

Overcoming inconsistencies in sub-national mining data

The case of copper and silver mining in Peru

Authors: Sebastian Luckeneder, Mirko Lieber & Stefan Giljum
Institute for Ecological Economics, Vienna University of Economics and Business (WU)

FINEPRINT Brief No. 4, January 2019

Sub-national mining data from different sources is not fully consistent. In this Brief, we analyse these inconsistencies using example data sets for the case of Peru by SNL Metals & Mining and the Peruvian Ministerio de Energía y Minas. We illustrate the nature of these inconsistencies and introduce measures how to overcome them.

The FINEPRINT objective of linking extractivist economic practice with its environmental and social impacts relies heavily on the availability of data on world-wide spatiotemporal raw material extraction. With regard to mapping mining activities, we have elaborated a methodological approach (see [Brief No. 2](#)), which allows us to obtain a consistent data foundation by integrating information from various sources.

When compiling sets of sub-national extraction accounts, inconsistencies between various data sources are a common issue. In the case of mining, data from private sources, such as those from the Global Metals & Mining database provided by S&P Global Market Intelligence (SNL Metals & Mining), do not necessarily accord with publicly available data, such as those from official national statistics. Common issues are the availability of different variables, nonmatching numbers of observations as well as ambiguous results on the amounts of extraction and metal ore grades.

To promote a deeper understanding of why it is crucial to exploit the information provided by multiple data sources and to reveal their weaknesses and strengths, this Brief gives insights on a case study performed for Peru and two of the country's major commodities, copper and silver. Based on our analysis, we develop generalised approaches for overcoming observed inconsistencies. These will be tested in future FINEPRINT work.

The Peruvian case

The mining industry is one of Peru's key economic sectors. In 2013, Peru was placed 7th in terms of global metal ore extraction volumes, being one of the world's largest producers of copper, gold, silver, tin, zinc, and lead [1]. This analysis selects two commodities for further investigation. On the one hand, focus is set on copper, as copper ores and concentrates accounted for a quarter of Peru's total monetary export value in 2017 [2]. On the other hand, Peru is one of the world's hotspots regarding silver production.

In the following, we summarise and compare data on material extraction reported by two different and independent sources. First, from the Peruvian *Ministerio de Energía y Minas* (MINEM), which provides extensive data on mining covering monthly and annual volumes of production as well as ore grades of every registered mine since 2009 [3]. Second, we evaluate all available data on Peru contained in the SNL *Metals & Mining database* [4].

National aggregates and coverage of mines

In a first step, we aggregate spatial information on production (i.e. the production of each mine listed by SNL and MINEM, respectively) to a national level and compare these aggregates to the data on net primary production included in the Global Material Flows Database of the UN Environment International Resource Panel [1] (Figure 1).

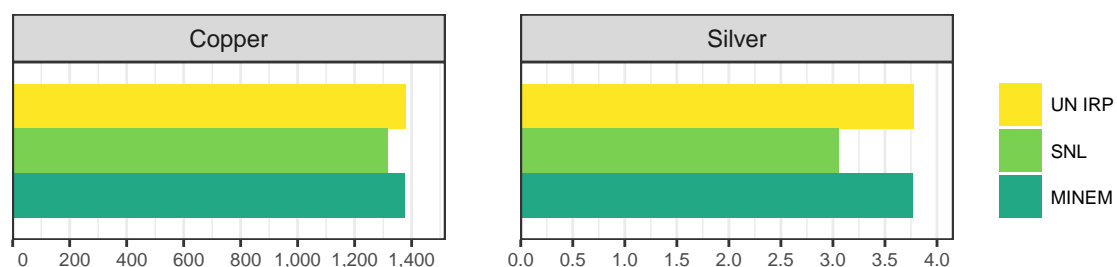


Figure 1: Aggregate production of copper and silver, Peru 2014 (in kilotonnes)

While the reporting by MINEM and data from the UN almost match, reporting by SNL shows lower numbers, i.e. about 95% of the amount that is stated by either of the other sources for the case of copper and 80% for silver.

