

A geostatistical analysis of geostatistics

TOMISLAV HENGL,^a BUDIMAN MINASNY,^b MICHAEL GOULD^c

^a *Institute for Biodiversity and Ecosystem Dynamics,
University of Amsterdam,
Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands*
^b *Faculty of Agriculture, Food & Natural Resources,
The University of Sydney, Sydney, Australia*
^c *Michael Gould Associates, Arnhem, The Netherlands*

The bibliometric indices of the scientific field of geostatistics were analyzed using statistical and spatial data analysis. The publications and their citation statistics were obtained from the Web of Science (4000 most relevant), Scopus (2000 most relevant) and Google Scholar (5389). The focus was on the analysis of the citation rate (CR), i.e. number of citations an author or a library item receives on average per year. This was the main criterion used to analyze global trends in geostatistics and to extract the Top 25 most-cited lists of the research articles and books in geostatistics. It was discovered that the average citation rate for geostatisticians has stabilized since 1999, while the authors' *n*-index seems to have declined ever since. One reason for this may be because there are more and more young authors with a lower *n*-index. We also found that the number of publications an author publishes explains only 60% of the variation in the citation statistics and that this number progressively declines for an author with a lower number of publications. Once the geographic location is attached to a selection of articles, an isotropic Gaussian kernel smoother weighted by the CR can be used to map scientific excellence around the world. This revealed clusters of scientific excellence around locations such as Wageningen, London, Utrecht, Hampshire, UK, Norwich, Paris, Louvain, Barcelona, and Zürich (Europe); Stanford, Ann Arbor, Tucson, Corvallis, Seattle, Boulder, Montreal, Baltimore, Durham, Santa Barbara and Los Angeles (North America); and Canberra, Melbourne, Sydney, Santiago (Chile), Taipei, and Beijing (other continents). Further correlation with socio-economic variables showed that the spatial distribution of CRs in geostatistics is independent of the night light image (which represents economic activity) and population density. This study demonstrates that the commercial scientific indexing companies could enhance their service by assigning the geographical location to library items to allow spatial exploration and analysis of bibliometric indices.

Introduction

Geostatistics has its origins as a scientific field in the work of YODEN [1951], MATHERON [1962] and GANDIN [1963]. It was later established as a scientific field by

Received May 16, 2008; Published online June 26, 2009

Address for correspondence:
TOMISLAV HENGL
E-mail: T.Hengl@uva.nl

0138–9130/US \$ 20.00
Copyright © 2009 Akadémiai Kiadó, Budapest
All rights reserved

JOURNAL [1978], JOURNAL [1986], ISAAKS [1989], CRESSIE [1993], STEIN [1999], WEBSTER [2007], and numerous other authors. Because geostatistics evolved in the mining industry, it for a long time meant statistics applied to geology [CRESSIE, 1990]. So far, geostatistical techniques have been successfully applied to applications in numerous fields, ranging from soil mapping, meteorology, ecology, oceanography, geochemistry, epidemiology, and human geography, to geomorphometry. Modern geostatistics, which largely overlaps with the term “*spatio-temporal statistics*”, can be defined as a branch of statistics that specializes in the analysis and interpretation of any spatially (and temporally) referenced data [JOURNAL, 1986]. It is a collection of techniques and theories that can be used to generate sampling designs, build statistical models, make spatio-temporal predictions at unvisited locations, extract spatio-temporal patterns in the data and analyze the associated uncertainties, just to mention the most frequent uses.

Most recently, ZHOU [2007] applied multivariate statistical analysis to 2,866 articles (1967–2005) extracted from the ISI’s database and draw several interesting conclusions. We discussed this issue at length and concluded that this field deserves a more informative analysis than that made in the paper written by ZHOU [2007]. First, we felt that the authors narrowed the definition of the field by only taking into account research articles indexed by Thompson Scientific. Second, the authors considered only the yearly Journal Impact Factor as a measure of quality, and it is known that there are large inter-annual fluctuations in the impact factor of a journal and that there is no relationship between the journal’s impact factor and that of individual papers. Third, ZHOU [2007] practically concluded that the field has already reached its peak in 1996–1998 and that, ever since, both the citations and impact of the articles have been in decline (we wanted to be sure that this is really true). And finally, we decided to conduct these investigations because we felt that any analysis of bibliometric indices of geostatistics should also include geostatistical analysis.

To provide a more detailed and more accurate picture of the field, we will first expand our search to include all available search engines and then focus on the individual library items and authors. This will allow us to list the ‘*hottest*’ publications and authors within the field of geostatistics. We will then make our article even more provocative by asking where the geostatistics ‘*hot-spots*’ are in the world. As a final product of these analyzes, we will provide a global map of the density of publication and annual citation rates.

Bibliometrics: The basics

Researchers report on new discoveries and on the results of experiments with the intention of making a significant contribution to human knowledge and solving research problems of theoretical and practical importance. One way to evaluate their success on this path is to observe how frequently their output is cited after publication. In a way, each publication has a life, which means that its citations will first grow exponentially, reach a peak and then decay (Figure 1a). Unlike books, which can be updated periodically in new editions, research articles are permanent and can have a limited life. Unfortunately, many publications have almost no life at all, i.e. they never get cited.

