CODEX PLANETARIUS

Codex Planetarius Pilot Phase Report: Generating Distributions of Impacts Using HESTIA

Juliet Telfer*, Guillaume Royer*, Joseph Poore*

*HESTIA Project, Department of Biology / Oxford Martin School, University of Oxford

About *Codex Planetarius*

Codex Planetarius is a proposed system of minimum environmental performance standards for producing globally traded food. It is modeled on the *Codex Alimentarius*, a set of minimum mandatory health and safety standards for globally traded food. The goal of Codex Planetarius is to measure and manage the key environmental impacts of food production, acknowledging that while some resources may be renewable, they may be consumed at a faster rate than the planet can renew them.

The global production of food has had the largest impact of any human activity on the planet. Continuing increases in population and per capita income, accompanied by dietary shifts, are putting even more pressure on the planet and its ability to regenerate renewable resources. We need to reduce food production's key impacts.

The impacts of food production are not spread evenly among producers. Data across commodities suggest that the bottom 10-20% of producers account for 60-80% of the impacts associated globally with producing any commodity, even though they produce only 5-10% of the product. We need to focus on the bottom.

Once approved, *Codex Planetarius* will provide governments and trade authorities with a baseline for environmental performance in the global trade of food and soft commodities. It won't replace what governments already do. Rather, it will help build consensus about key impacts, how to measure them, and what minimum acceptable performance should be for global trade. We need a common escalator of continuous improvement.

These papers are part of a multiyear proof of concept to answer questions and explore issues, launch an informed discussion, and help create a pathway to assess the overall viability of *Codex Planetarius*. We believe *Codex Planetarius* would improve food production and reduce its environmental impact on the planet.

This proof-of-concept research and analysis is funded by the Gordon and Betty Moore Foundation and led by World Wildlife Fund in collaboration with a number of global organizations and experts.

For more information, visit www.codexplanetarius.org

Codex Planetarius Pilot Phase Report: Generating Distributions of Impacts Using HESTIA

Juliet Telfer*, Guillaume Royer*, Joseph Poore*

*HESTIA Project, Department of Biology / Oxford Martin School, University of Oxford

Abstract

The aim of this project was to use HESTIA to contribute fundamental information to the *Codex Planetarius* project, helping define thresholds for environmental impacts for import and trade policy. Distributions were generated for five crop-country combinations specified, making these available through the data aggregations on the HESTIA platform. This report also includes contribution analysis of the highest impact farms, based on six key indicators, to the total production and total impact. Our analysis showed that a small number of farms generally contribute a disproportionate amount of the environmental impact for most impact indicators (e.g., for cocoa in Ghana, 25% of farms contribute 42% of the total GWP100 impact and 13% of the total production).

Introduction

Background

The aim of this project was to use HESTIA to contribute fundamental information to the *Codex Planetarius* project, helping define thresholds for environmental impacts for import and trade policy.

HESTIA is a platform that stores harmonized data describing ~140,000 farms globally and calculates environmental impacts using harmonised methods. It was initiated at Oxford University and has been funded and supported by WWF since its initiation.

This project aimed to implement new functionality on HESTIA to create distributions of environmental impacts across farms for key commodities in key geographies, as well as identifying the most impactful farms. This project also aimed to expand data on these commodities.

This report covers the methodology used for data collection of the key commodities, as well as documenting the data availability, sources, and any gaps that may influence the results. There is also an outline of the methods used for generating the distributions.

Overall Methodology

Data collection

Data were collected through literature reviews and by working with our network of collaborators within research and government organizations. We only included data that represents commercial practices. We detail sources for data collection below.

Data aggregation

The data were then aggregated to produce an average production cycle, considering the share of farmed area under organic and conventional practices (data from FiBL) and the farmed area under irrigated/rainfed conditions (data from AQUASTAT). The weightings used are country-specific.

Distributions and percentiles

Distributions were generated by sampling from the data in the production cycles behind each aggregation 1000 times. Sometimes, the data in the production cycles were in the form of mean + standard deviation. This was the case where a single study represented many farms but did not provide data for each farm and instead provided just summary statistics. To ensure we captured this variation, we sampled from the data assuming they were distributed normally (with a minimum of 0 where relevant).

In simple terms, we then calculated a full Life Cycle Assessment for each element in the array (so 1000 LCAs were performed for each crop-country combination). This then allows us to calculate distributions of each environmental impact indicator.

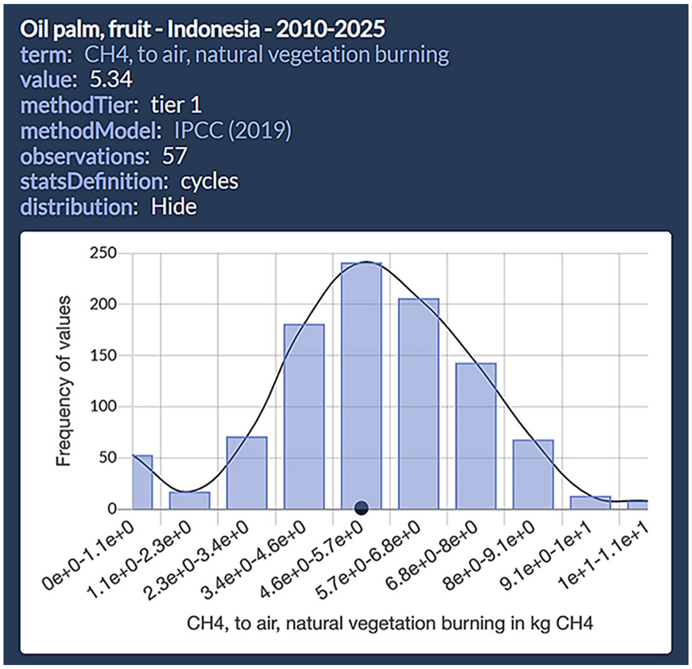
The percentile graphs below show the contribution of the highest impact farms for each indicator to the total impact and total production. For some of the total production bars, the contribution is very small, so it appears as 0 on the graph. The percentiles section also includes summary tables showing the mean, median, lower 10th percentile, and upper 10th percentile (90th percentile) for each indicator.

