Regionalized LCI modeling: The Case of Regionalized Cotton Datasets

Jürgen Reinhard[1], Mireille Faist-Emmenegger[2], Rainer Zah[3], Lorenz M. Hilty[4]

1. Introduction

Companies in the fashion industry are increasingly looking for reliable data to make informed decisions and to prioritize their sustainability efforts. Life Cycle Assessment (LCA) provide a comprehensive and holistic way to assess environmental impacts over the full life cycle; yet, credible LCA data on the cultivation and processing of textiles is still limited. The World Apparel and Footwear Life Cycle Assessment Database (WALDB[5]) was founded to solve this data challenge and to deliver robust data for environmental impact assessment and footprinting. One key limitation concerns LCA data on cotton cultivation. To date, the entire cultivation of cotton is represented in the form of two data points, i.e., country-generic LCA datasets which uniformly describe cotton cultivation in the context of the USA and in China. In other words, there is a lack in the coverage and geographical representativeness of LCA data representing global cotton cultivation.

We believe that the integration of spatial data into LCA calculations can deliver more representative data for the assessment of cotton cultivation, when combined with a computerized method for regionalized LCI modeling. Regionalized LCI modeling is the procedure that generates and links process datasets to the location where they occur (Mutel et al. 2012). Spatial explicit data on various context conditions (precipitation, soil properties, etc.) and production parameters (crop-specific fertilizer input, yield, etc.) is now available, in decent resolution and on a global scale (Hengl et al. 2014; Monfreda et al. 2008; Mueller et al. 2012). However, the consideration of such data in the generation of LCA datasets is too labor-intensive with the classical means of data processing. Agricultural LCA datasets are mainly generated manually, according to specific guidelines[6] and emissions models and involving a wide array of raw data sources, ranging from public available databases (FAOSTAT, EUROSTAT, etc.), company data, surveys, case studies, publications, measurements, etc. (Nemecek et al. 2015). They typically are site-generic meaning that one datasets represent an entire country.

Reinhard et al. (2017) have developed a regionalization framework that is capable of processing the spatial explicit data on various context conditions and production parameters into comprehensive LCA datasets. This work-in-progress article examines the extensions of the framework for the generation of robust and geographically representative cotton datasets.

---

[1] University of Zurich, Informatics and Sustainability Research, Switzerland, juergen.reinhard@access.uzh.ch
[2] Quantis, LCA research and consulting, Switzerland, mireille.faist@quantis-intl.com
[3] Quantis, LCA research and consulting, Switzerland, rainer.zah@quantis-intl.com
[4] University of Zurich, Informatics and Sustainability Research, Switzerland, hilty@ifi.uzh.ch
[5] Quantis founded the WALDB together with a pre-competitive consortium of leading organizations and companies from the apparel and footwear sector.
[6] Such as the data quality guidelines from ecoinvent (Weidema et al. 2013), the Agri-footprint database (Blonk 2014) or the WFLDB (Nemecek et al. 2015).

We first review the core elements of the method for regionalized LCI modeling developed by (Reinhard et al. 2017) and show how it is used for the generation of regionalized cotton datasets. We next highlight geographically explicit results for cotton cultivation in China and in Turkey and show exemplarily how we compute country specific average datasets. We conclude with a discussion of the advancement provided by and improvement options of the framework.

2. Method

2.1 Framework for regionalized LCI modeling

We compiled a repository of publicly available raster data indicating harvested area, yield, fertilizer application rates, irrigation requirement of all major crops as well as data on precipitation, soil properties and terrain. We use this repository to produce comprehensive cultivation dataset (CD) on cotton cultivation. For each specific location or grid-cell, we translate the contextual information (such as production volume, input of mineral fertilizer, precipitation, soil organic carbon content, yield, etc.) into a regionalized CD in the nomenclature of the ecoinvent database (Figure 1). Such CDs typically comprise 40-80 exchange flows describing the type and amount of resources used (e.g. water, land), the intermediate flows required (e.g. the application of mineral and organic fertilizer or the use of field operations) and the accompanying release of emissions into soil, air and water (e.g. nitrate, di-nitrogen monoxide, phosphate emissions).

