



geoFootprint

TECHNICAL DOCUMENTATION

Team: Jürgen Reinhard + Xavier Bengoa + Adrien Liernur
Contact: geofootprint@quantis-intl.com
Last update: 04 May 2021

Recommended citation:

Reinhard J., Bengoa X. & Liernur A. (2021): geoFootprint, Technical Documentation. Version 1, May 2021. Quantis, Lausanne, Switzerland.

By: **Quantis**

Supported by:



Climate-KIC

Climate-KIC is supported by the
EIF, a body of the European Union



Table of Contents

TECHNICAL REVIEW	6
GLOSSARY	7
ABBREVIATIONS	10
1. INTRODUCTION	11
1.1. BACKGROUND.....	11
1.2. CURRENT LIMITATIONS OF AGRICULTURAL LCA	11
1.3. GEOFOOTPRINT	12
1.4. GOAL AND STRUCTURE OF THIS REPORT.....	15
2. SCOPE	19
2.1. UNIT OF ANALYSIS.....	19
2.2. FUNCTIONAL UNIT.....	20
2.3. SYSTEM BOUNDARIES.....	21
2.3.1. SYSTEM BOUNDARIES FOR CROP PRODUCTION SYSTEMS.....	21
2.3.2. CROP COVERAGE	22
2.3.3. ALLOCATION OF CROP CO-PRODUCTS AT FARM.....	23
2.4. INVENTORY DATA ON AGRICULTURAL ACTIVITIES	25
2.5. DATA REPRESENTATIVENESS	27
2.5.1. GEOGRAPHICAL COVERAGE	28
2.5.2. TEMPORAL COVERAGE.....	29
2.5.3. TECHNOLOGICAL COVERAGE.....	29
3. KEY ELEMENTS IN REGIONALIZED LCA	30

3.1. SPATIAL ATTRIBUTE TABLE.....	30
3.1.1. EXTRAPOLATION OF SPATIAL ATTRIBUTES.....	33
3.1.2. PREPROCESSING OF SPATIAL ATTRIBUTES.....	37
3.2. DEFAULT CROP INVENTORY TABLE	45
3.2.1. IDENTIFYING AVAILABLE AUPD.....	46
3.2.2. LINKING AVAILABLE DATASETS TO GEOGRAPHIES.....	47
3.2.3. DETERMINING UNIQUE INVENTORY FLOWS PER CROP	48
3.2.4. GEOGRAPHICAL DIFFERENTIATION	49
3.3. REGIONALIZED INVENTORY TABLE	50
3.4. REGIONALIZED FOOTPRINT TABLE.....	60
3.4.1. ASSESSMENT OF ENVIRONMENTAL IMPACT IN GEOFOOTPRINT	61
3.4.2. PROXY & EFFICIENCY INDICATORS	63
3.4.3. GENERIC (SPATIALLY-INDEPENDENT) FOOTPRINTS.....	66
3.4.4. REGIONALIZED FOOTPRINTS	69
<u>4. RECALCULATING SCENARIOS.....</u>	<u>72</u>
<u>5. REGIONALIZING RESOURCE FLOWS</u>	<u>75</u>
5.1. LAND OCCUPATION	75
5.2. LAND TRANSFORMATION	76
<u>6. REGIONALIZING MINERAL AND ORGANIC FERTILIZER APPLICATION RATES.....</u>	<u>78</u>
6.1. ORGANIC FERTILIZER.....	78
6.2. MINERAL FERTILIZER.....	81
6.3. NITROGEN USE EFFICIENCY (NUE).....	82
<u>7. REGIONALIZING IRRIGATION WATER CONSUMPTION</u>	<u>86</u>
7.1. CURATION OF IRRIGATION VOLUME	87

7.1.1.	TREATMENT OF VERY LOW IRRIGATION VOLUMES	88
7.1.2.	TREATMENT OF VERY HIGH IRRIGATION VOLUMES	88
7.2.	INVENTORY FLOWS TO MODEL IRRIGATION WATER USE.....	89
7.3.	MODELING OF IRRIGATION	90
7.3.1.	WATER INPUT	90
7.3.2.	IRRIGATION TECHNOLOGIES.....	91
7.3.3.	WATER EMISSIONS	92
<u>8.</u>	<u>REGIONALIZING SOIL EROSION</u>	<u>93</u>
<u>9.</u>	<u>REGIONALIZING N & P EMISSIONS</u>	<u>97</u>
9.1.	NITRATE EMISSIONS	97
9.2.	PHOSPHORUS EMISSIONS.....	99
9.2.1.	PHOSPHORUS LOSS IN PARTICULATE FORM TO SURFACE WATER: SOIL EROSION	99
9.2.2.	PHOSPHORUS LOSS TO SURFACE WATER (RUN-OFF AND DRAINAGE).....	100
9.2.3.	PHOSPHORUS LOSS IN DISSOLVED FORM (PHOSPHATES) TO GROUNDWATER: LEACHING	102
<u>10.</u>	<u>REGIONALIZING EMISSIONS FROM LAND USE CHANGE FROM TREE COVER LOSS.....</u>	<u>104</u>
10.1.	GRID CELL BASED CALCULATION OF LUC IMPACT	104
10.1.1.	AGGREGATING LUC.....	108
10.1.2.	EXPANDING LUC.....	109
10.1.3.	DETERMINE CROP-SPECIFIC ALLOCATION FACTORS	110
10.1.4.	ALLOCATION LUC TO KEY CROPS.....	111
10.1.5.	NORMALIZATION TO ONE HECTARE.....	112
10.2.	LIMITATIONS.....	113
10.3.	OUTLOOK	116
<u>11.</u>	<u>REGIONALIZING EMISSIONS FROM LAND USE.....</u>	<u>117</u>

11.1. CHANGES IN SOC DUE TO MANAGEMENT.....	118
11.1.1. CO2 EMISSIONS FROM SOC	118
11.1.2. N2O EMISSIONS FROM SOC	120
11.2. SOC LEACHING DUE TO SOIL EROSION	120
11.3. RESIDUE MANAGEMENT	121
11.3.1. SOC INCREASES (CO2 ACCUMULATION) DUE TO ADDITION OF ORGANIC MATTER	121
11.3.2. EMISSION OF N ₂ O AND CH ₄ DUE TO CROP RESIDUE MANAGEMENT	124
11.3.3. N2O EMISSIONS DUE TO INCORPORATED CROP RESIDUE	124
11.4. N₂O, NO_x & NH₃ EMISSIONS DUE TO MINERAL AND ORGANIC FERTILIZER APPLICATION	125
11.4.1. INPUT FACTORS (DEFAULT AND USER DEFINED).....	125
11.4.2. CALCULATION OF DIRECT AND INDIRECT N2O EMISSIONS.....	131
11.4.3. CALCULATION OF NO.....	133
11.4.4. CALCULATION OF NH ₃	133
11.5. CH₄ EMISSIONS FROM RICE.....	133
11.6. CO₂ EMISSIONS FROM LIMING, UREA AND AMMONIUM BICARBONATE	135
11.7. LIMITATIONS AND OUTLOOK.....	136
<u>12. REFERENCES.....</u>	<u>137</u>
<u>13. APPENDIX</u>	<u>142</u>
13.1. DATA INPUTS	142
13.2. RASTERIZATION OF REGIONALIZED FOOTPRINTS	149

Technical Review

The following people contributed to reviewing these guidelines as part of a closed consultation procedure (in alphabetical order). We are grateful for their valuable inputs.

- Angel Avadi, PhD, CIRAD, France
- Assumpció Antón, PhD, IRTA, Spain
- Edilene Pereira, IRTA, Spain
- Erica Montemayor, IRTA, Spain
- Marta Ruiz PhD, IRTA, Spain
- Montse Núñez PhD, IRTA, Spain

Reviewers expressed their views and do not necessarily always agree with methods and data that have been applied in this technical documentation.

Glossary

AUPD	Agricultural unit process dataset. A dataset describing, in distinct reference to the yield, all inventory flows associated with the cultivation of a crop either on the country or global level.
Characterization factor	The environmental characterization of the potential environmental impact associated with the extraction of a resource or the release of an emission.
Footprint	We use the term “footprint” rather loosely in this documentation, namely to refer to an aggregation mechanism that expresses one or several environmental characteristics of the agricultural system (our regionalized inventory) under study in one number.
Generic footprint	A generic footprint is spatially independent, i.e., characterizes a particular resource extraction or emission, always with the same characterization factor.
Georeferenced data	Data associated with a ground system of geographic coordinates and can therefore be precisely localized on the earth surface.
Horizontal regionalization	The regionalization of certain data elements of <i>a particular product stage</i> in the supply chain (in geoFootprint’s case, agricultural cultivation).
Impact factor	The environmental impact associated with the cradle-to-gate provision of a product or service, i.e., an output

	<p>from technosphere. We compute impact factors of technosphere flows from the WFLDB 3.5 with SimaPro version 9.0. We use the impact factors published by ecoinvent to represent environmental impacts of all flows taken from the ecoinvent database.</p>
Inventory flows	<p>Physical consumption of resources, man-made stuff (technosphere flows) and corresponding emissions associated with the provision of an agricultural commodity and recorded in the form of an AUPD.</p>
LCIA	<p>Life cycle impact assessment. The LCA phase where the elementary flows in the life cycle inventory are characterized according to their environmental relevance.</p>
Efficiency indicator	<p>Summarizes physical fluxes on the inventory level—nitrogen use efficiency and water withdrawal—in order to indicate the efficiency of the agricultural system under study.</p>
Proxy indicator	<p>Uses the physical properties of the agricultural system under study—e.g., soil erosion and soil organic carbon stock change—to approximate an important environmental characteristic (e.g., soil health).</p>
Regionalized footprint	<p>Characterizes a resource extraction or emission under consideration of the local conditions, i.e., it considers the fact that the impact of one and the same amount of an emission or resource extraction can vary with the location. Also called a spatially differentiated footprint.</p>
Regionalized LCA modeling	<p>Regionalized life cycle assessment modeling. Refers to the modeling procedure that automatically generates</p>

site-specific (i) AUPDs under consideration of various spatial input attributes and existing default crop inventory data and (ii) transforms them into regionalized footprints.

Regionalized LCI modeling

Regionalized life cycle inventory modeling. Automatically generates comprehensive, site-specific AUPDs under consideration of various spatial input attributes and existing default crop inventory data.

Site-specific

In geoFootprint, site-specific refers to the native geoFootprint grid cell resolution of 5 x 5 arc-minutes resolution (i.e., 10 x 10 km at the equator)

Vertical regionalization

In contrast to horizontal regionalization, vertical regionalization focuses on the regionalization of certain data elements *along the entire supply chain*.

Abbreviations

ABC	Ammonium bicarbonate
AGN	Above-ground nitrogen
AUPD	Agricultural unit process dataset
BGN	Below-ground nitrogen
C	Carbon
CFT	Cool Farm Tool
CH ₄	Methane
CO ₂	Carbon dioxide
DMF	Dry matter fraction
GAEZ	Global Agro-Ecological Zones
GFW	Global Forest Watch
GHG	Greenhouse gas
IPCC	Intergovernmental Panel on Climate Change
ISRIC	International Soil Reference and Information Centre
LCA	Life cycle assessment
LCI	Life cycle inventory
LCIA	Life cycle impact assessment
LU	Land use
LUC	Land use change
N	Nitrogen
N ₂ O	Nitrous oxide
NH ₃	Ammonia
NO	Nitrogen monoxide
NO _x	Nitrogen oxides
pv	Production volume
SOC	Soil organic carbon
TAL	Tree cover area loss
WFLDB	World Food LCA Database

1. Introduction

1.1. Background

Agriculture and forestry together contribute over 20% of global greenhouse gases emissions. For many companies dependent on agricultural supply chains, continued business growth relies on the availability and quality of crop production. Consequently, they see an absolute need to shift towards more sustainable agricultural practices that preserve land and water resources and limit greenhouse gas (GHG) emissions. A growing number of companies worldwide have recently committed to reducing their environmental footprint at the crop production level, setting quantified targets. To achieve these, they have identified a need for more comprehensive site-specific data and more user-friendly tools to better monitor and manage local agricultural practices. These targets and their underlying data need to be supported by robust science.

Life Cycle Assessment (LCA) focuses on the assessment of environmental impacts of products and services along their entire life cycle. It is likely the most prominent technique to quantify the environmental footprint of products, and strongly supported by the European Commission and numerous national environmental agencies.

1.2. Current limitations of agricultural LCA

To date, however, LCA is limited in its capability to represent the context dependency of agricultural activities—such as crop production or deforestation—as it focuses on averages at the expense of specificity. Due to the high cost of data collection, it is common practice to consider the average meteorological and ecological conditions of a whole country and geographical region when compiling the required (country-level) datasets. However,

agricultural activities are strongly affected by the local meteorological and ecological conditions and are therefore highly context dependent¹.

Although country-level datasets provide a valid foundation for the identification of key drivers of environmental impacts, they are very limited in their capability to monitor and manage local agricultural practices. Moreover, models used to estimate direct field emissions may not be well adapted to yield accurate predictions in all agricultural situations, for example, different pedo-climatic conditions, different fertilization strategies, different crop types.

In other words, the common agricultural LCA practice provides limited decision support for close examination and implementation of site-specific mitigation measures.

1.3. geoFootprint

The geoFootprint initiative alleviates these limitations. geoFootprint facilitates the consideration and automated integration of spatial-explicit data into LCA calculation, thereby offering a radically more efficient and more accurate way for the generation of regionalized environmental footprints of agricultural products at an unprecedented resolution of 5x5 arc minutes (i.e., 10x10 km at the equator). A web-based platform allows exploration and manipulation of these “higher-resolution” crop-production datasets for all major crops, all regions in the world and several footprints of relevance, including carbon, water, soil, land use change and others.

We briefly elaborate on the geoFootprint architecture in order to introduce the key elements in regionalized LCA modeling. (See chapter 3 for a thorough examination of these elements.) Figure 1 shows the regionalization engine in the context of the overall geoFootprint architecture.

¹ The type and amount of resources used (water, land, etc.), the inputs required (the application of fertilizers and crop protection agents, the use of machinery, etc.) and the corresponding emissions into soil, air and water (carbon dioxide, nitrate, dinitrogen monoxide, phosphate, etc.) are determined by small-scale spatial parameters (precipitation, soil properties, slope, etc.) and are therefore highly context dependent.

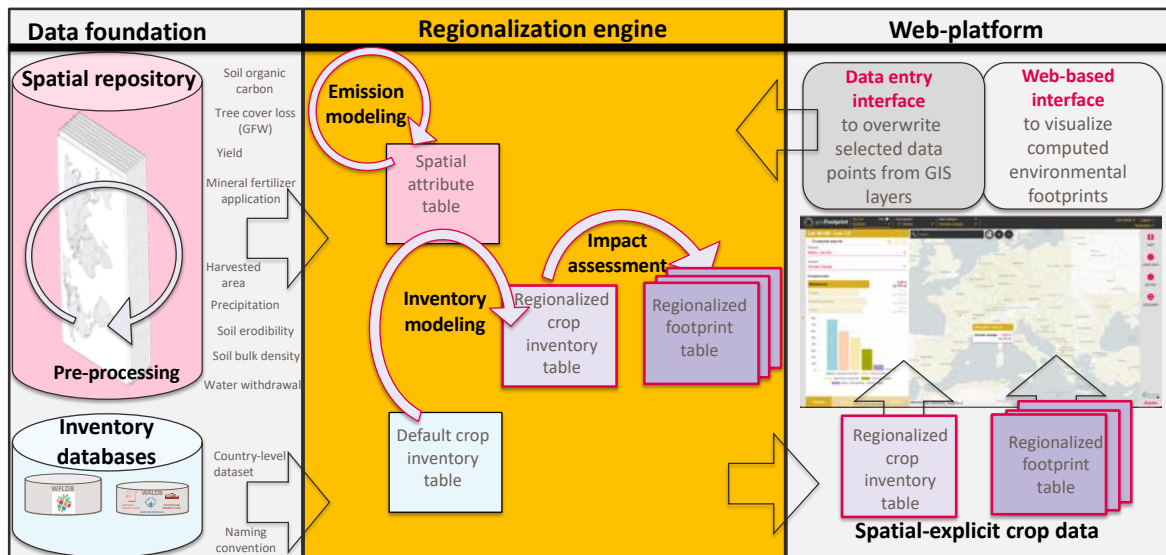


Figure 1: Key elements in the overall geoFootprint architecture.

We compiled a *spatial repository* of publicly available raster data² indicating harvested area, yield, fertilizer application rates, and irrigation requirement of major crops, as well as data on contextual conditions such as precipitation, tree cover area loss, soil properties and terrain on a global scale. We use this spatial data repository to produce a *spatial attribute table*. This table records all spatial attributes in distinct spatial (geo) reference to a global raster of 10x10-km grid cells and serves as a foundation for *emission modeling*, i.e., the computation of new spatial attributes on the basis of existing ones.

Moreover, we use available inventory databases (World Food LCA Database and ecoinvent) to retrieve and spatially differentiate available default country-level agricultural unit process datasets (AUPD) for all major crops. The spatially differentiated default crop inventory table records all inventory flows associated with the global production of a particular crop in distinct spatial reference to a grid cell and provides a powerful template for regionalized inventory modeling. It typically comprises 80–160 inventory flows describing, across all grid cells, the type and amount of:

- Natural resources used (e.g., in the form of water, land, energy)

² A representation of the world as a surface divided into a regular grid of cells. Raster models are useful for storing data that varies continuously, as in an aerial photograph, a satellite image, a surface of chemical concentrations, or an elevation surface.

- Intermediate flows required (e.g., the use of man-made products such as the application of mineral and organic fertilizer or the use of field operations)
- Accompanying release of emissions into soil, air and water (e.g., carbon dioxide, nitrate, nitrous oxide, phosphate emissions). See section 1.1 for a detailed explanation of the data elements used to model agricultural activities in geoFootprint.

We use the spatial attribute table and the default crop inventory table to produce a *comprehensive regionalized crop inventory table*. For each specific grid cell, we use the contextual and emission information in the spatial attribute table and *overwrite* selected flows in the default crop inventory table—that is what we mean by *regionalization*, i.e., the modeling procedure that automatically manipulates the default crop inventory table to generate regionalized agricultural unit process datasets (AUPD) on the grid-cell level under consideration of various spatial input parameters³. (See Table 7 for a comprehensive list of inventory and emission flow categories in the regionalized crop inventory table and their degree of regionalization.) For example, in order to regionalized N fertilizer application rate, the crop-specific and spatially explicit N application rate (retrieved from EarthStat) is assigned to the fertilizer types present in the (closest) default inventory in proportion to their quantitative relevance.

Operating on the technical foundation and naming convention of existing inventory databases, we can characterize the flows in the regionalized crop inventory table according to their environmental importance. That is, we transform the regionalized crop inventory table into various *regionalized footprint tables* (e.g., carbon footprint, water footprint, eutrophication, acidification) by multiplying each flow in the regionalized crop inventory table with its corresponding characterization factor. In principle, we can apply all commonly used LCIA metrics.

We visualize and aggregate selected data elements of the regionalized crop inventory and the regionalized footprint tables on the *geoFootprint web platform*. The platform offers two core functionalities:

³ We call the regionalization of certain data elements of a particular product stage in the supply chain *horizontal regionalization*. In contrast, *vertical regionalization* focuses on the regionalization of certain data elements along the entire supply chain.

- *Exploration of all key data elements* associated with the environmental footprint. Specifically, this is a detailed analysis of the composition of the environmental footprints (e.g., analysis of a particular grid cell or region the typical environmental relevance of the irrigation or mineral fertilizer application) and also an *exploration* of the underlying inventory data, such as contextual conditions (soil type, precipitation) and the default management practices used (e.g., tillage practice, input and type of mineral fertilizer applied).
- *Scenario assessment module*, i.e., a data entry interface that will allow the integration of user-specific data. The user can deploy and explore large-scale scenarios by overwriting and refining selected data elements, recomputing all associated metrics, and exploring the implications in visual form on various level of aggregation. The re-computation of the results is handled via the regionalization engine. The refined data is sent to the regionalization engine, which, after processing the changed data, sends the recomputed results back to the web-based platform.

Note the twofold purpose of the regionalization engine: First, it is used to compute the *spatially explicit crop data* (regionalized crop inventory and regionalized footprint) for all major crops on a global scale. This data builds the foundation of the web-based platform. Second, the regionalization engine is used by the web-based platform to directly *re-compute results for selected grid cells or regions*.

1.4. Goal and structure of this report

The goal of this report is a detailed description of the modeling procedures in the regionalization engine. More specifically, this includes a detailed treatment of the following topics:

Chapter 2 - Scope of the regionalization engine

In this chapter, we explain the unit of analysis (i.e., at which level of detail we perform all calculations) and the functional unit of the agricultural systems (i.e., the reference for evaluating products obtained from crop production). We detail the general system boundaries and the applied allocation methods used to treat co-products. Finally, we treat data

representativeness in regard to current crop, geographical, temporal and technological coverage.

Chapter 3 - Key elements in regionalized LCA modeling

This chapter elaborates the technical foundation in regionalized LCA modeling—basically all key elements highlighted in Figure 1. We first explain the detailed data and processing structure of our *spatial attribute table* and important pre-processing steps and extrapolation procedures executed on the spatial data before it is used. We next highlight the basic structure and the generation of the *default crop inventory table* (i.e., how we use available *inventory databases* from the World Food LCA Database and ecoinvent to retrieve and spatially differentiate available default country-level AUPD for all major crops). Further, we explain how we use the spatial attribute table and the default crop inventory table to produce a *comprehensive regionalized crop inventory table*. Finally, we clarify how we characterize the flows in the *regionalized crop inventory table* according to their environmental importance using various general and regionalized environmental footprints.

Chapter 4 – Recalculating scenarios

The scenario assessment module facilitates the customization of important cultivation and management practices. The user can interactively deploy and explore the environmental implications of large-scale scenarios (i.e., overwrite and refine select data elements, recompute all associated metrics, and explore the implications in visual form on various levels of aggregation). In this chapter, we explain the scope of the recalculation procedure.

Chapter 5 – Regionalizing resource flows

We explain how we model land occupation and land transformation flows on the basis of spatial data on yield, cultivation period and tree cover loss.

Chapter 6 – Regionalizing organic and mineral fertilizers

This chapter explains how we use the default crop inventories to translate spatially explicit manure application data in the form of kg N from FAO (J. S. Gerber et al., 2016) into individual crop-specific manure application types and their corresponding P content. Further, we detail the procedure to convert the total N, P and K fertilizer application rates from EarthStat into individual fertilizer types, once again using the default crop inventories. We finally explain how we calculate the nitrogen use efficiency (NUE) proxy footprint.

Chapter 7 – Regionalizing irrigation water consumption

In this chapter, we elaborate how we derive the total amount of annual, crop-specific irrigation from the blue water requirement of the Water Footprinting Network data layer from Mekonnen & Hoekstra (2011), how we distribute the total crop-specific irrigation to the irrigation technologies present in the default crop inventories and how our modeling of water emissions ensures that we produce water-balanced regionalized AUPD.

Chapter 8 – Regionalizing emissions from land use change

We explain the computational steps to calculate and allocate LUC impact (associated with tree cover loss) in a crop-specific manner and on a global scale using the shared-responsibility approach, its present limitations and gives an outlook which highlights how the approach can be further refined.

Chapter 9 – Regionalizing soil erosion

Next, we explain the Universal Soil Loss Equation (USLE) model to calculate spatially explicit soil erosion in metric ton per hectare. We show how we combine novel input data from Borrelli et al. 2017 (length-slope, erodibility and erosivity factor) with default/user defined data on the tillage and anti-erosion cultivation practice.

Chapter 10 – Regionalizing emission from N & P emissions

This chapter explains the procedures to calculate regionalized N and P leaching. We first show how we use a regression model—based on the formula developed by De Willigen (De Willigen, 2000)—to compute Nitrate leaching under consideration of water availability, soil and crop properties and N input. We next highlight how we calculate P leaching in the form of phosphorus to surface water, phosphate to ground water, and phosphate to surface water using SALCA-P (Prahsun, 2006) and selected refinement published in Scherer and Pfister (2015).

Chapter 11 – Regionalizing emission from land use

In this final chapter, we explain the modeling of all emissions associated with land use. The modeling of LU focuses mostly on changes of carbon in the soil, i.e., decrease or accumulation of soil organic carbon (SOC), due to the different management practices, and corresponding emissions in the form of carbon dioxide (CO₂), nitrous oxide (N₂O) and methane (CH₄).

Management practices considered in the LU model include the effect of land use change (LUC), tillage and input practice on SOC which can lead to CO₂ and N₂O emissions; carbon leaching due to soil erosion in the form of CO₂; residue management practices which influence CH₄ and N₂O emissions; mineral and organic fertilizer application which increase soil N availability and therefore also N₂O emissions from soils as a by-product of nitrification and denitrification; the decomposition of dead organic matter and soil organic matter under flooded conditions (e.g., rice cultivation) in the form of CH₄ and liming which transfers C from the earth crust to the atmosphere by removing calcium carbonate from limestone and dolomite deposits.

2. Scope

2.1. Unit of analysis

Our *unit of analysis*—the most fundamental level at which hold and process data—is on the grid cell level. A grid cell is an artificial boundary of arbitrary size used to deal with the continuous characteristic of spatial data in a discrete manner. The total number of all grid cells sub-divide the total world surface into a regular grid. geoFootprint operates on a grid cell resolution 5 x 5 arc-minutes (i.e., 10 x 10 km at the equator) in the World Geodesy system (WGS1984, EPSG:4326⁴)⁵.

Operating on this level of detail effectively means that we process all spatial, inventory and footprint data in the form of geo-referenced data tables that indicate, for each latitude (lat) and longitude (lon) combination, all corresponding attributes (see Figure 2 for an exemplary extract of such a table). Georeferenced data is data which is associated to a ground system of geographic coordinates and, hence, can be precisely localized on the earth surface. An attribute in a geo-referenced table can be

- A *spatial data element* (e.g., yield, harvested area, fertilizer application rate, precipitation, soil organic carbon stock)
- An *inventory flow in the naming convention of an LCI database* (e.g., the input of groundwater in m³, the input of a particular mineral fertilizer type in kg or the output of CO₂ or N₂O emissions in kg)
- Its corresponding *environmental footprint* (in kg CO₂ equivalent, for example)

⁴ <https://epsg.io/4326>

⁵ In this system, the world is sub-divided into roughly 9 million grid cells. Since more than 2/3 of these grid cells are above water—and therefore of no interest to the calculation of crop production systems—we limit the size to about 3 million grid cells. Note that the actual size of the grid cells varies because longitudinal lines converge at the poles. That is, the closer a grid cell to the pole, the smaller the actual area of the grid cell.

A particular lat/lon combination is uniquely identified by a corresponding cell_id. We also store the iso_a2 country abbreviation and the country_name the grid cell belongs to.⁶

Geo-referenced grid-cells.			Country grid-cell belongs to		Attributes (spatial data, inventory flow, or footprint)		
cell_id	lat	lon	iso_a2	country	attribute_1	attribute_2	...
...
2197353	47.625000	10.416667	DE	Germany	att_value	att_value	...
2197354	47.625000	10.500000	DE	Germany	att_value	att_value	...
...
2460844	42.541667	2.916667	FR	France	att_value	att_value	...
2460845	42.541667	3.000000	FR	France	att_value	att_value	...
..

Figure 2: The basic processing structure of geoFootprint; a geo-referenced attribute table.

This data structure facilitates a very efficient, column (or array)-wise processing of the data, i.e., the input to an equation is not an individual grid cell but all grid cell values associated with a particular attribute. Furthermore, the resulting data can be aggregated and analyzed to/at any scale of interest (e.g., political boundaries such as countries, regions or districts, or self-defined artificial boundaries).

2.2. Functional unit

In LCA (life cycle assessment), the **functional unit** is the reference for evaluating products, services and activities on a common basis. The reference flow is the amount of product or activity required to fulfil or satisfy the functional unit. Typically, life cycle inventory (LCI) data rely on a chosen reference flow.

In geoFootprint, results are computed (in the above-mentioned format) with different functional units along the data processing chain:

- First, we process all inventory flows and corresponding metrics (results) with a non-allocated area-based reference which indicates the output of the crop cultivated

⁶ This information helps to spatially-differentiate default AUPD into the grid cell structure but also provide powerful possibilities to aggregate and analyze results in regard to individual countries.

unpackaged, at farm exit gate *per hectare of main-product*. This means that all inventory flows (and corresponding environmental impacts) are exclusively attributed to the main product (barley grains, fresh fruit bunches, etc.), while co-products are not (yet) considered. *If not explicitly state, all computations highlighted in this documentation refer to this functional unit.*

- Next, we transform all inventory flows and corresponding metrics (results) into an allocated, area-based functional unit which indicates the output of the (main-or co-) product cultivated *per hectare of product*. This transformation happens on the web-based platform.
- Finally, we transform all inventory flows and corresponding metrics (results) into an allocated, mass-based functional unit which indicates the output of fresh product, unpackaged, at farm exit gate *per metric ton* of (main-or co-) product. We display this functional unit along the area-based unit on the web-based platform.

2.3. System boundaries

2.3.1. System boundaries for crop production systems

The system boundaries for crop production systems are illustrated in Figure 3.

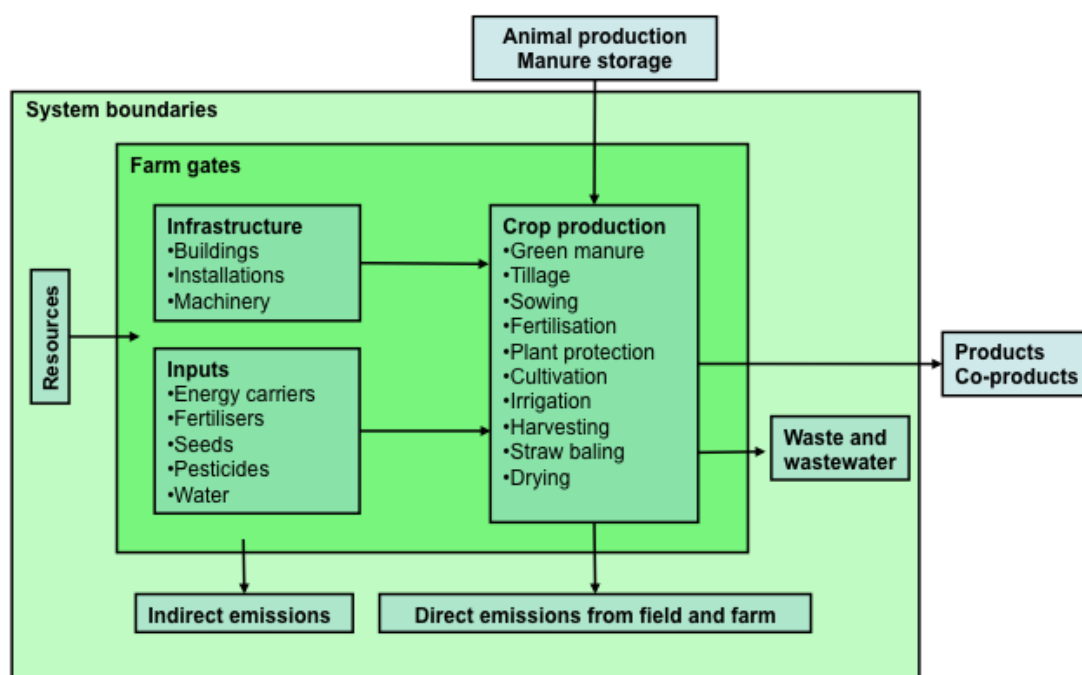


Figure 3: System boundaries for crop production in GeoFootprint. Source: WFLDB Methodological guidelines (2019).

As shown by Figure 3, we include irrigation water energy and infrastructure, as well as the water extraction from nature. Transport of material inputs and on-farm transport are also included. Drying of cereals before storage is included whether it takes place at the farm or not. Also included are wastes and wastewater treatment. Note that by default conventional seeds are used. On a case-by-case basis, treated seeds are modelled (e.g., coating).

We exclude animal traction and post-harvest processes, except drying and post-harvest pest treatment when relevant for proper storage of crop products and specific post-harvest activities taking place at the farm. Further, we do not consider the production and storage of animal manure, CH₄ emissions associated with crop storage, the packaging of output products, the labor, commuting and travels of seasonal workers, administrative work and processes that can reasonably be assumed to contribute to less than 1% of the environmental impact (cut-off criterion), when no data are available (T. Nemecek, 2019).

2.3.2. Crop coverage

Crops integration within the geoFootprint platform directly depends on the availability of spatial data. Spatially explicit and crop-specific data (e.g., average yield, production volume, fertilizer application rate) are indeed all necessary input data for a crop to be integrated within the geoFootprint platform. These input spatial data, being pre-processed data such as satellite images transformed into meaningful variables, require highly specific algorithms and skills to be generated which, hence, directly influence the quantity of available data.

Based on the current data availability detailed in Table 42, we launched geoFootprint in 2020 with 15 crops (Table 1). Using an economic allocation approach with crop specific allocation factors, co-products of the selected crops (e.g., barley straw, barley grain) are integrated in the platform as well. Further developments after the launch of geoFootprint will focus on the integration of additional crops provided required data to be available. In particular, the focus will primarily be on the integration of green coffee and cocoa beans, two key worldwide commodities, for which sources of potential input data could be identified through a first exploratory analysis.

Table 1: geoFootprint crop coverage

Crop (per hectare)	Products (per ton)	Date of integration
Barley	Barley grain	2019

	Barley straw	
Cotton	Cotton fiber Cotton seed	2019
Maize	Maize grain Maize stover	2019
Oil palm	Oil palm fruit bunch	2019
Rice	Rice grain Rice straw	2019
Soybean	Soybean	2019
Peanut	Peanut	2020
Potato	Potato	2020
Rapeseed	Rapeseed	2020
Rye	Rye grain Rye straw	2020
Sorghum	Sorghum grain Sorghum stem	2020
Sugar beet	Sugar beet	2020
Sugarcane	Sugarcane	2020
Sunflower	Sunflower seed	2020
Wheat	Wheat grain Wheat straw	2020
Cocoa	Cocoa bean	post-2020
Green coffee (arabica)	Green coffee bean (arabica)	post-2020
Green coffee (robusta)	Green coffee bean (robusta)	post-2020

2.3.3. Allocation of crop co-products at farm

The use of agricultural products and co-products from crop production systems is not defined in geoFootprint. For example, wheat can be used as food, feed or for production of bioethanol. Straw can either be used as bedding material, for combustion or for production of second-generation bioethanol. Different physical causalities would need to be applied in each of these cases; hence it is not possible to develop a single “physical causality” that fits for all potential applications (T. Nemecek, 2019).

