The Ethnoregional Mineral Resources Dataset

Jason Sorens, Principal Investigator[[1]](#footnote-1)

University at Buffalo, Department of Political Science

jsorens@buffalo.edu

# Description

This dataset presents the estimated annual volume and value of mined resources at the level of substate “ethnoregional groups” over the years 1950-2006. Ethnoregional groups are defined by the following criteria from the Minorities at Risk Project, Phase IV (MAR) (<http://www.cidcm.umd.edu/mar/>):

1. An ethnopolitical group is a non-state communal group that collectively suffers, or benefits from, systematic discriminatory treatment vis-à-vis other groups in a society; and/or collectively mobilizes in defense or promotion of its self-defined interests.
2. Groups that constitute less than 100,000 people or less than 1% of their country’s population, or that reside in countries with populations under 500,000 in every year over the 1950-2006 period are excluded.
3. An ethnoregional group is an ethnopolitical group that possesses a regional base. An ethnopolitical group possesses a regional base if it is the predominant part of population in a region larger than an urban area and at least 25% of the group lives in the region (MAR variable *GC2*=1).

Variable *group* in this dataset gives the name of the ethnoregional group, *country*, the name of the country, *scode* the three-letter text ISO-code for each country, and *numcode* a numerical code for each group, matching the MAR designation.

Two datasets are available. “resources3.xls” is the full dataset. “np-resources.xls” contains national production figures for each mineral. Before inclusion in the full dataset, missing production values are imputed, but these observations are left missing in the production dataset. The percentage of observations missing for each production variable are given below.

# Process of Mapping Ethnoregions and Allocating Resources

The MAR dataset includes a battery of variables intended to describe the location of each group’s regional base. Variable *GC3* names the regional base, and in many instances this name corresponds with an administrative unit, the boundaries of which may be easily determined by consulting a comprehensive atlas. (The *Times Atlas of the World*, 10th comprehensive edition was used for this purpose.) Ethnoregional groups with a distinct language base can have their territories of residence more precisely delineated by reference to the linguistic maps available at Ethnologue.com. Where ethnoregional groups are not linguistically defined (e.g., Yugoslavia), the online thematic maps section of the Perry-Castañeda Library at the University of Texas (<http://www.lib.utexas.edu/maps/thematic.html>) were consulted. Finally, for the United States’ ethnoregional groups, Census 2000 data were used to determine ethnic plurality status at the county level for contiguous clusters of counties. These techniques were sufficient to determine, to one-degree precision, the boundaries of every ethnoregion. Variable *region* in this dataset gives the name and/or description of the region as finally determined.

The minerals included in this study include hydrocarbons (natural gas, crude petroleum [including natural gas liquids], and coal), metals (lead, zinc, uranium, antimony, manganese, tantalum, chromium, molybdenum, titanium rutile, titanium ilmenite, copper, nickel, tungsten, gold, iron, silver, and tin), and diamonds (gem and industrial, natural only).

For all of the minerals above except natural gas, petroleum, and diamonds, I consulted Derry (1980), a comprehensive atlas of world mineral deposits then known, and implemented the following procedures. The atlas places symbols for each of these minerals at approximate sites where they are found on the map. Using the available information on the approximate boundaries of ethnoregions, I counted up the number of occurrences of each mineral within each ethnoregion and country.

Variables with the *reg* prefix record the number of occurrences of each mineral in each ethnoregion. Variables with the *nat* prefix record the number of occurrences of each mineral in each country, with the following exception: if there are no occurrences, these variables are coded “1” so as to avoid division by zero in subsequently discussed variables. If a mineral deposit appears to straddle regional or state boundaries, a half is recorded. In some cases, the atlas gives additional information allowing the size of the deposit to be roughly guessed. If the deposit is termed “globally significant,” it counts for five. If it is determined “significant” or “major,” or the name of the mine is given (presumably indicating large, ongoing output), it counts for two.

For natural gas, petroleum, and diamonds, I consulted different sources. Thieme, Rød, and Lujala (2009) have produced a geocoded, time-series dataset of natural gas and petroleum deposits, and Gilmore, Lujala, Gleditsch, and Rød (2005) have done the same for diamonds. The number of deposits are totaled up by region and country in the same manner, but for these minerals deposits are counted only if production is recorded at that site in that year (if year of first production is not available, it is assumed that production occurred throughout the period). Information on the significance of individual mines in total national production, expressed as a percentage, are occasionally available in these datasets, and the numbers are adjusted accordingly.