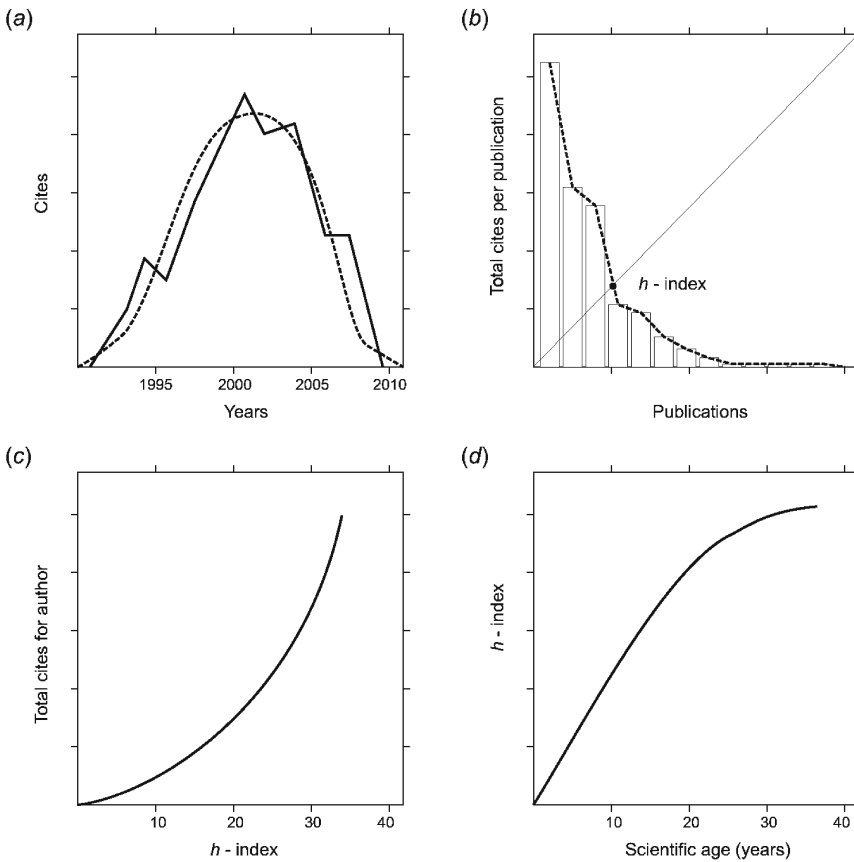


Figure 1. Some key principles of bibliometrics: (a) a publication has a life that approximately follows the asymmetric Hubbert curve; (b) the *h*-index is determined by sorting author's publications by number of citations; (c) *h*-index and total number of citations per author are usually highly correlated; (d) author's *h*-index is a function of the author's scientific age – the year of the first publication after MINASNY & AL. [2007]

In fact, typically senior researchers will only have a handful of publications to their name that made an impact in their field. If we sort an author's publications by the number of citations, we obtain a graph as shown in Figure 1b.

Today, the impact and importance of research publications and authors can be successfully followed through web services, of which the three most known are:

- *Web of Science* (isiknowledge.com) – WoS is a Thompson's Scientific subscription-based multidisciplinary database that covers about 8,700 peer-reviewed journals clustered into the following three databases: the Science Citation Index Expanded, Social Sciences Citation Index Expanded and Arts and Humanities Citation Index. The most recent version of WoS (v4.0) indicates that it contains about 23 million full documents from the period 1988–2007. WoS can generate citation reports and be used to analyze the results, e.g. to compare the success of authors, institutes or countries, to see how citations of papers change over time. The results of searches can be exported in a variety of formats and further used to generate reports. WoS is simply the most detailed and most accurate scientific database of peer-reviewed articles published in the English language.
- Elsevier's *SCOPUS* (scopus.com) – SCOPUS contains about 33 million of abstracts from approximately 15,000 peer-reviewed journals, but also shows the status of non-registered SCOPUS publications, including 386 million web-based publications. SCOPUS has recently offered a Citation Tracker service that makes it possible to assess the impact of an individual author/publication. The *h*-index, which is now also incorporated in SCOPUS, can be derived for each selected author/group. A limitation of SCOPUS is that it does not have complete citation information for articles published before 1996.
- *Google Scholar* (scholar.google.com) – GS is Google's non-commercial academic search service that registers any publication available on the web. It indexes all on-line materials including Power Point presentations, mailing lists, blogs and the like, but also all peer-reviewed journals available on-line (except those published by Elsevier). The advantage of Google Scholar is that it allows free search for publications written in any language and from any publisher and hence supports the democratization of citation analysis [NORUZI, 2005; HARZING, 2008]. Google Scholar has been available since 2004 and ever since its quality has been constantly improving. Its biggest limitations are 'noisy' data with many inconsistencies, duplicate publications and errors. More recently, Microsoft started its own academic search service, which is called *Windows Live Academic* (academic.live.com).

Each of the three web services listed above has its own particular advantages and includes some specific new information that its competitor does not have. On the other hand, they do compete in providing the key information about the citation records and there will always be differences – sometimes minor, sometimes significant. BAUER [2005] showed that there is not much difference between GS and WoS in terms of their accuracy in assessing the number of citations for the highly-cited publications, but there is indeed a significant difference between the GS and WoS/SCOPUS in assessing all publications/authors within a certain field [BAR-ILAN, 2008]. MEHO [2007] estimated that the overlap between the results of GS and WoS/SCOPUS is only 30–50%. This typically happens because GS indexes about twice as many publications as WoS/SCOPUS, most of which are conference papers, dissertations, theses and book chapters. Likewise, SCOPUS counts only citations from articles published after 1996, so that there will also be differences between SCOPUS, WoS and GS.

Although there has been considerable pressure for academic organizations to start using GS, many librarians advise sticking to WoS or SCOPUS if one needs to obtain an accurate citation count [GILES, 2005; JACSÓ, 2005; Kousha & Thelwall, 2007]. This is mainly because GS is often incomplete and noisy. In addition, Google still does not want to reveal details of its search algorithm, but this might change in the near future. In this article we have decided to treat all three services equally, although we will mainly use the results of WoS or SCOPUS to evaluate highly-cited publications and to extract specific information, e.g. about the geographical location of the authors.

