



Building capacities on small area estimation for SDGs: A Toolkit & an eLearning Course

Yongyi Min, Chief, Global SDG Monitoring Section, UNSD

Haoyi Chen, Coordinator, Inter-Secretariat Working Group on Household Surveys

Andres Gutierrez Rojas, Regional Advisor, UN-ECLAC Statistics Division

Improve national capacity on using SAE for regular production

SAE Toolkit for SDGs, UNSD working with experts on SAE, under the guidance of the IAEG-SDGs and ISWGHS

- Using SAE methods to improve SDG data availability for vulnerable population groups – requested by IAEG-SDGs
- Offering practical guidance and country case studies
- Guiding on the enabling environment for using SAE for official data production
- Providing a space for partners to document and disseminate their SAE methodologies: transparency

eLearning on SAE, UNSD in collaboration with UN-ECLAC & UNFPA

- a set of eLearning courses on small area estimation for SDGs data disaggregation
- Organized in 10 modules
- For self-paced and guided learners

Work modality

- A group of experts providing guidance
- Wiki-platform

UN Statistics Wiki Spaces People Analytics Cockpit Create ... Search

SAE4SDG 440 views

SAE4SDG

Created by UNSD Clarence Lio, last modified by Haoyi Chen on Mar 03, 2021

Welcome to the Toolkit for using Small Area Estimation for the SDGs!

In committing to the realization of the 2030 Agenda for Sustainable Development, Member States recognized that the dignity of the individuals is fundamental and that the Agenda's Goals and targets should be met for all nations and people and for all segments of society. Ensuring that these commitments are translated into effective action requires a precise understanding of the target populations and progress made in addressing their particular priorities.

To properly measure this, statistics need to be presented for different population groups and geographical areas. The Sustainable Development Goal (SDG) indicator framework has included an overarching principle of data disaggregation: SDG indicators should be disaggregated, where relevant, by income, sex, age, race, ethnicity, migratory status, disability and geographic location, or other characteristics, in accordance with the Fundamental Principles of Official Statistics.

To enable national statistical offices to estimate disaggregated indicators, guidelines are needed to support the process. The idea of writing guidelines on how to use statistical methods and, in particular small area estimation (SAE), to receive disaggregated statistical indicators is not new. Some focus on methodological aspects, others provide methodology in a specific program language or focus on a specific topic as poverty mapping. Several statistical institutions conducted projects on the evaluation of the usability of SAE for official statistics. In 2020, the Asian Development Bank even published practical guidelines especially focusing on the monitoring process of the Sustainable Development Goal with SAE. So how do these guidelines differ from the existing work?

The idea of the **SAE4SDG Toolkit** in Wiki is to complement and use the existing methodological work and case studies to encourage and enable national statistical offices

Space tools

Grid of 24 colored squares (3x8) and Sustainable Development Goals logo.

SAE by SDGs

Created by Ann-Kristin Kreutzmann, last modified by Haoyi Chen just a moment ago

This page gives a small guide on how to start a SAE case study and collects case studies for the Sustainable Development Goals. Case studies are not available for all SDGs yet, but more cases can be added continuously.

- [How to start a SAE case study](#)
- [Case studies for the estimation of disaggregated SDG indicators](#)
- [Case study submission](#)
- [References](#)

How to start a SAE case study

The guidelines give an overview of literature, available software and basic topics in small area estimation. While the concrete specification of a case study will vary among different applications, the same questions need to be answered.

	Questions
User needs <ul style="list-style-type: none"> • Goal • Indicator of interest • Disaggregation level 	<ul style="list-style-type: none"> • What are the key policies or funding decisions? • What questions need to be answered? • What are you trying to measure? • What type of indicator is the indicator of interest? • What is the relevant dimension of disaggregation?
Data availability	<ul style="list-style-type: none"> • Which survey data is available for the estimation of the indicator? • What are the data challenges? • Which additional data sources can be used?
SAE methods/Specification	<ul style="list-style-type: none"> • Which SAE approach can be used based on the inputs above? • Which approaches are available in statistical software? • What are the available expertise to do the computation, analysis and interpretation?
Model validation	<ul style="list-style-type: none"> • What is the plan for data validation? • What data sources for benchmarking? • Any plans for external review process?
Model refining	<ul style="list-style-type: none"> • Plan to refine the model
Extend case study to official production	<ul style="list-style-type: none"> • Plan or a roadmap to extend the case study for official data production?

