, it is imperative to have an ethics approval process in place that lays down the conditions under which such research can take place. There is a need for clear ethical standards for big data research and studies should be monitored by the Institutional Review Boards (IRBs).

Another ethical criterion when using big data can be concerned with the assessment of risks, most common being privacy breaches leading to identity theft or other cybersecurity risks. The possibility of re-identification of any individual user from poorly anonymised datasets adds to the concerns over anonymity of subjects. When combined with other sources, such datasets can be used to gain detailed insights about people without their knowledge. Such precise inferences may create the capacity for discrimination or mass manipulation. Sometimes data obtained for one purpose in social data research is used for secondary analyses but the associated risks may not be well understood. For example, Facebook data in the past has been used for ads targeting, as well as to tailor propaganda (Horowitz et al., 2018).

Big data may also inadvertently exclude certain sections of the population. For example, this bias can be observed in the case of “Street bump”, a mobile app that notifies the Boston City Hall whenever the user hits a bump on the road (Carrera et al., 2013). The data includes information only from the app users who often use their cars and the app, and this might inadvertently exclude poorer parts of the city that the app users may not frequent. Policy based on such big data sources may have unintended consequences on the people excluded.

We assessed the studies on the following:

* Ethical approval obtained
* Consent for secondary and other use of data discussed explicitly
* Discussion on data privacy
* Discussion on data security and governance arrangement
* Discussion of any potential unintended exclusion
* Discussion of potential unintended consequences for any group of people or individuals.

## **Exclusion criteria**

This report includes papers written in English and published between 2005 and 2019. This map does not include studies that develop algorithms or methodologies for using big data without explicit application to measuring or evaluating a development outcome. This map does not include studies that describe how big data and machine learning has been used in development programming helping in programme implementation, coordination, and management for designing and scaling new development solutions. Neither does it include the studies that show how big data methods are used in randomised controlled trials (RCTs) to identify the differential impact of sub-groups, and in improving survey data collection such as defining sampling frames.

# Results

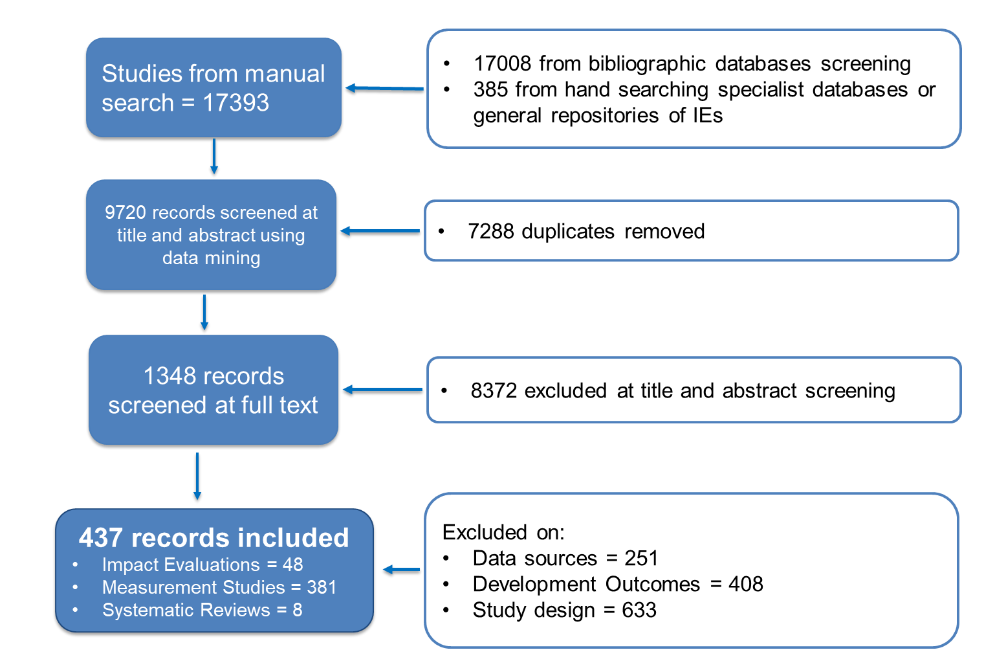
## **Search, screening and coding**

* + 1. **Searching and screening**

We systematically searched academic databases with the help of an information specialist (see appendix 3 for the list of academic data bases and search strings), and manually searched specialist organisational websites and grey literature sources. The initial searches resulted in 17,393 studies, of which 17,008 studies were identified through bibliographic databases search and 385 studies were identified through hand searching specialist databases, and IE and grey literature repositories (see figure 1). All the results were uploaded onto EPPI-Reviewer 4 for screening and coding. Screening of the studies was done in three stages.

We first employed EPPI-Reviewer’s built-in text mining, a machine learning technique, to sort the studies based on the inclusion and exclusion criteria at the title and screening stage (see Appendix 4 for more details on how text mining was used for screening). This reduced the number of studies to be screened at the title and abstract level to 9720. At the second stage, three researchers screened the studies for eligibility based on inclusion and exclusion criteria defined in the study protocol. At the end of stage 2, we identified 1348 studies to be screened at the full text level.

At stage three, four researchers coded the studies at the full text level. We followed a single screening with safety first approach[[7]](#footnote-8), were only one reviewer screened each record and coded doubtful studies as ‘to discuss’ to avoid excluding the studies that are potential includes. About 25 per cent of the studies at this stage were double coded for consistency, and any disagreement were resolved with the lead author. About 20 per cent of the studies were spot checked by the lead author. At the full text screening stage, we excluded 1292 studies. We excluded 251 studies based on irrelevant data sources, 408 studies on outcomes and 633 studies on study designs. Studies excluded on the design were primarily policy papers or computer science technical research papers that aimed at developing new algorithms rather than measuring development outcomes.

**Figure 1: The PRISMA flowchart**

Source: authors’ own calculation

* + 1. **Coding of included studies**

The three stage process of screening resulted in a final list of 437 studies which included 48 IEs, 381 measurement studies and 8 SRs. Through a consultation process, we identified the meta-data to be extracted using a standardised data extraction tool and defined them in the study protocol. During the final studies coding, we collected meta-data such as outcomes studies, outcome sub-categories, data sources used, geographical location of the intervention, country, evaluation design, target population, data transparency, ethics and other bibliographic information from the included studies (see appendix 4 for the data extraction tool). We also critically appraised the SRs (see appendix 5 for the SR appraisal tool and summary of the included SRs). Due to size and the nature of included studies we did not conduct critical appraisal of IEs and measurement studies.

* + 1. *Mapping of the studies*

The studies were mapped using 3ie’s evidence gap map platform that is organised as rows and columns. Various big data sources are placed on the rows and the development themes are on the columns. Any intersecting cell represent the development outcome measured or evaluated using the particular type of big data. Different colour bubbles represent the type of study where grey bubbles denote IEs, blue for measurement studies, green for high-quality SRs and red for low quality SRs. Hovering over the bubbles will show the links to studies. There are also filters for different regions, countries, study design, fragile context and the target population of the studies.

Given the large number of studies included in the report, there are development themes such as environmental sustainability, economic development and livelihoods, etc. include a large number of studies. We have provided maps within the main map to show how the studies are distributed across the sub-themes under these broad themes. In the sub-maps, the big data sources are mapped against the level 2 or level 1 sub-classification provided in the SDGs as relevant. For example, the sub-map for economic development and livelihoods has been coded against the level 2 indicators of eradicating poverty (SDG 1) and employment and economic growth (SDG 8), and the sub-map for environmental sustainability is coded against the level 1 classification of Goals 12, 13, 14 and 15. The following development themes have sub-maps:

* Economic development and livelihoods
* Health and well-being
* Governance and human rights
* Urban development
* Environmental sustainability

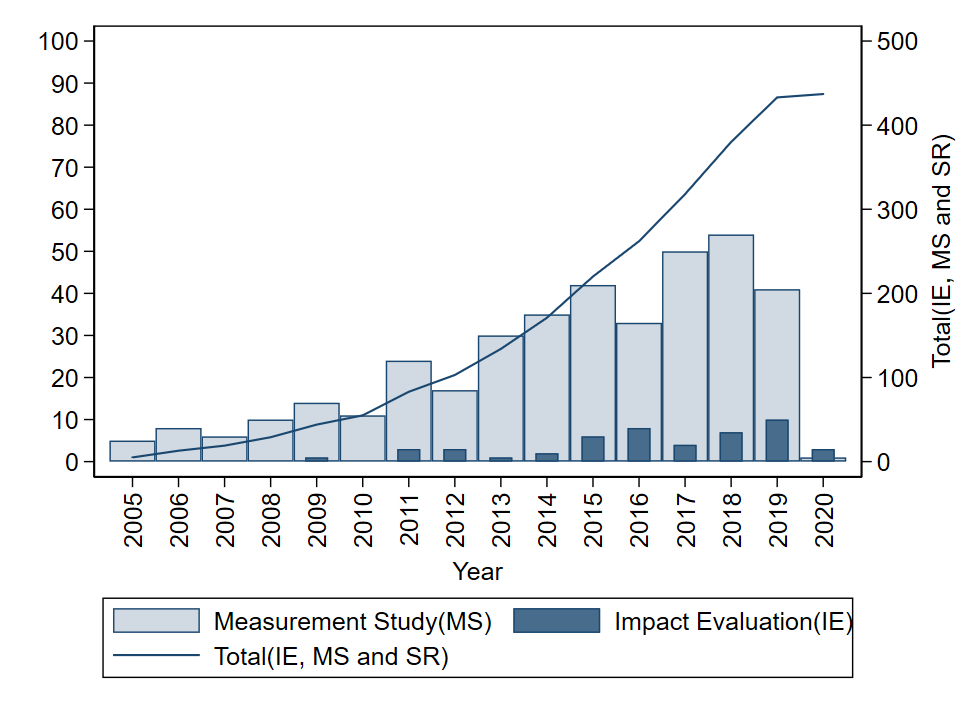
## **Characteristics and trends of the evidence base**

The map contains 437 studies, of which 48 are IEs, 381 are measurement studies and 8 are SRs. Of the 48 IEs, 8 are randomised controlled trials and the remaining are quasi-experimental studies.

Figure 2 displays the number of studies published each year from 2005 to 2020. The light blue bar shows the measurement studies, the dark blue bar shows IEs, and the line indicates the cumulative number of measurement studies, IEs and SRs. The number of measurement studies has grown gradually from 2005 to 2012, then increased substantially every year since then with the maximum numbers in 2017 and 2018. The past five years alone account for more than 60 per cent of the studies, indicating the increasing availability of big data, improved computational capacity, and the interest among researchers and journals.

The figure also shows that applying big data in IEs is a new phenomenon. The first IE using big data was published in 2009. While almost all the IEs were published after 2013, more than three quarters of the IEs were published in the last 5 years. We expect that the measurement studies will be proofs of concept, leading IEs to adopt to these approaches to innovatively measure development outcomes in evaluations. The map also points to the gap between the growth of measurement studies and use of big data in IEs.

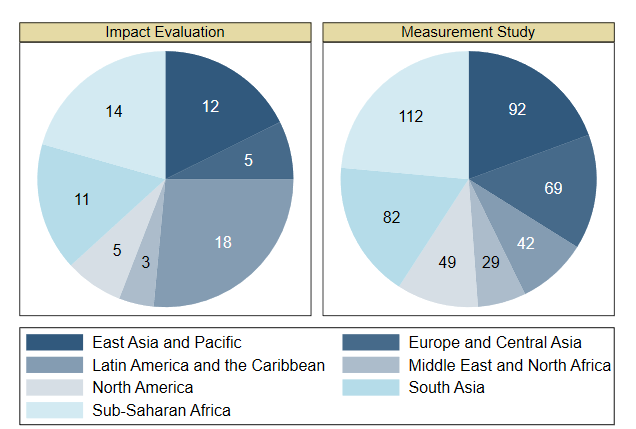
**Figure 2: Number of studies published per year**



Source: authors’ own calculation

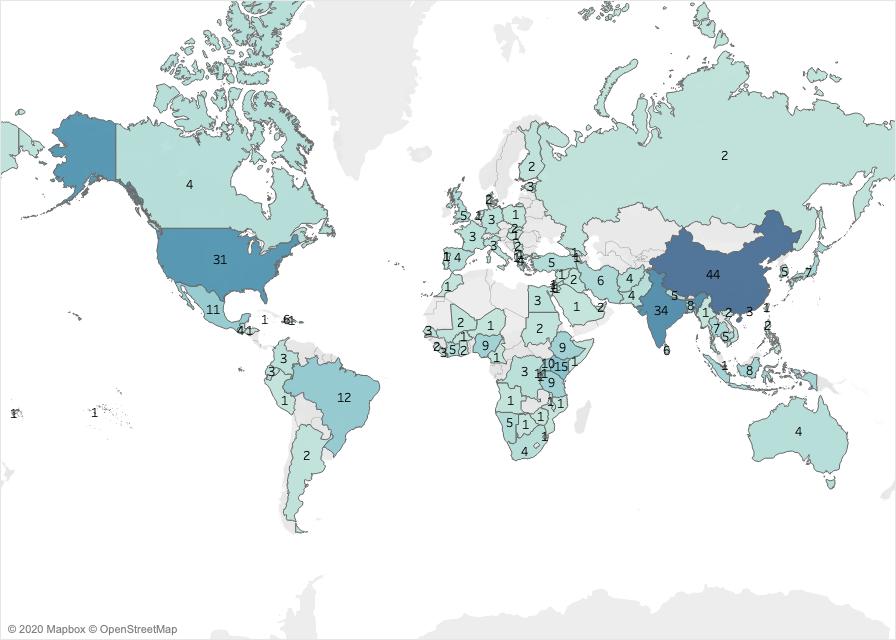
Figure 3 shows the geographical distribution of the included studies. About 50 per cent of the studies (n = 210) are from Asia and close to 30 per cent (n = 132) are from Sub-Saharan Africa. The distribution of IEs and measurement studies are roughly similar to the overall distribution. One notable exception here is Latin America and the Caribbean where the region accounts for 15 per cent of total studies (n = 65), but substantially more number of IEs (38 per cent, n = 18).

**Figure 3: Distribution of studies over regions**



Source: authors own calculation

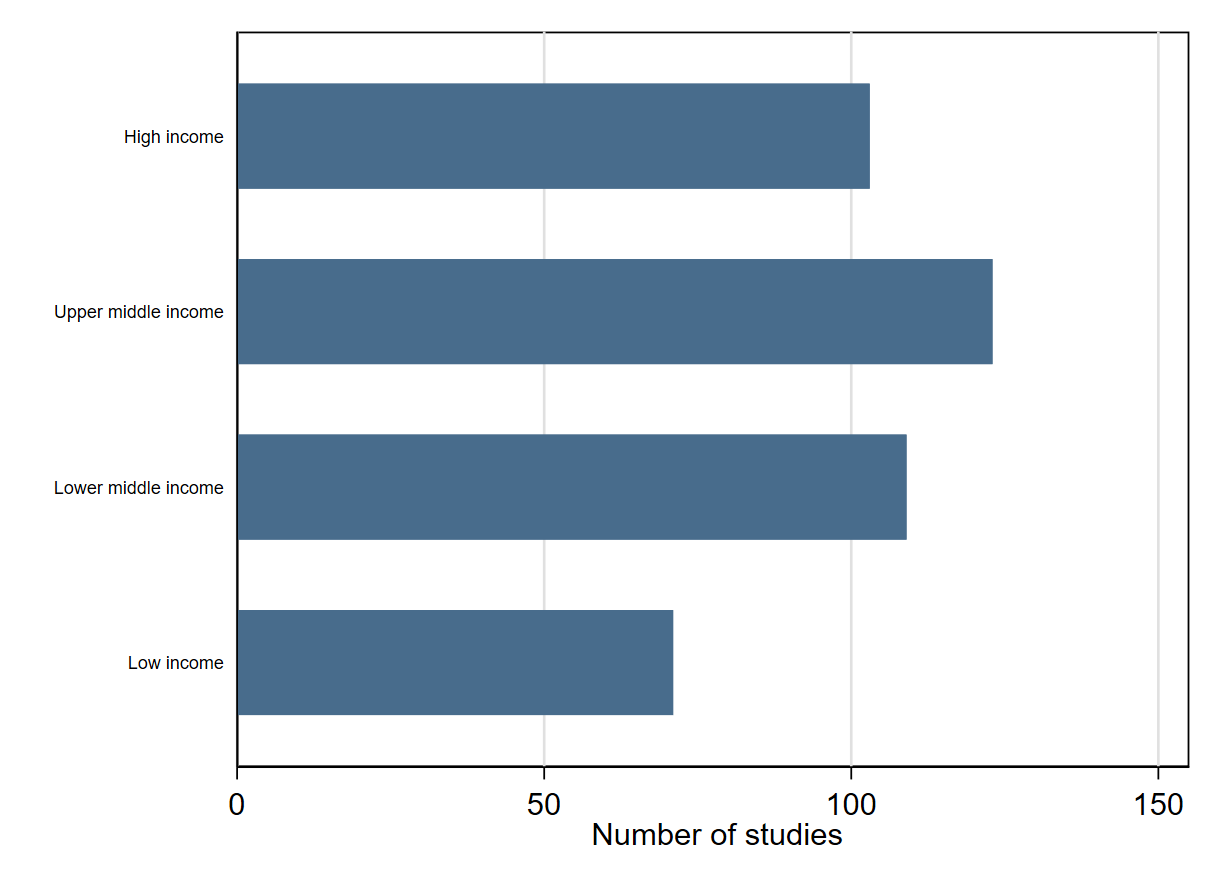
China (43) and India (34) are the most studied countries, and Kenya has highest number of studies (15) in Sub-Saharan Africa (figure 4). While East Africa is very well represented in the map, other African countries have fewer entries and several countries in West and North Africa have no studies at all. About 35 of the studies are multi-country studies and 11 studies did not specify the country name primarily to conceal the identity of the data provider. The distribution of IEs and measurement studies are again roughly similar to the overall distribution, but Latin America countries, particularly Mexico (5), account for a higher proportion of IEs. See table 4 for a list of top 20 countries with the maximum number of studies and table 5 for the geographical distribution of studies across the regions in Appendix 6.

**Figure 4: Geographical distribution of studies**

Source: authors own calculation

We find that most of the studies are concentrated in middle income countries. There are 232 studies (53 per cent) in the middle income group followed by 103 studies (24 per cent) in the high income group and 71 studies (16 per cent) in the low income group. Overall, about 69 per cent (n = 303) of the total studies are from L&MICs, but the IEs are distributed more in favour the L&MICs as 83 per cent of them (n = 40) are from L&MICs. One of the notable features of the studies in the map is that about 82 per cent (n = 359) of the total studies are published in peer-reviewed journals and the remaining of them are working papers (18 per cent, n = 78).

**Figure 5: Number of studies by income classification**

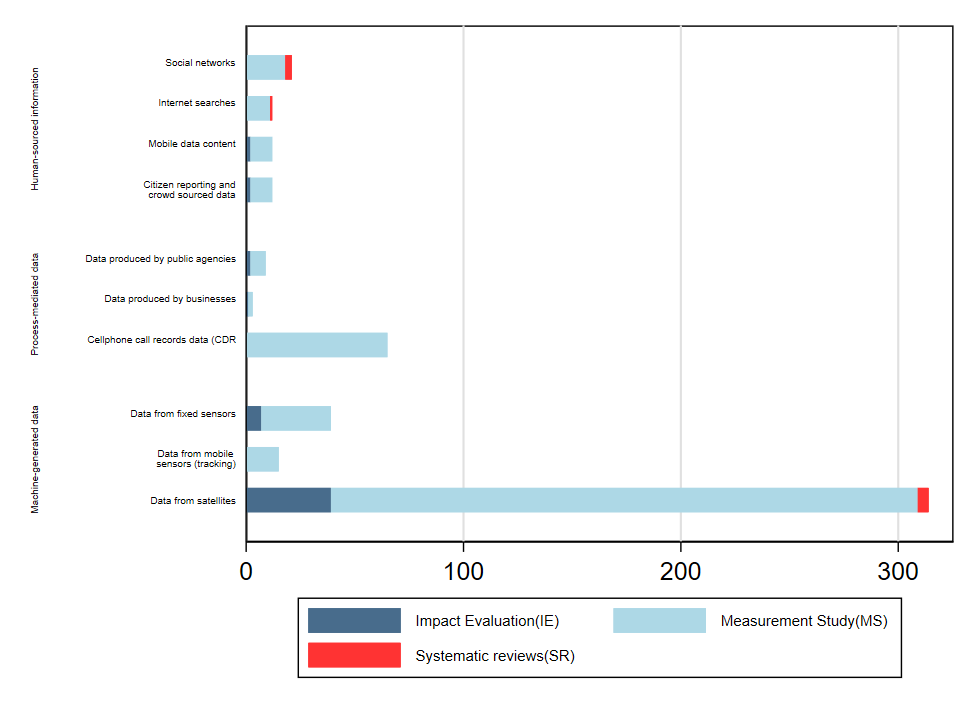


Source: authors own calculation

* + 1. **Distribution of studies across data sources**

As discussed in section 2.2, big data can be generated by human interaction on social media, process-mediated data recorded by governments and the business, and machine-generated data that is recorded by the automated systems. Figure 5 shows that machine-generated data is used the most. Of the total number of studies, close to 84 per cent (n = 380) of the studies used some form of machine-generated data, while 12 per cent (n = 53) of the studies used human generated sources, and 17 per cent of the studies (n = 77) used process-mediated data.

**Figure 6: Number of studies per different types of big data**



Source: authors own calculation

Table 6 below provides a detailed break-up of number of studies per data sources.

*Data from satellites and fixed sensors*: Satellite data is the most used source of big data as it accounts for 71 per cent of the measurement studies (n = 210), and close to 81 per cent of the IEs (n = 39). Data from fixed sensors such as weather and pollution sensors, traffic sensors, electricity meters that provide high-frequency, localised measured could also be readily used in IEs. This is the second most used data source, with 15 per cent of the IEs using these sources. This shows that the data from satellite and in-situ sensors that help measure spatial outcomes are used most in IEs. Other big data sources have been seldom used for IEs despite measurement studies showing proof of concept.

*Cell Phone Call Record Details*: A good number of measurement studies have used CRDs (17 per cent n = 65) for measuring population movement, migration, disease spread and even to understand the literacy level of the subscribers. Surprisingly, we found no impact evaluations which used this source of big data despite the availability of a good number of proof-of-concept papers in measuring key development outcomes.

**Table 6: Number of IEs and measurement studies across data sources**

|  |  |  |  |
| --- | --- | --- | --- |
| Data Source | Impact Evaluations | Measurement Studies | Systematic reviews |
| Human-Sourced |  |  |  |
| Social Networks | 0 | 18 | 3 |
| Internet Searches | 0 | 11 | 1 |
| Mobile Data Content | 2 | 10 | 0 |
| Citizen Reporting or Crowdsourced Data | 2 | 10 | 0 |
| Process-Mediated |  |  |  |
| Data Produced by Public Agencies | 1 | 2 | 0 |
| Data Produced By Businesses | 2 | 7 | 0 |
| Cell Phone Call Record Details | 0 | 65 | 0 |
| Machine-Generated Data |  |  |  |
| Data from Fixed Sensors | 7 | 32 | 0 |
| Data from Mobile Sensors (Tracking) | 0 | 15 | 0 |
| Data from Satellites | 39 | 270 | 5 |
| Total studies | 48 | 381 |  |

Note: per cent of sub-category total in the parentheses. Columns do not add up due to multiple entries.

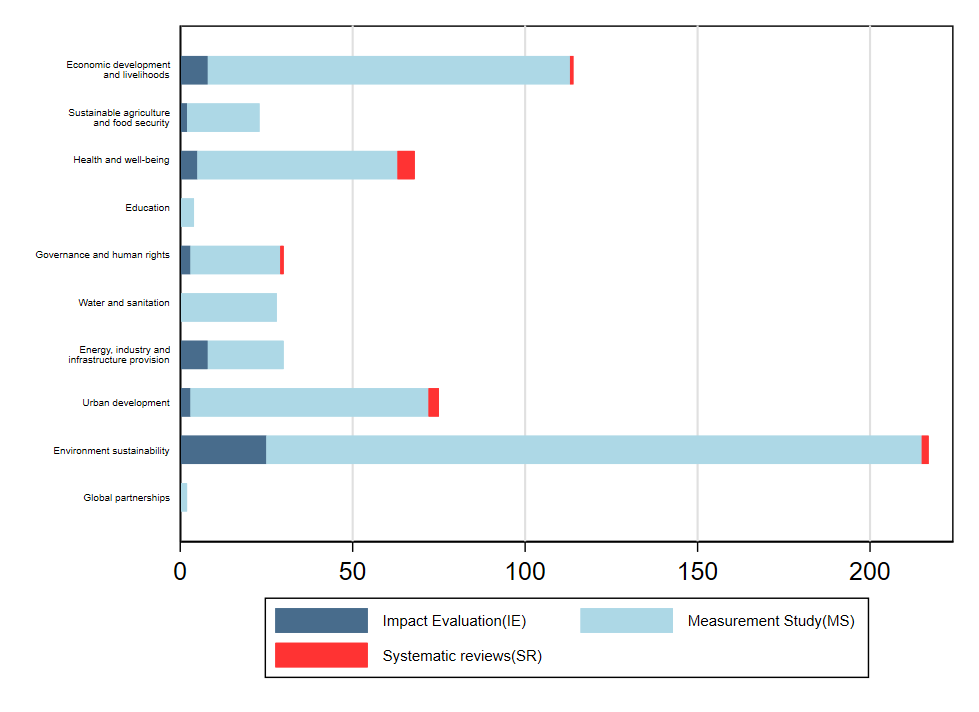
*Human-sourced data*: Social networks including Facebook, Twitter, and Wiki pages were used to measure development outcomes in 18 measurement studies. Internet searches like Google trends and other search engine queries were used in 11 studies. Mobile data content and crowdsourced data were used in 10 studies, primarily to measure disease outbreak, price data or people’s opinion on issues like development services. This source has not been used in IEs with two notable exceptions using crowdsourced data (Van der Windt, 2016; Edjekumhene, 2019).

*Complementarity between data sources*: There are about 57 studies in the map that have combined at least two sources of data and about 7 of them have combined three or more sources (table 7 in Appendix 6). Data from fixed sensors and satellites data seem to complement each other well — 13 measurement studies and 2 IEs have combined these two sources. Cell phone call records data (CDR) and satellite data is the other combination that has been used repeatedly. About 10 of the measurement studies have combined CRD data and satellite data in their analysis (see section 5 for a discussion and example on how satellite data and CRD can be combined together for better results). However, IEs seem not to have exploited this complementarity. See table 7 in Appendix 7 for a list of studies using multiple sources of big data.

* + 1. **Distribution of studies across development themes**

Section 2.3 identifies ten broad development themes based on SDGs. Figure 6 and Table 8 show the number of studies across the development themes. About 50 per cent of studies (n = 217) focus on environmental sustainability, which includes sustainable consumption and production, climate change, underwater life, and life on land. Economic development and livelihoods accounts for about 26 per cent of the total studies (n = 114). Urban development and health account for 16 per cent each (n = 68). Governance and human rights (7 per cent, n = 30), and energy, industry and infrastructure (7 per cent, n = 30) account for the remaining studies.

**Figure 6: Number of studies against development outcomes**



Source: authors own calculation

While the distribution of IEs and measurement studies across the development themes remains the same as the overall distribution, there is substantial number of IEs on economic development and livelihoods (17 per cent), and energy, industry and infrastructure (17 per cent). While most of the SRs looked at cross sectoral themes, health is the most studied sector (n = 5), followed by urban development (n = 3), and environment sustainability and economic development (n = 2). See appendix 6 for critical appraisal of the SRs.

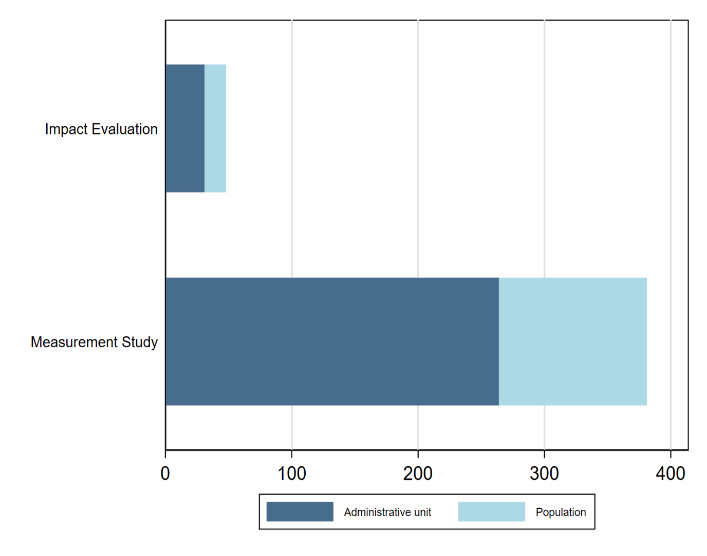
**Table 8: Distribution of studies across development themes**

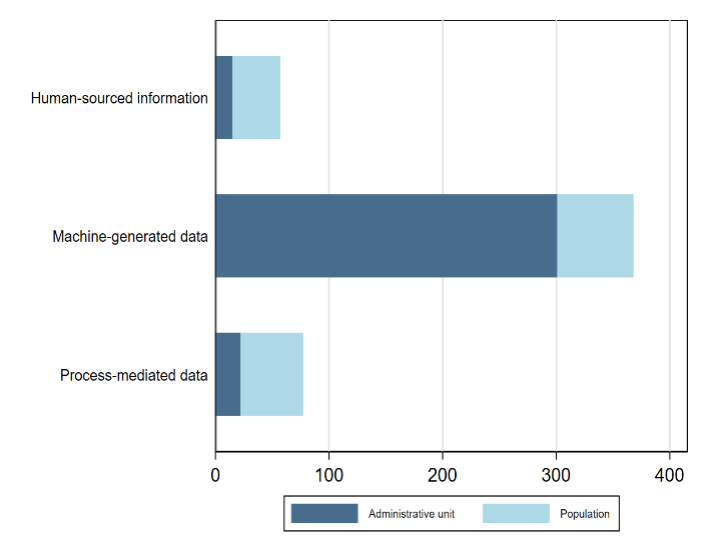
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Development themes | Impact Evaluations | Measurement Studies | Systematic Reviews | Total |
| Economic development and livelihood | 8 | 105 | 1 | 114 |
| Sustainable agriculture and food security | 2 | 21 | 0 | 23 |
| Health and well-being | 5 | 58 | 5 | 68 |
| Education | 0 | 4 | 0 | 0 |
| Water and sanitation | 0 | 28 | 0 | 28 |
| Governance and human rights | 3 | 26 | 1 | 30 |
| Energy, industry and infrastructure | 8 | 22 | 0 | 30 |
| Urban development | 3 | 69 | 3 | 75 |
| Environment sustainability | 25 | 190 | 2 | 217 |
| Global partnership | 0 | 2 | 0 | 2 |
|  |  |  |  |  |

* + 1. **Units of observation**

The unit of observation (or unit of analysis) is the class of elemental units that constitute the population and the units of measurement. Typically, in impact evaluations, the units of observation are individuals, households, facilities (in facility surveys), or various level of administrative units such as villages, counties or districts. We have classified the unit of observation as population (including both individuals and households), or administrative units (villages, land parcels, or any other units with a spatial element). The unit of observation seems to be an important element in analysing the use of big data in measuring development outcomes. Figure 7 (panel 1) shows that about 70 per cent of the measurement studies (n = 267) and 65 per cent of the IEs (n = 31) have administrative units as their unit of observation. There is a clear distinction between different sources of big data, as shown in Figure 7, panel 2. The unit of analysis for satellite data based studies is predominantly administrative units (n = 259, 83 per cent), while CRD-based studies are usually based on population units (n = 53, 82 per cent). This difference shows that satellite data is more applicable when the outcome of interest has some spatial dimension such as local economic development, agricultural land productivity, forest cover, or urban development.