Most international publications can also be browsed using the *WorldCat* service (<http://worldcatlibraries.org>), provided by the OCLC Online Computer Library Center, Inc. This web service contains over 1 billion items from more than 10,000 libraries. It makes it possible to search for bibliographic items by title, subject and/or author's name. The query results are grouped by authors, research fields, formats, languages and year of publication. WorldCat can also sort the query results by publication date and relevance. However, it does not provide any bibliometric evaluation of the publications, hence we will use it only to investigate summary statistics.

Methods

Data collection

The field of geostatistics can best be defined by listing a number of keywords that are unique to the field and can only be associated with a limited number of authors. After some discussion, we generated the following list: “*spatial*”, “*(geo)statistics*”, “*variogram*”, “*kriging*”, “*spatial interpolation*”, “*spatial prediction*”, and “*spatial sampling*”. We further decided that, from this list, the only essential words (these must be in the abstract or/and title) are “*spatial*”, “*statistic**”, “*variogram*” and/or “*kriging*”, while all other words merely increase the likelihood that a publication belongs to the

geostatistics domain. After some preliminary testing, we designed a list of queries which we used to extract balanced sets of publications that can be classified as *geostatistical*. In the case of WoS, the query was:

```
topic=(kriging OR variogram OR "spatial statistic"
OR "spatial interpolation" OR "spatial predict"
OR "spatial sampling" OR geostatistic*)
```

and in the case of SCOPUS:

```
TITLE-ABS-KEY (kriging OR variogram OR "spatial statistic"
OR "spatial interpolation" OR "spatial predict"
OR "spatial sampling" OR geostatistic*)
```

Once we retrieve the results of query, we can sort them by relevance (the number of times specified words appear in the text) and then export the first *n* from the list. This way we are sure that we will be processing truly representative articles (and not just any article with the word “*geostatistic**”). In the case of GS we are not able to sort the results based on relevance so we searched citations with ANY of the words: “*kriging*”, “*interpolation*”, and “*sampling*”, and with all of the words: “*spatial*”, “*statistic**” and “*variogram*”.

The results of queries from WoS or SCOPUS can be imported into bibliographic management software such as Endnote or exported directly to any other table editor. GS publications can be extracted using *Publish or Perish* (PoP) freeware (www.harzing.com), which automates the extraction of scientometric indices. The PoP software limits searches to 1000 results, so we repeated the queries for each of the last 27 years (1980–2007). The results of this search can be seen in Figure 4.

In principle, we focused on two bibliometric indices: (a) for articles: average number citations per year or *citation rate* (CR) and (b) for authors: *h-index* – number of author’s papers that are cited at least that number of times (Figure 1b). CR can distinguish the most influential publications, while the *h-index* is more suitable for evaluating authors because it corrects for the “one hit wonders” – academics who have authored only a small number of highly-cited papers [ROEDIGER, 2006; HARZING, 2008]. An author with a high *h-index* is, by definition, a sustained, long lasting record of good academic performance. Both indices are designed to distinguish truly influential scientists from those who simply publish many papers. For more details about the state-of-the-art bibliometric indices see [BATISTA, 2006] and [BIHUI, 2007].

Data processing and statistical analysis

The queries listed above gave us 6,393 publications from WoS, 10,491 from SCOPUS and 5,389 publications from GS. The WoS and SCOPUS publications were first sorted by relevance and then the first 4,000 entries were exported, filtered and reorganized to allow for further statistical analysis and processing. The GS database,

which is noisy, requires filtering before it can be used. We often found duplicate or triplicate publications in the systems, but there are also many publications with misspelling (special symbols) of authors' names. However, most of these can be easily filtered out, either by visually examining the results or by systematic filtering. We also decided to reduce the number of publications in the results of WoS and GS. In the case of WoS, publications can be sorted by relevance (the number of times keywords appear in the text); in the case of GS, we omitted publications that are over four years old and have still not been sighted. With SCOPUS items can also be sorted by relevance; however, it limits the number of items that can be exported to 2000.

From all search results, we prepared four tables with the following structure:

- *publications_wos* ($n = 4000$): publication title, year, authors, journal/publisher, number of citations, latitude and longitude.
- *publications_sc* ($n = 2000$): publication title, year, authors, journal/publisher, number of citations.
- *publications_gs* ($n = 4724$): publication title, year, authors, journal/publisher, number of citations.
- *authors* ($n = 200$): author, institution/company, city, latitude, longitude, total publications, SCOPUS *h*-index for authors, WoS *h*-index for their geostatistics publications.

The table (*authors*) was produced by manually recording the *h*-index and publication/citation statistics in WoS and SCOPUS for a list of 200 representative authors. We first located all articles from a single author and then looked at the citation reports. From the 200 authors, 158 authors were selected as having >5 geostatistics publications and the remaining 42 were added to represent equally all populated geographical regions (India, Brazil, Africa, China, etc.). This way we reduced the bias towards western countries, while still being able to focus on the most representative geostatistics authors. The table *publications_wos* was used for geostatistical analysis because WoS also records the addresses of the corresponding authors. The cities associated with the corresponding authors were matched with a point file representing 88,564 world cities. About 30% of cities could not be matched automatically, so we located their coordinates (latitude/longitude) using Google Earth database. Many publications had the same coordinates, which might represent a computational problem for further geostatistical analysis (duplicate observations), so we randomly added a positional error of 0.001 arc-degrees.

The four tables were imported into the R computing environment and further processed using various R packages. We first look for summary statistics for various parameters and then run three groups of analysis: (1) purely statistical analysis to observe summary and temporal trends and (2) geostatistical analysis to detect spatial patterns and map areas of scientific excellence. For trend analysis we use the GS database for global results and the WoS database to zoom into the specific publications.

For geostatistical analysis we first run the kernel smoother that is implemented in the *spatstat* package of R [DIGGLE, 2003]. This can be used to map the density of publications all over the world. We then attach the CR value at each point and then rerun the kernel smoother. The resulting map shows the density of CR, i.e. the clusters of scientific excellence in geostatistics.