Case studies for the estimation of disaggregated SDG indicators

In the following, case studies are summarized for the different SDGs. The descriptions are short and usually refer to a publicly available longer description of the study. The tables sum up user needs (indicators and disaggregation level), data availability and the specified estimation approach. The idea is to learn from other cases since some problems occur in different applications and thus for some problems, solutions may be found in another application.

Goal 1. End poverty in all its forms everywhere

[> Case studies](#)

SAE for official statistics

Dashboard / SAE4SDG   29 views

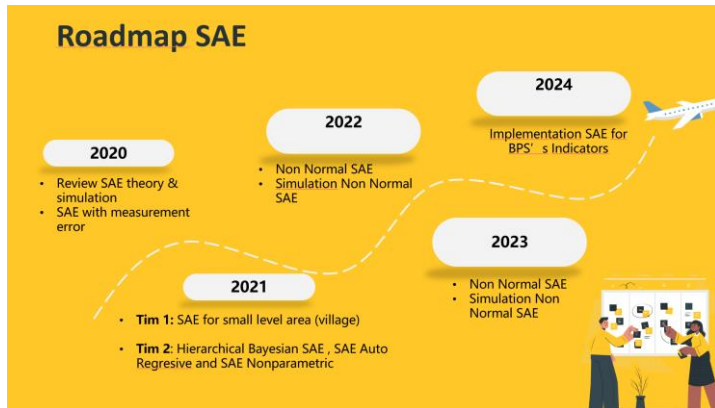
From SAE experiment to production: the enabling environment

Created by Haoyi Chen, last modified on Apr 21, 2021

Small area estimation has been in the field for many years but using it for official data production is still uncommon. It is important to understand the underlying reasons for the slow onset of SAE in the official data arena and identify "non-tech" areas that should be emphasized as creating an "enabling environment" for small area estimation.

- Challenges in using SAE for official data production
- Enabling environment to enable the use of SAE for official data production
 - Establishing a clear and focused objective that links SAE to data use for policymaking
 - Fostering an environment for research and development
 - Government commitment and sustainable financial support to SAE experimentation and production
 - Design-based versus model-based estimates: a changing culture in the national statistical offices
 - Usable input data for SAE
 - [Maintaining a high and fit-for-purpose quality standard](#)
 - Collaboration
 - Capacity building
 - Disclosure control
 - Transparency in releasing methodology and communicating quality
- Practical way forward: from experimental statistics to official statistics

SAE for official statistics – national examples



Example: United States SAIPE Program

In September 1994, the Congress passed the Improving America's Schools Act and signed it into law (PL 103-382). Title I of the law specifies the distribution of Federal funds to school districts based largely on "the number of children aged 5 to 17, inclusive, from families below the poverty level on the basis of the most recent satisfactory data, ..., available from the Department of Commerce."

This law further requires that in Fiscal Year 1997, the Secretary of Education use updated data on poor children for counties and, beginning in Fiscal Year 1999, updated data for school districts, published by the Department of Commerce, unless the Secretaries of Education and Commerce determine that the use of updated population data would be "inappropriate or unreliable."

It also directs the Secretary of Education to fund a National Academy of Sciences panel to provide advice on the suitability of the Census Bureau estimates for use in allocating funds.

Source: [Small Area Income and Poverty Estimates \(SAIPE\) Program, Origins of the Project](#)

Challenges in using SAE for official data production From National Statistical Offices

- "We did an experiment using small area estimation method for poverty but the results were not consistent with our own estimates so we did not pursue it again."
- "We do not have good input data source for SAE - census data are outdated and administrative data sources do not have good coverage and are lack of proper auxiliary variables."
- "SAE method is complicated and we are not comfortable with independently developing the method"
- "It is very difficult to convince the managers to use model-based estimates."