**Figure 7: Units of observation**





Source: authors own calculation

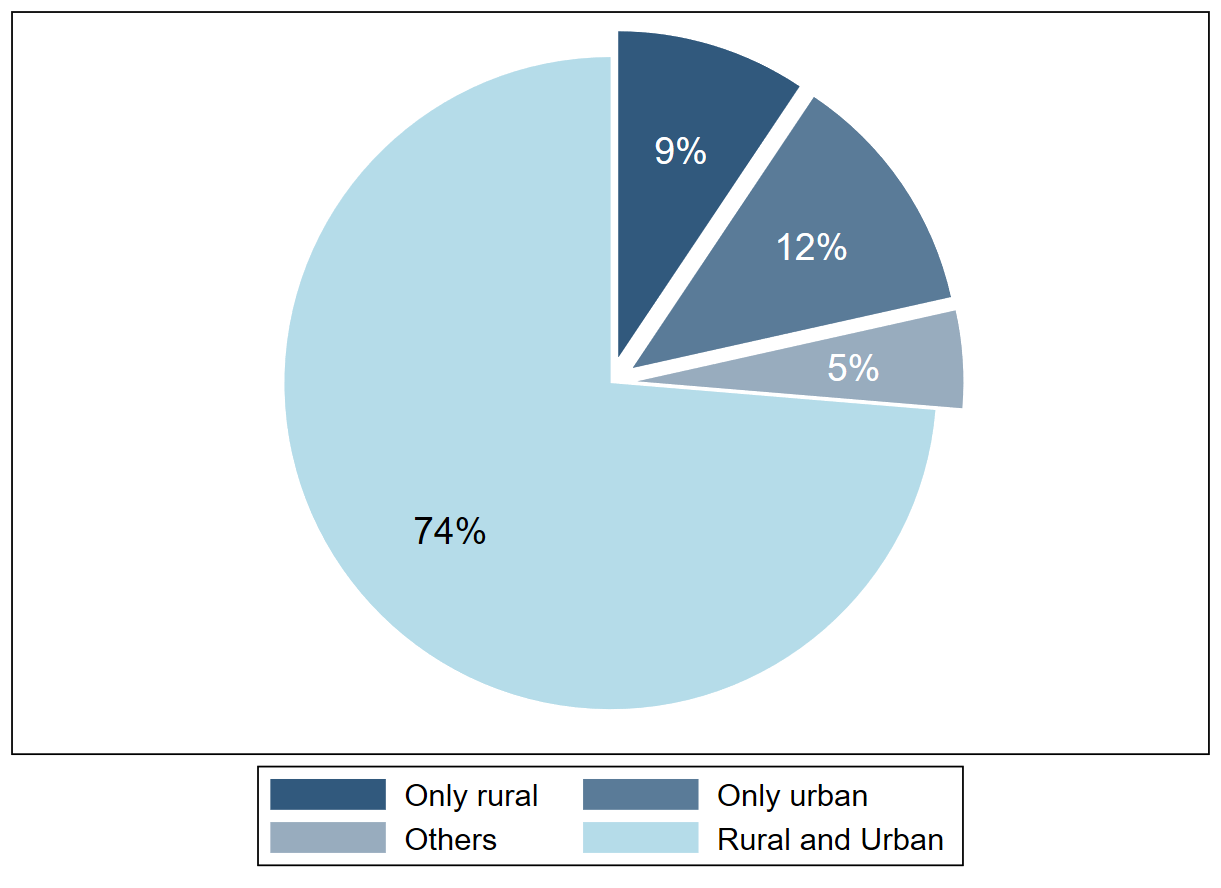
* + 1. **Studies using mixed-methods**

Mixed-methods IEs that combine qualitative and quantitative analyses help assess the quality implementation and reliability of data, and understand the mechanism of programme impact (Bamberger, 2012). Big data IEs can be combined with qualitative methods. However, there are only 3 IEs and 5 measurement studies report to have used mixed methods.

* + 1. **Studies with a rural or urban focus**

Figure 8 shows the proportion of studies focused on rural areas, urban areas, or both. Most studies looked at both rural and urban areas (74 per cent, n = 325). About 9 per cent of the studies (n = 41) were focused on rural areas, 12 per cent (n = 51) on urban areas. Among the remaining studies, 14 studies looked at conflict affected population, 4 studies at ethnic minorities and 2 at refugees.

**Figure 8: Distribution of population subgroups**

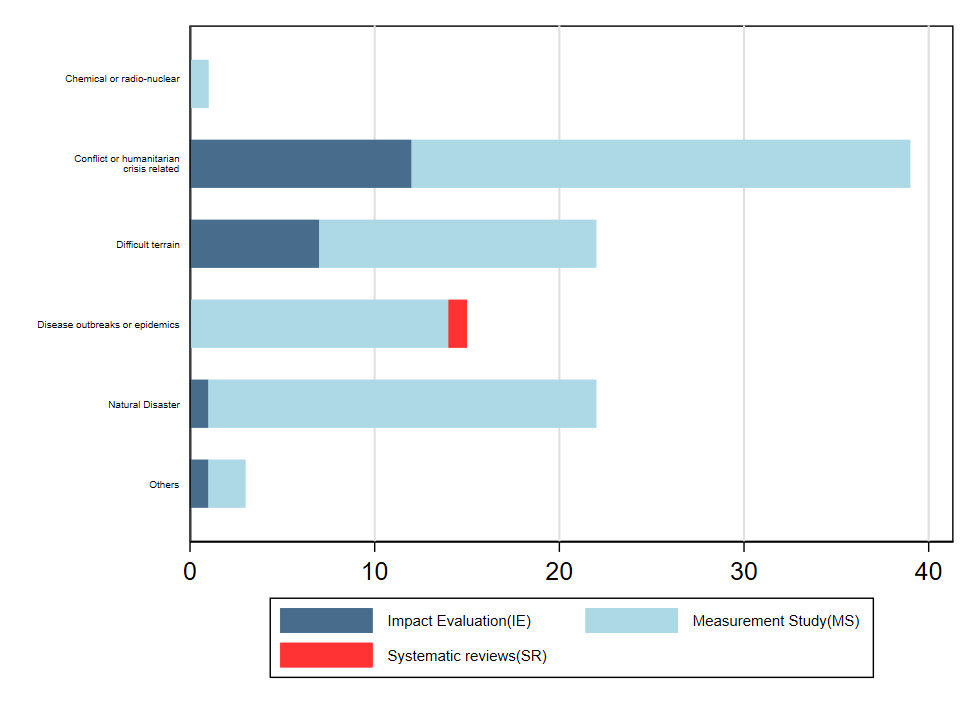


Source: authors own calculation

* + 1. **Studies by fragile context**

We use the OECD definition of fragile context that includes conflict, institutional, social fragility, environmental, health and climatic risks (OECD (2018), States of Fragility 2018). Figure 9 shows that 91 studies included in the map (21 per cent) are from countries that are considered to be fragile context. About 39 studies were conducted in conflict or humanitarian crisis context, 22 studies each in the contexts of difficult terrain and natural disasters, 15 studies in the context of 15 epidemics or disease outbreaks. There is one measurement study in the context of chemical, radio-nuclear disaster. IEs follow the same pattern, except for one notable gap as there is no IEs in the context of epidemics or disease outbreaks despite reasonably good number of measurement studies.

**Figure 9: Number of studies in fragile contexts**



Source: authors own calculation

Table 9 in appendix 6 shows that the satellite data is the most used in fragile contexts, followed by CRD data and then the senor data. The table also shows that all most all the big data sources have been used in one or two fragile context indicating the importance of big data in fragile contexts.

## **Appraising potential methodological biases, risks and limitations**

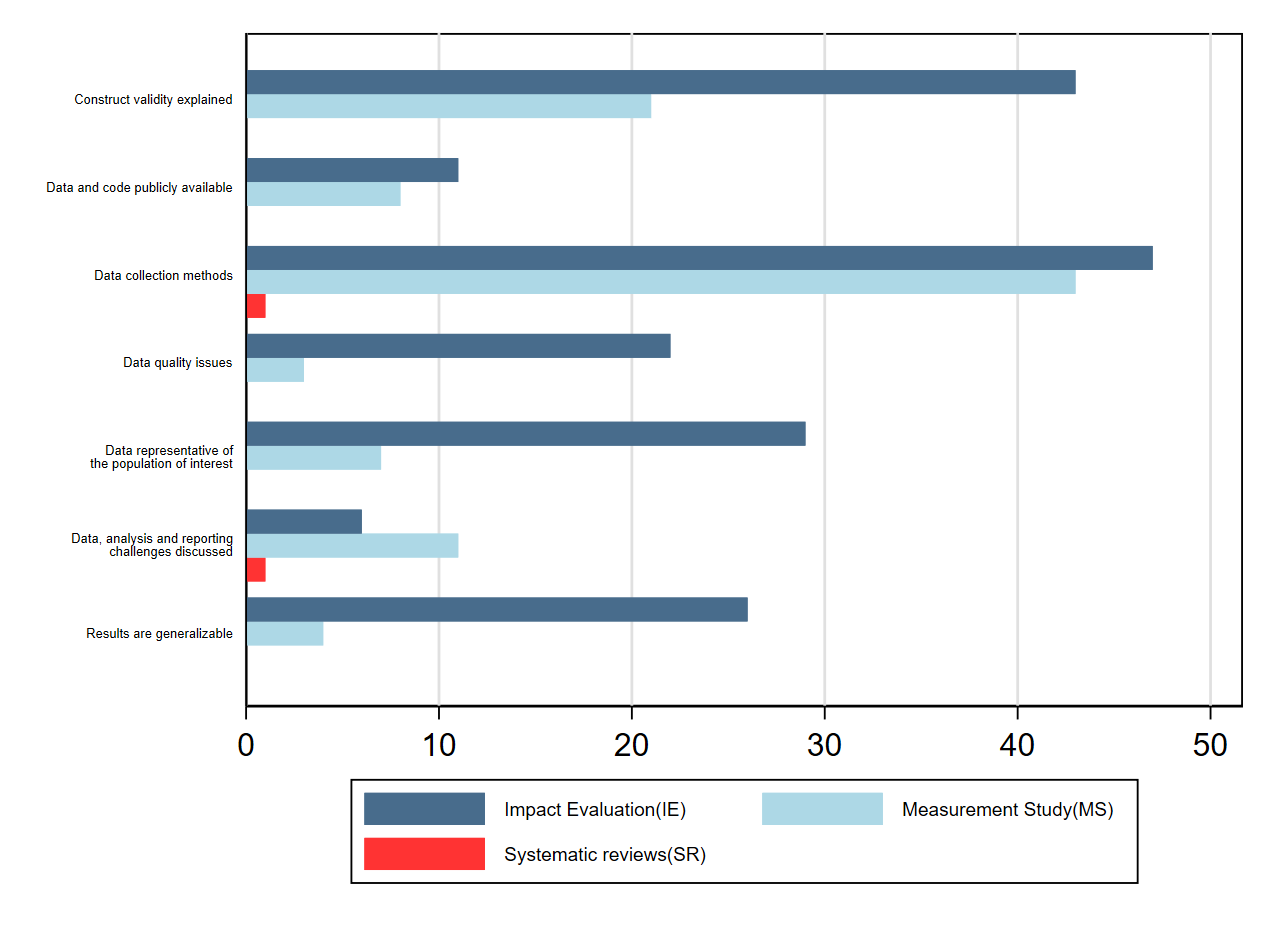
* + 1. **Reporting on methodological challenges and data transparency**

Figure 10 shows that very few studies meet any of the following methodological quality markers:

* Is the construct validity explained? i.e. is there a discussion on how the big data based indicator measures what the study claims to measure?
* Are data and codes publically available for replication?
* Are data collection methods discussed?
* Are there data quality issues in the dataset used and how are they addressed?
* Is the data representative of the population of interest?
* Are challenges in the analysis and reporting process discussed?
* Are the results generalizable? For example, are the research findings generalizable to other situations such as other platforms (data source), communities, or over time?

Only 95 studies (22 per cent) have reported on at least one of the above transparency criteria. For example, only 20 per cent (n = 91) of the total studies reported on data collection methods, 6 per cent (n = 25) on data quality issues, 8 per cent (n = 36) on data representativeness, 14 per cent (n = 64) on construct validity and 7 per cent (n = 30) on generalizability. Only 4 per cent (n = 19) of the studies have data and codes publically available or available upon request.

**Figure 10: Number of IEs, MS against data quality and transparency**



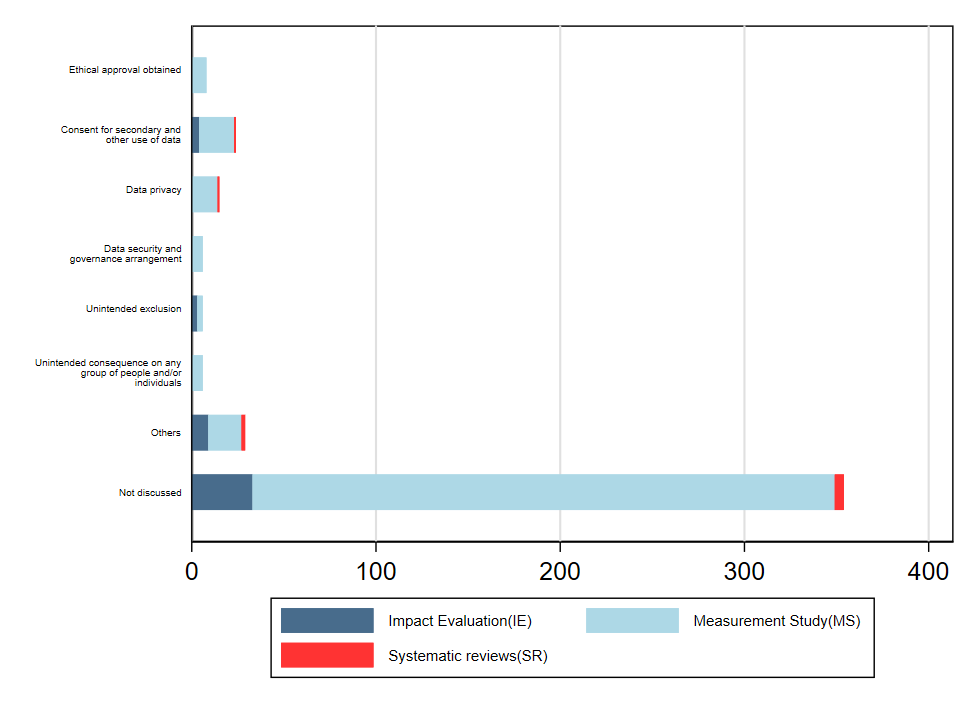
Source: authors own calculation

There, however, is considerable difference between IEs and measurement studies in terms of reporting on data quality issues and transparency. Table 9 in appendix 6 shows that IEs report a lot better on all these parameters. Of the total 48 IEs in the map, 46 of them report at least one aspect of transparency. Almost all the IEs report on data collection methods, 90 per cent (n = 43) report construct validity, 60 per cent (n = 29) discuss representativeness of data, 54 per cent (n = 26) discuss generalizability and 45 per cent (n = 22) discuss various data quality issues. However, only 23 per cent (n = 11) of them have data and codes available and 13 per cent (n = 6) discuss key data analysis and reporting challenges.

* + 1. **Reporting on privacy and ethical challenges**

Figure 11 shows that most studies – 81 per cent – do not report on ethical challenges and privacy issues. Of the few that do discuss such challenges, the most frequently discussed issue is that of consent for data use.

**Figure 11: Number of studies reporting on ethics issues**



# Key findings, discussion and lessons

The use of big data in measuring development outcomes is on the rise in the past 5 years. This rising trend is powered by the availability of and our capacity to process big data. In this section, we discuss the key findings, some of the notable gaps and the potential for future SRs.

## **Evidence base and key gaps**

**There is a considerable potential for measuring various development indicators using big data.** We identify a significant and growing evidence base of measurement studies that use some form of big data to measure a development outcome. Some outcomes are more amenable to the use of big data than others, and environmental sustainability, economic development and livelihoods, health and well-being, and urban development are where the majority of studies are concentrated. Education, sanitation, governance and human rights seem to be less responsive to big data use.

Multiple entries for most development theme indicates the potential of big data in contributing to measuring development indicators. Identifying measurement studies will be a valuable addition to development evaluators who look for innovative new ways to measure a development outcome which was difficult to measure at all required spatial and temporal scale using conventional data collection methods.

**There is potential for more IEs using big data on development interventions**. The map contains 48 impact evaluations. Use of big data measures in IEs as main outcomes or for controlling key covariates is fast growing, but the IEs are fewer comparing to the measurement studies in terms of numbers and the extent of their thematic and geographical coverage. IEs seem to be concentrated more around environmental sustainability, economic development and urban development. This complements the existing efforts to build evidence base in international development as these sectors have much less rigorous evaluations (Sabet and Brown, 2018). The IEs are also concentrated on using satellite data.

**Satellite data is used the most.** The map shows that 71 per cent of the measurement studies and 81 per cent of the IEs use satellite data. This is also one of the sources that has been used since the early 2000’s. The prominence of satellite data is primarily due to the fact that satellite images offer unique possibilities for measuring and evaluating development outcomes. Given the vast number of Satellites covering almost every location on Earth, it is possible to collect data at a high granularity (spatial resolution) and for multiple temporal frequencies for the past 30 years. Satellite data is freely available from several sources such as NASA's Landsat and MODIS, and the European Space Agency's (ESA) Sentinel, etc. and more importantly several pre-processed data bases such as AidData’s Geoquery[[8]](#footnote-9), Yale University’s G-Econ Project, the United Nations Environment Programme's Environmental Data Explorer[[9]](#footnote-10), NASA’s Socioeconomic Data and Applications Center[[10]](#footnote-11) (SEDAC), Global Forest Change 2000–2018 (Hansen et al., 2013), and several others. This pre-processed data or the image data then could be processed and converted into meaningful outcomes to measure economic activity at local level, urban development, forest cover, land productivity, distribution of the population, etc. These indicators can also be used for controlling for covariates.

**Spatial dimension matters**. One of the key findings of the map is that most of the big data studies are applied in the context where the phenomenon studied has spatial dimension, meaning the outcome and other covariates are measured in a spatial scale. Close to 70 per cent of the studies in the map report having administrative units as their unit of measurement. This is particularly true for satellite and sensor data based studies as 82 per cent have administrative units as their unit of measurement such as local economic development, agricultural land productivity, forest cover, or urban development. This is referred to as geospatial impact evaluation (BenYishay, 2017). However, there is considerable difference across data sources as CRD data is used to measure changes at the population level.

**CRD data has great potential for measuring and evaluating development outcomes but not used in IEs yet**. CRD data is one of the most widely used sources in measurement studies. This is used for measuring population movement, migration, disease spread, etc. Despite a number of high-profile measurement studies, our systematic search did not find even one IE that has used CRD data for rigorously evaluating a development outcome. This is a notable gap and a potential area for future exploration. It should be noted that the CRD data is also fraught with multiple methodological challenges such as non-representativeness, lack of completeness, etc., and ethical challenges such as consent, unintended exclusion, etc. Further, CRD data has been difficult to obtain as it is proprietary and hence difficult to maintain data transparency.

**Other big data sources such as human-sourced and process-mediated data have good proof of concepts**. Human-sourced data such as social networks, internet searches, mobile data content citizen reporting or crowdsourced data, and process-mediated data such as data produced by public agencies and by businesses, have good number of measurements studies as proof of concept for using these sources to measure various development outcomes, but there are not many IEs using these sources. This also shows the possibility of potential use of these sources in future IEs. Similar caveats on methodological and ethical challenges discussed above in relation to CDR data will apply.

**East Africa is well represented, but not the rest of Africa.** The geographical distribution of measurement studies and IEs show that the studies are evenly spread across the continents. However, Ethiopia, Kenya, Rwanda, Tanzania and Uganda are well represented in terms of number of measurement studies and IEs, but the only non-East African country that seem to have well represented in the map is Nigeria. There are very few studies in the rest of Africa. This gap is particularly serious given Africa’s data challenges (Serajuddin et.al. 2015).

**Big data holds great potential for conducting impact evaluations in fragile contexts, including during conflicts, humanitarian crises, epidemics and natural disasters.** Conducting rigorous evaluation in fragile contexts such as natural disasters, disease outbreaks and other crises contexts, can be costly, risky to the beneficiaries and the evaluators, and in some cases outright infeasible. We identified 73 measurement studies, 17 IEs and one SR in such fragile contexts. Measurement studies are spread evenly across conflict or humanitarian crisis, disease outbreaks or epidemics, natural disaster and difficult terrain to reach. However, the IEs are concentrated around conflict and difficult terrain. The number of measurement studies indicate the potential for more IEs in the fragile contexts.

**There are potential sectors and themes where systematic reviews will be useful**. Though the number of IEs are fewer, the map highlights a few potentials thematic areas where systematic reviews will help answer key questions on policy and research methods. For example, there is a concentration of IEs using satellite data, referred to as geospatial IEs. But we know little about how satellite data can help evaluate development programmes better, where it can add value, what type of interventions could be better evaluated and the technical challenges in using satellite data. A systematic review of all geospatial IEs that have used satellite data across the sectors will help understand the potential and challenges in using satellite data for IEs.

Similarly, there is a concentration of IEs on environmental sustainability and within that climate action and forest management. Though there a few SRs and evidence gap maps on forest management (Pelletier et.al., 2016; Puri, et.al, 2016), a new review with reference to innovative, new data sources used in rigorously evaluating forest cover, and the advantages and challenges thereof may be useful.

## **Analytical, ethical and practical challenges**

**Satellite data also presents misclassification problems**. Researchers, however, point to several technical challenges in using satellite images that may lead misleading conclusions. For example, Jain (2020) argues for the need for ground validation of satellite data as sometime the images could be misclassified i.e. flood irrigation may be classified as flooded area; and often this misclassification is systematic i.e. forest cover is almost always misclassified as agriculture which will bias the study results. Further, there could be difference in the data coming from different satellites and from the same satellite constellation but using different sensors i.e. Normalised Difference Vegetation Index (NDVI) varies for different satellite sources and for same satellites across different versions such as Landsat 5 and Landsat 8.

**Data quality and transparency is paramount**. The map also points to the need for setting standards for better reporting as about 87 per cent of the measurement studies did not report on data quality issues, representativeness, construct validity and generalisability. This would lead to questioning the internal and external validity of the findings. Further, there is also a need for setting standards for data transparency taking into consideration of the challenges in sharing proprietary data, data storing and capacity of the Dataverse (see box 2 for more details on data transparency).

# Box 1: Transparency in data analysis, use and sharing

In the recent times, the use of big data for in-depth comprehension of developments in the social sector has gained traction. Transparency refers to publishing all relevant materials, including data and code, used in a research study in the public domain for independent verification. Transparency in research encompasses a number of elements that are no different while using big data as opposed to traditional data sources. Certain challenges that might arise are discussed below:

1. *De-identification*: De-identification is related to preserving the identity of the study subject before it is made available for any sort of analysis. A few concepts used in de-identification of big data are K-anonymity, L-diversity, T-closeness. A dataset is said to have k-anonymity property if each person in the dataset cannot be identified by information of the other k-1 individuals in the dataset. In contrast, L-diversity is a group based anonymization techniques that reduces the granularity in the dataset. T-closeness is a refinement over the L-diversity and is used to decrease granularity over and above L-diversity. However, there are several examples of combining different source to re-identify the respondents in the data set (Archie et.al, 2018).
2. Scale and storage: In today’s world, storage of high volume of data is not a challenge owing to developments in cloud computing. A fairly new system that provides solution to the scalability and storage issue is Storage virtualization. In simple terms, it is a network of storage devices that are combined to create a single storage space. A few ways of safeguarding the data storage is encrypting all processes and usage of hybrid clouds. Data repositories such as Harvard dataverse, figshare have limited capacity to handle big data and are often restricted by the size of the data uploaded. Cloud storages such AWS and other similar such storage will be a better option for the same.

**Ethical concerns are substantial**. Ethical challenges such as consent, data privacy, data security and unintended exclusion, etc. are well documented in the literature (Lokanathan et.al, 2017; York and Bamberger, 2020), a brief analysis of the studies in the map shows that very few studies report on any of these ethical challenges. However, the challenges are different for different sources of big data. For example, satellite data that involve little human interaction may not need an institutional review board (IRB) review, but most other big data source that use human-generated data without their explicit consent for secondary use should be reviewed by IRBs. We also recommend big data evaluations to be more mixed methods to mitigate the potential disconnect between the development stakeholders and the big data researchers. Any mixed method research need to be reviewed by IRBs.

The map shows that most IEs have done well on reporting data quality issues, but not on the ethical issues. since big data requires ethical issues such as consent for secondary data use, unintended exclusion which are new to conventional ethical standards, there is a need for updating the current ethical standards practice to include big data use as well.

**Big data may be growing in use and popularity, but the need for independent auxiliary data for ‘ground truthing’ remain**. Many sources of big data are partial in terms of coverage, and prone to biases which are difficult to measure, control, and correct for in the absence of secondary data. Despite growing awareness and acknowledgement of its limitations, the household sample survey remains the dominant source of development policymaking. Big data often require survey data as ‘ground truth’ data to validate the findings. In practice traditional and big data exhibit important complementarities.

**Some capacity constraints are acute**. Development organizations need to build staff capacity in order to use big data as a strategic asset (Perera-Gomez & Lokanathan, 2017). They need to build multidisciplinary teams consisting of data experts and subject matter professionals, and also compete with the private sector to recruit the staff. Other major costs involve scaling up the technical infrastructure to enable data storage and processing at a large scale and data accessibility costs. The latter can be more difficult to predict considering that big data sources that are currently public may involve licensing in the future. Besides, ensuring the sustainability of data can be a cause of concern. As suggested by Hammer et al. (2017), as most of the big data is produced as a by-product by the private sector, continuity of data provision cannot be guaranteed in this age of evolving technology and market conditions. These concerns will question the sanity of committing resources upfront to build capacity.

**Need for better coordination between data scientists and evaluators.** Big data analysts and evaluators follow different framework and analytical tools, particularly the big data measurement studies look for identifying hidden patterns in the data with little support from theory and aim at prediction rather than causality (York and Bamberger, 2020). Further, the expertise needed to analyse big data remains largely localised and siloed. Outside of a small and highly specialised group of data scientists, there is uncertainty about how best to do large-scale big data analysis. The degree of technical specialisation, combined with strict access restrictions to many types of big data, has hindered big data applications in development evaluation. Hence, there is a need for promoting interaction between development evaluators and data scientists for better cross-learning and adoption of big data in measuring and evaluating development outcomes.

**The cost of collecting, analysis, storing and reporting big data is largely unknown**. There is very little publically available information on the cost of collecting, analysing and reporting big data. Blumenstock, Cadamuro and On (2015) reported that the phone survey for ground truthing the CRD data costed $12,000 and took 4 weeks to administer. This is however only the variable cost of data collection in this study. There is multiple hidden cost such as staff cost, cost of necessary computing infrastructure including storage, the opportunity cost of time involved in developing partnership with data providers in some cases is not known. BenYishay et.al (2017) report that the cost of geospatial IE is around USD 150k. One of the 3ie funded studies using satellite data reported to have spent USD 3,300 on data collection which is about 1 per cent of the total study budget, but have spent USD 103,864 on data analysis and reporting which is about 32 of the total cost[[11]](#footnote-12). Similarly, another 3ie funded study that has used in situ fixed sensors has reported to spend USD 6,152 on data collection or acquisition (11 per cent) and USD 54,444 staff cost for analysis and reporting (55 per cent). However, studies that have combined satellite data with their household survey have reported higher cost of data collection (USD 171,582; 43 per cent) and analysis and reporting (USD 109,495; 27 per cent). This is however, roughly comparable to data collection cost of an average 3ie-fundned multi-year, multi-round survey IE which costs about USD 176k (Puri and Rathinam, 2019).

# Using big data in IEs: potentials and challenges

Rigorous IEs require a valid counterfactual. Randomising programme placement ensures pre-programme comparability of the treatment and control groups in most cases, and quasi-experimental studies employ statistical procedures to identify a valid comparison group. In either cases, evaluators collect require a vast array of data on the outcomes, covariates and other contextual factors. There is almost always a trade-off between collecting complete array of necessary data and cost effectiveness, and in a few cases, it may not be feasible to collect some of the covariates and confounders.

Big data, with the help of improved machine learning techniques and analytical capacity, can now be manipulated to evaluate development outcomes. The potential advantages for big data are, to date, most discernible in contexts where the immediacy, scale, and/or reach of data is highly prised and alternative sources of data are absent or inadequate to the task. The ability to “zoom in” on particular zones of interest, and to produce estimates for small areas, is an oft-cited advantage of many types of big data (e.g. satellite and building footprint data, mobile phone CDR and signalling data and app-based location data), and one with particular relevance to evaluative contexts, and SDGs related urbanisation, climate change and infrastructure. This holds particular promise for settings where census data render small area estimation (SAE) methods unsuitable. Big data have also been shown to be particularly advantageous for the analysis of disaster-induced displacements and disease outbreaks. In each of these cases, the advantage of big data is that it can support rapid appraisal and introduction or adjustment of policies/interventions on the basis of near real time information.

In this section, we highlight a few examples from the map to show the steps involved in collecting, processing and using satellite and CRD data for measuring development outcomes. We draw on recent projects from 3ie and Flowminder to illustrate the processes.

