# Small area estimation using earth observation data

## Regional Workshop for Asia and the Pacific
## 21-25 October 2024, Bangkok, Thailand

17 October 2024

## Background

Better data and statistics are a key enabler for building back a better world and addressing the development divide, as well as ways to hold governments to account and to improve decision making. In particular, data on key development priorities that are produced with greater frequency, timeliness and granularity are in high demand. In response, national statistical offices (NSOs) are increasingly exploring innovative data sources, tools and methods to help address these user needs.

The ESCAP Committee on Statistics therefore decided at its 7th session to "feature big data for official statistics in its future work, with an emphasis on sharing country research, experiences and good practices and facilitating capacity development" and to "strengthening legislative provisions and institutional mechanisms to enable national statistical systems to take full advantage of new and innovative technologies while respecting the Fundamental Principles of Official Statistics."

The ESCAP secretariat implements several initiatives to implement these decisions, including the capacity development project entitled the '*2030 Data Decade - Strengthening the institutional capacity of national statistical offices in Asia and the Pacific to use innovative, new and big data sources for official statistics in support of the 2030 Agenda for Sustainable Development*' (the Big Data Project). The project is funded through the 2030 Agenda for Sustainable Development Sub-Fund of the UN Peace and Development Trust Fund. Through this project, ESCAP is providing technical assistance and related support to countries, as well as developing new knowledge products and facilitating the provision of opportunities for the sharing of achievements among countries in the region.

## Small area estimation and EO data

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 priorities. To properly measure this,

statistics need to be presented for different population groups and geographical areas irrespective of their size.

Small area estimation (SAE) techniques have been proven to be very useful in providing reliable disaggregated data for some SDG indicators, such as poverty estimation, food insecurity and undernutrition, health-related indicators, unemployment rate, etc. There have been great demands from countries on providing technical guidance and training in this area, as requested by the 52nd session of the UN Statistical Commission which "*encouraged further work on data disaggregation and small area estimation to provide additional comprehensive guidelines and tools for countries*".

In response to this request, a Toolkit[1] and eLearning course[2] were developed by DESA Statistics Division (UNSD) and its partners on small area estimation for SDGs. These tools provide guidance to countries on the use of SAE methods to develop national capacity to produce SAE estimates independently.

In recent years, there has been increasing interest by NSOs to expand their use of SAE, including moving beyond traditional data sources for covariates such as census or administrative data, to incorporate other forms of geospatial data such as earth observation.

As part of the project on Big Data for Official Statistics, ESCAP, together with UNICEF South Asia and UNSD, are providing capacity support to countries that have indicated the use of earth observation data with SAE as a priority. This support includes the ongoing facilitated e-learning course on SAE, which is taking place during Q2-Q3 2024 as well as this in-person workshop relating to the use of EO data, to be held in Bangkok, Thailand between 21st and 25th October 2024.

## Objectives and expected outcomes

The overall aim of the workshop is to develop skills for using Earth Observation data together with Small Area Estimation methods and understand how these can be used for the production of geographically disaggregated indicators. By the end of the workshop, it is expected that participants will:

- Consolidate the knowledge and skills in the application of small area estimation using R gained from the guided e-learning course.
- Understand different forms of geospatial data and how to work with them and be able to access key sources of publicly available earth observation data.

[1] https://unstats.un.org/wiki/display/SAE4SDG/SAE4SDG

[2] Launched during the 54th session of the UN Statistical Commission (https://unstats.un.org/UNSDWebsite/events-details/un54sc-17022023-M-Start-your-Small-Area-Estimation-eLearning-journey-today )

- Gain practical experience of developing SAE models in R using real-world geospatial data.
- Have a deeper appreciation of what is required to move from experimentation to the production of Official Statistics in this area.

## Participants and prerequisites

The target audience for these capacity development activities are statisticians or methodologists within the NSO (or other National Statistical System agency) with responsibility for producing geographically disaggregated data as well as conversant with using surveys for estimation of domain-based aggregates.

To participate effectively in the workshop, it is essential that participants are familiar with and able to apply the basics of SAE in R. Accordingly, **all participants are expected to have completed the guided e-learning course on SAE, including submitting the required assignments.**

## Date, Location and Duration

The workshop will be conducted over five days, from 21-25 October 2024 at Novotel Bangkok on Siam Square hotel, Thailand.

The workshop will take place in the Alpha Room, the 2nd floor near the Square restaurant area.

