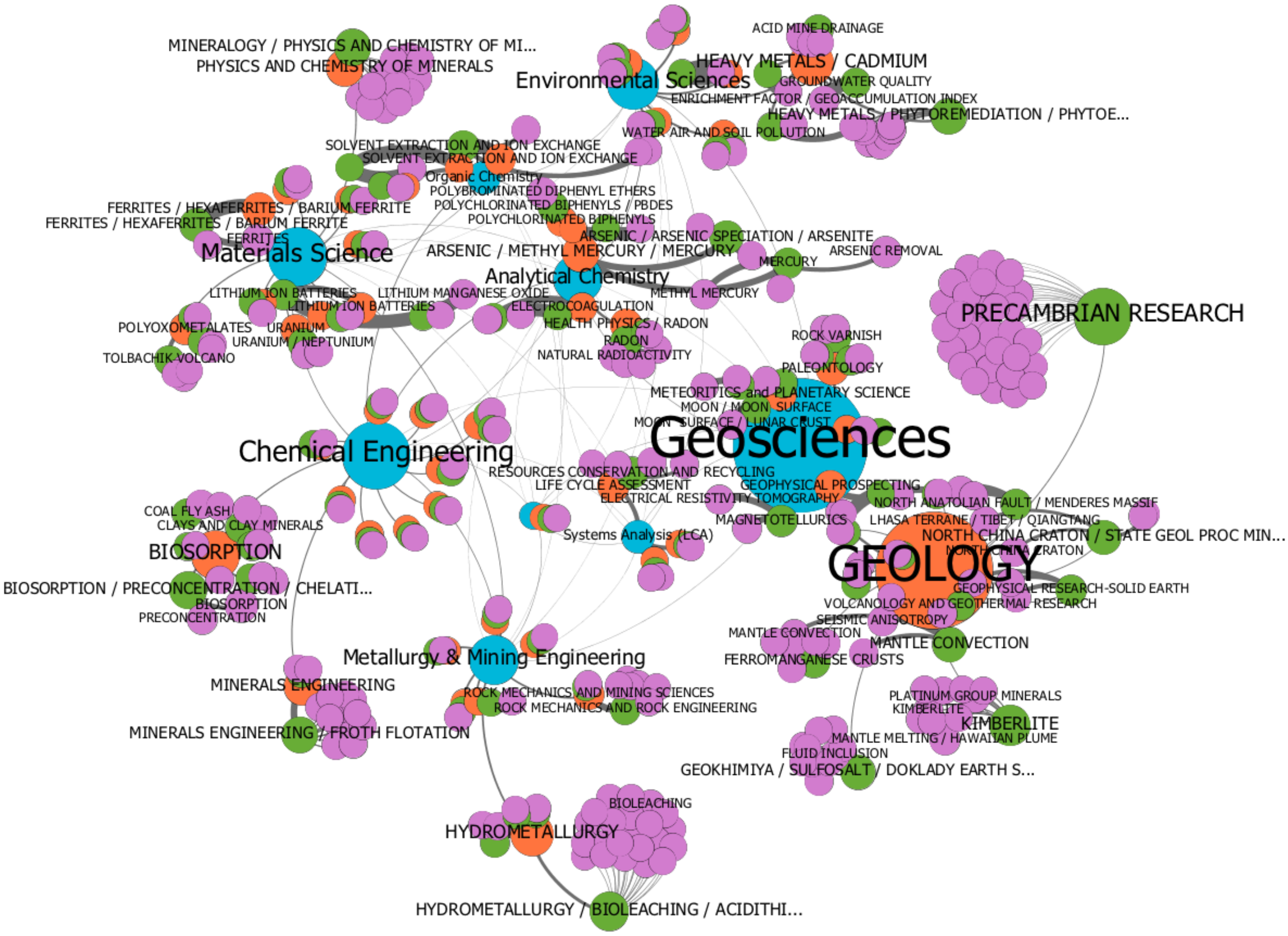
# Bibliometric study of Swedish research in mining and minerals

Ulf Kronman, Peter Sjögårde, Per Ahlgren.

KTH Royal Institute of Technology bibliometrics group. 2016-05-19, version 1.1



## Contents

## 1 Svensk sammanfattning

Denna rapport presenterar en bibliometrisk studie av svensk forskning inom området gruv- och mineralforskning. Studien har utförts med hjälp av en klustringsbaserad teknik som delar in forsknings-publikationer i hierarkiska klasser. Studien har genomförts som ett projekt för att bedöma användbarheten av den klustringsbaserade tekniken och utnyttjar delvis resultaten från en tidigare studie baserad på framsökning av publikationer via nyckelord som genomförts av Vetenskapsrådet.

För studiens genomförande klustrades 28 miljoner publikationer i KTHs bibliometriska databas *Bibmet* i 35 192 klasser och av dessa klasser valdes 243 ut som relevanta för gruv- och mineralforskning. I dessa klasser fanns 4 118 svenska publikationer mellan åren 1990-2014, utgörande 1,8 % av världsproduktionen inom området, vilket är cirka 30 % högre än den genomsnittliga svenska produktionen av vetenskapliga artiklar, sett över alla områden i databasen Web of Science. Den genomsnittliga citeringsgraden för svenska artiklar inom området var i början av 2000-talet i paritet med världsgenomsnittet inom området (1,16), men har under de senaste 5 åren sjunkit under världsgenomsnittet (0,92).

USA och Kina är de länder som har störst artikelproduktion inom området och där har Kina vuxit snabbt och gått förbi USA i publikationsvolym de senaste 5 åren. Sverige hamnar på 17:e plats i en lista över de mest publicerande länderna. De mest citerade länderna inom området är USA och Schweiz, där publikationerna citeras 40-50% mer än genomsnittet. De största publicerande organisationerna inom området internationellt sett är vetenskapsakademierna i Kina och Ryssland.

Studien indikerar att svensk forskning inom området är bibliometriskt stark inom klasserna[1] *High Pressure Research*[2] och *Lead Isotopes*[3], där Sverige publicerar 3 gånger mer och citeras 25-90 % mer än genomsnittet. Som svaga områden indikeras publikationsklasserna *Acid Mine Drainage*[4], *Tundish*[5], och *Impact Cratering*[6], där Sverige publicerar mycket men får ett svagt genomslag i form av citeringar.

De största svenska organisationerna inom området är Uppsala universitet, Luleå tekniska universitet, KTH och Naturhistoriska riksmuseet. Alla dessa organisationer har ökat sin publikationsproduktion under mätperioden, men den fältnormerade citeringsgraden för publikationerna inom området har sjunkit för i stort sett alla svenska organisationer sedan år 2000.

Den sammanfattande bedömningen av den i projektet tillämpade klassifikationsmetoden i relation till den nyckelordsbaserade metod som använts av Vetenskapsrådet är att den nya metoden ger möjlighet att fånga in fler publikationer än nyckelorden, samt att indelningen i publikationsklasser ger möjligheter till mer fokuserade analyser än de traditionella ämnesklassningarna som baserar sig på Web of Sciences ämneskategorier. Den klustringsbaserade metoden fångade in 4118 svenska publikationer i jämförelse med den nyckelordsbaserade teknikens 1100 publikationer.

Den klusterbaserade indelningen av publikationer skapar dock en viss osäkerhet om innehållet i publikationsklasserna, både genom att det är svårt att sätta etiketter på klasserna och att citeringarna

---

[1] Namnen på publikationsklasserna som anges i texten är manuella extraktioner och tolkningar av automatiserade etiketter som tagits fram med statistiska metoder. Den automatiserade etiketten anges nedan i fotnot kopplad till respektive tolkad etikett.

[2] Shock Wave Detonat Phys / High Pressure Research / Sci Extreme Condit

[3] Lead Isotopes / Stable Lead Isotopes / Unite Format Rech Phys

[4] Acid Mine Drainage / Acid Sulfate Soil / Acid Sulfate Soils

[5] Tundish / Continuous Casting / Electromagnetic Brake

[6] Meteoritics & Planetary Science / Impact Cratering Grp / Shock Metamorphism

mellan publikationer kan basera sig på andra kriterier än gemensam ämnestillhörighet samt att innehållet i klasserna dessutom kan variera över tiden. Möjligheten att analysera många små publikationsklasser skapar också ett mycket omfattande statistiskt datamaterial som kan vara svårt att presentera och visualisera. Projektet har löst detta genom att presentera statistik för utvalda publikationsklasser och visualiseringar av publiceringsrelationer på en webbplats som ett komplement till denna rapport: http:/www.kth.se/bibliometrics/projects/mine_and_minerals/

## 2 Summary in English

This report describes a method to analyse Swedish research in mining and minerals using a clustering-based hierarchical classification system and some results from the analysis. The method partly utilises the results from a keyword-based study om mining and minerals research performed by the Swedish Research Council.

In order to find publications related to mining and minerals research, 28 million publications in the bibliometric database *Bibmet* at the Royal Institute of Technology were clustered into 35 192 publication classes using a technique based on citations between the publications. 243 classes were selected as being relevant to mining and minerals research, using a statistical content labelling system.

For Sweden, a total of 4118 publications were found in the selected publication classes for the years 1990 to 2014. For the world as a total, the number of publications in the same publication classes is 225 366 which means that Swedish research in mining and minerals here is estimated to 1.8 percent, which is 30 percent above the average Swedish research presence, counted over all fields covered by the Web of Science database. The normalized citation impact at 1.16 of Swedish research in in parity with the world average for the mining and minerals field during the years 2000-2004, but has gone down to 0.92, below the world average, during the last 5-year period.

The most publishing countries in mining and minerals are USA and China, where China has had a large growth in publishing and surpassed USA in the last 5-year period. Sweden ends up at rank 17 in the list of countries publishing mining and minerals research. The most cited countries in mining and minerals research are USA and Switzerland at a level 30-45 % higher than the world average. The largest publishing organizations internationally are the Chinese and Russian academies of sciences.

The publication classes[7] where Sweden is found to be strong in mining and minerals research are *High Pressure Research*[8] and *Lead Isotopes*[9], where the Swedish publication output is 3 times higher than normal in mining and minerals classes and the citation impact is 25-90% above the class average.

The areas where Sweden's bibliometric performance in mining and minerals research is found to be weak are *Acid Mine Drainage*[10], *Tundish*[11], and *Impact Cratering*[12], where Sweden has a high publication presence in the classes, but a low citation impact.

---

[7] The names of the publication classes given in the text are manually interpreted labels extracted from statistically generated labels. The statistical labels are given in footnotes below, linked to the manual labels in the text.

[8] Shock Wave Detonat Phys / High Pressure Research / Sci Extreme Condit

[9] Lead Isotopes / Stable Lead Isotopes / Unite Format Rech Phys

[10] Acid Mine Drainage / Acid Sulfate Soil / Acid Sulfate Soils

[11] Tundish / Continuous Casting / Electromagnetic Brake

[12] Meteoritics & Planetary Science / Impact Cratering Grp / Shock Metamorphism

The organizations with highest publication volumes in Sweden are Uppsala University, Luleå University of Technology, KTH Royal Institute of Technology, and the Swedish Museum of Natural History. All these organizations have increased their publishing in mining and minerals research during the last 5-year period, but we also note that the citation rate for mining and minerals publications has fallen for most Swedish organisations since 2000.

The conclusions about the used classification methodology are that it brings forward possibilities to retrieve more publications than a keyword-based method and possibilities to analyse on a more detailed level than for studies based on the Web of Science subject categories. The drawbacks with the method are that the grouping of publications may not be fully in accordance with a subject classification and that it is difficult to describe the subject content of the publications in the classes, especially over a longer period of time. Analysis at lower classification levels also gives a vast amount of statistical data that can be difficult to digest and visualize. The project has solved this issue by publishing detailed results on a web site, as a supplement to this report: http:/www.kth.se/bibliometrics/projects/mine_and_minerals/

## 3 Background

The Swedish innovation agency Vinnova has been commissioned by the Swedish government to do an investigation of strengths and challenges in Swedish research in mining and minerals. The assignment can be summarized as follows in Swedish:

*Vinnova ska genomföra en ämnesöversikt och kartläggning av svenska styrkor och utmaningar inom gruv- och mineralforskningsområdet. Forskningsområdena Urban Mining, återvinning av metaller (särskilt kritiska metaller) och substitution av kritiska metaller ska inkluderas i uppdraget.*

*Sveriges styrkeområden och utmaningar inom gruv- och mineralforskningen i ett internationellt perspektiv ska identifieras och den strategiska nyttan av att genomföra satsningar på dessa ska bedömas.*

In parallel with the Vinnova commission, the Swedish Research Council (SRC) got assigned to do a bibliometric study of Swedish research in mining and minerals as a supplement and an input to the Vinnova study. The SRC bibliometric study has been finalized and results have been reported (Vetenskapsrådet, 2015). The study was based on keyword searches in the SRC version of the Web of Science (WoS) database.

Within the Vinnova-funded project *BibCap*, with aim to develop the bibliometric system at the KTH Royal Institute of Technology (KTH), a new class-based methodology for classifying research publications has been implemented. The analysts at Vinnova commissioned the KTH bibliometrics group to perform a project with a second bibliometric study on Swedish research in mining and minerals to put the new method to test. This report represents a documentation of the method and the results from the project.

An interpretation of the basic questions to answer in relation to the assignment from the government can be as follows:

- Which are the Swedish areas of strengths and weakness within research in mining and minerals as seen from a bibliometric perspective?
- In which countries and at what organisations is mining and minerals research located, in Sweden and globally?
- Where are the Swedish mining and minerals research environments located and what are their research focuses and bibliometric relations?

The aim of the KTH project has not only been to answer these questions in particular, but also to develop general methods to be able to answer similar questions within other research fields.

## 4 Data and methods

The KTH bibliometrics group had to address a number of challenges to be able to generate data to answer the questions stated for the study. These involved the grouping of publications into classes, the description of these classes, the selection of relevant publication classes for the study, and to compile relevant statistics for each publication class. Since the amount of information turned out to be extensive, there was also a need to make the statistical results navigational and to add options to zoom between different levels of aggregation.

To facilitate the interpretation, the project group also created visualizations. The visualizations also came with challenges regarding comprehensiveness, navigation and "zoomability". Visualization has also been the main approach to create means to identify research environments. Relations between researchers (co-authorship) and between researchers and publication classes were used in the visualizations to create a picture of the Swedish researchers' activity within the analysed subject area. However, a set of publications belonging to a researcher is not easily identified since author names are not unique. This problem of author ambiguity also needed to be addressed within the project.

The sub-headings of this section describe the most important steps in the publication selection and analysis process.

## 4.1. Data source

The data source used for the analysis is the KTH bibliometrics group's in-house database *Bibmet*, containing the same publication data as the Thomson Reuters database Web of Science (WoS).[13] A limitation to WoS publication types “articles” and “reviews” was made. Conference papers do not contain full address information and were therefore omitted.

It is important to bear in mind that WoS covers some subject areas better than others. For instance, research publications regarding social aspects in relation to mining and minerals are likely to be covered to a lower extent than publications regarding chemical, physical and biomedical aspects, since the social sciences publish a higher degree of publication types other than peer reviewed articles. Some research areas and topics within mining and minerals research may therefore be underrepresented in the results.

## 4.2. Creation of publication classes through clustering

To be able to identify publications within the area of mining and minerals research, a classification of the publications is needed, so that all publications within the data source are classified into a distinct publication class.

The WoS database contains a subject classification of journals. This classification is however broad and lack hierarchy, which makes zooming in to more narrow subareas impossible. Another problem with this kind of classification is that it only classifies journals, and consequently all publications in a journal will belong to the same field. This is problematic since not all articles in a journal belong to the same subject area, and this is especially the case for large broad journals such as for instance Nature, Science, PLoS ONE and Physical Review letters.

To solve this problem, a methodology–originated from the Centre for Science and Technology Studies (CWTS), Leiden University–for creating a publication-level hierarchical classification system was used (Waltman & van Eck, 2012). Within this methodology, the relatedness of publications is measured on the basis of citation links between the publications. Further, when the relatedness scores have been obtained, a network is constructed, and the publications are grouped into classes by use of a modularity-based clustering technique (Newman & Girvan, 2004) (Waltman & van Eck, 2013). Fig. 4.1 shows an example of a network partitioned by the use of the clustering technique.

To create a hierarchy, we used the same clustering technique to group the resulting publication classes into broader classes. By using this method, we obtained 4 levels of hierarchy, from 35 192 small, finely grained publication classes at level 1 up to 21 to broad large publication classes at level 4.