We finally correlated the bibliometric indices with the global socio-economic maps: the global gridded population density map, the night light image and the mean global Vegetation Index for the period 1981–2001. The 2.5 arc-minutes population density map for year 2000 was obtained from the Socioeconomic Data and Applications Center (SEDAC) of Columbia University (<http://sedac.ciesin.org/gpw/>), the filtered night light image from 1992 through 2003 was obtained from NASA's Defense Meteorological Satellite Program (<http://www.ngdc.noaa.gov/dmsp/>), the elevation map of the world was obtained from the NOAA's National Geophysical Data Center (<http://www.ngdc.noaa.gov>), and the 10 km AVHRR NDVI time-series (232 monthly images; July, 1981 through September, 2001) were obtained from the International Water Management Institute (<http://iwmidsp.org>). The NDVI time series were used to extract first, second, third and the fourth Principal Components, where the first component usually illustrates the long term mean value, and the second component indicates seasonal variations which are due to the climatic differences [EASTMAN, 1993; PIWOWAR, 2002]. Further components can be connected to the long-term changes of biomass, i.e. they can be used to quantify degradation processes in heterogeneous landscapes [LASAPONARA, 2005]. The final list of predictors in use was:

- population density map for the year 2000;
- mean night light image for the period 1992–2003;
- change in night light images 2003/1992;
- elevation map of the world;
- 1–4 PC components of the NDVI images for the period 1981–2001;

These predictors were deliberately selected to investigate whether certain indices can be explained by local socio-economic conditions. DOLL [2006], for example, showed that the night light image is closely correlated with Gross Domestic Product. Here we also assume that the relevance of a region will increase as its economic wealth grows. In addition, we derived biomass change index for whole world and investigated whether it can be connected with the scientific quality. All raster maps were re-sampled to the 2.5 arc-minutes grid (about 10 km at equator), with LatLongWGS84 coordinates. For practical purposes, the area of interest has been subset to the latitudes 65°S–65°N. For those interested in getting more insight into the data processing steps (R script) and in obtaining the input data sets and maps used in this paper, please visit the contact author's website.

Results

Geostatistics and its future

Let us first look at the results of searching the WorldCat service (Table 1). This will give us some general idea about the summary structure of library items, i.e. which are the most dominant formats, fields, languages and what the trends are.

Table 1. Summary statistics for the library items obtained using keywords in geostatistics. Results based on WorldCat (provided by the OCLC Online Computer Library Center, Inc.)

	Spatial interpolation	Spatial statistics	Spatial sampling	Variogram	Kriging	Geostatistics	TOTAL
Total library items							
2000–2007	195	1169	628	169	567	509	3237
1995–2000	51	467	180	56	213	231	1198
1990–1995	24	379	77	10	52	67	609
1985–1990	6	233	25	7	68	36	375
<1985	8	192	36	7	30	44	317
Format							
Article	99	1252	376	180	555	510	47%
Book	110	1138	285	44	274	323	35%
Internet resource	75	367	372	25	121	96	17%
Computer file	3	39	2	2		6	1%
Map		25					0%
Language							
English	266	2492	901	237	840	868	93%
German		31	7			5	0.7%
French		46					0.8%
Chinese	2		20	3	27		0.9%
Portuguese	15	49	70	7	37	30	3.5%
Spanish		31	6				0.6%
Japanese				3		5	0.1%
Content category							
Earth Sciences	15	239	27	8	26	77	37%
Engineering & Tech.	12		24	8	25	48	11%
Mathematics	3	154	18	3	13		18%
Biologicalsciences			47				4%
Agriculture	3	48			7	8	6%
Sociology		78					7%
Business & Economics		151				5	15%

Surprisingly, about 53% of publications are not articles – mainly books (35%) and Internet resources (17%). This indicates that indexing only research articles for geostatistics would give an incomplete picture of the field. We were also surprised to find that there are about twice as many publications containing the words “*spatial AND statistic**” as those with words “*geostatistic* OR kriging*”. This shows the importance of selecting precisely the keywords that define a research field. It is very possible that ZHOU [2007], by defining the field simply, with just one keyword (“*geostatistic**”) have inadvertently missed many important library items in their analysis of geostatistics.

Furthermore, WorldCat shows that most publications belong to the group Geography & Earth Sciences (37%), second is Mathematics (18%), followed by Business & Economics (15%). About 93% of publications are in English; the second on the list is Portuguese (3.5%), then Chinese (0.9%), French (0.8%) and Spanish (0.6%). WorldCat sorts search results by relevance, which can also be of interest. For example, the books listed as the most relevant for spatial AND interpolation is “*Spatial Statistics*” by Ripley, B. D.; for spatial AND statistics “*Statistics for Spatial Data*” by Cressie, N. A. C.; for spatial AND sampling “*Spatial Sampling and Regression Strategies*” by Lee, B. C. and Brooks, D. M.; for geostatistics “*Principles of Environmental Sampling*” by Keith, L. H., “*Applied Geostatistics*” by Isaaks, E. H. and Srivastava, R. M., “*Model-based Geostatistics*” by Diggle, P. and Ribiero, P. J. and “*Multivariate Geostatistics*” by H. Wackernagel. Compare further with the Top 25 list of books in Table 2.