Therefore, *economic allocation* is used by default for crop co- products at the farm. Since only traded products and co-products are addressed, price information is available. We use the price information of the WFLDB database where prices are calculated as average values of the years 2009-2012, when available. This allocation rule applies to main products and co-product (e.g., for co-products at farm grains and straw, oil and press cake) (T. Nemecek, 2019). Table 1 shows the allocation factor and mass relationships between the main and co-products in geoFootprint.

Table 2: Allocation factor, mass relationships and water content of main-and co-products in geoFootprint. The water content refers to both, main- and co-product. Source: ALCIG (2020). * approximated with Wheat, +retrieved from (Myers, 1980).

Crop	Main product	Allocation main product	Co-product	Allocation co-product	Mass ratio (-): main over co-product	Water content (-)	Carbon content (kg C / kg DM)	Rooting depth (m)
Barley	Barley grain	0.76	Barley straw	0.24	1.45	0.11	0.42*	0.7*
Cotton	Cottonseed	0.80	n/a	0	n/a	0.09	0.45	1.35
Maize	Maize grain	0.95	Maize stover	0.05	0.65	0.14	0.44	1.35
Oil palm	Oil palm fruit bunch	1.00	n/a	0	n/a	0.47	0.60	1.00
Rice	Rice grain	0.96	Rice straw	0.04	2.00	0.13	0.48	0.60
Soybean	Soybean	1.00	n/a	0	n/a	0.11	0.68	0.95
Peanut	Peanut	1.00	n/a	0	n/a	0.06	0.58	0.50
Potato	Potato	1.00	n/a	0	n/a	0.78	0.42	0.30
Rapeseed	Rapeseed	1.00	n/a	0	n/a	0.12	0.47	0.90
Rye	Rye grain	0.90	Rye straw	0.10	1.45	0.11	0.42*	0.7*
Sorghum	Sorghum grain	0.43	Sorghum stem	0.57	0.08	0.11	0.42*	1+
Sugar beet	Sugar beet	1.00	n/a	0	n/a	0.77	0.38	0.90
Sugarcane	Sugarcane	1.00	n/a	0	n/a	0.77	0.48	1.50
Sunflower	Sunflower seed	1.00	n/a	0	n/a	0.06	0.84	1.35
Wheat	Wheat grain	0.79	Wheat straw	0.21	1.79	0.15	0.42	0.70

2.4. Inventory data on agricultural activities

Agricultural unit process datasets (AUPD) represent inventories of resources, material and energy flows and other global parameters characterizing the different agricultural activities and practices. In geoFootprint, we use AUPD from two widely used Life Cycle Inventory (LCI) databases:

- **ecoinvent⁷**: ecoinvent is recognized as one of the most complete background LCI databases available, from a quantitative (number of included processes) and a qualitative perspective (e.g., quality of the validation processes, data completeness) (Weidema et al., 2013). This database covers thousands of commodities and industrial processes and is characterized by a high credibility and transparency. Data taken from the ecoinvent database and used in geoFootprint are mostly related to non-agricultural background processes (e.g., electricity, energy, transportation) and pesticide production. We currently work with version 3.5 of the ecoinvent database, published in August 2018.
- **World Food LCA Database (WFLDB)⁸**: The World Food LCA Database (WFLDB) is a global initiative led by Quantis in partnership with leaders in the private food sector aiming to generate emission factors and environmental footprint data. The database includes more than 2,000 datasets—such as fertilizers, cereals, fruits, nuts, coffee, tea and more—covering more than 40 countries. Data taken from the WFLDB and used in geoFootprint are mostly related to agricultural processes (e.g., fertilizer production, field practices, irrigation technologies). We currently work with version 3.5 of the WFLDB, published in March 2020.

An AUPD typically records, in distinct reference to the yield, the following exchange flows.

- **Resource flows**: These cover all direct resource inputs from nature such as
 - Land occupation flows, i.e., the amount of land occupied per year to produce a certain crop yield.

⁷ www.ecoinvent.org

⁸ www.quantis-intl.com/wflldb

- Land transformation flows, i.e., the amount of land transformed to produce a certain crop yield.
- CO₂ uptake from air flow, i.e., the amount of CO₂ uptake during plant growth.
- Energy in biomass, i.e., the amount of energy stored in the harvested crop yield.
- Water flows, i.e., the amount of water from river, well, or unspecified origin obtained from nature.
- Inputs from technosphere: These cover all man-made input flows such as
 - Mineral fertilizers flows, i.e., the amount and type of nitrogen (N), phosphorous (P) and potassium (K) applied in the form of chemical fertilizers.
 - Organic fertilizers flows, i.e., the amount and type of organic fertilizer applied (e.g., manure from different animal categories, compost, green waste).
 - Supplement flows, i.e., the application of minerals such as dolomite, lime, sulfur, etc.
 - Field operations flows, i.e., the type (ploughing, sawing, harvesting, tilling, etc.) and amount of field operations applied.
 - Irrigation flows, i.e., the amount of water used for irrigation and the technology used (e.g., surface irrigation, drip irrigation).
 - Pesticide flows, i.e., the amount and type of pesticide applied.
 - Energy requirements flows in the form of heat, gas or electricity associated with drying of feedstocks.
- Outputs to technosphere: These cover the man-made output flows such as
 - Waste treatment flows, i.e., the removal of organic and general waste from the farm and its treatment pathway.
 - The output of the main- and co-product.
- Emissions (output) flows. Emissions are distinguished according to the following compartments.
 - Emission flows to air, e.g., dinitrogen monoxide (N₂O) emissions associated with the fertilizer application, CO₂ emissions associated with land use change, etc.
 - Emission flows to soil, e.g., pesticide emissions, heavy metal emissions related to the application of mineral and organic fertilizers, etc.

- Emission flows to water flows, e.g., nitrate (NO_3^-) or phosphate (PO_4^{3-}) emissions to groundwater.

Figure 4 shows a screenshot of an AUPD of barley grain production in France taken from SimaPro software. Note that emission flows are present but not visible in the screenshot.

Products									
Outputs to technosphere: Products and co-products		Amount	Unit	Quantity	Allocation %	Comment			
Barley grain, non-irrigated, at farm (WFLDB 3.5)/FR U		6238	kg	Mass	75.6 %	Economic allocation based on FEFAC (2015)			
Barley straw, non-irrigated, at farm (WFLDB 3.5)/FR U		4302	kg	Mass	24.4 %	Economic allocation based on FEFAC (2015)			
Add									
Outputs to technosphere: Avoided products		Amount	Unit	Distribution	SD2 or				
Add									
Inputs									
Inputs from nature		Sub-compartment	Amount	Unit	Distribution	S	I	I	Comment
Carbon dioxide, in air		in air	8239.801	kg	Lognormal	-	-	-	(2,2,1,1,na)
Energy, gross calorific value, in biomass			96024.653	MJ	Lognormal	-	-	-	(2,2,1,1,na)
Occupation, annual crop, non-irrigated		land	10000	m2a	Lognormal	-	-	-	(2,1,1,1,na)
Transformation, from annual crop, non-irrigated		land	10000	m2	Lognormal	-	-	-	(2,1,1,1,na)
Transformation, to annual crop, non-irrigated		land	10000	m2	Lognormal	-	-	-	(2,1,1,1,na)
Add									
Inputs from technosphere: materials/fuels		Amount	Unit	Distribution	S	I	I	Comment	
Barley seed, for sowing (GLO) market for Cut-off, U		133.9	kg	Lognormal	-	-	-	(2,1,1,1,na) - LEAP Database (FAO 2017)	
Tillage, currying, by weeder (GLO) market for Cut-off, U		1.5	ha	Lognormal	-	-	-	(2,1,1,1,na) - LEAP Database (FAO 2017)	
Tillage, ploughing (GLO) market for Cut-off, U		1.0	ha	Lognormal	-	-	-	(2,1,1,1,na) - LEAP Database (FAO 2017)	
Sowing (GLO) market for Cut-off, U		1.0	ha	Lognormal	-	-	-	(2,1,1,1,na) - LEAP Database (FAO 2017)	
Fertilising, by broadcaster (GLO) market for Cut-off, U		2.3	ha	Lognormal	-	-	-	(2,1,1,1,na) - LEAP Database (FAO 2017)	
Combine harvesting (GLO) market for Cut-off, U		1.0	ha	Lognormal	-	-	-	(2,1,1,1,na) - LEAP Database (FAO 2017)	
Pesticide, unspecified, mix for cereal crops, at plant (WFLDB 3.5)/GLO U		1.9	kg	Lognormal	-	-	-	(2,1,1,1,na) - LEAP Database (FAO 2017)	
Application of plant protection product, by field sprayer (GLO) market for Cut-off, U		1.5	ha	Lognormal	-	-	-	(2,1,1,1,na) - LEAP Database (FAO 2017)	
Manure, solid, cattle (GLO) market for Cut-off, U		3882.353	kg	Lognormal	-	-	-	(2,1,1,1,na) - LEAP Database (FAO 2017)	
Solid manure loading and spreading, by hydraulic loader and spreader (GLO) market for Cut-off, U		3882.353	kg	Undefined	-	-	-	(2,1,1,1,na) - LEAP Database (FAO 2017)	
Urea, as N, at plant (WFLDB 3.5)/RER U		20.9	kg	Lognormal	-	-	-	(2,1,1,1,na) - LEAP Database (FAO 2017)	
Ammonium nitrate (AN), as N, at plant (WFLDB 3.5)/RER U		19.8	kg	Lognormal	-	-	-	(2,1,1,1,na) - LEAP Database (FAO 2017)	
Urea ammonium nitrate (UAN), as N, at plant (WFLDB 3.5)/RER U		15.4	kg	Lognormal	-	-	-	(2,1,1,1,na) - LEAP Database (FAO 2017)	
Monoammonium phosphate (MAP), as N, at plant (WFLDB 3.5)/RER U		2.2	kg	Lognormal	-	-	-	(2,1,1,1,na) - LEAP Database (FAO 2017)	
Calcium ammonium nitrate (CAN), as N, at plant (WFLDB 3.5)/RER U		26.4	kg	Lognormal	-	-	-	(2,1,1,1,na) - LEAP Database (FAO 2017)	
Ammonium sulfate, as N (GLO) market for Cut-off, U		3.3	kg	Lognormal	-	-	-	(2,1,1,1,na) - LEAP Database (FAO 2017)	

Figure 4: Screenshot illustrating the AUPD associated with barley cultivation in France from the WFLDB. The first lines correspond to the name and amounts of the different co-products for one hectare of cultivated land over one year, while the following lines correspond to all datasets required to build the inventory of the cultivation system (e.g., seeds, agricultural machinery, pesticides, fertilizers) with their respective quantity.

2.5. Data representativeness

To provide crop-specific and regionalized AUPDs of key commodities with a global coverage, geoFootprint relies on a large set of input data. These data come from several different public sources and other life-cycle inventory databases with the aim to form the most comprehensive and consistent set of data from an inventory perspective (i.e., to account for all possible material, energy and emission flows) and from a spatial perspective (i.e., maximizing the use of spatially explicit data as much as possible) while using the most up-to-date and globally harmonized possible data.

Data representativeness is a major challenge given the amount of data needed to feed the geoFootprint regionalization engine. Ideally, all data should be gathered at the working resolution of ~10x10 km with global coverage, and representative of multi-annual averages over the 4 to 5 the most recent years, as to account for time variability of agricultural practices, productivity and climate conditions. However, due to the lack of data, this is currently not possible.

Therefore, some strategic decisions regarding the selection of input data were taken. We made the choice to prioritize geographically homogeneous and consistent data with global coverage. While more precise and complete data are available for specific regions and could potentially be used as input data (and regionally increase the data accuracy), their limited spatial extent did not fit with the objective of having a globally consistent and homogenous data coverage.

Aware of these limitations, the next developments of geoFootprint will aim at first improving both the spatial and temporal resolutions by including other more recent data sources. By exploring new data sources and new data acquisition methodologies, future developments will aim to increase the spatial and temporal representativeness of key input data variables, while at the same time ensuring a globally consistent and homogenous data coverage.

The Appendix (see 13.1) aims to provide an overview of all data inputs being used in geoFootprint and highlights the diversity of data types, data resolutions, known limitations and potential future developments.

2.5.1. Geographical coverage

Regarding the geographical resolution, a current limitation of the model is the lack of available regionalized data on agricultural practices (field operations, macro-nutrients, pest treatments, etc.). Best available data are currently available at country level for some countries and, otherwise, at global level. Hence, in order to satisfy the global coverage requirement, country-level and global-level data have been extrapolated and applied to all corresponding pixels depending on the available data (see chapter 3.2.2).

2.5.2. Temporal coverage

Regarding the temporal resolution, some of the input data, such as those derived from EarthStat, are representative of year 2000. While this is not necessarily a concern for all crops, it can create a certain bias for crops whose cultivated area has largely evolved since then. Bearing this in mind, we have extrapolated the spatially explicit yield and area data (and consequently production volume data) with a country-centric approach using data from the FAOSTAT database (see chapter 3.1.1).

2.5.3. Technological coverage

The available spatial data and the default AUPD in LCI databases, both determine our ability to model different agricultural production technologies or cultivation practices such as organic, conventional, extensive, or intensive crop production.

While AUPDs in LCI databases would sometimes allow for a rudimentary distinction of agricultural cultivation practices, EarthStat—the data sources which provides the spatial distribution of crops—does not distinguish between different cultivation practices. Therefore, agricultural cultivation practices modeled in geoFootprint all refer to conventional production.

3. Key elements in regionalized LCA

Regionalized life cycle assessment (LCA) modeling refers to the modeling procedure that automatically generates regionalized AUPDs on the grid cell level—in our case at a resolution of ~10x10km—under consideration of various spatial input parameters and transforms them into regionalized footprints.

Due to the current limitation in the availability of spatial data, not all inventory flows can be computed on the basis of spatially explicit data. Therefore, regionalized inventory modeling often requires a close pairing of the information already present in existing AUPD retrieved from existing LCI databases (e.g., fertilizer types applied, irrigation technologies used, manure types applied) with the novel information provided in the form of spatially explicit data (amount of N applied, cubic meter of irrigation water, N of manure).

This chapter elaborates the data elements required for regionalized LCA modeling in the context of geoFootprint and associated pre-processing steps.

3.1. Spatial attribute table

The spatial data used in geoFootprint are based on a raster data format, i.e., an image composed of very large number of pixels. Each pixel or grid cell refers to a precise set of geographic coordinates and covers a defined area of the earth—the spatial resolution—for a specific attribute (e.g., crop yield, N fertilizer input). Figure 5 shows an example of spatial input data retrieved from EarthStat⁹, which serves spatial datasets for harvested area and yield, nutrient application, etc. on a global scale. Note that the data from EarthStat is crop-specific, i.e., always refers to a particular crop.

⁹ For further details please visit: <http://www.earthstat.org/>

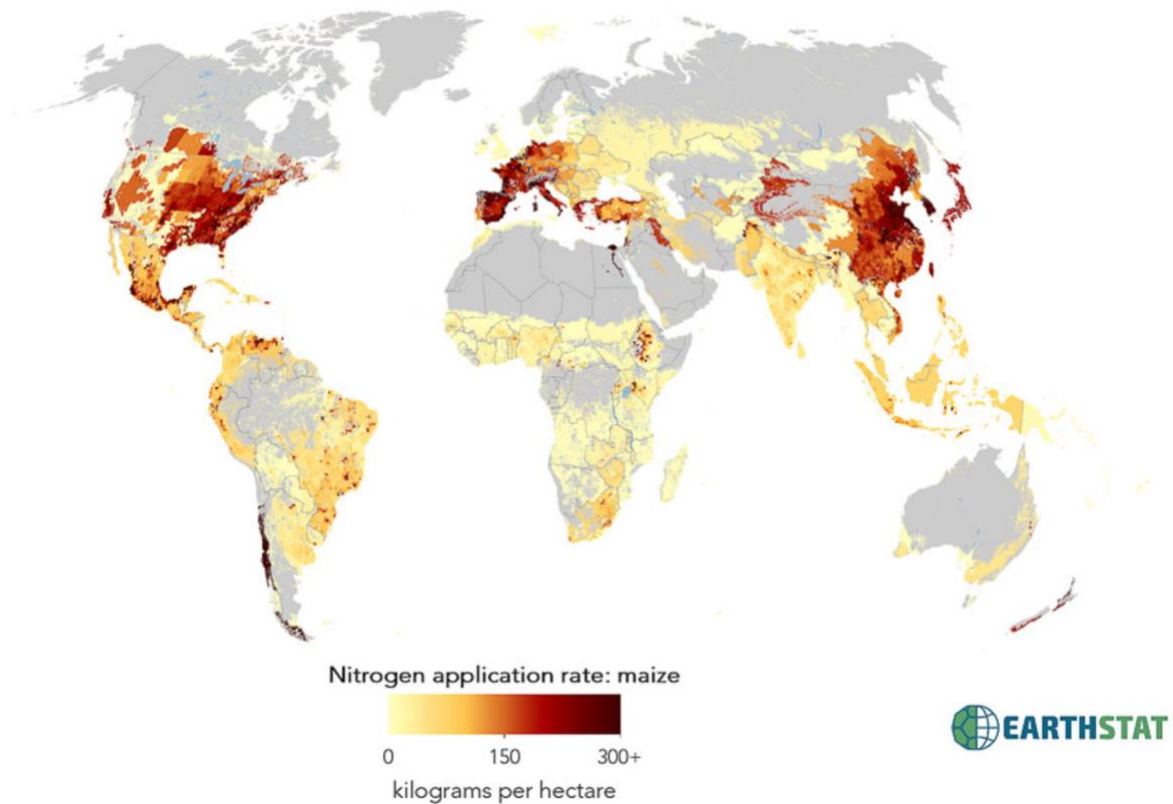


Figure 5: Screenshot from the EarthStat website displaying the Nitrogen application rate for maize data layer used as input data in geoFootprint. Spatial resolution: 5 arc-minutes (~10x10 km); Spatial and temporal extent: global / 2000. Source: Mueller et al. (2012).

We transform all spatial raster data into a geo-referenced tabular data structure which we call “Spatial attribute table” (see Figure 6). Each *row* of the spatial attribute table represents the corresponding grid cell values of all spatial datasets. Each *column* of the spatial attribute table represents all values of a spatial attribute for the entire globe, i.e., for all relevant latitude-longitude combinations.

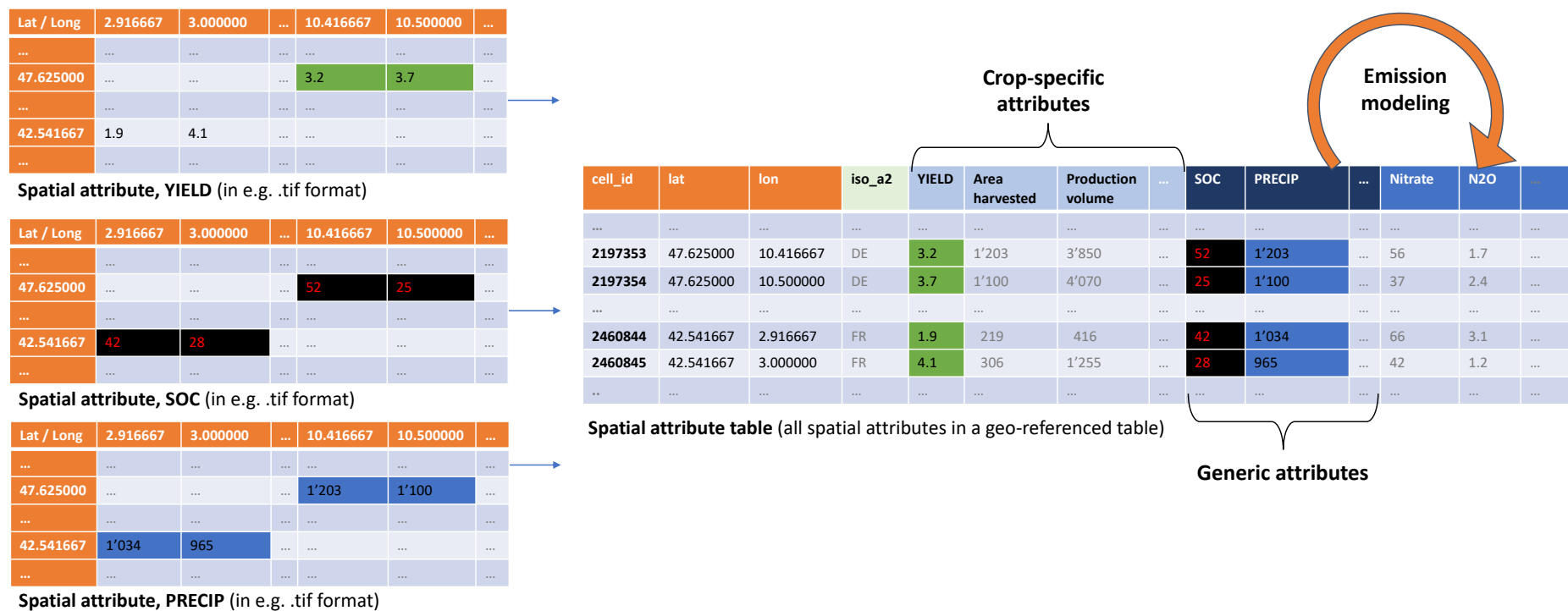


Figure 6: From individual spatial raster data (in mesh grid form) to a geo-reference spatial attribute table. The example shows how YIELD (green), PRECIP (precipitation-blue) and SOC (soil organic carbon stock, black) appears in the two different formats. The actual spatial attribute table accommodates more than 60 attributes prior to any emission modeling.

We distinguish between two kind of spatial attributes, namely

- *Crop-specific attributes*, which are associated with a particular crop. Yield, area harvested, production volume, mineral fertilizer application rate in the form of NPK, cultivation period, etc. are examples of crop-specific spatial attributes.
- *Generic attributes*, which are not associated with a particular crop. Precipitation, soil organic carbon, soil bulk density and soil clay content are examples of generic spatial attributes. (See Annex 13.1.)

In addition, we use *global default values* (non-spatial data) in the default computation—which works without any user input—in order to consider average management practices.

The spatial attribute table serves as foundation for emission modeling. That is, we compute new columns—emissions or intermediate flows—on the basis of emission models (see chapter 3.3) which use existing—crop-specific and generic—spatial attributes and global default values as an input. Annex 13.1 provides a comprehensive overview of the spatial data and the global default values used in geoFootprint.

3.1.1. Extrapolation of spatial attributes

As mentioned in the data representativeness chapter (see chapter 2.5) input data for spatial attributes were collected with the objective to prioritize geographically homogeneous and consistent data with global coverage. Following and applying this principle also meant - in some cases - considering and including less recent data sources in our set of input data to prioritize data comprehensiveness and consistency.

While several input data do not or very slightly change over time, some other data might be more subject to temporal variability, and hence, potentially influence the crop environmental footprint over time. To provide the most recent possible footprint picture of our 15 key crops, we developed an extrapolation procedure using long-time monitored and latest available country level data from the Food and Agriculture Organization of the United Nations (FAO) database FAOSTAT (FAO, 2020) together with grid cell level data from EarthStat (Monfreda et al., 2008) to improve the temporal representativeness of some key spatial attributes. We applied the procedure to two key crop-specific spatial attributes: the yield and the harvested area.

3.1.1.1. Description of the extrapolation procedure

Figure 7 describes the applied extrapolation process in an illustrative manner.

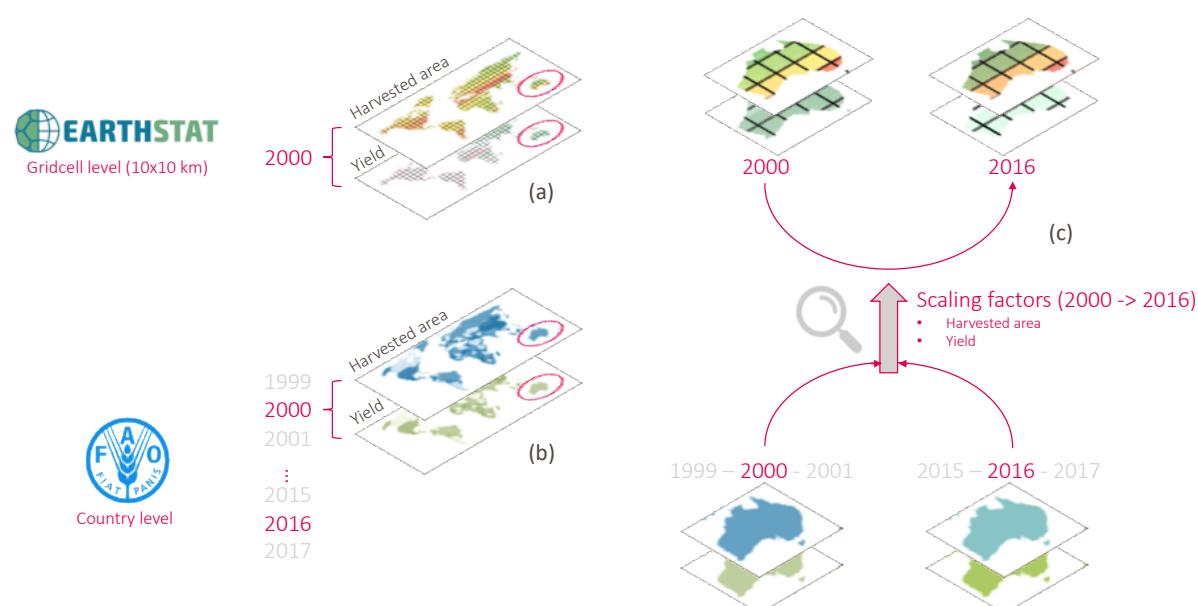


Figure 7: Yield and harvested area extrapolation procedure – illustrative description.

For a given crop and a given attribute, i.e., the yield or the harvested area, we combine grid cell level data from EarthStat representative of the year 2000 (Figure 7a) with annual country level data from the FAOSTAT database for the years 2000 to 2017 (Figure 7b). We use both data sources together and combine them to know 1) in which grid cell a specific crop is produced in the year 2000 (EarthStat data) and 2) to compute changes (i.e., increase or decrease) for the given attribute between the year 2000 and more recent years at country level (FAO data).

We use then both information to compute the extrapolated yield and harvested area values. Figure 8 details the extrapolation procedure step-by-step for the example of barley production in Switzerland.

1) FAO data extraction

	Area Code	Area	Item Code	Item	Element Code	Element	Year Code	Year	Unit	Value	Flag
1502376	211	Switzerland	44	Barley	5312	Area harvested	1999	1999	ha	48941.0	NaN
1502377	211	Switzerland	44	Barley	5312	Area harvested	2000	2000	ha	45741.0	NaN
1502378	211	Switzerland	44	Barley	5312	Area harvested	2001	2001	ha	43845.0	NaN
1502392	211	Switzerland	44	Barley	5312	Area harvested	2015	2015	ha	28753.0	NaN
1502393	211	Switzerland	44	Barley	5312	Area harvested	2016	2016	ha	28938.0	NaN
1502394	211	Switzerland	44	Barley	5312	Area harvested	2017	2017	ha	28455.0	NaN

2) Computation of the scaling factors

Crop	AreaCode	Country	iso_a2	cropiso_a2	max_year	A ScalingFactor (FAO: 2000-2016)	Actual A [ha]	Y ScalingFactor (FAO: 2000-2016)	Actual Y [tonnes/ha]
179	Barley	211.0	Switzerland	CH	BarleyCH	2017.0	0.622	1.18	6.5738

$SF \text{ (Harvested Area)}$ $SF \text{ (Yield)}$

3) Link scaling factors to pixels

cell_id	long	lat	iso_a2	country_name	yield_barley	ha_barley	hf_barley	pv_barley	Crop	cropiso_a2	A ScalingFactor (FAO: 2000-2016)	Y ScalingFactor (FAO: 2000-2016)	yield_barley_New	ha_barley_New
2188688	8.500000	47.791667	CH	Switzerland	4.598353	112.772438	0.019567	518.567505	Barley	BarleyCH	0.622	1.18	5.428057	70.144456
2188689	8.583333	47.791667	CH	Switzerland	4.859858	106.564178	0.018492	504.879089	Barley	BarleyCH	0.622	1.18	5.734832	66.282919
2188690	8.666667	47.791667	CH	Switzerland	5.121361	155.413727	0.026963	795.481995	Barley	BarleyCH	0.622	1.18	6.043206	96.667338
2193008	8.416667	47.708333	CH	Switzerland	4.598353	65.410408	0.011330	300.780182	Barley	BarleyCH	0.622	1.18	5.428057	40.685274
2193009	8.500000	47.708333	CH	Switzerland	4.859858	151.125793	0.026170	771.512085	Barley	BarleyCH	0.622	1.18	5.734832	94.002444

$New = SF \times old$

e.g. yield extrapolation

Figure 8: Yield and harvested area extrapolation procedure – detailed description.

As shown on Figure 8, we first extract country specific (e.g., Switzerland) and crop specific (e.g., Barley) data from the FAOSTAT database for the given attribute (e.g., harvested area). Averages are computed for the years 2000 and 2016 by considering the average over the years 1999-2001 and 2015-2017 respectively (note that 2017 was the last available year at the time of the data extraction). Then, we compute a scaling factor for the given crop and the given attribute as the ratio of the three-year average of the year 2016 over the average for the year 2000. Finally, the extrapolated attribute value is computed for each grid cell falling within the considered country by multiplying the initial 2000 EarthStat grid cell level value with the computed scaling factor. We apply this extrapolation procedure for across all 15 key crops available in geoFootprint.

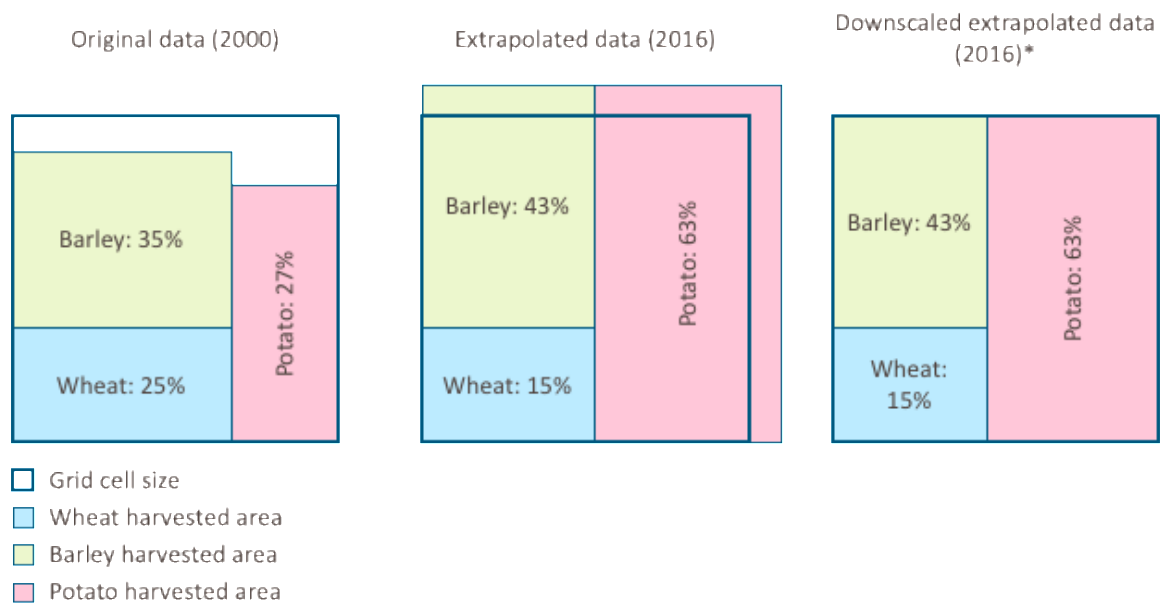
We finally compute the extrapolated production volume by multiplying the extrapolated yield with the extrapolated harvested area.

Note that we also apply this extrapolation procedure to most¹⁰ other crops in order to consistently compute the relative relevance of the harvested area of each crop for all grid cells. We use this information to consistently allocate tree cover losses (see section 10.1.4) and generic manure availability (see section 6.1) according to the relative relevance of each crop per grid cell, i.e., the shared-responsibility approach.

3.1.1.2. Post-processing: rescaling of extrapolated data

We verify the extrapolated data to identify potential discrepancies. For less than 1% of all grid cells (about 11,000 grid cells) where production of at least 1 of the 15 crops is taking place, the sum of the harvested area of all 15 key crops exceeded a total harvested area fraction of 1, i.e., a harvested area larger than the actual size of the grid cell. Although this is at least in part a consequence of the modeling assumption in the initial data set—double cropping of a particular crop in one year is accounted by doubling the harvested area of that crop and therefore can lead to a harvested area fraction greater than 1 (EarthStat, 2020)—we decided to re-scale the harvested area of each crop produced in each of those grid cells proportionally to the total harvest crop area. Figure 9 shows how we rescaled the total harvested area for all 15 crops.

¹⁰ Out of the 175 crops available from EarthStat, we computed the extrapolated area and yield for 156 crops for which FAO data are available. These crops account for more than 99% of the total harvested area.



*only if the total harvested area fraction of extrapolated data leads to a total harvested fraction area larger than 100% of the pixel size

Figure 9: Downsampling principle applied to the extrapolated data with total harvested area fraction larger than 100% of the grid cell size.

3.1.2. Preprocessing of spatial attributes

We preprocess the spatial attribute table in order to exclude unrealistic and irrelevant data, and to enhance the visual rendering of the data on the web-based platform.