Accessing the data directly from HESTIA

For each crop-country combination, we provide a link to the aggregated data. To access this:

1. Create a HESTIA account and login.
2. The links will take you to our data explorer. Click through to the relevant aggregation.
3. In the aggregation page, click on a data item, e.g., an emission, an input.
4. Click where it says “Show” on the distribution. You will see a histogram (Example #1, below) of the distribution (blue bars), a smoothed curve, and a dot (the mean).
5. Some indicators have multiple models which calculate them. In this case, the “distribution” will not immediately appear. Press the gear in the top left of the popup box and tick the box next to “distribution” to add the distribution.
6. You can also download data from the section at the bottom of the page.

Example 1.



Results

Rice, India

Overview of data

- Link to aggregation on HESTIA: [Rice grain (in husk), flooded - India - 2010-2025 | HESTIA]
- Underlying cycles: 30,397
- Underlying sources: 7
- Representativeness = HIGH (our yield = 4,180 kg/h FAOSTAT yield = 3,850 kg/ha; very strong sub-national coverage)

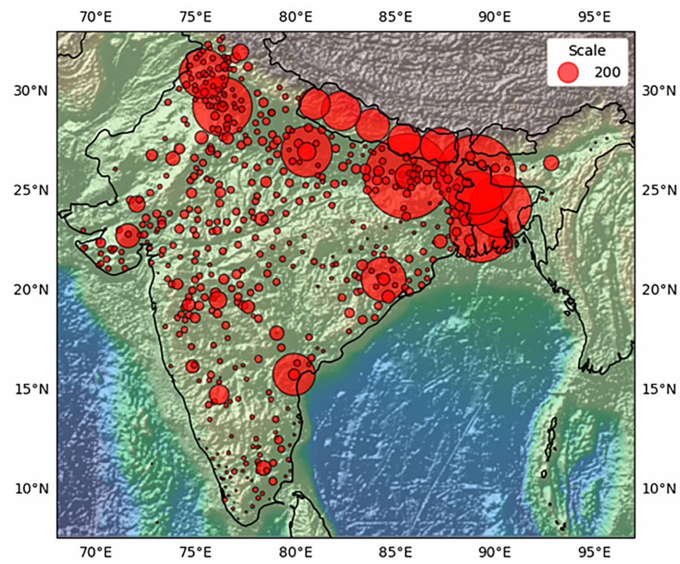


Figure 1. Map of all the farms surveyed in India.

The two main sources are CIMMYT (2017) and Hossain (2024) (IRRI dataset). These were both large-scale farm surveys, containing primary information on fertilizer use, yield, machinery use, tillage, and crop residue practices. They also contained the locations of the farms surveyed. The map to the right (Figure 1) shows all the farms surveyed (rice and non-rice farms – mainly grain). This shows how the surveys strongly cover the Indo-Gangetic plain of India (a major producing region). These two sources did not contain information on irrigation water or pesticide use, but the other studies we identified and uploaded did include this information, hence our results for pesticide and water-related indicators are based on a smaller sample. All the data represents conventional production (organic wheat production is not a major system in India). The weighting we implemented was 57.2% to rainfed and 42.8% to irrigated.[1]

Distributions and share of production in most impactful decile/quartile - Rice *(Figure 2, next page)*

Drivers of high impacts

For GWP100, 57% of the contributions are methane emissions from rice flooding, high impact farms are likely incorporating more crop residue and using more organic fertiliser which increases this emission source. With regards to eutrophication, high impact farms are likely to have a higher slope. While paddy fields are generally terraced, there is still large potential for soil erosion, which contributes to higher nitrogen losses[4]. For the soil quality index and the biodiversity impact, these mostly depend on the yield of the farm, with high-yielding farms having lower biodiversity impacts and soil quality per kg of product (note – these farms may have high biodiversity and quality per hectare which is worth exploring further as the *Codex* project develops).

Data gaps and how to fill them

The data for rice in India are very strong and representative. However, we are aware that there are new surveys from CIM-MYT, IRRI, and the Indian government for more recent years. Therefore, our suggestions to improve the data are:

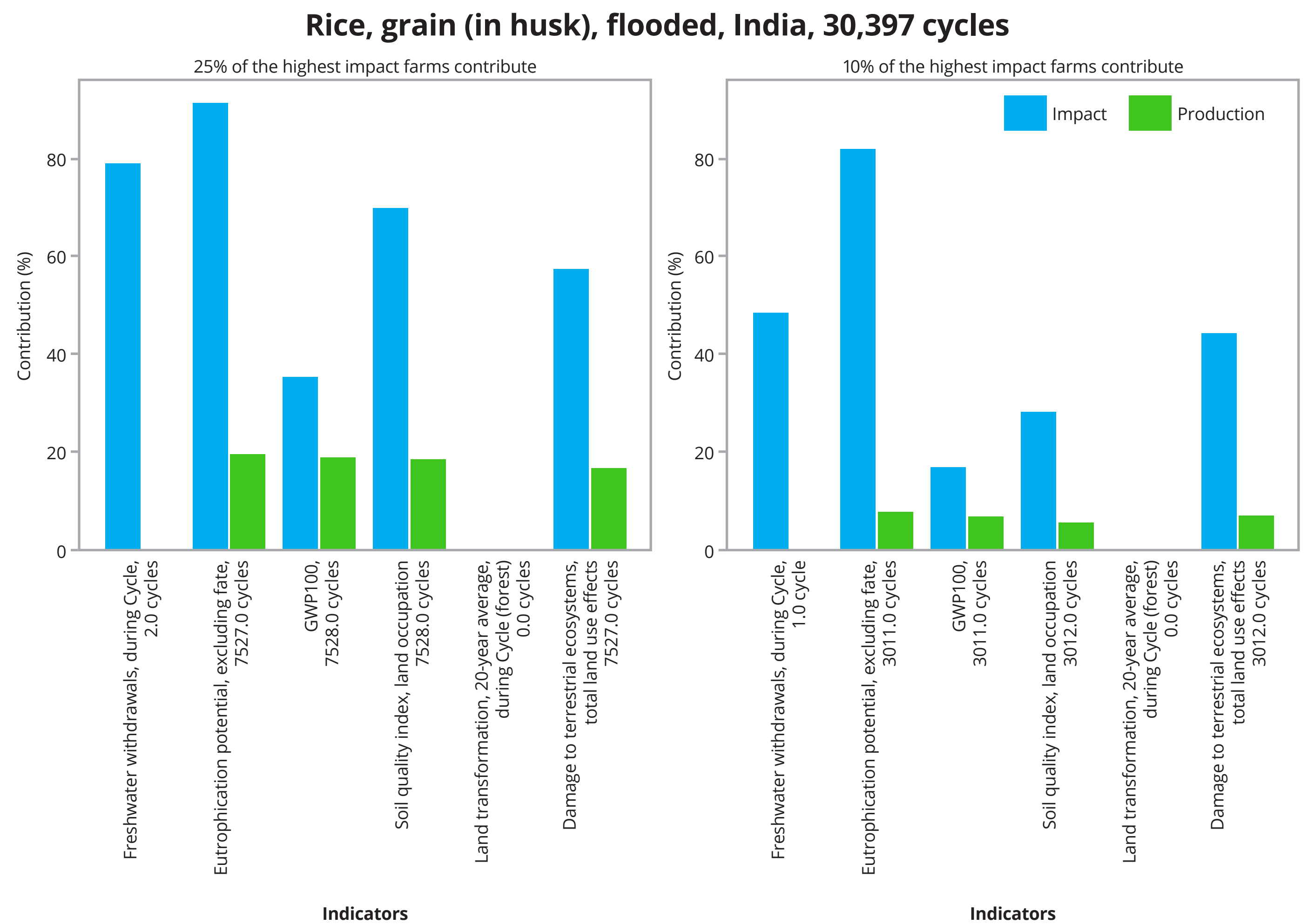
1. Integrate the latest survey datasets from CIMMYT, IRRI, and the Indian government.

2. Undertake a new literature review with broader search terms to further improve pesticide and water use data.

Additionally, we would also propose to:

3. Improve our weighting structures, so that we more explicitly weight based on the share of production per sub-region.

Figure 2. ***Oil Palm: Distributions and share of production in most impactful decile/quartile***



	Freshwater withdrawals, during Cycle (L)	Eutrophication potential, excluding fate (kg PO43-eq)	GWP100 (kg CO2-eq)	Soil quality index, land occupation (points)	Land transformation, 20-year average, during Cycle (forest) (m2)	Damage to terrestrial ecosystems, total land use effects (PDF* year)
Mean	1460	0.0637[2]	0.703	15.9	0	1.66E-13
Median	754	0.00508	0.590	N/A[3]	0	8.95E-14
Upper 10th Percentile	3490	0.0630	1.109	46.2	0	2.35E-13
Lower 10th Percentile	147	0.000976	0.285	N/A	0	2.29E-14

Oil palm fruit, Indonesia

Overview of data

- Link to aggregation on HESTIA: Oil palm, fruit - Indonesia - 2010-2025 | HESTIA [note, see the 2010-2025 aggregation, not the 1990-2009 aggregation]
- Underlying cycles: 57
- Underlying sources: 12
- Representativeness = MEDIUM/HIGH (our yield = 21,000 kg/ha, FAOSTAT yield = 17,900 kg/ha; strong sub-national coverage)

The sources used are mostly LCA-focused papers looking at the environmental impact of palm oil. All the data are representative of conventional, rainfed production. Fifty-five percent of the sites are located in the Central Kalimantan province, 5% in Riau, 9% in Central Java, 4% in Jambi, the rest are country-level sites. This is not fully representative of the regions as Riau produces 20% of Indonesia's total palm oil, and Central Kalimantan produces 18%[5]. RSPO and non-certified RSPO production is represented as 14% of cycles are certified.

Distributions and share of production in most impactful decile/quartile - Oil Palm *(Figure 3, next page)*

Drivers of high impacts

The reason for farms being high impact with respect to GWP100, is likely to do with CO2 emissions from organic soil cultivation, which contribute around 22% to the overall value, but only appear in 15 cycles. Currently, we don't have a spatial gap-fill for the soil measurements that are used to calculate this emission, so it only comes from uploaded data. Around 50% of the eutrophication impact comes from N and P erosion, however this emission is only running for a few cycles. This is due to most of the cycles being country-level, and so we cannot reliably gap-fill the site measurements needed to allow the model to run. For the biodiversity impact, 88% of the contributions are from land transformation, and only 12% are from the land occupation. For this and the soil quality index, the high impact farms are likely young palm plantations that have significantly lower yield than fully matured plantations (few thousand kg/ha compared to 20,000 kg/ha).