## **Using satellite data in IEs**

Pande and Sudarshan (2019), a 3ie funded evaluation conducted by the Institute for Financial Management and Research evaluated the recent environmental clearance (EC) reforms in India. Before the 2006 reform, mines of area over 25 hectares (ha) were required to hold a public hearing before approval. The new EC reform required mines of area between 5 and 25 ha to hold a public hearing as well. This study exploits this historical discontinuity in clearance requirements to evaluate the impact of public hearings on mines’ environmental compliance. Apart from rigorously evaluating the EC process in India, this study also provides a proof of concept for the use of remote-sensing data and other publically available data to monitor mines’ environmental compliance. Using satellite data to assess the impact of EC process require data on the timing of the intervention, the geographical scope of the intervention (i.e. the individual mines in this case) and on the outcome of interest such as air pollution, land cover and water quality for the corresponding intervention and control areas for the years before and after the intervention.

The following were the key steps involved in the big data IE:

**Step 1:** the researchers used web scraping techniques to collect information on the mines from their EC application for the years from 2006 to 2016, available online in a database published by Ministry of Environment, Forests and Climate Change (MoEFCC). These are all mostly scanned pdf documents. The researchers scrapped for the information on project name, location (tehsil and village where the mine is located) and the dates of key EC stages of submission, review and approval, and mine characteristics such as minerals mined, mine production capacity, size of the mine; they also scrapped the clearing letters available in the same database for cross checking the data. They collected information about all 934 relevant mines and used 134 of them in their regression discontinuity analysis. Finally, the researchers hired ML Infomap, a local company, for geocoding all the mines identified.

**Step 2**: Satellite data on various key environmental outcomes such as air pollution, land cover and water quality were collected from different sources[[12]](#footnote-13):

* The researchers used the data provided by Dalhousie University on the fine particulate matter concentration as a proxy for air pollution. This database contains average annual PM concentration for every one kilometre cell for the study period.
* They have used the Enhanced Vegetation Index (EVI) data from NASA’s MODIS satellite to measure deforestation around the mining areas. EVI is available at the resolution of 250 meters for the entire globe, and the researchers calculated annual maximum, median and mean EVI at mine sites. EVI data was used to measure the extend of and the date of beginning of deforestation (i.e. structural break in the time series) for each mine.
* Data on water quality from the site monitor nearest to each mines were collected from the Central Pollution Control Board’s ENVIS database. They used Biological Oxygen Demand (BOD), a measure of organics pollution, as the proxy for water pollution.

**Step 3:** The researchers then linked the geocoded mine sites to the corresponding cells of environmental outcomes. Of the total 934 mines in their database, they could link 889 of 1km cells of EVI data and 882 of 250 meter cells with corresponding geocoded mine sites. They could also link 538 site monitors to the mines.

**Step 4**: The new EC reform required the mines of area between 5 and 25 ha to hold a public hearing which were not considered big enough to hold public hearing during the previous regime i.e. only the mines of areas above 25 ha were required to hold the hearings. This study exploited the discontinuity around the 25 ha mark and compared the mines marginally above 25 ha with the ones marginally below 25 ha. The final sample included 134 mines, of which 68 were treatment mines (less than 25 ha) and 66 were control (greater than 25 ha). Using data before and after the EC applications, they have estimated a difference-in-difference model.

This study, utilising web scraping to collect data on project characteristics, and various sources of satellite data for measuring the outcomes of interest, is an excellent example of innovative data collection methods in a sector where the evidence base is very small (Rathinam, et.al., 2019b).

## **Using CDR analytics to inform disaster management**

In this section, we briefly outline the process for undertaking CDR analytics to measure, characterise, and predict population displacement and returnee/resettlement patterns in post-disaster settings. While, to date, applications of CDR data analytics have lacked an evaluative component, their potential in this regard is evident. In this section, we briefly outline the process for undertaking CDR analytics to inform the disaster management cycle. We draw on a recent project at Flowminder, which revisited three sudden-onset disaster events to investigate drivers of displacements (individual and contextual) and the feasibility of predicting displacement locations from CDR data and data on disaster intensity and damage, on population density and on the humanitarian response. The three events were: the 2010 earthquake in Haiti, the 2015 Gorkha earthquake in Nepal, and 2016’s Hurricane Matthew in Haiti (REF).

Flowminder has long-established partnerships with the major mobile phone network operators in Haiti and Nepal. Historically, data access has been a major barrier to the scale-up of CDR analytics for humanitarian and development applications. Mobile network operators (MNOs) are, justifiably, hesitant to authorise third-party access, given the need to safeguard subscriber’s personal data[[13]](#footnote-14).

Prior to analysis, MNO data undergo a long series of cleaning and pre-processing steps as part of quality assurance and to support the generation of standardised metrics. A first stage of analysis is undertaken to structure the data in a usable format and to detect data anomalies. Once data are cleaned, quality assured, and converted into an analysable format, a number of preliminary processing steps must be undertaken, including:

* Clustering of cell tower locations
* Assessment of each subscriber’s phone usage behaviours (number of events, frequency, regularity)
* Determination of pre-disaster ‘home’ location

**MNO Data: Pre-processing steps**

Here are some commonly-occurring issues in MNO data. Once identified, corrections and / or accommodations can be performed prior to and/or during the preliminary processing and analysis phases.

1. **Standard data quality issues applicable to MNO data:**

* Item missing data (incomplete data records i.e. missing fields)
* Invalid entries for fields
* Duplicate records
* Interrupted data series’ (e.g. no data for a particular time period)
* Inconsistent values (either in format, or definition) for keys that are used to join multiple datasets together
* Inconsistent entries for the ‘same’ value (e.g. different spellings of the same place name)

1. **Issues specific to MNO datasets, CDR:**

* Inconsistencies or errors in method used for ‘hashing’ (a form of pseudonymisation) subscribers IDs
* Inconsistent ‘hashing’ of sender and recipient IDs for communication events (e.g. standard SMS or phone calls)

1. **Issues specific to MNO datasets: cell location and coverage maps**

* Cell locations occur outside national borders
* Updated cell locations are not consistent with previous cell locations
* Small deviations in updated cell locations, possibly due to inexact GPS measurements
* Datum, projection and coordinate system information are often missing in coverage datasets
* Inconsistent output formats
* For best server or cell-in-isolation maps, need to ensure polygons are labelled in a manner consistent with cell table and/or CDR dataset

1. **Common data anomalies, indicative of a network issue or a sudden change in subscribers’ behaviour due to an event or ‘shock’:**

* Individual cell towers have significantly more/less traffic than normal
* Overall network traffic is significantly higher/lower than expected
* Traffic from a particular region is significantly higher/lower than expected.

The processing steps undertaken to discern at individual level disaster-induced displacements from pseudonymised, time series CDR data are presented below in figure 12.

**Figure 12: Steps in CDR processing for displacement and return/resettlement/recovery pattern analysis**

|  |  |
| --- | --- |
| **Step one:** Calculate ‘distance curves’  Daily, travel distances are calculated to assess departures from an individual’s ‘home’ location, or pre-disaster ‘reference location’. Missing values are imputed to create a regularly-spaced time series. |  |
| **Step two:** Identifying times of changes in location  Displacements are modelled as a shift from one stable location to another. |  |
| **Step three:** Identifying IDPs and ‘stay’ locations  IDPs are distinguished from non-displaced persons by shifts in stable location coinciding with disaster-onset. The post disaster temporary ‘stay’ locations are calculated for each IDP, as well as returns to the pre-disaster home or resettlement at a new stable location. |  |

The analysis disclosed striking commonalities in internally displaced persons’ (IDP) return / resettlement rates, with the fraction of IDPs who remain displaced exhibiting a common rate of decay across all three post-disaster settings studied.

In a further step, the team developed new mobility and social network metrics to permit analysis of the relationships between contextual and individual variables and displacement duration, distance, and trajectories, controlling for the severity of impacts and humanitarian response. The results suggest that the dispersal of an individual’s social contacts and travel history, pre-disaster are highly predictive of their post-disaster displacement trajectories. Individuals with localised travel patterns and social contacts were more likely to be displaced in the vicinity of their usual residence, compared with those with more dispersed travel patterns and social contacts. A majority of IDPs remained within a 10km radius of their usual place of residence. Across the three disasters, 60% - 70% of long-distance displacements (in excess of 100km) involved travel to a familiar location and/or proximate to one or more contacts discernible in the pre-disaster CDR data. This pattern holds controlling for the severity of impacts at local area level, and is consistent across all three disasters.

Results were validated via comparisons with reports retrospectively quantifying population displacements, produced by the International Organisation for Migration (IoM), as well as with reference to data on the intensity of each disaster’s impacts on affected areas. The results indicate that CDR data analysis can be used to predict the estimated number and spatial distribution of IDPs at different time points, based on initial estimates of the number of persons displaced in the immediate wake of a disaster, and to predict recovery / resettlement timelines. This has important implications for post-disaster humanitarian response and resettlement efforts. The same methods can support disaster resilience assessments and planning, and provide a means to compare recovery and resettlement rates across different disaster events.

# Limitations of this study and potential next steps

This map covers large thematic areas and outcomes corresponding to SDGs. Given the wide scope of the outcomes, the evidence is sparse and bunched around a few themes. The thematic gaps here may not be read as actual gaps, but these areas may be not readily relevant to using big data. This map rather shows what evidence or proof of concepts available to measure and evaluate development outcomes using big data.

This map followed a systematic process of searching, screening and coding of studies based on a pre-defined set of criteria in the study protocol developed with inputs from key stakeholders. However, despite the best efforts to search and screen the studies, given the wide scope of big data sources and their application across all developmental themes, and the pace at which the literature growing, it is possible that some relevant studies, especially the measurement studies, could have been missed out.

It was beyond the scope of the study to provide critical quality appraisal of the IEs or the measurement studies given the large number of studies included in the map. Neither the report looked at the details of machine learning methods used in the included studies.

Given the wide scope of development applications, it was not possible to code the studies for all sub-classifications. Though the sub-maps, especially for economic development and livelihoods, health and well-being, and urban development, provide coding at the level 2 indicators, it was not possible to provide granular analysis of development themes corresponding to SDG indicators at level 3. Future systematic maps should aim to produce more granular classifications on the use of big data the indicator level.

There are a few studies that have used machine learning techniques for treatment effect heterogeneity in randomised control trails (Chernozhukov et.al., 2019). However, it was beyond the scope of this report to include the role of big data analytical methods in conventional IE designs such as RCT and other quasi-experimental designs. This is a nascent, but growing literature and could be considered for inclusion in the future maps.

Several studies suggest that the key prospects of big data sources, especially the satellite data, is the long term availability of data which will help evaluate the long term impact of development interventions; possibilities of collecting vast array of information on several contextual factors using big data can help evaluate complex interventions (Bamberger, 2016). However, this map did not code the studies for long term impact nor for complex interventions. Future maps may code and analyse the role of big data for measuring long term impact s and in evaluating complex interventions.

## **Next steps**

The systematic map shows that the studies, both IEs and measurement studies, dramatically increased in the past five years at an increasing rate. Given the potential for faster growth in the availability and computational capacity, it is very likely that the number of studies will grow faster than what we witnessed in the past 5 years. Hence, we recommend that this map to be updated within the next two years.

The fact that more than 80 per cent of the included studies are peer-reviewed shows the growing number of journals interested in big data application in international development, it will be useful to include a more exhaustive grey literature search to identify the full extent of the literature.

This map shows the potential for big data to measure and evaluate various development themes. However, most of these studies are supported by universities and specialist organisations, and conducted by researchers associated with these organisations. Widely disseminating the findings of the map among development researchers, evaluators, practitioners and donors will help promote adoption of big data measures into future IEs.

# Conclusions

Big data has great potential to help address questions of relevance to international development, including for evaluating the effects of interventions. This systematic map compiles IEs, SRs and measurement studies that incorporate big data to highlight how this innovative, new data source is being used to evaluate development outcomes, and more importantly where there is more potential to use big data in the future evaluations. We found 437 studies, of which 48 are IEs, 381 are measurement studies and 8 are SRs. Roughly half the studies are from Asia and another 30 per cent are from Africa, and about 70 per cent is from LMICs. Of the 48 IEs, 8 are randomised controlled trials and the remaining are quasi-experimental studies.

Our results highlight there is considerable potential for using big data for measuring various development outcomes across SDG themes, but big data is more relevant to environmental sustainability, economic development and livelihoods, health and well-being, and urban development. This map also highlights that big data can contribute to evidence base in development sectors where evaluations are not generally feasible due to a lack of data, in particular due to fragile contexts.

One of the key ‘absolute gap’ that the map has identified is that the number of IEs is fewer comparing the measurement studies. Given the fast growing availability of big data and improving computation capacity, there is a great potential for using big data in the future IEs. This is however may not be straightforward as there are several analytical, ethical and logistical challenges that may hinder the use of big data in evaluations. The development community that help set the standards and best practices, and development stakeholders including the donors who facilitate rigorous evaluations and learning, have a strong role to play in facilitating this process. The report highlights the need for setting standards for better reporting on data quality issues, representativeness, construct validity and generalisability, and the need for data transparency and sharing. The report also calls for facilitating better interaction between big data analysts, remote sensing scientists and evaluators.

One of the key findings of the report is that satellite and sensor data are the most used data sources for both measurements studies and impact evaluations. There are several sources of pre-processed satellite data that could be used in evaluations directly without the evaluators having to process them using complex machine learning models themselves. Satellite data seems to be particularly useful in the context where the development interventions and the outcomes studied have spatial dimension economic activity at local level, urban development, forest cover, land productivity, distribution of the population, etc. where the outcome and other covariates are measured in a spatial scale i.e. villages, counties, districts, plots, protected areas. CRD data, on the other hand, despite being used widely in measurement studies is not used in IEs yet. Data deficiency in international development is partly due to the fragile contexts such as diseases spread, violence, natural calamities, difficult terrains, etc. This map highlights the potential of big data in fragile contexts as one quarter of the studies were conducted in fragile contexts.

For the evaluators and researchers, the report calls for better reporting on data quality, ethics and transparency. There is also an absolute gap in using mixed methods jointly with big data and cost effectiveness. For the donors, this report calls for more efforts on setting up best practices and ethical standards, and facilitating more interaction among remote sensing scientist, big data analysts and development evaluators.

# References

Archie, M., Gershon, S., Katcoff, A., & Zeng, A. (2018). Who’ s Watching ? De-anonymization of Netflix Reviews using Amazon Reviews.

Bamberger, M, 2088. *Dealing With Complexity in Development Evaluation A Practical Approach.*Sage Publications.

BenYishay, A., Trichler, R., Runfola, D., & Goodman, S. (2018). Final Report: Evaluation of the Infrastructure Needs Program II. Williamsburg, VA: AidData at William & Mary.

Blondel, V.D., Decuyper, A. and Krings, G., 2015. A survey of results on mobile phone datasets analysis. EPJ data science, 4(1), p.10.

Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. Science, 350(6264), 1073-1076.

Burke, M., Lobell, D. 2017 Satellite-based assessment of African yields, Proceedings of the National Academy of Sciences Feb 2017, 114 (9) 2189-2194; DOI: 10.1073/pnas.1616919114

Carrera, F. & Guerin, Stephen & Thorp, Joshua. (2013). By the people, for the people: the crowdsourcing of “streetbump”,an automatic pothole mapping app. UDMS2013-Street Bump Fina. 52. 10.5194/isprsarchives-XL-4-W1-19-2013.

Chernozhukov V, Demirer M, Duflo E, Fernadezval I. Generic machine learning inference on heterogeneous treatment effects in randomized experiments. 2019. arXiv:1712.04802v4.

Decuyper, A., Rutherford, A., Wadhwa, A., Bauer, J. M., Krings, G., Gutierrez, T., Luengo-Oroz, A. (2014). Estimating food consumption and poverty indices with mobile phone data, CoRR, abs/1412.2595.

DFID (2016) Memorandum by the Department for International Development. DFID. Available at <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/international-development-committee/dfids-allocation-of-resources/written/28276.pdf>

Fu, Haishan (2019) Data for development impact: Why we need to invest in data, people and ideas, https://blogs.worldbank.org/voices/data-for-development-impact-why-we-need-to-invest-in-data-people-and-ideas. Accessed on 16 Aug 2019.

Gaarder, M., Annan J (2013). Impact Evaluation of Conflict Prevention and Peacebuilding Interventions. The World Bank Policy Research Working Paper 6496, <http://documents.worldbank.org/curated/en/445741468177842482/pdf/WPS6496.pdf>. Accessed on May 29.

Hammer, Cornelia & Kostroch, Diane & Quiros, Gabriel. (2017). Big Data: Potential, Challenges and Statistical Implications. Staff Discussion Notes. 17. 1. 10.5089/9781484310908.006.

Jain, Meha 2020. The Benefits and Pitfalls of Using Satellite Data for Causal Inference, Review of Environmental Economics and Policy, Volume 14, Issue 1, Winter, Pages 157–169, https://doi.org/10.1093/reep/rez023

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. Science, 353(6301), 790-794.

Jayachandran, S, et al (2017). Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation. Science Vol. 357, Issue 6348, pp. 267-273.

Letouzé, Emmanuel 2016 Big Data and Development: An Overview

Lokanathan, S., Gomez, T., and Zuhyle, S 2017 Mapping Big Data Solutions for the Sustainable Development Goals, LIRNEasia

Lu, X., Bengtsson, L., & Holme, P. (2012). Predictability of population displacement after the 2010 Haiti earthquake. Proceedings of the National Academy of Sciences, 109(29), 11576-11581.

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A. H. Byers, 2011 “Big data: The next frontier for innovation, competition, and productivity.,” McKinsey & Company.

Olteanu, A., Castillo, C., Diaz, F., and Kıcıman, E., (2019) Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. Front. Big Data, https://doi.org/10.3389/fdata.2019.00013

Osorio, J. A. V. I. E. R. "Numbers under fire: The challenges of gathering quantitative data in highly violent settings." Social Science Research Council, Drugs, Security and Democracy Program (DSD) Working Papers on Research Security 6 (2014).

Pande, R and Sudarshan, A, 2019, Harnessing transparency initiatives to improve India's environmental clearance process for the mineral mining sector, 3ie Impact Evaluation Report 92. New Delhi: International Initiative for Impact Evaluation (3ie). Available at: https://doi.org/10.23846/TW8IE92

Perera-Gomez, Thavisha & Lokanathan, Sriganesh. (2017). Leveraging Big Data to Support Measurement of the Sustainable Development Goals. SSRN Electronic Journal. 10.2139/ssrn.3058530.

Pellegrini, L, 2019. Community monitoring of socio-environmental liabilities with advanced technologies in the Ecuadorian and Peruvian Amazon, 3ie Grantee Final Report. New Delhi: International Initiative for Impact Evaluation (3ie).

Pelletier, J.; Gélinas, N.; Skutsch, M. The Place of Community Forest Management in the REDD+ Landscape. Forests 2016, 7, 170.

Phillips D, Coffey C, Tsoli S, Stevenson J, Waddington H, Eyers J, White H and Snilstveit B, 2017 A Map of Evidence Maps Relating to Sustainable Development in Low- and MiddleIncome Countries Evidence Gap Map Report. CEDIL Pre-Inception Paper.

Puri, J, Nath, M, Bhatia, R and Glew, L. 2016. Examining the evidence base for forest conservation interventions, Evidence Gap Map Report 4. International Initiative for Impact Evaluation (3ie): New Delhi.

Puri, J., & Rathinam, F. 2019. Challenges in real-world impact evaluations: Some learning on costs and timeliness. IEU Working Paper No. 03, Green Climate Fund. Songdo, South Korea

Rathinam, F, Cardoz, P, Siddiqui, Z and Gaarder, M, 2019a. Transparency and accountability in the extractives sector: a synthesis of what works and what does not, 3ie Working Paper 33. New Delhi: International Initiative for Impact Evaluation (3ie). Available at: doi: https://doi.org/10.23846/WP0033

Rathinam, F, Finetti, J, Snilstveit, B, Siddiqui, Z, Chirgwin, H, Appell, R, Dickens, E and Gaarder, M, 2019b. The effect of transparency and accountability interventions in the extractive sectors: an evidence gap map, 3ie Evidence Gap Map Report 14. New Delhi: International Initiative for Impact Evaluation (3ie). Available at: https://doi.org/10.23846/EGM014

Salganik, M. J. (2017). Bit by Bit: Social Research in the Digital Age. Princeton, NJ: Princeton University Press.

Serajuddin, Umar; Uematsu, Hiroki; Wieser, Christina; Yoshida, Nobuo; Dabalen, Andrew L. 2015. Data deprivation: another deprivation to end (English). Policy Research working paper; no. WPS 7252. Washington, DC: World Bank Group

Shayda Mae Sabet & Annette N. Brown (2018) Is impact evaluation still on the rise? The new trends in 2010–2015, Journal of Development Effectiveness, 10:3, 291-304, DOI: 10.1080/19439342.2018.1483414UN (2018) The Sustainable Development Goals Report 2018.

UN (2019) Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development.

UN Global Pulse (2012) Big Data for Development: Challenges & Opportunities.

UN Global Pulse (2013) Big Data for Development: A primer.

UN Global Pulse. 2016 'Integrating Big Data into the Monitoring and Evaluation of Development Programmes,' Available at https://beta.unglobalpulse.org/wp-content/uploads/2016/12/integratingbigdataintomedpwebungp-161213223139.pdf

Vaitla, B. (2014) The Landscape of Big Data for Development: Key Actors and Major Research Themes, UN Foundation/Data2X

Van der Windt, P and Humphreys, M, 2016. Crowdseeding in Eastern Congo: Using Cell Phones to Collect Conflict Events Data in Real Time. Journal of Conflict Resolution, 60(4), 748-781.

Webster, J, Landegger, J, Bruce, J, Malunda, D, Chantler, T, Kumakech, E, Schmucker, L, Kiapi, L, Kozuki, N, Olorunsaiye, C and Byrne, E, 2019. Impacts of IRC’s Fifth Child community engagement strategy to increase immunisation in northern Uganda, 3ie Grantee Final Report. New Delhi: International Initiative for Impact Evaluation (3ie).

Wassenich, Paul. 2007. Data for impact evaluation (English). Doing impact evaluation series; no. 6. Washington, DC: World Bank. <http://documents.worldbank.org/curated/en/332891468313760995/Data-for-impact-evaluation>

Webster, J, Landegger, J, Bruce, J, Malunda, D, Chantler, T, Kumakech, E, Schmucker, L, Kiapi, L, Kozuki, N, Olorunsaiye, C and Byrne, E, 2019. Impacts of IRC’s Fifth Child community engagement strategy to increase immunisation in northern Uganda, 3ie Grantee Final Report. New Delhi: International Initiative for Impact Evaluation (3ie).

Wilson, R., zu Erbach-Schoenberg, E., Albert, M., Power, D., Tudge, S., Gonzalez, M., Pitonakova, L. (2016). Rapid and near real-time assessments of population displacement using mobile phone data following disasters: the 2015 Nepal earthquake. PLoS currents, 8.

York, P. and Bamberger, M. 2020. Measuring results and impact in the age of big data: The nexus of evaluation, analytics, and digital technology, the Rockefeller Foundation. Available at https://www.rockefellerfoundation.org/wp-content/uploads/Measuring-results-and-impact-in-the-age-of-big-data-by-York-and-Bamberger-March-2020.pdf

# Appendix A: OECD definition of fragile states

OECD defines fragility as the combination of exposure to risk and insufficient coping capacity of the state, system and/or communities to manage, absorb or mitigate those risks. Fragility can lead to negative outcomes including violence, the breakdown of institutions, displacement, humanitarian crises or other emergencies. The OECD fragility framework considers not only current exposure to negative events such as natural disasters and armed conflict but also capacity to cope with likely future negative events.

The new framework considers five dimensions of fragility:

* **Economic fragility** is vulnerability to risks stemming from weaknesses in economic foundations and human capital including macroeconomic shocks, unequal growth and high youth unemployment
* **Environmental fragility** is vulnerability to environmental, climatic and health risks that affect citizens’ lives and livelihoods. Risk factors can be external or internal, including exposure to natural disasters; air, water and sanitation quality; prevalence of infectious disease; number of uprooted people; and vulnerability of household livelihoods.
* **Political fragility** is vulnerability to risks inherent in political processes, events or decisions; political inclusiveness (including of elites); and transparency, corruption and society’s ability to accommodate change and avoid repression. Risk factors include regime persistence, state-sponsored violence or political terror, and levels of corruption.
* **Security fragility** is the vulnerability of overall security to violence and crime, including both political and social violence. Risks are measured by the homicide rate, level of violent organised crime, number of deaths from non-state actors or terrorism, number of battle deaths from conventional warfare, and levels of domestic violence.
* **Societal fragility** is vulnerability to risks affecting societal cohesion that stem from both vertical and horizontal inequalities, including inequality among culturally defined or constructed groups and social cleavages. Risk indicators include income inequalities (vertical) and social inequalities related to gender, growth in urbanisation and numbers of displaced people.

In the past few years, OECD has moved away from the ‘fragile states list’, towards measuring each of those five dimensions on a spectrum of intensity for 58 fragile contexts. This comes as a part of their effort to move towards a universal concept of fragility, recognizing that it affects not only developing countries but all countries to some extent.

Table A1-1: Fragile contexts provided in the OECD fragility framework 2018 (decreasing order of severity):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Economic** | **Environmental** | **Political** | **Security** | **Societal** |
| Central African Rep. | Somalia | Eritrea | Syria | South Sudan |
| South Sudan | Central African Rep. | Sudan | Libya | Yemen |
| Liberia | South Sudan | DPRK | Yemen | Syria |
| Somalia | Chad | Yemen | Somalia | Egypt |
| Solomon Islands | DRC | Syria | Afghanistan | Somalia |
| Afghanistan | Burundi | Somalia | South Sudan | Sudan |
| Niger | Mozambique | South Sudan | Iraq | Burundi |
| Sierra Leone | Guinea-Bissau | Burundi | Sudan | Eritrea |
| Comoros | Niger | Chad | Central African Rep. | DPRK |
| Guinea-Bissau | Liberia | Congo | Nigeria | DRC |
| Mozambique | Burkina Faso | Gambia | Mali | Congo |
| Haiti | Guinea | Ethiopia | Pakistan | Chad |
| Mali | Sierra Leone | Guinea | Chad | Equatorial Guinea |
| Congo | Mali | DRC | DRC | Pakistan |
| Libya | Malawi | Mauritania | West Bank & Gaza | Iran |
| Syria | Zambia | Bangladesh | Cameroon | Central African Rep. |
| Yemen | Cameroon | Angola | Egypt | Afghanistan |
| Tajikistan | Eswatini | Djibouti | Niger | Zimbabwe |
| Djibouti | Côte d'Ivoire | Venezuela | Haiti | Ethiopia |
| Gambia | Eritrea | Iran | Ethiopia | Kenya |
| Mauritania | Ethiopia | West Bank & Gaza | Kenya | Eswatini |
| Iraq | Afghanistan | Zimbabwe | Myanmar | Mauritania |
| Honduras | DPRK | Madagascar | Nepal | West Bank & Gaza |
| Timor-Leste | Yemen | Afghanistan | Burkina Faso | Angola |
| Eswatini | Haiti | Central African Rep. | Congo | Iraq |
| Pakistan | Syria | Libya | Eritrea | Guatemala |
| DRC | Tanzania | Egypt | Burundi | Gambia |
| Eritrea | Zimbabwe | Guinea-Bissau | Mauritania | Cameroon |
| Burkina Faso | Madagascar | Iraq | Iran | Haiti |
| Burundi | Congo | Sierra Leone | Comoros | Uganda |
| Zimbabwe | Rwanda | Mali | Bangladesh | Rwanda |
| Ethiopia | Myanmar | Haiti | Venezuela | Lao PDR |
| Chad | Uganda | Côte d'Ivoire | Tanzania | Myanmar |
| Guinea | Pakistan | Mozambique | Uganda | Venezuela |
| Malawi | Comoros | Myanmar | Mozambique | Tajikistan |
| Madagascar | Bangladesh | Rwanda | Guatemala | Bangladesh |
| Rwanda | Tajikistan | Liberia | Côte d'Ivoire | Guinea |
| DPRK | Lao PDR | Nigeria | Sierra Leone | Libya |
| Zambia | Djibouti | Pakistan | Liberia | Djibouti |
| Uganda | Mauritania | Burkina Faso | Honduras | Guinea-Bissau |
| Tanzania | Kenya | Uganda | Guinea | Tanzania |
| Papua New Guinea | Papua New Guinea | Kenya | Guinea-Bissau | Honduras |
| Nepal | Angola | Honduras | Angola | Mozambique |
| Equatorial Guinea | Iraq | Niger | Gambia | Côte d'Ivoire |
| Lao PDR | Gambia | Tajikistan | Tajikistan | Madagascar |
| Myanmar | Sudan | Equatorial Guinea | Equatorial Guinea | Nigeria |
| Cameroon | Nepal | Cameroon | Lao PDR | Papua New Guinea |
| Angola | Timor-Leste | Eswatini | Solomon Islands | Zambia |
| Bangladesh | Nigeria | Lao PDR | Madagascar | Mali |
| Guatemala | Guatemala | Guatemala | Papua New Guinea | Sierra Leone |
| Côte d'Ivoire | Solomon Islands | Zambia | Zimbabwe | Nepal |
| Kenya | Venezuela | Timor-Leste | Zambia | Burkina Faso |
| Venezuela | Honduras | Nepal | Timor-Leste | Malawi |
| Nigeria | West Bank & Gaza | Tanzania | Malawi | Niger |
| Sudan | Iran | Papua New Guinea | Eswatini | Timor-Leste |
| Iran | Egypt | Comoros | Djibouti | Liberia |
| West Bank & Gaza | Libya | Malawi | Rwanda | Solomon Islands |
| Egypt | Equatorial Guinea | Solomon Islands | DPRK | Comoros |

# Appendix 2: Details of Sub-maps

Outcome categories and definitions

|  |  |  |
| --- | --- | --- |
| **Category** | **Definition** | **Sub-categories** |
| **Economic development and livelihoods**  **(SDG 1 & 8)** | End poverty in all its forms everywhere;  Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all | Poverty mapping/measurement |
| Measuring wealth/GDP |
| Access to basic services |
| Social Protection including financial inclusion |
| Vulnerability reduction |
| Economic growth and productivity |
| Full and productive employment including youth unemployment |
| Eradicate forced and underage labour, and protect labour rights and promote safe workspace |
| Promote sustainable tourism |
| Strengthen capacity of domestic financial institutions |
| **Health and well-being**  **(SDG 3)** | Ensuring healthy lives and promoting well-being for all ages | Reduce mortality |
| End epidemics of communicable diseases |
| Strengthen prevention and treatment of substance abuse |
| Decrease fatalities due to road accidents |
| Universal access to sexual and reproductive healthcare services |
| Achieve universal health coverage |
| Reduce fatalities due to pollution |
| **Governance and human rights**  **(SDG 5, 10 & 16)** | Achieving gender equality and empower all women and girls; | |
| Reduce inequality within and among countries; | |
| Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels | |
| **Urban development**  **(SDG 11)** | Make cities and human settlements inclusive, safe, resilient and sustainable | Access to affordable housing |
| Access to affordable transport systems |
| Enhance inclusive and sustainable urbanization |
| Strengthen efforts to protect cultural and natural heritage |
| Better disaster management |
| Reduce environmental impact of cities |
| Universal access to green public spaces |
| **Environmental sustainability**  **(SDG 12, 13, 14 & 15)** | Ensure sustainable consumption and production patterns; | |
| Take urgent action to combat climate change and its impacts;  Conserve and sustainable use the oceans, seas and marine resources for sustainable development; | |
| Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss | |

# Appendix 3: Search strategy and the data bases searched

We developed a systematic search strategy in consultation with an information specialist after finalising the protocol.