3.1.2.1. Excluding grid cells with a non-significant production volume

Grid cells with a very small production volume—sometimes even lower than 1 metric ton per grid cell—distort the views over the different metrics on the web-based platform. Due to limited possibilities to treat these distortions via rendering on the web-based platform, we removed grid cells whose production volume (pv) is below a certain threshold (see Table 3).

We performed a detailed analysis of the global, crop-specific production volume data from EarthStat (Monfreda, 2008). Our analysis shows that the majority of grid cells only provide a very small percentage of the global production volume. For example, 5% of the grid cells produce about 50% of the global production volume for barley (see Figure 10).

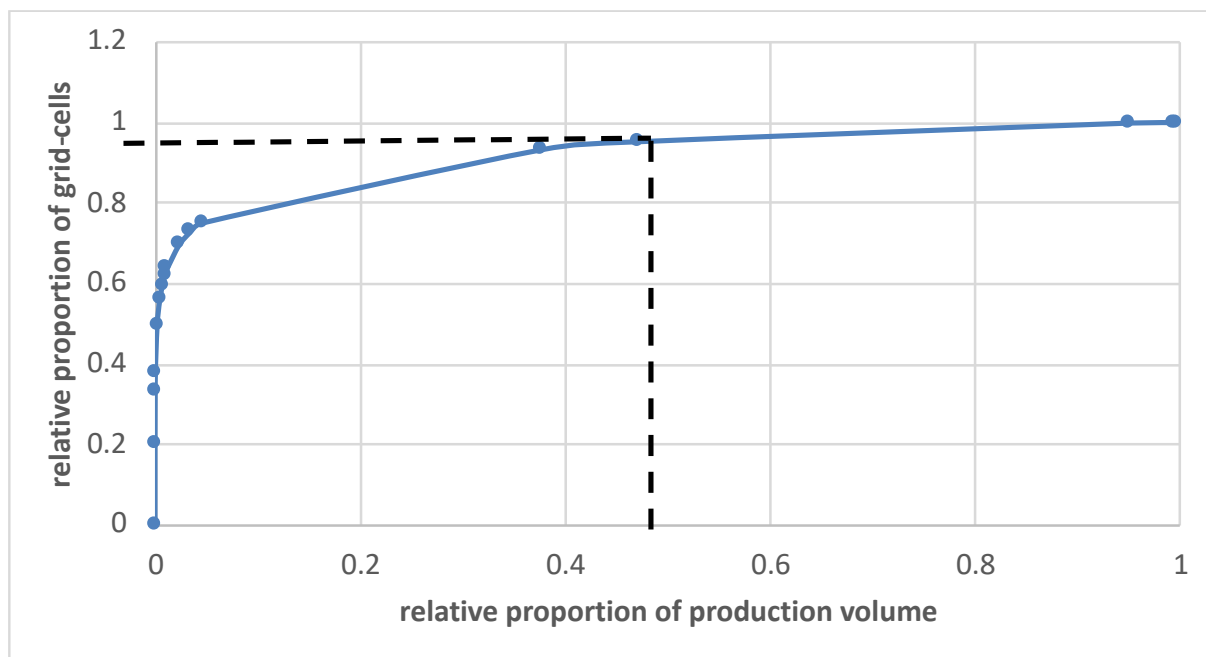


Figure 10: Lorenz curve showing how the total production volume is distributed across all grid cells. The dashed line shows that about 95% of the grid cells produce only about 50% of the global pv for barley, while 5% of the grid cells produce about 50% of the production volume.

Also for global barley cultivation, Figure 11 shows the relative number of grid cells and production volume excluded from the analysis as a function of a threshold on the production volume produced per grid cell. The larger the threshold, the larger the number of grid cells and corresponding production volume excluded.

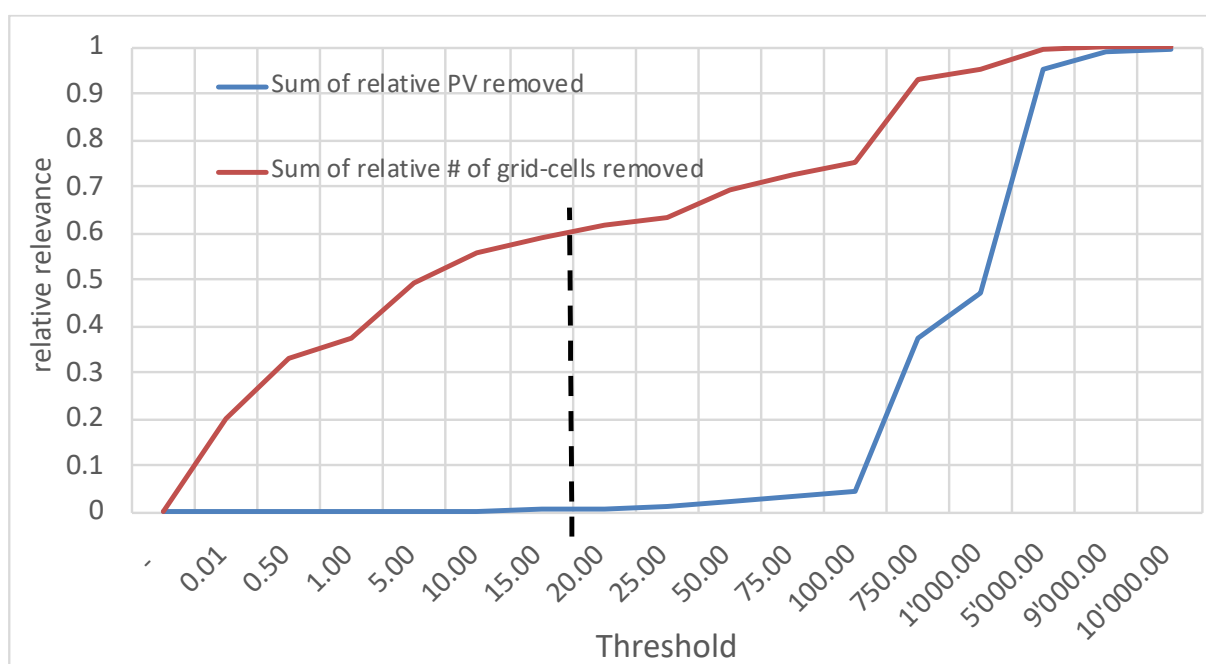


Figure 11: The relative relevance of the grid cells and production volume removed as a function of the threshold in metric ton production volume. Excluding all grid cells with a production smaller than 20 metric ton removes about 60% of the grid cells but virtually does not affect the global production volume.

As shown by Figure 11, about one third of the global grid cells have a production volume that is less than or equal to 1.00 metric ton of barley. To put this into perspective, a size of a grid cell at the equator is about 8,585 hectares. Such grid cells have virtually no global relevance in terms of production volume (< 0.01%) and no important constraints on local supply. We observe these patterns across all crops.

Table 3 details the final crop-specific threshold criterion (in pv per grid cell) and the corresponding percentage of grid cells and production volume we remove prior to the bulk calculation.

Table 3: Crop-specific threshold for the removal of grid cell based on their production volume (pv) and the corresponding effect on the total number of grid cells and the global production volume.

Crop	Threshold [metric ton per grid cell]	Relative PV removed [%]	Relative # of grid cells removed [%]
Barley	20.00	0.92%	54%
Wheat	20.00	0.17%	42%
Maize	50.00	0.42%	44%
Potato	20.00	0.38%	43%
Sugarcane	20.00	0.12%	36%
Sunflower	20.00	2.63%	79%
Rapeseed	20.00	1.81%	67%
Oil palm	20.00	0.27%	62%
Cotton	20.00	1.88%	72%
Rice	50.00	0.40%	56%
Rye	20.00	9.03%	80%
Sugar beet	20.00	0.18%	43%
Soybean	20.00	0.41%	56%
Groundnut (peanut)	20.00	2.99%	71%
Sorghum	20.00	2.73%	73%

We typically remove about 50% of the grid cells but rarely more than 1% of the production volume with the selected thresholds. Notable exceptions are rye, sunflower, groundnut, sorghum and rapeseed. This significantly reduces the number of grid cells to be visualized (rendered) on the web-based platform. As a result, we do not display the entire global production pattern of a crop but only an excerpt.

3.1.2.2. Outlier identification and manipulation for N, P and K application rates

A close analysis of the spatial data shows that N, P and K application rate data from EarthStat contain extremely high values for a very small proportion of grid cells, e.g., values which exceed application rates of 1,000 kg of N per hectare. In order to avoid these unrealistic application rates distorting results—particularly N and P application rate drive N and P leaching and corresponding emissions—we developed an outlier identification and manipulation module, which overwrites identified outlier values with the arithmetic mean value of the closest 20 grid cells. The rather large number of 20 grid cells is used to compute a corrected value that is rather robust.

We identify every value larger than or equal to the 99.9th quantile as an outlier. Table 4 shows the size of the threshold for N, P and K fertilizer application rates and the corresponding total and relative number of corresponding grid cells that qualify as outliers.

Table 4: Crop-specific correction of grid cell based on their 99.9th quantile. Every value larger than the 99.9th quantile is corrected, i.e., overwritten with the arithmetic mean value of the closest 20 grid cells.

Crop	Metric	Threshold [in kg per hectare]	Grid cells corrected [#]	Grid cells corrected [%]	Grid cells with a value [#]
Barley	N app rate	381	952	0.10%	951,040
Wheat	N app rate	369	952	0.10%	951,040
Maize	N app rate	408	952	0.10%	951,040
Potato	N app rate	485	952	0.10%	951,040
Sugarcane	N app rate	342	952	0.10%	951,040
Sunflower	N app rate	223	952	0.10%	951,040
Rapeseed	N app rate	330	952	0.10%	951,040
Oil palm	N app rate	177	952	0.10%	951,040
Cotton	N app rate	314	952	0.10%	951,040
Rice	N app rate	377	952	0.10%	951,040
Rye	N app rate	283	952	0.10%	951,040
Sugar beet	N app rate	375	952	0.10%	951,040
Soybean	N app rate	225	952	0.10%	951,040
Groundnut	N app rate	245	952	0.10%	951,040
Sorghum	N app rate	284	952	0.10%	951,040
Barley	P app rate	136	527	0.10%	542,088
Wheat	P app rate	120	636	0.10%	644,179
Maize	P app rate	148	627	0.09%	728,704
Potato	P app rate	210	631	0.09%	680,370
Sugarcane	P app rate	128	335	0.10%	334,223
Sunflower	P app rate	58	449	0.10%	464,690
Rapeseed	P app rate	112	327	0.08%	410,204

Oil palm	P app rate	67	200	0.10%	203,008
Cotton	P app rate	134	374	0.10%	377,878
Rice	P app rate	180	447	0.10%	449,869
Rye	P app rate	98	347	0.10%	358,771
Sugar beet	P app rate	248	226	0.10%	229,945
Soybean	P app rate	92	467	0.10%	489,126
Groundnut	P app rate	129	447	0.10%	449,661
Sorghum	P app rate	150	502	0.10%	510,362
Barley	K app rate	124	0	0.00%	409,479
Wheat	K app rate	124	0	0.00%	520,558
Maize	K app rate	124	173	0.03%	617,621
Potato	K app rate	282	177	0.03%	553,155
Sugarcane	K app rate	118	196	0.07%	296,924
Sunflower	K app rate	142	145	0.03%	437,781
Rapeseed	K app rate	86	357	0.09%	386,587
Oil palm	K app rate	248	0	0.00%	192,596
Cotton	K app rate	113	145	0.05%	311,407
Rice	K app rate	66	0	0.00%	391,058
Rye	K app rate	124	0	0.00%	327,355
Sugar beet	K app rate	190	0	0.00%	195,595
Soybean	K app rate	131	118	0.03%	451,810
Groundnut	K app rate	76	154	0.04%	376,243
Sorghum	K app rate	124	0	0.00%	349,087

As shown by Table 4, the number of corrected grid cells correspond to about 1 per million or less which is reasonable considering we use the 99.9th quantile to identify outliers. For N application rate, the number of grid cells corrected always amounts to exactly 952 grid cells. We speculate that this results from the fact that EarthStat used the same distribution functions to model the spatially explicit N fertilizer application rate across all crops.

3.1.2.3. Preprocessing of soil property information

We use the spatially explicit soil property information from ISRIC soil grids at the pre-aggregated 5 arc-minutes (~10x10 km) resolution (Hengl et al., 2014) for the global consideration of soil property information. ISRIC provides individual data layers for different soil depth (see Table 5).

We consider the first 30 cm of the soil surface in our calculations. In order to arrive at a representative value for this depth, we calculate the arithmetic mean of the 3 to 4 individual

data layers outlined in Table 5. Equation 1 shows the general equation used to calculate a representative value for a soil depth of 30 cm.

$$S_p = \frac{1}{n} \sum_i^n x_{ip} \quad (1)$$

with

- p representing the soil property (coarse fragments, bulk density, etc.). The detailed unit of the soil properties are detailed in Table 5 and n representing “4” for all soil properties, since the four values of the four different depths are averaged.

Note that we do not precompute soil organic carbon stock because ISRIC offers a ready-to-use soil organic carbon stock value for the upper 30 cm of the soil pre-aggregated to our resolution of 5 arc minutes. Further, we compute the fraction of the soil that is free of any coarse fragments, $SfoCf$, i.e., the inverse fraction of the property coarse fragments (see equation 2). We use this information for the calculation of Norg in section 9.1.

$$SfoCf = \frac{100 - Cf}{100} \quad (2)$$

with

- Cf representing the average coarse fragments in percent.

Following the ISRIC soil grids recommendation¹¹, we convert the average soil $ph \times 10$ in KCl to ph values by dividing it by 10.

¹¹ See <https://data.isric.org/geonetwork/srv/api/records/d6633978-ae41-4529-b2c7-22560213ad76>

Table 5: Raw soil property data from ISRIC soil grids used in the regionalization engine (Hengl, 2014). According to common practice, we consider the soil properties of the first 30 cm, and sometimes the first 50 cm of the topsoil only. We pre-process and aggregate the layers available for each attribute into one layer.

ATTRIBUTE_TITLE	ATTRIBUTE_UNITS_OF_MEASURE
Bulk density (fine earth) in kg / cubic-meter at depth 0.00 m	kg / cubic-m
Bulk density (fine earth) in kg / cubic-meter at depth 0.05 m	kg / cubic-m
Bulk density (fine earth) in kg / cubic-meter at depth 0.15 m	kg / cubic-m
Bulk density (fine earth) in kg / cubic-meter at depth 0.30 m	kg / cubic-m
Bulk density (fine earth) in kg / cubic-meter at depth 0.50 m	kg / cubic-m
Clay content (0-2 micro meter) mass fraction in % at depth 0.00 m	percent
Clay content (0-2 micro meter) mass fraction in % at depth 0.05 m	percent
Clay content (0-2 micro meter) mass fraction in % at depth 0.15 m	percent
Clay content (0-2 micro meter) mass fraction in % at depth 0.30 m	percent
Clay content (0-2 micro meter) mass fraction in % at depth 0.50 m	percent
Coarse fragments volumetric in % at depth 0.00 m	percent
Coarse fragments volumetric in % at depth 0.05 m	percent
Coarse fragments volumetric in % at depth 0.15 m	percent
Coarse fragments volumetric in % at depth 0.30 m	percent
Coarse fragments volumetric in % at depth 0.50 m	percent
Soil organic carbon stock in tons per ha for depth interval 0.00 m - 0.30 m	tonnes / ha
Soil organic carbon content in g fine earth fraction at depth 0.00 m	g / kg
Soil organic carbon content in g fine earth fraction at depth 0.05 m	g / kg
Soil organic carbon content in g fine earth fraction at depth 0.15 m	g / kg
Soil organic carbon content in g fine earth fraction at depth 0.30 m	g / kg
Soil organic carbon content in g fine earth fraction at depth 0.50 m	g / kg
Soil pH x 10 in KCl at depth 0.00 m	index*10
Soil pH x 10 in KCl at depth 0.05 m	index*10
Soil pH x 10 in KCl at depth 0.15 m	index*10

Soil pH x 10 in KCl at depth 0.30 m	index*10
Soil pH x 10 in KCl at depth 0.50 m	index*10
Silt content (2-50 micro meter) mass fraction in % at depth 0.00 m	percent
Silt content (2-50 micro meter) mass fraction in % at depth 0.05 m	percent
Silt content (2-50 micro meter) mass fraction in % at depth 0.15 m	percent
Silt content (2-50 micro meter) mass fraction in % at depth 0.30 m	percent
Silt content (2-50 micro meter) mass fraction in % at depth 0.50 m	percent
Sand content (50-2000 micro meter) mass fraction in % at depth 0.00 m	percent
Sand content (50-2000 micro meter) mass fraction in % at depth 0.05 m	percent
Sand content (50-2000 micro meter) mass fraction in % at depth 0.15 m	percent
Sand content (50-2000 micro meter) mass fraction in % at depth 0.30 m	percent
Sand content (50-2000 micro meter) mass fraction in % at depth 0.50 m	percent

3.2. Default crop inventory table

The AUPDs from LCI database represent crop production on a country or global level. This does *not* match the native geoFootprint grid cell resolution of $\sim 10 \times 10$ km. We therefore “cast” the available dataset (or geographic instances representing crop production at the country level) associated with a particular crop into the common geoFootprint grid cell structure. We call this procedure *spatial differentiation*. The process of spatial differentiation does not generate new data. In fact, it duplicates the inventory flow values associated with a geographical domain (e.g., country) for all grid cells located in that geographical domain.

The resulting default crop inventory table accommodates all inventory flows used across all geographical instances (AUPD) associated with a crop, and consequently we generate such a table for every major crop in geoFootprint. The grid cells, located in a particular country, hold the exact same value for a particular inventory flow (see Figure 12).

Geo-referenced grid-cells (simplified)

Grid-cell belongs to

Natural resource flows

Technosphere flows

Emissions flows

id	lat	lon	iso_a2	Country	Land occupation, arable [m2*a]	Water, well [m3]	...	Ammonium nitrate, as N [kg]	Irrigation, sprinkler [m3]	Tillage [ha]	...	Nitrate [kg]	...	N ₂ O [kg]	CO ₂ [kg]	...
...
2	45	5	FR	France	10'000	400	...	134	400	1	...	32	...	2.2	34	...
3	45	4	FR	France	10'000	400	...	134	400	1	...	32	...	2.2	34	...
...
8	47	10	DE	Ger.	10'000	0	...	50	0	1	...	12	...	1.1	20	...
9	47	9	DE	Ger.	10'000	0	...	50	0	1	...	12	...	1.1	20	...
...

All grid-cells

All inventory flows used across all AUPDs of a crop

Figure 12: The structure of the default crop inventory table. We generate this table by spatial differentiation of available country-level datasets into the common geoFootprint grid cell structure. A detailed explanation is provided in the subsections of this chapter.

The default crop inventory table simplifies regionalized LCI modeling because it provides

- A consistent framework for inventory modeling. Both, spatially explicit and inventory data are operating on the exact same scale and have the exact same form.

- A *powerful template* for data manipulation. With the default crop inventory, we already start with the most comprehensive and globally representative data, i.e., a regionalized data structure that covers all exchange flows associated with a crop across all of its available (geographic) instances. We simply manipulate (overwrite) selected flows and leave default flows unchanged. That is, we can fully focus our efforts on the regionalization of selected inventory flows and leave the remaining data unchanged.

In the next sub-chapters, we elaborate how we generate the spatially differentiated default crop inventory table.

3.2.1. Identifying available AUPD

We systematically screen the above-mentioned LCI databases for available AUPDs in order to identify all AUPDs available to represent the cultivation of a particular crop. To avoid inconsistencies, global cultivation of a crop is modeled exclusively on the basis of the AUPDs from one LCI database. We do not represent the cultivation of a particular crop with AUPD's from different LCI databases.

Moreover, since the WFLDB 3.5 is a dedicated database for agricultural activities—and in general more up to date than ecoinvent in that regard—we prioritize AUPD from the WFLDB 3.5 whenever possible. Table 6 shows which LCI database we use for each crop in geoFootprint.

Table 6: LCI database source for AUPDs to represent crop cultivation in geoFootprint.

Crop	Database
Barley	WFLDB 3.5 (March 2020)
Wheat	
Maize	
Soybean	
Oil palm	
Rice	
Sugarcane	
Potato	
Rapeseed	
Sugar beet	
Sunflower	
Peanut	
Cocoa	
Coffee (Arabica)	
Coffee (Robusta)	
Cotton	Ecoinvent 3.5 (August 2018)
Rye	

Sorghum	
---------	--

3.2.2. Linking available datasets to geographies

Next we map each available AUPD to its geographical domain. That is, we define the geographical validity of each available dataset (see Figure 13).

name	shortname	barley
Argentina	AR	Barley grain, irrigated, at farm (WFLDB 3.5) # AR
Austria	AT	Barley grain, irrigated, at farm (WFLDB 3.5) # DE
Australia	AU	Barley grain, irrigated, at farm (WFLDB 3.5) # AU
Brazil	BR	Barley grain, irrigated, at farm (WFLDB 3.5) # AR
Canada	CA	Barley grain, irrigated, at farm (WFLDB 3.5) # CA
Germany	DE	Barley grain, irrigated, at farm (WFLDB 3.5) # DE
France	FR	Barley grain, irrigated, at farm (WFLDB 3.5) # FR
United Kingdom	GB	Barley grain, irrigated, at farm (WFLDB 3.5) # GB
Italy	IT	Barley grain, irrigated, at farm (WFLDB 3.5) # IT
Poland	PL	Barley grain, irrigated, at farm (WFLDB 3.5) # PL
Russia	RU	Barley grain, irrigated, at farm (WFLDB 3.5) # RU
Ukraine	UA	Barley grain, irrigated, at farm (WFLDB 3.5) # UA
United States	US	Barley grain, irrigated, at farm (WFLDB 3.5) # CA

Figure 13: Mapping available AUPDs to their geographical domain.

As shown in Figure 13, we can also use this procedure to determine proxy datasets. For example, since we do not have a specific dataset for Brazil or the United States, we extend the geographical validity of existing datasets. We use the Argentinian AUPD to represent barley cultivation in Brazil and the Canadian dataset to represent barley cultivation in the United States. What we use as a proxy is the dataset template, i.e., the type of flows recorded in the proxy datasets, not their values. That is, we overwrite all important flow values under full consideration of the spatial context (precipitation, soil properties, etc.) See Table 7.

We represent all geographical domains without a link to an existing dataset with a global default dataset. Consequently, all geographical domains (countries) not shown in Figure 13 are modeled with the global default dataset, which is the production volume weighted average of the existing geographical AUPDs. For example, we use the global dataset to represent barley cultivation in Kyrgyzstan since we do not have a dataset (or proxy) for this geographical domain.

3.2.3. Determining unique inventory flows per crop

In order to comprehensively describe agricultural cultivation of a crop on a global scale the default crop inventory table has to accommodate all inventory flows associated with the available AUPDs.

We describe all AUPDs in the form of a crop inventory matrix which lists, for each AUPDs, all corresponding inventory flows and the reference flow.

		AUPD's, available for barley cultivation			AUPD's, available for Maize cultivation		
Inventory flows		Barley grain, non-irrigated, at farm (WFLDB3.5) # UA	Barley grain, non-irrigated, at farm (WFLDB3.5) # DE	...	Maize grain, irrigated, at farm (WFLDB 3.5) # CN	Maize grain, non-irrigated, at farm (WFLDB 3.5) # ES	...
Technosphere flows
	Urea ammonium nitrate (UAN), as N, at plant (WFLDB 3.5) # RER	1.17	19.27	...	7.02	0	...
	Monoammonium phosphate (MAP), as N, at plant (WFLDB 3.5) # RER	1.17	2.75	...	1.00	0	...
	Calcium ammonium nitrate (CAN), as N, at plant (WFLDB 3.5) # RER	0.23	33.04	...	12.04	0	...
	market for ammonium sulfate, as N # GLO	0.93	4.13	...	1.50	30	...

Natural resources	Carbon dioxide, in air	4047	7629	...	5555	2567	...
	Energy, gross calorific value, in biomass	47165	88912	...	64550	29832	...
Emissions	Nitrate	231	334	...	121	205	...
	Phosphorus	0.22	0.30	...	0.43	0.39	...

Reference flow	Production amount	4052	7640	...	13103	6055	...

Figure 14: The crop inventory matrix holds all inventory flows for all AUPD's of relevance.

All inventory flows in the crop inventory matrix refer to the cultivation of one hectare of the crop, i.e., are fully (100%) allocated to the particular product. Recording the inventory flows in this form, simplifies the later integration of allocation.

We use this matrix to determine the total number of unique flows in use across all geographical instances (AUPDs) of a crop and to retrieve the individual flows associated with a particular AUPD. Depending on the crop of interest, the number of unique flows varies between 73 (rice) and 155 flows (potato)—and so does the number of columns of the default crop inventory table.

3.2.4. Geographical differentiation

We use the number of unique flows in use across all geographical instances (AUPDs) of a crop to determine the extent—columns and their unique name—of the default crop inventory table. The number of rows of the table represent all grid cells which cover land surface.

Using the country code, we merge all inventory flows associated with a particular AUPD's into the default crop inventory table (see Figure 8).

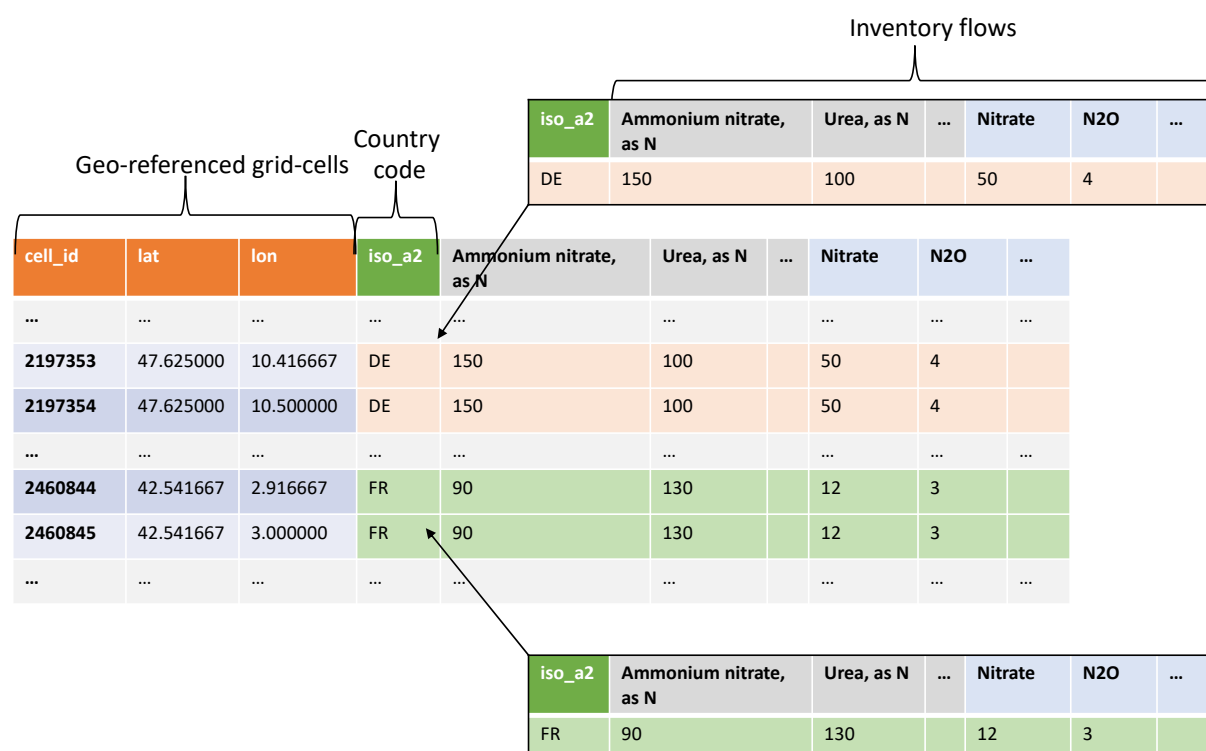


Figure 15: How we generate the default crop inventory table. We merge all inventory flows associated with the cultivation of a certain crop according to its geographical validity into the default crop inventory table.

The example in Figure 15 shows how we merge inventory flows associated with the cultivation of barley in Germany to all grid cells which are located in Germany and how we merge inventory flows associated with the cultivation of barley in France to all grid cells which are located in France. We perform this operation for all countries. Note that countries where we do not have an explicit AUPDs are approximated with the inventory flows of the production volume weighted global average dataset.

The final outcome of this procedure is a globally valid default crop inventory table that accommodates the best available AUPDs for each country.

3.3. Regionalized inventory table

We manipulate (overwrite) selected columns in the default crop inventory table using numerous inventory and emissions models. Table 7 shows and describes these models and cross-references to the respective chapter of the report, which treats them in detail. It also details the *degree of regionalization* of each inventory/emissions flow (i.e., the spatial resolution) for each flow category in this table.

The final outcome of regionalized LCI modeling is the regionalized inventory table. This table has the same structure as the default crop inventory table but also includes regionalized crop information such as yield and area harvested. (See Figure 16.)

Geo-referenced grid-cells (simplified)

Grid-cell belongs to

Crop information

Natural resource flows

Technosphere flows

Emissions flows

id	lat	lon	iso_a2	Country	Yield [t]	Area harvested [ha]	Production volume [t]	Land occupation, arable [m2*a]	Water, well [m3]	...	Ammonium nitrate, as N [kg]	Irrigation, sprinkler [m3]	Tillage [ha]	...	Nitrate [kg]	...	N ₂ O [kg]	CO ₂ [kg]	...	
...
2	45	5	FR	France	3.2	1'203	3'850	10'000	400	...	134	400	1	...	32	...	4.1	34	...	
3	45	4	FR	France	3.7	1'100	4'070	10'000	430	...	122	430	1	...	67	...	2.2	22	...	
...	
8	47	10	DE	Ger.	1.9	219	416	10'000	670	...	50	670	1	...	54	...	1.9	56	...	
9	47	9	DE	Ger.	4.1	306	1'255	10'000	320	...	50	320	1	...	33	...	0.8	20	...	
...	

All grid-cells

All inventory flows used across all AUPDs of a crop

Figure 16: The structure of the regionalized crop inventory table.

In contrast to the default crop inventory table (Figure 12), the regionalized table holds regionalized values computed with and retrieved from the spatial attribute table.

The degree of regionalization in Table 7 indicates the resolution of the final inventory and emission flow categories. For example, when we indicate *grid cell (10x10 km)*, this means that the respective flow in this category will actually vary from grid cell to grid cell at a resolution of 10x10 km. For the emission category *dinitrogen monoxide (N₂O), from land use*, the value will actually change at a scale of 10x10 km. Note that most of the inventory and emission categories are regionalized at the highest possible resolution, 10x10 km (see No. 1-17, except No. 9). These inventory and emissions flows typically constitute more than 80% of the total footprint for all of the footprints currently applied in geoFootprint (see section 3.4).

In addition to the highly regionalized climate and soil data, the default calculation of the regionalized inventory and emission flow categories also rely on the use of generic attributes at country or global scale. Most of these generic attributes can be customized to a grid cell resolution of 10x10 km in the Scenario assessment module of the geoFootprint platform. (See No. 18-20, for example.) However, to date we are missing the spatial data to regionalize flows relating to field operations, pesticides and supplements and therefore model these flow categories on the basis of the flow (type and value) recorded in the default crop inventory table. (See No. 21 & 22.) Table 37 in Annex 13.1 indicates the spatial scale for all input data elements.

Table 7 also indicates the spatial validity of the inventory and emissions models, i.e., the region for which a model is expected to produce reliable results. Most models have a global validity. However, notably, the model used for N leaching originates from a temperate context and does therefore not necessarily produce reliable results on a global scale—though it was applied successfully in sub-Saharan Africa (Faist Emmenegger et al., 2009). Due to the absence of alternatives, we currently apply this model on a global scale, an assumption we plan to refine with the next iteration of the geoFootprint platform. Additional insights on the spatial validity are provided below Table 7.

Table 7: Inventory/emissions models applied in the regionalization engine to compute comprehensive inventory flows and their degree of regionalization. Soil erosion (No. 17) is in itself not a recorded inventory flow but an important proxy indicator for soil health and an intermediate result to compute leaching of N and P flows.