Data gaps and how to fill them

The data we have for Indonesia are very strong. However, they are weighted towards large plantations, and we could expand to include more smallholder data. We explored collaborations with RSPO, BRIN, and Cambridge University to achieve this, but have not had success at bringing their data on to the platform for a mix of reasons including: data confidentiality (RSPO), poor response rates and lack of time on the collaborators side (BRIN) and insufficient data on key items of activity where data were not collected with LCA in mind (Cambridge University).

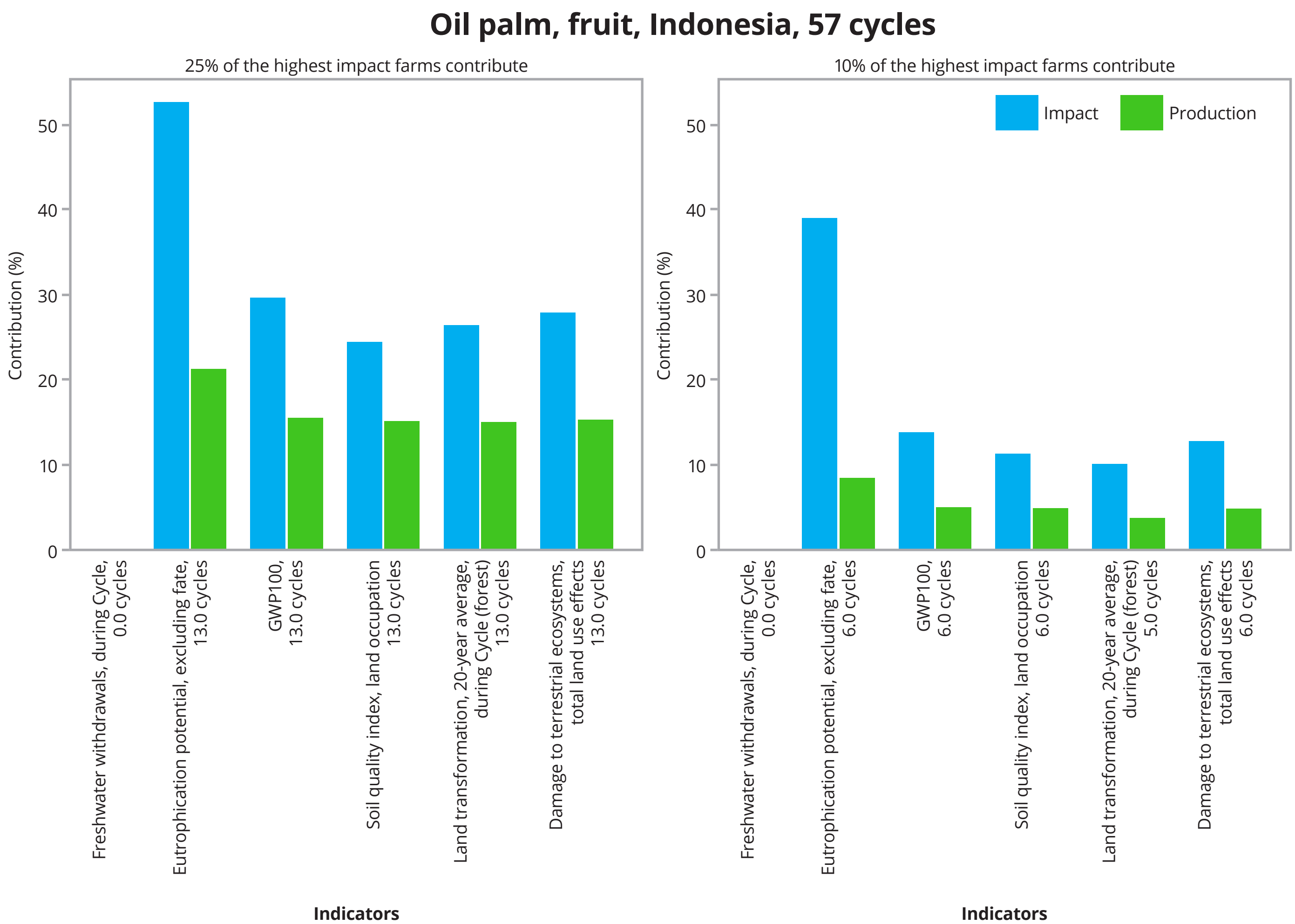
Therefore, our suggestions to improve the data are:

1. Work with RSPO on a model to get access to their data.
2. Retry our collaboration with BRIN (we may need research funding to flow to them, or an in-person visit[6] to unlock this).
3. Consider primary data collection.

We could also improve our weighting structure:

4. Improve our weights to better account for the share of production from smallholders vs large plantations, as well as certified vs non-certified production. Identify data to do this.