---

[13] Certain data included herein are derived from the Web of Science ® prepared by THOMSON REUTERS ®, Inc. (Thomson®), Philadelphia, Pennsylvania, USA: 

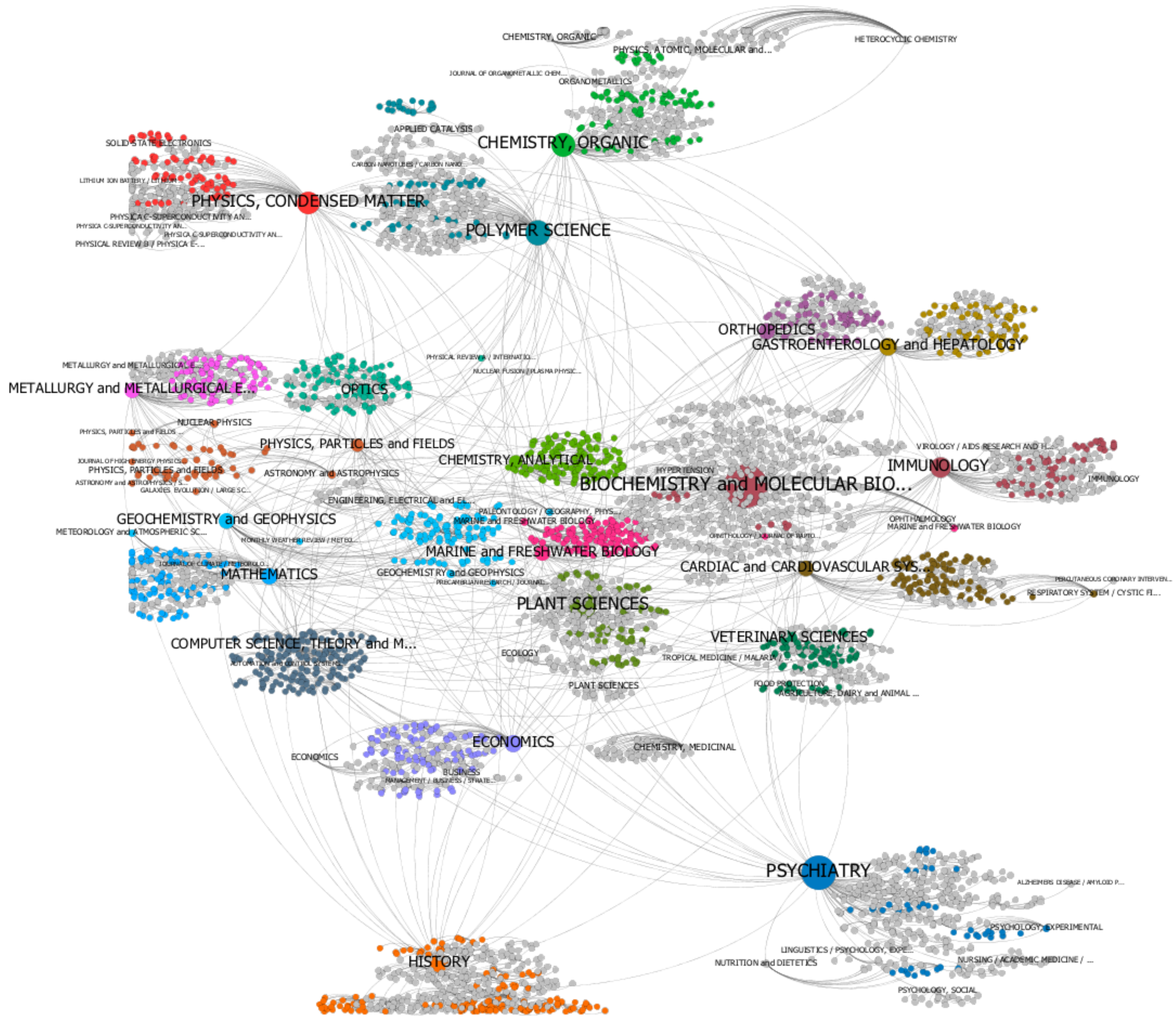



*Fig. 4.1* *Example of network where node colours show partitioning of nodes as a result of automatic clustering techniques. An interactive web presentation of details for publication classes can be viewed at address [http://www.kth.se/bibliometrics/classification/2015Q3/network/](http://www.kth.se/bibliometrics/classification/2015Q3/network/) and statistical data for the classes can be found at address [http://www.kth.se/bibliometrics/classification/2015Q3/](http://www.kth.se/bibliometrics/classification/2015Q3/)*

## 4.3. Description and labelling of publication classes

When the classification system had been created, it remained to describe the classes with regard to the subject area they represent. However, one has to note that using citations as the relation between publications to create classes does not always result in classes that hold together solely by the subject similarity of the publications. The common denominator could be the geographical location (e.g. when researchers within a country primarily cite each other), language (e.g. when publications mostly cite other publications written in the same language) or common theory, methods or data (e.g. publications may be grouped into the same class since they use the same method even though the application has been done within different subject areas).

To get descriptive information about each class, statistics of the publication data was compiled. The statistics included keywords, journal names, WoS journal categories, and titles of review articles. For a given class, a relevance score has been calculated for each such descriptive piece of information. The relevance score shows how informative a term is when it comes to describe the class. The terms with the highest relevance scores can then be used for description and labelling of classes.

In some classes author keywords most effectively describe the class, while in others a journal name most effectively describes the class. At the most aggregated level the WoS journal categories dominate as descriptors and at the most disaggregated levels author keywords have the best performance. The BibCap project plans to refine and document this newly developed method of labelling classes later as a part of the main project. Descriptive information of this kind has been generated for each publication class in the classification system. This has been done for all articles in the database and can be navigated at the web pages of the KTH bibliometrics group.[14]

During the course of the project for the bibliometric study of Swedish mining and minerals research, some drawbacks with the hierarchical nature of the publication classes have been spotted with regard to the labelling of the classes. When classes at lower levels (1 or 2) are aggregated to higher levels (2 or 3), sometimes the labels at the higher level can be misleading regarding the content of the lower classes, since the labels describe the content of *all* sub-classes, not only the selected classes for mining and minerals.

This effect is especially significant when a single class at level 1 is "aggregated" into a class at a higher level. In this case, it is more appropriate to keep both the analysis and the description of the content at level 1. This strategy was utilized in the analysis performed in sections 5.5 and 5.6, where the analysis has been performed on a mix of classes from level 2 and 1, depending on the number of subclasses in mining and minerals.

For the description of labels at the highest level 4, a manual interpretation and mapping was made, to make visualizations and results at this level more readable. This is the translation used between automatic labelling and manual labelling:

| Interpreted content label | Automatically generated label |
|---|---|
| **Applied Mathematics** | Mathematics / Mathematics, Applied / Math |
| **Systems Analysis (LCA)** | Economics / Operations Research & Management Science / Business |
| **Organic Chemistry** | Chemistry, Organic / Chemistry, Inorganic & Nuclear / Chemistry, Multidisciplinary |
| **Metallurgy & Mining Engineering** | Metallurgy & Metallurgical Engineering / Materials Science, Multidisciplinary / Engineering, Mechanical |
| **Chemical Engineering** | Polymer Science / Engineering, Chemical / Chemistry, Physical |
| **Environmental Sciences** | Plant Sciences / Agronomy / Horticulture |
| **Materials Science** | Physics, Condensed Matter / Physics, Applied / Materials Science, Multidisciplinary |
| **Geosciences** | Geochemistry & Geophysics / Geosciences, Multidisciplinary / Meteorology & Atmospheric Sciences |
| **Analytical Chemistry** | Chemistry, Analytical / Environmental Sciences / Atmospheric Environment |

*Table 4.1 Mapping of statistically generated labels for publication classes at level 4 to manually interpreted content labels. LCA stands for Life Cycle Analysis.*

[14] Description of publication classes: http:/www.kth.se/bibliometrics/classification/2015Q3/

## 4.4. Selection of publication classes for mining and minerals research

At this point in the process we had a four-level classification system containing information about publication classes and statistically generated labels for the classes. However, we did not know which of the classes that belong to the mining and minerals research area. To make such a selection, we used the publications previously retrieved by SRC with the use of keyword search terms. Our assumption was that if a class contained a large number of publications from the SRC study[15], and these publications constituted a large share of the total number of publications within this class, then this class most likely should be categorized as a mining and mineral publication class. Therefore, we compiled statistics for each class about the number of SRC-study-publications. This was done at level 1, the lowest and most disaggregated level.

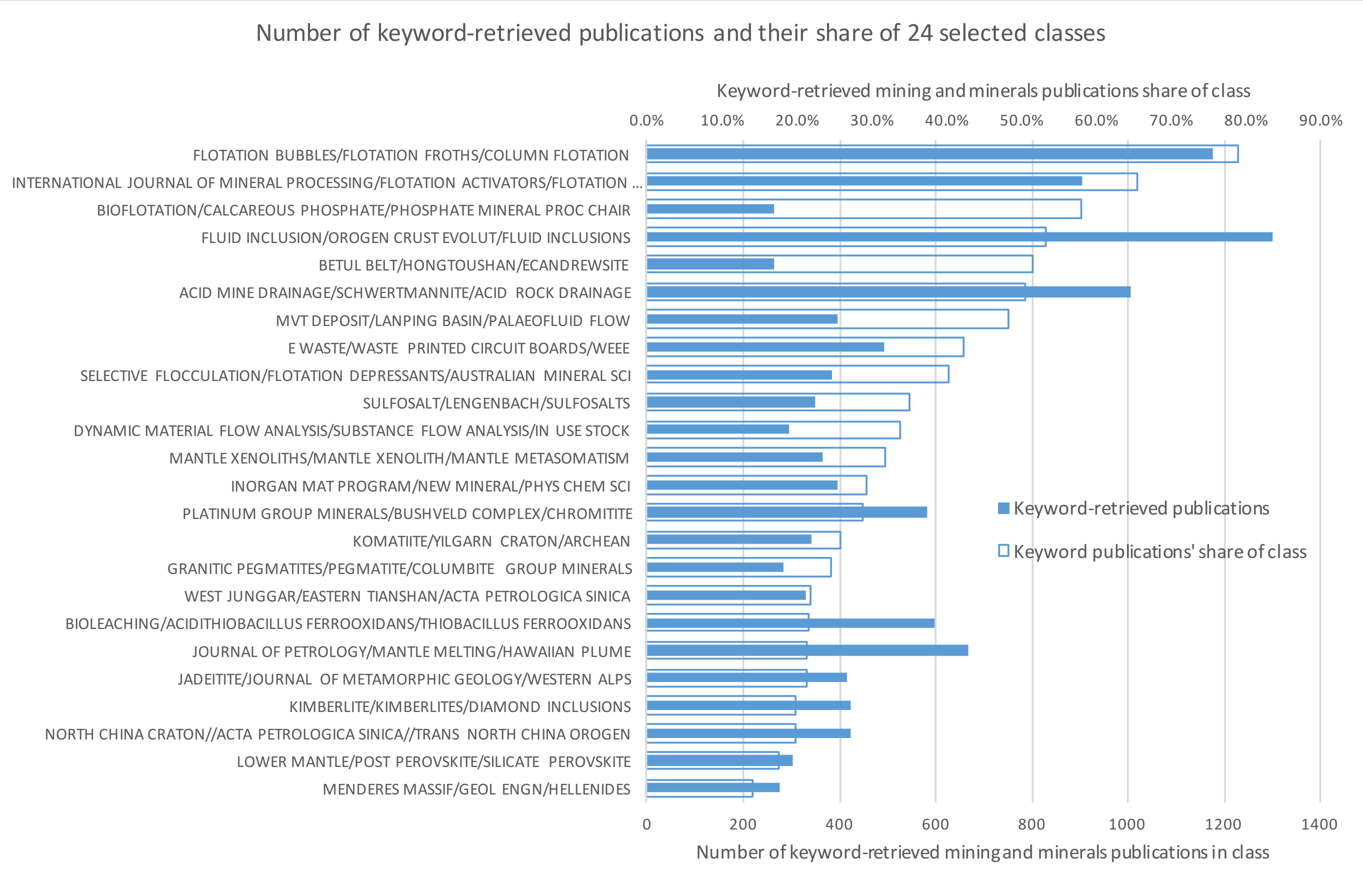



*Fig. 4.2 Number of publications retrieved by the SRC keyword search for publication classes at level 1 during years 1990-2014, together with the keyword searched publications' share of each class. 24 classes with more than 250 publications listed of altogether 243 selected classes at level 1. Classes are listed in descending order according to the share of SRC-publications.*

We then set limits regarding number of SRC-study-publications and share of SRC-study-publications of the total number of publications in class to determine if a class was to be included in the study or not.

[15] Called SRC-study-publications below

The following 4 criteria were used to include classes relevant for mining and minerals research from the 35 192 publication classes at level 1. If *all* of the following criteria were satisfied, the class was included in the automatic class selection:

1. The class contains at least 10 SRC-publications OR the SRC-publications constitute at least 10% of the total number of publications in the class.
2. The class contains at least 50 SRC-publications OR the SRC-publications constitute more than 10% of the total number of publications in the class OR the total number of publications in the class is at least 150.
3. The class contains at least 50 SRC-publications OR the SRC-publications constitute more than 3% of the total number of publications in the class.
4. There are at least 4 Swedish SRC-publications in the class.

The classes were then looked through and some manual adjustments were made, where about 55 classes below the levels set for inclusion were included manually and about 25 classes above this level were excluded. The manual adjustment was done in collaboration with a subject expert from Vinnova.

About 165 classes were included exclusively on statistical grounds, making up to a sum of 245 selected classes at level 1. Two classes were removed later in the project by examining the titles of the publications and considering them to be mostly out of scope, resulting in a final result of 243 analysed classes at level 1.

During the manual selection of classes, some decisions regarding the balance between *precision* and *recall*[16] were made, so that recall was favoured over precision. This means that there is a risk that the selected classes cover more aspects of mining and minerals research than could be expected if the publications had been selected individually or by keyword search. This also means that more than the commercial "core" of mining and minerals research are covered. For instance, some geological science and environmental research regarding effects from mining and minerals are covered by the selected publication classes.

The aggregated publication classes at levels 2-4 above classes from level 1 in the hierarchy were delimited to contain publications only from the selected mining and minerals classes at level 1. A list of labels for the selected 243 classes at level 1 and their position in the class hierarchy can be viewed at the project results website.[17]

---

[16] *Precision* is the fraction of retrieved instances that are relevant, while *recall* is the fraction of relevant instances that are retrieved.

[17] Selected classes in mining and minerals research http:/www.kth.se/bibliometrics/projects/mine_and_minerals/

The final result of class selection process in relation to the SRC keyword selection can be summed up as follows:

| | |
|---|---|
| Number of  included publication classes | 243 |
| Number of included publications, worldwide | 225 366 |
| Number of included Swedish publications[18] | 4118 |
| Number of included SRC keyword publications, worldwide | 33 991 |
| Number of *excluded* SRC keyword publications, worldwide | 19 046 |
| Number of included Swedish SRC keyword publications | 976 |
| Number of *excluded* Swedish SRC keyword publications | 124 |

*Table 4.2 Publication statistics for the 243 selected publication classes in the study.*

Labels and statistical information for all the 243 selected publication classes at level 1 can be viewed at the project website at address: http:/www.kth.se/bibliometrics/projects/mine_and_minerals/ in a class hierarchy, together with links to more detailed information for each publication class.

## 4.5. Identification of authors and research environments

Information about research environments is not easily retrieved from the database since there is no such data. Addresses in author affiliations normally contains information about university or other organization and country. Information on a more detailed level, such as department or centre, is not always included. Therefore, we considered publication addresses as unreliable to identify research environments. Instead, we used data on the researcher level to indicate research environments.

We assumed that researchers who co-publishes frequently can be said to represent a research environment. However, individual researchers are not easily identified in the database since authors often share the same name. To identify publications written by the same person we used a disambiguation algorithm developed by CWTS, described in (Caron & van Eck, 2014).

To determine pairs of publications which have been authored by the same person, the approach uses both similarity of author information connected to the person in the publications and similarity between the general publication information. The approach is not 100 percent correct but it has performed well[19] when compared to verified data from KTH and the Karolinska Institute. We therefore considered the approach appropriate for this study and implemented it on Swedish researchers. The result of the author disambiguation approach is a publication set for each identified researcher.

To identify research environments, we used this information to create networks based on co-publishing between researchers (see the visualization "Co-publishing between Swedish researchers within mining and minerals research" below). To facilitate the identification of research environments we used the modularity-based clustering technique of the open-source program *Gephi* to group researchers into potential "environments" based on their co-publishing relations. The grouping is not organization-based, which means that co-publishing researchers within different organizations can be considered as belonging to the same research environment.

[18] A Swedish publication is defined as a publication where at least one author is affiliated with a Swedish organization.

[19] The method has been tested on 63 000 author-verified articles from KTH and the Karolinska Institute and showed a recall of 93% and a precision of 97%.

## 4.6. Compilation and presentation of statistics

The ambition was to present the statistics in a manner that makes it possible to zoom in from an aggregated level to a less aggregated level. Therefore, we have presented the statistics as web pages with links between classification levels and to pages with country and organization statistics for each publication class. The web pages are also linked to the classification system that we have created to make it possible to get further information about the contents of each class.

