Data quality control and data quality reporting in large scale background LCI databases: procedures, effects and challenges

LCA Discussion Forum 77, Zurich

21st April 2021

Guillaume Bourgault

Project manager

ecoinvent Association

Overview

- The problems of scale
  - Þ Curse/blessing of crowd-sourcing
  - Þ prioritization
- Strategies:
  - Þ Balances
  - Þ Cross-cutting
  - Þ Versions comparison

The problems of scale

- 17 471 datasets
    - Þ +350 000 exchanges
    - Þ +36 00 production volumes
    - Þ +3 000 prices
    - Þ +39 000 properties
- 51 LCIA methods
    - Þ 870 indicators
    - Þ +220 000 characterisation factors
- Not a task for a small team like ecoinvent!

The problems of scale

- A large user base, many are LCA experts
- ~14 versions over ~20 years -> many review cycles
- Crowd-sourced review from people with different:
    - Þ Background
    - Þ Biases
    - Þ Lens / interest
- But every year, we publish a new version, BEFORE the benefit of a crowd-sourced review

The problems of scale

- Sensitivity coefficient
  - Þ relative change of score, divided by relative change of value in the data point
  - Þ “if the sensitivity of a datapoint is 0.1 and I increase the value of the datapoint by 50%, the score will increase by 5%”
  - Þ 1 sensitivity coefficient per dataset per data point per indicator
  - Þ 17 000 datasets x 400 000 data points x 870 indicators = ~6x10^{12} coefficients
- Large proportion of data points are not sensitive at all
- A handful of data points are sensitive for most datasets and indicators
- Some data points are not sensitive for most datasets, but very sensitive for a handful of datasets
- Tension between the last two points

The problems of scale

- Ubiquitous sectors like electricity/heat, transport, mining oil&gas, receive a lot of attention
- A handful of “special interest” sectors, sometimes overseen by industrial associations
- We are left with “everything can be important”, depending who you ask

Strategies: dataset balances

- Exchanges have properties: carbon/water/metal content, and prices
- Possibility to balance datasets relative to those content
  - Þ Easy to automate
- Change exchange amounts or properties to correct imbalances
- For prices, we neglect many factors, but these cases raise a flag:
  - Þ value in > value out
  - Þ value in << value out