Model-based estimates at Statistics Netherlands

In a more recent paper from Statistics Netherlands (Buelens, Wolf and Zeelenberg, 2016), a set of guidelines were provided that can be used to evaluate. Those interested in more details should refer to the original paper.

1. General principle. The general principle when using model based estimation in official statistics, is the principle that official statistics give a definite answer to a question.
 - a. Objectivity: data used to estimate the model should be related to the subject of the statistic of interest. The model should only be used to estimate the present, but estimation should not exceed the present.
 - b. Reliability: failure of the model should not lead to changes in the (conclusions based on the) estimate of the statistical phenomenon. The model should be robust to small changes in the data.
2. The use of models.
 - a. Goal. The goal of using model based estimation should be to estimate data that is not available, and as such to improve the overall estimation.
 - b. Data. Models are used to estimate missing data. Both for fitting the model as well as for the final estimation procedure, only data that are available should be used.
 - c. Standard. Model based methods that are used at Statistics Netherlands should follow any general consensus in the literature on similar methods.
 - d. Model selection. Alternative models should be considered, in order to find the most appropriate model. With model selection, the aim is to find a model that is both simple and accurate.

Consultations

- Key SAE experts: consultation meeting organized by JPSM Technical Group on SAE, May 2021
- Emails and focus-group discussions
 - Australia, Canada, Chile, Colombia, Indonesia, Jamaica, Moldova, Philippines, South Africa, US, Viet Nam
- Next steps:
 - Approaching more countries and document the challenges/lessons learned
 - Finalise the first stage of the Toolkit; advocating the usefulness of SAE while underlining the important aspects to be considered
 - Incorporate an element on the use of geospatial data for SAE
 - Organise small technical group discussion (countries + academic)

eLearning modality

UNSD, UNECLAC,
UNFPA

- Course format
 - Reading materials
 - Recorded videos (50 videos with about 10-15 minutes for each video), organized in 10 modules
 - Evaluation materials including weekly computer-graded assessments, two mid-term projects, and a final project
 - R program language code that can be used for SAE modelling
- Self-paced learners:
 - Learn at their own pace
 - Access all the above learning materials
 - Machine graded weekly assessments
 - Access to projects, R script and data – not graded
- Guided learners
 - Guided learning and need to follow a fixed schedule that entails about 1-hour of work per day for ten weeks, reading assigned materials, watching course videos, and completing the assigned projects
 - 2-hour interactive workshop per week for ten weeks that will cover a summary of the weekly learning materials and instructions of R code that can be used for SAE modeling
 - Feedback and grading for all three projects

Inter-Secretariat Working Group on Household Surveys

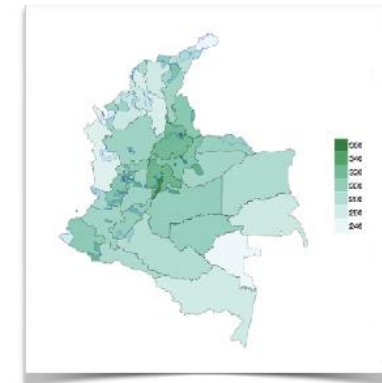
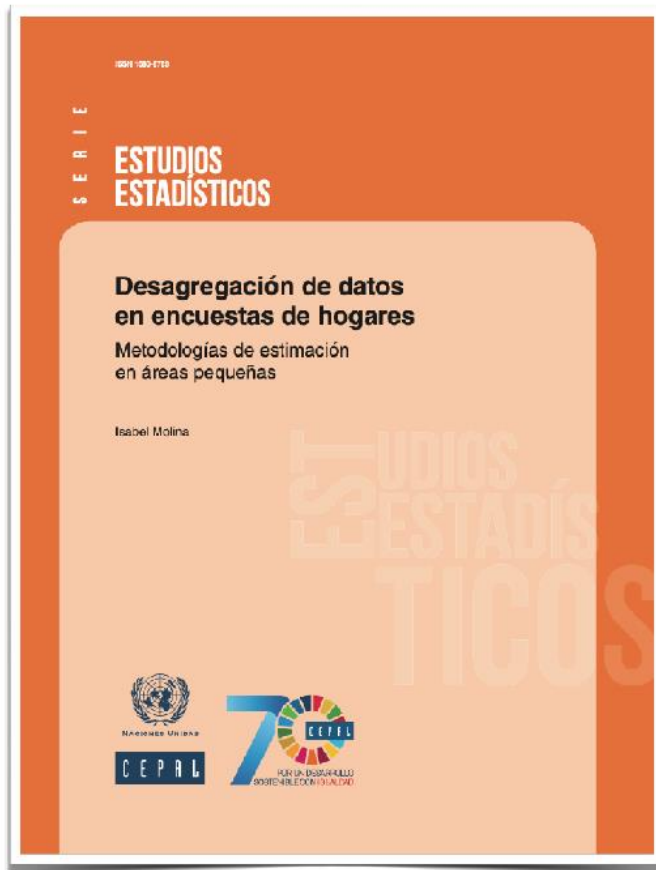
- Improve **coordination**: surveys within the country and efforts at the global level
- Advance (cross-cutting) survey **methodology**
- Enhance **communication and advocacy**