Figure 3. ***Oil Palm: Distributions and share of production in most impactful decile/quartile***



	Freshwater withdrawals, during Cycle (L)	Eutrophication potential, excluding fate (kg PO43-eq)	GWP100 (kg CO2-eq)	Soil quality index, land occupation (points)	Land transformation, 20-year average, during Cycle (forest) (m2)	Damage to terrestrial ecosystems, total land use effects (PDF* year)
Mean	0	0.000529	2.26	134.0	0.0266	2.80E-12
Median	0	0.000328	1.68	102.0	0.0203	2.36E-12
Upper 10th Percentile	0	0.000731	3.23	168.0	0.0309	3.56E-12
Lower 10th Percentile	0	0.000212	1.25	84.7	0.0159	4.11E-13

Soy, Brazil

Overview of data

- Link to aggregation on HESTIA: [Soybean, seed (whole) - Brazil - 2010-2025 | HESTIA]
- Underlying cycles: 194
- Underlying sources: 14
- Representativeness = HIGH (our yield = 3,020 kg/ha; FAOSTAT yield = 3,110 kg/ha; very strong sub-national coverage)

The main source is Ramos et al. (2013) which used data from farm surveys and contains 14 cycles of soybean production. The other sources are all LCA-focused papers. All the data represents conventional, rainfed production. There is one cycle with a water input, but this is not used for irrigation and likely is for diluting fertilisers or pesticides. Twenty-five percent of sites represent country-level production, the rest are representative of Brazil's key producing regions[7] – Mato Grosso (13%), Paraná (6%), Rio Grande do Sul (13%), Goiás (8%), Bahia (4%).

Distributions and share of production in most impactful decile/quartile - Soy (Figure 4, next page)

Drivers of high impacts

The main reason for farms being high impact with regards to GWP100 is due to emissions from land use change (LUC), which only appears in the minority of the underlying cycles but contributes to 40% of the GWP. This is due to an issue with our historical land cover model, which is not running correctly for cycles that occur over two different years, and means these cycles do not have any LUC emissions. Once corrected, the emissions from LUC should contribute even more to the GWP100 value. This issue also affects the biodiversity impact; once fixed the impact should depend mostly on the yield like for palm. For eutrophication, the biggest contributor is NO3 to groundwater, from crop residue decomposition, followed by N and P erosion. The high impact farms for the soil quality index are there due to low yield, as you can see by the low contribution they have to overall production.

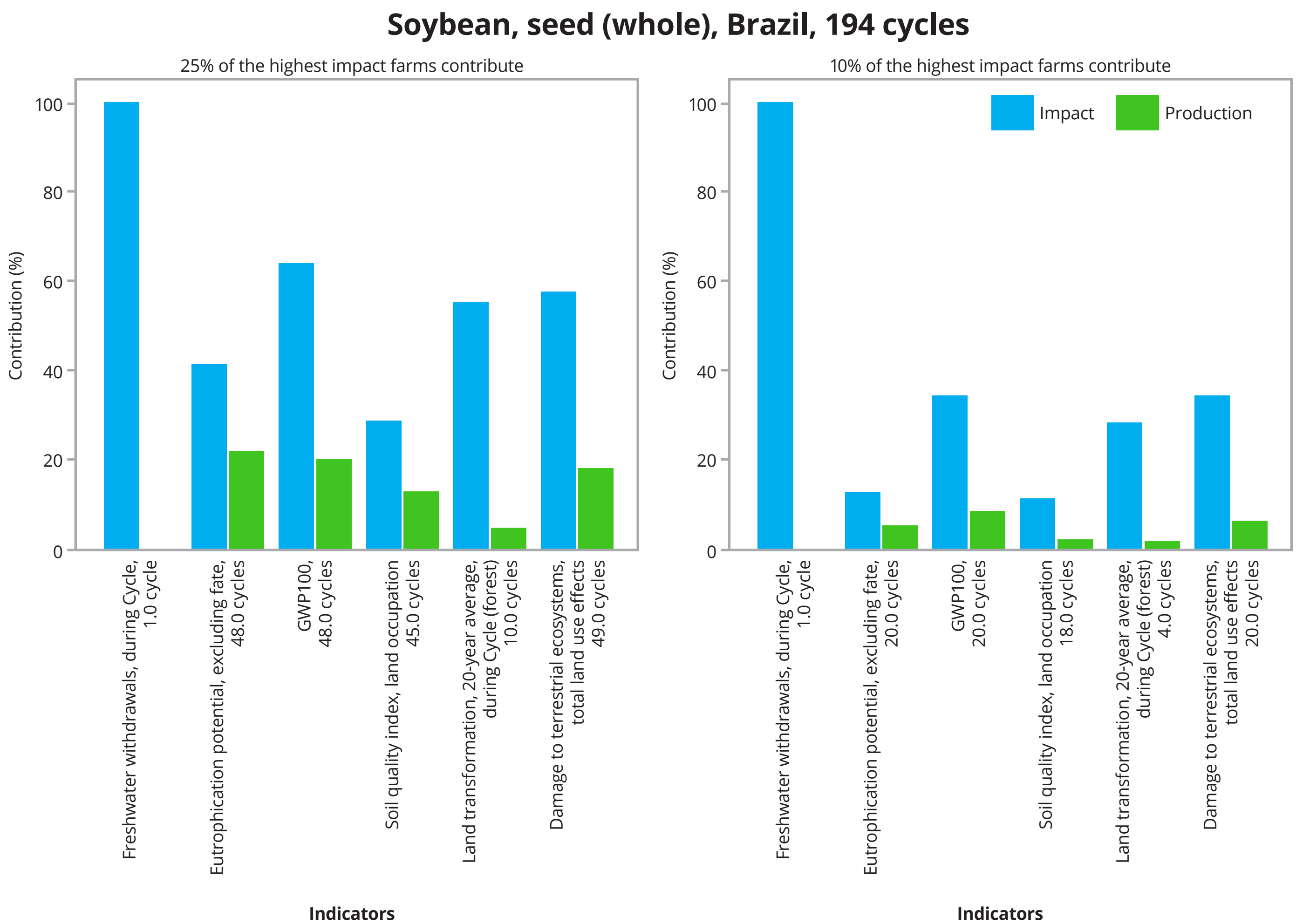
Data gaps and how to fill them

An improvement would be to include data on irrigated cultivation which accounts for around 10% of production[8]. An additional improvement would be to get a larger number of Cycles to better characterise the distribution. We have exhausted the published literature, and hence the only ways to achieve this would be to identify industry datasets or undertake/join primary data collection efforts.