No.	Inventory flow category / emission flow category	Degree of regionalization	Can be manipulated with the scenario module	Ref.	Input data required	Model	Spatial validity of the model	Brief description
1	Land occupation	Grid cell (10x10km)	Yes, indirectly via yield	5.1	- Cultivation period	geoFootprint regionalized land occupation model.	global	Land occupation is measured in area and time a piece of land is occupied. We compute land occupation under consideration of the spatially explicit and crop-specific cultivation period taken from GAEZ.
2	Land transformation	Grid cell (10x10km)	Yes, indirectly, via yield	5.2	- Allocated and annualized tree cover area loss per ha crop	geoFootprint regionalized land transformation model.	global	We use the spatially explicit tree cover loss data from GFW to determine the former (non-arable) land use and to inventory the corresponding land transformation type (primary and/or secondary forest) to arable land.
4	Irrigation water consumption (covers how we model water inputs, irrigation technologies and water emissions flows based on spatially explicit data)	Grid cell (10x10km)	Yes, directly.	7	-Water Footprint Network blue water footprint -Irrigation efficiency factor -Irrigation technologies applied in the default inventory dataset	geoFootprint regionalized irrigation water consumption model (which is based on the principles of irrigation water modeling of the WFLDB Methodological Guidelines (Nemecek et al. 2019).	global	Water use is modelled according to the WFLDB Methodological Guidelines. The water use calculation is based on the consumed water (blue water footprint or evapotranspiration from irrigation) (Mekonnen, 2011). The spatially explicit irrigation is assigned to the split of irrigation technologies present in the (closest) default crop inventory dataset in proportion to their quantitative relevance.
5	Organic manure application rate	Grid cell (10x10km)	Yes, directly.	6.1	-N manure application rate	geoFootprint regionalized manure model (which is based on	global	The spatially explicit N manure application rate is not crop-specific and we use the relative occupied crop area per grid cell to allocate the total N input to the crop of

No.	Inventory flow category / emission flow category	Degree of regionalization	Can be manipulated with the scenario module	Ref.	Input data required	Model	Spatial validity of the model	Brief description
	(N and P applied in the form of manure)				-relative area occupied by a crop (allocation factor) -default N content of solid and liquid manure -default P content of solid and liquid manure	the modeling principles developed in (J. Reinhard et al., 2017; Juergen Reinhard et al., 2017)).		interest. Next, we use the distribution of manure types (liquid and solid) of the crop default inventory to derive the corresponding amount of liquid and solid manure and also the corresponding amount of P input (the latter is required for the calculation of P leaching) from the crop-specific N manure application rate.
6	N fertilizer type application rate	Grid cell (10x10km)	Yes, directly.	6.2	-N application rate -N fertilizers applied in default inventory	geoFootprint regionalized N fertilizer model (which is based on the modeling principles developed in (J. Reinhard, 2017; Juergen Reinhard, 2017)).	global	The individual N fertilizer types and their application rate is derived from the default crop inventory. The crop-specific and spatially explicit N application rate is assigned to the fertilizer types present in the (closest) default inventory in proportion to their quantitative relevance.
7	P fertilizer type application rate	Grid cell (10x10km)	Yes, directly.	6.2	-P application rate -P fertilizers applied in default inventory	geoFootprint regionalized P fertilizer model (which is based on the modeling principles developed in (J. Reinhard, 2017; Juergen Reinhard, 2017)).	global	The individual P fertilizer types and their application rate is derived from the default crop inventory. The crop-specific and spatially explicit P application rate is assigned to the fertilizer types present in the (closest) default inventory in proportion to their quantitative relevance.
8	K fertilizer type application rate	Grid cell (10x10km)	Yes, directly.	6.2	-K application rate -K fertilizers applied in default inventory	geoFootprint regionalized K fertilizer model (which is based on the modeling	global	The individual K fertilizer types and their application rate is derived from the default inventory. The crop-specific and spatially explicit K application rate is assigned to the fertilizer types present in

No.	Inventory flow category / emission flow category	Degree of regionalization	Can be manipulated with the scenario module	Ref.	Input data required	Model	Spatial validity of the model	Brief description
						principles developed in (J. Reinhard, 2017; Juergen Reinhard, 2017)).		the (closest) default inventory in proportion to their quantitative relevance.
9	Soil erosion	Grid cell (25x25km)	Yes, partly via tillage and anti-erosion practices	8	-length-slope factor -erodibility factor -erosivity factor -crop factor -tillage factor -practice factor	geoFootprint soil erosion model which uses the USLE (Universal Soil loss equation) model (Faist Emmenegger et al. 2009) in combination with new global input data from (Borrelli et al., 2017).	global	The models compute spatially explicit soil erosion in metric ton per hectare. It uses novel input data from Borelli et al. 2017 (length-slope, erodibility and erosivity factor) in combination with default/user defined data on the tillage and cultivation practice. The data from Borelli et al. 2017 (length-slope, erodibility and erosivity factor) comes at a resolution of 25x25 km, i.e., these input factors only vary
10	N leaching in the form of Nitrate	Grid cell (10x10km)	Yes, indirectly via input/management practice	9.1	-precipitation -irrigation -clay content of soil -N application rate -N in soil -crop-specific rooting depth	geoFootprint regionalized nitrate model (which builds on the SQCB model (Faist Emmenegger et al. 2009; Nemecek et al. 2019).	Europe, sub-Saharan Africa*	The model computes N leaching in the form of Nitrate. The SQCB model was specifically developed for the spatially explicit assessment of biofuel crops in an LCA perspective, i.e., the Sustainability Quick Check for Biofuels. It is a simple regression model based on the formula developed by de Willigen (2000) and computes Nitrate leaching under consideration of water availability, soil and crop properties and N input. We calculate N ₂ O associated with Nitrate leaching according to the Cool Farm Alliance (Kayatz et al., 2020) (which is based on (IPCC 2006))

No.	Inventory flow category / emission flow category	Degree of regionalization	Can be manipulated with the scenario module	Ref.	Input data required	Model	Spatial validity of the model	Brief description
11	P leaching in the form of Phosphorus and Phosphate	Grid cell (10x10km)	Yes, indirectly via input/management practice	9.2	-amount of eroded soil -P content in soil -P applied in the form of organic and mineral fertilizer	SALCA-P with refinements published in Scherer and Pfister (Prasuhn 2006; Scherer and Pfister 2015).	global+	The model computes P leaching in the form of phosphorus to surface water, phosphate to ground water, and phosphate to surface water. The basic model structure is based on SALCA-P. We use the refinement published in Scherer and Pfister (2015).
12	Dinitrogen monoxide (N ₂ O), from land use	Grid cell (10x10km)	Yes, indirectly via input/management practice	11.4 11.3.2 11.3.3	-N fertilizer application rate -fertilizer type -crop type -soil texture -soil organic carbon stock -soil pH -drainage -residue management method -emission factors	Bouwman model as implemented by the Cool Farm Alliance (Kayatz et al. 2020)	global ^f	N ₂ O emissions associated with land use are modelled according to the Bouwman model as implemented by the Cool Farm Alliance. The model calculates N ₂ O emissions on the basis of an exponential regression model that considers various environmental and management condition parameters. The detailed implementation of the model in the regionalization engine is described in the deliverable "D1.1 GeoFootprint LU management calculation module". In addition, we model N ₂ O emissions associated with various residue management practices and with residue application (see section 11.3.2 and 11.3.3) in the land use category.
13	Nitrous oxides (NO _x)	Grid cell (10x10km)	Yes, indirectly via input/management practice	11.4.3	Same as in (12) but without crop type, soil texture and soil pH.	Bouwman model as implemented by the Cool Farm Alliance (Kayatz et al. 2020).	global ^f	Emissions from NO _x are modelled with the same functional form than emissions of N ₂ O, i.e., on the basis of an exponential regression model. A conversion factor of 0.01 is used to calculate the resulting N ₂ O from NO _x emissions. The detailed implementation of the model in the regionalization engine is described in the deliverable "D1.1 GeoFootprint LU management calculation module".

No.	Inventory flow category / emission flow category	Degree of regionalization	Can be manipulated with the scenario module	Ref.	Input data required	Model	Spatial validity of the model	Brief description
14	Ammonia (NH ₃)	Grid cell (10x10km)	Yes, indirectly via input/management practice	11.4.4	-N fertilizer application rate -fertilizer type -crop type -soil moisture -soil pH -application method	Bouwman model as implemented by the Cool Farm Alliance (Kayatz et al. 2020).	global ^f	Volatilization of NH ₃ emissions are computed with the same functional form than emissions of N ₂ O, i.e., on the basis of the exponential regression model also used for N ₂ O emissions. A conversion factor of 0.01 is used to calculate the resulting N ₂ O from NH ₃ .
15	Carbon dioxide (CO ₂), from land use change	Grid cell (10x10km)	No	1	-total tree cover area loss -total biomass loss -root-to-shoot ratio -dead organic matter -relative area occupied by a crop (allocation factor) -time period under observation	geoFootprint regionalized land use change model.	global	CO ₂ emissions associated with the annualized tree cover loss (and corresponding biomass loss) allocated to a particular crop.
16	Dinitrogen monoxide (N ₂ O), from land use change	Grid cell (10x10km)	Yes, indirectly via inputs / management practices	11.1.2	-annual loss of soil organic carbon -type of land transformation -C/N ratio	Bouwman model as implemented by the Cool Farm Alliance (Kayatz et al. 2020).	global	N ₂ O emissions associated with land use change are modelled in line with the Cool Farm Alliance. The model is equivalent to the IPCC model of N mineralized in mineral soils (IPCC 2006) – see chapter 11, equation 11.8.- and makes use of the default emission factor of 0.01 to estimate direct N ₂ O emissions – see chapter 11, table 11.1 (IPCC 2006).
17	Carbon dioxide (CO ₂), from land use	Grid cell (10x10km)	Yes, indirectly via input/management practice	11.1.1 11.2 11.3.1	-soil erosion -soil organic carbon stock -organic additions in the form of manure, residues and straw. -tillage and input practice	-geoFootprint regionalized CO ₂ loss associated with soil erosion. -CO ₂ accumulation associated with organic additions as implemented by	global europe s	Includes the CO ₂ loss associated with the loss of soil via soil erosion (computed by multiplying the soil erosion with the corresponding loss of organic matter), the CO ₂ accumulation associated with the annual relative SOC increase due to the application of organic additions (Smith et al. 1997) and the CO ₂ emissions

No.	Inventory flow category / emission flow category	Degree of regionalization	Can be manipulated with the scenario module	Ref.	Input data required	Model	Spatial validity of the model	Brief description
					-crop type of new land use	the Cool Farm Alliance (Kayatz et al. 2020) which is based on (Smith et al., 1997). -CO ₂ emissions associated with changes in SOC due to changes in management practice as explained in (IPCC 2006), chapter 5, table 5.5.	global	/accumulations associated with changes in soil organic carbon.
18	Carbon dioxide (CO ₂), supplements	Country-level flow composition for limestone / dolomite input from the closest AUPD is used.	Yes, directly	11.6	-limestone / dolomite amount if/as present in the default inventory -urea amount if/as present in the default inventory -ammonium bicarbonate if/as present in the default inventory	CO ₂ emissions from hydrolysis of urea and ammonium bicarbonate model as implemented by the Cool Farm Alliance (Kayatz et al. 2020) (which is based on (IPCC 2006)).	global ^L	Computes CO ₂ emission associated with the application of supplements such as urea, limestone, and ammonium bicarbonate. We use the flows present in the (closest) default inventory to compute emissions.
19	Methane (CH ₄) from rice	Global default settings for management practices are used, can be refined in the Scenario Module.	Yes, directly	11.5	-water management practice -preseason water management practice -organic amendment	CH ₄ emissions from rice cultivation model as implemented by the Cool Farm Alliance (Kayatz et al. 2020) (which is based on (IPCC 2006)).	global ^M	The model considers the formation of methane in dependence on organic amendment and water management practice in rice cultivation.

No.	Inventory flow category / emission flow category	Degree of regionalization	Can be manipulated with the scenario module	Ref.	Input data required	Model	Spatial validity of the model	Brief description
20	Methane (CH ₄) from crop residues	Global default settings for residue management method is used.	Yes, directly	11.3.2	-residue management method -emission factors	CH ₄ emissions from residue model as implemented by the Cool Farm Alliance (Kayatz et al. 2020) (which is based on (IPCC 2006))	global	The model considers the formation of methane in dependence on residue management method.
21	Field operations / Supplements / Pesticide provision	Country-level field operations / supplements from the closest AUPD is used.	Yes, directly	n.a.	n.a.	We use the field operations / supplement / pesticide flows (type and value) recorded in the default crop inventory table.	global	We do not yet have spatially explicit input data which would allow a regionalized modeling of field operations, supplements or pesticides.
22	Pesticide / Heavy metal emissions	Not considered	Not considered	n.a.	n.a.	We exclude pesticide and heavy metal (associated with fertilizer application and influence by crop uptake) emission from the inventory.	n.a.	We do not yet have spatially explicit input data which would allow a regionalized modeling of pesticide and heavy metal emissions. We also do not apply a footprint which considers these emission categories in the impact assessment.

* The model has been used successfully for a nitrogen balance assessment in the whole Sub-Saharan Africa (Faist Emmenegger, 2009). In general, a regression equation such as in (De Willigen 2000) should be applied only for interpolation, i.e., within the ranges of the data used for the regression. We are often, but not always, in the range of the given data for most of the variables. Due to the absence of alternatives, we currently apply the model on a global scale. We plan to replace the model with the next iteration of the geoFootprint platform.

+As shown by the application of Scherer and Pfister (2015) and Faist Emmenegger et al. (2009) the model has been applied under a wide pedo-climatic condition and therefore is considered to be globally valid.

S The model was developed in a European context. We rely on this model in order to match the influence on SOC associated with organic amendments to the calculation performed in the Cool Farm Tool.

^F The model is used on a global scale (Bouwman et al., 2002; Stehfest & Bouwman, 2006). Yet, the dataset used to generate the Bouwman model is dominated by measurements in industrialized countries with high atmospheric-N deposition rates. The model might therefore overestimate emissions in countries with low deposition rates.

^T The model is applicable on a global scale but we ignore climatic variations in the above- to below-ground-biomass ratio and work with a global average.

^L We work with global default emission factors (Tier1) which represent the maximum amount of CO₂ emissions. The uncertainty is about -50% based on approximations suggesting emissions may be less than half of the maximum value (IPCC 2006).

^M The model is applicable on a global scale. However, it is known that other contextual and management factors, such as soil type, rice cultivar, sulphate containing amendments, etc., can significantly influence CH₄ emissions (IPCC 2006). We are reviewing options to integrate these parameters into the model.

3.4. Regionalized footprint table

The geo-referenced inventory table provides a comprehensive data foundation for the consideration of all environmental footprints typically applied in the field of life cycle assessment (LCA). In geoFootprint, we focus on the most relevant footprints to assess environmental performance of agricultural systems. We use the term “footprint” rather loosely, namely to express an aggregation mechanism which amplifies one or several environmental characteristics of the agricultural system (our regionalized inventory) under study. A footprint can represent a classical life cycle impact assessment (LCIA) method and/or on of its impact categories (such as climate change or acidification), or a proxy indicator (such as soil erosion or the nitrogen use efficiency), which are very relevant to assess the environmental performance of agricultural systems but are not frequently applied in the typical LCA framework.

Further, we distinguish between regionalized and generic footprints. A regionalized (or spatially differentiated) footprint characterizes a resource extraction or emission under consideration of the local conditions. For example, a regionalized LCIA method for water consumptions—such as the AWARE 100 model (Boulay et al., 2018)—derives the regionalized characterization factor from the potential of water deprivation, to either humans or ecosystems, building on the assumption that the less water remaining available per area, the more likely another user will be deprived. Note that we use regionalized LCIA methods exclusively for the characterization of direct impacts happening on the agricultural stage, such as water consumption and land use. That is, we do not characterize water and land use impacts associated with upstream¹² supply chain processes such as fertilizer production, electricity production, etc. Due to the relative quantitative relevance of the agricultural stage, we expect to capture effectively all relevant impacts of water and land use with this approach.

A generic footprint is spatially independent, i.e., characterizes a particular resource extraction or emission always with the same characterization factor. We use spatially independent LCIA

¹² That is, we do not perform a vertical regionalized assessment, which would include the entire product system, but apply a horizontal regionalized impact assessment which focuses exclusively on the agricultural stage.

methods either because the impact of an emission is not spatially dependent or a spatially explicit method does not exist or is not yet implemented. For example, our climate change metric is spatially independent because the capacity of a greenhouse gas to influence radiative forcing (the mechanism responsible for global warming) can, at least for processes on the surface¹³, be suitably expressed in terms of a globally valid reference substance, namely carbon dioxide equivalents.

3.4.1. Assessment of environmental impact in geoFootprint

To assess the global environmental impacts of a particular crop, we transform the geo-referenced, regionalized crop inventory table into a regionalized footprint table for any footprint of interest (see Figure 17).

To compute the environmental impact associated with each regionalized inventory flow, we multiply each inventory flow (column) in the regionalized crop inventory table with its corresponding impact¹⁴ and/or characterization¹⁵ factor. Man-made technosphere inventory flows are multiplied with their impact factor (and there is always one distinct impact factor per technosphere flow). Resource extraction and emission flows are multiplied with their characterization factor. We multiply a grid cell specific resource or emission flow with a constant (or grid cell generic) characterization factor for spatially independent LCIA methods, whereas we multiply with a grid cell specific (spatially dependent) characterization factor for regionalized LCIA methods.

¹³ Indeed, the radiative forcing of GHGs change with the altitude of the emission. The change in radiative forcing of GHGs in dependence of the altitude of the emissions—the albedo effect—requires consideration for the appropriate assessment of the climate change impact of plane flying.

¹⁴ An impact factor is the environmental impact associated with the provision of a product or service, i.e., an output from technosphere. We compute impact factors of technosphere flows from the WFLDB 3.5 with SimaPro version 9.0. We use the impact factors published by ecoinvent to represent environmental impacts of all flows taken from the ecoinvent database.

¹⁵ A characterization factor is the environmental characterization of the potential environmental impact associated with the extraction of a resource or the release of an emission. We use the characterization factors of the LCIA methods, as implemented in SimaPro version 9.0.

The result of the multiplication is a geo-referenced footprint table that indicates environmental footprints per hectare of crop at farm gate for each inventory flow of relevance and for all grid cells. Note that the assessment of one regionalized inventory table with N environmental footprints produces N regionalized footprint tables.

Figure 17 shows an excerpt of a regionalized footprint table for the climate change footprint, i.e., all inventory flows. All columns marked as Natural resource flows, Technosphere flows and Emission flows, are expressed in kg CO₂-eq.

Geo-referenced grid-cells (simplified)			Grid-cell belongs to		Production characteristics			Natural resource flows			Technosphere flows				Emissions flows			
id	lat	lon	iso_a2	Country	Yield [t]	Area harvested [ha]	Production volume [t]	Land occupation, arable [m2*a]	Water, well [m3]	...	Ammonium nitrate, as N [kg]	Irrigation, sprinkler [m3]	Tillage [ha]	...	Nitrate [kg]	...	N ₂ O [kg]	CO ₂ [kg]
...
2	45	5	FR	France	3.2	1'203	3'850	0	0	...	398	40	120	...	0	...	470	34
3	45	4	FR	France	3.7	1'100	4'070	0	0	...	240	14	120	...	0	...	398	9
...
8	46	5	FR	France	1.9	219	416	0	0	...	450	0	120	...	0	...	301	20
9	46	4	FR	France	4.1	306	1'255	0	0	...	560	22	120	...	0	...	832	1
...

Exchange or inventory flows

Figure 17: Regionalized footprint table for climate change footprints. Since all flows refer to the cultivation of one hectare of a crop, the column-wise sum expresses the total environmental impact per hectare of crop cultivated. Note that we do not consider impacts located in the upstream (or background) system for regionalized footprints. The impacts for tillage are identical across all grid cells because we do not overwrite the default data.

The table allows either to aggregate along inventory flows—e.g., compute the total contribution of all emission, irrigation, field operation flows—or along grid cells—e.g., compute the production volume weighted average impact of a flow for an artificial or political boundary. For example, the columns-wise sum of all inventory flows expresses the total environment impact (e.g., carbon footprint) per hectare of crop at farm gate across all grid cells. All of these possibilities are applied on the web-based platform to facilitate advance contribution analysis of footprint metrics across various artificial and political boundaries.

We next discuss the key characteristics of the implemented footprints. We distinguish between proxy indicators, generic (spatially independent) and regionalized footprints.

3.4.2. Proxy & Efficiency indicators

A proxy indicator uses the physical properties of the agricultural system under study—soil erosion and soil organic carbon stock change—to approximate an important environmental characteristic (soil health). An efficiency indicator summarizes physical fluxes on the inventory level which can be used to approximate the efficiency of the agricultural system under study. Both, proxy and efficiency indicators are always regionalized because they operate on the (already regionalized) inventory table. Table 9 outlines the efficiency indicators and models implemented in geoFootprint as of end of 2020.

Table 8: Efficiency indicators in geoFootprint.

Metric	Unit	Model	Footprint characteristics	Spatial resolution	Description
Soil erosion	Metric ton soil	See Table 7 “Soil erosion”	Proxy indicator for soil health	Regionalized	We use soil erosion (as computed in Table 7) as a proxy indicator for soil health. The USLE model compute spatially explicit soil erosion in metric ton per hectare. We use novel input data from Borelli et al. 2017 (length-slope, erodibility and erosivity factor) in combination with default/user defined data on the tillage and cultivation practice.
SOC change	Metric ton soil organic carbon	See Table 7 “Carbon dioxide (CO ₂), from land use.”	Proxy indicator for soil health	Regionalized	We use the CO ₂ fluxes associated with soil organic carbon stock changes (see Table 7) as an efficiency indicator for soil health. The flux accounts for the CO ₂ loss associated with the loss of soil via soil erosion (computed by multiplying the soil erosion with the corresponding loss of organic matter), the CO ₂ accumulation associated with the annual relative SOC increase due to the application of organic additions (Smith et al. 1997) and the CO ₂ loss/accumulation associated with changes in soil organic carbon. We use a temporal characterization factor of - 0.05 to account for accumulations in the soil but treat any loss of SOC with a characterization factor of 1, i.e., as a permanent loss.
Water use (withdrawal)	m ³ water used	See Table 7 “Irrigation water”	Efficiency indicator for water impacts (without consideration of actual scarcities)	Regionalized	The water use indicates the actual water input of the agricultural system (without consideration of local scarcity). We express water use in the form of the total water withdrawal, i.e., blue water divided by the irrigation efficiency (equal to Irrig in section 7).

Metric	Unit	Model	Footprint characteristics	Spatial resolution	Description
Nitrogen Use Efficiency	%	(EU Nitrogen Expert Panel, 2016)	Efficiency indicator for efficiency of N-based fertilization	Regionalized	Recommended by the EU Nitrogen Expert Panel this indicator aims to contribute to improving efficient nitrogen use in agricultural production. The Nitrogen Use Efficiency (NUE) is based on the mass balance principle, i.e., using N input and N output data for its calculation: $NUE = N \text{ output} / N \text{ input}$. NUE values have to be interpreted in relation to productivity (N output) and N surplus (i.e., the difference between N input and harvested N output). An $NUE > 90\%$ is considered too high and creates risk of "soil mining", where an $NUE < 50\%$ is considered too low and reflects inefficient nitrogen use.

3.4.3. Generic (spatially-independent) footprints

Generic footprints implemented in geoFootprint are either midpoint- or damage-oriented. Midpoint-oriented impact categories aggregate environmental characteristics according to a shared physical trait of the resource extractions or emissions in the regionalized inventory table, e.g., climate change uses the radiative forcing potential as the common “currency” to translate all GHG emissions into CO₂-eq, freshwater eutrophication uses P-eq as a common unit. Damage-oriented impact categories assess the damage caused by a resource extraction or emission profile to a selected area of protection (e.g., human health, ecosystem quality, resources). We use the Ecosystem quality impact category indicator of Impact World+ (Bulle et al., 2019) to assess the cumulated impacts of the regionalized inventory on ecosystem quality. Table 9 outlines the generic footprints implemented in geoFootprint as of the end of 2020.

Table 9: Generic impact assessment (footprint) metrics and models in geoFootprint.

Metric	Unit	Model	Footprint characteristics	Spatial resolution	Description
Climate change	kg CO ₂ -eq	Bern model – Global Warming potentials (GWP) over a 100-year time horizon (IPCC 2013)	Midpoint	Generic	Impact category that accounts for radiative forcing caused by greenhouse gas (GHG) emissions such as carbon dioxide (CO ₂), methane (CH ₄) or nitrous oxide (N ₂ O). The capacity of a greenhouse gas to influence radiative forcing is expressed in terms of a reference substance (carbon dioxide equivalents) and considers a time horizon of 100 years following the guidelines from the Intergovernmental Panel on Climate Change (IPCC 2013). Radiative forcing is the mechanism responsible for global warming. Note that we introduce a dedicated flow “Carbon dioxide, to soil stock” for the appropriate consideration and characterization carbon accumulations in the soil. We characterize this flow with a temporal characterization factor of -0.05.
Freshwater eutrophication	kg P-eq	EUTREND model (Struijs et al., 2009)	Midpoint	Generic	Impact category that addresses impacts from nutrients (mainly phosphorus) from sewage outfalls and fertilized farmland which accelerate the growth of algae and other vegetation in freshwater. The degradation of organic material consumes oxygen resulting in oxygen deficiency. In freshwater environments, phosphorus is considered the limiting factor. The impact metric is expressed in kg P-eq (kg phosphorous equivalent to freshwater).
Marine eutrophication	kg N-eq	EUTREND model (Struijs, 2009)	Midpoint	Generic	Impact category that addresses impacts from nutrients (mainly nitrogen) from sewage outfalls and fertilized farmland which accelerate the growth of algae and other vegetation in marine water. The degradation of organic material consumes oxygen resulting in oxygen deficiency. In marine environments, nitrate (N) is considered the limiting factor. The impact metric is expressed in kg N-eq (kg nitrogen equivalent to water).

Metric	Unit	Model	Footprint characteristics	Spatial resolution	Description
Acidification	mol H ⁺ -eq	Accumulated Exceedance model (Seppälä et al., 2006)	Midpoint	Generic	Impact category that addresses impacts due to acidifying substances in the environment. Emissions of nitrogen oxides (NO _x), ammonia (NH ₃) and sulphur oxides (SO _x) lead to releases of hydrogen ions (H ⁺) when the gases are mineralized. The protons contribute to the acidification of soils and water when they are released in areas where the buffering capacity is low, resulting in forest decline and lake acidification. The impact metric is expressed in mole H ⁺ -eq (hydrogen ions to soil and water equivalents).
Ecosystems quality	PDF*m ² *year	IMPACT World+ (Bulle, 2019)	Damage oriented	Generic	Impact category that measures the potential effect (i.e., damage) of air, soil and water pollution on the quality of ecosystems in the form of potentially disappeared fraction of species per m ² and year. Our implementation of the impact category indicator is based on a SimaPro export of the LCIA indicators/impact factors. We mapped elementary characterization factors and technosphere impacts factors to the flows used in geoFootprint to fully integrate the method into geoFootprint.

3.4.4. Regionalized footprints

We use three regionalized footprints to characterize environmental impacts of water and land use. We characterize regionalized impacts on biodiversity related to land use (Chaudhary et al., 2015) and water scarcity impact (Boulay, 2018) as recommended by the UNEP-SETAC Life Cycle Initiative (*Global Guidance for Life Cycle Impact Assessment Indicators Volume 1*, 2016). In addition, we integrate the irrigation stress intensity score, derived from the regionalized baseline water stress score from the Aqueduct 3.0 database developed by the World Resource Institute (WRI) (WRI, 2020). We rasterize these LCIA methods from their initial spatial resolution (e.g., watershed, eco-region) into the common geoFootprint grid cell structure using QGIS v3.0 (QGIS, 2020). See Annex section 13.2 for a detailed explanation of the rasterization procedure and corresponding assumptions.

The footprint on biodiversity loss due to land use *“represents regional species loss taking into account the effect of land occupation displacing entirely or reducing the species that would otherwise exist on that land, the relative abundance of those species within the ecoregion, and the overall global threat level for the affected species”* (*Global Guidance for Life Cycle Impact Assessment Indicators Volume 1*, 2016). We use the taxa-aggregated characterization factors for Regional Potential Species Loss (PSL) available from the *Global Guidance for Life Cycle Impact Assessment Indicators Volume 1* at the ecoregion level. We compute the biodiversity loss due to land use impact [PDF*m²*year] by multiplying the “land transformation, from forest” (area with tree cover loss) and “land occupation” elementary flows in the regionalized inventory (see section 5.1 and 5.2, respectively) with the corresponding regionalized characterization factor.

The water scarcity (WS) footprint developed in the AWARE methodology quantifies the Available Water Remaining (AWARE) after consumption i.e., it quantifies *“the relative available water remaining per area once the demand of humans and aquatic ecosystems has been met, answering the question What is the potential to deprive another user (human or ecosystem) when consuming water in this area?”* (Boulay, 2018). We use the characterization factors at watershed level and with annual aggregation. We compute the water scarcity impact [m³ water deprived-eq] by multiplying the consumed (blue) water [m³] in the

regionalized inventory (equal to ET_{irr} in section 7) with the corresponding regionalized characterization factor [-].

We calculated the irrigation stress intensity (ISI) score based on the **baseline water stress score (BWS)** developed by the WRI and used in the **Aqueduct Water Risk Atlas tool**, and the irrigation intensity (blue water). The BWS “*measures the ratio of total water withdrawals to available renewable surface and groundwater supplies. Water withdrawals include domestic, industrial, irrigation, and livestock consumptive and non-consumptive uses. Available renewable water supplies include the impact of upstream consumptive water users and large dams on downstream water availability. Higher values indicate more competition among users.*” (Hofste et al., 2019). To have an irrigation stress intensity score covering the same value domain as the AWARE water scarcity indicator (0-100%), we transfer (normalize) the value domain of the **initial baseline water stress score** (ranging between 0 and 5) to a scale of 0 and 1. We compute the irrigation stress intensity score [m³ water stress-eq] by multiplying the consumed (blue) water [m³] in the regionalized inventory (equal to ET_{irr} in section 7) with the corresponding regionalized and normalized baseline water stress score [-].

Table 10 provides summarizes key characteristics of all 3 considered regionalized footprints.

Table 10: Regionalized footprint metrics and models in geoFootprint.

Metric	Unit	Model	Spatial resolution	Description
Water scarcity (WS)	m ³ water deprived-eq	AWARE 100 model (Boulay, 2018) ¹	regionalized	Regionalized LCIA method recommended by the UNEP-SETAC Life Cycle Initiative that characterizes the potential of water deprivation, to either humans or ecosystems , building on the assumption that the less water remaining available per area, the more likely another user will be deprived. It is based on the AWARE 100 model (Boulay, 2018) the recommended method from WULCA for water consumption impact assessment in LCA. The footprint metric is expressed in cubic meters of water deprived equivalents (m ³ water deprived-eq). In order to compute impacts we multiply the consumed (blue) water from irrigation with corresponding water deprived equivalents . We do not consider impacts located in the upstream (or background) system.
Irrigation stress intensity (ISI)	m ³ water stress-eq	(Hofste, 2019) ³	regionalized	Derived from the regionalized footprint developed by the WRI that characterizes the potential water stress. The baseline water stress score is directly available in a shapefile format (polygons and score already merged) and measures the ratio of total water withdrawals to available renewable surface and groundwater supplies . Water withdrawals include domestic, industrial, irrigation, and livestock consumptive and no consumptive use. Available renewable water supplies include the impact of upstream consumptive water users and large dams on downstream water availability. Higher values indicate more competition among users. In order to assess the irrigation stress intensity, we multiply the consumed (blue) water from irrigation with corresponding baseline water stress score. We do not consider impacts located in the upstream (or background) system.
Potential Species Loss from Land Use (PSL)	PDF*year	(Chaudhary, 2015) ²	regionalized	Regionalized LCIA method recommended by the UNEP-SETAC Life Cycle Initiative that measures the potential loss of species due to land use and land use change (Chaudhary, 2015). Characterization factors are available from: https://www.lifecycleinitiative.org/applying-lca/lcia-cf/ In order to assess impacts we multiply the inventoried amount of (i) land occupation (of either annual and perennial crops) and (ii) land transformation from natural habitat (area with tree cover loss which was transformed to either annual or perennial crops) with the corresponding potential species loss. We do not consider impacts located in the upstream (or background) system.

¹ <http://www.wulca-waterlca.org/aware.html>, ² <https://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world> ³ <https://www.wri.org/resources/data-sets/aqueduct-global-maps-30-data>

4. Recalculating scenarios

As mentioned in section 1.3 (and in Figure 1) the regionalization engine is used in two distinctive modes.

First, it is used to bulk compute the regionalized crop data (regionalized crop inventory and regionalized footprint) for all major crops on a global scale. It is a subset of this data which is deployed on the web-based platform.

Second, however, the regionalization engine is used by the web-based platform to directly re-compute results for selected grid cells or regions. The scenario assessment module facilitates the customization of important cultivation and management practices. The user can interactively deploy and explore the environmental implications of large-scale scenarios, i.e., overwrite and refine selected data elements, recompute all associated metrics and explore the implications in visual form on various level of aggregation.

We distinguish between simple and advanced recalculation scenarios. Advanced scenarios allow for a higher level of detail in the manipulation of cultivation characteristics, particularly the individual N fertilizer types and soil amendments. Table 11 shows the key differences between a simple and advanced scenario assessment.

Table 11: Cultivation and management practice manipulations in the simplified and advanced scenario module.

Attribute type	Attribute	Simple	Advanced
Management characteristic	Tillage practice	-All	-All
	Tillage method	-All	-All
	Anti-erosion practice	-All	-All
	Input practice	-All	-All
	Residue management	-All	-All
	Rice cultivation practices	-All	-All
Cultivation characteristic	Yield	-Yield	-Yield
	Irrigation	-Irrigation water withdrawal -Mix between irrigation technologies	-Irrigation water withdrawal -Mix between irrigation technologies
	Organic fertilizer	-Solid manure application rate	-Solid manure application rate
	Field operations	-Sowing/Planting -Mineral fertilizer application practice -Harvesting	-Sowing/Planting -Mineral fertilizer application practice -Harvesting

Attribute type	Attribute	Simple	Advanced
	Mineral fertilizers	-Total N fertilizer -Total P fertilizer -Total K fertilizer	-select individual N fertilizers (10) -select individual P fertilizers (7) -select individual K fertilizers (4)
	Soil amendments	-n.a.	-Lime application rate -Dolomite application rate

Table 12 details the scope of the scenario recalculation, i.e., which cultivation and management practices a user can change and their corresponding impact pathway, i.e., the effects of a particular practice on results.

Table 12: Cultivation & management practices, their value domain, and their effect on regionalized results. For the management characteristics, we also indicate the parameter setting used in the default (bulk) computation.

Attribute type	Attribute	What can be entered?	Effect on results	Parameter setting used in bulk calculation
Management characteristic	Tillage practice	Category: “Full”, “Reduced”, “No”	Changes SOC stock and therefore can change CO2 emissions from LUC	“No” for oil palm and sugarcane, “Full” all others.
	Tillage method	Category: see Table 24	Changes soil erosion, and therefore P leaching and C losses due to soil erosion	See Table 25
	Anti-erosion practice	Category: see Table 26	Changes soil erosion, and therefore P leaching and C losses due to soil erosion	“No anti-erosion practices for all grid cells whose LS (length-slope) factor does not exceed 3. For all grid cells which exceed a certain level of steepness, i.e., with a LS factor > 3, we apply “Contour farming”.
	Input practice	Category: “Low” or “Medium”	Changes SOC stock and therefore can change CO2 emissions from LUC	“Medium”
	Residue management	Category: see Table 29	Changes N2O and CH4 emissions from residues	“Let on field; incorporated or mulch”
	Rice cultivation practices	Category: see Table 39, Table 40, Table 41	Changes CH4 emissions from rice	-“Intermittently flooded – multiple aeration” -“Not flooded pre-season (within less than 180 days)” -“Green manure”
Cultivation characteristic	Yield	Value or percent increase/decrease	Changes all results per metric ton	n.a.
	Irrigation	Irrigation volume & technology	Changes water consumption, emissions associated with technology & N leaching.	n.a.
	Organic fertilizer	App. rate value or percent increase/ decrease	Changes N2O emissions, N & P leaching N2O emissions, NH3 emissions, scales field operation for fertilizer application.	n.a.
	Field operations	“Manual” or “Mechanical”	Changes upstream impacts / emission of field operation.	n.a.
	Mineral fertilizers	App. rate value / percent increase / decrease	Changes upstream impacts, N & P leaching, N2O emissions, NH3 emissions, NOx emissions.	n.a.
	Soil amendments	App. rate value / percent increase / decrease	Changes upstream impacts & CO2 emissions from amendments.	n.a.