The results of the statistical analysis of the 243 selected publication classes are presented at the KTH bibliometrics group's website for the project of the bibliometric study of Swedish mining and minerals research.[20] The start page of the web presentation contains links to visualizations of relations between publication classes and authors affiliated to Swedish organizations, and a list of basic statistics for all the selected publication classes at the lowest level together with an outline of the hierarchy of classes, so that publication classes at higher levels are presented together with each selected class. The list contains links to webpages with more detailed information.

For each of the selected publication classes a web page was created with statistics regarding publication output and citation impact. This was done at the level of countries and organizations. Since the material was sizeable we limited the results to be displayed for each class to 50 countries and 100 organizations. Both limitations were based on publication frequencies for the countries and organizations respectively. In addition to this information separate result pages were compiled for Swedish organizations.

The following statistics are included on each web page:

- **Number of publications, full counts** – This indicator expresses the number of publications that have been authored or co-authored by a country or an organization. Since full counts are used this means that a publication can be counted multiple times if co-authored between countries or organizations. For instance, a publication co-published by Swedish researchers and Chinese researchers will be counted both for Sweden and China.
- **Share of the total number of publications in the field, full counts** – The share of all publications in each field that have been authored or co-authored by a country or an organization is expressed by this indicator. The same as mentioned above apply regarding multiple counting if a publication is co-authored by two or more countries or organizations.
- **Number of publications among 10% most cited in field, full counts** – This indicator expresses the number of publications that have been authored or co-authored by a country or an organization that are among the 10% most cited publications in the same WoS subject category, published the same year and of the same publication type (article or review). Note that the WoS subject categories are used as reference groups, not the publication class itself.
- **Share of publications among 10% most cited in field, fractionalized** – The same references groups are used for this indicator. It is calculated as weighted means. For a country or an organization and a publication, the number of addresses containing the country name or organization name is calculated. The value is then divided with the total number of addresses in the publication, resulting in a weight of between 0 and 1. This weight is then used for the calculation of the weighted mean. Publications co-authored by many countries or organizations are given less weight in the calculation than publications by a single author/organization or co-authored by just a

[20] http:/www.kth.se/bibliometrics/projects/mine_and_minerals/

few countries or organizations. This is done to avoid an inflation of the indicator values, caused by an increasing degree of co-authorship. The world average of the indicator is exactly 10 percent. Once again note that the WoS subject categories are used as reference groups, not the publication class itself, which means that the world average within a publication class can differ from 10 percent. For a formal definition of the indicator, see (Ahlgren & Sjögårde, 2015)

- **Field normalized citation rate ($c_f$), fractionalized** – This indicator, like the top 10% indicators described above, normalizes for the variation of citation pattern between subject fields. Each publication is compared to a reference group of publications, i.e. publications within the same WoS subject category, published the same year and of the same document type (article or review). This is done by dividing the number of citations of each publication with the average number of citations for the reference group of publications, which results in a normalized citation rate. The indicator value expresses the field normalized mean citation rate of the country's or organization's publications. It follows from the definition of cf that the mean field normalized citation rate of all records in the WoS database is 1. A citation rate above one indicates that the set of publications is cited above world mean, e.g. a citation rate of 1.2 indicates that the publications are cited 20 percent above world average. Fractional weighting is calculated by the same method mentioned above and used for the same reasons given above. Once again note that the WoS subject categories are used as reference groups, not the publication class itself, which means that the world average within a publication class can differ from 1.0. For a formal definition of the indicator, see (Ahlgren & Sjögårde, 2015).

# 5 Results

The approximately 30 000 scientific publications per year from Sweden that can be found in the WoS database can be estimated to be around 1.25 - 1.5 percent of the yearly world production of scientific articles, depending on whether articles are counted in full (1.5) or address fractional counts (1.25), calculated over all fields of science in the WoS database. This average share can be taken as some sort of baseline when trying to identify research areas where Sweden can be considered to be strong in publication volume. If the Swedish share of publications in a class for instance is larger than 2 percent, this could be considered to be an area where Sweden has a high publication volume.

Likewise, the mean field normalized citation rate ($c_f$) for Swedish research publications the last 10 years is about 1.15 and the mean share of publications among the 10 percent most cited in each field (top10%) is about 11.5 percent (both mean values calculated on address fractionalized publication weights). These values can thus be considered as some kind of indication on whether Swedish publications within a research area has high citation impact or not. For instance, a $c_f$ value below 0.8 can be considered average or low whereas a $c_f$ value above 1.2 can be considered fairly high (Visser & Nederhof, 2011). Equally, a top10% value below 8 percent can be considered low and a value over 15 percent can be considered high.

## 5.1. Mining and minerals research in Sweden in relation to the rest of the world

The data analysis starts with an overview of the trends for publication volume and citation impact for Sweden for *all* the 243 selected publication classes for mining and minerals *as a group* in relation to publications from the whole world in the same group.

### *5.1.1. Publication volume*

For Sweden, a total of 4118 publications[21] can be found in the 243 selected publication classes for the years 1990 to 2014, which means that the analysed number of Swedish publications in the selected publication classes are less than 170 per year.

For the world as a total, the number of publications in the same publication classes is 225 366. The Swedish share of publications in the mining and minerals selection is thus 1.82 percent, which is about 20 percent higher than the average Swedish production of scientific articles over all areas in WoS (counted as full publication counts).

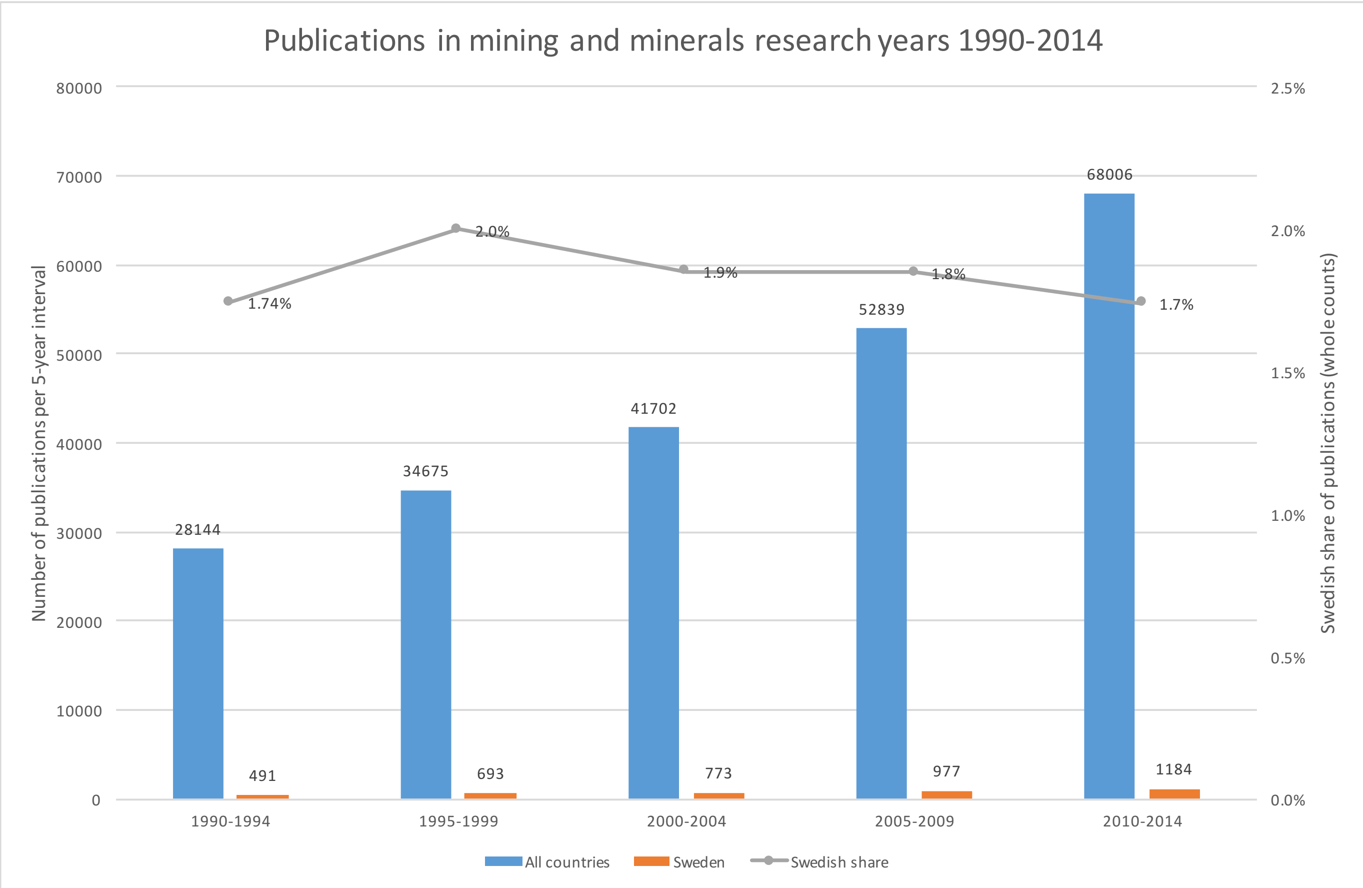



*Fig. 5.1 Publication volume in mining and minerals research for all countries and Sweden, together with Swedish share of publications (full counts) for 5-year periods years 1990-2014.*

From Fig. 5.1 it can be noted that during the years 1995 to 2009 the Swedish share of publications in mining and minerals research was higher than Sweden's average 1.5 percent share of scientific publications. But it can also be noted that the share has fallen from 2.0 percent to 1.7 percent during the most recent years, even if this is still above the Swedish average share. The Swedish average share over the whole period is 1.82 percent.

Some of the decline in Swedish share of global publications may be attributed to the rise of China's publishing in this field, which has grown significantly during the recent years. It can be noted that the

---

[21] 4118 publications full counts and 2874 publications, address fractional counts

number of Swedish publications within the field still is rising from 977 the previous 5-year period to 1184 the last period.

Next, we look at the development of the share of publications for Sweden, compared to a set of selected "benchmark nations". The chosen countries are: USA, China, Russia, Australia, Switzerland, Canada, Brazil, Germany, and Finland.

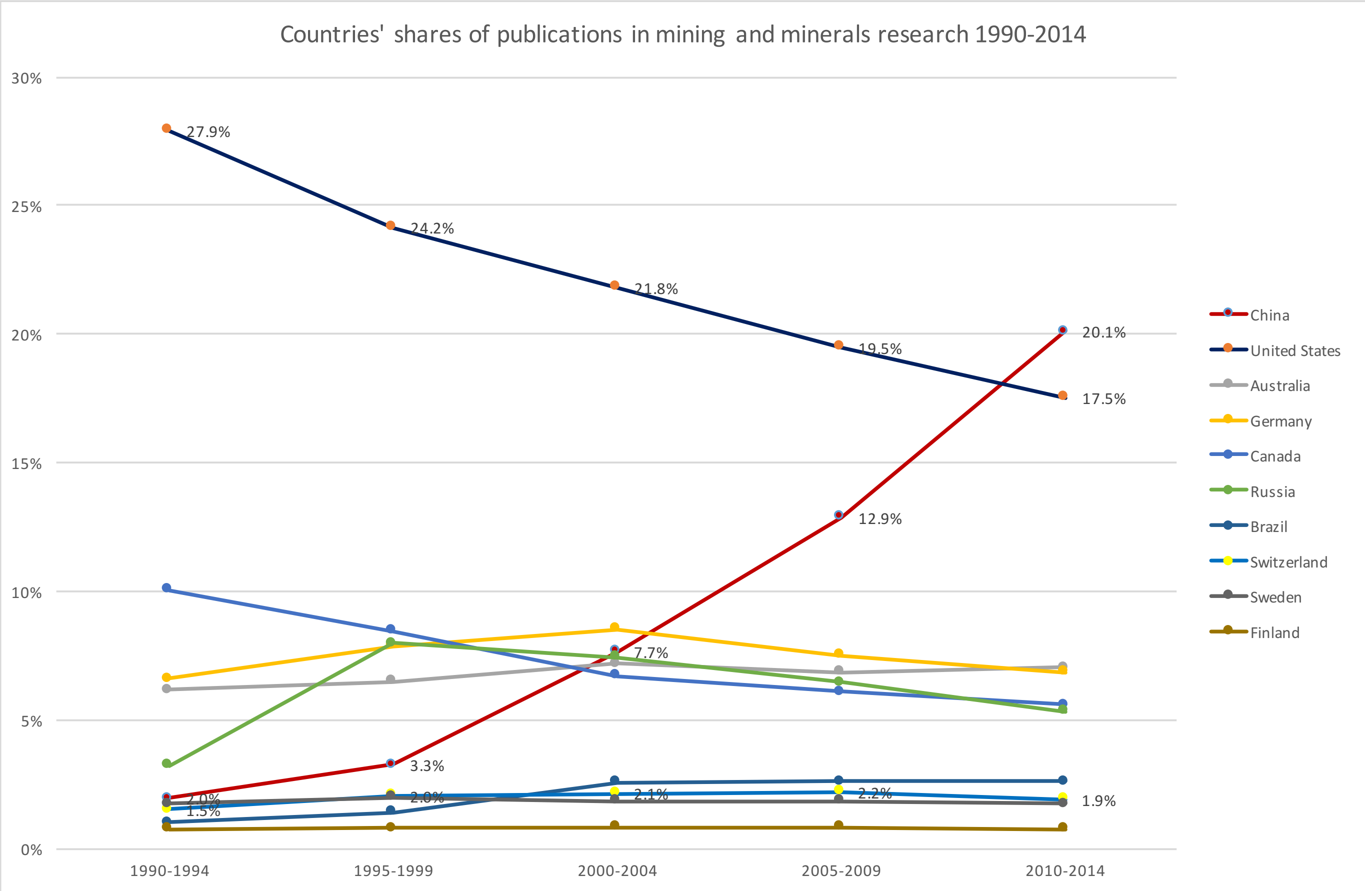



*Fig. 5.2 Shares of publications in mining and minerals research 1990-2014 for a set of selected mining and minerals countries. Lines for USA, China and Sweden are marked with percentage values. A web presentation of the data behind the diagram can be found at the address* *http://www.kth.se/bibliometrics/projects/mine_and_minerals/Overall_Level4/all_countries.html*

Fig. 5.2 shows that China has had a huge rise in publishing in mining and minerals research from 2000 and onwards. The share goes up from around 3 percent in 1995-1999 to over 20 percent in 2010-2014. This huge rise affects all other countries, which have shown a decline in their shares since 2000; this is especially notable for the US production, which has declined from about 28 percent down to 17.5 percent.

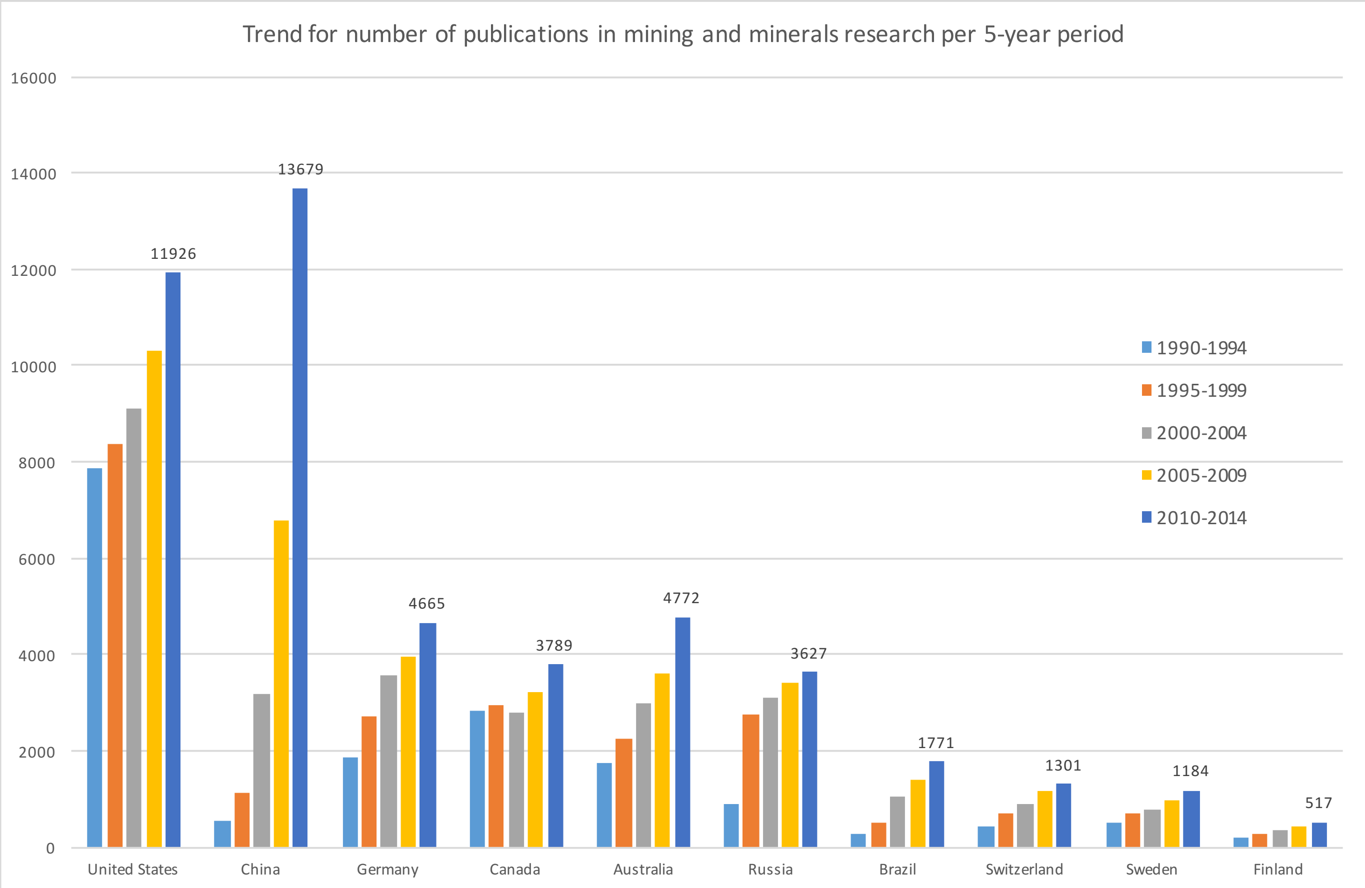



*Fig. 5.3 Number of publications per 5 year period for a selection of countries in mining and minerals research over the years 1990-2014. The last period is marked with the actual publication numbers (full counts). A web presentation of the data behind the diagram can be found at the address* *http://www.kth.se/bibliometrics/projects/mine_and_minerals/Overall_Level4/all_countries.html*

From Fig. 5.3 the huge rise of publications in mining and minerals research in China once again can be noted, surpassing USA in the last analysis period. Even though the publication volume is growing for all countries, the extreme rise of Chinese publishing makes the shares of all other countries diminish.