Therefore, our suggestions to improve the data are:

1. Systematically aim to work with industry organizations. For HESTIA (which is geared to working with public open data sets) this will require a new data strategy.
2. Identify primary data collection efforts in Brazil soy. Currently we are applying to a Horizon grant which includes a partner collecting such data.

Figure 4. ***Soy: Distributions and share of production in most impactful decile/quartile***



	Freshwater withdrawals, during Cycle (L)	Eutrophication potential, excluding fate (kg PO43-eq)	GWP100 (kg CO2-eq)	Soil quality index, land occupation (points)	Land transformation, 20-year average, during Cycle (forest) (m2)	Damage to terrestrial ecosystems, total land use effects (PDF* year)
Mean	0.00102	0.00379	0.939	629	0.0151	6.01E-13
Median	0	0.00273	0.450	386	0.0133	2.85E-13
Upper 10th Percentile	0	0.00684	2.60	799	0.0396	1.49E-12
Lower 10th Percentile	0	0.00198	0.169	165	0	9.17E-14

Wheat, Australia

Overview of data

- Link to aggregation on HESTIA: [Wheat, grain - Australia - 2010-2025 | HESTIA]
- Underlying cycles: 85
- Underlying sources: 9
- Representativeness = MEDIUM/HIGH (our yield = 3,490 kg/ha; FAOSTAT yield = 2,42 kg/ha; strong sub-national coverage)

The two main sources were AusLCI data, which contributed 24 cycles and used data from farm surveys, and Sevenster et al. (2024), which contributed 30 cycles and covered a long-term field experiment in New South Wales. The other sources were a mix of experimental sites and LCA-focused papers studying farms. The yield is higher than FAOSTAT, this could be due to the amount of experimental data included in the aggregation. All the data are representative of conventional, rainfed production. Twenty percent of the sites are in Western Australia, the main wheat-producing region[9], 44% are in New South Wales, 15% are in Queensland, 9% are in South Australia, 4% are in Victoria and 2% are in Tasmania, the rest of the sites represent country-level production.

Distributions and share of production in most impactful decile/quartile - Wheat *(Figure 5, next page)*

Drivers of high impacts

As the yield varies quite significantly, 1200 kg/ha – 7300 kg/ha, the high impact cycles are likely a combination of low yield and high inputs. This correlates with the biggest contributors to GWP100 being fertilizer application, fertilizer production, and fuel combustion. Also, for eutrophication, the biggest contributor is nitrogen fertilizer application at 32%. For the biodiversity impact, 86% of the contributions are from land occupation and only 14% from land transformation. Therefore, low yield is the main reason for the high impact farms, similarly for the soil quality index.

