# The Global Subnational Agricultural Production (GSAP) database

**Data for empowering food security worldwide**

**SEI project fact sheet**
**April 2024**

Rafaela Flach

Any efforts to understand or address agricultural production need to start with fundamental questions: Which crops are being planted every year? Where exactly are they being planted? How much is produced? And how much land is being used to produce it? The Global Subnational Agricultural Production (GSAP) database aims to gather the answers to these questions, and to make them available to policymakers and decision-makers working to improve food security at all levels.

Databases that provide the global view of agricultural production at smaller scales have come up short, despite widespread recognition of such information's importance. Not having this type of information available results in the reliance on outdated, costly or inaccurate modelling efforts, which leads to poor understanding of current agricultural production at a global scale and hampers efforts to improve food security worldwide. While most countries collect reliable local data on their agriculture sectors and publish them through their national statistical offices, this information has yet to be tapped into for a global outlook.

Agricultural production is changing in the face of increasing demographic pressures, pervasive effects of climate change, growing water and fertilizer scarcity, soil erosion, and geopolitical tensions leading to increased risks for food-importing countries. A global view of smaller scales of agricultural production will allow policymakers and other decision-makers to respond to these challenges empowered with spatially explicit, accurate and up-to-date information on food systems.

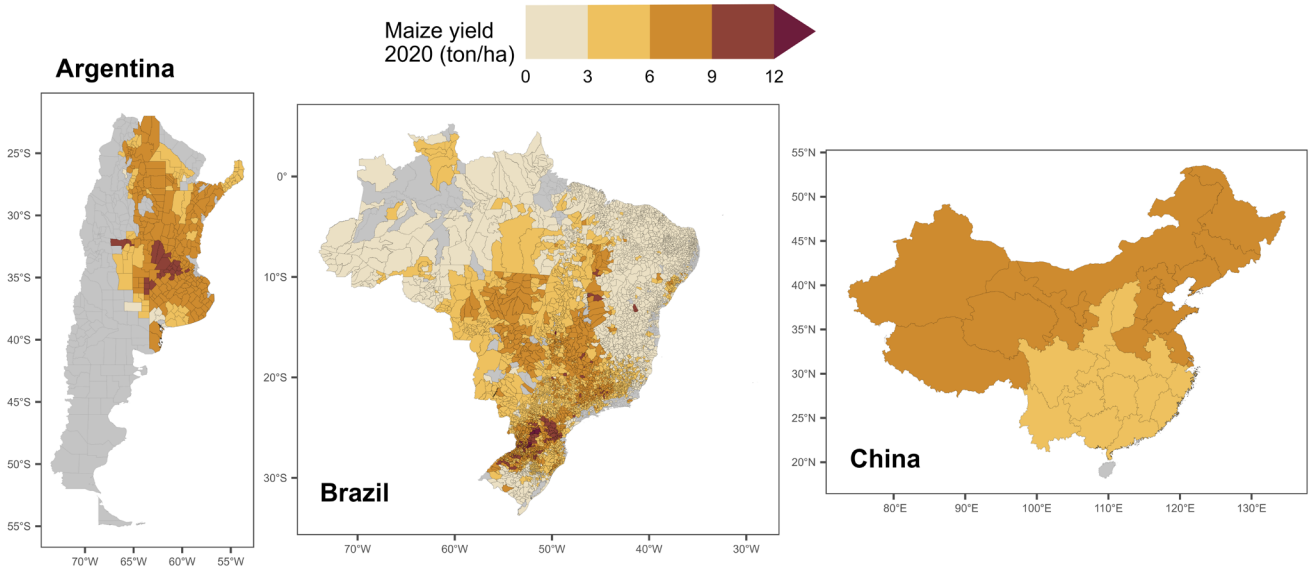
## 1. What is GSAP?

The GSAP dataset is the first standardized global subnational agricultural production database that covers key global crops and is both updated (2010–21) and comprehensive in global country coverage. We collect past and current data from national agricultural censuses and surveys, at the highest resolution available (districts, provinces, regions), for the world's 50 most relevant crops. These include the world's main grains, such as wheat, maize, rice and soybean, as well as crops important for local economies or food security such as cassava, yams, coffee and cocoa.

The finer scale of GSAP data versus national crop production data and models allows for a better integration with Earth-observation and socio-economic data. This should lead to more accurate analyses of drivers and risks of crop production and associated socio-economic impacts, such as financial and humanitarian risks.

IMAGE (ABOVE): Dried beans and legumes

Figure 1. Example of subnational yield data. Maize yields in Argentina, Brazil and China at a subnational scale, in 2020. Along with the US, these are the world's main producers of maize.




*Source: Author's own*

## 2. Filling the gaps with GSAP

Without this global view of smaller scales of agricultural production, we face an information gap that leads to uncertainty in the prioritization of investments in agricultural research and development, forecasts of local food production and associated variability caused by changes in climate and productive conditions, and future regional geopolitical and humanitarian scenarios. Moreover, this critical information gap hampers efforts to achieve the Sustainable Development Goals (SDGs) 1 and 2.

Currently, the main source of information for global agricultural production data is FAOSTAT, which provides agricultural data at a national resolution. The available global crop maps that are widely used across different communities are either outdated (Monfreda et al., 2008) or based on estimated crop presence using suitability models (e.g. MapsPAM, GAEZ). Current attempts to bypass the basic need for subnational agricultural data are costly and yield large uncertainty.

Having this information would increase the accuracy of analyses of agricultural production and global yield gaps. It would also boost the work of researchers, practitioners and policymakers in fields as diverse as rural development, food security, climate change, land use, development finance and supply chain sustainability.

## 3. How we work

The approach used by the GSAP team can be summarized as follows:

1. **Obtain official national data.** Our team obtains data directly from national statistical offices and ministries of agriculture, from agricultural censuses or surveys.