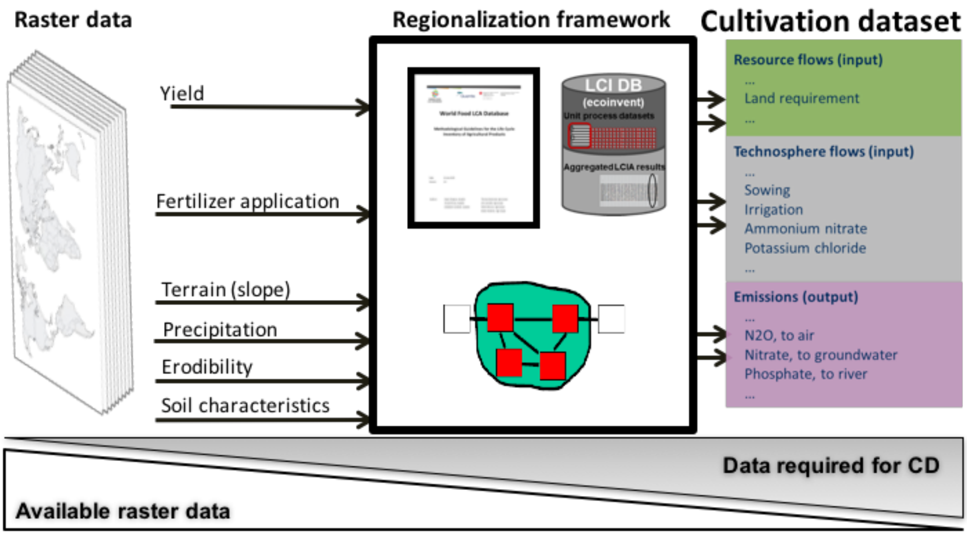

Figure 1: Regionalized LCI modeling. Bridging the gap between spatial data and cultivation datasets.

The processing of the spatial data into a comprehensive CDs is a non-trivial task. On the one hand, the spatial parameters cannot be used directly but require a high degree of manipulation until they represent one or several exchange flows of relevance in an agricultural CD. We developed functions for the harmonization of raster data resolutions (e.g. 30x30 arc second, 5x5 min, etc.) and for processing different spatial data formats (Geotiff, NetCDF, etc.). into Pandas data frames, a Python Data Analysis Library. We can merge different data frames on the basis of a shared index consisting of latitude-longitude combinations. Using this index as a basis, we can efficiently join data frames of different resolutions into one and the same data table.

In addition, because not all exchange flows can be computed on the basis of spatial data, the datasets need to be complemented with default data from background databases. We process and expand spatial data by using background data from the ecoinvent database (version 3.2). The integration of ecoinvent data allows the consideration of the entire background life-cycle (field operations, production of fertilizers, etc.). It also facilitates the generation of complete CDs, even when detailed data is missing or cannot be computed. A detailed description of the framework is given in Reinhard et al. (2017).

2.2 Generating and assessing regionalized cotton datasets

We generate and assess cotton datasets for a resolution of ~10 x 10 km grid scale. We focus on the 10 largest cotton producers, i.e., India, China, USA, Pakistan, Brazil, Uzbekistan, Australia, Turkey, Burkina Faso & Mali, which already cover 88% of the world's cotton production. We compute CDs for each grid-cell where cotton cultivation takes place. The final output is a geo-referenced CD table which indicates, for each relevant latitude-longitude combination, all exchange flows (e.g. input of natural resources; input of fertilizers, irrigation and machinery; and the output of emissions) associated with a particular CD. All exchange flows refer to the cultivation of one hectare of cotton in a cradle-to-gate perspective[7].

We can transform this geo-referenced CD table into a geo-referenced impact table for any LCIA indicator of interest. In this work-in-progress article, we assess environmental impacts for each CD in the inventory table only according to the midpoint indicator climate change (CC, IPCC2013 GWP100a). The result is a geo-referenced table that indicates GHG emissions on a per hectare basis for each ~10 x 10 km latitude-longitude combination, i.e., grid-cell.

2.3 Aggregating to country average CDs

To date, LCI database operate mainly with datasets on the global (site-generic) and/or the country level (site-dependent level). Therefore, the geo-referenced CD table in our resolution cannot be used directly in LCI databases but require an aggregation into representative averages. We aggregate the geo-referenced CD table into a production-volume weighted country average by first multiplying all exchange flows associated with a particular CD with the relative contribution to the overall production volume of the country—the production volume produced by each grid-cell also comes with the spatial explicit data on cotton cultivation (Monfreda et al. 2008). The aggregation to a weighted average is performed by summing up all exchange flows.

[7] That is, all upstream interventions are included. The further usage of the cotton (e.g. as feed and fiber) is not considered.

3. Results

3.1 Geographically differentiated GHG emissions

Figure 2 shows the spatially explicit environmental impacts per hectare cotton cultivated for CC for Turkey and for China.