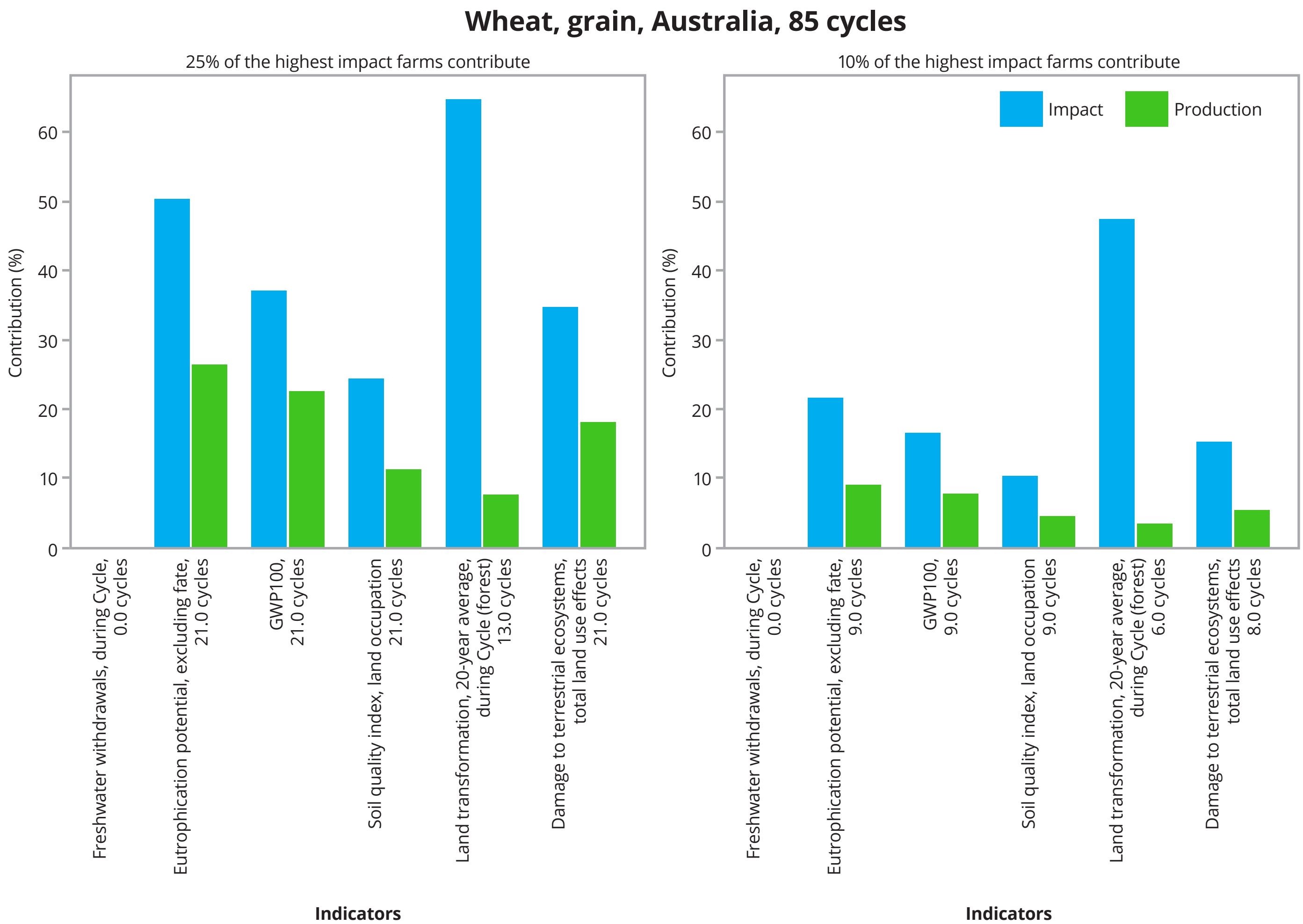
Data gaps and how to fill them

While the data are generally good, we could benefit from more Cycles to allow us to better characterize the tails of the distributions. During our work, we became aware of new datasets available in Australia, which had licensing-related issues to work through.

Therefore, our suggestions to improve the data are:

1. Create a generic plan for working with licensed datasets, and identify how we can make these available on HESTIA.
2. Strengthen our collaborations with Australian researchers (e.g., by jointly publishing results with them) and getting access to updates of their datasets as well as data from more field stations.

Figure 5. ***Wheat: Distributions and share of production in most impactful decile/quartile***



	Freshwater withdrawals, during Cycle (L)	Eutrophication potential, excluding fate (kg PO43-eq)	GWP100 (kg CO2-eq)	Soil quality index, land occupation (points)	Land transformation, 20-year average, during Cycle (forest) (m2)	Damage to terrestrial ecosystems, total land use effects (PDF* year)
Mean	0	0.00254	0.627	125	0.00959	1.09E-12
Median	0	0.00267	0.583	116	0.00761	9.47E-13
Upper 10th Percentile	0	0.00543	1.01	207	0.0161	1.70E-12
Lower 10th Percentile	0	0.000635	0.318	52	0	5.82E-13

Cocoa, Ghana

Overview of data

- Link to aggregation on HESTIA: [Cocoa, seed (whole) - Ghana - 2010-2025 | HESTIA]
- Underlying cycles: 576
- Underlying sources: 6
- Representativeness = HIGH (our yield = 516 kg/ha; FAOSTAT yield = 484 kg/ha; strong sub-national coverage)

The two main sources used are farm surveys; Schader et al. (2021) contains 391 cycles of organic farms, and Abdulai et al. (2018) contains 149 cycles of farms along the climate gradient in the cocoa-producing region. One hundred and sixty-nine of the sites are located in the Ashanti, Eastern, Western and Central regions of Ghana, which are the main cocoa producing regions[10]. The weighting towards rainfed production is 99.2%, and the weighting towards organic production is 0.273%. Twenty percent of the data represents Fair Trade Certified farms, and 10% represents Rainforest Alliance Certified. Approximately 10%[11] of cocoa farmers in Ghana are Fair Trade certified.

Distributions and share of production in most impactful decile/quartile - Cocoa *(Figure 6, next page)*

Drivers of high impacts

As emissions from land use change contribute over 90% to the total GWP, the main reason for certain farms being high impact is likely due to lower yield. For eutrophication, NO3 to groundwater from crop residue decomposition contributes 40% to the total impact. Low yield is also the main reason for farms being high impact with regards to the soil quality index and the biodiversity impact.