The variable suffixes for the minerals covered in this dataset are as follows:

|  |  |
| --- | --- |
| *gas* | natural gas |
| *oil* | crude petroleum & natural gas liquids |
| *pb* | lead |
| *zn* | zinc |
| *coal* | coal, all forms |
| *u* | uranium |
| *sb* | antimony |
| *mn* | manganese |
| *ta* | tantalum |
| *cr* | chromium |
| *mo* | molybdenum |
| *ti* | titanium |
| *tiil* | titanium-ilmenite |
| *tirut* | titanium-rutile |
| *cu* | copper |
| *ni* | nickel |
| *w* | tungsten |
| *au* | gold |
| *fe* | iron |
| *ag* | silver |
| *sn* | tin |
| *diam* | diamonds, gem and industrial, natural only |
| *gem* | diamonds, gem, natural only |
| *ind* | diamonds, industrial, natural only |

For titanium and diamonds, the deposit information does not distinguish between different ores/qualities, but the available production and price figures do make this distinction. The handling of this issue is described below.

# Country Production Figures: Sources and Interpretation

The next step in the process was to obtain annual volume measures of country mine production of each mineral. These variables each begin with the prefix *np*. In all cases, only mine production, not refinery production or metals recovered from scrap, is counted. For most metals, only metal content is counted, rather than gross ore tonnage. The measurement units and type of material (ore versus metal) for each mineral can be found in the comments on the column headings in the “np-resources.xls” spreadsheet. The key concern was that the production and price variables measure the same materials in the same units. In the “resources3.xls” spreadsheet, an *i* suffix appears in variable names for which some observations were imputed.

For the United States, annual data on all minerals except uranium and the hydrocarbons are available from the U.S. Geological Survey (USGS) in spreadsheet form at <http://minerals.usgs.gov/ds/2005/140/> (accessed October 2, 2009). These spreadsheets also contain price data in constant 1998 U.S. dollars. For global production figures of all metals and the “all diamonds” variable from 1993-2006, I received spreadsheets via personal communication with the British Geological Survey (BGS). Data for the remaining years were entered by hand from the Bureau of Mines Minerals Yearbooks, available as scanned images at <http://minerals.usgs.gov/minerals/pubs/usbmmyb.html> (accessed October 2, 2009). Those yearbooks also include data on hydrocarbon production (after 1976 in the “Area Profiles” sections). Note that because data on gem and industrial diamonds come from the USGS, while data on “all diamonds” come from the BGS, the sum of gem and industrial diamond production does not always equal “all diamonds” production. For natural gas production for all countries for 1980-2006, data were available from the Energy Information Administration (EIA) at <http://tonto.eia.doe.gov/cfapps/ipdbproject/IEDIndex3.cfm?tid=3&pid=3&aid=1> (accessed October 2, 2009). For crude petroleum and natural gas liquids production for all countries from 1980 to 2006, data were available from this EIA website: <http://tonto.eia.doe.gov/cfapps/ipdbproject/IEDIndex3.cfm?tid=5&pid=53&aid=1> (accessed October 2, 2009). Finally, for coal production for all countries from 1980 to 2006, data were available from this EIA website: <http://tonto.eia.doe.gov/cfapps/ipdbproject/IEDIndex3.cfm?tid=1&pid=7&aid=1> (accessed October 2, 2009).

# Mineral Price Figures

The next step was to obtain annual estimates of prevailing international prices for each of the minerals. In general, f.o.b. spot prices were used. A minor limitation of the data is that they are all based on U.S. prices, but since almost all mineral markets are fully globalized, international prices correlate very closely with U.S. import prices. The only exception is natural gas, for which markets are more accurately described as continental, but even here, the availability of fuel substitutes causes natural gas prices to converge toward a global average. The USGS data already described were used for all non-energy minerals except diamonds. Rough natural diamond prices, gem and industrial quality, come from the USGS/Bureau of Mines Mineral Yearbooks and are estimated by dividing total value of U.S. imports by volume of U.S. imports. Uranium price data come from <http://www.eia.doe.gov/emeu/aer/nuclear.html> for 1981-2006 (accessed October 2, 2009), British Geological Survey (2007), *Mineral Profile: Uranium* for 1972-1980, and the Minerals Yearbooks for 1950-1971. Coal price data (average for all types of coal) come from the EIA: <http://www.eia.doe.gov/emeu/aer/coal.html> , as do oil (<http://www.eia.doe.gov/emeu/aer/petro.html>) and natural gas price data (<http://www.eia.doe.gov/emeu/aer/natgas.html>) (all accessed October 2, 2009). All these price data are expressed in constant 1998 U.S. dollars, with adjustments to nominal data based on the U.S. Consumer Price Index. The mineral price variables begin with the prefix *VAL*.