5. Regionalizing resource flows

The direct input of natural resources also covers the input of water from river, well and groundwater. For consistency reasons, however, we elaborate the modeling of these water flows in chapter 7, i.e., in combination with all other water input and output flows. The modeling of all other considered resource flows is elaborated in this chapter.

5.1. Land occupation

Land occupation prevents the currently occupied land from changing to a more natural state. Therefore, both the area and the duration required for the production of a certain crop are important and land occupation is measured by multiplying the occupied area by time, i.e., in m²*year (Weidema, 2013).

Land occupation starts after the harvest of the previous crop (average harvested date) and ends with the harvest of the considered crop. In geoFootprint, we use the spatially explicit and crop-specific data from Global Agrological-ecological Zones (GAEZ, (IIASSA/FAO, 2012)) to model the cultivation period while we use the spatially explicit harvested area or yield data from EarthStat (Monfreda, 2008) to calculate the area use. Land occupation in m² per hectare crop cultivated is then calculated with equation 1. (The yield is not considered in this equation but is used when results are transformed from hectare to metric ton of crop on the web-based platform.)

$$LandOccupation = ha * \frac{CP}{12} \quad (1)$$

where

- ha represents 10,000 m²
- CP represents the cultivation period of annual crops [months] retrieved from the crop-specific GAEZ dataset. The CP of perennial crops is set to 12 months. GAEZ does not offer data on the cultivation period of sunflower. In line with Agrifarming (Agrifarming, 2020), we work with a global constant cultivation period of four months.

The environmental impacts associated with land occupation are characterized by the impact category Ecosystem Quality (ImpactWorld+) and in the regionalized PSL.

5.2. Land transformation

Land transformation refers to the change in land use, e.g., the conversion of a forest to agricultural area or the conversion of agricultural area to grassland, etc. In an AUPD, land transformations are inventoried with two entries (Weidema, 2013):

1. Land transformation, *from* land use class X
2. Land transformation, *to* land use class Y

Land transformation is measured in area [m^2] and has no explicit time dimension, though common practice has shown that land transformations are typically inventoried annually, i.e., are annualized over crop-production cycles or cultivation periods, in order to ensure that the transformation of a forest for e.g., palm oil production is not allocated to exclusively one year but distributed over the entire cultivation period of the crop of interest (Thomas Nemecek et al., 2014).

In the LCIA phase, land transformation flows (as a direct resource input from nature) are used to characterize biodiversity impacts, i.e., the decrease or increase in biodiversity associated with a certain type of land transformation, while their greenhouse-gas relevance, i.e., the corresponding change in carbon pools in the form of GHG fluxes associated with a particular type of land transformation, is typically not characterized. In fact, the GHG relevance of land transformations is typically considered on the inventory level, namely by explicitly modeling the corresponding loss/gain in the form of CO_2 , N_2O and CH_4 in the AUPD. We follow this approach in geoFootprint (see chapter 1).

Bearing this in mind, the modeling of land transformation flows (as a direct resource input from nature) exclusively serves the assessment biodiversity impacts in the form of PSL. We model land transformation flows according to the results of the LUC modeling (see chapter 1). This means that we distinguish between two types of land transformation:

In the absence of any LUC, i.e., an annualized tree cover area loss (TAL) equal to zero, we assume that all of the land transformation is from arable to arable land and consequently inventory

- Transformation, from arable land 10,000 m²
- Transformation, to arable land 10,000 m²

We further discriminate arable land into annual and perennial crops based on the crop under study. These kind of land transformations cancel each other out, i.e., have no impact on biodiversity.

If our LUC approach indicates an annualized TAL greater zero, we inventory the annualized TAL as a land transformation from forest. For example, for an annualized TAL of 300 m² we would inventory

- Transformation, from forest 300 m²
- Transformation, from arable 9,700 m²
- Transformation, to arable 10,000 m²

Due to the lack of more refined LUC data, we can only inventory the land transformations from forest to arable (and not from e.g., grassland). This limits the resolution of any biodiversity assessment to the most important transformations.

The environmental impacts associated with the land transformation from forest are characterized by EQL (ImpactWorld+) and in the regionalized PSLreg LCIA method.

6. Regionalizing mineral and organic fertilizer application rates

6.1. Organic fertilizer

The application of organic fertilizer in the form of liquid and solid manure affects the N and P balance of the agricultural system and can therefore contribute to N and P based emissions (leaching). In addition, the production, provision and application of liquid and solid manure can relate to environmental impacts.

geoFootprint uses generic spatially explicit manure application data (~10x10 km) in the form of kg N per hectare of arable land from FAO (J. S. Gerber, 2016) to model crop-specific manure application rates. To make the generic N application rate crop-specific, we use the relative area occupied by each crop (shared-responsibility approach) as an indication to distribute the total N in manure to the crop of interest.

In order to assess the total amount of N and P¹⁶ applied in the form of manure, we have to map it to the manure types listed in the default crop inventory, i.e., the flows for which we know the environmental impacts associated with their provision. We use the information in the default inventories to perform this mapping, basically by using the following procedure:

1. To make the generic N application rate crop-specific, we allocate total N input to all crops produced within a grid cell. Like with LUC (see section 10.1.3), we use the relative area occupied by each crop (shared-responsibility approach) as an indication to distribute the total N in manure to the crop of interest. With this allocation approach, the magnitude of the crop-specific manure N application rate is proportional to the relative harvested area of a crop per grid cell.

$$A_{Norg_c} = Norg * AO_{rel_c} \quad (1)$$

¹⁶ We require the P input for the comprehensive modeling of P leaching.

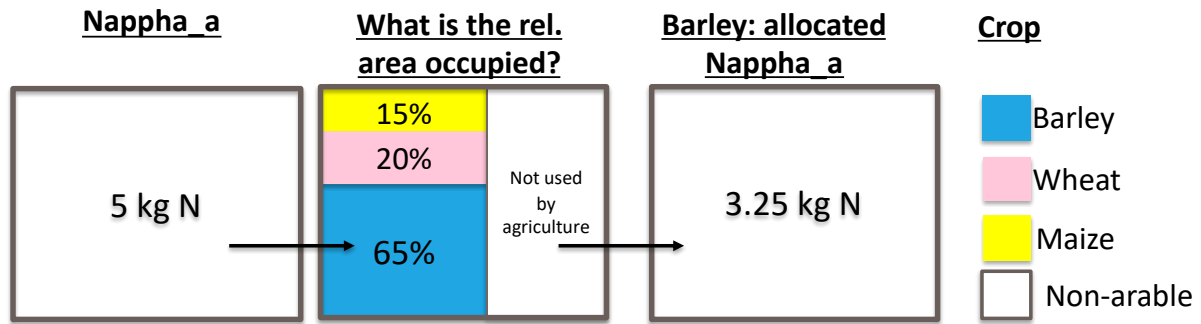


Figure 18: Exemplary illustration of the allocation of the generic organic N application rate to individual crops for the example of barley.

- Next, we use the manure flows listed in the default inventory to calculate the fraction of the total N which is applied in solid and in liquid form, i.e., the relative contribution per liquid and solid manure. (See Table 13 for a simplified example.)

$$Norg_{rel_liquid} = \frac{Norg_{liquid_i}}{\sum_{i=0}^n Norg_{liquid_i} + \sum_{i=0}^n Norg_{solid_i}} \quad (2)$$

$$Norg_{rel_solid} = \frac{Norg_{solid_i}}{\sum_{i=0}^n Norg_{liquid_i} + \sum_{i=0}^n Norg_{solid_i}} \quad (3)$$

- We then compute the corresponding amount of P application rate by multiplying the actual N content added in liquid and solid form with the corresponding P / N ratio. (See Table 14 for a simplified example.)

$$Porg_{liquid} = A_{Norg_c} * Norg_{rel_liquid} * \frac{Pcontent_liquid}{Ncontent_liquid} \quad (4)$$

$$Porg_{solid} = A_{Norg_c} * Norg_{rel_solid} * \frac{Pcontent_solid}{Ncontent_solid} \quad (5)$$

- Finally, we calculate the actual input flow of each type of manure in the regionalized inventory (see Table 15 for a simplified example) by multiplying the total amount of manure allocated to the crop of interest (column 1, Table 15) with the relative contribution of each type of manure (column 5, Table 13) and with the amount of manure per kg of N (column 5, Table 15, obtained by dividing 1,000 kg of manure by its corresponding N content).

$$Manure_type_i = A_{Norg_c} * Norg_{rel_type_i} * \frac{1000}{Ncontent_manure_type_i} \quad (6)$$

The computation in equation 6 focuses on individual manure types, i.e., is manure type specific, rather than distinguishing just between liquid and solid manure. This is intentional as we expect to use more refined factors for the N content (column 3, in Table 13), which will also facilitate a more refined distinction according to the source of manure (cattle, pig, broiler, etc.).

Table 13: Calculating the relative contribution of each type of manure (liquid and solid)

(1) Manure flows listed in default inventory	(2) Amount [kg]	(3) N content in kg per metric ton of manure	(4) Total N content provided [kg]	(5) Relative N contribution per manure type [-]	(6) Relative N contribution per liquid and solid manure [-]
manure, solid, cattle	4000	3.2	12.8	0.64	0.36
manure, liquid, cattle	4000	1.2	4.8	0.24	
manure, liquid, swine	2000	1.2	2.4	0.12	

Table 14: Calculation of P application rate from the information in the default inventory (example)

(1) Total manure allocated to Barley [kg N]	(2) Manure flows in def. inventory	(3) Fraction [-]	(4) N content [kg]	(5) Ratio P : N	(6) P content [kg]
3.25 kg	Solid	0.64	2.08	1.8 / 3.2	1.17
	Liquid	0.36	1.17	0.5 / 1.2	0.49

Table 15: Calculation of regionalized manure flows

(1) Total manure allocated to Barley [kg N]	(2) Manure in def. inventory	(3) Relative N contribution per manure type [-]	(4) Total N contribution per manure type [-]	(5) Manure per kg of N [kg]	(6) Amount of specific manure input flow in regionalized inventory [kg]
3.25 kg	manure, solid, cattle	0.64	2.08	1000 / 3.2	650
	manure, liquid, cattle	0.24	0.78	1000 / 1.2	650
	manure, liquid, swine	0.12	0.39	1000 / 1.2	325

6.2. Mineral fertilizer

The application of mineral fertilizer in the form of N, P and K affects the N and P balance of the agricultural system and can therefore contribute to N and P based emissions. In addition, we account for the environmental impacts associated with the production and provision of mineral fertilizers.

We use the spatially explicit N, P and K fertilizer application rates from EarthStat (Mueller et al., 2012) to regionalize mineral fertilizer application rates in our default crop inventories. The N and P fertilizer application rates represent both N input in mineral and organic form. Therefore, in order to retrieve the N and P input exclusively associated with mineral fertilizer, we subtract the total N and P associated with organic fertilizer. (See equations 1 & 2.)

$$N_{mineral} = N_{total} - N_{organic} \quad (1)$$

$$P_{mineral} = P_{total} - P_{organic} \quad (2)$$

Note that negative values are avoided by testing the final result with equations 3 and 4, respectively.

$$N_{mineral} = \begin{cases} N_{mineral} & \text{if } N_{mineral} \geq 0 \\ 0 & \text{if } N_{mineral} < 0 \end{cases} \quad (3)$$

$$P_{mineral} = \begin{cases} P_{mineral} & \text{if } P_{mineral} \geq 0 \\ 0 & \text{if } P_{mineral} < 0 \end{cases} \quad (4)$$

N, P and K fertilizer application rates in the default crop inventories are typically modeled with several different types of fertilizers exchange flows. (See Table 31 for a list of fertilizer types used in the default inventories.) That is, we have to map (or distribute) the total amount of spatially explicit mineral fertilizer rates to the mineral fertilizer types present in the default inventory. As with manure modeling, we use the information in the default inventories to perform this mapping.

1. We first calculate how much of the total N each type of N mineral fertilizer in the default inventory contributes in order to determine the relative contribution of each type of N fertilizer. (See column 3, Table 16.)

$$N_{rel_MinFertType_i} = \frac{N_{MinFertType_i}}{\sum_{i=0}^n N_{MinFertType_i}} \quad (5)$$

- We then use the relative contribution of each N fertilizer type to compute the regionalized amount of N input associated with each N-based mineral fertilizer type (see column 4, Table 16), by multiplying the total N input associated with all mineral fertilizers (column 1) with the relative contribution of each N-based mineral fertilizer type (column 3).

$$N_{reg_MinFertType_i} = N_{rel_MinFertType_i} * N_{mineral} \quad (6)$$

The procedure is identical for all types of mineral fertilizers (N, P, and K) and therefore only shown for N.

Table 16: Calculation of regionalized N-based mineral fertilizer application rates per fertilizer type in the regionalized inventory from the information present in the default inventory (example).

(1) Total N applied in mineral form ($N_{mineral}$) [kg]	(2) Fertilizer type ($N_{MinFertType_i}$)	(3) Relative relevance according to default inventory ($N_{rel_MinFertType_i}$) [-]	(4) Regionalized N-based fertilizer in regionalized inventory [m3]
200	Diammonium phosphate (DAP), as N, at plant (WFLDB 3.5) # RoW	0.7	140
	Ammonium nitrate (AN), as N, at plant (WFLDB 3.5) # RER	0.1	20
	Urea ammonium nitrate (UAN), as N, at plant (WFLDB 3.5) # RER	0.2	40

6.3. Nitrogen use efficiency

The nitrogen use efficiency (NUE) is not an inventory flow but a proxy indicator for the assessment of the efficiency of nitrogen use (EU Nitrogen Expert Panel, 2016). The concept for NUE used here is based on the mass balance principle, i.e., using N input and N output data for its calculation.

The mass balance reads as $N \text{ input} = N \text{ output} + N \text{ losses} + \text{changes in stocks}$. We compute the NUE indicator as the percentage of N output to N input (see equation 1).

$$NUE = \frac{N_{output}}{N_{input}} * 100 \quad (1)$$

with

$$N_{input} = N_{mineral} + N_{organic} + N_{deposition} + N_{seed} + N_{BND} \quad (2)$$

and

$$N_{out} = N_{harvMainProduct} + N_{harvCoProduct} \quad (3)$$

where

- $N_{mineral}$ represent the spatially explicit N input in the form of mineral fertilizers [kg N/(hectare*year)]. $N_{mineral}$ is equal to $N_{mineral}$ and derived from the calculation in chapter 6.2.
- $N_{organic}$ represents the spatially explicit N input in the form of organic fertilizer [kg N/(hectare*year)]. $N_{organic}$ is equal to A_Norg_c and derived from the calculation in chapter 6.1.
- $N_{deposition}$ represents the N input in the form of natural deposition [kg N/(hectare*year)]. For now, we work with an average global deposition rate of 14 kg [N per ha and year], which is in the range of the average deposition in central Europe (rather agricultural intense areas) (EU Nitrogen Expert Panel, 2016). We intend to replace this value by more refined spatially explicit data with the next iteration of geoFootprint¹⁷.
- N_{seed} represent the N input in the form of seeds and planting material [kg N/(hectare*year)] and is calculated by multiplying the seed input [kg/hectare*year] (retrieved from the default crop inventory) with its corresponding N content [kg N/metric ton of seed] and dividing by 1000. Default values for the N content of harvested products are taken from Table 17 (EU Nitrogen Expert Panel, 2016).

¹⁷ e.g. <https://conservancy.umn.edu/handle/11299/197613>

- N_{BND} represent the biological nitrogen fixation (BNF) by the crop under study [kg N/(hectare*year)]. To date (2020), the only nitrogen fixating crop in geoFootprint is soybean. We use the Tier 2 approach proposed by (EU Nitrogen Expert Panel, 2016) which estimates the amounts of N_2 fixed from linear relationships with the amount of exported (harvested) dry matter (DM).

$$BNF = \left(a + b * \frac{Yield}{HI} \right) * BGN \quad (4) \quad \text{where:}$$

- For a and b, we use the default values of pea given in (EU Nitrogen Expert Panel, 2016), i.e., 7 and 17.6 respectively.
- HI is the harvest index, defined as the ratio of the harvested (exported) biomass to the total aboveground biomass. For beans, lentils and peas, the harvest index (HI) is set at 0.7, unless more precise estimates are available.
- BGN is a correction factor for belowground N_2 fixation, comprising N associated with roots, nodules and rhizodeposition, set at 1.3 for lentils and peas, 1.4 for fava bean and 1.7 for forage legumes. We use the default factor of peas, i.e., 1.3.
- $N_{harvCoProduct}$ represents the N output in the form of removed residues [kg N/(hectare * year)]. We calculate the quantity of N removed in the form of residues ($N_{harvCoProduct}$) by multiplying the crop specific n uptake (N_{crop}) with the harvested quantity of residues. We compute the latter by multiplying the spatially explicit (and extrapolated) yield (y) from EarthStat (Monfreda, 2008) with the main-to-coproduct ratio (MPR, see Table 2) and the dry matter fraction (DMF) for the crop of interest.

$$N_{harvCoProduct} = N_{crop} * y * \frac{1}{MPR} * DMF \quad (5)$$

- $N_{harvMainProduct}$ represent the N output in the form of the harvested product [kg N/(hectare*year)]. In many cases, N output is linearly related to harvested yield (the N content of the harvested crop is more or less independent of N input). We compute the N in harvested main product by multiplying the N content of a particular crop with the spatially explicit (and extrapolated) yield from EarthStat (Monfreda, 2008).

$$N_{harvMainProduct} = N_{crop} * y * DMF \quad (6)$$

Default values for the N content of harvested products (N_{crop}) are taken from Table 17 (EU Nitrogen Expert Panel, 2016).

Table 17: Crop-specific N content used for the calculation of the total N in seed, harvested Product and Residues.

Source: (EU Nitrogen Expert Panel, 2016).

Crop	N content [kg/metric ton]
Potato	2.9
Sugar beet	2.3
Sugarcane	1.3
Sorghum	16.1
Rapeseed	35.2
Soybean	58.6
Sunflower	28.8
Oil palm	4.9
Barley	17
Wheat	20.8
Maize	15.3
Rye	17.5
Rice	15
Cotton	26.8
Groundnut	41.9

The NUE addresses the N outputs in harvested products only. This represents only a fraction of the N output; the remainder is dissipated into the atmosphere or water bodies (N losses) or is accumulated in the system (temporarily, as a stock). The NUE indicator does not consider these N losses and stocks. In fact, it reveals the relative relevance of these “outputs” in relation to the total N input when we divide the (incomplete) N output by the sum of all N inputs.

7. Regionalizing irrigation water consumption

Water use is modeled following the WFLDB Methodological Guidelines (T. Nemecek, 2019). In crop production, all consumed water—water withdrawal where release back to the source does not occur due evaporation, evapotranspiration, product integration or discharge into a different drainage basin—is considered as irrigation water. Green water, i.e., the volume of rainwater consumed during the production process, is not accounted for since we do not yet have the LCIA methods to account for the deplete effects associated with green water consumption.

We derive the total amount of annual, crop-specific irrigation from the blue water requirement of the Water Footprinting Network (WFN) data layer from Mekonnen & Hoekstra (Mekonnen, 2011) which comes in a resolution of 10x10 km. All calculations are performed on the grid-cell level. That is, we compute irrigation water consumption, corresponding irrigation technology usage and associated emissions for each and every grid cell.

We divide the blue water requirement per crop c by an efficiency value computed as the weighted average efficiencies of the applied irrigation technologies in the spatially relevant default crop inventory. Input irrigation water ($Irrig$, i.e., water withdrawal [$m^3/ha*yr$]) is then calculated across all grid cells as:

$$Irrig_c = \frac{ET_{irr_c}}{E_{Firr}} * F_{harv_area_c} * \frac{10000}{1000} \quad (1)$$

Where

- ET_{irr} represents the evapotranspiration from irrigation (blue water) [mm/yr]. The data of the WFN has been calculated by dividing the total water footprint per grid cell [m^3/yr] by the total area of the grid cell (Mekonnen, 2011). In order to (re-)scale this to the harvested area of the crop of interest, we have to divide it by the fraction of the grid cell, which is cultivated with the crop of interest.

- $F_{harv_area_c}$ represents the fraction of the harvested area of crop c . We use the extrapolated fraction of harvested area taken from EarthStat (Monfreda, 2008).
- The division by 1,000 converts from mm to meter and the multiplication by 10,000 converts m^2 to hectare so that the final unit is in $m^3/ha*yr$.
- E_{firr} represents the dimensionless irrigation efficiency factor [-]. Table 18 shows E_{firr} for three irrigation technologies. Because crop production datasets typically reflect average practices at country level, we compute the irrigation efficiency factor as a weighted average of the water volume applied by each irrigation technologies in the default crop inventory (see equation 2).

$$EF_{irr} = \sum_i^n Water_{rel_irrigation_type_i} * EF_{irr_i} \quad (2)$$

Where:

- $Water_{rel_irrigation_type_i}$ is the relative water volume contributed by irrigation technology i . See section 7.3.2.)
- EF_{irr_i} is the efficiency factor of irrigation technology i . (See Table 18.)

Table 18: Irrigation efficiency EF_{irr} for different irrigation technologies (adapted from (T. Nemecek, 2019))

Irrigation technology	Irrigation efficiency factor EF_{irr} [-]
Surface irrigation	0.45
Sprinkler irrigation	0.75
Drip irrigation	0.90

7.1. Curation of irrigation volume

The WFN blue water value characteristics, i.e., their reference to the entire grid cell area, makes them very sensitive to the gridded crop information used to interpret the WFN data. As a consequence, a very low harvested crop area in one grid cell and a very high blue water value can result in unrealistically high irrigation volume per hectare while a very low blue water value and a very high harvested crop area will result in an unrealistically low irrigation volume per hectare. Bearing this in mind, we filter or correct for particular low or particular high irrigation volume values.

7.1.1. Treatment of very low irrigation volumes

We replace any irrigation volume lower than 100 m³/ha*yr with zero because an irrigation volume of this size means that one m² is irrigated with about 10 liter per year.

Table 19: Crop-specific threshold for the treatment of very low irrigation volumes and the corresponding effect on the total number of grid cells and the global irrigation volume

Crop	Threshold [m ³ /ha*yr]	Relative # of grid cells irrigated	Relative irrigation volume removed [%]	Relative # of irrigated grid cells set to zero [%]
Barley	100	51%	0.61%	64%
Wheat	100	61%	0.23%	53%
Maize	100	74%	0.67%	62%
Potato	100	51%	0.32%	40%
Sugarcane	100	40%	0.03%	14%
Sunflower	100	41%	0.66%	46%
Rapeseed	100	18%	0.54%	45%
Oil palm	100	n.a.	n.a.	n.a.
Cotton	100	58%	0.06%	33%
Rice	100	62%	0.03%	15%
Rye	100	n.a.	n.a.	n.a.
Sugar beet	100	n.a.	n.a.	n.a.
Soybean	100	45%	1.01%	60%
Groundnut (peanut)	100	25%	0.46%	35%
Sorghum	100	37%	1.00%	50%

The exclusion of very low irrigation volume affects quite a large proportion of grid cells (14% up to 64%) but reduces overall irrigation volume by typically less than 1%.

7.1.2. Treatment of very high irrigation volumes

We cap all irrigation volume values which exceed the 98th percentile to the value of the 98th percentile in order to avoid unrealistically high irrigation volume values.

Table 20: Crop-specific threshold for the treatment of very high irrigation volumes and the corresponding effect on the total number of grid cells and the global irrigation volume

Crop	Threshold, 98 th percentile [m ³ /ha*yr]	Relative irrigation volume removed [%]
Barley	6'819	69%
Wheat	18'339	49%
Maize	10'701	30%
Potato	19'192	37%
Sugarcane	38'124	51%

Sunflower	5'981	34%
Rapeseed	4'475	18%
Oil palm	n.a.	n.a.
Cotton	66'426	29%
Rice	55'217	36%
Rye	n.a.	n.a.
Sugar beet	n.a.	n.a.
Soybean	3'842	48%
Groundnut (peanut)	4'206	48%
Sorghum	4'681	40%

This operation does not reduce the number of grid cells with irrigation but has an important influence on the total irrigation volume because the capped outliers are very large. Due to its importance, we will review and refine this approach with the next iteration of geoFootprint.

7.2. Inventory flows to model irrigation water use

The total amount of annual irrigation affects the entire modeling of water in the crop datasets. Consequently, we have to adapt (regionalize) all water flows listed in the default inventory to develop a water balanced, regionalized inventory dataset.

The following water flows are used across all default crop inventories:

- **Water input from nature:** Across all crop datasets, direct water inputs originating from river, groundwater, or unspecified origin (river or ground water) can be present.
 - Water river, natural resource, from water.
 - Water well, natural resource, from water.
 - Water, unspecified natural origin.
 - Natural resource, from water
 - Natural resource, from ground
- **Irrigation datasets, WFLDB (input from technosphere).** The WFLDB distinguishes between three irrigation technologies: drip, surface and sprinkler. The corresponding datasets include all relevant material and energy requirements but *not* the actual water input—the input of water is modelled in the actual crop datasets, i.e., as a separate water input from nature.

- **Water treatment (output to technosphere).** Some crop datasets may produce polluted water which is then treated with a water treatment process that transforms the polluted water into potable water. To date, we do not offer the possibility to manipulate these flows, i.e., they are part of the regionalized inventory but left unchanged. Note that water treatment flows are not considered in the overall water balance.
- **Water output to nature:** We follow the WFLDB guidelines and distinguish between the following water outputs:
 - Water, to surface water (water emission in the form of surface water)
 - Water, to ground water (water emission in the form of ground water)
 - Water, unspecified to air (the evapotranspiration from irrigation to air).

7.3. Modeling of irrigation

Recall that all calculations are applied on the grid cell level. In order to develop a water balanced, regionalized irrigation dataset, we apply the following modeling procedure.

7.3.1. Water input

There can be several water inputs present per crop dataset and consequently, our algorithm needs to distribute the total water withdrawal to all water input flows present in the default inventory, basically by using the following procedure:

1. We first calculate how much of the total water withdrawal each type of water input in the default inventory contributes in order to determine the relative contribution of each type of water input (see column 3, Table 22).

$$Water_{rel_withdrawal_type_i} = \frac{Water_{withdrawal_type_i}}{\sum_{i=0}^n Water_{withdrawal_type_i}} \quad (1)$$

2. We then use the relative contribution of each water input to compute the total amount of water associated with each water input (see column 4, Table 22), by multiplying the total irrigation water withdrawal (column 1) with the relative contribution (column 3).

$$Water_{withdrawal_type_i} = Irrig * Water_{rel_withdrawal_type_i} \quad (2)$$

Table 21: Calculation of water inputs from nature in the regionalized inventory from the information present in the default inventory (example).

(1) Irrigation water (<i>Irrig</i>) [m3]	(2) Water input according to default inventory	(3) Relative relevance according to default inventory (<i>Water_{rel_withdrawal_type}</i>) [-]	(4) Amount of water input from nature listed in regionalized inventory (<i>Water_{withdrawal_type}</i>)[m3]
500	Water, unspecified, natural origin	0.9	450
	Water, well	0.1	50

7.3.2. Irrigation technologies

Typically, there is often more than one irrigation technology per crop present in the default inventory, mainly because crop production datasets in WFLDB typically reflect average practices at country level. Therefore, our algorithm needs to distribute the total water withdrawal to all the irrigation technologies present in the default inventory, using the following procedure:

1. We first calculate how much of the total water withdrawal each type of irrigation flow in the default inventory contributes in order to determine the relative contribution of each type of irrigation technology (see column 3, Table 22).

$$Water_{rel_irrigation_type_i} = \frac{Water_{irrigation_type_i}}{\sum_{i=0}^n Water_{irrigation_type_i}} \quad (3)$$

2. We then use the relative contribution of each irrigation type to compute the total amount of irrigation associated with each irrigation technology (see column 4, Table 22), by multiplying the water withdrawal (column 1) with the relative contribution (column 3).

$$Water_{irrigation_type_i} = Irrig * Water_{rel_irrigaton_type_i} \quad (4)$$

Table 22: Calculation of water withdrawal in the regionalized inventory from the information present in the default inventory (example).

(1) Irrigation water (water withdrawal) [m3]	(2) Irrigation technologies according to default inventory	(3) Relative relevance according to default inventory [-]	(4) Amount of irrigation listed in regionalized inventory [m3]
500	Irrigating, sprinkler, electricity powered (WFLDB 3.5) # CA	0.64	320
	Irrigating, surface, electricity powered (WFLDB 3.5) # CA	0.36	180

7.3.3. Water emissions

For each regionalized inventory, we calculate three output flows:

- **Water emitted to air:** The water emitted to air through the natural process of evapotranspiration, which depends on climate conditions and growth stage of the crop is equal to the amount of water from evapotranspiration from irrigation, i.e., the loss due to inefficiency in irrigation. It is therefore equal to ET_{irr} (see equation 1).

$$WaterToAir = ET_{irr} \quad (5)$$

- **Water emitted to surface water:** We assume that 80% of the water applied in excess in relation to crop requirements (evapotranspiration) would be lost in the form of surface water (T. Nemecek, 2019).

$$WaterToSurfaceWater = 0.8 * \left(\left(\frac{ET_{irr}}{EF_{irr}} \right) - ET_{irr} \right) \quad (6)$$

- **Water emitted to ground water:** We assume that 20% of the water applied in excess in relation to crop requirements (evapotranspiration) would be lost in the form of surface water (T. Nemecek, 2019)

$$WaterToSurfaceWater = 0.2 * \left(\left(\frac{ET_{irr}}{EF_{irr}} \right) - ET_{irr} \right) \quad (7)$$

8. Regionalizing soil erosion

Soil erosion consists of the detachment, transport and deposition of soil particles and is considered a major problem across the world (Faist Emmenegger, 2009). Soil erosion in itself is not recorded as an exchange flow in the AUPD. In fact, it is a key input into the modeling of P leaching and in itself an important proxy on the environmental performance—soil quality—of crop-production systems.

There are two drivers of erosion: water and wind. Due to the lack of appropriate data, wind erosion is currently not considered in our erosion model. Consequently, the eroded volume is underestimated, particularly in regions where the wind erosion potential is high (see Figure 5 in (Faist Emmenegger, 2009)).

We model soil erosion on the basis of the widely used Universal Soil Loss Equation (USLE) (Wischmeier & Smith, 1978) (see equation 1). The linear equation calculates the annual long-term erosion caused by water. Note, however, that it does not predict the erosion associated with extreme events.

We perform all calculations at the grid cell level. That is, we calculate soil erosion (A) for each and every grid cell where a crop is cultivated under consideration of spatially explicit data for R , k , LS , & CP and default data for $c1$, $c2$, P .

$$A = R * k * LS * c1 * c2 * P * CP \quad (1)$$

Where:

- A represents the potential long-term annual soil loss [metric ton/(ha*a)].
- R represents the erosivity factor [MJ mm / (ha * h * a)]. The erosivity factor represents the erosive force of rainfall. The numerical value of the R factor should quantify the raindrop impact effect and provide information about the related rate and amount of runoff (Wischmeier, 1978). We retrieve the spatially explicit R factor from Borrelli et al (2017))
- k represents the erodibility factor [t*h/ MJ*mm]. The erodibility represents the soil susceptibility to erosion (Wischmeier, 1978). It corresponds to the difference in the eroded soil quantity between two soils under the same conditions (same rainfall, same

slope, etc.) due to soil properties. We retrieve the spatially explicit k factor from Borrelli et al (2017)).

- LS represents the length-slope factor [-]. Is computed on the basis of the slope length and the slope steepness (in %). We retrieve the spatially explicit LS factor from Borrelli et al (2017).
- c1 represents the crop factor [-]. The original USLE method combines the crop and the tillage factors in a single factor. We distinguish between the influence of the crop (c1) and the applied tillage (c2). We assume that this provides a more powerful foundation for the discrimination of cultivation practices than the combined factor. The crop factor are taken from Faist Emmenegger et al. (Faist Emmenegger, 2009) and complemented with data from Scherer & Pfister (Scherer, 2015). (See Table 23.)
- c2 represents the tillage factor [-]: The tillage factor (C2) compares the effectiveness in preventing soil loss of different crop management systems. We use the tillage factors detailed in (T. Nemecek, 2019) to distinguish between relevant tillage factors—see Table 24. We use the tillage factors detailed in Table 25 for the default calculation.
- P represents the practice factor [-]: The practice factor reflects the effectiveness of different cropping practices in reducing runoff and thereby erosion (Scherer, 2015). We use the practice factors from Faist Emmenegger et al. (Faist Emmenegger, 2009) to distinguish between relevant anti-erosion practices (see Table 26). We apply an anti-erosion practice of 1 (“no anti-erosion practice) for all grid cells whose LS factor does not exceed 3. For all grid cells which exceed a certain level of steepness, i.e., with a LS factor > 3, we apply *contour farming*, i.e., a factor of 0.5.
- CP represents the cultivation period.

The open source version of the spatially explicit data from Borrelli et al (2017) is only available in a grid cell resolution of ~25 x 25 km and therefore transformed (warped) into our native resolution of ~10 x 10 km.