### *5.1.2. Citation impact*

As a supplementary information to the publication volume, the impact measured as the field normalized citation rate ($c_f$) and the share of publications among the top 10 percent most cited in the field (top10%) can be studied.

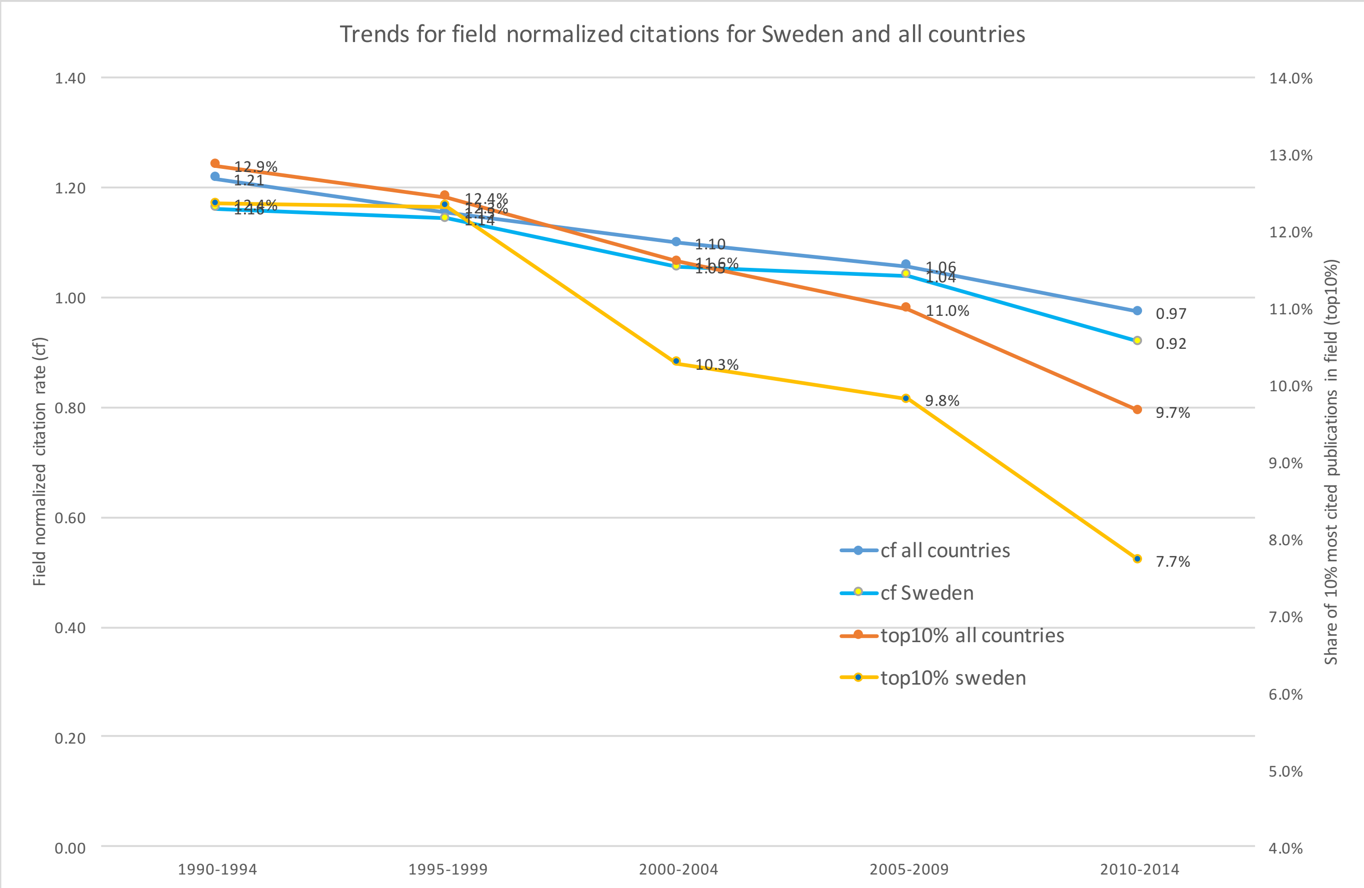



*Fig. 5.4 Field normalized citation rate ($c_f$) and share of publications among top 10 percent most cited (top10%) for Sweden and all countries during years 1990-2014. Address fractionalized and weighted citation counts.*

Fig. 5.4 shows that Swedish research in mining and minerals had a citation impact at par with the world mean during years 1995 to 2009, but that the citation impact has fallen below the world mean in the last 5 years. The effect is especially visible for the share of top 10 % most cited publications, where Sweden has fallen from a share of 12.4% down to 7.7%.

It is also interesting to note that the mean normalized citation rate and the share of publications among the 10 percent most cited in the WoS journal fields has been declining for the whole world the last 10 years for mining and minerals research as defined by the selected publication classes. Since the field normalization is done against WoS subject fields this means that the publications in the selected classes have had diminishing citation impact in their fields the last 10 years. This needs to be further investigated, maybe by looking more into details about the publication classes at lower levels in the hierarchy.

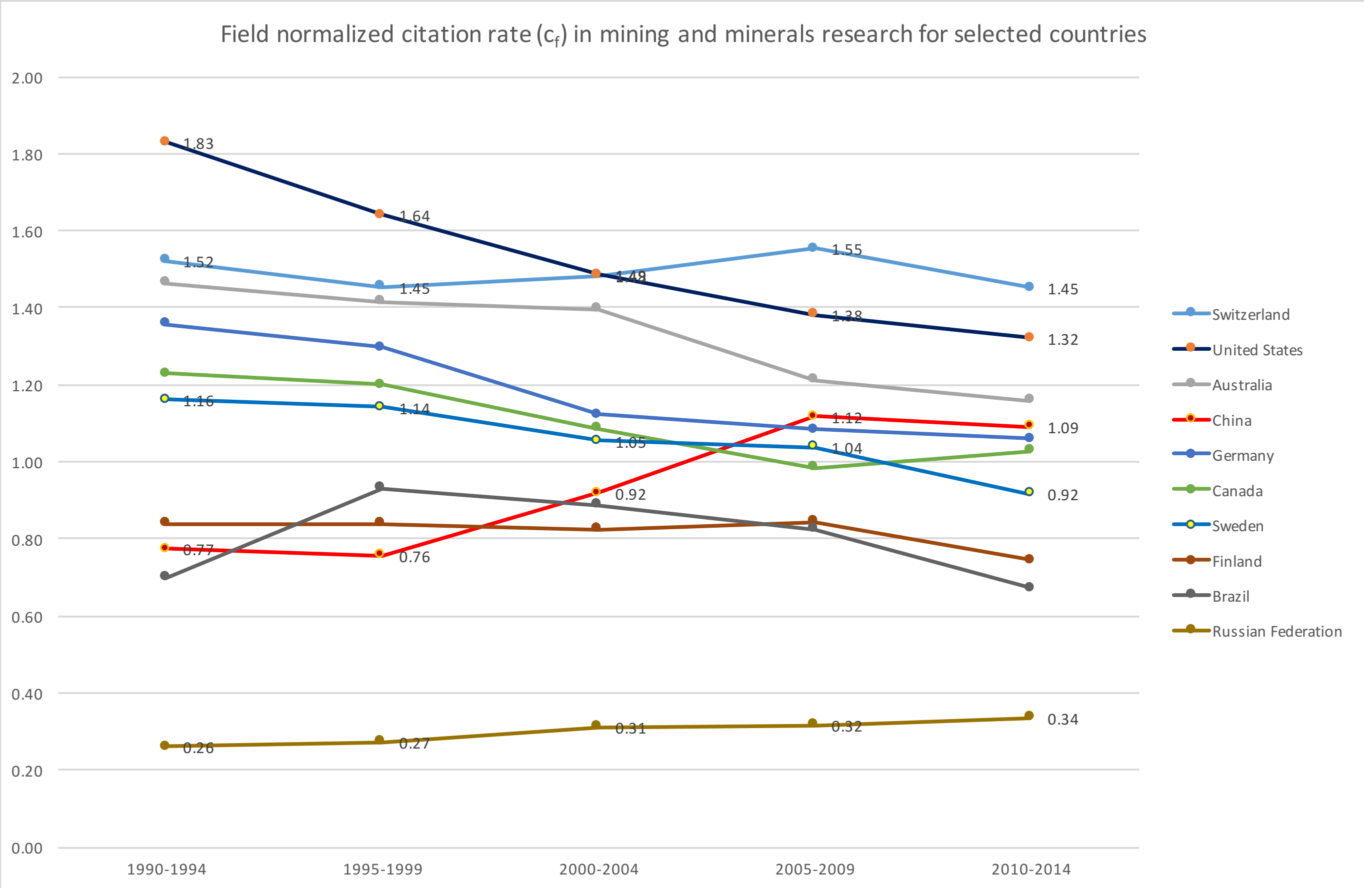



*Fig. 5.5 Field normalized citation rate ($c_f$) for selected countries in mining and minerals. Switzerland, USA, China, Russia and Sweden lines are marked with values. A web presentation of the data for the graph can be found at the address* *http:/www.kth.se/bibliometrics/projects/mine_and_minerals/Overall_Level4/all_countries.html*

From Fig. 5.5 it can be noted the decline of citation impact (1.83 – 1.32) of US publications and the rise of impact (0.77 – 1.09) of Chinese publications. It can also be noted that Switzerland has maintained a high-impact position during the period and is now the country with highest impact in mining and minerals research (about 1.45).

We can also see that the fairly large amount of publications coming from Russia has a very low international citation impact. This may be due to the articles being written mainly in Russian, or dealing with domestic aspects of mining.

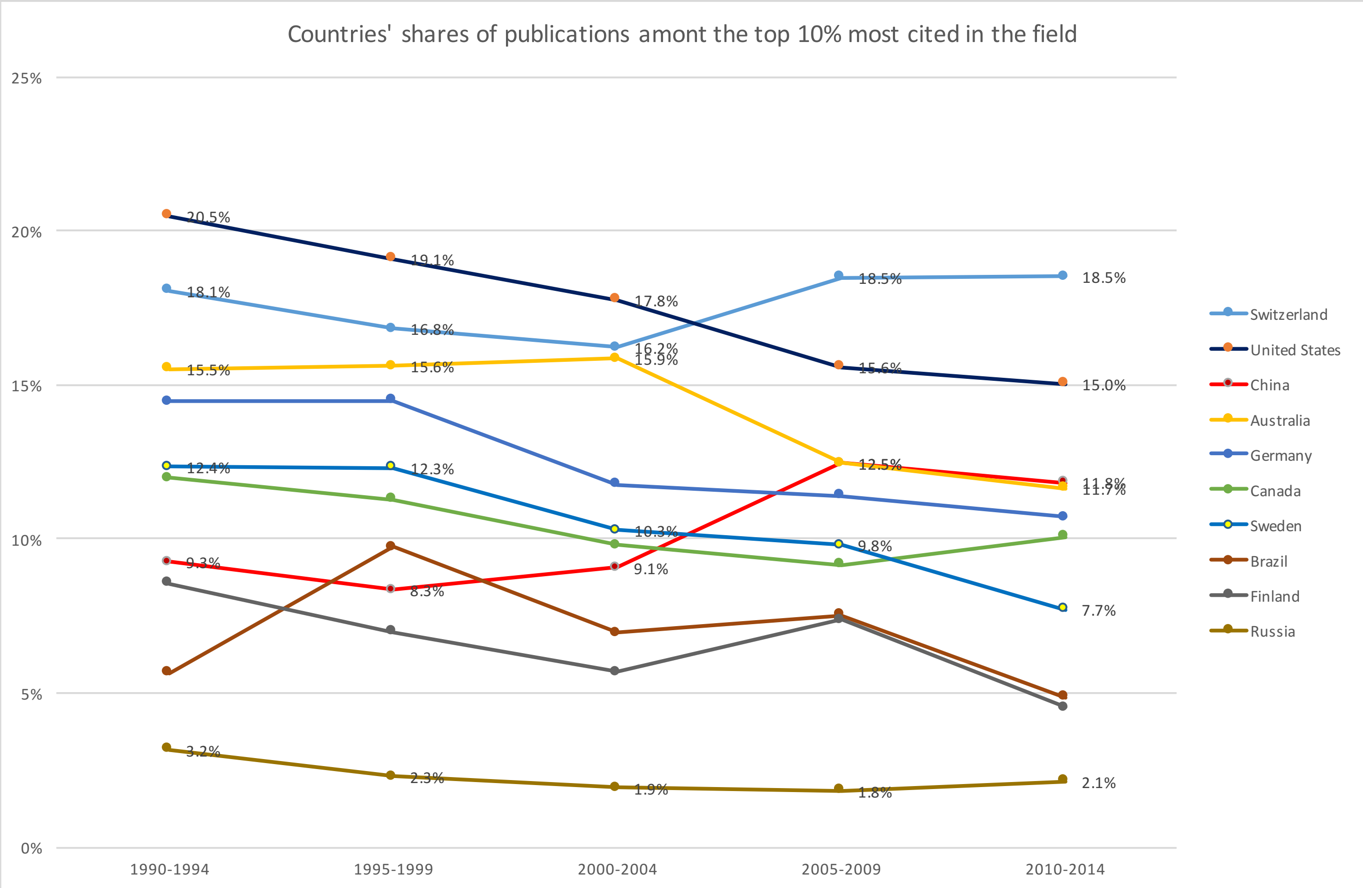



*Fig. 5.6 The selected mining and minerals countries' shares of publications among the top 10% most cited in the field. Numbers are based on weighted address-fractional counts.*