We searched the following general databases:

* CAB Abst: <https://www.cabi.org/publishing-products/online-information-resources/cab-abstracts/>
* Econlit (Ovid): <http://www.ovid.com/site/catalog/databases/52.jsp>
* Ebsco Discovery: <https://www.ebscohost.com/discovery>
* Scopus: <https://www.scopus.com/>
* Social Sciences Citation Index (SSCI) (via Web of Science): <https://library.maastrichtuniversity.nl/collections/databases/ssci/>

Bilateral and multilateral agencies and general repositories of impact evaluations in international development:

* 3ie Repository of Impact Evaluations <http://www.3ieimpact.org/en/evidence/impactevaluations/>
* 3ie RIDIE (Registry for International Development Impact Evaluations): <http://ridie.3ieimpact.org/>
* USAID Evaluation Clearing House: <https://dec.usaid.gov/dec/content/evaluations.aspx>
* Innovations for Poverty Action (IPA) [www.poverty-action.org/project-evaluations](http://www.poverty-action.org/project-evaluations)
* J-Poverty Action Lab (JPAL) :[www.povertyactionlab.org](http://www.povertyactionlab.org/)
* The World Bank: www.worldbank.org/
* AEA RCT Registry : <https://www.socialscienceregistry.org/>
* DFID Research for Development (R4D): <http://r4d.dfid.gov.uk/>
* Campbell Collaboration, [www.campbellcollaboration.org](http://www.campbellcollaboration.org/)
* African Development Bank (AfDB): <https://www.afdb.org/en/documents/publications/>
* BREAD: http://ibread.org/bread/papers
* Center for Effective Global Action (CEGA): <http://cega.berkeley.edu/evidence/>

We searched specialist organizational databases, which might include big data measurement studies and impact evaluations.

* AidData — A Research Lab at William & Mary: <https://www.aiddata.org/>
* Flowminder: <https://web.flowminder.org/>
* Stanford University, Center on Food Security and the Environment: <https://fse.fsi.stanford.edu/>
* UN Global Pulse: <https://www.unglobalpulse.org/>
* Data-Intensive Development Lab: <http://didl.berkeley.edu/>
* Global partnership for sustainable development: [http://www.data4sdgs.org](http://www.data4sdgs.org/)
* Arxiv database (<https://arxiv.org>).
* Eldis: <https://www.eldis.org>

We also searched grey literature via Google Scholar, and check references of any systematic review that we find in our searches and meet our inclusion criteria.

Given the relative recent state of the evidence base in this field and the fact that most programs/ initiative have started to flourish in the late 2000s, we conducted the searches from 2005 onwards.

**Web of Science (Social Sciences Citation Index & Science Citation Index) – Searched 22nd September 2019**

# 1 34,746

TS=( "big data" or metadata or "meta data" or meta-data or "macro data" or macrodata or "mass data" or massdata or "large data" or "bulk data" )

# 2 45,418

TS=( ((satellite\* NEAR/3 imag\*) or sensor\* or surveillance or meter\* or drone\* or ((mobile\* or cell) NEAR/3 (phone\* or telephone\*)) or mobiles or gps or "global positioning" or gis or "global information system\*" or self-tracking or "self tracking") NEAR/3 data)

# 3 41,524

TS=( ((e-commerce or commerc\* or business\* or "credit card\*" or ATM\* or "automated teller" or "cash machine\*" or "money transfer\*") and (transaction\* or record or records)) or ((savings or loan or loans) NEAR/3 repay\*) NEAR/3 data )

# 4 109

TS=( ((((toll or tolls) NEAR/2 (road\* or highway\* or motorway\*)) or "public transport") NEAR/3 data) )

# 5 1,036

TS=( (((online or internet or web or virtual or google) NEAR/3 (search\* or log or logs)) NEAR/3 data) )

# 6 14,088

TS=( (("social network\*" or "social media" or "opinion platform\*" or blog\* or Twitter or Facebook or LinkedIn or YouTube or Wiki\* or Open or "text messag\*" ) NEAR/3 data) )

# 7 596

TS=( (((citizen\* NEAR/2 (report\* or derived or submit\* or communicat\* or inform or informed or informing or feedback)) or hotline\* or crowdsourc\*) NEAR/3 data) )

# 8 133,635

#7 OR #6 OR #5 OR #4 OR #3 OR #2 OR #1

# 9 35,994

TS=( ((poverty NEAR/5 (eradicat\* or measur\* or map\*)) or ((wealth or GDP) NEAR/3 measur\*) or (access\* NEAR/5 (service\* or financ\*)) or "social protection" or (reduc\* NEAR/3 vulnerab\*)) )

# 10 36,506

TS=( (((hunger or malnutrition or malnourished) NEAR/3 (eradicat\* or decreas\* or end\*)) or "food security" or (improv\* NEAR/3 nutrition\*) or "sustain\* agriculture" or (access\* NEAR/3 food\*) or ((productiv\* or income\*) NEAR/3 ("small scale" NEAR/2 produc\*)) or (sustain\* NEAR/3 "food produc\*") or ((maintain\* or maintenance) NEAR/2 "genetic diversity")) )

# 11 57,651

TS=( (((health or well-being or wellbeing or "well being") NEAR/2 (promot\* or ensur\*)) or (reduc\* NEAR/3 (maternal or child\* or "under 5\*" or under-5\* or "under five\*") NEAR/2 (mortality or death\*)) or (("communicable disease\*" or "infectious disease\*") NEAR/3 (epidemic\* or pandemic\*)) or (reduc\* NEAR/2 ("premature mortality" or "premature death\*")) or ((prevent\* or treatment) NEAR/3 "substance abuse") or ((reduc\* or prevent\*) NEAR/3 (road\* or traffic or vehicle\*) NEAR/2 (accident\* or death\* or fatal\*)) or (access\* NEAR/2 (reproductive or sexual) NEAR/2 service\*) or (universal NEAR/2 health NEAR/2 coverage) or (pollut\* NEAR/3 (death\* or mortality or fatal\*) NEAR/2 reduc\*)) )

# 12 26,746

TS=( ((education or "life-long learning" or (skill\* NEAR/3 (build\* or acquir\* or learn\* )) or literacy or numeracy) NEAR/3 (access\* or complet\* or graduat\* or gender\* or opportunit\*)) )

# 13 40,719

TS=( ((gender\* or women or girl\* or female\*) NEAR/3 (equalit\* or inequalit\* or discriminat\* or empower\* or violen\* or harm\* or (unpaid NEAR/2 (domestic or house\*)) or participat\* or ((sexual or reproductive) NEAR/3 right\*))) )

# 14 100,434

TS=( ((water or WASH or sanitation) NEAR/3 (access\* or quality or sustainable or manag\* or ecosystem\*)) )

# 15 139,810

TS=( ((energy NEAR/3 (access\* or renewable or afford\* or clean or sustainable or efficien\*)) ))

# 16 210,410

TS=( (((econom\* NEAR/2 (growth or productivity)) or employment or labour or work or job or jobs or development or "global resource\*" or ((youth or "young people" ) NEAR/3 unemploy\*) or tourism or (financ\* NEAR/3 institution\*) or bank or banks or banking) NEAR/3 (sustain\* or increas\* or promot\* or efficien\* or reduc\* or forced or underage or "under age" or rights or safe or strengthen\* or (capacity NEAR/2 build\*))) )

# 17 19,679

TS=( ((industr\* or innovation or infrastructure or (financ\* NEAR/2 service\*) or ((scien\* or techn\*) NEAR/2 research)) NEAR/3 (resilen\* or sustain\* or promot\* or foster\* or access\* or upgrad\* or modern\*)) )

# 18 97,676

TS=( ((inequalit\* NEAR/2 (reduc\*)) or ((income NEAR/2 growth) or "equal opportunit\*" ) NEAR/2 (sustain\* or increas\* or promot\*)) or (inclusivity NEAR/2 (promot\* or sustain\* or empower\*)) or (equality NEAR/2 (increas\* or promot\* or engender\*)) or ((global NEAR/2 (financ\* or market\* or institution\*) NEAR/3 (regulat\* or monitor\* or legislat\* or law or laws)) or (("developing countr\*" or LMIC\* or "low and middle-income" or "third world" or "global south") NEAR/3 (represent\* NEAR/3 (enhanc\* or ensur\* or increas\* or promot\*)) or (migration NEAR/3 (policy or policies) NEAR/3 plan\*) or (different\* or special or positive) NEAR/2 (treat\* or discriminat\*)))

# 19 36,884

TS=( ((city or cities or urban\* or communit\* or village\* or housing or dwelling\* or transport\* or ((cultural or natural) NEAR/3 heritage) or ((disaster\* or emergenc\*) NEAR/3 manag\*) or "green space\*") NEAR/3 (sustain\* or safe or inclusive or resilien\* or afford\* or protect\* or ((reduc\* or limit\* or mitigat\*) NEAR/3 "environmental impact" NEAR/3 (city or cities or urban\*) or access\*)) ))

# 20 149,994

TS=( ((consum\* or production or "natural resource\*" or waste or wastes or procurement or (information NEAR/2 (access\* or availab\*))) NEAR/3 (sustain\* or responsible or efficien\* or environment\* or reduc\*)) )

# 21 201,412

TS=(("climat\* change\*" or "global warming" or (greenhouse NEAR/2 (effect\* or gas\*)) NEAR/3 (action or combat\* or mitigat\* or impact\* or resilien\* or adapt\* or (national NEAR/2 (policy or policies)) or ((promot\* or rais\* or improv\*) NEAR/3 aware\*)) ))

# 22 54,354

TS=( ((ocean\* or sea or seas or marine or coastal or fishing or fisher\* or "small island\*") NEAR/3 (sustain\* or conserv\* or manag\* or protect\* or regulat\* or ecosystem\* or (reduc\* NEAR/2 pollut\*) or acidity) ))

# 23 22,000

TS=( ((terrestrial or land or land-based or forest\* or mountain\* or (natural NEAR/2 (habitat or ecosystem\*)) or poaching or poacher\* or traffick\*) NEAR/3 (sustain\* or conserv\* or protect\* or (degrad\* NEAR/3 (neutral or prevent\* or protect\* or minimi\* or reduc\*)) or ("genetic resources" NEAR/2 (utilis or utiliz\* or use or using or promot\*)) or prohibit\* or ban or banning or banned) or (("alien species" NEAR/2 (invasive or introduced or imported)) NEAR/3 (prevent\* or impact\*)) or ((ecosystem\* or biodiversity) NEAR/3 (national or local\*) NEAR/3 (plan\* or policy or policies))) )

# 24 2,378

TS=(((societ\* NEAR/3 (peaceful or inclusive)) or justice or "rule of law") NEAR/3 (sustain\* or access\* or effective\* or accountab\* or strong or robust or promot\*))

# 25 105,185

TS=((violen\* or corrupt\* or ((illicit or illegal) NEAR/2 financ\* NEAR/2 (activit\* or dealing\*)) NEAR/2 (reduc\* or mitigat\* or eliminat\*)) or (institution\* NEAR/2 (accountab\* or transparen\* or effective\*)))

# 26 17,951

TS=(((("decision making" or decision-making or participati\*) NEAR/2 (representative or democra\*)) or (("global governance" NEAR/3 (participati\* or involv\* or includ\*)) NEAR/3 ("developing countr\*" or LMIC\* or "low and middle-income" or "global south" or "third world")) or ("legal identit\*" NEAR/3 (provi\* or promot\* or regist\*)) or (information NEAR/3 access\*)) )

# 27 141

TS= (("sustainable development goal\*" NEAR/3 (partnership\* or implement\* or revitali\* or strengthen\* or sustainability or (capacity NEAR/2 build\*) or ((development or financial or domestic) NEAR/2 (resource\* or assistance)) or ((investment or technolog\* or ("trading system\*" NEAR/2 universal)) NEAR/3 (promot\* or innovation\*)) or (knowledge NEAR/2 shar\*) or "technology bank\*" or ((exports or market\*) NEAR/3 (increas\* or promot\*) NEAR/3 ("developing countr\*" or LMIC\* or "low and middle-income" or "third world" or "global south")) or "global macroeconomic" or ((policy or policies) NEAR/2 coheren\*) or "global partnership\*" or "public-private partnership\*" or (progress NEAR/3 (monitor\* or assess\* or evaluat\* or review\*)))) )

#27 OR #26 OR #25 OR #24 OR #23 OR #22 OR #21 OR #20 OR #19 OR #18 OR #17 OR #16 OR #15 OR #14 OR #13 OR #12 OR #11 OR #10 OR #9

# 28 1,196,499

#28 AND #8

# 29 16,624

# 30 4,365

TS=(sdg or sdgs or "sustainable development goal\*")

#30 AND #1

# 31 35

#31 OR #29

# 32 **16,625** (Indexes=SCI-EXPANDED, SSCI Timespan=2005-2019)

#28 AND #8

# 33 **3,519** (Indexes=SSCI Timespan=2005-2019)

# Appendix 4: Screening protocol and data extraction

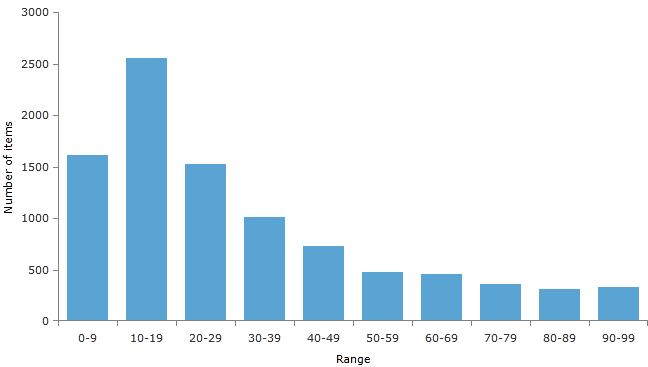
|  |  |
| --- | --- |
| **Variable Name** | **Variable Description** |
| study id | Unique id ascribed to each record |
| title name | Use only the English version of the publication's main title. If paper is not written in English and has the title translated, use the translated version of the title. If the publication does not provide an English version, include the title in its original language. Please enter title in sentence case. Ensure there are no line breaks. |
| Language | Select full text language that applies: English |
| Open access | If the study's (full-text) content is available, code as "Yes". If study has paywalls code as "No".  Please save the pdf in the Dropbox folder called “Full Text PDFs” using the following format Firstauthorsurname\_year\_record id    If study has multiple versions, in other words, if the study has been published as both a journal article and a working paper, both versions may be included in the IER. |
| **Type of big data** | Select one or more from the list based on the type of big data being used:   1. Human-sourced information (Social Networks)    1. Social Networks    2. Internet searches    3. Mobile data content    4. Citizen Reporting or Crowdsourced data 2. Process-mediated data (Traditional Business systems and Websites)    1. Data produced by public agencies    2. Data produced by businesses    3. Cell phone call record details 3. Machine-generated data (Automated Systems)    1. Data from fixed sensors    2. Data from mobile sensors (tracking)    3. Data from Satellites |
| **Outcome** | Select ONE outcome that applies according to the intervention being evaluated:  Economic development and livelihood  Sustainable agriculture and food security  Health and well-being  Education  Water and sanitation  Governance and human rights  Energy, industry and infrastructure  Urban development  Environment sustainability  Global partnership |
| **Sub**-**outcomes** | Poverty mapping/measurement  Measuring wealth/GDP  Access to basic services  Social Protection including financial inclusion  Vulnerability reduction  Economic growth and productivity  Full and productive employment including youth unemployment  Eradicate forced and underage labour, and protect labour rights and promote safe workspace  Promote sustainable tourism  Strengthen capacity of domestic financial institutions  Reduce mortality  End epidemics of communicable diseases  Strengthen prevention and treatment of substance abuse  Decrease fatalities due to road accidents  Universal access to sexual and reproductive healthcare services  Achieve universal health coverage  Reduce fatalities due to pollution  Achieving gender equality and empower all women and girls;  Reduce inequality within and among countries;  Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels  Access to affordable housing  Access to affordable transport systems  Enhance inclusive and sustainable urbanization  Strengthen efforts to protect cultural and natural heritage  Better disaster management  Reduce environmental impact of cities  Universal access to green public spaces  Ensure sustainable consumption and production patterns;  Take urgent action to combat climate change and its impacts;  Conserve and sustainable use the oceans, seas and marine resources for sustainable development;  Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss |
| **Gender** Equity focus | Does this study consider gender and/or\* equity?  Yes  No |
| **Cost** | Whether cost data is provided?  Yes  No |
| **Population** | What population sub-groups does the study target?  Rural  Urban  Refugees  Conflict affected persons  Ethnic minorities  Sex |
| **Geographical Coverage** | Select the continent/region in which the study was conducted:  Asia  Africa  North America  South America  Europe  Australia  Antarctica |
| **Fragile context** | Select the type based on fragile context definition   * + Conflict related difficulty   + Logistical difficulty   + All others |
| Evaluation design | Select one of three options defined as:  1. Experimental:  a) RCT defined as prospective randomised assignment, where randomisation is implemented by researchers (or by decision makers in the context of an evaluation study)  2. Quasi-experimental:  a) Quasi-random assignment: i) regression discontinuity design (sharp designs) or ii) natural experiment in which exposure to treatment is random  b) Non-random assignment: i) Studies that control for unobservables (DID, FE, IV, Fuzzy RDD, ITS) or ii) studies that control for observables only (e.g. statistical matching, synth control, regression adjustment)  3.Measurement Studies  The studies that have innovatively used big data to measure and validate an SDG indicator such as poverty mapping, food security, forest cover, etc. |
| Evaluation method | If Experimental then select (only if one or two selected under Evaluation design):  Randomised controlled trials  If Quasi-experimental then select:  Sharp RDD  DID  FE estimation  IV estimation  Fuzzy RDD  Statistical matching (includes PSM) |
| Methods for analysing big data | Select one or more methods used in the analysis of ‘big data’:  Machine Learning: Supervised learning, Unsupervised learning, Clustering, Anomaly detection, Random Forests, Artificial Neural Network, Convolutional Neural Network, Support vector regression.  Bayesian Geostatistical models: Latent Gaussian models, Integrated Nested Laplace Approximations.  Natural Language Processing: Vectorisation, Embedding, Categorisation, Sentiment Analysis**.**  Any other: Please specify. |
| **Big data validation** | Select one or more methods used to calibrate / validate results of big data analysis using survey data:  Administrative/survey data used for calibration or validation? Yes/no  Type of data   * Administrative data * Survey data   + Secondary survey data   + Primary survey data   Survey name   * Demographic Health Survey * Living Standards Measurement Study (LSMS) * Other, Specify   Mode of data collection   * Face-to-face * Telephone-mode * Mobile-phone telephone mode * Duel-frame * Other, Specify   Sample size  Sample design  Link to survey details (if available) |
| Mixed methods | Select YES if study includes quantitative and qualitative analyses, otherwise select NO. |
| **Transparency in data collection, analysis and reporting** | Choose if the study discusses the following:   * Are the data collection methods described? * Are data quality issues such as completeness (missing or incomplete entries, and empty cells) and noise discussed? * Is the data representative of the population of interest? * Is the construct validity explained? i.e. is there a discussion on how the big data based indicator measures what the study claims to measure? * Are the results generalizable? For example, are the research findings generalizable to other situations such as other platforms (data source), communities, or over time? * Are data and codes publically available for replication? * Any other data, analysis and reporting challenge discussed? |
| **Ethical approval and discussion** | Choose if the study discusses the following:   * Ethical approval obtained? * Consent for secondary and other use of data * Data privacy * Data security and governance arrangement * Unintended exclusion * Unintended consequence on any group of people and/or individuals. * Others, specify. |
| Unit of observation | Enter all the levels of observation of the variables used for the analysis:   * Community * Cohort (includes schools or clinics) * Household * Individual   If more than one, include in separate rows   * Country * Districts * Sub-districts * Village/city |
| **Funding Agency** | What category of funding agency funded the research?  *Note: only code if reported in the study, no need to do additional research to find.*  Select one of the following:  · Government agency  · International Aid Agency  · International Financial Institution  · Non-Profit Organization  · For-Profit Firm  · Academic Institution  · Charitable Foundation or Private Foundation  · Not specified |

# 

# Appendix 5: Machine learning and manual screening

All the studies obtained from manual and automatic search were imported to EPPI-Reviewer 4, a software facilitating management of references, identification and removal of duplicates, and screening of studies at both title and abstract and full-text stages. The title and abstract screening was done using the machine learning functionality available in EPPI, in order to make the process more efficient. The model built using this functionality learns from the set of manually screened studies to guess the inclusion criteria and reorganizes the list of studies based on the likelihood of being included. Using a training dataset consisting of 3300 manually screened studies, the algorithm ranked studies prioritizing them in order of likelihood of inclusion and excluding more than 1500 studies in a row. Based on this likelihood, the studies are grouped into ten groups (figure A4-1), with the set of excluded studies classified in the 0-10% range.

Figure A4-1: Order of relevance of studies at the title and abstract screening stage.



Source: Author’s own calculation.

This systematic map includes studies that are extremely diverse, and hence machine learning may not be accurate. To undertake a more nuanced approach and account for this diversity in studies, we conducted an additional manual screening within the groups. Since this manual screening gave us only 5 per cent includes from the set of studies within 10-40% range, those were not considered in the subsequent screening process. Once we had a set of studies that were included in the title and abstract screening stage, it was followed by full text screening. Here, we screened studies that had more than 50% likelihood of inclusion based on the machine learning model. A random check on twenty percent of studies belonging to 40-49% group provided very limited number of relevant studies that could be included, due to which the lower likelihood studies were not screened.

# Appendix 6: Systematic review appraisal tool and summary

**Summary of the systematic reviews**

**Utility and potential of rapid epidemic intelligence from internet-based sources. Yan and colleagues (2017)**

The study aimed to summarise internet-based methods that use freely accessible, unstructured data for epidemic surveillance and explore their timeliness and accuracy outcomes. The study is based on 84 articles published between 2006-2016 which relate to internet-based public health surveillance methods. These studies employ search queries, social media posts and approached derived from existing internet-based systems for early epidemic alerts and real-time monitoring. Primary methodology used for this review is the preferred reporting items for systematic review and meta analyses. In this method, the authors can assess the benefits and challenges of a health care intervention through an evidence-based minimum set of items. The study does not clearly demonstrate the inclusion criteria for the study designs making it difficult for the readers to interpret the findings.

**Ebola virus disease and social media: a systematic review. Fung and colleagues (2016)**

The study is a systematic review of the existing research pertinent to the Ebola virus disease and social media, especially to identify the research questions and methods used to collect and analyze social media. The study searched 6 databases for research articles relevant to Ebola virus disease and social media. Twelve articles were included in the main analysis: 7 from Twitter with 1 also including Weibo, 1 from Facebook, 3 from YouTube, and 1 from Instagram and Flickr. All the studies were cross-sectional. The study uses a standardized form to extract the data. A key challenge of the review is that the methods used by the review authors to analyze the findings of the included studies are not clearly defined.

**Satellite remote sensing in shark and ray ecology, conservation and management. Williamson and colleagues (2019)**

The study aims to present a summary of the existing state of knowledge on the application of Satellite Remote Sensing (SRS). SRS is seen as a key opportunity to analyse important environmental drivers in elasmobranch ecology and to aid management decisions for the conservation of declining population. The review included 71 papers and made use of several academic tools in order to conduct the bibliographic search. These included ISI Web of Science, Scopus, and Google Scholar. One of the limitations of the review is that it takes a wide variety of studies but does not combine the results of these studies explicitly.

**Concurrence of big data analytics and healthcare: A systematic review. Mehta and Pandit (2018)**

This systematic review of literature aims to determine the scope of Big data analytics in the field of healthcare including its applications and challenges in its adoption in healthcare. The review employs a systematic search of articles on five major scientific databases namely: Science Direct, PubMed, Emerald, IEEE Xplore and Taylor& Francis. Two reviewers independently extracted information on definition of big data analytics; sources and application s of big data. A total of 58 articles were selected. The authors did not explicitly state the methods used to analyze the quality of included studies and also do not report results for each of the study in the review.

**Data mining and machine learning to promote smart cities: a systematic review from 2000 to 2018. De Souza and colleagues (2019)**

This study aimed to present a systematic review regarding data mining (DM) and machine learning (ML) approaches adopted in the promotion of smart cities. The study seeks to provide, from a literature review in journals belonging to the Web of Science and Scopus databases, the different DM and ML techniques used, as well as to present the sectors most engaged in the promotion of smart cities. A total of 39 studies were included in the map which were further analyzed to assess the most commonly used DM and ML techniques to promote smart cities. While the report describes individual results of the studies clearly, it does not combine the results of all the studies.

**Using social media for actionable disease surveillance and outbreak management: a systematic literature review. Charles and Smith (2015)**

The studies in this review demonstrate how social media may be a valuable tool in disease surveillance toolkit used for improving public health professionals; ability to detect disease outbreaks faster than traditional methods and to enhance outbreak response. A social media application was defined for this review “an internet based application where people can communicate and share resources and information, and where users can activate and set their own profiles, have the ability to develop and update them constantly and have the opportunity to make such profiles totally or partially public and linked with other profiles in the network”. A total of 60 articles were selected for this systematic review which addressed the two research questions: can social media be integrated into disease surveillance and can it be used to improve health outcomes.

**Use of global positioning systems to study physical activity and the environment: a systematic review. Krenn and colleagues (2011)**

The aims of this systematic review was to determine the capability of GPS to collect hugh-quality data on the location of activities in research on the relationship between physical activity and the environment. Studies were eligible for inclusion if they were undertaken in humans, used GPS to measure the locations where PA occurred, and included analysis of the relationship between the characteristics of the environment and any form of PA behavior including leisure-time PA, sport or active travel. The capability of GPS was expressed in terms of data quality, which in turn was defined as the proportion of GPS data lost in each study. The authors do not mention the proper risk of bias assessment of the included study.

**Advances in using multitemporal night-time lights satellite imagery to detect and estimate and monitor socioeconomic dynamics. Bennett and colleagues (2016)**

This paper reviews progress in using multi-temporal Defense Meteorological Satellite Program-Operational Linescan System (DMSP-OLS), for which digital imagery was available from 1992 to 2013. It also reviews the Visible Infraref Imaging Radiometer (VIIRS) imagery to analyse urbanization, economic and population dynamics across range of geographic scales. The review provides an overview of data corrections and processing for comparison of multi-temporal nightlight imagery. The review includes a total of 17 studies. The results of all the studies were not combined and reported in the review properly. Risk of bias assessment of the included study has also not been reported in the review.