- Established at 46th Session of UNSC in 2015
- Current (rotating) co-chairs: UN Women and WB
- Secretariat: UNSD
- Members: 11 international agencies and 8 countries

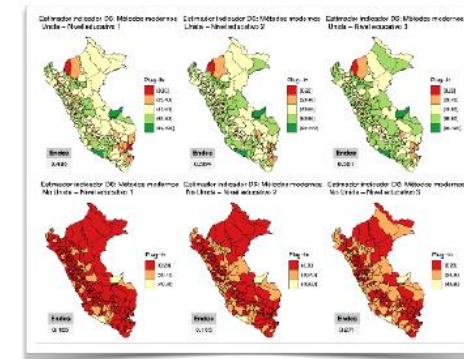


-  Colombia - Departamento Administrativo Nacional de Estadística
-  Ghana - Ghana Statistical Services
-  India - Ministry of Statistics and Programme Implementation
-  Malaysia - Department of Statistics Malaysia
-  Niger - Institut National de la Statistique
-  Samoa - Samoa Bureau of Statistics
-  State of Palestine - Palestinian Central Bureau of Statistics
-  Sweden - Statistics Sweden

ECLAC approach to capacity building in LAC



```
13 sigma <- 100
16 beta1 <- 0
17 beta0 <- 200
18 x1 <- runif(N, 0, 25)
19
20 y1 <- beta0 + 20 + beta1 * x1 + rnorm(N, 0, sigma)
21 y2 <- beta0 + 120 + beta1 * x1 + rnorm(N, 0, sigma)
22 y3 <- beta0 + 220 + beta1 * x1 + rnorm(N, 0, sigma)
23 y4 <- beta0 + 320 + beta1 * x1 + rnorm(N, 0, sigma)
24 y5 <- beta0 + 420 + beta1 * x1 + rnorm(N, 0, sigma)
25 ID <- rep(LETTERS[1:5], each = N)
26
27 test <- data.frame(Index = c(x1),
28                   Ingreso = c(y1, y2, y3, y4, y5), ID = ID)
29
30 p1 <- ggplot(data = test, aes(Index, Ingreso)) +
31   geom_point() + geom_smooth(method=lm) + theme_bw()
32
```



ECLAC approach to capacity building in LAC

- Before the pandemic, ECLAC Statistics Division had carried out F2F trainings to colleagues from Uruguay, El Salvador, Colombia and Chile.
- By May 2021, ECLAC gave a virtual training on SAE to 18 NSO from all around LAC
- During 2021, ECLAC also has offered and completed four versions of comprehensive virtual trainings on SAE to Argentina, Perú, Ecuador, Bolivia.

Projects on SAE supported by ECLAC

- Ecuador: malnutrition rates at municipality level (SDG 2).
- Chile: poverty rate at municipality level (SDG 1).
- Chile: victimization rates at the municipality level (SDG 5 and 16).
- Perú and Colombia: family planning indicators at the municipality level at the municipality level (SDG 3).
- Colombia: access to justice indicators (SDG 16).

Thank you!
chen9@un.org