Strategies: LCI balances

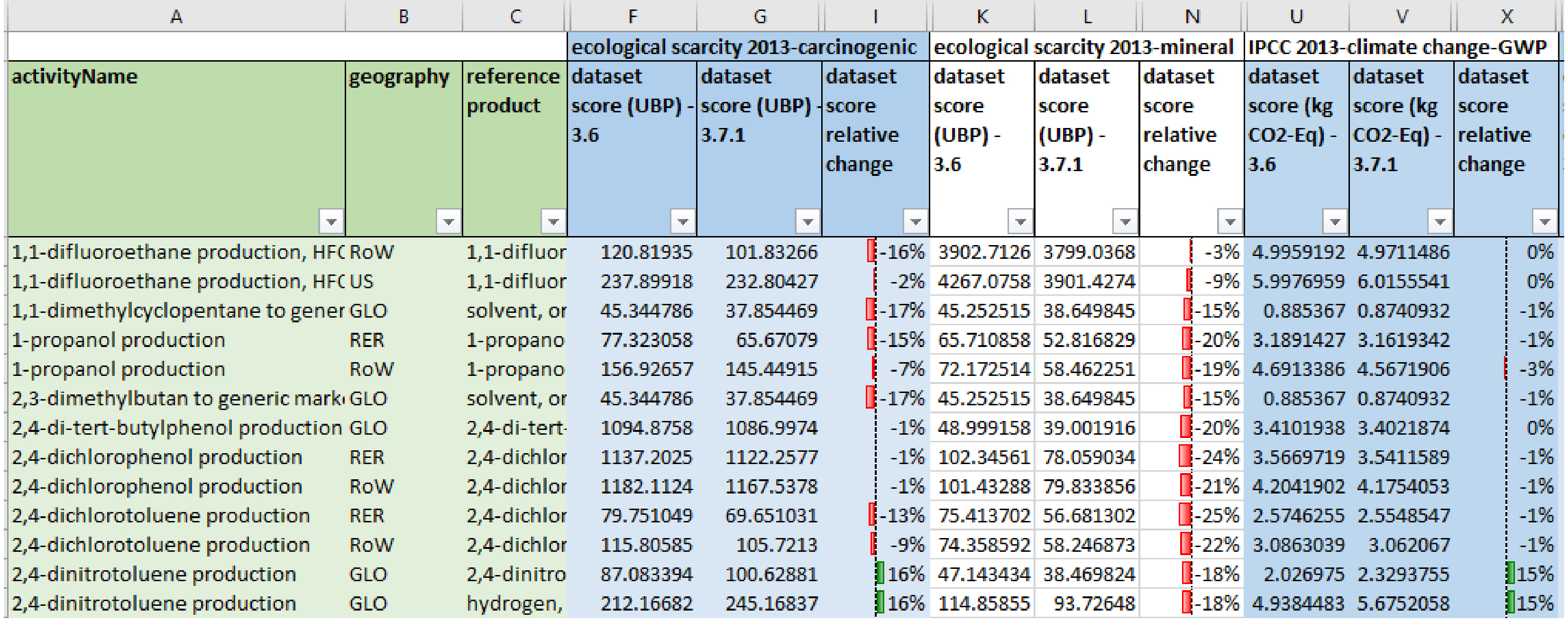
- In theory: from environment = to environment + reference product
- In practice, there are some limitations
    - Þ allocation/subdivision distort balances
    - Þ Recycling and waste treatment create sources/sinks of matter
- However, water is rarely a coproduct, less affected by that issue
- Balancing LCIs was used to correct water exchanges in datasets

Strategies: cross-cutting

- Identify a quantity present in many datasets
  - Þ Exchange amount, property, LCI or LCIA score…
- Check its distribution to spot outliers and raise flag
- Examples:
  - Þ Water/fertilizer consumption per kg of crop
  - Þ Loss due to transport in electricity markets
  - Þ CO2-eq per kg of the same metal for different production processes

Strategies: version n VS version n-1

- Because we start with a well-reviewed version, it makes sense to use it as a benchmark for a new version