We proceeded by analyzing the summary results for WoS, SCOPUS and GS. The results of searching WoS (4,000 hits, time span = 1988–2007) clearly show that both the number of published research papers and citations in the field of geostatistics are still increasing. Most papers are published in the field of multidisciplinary geosciences (23%), followed by environmental sciences (19%), water resources (14%), statistics (14%), soil science (14%), interdisciplinary applications of mathematics (13%) and ecology (6%). The most productive journals are *Water Resources Research* (64), *Mathematical Geology* (43), *Monthly Weather Review* (42), *Journal of Geophysical Research* (40), *Journal of Hydrology* (40), *Geoderma* (37), etc. According to SCOPUS (2000 hits), the most productive journals are *Geoderma* (143), *Mathematical Geology* (134), *Water Resources Research* (91), *Journal of Hydrology* (82), *Environmetrics* (77), *Soil Science Society of America Journal* (60), *Computers and Geosciences* (49), etc. GS contains many more unique bibliographic sources (1376 in total) and has a somewhat different order: *Mathematical Geology* (320), *Geoderma* (317), *Journal of Hydrology* (130), *International Journal of Remote Sensing* (113), *Computers and Geosciences* (96), etc.

Table 2. The top 25 books in geostatistics, based on results in January 2008 (SCOPUS/GS).
General reference books and encyclopedia have been excluded

#	Book	CR
1	NAC Cressie, " <i>Statistics for Spatial Data</i> ", Wiley	(211/219)
2	EH Isaaks and RM Srivastava, " <i>An Introduction to Applied Geostatistics</i> ", Oxford University Press	(99/102)
3	CV Deutsch and AG Journel, " <i>GSLIB: Geostatistical Software Library and User's Guide</i> ", Oxford University Press	(65/128)
4	P Goovaerts, " <i>Geostatistics for Natural Resources Evaluation</i> ", Oxford University Press	(79/86)
5	AG Journel and CJ Huijbregts, " <i>Mining Geostatistics</i> ", Academic Press	(55/58)
6	TC Bailey et al., " <i>Interactive Spatial Data Analysis</i> ", Prentice Hall	(40/71)
7	JP Chiles and PJ-P Delfiner, " <i>Geostatistics: Modeling Spatial Uncertainty</i> ", Wiley	(54/57)
8	R Webster and MA Oliver, " <i>Geostatistics for Environmental Scientists</i> ", Wiley	(42/36)
9	S Banerjee et al., " <i>Hierarchical Modeling and Analysis for Spatial Data</i> ", CRC Press	(23/49)
10	PJ Diggle, " <i>Statistical Analysis of Spatial Point Patterns</i> ", A Hodder Arnold	(22/41)
11	ML Stein, " <i>Interpolation of Spatial Data: Some Theory for Kriging</i> ", Springer	(27/33)
12	H Wackernagel, " <i>Multivariate Geostatistics: an Introduction with Applications</i> ", Springer	(17/42)
13	AB Lawson, " <i>Statistical Methods in Spatial Epidemiology</i> ", Wiley	(13/43)
14	MRT Dale, " <i>Spatial Pattern Analysis in Plant Ecology</i> ", Cambridge University Press	(24/28)
15	AS Fotheringham et al., " <i>Quantitative Geography: Perspectives on Spatial Data Analysis</i> ", Sage Publications Ltd	(17/31)
16	R Haining, " <i>Spatial Data Analysis: Theory and Practice</i> ", Cambridge University Press	(18/29)
17	P Elliott, " <i>Spatial Epidemiology: Methods and Applications</i> ", Oxford University Press	(26/20)
18	R Webster and MA Oliver, " <i>Statistical methods in soil and land resource survey</i> ", Oxford University Press	(20/24)
19	LA Waller and CA Gotway, " <i>Applied Spatial Statistics for Public Health Data</i> ", Wiley	(19/23)
20	DL Peterson, " <i>Ecological Scale: Theory and Application</i> ", Columbia University Press	(26/14)
21	MJ Fortin and M Dale, " <i>Spatial Analysis: A Guide for Ecologists</i> ", Cambridge University Press	(20/19)
22	PJ Diggle and PJ Ribeiro, " <i>Model-Based Geostatistics</i> ", Springer	(19/19)
23	GBM Heuvelink, " <i>Error Propagation In Environmental Modelling With GIS</i> ", CRC Press	(14/22)
24	PK Kitanidis, " <i>Introduction to Geostatistics: Applications in Hydrogeology</i> ", Cambridge University Press	(13/22)
25	BD Ripley, " <i>Statistical Inference for Spatial Processes</i> ", Cambridge University Press	(11/21)

WoS shows that the most productive geostatistics authors are Stein, A. (62), Goovaerts, P. (51), Webster, R. (45), Deutsch, C. V. (43), Lark, R. M. (41), Atkinson, P. M. (38) and McBratney, A. B. (36). The most productive authors by SCOPUS are

McBratney, A. B. (39), Stein, A. (36), Lark, R. M. (35), Goovaerts, P. (35), Atkinson, P. M. (29), Cressie, N. (26), Christakos, G. (25) etc. The most productive authors according to GS are: Atkinson, P. M. (50), Stein, A. (41), McBratney, A. B. (37), Cressie, N. (36), Lark, R. M. (35), Webster, R. (34) etc. These results coincide with those found by ZHOU [2007].

At this stage, it seems useful to see how well the results of GS, WoS and SCOPUS match and whether there are statistically significant differences between the different sources. Figure 2 shows that the citation distributions estimated for the five most productive journals do differ slightly. In both cases, the articles published in *Water Resources Research* are the most influential. To determine the second most influential journal from the five listed seems to be trickier: for WoS this is *Geoderma*, for GS *Journal of Hydrology*. If we run a t-test on distributions estimated by various sources, there is no significant difference: e.g. The CRs estimated for *Geoderma* by WoS and GS do not differ significantly ($p = 0.901$). This indicates that both WoS/SCOPUS and GS are equally good for evaluating the *good players*. Note also that the differences between the journals are significant: *Water Resources Research* does receive significantly more total citations than *Computers and Geosciences* ($p = 0.0021$). The same conclusions can be derived using both WoS/SCOPUS and GS.