In a next step, we analyse how many mining properties are covered by the respective data sets. Restricting the data to all mines that are related to copper and/or silver, SNL lists 49 active mines for the year 2014, 43 for 2015 and 45 for 2016 that report production numbers. In contrast, MINEM publishes annual Excel files, counting 189 mining properties with production values in 2014, 182 in 2015 and 169 in 2016. Hence, although MINEM reports more than three times as many mines as SNL, national production aggregates match well. SNL thus covers the largest mines in terms of production, whereas MINEM additionally includes many smaller mines. This is illustrated in Figure 2, which shows the cumulative production of the largest 10 copper mines in 2014 as well as the contribution of all other mines to the total national aggregate (calculated on the basis of MINEM data). The figure also visualises that the largest ten mines make up 91% of overall production, while



the 179 smaller mining sites additionally reported by MINEM contribute only 9%. For silver mining, we find a similar picture, but with the largest 10 mines' contribution to the total being only 50%.

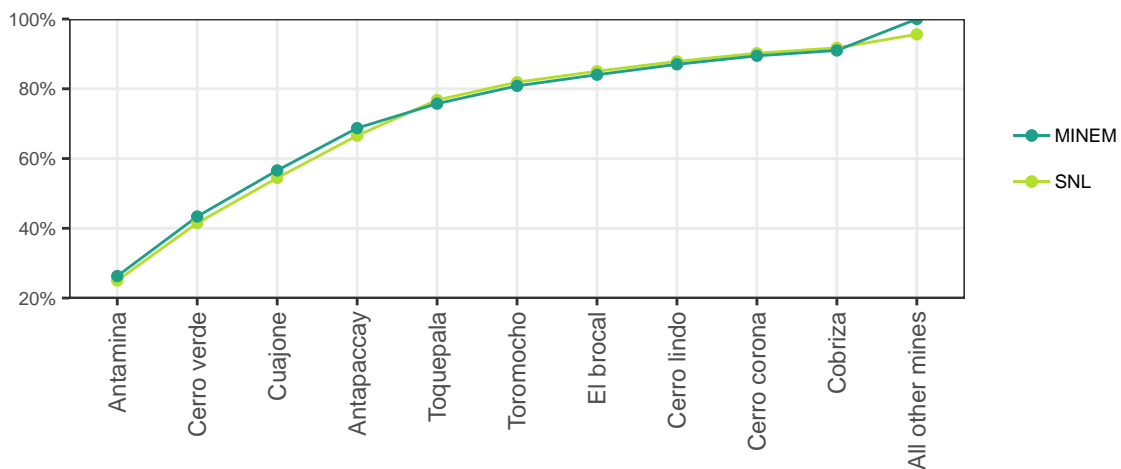


Figure 2: Cumulative Peruvian copper production (in % of MINEM total)

Identifying consistent combinations of data sets

When integrating two data sets, we eventually end up with three types of mines: mines that are only covered by data source A, those only contained in B, and mining properties covered by both. When aiming at creating a complete dataset on mining activities on the national or global level, we need to filter information for the set of intersecting mines and at the same time include all mines that are uniquely reported by only one of the available sources.

In order to obtain the intersecting fraction, we integrate both data sets into one by matching the mines' names. Matching names is vulnerable to double counting because of spelling issues or non-uniform naming. We therefore match mines by the exact location whenever geographical information is available. The intersection between SNL's data and the national data set is surprisingly small in our case study, especially regarding silver mining. For the year 2014, only 30 out of the 46 silver mines listed by SNL and the 161 silver mines reported by MINEM match by names. The overlapping data reduces to 151 observations of 38 mines either producing copper, or silver, or both for the years 2014 to 2016. This highlights the importance of integrating several sources, in order to ensure maximum coverage of the final data set.