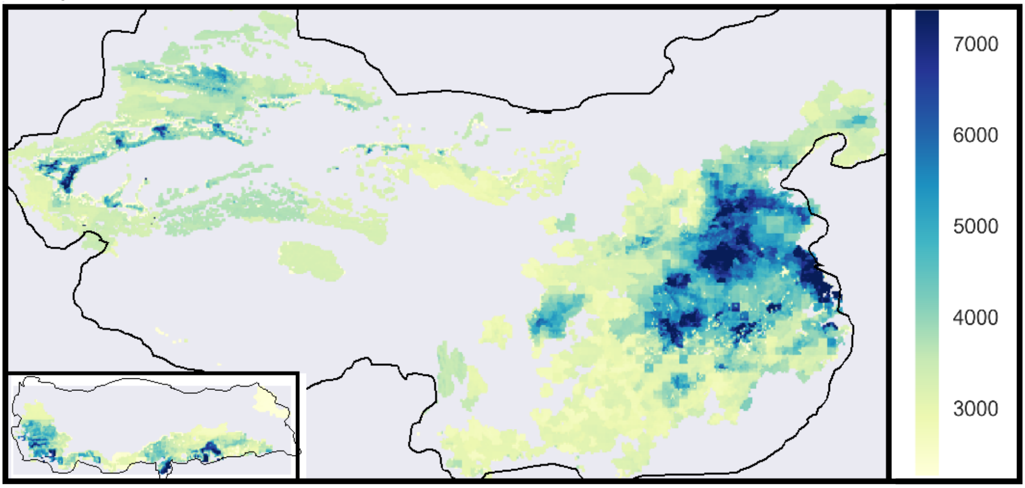

Figure 2: Heatmap of GHG emissions (in kg CO2 eq. per hectare) associated with cotton cultivation in Turkey (left corner) and China. The size of each grid-cell represents ~10 x 10 km.

The spatial distribution of GHG emissions correlates largely with the application intensity of N-based mineral fertilizer. GHG emissions are dominated by the energy intensive production of N-based mineral fertilizer and resulting N_2O emissions; both typically cause around 70% of the impacts.

3.2 Country averages CDs

Table 1 shows selected exchange flows of the production-volume weighted country average for Turkey.

Exchange flow	Unit	Turkey	
		Mean	CoV
Nitrogen fertilizer	kg N/ha	140.74	0.45
Potassium fertilizer	kg P_2O_5/ha	46.31	0.74

Potash fertilizer	kg K_2O/ha	24.30	0.74
N_2O emissions	kg N_2O/ha	3.03	0.39
Phosphorus emissions	kg P/ha	1.86	1.65
Nitrate emissions	kg NO_3-N/ha	125.53	0.32

Tab. 1. Production-volume weighted average of important exchange flows of the Turkish and the Chinese datasets and their variability measured as the coefficient of variation (CoV).

The coefficient of variation (CV) indicates a high spatial sensitivity. This confirms that the spatially explicit computation of these flows is important to obtain geographically representative CDs.

4. Discussion

The goal of this work-in-progress article was to show how the integration of spatial data into LCA calculations can improve the coverage and geographical representativeness of cotton cultivation. The case study confirms that regionalized LCI modeling matters. Emissions of high environmental relevance such as N_2O, nitrate and phosphorous also show great spatial variability. In this regard, our framework increases geographical representativeness of agricultural datasets by making feasible the consideration of spatial conditions which cannot be accounted for in site-dependent (country-generic) datasets. Furthermore, it can increase reproducibility by enforcing consistent use of assumptions and methods, a topic of particular importance with regard to emission modeling[8]. Finally, the framework can increase completeness because it allows for calculation and consideration of all relevant data points for a particular spatial scale of interest.

In this regard, our research will probably deliver the most comprehensive dataset on cotton cultivation produced so far; a set of roughly one million data points (CDs) will be aggregated into 10 country-weighted average cotton datasets representing 88% of the world's cotton production. This will enhance the current situation significantly (two data points representing roughly 38% of the world's cotton production) and provide much more reliable data basis to make informed decisions and to prioritize sustainability efforts in the apparel industry in the framework of the WALDB.

Furthermore, the case study shows that the regionalization framework can be used as a tool to improve cultivation dataset coverage and representation in LCI databases. It also offers novel possibilities for the aggregation and analysis of agricultural process data. Our aggregation was focusing on the generation of country averages, but could have been performed according to other political boundaries, such as a particular state or city district, or according to relatively homogeneous regions. However, what

[8] Due to inconsistent use of assumptions and methods, recent CDates in the ecoinvent database focused on the harmonization of the emissions modeling in agricultural datasets (Nemecek et al. 2014).

constitutes a homogeneous region, “is a matter for scientific inquiry“ (Mutel 2012) meaning that the ideal spatial scale of a CD is dependent on focused characteristics. For example, the spatial scale of a CD optimized for the reduction in the variability of nitrate emissions will be much smaller than the spatial scale of a CD optimized for the reduction of the variability in fertilizer application rate. Future research should therefore investigate multi-objective aggregation procedures that trade off exchange flow relevance (in terms of environmental impacts), variability, data quality and spatial proximity to build representative UDP from the body of data produced by our framework. This would increase the representativeness of agricultural datasets in LCI databases and improve the general utility of the framework for the domain of LCA. The performance of these approaches should be compared with the approach of spatial autocorrelation proposed by Mutel et al. (2012).