## 5.2. Important organisations with regard to publication volume and citation impact

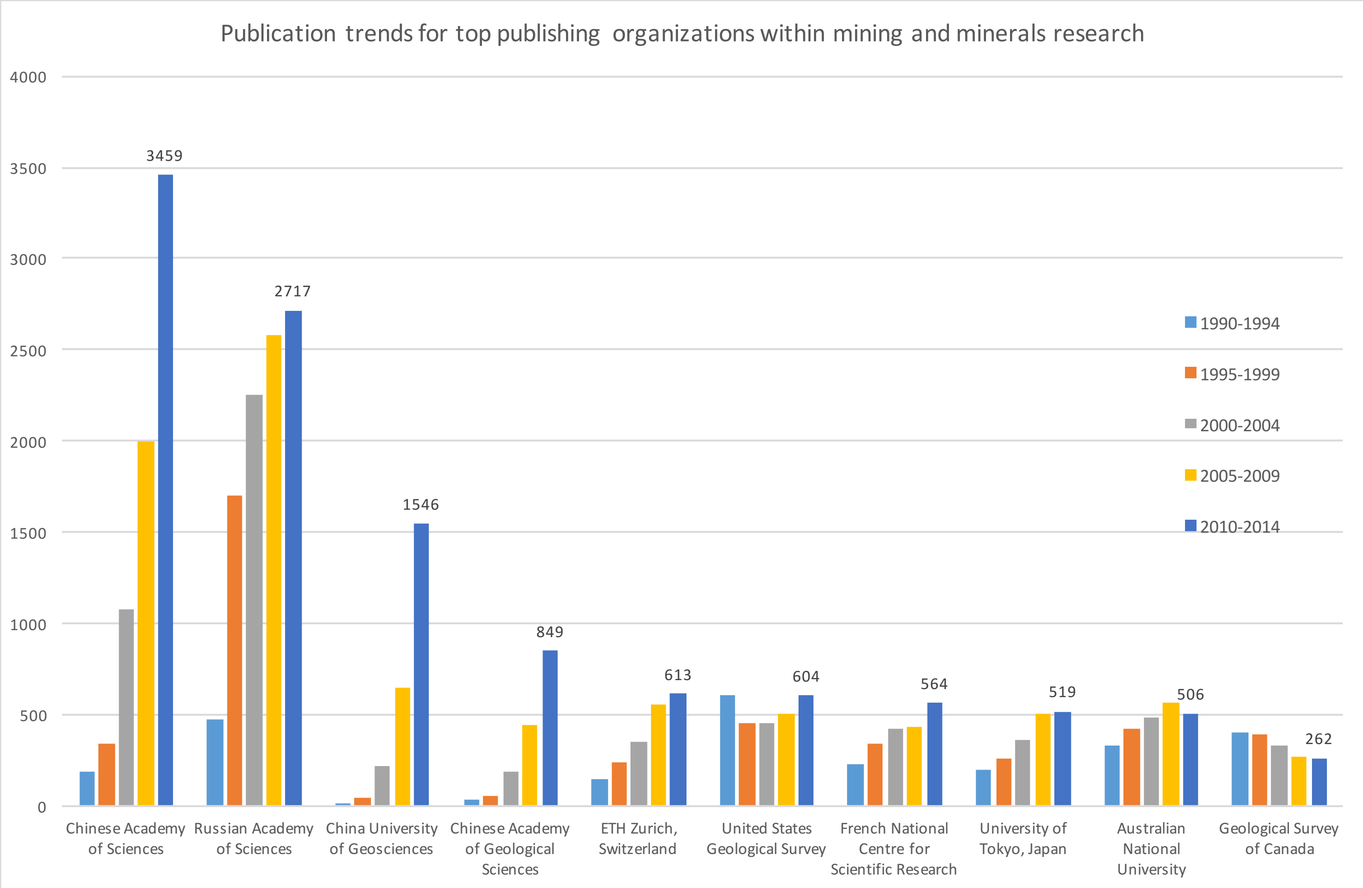



*Fig. 5.7 Number of publications per 5-year period for top publishing organizations within mining and minerals research. A web presentation of the data for the graph can be found at the address http:/www.kth.se/bibliometrics/projects/mine_and_minerals/Overall_Level4/all_org.html*

Once again a huge rise in publications for the Chinese organizations can be noted, surpassing the largest Russian organization in the last period. No Swedish organization is represented among the top publishing organizations; Uppsala University is at the Swedish top in slot 76 in the international list.

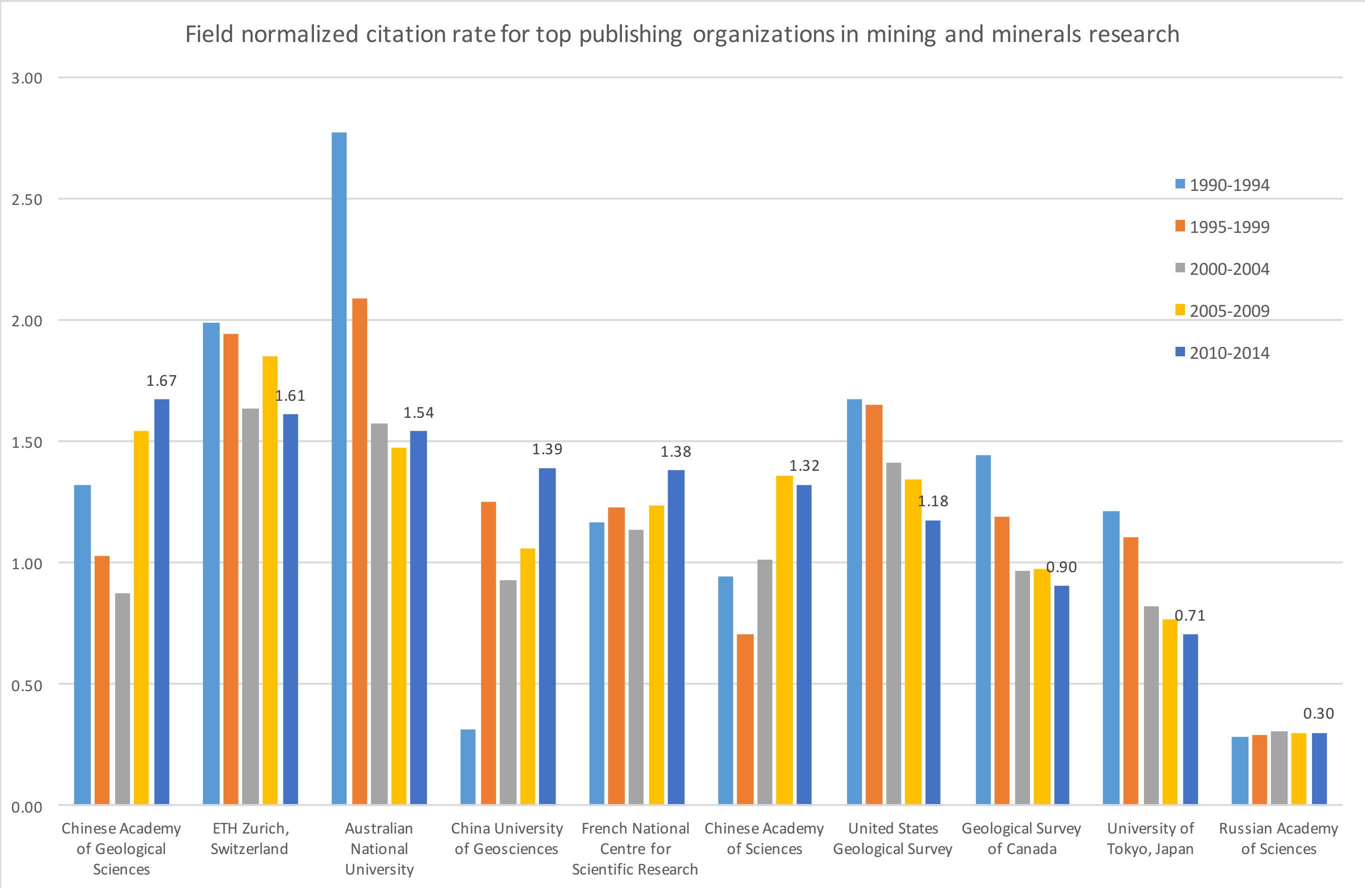



*Fig. 5.8 Field normalized citation rate ($c_f$) for top publishing organizations in mining and minerals research. The graph is limited to organizations publishing more than 1500 publications between 1990 and 2014 and based on address-fractional publication counts. A web presentation of the data for the graph can be found at the address* [http:/www.kth.se/bibliometrics/projects/mine_and_minerals/Overall_Level4/all_org.html](http:/www.kth.se/bibliometrics/projects/mine_and_minerals/Overall_Level4/all_org.html)

As can be noted from Fig. 5.8, the Chinese Academy of Geological Sciences now is producing the most cited publications, followed by ETH Zürich and the Australian National University, which though has had a decline in citation impact over the analysis period, but recovered a bit in the last 5-year period. ETH Zurich has maintained a high citation impact over the full analysis period. We can also note that the Russian Academy of Sciences, although having a large publication volume, has a very low citation impact on their publications, which is in accordance with the overall low citation rate for Russian research in mining and minerals.

## 5.3. Important Swedish organizations with regard to publication volume and citation impact

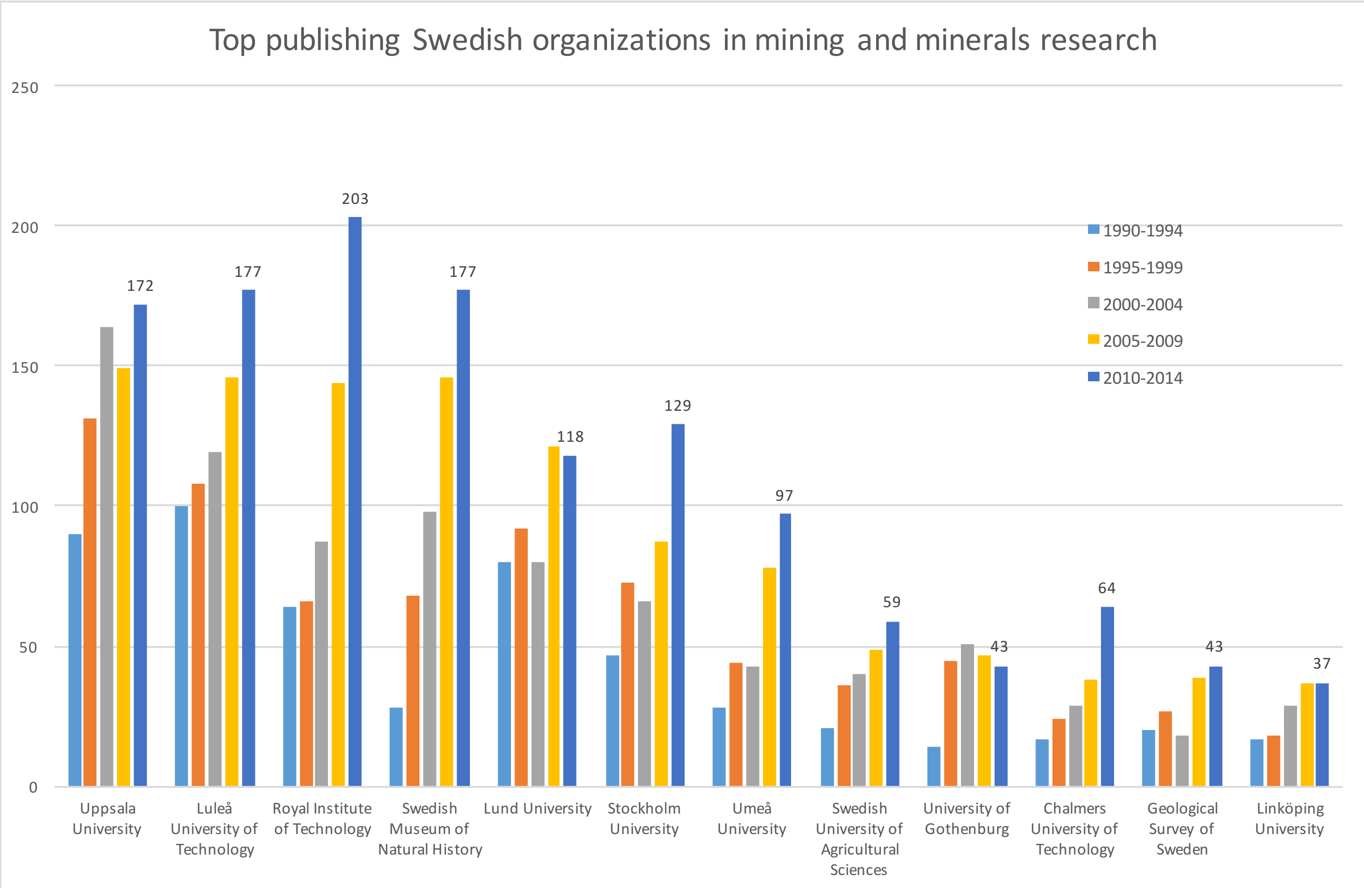



*Fig. 5.9 Top publishing Swedish organizations in mining and minerals research. Publications per 5-year period with the actual number of publications (full counts) given for the last period. Organizations which have published more than 100 publications between 1990 and 2014 are included in the graph. A web presentation of the data for the graph can be found at the address http:/www.kth.se/bibliometrics/projects/mine_and_minerals/Sweden_Level4/all_org.html*

From Fig. 5.9 it can be noted that the main Swedish actors in the mining and minerals field are Uppsala University, Luleå University of Technology, Stockholm University, KTH Royal Institute of Technology (KTH), and the Swedish Museum of Natural History and that all these organizations have increased their publishing in mining and minerals during the last 5-year period. For KTH and Swedish Museum of Natural History we can see a large increase in the publishing within the field over the analysis period.

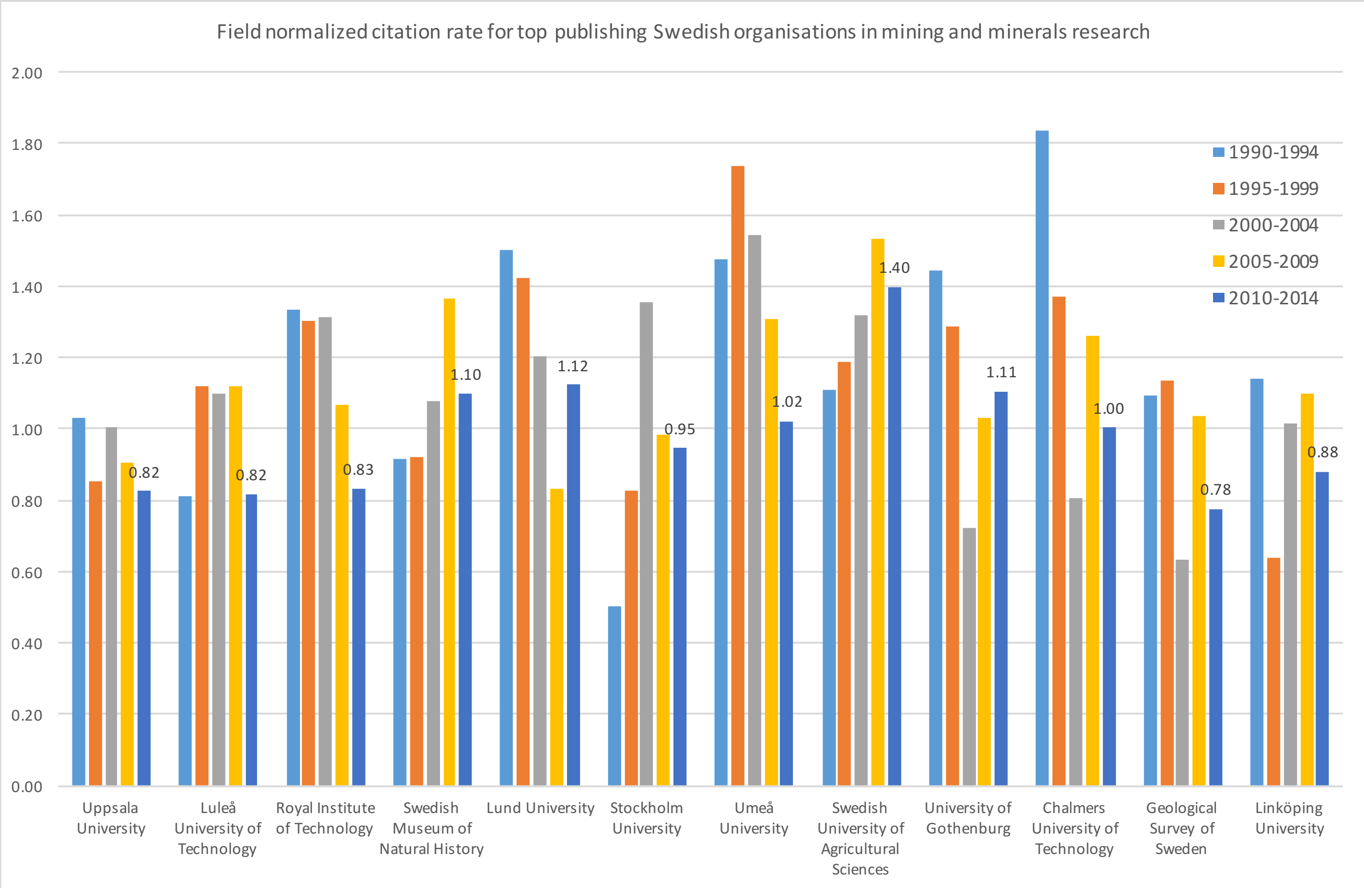



*Fig. 5.10 Field normalized citation rate for top publishing Swedish organizations in mining and minerals research. The normalized citation rate based on address-fractional counts of publications for the last period is given in the diagram. Organizations which have published more than 100 publications between 1990 and 2014 are included in the graph. A web presentation of the data for the graph can be found at the address http:/www.kth.se/bibliometrics/projects/mine_and_minerals/Sweden_Level4/all_org.html*

From Fig. 5.10 it can be noted that even though many Swedish organizations have a normalized citation rate above the world average, the Swedish University of Agricultural Sciences stands out as a producer of high-impact publications in the mining and minerals field. We can also note that the citation rate for mining and minerals publications has fallen for many Swedish organisations since 2000. It should also be noted that the number of publications per 5-year period for University of Gothenburg, Chalmers University of Technology, Geological Survey of Sweden, and Linköping University are below 20 for some of the periods, so these citation values should be interpreted with caution.