# Systematic review appraisal

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Questions | **Yan and colleagues (2017)** | **Fung and colleagues (2016)** | **Williamson and colleagues (2019)** | **Mehta and Pandit (2018)** | **De Souza and colleagues (2019)** | **Charles and Smith (2015)** | **Krenn and colleagues (2011)** | **Benett and Smith (2016)** |
| **Section A: Methods used to identify, include and critically appraise studies** | | | | | | | | |
| A.1 Were the criteria  used for deciding which studies to include in the review reported? | Yes | Yes | Yes | Partially | Yes | Yes | Yes | Yes |
| A.2 Was the search for evidence reasonably comprehensive? | No | No | Yes | No | Partially | No | Partially | No |
| A.3 Does the review cover an appropriate time period? | Not clear | Not clear | Not clear | Not clear | Not clear | Not clear | Not clear | Not clear |
| A.4 Was bias in the selection of articles avoided? | No | Partially | No | No | Partially | No | No | Not clear |
| A.5 Did the authors use appropriate criteria to assess the quality and risk of bias in analysing the studies that are included? | No | Not clear | Not clear | No | No | No | No | No |
| A.6 Overall, how much confidence do you have in the methods used to identify, include and critically appraise studies? | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) |
| **Section B Methods used to analyse the findings** | | | | | | | | |
| B.1 Were the characteristics and results of the included studies reliably reported? | Yes | No | No | Partially | No | No | No | Partially |
| B.2 Are the methods used by the review authors to analyse the findings of the included studies clear, including methods for calculating effect sizes if applicable? | No | Yes | No | No | Yes | No | No | No |
| B.3 Did the review describe the extent of heterogeneity? | No | No | No | No | Yes | Yes | No | No |
| B.4 Were the findings of the relevant studies combined (or not combined) appropriately relative to the primary question the review addresses and the available data? | Yes | Yes | No | No | Partially | No | No | Partially |
| B.5 Does the review report evidence appropriately? | No | No | No | Partially | No | No | Yes | Yes |
| B.6 Did the review examine the extent to which specific factors might explain differences in the results of the included studies? | No | No | No | No | No | No | No | No |
| B.7 Overall, how much confidence do you have in the methods used to analyse the findings relative to the primary question addressed in the review? | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) |
| **Section C: Overall assessment of the reliability of the review** | | | | | | | | |
| C.1 Are there any other aspects of the review not mentioned before that lead you to question the results? | No | No | No | No | No | No | No | No |
| C.2 Are there any mitigating factors that should be taken into account in determining the review’s reliability? | Not clear | Not clear | Not clear | Not clear | Not clear | Not clear | Not clear | Cant tell |
| C.3 Based on the above assessments of the methods, how would you rate the reliability of the review? | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) | Low (limitations are important enough that the results of the review are not reliable) |

# Appendix 7: Additional tables and figures

Table 4: Top 20 countries with maximum number of studies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Country** | **Total** | **Impact Evaluation** | **Measurement Study** | **Systematic reviews** |
| China | 43 | 6 | 37 |  |
| India | 34 | 4 | 30 |  |
| United States | 31 | 3 | 28 |  |
| Multi country | 31 | 3 | 25 | 3 |
| Unidentified | 14 | 1 | 11 | 2 |
| Kenya | 13 |  | 13 |  |
| Brazil | 11 | 3 | 8 |  |
| Mexico | 10 | 5 | 5 |  |
| Ethiopia | 9 | 1 | 8 |  |
| Bangladesh | 8 | 1 | 7 |  |
| Indonesia | 8 | 1 | 7 |  |
| Rwanda | 8 | 2 | 6 |  |
| Uganda | 8 | 2 | 6 |  |
| Japan | 7 |  | 7 |  |
| Nigeria | 7 |  | 7 |  |
| Tanzania | 7 | 1 | 6 |  |
| Iran | 6 |  | 6 |  |
| Sri Lanka | 6 |  | 6 |  |

Table 5: Distribution of studies across the region.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **East Asia and Pacific** | **Europe and Central Asia** | **Latin America and the Caribbean** | **Middle East and North Africa** | **North America** | **South Asia** | **Sub-Saharan Africa** |
| **Impact Evaluation** | 12 | 5 | 18 | 3 | 5 | 11 | 14 |
| **Measurement Study** | 92 | 69 | 42 | 29 | 49 | 82 | 112 |
| **Systematic reviews** | 7 | 6 | 5 | 6 | 6 | 6 | 6 |
| **Total** | 111 | 80 | 65 | 38 | 60 | 99 | 132 |

Table 7: Studies using multiple sources of big data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Combination of big data** | **IEs** | **Measurement Study** | **SRs** | **Grand Total** |
| Cell phone call records data (CDR) || Data from fixed sensors |  | 1 |  | 1 |
| Cell phone call records data (CDR) || Data from satellites |  | 10 |  | 10 |
| Citizen reporting and crowd sourced data || Data from fixed sensors || Data from satellites |  | 1 |  | 1 |
| Citizen reporting and crowd sourced data || Data from mobile sensors (tracking) |  | 1 |  | 1 |
| Citizen reporting and crowd sourced data || Data from satellites |  | 1 |  | 1 |
| Data from fixed sensors || Data from satellites | 2 | 13 |  | 15 |
| Data from mobile sensors (tracking) || Data from satellites |  | 5 |  | 5 |
| Data produced by public agencies || Data from fixed sensors || Data from satellites | 1 | 1 |  | 2 |
| Data produced by public agencies || Data from satellites | 1 | 3 |  | 4 |
| Data produced by public agencies || Data produced by businesses |  | 1 |  | 1 |
| Mobile data content || Cellphone call records data (CDR) |  | 2 |  | 2 |
| Mobile data content || Citizen reporting and crowd sourced data |  | 2 |  | 2 |
| Mobile data content || Data from mobile sensors (tracking) |  | 1 |  | 1 |
| Social networks || Citizen reporting and crowd sourced data |  | 2 |  | 2 |
| Social networks || Data from satellites |  | 1 |  | 1 |
| Social networks || Data produced by public agencies || Data from fixed sensors || Data from satellites |  | 1 |  | 1 |
| Social networks || Internet searches |  | 4 | 1 | 5 |
| Social networks || Internet searches || Mobile data content || Cell phone call records data (CDR) |  | 1 |  | 1 |
| Social networks || Mobile data content || Cell phone call records data (CDR) |  | 1 |  | 1 |
| **Grand Total** | **4** | **52** | **1** | **57** |

Table 9: Big data source wise studies in fragile context

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Chemical or radio-nuclear** | **Conflict or humanitarian crisis** | **Difficult terrain** | **Disease outbreaks or epidemics** | **Natural Disaster** | **Others** | **Total (row)** |
| **Cell phone call records data (CDR)** | 1 | 6 | 0 | 11 | 10 | 0 | 28 |
| **Citizen reporting and crowdsourced data** | 0 | 7 | 1 | 0 | 1 | 0 | 9 |
| **Data from fixed sensors** | 0 | 5 | 7 | 0 | 1 | 1 | 14 |
| **Data from mobile sensors (tracking)** | 0 | 0 | 2 | 0 | 0 | 0 | 2 |
| **Data from satellites** | 0 | 25 | 18 | 4 | 10 | 3 | 60 |
| **Data produced by businesses** | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Data produced by public agencies** | 0 | 0 | 0 | 1 | 3 | 0 | 4 |
| **Internet searches** | 0 | 0 | 0 | 1 | 0 | 1 | 2 |
| **Mobile data content** | 0 | 1 | 0 | 5 | 0 | 0 | 6 |
| **Social networks** | 0 | 2 | 0 | 2 | 0 | 1 | 5 |
| **Total (column)** | 1 | 46 | 28 | 24 | 25 | 6 |  |

Table 10: How studies have dealt with data quality and transparency

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Data collection methods** | **Data quality issues** | **Data representative of the population of interest** | **Construct validity explained** | **Results are generalizable** | **Data and code publicly available** | **Data, analysis and reporting challenges discussed** |
| **Impact Evaluation** | 47 | 22 | 29 | 43 | 26 | 11 | 6 |
| **Measurement Study** | 43 | 3 | 7 | 21 | 4 | 8 | 11 |
| **Systematic reviews** | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| **Grand Total** | 91 | 25 | 36 | 64 | 30 | 19 | 18 |

# Appendix 8: The systematic map of big data sources and outcomes



The online map can be accessed here: <https://gapmaps.3ieimpact.org/evidence-maps/big-data-systematic-map>

# Appendix 9: List of included studies

**Impact evaluations**

1. Alix-Garcia Jennifer, Aronson Glen, Radeloff Volker, Ramirez-Reyes Carlos, Shapiro Elizabeth, Sims Katharine, Yanez-Pagans Patricia, 2015. Impacts of payments for ecosystem services programme in Mexico
2. Alix-Garcia Jennifer Marie, Shapiro Elizabeth N, Sims Katharine R.E, 2012. Forest Conservation and Slippage: Evidence from Mexico’s National Payments for Ecosystem Services Program, Staff Paper Series.
3. Allcott Hunt, 2011. Social norms and energy conservation. Journal of Public Economics 95, 1082–1095. https://doi.org/10.1016/j.jpubeco.2011.03.003
4. Allcott Hunt, Rogers Todd, 2014. The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. American Economic Review 104, 3003–37. https://doi.org/10.1257/aer.104.10.3003
5. Ayres Ian, Raseman Sophie, Shih Alice, 2012. Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage. The Journal of Law, Economics, and Organization 29, 992–1022. https://doi.org/10.1093/jleo/ews020
6. BenYishay Ariel, Glenn Carey, Trichler Rachel, Runfola Dan, Goodman Seth, 2018a. Evaluation of the On-Farm Water Management Program.
7. BenYishay Ariel, Parks Bradley, Trichler Rachel, Runfola Dan, 2016a. Forest Cover Impacts of Chinese Development Projects in Ecologically Sensitive Areas.
8. BenYishay Ariel, Trichler Rachel, Runfola Dan, Goodman Seth, 2018b. Evaluation of the Infrastructure Needs Program II.
9. BenYishay Ariel, Trichler Rachel, Runfola Dan, Heuser Silke, 2016b. Indigenous Land Rights and Deforestation: Evidence from the Brazilian Amazon.
10. Blackman Allen, 2015. Strict versus mixed-use protected areas: Guatemala’s Maya Biosphere Reserve. Ecological Economics 112, 14–24. https://doi.org/10.1016/j.ecolecon.2015.01.009
11. Blackman Allen, Pfaff Alexander, Robalino Juan, 2015. Paper park performance: Mexico’s natural protected areas in the 1990s. Global Environmental Change 31, 50–61. https://doi.org/10.1016/j.gloenvcha.2014.12.004
12. Blackman Allen, Villalobos Laura, 2019. Clear, but don’t invest: protected areas discourage some land uses more than others. Environmental Research Letters 14. https://doi.org/10.1088/1748-9326/ab3ca1
13. Blumenstock Joshua, Callen Michael, Ghani Tarek, Koepke Lucas, 2015. Promises and Pitfalls of Mobile Money in Afghanistan: Evidence from a Randomized Control Trial. In Proceedings of the Seventh International Conference on Information and Communication Technologies and Development, edited by Arul Chib, Matthew Kam, and Jenna Burrell.
14. Blumenstock Joshua, Eagle Nathan, Fafchamps Marcel, 2011. Risk and Reciprocity Over the Mobile Phone Network: Evidence from Rwanda, CSAE Working Paper Series. Centre for the Study of African Economies, University of Oxford.
15. Bonnier Evelina, Poulsen Jonas, Rogall Thorsten, Stryjan Miri, 2016. Preparing for Genocide: Quasi-Experimental Evidence from Rwanda, SITE Working Paper Series. Stockholm School of Economics, Stockholm Institute of Transition Economics.
16. Buchanan Graeme M, Parks Bradley C, Donald Paul F, O’Donell Brian F, Runfola Daniel, Swaddle John P, Tracewski Lukasz, Butchart Stuart H.M, 2015. The Impacts of World Bank Development Projects on Sites of High Biodiversity Importance.
17. Buchanan Graeme M, Parks Bradley C, Donald Paul F, O’Donnell Brian F, Runfola Daniel, Swaddle John P, Tracewski Łukasz, Butchart Stuart H. M, 2018. The Local Impacts of World Bank Development Projects Near Sites of Conservation Significance. The Journal of Environment & Development 27, 299–322. https://doi.org/10.1177/1070496518785943
18. Buntaine Mark T, Hamilton Stuart E, Millones Marco, 2015. Titling community land to prevent deforestation: An evaluation of a best-case program in Morona-Santiago, Ecuador. Global Environmental Change 33, 32–43. https://doi.org/10.1016/j.gloenvcha.2015.04.001
19. Bunte Jonas B, Desai Harsh, Gbala Kanio, Parks Brad, Runfola Danel Miller, 2017. Natural Resources Sector FDI and Growth in Post Conflict Settings: Subnational Evidence from Liberia.
20. Burwen Jason, Levine David I, 2012. A Rapid Assessment Randomized-Controlled Trial of Improved Cook Stoves in the Tumu Region of Ghana.
21. Carbi et al, 2014. Federal transfer multipliers. Quasi-experimental evidence. NBER Working Paper No. 20751.
22. Chen Yuyu, Ebenstein Avraham, Greenstone Michael, Li Hongbin, 2013. Evidence on the impact of sustained exposure to air pollution on life expectancy from China’s Huai River policy. Proceedings of the National Academy of Sciences 110, 12936. https://doi.org/10.1073/pnas.1300018110
23. Daniel Ayalew Ali Klaus Deininger Daniel Monchuk, 2018. Using Satellite Imagery to Assess Impacts of Soil and Water Conservation Measures: Evidence from Ethiopia’s Tana-Beles Watershed, Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-8321
24. Gabriel Lara Ibarra David McKenzie Claudia Ruiz Ortega, 2017. Learning the Impact of Financial Education When Take-Up Is Low, Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-8238
25. Gaveau David L. A, Epting Justin, Lyne Owen, Linkie Matthew, Kumara Indra, Kanninen Markku, Leader-Williams Nigel, 2009. Evaluating whether protected areas reduce tropical deforestation in Sumatra. Journal of Biogeography 36, 2165–2175. https://doi.org/10.1111/j.1365-2699.2009.02147.x
26. Jain Meha, Balwinder-Singh, Rao Preeti, Srivastava Amit K, Poonia Shishpal, Blesh Jennifer, Azzari George, McDonald Andrew J, Lobell David B, 2019. The impact of agricultural interventions can be doubled by using satellite data. Nature Sustainability 2, 931–934. https://doi.org/10.1038/s41893-019-0396-x
27. Jaiswal Sreeja, Bensch Gunther, Navalkar Aniket, Jayaraman T, Murari Kamal, Patnaik Unmesh, 2020. Evaluating the impact of infrastructure development: case study of the Konkan Railway in India.
28. Jayachandran Seema, de Laat, Joost, Lambin Eric F, Stanton Charlotte Y, 2016. Cash for Carbon: A Randomized Controlled Trial of Payments for Ecosystem Services to Reduce Deforestation. National Bureau of Economic Research Working Paper Series No. 22378. https://doi.org/10.3386/w22378
29. Jianing Zhao, Runfola Daniel, Peter Kemper, 2017. Quantifying Heterogeneous Causal Treatment Effects in World Bank Development Finance Projects.
30. Li Shanjun, Liu Yanyan, 2020. Using big data to evaluate the impacts of transportation infrastructure investment: the case of subway systems in Beijing.
31. Li Shanjun, Liu Yanyan, Purevjav Avralt-Od, Yang Lin, 2019. Does subway expansion improve air quality? Journal of Environmental Economics and Management 96, 213–235. https://doi.org/10.1016/j.jeem.2019.05.005
32. Liana O Anderson, Samantha De Martino, Torfinn Harding, Karlygash Kuralbayeva, Andre Lima, 2016. The Effects of Land Use Regulation on Deforestation: Evidence from the Brazilian Amazon, OxCarre Working Papers. Oxford Centre for the Analysis of Resource Rich Economies, University of Oxford.
33. Long et al, 2018. High-speed rail and urban expansion: An empirical study using a time series. Journal of Transport Geography 106–118.
34. Marty Robert, Goodman Seth, LeFew Michael, Dolan Carrie, BenYishay Ariel, Runfola Daniel, 2019. Assessing the causal impact of Chinese aid on vegetative land cover in Burundi and Rwanda under conditions of spatial imprecision. Development Engineering 4, 100038. https://doi.org/10.1016/j.deveng.2018.11.001
35. Muneeza Alam Matias Herrera Dappe Martin Melecky Ran Goldblatt, 2019. Wider Economic Benefits of Transport Corridors: Evidence from International Development Organizations, Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-9057
36. Pande Rohini, Sudarshan Anant, 2019. Harnessing transparency initiatives to improve India’s environmental clearance process for the mineral mining sector.
37. Parks Brad, Baehr Christian, Aboagye Daniel, Trichler Rachel, BenYishay Ariel, Runfola Daniel, Puram Punwath, 2019. Building on a Foundation Stone: The Long-Term Impacts of a Local Infrastructure and Governance Program in Cambodia.
38. Pellegrini Lorenzo, 2019. Impacts of community monitoring of socio-environmental liabilities in the Ecuadorian and Peruvian Amazon.
39. Persha Lauren, Meshack Charles, 2016. A triple win? The impact of Tanzania’s Joint Forest Management programme on livelihoods, governance and forests.
40. Sam Asher Teevrat Garg Paul Novosad, 2018. The Ecological Impact of Transportation Infrastructure, Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-8507
41. Scullion Jason, Thomas Craig W., Vogt Kristina A., Perez-Maqueo Octavio, Logsdon Miles G, 2011. Evaluating the environmental impact of payments for ecosystem services in Coatepec (Mexico) using remote sensing and on-site interviews. Environmental Conservation 38, 426–434. https://doi.org/10.1017/S037689291100052X
42. Tanner J, Goodman Seth, Leu Matthias, Trichler Rachel, BenYishay Ariel, Runfola Daniel, Marty Robert, Nagol J, 2017. A Top-Down Approach to Estimating Spatially Heterogeneous Impacts of Development Aid on Vegetative Carbon Sequestration.
43. Van der Windt, Peter, Humphreys Macartan, 2016. Crowdseeding in Eastern Congo: Using Cell Phones to Collect Conflict Events Data in Real Time. The Journal of Conflict Resolution 60, 748–781.
44. Velilla Raul, Braganca Alexandre, 2019. Coffee Price Shock and Local Economic Growth: Evidence from Colombia.
45. Villa Juan M, 2016. Social Transfers and Growth: Evidence from Luminosity Data. Economic Development and Cultural Change 65, 39–61. https://doi.org/10.1086/687548
46. Webster Jayne, Landegger Justine, Bruce Jane, Malunda Jackson, Chantler Tracey, Kumakech Edward, Schmucker laura, Kiapi Lilian, Kozuki Naoko, Olorunsaiye Comfort, Bryne Erin, 2019. Impacts of IRC’s Fifth Child community engagement strategy to increase immunisation in northern Uganda.
47. Yang Jun, Chen Shuai, Qin Ping, Lu Fangwen, Liu Antung A, 2018. The effect of subway expansions on vehicle congestion: Evidence from Beijing. Journal of Environmental Economics and Management 88, 114–133. https://doi.org/10.1016/j.jeem.2017.09.007
48. Yang Jun, Purevjav Avralt-Od, Li Shanjun, 2020. The Marginal Cost of Traffic Congestion and Road Pricing: Evidence from a Natural Experiment in Beijing. American Economic Journal: Economic Policy 12, 418–53. <https://doi.org/10.1257/pol.20170195>

**Systematic reviews**

1. Bennett Mia M, Smith Laurence C, 2017. Advances in using multitemporal night-time lights satellite imagery to detect, estimate, and monitor socioeconomic dynamics. Remote Sensing of Environment 192, 176–197. https://doi.org/10.1016/j.rse.2017.01.005
2. Charles-Smith Lauren E, Reynolds Tera L, Cameron Mark A, Conway Mike, Lau Eric H Y, Olsen Jennifer M, Pavlin Julie A, Shigematsu Mika, Streichert Laura C, Suda Katie J, Corley Courtney D, 2015. Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review. PloS one 10, e0139701–e0139701. https://doi.org/10.1371/journal.pone.0139701
3. de Souza, J T, de Francisco, A C, Piekarski C M, do Prado, G F, 2019. Data Mining and Machine Learning to Promote Smart Cities: A Systematic Review from 2000 to 2018. Sustainability 11. https://doi.org/10.3390/su11041077
4. Fung ChunHai, Duke C H, Finch K C, Snook K R, Tseng PeiLing, Hernandez A C, Manoj Gambhir, Fu KingWa, Tse TszHo, 2016. Ebola virus disease and social media: a systematic review. AJIC - American Journal of Infection Control 44, 1660–1671.
5. Krenn P J, Titze S, Oja P, Jones A, Ogilvie D, 2011. Use of global positioning systems to study physical activity and the environment: a systematic review. American Journal of Preventive Medicine 41, 508–515.
6. Mehta Nishita, Pandit Anil, 2018. Concurrence of big data analytics and healthcare: A systematic review. International Journal of Medical Informatics 114, 57–65. https://doi.org/10.1016/j.ijmedinf.2018.03.013
7. Williamson M J, Tebbs E J, Dawson T P, Jacoby D M. P, 2019. Satellite Remote Sensing in Shark and Ray Ecology, Conservation and Management. Frontiers in Marine Science 6. https://doi.org/10.3389/fmars.2019.00135
8. Yan S J, Chughtai A A, Macintyre C R, 2017. Utility and potential of rapid epidemic intelligence from Internet-based sources. International Journal of Infectious Diseases 63, 77–87.