Table 23: Crops-specific c1 factor for major crops in geoFootprint. Source: (Faist Emmenegger, 2009)

Crop	c ₁ factor	Source
Potato	0.40	Nemecek et al. 2019
Sugar beet	0.50	
Sugarcane	0.30	
Sorghum	0.40	

Rapeseed	0.50	(Scherer, 2015)
Soybean	0.40	
Oil palm	0.20	
Cotton	0.50	
Maize	0.40	
Barley	0.35	
Wheat	0.35	

Table 24: Tillage factor for each tillage method. The global factor used for the default calculation are marked in grey. Source: (Faist Emmenegger, 2009). Note that we adapted the factor for “No tillage” from 0 to 0.1 because of the potential presence of water driven soil erosion even in the absence of tillage.

Tillage method	C ₂ factor	Description
Fall plow	1	Traditional plowing after the harvest. Aim: turn over the upper layer of the soil, bury the crops residues of the harvest and bring fresh nutrients to the surface. It can be performed with some sophisticated machines or with a rudimentary plow pulled by horses or oxen.
Spring plow	0.9	Traditional plowing carried out before a new sowing, after the cold months. Aim: break up the soil for planting the seeds among others. It can be performed with some sophisticated machines or with a rudimentary plow pulled by horses or oxen too.
Mulch tillage	0.6	Mulch tillage is also known as conservation tillage, reduced tillage or minimum tillage. It leaves at least 30% of the soil surface covered by residues. Mulch tillage can be achieved with a chisel plow, with discs or with rotary till implements, for instance.
Ridge tillage	0.35	Ridge tillage is a tillage method maintaining permanent ridges on which row crops are grown. After harvest, the crop residues are left. The new seeds are planted in the ridge, after pushing residues aside.
Zone tillage	0.25	Zone tillage is also called strip tillage. It is a reduced tillage method that limits soil disturbance to the area (strips of 5 to 20 cm) of the planting row and leaves the areas between the crop rows undisturbed.
No tillage	0.1	No-till systems, or zero tillage systems, do not use tillage for establishing a seedbed. Crops are simply planted into the previous year’s crop residue. The seeds are drilled into the soil without prior land preparation.

Table 25: Crops-specific C₂ factor used for default calculation.

Crop	C ₂ factor
Potato	0.9
Sugar beet	0.9
Sugarcane	0.1
Sorghum	0.9
Rapeseed	1
Soybean	0.9
Oil palm	0.1
Cotton	1
Barley	1
Wheat	1
Maize	0.9

Sunflower	0.9
Groundnut	0.9
Rye	1
Rice	0.9

Table 26: Practice factor for each anti-erosion practice. The global factor used for the default calculation are marked in grey. Source: (Faist Emmenegger, 2009)

Anti-erosion practice	P factor	Description
Up & down slope (no practice)	1	Crop is cultivated in the direction of the slope of the hill. It is the traditional cropping system. It corresponds to a “no anti-erosion practice” system.
Cross slope	0.75	Cultivating (planting, tilling and other farming operation) is done perpendicular to the direction of the hill slope.
Contour farming	0.5	Crop rows follow the contours. Contour farming entails performing all planting and tillage operations on or near the same elevation.
Strip cropping, cross slope	0.37	Cross slope cropping system where two or more crops are cultivated in strips alternatively. The crop stripes are perpendicular to the slope.
Strip cropping, contour	0.25	Contour farming where two or more crops are cultivated in strips alternatively. The crop stripes follow the contours.

9. Regionalizing N & P emissions

Nutrients, like phosphorus and nitrogen, play key roles in plants (oxygen transfer, protein building, etc.) and inputs of these nutrients are essential for profitable crop agriculture. They can, however, have some adverse effects when they are present in high concentration in some environmental compartments and are of major concern in water bodies where they cause eutrophication.

9.1. Nitrate emissions

We compute nitrate leaching with the regression model developed by (De Willigen 2000) and (Roy et al., 2003).

$$A_{Nitrate} = \left[21.37 + \frac{P}{c*L} * [0.0037 * S + 0.0000601 * N_{org} - 0.00362 * U] \right] \quad (1)$$

Where:

- $A_{Nitrate}$ represents the nitrate loss through leaching [kg N/(ha*a)].
- P represents the total annual precipitation and irrigation for the whole year [mm/a]. We use spatially explicit data from GAEZ for precipitation. Note that using the total, instead of a cultivation period specific, value for annual precipitation is rather conservative since the actual precipitation for most crops during the cultivation period will be lower. We use the spatially explicit amount of irrigation water in mm.
- c represents the clay content [%]. We use the spatially explicit data from ISRIC soil grids (Hengl, 2014).
- L represents the root depth [m]. We use the crop-specific default values from the WFLDB guidelines. (See Table 2.)
- S represents the total nitrogen supply [kg N/(ha*a)]. We consider the total nitrogen supplied through organic and mineral fertilizer.
- N_{org} represents the mass of organic nitrogen contained in the upper 50 cm of soil [kg/ha]. Naturally only a fraction of this mass is mineralized and, hence, available for

uptake by plants and leaching to the ground water. This fraction is determined by the mineralization rate, which is 1.6% here and implicitly included in the regression coefficient (0.0000601) of the term N_{org} (T. Nemecek, 2019). We calculate N_{org} with equation 2.

- U represents the Nitrogen uptake [kg N/(ha*a)]. We use the crop-specific default values from EU Nitrogen Expert Panel. (See EU Nitrogen Expert Panel, 2016.)

$$N_{org} = \frac{\left(\frac{C_{org}}{1000} * V * B_d * SfoCf\right)}{r_{C/N} * r_{N_{org}}} \quad (2)$$

Where:

- C_{org} represents the soil organic carbon content (fine earth fraction) [g/kg]. Note that we use the soil organic carbon content up to a depth of 50cm in order to be consistent with V . A division by 1000 yields the mass fraction in %.
- V represents the soil volume [m³/ha]. V is taken to be 5000 m³, which means that the upper 50 cm of soil are considered.
- B_d the bulk density of the soil [kg/m³]. We use the spatially explicit data preprocessed from ISRIC soil grids (Hengl, 2014). Note that we use the bulk density up to a depth of 50 cm in order to be consistent with V .
- $SfoCf$ represents the soil free of any coarse fragments in the form of a fraction. We use the spatially explicit data preprocessed from ISRIC soil grids (Hengl, 2014). The multiplication with the fraction removes the coarse fraction of the soil without a soil organic carbon content.
- $r_{C/N}$ represents the ratio between C/N and is taken to be 11 (T. Nemecek, 2019).
- $r_{N_{org}}$ represents the ratio of N_{org} to N_{tot} (total soil nitrogen) [-]. The ratio $r_{N_{org}}$ is needed to calculate N_{org} from N_{tot} . $r_{N_{org}}$ is assumed to be 0.85 (T. Nemecek, 2019).

Note that negative values are avoided by testing the final result with equation 3.

$$N = \begin{cases} N & \text{if } N \geq 0 \\ 0 & \text{if } N < 0 \end{cases} \quad (3)$$

9.2. Phosphorus emissions

Phosphorus emissions refer to the phosphorus quantities transported from soil to waters and further causing potential eutrophication. We use the emission models described in the WFLDB guidelines (T. Nemecek, 2019) with selected (spatially-relevant) extensions published in Scherer & Pfister (Scherer, 2015). The model distinguishes the losses in surface water from those in groundwater and the transport in particulate and dissolved form:

- Phosphorus loss in particulate form to surface water: soil erosion
- Phosphorus loss in dissolved form (phosphates) to surface water: run-off and drainage
- Phosphorus loss in dissolved form (phosphates) to groundwater: leaching

9.2.1. Phosphorus loss in particulate form to surface water: soil erosion

Phosphorus loss in particulate form to surface water occurs through soil erosion and is computed with the same mathematical form, than in the WFLDB guidelines (T. Nemecek, 2019). However, we use spatially explicit input data for soil erosion, Phosphorus concentration in soil, and for the cultivation period. (See the explanation of the respective terms below.)

$$P_e = A * 1000 * r * e * P_{conc} * \frac{days}{365} \quad (1)$$

Where:

- P_e represents the Phosphorus loss to surface water [kg/(ha*a)]. We inventory the Phosphorus loss as an emission to water with the flow *Phosphorus, to surface water*.
- A represents potential long-term annual soil loss [metric ton/(ha*a)].
- r represents the fraction of phosphorus loss which effectively reaches the surface water system [-]. We work with a global default value of 0.2. The value for the coefficient expresses the fact that only a certain proportion of the eroded phosphorus effectively reaches the water system is taken from (Prahsun, 2006).
- e represents the enrichment factor [-]. The value for the enrichment factor is 2 and taken from (Roy, 2003). This factor takes account of the fact that the eroded soil particles contain more P than the average soil.
- P_{conc} represents the P content in the topsoil [kg P/kg soil]. We derive the concentration from the spatially explicit P area density, P_{ad} [g/m²] from Yang et al. (YANG, 2014). In

line with Scherer and Pfister (2015), we disregard (subtract) occluded P as it is assumed to be biologically unavailable. We compute the phosphorus concentration by dividing the P area density by the bulk density of the soil [kg/m³] obtained from ISRIC soil grids (ISRIC). The multiplication with 1/1000 and with 0.5 (soil depth in meters considered), respectively, transforms both the denominator and the numerator, into the common unit of [kg/m²].

$$P_{conc} = \frac{(P_{ad} - P_{occ}) * \frac{1}{1000}}{0.5 * B_d} \quad (2)$$

- days represents the duration of cultivation [days]. The coefficient related to the duration of cultivation is applied because the phosphorus loss through erosion is computed with an annual soil loss. Without this correction for the cultivation period, we would also account for the phosphorus loss outside of the crop cultivation period.

The most influential parameter in this equation is the quantity of eroded soil.

9.2.2. Phosphorus loss to surface water (run-off and drainage)

9.2.2.1. Run-off to surface water

In order to calculate P run-off to surface water, we expand the classical equation of the WFLDB guidelines by the slope factor s:

$$P_{ro} = s * P_{rol} * F_{ro} \quad (3)$$

Where:

- P_{ro} represents the quantity of P lost through run-off to rivers [kg/(ha*a)]. We inventory the P loss as an emission to water with the flow *Phosphate, to surface water*.
- s represents the dimensionless slope factor [-]. In line with Scherer & Pfister, the binary variable was set to zero for slopes less than 2%¹⁸—assuming that under this condition, not run-off to surface water occurs—and to 1 for slopes equal or greater than 2% (Prahsun, 2006). We use the spatially explicit slope class data from the harmonized

¹⁸ This corresponds to a slope class equal or smaller than 2. Available slope classes (sc) in the dataset are sc1=0-0.5%, sc2=0.5-2%, sc3 = 2-5%, sc4 = 5-10%, sc5 = 10-15%, sc6 = 15-30%, sc7 = 30-45%, sc8 = 45%.

world soil database (HWSD) to determine the slope (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009).

$$s = \begin{cases} 1 & \text{if slope class} > 2 \\ 0 & \text{if slope class} \leq 2 \end{cases} \quad (4)$$

- P_{rol} represents the average quantity of P lost through run-off for a land use category [kg/(ha*a)], which is 0.175 kg P/(ha*a) for arable land (Scherer, 2015).
- F_{ro} represents the dimensionless correction factor for fertilization with P [-], calculated as:

$$F_{ro} = 1 + \frac{0.2}{80} * P_{min} + \frac{0.7}{80} * P_{liq} + \frac{0.4}{80} * P_{man} \quad (5)$$

Where:

- P_{min} represents the quantity of P₂O₅ contained in mineral fertilizer [kg/ha]
- P_{liq} represents quantity of P₂O₅ contained in slurry or liquid sewage sludge [kg/ha]
- P_{man} represents the quantity of P₂O₅ contained in solid manure [kg/ha].

The values of P₂O₅-content for slurry and manure were adapted from Flisch et al. (Flisch et al., 2009).

Table 27: N and P₂O₅ content for different types of farm-yard manure. Source: Flisch et al. (2009).

Crop	N content [kg N/metric ton]	P ₂ O ₅ content [kg P ₂ O ₅ /metric ton]
Solid manure, cow	3.7	1.8
Liquid manure cow	3.7	1.8
Solid manure, pig	3.7	1.8
Liquid manure, pig	3.7	1.8
Solid manure broiler	3.7	1.8

9.2.2.2. Drainage to surface water

To calculate P losses due to drainage to surface water, we expand the classical WFLDB guideline equation with the dimensionless drainage factor proposed in Scherer & Pfister (Scherer, 2015):

$$P_d = P_{gwl} * F_{gws} * F_d \quad (6)$$

Where:

- P_d represents P losses due to drainage to surface water [kg/(ha*a)]. We inventory the P loss as an emission to water with the flow *Phosphate, to surface water*, i.e., we add P_{gw} and P_{ro} to compute total Phosphate loss to surface water.
- P_{gwl} represents the average quantity of P leached to ground water for a land use category [kg/(ha*a)], which is 0.07 kg P/(ha*a) for arable land (Scherer, 2015).
- F_{gws} represents the correction factor for fertilization by slurry [dimensionless], calculated as:

$$F_{gws} = 1 + \frac{0.2}{80} * (P_{man} * fract_man_liquid) \quad (7)$$

Where: ($P_{man} * fract_man_liquid$) represents the quantity of P₂O₅ contained in the slurry or liquid sewage sludge [kg/ha]. (See Table 27.)

- F_d representing the dimensionless drainage factor [-], calculated as:

$$F_d = \begin{cases} 6 & \text{if } Drained > 3 \\ 0 & \text{if } Drained \leq 3 \end{cases} \quad (8)$$

Where *Drained* represents the drainage class. Information on drainage was given in the HWSD (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009). Conventional drainage was assumed for drainage classes above 3, which is at least *moderately well drained*. (Scherer, 2015).

9.2.3. Phosphorus loss in dissolved form (phosphates) to groundwater: leaching

We calculate P leaching to the ground water with the same mathematical form than P losses due to drainage to surface water, but use a regionalized correction factor proposed by Scherer & Pfister (Scherer, 2015) (here F_{gw} , above F_d) derived from the spatially explicit drainage data of HWSD. This factor acts diametral to F_d , i.e.,

$$F_{gw} = \begin{cases} 0 & \text{if } F_d > 0 \\ > 0 & \text{if } F_d = 0 \end{cases} \quad (9)$$

That is, either we have a phosphorus loss in dissolved form to groundwater or we have a phosphorus loss due to drainage to surface water, but never both. Equation 13 shows the calculation of P leaching to groundwater:

$$P_{gw} = P_{gwl} * F_{gws} * F_{gw} \quad (10)$$

Where:

- P_{gw} represents P leached to ground water [kg/(ha*a)]. We inventory the P leached as an emission to water with the flow *Phosphate, to groundwater*
- F_{gw} represents the dimensionless groundwater factor [-], calculated as:

$$F_{gw} = \begin{cases} 0 & \text{if } Drained > 3 \\ 1 & \text{if } Drained \leq 3 \end{cases} \quad (11)$$

10. Regionalizing emissions from land use change from tree cover loss

Impacts from land use change (LUC) are often a key driver of environmental impacts associated with crop production. Bearing this in mind, we developed a dedicated module for geoFootprint to compute LUC impacts based on the tree cover loss datasets of Global Forest Watch (GFW)¹⁹.

The sections below outline the computational framework to calculate and allocate LUC impact (associated with tree cover loss) in a crop-specific manner and on a global scale, its present limitations and gives an outlook which highlights how the approach can be further refined.

10.1. Grid-cell based calculation of LUC impact

Due to our grid cell size of 5 arc minutes or ~10x10 km, we will typically find much more than one agricultural commodity in one grid cell. This induces the need to allocate the LUC associated with a grid cell to all agricultural commodities produced by that grid cell. We use the shared responsibility approach (SRA) (T. Nemecek, 2019) to perform the allocation of tree cover loss to all crops cultivated within a grid cell.

The SRA allocates tree cover loss or LUC on the basis of the physical area occupied by a particular crop in relation to the physical area occupied by all other crops in a grid cell. That is, LUC is allocated to all crops which occupy area in a grid cell, even if they are (or have been) decreasing. We use spatially explicit data on 175 crops from EarthStat (Monfreda, 2008) to determine the relative importance (harvested area) of each crop throughout all grid cells on a global scale

The detailed procedure to perform the allocation of GFW tree cover loss data involves five steps:

¹⁹ <https://www.globalforestwatch.org/>

- **Step 1: Aggregate LUC:** Overlay grid cell with GFW tree cover loss dataset (which is in a resolution of roughly 30x30 meter) and compute total tree cover loss for each grid cell in a 5 arc minute resolution on a global scale for the available temporal tree cover loss data, i.e., 18 years. This step was executed by the Global Forest Watch team at the World Resources Institute (WRI).
- **Step 2: Expand LUC to include below-ground biomass (BGB) and dead organic matter (DOM).** Expand GFW data to cover all relevant biomass pools and into corresponding CO₂ fluxes, i.e., integrate below-ground biomass and dead organic matter and transform the total biomass into the corresponding CO₂ flux.
- **Step 3: Determine allocation factor.** Compute the relative proportion of the occupied area for each grid cell and across all 175 crops provided by EarthStat (Monfreda, 2008).
- **Step 4: Allocate LUC to crops.** Perform the shared-responsibility allocation per grid cell on the basis of the relative proportion of the agricultural area occupied by each crop.
- **Step 5: Normalize to one hectare & annualize.** Allocated LUC impacts per crop refer to the entire grid cell and are divided by the total area cultivated with a particular crop in order to relate LUC impacts to one hectare. In addition, LUC impacts are equally distributed through time, i.e., annualized.

This processing sequence basically generates and manipulates four data elements.

- **Total Area Loss (TAL),** i.e., the total tree cover area in hectare lost in the last 18 years, i.e., since 2001 (we will work with 20 years once the temporal time series data becomes available). Note that tree cover loss covers the most important land use change category—the transformation from forest and potentially also shrubland—but ignores other important land use changes such as the transformation from natural grassland. The annualized area loss (provided by step 5) is required to compute the changes in soil organic carbon (SOC) associated with the change in land use from forest to arable land. In addition, it is provided for interpretive purpose, namely, to give the users of geoFootprint an indication of the total tree area loss (TTAL) associated with a particular crop in a particular or across several grid cells.
- **Total Biomass Loss (TBL),** i.e., the total biomass (AGB, BGB and DOM) lost in the form of tree cover biomass loss in the last 18 years. The normalized and annualized biomass loss is provided for interpretive purpose, namely, to give the users of geoFootprint an

indication of the total annualized biomass loss (TBL) associated with the corresponding area loss of a particular crop in a particular or across several grid cells.

- **Total CO₂ Loss (TCL)**, i.e., the total CO₂ fluxes in the last 18 years associated with the TAL and the TBL. The annualized TCL represents the “annual CO₂ emissions from tree cover loss per hectare crop cultivated” and basically combines the TAL with the TBL.
- **Allocation factors**, i.e., the area occupied by a particular crop in relation to the area occupied by all other crops—the total area under agricultural production— in a grid cell.

The overall procedure to aggregate, expand, allocate, normalized and annualize LUC impacts is summarized in Figure 19 below for an exemplary set of eight grid cells.

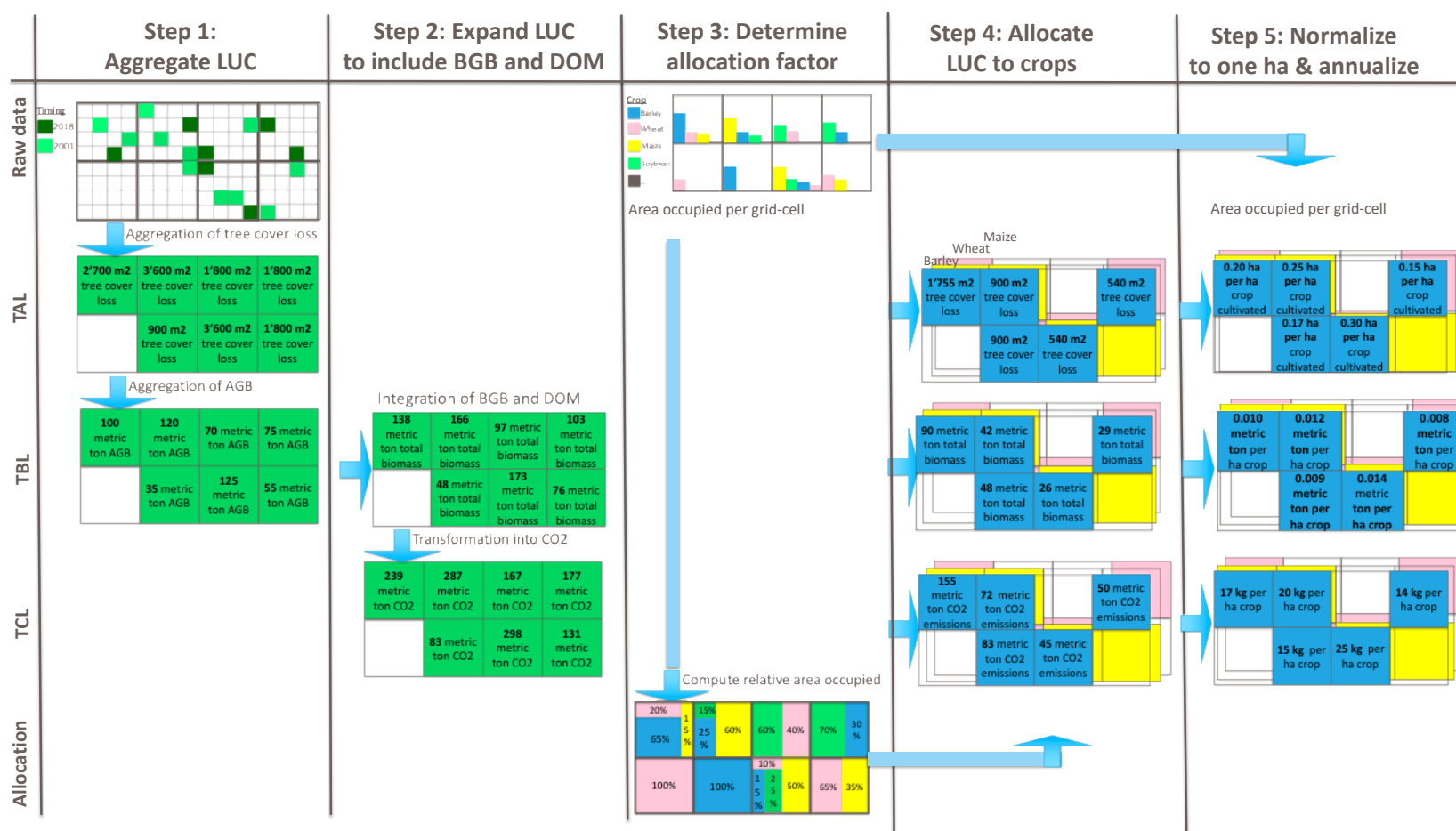


Figure 19: Procedure to aggregate, expand, allocate, normalize and annualize LUC impacts associated with tree cover loss to crops according to the shared-responsibility approach for an exemplarily set of eight grid cells. The columns indicate the steps and the rows the raw data used and the data elements processed/ generated. Note that step 4 produces three additional layers for each and every key crop under consideration. For simplicity, the graph only shows three key crops (barley, wheat and maize). Abbreviations: AGB; Above Ground Biomass, BGB; Below Ground Biomass, DOM; Dead Organic Matter, LUCM; Land use change, TAL; Total Area Loss, TBL, Total Biomass Loss, TCL; Total Carbon Dioxide Loss.

In the following section, we detail each of the 5 steps highlighted above. If not stated otherwise, the computation is always applied on a global scale, i.e., to all grid cells.

10.1.1. Aggregating LUC

We use selected data elements of the global WRI dataset, i.e., with the

- threshold
- area_loss_year
- biomass_loss_year

We apply some further data processing to expand the biomass to also include BGB and DOM and to compute corresponding CO₂ fluxes. We do not work with the carbon_emissions_year since it only quantifies the CO₂ fluxes associated with AGB. Following the advice from WRI, we do not consider tree_cover_gain since gain and loss do not follow the same temporal resolution and are therefore difficult to relate to each other.

Note that the data processing operates on one particular threshold. The GFW dataset includes three thresholds, which can be interpreted as three variable forest definitions, for the detection of tree cover loss: 30%, 50% and 70%. All canopy cover above a threshold is reported as tree cover loss, whereas tree cover loss below the threshold is not reported. Consequently, the total tree cover loss reported decreases with the increase in threshold. For example, if we select a threshold of 70% canopy cover—the most stringent forest definition available—all tree cover loss associated with a lower canopy cover is not reported as tree cover loss. Currently, we work with a threshold of 50%.

To give a rough idea of the differences between the thresholds, Figure 20 shows the global carbon dioxide emissions per grid cell for a threshold of 30% (top) and 70%. (bottom).

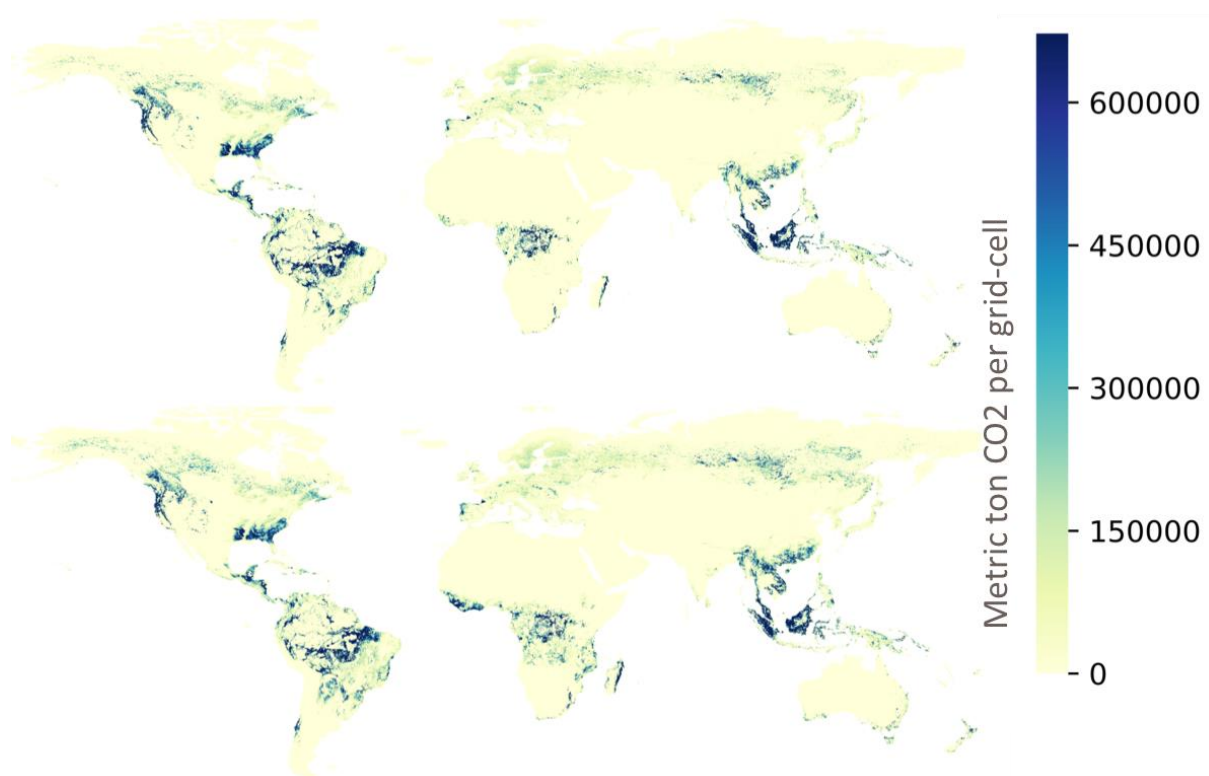


Figure 20: Global CO₂ emissions per grid cell in native geoFootprint resolution of ~10x10km for a canopy cover threshold of 30% (top) and 70% (bottom). One can spot large differences in Africa (e.g., Congo, Ivory Coast) but also in Russia.

10.1.2. Expanding LUC

In order to establish a comprehensive accounting which considers all relevant carbon pools associated with the tree cover loss, we have to expand the biomass_loss_year with the root-to-shoot (RtS) ratio into below ground biomass (BGB). According to the IPCC 2006 (see chapter 4, Table 4.4) the RtS varies between 0.20 and 0.56 according to

- The ecological zone
- The above-ground biomass content of the forest
- The dominant tree species (IPCC, 2006)

Due to the lack of a spatially explicit ecological zone map, we cannot discriminate appropriately between the different RtS value. We therefore apply the RtS ratio of tropical forests, i.e., of 0.37, on a global scale. This captures BGB values of tropical forests correctly but causes an underestimation of the BGB in dry forests and shrubland and an overestimation of BGB in all other forest biomes.

In addition, we add the biomass typically found in the form of dead organic matter (DOM), which includes dead wood (lying on the ground) as well as litter and leaves, in the form of a global default factor of 5.2 metric ton of (dry) biomass per ha of tree cover lost (IPCC, 2006).

Overall, the annual total biomass loss (TBL) associated with the tree cover loss in a particular year y is calculated with equation 1 (IPCC, 2006). See equation 2.13.

$$BL_y = AGB_y + (AGB_y * RtSR_c) + DOM \quad (1)$$

Where BL is the biomass loss of year y and RtSR is the root-to-shoot ratio of climate zone C (as mentioned above, we currently work with a global default value of 0.37).

Once we have the BL computed for each grid cell and loss years, we transform it into carbon dioxide (CO₂) by multiplying with the carbon fraction (cf) in dry mass—we use a global default factor of 0.47 (see IPCC2006, chapter 4, table 4.3)—and by multiplying with the mol ratio between C and CO₂—44/12 (IPCC, 2006). The total loss of biomass in the form of CO₂ is calculated with equation 2 (IPCC, 2006). See equation 2.13 in the guidelines.

$$CL_y = BL_y * cf * \frac{44}{12} \quad (2)$$

Next, we sum over all available loss years of the AL (area loss per year), the BL and the CL in order to obtain the cumulated or total loss, since the year 2001, of tree cover area (in ha), TAL, biomass (in metric tons oven dry mass), TBL and CO₂ (in metric ton CO₂), TCL.

$$TAL = \sum_y AL_y \quad (3)$$

$$TBL = \sum_y BL_y \quad (4)$$

$$TCL = \sum_y CL_y \quad (5)$$

Optionally, we could also work with the annual (instead of the total) tree cover, biomass, and CO₂ loss. This is of relevance if we want to distribute CO₂ emission in a non-equal fashion throughout the years, e.g., if we want to apply a linear (instead of an equal) depreciation/annualization scheme.

10.1.3. Determine crop-specific allocation factors

In order to compute the relative area occupied by a crop in each grid cell we execute the following steps:

1. Load the harvested area of all 175 crops provided by EarthStat into our grid cell structure, i.e., a pandas dataframe (Monfreda, 2008). We interpret harvested area as area occupied although the actually harvested area is typically smaller than the actually occupied or physically cultivated area. Note that we use the extrapolated and rescaled harvested area (see section 3.1.1).
2. Compute the total harvested agricultural area (TAO) across all crops for each grid cell. Note that the total harvested agricultural area can exceed the size of a grid cell since EarthStat treats one hectare of a crop which is harvested twice per year as a harvested area of 2 hectare, i.e., in the form of a physical expansion of area. Note that this accounting principle increases the potential punishment (relative area occupied which is used as an allocation factor) of crops which are harvested more than once in one year.
3. Compute the relative proportion of harvested area of each crop (AO_{rel}) by dividing the harvested area of a particular (key) crop by the total harvested agricultural area of all crops cultivated in the grid cell. We perform this computation only for the 15 key crops currently focused in geoFootprint.

We implement the computation of step 2 and 3 on the basis of equation 6 and 7.

$$TAO = \sum_c AO_c \quad (6)$$

$$AO_{rel_c} = \frac{AO_c}{TAO} \quad (7)$$

10.1.4. Allocation LUC to key crops

Once we have computed (i) the TAL, TBL and the TCL and (ii) the relative proportion of harvested area of each crop, we can perform the allocation of the TAL, TBL and TCL of a particular grid cell, by multiplying it with the relative harvested area of the crop of interest in that grid cell (see equation 9, 10 and 11).

$$A_{TAL_c} = TAL * AO_{rel_c} \quad (8)$$

$$A_{TBL_c} = TBL * AO_{rel_c} \quad (9)$$

$$A_{TCL_c} = TCL * AO_{rel_c} \quad (10)$$

The results are crop-specific meaning that for each considered key crop, we receive one global layer of allocated LUC impacts.

10.1.5. Normalization to one hectare

We next normalize the grid cell specific results to the cultivation of one hectare of each crop. We first divide the total allocated tree cover loss, associated with a particular crop, by the number of hectares cultivated with a crop in that grid cell. This results in the normalized total loss of tree cover area, biomass and CO₂ per hectare over the last 18 years (see equation 11, 12, 13)

$$NA_TAL_c = \frac{A_TAL_c}{(TAO * AO_c)} \quad (11)$$

$$NA_TBL_c = \frac{A_TBL_c}{(TAO * AO_c)} \quad (12)$$

$$NA_TCL_c = \frac{A_TCL_c}{(TAO * AO_c)} \quad (13)$$

Where TAO denotes the total area occupied by all crops in a grid cell, meaning that we divide by the number of hectares cultivated with a crop in a grid cell.