Address: 392/44 Siam Square Soi 6, Rama I Road, 10330 Pathumwan, Bangkok
Telephone: +66 2 209 8888
Website: https://www.novotelbkk.com/

## Preparatory activities

In advance of the workshop, please:

- Ensure R/RStudio is installed on the laptop you are bringing to Bangkok
- Install the R package *terra* if possible in order to save time during the workshop itself

We also request that you bring to the workshop national survey data containing the variable of interest for which you wish to produce disaggregated estimates. As we will be working with geospatial data, this survey data should either:

1. Be geocoded with x- and y- coordinates
2. Contain some form of administrative identifier which can be used to link the data to a low-level geographic functional area (e.g. district or census enumeration area)

Please also bring a shapefile containing these low-level functional areas, particularly if you are following option 2, above.

## Contact information

**ESCAP:**

- Mr. Richard Tonkin, Statistician – Richard.tonkin@un.org (Programmatic and substantive issues)
- Ms. Nannapas Sukwattananipaat, Programme Assistant - nannapas.sukwattananipaat@un.org (Administrative issues for ESCAP supported participants & questions relating to Thai visas and venue)

**UNICEF:**

- Mr. Gustavo Nicolas Paez Salamanca, Data Analyst, gnpaez@unicef.org (UNICEF-supported participants)

## Programme

| Date | Time | Description |
|---|---|---|
| **Monday 21 October** | 9.00 – 10.30 | **Welcome and Intro**<br>• Introductory remarks from ESCAP, UNSD and UNICEF<br>• Country introductions<br>• Remind of objectives of the workshop<br>• Understand goals of participants |
| | 10:30 – 10:45 | *Coffee Break* |
| | 10:45 – 12:30 | **Session**<br>• Why small area estimation? Why geospatial data?<br>• Introduction to GitHub page containing data & slides<br>• Review experience with R and SAE from virtual training |
| | 12:30 – 13:30 | *Lunch* |
| | 13:30 – 17:00 | **Session**<br>• The use of R/RStudio (continued)<br>  ○ Tidyverse<br>  ○ Aggregating data<br>  ○ ggplot<br>• Practice with data<br><br>*Coffee break at 15:00 – 15.15* |
| **Tuesday 22 October** | 09:00 – 12:30 | **Session**<br>• Introduction to geospatial data<br>• Introduction to coordinate reference systems<br>• Shapefiles (vector data)<br>  ○ The terra and tidyterra packages<br>  ○ Polygons, lines and points<br>  ○ Creating maps<br>  ○ Overlaying shapefiles<br><br>*Coffee break at 10:30 – 10:45* |
| | 12:30 – 13:30 | *Lunch* |
| | 13:30 – 17:00 | **Session**<br>• Shapefile relations<br>  ○ Areas/distances<br>  ○ Intersections |

| | | |
|---|---|---|
| | | ○ Unions<br>○ Buffers<br>• Practice with data<br><br>*Coffee break at 15:00 – 15.15* |
| **Wednesday 23 October** | 09:00 – 12:30 | **Session**<br>• Rasters – what is raster data?<br>○ The terra and tidyterra packages<br>○ Plotting rasters<br>• Raster extraction<br>○ Extracting raster data to shapefiles<br>○ Creating grids<br>○ Matching grids to shapefiles<br>• Preparation for later practical sessions<br><br>*Coffee break at 10:30 – 10:45* |
| | 12:30 – 13:30 | *Lunch* |
| | 13:30 – 17:00 | **Session**<br>• Accessing rasters<br>○ Finding rasters online<br>○ Geolink package<br>○ Downloading rasters from Google Earth Engine and other repositories<br><br>*Coffee break at 15:00 – 15.15* |
| **Thursday 24 October** | 09:00 – 12:30 | **Session**<br>• Group activity<br>○ Grids, rasters and aggregation<br>• Review of SAE methods<br>○ Focus on sub-area level models<br>• Estimating SAE models in *povmap* in R<br><br>To be delivered jointly with David Newhouse, World Bank.<br><br>*Coffee break at 10:30 – 10:45* |
| | 12:30 – 13:30 | *Lunch* |
| | 13:30 – 17:00 | **Session**<br>• Estimating SAE models in *povmap* in R (continued)<br>• Introduction to team task – implementing a (small) SAE model using national data<br><br>*Coffee break at 15:00 – 15.15* |
| **Friday 25 October** | 09:00 – 12:30 | **Session**<br>• Completion of team task |

| | | |
|---|---|---|
| | | • Presenting results to group and discussion<br><br>*Coffee break at 10:30 – 10:45* |
| | 12:30 – 13:30 | *Lunch* |
| | 13:30 - close | **Session**<br>• From experimentation to official statistics<br><br>**Closing**<br>• Evaluation of training programme<br>• Next steps and close<br><br>*Coffee break at 15:00 – 15.15* |