**Measurement studies**

1. A Study on Urban Mobility and Dynamic Population Estimation by Using Aggregate Mobile Phone Sources, 2010.
2. AbdelRahman Mohamed A. E, Natarajan A, Hegde Rajendra, Prakash S S, 2018. Assessment of land degradation using comprehensive geostatistical approach and remote sensing data in GIS-model builder. The Egyptian Journal of Remote Sensing and Space Sciences. https://doi.org/10.1016/j.ejrs.2018.03.002
3. Abt Associates, 2017. Haryana Health GIS:Leveraging Technology to strengthen evidence based decision making in public health.
4. Acosta Mireles, M, Pérez Miranda, R, Romero Sánchez, M E, González Hernández, A, Martínez Ángel, L, 2017. Landsat ETM+ imaging for the estimation of the forest density in the southern region of the State of Mexico. Revista Mexicana de Ciencias Forestales 8, 30–55.
5. Adeline Decuyper, Alex Rutherford, Amit Wadhwa, Jean-Martin Bauer, Gautier Krings, Thoralf Gutierrez, Vincent D Blondel, Miguel A Luengo-Oroz, 2014. Estimating Food Consumption and Poverty Indices with Mobile Phone Data.
6. Adhikary Partha Pratim, Barman D, Madhu M, Dash Ch J, Jakhar P, Hombegowda H C, Naik B S, Sahoo D C, Beer Karma, 2019. Land use and land cover dynamics with special emphasis on shifting cultivation in Eastern Ghats Highlands of India using remote sensing data and GIS. Environmental Monitoring & Assessment 191, 1–15. https://doi.org/10.1007/s10661-019-7447-7
7. Aditya Aladangady, Shifrah Aron-Dine, Wendy E Dunn, Laura Feiveson, Paul Lengermann, Claudia R Sahm, 2019. From Transactions Data to Economic Statistics: Constructing Real-time, High-frequency, Geographic Measures of Consumer Spending.
8. Agidew A A, Singh K N, 2017. The implications of land use and land cover changes for rural household food insecurity in the northeastern highlands of Ethiopia: the case of the Teleyayen sub-watershed. Agriculture and Food Security 6, (31 October 2017)-(31 October 2017).
9. Aguda A S, Farinde T A, Adegboyega S A, Olawole M O, 2013. Spatio-temporal assessment of urban growth of medium-size and nodal towns for sustainable management: using GIS. Management of Environmental Quality: An International Journal 24, 94–106. https://doi.org/10.1108/14777831311291159
10. Ahas Rein, Saluveer Erki, Tiru Margus, Silm Siiri, 2008. Mobile Positioning Based Tourism Monitoring System: Positium Barometer, in: O’Connor Peter, Höpken Wolfram, Gretzel Ulrike (Eds.), . Presented at the Information and Communication Technologies in Tourism 2008, Springer Vienna, Vienna, pp. 475–485.
11. Aiello A, Adamo M, Canora F, 2015. Remote sensing and GIS to assess soil erosion with RUSLE3D and USPED at river basin scale in southern Italy. Catena 131, 174–185.
12. Ajin R S, Loghin A M, Vinod P G, Menon A R. R, Jacob M K, 2018. Forest fire risk assessment using geospatial techniques: a study in Mannarkkad forest division of Palakkad District, Kerala, India. Ecoterra 15, 1–9.
13. Aklibasinda Meliha, 2019. Determining the active green areas and their adequacy by using satellite images and gis: the case of Nevsehir city (Turkey). Fresenius Environmental Bulletin 28, 7274–7281.
14. Al-Bahrani H S, 2014. Spatial prediction and classification of water quality parameters for irrigation use in the Euphrates River (Iraq) using GIS and satellite image analyses. International Journal of Sustainable Development and Planning 9, 389–399.
15. Albuquerque Miguel, Espinoza Jean, Teixeira Priscila, de Oliveira, Alan, Corrêa Iran, Calliari Lauro, 2013. Erosion or Coastal Variability : An Evaluation of the DSAS and the Change Polygon Methods for the Determination of Erosive Processes on Sandy Beaches. Journal of Coastal Research 1710.
16. Alegana V A, Atkinson P M, Pezzulo C, Sorichetta A, Weiss D, Bird T, Erbach-Schoenberg E, Tatem A J, 2015. Fine resolution mapping of population age-structures for health and development applications. Journal of the Royal Society, Interface 12, 20150073. https://doi.org/10.1098/rsif.2015.0073
17. Allevato Emilia, Saulino Luigi, Cesarano Gaspare, Chirico Giovanni Battista, D’Urso Guido, Falanga Bolognesi, Salvatore, Rita Angelo, Rossi Sergio, Saracino Antonio, Bonanomi Giuliano, 2019. Canopy damage by spring frost in European beech along the Apennines: effect of latitude, altitude and aspect. Remote Sensing of Environment 225, 431–440. https://doi.org/10.1016/j.rse.2019.03.023
18. Alshaikh Amal Yahya, 2015. Space applications for drought assessment in Wadi-Dama (West Tabouk), KSA. The Egyptian Journal of Remote Sensing and Space Sciences 18, S43–S53. https://doi.org/10.1016/j.ejrs.2015.07.001
19. Althouse Benjamin M, Ng Yih Yng, Cummings Derek A. T, 2011. Prediction of Dengue Incidence Using Search Query Surveillance. PLOS Neglected Tropical Diseases 5, e1258.
20. Alwadi Mohammad, Chetty Girija, 2015. Energy Efficient Data Mining Scheme for High Dimensional Data. Procedia Computer Science 46, 483–490. https://doi.org/10.1016/j.procs.2015.02.047
21. Amy Wesolowski, Wendy Prudhomme O’Meara, y Tatem, Samson Ndege, Nathan Eagle, Caroline Buckee, 2015. Quantifying the Impact of Accessibility on Preventive Healthcare in Sub-Saharan Africa Using Mobile Phone Data.
22. Andrés Viña, Xiaodong Chen, William J McConnell, Wei Liu, Weihua Xu, Zhiyun Ouyang, Hemin Zhang, Jianguo Liu, 2011. Effects of Natural Disasters on Conservation Policies: The Case of the 2008 Wenchuan Earthquake, China. Ambio 40, 274.
23. Armenteras Dolors, Gibbes Cerian, Anaya Jesús A, Dávalos Liliana M, 2017. Integrating remotely sensed fires for predicting deforestation for REDD+. Ecological Applications 27, 1294–1304. https://doi.org/10.1002/eap.1522
24. Ashraf S, Afshari H, Ebadi A G, 2011. Application of GIS for determination of groundwater quality suitable in crops influenced by irrigation water in the Damghan region of Iran. International Journal of Physical Sciences 6, 843–854.
25. Babaie-Kafaky S, Mataji A, Sani N A, 2009. Ecological capability assessment for multiple-use in forest areas using GIS-based multiple criteria decision making approach. American Journal of Environmental Sciences 5, 714–721.
26. Badarinath K V. S, Chand T R. Kiran, Prasad V Krishna, 2009. Emissions from grassland burning in Kaziranga National Park, India - analysis from IRS-P6 AWiFS satellite remote sensing datasets. Geocarto International 24, 89–97. https://doi.org/10.1080/10106040701207225
27. Badarinath K V. S, Chand T R. Kiran, Prasad V Krishna, 2006. Agriculture crop residue burning in the Indo-Gangetic Plains – A study using IRS-P6 AWiFS satellite data. Current Science 91, 1085.
28. Bahuguna Anjali, Nayak Shailesh, Roy Dam, 2008. Impact of the tsunami and earthquake of 26th December 2004 on the vital coastal ecosystems of the Andaman and Nicobar Islands assessed using RESOURCESAT AWiFS data. International Journal of Applied Earth Observations and Geoinformation 10, 229–237. https://doi.org/10.1016/j.jag.2008.02.010
29. Baltrusaitis K, Brownstein J S, Scarpino S V, Bakota E, Crawley A W, Conidi G, Gunn J, Gray J, Zink A, Santillana M, 2018. Comparison of crowd-sourced, electronic health records based, and traditional health-care based influenza-tracking systems at multiple spatial resolutions in the United States of America. BMC Infectious Diseases 18, (15 August 2018)-(15 August 2018).
30. Banerjee R, Srivastava P K, 2013. Reconstruction of contested landscape: Detecting land cover transformation hosting cultural heritage sites from Central India using remote sensing. Land Use Policy 34, 193–203. https://doi.org/10.1016/j.landusepol.2013.03.005
31. Barbieri T, Despini F, Teggi S, 2018. A multi-temporal analyses of land surface temperature using Landsat-8 data and open source software: the case study of Modena, Italy. Sustainability 10, 1678–1678.
32. Barros N, Fontes T, Silva M P, Manso M C, 2013. How wide should be the adjacent area to an urban motorway to prevent potential health impacts from traffic emissions? Transportation Research Part a-Policy and Practice 50, 113–128. https://doi.org/10.1016/j.tra.2013.01.021
33. Bartlett J, Krasodomski-Jones A, Daniel N, Fisher A, Jesperson S, 2015. Social Media for Election Communication and Monitoring in Nigeria. Demos for Department for International Development.
34. Baus Peter, Kováč Urban, Pauditšová Eva, Kohutková Ivana, Komorník Jozef, 2014. Identification of interconnections between landscape pattern and urban dynamics—Case study Bratislava, Slovakia. Ecological Indicators 42, 104–111. https://doi.org/10.1016/j.ecolind.2013.12.011
35. Bekele B, Wu W, Legesse A, Temesgen H, Yirsaw E, 2018. Socio-environmental impacts of land use/cover change in Ethiopian Central Rift Valley Lakes Region, East Africa. Applied Ecology and Environmental Research 16, 6607–6632.
36. Ben Klemens Andrea Coppola Max Shron, 2015. Estimating Local Poverty Measures Using Satellite Images: A Pilot Application to Central America, Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-7329
37. Bengtsson L, Gaudart J, Lu Xin, Moore S, Wetter E, Sallah K, Rebaudet S, Piarroux R, 2015. Using mobile phone data to predict the spatial spread of cholera. Scientific Reports 5, sre-sre.
38. Bengtsson Linus, Lu Xin, Thorson Anna, Garfield Richard, von Schreeb, Johan, 2011. Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti. PLOS Medicine 8, e1001083.
39. Bharti Nita, Djibo Ali, Tatem Andrew J, Grenfell Bryan T, Ferrari Matthew J, 2016. Measuring populations to improve vaccination coverage. Scientific Reports 6, 34541. https://doi.org/10.1038/srep34541
40. Bharti Nita, Lu Xin, Bengtsson Linus, Wetter Erik, Tatem Andrew J, 2015. Remotely measuring populations during a crisis by overlaying two data sources. International Health 7, 90–98. https://doi.org/10.1093/inthealth/ihv003
41. Bick I A, Bardhan R, Beaubois T, 2018. Applying fuzzy logic to open data for sustainable development decision-making: a case study of the planned city Amaravati. Natural Hazards 91, 1317–1339. https://doi.org/10.1007/s11069-018-3186-2
42. Blumenstock Joshua, 2018. Estimating Economic Characteristics with Phone Data. blumen phonedata. https://doi.org/10.1257/pandp.20181033
43. Blumenstock Joshua, 2014. Calling for Better Measurement: Estimating an Individual’s Wealth and Well-Being.
44. Blumenstock Joshua, Cadamuro Gabriel, On Robert, 2015. Predicting poverty and wealth from mobile phone metadata. Science 350, 1073. https://doi.org/10.1126/science.aac4420
45. Blumenstock Joshua, Chokkalingam Ramkumar, Gaikwad Vijay, 2014. Probabilistic Inference of Unknown Locations: Exploiting Collective Behavior when Individual Data is Scarce.
46. Blumenstock Joshua E, 2012. Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda. Information Technology for Development 18, 107–125. https://doi.org/10.1080/02681102.2011.643209
47. Blumenstock Joshua, Eagle Nathan, 2010. Mobile Divides: Gender, Socioeconomic Status, and Mobile Phone Use in Rwanda.
48. Blumenstock Joshua, Keleher Niall, 2015. The Price is Right? Statistical evaluation of a crowd-sourced market information system in Liberia.
49. Blumenstock Joshua, Maldeniya Danaja, Lokanathan Sriganesh, 2017. Understanding the Impact of Urban Infrastructure: New Insights from Population-Scale Data.
50. Blumenstock Joshua, Shen Ye, Eagle Nathan, 2010. A Method for Estimating the Relationship Between Phone Use and Wealth.
51. Bogomolov Andrey, Lepri Bruno, Staiano Jacopo, Letouzé Emmanuel, Oliver Nuria, Pianesi Fabio, Pentland Alex, 2015. Moves on the Street: Classifying Crime Hotspots Using Aggregated Anonymized Data on People Dynamics. Big Data 3, 148–158. https://doi.org/10.1089/big.2014.0054
52. Bosco Claudio, Tejedor-Garavito Natalia, de Rigo, Daniele, Pezzulo Carla, Bengtsson Linus, Tatem Andrew, Bird Tomas, 2017. Mapping the interaction between development aid and stunting in Nigeria.
53. Bosco Claudio, Tejedor-Garavito Natalia, de Rigo, Daniele, Tatem Andrew J, Pezzulo Carla, Wood Richard, Chamberlain Heather, Bird Tom, 2018. Geostatistical Tools to Map the Interaction between Development Aid and Indices of Need.
54. Botha R, Labuschagne C, Williams A G, Bosman G, Brunke E G, Rossouw A, Lindsay R, 2018. Characterising fifteen years of continuous atmospheric radon activity observations at Cape Point (South Africa). Atmospheric Environment 176, 30–39. https://doi.org/10.1016/j.atmosenv.2017.12.010
55. Brdar Sanja, Gavrić Katarina, Ćulibrk Dubravko, Crnojević Vladimir, 2016. Unveiling Spatial Epidemiology of HIV with Mobile Phone Data. Scientific Reports 6, 19342. https://doi.org/10.1038/srep19342
56. Brewin R J. W, Mora L de, Jackson T, Brewin T G, Shutler J, 2015. On the potential of surfers to monitor environmental indicators in the coastal zone. PLoS ONE 10, e0127706–e0127706.
57. Burke Marshall, Lobell David B, 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. Proceedings of the National Academy of Sciences 114, 2189. https://doi.org/10.1073/pnas.1616919114
58. C Bosco, V Alegana, T Bird, C Pezzulo, L Bengtsson, A Sorichetta, J Steele, G Hornby, C Ruktanonchai, N Ruktanonchai, E Wetter, A J Tatem, 2017. Exploring the high-resolution mapping of gender-disaggregated development indicators.
59. Cabral Ana I. R, Saito Carlos, Pereira Henrique, Laques Anne Elisabeth, 2018. Deforestation pattern dynamics in protected areas of the Brazilian Legal Amazon using remote sensing data. Applied Geography 100, 101–115. https://doi.org/10.1016/j.apgeog.2018.10.003
60. Cahoon D R, Olker J H, Yeates A G, Guntenspergen G R, Grace J B, Adamowicz S C, Anisfeld S C, Baldwin A H, Barrett N, Beckett L, Benzecry A, Blum L K, Burdick D M, Crouch W, Ekberg M C, Fernald S, Grimes K W, Grzyb J, Hartig E K, Kreeger D A, Larson M, Lerberg S, Lynch J C, Maher N, Maxwell-Doyle M, Mitchell L R, 2019. Hurricane Sandy impacts on coastal wetland resilience. Open-File Report - US Geological Survey.
61. Cai Yaping, Guan Kaiyu, Lobell David, Potgieter Andries B, Wang Shaowen, Peng Jian, Xu Tianfang, Asseng Senthold, Zhang Yongguang, You Liangzhi, Peng Bin, 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agricultural and Forest Meteorology 274, 144–159. https://doi.org/10.1016/j.agrformet.2019.03.010
62. Candelieri Antonio, Soldi Davide, Archetti Francesco, 2015. Short-term forecasting of hourly water consumption by using automatic metering readers data. Procedia Engineering 119, 844–853. https://doi.org/10.1016/j.proeng.2015.08.948
63. Carreiras J M. B, Jones J, Lucas R M, Gabriel C, 2014. Land use and land cover change dynamics across the Brazilian Amazon: insights from extensive time-series analysis of remote sensing data. PLoS ONE 9, e104144–e104144.
64. Caughlin T Trevor, Ruktanonchai Nick, Acevedo Miguel A, Lopiano Kenneth K, Prosper Olivia, Eagle Nathan, Tatem Andrew J, 2013. Place-Based Attributes Predict Community Membership in a Mobile Phone Communication Network. PLOS ONE 8, e56057.
65. Cavallo Alberto, 2013. Online and official price indexes: Measuring Argentina’s inflation. Journal of Monetary Economics 60, 152–165. https://doi.org/10.1016/j.jmoneco.2012.10
66. Chae SangWon, Kwon SungJun, Lee DongHyun, 2018. Predicting infectious disease using deep learning and big data. International Journal of Environmental Research and Public Health 15, 1596–1596.
67. Chakraborty A, Seshasai M V. R, Reddy C S, Dadhwal V K, 2018. Persistent negative changes in seasonal greenness over different forest types of India using MODIS time series NDVI data (2001-2014). Ecological Indicators 85, 887–903. https://doi.org/10.1016/j.ecolind.2017.11.032
68. Chen Xi, Nordhaus William D, 2011. Using luminosity data as a proxy for economic statistics. Proceedings of the National Academy of Sciences 108, 8589. https://doi.org/10.1073/pnas.1017031108
69. Chen Yuyu, Jin Ginger Zhe, Kumar Naresh, Shi Guang, 2013. The promise of Beijing: Evaluating the impact of the 2008 Olympic Games on air quality. Journal of Environmental Economics and Management 66, 424–443. https://doi.org/10.1016/j.jeem.2013.06.005
70. Choi Hyunyoung, Varian Hal, 2012. Predicting the Present with Google Trends. The Economic Record 88, 2–9.
71. Chong Wai K, Naganathan Hariharan, Liu Huan, Ariaratnam Samuel, Kim Joonhoon, 2018. Understanding Infrastructure Resiliency in Chennai, India Using Twitter’s Geotags and Texts: A Preliminary Study. Engineering 4, 218–223. https://doi.org/10.1016/j.eng.2018.03.010
72. Civelli Andrea, Horowitz Andrew, Teixeira Arilton, 2017. Foreign Aid and Growth at the Subnational level.
73. Clifford E, Mulligan S, Comer J, Hannon L, 2018. Flow-signature analysis of water consumption in nonresidential building water networks using high-resolution and medium-resolution smart meter data: two case studies. Water Resources Research 54, 88–106.
74. Convergne Elodie, Snyder Michael R, 2015. Making Maps to Make Peace: Geospatial Technology as a Tool for UN Peacekeeping. International Peacekeeping 22, 565–586. https://doi.org/10.1080/13533312.2015.1094193
75. Coskun H G, Alganci U, Usta G, Celik H, 2009. Analysis of the provincial structure of Sariyer/Istanbul using Remote Sensing and GIS. Remote sensing for a changing Europe. Proceedings of the 28th Symposium of the European Association of Remote Sensing Laboratories, Istanbul, Turkey, 2-5 June 2008.
76. D Pastor-Escuredo, A Morales-Guzmán, Y Torres-Fernández, J Bauer, A Wadhwa, C Castro-Correa, L Romanoff, Jong Gun Lee, A Rutherford, V Frias-Martinez, N Oliver, E Frias-Martinez, M Luengo-Oroz, 2014. Flooding through the lens of mobile phone activity. Presented at the IEEE Global Humanitarian Technology Conference (GHTC 2014), pp. 279–286. https://doi.org/10.1109/GHTC.2014.6970293
77. Dadhich P N, Hanaoka S, 2011. Spatio-temporal Urban Growth Modeling of Jaipur, India. Journal of Urban Technology 18, 45–65. https://doi.org/10.1080/10630732.2011.615567
78. Dalezios N R, Mplanta A, Domenikiotis C, 2011. Remotely sensed cotton evapotranspiration for irrigation water management in vulnerable agriculture of central Greece. Journal of Information Technology in Agriculture 4, un-un.
79. Damanik-Ambarita Minar Naomi, Boets Pieter, Nguyen Thi, Hanh Tien, Forio Marie Anne Eurie, Everaert Gert, Lock Koen, Musonge Peace Liz Sasha, Suhareva Natalija, Bennetsen Elina, Gobeyn Sacha, Ho Tuan Long, Dominguez-Granda Luis, Goethals Peter L. M, 2018. Impact assessment of local land use on ecological water quality of the Guayas river basin (Ecuador). Ecological Informatics 48, 226–237. https://doi.org/10.1016/j.ecoinf.2018.08.009
80. David B Lobell George Azzari Marshall Burke Sydney Gourlay, 2018. Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis in Uganda, Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-8374
81. Davies Kevin P, Murphy Richard J, Bruce Eleanor, 2016. Detecting historical changes to vegetation in a Cambodian protected area using the Landsat TM and ETM+ sensors. Remote Sensing of Environment 187, 332–344. https://doi.org/10.1016/j.rse.2016.10.027
82. de Jong, Ben, Anaya Carlos, Masera Omar, Olguín Marcela, Paz Fernando, Etchevers Jorge, Martínez René D, Guerrero Gabriela, Balbontín Claudio, 2010. Greenhouse gas emissions between 1993 and 2002 from land-use change and forestry in Mexico. Forest Ecology & Management 260, 1689–1701. https://doi.org/10.1016/j.foreco.2010.08.011
83. Dehkordi L F, Sohrabi T A, Ghanavizbaf M H, Ghazavi R, 2016. Drought monitoring by using of MODIS satellite images in dry lands (case study: Isfahan rangelands). Geography and Environmental Planning 27, 177–190.
84. De-min Zhou, Radke John, Yong-qian Tian, Jian-chun Xu, Lan Mu, 2005. A model supported by GIS for locating and quantifying PM\_{2.5} emission originated from residential wood burning. Journal of Environmental Sciences (IOS Press) 17, 861–865.
85. Demisse Getachew, Tadesse Tsegaye, Atnafu Solomon, Hill Shawndra, 2011. Drought Monitoring in Food-Insecure Areas of Ethiopia by Using Satellite Technologies. pp. 183–200. https://doi.org/10.1007/978-3-642-22315-0\_11
86. Deng XiangZheng, Gibson J, Wang Pei, 2017. Relationship between landscape diversity and crop production: a case study in the Hebei Province of China based on multi-source data integration. Journal of Cleaner Production 142, 985–992.
87. Dengiz O, Baskan O, 2009. Land quality assessment and sustainable land use in Salt Lake (Tuz Gölü) specially protected area. Environmental Monitoring and Assessment 148, 233–243.
88. Dennison Philip E, Thorpe Andrew K, Pardyjak Eric R, Roberts Dar A, Qi Yi, Green Robert O, Bradley Eliza S, Funk Christopher C, 2013. High spatial resolution mapping of elevated atmospheric carbon dioxide using airborne imaging spectroscopy: Radiative transfer modeling and power plant plume detection. Remote Sensing of Environment 139, 116–129. https://doi.org/10.1016/j.rse.2013.08.001
89. Development Alternatives Inc, 2008. Liberia environmental threats an opportunities assessment (LETOA): GIS and spatial data infrastructure final assessment.
90. Deville Pierre, Linard Catherine, Martin Samuel, Gilbert Marius, Stevens Forrest R, Gaughan Andrea E, Blondel Vincent D, Tatem Andrew J, 2014. Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences 201408439. https://doi.org/10.1073/pnas.1408439111
91. Dewan A M, Yamaguchi Y, Rahman M Z, 2012. Dynamics of land use/cover changes and the analysis of landscape fragmentation in Dhaka Metropolitan, Bangladesh. GeoJournal 77, 315–330.
92. Dingel Jonathan I, Miscio Antonio, Davis Donald R, 2019. Cities, lights, and skills in developing economies. Journal of Urban Economics 103174. https://doi.org/10.1016/j.jue.2019.05.005
93. Dobra A, Williams N E, Eagle N, 2015. Spatiotemporal Detection of Unusual Human Population Behavior Using Mobile Phone Data. Plos One 10. https://doi.org/10.1371/journal.pone.0120449
94. Dolan Antenucci, Michael Cafarella, Margaret C Levenstein, Christopher Ré, Matthew D Shapiro, 2014. Using Social Media to Measure Labor Market Flows.
95. Doll Christopher N.H, Muller Jan-Peter, Morley Jeremy G, 2006. Mapping regional economic activity from night-time light satellite imagery. Ecological Economics 57, 75–92. https://doi.org/10.1016/j.ecolecon.2005.03.007
96. Doll Christopher N.H, Pachauri Shonali, 2010. Estimating rural populations without access to electricity in developing countries through night-time light satellite imagery. Energy Policy 38, 5661–5670. https://doi.org/10.1016/j.enpol.2010.05.014
97. Dushaj L, Sallaku F, Tafaj S, Rrapo S, 2011. Application on GIS for land use planning in central part of Albania, Maminas commune. Albanian Journal of Agricultural Sciences 10, 23–29.
98. Eagle Nathan, Macy Michael, Claxton Rob, 2010. Network Diversity and Economic Development. Science 328, 1029. https://doi.org/10.1126/science.1186605
99. Ebener Steeve, Murray Christopher, Tandon Ajay, Elvidge Christopher, 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, 5. https://doi.org/10.1186/1476-072X-4-5
100. El-Zeiny Ahmed, El-Kafrawy Sameh, 2017. Assessment of water pollution induced by human activities in Burullus Lake using Landsat 8 operational land imager and GIS. The Egyptian Journal of Remote Sensing and Space Sciences 20, S49–S56. https://doi.org/10.1016/j.ejrs.2016.10.002
101. Emadodin Iraj, Taravat Alireza, Rajaei Masih, 2016. Effects of urban sprawl on local climate: A case study, north central Iran. Urban Climate 17, 230–247. https://doi.org/10.1016/j.uclim.2016.08.008
102. Erhan L, Ndubuaku M, Ferrara E, Richardson M, Sheffield D, Ferguson F J, Brindley P, Liotta A, 2019. Analyzing Objective and Subjective Data in Social Sciences: Implications for Smart Cities. Ieee Access 7, 19890–19906. https://doi.org/10.1109/access.2019.2897217
103. Evidence Chinedu, Enoguanbhor, Florian Gollnow, Jonas Ostergaard, Nielsen, Tobia Lakes, Blake Byron, Walker, 2019. Land Cover Change in the Abuja City-Region, Nigeria: Integrating GIS and Remotely Sensed Data to Support Land Use Planning. Sustainability 1.
104. Faber Benjamin, Gaubert Cecile, 2019. Tourism and Economic Development: Evidence from Mexico’s Coastline. American Economic Review 109, 2245–93. https://doi.org/10.1257/aer.20161434
105. Faisal Nadeem, Saher, Nasly M A, Tuty Asmawaty, B A K, Yahaya N K. E. M, Wan Mohd, F W I, 2014. Harnessing floodwater of hill torrents for improved spate irrigation system using geo-informatics approach. Research Journal of Recent Sciences 3, 14–22.
106. Feyera Senbeta, 2018. Community perception of land use/land cover change and its impacts on biodiversity and ecosystem services in northwestern Ethiopia. Journal of Sustainable Development in Africa 20, 108–126.
107. Feyisa Gudina L, Henrik Meilby, Rasmus Fensholt, Simon R Proud, 2014. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. Remote Sensing of Environment 140, 23–35. https://doi.org/10.1016/j.rse.2013.08.029
108. Finger Flavio, Genolet Tina, Mari Lorenzo, de Magny, Guillaume Constantin, Manga Noël Magloire, Rinaldo Andrea, Bertuzzo Enrico, 2016. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks. Proceedings of the National Academy of Sciences 113, 6421. https://doi.org/10.1073/pnas.1522305113
109. Firchow Pamina, Mac Ginty, Roger, 2017. Including Hard-to-Access Populations Using Mobile Phone Surveys and Participatory Indicators. Sociological Methods & Research 49, 133–160. https://doi.org/10.1177/0049124117729702
110. Firoz Ahmad, Laxmi Goparaju, 2017. Assessment of threats to forest ecosystems using geospatial technology in Jharkhand state of India. Current World Environment 12, 355–365.
111. Firoz Ahmad, Laxmi Goparaju, 2016. Geospatial technology in urban forest suitability: analysis for Ranchi, Jharkhand, India. Ecological Questions 24, 45–57.
112. Flavio Lupia, Valerio Baiocchi, Keti Lelo, Giuseppe Pulighe, 2017. Exploring Rooftop Rainwater Harvesting Potential for Food Production in Urban Areas. Agriculture 1.
113. Foster Andrew, Gutierrez Emilio, Kumar Naresh, 2009. Voluntary Compliance, Pollution Levels, and Infant Mortality in Mexico. American Economic Review 99, 191–97. https://doi.org/10.1257/aer.99.2.191
114. Frías-Martínez Vanessa, Soguero Ruiz, Cristina, Frias-Martinez Enrique, 2012. Estimation of urban commuting patterns using cellphone network data. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. https://doi.org/10.1145/2346496.2346499
115. Frias-Martinez Vanessa, Virseda-Jerez Jesus, Frias-Martinez Enrique, 2012. On the relation between socio-economic status and physical mobility. Information Technology for Development 18, 91–106. https://doi.org/10.1080/02681102.2011.630312
116. Gabriel Kreindler, Yuhei Miyauchi, 2013. Commuting and Productivity: Quantifying Urban Economic Activity using Cell Phone Data.
117. Garcia Andres J, Pindolia Deepa K, Lopiano Kenneth K, Tatem Andrew J, 2014. Modeling internal migration flows in sub-Saharan Africa using census microdata. Migration Studies 3, 89–110. https://doi.org/10.1093/migration/mnu036
118. Georgiadis G, Spanou S, Kokkoris I, Tiniakou A, Georgiadis Th, 2014. Introducing an integrated monitoring system for natural ecosystems: the example of Strofilia wetlands in Western Peloponnese, Greece. International Journal of Environmental Research 8, 1195–1202.
119. Gillespie Thomas W, Chu Jasmine, Frankenberg Elizabeth, Thomas Duncan, 2007. Assessment and prediction of natural hazards from satellite imagery. Progress in Physical Geography: Earth and Environment 31, 459–470. https://doi.org/10.1177/0309133307083296
120. Goab J, uczkowski B, 2017. GIS maps and analysis in designing forest road system on mountainous areas. Electronic Journal of Polish Agricultural Universities 20, art 14-art 14.
121. Gong HuiLi, Tang T, Gong ZhaoNing, Li XiaoJuan, Chen YuZhu, Zhao WenJi, 2015. Spatial and temporal trend of water resources in Beijing, China during 1999-2012 and its impact analysis. British Journal of Environment and Climate Change 5, 176–188.
122. González Marta C, Hidalgo César A, Barabási Albert-László, 2008. Understanding individual human mobility patterns. Nature 453, 779–782. https://doi.org/10.1038/nature06958
123. Grant A T. J, McKinney N L, Ries R, 2018. An approach to quantifying rainwater harvesting potential using imagery, geographic information systems (GIS) and LiDAR data. Water Science and Technology: Water Supply 18, 108–118.
124. Guan Kaiyu, Wu Jin, Kimball John S, Anderson Martha C, Frolking Steve, Li Bo, Hain Christopher R, Lobell David B, 2017. The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields. Remote Sensing of Environment 199, 333–349. https://doi.org/10.1016/j.rse.2017.06.043
125. Gumma Murali Krishna, Birhanu Birhanu Zemadim, Mohammed Irshad A, Tabo Ramadjita, Whitbread Anthony M, 2016. Prioritization of Watersheds across Mali Using Remote Sensing Data and GIS Techniques for Agricultural Development Planning.
126. Guo Yan, Su XueMei, 2012. Mobile device-based reporting system for Sichuan earthquake-affected areas infectious disease reporting in China. Biomedical and Environmental Sciences 25, 724–729.
127. Han WangYa, Liu GuoHua, Su XuKun, Wu Xing, Chen Li, 2019. Assessment of potential land degradation and recommendations for management in the south subtropical region, Southwest China. Land Degradation & Development 30, 979–990.
128. Hartter Joel, Southworth Jane, 2009. Dwindling resources and fragmentation of landscapes around parks: wetlands and forest patches around Kibale National Park, Uganda. Landscape Ecology 24, 643–656. https://doi.org/10.1007/s10980-009-9339-7
129. Hayano Ryugo S, ADACHI Ryutaro, 2013. Estimation of the total population moving into and out of the 20km evacuation zone during the Fukushima NPP accident as calculated using “Auto-GPS” mobile phone data 89, 196–199. https://doi.org/10.2183/pjab.89.196
130. Head Andrew, Manguin Melanie, Tran Nhat, Blumenstock Joshua, 2017. Can Human Development be Measured with Satellite Imagery?
131. Heathman G C, Larose M, Ascough J C, II, 2009. Soil and Water Assessment Tool evaluation of soil and land use geographic information system data sets on simulated stream flow. Journal of Soil and Water Conservation (Ankeny) 64, 17–32.
132. Heft-Neal Sam, Burney Jennifer, Bendavid Eran, Voss Kara, Burke Marshall, 2019. Air Pollution and Infant Mortality: Evidence from Saharan Dust. National Bureau of Economic Research Working Paper Series No. 26107. https://doi.org/10.3386/w26107
133. Heinzelman Jessica, Brown Rachel, Meier Patrick, 2011. Mobile Technology, Crowdsourcing and Peace Mapping: New Theory and Applications for Conflict Management, in: Poblet Marta (Ed.), Mobile Technologies for Conflict Management: Online Dispute Resolution, Governance, Participation. Springer Netherlands, Dordrecht, pp. 39–53. https://doi.org/10.1007/978-94-007-1384-0\_4
134. Hively W D, Lang M, McCarty G W, Keppler J, Sadeghi A, McConnell L L, 2009. Using satellite remote sensing to estimate winter cover crop nutrient uptake efficiency. Journal of Soil and Water Conservation (Ankeny) 64, 303–313.
135. Honarvar Ali Reza, Sami Ashkan, 2019. Towards Sustainable Smart City by Particulate Matter Prediction Using Urban Big Data, Excluding Expensive Air Pollution Infrastructures. Big Data Research 17, 56–65. https://doi.org/10.1016/j.bdr.2018.05.006
136. Hristova Desislava, Rutherford Alex, Anson Jose, Luengo-Oroz Miguel, Mascolo Cecilia, 2016. The International Postal Network and Other Global Flows as Proxies for National Wellbeing. PLOS ONE 11, e0155976.
137. Huan Juan, Cao WeiJian, Liu XingQiao, 2017. A dissolved oxygen prediction method based on K-means clustering and the ELM neural network: a case study of the Changdang Lake, China. Applied Engineering in Agriculture 33, 461–469.
138. Huang H P, Li Q Z, Zhang Y, 2019. Urban Residential Land Suitability Analysis Combining Remote Sensing and Social Sensing Data: A Case Study in Beijing, China. Sustainability 11. https://doi.org/10.3390/su11082255
139. Huang Z R, Ling X M, Wang P, Zhang F, Mao Y P, Lin T, Wang F Y, 2018. Modeling real-time human mobility based on mobile phone and transportation data fusion. Transportation Research Part C-Emerging Technologies 96, 251–269. https://doi.org/10.1016/j.trc.2018.09.016
140. Ibrahim M, Koch B, 2015. Assessment and mapping of groundwater vulnerability using SAR concentrations and GIS: a case study in Al-Mafraq, Jordan. Journal of Water Resource and Protection 7, 588–596.
141. Ikemi H, 2017. Geologically constrained changes to landforms caused by human activities in the 20th century: a case study from Fukuoka Prefecture, Japan. Applied Geography 87, 115–126.
142. J Tian, N Zhao, E L Samson, S Wang, 2014. Brightness of Nighttime Lights as a Proxy for Freight Traffic: A Case Study of China. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 206–212. https://doi.org/10.1109/JSTARS.2013.2258892
143. Jahan Saherah, Kalita Sarbeswar, Kumar B B, 2015. An assessment of land use-land cover change using geoinformatics in Sonai-Rupai Wildlife Sanctuary, Assam, India. Journal of Environmental Research and Development 9, 1257–1263.
144. Jahani Eaman, Sundsøy Pål, Bjelland Johannes, Bengtsson Linus, Pentland Alex ‘Sandy,’ de Montjoye, Yves-Alexandre, 2017. Improving official statistics in emerging markets using machine learning and mobile phone data. EPJ Data Science 6, 3. https://doi.org/10.1140/epjds/s13688-017-0099-3
145. Jain M, Singh Balwinder, Srivastava A A K, Malik R K, McDonald A J, Lobell D B, 2017. Using satellite data to identify the causes of and potential solutions for yield gaps in India’s Wheat Belt. Environmental Research Letters 12, 094011. https://doi.org/10.1088/1748-9326/aa8228
146. Jain Rishee K, Smith Kevin M, Culligan Patricia J, Taylor John E, 2014. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. Applied Energy 168.
147. Jäppinen Sakari, Toivonen Tuuli, Salonen Maria, 2013. Modelling the potential effect of shared bicycles on public transport travel times in Greater Helsinki: An open data approach. Applied Geography 43, 13–24. https://doi.org/10.1016/j.apgeog.2013.05.010
148. Järv Olle, Ahas Rein, Saluveer Erki, Derudder Ben, Witlox Frank, 2012. Mobile Phones in a Traffic Flow: A Geographical Perspective to Evening Rush Hour Traffic Analysis Using Call Detail Records. PLOS ONE 7, e49171.
149. Jawdar Mohammad Yousif Al, Shiobara Manabu, Onuma Takumi, 2005. Chapter 2 Monitoring of coastal environment using satellite images in the United Arab Emirates. Developments in Earth and Environmental Sciences 3, 13–29. https://doi.org/10.1016/S1571-9197(05)80026-6
150. Jayasekara M J. P. T. M, Kadupitiya H K, Vitharana U W. A, 2018. Mapping of soil erosion hazard zones of Sri Lanka. Tropical Agricultural Research 29, 135–146.
151. Jean Neal, Burke Marshall, Xie Michael, Davis W Matthew, Lobell David B, Ermon Stefano, 2016. Combining satellite imagery and machine learning to predict poverty. Science 353, 790. https://doi.org/10.1126/science.aaf7894
152. Jeremiah J Nieves, Forrest R Stevens, rea E Gaughan, Catherine Linard, Aless, ro Sorichetta, Graeme Hornby, Nirav N Patel, rew J Tatem, 2017. Examining the correlates and drivers of human population distributions across low- and middle-income countries.
153. Jiang Yunfang, Shi Tiemao, Gu Xixi, 2016. Healthy urban streams: The ecological continuity study of the Suzhou creek corridor in Shanghai. Cities 59, 80–94. https://doi.org/10.1016/j.cities.2016.06.002
154. Jin Zhenong, Azzari George, Burke Marshall, Aston Stephen, Lobell B David, 2017. Mapping Smallholder Yield Heterogeneity at Multiple Scales in Eastern Africa. Remote Sensing 9. https://doi.org/10.3390/rs9090931
155. Jones M O, Allred B W, Naugle D E, Maestas J D, Donnelly P, Metz L J, Karl J, Smith R, Bestelmeyer B, Boyd C, Kerby J D, McIver J D, 2018. Innovation in rangeland monitoring: annual, 30 m, plant functional type percent cover maps for U.S. rangelands, 1984-2017. Ecosphere 9, e02430–e02430.
156. Jones M O, Kimball J S, Jones L A, 2013. Satellite microwave detection of boreal forest recovery from the extreme 2004 wildfires in Alaska and Canada. Global Change Biology 19, 3111–3122.
157. Juniper S K, Thornborough K, Douglas K, Hillier J, 2019. Remote monitoring of a deep-sea marine protected area: The Endeavour Hydrothermal Vents. Aquatic Conservation-Marine and Freshwater Ecosystems 29, 84–102. https://doi.org/10.1002/aqc.3020
158. Kabisch N, Selsam P, Kirsten T, Lausch A, Bumberger J, 2019. A multi-sensor and multi-temporal remote sensing approach to detect land cover change dynamics in heterogeneous urban landscapes. Ecological Indicators 99, 273–282.
159. Kaiser M F, Aboulela H, El Serehy, H, Ezz Edin, H, 2010. Spectral enhancement of SPOT imagery data to assess marine pollution near Port Said, Egypt. International Journal of Remote Sensing 31, 1753–1764. https://doi.org/10.1080/01431160902926624
160. Kaliraj S, Chandrasekar N, Ramachandran K K, 2017. Mapping of coastal landforms and volumetric change analysis in the south west coast of Kanyakumari, South India using remote sensing and GIS techniques. The Egyptian Journal of Remote Sensing and Space Sciences 20, 265–282. https://doi.org/10.1016/j.ejrs.2016.12.006
161. Kalvani S R, Sharaai A, Manaf L, Hamidian A, 2019. Assessing ground and surface water scarcity indices using ground and surface water footprints in the Tehran province of Iran. Applied Ecology and Environmental Research 17, 4985–4997.
162. Kan Z H, Tang L L, Kwan M P, Ren C, Liu D, Pei T, Liu Y, Deng M, Li Q Q, 2018a. Fine-grained analysis on fuel-consumption and emission from vehicles trace. Journal of Cleaner Production 203, 340–352. https://doi.org/10.1016/j.jclepro.2018.08.222
163. Kan Z H, Tang L L, Kwan M P, Zhang X, 2018b. Estimating Vehicle Fuel Consumption and Emissions Using GPS Big Data. International Journal of Environmental Research and Public Health 15. https://doi.org/10.3390/ijerph15040566
164. Kangbéni Dimobe, Jean Léandre N’djoré, Kouakou, Jérôme E Tondoh, Benewinde J B. Zoungrana, Gerald Forkuor, Korotimi Ouédraogo, 2018. Predicting the Potential Impact of Climate Change on Carbon Stock in Semi-Arid West African Savannas. Land 1.
165. Karan S K, Samadder S R, 2016. Accuracy of land use change detection using support vector machine and maximum likelihood techniques for open-cast coal mining areas. Environmental Monitoring and Assessment 188, 486–486.
166. Karapetyan D, d’Adda G, 2014. Determinants of conservation among the rural poor: A charitable contribution experiment. Ecological Economics 99, 74–87. https://doi.org/10.1016/j.ecolecon.2014.01.009
167. Katerina Tzavella, Alexander Fekete, Frank Fiedrich, 2018. Opportunities provided by geographic information systems and volunteered geographic information for a timely emergency response during flood events in Cologne, Germany. Natural Hazards: Journal of the International Society for the Prevention and Mitigation of Natural Hazards 29. https://doi.org/10.1007/s11069-017-3102-1
168. Kaushalya Madhawa, Sriganesh Lokanathan, Danaja Maldeniya, Rohan Samarajiva, 2015. Using mobile network big data for land use classification.
169. Keeratikasikorn C, Bonafoni S, 2018. Urban heat island analysis over the land use zoning plan of Bangkok by means of Landsat 8 imagery. Remote Sensing 10, 440–440.
170. Khakh Amritpal Kaur, Fast Victoria, 2017. Measuring Spatial Accessibility of Healthcare Services in Calgary (poster). Journal of Transport & Health 7, S13–S14. https://doi.org/10.1016/j.jth.2017.11.023
171. Khalil A, Hanich L, Hakkou R, Lepage M, 2014. GIS-based environmental database for assessing the mine pollution: A case study of an abandoned mine site in Morocco. Journal of Geochemical Exploration 144, 468–477. https://doi.org/10.1016/j.gexplo.2014.03.023
172. Khomba Daniel Chris, Trew Alex, 2017. Aid and Growth in Malawi.
173. Khorchani Makki, Vicente-Serrano Sergio M, Azorin-Molina Cesar, Garcia Monica, Martin-Hernandez Natalia, Peña-Gallardo Marina, El Kenawy, Ahmed, Domínguez-Castro Fernando, 2018. Trends in LST over the peninsular Spain as derived from the AVHRR imagery data. Global and Planetary Change 166, 75–93. https://doi.org/10.1016/j.gloplacha.2018.04.006
174. Kim Hwan Yong, Lee Hye Kyung, 2014. Enhanced validity and reliability of spatial decision support systems (SDSS) for sustainable transportation decision-making. Applied Geography 51, 65–71. https://doi.org/10.1016/j.apgeog.2014.03.009
175. Kim JeongKon, Noh JoonWoo, Son KyungHo, Kim IkJae, 2012. Impacts of GIS data quality on determination of runoff and suspended sediments in the Imha watershed in Korea. Geosciences Journal 16, 181–192.
176. Kim S J, Marsch L A, Hancock J T, Das A K, 2017. Scaling Up Research on Drug Abuse and Addiction Through Social Media Big Data. Journal of Medical Internet Research 19. https://doi.org/10.2196/jmir.6426
177. Kim Y H, Im Jungho, Ha HoKyung, Choi JongKuk, Ha SungHyun, 2014. Machine learning approaches to coastal water quality monitoring using GOCI satellite data. GIScience and Remote Sensing 51, 158–174.
178. Kinyanjui M J, 2011. NDVI-based vegetation monitoring in Mau forest complex, Kenya. African Journal of Ecology 49, 165–174.
179. Knadel M, Thomsen A, Schelde K, Greve M H, 2015. Soil organic carbon and particle sizes mapping using vis-NIR, EC and temperature mobile sensor platform. Computers and Electronics in Agriculture 114, 134–144.
180. Knudby A, Roelfsema C, Lyons M, Phinn S, Jupiter S, 2011. Mapping fish community variables by integrating field and satellite data, object-based image analysis and modeling in a traditional Fijian fisheries management area. Remote Sensing 3, 460–483.
181. Kong C F, Lan H, Yang G, Xu K, 2016. Geo-environmental suitability assessment for agricultural land in the rural-urban fringe using BPNN and GIS: a case study of Hangzhou. Environmental Earth Sciences 75. https://doi.org/10.1007/s12665-016-5956-z
182. Kong Chunfang, Xu Kai, Wu Chonglong, 2006. Classification and Extraction of Urban Land-Use Information from High-Resolution Image Based on Object Multi-features. Journal of China University of Geosciences 17, 151–157. https://doi.org/10.1016/S1002-0705(06)60021-6
183. Kraemer Moritz U G, Faria Nuno R, Reiner Robert C, J., Golding Nick, Nikolay Birgit, Stasse Stephanie, Johansson Michael A, Salje Henrik, Faye Ousmane, Wint G R William, Niedrig Matthias, Shearer Freya M, Hill Sarah C, Thompson Robin N, Bisanzio Donal, Taveira Nuno, Nax Heinrich H, Pradelski Bary S R, Nsoesie Elaine O, Murphy Nicholas R, Bogoch Isaac I, Khan Kamran, Brownstein John S, Tatem Andrew J, de Oliveira, Tulio, Smith David L, Sall Amadou A, Pybus Oliver G, Hay Simon I, Cauchemez Simon, 2017. Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015–16: a modelling study. The Lancet Infectious Diseases 17, 330–338. https://doi.org/10.1016/S1473-3099(16)30513-8
184. Krishnadas Meghna, Agarwala Meghna, Sridhara Sachin, Eastwood Erin, 2018. Parks protect forest cover in a tropical biodiversity hotspot, but high human population densities can limit success. Biological Conservation 223, 147–155. https://doi.org/10.1016/j.biocon.2018.04.034
185. Kropáček Jan, Braun Andreas, Kang Shichang, Feng Chen, Ye Qinghua, Hochschild Volker, 2012. Analysis of lake level changes in Nam Co in central Tibet utilizing synergistic satellite altimetry and optical imagery. International Journal of Applied Earth Observations and Geoinformation 17, 3–11. https://doi.org/10.1016/j.jag.2011.10.001
186. Krtička L, Tomčíková I, Rakytová I, 2018. Development versus conservation: evaluation of landscape structure changes in Demänovská Valley, Slovakia. Journal of Mountain Science 15, 1153–1170.
187. Kubitza Christoph, Krishna Vijesh V, Urban Kira, Alamsyah Zulkifli, Qaim Matin, 2018. Land Property Rights, Agricultural Intensification, and Deforestation in Indonesia. Ecological Economics 312.
188. Kwasi Appeaning, Addo, 2015. Monitoring sea level rise-induced hazards along the coast of Accra in Ghana. Natural Hazards: Journal of the International Society for the Prevention and Mitigation of Natural Hazards 1293. https://doi.org/10.1007/s11069-015-1771-1
189. L Pappalardo, D Pedreschi, Z Smoreda, F Giannotti, 2015. Using big data to study the link between human mobility and socio-economic development. Presented at the 2015 IEEE International Conference on Big Data (Big Data), pp. 871–878. https://doi.org/10.1109/BigData.2015.7363835
190. Ladd B, Laffan S W, Amelung W, Peri P L, Silva L C. R, Gervassi P, Bonser S P, Navall M, Sheil D, 2013. Estimates of soil carbon concentration in tropical and temperate forest and woodland from available GIS data on three continents. Global Ecology and Biogeography 22, 461–469.
191. Lai Shengjie, Erbach-Schoenberg Elisabeth zu, Pezzulo Carla, Ruktanonchai Nick W, Sorichetta Alessandro, Steele Jessica, Li Tracey, Dooley Claire A, Tatem Andrew J, 2019. Exploring the use of mobile phone data for national migration statistics. Palgrave Communications 5, 34. https://doi.org/10.1057/s41599-019-0242-9
192. le Blanc, David, Perez Romain, 2008. The relationship between rainfall and human density and its implications for future water stress in Sub-Saharan Africa. Ecological Economics 319.
193. Le Menach, Arnaud, Tatem Andrew J, Cohen Justin M, Hay Simon I, Randell Heather, Patil Anand P, Smith David L, 2011. Travel risk, malaria importation and malaria transmission in Zanzibar. Scientific Reports 1, 93. https://doi.org/10.1038/srep00093
194. Leauthaud C, Belaud G, Duvail S, Moussa R, Grünberger O, Albergel J, 2013. Characterizing floods in the poorly gauged wetlands of the Tana River Delta, Kenya, using a water balance model and satellite data. Hydrology and Earth System Sciences 17, 3059–3075.
195. Lee E C, Asher J M, Goldlust S, Kraemer J D, Lawson A B, Bansal S, 2016. Mind the scales: harnessing spatial big data for infectious disease surveillance and inference. Journal of Infectious Diseases 214, S409–S413.