Finally, we annualize (distribute) the total loss by dividing it by the total time period under observation, i.e., currently 18 years. That is, each year receives the same proportion of the total impacts (see equation 14, 15, 16). Alternative distribution of impacts through time, such as linear depreciation, are also feasible but currently not applied.

$$ANA_TAL_c = \frac{NA_TAL_c}{y} \quad (14)$$

$$ANA_TBL_c = \frac{NA_TBL_c}{y} \quad (15)$$

$$ANA_TCL_c = \frac{NA_TCL_c}{y} \quad (16)$$

In addition, since CO₂ emissions in the inventory are recorded in kilograms, we multiply the metric ton of normalized and annualized CO₂ loss per hectare by a factor of 1000.

$$ANA_KG_{TCL_c} = ANA_{TCL_c} * 1000 \quad (17)$$

The final result of the computation is the

- Total area loss in hectare per hectare crop and year

- The total biomass loss in metric tons per hectare crop and year
- The corresponding CO₂ emissions in kg CO₂ equivalent per hectare crop per year

10.2. Limitations

Currently, the presented calculation framework is subject to the following limitations:

- Due to the absence of a complete²⁰ global LUC dataset, we equate LUC with the tree cover loss dataset from Global Forest Watch. This dataset allows the spatially explicit consideration of the most important LUC (i.e., the transformation from forest and potentially also from shrubland) but ignores other important land use changes such as the transformation from natural grassland. However, it's considered the most complete and ready-to-use dataset on LUC on a global scale. We will regularly review alternative data sources that would facilitate the refinement of our LUC computation.
- We use the *shared responsibility approach* (SRA) to perform the allocation of tree cover loss to all crops cultivated within a grid cell, mainly due to the current lack of spatially- and crop-specific time series data. The alternative *crop-specific approach* (CSA)—our initial preference—only allocates tree cover loss to the crops which increased their area. In this perspective, only the crops with increasing area would be allocated with tree cover loss. However, the CSA would require spatial data that facilitates the assessment of changes in the land use of each crop in a spatially explicit manner. Although MapSPAM²¹ offers a first time series of major commodities in a spatially explicit manner, it is not yet temporally complete and consistent with the fertilizer input data used in geoFootprint. However, we intend to refine the LUC allocation approach once this data becomes available.
- In addition, it is worth noting that the allocation of the SRA is not of a causal nature, i.e., it does not necessarily identify the actual driver of the tree-cover loss but allocated on the basis of spatial proximity. In fact, it only distributes the total tree cover loss

²⁰ Complete meaning that the dataset allows to track all relevant land use changes (e.g., also from grassland) and not only tree cover loss.

²¹ <https://www.mapspam.info/>

associated with a grid cell according to the relative area occupied by all crop occupying that grid cell.

- We do not consider any carbon accumulation for annual but also perennial crops (e.g., oil palm). While the carbon accumulated during plant growth is typically negligible for annual plants, it can be of relevance for perennial plants, which can remove carbon from the atmosphere for periods up to 25 years. However, assessing the average retention time of carbon in a plant and accounting for the associated benefit is related to high uncertainties because the permanence of the carbon accumulation is difficult to assess. We intend to consider the benefit associated with the carbon accumulation of perennial crops in the next version of geoFootprint by introducing time-specific carbon flows and corresponding characterization factors.
- The SRA allocates tree cover loss or LUC on the basis of the physical area occupied by a particular crop in relation to the physical area occupied by all other crops in a grid cell. That is, LUC is allocated to all crops that occupy area in a grid cell, even if they are (or have been) decreasing. This approach ignores tree cover losses induced by forestry, infrastructure (roads, dams, etc.) or subsistence farming due to the current lack of available spatial data. That is, performing the SRA exclusively with agricultural crop commodity data will overestimate the LUC allocated to the agricultural domain. In order to avoid unreasonable LUC allocation, we limit or cap the total tree cover area loss (TAL) that are physically impossible (or at least very unlikely) to a reasonable size, i.e., per annum we do not allow for more than one hectare of tree cover loss per ha of crop cultivated. We adapt the TAL based on a country presence in a country list²². For all countries in the country list, we correct all grid cells that have a $TAL > 0.01$ to 0.01. That is, we assume that 99% of the tree cover loss caused in each grid cell throughout these countries is caused by other drivers than crop cultivation. For the rest of the world countries not in the country list, we set the $TAL > 1 = 1$, only cap grid cells with a

²² ['Georgia', 'Slovenia', 'Slovakia', 'Latvia', 'Canada', 'United States of America', 'Russia', 'Australia', 'Mongolia', 'Sweden', 'New Zealand', 'Portugal', 'Finland', 'France', 'Spain', 'Japan', 'Norway', 'Austria', 'Romania', 'Ireland', 'Australia', 'Germany', 'Ukraine', 'Poland', 'Switzerland', 'Italy', 'Greece', 'Netherlands', 'Bosnia and Herz.', 'Turkey', 'Bulgaria', 'Serbia', 'Hungary', 'Montenegro', 'Croatia', 'Luxembourg', 'Czechia', 'Lithuania', 'United Kingdom', 'Estonia', 'Belgium', 'Latvia'].

TAL greater than one. As mentioned above, this is still quite a lot of LUC-related CO₂ per ha, which makes LUC in “rest of the world” countries often the most dominant source of climate change impacts. This approach is quite coarse and will be further refined using data from Curtis et al. (2018).

10.3. Outlook

The impacts from LUC are extremely relevant to the agricultural sector. However, due to the lack of data, we currently ignore other relevant drivers of tree cover loss such as forestry, infrastructure development and mining, and consequently overestimate LUC impacts associated with agricultural production.

A recent publication in *Nature* provides initial insight into the relevance of these drivers (Erb et al. 2018). It shows that cropland accounts for roughly a quarter of global tree cover loss, while forestry, grassland and infrastructure account for the remainder. More importantly, the publication provides some hints on how we can integrate additional drivers into geoFootprint. In a nutshell, we would require the following maps/spatial data to add additional drivers into the equation.

- Carbon stocks in *potential natural vegetation*, i.e., the potential (theoretical) natural vegetation (and its corresponding carbon stock) that would prevail in the absence of any human activity see e.g., <http://www.earthstat.org/carbon-stocks-potential-natural-vegetation/>
- Carbon stock in *actual vegetation*, i.e., the actual (measurable) natural vegetation and its corresponding carbon stock.
- Map that shows the *actual land use* on a global scale.

Combining these datasets would allow us to allocate the difference between the potential and the actual carbon stock to the actual land use. For example, if a grid cell in Russia shows a difference between potential and actual carbon stock and we know that the current land use is forest, the difference would be allocated to forestry. Vice versa, if we know that the land use is agriculture, the difference would be allocated to agriculture. This approach would probably also facilitate the integration of additional LUC changes. Alternatively, we can concentrate on tree cover loss and only allocated to the actual land use the actual tree cover loss derived from the GFW dataset. Somehow or other, the key ingredient is a map which shows actual land use on a global scale. It seems that such a dataset is available from a the recent study of Erb et al. 2018.

11. Regionalizing emissions from land use

Impacts from land use (LU) represent an important contribution to the environmental footprint associated with crop production. Bearing this in mind, we developed a dedicated module for geoFootprint to compute LU impacts based on the basis of the tree cover loss datasets of Global Forest Watch (GFW 2019).

The modeling of LU focuses mostly on changes of carbon in the soil, i.e., decrease or accumulation of soil organic carbon (SOC) due to the different management practices, and corresponding emissions in the form of carbon dioxide (CO_2), nitrous oxide (N_2O) and methane (CH_4). Management practices of relevance under consideration in the LU model include the effect of

- Land use change (LUC), effects on SOC change: tillage and input practice on SOC which can lead to CO_2 and N_2O emissions
- Carbon leaching due to soil erosion in the form of CO_2
- Residue management practices that regulate carbon (C) input from plant residues and C emission from decomposition, and therefore influence CH_4 and N_2O emissions
- Mineral and organic fertilizer application, which increases soil N availability and therefore also N_2O emissions from soils as a by-product of nitrification and denitrification
- The decomposition of dead organic matter and soil organic matter under flooded conditions (e.g., rice cultivation) in the form of CH_4
- Liming, which transfers C from the Earth's crust to the atmosphere by removing calcium carbonate from limestone and dolomite deposits

geoFootprint utilizes the emission models from the IPCC together with the specific refinements proposed by the Cool Farm Tool²³ to compute these changes (IPCC 2006; Kayatz et al. 2020).

The following sections outlines the computational framework to calculate LU impact on a global scale and its present limitations, and gives an outlook that highlights how the approach can be further refined.

11.1. Changes in SOC due to management

11.1.1. CO2 emissions from SOC

We use the standard IPCC 2006 equation and factors to account for changes in soil organic carbon (SOC) for a depth of 30 cm. However, we calculate two distinct delta SOC: one for the land use change from forest to arable land (equation 1) and one for the land use change from arable to arable land (equation 2). In equation 2, the land use factor is set to one so that its influence is neutral and only the management practice can influence SOC. Total SOC delta is then computed by adding the result of equation 1 and 2 (see equation 3). All of the computations are crop specific.

$$SOC_{delta_fromforest} = (SOC_{init} - (SOC_{init} * LUf * If * Tf)) * FFL \quad (1)$$

$$SOC_{delta_fromarable} = (SOC_{init} - (SOC_{init} * LUf * If * Tf)) * FAL \quad (2)$$

$$SOC_{delta_total} = SOC_{delta_fromforest} + SOC_{delta_fromarable} \quad (3)$$

Where:

- SOC_{delta_total} is the total difference between the initial and the new soil organic carbon stock resulting from the land use change from forest and from arable to arable land [metric ton C/ha].
- $SOC_{delta_fromforest}$ is the difference between the initial and the new soil organic carbon stock resulting from the land use change from forest to arable land [metric ton C/ha].

²³ <https://coolfarmtool.org/>

- $SOC_{\text{delta_fromarable}}$ is the difference between the initial and the new soil organic carbon stock for the land use change from arable to arable land [metric ton C/ha]. Due to the nature of this transformation (i.e., from arable to arable land) and the lack of more specific information on the type of arable land use, the LU_f is always set to one.
- SOC_{init} is the initial soil organic carbon stock taken from the ISRIC soil grids spatial data layer [metric ton C/ha].
- SOC_{new} (not shown in equation 1 & 2) is the new SOC stock induced by the land use (LU_f) and management practice (If & Tf) and is obtained by multiplying the initial SOC with the factor for land use and management practice [metric ton C/ha].
- LU_f is the dimensionless land use change factor and is set according to the crop type of the new land use (perennial crop, annual crop or rice) and, if the crop is an annual crop, according to the climate and moisture regime [-].
- If is the dimensionless input factor and can take values of “low” or “medium” [-]. Following the CFT methodology, we do not consider (the additional IPCC) values of “high” or “high, with manure” since these are supposed to be captured in the SOC changes due to the addition of organic matter. In line with the CFT, the If is thus exclusively used for the addition or removal of cover cropping and uses the low-to-medium or medium-to-low conversion factors of the IPCC method respectively. The default input factor is set to “medium” (i.e., 1 or no influence) for the default computation, but users can change this default value via the geoFootprint platform.
- Tf is the dimensionless tillage factor and can take values of either “full,” “reduced” or “none” [-]. The default tillage factor is set to “full” (i.e., 1 or no influence) for the default computation, but users can change this default value via the geoFootprint platform.
- FFL is the dimensionless fraction of forest land, i.e., the fraction of a hectare cultivated with a particular crop which was formerly forest land, i.e., covered by tree cover [-]. FFL is obtained from the land use change calculation module, i.e., it is equal to ANA_TAL in equation 14, in chapter 10.1.5.
- FAL is the dimensionless fraction of agricultural land, i.e., the hectare cultivated with a particular crop that was formerly already agricultural land [-]. FAL is obtained by computing $1 - FFL$ for each grid cell.

We compute annualized CO₂ emissions [kg CO₂/ha] with equation 4.

$$CO2_{SOC} = \frac{SOC_{delta_total} * 1000 * \frac{44}{12}}{y} \quad (4)$$

Where y is equal to the time period under observation. Since the Global Forest Watch dataset only covers a time period of 18 years, y is equal to 18.

We finally assign the loss/gain of carbon dioxide from/to the soil to the appropriate inventory flow, i.e., either “carbon dioxide, from soil management” in case of a carbon dioxide emission (CO₂_{SOC} is > 0) or “carbon dioxide, to soil stock” in case of a carbon dioxide accumulation (CO₂_{SOC} < 0). A gain in SOC stock is assessed with a characterization factor of -0.05, based on the recommendation of the C-sequestration guidelines (International Dairy Federation 2021a). (See section 11.3.1.)

11.1.2. N₂O emissions from SOC

We compute the N₂O emissions [kg/ha] caused by the corresponding denitrification of the soil with equation 5-7. Because the default factors for the C:N ratio are different for land transformed from forest and from arable land (15 or 10, respectively), the following equations (5-7) use SOC_{delta_fromforest} or SOC_{delta_fromarable} as a starting point.

$$N2O_{SOC_forest} = \frac{SOC_{delta_fromforest} * \frac{1}{15} * \frac{44}{28} * 0.01 * 1000}{y} \quad (5)$$

$$N2O_{SOC_arable} = \frac{SOC_{delta_fromarable} * \frac{1}{10} * \frac{44}{28} * 0.01 * 1000}{y} \quad (6)$$

$$N2O_{SOC_total} = N2O_{SOC_forest} + N2O_{SOC_arable} \quad (7)$$

Where 44/28 is the ratio of the relative molecular mass of N₂O to the relative atomic mass of nitrogen and y is the time period under observation, i.e., currently 18 years.

11.2. SOC leaching due to soil erosion

Soil erosion can cause a significant loss and redistribution of fertile soil. Because soil organic matter content is concentrated in the surface soil, accelerated erosion leads to its progressive

depletion. Erosion pattern and original SOC content are two main factors by which different soils can influence SOC loss (ZHANG et al. 2006; Li et al. 2016).

Bearing this in mind, we account for the corresponding decrease in SOC in the form of CO₂ emissions using equation 8 and 9. In the absence of detailed data, we assume that one half of the SOC in the eroded soil is lost in the form of CO₂, while the other half is presumably relocated. We will refine this soil loss factor of ½ once more detailed data becomes available.

$$SOC_{soil_erosion} = \frac{SOC_{init}}{BulkD*3000} * A * \frac{1}{2} \quad (8)$$

$$SOC_{SE_CO2} = SOC_{soil_erosion} * 1000 * \frac{44}{12} \quad (9)$$

We first divide the initial carbon stock [metric ton/ha] (SOC_{init} in the 30 cm of the top soil) by the total mass (bulk density, BulkD [kg/m³]) of the soil in this top soil layer (i.e., 3,000 m³/ha) to receive the SOC content [metric ton C/metric ton of soil]. By multiplying this with the annual soil erosion [metric ton/(ha * a)], A, and the loss factor (½) we receive the SOC lost via soil erosion [metric ton/(ha*a)]. Finally, we transform this into kg CO₂/(ha*a)] by multiplying with the C: CO₂ mol ratio and a factor of 1000. Since soil erosion is crop dependent, these equations are applied to each crop and grid cell.

The final result is the SOC loss (and relocated) due to soil erosion in the form of kg CO₂ per hectare. We inventory the loss of CO₂ due to erosion in the form of “CO₂, fossil,” i.e., as a permanent loss of SOC. We inventory the SOC relocated in the form of “CO₂, relocated,” a dedicated flow which is assessed with a characterization factor of 0.

We will refine the fraction of SOC loss used in equation 8, once more detailed information is identified or becomes available.

11.3. Residue management

11.3.1. SOC increases (CO₂ accumulation) due to addition of organic matter

In regard to the quantification of SOC increases due to the addition of organic matter, we follow the modeling of the CFT (Kayatz et al. 2020). That is, in order to differentiate between different levels of organic amendment—manures, straw and residue—returns (or benefits)

are not considered via the impact factor (If) (see section 11.1) but are modelled according to (Smith et al. 1997). The relationship between organic matter added and annual percentage change in SOC stocks is given by equation 10-13.

$$SOC_{\%Total} = SOC_{\%manure} + SOC_{\%straw} + SOC_{\%residue} \quad (10)$$

With:

$$SOC_{\%manure} = 0.036 * A_{manure} \quad (11)$$

$$SOC_{\%straw} = 0.142 * A_{straw} \quad (12)$$

$$SOC_{\%residue} = 0.142 * A_{residue} \quad (13)$$

Where:

- A_{manure} and A_{straw} are the dry weights of added farmyard manure and straw, respectively [metric ton/(ha*a)]. We retrieve the amount of manure from the regionalizing the N application in the form of manure retrieved from (Gerber et al. 2013) (see procedure in chapter 6.1). The dry weight of straw is set to zero, i.e., not considered, due to the lack of appropriate input data.
- $A_{residue}$ is the quantity of residue incorporated [metric ton/(ha*a)]. We compute default values for $A_{residue}$ in correlation to the yield according to the crop-specific default values / equation from the IPCC. (See Table 28.)

$$A_{residue} = Y * DMF * S * I \quad (14)$$

Where:

- Y represents the crop yield [metric ton/(ha*a)]
- DMF the dry matter fraction retrieved from Table 28 [-].
- S the slope retrieved from Table 28.
- I the intercept retrieved from Table 28.

The final result of the equation is the annualized relative change in SOC stocks due to the addition of residues. Multiplying this value with the SOC stock, we can derive the C accumulated in the SOC and, by multiplication with the stoichiometric factor of 44/12, the amount of CO₂ to soil stock [kg CO₂/(ha * a)]. Note that we add CO₂_{SOC}, if (CO₂_{SOC} < 0).

$$CO2_{ToSoilStock} = SOC_{\%Total} * SOC_{init} * 1000 * \frac{44}{12} + CO2_{SOC} * -1 \quad (15)$$

We characterize this flow with a characterization factor of -0.05 to account for the expected duration of the storage time (International Dairy Federation, 2021).

Table 28: IPCC default values for the computation of residues (by multiplying the fresh yield of a crop with the corresponding dry matter fraction (DMF) and the Slope + by adding the intercept value). AGN is the corresponding above-ground nitrogen and BGN the below-ground nitrogen in kg per metric ton of crop residue added annually. In the absence of crop-specific data (column 'Specific' = Other), we use the generic average computed and used by the CFT (IPCC, 2006; Kayatz, 2020). mt = metric ton. The DMF is not taken from the WFLDB Guidelines (Nemecek et al. 2019).

Label	ID	Unit	DMF	Specific	Slope	Intercept	AGN	BGN
Barley	1	mt	0.89	yes	0.98	0.59	0.007	0.014
Wheat	2	mt	0.85	yes	1.51	0.52	0.006	0.009
Sugar cane	3	mt	0.23	Other	0.665	0.7217	0.0152	0.0132
Palm fruit bunch	4	mt	0.53	Other	0.665	0.7217	0.0152	0.0132
Potato	5	mt	0.22	yes	0.1	1.06	0.019	0.014
Maize	6	mt	0.86	yes	1.03	0.61	0.006	0.007
Rice	7	mt	0.87	yes	0.95	2.46	0.007	0
Soybean	8	mt	0.89	yes	0.93	1.35	0.008	0.008
Peanut	9	mt	0.94	yes	1.07	1.54	0.016	0
Sunflower	10	mt	0.94	Other	0.665	0.7217	0.0152	0.0132
Sorghum	11	mt	0.89	yes	0.88	1.33	0.007	0.006
Rye	12	mt	0.89	yes	1.09	0.88	0.005	0.011
Cotton	13	mt	0.91	Other	0.665	0.7217	0.0152	0.0132
Rapeseed	14	mt	0.88	Other	0.665	0.7217	0.0152	0.0132
Sugar beet	15	mt	0.23	Other	1.07	1.54	0.019	0.014

11.3.2. Emission of N₂O and CH₄ due to crop residue management

The treatment of residues from crops (such as straw or leaf litter) can result in emissions of N₂O and CH₄. We follow the modeling of the CFT in order to estimate these emissions (Kayatz, 2020). Accordingly, emissions of N₂O and CH₄ are computed with equation 16 and 17.

$$CH4_{residue} = EF_{residue_{CH_4}} * A_{residue} \quad (16)$$

$$N2O_{residue} = EF_{residue_{N2O}} * A_{residue} \quad (17)$$

Where

- $A_{Residue}$ is the amount of residue dry matter [metric ton/(ha*a)] and
- $EF_{Residue}$ for each method of residue management is given in Table 9.

Table 29: Emission factor for common crop residue management methods (adapted from CFT documentation (Kayatz, 2020). Note the emission factor are given in kg N₂O or CH₄ per kg of dry mass residue, respectively.

Crop residue management method	EF_ CH ₄	EF_ N ₂ O
Removed, left untreated in heaps or pits	0.002611	1.70E-6
Removed; non-forced-aeration compost	0.0002000	1.70E-6
Removed; forced-aeration compost	0.00012	1.13E-6
Let on field; incorporated or mulch	0	0
Burned	0.00011	2.35 E-7
Exported off farm	0	0

11.3.3. N₂O emissions due to incorporated crop residue

We also compute the N₂O emissions resulting from incorporated crop residues. Following the CFT approach, for incorporated crop residues the N content is estimated using the IPCC default values. We first compute the total N added in the form of residues by multiplying the typical

N content of a crop-specific residue (above and below ground) with the total amount of residues (A_{residue}) [metric ton/(ha*a)]. Next, we apply the IPCC emission factor of 1% to compute N_2O emissions.

$$N_{\text{residue}} = N_{\text{residue_cont}} * A_{\text{residue}} \quad (18)$$

$$N_2O_{\text{residue}} = N_{\text{residue}} * Ef \quad (19)$$

Where N_{residue} is the annual amount of N added in the form of residues and Ef is the IPCC default emission factor of 0.01. For the incorporation of manure and straw, N_2O emissions are estimated with the Bouwman model (see section 11.3.2).

11.4. N_2O , NO_x & NH_3 emissions due to mineral and organic fertilizer application

We model the array of N_2O , NO_x & NH_3 emissions arising from the processes of nitrification and denitrification according to the Bouwman model as implemented by the Cool Farm Alliance (Kayatz, 2020). The model calculates N_2O emissions on the basis of an exponential regression model that considers various environmental condition and management parameters. (See Table 30 for an overview.) The detailed implementation of the model in the regionalization engine is described in the following section.

11.4.1. Input factors (default and user defined)

The Bouwman model uses various environmental condition and management parameters (input factors) to calculate N_2O emissions. We discuss the use of these input factors for the calculation of total N_2O emissions according to

- Background N_2O emissions
- Direct N_2O emissions from addition of N-containing fertilizers
- Indirect N_2O emissions arising from volatilization of NO and NH_3
- Indirect N_2O emissions arising from leaching of applied N into groundwater

Note that we follow the explicit distinction and consideration of background and foreground emissions introduced by the Cool Farm Tool (CFT) to discriminate between emissions that are

independent of the rate of N application but depended on the contextual conditions. These are the “background” emissions ($N_2O_{N_B}$; see section 11.4.2.1) and between fertilizer-induced emissions ($N_2O_{N_F}$; see section 11.4.2.2).

We use a similar functional form to also calculate NO_x and NH₃ emissions (see section 11.4.3 and 11.4.4, respectively). Table 30 gives an overview of the input factors affected/considered in the N₂O, NO_x and NH₃ computation and how we derive them.

Table 30: Input factors affecting (grey highlighted cells) and not affecting (white cells) N₂O, NO_x and NH₃ estimates in the Bouwman model (adapted from CFT technical documentation (Kayatz, 2020)).

Input factor	Notation	N ₂ O	NO _x	NH ₃	Comment
Fertilizer application rate					According to quantitative amount of fertilizer applied
Fertilizer type	Ff_{AVGW}				According to fertilizer types defined in the (default or user-defined) LCI inventory.
Crop type	Ctf				Mapping according to Table 32.
Soil texture	Stf				Derived from spatial data, see Table 33.
SOC	$SOCf$				Derived from spatial data, see Table 34.
Moisture factor (CEC)	Mf				Derived from spatial data, see Table 35.
Soil pH	pHf				Derived from spatial data, see Table 36.
Drainage	Df				Derived from spatial data, see Table 38.
Appl. method	Amf				According to options given in Table 37.

In the following subsections, we explain how we derive all of the input factors before we apply them in section 11.4.2.

11.4.1.1. Fertilizer type factor

The fertilizer type factor covers the input of both mineral and organic fertilizer. We derive the fertilizer type factor associated with a particular crop life cycle inventory by computing the weighted arithmetic mean of all fertilizer types i according to their application rate (Ff_{AVGW} - see equation 20). Note that we execute equation 20 separately for each emission type, i.e., once for N₂O, once for NO and once for NH₃.

$$Ff_{AVGW} = \sum_i Ff_i * Ar_i \quad (20)$$

We utilize the values given per fertilizer type *i* in the technical CFT documentation according to the mapping in Table 31.

*Table 31: Mapping of CFT fertilizer type factors to the fertilizer types available and used in geoFootprint. Fertilizers that do not appear in the inventories of geoFootprint are denoted with n.a. The * is approximated with values from diammonium phosphate. Source: adapted from CFT, technical documentation (Kayatz, 2020).*

Fertilizer	Fertilizer (geoFootprint naming)	N ₂ O factor	NO _x factor	NH ₃ factor
Ammonium Bicarbonate	n.a.			
Ammonium nitrate	Ammonium nitrate (AN), as N, at plant; market for ammonium nitrate, as N; market for nitrogen fertilizer, as N	0.0049	0.0051	-0.35
Ammonium sulphate	Market for ammonium sulphate, as N	0.0051	0.0056	0.429
Ammonium sulphate nitrate - 26%N	Market for ammonium sulphate, as N	0.0051	0.0056	0.429
Anhydrous ammonia	Ammonia (with 100% NH ₃), steam reforming process, at plant	0.0056	0.0051	-1.515
Calcium ammonium nitrate	Calcium ammonium nitrate (CAN), as N, at plant	0.0049	0.0051	-1.064
Calcium nitrate	n.a.			
Compound NPK 15%N 15% K ₂ O 15% P ₂ O ₅	NPK (15-15-15), as N, at plant	0.0039	0.0055	-1.585
Diammonium phosphate	Diammonium phosphate (DAP), as N, at plant	0.0039	0.0055	0.182
Monoammonium phosphate	Monoammonium phosphate (MAP), as N, at plant	0.0039	0.0055	-0.622
Urea	Urea, as N, at plant; market for urea, as N	0.0051	0.0061	0.666
Urea ammonia nitrate solution	Urea ammonium nitrate (AN), as N, at plant	0.0053	0.0004	0
n.a.	Single superphosphate production*	0.0039	0.0055	0.182
Manure	market for manure, solid, cattle; market for poultry manure, fresh; market for manure, liquid, cattle; market for manure, liquid, swine; market	0.0021	0.0016	0.995

The CFT models uses the reduction factors from the meta-analysis of Akiyama et al. (2010) in order to model the influence of nitrification inhibitors and polymer-coated fertilizer. Since the meta-analysis deals with total emissions and does not attempt to differentiate between background emissions and fertilizer induced emissions, the CFT has taken the conservative approach of applying emissions to the fertilizer-induced part only. As a consequence, the

mitigation potential of such options may be underestimated. It should be noted that these factors are only applicable to nitrogen (N) containing fertilizers.

11.4.1.2. Crop type factor

We derive the crop type factor from the crop under study. We map the crop in geoFootprint to the CFT crop types according to Table 32.

Table 32: Crop type factors used in geoFootprint. Source: adapted from CFT, technical documentation (Kayatz, 2020).

Crop type	Crops (geoFootprint)	N ₂ O factor	NH ₃ factor
Grass	Not applied	n.a.	n.a.
Grass-clover	Not applied	n.a.	n.a.
Rice	Rice	-2.536	0
Legumes	Soybean	-0.023	-0.045
Other upland	All other	0	-0.045

11.4.1.3. Soil texture factor

Soil texture is important for the formation of N₂O emission. The CFT distinguishes between coarse, medium and fine soil textures (see Table 33). Users of the CFT can select the appropriate category on the basis of the soil composition.

Table 33: Soil texture factors used in geoFootprint. The soil texture factor is only relevant for N₂O computation. Source: adapted from CFT, technical documentation (Kayatz, 2020).

Soil texture (CFT naming)	CFT definition (user defined input)	geoFootprint definition (derived from “coarse fragments” ISRIC soil grids layer)	N ₂ O factor
Fine	includes sandy, clay, silty clay, and clay	If coarse fragments value is < 10%	0
Medium	includes sandy clay loam, clay loam, and silty clay loam	If coarse fragments value is >= 10% but <= 30%	-0.472
Coarse	includes sand, loamy sand, loamy silt loam and silt.	If coarse fragments value is > 30%	-0.008

We use the spatially explicit “coarse fragments” layer from ISRIC Soil Grids (Hengl, 2014) to derive the appropriate soil texture factor automatically for each grid cell. In order to appropriately distinguish between the soil texture categories of the CFT, we use the ISRIC soil type classification system.

11.4.1.4. Soil organic carbon factor

The soil organic carbon content influence N₂O and NO emissions. The CFT distinguishes among 4 categories covering the range between smaller than 1%, larger than 1% and smaller or equal than 3%, larger than 3% and smaller or equal than 6% and greater 6% (see Table 34). Users of the CFT are advised to either select the appropriate category, or “if known exactly, you may enter the relevant [SOM] value” (...) directly. To transform from soil organic matter (SOM) to soil organic carbon they suggested to multiply the SOC value (in metric ton C / ha) with a generic factor of 1.72.

Table 34: Soil organic carbon factors used in geoFootprint. The SOC factor is only relevant for N₂O and NO computation. Source: adapted from CFT, technical documentation (Kayatz, 2020).

SOC categories (CFT)	geoFootprint definition (derived from “SOC” ISRIC soil grids layer)	N ₂ O factor	NO factor
<= 1%	If SOC value * 1.72 is < 10%	0	0
>1% -<= 3 %	If SOC value* 1.72 is >= 1% but < = 3%	0.14	0
>3 - <= 6%	If SOC value *1.72 is > 3% but <= 6%	0.58	2.571
>6%	If SOC value *1.72 is > 6%	1.045	2.571

11.4.1.5. Soil moisture factor

The soil moisture factor content influence N₂O and NH₃ emissions. The CFT distinguishes between “dry” and “moist” soils and defines “dry” soils as soils for which “water is limited (evaporation²⁴ exceeds rainfall) during a significant period of the growing seasons” and “moist” soils as “those without any significant water constraints (including irrigated soils)”. We use these definitions to derive the soil moisture factor from the evapotranspiration layer Mekonnen & Hoekstra (Mekonnen, 2011) and the precipitation layer from GAEZ (IIASSA/FAO, 2012). (See Table 35.)

Table 35: Soil moisture factors used in geoFootprint. The soil moisture factor is relevant for N₂O and NH₃ computation. Source: adapted from CFT, technical documentation (Kayatz, 2020).

SOC categories (CFT)	geoFootprint definition (derived from evapotranspiration and precipitation layer)	N ₂ O factor	NH ₃ factor
----------------------	---	-------------------------	------------------------

²⁴ It is not clear to which form of evaporation CFA is referring to. We use evapotranspiration as a proxy to evaporation but since the former is the sum of evaporation of water surface evaporation, soil moisture evaporation and plant transpiration, we probably overestimate the presence of dry soils with this approach.

Dry	If annual evapotranspiration is > precipitation	0.82	0
Moist	Else	0	-0.402

11.4.1.6. Soil pH factor

The soil pH factor influences N₂O and NH₃ emissions. The CFT distinguishes between 4 pH classes (see Table 36). We can directly use these distinctions to derive the corresponding N₂O and NH₃ factor from the ISRIC soil grids pH layer.

Table 36: Soil pH factors used in geoFootprint. The soil pH factor is only relevant for N₂O and NO computation. Source: adapted from CFT, technical documentation (Kayatz, 2020).

pH categories (CFT)	geoFootprint definition (derived from evapotranspiration and precipitation layer)	N ₂ O factor	NH ₃ factor
<= 5	If pH value is < =5.5	0	-1.702
>5 & <= 7.3	If pH value is >= 5.5 but < = 7.3	0.109	-0.933
>7.3 & <= 8.5	If pH value is >= 7.3 but < = 8.5	-0.352	-0.608
> 8.5	If pH value is 8.5	-0.352	0

11.4.1.7. Application method factor

The application method factor is only relevant for the calculation of NH₃ and derived from the fertilizer application method. We use the application methods indicated by the CFT (Kayatz, 2020) (See Table 37.)

Table 37: Application method factors used in geoFootprint for the calculation of NH₃ emission factor. Source: adapted from CFT, technical documentation (Kayatz, 2020).

Application method	NH ₃ factor
Apply in solution	-1.292
Broadcast	-1.305
Broadcast or incorporate then flood	-1.844
Broadcast to floodwater at panicle initiation	-2.465
Incorporate	-1.895
Surface drip	0

11.4.1.8. Soil drainage factor

The soil drainage factor content influence N₂O and NO emissions. The CFT distinguishes between “poor” and “good” soils and defines “poor” soils as “(typically) clay soils with restricted drainage”. All other cases correspond to “good”.

We use these definitions to derive the soil drainage factor from the ISRIC soil grids clay content (Hengl, 2014) and the harmonized world soil database drainage layer (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009). (See Table 38.)

Table 38: Soil drainage factors used in geoFootprint. The soil drainage factor is only relevant for N₂O and NO computation. Source: adapted from CFT, technical documentation (Kayatz, 2020).

SOC categories (CFT)	geoFootprint definition (derived from clay content and soil drainage factor layer)	N ₂ O factor	NO factor
Poor	If drainage value is > 3 and clay content is > 50%	0	0
Good	Else	-0.42	0.946

We have now shown how we derive all input factors on the basis of default (or user input data) and spatial data and can now proceed to the use of these input factors in the calculation of N₂O, NO and NH₃.

11.4.2. Calculation of direct and indirect N₂O emissions

11.4.2.1. Background N₂O emissions

Background N₂O emissions from soil [kg/(ha*a)] are computed with equation 21.

$$N2O_{N_B} = \exp(k + Ctf + Stf + pHf + Df + Mf + SOCf) * N2Of_{Inhibitor} \quad (21)$$

Where k represents the constant N₂O factor, Ctf the crop type factor, Stf the soil texture factor, pHf the pH factor, Df the drainage factor, Mf the soil moisture factor and $SOCf$ the soil organic carbon factor. $N2Of_{Inhibitor}$ represents the general N₂O inhibitor index which represents the N₂O inhabitation, i.e., the chemical compounds that slow the nitrification of ammonia, ammonium-containing, or urea-containing fertilizers. The index can take values of 0.63 (Nitrification inhibitor,) and 1 (Polymer coating, None).