We proceed by determining whether the number of references or the year of publication have any effect on the CR of an article. For this we used the results in the WoS database. Figure 3 shows that articles with more references do perform better, although the plot is relatively noisy ($r = 0.29$). Note also that the CR is a relatively constant parameter, as it does not change much over the years. The graph in Figure 3 indicates that it takes at least five years for an article to get a relatively stable CR. There is practically no correlation between the number of pages and the CR ($r = 0.15$), although the scatter-smooth plot shows that the optimal number of pages is about 23.

We further focused on analyzing the trends in the field. The results of querying GS (Figure 4) confirmed those of ZHOU [2007]: the average h -index of the authors reached a peak in 1998 and has declined since then. However, the per-author citation rate does not show this decline. Why this discrepancy between the two? We assume that the average h -index has been lower in the last few years because the structure of authors has changed. If the average age of authors is declining, the h -index would also decrease because younger authors have an h -index that is up to several times smaller, even if their articles perform well (see also Figure 1d).

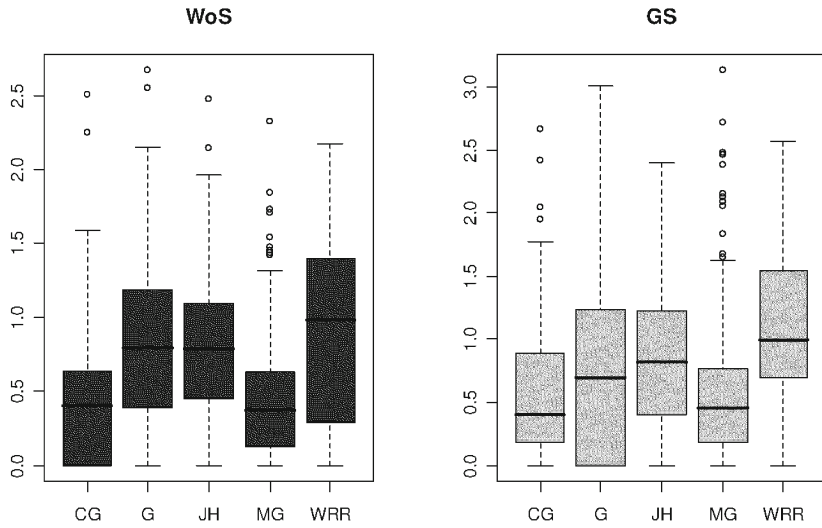


Figure 2. Box plots of the log of citations for the five most productive geostatistics journals. This shows that both WoS/SCOPUS and GS provide similar citation statistics. CG – *Computers and Geosciences*, G – *Geoderma*, JH – *Journal of Hydrology*, MG – *Mathematical Geology*, WRR – *Water Resources Research*

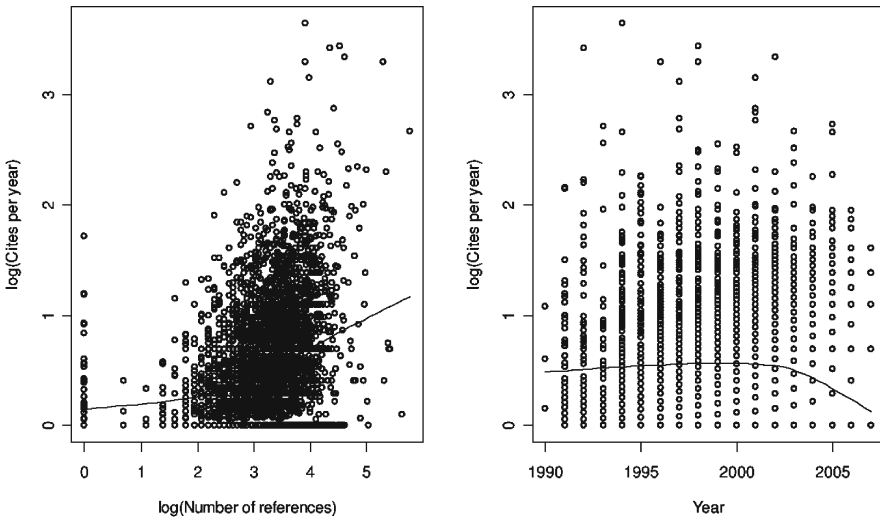


Figure 3. The relationship between the number of references and year of publication and the citation rate for geostatistics papers

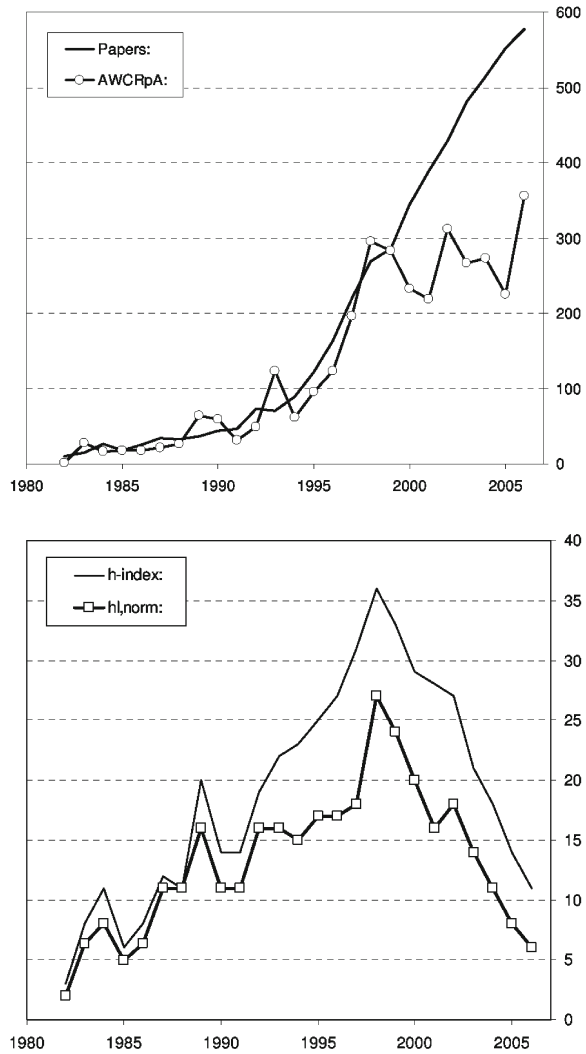


Figure 4. Trends in geostatistics, considering the number of publications and citations per year (above) and authors' h -indices (below). Results obtained using PoP software. See in the 'Data collection' section for a description of the data collection steps