2. **Scrape, process and clean the data.** Using a variety of tools, including AI-related optical character recognition and pattern matching, we scape, clean, and reorganize the data obtained.

**A glimpse into our data**

**Obtained data:**

185 countries

Containing upwards of 30 000 subnational jurisdictions

**Processed data:**

Around 15 000 subnational jurisdictions

These mapped jurisdictions produce:

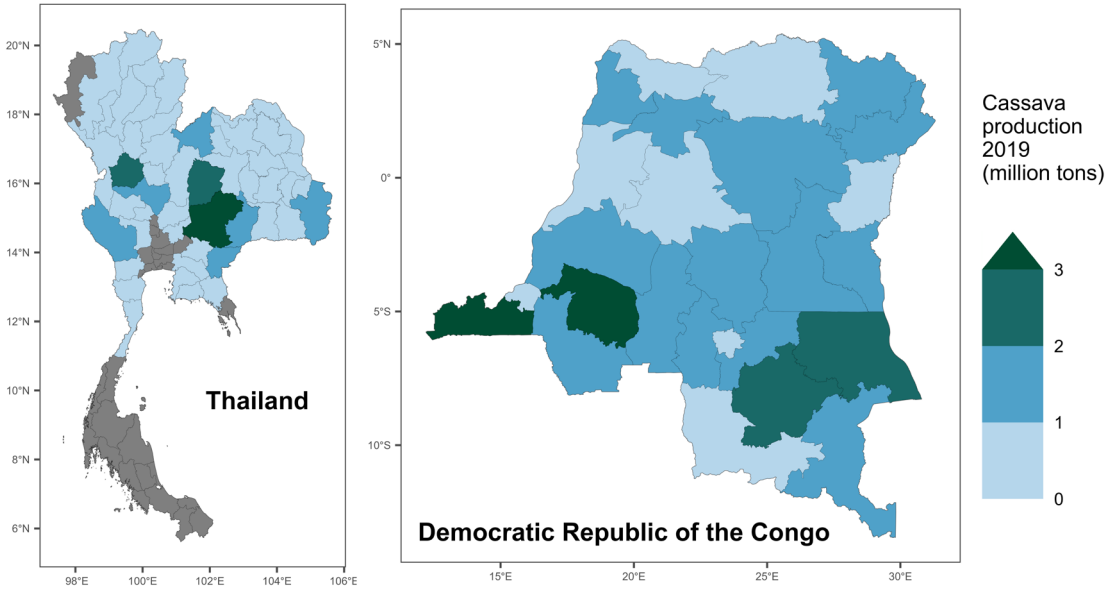
- 75% of global production of wheat and rice
- 85% of global production of maize and sugarcane
- 92% of global production of soybean

3. **Build harmonized dictionaries.** We harmonize the geographical units, crops and units based on each global jurisdiction, with its own unique code from a standardized system: the Database of Global Administrative Area for jurisdictions and the FAO crop coding system for crops. We convert all values to the same units: hectares and metric tons.

4. **Quality assurance.** This process includes a full data availability scan, finding duplicates and missing values, and checking for correspondence with values previously reported.

The first phase of GSAP demonstrated the viability of obtaining the necessary data, as well as the efficacy of the method developed by the GSAP team. Our team found that subnational agricultural information is widely available from the same national entities that annually report to FAO, and we confirmed that the information we collected is consistent with the figures reported by FAO.

Figure 2. Example of subnational production data. Cassava production in Thailand and the Democratic Republic of Congo at a subnational scale, in 2019. Along with Nigeria and Brazil, these are the world's main producers of cassava.




*Source: Author's own*

## POTENTIAL APPLICATIONS

GSAP data have the potential to improve the accuracy, relevance and applicability of several avenues of investigation. Some of the sectors and research fields that are set to benefit most include:

- **food security and rural development**
- **yield gaps and understanding future food availability vis-a-vis expected demand**
- **supply chain risks and sustainability**
- **global security and geopolitics**
- **yield and land use modelling, to incorporate effects of technology, climate and environment on agriculture**
- **environmental and economic modelling for climate, hydrological, inter-sectoral, crop, deforestation, and land use**
- **climate adaptation**
- **biodiversity**.

## 4. Future prospects

While the first phase showed viability of the GSAP project, we need to expand on this success to make the best use of the available data. The potential path to achieve that would include:

- **Build an online platform.** Make the database available publicly, enabling access for researchers and practitioners.
- **Engage.** Promote the platform, the database, and secondary products (see below). We aim to engage with key institutions for ready uptake, as well as with a more general audience to share insights.
- **Cover 100% of production.** Expand the database to cover a larger sample of crops and cover all the countries that produce each crop of interest.
- **Produce research and secondary products.** The GSAP data can enable a wide range of avenues of investigation, including on yield gaps, supply chains and agricultural system trends, among others. We envision the creation of a **Global Agriculture Lab**, where the data can be better understood, communicated and applied to as many applications as possible.
- **Expand beyond crop production.** We see potential to include additional forms of agricultural data in GSAP including data on livestock, forestry, and agricultural management practices such as irrigation, cultivars and crop calendars. The process of obtaining the data for GSAP showed that this additional data are largely available.

World heritage Ifugao rice terraces, the Philippine Cordilleras © PRAPASS PULSUB / GETTY IMAGES

SEI Stockholm Environment Institute

**Published by**
Stockholm Environment Institute
Linnégatan 87D, Box 24218
104 51 Stockholm, Sweden
Tel: +46 8 30 80 44

**Author contact**
rafaela.flach@sei.org

Visit us: sei.org
Twitter: @SEIresearch
@SEIclimate

Editor: Naomi Lubick
Layout: Richard Clay