Data gaps and how to fill them

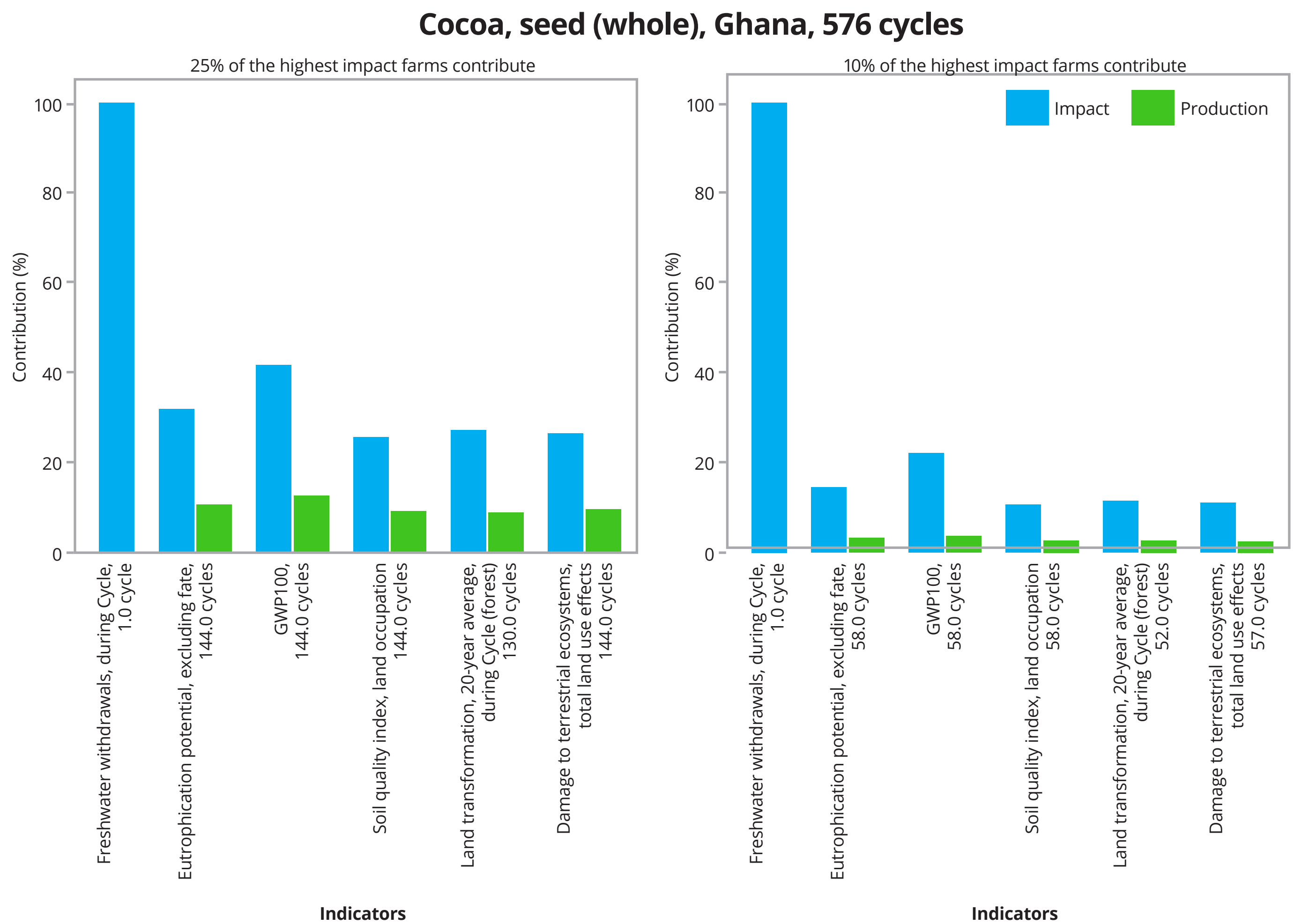
The majority of underlying cycles are from organic farms, so increasing the data volume for conventional farms would add to the representativeness of the data. We are in the process of looking at some farm surveys from a researcher at Oxford that may be suitable to upload onto the platform.

Therefore, our suggestions to improve the data are:

1. Create a collaboration with an Oxford researcher to have access to more farm data to upload onto the platform.

2. Look at including weightings for Fair Trade certified and Rainforest Alliance certified production.

Figure 6. ***Cocoa: Distributions and share of production in most impactful decile/quartile***



	Freshwater withdrawals, during Cycle (L)	Eutrophication potential, excluding fate (kg PO43-eq)	GWP100 (kg CO2-eq)	Soil quality index, land occupation (points)	Land transformation, 20-year average, during Cycle (forest) (m2)	Damage to terrestrial ecosystems, total land use effects (PDF* year)
Mean	6.47	0.0103	30.2	3260	0.375	1.73E-11
Median	0	0.00656	16.7	2290	0.264	1.20E-11
Upper 10th Percentile	0	0.0183	53.2	5980	0.643	3.02E-11
Lower 10th Percentile	0	0.00345	7.98	1260	0.139	6.57E-12

Conclusion & Next Steps

We've achieved good data quality for all the crop-country combinations specified, using survey data whenever possible to accurately reflect commercial practices. For the distributions, we've succeeded in generating 1000-point distributions of LCI data and indicators, making these available on our front-end as well as in the download formats. We have generated distributions and proven that HESTIA can support WWF to deliver the *Codex Planetarius* project. ■

Footnotes

[1] Our weighting structure here may need improving, as more recent data suggests rainfed production is now just 40% of area in India (https://doi.org/10.1016/bs.agron.2015.05.004).

[2] Here the mean is higher than the upper 10th percentile, indicating very high skew. This is driven by a few sites with high slope affecting the paddy rice nutrient loss model. We will implement a fix for this, but note that a new model is being developed here.

[3] One of the larger datasets (IRRI) has not been recently recalculated, and this indicator is missing for some of the rice cycles.

[4] The model we use is based on the Universal Soil Loss Equation, and implemented in Scherer & Pfister (2015) Int J. LCA. We are aware that this model is being updated to better account for terracing in paddy fields so this source of nitrogen loss may change.

[5] Indonesian Oil Palm Statistics 2023 - BPS-Statistics Indonesia

[6] As a project, we generally do not fly and would need a strong justification and clarity that this would unlock data for us to proceed.

[7] Brazil's Soybean Production: Key Regions & Economic Impact

[8] https://doi.org/10.1016/j.agwat.2019.03.003

[9] Wheat Belt | Grain Farming, Wheat Production & Wheat Harvesting | Britannica

[10] Here is a list of Ghana's top 5 cocoa-growing communities | GhanaRemembers

[11] The Ghana Cocoa Report 2024: Fairtrade Cocoa in Ghana: Economic and Social Impacts