## 5.4. Distribution of publications within mining and minerals research over major publication classes

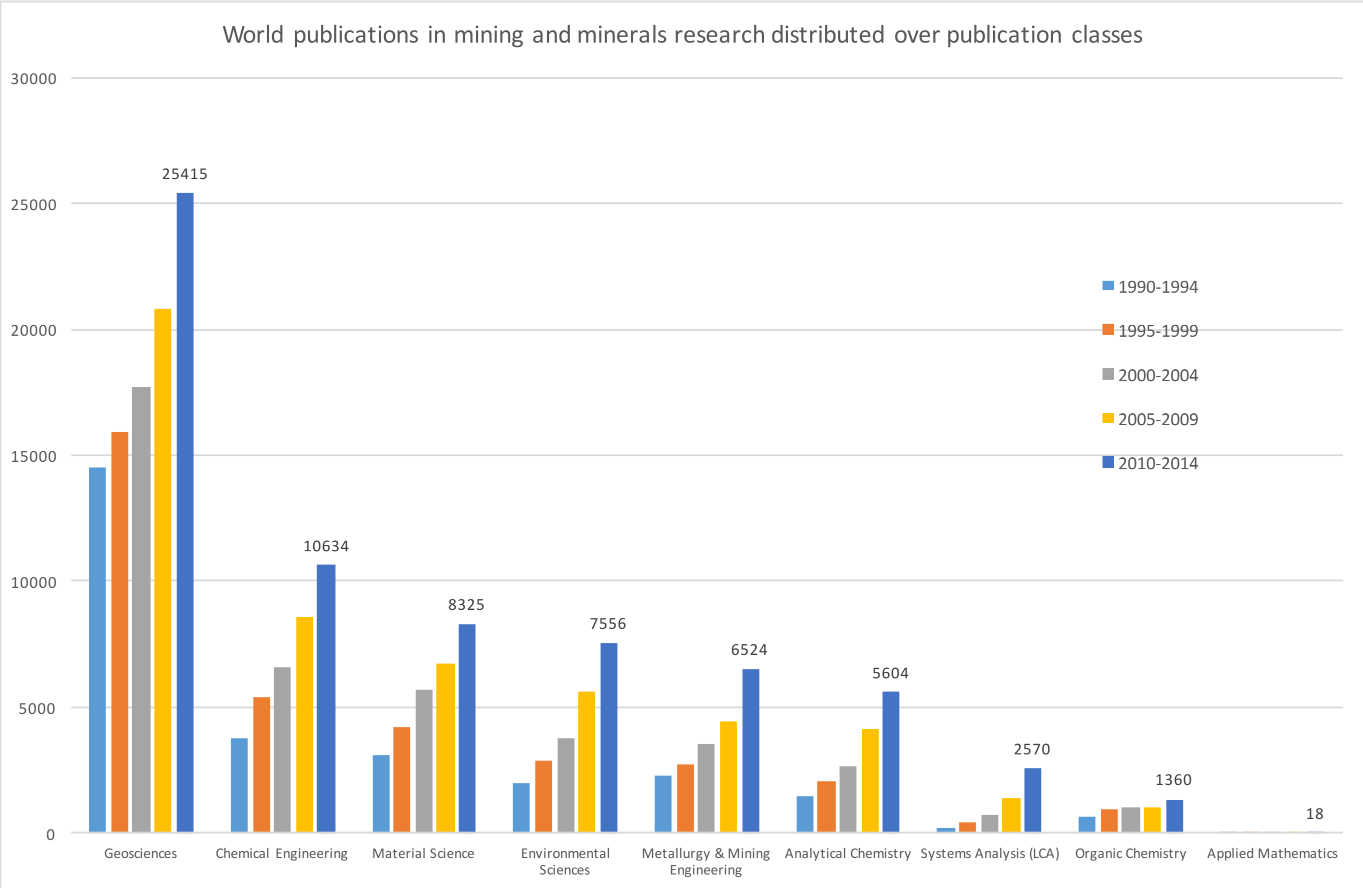



*Fig. 5.11 The number of publications per 5-year period for the whole world for different publication classes at level 4 in the classification system used in this study. LCA stands for Life Cycle Analysis, within the field of Systems Analysis.*

As expected, it can be seen in Fig. 5.11 that the field of Geosciences is dominating for publications within mining and minerals research, and that the number of publications is constantly growing within the field.

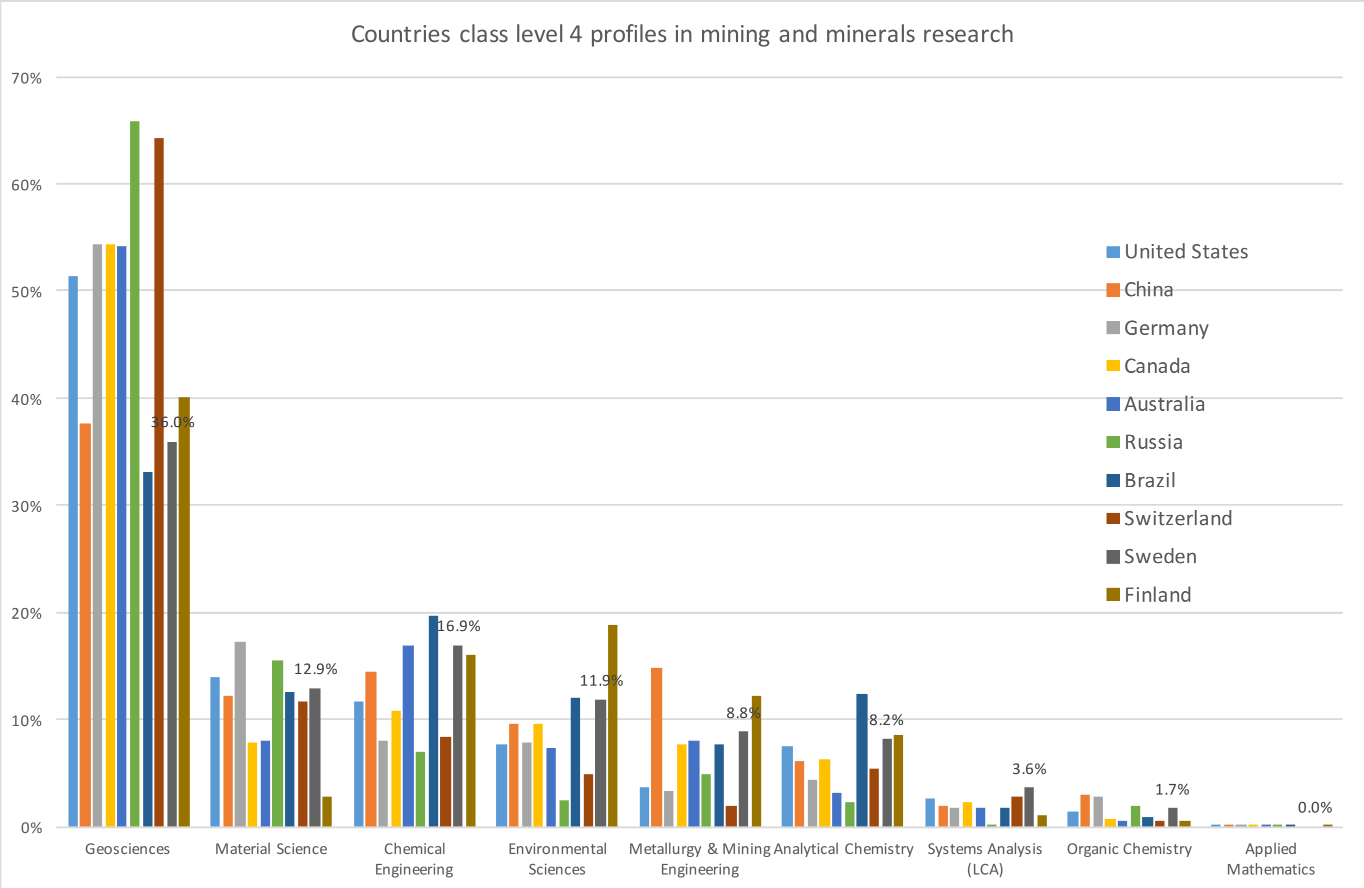



*Fig. 5.12 The selected mining and minerals countries distribution of publications over classes at level 4. The bars represent each country's share in a class of its total number of publications publications in mining and minerals. The actual percentages are given for Sweden. The data behind the graph can be studied in detail at* *http://www.kth.se/bibliometrics/projects/mine_and_minerals/Overall_Level4/all_countries.html*

From Fig. 5.12 it can be seen that the Swedish subject profile in mining and minerals research follows the general country profile quite well. Sweden produces about 5 percent less publications in Geosciences and about 2 percent more publications in Environmental Sciences than the world average. Sweden also stands out as having the largest share of publications in the Systems Analysis class.

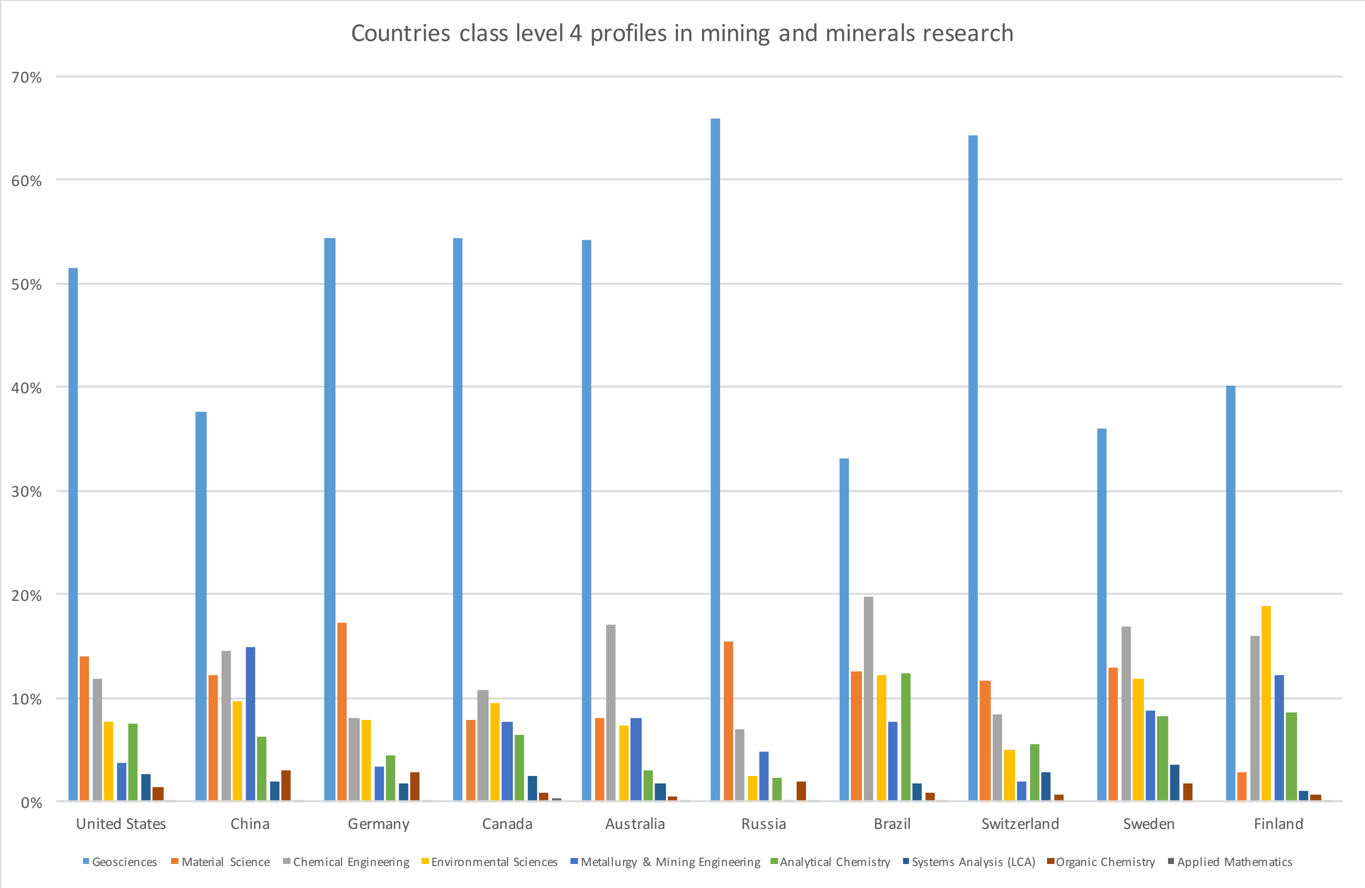



*Fig. 5.13 Country class profiles for the selected mining and minerals countries, as seen per country. The bars represent each country's share in a class of its total number of publications publications in mining and minerals. The data behind the graph can be studied in detail at [http://www.kth.se/bibliometrics/projects/mine_and_minerals/Overall_Level4/all_countries.html](http://www.kth.se/bibliometrics/projects/mine_and_minerals/Overall_Level4/all_countries.html)*

## 5.5. Areas within mining and minerals research where Sweden is strong with regard to publication volume and/or citation impact

If we go down the classification hierarchy and look at individual publication classes selected as relevant for mining and minerals research, we have to decide which aggregation level that is appropriate for a detailed analysis. Since there are 243 classes at the lowest level (1), an analysis of each of these would render a very voluminous report. We have to do some kind of selection in the class hierarchy. A suggestion is to limit the selection to the 90 classes at level 2.

But some aggregated classes of publications at level 2 only consists of one single sub-class at level 1. In this circumstance, it is more appropriate to do the analysis at the most specific level, i.e. level 1. Doing this discrimination of classes for analysis, we get 51 classes at level 1 and 39 aggregated classes at level 2 that are qualified for further analysis.

The next step is to define what is to be considered as "strong" in a bibliometric sense. Since the average publication share for Sweden among the mining and minerals classes as a whole is 1.82 percent, a class where the Swedish presence is above 3 percent units (more than 50 percent above average) can be considered a suitable choice to indicate a strong Swedish presence with regard to publication volume.

Regarding normalized citations, Sweden has a value of 1.15 over all WoS fields, so a value over 1.5 could be considered as high. But in this selection of publications, which is based on classes, the mean citation rate within an analysed class can be higher or lower than the world average for the WoS subject categories the publications belong to. Hence the normalized citation average for the whole class can differ from 1. An example of this can be seen in Fig. 5.4 where the normalized citation value for publications from all countries in the mining and minerals classes is below 1 in the end of the series.

We have therefore introduced a class-normalized indicator for the normalized citation rate which we call the *class normalized citation rate* ($c_{fcn}$). This indicator is the ordinary field normalized $c_f$ value for a country in a class in relation to the $c_f$ value for the whole class. If a country has a class normalized citation rate above 1.2 (more than 20% above the class mean), we consider its citation performance to be strong. If the class normalized citation rate is below 0.8 (lower than 80% of the class mean), the citation-based bibliometric performance can be considered to be weak.

For the share of top 10 percent most cited publications, the same reasoning holds true. The whole class share of publications among the 10 percent most cited can be above or below 10 percent. So we introduce an indicator we call the *class normalized share of top 10 percent most cited publications* ($top10_{cn}$). This indicator is the number of publications (full counts) for a country in a class that is among the 10 percent most cited in the field, divided by the number of publications among the 10 percent most cited in the field for the whole class, and then normalized (divided) by the total share of publications for the country in the class.[22] The result is a normalized number, where a value above 1 is a top10 share above the class average and a value below 1 is below class average. We then consider a $top10_{cn}$ value above 1.2 to indicate a strong top10 citation performance and a value below 0.8 to be a weak top10 citation performance. It is worth noting that for this indicator, the numbers both in the numerator and the denominator can sometimes be very small, so the result can be quite instable and dependent on a few highly-cited publications in the dataset.

If we make a selection of publication classes where Sweden can be considered bibliometrically strong in accordance with the above suggested boundaries, the following criteria can be used[23]:

- More than 50 publications in years 1990-2014 (i.e. on average more than 2 publications per year, to get sufficient stable citation values for the 5-year intervals)

AND

- A class share of publications > 3 %
- OR a class normalized citation impact > 1.2
- OR a class normalized share of publications among 10 % most cited > 1.2

[22] Top10% class normalized = Ptop10 (country) / Ptop10 (class) / Pshare (country). P stands for Publications.