196. Lee SunMin, Lee SaRo, Lee MoungJin, Jung HyungSup, 2018. Spatial assessment of urban flood susceptibility using data mining and geographic information system (GIS) tools. Sustainability 10, 648–648.
197. Lee Taejun, Park Hyojung, Lee Junesoo, 2019. Collaborative accountability for sustainable public health: A Korean perspective on the effective use of ICT-based health risk communication. Government Information Quarterly 36, 226–236. https://doi.org/10.1016/j.giq.2018.12.008
198. Lehtomäki Joona, Tomppo Erkki, Kuokkanen Panu, Hanski Ilkka, Moilanen Atte, 2009. Applying spatial conservation prioritization software and high-resolution GIS data to a national-scale study in forest conservation. Forest Ecology and Management 258, 2439–2449. https://doi.org/10.1016/j.foreco.2009.08.026
199. Li HongMei, Ma YouXin, Aide T M, Liu WenJun, 2008. Past, present and future land-use in Xishuangbanna, China and the implications for carbon dynamics. Forest Ecology and Management 255, 16–24.
200. Li Shen, Jared M Kyllo, Xulin Guo, 2013. An Integrated Model Based on a Hierarchical Indices System for Monitoring and Evaluating Urban Sustainability. Sustainability 1.
201. Li Xi, 2014. Can night-time light images play a role in evaluating the Syrian Crisis? International Journal of Remote Sensing 35. https://doi.org/10.1080/01431161.2014.971469
202. Linard Catherine, Kabaria Caroline W, Gilbert Marius, Tatem Andrew J, Gaughan Andrea E, Stevens Forrest R, Sorichetta Alessandro, Noor Abdisalan M, Snow Robert W, 2017. Modelling changing population distributions: an example of the Kenyan Coast, 1979–2009. International Journal of Digital Earth 10, 1017–1029. https://doi.org/10.1080/17538947.2016.1275829
203. Ljiljana Živković, Aleksandar Đorđević, 2016. Building a GIS Platform for Sustainable Land Management: A Case Study of the City of Čačak, Serbia. Journal of Urban Technology 29. https://doi.org/10.1080/10630732.2015.1102420
204. Llorente Alejandro, Garcia-Herranz Manuel, Cebrian Manuel, Moro Esteban, 2015. Social Media Fingerprints of Unemployment. PLoS ONE 10, e0128692. https://doi.org/10.1371/journal.pone.0128692
205. Lobell David B, Ortiz-Monasterio J Ivan, Lee Anna S, 2010. Satellite evidence for yield growth opportunities in Northwest India. Field Crops Research 118, 13–20. https://doi.org/10.1016/j.fcr.2010.03.013
206. Louail Thomas, Lenormand Maxime, Cantu Ros, Oliva G, Picornell Miguel, Herranz Ricardo, Frias-Martinez Enrique, Ramasco José J, Barthelemy Marc, 2014. From mobile phone data to the spatial structure of cities. Scientific Reports 4, 5276. https://doi.org/10.1038/srep05276
207. Lu D, Hetrick S, Moran E, Li G, 2012. Application of Time Series Landsat Images to Examining Land-use/Land-cover Dynamic Change.
208. Lu Xin, Bengtsson Linus, Holme Petter, 2012. Predictability of population displacement after the 2010 Haiti earthquake. Proceedings of the National Academy of Sciences. https://doi.org/10.1073/pnas.1203882109
209. Lu Xin, Brelsford Christa, 2014. Network Structure and Community Evolution on Twitter: Human Behavior Change in Response to the 2011 Japanese Earthquake and Tsunami. Scientific Reports 4, 6773. https://doi.org/10.1038/srep06773
210. Lu Xin, Wetter Erik, Bharti Nita, Tatem Andrew J, Bengtsson Linus, 2013. Approaching the Limit of Predictability in Human Mobility. Scientific Reports 3, 2923. https://doi.org/10.1038/srep02923
211. Lu Xin, Wrathall David J, Sundsøy Pål Roe, Nadiruzzaman Md, Wetter Erik, Iqbal Asif, Qureshi Taimur, Tatem Andrew, Canright Geoffrey, Engø-Monsen Kenth, Bengtsson Linus, 2016a. Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh. Global Environmental Change 38, 1–7. https://doi.org/10.1016/j.gloenvcha.2016.02.002
212. Lu Xin, Wrathall David, Sundsøy Pål, Nadiruzzaman Md, Wetter Erik, Iqbal Asif, Qureshi Taimur, Tatem Andrew, Canright Geoffrey, Engø-Monsen Kenth, Bengtsson Linus, 2016b. Detecting climate adaptation with mobile network data in Bangladesh: anomalies in communication, mobility and consumption patterns during cyclone Mahasen. Climatic Change 138, 505–519. https://doi.org/10.1007/s10584-016-1753-7
213. Luis A Andres Samir Bhatt Basab Dasgupta Juan A. Echeniqu, 2018. Geo-Spatial Modeling of Access to Water and Sanitation in Nigeria, Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-8357
214. Madden M, Ross A, 2009. Genocide and GIScience: Integrating Personal Narratives and Geographic Information Science to Study Human Rights. Professional Geographer 61, 508–526. https://doi.org/10.1080/00330120903163480
215. Madu I A, 2011. Spatial vulnerability of rural Nigeria to climate change: implications for internal security. International Journal of Climate Change: Impacts and Responses 3, 79–98.
216. Mahler Halley, Searle Sarah, Plotkin Marya, Kulinda Yusuph, Greenberg Seth, Mlanaga Erick, Njeuhmeli Emmanuel, Lija Gissenje, 2015. Covering the Last Kilometer: Using GIS to Scale-Up Voluntary Medical Male Circumcision Services in Iringa and Njombe Regions, Tanzania.
217. Maina Joseph, Venus Valentijn, McClanahan Timothy R, Ateweberhan Mebrahtu, 2008. Modelling susceptibility of coral reefs to environmental stress using remote sensing data and GIS models. Ecological Modelling 212, 180–199. https://doi.org/10.1016/j.ecolmodel.2007.10.033
218. Malleson N, Andresen M A, 2015. The impact of using social media data in crime rate calculations: shifting hot spots and changing spatial patterns. Cartography and Geographic Information Science 42, 112–121. https://doi.org/10.1080/15230406.2014.905756
219. March R G, Smith E H, 2012. Modeling potential coastal vegetation response to sea level rise and storm surge on estuarine peninsulas. Journal of Coastal Research 28, 993–1007.
220. Marco Hernandez Lingzi Hong Vanessa Frias-Martinez Enrique Frias-M, 2017. Estimating Poverty Using Cell Phone Data: Evidence from Guatemala, Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-7969
221. Marfai Muh, Aris, 2014. Impact Of Sea Level Rise To Coastal Ecology: A Case Study On The Northern Part Of Java Island, Indonesia. Quaestiones Geographicae 107.
222. Martins F R, Pereira E B, Abreu S L, 2007. Satellite-derived solar resource maps for Brazil under SWERA project. Solar Energy 81, 517–528. https://doi.org/10.1016/j.solener.2006.07.009
223. Mathew Aneesh, Khandelwal Sumit, Kaul Nivedita, 2018. Spatio-temporal variations of surface temperatures of Ahmedabad city and its relationship with vegetation and urbanization parameters as indicators of surface temperatures. Remote Sensing Applications: Society and Environment 11, 119–139. https://doi.org/10.1016/j.rsase.2018.05.003
224. Melchiorri M, Pesaresi M, Florczyk A J, Corbane C, Kemper T, 2019. Principles and Applications of the Global Human Settlement Layer as Baseline for the Land Use Efficiency Indicator-DSDG 11.3.1. Isprs International Journal of Geo-Information 8. https://doi.org/10.3390/ijgi8020096
225. Merem C E, Isokpehi P, Wesley J, Nwagboso E, Richardson C, Fageir S, Iwehabura S, Crisler M, 2014. The analysis of coal mining impacts on West Virginia’s environment. British Journal of Applied Science & Technology 4, 1171–1197.
226. Milena Lopreite, Michelangelo Puliga, Massimo Riccaboni, 2018. The Global Health Networks: A Comparative Analysis of Tuberculosis, Malaria and Pneumonia Using Social Media Data.
227. Miles V V, Esau I, 2016. Spatial heterogeneity of greening and browning between and within bioclimatic zones in northern West Siberia. Environmental Research Letters 11, 115002–115002.
228. Min Brian, Gaba Kwawu Mensan, Sarr Ousmane Fall, Agalassou Alassane, 2013. Detection of rural electrification in Africa using DMSP-OLS night lights imagery. International Journal of Remote Sensing 34, 8118–8141. https://doi.org/10.1080/01431161.2013.833358
229. Moffitt D L, Lalit Kumar, 2018. Remote sensing of a shallow, fringing reef platform for analysis of island sector susceptibility and development of a coastal vulnerability index. Journal of Coastal Research 34, 122–135.
230. Mohammad Mehedy, Hassan, Mohhamad Nurul Islam, Nazem, 2016. Examination of land use/land cover changes, urban growth dynamics, and environmental sustainability in Chittagong city, Bangladesh. Environment, Development and Sustainability: A Multidisciplinary Approach to the Theory and Practice of Sustainable Development 697. https://doi.org/10.1007/s10668-015-9672-8
231. Mohammad Redowan, Sharmin Akter, Nusrat Islam, 2014. Analysis of forest cover change at Khadimnagar National Park, Sylhet, Bangladesh, using Landsat TM and GIS data. Journal of Forestry Research 25, 393–400.
232. Mokgedi Leokeditswe, Nobert Joel, Munishi Subira, 2019. Assessment of lake surface dynamics using satellite imagery and in-situ data; case of Lake Ngami in North-West Botswana. Physics and Chemistry of the Earth 112, 175–186. https://doi.org/10.1016/j.pce.2018.12.008
233. Molla M B, 2015. Land Use/Land Cover dynamics in the Central Rift Valley region of Ethiopia: case of Arsi Negele District. African Journal of Agricultural Research 10, 434–449.
234. Mondal Pinki, Tatem Andrew J, 2012. Uncertainties in Measuring Populations Potentially Impacted by Sea Level Rise and Coastal Flooding. PLOS ONE 7, e48191.
235. Mori H, Sah A K, Sah B P, Yamaya M, Senthil Selvaraj, 2015. An approach for monitoring the reforestation and conservation efforts by local communities. International Journal of Environmental and Rural Development 6, 91–96.
236. Moss R, Naghizade E, Tomko M, Geard N, 2019. What can urban mobility data reveal about the spatial distribution of infection in a single city? BMC Public Health 19, (29 May 2019)-(29 May 2019).
237. Moufaddal Wahid, 2005. Use of Satellite Imagery as Environmental Impact Assessment Tool: A case Study from the Nw Egyptian Red Sea Coastal Zone. Environmental Monitoring & Assessment 107, 427–452. https://doi.org/10.1007/s10661-005-3576-2
238. Msoffe F U, Said M Y, Ogutu J O, Kifugo S C, Leeuw J de, Gardingen P van, Reid R S, 2011. Spatial correlates of land-use changes in the Maasai-Steppe of Tanzania: implications for conservation and environmental planning. International Journal of Biodiversity and Conservation 3, 280–290.
239. Muhammad Jehanzeb Masud Cheema, Allah Baksh, Talha Mahmood, Muhammad Usman Liaqat, 2016. Assessment of water allocations using remote sensing and GIS modeling for Indus Basin, Pakistan.
240. Musaoglu N, Tanik A, Dikerler T, Buhur S, 2015. Use of remote sensing and geographic information systems in the determination of high-risk areas regarding marine traffic in the Istanbul Strait. Environmental Hazards-Human and Policy Dimensions 14, 54–73. https://doi.org/10.1080/17477891.2014.986042
241. Musaoglu Nebiye, Gurel Melike, Ulugtekin Necla, Tanik Aysegul, Seker Dursun Zafer, 2006. Use of Remotely Sensed Data for Analysis of Land-Use Change in a Highly Urbanized District of Mega City, Istanbul. Journal of Environmental Science & Health, Part A: Toxic/Hazardous Substances & Environmental Engineering 41, 2057–2069. https://doi.org/10.1080/10934520600780719
242. Nackoney Janet, Rybock Demian, Dupian Jef, Facheux Charly, 2013. Coupling participatory mapping and GIS to inform village-level agricultural zoning in the Democratic Republic of the Congo.
243. Nawapramote W, Prabudhanitisarn S, Sangawongse S, Promburom P, 2014. Co-management in land demarcation to reduce forest utilization conflicts: a case study of Mae Tia-Mae Tae watershed, Chom Thong district, Chiang Mai province. Environment and Natural Resources Journal 12, 47–57.
244. Nigatu Wondrade, Dick Øystein, Tveite Havard, 2014. GIS based mapping of land cover changes utilizing multi-temporal remotely sensed image data in Lake Hawassa Watershed, Ethiopia. Environmental Monitoring & Assessment 186, 1765–1780. https://doi.org/10.1007/s10661-013-3491-x
245. Nikolaos Askitas, Klaus F Zimmermann, 2009. Google Econometrics and Unemployment Forecasting.
246. Niranjana K V, Sathish A, 2011. Remote sensing and GIS approach for watershed based resources management in the Eastern Dry Zone of Karnataka. Mysore Journal of Agricultural Sciences 45, 316–321.
247. Noori A M, Biswajeet Pradhan, Ajaj Q M, 2019. Dam site suitability assessment at the greater zab river in northern Iraq using remote sensing data and GIS. Journal of Hydrology (Amsterdam) 574, 964–979.
248. Novak Jakub, Ahas Rein, Aasa Anto, Silm Siiri, 2013. Application of mobile phone location data in mapping of commuting patterns and functional regionalization: a pilot study of Estonia. Journal of Maps 9, 10–15. https://doi.org/10.1080/17445647.2012.762331
249. Ojoyi M M, Antwi-Agyei P, Mutanga O, Odindi J, Abdel-Rahman E M, 2015. An analysis of ecosystem vulnerability and management interventions in the Morogoro region landscapes, Tanzania. Tropical Conservation Science 8, 662–680.
250. Oloo F, 2018. Mapping Rural Road Networks from Global Positioning System (GPS) Trajectories of Motorcycle Taxis in Sigomre Area, Siaya County, Kenya. Isprs International Journal of Geo-Information 7. https://doi.org/10.3390/ijgi7080309
251. Ozbay G, Fan ChunLei, Yang ZhiMing, 2017. Relationship between land use and water quality and its assessment using hyperspectral remote sensing in Mid-Atlantic Estuaries. Water quality.
252. Papastergiadou E, Kagalou I, Stefanidis K, Retalis A, Leonardos I, 2010. Effects of Anthropogenic Influences on the Trophic State, Land Uses and Aquatic Vegetation in a Shallow Mediterranean Lake: Implications for Restoration. Water Resources Management: An International Journal, Published for the European Water Resources Association (EWRA) 415. https://doi.org/10.1007/s11269-009-9453-y
253. Pásztor László, Laborczi Annamária, Takács Katalin, Szatmári Gábor, Fodor Nándor, Illés Gábor, Farkas-Iványi Kinga, Bakacsi Zsófia, Szabó József, 2017. Chapter 9 - Compilation of Functional Soil Maps for the Support of Spatial Planning and Land Management in Hungary. Soil Mapping and Process Modeling for Sustainable Land Use Management 293–317. https://doi.org/10.1016/B978-0-12-805200-6.00009-8
254. Patel Nirav N, Angiuli Emanuele, Gamba Paolo, Gaughan Andrea, Lisini Gianni, Stevens Forrest R, Tatem Andrew J, Trianni Giovanna, 2015. Multitemporal settlement and population mapping from Landsat using Google Earth Engine. International Journal of Applied Earth Observation and Geoinformation 35, 199–208. https://doi.org/10.1016/j.jag.2014.09.005
255. Patel Nirav N, Stevens Forrest R, Huang Zhuojie, Gaughan Andrea E, Elyazar Iqbal, Tatem Andrew J, 2017. Improving Large Area Population Mapping Using Geotweet Densities. Transactions in GIS 21, 317–331. https://doi.org/10.1111/tgis.12214
256. Patode R S, Pande C B, Nagdeve M B, Moharir K N, Wankhade R M, 2017. Planning of conservation measures for watershed management and development by using geospatial technology - a case study of Patur watershed in Akola district of Maharashtra. Current World Environment 12, 708–716.
257. Peak Corey M, Wesolowski Amy, zu Erbach-Schoenberg, Elisabeth, Tatem Andrew J, Wetter Erik, Lu Xin, Power Daniel, Weidman-Grunewald Elaine, Ramos Sergio, Moritz Simon, Buckee Caroline O, Bengtsson Linus, 2018. Population mobility reductions associated with travel restrictions during the Ebola epidemic in Sierra Leone: use of mobile phone data. International Journal of Epidemiology 47, 1562–1570. https://doi.org/10.1093/ije/dyy095
258. Pellikka Petri K. E, Clark Barnaby J. F, Gosa Alemu Gonsamo, Himberg Nina, Hurskainen Pekka, Maeda Eduardo, Mwang’ombe James, Omoro Loice M. A, Siljander Mika, 2013. Chapter 13 - Agricultural Expansion and Its Consequences in the Taita Hills, Kenya. Developments in Earth Surface Processes 16, 165–179. https://doi.org/10.1016/B978-0-444-59559-1.00013-X
259. Peng J, Wu J S, Yin H, Li Z G, Chang Q, Mu T L, 2008. Rural Land Use Change during 1986-2002 in Lijiang, China, Based on Remote Sensing and GIS Data. Sensors 8, 8201–8223. https://doi.org/10.3390/s8128201
260. Pfeifer M, Boyle M J. W, Dunning S, Olivier P I, 2019. Forest floor temperature and greenness link significantly to canopy attributes in South Africa’s fragmented coastal forests. PeerJ 7, e6190–e6190.
261. Picoli Michelle Cristina Araujo, Camara Gilberto, Sanches Ieda, Simões Rolf, Carvalho Alexandre, Maciel Adeline, Coutinho Alexandre, Esquerdo Julio, Antunes João, Begotti Rodrigo Anzolin, Arvor Damien, Almeida Claudio, 2018. Big earth observation time series analysis for monitoring Brazilian agriculture. ISPRS Journal of Photogrammetry and Remote Sensing 145, 328–339. https://doi.org/10.1016/j.isprsjprs.2018.08.007
262. Pillarisetti Ajay, Allen Tracy, Ruiz-Mercado Ilse, Edwards Rufus, Chowdhury Zohir, Garland Charity, Hill L Drew, Johnson Michael, Litton Charles D, Lam Nicholas L, Pennise David, Smith Kirk R, 2017. Small, Smart, Fast, and Cheap: Microchip-Based Sensors to Estimate Air Pollution Exposures in Rural Households. Sensors (Basel, Switzerland) 17, 1879. https://doi.org/10.3390/s17081879
263. Pindolia Deepa K, Garcia Andres J, Huang Zhuojie, Fik Timothy, Smith David L, Tatem Andrew J, 2014. Quantifying cross-border movements and migrations for guiding the strategic planning of malaria control and elimination. Malaria Journal 13, 169. https://doi.org/10.1186/1475-2875-13-169
264. Ping Liu, Jin Wang, Arun Kumar, Sangaiah, Yang Xie, Xinchun Yin, 2019. Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment. Sustainability 1.
265. Pokhriyal N, Jacques D C, 2017. Combining disparate data sources for improved poverty prediction and mapping. Proceedings of the National Academy of Sciences of the United States of America 114, E9783–E9792. https://doi.org/10.1073/pnas.1700319114
266. Potter Christopher, 2014a. Monitoring the production of Central California coastal rangelands using satellite remote sensing. Journal of Coastal Conservation (Springer Science & Business Media B.V.) 18, 213–220. https://doi.org/10.1007/s11852-014-0308-1
267. Potter Christopher, 2014b. Ten years of forest cover change in the Sierra Nevada detected using Landsat satellite image analysis. International Journal of Remote Sensing 35, 7136–7153. https://doi.org/10.1080/01431161.2014.968687
268. Quantifying socio-economic indicators in developing countries from mobile phone communication data: applications to Côte d’Ivoire, 2015. https://doi.org/10.1140/epjds/s13688-015-0053-1
269. R Zhang, H Su, J Tian, z Li, S Chen, J Zhan, X Deng, X Sun, J Wu, 2008. Drought Monitoring in Northern China based on Remote Sensing Data and Land Surface Modeling. Presented at the IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium, pp. III–860. https://doi.org/10.1109/IGARSS.2008.4779485
270. Rahman M M, Csaplovics E, Koch B, 2008. Satellite estimation of forest carbon using regression models. International Journal of Remote Sensing 29, 6917–6936.
271. Rangel Marcos A, Vogl Tom, 2016. Agricultural Fires and Infant Health. National Bureau of Economic Research Working Paper Series No. 22955. https://doi.org/10.3386/w22955
272. Ranvir Singh, 2005. Water productivity analysis from field to regional scale: integration of crop and soil modelling, remote sensing and geographical information. Wageningen; Netherlands.
273. Raucoules Daniel, Le Cozannet, Gonéri, Wöppelmann Guy, de Michele, Marcello, Gravelle Médéric, Daag Arturo, Marcos Marta, 2013. High nonlinear urban ground motion in Manila (Philippines) from 1993 to 2010 observed by DInSAR: Implications for sea-level measurement. Remote Sensing of Environment 139, 386–397. https://doi.org/10.1016/j.rse.2013.08.021
274. Reddy C S, Alekhya V V. L. P, Saranya K R. L, Athira K, Jha C S, Diwakar P G, Dadhwal V K, 2017. Monitoring of fire incidences in vegetation types and Protected Areas of India: implications on carbon emissions. Journal of Earth System Science 126, Article ID 11-Article ID 11.
275. Robin M, Chapuis J L, Lebouvier M, 2011. Remote sensing of vegetation cover change in islands of the Kerguelen archipelago. Polar Biology 34, 1689–1700.
276. Robinson Timothy, Emwanu Thomas, Rogers David, 2007. Environmental Approaches to Poverty Mapping: an example from Uganda. Information Development - INF DEV 23, 205–215. https://doi.org/10.1177/0266666907079077
277. Rodriguez Fernandez, M, Garcia A C, Alonso I G, Casanova E Z, 2016. Using the Big Data generated by the Smart Home to improve energy efficiency management. Energy Efficiency 9, 249–260. https://doi.org/10.1007/s12053-015-9361-3
278. Rogers David, Emwanu Thomas, Robinson Timothy, 2006. Poverty Mapping in Uganda: An Analysis Using Remotely Sensed and Other Environmental Data.
279. Rongfeng Yang, Yi Luo, Kun Yang, Liang Hong, Xiaolu Zhou, 2019. Analysis of Forest Deforestation and its Driving Factors in Myanmar from 1988 to 2017. Sustainability 1.
280. Roy K C, Cebrian M, Hasan S, 2019. Quantifying human mobility resilience to extreme events using geo-located social media data. Epj Data Science 8. https://doi.org/10.1140/epjds/s13688-019-0196-6
281. Rui YuHan, Fu DaFang, Ha Do, Minh, Radhakrishnan M, Zevenbergen C, Pathirana A, 2018. Urban surface water quality, flood water quality and human health impacts in Chinese cities. What do we know? Water 10, 240–240.
282. Ruktanonchai Nick W, DeLeenheer Patrick, Tatem Andrew J, Alegana Victor A, Caughlin T Trevor, zu Erbach-Schoenberg, Elisabeth, Lourenço Christopher, Ruktanonchai Corrine W, Smith David L, 2016. Identifying Malaria Transmission Foci for Elimination Using Human Mobility Data. PLOS Computational Biology 12, e1004846.
283. Ryan Engstrom Jonathan Hersh David Newhouse, 2017. Poverty from Space: Using High-Resolution Satellite Imagery for Estimating Economic Well-Being, Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-8284
284. Saha Mithun, Eckelman Matthew J, 2017. Growing fresh fruits and vegetables in an urban landscape: A geospatial assessment of ground level and rooftop urban agriculture potential in Boston, USA. Landscape and Urban Planning 165, 130–141. https://doi.org/10.1016/j.landurbplan.2017.04.015
285. Salhab J, Wang JianLi, Anjum S A, Chen YunXuan, 2010. Assessment of the grassland degradation in the southeastern part of the source region of the Yellow River from 1994 to 2001. Journal of Food, Agriculture & Environment 8, 1367–1372.
286. Sanga-Ngoie K, Iizuka K, Kobayashi S, 2012. Estimating CO2 sequestration by forests in Oita Prefecture, Japan, by combining LANDSAT ETM+ and ALOS satellite remote sensing data. Remote Sensing 4, 3544–3570.
287. Sani N A, Kafaky S B, Pukkala T, Mataji A, 2016. Integrated use of GIS, remote sensing and multi-criteria decision analysis to assess ecological land suitability in multi-functional forestry. Journal of Forestry Research 27, 1127–1135.
288. Santa Man, Rai, Bishal Nath, Upreti, Subodh Dhakal, Tara Nidhi, Bhattarai, Basanta Raj, Adhikari, Samjwal Ratna, Bajracharya, Masaru Yoshida, 2017. Climate Change Impact on Glacier Retreat and Local Community in the Langtang Valley, Central Nepal. Journal of Development Innovations 45.
289. Sayema Jamal, Akram Javed, Yousuf Khanday, 2016. Evaluation of land degradation and so-cio-environmental issues: a case study of semi arid watershed in Western Rajasthan. Journal of Environmental Protection 7, 1132–1147.
290. Schneider A, Mertes C M, Tatem A J, Tan B, Sulla-Menashe D, Graves S J, Patel N N, Horton J A, Gaughan A E, Rollo J T, Schelly I H, Stevens F R, Dastur A, 2015. A new urban landscape in East–Southeast Asia, 2000–2010. Environmental Research Letters 10, 034002. https://doi.org/10.1088/1748-9326/10/3/034002
291. Scholte R G. C, Freitas C C, Dutra L V, Guimaraes R J. P. S, Drummond S C, Oliveira G, Carvalho O S, 2012. Utilizing environmental, socioeconomic data and GIS techniques to estimate the risk for ascariasis and trichuriasis in Minas Gerais, Brazil. Acta Tropica 121, 112–117.
292. Sedda Luigi, Tatem Andrew J, Morley David W, Atkinson Peter M, Wardrop Nicola A, Pezzulo Carla, Sorichetta Alessandro, Kuleszo Joanna, Rogers David J, 2015. Poverty, health and satellite-derived vegetation indices: their inter-spatial relationship in West Africa. International Health 7, 99–106. https://doi.org/10.1093/inthealth/ihv005
293. See L, McCallum I, Fritz S, Perger C, Kraxner F, Obersteiner M, Baruah U D, Nitashree Mili, Kalita N R, 2013. Mapping cropland in Ethiopia using crowdsourcing. International Journal of Geosciences 4, 6–13.
294. Seitz Nicole E, Westbrook Cherie J, Dubé Monique G, Squires Allison J, 2013. Assessing large spatial scale landscape change effects on water quality and quantity response in the lower Athabasca River basin. Integrated Environmental Assessment & Management 9, 392–404. https://doi.org/10.1002/ieam.1336
295. Shadumyan Hartutyun, 2006. Application of geographic information systems(GIS) Armenia: opportunities for expanded use in the health sphere.
296. Shailendra Singh, Abdulsalam Yassine, 2018. Big Data Mining of Energy Time Series for Behavioral Analytics and Energy Consumption Forecasting. Energies 1.
297. Shanableh A, Al-Ruzouq R, Yilmaz A G, Siddique M, Merabtene T, Imteaz M A, 2018. Effects of land cover change on urban floods and rainwater harvesting: a case study in Sharjah, UAE. Water 10, 631–631.
298. Shomar B, Fakher S A, Yahya A, 2010. Assessment of groundwater quality in the Gaza Strip, Palestine using GIS mapping. Journal of Water Resource and Protection 2, 93–104.
299. Shukla Anugya, Jain Kamal, 2019. Critical analysis of rural-urban transitions and transformations in Lucknow city, India. Remote Sensing Applications: Society and Environment 13, 445–456. https://doi.org/10.1016/j.rsase.2019.01.001
300. Shyamantha Subasinghe, Estoque R C, Murayama Y, 2016. Spatiotemporal analysis of urban growth using GIS and remote sensing: a case study of the Colombo Metropolitan Area, Sri Lanka. ISPRS International Journal of Geo-Information 5, 197–197.
301. Siddiqui Z, 2011. Holistic approach to mitigate the pollution impacts in the coastal ecosystem of Thailand using the remote sensing techniques. International Journal of Environmental Research 5, 297–306.
302. Siyal Altaf A, Misrani Dhanji M, Dars Ghulam H, Ahmad Sajjad, Kamojjala S, 2018. Application of GIS and Remote Sensing for Identification of Potential Runoff Harvesting Sites: A Case Study of Karoonjhar Mountainous Area, Pakistan.
303. Sofeska Emilija, 2017. Understanding the Livability in a City Through Smart Solutions and Urban Planning Toward Developing Sustainable Livable Future of the City of Skopje. Procedia Environmental Sciences 37, 442–453. https://doi.org/10.1016/j.proenv.2017.03.014
304. Sommarat Chantarat, Preesan Rakwatin, Chutatong Charumilind, 2017. Farmers and Pixels: Toward Sustainable Agricultural Finance with Space Technology.
305. Song Jinchao, Zhao Chunli, Lin Tao, Li Xinhu, Prishchepov Alexander V, 2019. Spatio-temporal patterns of traffic-related air pollutant emissions in different urban functional zones estimated by real-time video and deep learning technique. Journal of Cleaner Production 238. https://doi.org/10.1016/j.jclepro.2019.117881
306. Soytong P, Janchidfa K, Phengphit N, Chayhard S, Perera R, 2016. The effects of land use change and climate change on water resources in the eastern region of Thailand. International Journal of Agricultural Technology 12, 1697–1724.
307. Steele Jessica E, Sundsøy Pål Roe, Pezzulo Carla, Alegana Victor A, Bird Tomas J, Blumenstock Joshua, Bjelland Johannes, Engø-Monsen Kenth, de Montjoye, Yves-Alexandre, Iqbal Asif M, Hadiuzzaman Khandakar N, Lu Xin, Wetter Erik, Tatem Andrew J, Bengtsson Linus, 2017. Mapping poverty using mobile phone and satellite data. Journal of The Royal Society Interface 14, 20160690. https://doi.org/10.1098/rsif.2016.0690
308. Stein C, Ernstson H, Barron J, 2011. A social network approach to analyzing water governance: The case of the Mkindo catchment, Tanzania. Physics and Chemistry of the Earth 36, 1085–1092. https://doi.org/10.1016/j.pce.2011.07.083
309. Stergiadou A, Libello D, Cavalli R, Krč J, 2009. Estimating forest harvesting operations to achieve sustainable rural development in Samarina (Greece). Folia Forestalia Polonica. Seria A, Leśnictwo 51, 21–28.
310. Stevens Forrest R, Gaughan Andrea E, Linard Catherine, Tatem Andrew J, 2015. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. PLOS ONE 10, e0107042.
311. Sturrock Hugh JW, Cohen Justin M, Keil Petr, Tatem Andrew J, Le Menach, Arnaud, Ntshalintshali Nyasatu E, Hsiang Michelle S, Gosling Roland D, 2014. Fine-scale malaria risk mapping from routine aggregated case data. Malaria Journal 13, 421. https://doi.org/10.1186/1475-2875-13-421
312. Sujan Ghimire, Ravinesh C Deo, Nawin Raj, Jianchun Mi, 2019. Deep Learning Neural Networks Trained with MODIS Satellite-Derived Predictors for Long-Term Global Solar Radiation Prediction. Energies 1.
313. Sundsøy Pål, 2016. Can mobile usage predict illiteracy in a developing country? arXiv e-prints arXiv:1607.01337.
314. Sutton Paul, Elvidge Christopher, Tilottama Ghosh, 2007. Estimation of Gross Domestic Product at Sub-National Scales Using Nighttime Satellite Imagery. International Journal of Ecological Economics & Statistics 8.
315. Tanaka Kiyoyasu, Keola Souknilanh, 2017. Shedding Light on the Shadow Economy: A Nighttime Light Approach. The Journal of Development Studies 53, 32–48. https://doi.org/10.1080/00220388.2016.1171845
316. Taramelli A, Valentini E, Starlacchini S, 2015. A GIS-based approach for hurricane hazard and vulnerability assessment in the Cayman Islands.
317. Tatem Andrew J, Huang Zhuojie, Narib Clothilde, Kumar Udayan, Kandula Deepika, Pindolia Deepa K, Smith David L, Cohen Justin M, Graupe Bonita, Uusiku Petrina, Lourenço Christopher, 2014. Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. Malaria Journal 13, 52. https://doi.org/10.1186/1475-2875-13-52
318. Tatem Andrew J, Qiu Youliang, Smith David L, Sabot Oliver, Ali Abdullah S, Moonen Bruno, 2009. The use of mobile phone data for the estimation of the travel patterns and imported Plasmodium falciparum rates among Zanzibar residents. Malaria Journal 8, 287. https://doi.org/10.1186/1475-2875-8-287
319. Thomson R Dana, Kools Lieke, Jochem C Warren, 2018. Linking Synthetic Populations to Household Geolocations: A Demonstration in Namibia. Data 3. https://doi.org/10.3390/data3030030
320. Thoralf Gutierrez, Gautier Krings, Vincent D Blondel, 2013. Evaluating socio-economic state of a country analyzing airtime credit and mobile phone datasets.
321. Tom Bundervoet Laban Maiyo Apurva Sanghi, 2015. Bright Lights, Big Cities: Measuring National and Subnational Economic Growth in Africa from Outer Space, with an Application to Kenya and Rwanda, Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-7461
322. Toole Jameson L, Lin Yu-Ru, Muehlegger Erich, Shoag Daniel, González Marta C, Lazer David, 2015. Tracking employment shocks using mobile phone data. Journal of the Royal Society, Interface 12. https://doi.org/10.1098/rsif.2015.0185
323. Tracey Li, Jesper Dejby, Maximilian Albert, Linus Bengtsson, Veronique Lefebvre, 2019a. Detecting individual internal displacements following a sudden-onset disaster using time series analysis of call detail records.
324. Tracey Li, Jesper Dejby, Maximilian Albert, Linus Bengtsson, Veronique Lefebvre, 2019b. Estimating the resilience to natural disasters by using call detail records to analyse the mobility of internally displaced persons.
325. Trigg S N, Curran L M, McDonald A K, 2006. Utility of Landsat 7 satellite data for continued monitoring of forest cover change in protected areas in Southeast Asia. Singapore Journal of Tropical Geography 27, 49–66. https://doi.org/10.1111/j.1467-9493.2006.00239.x
326. Trinh Le, Hung, Vu Danh, Tuyen, Do Nhu, Hiep, 2015. Evaluation of soil erosion risk using remote sensing and GIS data (a case study: Lang Chanh district, Thanh Hoa province, Vietnam). Vestnik OrelGAU 57–64.
327. Uddin Kabir, Murthy MSR, Wahid Shahriar M, Matin Mir A, 2016. Estimation of soil erosion dynamics in the Koshi Basin using GIS and remote sensing to assess priority areas for conservation.
328. USAID, 2018. REAL-TIME MONITORING FOR IMPROVED WATER SERVICES IN THE ETHIOPIAN LOWLANDS.
329. Using Global Positioning System Travel Data to Assess Real-World Energy Use of Plug-In Hybrid Electric Vehicles, 2007. https://doi.org/10.3141/2017-04
330. Utazi C Edson, Thorley Julia, Alegana Victor A, Ferrari Matthew J, Takahashi Saki, Metcalf C Jessica E, Lessler Justin, Tatem Andrew J, 2018. High resolution age-structured mapping of childhood vaccination coverage in low and middle income countries. Vaccine 36, 1583–1591. https://doi.org/10.1016/j.vaccine.2018.02.020
331. Utz Pape, Luca Parisotto, Véronique Lefebvre, Sarchil Qader, Amy Ninneman, Tom Bird, 2019a. Estimating poverty in a fragile context: the high frequency survey in South Sudan.
332. Utz Pape, Philip Wollburg, Veronique Lefebvre, Sarchil Qader, Amy Ninneman, Dana Thomson, Tom Bird, 2019b. Estimation of poverty in Somalia using innovative methodologies.
333. V Frias-Martinez, V Soto, H Hohwald, E Frias-Martinez, 2012. Characterizing Urban Landscapes Using Geolocated Tweets. Presented at the 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing, pp. 239–248. https://doi.org/10.1109/SocialCom-PASSAT.2012.19
334. Valjarević Aleksandar, Djekić Tatjana, Stevanović Vladica, Ivanović Radomir, Jandziković Bojana, 2018. GIS numerical and remote sensing analyses of forest changes in the Toplica region for the period of 1953–2013. Applied Geography 92, 131–139. https://doi.org/10.1016/j.apgeog.2018.01.016
335. van Beijma, Sybrand, Chatterton Julia, Page Susan, Rawlings Chris, Tiffin Richard, King Henry, 2018. The challenges of using satellite data sets to assess historical land use change and associated greenhouse gas emissions: a case study of three Indonesian provinces. Carbon Management 9, 399–413. https://doi.org/10.1080/17583004.2018.1511383
336. Varshney Kush R, Chen George H, Abelson Brian, Nowocin Kendall, Sakhrani Vivek, Xu Ling, Spatocco Brian L, 2015. Targeting Villages for Rural Development Using Satellite Image Analysis. Big Data 3, 41–53. https://doi.org/10.1089/big.2014.0061
337. Vogel Kathryn Baragwanath, Goldblatt Ran, Hanson Gordon H, Khandelwal Amit K, 2018. Detecting Urban Markets with Satellite Imagery: An Application to India. National Bureau of Economic Research Working Paper Series No. 24796. https://doi.org/10.3386/w24796
338. W Xu, T Zheng, Z Li, 2011. A Neural Network Based Forecasting Method For the Unemployment Rate Prediction Using the Search Engine Query Data. Presented at the 2011 IEEE 8th International Conference on e-Business Engineering, pp. 9–15. https://doi.org/10.1109/ICEBE.2011.21
339. Wahyu Andy, Nugraha, Zainul Hidayah, Insafitri, 2013. Developing Coral Reef Conservation Zones in the Kangean Archipelago, Indonesia.
340. Wang Anna, Tran Caelin, Desai Nikhil, Lobell David, Ermon Stefano, 2018. Deep Transfer Learning for Crop Yield Prediction with Remote Sensing Data. https://doi.org/10.1145/3209811.3212707
341. Wang L Y, Fan H, Wang Y K, 2019. Fine-Resolution Population Mapping from International Space Station Nighttime Photography and Multisource Social Sensing Data Based on Similarity Matching. Remote Sensing 11. https://doi.org/10.3390/rs11161900
342. Wang Sherrie, Azzari George, Lobell David B, 2019. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. Remote Sensing of Environment 222, 303–317. https://doi.org/10.1016/j.rse.2018.12.026
343. Wang Sherry, Chen William, Xie San Michael, Azzari George, Lobell David B, 2020. Weakly supervised deep learning for segmentation of remote sensing imgery. Remote Sensing. https://doi.org/10.3390/rs12020207
344. Weinzierl T, Wehberg J, Böhner J, Conrad O, 2016. Spatial assessment of land degradation risk for the Okavango River catchment, southern Africa. Land Degradation & Development 27, 281–294.
345. Weng Qihao, Fu Peng, Gao Feng, 2014. Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data. Remote Sensing of Environment 145, 55–67. https://doi.org/10.1016/j.rse.2014.02.003
346. Wesolowski Amy, Buckee Caroline O, Bengtsson Linus, Wetter Erik, Lu Xin, Tatem Andrew J, 2014a. Commentary: containing the ebola outbreak - the potential and challenge of mobile network data. PLoS currents 6, ecurrents.outbreaks.0177e7fcf52217b8b634376e2f3efc. https://doi.org/10.1371/currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e
347. Wesolowski Amy, Buckee Caroline O, Pindolia Deepa K, Eagle Nathan, Smith David L, Garcia Andres J, Tatem Andrew J, 2013a. The Use of Census Migration Data to Approximate Human Movement Patterns across Temporal Scales. PLOS ONE 8, e52971.
348. Wesolowski Amy, Eagle Nathan, Noor Abdisalan M, Snow Robert W, Buckee Caroline O, 2013b. The impact of biases in mobile phone ownership on estimates of human mobility. Journal of The Royal Society Interface 10, 20120986. https://doi.org/10.1098/rsif.2012.0986
349. Wesolowski Amy, Eagle Nathan, Tatem Andrew J, Smith David L, Noor Abdisalan M, Snow Robert W, Buckee Caroline O, 2012. Quantifying the Impact of Human Mobility on Malaria. Science 338, 267. https://doi.org/10.1126/science.1223467
350. Wesolowski Amy, Metcalf C J. E, Eagle Nathan, Kombich Janeth, Grenfell Bryan T, Bjørnstad Ottar N, Lessler Justin, Tatem Andrew J, Buckee Caroline O, 2015. Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. Proceedings of the National Academy of Sciences 112, 11114. https://doi.org/10.1073/pnas.1423542112
351. Wesolowski Amy, Stresman Gillian, Eagle Nathan, Stevenson Jennifer, Owaga Chrispin, Marube Elizabeth, Bousema Teun, Drakeley Christopher, Cox Jonathan, Buckee Caroline O, 2014b. Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phones. Scientific Reports 4, 5678. https://doi.org/10.1038/srep05678
352. Wey Wann-Ming, Huang Jhong-You, 2018. Urban sustainable transportation planning strategies for livable City’s quality of life. Habitat International 82, 9–27. https://doi.org/10.1016/j.habitatint.2018.10.002
353. Wilson R, zu Erbach-Schoenberg E, Albert M, Power D, Tudge S, Gonzalez M, Guthrie S, Chamberlain H, Brooks C, Hughes C, Pitonakova L, Buckee C, Lu X, Wetter E, Tatem A, Bengtsson L, 2016. Rapid and Near Real-Time Assessments of Population Displacement Using Mobile Phone Data Following Disasters: The 2015 Nepal Earthquake.
354. Witmer Frank, O’Loughlin John, 2011. Detecting the Effects of Wars in the Caucasus Regions of Russia and Georgia Using Radiometrically Normalized DMSP-OLS Nighttime Lights Imagery. GIScience & Remote Sensing 48, 478–500. https://doi.org/10.2747/1548-1603.48.4.478
355. Wong Man Sing, Zhu Rui, Liu Zhizhao, Lu Lin, Peng Jinqing, Tang Zhaoqin, Lo Chung Ho, Chan Wai Ki, 2016. Estimation of Hong Kong’s solar energy potential using GIS and remote sensing technologies. Renewable Energy: An International Journal 99, 325–335. https://doi.org/10.1016/j.renene.2016.07.003
356. Wu W, Ren H Y, Yu M, Wang Z, 2018. Distinct Influences of Urban Villages on Urban Heat Islands: A Case Study in the Pearl River Delta, China. International Journal of Environmental Research and Public Health 15. https://doi.org/10.3390/ijerph15081666
357. Xia T Q, Song X, Zhang H R, Song X Y, Kanasugi H, Shibasaki R, 2019. Measuring spatio-temporal accessibility to emergency medical services through big GPS data. Health & Place 56, 53–62. https://doi.org/10.1016/j.healthplace.2019.01.012
358. Xiaoling Chen, Hong Song, Li Y S, Zhongzhi Yuan, Wai Onyx W. H, Zhilin Li, Zhanhua Xu, Baojun Zhang, 2006. Spatial analysis of water quality using a combination of GIS and statistical software. International Journal of Environment & Pollution 28, 274–296. https://doi.org/10.1504/IJEP.2006.011212
359. Xie Hualin, He Yafen, Xie Xue, 2017. Exploring the factors influencing ecological land change for China’s Beijing–Tianjin–Hebei Region using big data. Journal of Cleaner Production 142, 677–687. https://doi.org/10.1016/j.jclepro.2016.03.064
360. Xu QinNeng, Gel Y R, Ramirez Ramirez, L L, Nezafati K, Zhang QingPeng, Tsui KwokLeung, 2017. Forecasting influenza in Hong Kong with Google search queries and statistical model fusion. PLoS ONE 12, e0176690–e0176690.
361. Xu Yanyan, Jiang Shan, Li Ruiqi, Zhang Jiang, Zhao Jinhua, Abbar Sofiane, González Marta C, 2019. Unraveling environmental justice in ambient PM2.5 exposure in Beijing: A big data approach. Computers, Environment and Urban Systems 75, 12–21. https://doi.org/10.1016/j.compenvurbsys.2018.12.006
362. Yang Shihao, Santillana Mauricio, Kou S C, 2015. Accurate estimation of influenza epidemics using Google search data via ARGO. Proceedings of the National Academy of Sciences 112, 14473. https://doi.org/10.1073/pnas.1515373112
363. Yang XiHua, Smith P L, Yu Tao, Gao HaiLiang, 2011. Estimating evapotranspiration from terrestrial groundwater-dependent ecosystems using Landsat images. International Journal of Digital Earth 4, 154–170.
364. Yang Yun, Anderson M C, Gao Feng, Hain C R, Semmens K A, Kustas W P, Noormets A, Wynne R H, Thomas V A, Sun Ge, 2017. Daily landsat-scale evapotranspiration estimation over a forested landscape in North Carolina, USA, using multi-satellite data fusion. Hydrology and Earth System Sciences 21, 1017–1037.
365. Yangjian Zhang, Jian Yang, He Hong S, Dijak William D, Shifley Stephen R, Palik Brian J, 2009. Integration of Satellite Imagery and Forest Inventory in Mapping Dominant and Associated Species at a Regional Scale. Environmental Management 44, 312–323. https://doi.org/10.1007/s00267-009-9307-7
366. Yin Jie, Yin Zhane, Zhong Haidong, Xu Shiyuan, Hu Xiaomeng, Wang Jun, Wu Jianping, 2011. Monitoring urban expansion and land use/land cover changes of Shanghai metropolitan area during the transitional economy (1979–2009) in China. Environmental Monitoring and Assessment 177, 609–621. https://doi.org/10.1007/s10661-010-1660-8
367. Yoshiki Yamagata, Takahiro Yoshida, Daisuke Murakami, Tomoko Matsui, Yuki Akiyama, 2018. Seasonal Urban Carbon Emission Estimation Using Spatial Micro Big Data. Sustainability 1.
368. Zabala Aiora, 2013. Comparing Global Spatial Data on Deforestation for Institutional Analysis in Africa. Reference Module in Earth Systems and Environmental Sciences. https://doi.org/10.1016/B978-0-12-409548-9.09681-0
369. Zafar S, Zaidi A, 2019. Impact of urbanization on basin hydrology: a case study of the Malir Basin, Karachi, Pakistan. Regional Environmental Change 19, 1815–1827. https://doi.org/10.1007/s10113-019-01512-9
370. Zagatti Guilherme Augusto, Gonzalez Miguel, Avner Paolo, Lozano-Gracia Nancy, Brooks Christopher J, Albert Maximilian, Gray Jonathan, Antos Sarah Elizabeth, Burci Priya, zu Erbach-Schoenberg, Elisabeth, Tatem Andrew J, Wetter Erik, Bengtsson Linus, 2018. A trip to work: Estimation of origin and destination of commuting patterns in the main metropolitan regions of Haiti using CDR. Development Engineering 3, 133–165. https://doi.org/10.1016/j.deveng.2018.03.002
371. Zhang ChangZhen, Cai JiaHao, Xiao DeQin, Ye YaoWen, Chehelamirani M, 2018. Research on vegetable pest warning system based on multidimensional big data. Insects 9, 66–66.
372. Zhang Haoran, Song Xuan, Long Yin, Xia Tianqi, Fang Kai, Zheng Jianqin, Huang Dou, Shibasaki Ryosuke, Liang Yongtu, 2019. Mobile phone GPS data in urban bicycle-sharing: Layout optimization and emissions reduction analysis. Applied Energy 138.
373. Zhang JiaHua, Yao FengMei, Liu Cheng, Yang L M, Boken V K, 2011. Detection, emission estimation and risk prediction of forest fires in China using satellite sensors and simulation models in the past three decades - an overview. International Journal of Environmental Research and Public Health 8, 3156–3178.
374. Zhang P C, Zhao Q, Gao J, Li W R, Lu J M, 2019. Urban Street Cleanliness Assessment Using Mobile Edge Computing and Deep Learning. Ieee Access 7, 63550–63563. https://doi.org/10.1109/access.2019.2914270
375. Zhang Xuelei, Chen Jie, Tan Manzhi, Sun Yanci, 2007. Assessing the impact of urban sprawl on soil resources of Nanjing city using satellite images and digital soil databases. Catena 69, 16–30. https://doi.org/10.1016/j.catena.2006.04.020
376. Zhang Xun, Fu JingYing, Lin Gang, Jiang Dong, Yan XiaoXi, 2017. Switchgrass-based bioethanol productivity and potential environmental impact from marginal lands in China. Energies 10, 260–260.
377. Zheng Hang, Hong Yang, Long Di, Jing Hua, 2017. Monitoring surface water quality using social media in the context of citizen science. Hydrology and Earth System Sciences 21, 949–961.
378. Zheng Xiao, Zhu Jiaojun, Yan Qiaoling, 2013. Monthly Air Temperatures over Northern China Estimated by Integrating MODIS Data with GIS Techniques. Journal of Applied Meteorology and Climatology 52, 1987.
379. Zonghao Rao, Dongyuan Yang, Zhengyu Duan, 2013. Resident Mobility Analysis Based on Mobile-phone Billing Data. Procedia - Social and Behavioral Sciences 96, 2032–2041. https://doi.org/10.1016/j.sbspro.2013.08.229
380. Zope P E, Eldho T I, Jothiprakash V, 2017. Hydrological impacts of land use–land cover change and detention basins on urban flood hazard: a case study of Poisar River basin, Mumbai, India. Natural Hazards: Journal of the International Society for the Prevention and Mitigation of Natural Hazards 1267. https://doi.org/10.1007/s11069-017-2816-4
381. Zu Erbach-Schoenberg, Elisabeth, Alegana Victor A, Sorichetta Alessandro, Linard Catherine, Lourenço Christoper, Ruktanonchai Nick W, Graupe Bonita, Bird Tomas J, Pezzulo Carla, Wesolowski Amy, Tatem Andrew J, 2016. Dynamic denominators: the impact of seasonally varying population numbers on disease incidence estimates. Population Health Metrics 14, 35. https://doi.org/10.1186/s12963-016-0106-0