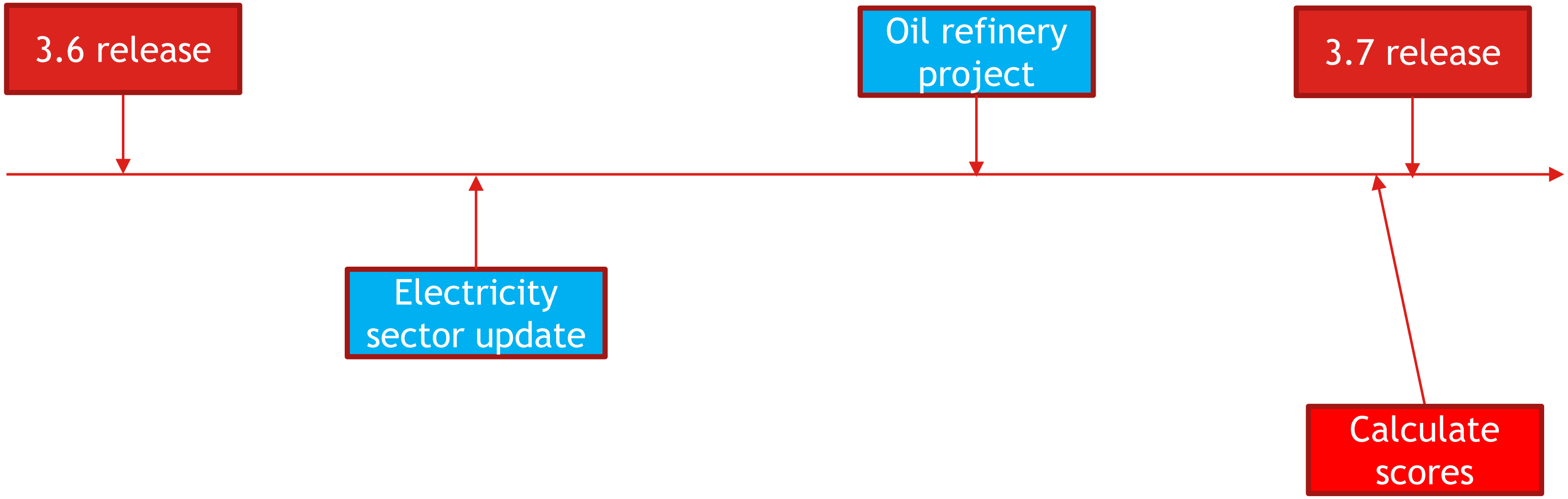
A	B	C	F	G	I	K	L	N	U	V	X
			ecological scarcity 2013-carcinogenic			ecological scarcity 2013-mineral			IPCC 2013-climate change-GWP		
activityName	geography	reference product	dataset score (UBP) - 3.6	dataset score (UBP) - 3.7.1	dataset score relative change	dataset score (UBP) - 3.6	dataset score (UBP) - 3.7.1	dataset score relative change	dataset score (kg CO2-Eq) - 3.6	dataset score (kg CO2-Eq) - 3.7.1	dataset score relative change
1,1-difluoroethane production, HFC	RoW	1,1-difluor	120.81935	101.83266	-16%	3902.7126	3799.0368	-3%	4.9959192	4.9711486	0%
1,1-difluoroethane production, HFC	US	1,1-difluor	237.89918	232.80427	-2%	4267.0758	3901.4274	-9%	5.9976959	6.0155541	0%
1,1-dimethylcyclopentane to gener	GLO	solvent, or	45.344786	37.854469	-17%	45.252515	38.649845	-15%	0.885367	0.8740932	-1%
1-propanol production	RER	1-propano	77.323058	65.67079	-15%	65.710858	52.816829	-20%	3.1891427	3.1619342	-1%
1-propanol production	RoW	1-propano	156.92657	145.44915	-7%	72.172514	58.462251	-19%	4.6913386	4.5671906	-3%
2,3-dimethylbutan to generic mark	GLO	solvent, or	45.344786	37.854469	-17%	45.252515	38.649845	-15%	0.885367	0.8740932	-1%
2,4-di-tert-butylphenol production	GLO	2,4-di-tert-	1094.8758	1086.9974	-1%	48.999158	39.001916	-20%	3.4101938	3.4021874	0%
2,4-dichlorophenol production	RER	2,4-dichlor	1137.2025	1122.2577	-1%	102.34561	78.059034	-24%	3.5669719	3.5411589	-1%
2,4-dichlorophenol production	RoW	2,4-dichlor	1182.1124	1167.5378	-1%	101.43288	79.833856	-21%	4.2041902	4.1754053	-1%
2,4-dichlorotoluene production	RER	2,4-dichlor	79.751049	69.651031	-13%	75.413702	56.681302	-25%	2.5746255	2.5548547	-1%
2,4-dichlorotoluene production	RoW	2,4-dichlor	115.80585	105.7213	-9%	74.358592	58.246873	-22%	3.0863039	3.062067	-1%
2,4-dinitrotoluene production	GLO	2,4-dinitro	87.083394	100.62881	16%	47.143434	38.469824	-18%	2.026975	2.3293755	15%
2,4-dinitrotoluene production	GLO	hydrogen,	212.16682	245.16837	16%	114.85855	93.72648	-18%	4.9384483	5.6752058	15%