Table 3. The top 25 research articles in geostatistics, based on results in January 2008 (WoS/SCOPUS/GS). Only articles older than three years and with at least four keywords in the title or abstract have been considered

#	Article	CR
1	RW Reynolds and TM Smith, "Improved global sea-surface temperature analyses using optimum interpolation", <i>Journal of Climate</i>	(101/106/106)
2	P Legendre, "Spatial autocorrelation: Trouble or new paradigm?", <i>Ecology</i>	(46/46/51)
3	C Daly et al., "A statistical-topographic model for mapping climatological precipitation over mountainous terrain", <i>Journal of Applied Meteorology</i>	(37/39/48)
4	P Legendre and MJ Fortin, "Spatial pattern and ecological analysis", <i>Vegetatio</i>	(40/35/41)
5	EJ Gustafson, "Quantifying landscape spatial pattern: What is the state of the art?", <i>Ecosystems</i>	(30/34/42)
6	A Kaplan et al., "Analyses of global sea surface temperature 1856-1991", <i>Journal of Geophysical Research</i>	(35/32/39)
7	MP Austin, "Spatial prediction of species distribution: an interface between ecological theory and statistical modelling", <i>Ecological Modelling</i>	(30/30/32)
8	RE Rossi et al., "Geostatistical tools for modeling and interpreting ecological spatial dependence", <i>Ecological Monographs</i>	(30/29/33)
9	WH Schlesinger et al., "On the spatial pattern of soil nutrients in desert ecosystems", <i>Ecology</i>	(26/29/33)
10	M Kulldorff, "A spatial scan statistic", <i>Communications in Statistics – Theory and Methods</i>	(19/23/36)
11	D Stockwell and D Peters, "The GARP modelling system: problems and solutions to automated spatial prediction", <i>International Journal of Geographical Information Science</i>	(22/23/32)
12	L Anselin, "Spatial dependence in linear regression models with an introduction to spatial econometrics", <i>Handbook of Applied Economic Statistics</i>	(-/13/36)
13	P Legendre et al., "The consequences of spatial structure for the design and analysis of ecological field surveys", <i>Ecography</i>	(23/23/24)
14	SP Rushton et al., "New paradigms for modelling species distributions?", <i>Journal of Applied Ecology</i>	(22/23/25)
15	P Legendre et al., "The consequences of spatial structure for the design and analysis of ecological field surveys", <i>Ecography</i>	(23/23/20)
16	CH Ettema and DA Wardle, "Spatial soil ecology", <i>Trends in Ecology and Evolution</i>	(20/21/25)
17	X Vekemans and OJ Hardy, "New insights from fine-scale spatial genetic structure analyses in plant populations", <i>Molecular Ecology</i>	(21/23/27)
18	P Segurado and MB Araujo, "An evaluation of methods for modelling species distributions", <i>Journal of Biogeography</i>	(20/21/20)
19	AJ Underwood and MG Chapman, "Scales of spatial patterns of distribution of intertidal invertebrates", <i>Oecologia</i>	(22/22/18)
20	R Engler et al., "An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data", <i>Journal of Applied Ecology</i>	(19/19/23)
21	JL Dungan et al., "A balanced view of scale in spatial statistical analysis", <i>Ecography</i>	(16/18/18)
22	SJ Jeffrey et al., "Using spatial interpolation to construct a comprehensive archive of Australian climate data", <i>Environmental Modelling & Software</i>	(15/17/16)
23	JN Perry et al., "Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data", <i>Ecography</i>	(15/15/18)
24	J Franklin, "Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients", <i>Progress in Physical Geography</i>	(14/15/17)
25	P Goovaerts, "Geostatistics in soil science: State-of-the-art and perspectives", <i>Geoderma</i>	(12/15/19)

The most influential publications/authors

Tables 2 and 3 show the Top 25 lists of research articles and books in geostatistics. These were produced by taking the average statistics obtained from various sources and then by sorting authors/publications by the best score. Note that there are often small differences between the numbers of citations estimated by various sources, but that such differences usually do not affect the final order.

The Top 25 list of books presented in Table 2 shows that, in geostatistics, many readers still refer to books over 15 years old: absolute winners are N. A. C. Cressie's "*Statistics for Spatial Data*" from 1993 and E. H. Isaaks and R. M. Srivastava's "*An Introduction to Applied Geostatistics*" from 1990, while the hottest new books on the list are "*Hierarchical Modeling and Analysis for Spatial Data*" (2004), "*Spatial Analysis: A Guide for Ecologists*" (2005) and "*Model-Based Geostatistics*" (2007). Also note that most of the Top 25 geostatistics books have been published in the Oxford University Press Applied Geostatistics Series.

For the Top 25 articles, all three sources give approximately the same citation statistics, which makes it easier to depict the best. The absolute winner is "Improved global sea-surface temperature analyses using optimum interpolation" by Reynolds, R. W. and Smith, T. M., with over 100 citations per year since 1994. Most of papers on the Top 25 list are connected with ecological applications and habitat mapping, the second set of papers on the list are those connected with interpolation of meteorological and soil data. Few authors (especially Legendre, P. and Rossi, R. E.) seem to have the necessary talent to produce highly-cited articles. An interesting discovery is that books seem on average to be cited more often than articles, although the difference is not significant ($p = 0.062$). Also note that several Top 25 articles do not come from journals with an extremely high impact factor (*International Journal of Geographical Information Science* has JIF = 1.4, *Environmental Modelling and Software* JIF = 2.0, *Journal of Geophysical Research* JIF = 2.8). On the other hand, JIF does play an important role, because most of the highly cited articles come from journals with a high impact factor: *Ecology* (JIF = 4.8), *Ecography* (JIF = 3.3) and *Journal of Applied Ecology* (JIF = 4.5).