11.4.2.2. Direct N₂O emissions from addition of N-containing fertilizers

Direct N₂O emissions [kg/(ha*a)], i.e., those directly associated with N-containing fertilizers, are computed with equation 22.

$$N2O_{N_F} = \left(\exp \left(k + Ctf + Stf + pHf + Df + Mf + SOCf + (Ff_{AVGW} * A_{Nfert}) \right) * N2Of_{Inhibitor} \right) - N2O_{N_B} \quad (22)$$

With Ff_{AVGW} as the weighted average of the applied fertilizer types and A_{Nfert} as the total amount of N applied.

11.4.2.3. Indirect N₂O emissions arising from volatilization of NO and NH₃

N₂O missions from of NO and NH₃ [kg/(ha*a)] are calculated using a IPCC default conversion factor of 0.01 (IPCC, 2006).

$$N2O_{N_{NO}} = 0.01 * NO_{Total} \quad (23)$$

$$N2O_{N_{NH3}} = 0.01 * NH3_{Total} \quad (24)$$

The calculation of NO_{Total} and NH₃_{Total} is shown in section 11.4.3 and 11.4.4, respectively.

11.4.2.4. Indirect N₂O emissions arising from leaching of applied N into groundwater

In the CFT, N₂O emissions arising from leaching are calculated if a soil is classified as “moist” by the user. The default leaching factors from the IPCC (IPCC, 2006) is used (30% leaching & and emission factor of 0.0075) to estimate the amount of N lost through this pathway and resulting as N₂O.

In geoFootprint, we rely on a dedicated model to calculate nitrate emissions from fertilizer application under consideration of contextual conditions, the SQCB nitrate model (Faist Emmenegger, 2009). Therefore, we compute the N₂O emissions associate with nitrate leaching [kg/(ha*a)] *not* on the basis of the default leaching factor of the IPCC (30% leaching), but instead directly use the quantitative amount of nitrate leaching ($A_{Nitrate}$) (see chapter 9.1) as an input (see equation 25).

$$N2O_{N_L} = \frac{A_{Nitrate}}{\frac{62}{14}} * 0.0075 * \frac{44}{28} \quad (25)$$

Where:

- $A_{Nitrate}$ represents the amount of nitrate emissions in kg per hectare obtained from the SQCB-model (see chapter 9.1)
- 62/14 represents the relative molar mass of nitrate to nitrogen
- 0.0075 represents the IPCC default emission factor
- 44/28 represents the ratio of the relative molecular mass of N₂O to the relative atomic mass of nitrogen

11.4.2.5. Total N₂O

We compute total N₂O emissions [kg/(ha*a)] by adding up all N₂O emissions (equation 26). See section 11.1.2 for the calculation of N₂O_SOC.

$$N2O_N_{Total} = N2O_N_F + N2O_N_B + N2O_N_L + N2O_N_{NO} + N2O_N_{NH3} + N2O_N_{SOC} + N2O_N_{Residue} \quad (26)$$

11.4.3. Calculation of NO

Emissions from nitrous oxides (NO) are modelled using the same functional form than the calculation of N₂O emissions (see equation 27-29). We also distinguish between NO emissions from the background (contextual conditions) and from fertilizer application.

$$NO_B = \exp(k + Df + SOCf) * NOf_{Inhibitor} \quad (27)$$

$$NO_F = \left(\exp \left(k + Df + SOCf + (Ff_{AVGW} * A_{Nfert}) \right) * NOf_{Inhibitor} \right) - NO_B \quad (28)$$

$$NO_{Total} = NO_B + NO_F \quad (29)$$

11.4.4. Calculation of NH₃

Emissions of ammonia (NH₃) [kg/(ha*a)] are modelled with a similar functional form than the calculation of N₂O emissions (see equation 27-29) but without a distinction between foreground and background emissions. In addition, A_{Nfert} is not treated as a factor class but appears outside of the exponent, while in the computation of N₂O and NO induced by fertilizer application, A_{Nfert} is a factor class inside the exponent (see equation 30).

$$NH3_N_{Total} = A_{Nfert} * \exp(Ctf + SMf + pHf + Df + Mf + AMf + Ff_{AVGW}) \quad (30)$$

Where AMf represents the application method factor.

11.5. CH₄ emissions from rice

The management of water saturation in rice cultivation has a large influence on the decomposition of organic matter and therefore methane (CH₄) emissions. We model CH₄

emissions from rice [kg/(ha*a)] according to the CFT (Kayatz, 2020), which is based on the IPCC (IPCC, 2006).

$$CH4_{rice} = CP * EF_c * SF_w * SF_p * SF_o \quad (31)$$

With

- CP represents the spatially explicit cultivation period in days retrieved from the GAEZ dataset.
- EF_c as 1.3, the daily methane emissions in kg for continuously flooded fields with no organic amendment. It is this factor which is scaled according to the management practice.
- SF_w is the scaling factor for practiced water regime for the cultivation period (see Table 39).
- SF_p is the scaling factor for practices water regime for the pre-season period (see Table 40).
- SF_o is the scaling factor for practiced organic amendment application (see Table 41).

Table 39: Water management methane scaling factors for rice cultivation (adapted from (Kayatz, 2020)).

Rice management practice	SF _w
Continuously flooded	1.00
Intermittently flooded	0.60
Intermittently flooded – multiple aeration	0.52
Rainfed	0.28
Draught prone	0.25
Deep water	0.31

Table 40: Pre-season water management methane scaling factors for rice cultivation (adapted from (Kayatz, 2020)).

Rice management practice	SF _p
Not flooded pre-season (within less than or equal to 180 days)	1.00
Not flooded pre-season (within more than 180 days)	0.68
Flooded pre-season	1.90

Table 41: Organic amendment methane scaling factors for rice cultivation (adapted from (Kayatz, 2020)).

Rice management practice	SF _w
Straw incorporated shortly before cultivation	1.00

Straw incorporated more than 30 days before cultivation	0.29
Compost	0.05
Farmyard manure	0.14
Green manure	0.50

11.6. CO₂ emissions from liming, urea and ammonium bicarbonate

The emissions due to liming²⁵, urea²⁶ and ammonium bicarbonate (ABC) are standard factor multiplication [kg/(ha*a)] and computed according to the CFT (Kayatz, 2020), which is based on the IPCC (IPCC, 2006). See equations 11.12 and 11.13, with equations 32, 33 and 34, respectively.

$$CO_{2CaCO_3} = [(A_{limestone} * EF_{limestone}) + (A_{dolomite} * EF_{dolomite})] * \frac{44}{12} \quad (32)$$

$$CO_{2urea} = (A_{urea} * EF_{urea}) * \frac{44}{12} \quad (33)$$

$$CO_{2ABC} = (A_{ABC} * EF_{ABC}) * \frac{44}{12} \quad (34)$$

Where:

- 44/12 represents the mol ratio between carbon dioxide and carbon
- EF_{limestone} represents the emission factor for limestone equal to 0.12 [kg of carbon/kg of limestone] (all calcium carbonate emissions are assumed to be due to limestone)

²⁵ Liming is used to reduce soil acidity and improve plant growth in managed systems, particularly agricultural lands and managed forests. Adding carbonates to soils in the form of lime (e.g., calcic limestone (CaCO₃), or dolomite (CaMg(CO₃)₂) leads to CO₂ emissions as the carbonate limes dissolve and release bicarbonate (2HCO₃⁻), which evolves into CO₂ and water (H₂O) (IPCC, 2006).

²⁶ Adding urea to soils during fertilisation leads to a loss of CO₂ that was fixed in the industrial production process. Urea (CO(NH₂)₂) is converted into ammonium (NH₄⁺), hydroxyl ion (OH⁻), and bicarbonate (HCO₃⁻), in the presence of water and urease enzymes. Similar to the soil reaction following addition of lime, bicarbonate that is formed evolves into CO₂ and water (IPCC, 2006).

- EF_{dolomite} represents the emission factor for dolomite equal to 0.20 [kg of carbon/kg of dolomite].
- EF_{urea} represents the emission factor for urea equal to 0.20 [kg of carbon/kg of urea].
- EF_{ABC} represents the emission factor for ABC equal to 0.65 [kg of carbon/kg of ABC].

We retrieve the amount of limestone, dolomite, urea and ABC from the default crop inventory [kg/(ha*a)], i.e., the crop-specific inventory dataset that has the highest spatial correlation with the respective grid cell under study.

11.7. Limitations and outlook

Due to the lack or limited availability of data, there are some noteworthy limitations associated with the LU module.

- Due to the lack of specific land use change information in the default computation, we use the pre-processed GFW data to determine the average proportion of a hectare cultivated with a particular crop in a particular grid cell that is transformed from forest land, i.e., namely all land area that was formerly covered with trees. While this is the land use change with the largest influence on SOC, we potentially miss the SOC impact of other important land use changes such as the one from grassland.
- The consideration of changes in SOC focus on mineral soils while the potential emissions associated with organic (i.e., particularly peat) soils is currently not considered. Emissions from peat can be very significant, particularly in the tropics. We will integrate spatially explicit peat emissions with the next version of geoFootprint.
- The increase in SOC due to organic amendments removes CO_2 from the atmosphere. In order to account for the temporal uncertainty of these removals, we use a temporally adapted characterization factor of -0.05 to account for these benefits (International Dairy Federation, 2021).

12. References

- Agrifarming (2020) Agrifarming. <https://www.agrifarming.in/sunflower-farming>
- AKIYAMA H, YAN X, YAGI K (2010) Evaluation of effectiveness of enhanced-efficiency fertilizers as mitigation options for N₂O and NO emissions from agricultural soils: meta-analysis. *Glob Change Biol* 16:1837–1846. <https://doi.org/10.1111/j.1365-2486.2009.02031.x>
- Borrelli P, Robinson DA, Fleischer LR, et al (2017) An assessment of the global impact of 21st century land use change on soil erosion. *Nat Commun* 8:2013. <https://doi.org/10.1038/s41467-017-02142-7>
- Boulay A-M, Bare J, Benini L, et al (2018) The WULCA consensus characterization model for water scarcity footprints: assessing impacts of water consumption based on available water remaining (AWARE). *Int J Life Cycle Assess* 23:368–378. <https://doi.org/10.1007/s11367-017-1333-8>
- Bouwman AF, Boumans LJM, Batjes NH (2002) Modeling global annual N₂O and NO emissions from fertilized fields. *Glob Biogeochem Cycles* 16:28–29. <https://doi.org/10.1029/2001GB001812>
- Bulle C, Margni M, Patouillard L, et al (2019) IMPACT World+: a globally regionalized life cycle impact assessment method. *Int J Life Cycle Assess* 24:1653–1674. <https://doi.org/10.1007/s11367-019-01583-0>
- Chaudhary A, Verones F, de Baan L, Hellweg S (2015) Quantifying Land Use Impacts on Biodiversity: Combining Species–Area Models and Vulnerability Indicators. *Environ Sci Technol* 49:9987–9995. <https://doi.org/10.1021/acs.est.5b02507>
- Curtis PG, Slay CM, Harris NL, et al (2018) Classifying drivers of global forest loss. *Science* 361:1108. <https://doi.org/10.1126/science.aau3445>
- De Willigen P (2000) An analysis of the calculation of leaching and denitrification losses as practised in the NUTMON approach. Report 18. Plant Research International, Wageningen, The Netherlands

EarthStat (2020) EarthStat

Erb K-H, Kastner T, Plutzer C, et al (2018) Unexpectedly large impact of forest management and grazing on global vegetation biomass. *Nature* 553:73–76. <https://doi.org/10.1038/nature25138>

EU Nitrogen Expert Panel (2016) Nitrogen Use Efficiency (NUE) – Guidance document for assessing NUE at farm level. Wageningen University

Faist Emmenegger M, Reinhard J, Zah R (2009) Sustainability Quick Check for Biofuels - intermediate background report. With contributions from T. Ziep, R. Weichbrodt, Prof. Dr. V. Wohlgemuth, FHTW Berlin and A. Roches, R. Feiermuth Knuchel, Dr. G. Gaillard, Agroscope Reckenholz Tänikon. EMPA, Dübendorf, Switzerland

FAO (2020) FAOSTAT Database. <http://www.fao.org/faostat/en/#data/QC>

FAO/IIASA/ISRIC/ISS-CAS/JRC (2009) Harmonized World Soil Database (version 1.1). FAO & IIASA, Rome, Italy & Laxenburg, Austria

Flisch R, Sinaj S, Charles R, Richner W (2009) GRUDAF 2009 - Grundlagen für die Düngung im Acker und Futterbau. *Agrarforschung* 16:1–97

Gerber JS, Carlson KM, Makowski D, et al (2016) Spatially explicit estimates of N₂O emissions from croplands suggest climate mitigation opportunities from improved fertilizer management. *Glob Change Biol* 22:3383–3394. <https://doi.org/10.1111/gcb.13341>

Gerber PJ, Steinfeld H, Henderson B, et al (2013) Tackling climate change through livestock: a global assessment of emissions and mitigation opportunities. Food and Agriculture Organization of the United Nations (FAO)

GFW (2019) Global Forest Watch: Forest monitoring designed for action. In: <http://www.globalforestwatch.org/map>. <http://www.globalforestwatch.org/map>

Hengl T, de Jesus JM, MacMillan RA, et al (2014) SoilGrids1km — Global Soil Information Based on Automated Mapping. *PLoS ONE* 9:e105992. <https://doi.org/10.1371/journal.pone.0105992>

- Hofste R, KUZMA S, Reig P, et al (2019) AQUEDUCT 3.0: UPDATED DECISION-RELEVANT GLOBAL WATER RISK INDICATORS. World Resources Institute
- IIASSA/FAO (2012) Global Agro-ecological Zones (GAEZ v3.0). IIASA, Laxenburg / FAO, Rome
- International Dairy Federation (2021a) C-Sequ: Development of LCA guidelines for the calculation of Carbon Sequestration in cattle production systems. Draft Guidelines for the Purposes of Public Consultation
- International Dairy Federation (2021b) C-Sequ: Development of LCA guidelines for the calculation of Carbon Sequestration in cattle production systems. Draft Guidelines for the Purposes of Public Consultation
- IPCC (2006) Guidelines for national greenhouse gas inventories. Volume 4: Agriculture, forestry and other land use. Kanagawa
- Kayatz B, van Tonder C, Hillier J, et al (2020) Cool Farm Tool Technical Documentation. Cool Farm Alliance, UK
- Li Z, Nie X, Chang X, et al (2016) Characteristics of Soil and Organic Carbon Loss Induced by Water Erosion on the Loess Plateau in China. PloS One 11:e0154591. <https://doi.org/10.1371/journal.pone.0154591>
- Mekonnen MM, Hoekstra AY (2011) The green, blue and grey water footprint of crops and derived crop products. Hydrol Earth Syst Sci Discuss 8:763–809. <https://doi.org/10.5194/hessd-8-763-2011>
- Monfreda C, Ramankutty N, Foley JA (2008) Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Glob Biogeochem Cycles 22:. <https://doi.org/10.1029/2007GB002947>.
- Mueller ND, Gerber JS, Johnston M, et al (2012) Closing yield gaps through nutrient and water management. Nature 490:254–257. <https://doi.org/10.1038/nature11420>
- Myers RJK (1980) The root system of a grain sorghum crop. Field Crops Res 3:53–64. [https://doi.org/10.1016/0378-4290\(80\)90007-6](https://doi.org/10.1016/0378-4290(80)90007-6)

- Nemecek T, Bengoa X, Lansche J, et al (2019) Methodological Guidelines for the Life Cycle Inventory of Agricultural Products. World Food LCA Database (WFLDB), Version 3.5
- Nemecek T, Schnetzer J, Reinhard J (2014) Updated and harmonised greenhouse gas emissions for crop inventories. *Int J Life Cycle Assess* 1–18. <https://doi.org/10.1007/s11367-014-0712-7>
- Prahsun V (2006) Erfassung der PO₄-Austräge für die Ökobilanzierung- SALCA-Phosphor. Agroscope FAL Reckenholz
- QGIS (2020) QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.org>
- Reinhard J, Zah R, Faist Emmenegger M, Hilty LM (2017a) Regionalized LCI modeling: The case of regionalized cotton datasets. In book: *EnviroInfo2017, From Science to Society: The Bridge provided by Environmental Informatics* Chapter: Regionalized LCI modeling: The case of regionalized cotton datasets Publisher: Shaker Verlag Editors: Benoit Otjacques, Patrick Hteltberger, Stefan Naumann, Volker Wohlgemuth. Springer Berlin Heidelberg
- Reinhard J, Zah R, Hilty LM (2017b) Regionalized LCI Modeling: A Framework for the Integration of Spatial Data in Life Cycle Assessment. In: Wohlgemuth V, Fuchs-Kittowski F, Wittmann J (eds) *Advances and New Trends in Environmental Informatics: Stability, Continuity, Innovation*. Springer International Publishing, Cham, pp 223–235
- Roy RN, Misra R V, Lesschen JP, Smaling EM (2003) Assessment of soil nutrient balance: Approaches and Methodologies. FAO, FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS, Rome
- Scherer L, Pfister S (2015) Modelling spatially explicit impacts from phosphorus emissions in agriculture. *Int J Life Cycle Assess* 20:785–795. <https://doi.org/10.1007/s11367-015-0880-0>
- Seppälä J, Posch M, Johansson M, Hettelingh J-P (2006) Country-dependent Characterisation Factors for Acidification and Terrestrial Eutrophication Based on Accumulated

- Exceedance as an Impact Category Indicator (14 pp). *Int J Life Cycle Assess* 11:403–416. <https://doi.org/10.1065/lca2005.06.215>
- Smith P, Powlson M, Glendining M, Smith J (1997) Potential for carbon sequestration in european soils: Preliminary estimates for five scenarios using results from long-term experiments. *Glob Change Biol* 3:67–79
- Stehfest E, Bouwman L (2006) N₂O and NO emission from agricultural fields and soils under natural vegetation: summarizing available measurement data and modeling of global annual emissions. *Nutr Cycl Agroecosystems* 74:207–228. <https://doi.org/10.1007/s10705-006-9000-7>
- Struijs J, Beusen A, van Jaarsveld H, Huijbregts MAJ (2009) Aquatic Eutrophication. Chapter 6 in: Goedkoop, M., Heijungs, R., Huijbregts, M.A.J., De Schryver, A., Struijs, J., Van Zelm, R. (2009). *ReCiPe 2008 A life cycle impact assessment method which comprises harmonised category indicators at the midpoint and the endpoint level*.
- UNEP/SETAC, Life Cycle Initiative (2016) *Global Guidance for Life Cycle Impact Assessment Indicators Volume 1*
- Weidema B, Bauer C, Hischer R, et al (2013) Overview and methodology. Data quality guideline for the ecoinvent database version 3
- Wischmeier WH, Smith DD (1978) Predicting rainfall erosion losses - a guide to erosion planning. U.S. Department of Agriculture, Agriculture Handbook No. 537
- WRI (2020) Aqueduct. <https://www.wri.org/aqueduct>
- YANG X; P WM; THORNTON, PE; JAIN, AK; (2014) Global Gridded Soil Phosphorus Distribution Maps at 0.5-degree Resolution
- ZHANG J, QUINE TA, NI S, GE F (2006) Stocks and dynamics of SOC in relation to soil redistribution by water and tillage erosion. *Glob Change Biol* 12:1834–1841. <https://doi.org/10.1111/j.1365-2486.2006.01206.x>

13. Appendix

13.1. Data inputs

Table 42 represent the geoFootprint input data classified according the following four categories:

- Global attributes: Input data that are not spatially explicit and represent one single value that is used throughout all computations (independent of the location). Some of the global attribute values are crop-specific, however the value of the attribute remains the same across the globe.
- Crop attributes: Input data that are both crop-specific and spatially explicit, i.e., the data value differs for each crop and each pixel. These data are key elements of the inventory computation as they enable us to identify in which locations a specific crop is grown, in which quantity and how much of the pixel area is covered by the crop.
- Inventory modeling data: Input data that are all crop-specific and defined at global level unless more precise data at country-level are available. These data mostly refer to all agricultural inputs such as fertilizer, irrigation and pest treatment products, as well as field operations.
- Contextual data: Input data that are spatially explicit but not crop-specific i.e., the data value differs for each pixel but is not related to a specific crop. These data typically provide information on soil properties (pH, soil organic carbon, etc.), which are used in several inventory models.

In addition to the division of the data according to the categories described above, input data are structured by the model in which they are used, i.e., which data are used for which model. In some cases, it can be noticed that a single model relies on a large amount of input data and requires data from different data categories.

For all data used and presented in Table 42, the following information are documented:

- Variable name: Name of the variable used as input for the geoFootprint computation engine.

- Institution: Institution hosting the data used. The data hosting institution is not necessarily the institution that developed the data as it sometimes only gathers several data arising from different data sources (e.g., EarthStat).
- Spatial resolution: The area covered by a single pixel of an image at the Earth surface (e.g., 10x10 km). This corresponds to the minimum possible distinctive area under which all information is aggregated to a single pixel value. The higher the resolution, the smaller the covered area per pixel, the more detailed the image.
- Spatial and temporal extent: Spatial and temporal range covered by the data. Respectively, this refers to the surface of the world covered by the data (e.g., entire globe, single country) and the timescale covered (e.g., several years, single year).
- Data type: Crop representativeness of the data. Some of the data are crop specific (i.e., they are specific to a particular crop) while other data are generic (i.e., crop-independent)
- Source: The text reference that refers to the data source (not to be confused with the hosting institution) and for which full reference details are available in the bibliography at the end of this report.

The highly heterogeneous input data undergo several processing steps in order to be exploitable in the geoFootprint computation engine. These steps, in particular, require harmonizing all input data so that they all share the same projection system and the same resolution (10 x 10 km), a mandatory precondition for any subsequent computation.

Table 42: geoFootprint input data database

	Variable name	Institution	Spatial resolution	Spatial and temporal extent	Data type	Source
GLOBAL ATTRIBUTES						
Land use practice						
	Residues management	Cool Farm Alliance	global default factor	global / n.a.	Generic	Kayatz et al. 2020
	Tillage practice	IPCC	global default factor	global / n.a.	Generic	IPCC, 2006
	Input practice	IPCC	global default factor	global / n.a.	Generic	IPCC, 2006
Soil erosion						
	Tillage factor	Agroscope	global default factor	global / n.a.	Generic	Faist Emmenegger et al. 2009
	Practice factor	Agroscope	global default factor	global / n.a.	Generic	Faist Emmenegger et al. 2009
Rice cultivation practices						
	Rice management practice	Cool Farm Alliance	global default factor	global / n.a.	Crop specific (only relevant for rice)	Kayatz et al. 2020
	Pre-season management practice	Cool Farm Alliance	global default factor	global / n.a.	Crop specific (only relevant for rice)	Kayatz et al. 2020
	Organic amendment practice	Cool Farm Alliance	global default factor	global / n.a.	Crop specific (only relevant for rice)	Kayatz et al. 2020
Bouwman model (N ₂ O)						
	Application method fertilizer	Cool Farm Alliance	global default factor	global / n.a.	Crop specific	Kayatz et al. 2020
	Nitrification inhibitor index	Cool Farm Alliance	global default factor	global / n.a.	Crop specific	Kayatz et al. 2020
Co-products allocation						
	Allocation to grain or seed	Quantis / ecoinvent	global default factor	global / n.a.	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
	Allocation to straw or stover	Quantis / ecoinvent	global default factor	global / n.a.	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
Land use change						
	Root-to-shoot ratio	IPCC	global default factor	global / n.a.	Generic - should be biome-specific	IPCC, 2006
	Dead organic matter in metric ton of C	IPCC	global default factor	global / n.a.	Generic	IPCC, 2006

Nitrate modeling						
	Rooting depth	Quantis	global default factor	global / n.a.	Crop specific	Faist Emmenegger et al. 2009
Soil erosion						
	Crop factor	Quantis	global default factor	global / n.a.	Crop specific	Faist Emmenegger et al. 2009
CROP ATTRIBUTES						
Crop information						
	Average yield	Earthstat	5 arc-minutes (~10x10 km)	global / 2000	Crop specific	Monfreda et al. 2008
	Production volume	Earthstat	5 arc-minutes (~10x10 km)	global / 2000	Crop specific	Monfreda et al. 2008
	Harvested crop area	Earthstat	5 arc-minutes (~10x10 km)	global / 2000	Crop specific	Monfreda et al. 2008
	Harvested crop area fraction	Earthstat	5 arc-minutes (~10x10 km)	global / 2000	Crop specific	Monfreda et al. 2008
	Cultivation period	GAEZ	5 arc-minutes (~10x10 km)	global / 2010	Crop specific	FAO/IIASA, 2010
INVENTORY MODELING						
Land Use and Land Use Change						
	Land Use Change					
	- annual tree cover loss in ha per year	GFW	5 arc-minutes (~10x10 km)	global / 2000-2018	generic	Global Forest Watch, 2014
	- annual (above-ground) biomass loss in metric tons of dry mass	GFW	5 arc-minutes (~10x10 km)	global / 2000-2018	generic	Global Forest Watch, 2014
	- LUC allocation factor	Earthstat / Quantis	5 arc-minutes (~10x10 km)	global / 2000	crop-specific	Monfreda et al. 2008
Irrigation						
	Water use					
	- Blue water	WFN	5 arc-minutes (~10x10 km)	global / 1996–2005	Crop specific	Mekonnen et al. 2011

	Technology	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
Mineral fertilizer						
	Nitrogen fertilizer	Earthstat	5 arc-minutes (~10x10 km)	global / 2000	Crop specific	Mueller et al. 2012
	Phosphorous fertilizer	Earthstat	5 arc-minutes (~10x10 km)	global / 2000	Crop specific	Mueller et al. 2012
	Potassium fertilizer	Earthstat	5 arc-minutes (~10x10 km)	global / 2000	Crop specific	Mueller et al. 2012
	Fertilizer types	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
Organic fertilizer						
	N applied in the form of manure	FAO (GLEAM 2.0)	5 arc-minutes (~10x10 km)	Countries covered in WFLDB / ecoinvent	Crop specific	FAO, 2013
	Manure types	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
Other macronutrients						
	Lime application rate	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
	Dolomite application rate	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
	Magnesium sulfate application rate	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
	Sulfur application rate	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
Pest treatment						
	Pesticide application rate	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
Field operations						
	Input practice	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
	Tillage practice	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
	Sowing or planting	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013

	Fertilizer application	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
	Fertilizer application practice	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
	Solid manure spreading	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
	Liquid manure spreading	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
	Pesticide application	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013
	Harvesting	Quantis / ecoinvent	country-level	Countries covered in WFLDB / ecoinvent	Crop specific	Nemecek et al. 2019 Weidema et al. 2013

CONTEXT						
Contextual information						
	Soil pH	ISRIC soil grids	5 arc-minutes (~10x10 km)	global / 2014	generic	Hengl et al. 2014
	Soil organic carbon	ISRIC soil grids	5 arc-minutes (~10x10 km)	global / 2014	generic	Hengl et al. 2014
	Clay content	ISRIC soil grids	5 arc-minutes (~10x10 km)	global / 2014	generic	Hengl et al. 2014
	Silt content	ISRIC soil grids	5 arc-minutes (~10x10 km)	global / 2014	generic	Hengl et al. 2014
	Sand content	ISRIC soil grids	5 arc-minutes (~10x10 km)	global / 2014	generic	Hengl et al. 2014
	Precipitation	GAEZ	5 arc-minutes (~10x10 km)	global / 2014	generic	FAO/IIASA, 2010
	Soil P content	NASA, Yang	5 arc-minutes (~10x10 km)	global / 2014	generic	Yang et al. 2014
	Temperature (min)	WorldClim	5 arc-minutes (~10x10 km)	global / 2014	generic	Fick et al. 2017
	Temperature (average)	GAEZ / WorldClim	5 arc-minutes (~10x10 km)	global / 2014	generic	FAO/IIASA, 2010
	Temperature (max)	WorldClim	5 arc-minutes (~10x10 km)	global / 2014	generic	Fick et al. 2017
	Soil erosion				generic	

	- Erosivity factor	JRC	1.25 degree (~25x25 km)	global / 2012	generic	Borrelli et al. 2017
	- Erodibility factor	JRC	1.25 degree (~25x25 km)	global / 2012	generic	Borrelli et al. 2017
	- Length slope factor	JRC	1.25 degree (~25x25 km)	global / 2012	generic	Borrelli et al. 2017

IPCC – Intergovernmental Panel on Climate Change. <https://www.ipcc.ch/>

Ecoinvent. <https://www.ecoinvent.org/>

Earthstat. <http://www.earthstat.org/>

GAEZ - Global Agro-Ecological Zones. <http://gaez.fao.org/Main.html#>

GFW – Global Forest Watch. <https://www.globalforestwatch.org/>

GLEAM – Global Livestock Environmental Assessment Model. <http://www.fao.org/gleam/en/>

WFN – Water Footprint Network. <https://waterfootprint.org/en/>

PIC - Potsdam Institute of Climate. <https://www.pik-potsdam.de/pik-frontpage>

FAO - Food and Agriculture Organization of the United Nations. <http://www.fao.org/home/en/>

ISRIC soil grids - International Soil Reference and Information Centre. <https://www.isric.org/>

NASA - National Aeronautics and Space Administration. <https://www.nasa.gov/>

WorldClim – Global Climate data. <https://www.worldclim.org/>

JRC – Joint Research Center. <https://ec.europa.eu/jrc/en>

13.2. Rasterization of regionalized footprints

We describe below the common rasterization process applied to all three regionalized footprints to retrieve the characterization factor values for each grid cell based on available characterization factor data. We apply the following steps to perform the rasterization highlighted in section 3.4.4:

1. For the AWARE and Chaudhary dataset, characterization factor and spatial zones come in two separate files: a spreadsheet listing the characterization factors for all spatial zones and a shapefile representing the geometric shape and location of the spatial zones. We first merge both files to distinctly link each characterization factor to its corresponding spatial zones, i.e., polygon.
2. To avoid any issue with further processing steps, we check the validity of all polygon geometries (*function: check validity*) and fix the geometries of the polygons when the validity criteria are not met (*function: fix geometries*).
3. We then overlay the default geoFootprint grid cell layer with the layer containing all polygons and adjust both layers so that they share the exact same projection system. We next join both layers based on the location and associate to each grid cell the value of the emission factor of the polygon in which the grid cell is falling into (*function: join attributes by location*) (Figure 4)
4. We finally give a characterization factor value to grid cells not falling in any polygons and grid cells falling in polygons with null value (i.e., NaN value) by retrieving the characterization factor value of the closest non-null grid cell (*function: Join attributes by Nearest*). In case several grid cells are equidistant from the considered grid cell, we then consider the average value of all selected closest grid cells (Figure 4). This correction only concerns less than 0.5% of all grid cells where production of at least 1 of the 15 crops is actually taking place.

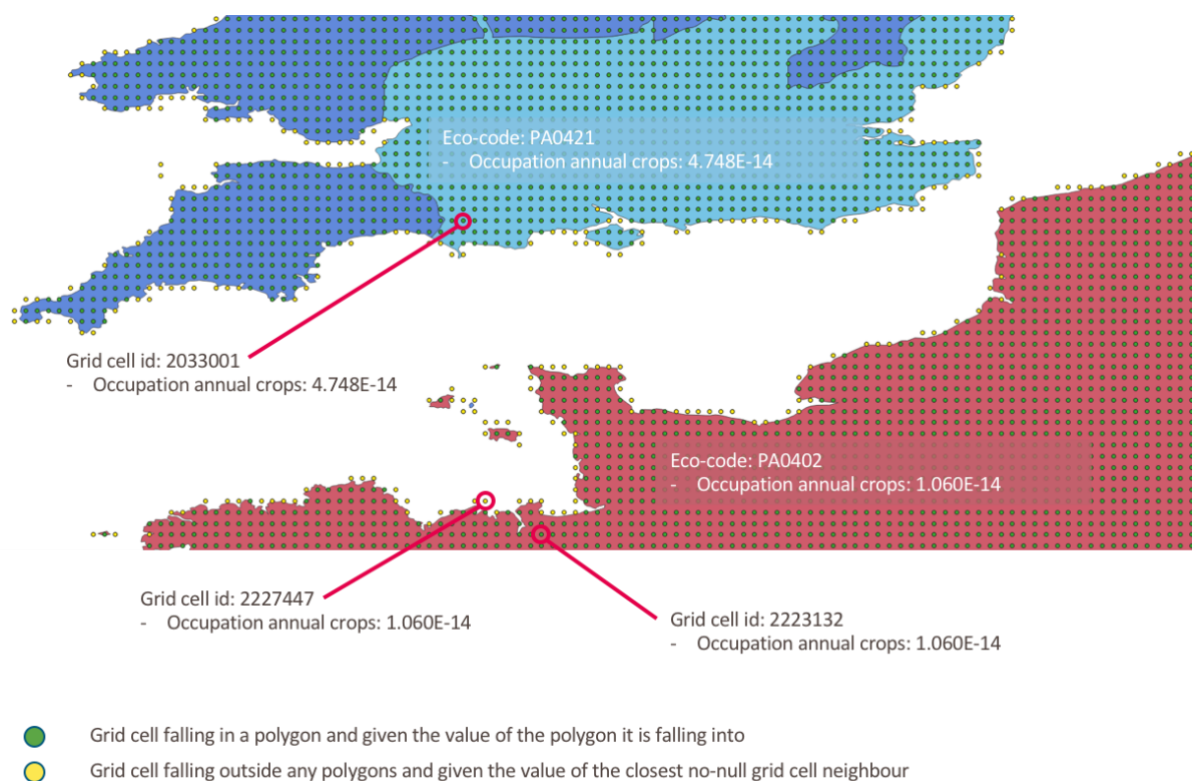


Figure 21: Example illustrating the rasterization process of spatially explicit LCIA methods to the geoFootprint grid cell structure. This example is for the LCIA method of biodiversity loss due to land use (Chaudhary et al. 2015) and illustrates only the characterization factor for occupation of annual crops (the transformation characterization factor is not displayed in this example) for the region of the Channel between France and the United Kingdom. Three different eco-regions are displayed in light blue, dark blue and red together with the geoFootprint grid cells. Green dots are grid cells given the emission factor value of the polygon in which they are falling into. Yellow dots are grid cells falling outside of any polygons and for which the characterization factor value of the closest non-null grid cell is given.