[23] P > 50 AND (Pshare > 3% OR $C_{fcn}$ > 1.2 OR $top10_{cn}$ > 1.2)

Using this selection criteria, we find the following 14 "strong" classes as candidates to investigate (sorted under each heading in descending order according to the Swedish publication share in the class):

| Class ID | Statistically generated class label | P | P share | $c_f$ | top10% | cf class norm. | top10 class norm. |
|---|---|---|---|---|---|---|---|
| | **Classes with both high Swedish publication volume *and* high citation impact** | | | | | | |
| L2:980 | High Pressure Research<br>(Shock Wave Detonat Phys / High Pressure Research / Sci Extreme Condit) | 132 | **6.6%** | 1.50 | 25.0% | **1.26** | **1.87** |
| L1:5235 | Lead Isotopes<br>(Lead Isotopes / Stable Lead Isotopes / Unite Format Rech Phys) | 83 | **6.1%** | 1.35 | 15.8% | **1.35** | **1.69** |
| L2:1308 | Mercury<br>(Mercury / Methyl Mercury / Atmospheric Mercury) | 149 | **4.8%** | 1.32 | 17.7% | 1.07 | **1.39** |
| | **Classes with high Swedish publication volume, but average or low citation impact** | | | | | | |
| L1:2594 | Surface Complexation<br>(Surface Complexation / Theoret Problems Adsorpt / Surface Complexation Model) | 135 | **7.5%** | 1.27 | 12.8% | 1.03 | 0.91 |
| L1:8055 | Tundish<br>(Tundish / Continuous Casting / Electromagnetic Brake) | 54 | **4.9%** | 0.73 | 2.2% | 0.76 | 0.21 |
| L2:2624 | Acid Mine Drainage<br>(Acid Mine Drainage / Acid Sulfate Soil / Acid Sulfate Soils) | 133 | **4.9%** | 0.74 | 1.8% | 0.89 | 0.26 |
| L1:3185 | Impact Cratering<br>(Meteoritics & Planetary Science / Impact Cratering Grp / Shock Metamorphism) | 77 | **4.7%** | 0.75 | 2.2% | 1.04 | 0.53 |
| L2:2315 | Blast Furnace<br>(Blast Furnace / Ironmaking / Iron Ore Sinter) | 60 | **3.5%** | 0.70 | 1.9% | 0.96 | 0.34 |
| L2:1146 | Resources Conservation and Recycling<br>(International Journal of Life Cycle Assessment / Resources Conservation and Recycling / Remanufacturing) | 123 | **3.0%** | 1.16 | 10.3% | 0.87 | 0.66 |
| | **Classes with Swedish high citation impact, but average or low publication volume** | | | | | | |
| L2:981 | Minerals Engineering<br>(Minerals Engineering / International Journal of Mineral Processing / Froth Flotation) | 240 | 2.9% | 0.98 | 7.7% | **1.27** | **1.38** |
| L2:779 | Magnetotellurics<br>(Magnetotellurics / Geophysics / Journal of Applied Geophysics) | 116 | 2.8% | 0.87 | 7.1% | **1.14** | **1.28** |
| L2:106 | Heavy Metals<br>(Heavy Metals / Phytoremediation / Phytoextraction) | 159 | 2.0% | 1.64 | 20.9% | **1.41** | **1.83** |
| L2:783 | Solvent Extraction and Ion Exchange<br>(Solvent Extraction and Ion Exchange / Solvent Extraction / Supported Liquid Membrane) | 55 | 1.5% | 1.55 | 16.3% | **1.72** | **2.99** |
| L2:577 | Hydrometallurgy<br>(Hydrometallurgy / Bioleaching / Acidithiobacillus Ferrooxidans) | 141 | 1.3% | 1.32 | 15.4% | **1.44** | **1.70** |

*Table 5.1 Publication classes where Sweden has more than 50 publications (full counts) during the years 1990-2014 and more than 3 percent share of publications, or a class normalized $c_f$ value above 1.2, or a class normalized top10% value above 1.2, ordered in descending order after the Swedish share of publications in the class for each group. The most describing part of the automatic label has been extracted and used as leading label. The numbers that made the class to be included among Swedish strong classes is marked in bold.*

The headings for Table 5.1 to Table 5.2 should be interpreted as follows:

- **Class ID**: A combination of level (L) and class number

- **Statistically generated class label**: A class label that has been constructed though statistical processing of WoS field names, journal titles, keywords and author address parts. The most describing part of the automatic label has been extracted and used as leading label.
- **P**: The number of Swedish publications in the class during the years 1990-2014.
- **P share**: The Swedish publications' share of the class.
- $c_f$: The field normalized citation rate for Swedish publications in the class. Address fractionalized.
- **top10%**: The share of Swedish publications among the 10 % most cited in the WoS field. Address fractionalized.
- $c_f$ **class norm.**: The field normalized citation rate for Swedish publications in the class, normalized (divided by) against the field normalized citation rate for all publications in the class. Address fractionalized.
- **top10% class norm.**: The Swedish number of publications among the top 10% most cited in the WoS field, normalized against the number of publications among the top 10% most cited for the whole class. Full publication counts.

It is worth noting that some of the classes considered as strong in publication volume here are weak in citation impact, which means that they also will turn up as weak areas in the analysis further down, and conversely some classes at the bottom of the list where Swedish publication volume is small are still considered to be strong due to the high citation rate. The knowledgeable interpreter has to decide if volume or impact is the measure to look for when identifying a strong class in the table above.

If we strengthen the conditions for the Swedish publishing to be considered strong to fulfil *all* the above stated criteria at the same time, i.e. more than 50 publications AND a share above 3% AND a class normalized citation rate above 1.2 OR a class normalized top10% above 1.2, we get a short list of the two publication classes at the top of Table 5.1; *High Pressure Research* and *Lead Isotopes*, tightly followed by the class *Mercury*, where both publication volume is high and share of top 10% most cited is high, but the mean citation rate is only slightly above average.

## 5.6. Areas within mining and minerals research where Sweden is weak with regard to publication volume or citation impact

The assignment from the government also mentions that it would be desirable to identify areas where Swedish research in mining and minerals could be considered "weak".

One suggestion for a selection would be to look in more detail at classes where Sweden fulfils the following criteria[24]:

- More than 50 publications in years 1990-2014 (i.e. on average more than 2 publications per year, to get a sufficient stable citation value)

AND

- A class share of publications < 1 %
- OR a class normalized citation impact < 0.8

[24] P > 50 AND (Pshare < 1% OR $C_{fcn}$ < 0.8 OR $top10_{cn}$ < 0.8)

- OR a class normalized share of publications among 10 % most cited < 0.8

Following the above criteria and doing a selection from publication classes, we get 7 classes where Sweden could be considered as having a weak bibliometric presence:

| Class ID | Statistically generated class label | P | P share | $c_f$ | top10% | $c_f$ class norm. | top10 class norm. |
|---|---|---|---|---|---|---|---|
| | **Classes with high Swedish publication volume, but low citation impact** | | | | | | |
| L2:2624 | Acid Mine Drainage (Acid Mine Drainage / Acid Sulfate Soil / Acid Sulfate Soils) | 133 | 4.9% | 0.74 | 1.8% | 0.89 | **0.26** |
| L1:8055 | Tundish (Tundish / Continuous Casting / Electromagnetic Brake) | 54 | 4.9% | 0.73 | 2.2% | **0.76** | **0.21** |
| L1:3185 | Impact Cratering (Meteoritics & Planetary Science / Impact Cratering Grp / Shock Metamorphism) | 77 | 4.7% | 0.75 | 2.2% | 1.04 | **0.53** |
| L2:2315 | Blast Furnace (Blast Furnace / Ironmaking / Iron Ore Sinter) | 60 | 3.5% | 0.70 | 1.9% | 0.96 | **0.34** |
| L2:1146 | Resources Conservation and Recycling (International Journal of Life Cycle Assessment / Resources Conservation and Recycling / Remanufacturing) | 123 | 3.0% | 1.16 | 10.3% | 0.87 | **0.66** |
| L2:471 | Mineralogy (Mineralogy / American Mineralogist / Physics and Chemistry of Minerals) | 228 | 2.6% | 0.68 | 4.7% | **0.76** | **0.79** |
| | **Class with low Swedish publication volume *and* low citation impact** | | | | | | |
| L2:349 | Kimberlite (Kimberlite / Journal of Petrology / Carbonatite) | 84 | **0.7%** | 0.73 | 3.5% | **0.62** | **0.39** |

*Table 5.2 Publication classes where Sweden has more than 50 publications during the years 1990-2014 and less than 1 percent share of publications in the class, or a class normalized $c_f$ value below 0.8, or a class normalized top10% value below 0.8, ordered in descending order after the Swedish share of publications in the class. The most describing part of the automatic label has been extracted and used as leading label.*

From Table 5.2 we can see that all publication classes except the last two in the list are classes where Sweden can be considered strong with regard to publication volume (share of class publications), but get low return in citation impact, hence considered to be weak classes here. Once again, it is up to the interpreter to decide if volume or impact is the measure to look for when identifying a weak class in the table above.

If we strengthen the conditions for a weak class to fulfil *both* the criteria for small publication volume and low citation impact, we only get the publication class labelled *Kimberlite* left.

## 5.7. Visualizations of bibliometric networks

Interactive visualisations for relations between publication classes and Swedish authors can be reached from links at the project's start page at the web pages of KTH bibliometrics group.[25]

All visualizations are delimited to the chosen publication classes at level 1 and the publication years 1990-2014. The open-source program Gephi has been used for the visualizations and Force Atlas 2 algorithm has been used for the layouts.

### *5.7.1. Visualization of relations between publication classes within mining and minerals research*

This visualization shows a network of publication classes within mining and minerals research and their relations to each other.[26] Each *node* (dot) in the visualization represents a publication class. The size of each node is relative to the number of mining and minerals publications belonging to the node. Only publications within the classes at level 1 chosen for the analysis have been included and counted as mining and minerals publications.

At the top level, the links between the nodes express citation relations between the classes as whole. Links between classes on different levels are hierarchical so that a class at level x has one relation to the level above and the weight of its sub classes (if any) adds up to 1. For other levels than the top level, the visualization does not express relations between classes at the same level.

The nodes representing the publication classes at the different hierarchical levels have been coloured according to their respective level. The large classes at the top level 4 are blue, classes at level 3 are orange, classes at level 2 are green, and the smallest classes at the bottom level 1 are coloured purple.

An interactive web presentation of the relations between the publication classes can be found at address http://www.kth.se/bibliometrics/projects/mine_and_minerals/network_total_MM_publ/

[25] http:/kth.se/bibliometrics/projects/mine_and_minerals/

[26] http:/kth.se/bibliometrics/projects/mine_and_minerals/network_total_MM_publ/

### 5.7.2. *Visualization of Swedish publications within mining and minerals research publication classes*

This visualization at the project web site is based on the same data as above.[27] However, in this visualization the node sizes are set according to the number of Swedish mining and minerals publications in each class.

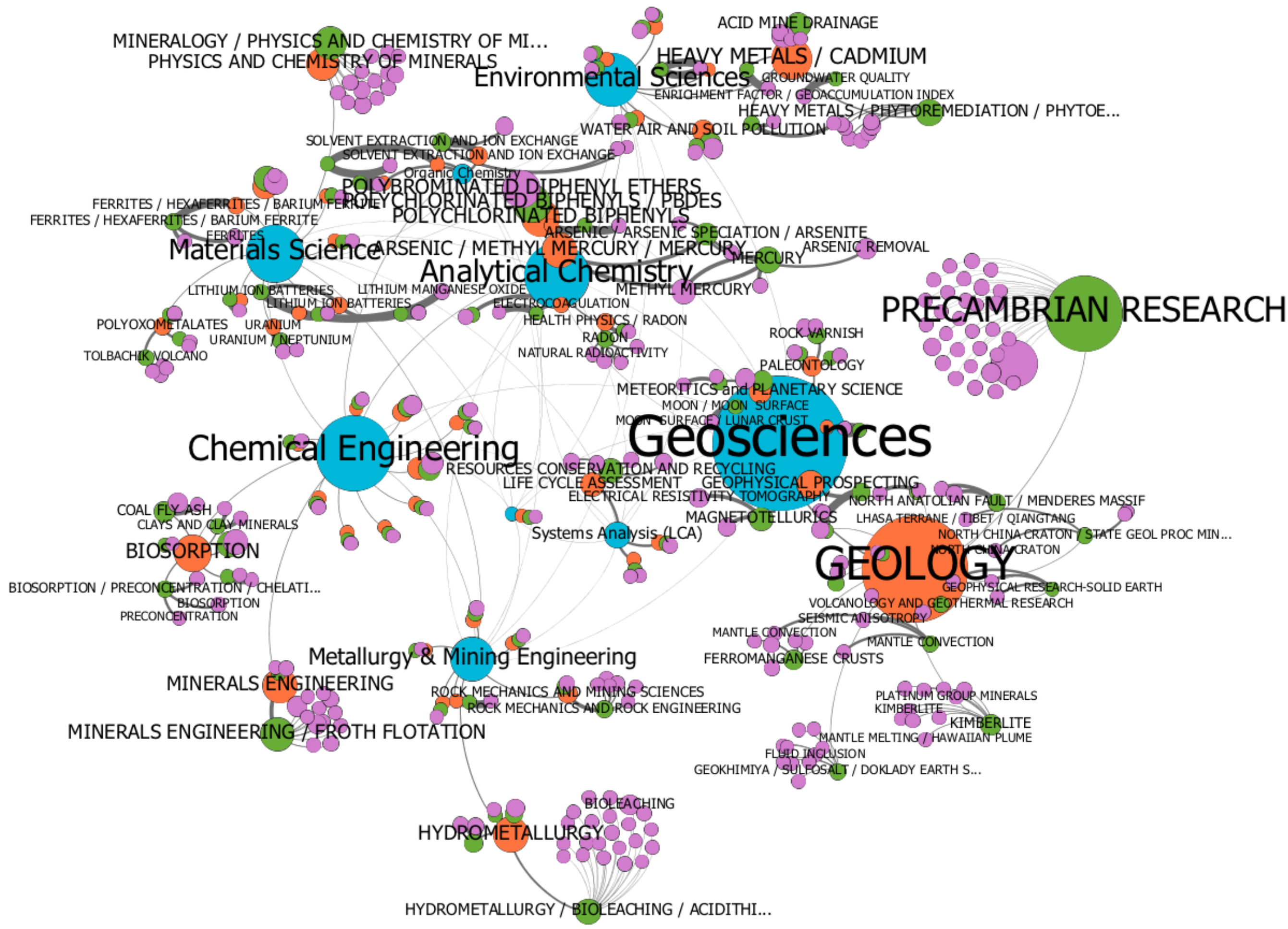



*Fig. 5.14 Visualization of relations network between Swedish publications within mining and minerals research. Node sizes represent the number of Swedish publications in mining and minerals and node colours represent the class level in the publication clustering. An interactive web presentation of the network can be seen at address* *http:/kth.se/bibliometrics/projects/mine_and_minerals/network_swe_MM_publ/*

---

[27] http:/kth.se/bibliometrics/projects/mine_and_minerals/network_swe_MM_publ/

In the interactive web visualization, the colours of the nodes are set according to according to the share of Swedish publications in each class (red signifies a high share and blue a low share).

### *5.7.3. Co-publishing between Swedish researchers within mining and minerals research*

The network in this visualization represents co-publishing relations between Swedish researchers in mining and minerals research as defined by the selected publication classes in this study.[28] All researchers publishing within the selected mining and minerals publication classes with Sweden as their main country of affiliation has been included in the graph. Node sizes are set to fractionalized number of publications. Links between classes express co-publishing. Each co-publication has been weighted according to the number of authors in the publication so that the contribution to the link's total weight is 1/no. of authors for each co-publication. The weight of a link between two nodes is the sum of the co-publication weights.

Authors have been grouped together based on the level of co-publishing and the groups have been coloured using different colours to indicate possible Swedish "research environments". An interactive web presentation of the network can be seen at the address
http:/www.kth.se/bibliometrics/projects/mine_and_minerals/network_swe_copub/

### *5.7.4. Relations between Swedish researchers and publication classes within mining and minerals research*

This network expresses relations between Swedish researchers and level 2 publication classes in the mining and minerals research dataset.[29] There are two types of nodes in the network: 1. Publication classes and 2. Researchers. The network is delimited to Swedish researchers (same definition as in the co-publishing network).

Node sizes express either the number of publications in the class (for publication class nodes) or the number of publications authored by a researcher (for researcher nodes). The node size is relative to other nodes of the same type. Hence, node sizes of researcher nodes should not be compared to the node sizes of the publication class nodes. Links between nodes expresses number of publications by a researcher that belong to a publication class. In this network there are only relations (links) between authors and classes, not between author-author or between class-class.