The framework is built upon spatial raster data, a rather new source of raw data in the domain of LCI modeling. The use of spatial raster data generates dependencies but also new opportunities. For example, spatially explicit data of crop production and fertilizer input is, to our knowledge, only available from EarthStat (Mueller et al. 2012). This means that the results of our regionalization framework are currently bound to the year 2000 and therefore subject to future-dependent CDates. Recent initiatives for open spatial data (Earth Observation Center 2017; FAO 2017; P. Panagos, Borelli, and Meusburger 2015; Panos Panagos et al. 2014) might diminish such dependencies in the long term. On the other hand, many of the spatial raster files in our repository come with a spatial-explicit rating of data quality (Mueller et al. 2012). This rating is not used in the current framework. Future work should therefore focus on the integration of such data quality ratings for the assessment of uncertainty.

5. References

Blonk, C.: Agri-Footprint - Part 2 - Description of data - Version 1.0., (2014)

Earth Observation Center: http://www.dlr.de/eoc/desktopdefault.aspx/tabid-5367/9013_read-16792/

FAO: Agro-Maps. Global Spatial Database of Agricultural Land-use Statistics, http://kids.fao.org/agromaps/

Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Gonzalez, M.R.: SoilGrids1km — Global Soil Information Based on Automated Mapping. PLoS ONE. 9, e105992 (2014). doi:10.1371/journal.pone.0105992

Monfreda, C., Ramankutty, N., Foley, J.A.: Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Glob. Biogeochem. Cycles. 22, (2008). doi:10.1029/2007GB002947.

Mueller, N.D., Gerber, J.S., Johnston, M., Ray, D.K., Ramankutty, N., Foley, J.A.: Closing yield gaps through nutrient and water management. Nature. 490, 254–257 (2012). doi:10.1038/nature11420

Mutel, C.: Framework and Tools for Regionlization in Life Cycle Assessment, https://e-collection.library.ethz.ch/view/eth:6340?q=(keywords_en:PRODUCT%20LIFE%20CYCLE), (2012)

Mutel, C.L., Pfister, S., Hellweg, S.: GIS-based regionalized life cycle assessment: how big is small enough? Methodology and case study of electricity generation. Environ. Sci. Technol. 46, 1096–1103 (2012). doi:10.1021/es203117z

Nemecek, T., Bengoa, X., Lansche, J., Mouron, P., Riedener, E., Rossi, V., Humbert, S.: Methodological Guidelines for the Life Cycle Inventory of Agricultural Products. World Food LCA Database (WFLDB), Version 3.0, http://www.quantis-intl.com/microsites/wfldb/files/WFLDB_MethodologicalGuidelines_v3.0.pdf, (2015) Nemecek, T., Schnetzer, J., Reinhard, J.: CDated and harmonised greenhouse gas emissions for crop inventories. Int. J. Life Cycle Assess. 1–18 (2014). doi:10.1007/s11367-014-0712-7

Panagos, P., Borelli, P., Meusburger, K.: A New European Slope Length and Steepness Factor (LS-Factor) for Modeling Soil Erosion by Water. Geosciences. 5, 117–126 (2015)

Panagos, P., Meusburger, K., Ballabio, C., Borrelli, P., Alewell, C.: Soil erodibility in Europe: A high-resolution dataset based on LUCAS. Sci. Total Environ. 479–480, 189–200 (2014). doi:10.1016/j.scitotenv.2014.02.010

Reinhard, J., Zah, R., Hilty, L.M.: Regionalized LCI Modeling: A Framework for the Integration of Spatial Data in Life Cycle Assessment. In: Wohlgemuth, V., Fuchs-Kittowski, F., and Wittmann, J. (eds.) Advances and New Trends in Environmental Informatics: Stability, Continuity, Innovation. pp. 223–235. Springer International Publishing, Cham (2017)

Weidema, B., Bauer, C., Hischier, R., Mutel, C., Nemecek, T., Reinhard, J., Vadenbo, C., Wernet, G.: Overview and methodology. Data quality guideline for the ecoinvent database version 3, http://www.ecoinvent.org/fileadmin/documents/en/Data_Quality_Guidelines/01_DataQualityGuideline_v3_Final.pdf, (2013)