Strategies: version n VS version n-1

- Where does the change come from?
  - Change inside the dataset?
  - Change in its supply chain?
  - Change in linking rules?
- For the digging: “Direct contribution comparison”
  - It would be information overload to show it here
  - Shows delta of exchange amount, score per unit of supply chain or CF, and how much of this change is responsible for the score change of the whole dataset
  - Allows quick identification of the source of a score change

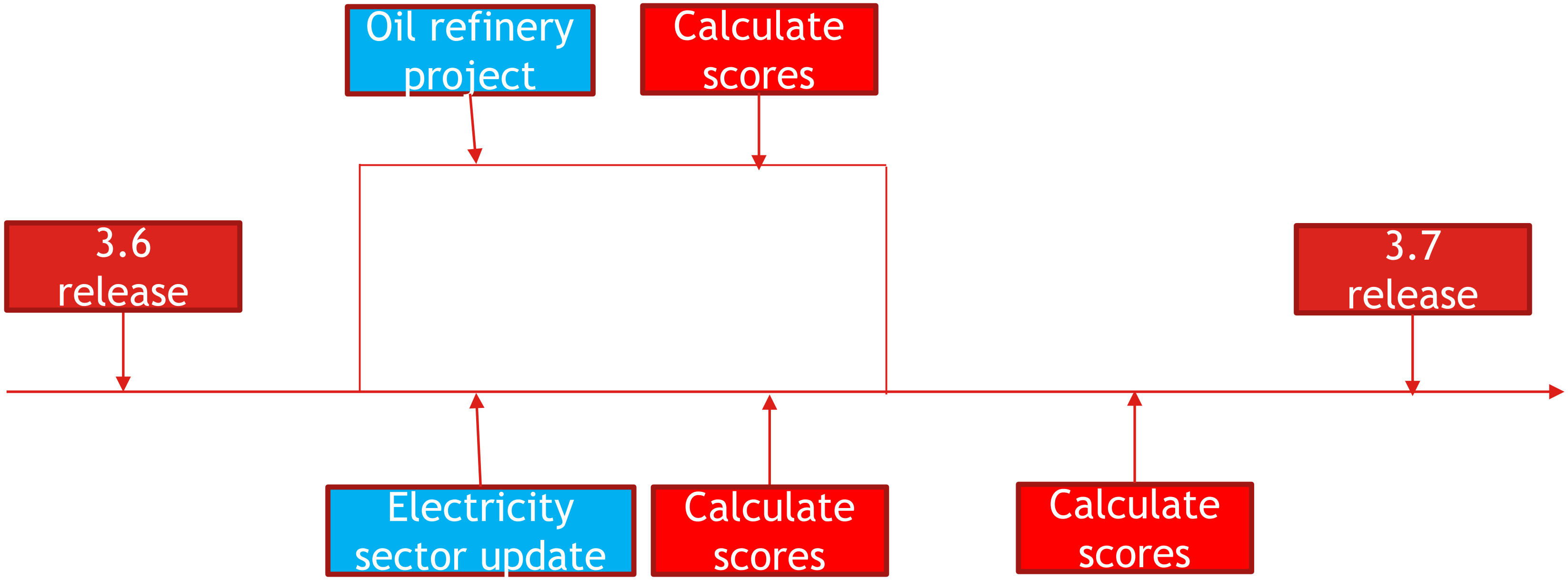
Strategies: version n VS version n-1

- What we used to do: calculate the scores after the completion of many projects



Strategies: version n VS version n-1

- What we do now:



Conclusion

- “Trust in transparency” is ecoinvent’s motto
    - Þ It unlocks the power of crowd-sourced review
- Prioritization is key
- Large volume of data can help to spot mistakes

Question?

Guillaume Bourgault

Project manager

bourgault@ecoinvent.org

www.ecoinvent.org