Two specific issues need to be addressed in the integration process. First, it is possible that mines are missing in either of the data sets just for certain years. In this case we keep the uniquely available observation. Second, multiple production values can be reported for identical mining properties. If the data sources report the same amounts of production, we select the publicly available data. If they report different amounts of production, we conduct the following assessments: (1) If significant differences occur, we investigate whether e.g. uneven product types or different ore compounds or metal elements being reported can be the reasons for the observed differences. (2) We check whether data points from one source can be valued as more reliable, e.g. reporting specific values for each year versus identical values for several years. Also, we investigate the primary source of SNL's data in order to possibly rule out either of the accounts. (3) Marking certain data points as "in



question” and keeping values from both sources will allow us to check the data against national aggregates at a later stage (to be examined in a second part of this case study).

A useful combination of datasets also implies that spatially explicit observations are delivered. Data by MINEM does not report the exact location of mines (i.e. coordinates), but we can use the SNL Metals & Mining database to complement this information.

Comparisons at the individual mine level

Ambiguous reporting of production values is a common issue. To investigate this in more detail on the level of individual mines, we narrow down the sample of mines to the intersection between the SNL and MINEM data matched by mine names. Figure 3 plots SNL’s production reporting on the x-axis versus the Peruvian ministry’s data on the y-axis (note the log-transformation). The black line inside both panels indicates hypothetical equal reporting. Due to the properties of the natural logarithm, we can linearly display deviations in percent. All observations within the inner dashed lines show cases where reported numbers do not deviate by more than 20 percent, the outer dashes indicate deviations by 50 percent. It is evident that, on the one hand, there is a general equivalence for many mines regarding their reported production. On the other hand, there are some observations with severely contradicting reporting. Furthermore, we can see that none of the mines with severe deviations belongs to the group of mining sites with very large production values.

MINEM tends to report slightly higher production values for many cases. However, highest percentage deviations occur towards the opposite direction, i.e S&P reporting significantly higher production values such as for the case of the Morococha mine (regarding both copper and silver production). For a further analysis, it is indispensable to investigate these severe mismatches by referring to additional points of reference. Pan American Silver, the majority owner of Morococha mine, for example, publishes annual reports [5]. These match SNL’s numbers accordingly to the 92.3% share of the mining company’s ownership, suggesting that SNL draws its information from the official company reports. No additional source, however, was found to confirm MINEM’s significantly lower numbers, which is why we keep SNL’s observation in this case and drop the reporting of MINEM.

Conclusions

In this Brief, we have illustrated how sub-national mining data can vary by source and thus raises the question how to deal with these inconsistencies. Regarding the case study on copper and silver mining in Peru, we find that although SNL reports significantly fewer mines with actual production than the Peruvian ministry, national aggregates of SNL’s data sum up to 80% and 95% for silver and copper production, respectively. We clearly see the opportunity to complement multiple data sources by combining the intersecting mines, which are listed in either of the sources’ data, with their symmetric difference. Whenever multiple sources report production values for the same mine, we either use the official data (if the numbers match) or, if deviations are detected, we conduct further investigations to ensure that the most appropriate data is chosen for the final FINEPRINT data on world-wide spatially explicit raw material extraction. These include a background check on the type of product, data reliability as well as an evaluation against national aggregates. The latter is going to be introduced as a subsequent section of this case study in an upcoming Brief.