# Total Ethnoregional Production Figures

The regional deposit count for each mineral is divided by the country deposit count for the same mineral, to yield a rough estimate of the percentage of country production that takes place within the region, and then multiplied by the corresponding annual country production figure and by the corresponding annual price per unit figure. The resulting variables, prefixed with *rp*, thus give our best annual estimates of the total value of regional production in constant 1998 U.S. dollars. For titanium-ilmenite and titanium-rutile, the regional percentage of recorded titanium deposits is multiplied by the respective country production and price variables to create the two regional production variables. For gem and industrial diamonds, the regional percentage of recorded diamond deposits is multiplied by the respective country production and price variables to create the two regional production variables.

The sum of all the regional production figures for all minerals constitutes the variable *resource*, which measures the total annual value of all mined metals and hydrocarbons included in this dataset in each region.

Two additional variables based on the Gilmore et al. (2005) dataset are also included. *regsecondary* is the estimated proportion of regional diamond deposits that are secondary or alluvial in nature, while *csecondary* is the estimated proportion of country diamond deposits that are secondary or alluvial in nature.

# Missing Data

The original sources did not have complete production data for every country-year in the dataset, but in general rates of missingness were extremely low. The following table gives rates of missingness for each national production variable.

|  |  |
| --- | --- |
| *npgas* | 198/4465 |
| *npoil* | 0/4465 |
| *nppb* | 21/4465 |
| *npzn* | 55/4465 |
| *npcoal* | 37/4465 |
| *npu* | 295/4465 |
| *npsb* | 37/4465 |
| *npmn* | 36/4465 |
| *npta* | 96/4465 |
| *npcr* | 44/4465 |
| *npmo* | 77/4465 |
| *nptiil* | 32/4465 |
| *nptirut* | 34/4465 |
| *npcu* | 23/4465 |
| *npni* | 17/4465 |
| *npw* | 44/4465 |
| *npau* | 143/4465 |
| *npfe* | 33/4465 |
| *npag* | 31/4465 |
| *npsn* | 19/4465 |
| *npdiam* | 41/4465 |
| *npgem* | 58/4465 |
| *npind* | 73/4465 |

# Future Improvements

The USGS makes available geocoded data of mineral deposits for each of the minerals included in this study. Coding each mineral globally for active production years would be roughly approximate to the effort involved in the Thieme et al. (2009) and Gilmore et al. (2005) projects, and completing them all would thus require a substantial investment of time and personnel. However, it should give us a more accurate picture of regional production. Finally, the regional coverage of the dataset could be expanded, to include ethnoregions not in the MAR dataset or perhaps to be re-oriented along the boundaries of legally defined, substate jurisdictions (thus broadening the number of potential applications).

# References

Derry, Duncan R. 1980. *World Atlas of Geology and Mineral Deposits*. London: Mining Journal

Books.

Gilmore, Elisabeth, Nils Petter Gleditsch, Päivi Lujala, and Jan Ketil Rød. 2005. Conflict

diamonds: A new dataset. *Conflict Management and Peace Science* 22 (3): 257-92.

Thieme, Nadja, Jan Ketil Rød, and Päivi Lujala. 2009. The petroleum dataset codebook,

version 1.1. http://www.prio.no/CSCW/Datasets/Geographical-and-

Resource/Petroleum-Dataset/, accessed July 2, 2009.

1. I would like to acknowledge helpful research assistance from Wesley Boling, Fait Muedini, Anna Diakun, Haydee Izaguirre, and Kunchok Youdon. All remaining errors are my own. [↑](#footnote-ref-1)