---

28 http:/www.kth.se/bibliometrics/projects/mine_and_minerals/network_swe_copub/

29 http:/www.kth.se/bibliometrics/projects/mine_and_minerals/network_swe_rel_auth_class/

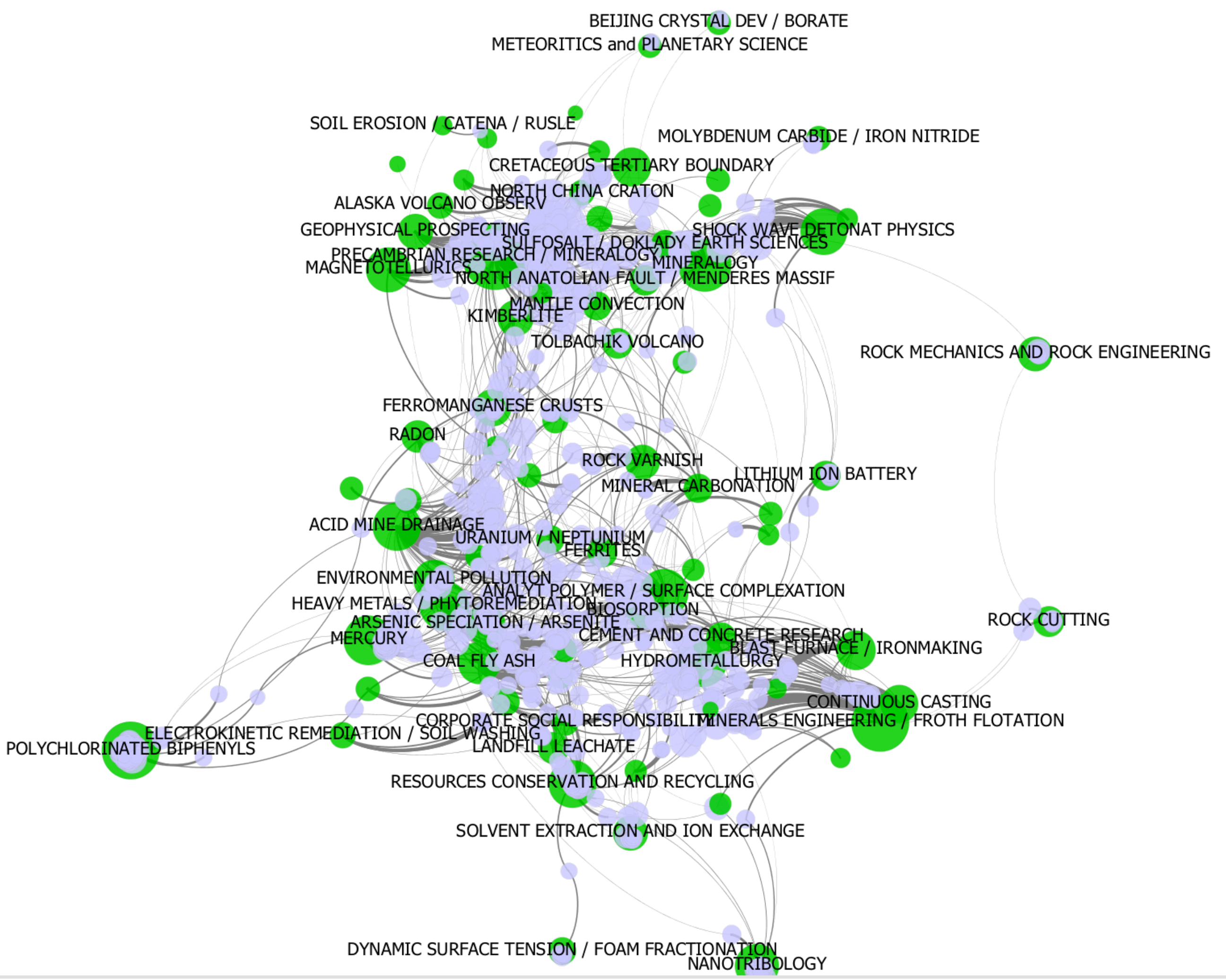



*Fig. 5.15 Visualization of relations between Swedish researchers and publication classes within mining and minerals research. Nodes representing publication classes are green and nodes representing researchers are blue. Node sizes are set in proportion to the number of publications in each class. Only the largest publication classes have labels. An interactive web presentation of the network can be seen at the address [http:/www.kth.se/bibliometrics/projects/mine_and_minerals/network_swe_rel_auth_class/](http:/www.kth.se/bibliometrics/projects/mine_and_minerals/network_swe_rel_auth_class/).*

## 5.8. Swedish research environments and networks

An identification of research environments is dependent on a visual and qualitative analysis of the co-publishing between Swedish researchers in mining and minerals and the relation of their publications to the labels of the publication classes that can be seen in Fig. 5.15.

This analysis should preferably be done by subject experts using the interactive visualizations "Co-publishing between Swedish researchers within mining and minerals research"[30] and " Relations between

[30] http:/www.kth.se/bibliometrics/projects/mine_and_minerals/network_swe_copub/

Swedish researchers and publication classes within mining and minerals research"[31] on the project website.

## 5.9. Comparing the class-based study with the Swedish Research Council keyword-based study

The newly developed method using clustered publication classes has some main advantages:

- The possibility to catch publications which are not retrieved using keywords, due to synonyms and problems with stemming (plurals, etc.) of words.
- The possibility to focus the analysis to narrower areas of research through the hierarchical grouping of publications in classes.
- The possibility to zoom out through the grouping of publication classes into groups at higher levels in the hierarchical classification.

The class-based method also comes with some drawbacks:

- It is not always certain that the method using clustering by following citation gives the desired grouping of publications. Citations may be done by other reasons than mere subject-based linking.
- The subject-content and focus of publications within in a given class may change over time, due to re-clustering, and may not always correspond to a subject term trying describe the content of the class.
- Comparing normalized citation levels in a class-based publication set can be complicated, since the standard method for normalization of citations is based on WoS subject classes, and not clustered publication classes.
- Since the publication classes are dynamic, due to changing citations, comparisons over time of class-based analyses can be difficult to perform.
- The labelling of the classes exhibits some challenges and needs to be developed further. When classes at lower levels (1 or 2) are aggregated to higher levels (2 or 3), sometimes the labels at the higher level can be misleading regarding the content of the lower classes. This effect is especially significant when a single class at level 1 is "aggregated" into a class at a higher level.

When comparing the present class-based study to the previous SRC keyword-based study, the conclusion is that the present study has favoured recall over precision, whereas the keyword-based study favoured precision over recall. In practice, this means that the present study did capture about four times the publications in the SRC study; 225 366 publications, compared to 53 037 publications in the SRC study. Looking at the numbers for Swedish publications, this study did capture 4118 publications compared to the SRC study's 1100 publications.

The larger number of publications in this study means that there is a risk that the publications in some classes are dealing with subjects that might be considered off-topic regarding mining and minerals, since the clustered publication classes can have diffuse boundaries that can go out of scope.

The selected 90 classes in this study covers 976 of the 1100 Swedish publications in the SRC study, which means that about 10% of the keyword-selected publications is missing, since these were being clustered into publication classes that not were selected in the semi-manual class selection process.

---

[31] http:/www.kth.se/bibliometrics/projects/mine_and_minerals/network_swe_rel_auth_class/

Table 5.3 shows how many more publications that were covered by the class-based study in relation to the keyword-based study for the publication classes where the keyword-publications were more than 10% of the class publications.

| Level: Class ID | Label | Class-selected publications | Keyword-selected publications | Difference class-keyword-selected | Keyword publications share of class |
|---|---|---|---|---|---|
| L2:2423 | Mathematical Geology (Mathematical Geology / Multiple Point Statistics / Training Image) | 397 | 240 | 157 | 60% |
| L2:503 | Organic Geochemistry (Organic Geochemistry / Petr Explorat Dev / Aapg Bulletin) | 975 | 536 | 439 | 55% |
| L2:1652 | Dynamic Surface Tension (Dynamic Surface Tension / Foam Fractionation / Dissolved Air Flotation) | 876 | 413 | 463 | 47% |
| L2:981 | Minerals Engineering (Minerals Engineering / International Journal of Mineral Processing / Froth Flotation) | 8146 | 3407 | 4739 | 42% |
| L2:2624 | Acid Mine Drainage (Acid Mine Drainage / Acid Sulfate Soil / Acid Sulfate Soils) | 2732 | 1084 | 1648 | 40% |
| L2:1012 | Geokhimiya (Geokhimiya / Sulfosalt / Doklady Earth Sciences) | 6785 | 2522 | 4263 | 37% |
| L2:945 | Meteoritics & Planetary Science (Meteoritics & Planetary Science / Meteoritics / Geochimica Et Cosmochimica Acta) | 1462 | 496 | 966 | 34% |
| L1:33277 | Inco Limited (Inco Limited / Remote Mining Technology / Solutions of Acrylic Acid) | 75 | 24 | 51 | 32% |
| L2:1146 | International Journal of Life Cycle Assessment (International Journal of Life Cycle Assessment / Resources Conservation and Recycling / Remanufacturing) | 4068 | 1271 | 2797 | 31% |
| L2:428 | Acta Petrologica Sinica (Acta Petrologica Sinica / North China Craton / State Geol Proc Mineral Ources) | 6820 | 2092 | 4728 | 31% |
| L2:1438 | Inorgan Mat Program (Inorgan Mat Program / New Mineral / Tolbachik Volcano) | 2199 | 596 | 1603 | 27% |
| L1:7415 | Mine Soils (Mine Soils / Reclamation / Reclaimed Mine Soil) | 902 | 230 | 672 | 25% |
| L2:349 | Kimberlite (Kimberlite / Journal of Petrology / Carbonatite) | 12275 | 2823 | 9452 | 23% |
| L2:298 | Journal of Volcanology and Geothermal Research (Journal of Volcanology and Geothermal Research / Bulletin of Volcanology / Alaska Volcano Observ) | 2118 | 476 | 1642 | 22% |
| L1:11025 | Artisanal and Small Scale Mining Asm (Artisanal and Small Scale Mining Asm / Artisanal Mining / Small Scale Mining) | 856 | 186 | 670 | 22% |
| L1:22359 | Solid State Recycling (Solid State Recycling / Solid Recycling Process / Machined Chips) | 275 | 59 | 216 | 21% |
| L2:444 | International Journal of Rock Mechanics and Mining Sciences (International Journal of Rock Mechanics and Mining Sciences / Rock Mechanics and Rock Engineering / Engineering, Geological) | 2548 | 505 | 2043 | 20% |
| L2:471 | Mineralogy (Mineralogy / American Mineralogist / Physics and Chemistry of Minerals) | 8667 | 1704 | 6963 | 20% |
| L2:2361 | Nasicon (Nasicon / Ets 10 / Ets 4) | 1450 | 261 | 1189 | 18% |
| L2:577 | Hydrometallurgy (Hydrometallurgy / Bioleaching / Acidithiobacillus Ferrooxidans) | 10590 | 1875 | 8715 | 18% |
| L2:280 | Journal of Structural Geology (Journal of Structural Geology / Tectonophysics / Bulletin De La Societe Geologique De France) | 1993 | 325 | 1668 | 16% |
| L2:8 | Precambrian Research (Precambrian Research / Journal of Metamorphic Geology / Mineralogy) | 26356 | 4221 | 22135 | 16% |
| L1:19707 | Steel Slag (Steel Slag / Ladle Furnace Slag / Expansive Compounds) | 373 | 57 | 316 | 15% |

| | | | | | |
|---|---|---|---|---|---|
| L2:2862 | Journal of Geochemical Exploration (Journal of Geochemical Exploration / Geochemistry-Exploration Environment Analysis / Compositional Data) | 1493 | 226 | 1267 | 15% |
| L1:29023 | Strontium Carbonate (Strontium Carbonate / Interface Coupled Dissolution Reprecipitation / Violarite) | 143 | 21 | 122 | 15% |
| L2:117 | New Zealand Journal of Geology and Geophysics (New Zealand Journal of Geology and Geophysics / Journal of Geophysical Research-Solid Earth / Bulletin of The Seismological Society of America) | 3528 | 501 | 3027 | 14% |
| L1:21606 | Ludwigite (Ludwigite / Borate Oxide / Ludwigites) | 278 | 39 | 239 | 14% |
| L1:18608 | Controlled Source Seismol Grp (Controlled Source Seismol Grp / Crooked Line / Sez Geol Plicata Geofis Plicata) | 394 | 54 | 340 | 14% |
| L1:6561 | Sulfate Reduction (Sulfate Reduction / Sulfate Reducing Bacteria / Sulphate Reducing Bacteria) | 1230 | 165 | 1065 | 13% |
| L2:1223 | Rock Varnish (Rock Varnish / Geoderma / Earth Environm Sci Geol Geophys) | 2095 | 278 | 1817 | 13% |
| L1:18589 | Bioleaching (Bioleaching / Thiobacilli / Bio Acidification) | 364 | 47 | 317 | 13% |
| L2:1624 | Geol Engn (Geol Engn / North Anatolian Fault / Menderes Massif) | 2996 | 378 | 2618 | 13% |
| L1:19220 | Computat Geosci (Computat Geosci / Carbonate Acidizing / Explorat Min) | 382 | 47 | 335 | 12% |
| L1:23318 | Crichtonite Group (Crichtonite Group / Senaite / Crichtonite) | 158 | 19 | 139 | 12% |
| L1:23332 | Gamma Si3n4 (Gamma Si3n4 / Tin Nitride / Cubic Silicon Nitride) | 263 | 31 | 232 | 12% |
| L1:19772 | San Benito County (San Benito County / Clear Creek Mine / Mercury Minerals) | 287 | 32 | 255 | 11% |
| L1:9020 | Electrochemical Reactors (Electrochemical Reactors / Reactive Electrodialysis / Solid Phase Conductivity) | 727 | 79 | 648 | 11% |
| L1:2450 | Central Andes (Central Andes / Sierras Pampeanas / Journal of South American Earth Sciences) | 1842 | 196 | 1646 | 11% |

*Table 5.3 Selected mining and minerals publication classes where the SRC keyword selected publications have a larger share than 10 %, listed in descending order on the share of keyword-selected publications in relation to the class-selected publications.*

# 6 Discussion and conclusion

The aim of this bibliometric study has been to put a new publication classification method to test and to see if it is possible to use the method to produce meaningful results. The method shows some promising characteristics, as the possibility to retrieve and group publications on a non-textual basis, and the possibilities to classify and limit the analysis to focused research areas.

During the project, a number of challenges regarding the method has been revealed. The largest challenges involve the description of the research topics of each article class and also boundary effects in the clustering and classification procedure that may lead to some articles in a class as being considered off-topic for the study. For instance, in some classes where topics regarding the handling of substances in urban mining was found to drift too much into environmental research and the classes had to be removed to reduce noise in the study.

One might argue that the method of following references and citations to cluster publications into classes in some sense reveals the inherent complexity of science and scientific publishing, and that the urge to divide research into distinct subject areas is a somewhat over-simplified construction. Sometimes citations pass subject boundaries in a way that is unexpected when just looking from above using a subject-focused view. During the project, we sometimes had to study the classes at article level, and the citation links between the publications usually made good sense at the micro level.

In cases where publication classes show a risk of having publications that are off-topic, a more text-based selection or exclusion could be combined with the clustering technique, so that classes that risk drifting out of scope can be limited to the part of the class that is relevant to the subject at study. Another possibility to exclude non-relevant publications could be to combine the class-based selection of publications with the WoS subject categories, so that publications in journals that not are considered to belong to the right subject areas are excluded from the class-based publication set.

A suggestion for improvement of the present method can be to start with a set of keywords and then retrieve new suitable keywords through statistical analysis of the publication classes matched by the initial set of keywords. The keyword-based publication retrieval could then be re-done with the new set of keywords. This would then mean that the classification method is used as an intermediate step in an otherwise keywords-based study. The iterative interaction between keyword searches and class-based class selection can also be used as a starting point to get better statistical data for selecting classes for the final analysis.

In conclusion, the class-based classification methodology brings forward some new promising possibilities, but also makes labelling, evaluation and use of the results more complex and uncertain. The methods for selection and description of publication classes need to be further developed and the results verified by subject experts, and the method could also be improved further by using text-based selection of publications in some classes.

## 7 References


Ahlgren, Per, & Sjögårde, Peter. (2015). *Formal definitions of field normalized citation indicators and their implementation at KTH Royal Institute of Technology*. Retrieved from http://www.kth.se/polopoly_fs/1.544479!/Formal definitions of field normalized citation indicators at KTH.pdf

Caron, E., & van Eck, N.J. (2014). *Large scale author name disambiguation using rule-based scoring and clustering*. Paper presented at the Proceedings of the 19th International Conference on Science and Technology Indicators.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*(2). doi:10.1103/PhysRevE.69.026113

Vetenskapsrådet. (2015). *Svar på uppdraget att genomföra en bibliometrisk utvärdering av gruv- och mineralforskningsområdet i Sverige (U2015/1361/F)*. Retrieved from http://www.vr.se/omvetenskapsradet/regeringsuppdrag.4.12fff4451215cbd83e4800020953.html

Visser, M.S., & Nederhof, A.J. (2011). *Appendix II to "Bibliometric Analysis of Uppsala University 2007–2010"*. Retrieved from http://uu.diva-portal.org/smash/get/diva2:461235/ATTACHMENT02.pdf

Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology, 63*(12), 2378-2392. doi:10.1002/asi.22748

Waltman, L., & van Eck, N. J. (2013). A smart local moving algorithm for large-scale modularity-based community detection. *European Physical Journal B, 86*(11). doi:10.1140/epjb/e2013-40829-0