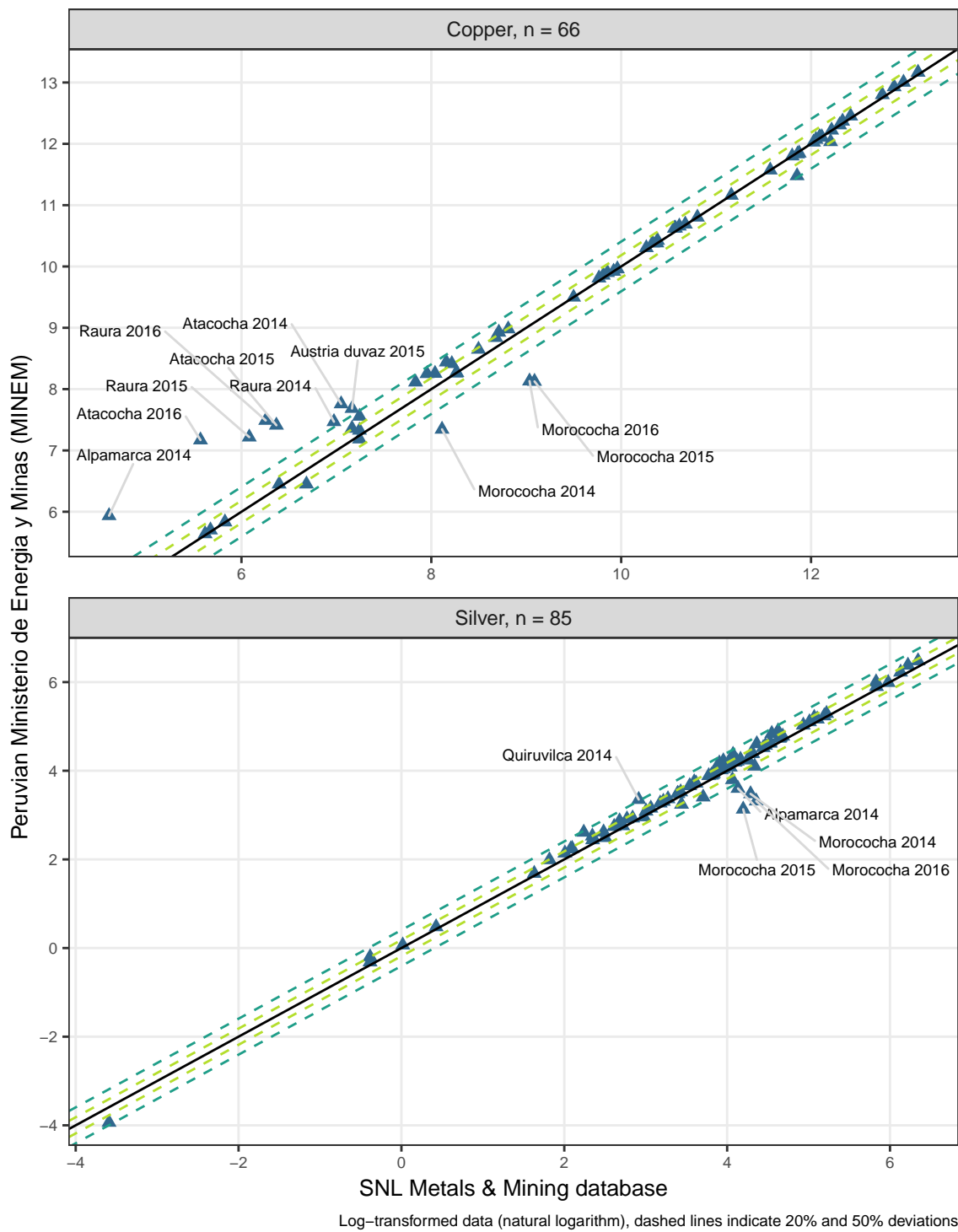


Figure 3: Mining activities in Peru 2014-2016 (production in tonnes, log-transformed): MINEM data vs. SNL Metals & Mining database



Citation

Luckeneder, S., Lieber, M., Giljum, S. 2019. Overcoming inconsistencies in sub-national mining data. The case of copper and silver mining in Peru. FINEPRINT Brief No. 4. Vienna University of Economics and Business (WU). Austria.

References

- [1] UN IRP. Global Material Flows Database: Version 2017, <http://www.resourcepanel.org/global-material-flows-database>. Paris: 2017.
- [2] COMTRADE. UN comtrade analytics. New York: United Nations Statistics Division; 2018.
- [3] MINEM. Estadísticas de producción minera. Lima: Ministerio de Energía y Minas; 2018.
- [4] SNL. Metals and mining database. New York: S&P Global Market Intelligence; 2018.
- [5] Pan American Silver. Reports and filings, <https://www.panamericansilver.com/investors/reports-and-filings/>. 2019.