1. The gap in data availability at the country level is partially driven by a lack of resources, limited capacity within governments and logistical difficulties in collecting the data. For example, the total cost of collecting data on all the 169 SDG targets was estimated to be around USD 254 billion, which is about 12.5 per cent of total official development assistance (ODA) to be committed for the post-2015 period (Jerven, 2014). A recent UN survey shows that there is existing capacity to collect data on only 40 SDG indicators and data sources for another 47 indicators are available in principle. There is little capacity and resources for collecting data on the remaining indicators (UN, 2018). [↑](#footnote-ref-2)
2. For the purpose of this report, big data measurement studies are defined as the studies that have innovatively used big data to measure and validate any development outcome such as poverty measurement, crop productivity, employment, mobility, forest cover, etc. These are not impact evaluations, but can inform future evaluations. [↑](#footnote-ref-3)
3. 3ie (2012) Impact Evaluation Glossary. Available at <https://www.3ieimpact.org/sites/default/files/2018-07/impact_evaluation_glossary_-_july_2012_3.pdf> [↑](#footnote-ref-4)
4. The online map can be accessed here: <https://gapmaps.3ieimpact.org/evidence-maps/big-data-systematic-map> [↑](#footnote-ref-5)
5. There are also other classifications of big data based on its structure: structured, semi-structured and unstructured data. Structured data is in the standard columns and rows form such as what sensors provide; semi-structured data contains free texts, but can be tagged such as Twitter; and the unstructured data include images and videos (Desouza and Jacob, 2017). Big data can also be classified as open and proprietary data (Maaroof, 2015). [↑](#footnote-ref-6)
6. Our map refers to SDGs since they are globally acknowledged and fairly inclusive list of development outcomes. All the studies relevant to this map are expected to fall under at least one of the 169 targets listed within SDGs. [↑](#footnote-ref-7)
7. See Rathinam, et.al (2019a) for more details on single screening with safety first approach. [↑](#footnote-ref-8)
8. <http://geo.aiddata.org/query/> [↑](#footnote-ref-9)
9. <https://geodata.grid.unep.ch/> [↑](#footnote-ref-10)
10. <https://sedac.ciesin.columbia.edu/> [↑](#footnote-ref-11)
11. It should be noted that the cost discussed here includes only the variable cost of data collection and the staff time but may not include the cost of fixed infrastructure and equipment. [↑](#footnote-ref-12)
12. There are several ways to collect the required satellite data. For a few key variables such as night lights, air pollution, land cover and water quality, elevation, slope, distance from certain services or infrastructure, geocoded data for various granularity and frequency is readily available in several data bases such as Aiddata DataQuary, SEDAC, etc. This study has utilized data from differences sources that provide readily useable data. Alternatively, the researchers use machine learning techniques to analyse satellite images and predict development outcomes (Jean et.al. 2016). Another useful source of areal images come from custom built drones that can provide very high resolution data for the exact spatial and temporal frequency (Pellegrini, 2019). [↑](#footnote-ref-13)
13. Flowminder’s original ‘data partnership’ model developed lasting collaborations with individual MNOs. The priority was the countries where substatial potential gains were available from novel, digital data, given the existing data landscape. Lengthy negotiations with MNOs followed, often spanning many years, and consuming extensive organisational resources, with no guarantee of a successful outcome. When this model ‘worked’, it led to strong and sustained partnerships with MNOs. In pursuit of impact at scale, Flowminder Foundation supplemented its original ‘data partnership’ model with a toolkit-based approach designed to break down silos between data and methods, in effect negating the need for Flowminder to access MNO data by transferring methods expertise to MNOs themselves via the ‘Flowkit’ suite of software. [↑](#footnote-ref-14)