Another way to look for the best papers in geostatistics is to compare only the articles published within the same year. This means that we can use the same queries as in 'Data collection' section, but then limit the publication years. WoS shows that the most cited articles for 2006 are "Ecological models of alcohol outlets and violent assaults: crime potentials and geospatial analysis" (13 citations), for 2005 "Ambient air pollution and atherosclerosis in Los Angeles" (68 citations), for 2004 "Multivariable geostatistics in S: the gstat package" (34 citations), for 2003 "On digital soil mapping" (67 citations), for 2002 "Spatial soil ecology" (121 citations), for 2001 "Adjusting for sampling density in grid box land and ocean surface temperature time series" (113

citations), and for 2000 “Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall” (87 citations), etc. As you can see, several lists of influential publications can always be generated even if only one bibliometric parameter is taken into account.

We further examined whether authors’ productivity is directly correlated with the impact they make. For this, we use the table authors, which contains both the number of articles authors publish per year (either as first author or as co-author) and citations per year for each author. From the sample of 200 representative geostatistics authors (>5 geostatistics publications), we can clearly see that there is a significant correlation ($r = 0.77$) between logs of their CR and published articles (Figure 5). However, this relationship is noisier than we expected – productivity explains only 60% of the authors’ impact. The correlation plot becomes especially noisy for authors with less than 20–30 publications: the correlation coefficient drops from 0.68 for authors with <60 publications, to 0.63 for authors with <60 publications and 0.58 for authors with <20 publications. The most extreme cases of geostatisticians who made an enormous impact with only a few publications are Rossi, R. E. (666 citations with only nine publications!) and Daly, C. (726 citations with 27 publications). The h -index has a slightly stronger correlation with the number of publications ($r = 0.81$) and the authors’ CR ($r = 0.83$).

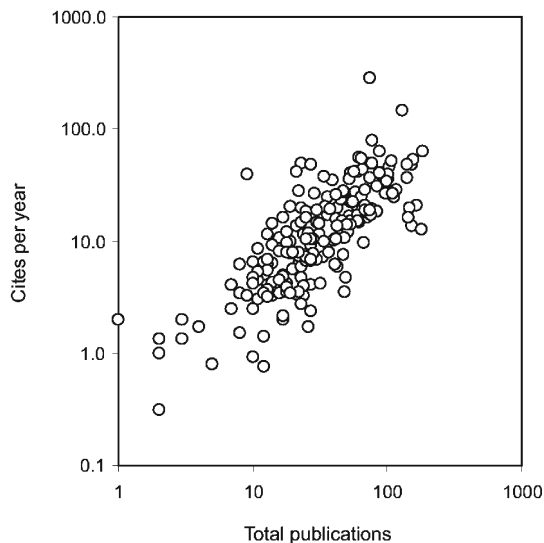


Figure 5. Relationship between the authors’ productivity (total number of published publications) and impact (citations per year) for 200 representative geostatistics authors. Based on SCOPUS author details

World maps of geostatistics

Figure 6 shows the summary results of the geostatistical analysis produced using the table `geostat_wos` and auxiliary maps. The world map of the density of publications (Figure 6a) has been made more straightforward by running a kernel smoother for a given grid. This shows several hotspots where geostatistics articles are produced in larger quantities: in Europe – University of Reading, Southampton, Rothmasted Research, INRA and Wageningen University; in the United States there are many more small clusters – Stanford University, University of Arizona, Michigan, Wisconsin, California, Alberta, Illinois, Texas, and North Carolina State University, etc; in South America, many geostatistics articles are produced near to Sao Paulo in Brazil and Santiago in Chile; in Australia four cities seem to be of equal importance – Sydney, Melbourne, Adelaide and Perth; in Asia, the obvious hotspots are Taipei, Beijing and Seoul.

The kernel-smoothed map of geostatistics publications with the CR used as a weight vector is shown in Figure 6b. As you can see, many areas with the CR of publications close to zero will disappear from the world map, while locations in Europe, United States and Australia are emphasized even more. This shows that there are clusters of scientific excellence around Wageningen, London, Utrecht, Hampshire, Norwich, Paris, Louvain, Barcelona, and Zürich (Europe); Stanford, Ann Arbor, Tucson, Corvallis, Seattle, Boulder, Montreal, Baltimore, Durham, Santa Barbara and Los Angeles (North America); Canberra, Melbourne, Sydney, Santiago (Chile), Taipei, and Beijing (other continents). The CR-weighted density map is hence more informative than the map in Figure 6a because it can be more closely connected with scientific excellence.

Initial attempts to interpolate the CR using standard geostatistical techniques such as kriging proved to be difficult. First, the variable shows a very large short-range variation – obviously because many institutions have both very successful and less successful publications. Second, the CR is heavily skewed, hence it requires a double log transformation before its histogram shows close-to-normal distribution. To account for these problems, we first average the values of the CR over the 6 arcminutes grid and then work with the smoothed values. Further correlation with socio-economic variables showed that the spatial distribution of the CR in geostatistics is not correlated with the night light image (economic activity, Figure 6c) nor with environmental conditions, but is negatively correlated with economic growth ($r = -0.11$) and population density ($r = -0.05$). The year of publication is also correlated with economic growth, indicating that developing countries are publishing more. In all cases we get a very weak correlation, which indicates that productivity and excellence in geostatistics is relatively independent of socio-economic conditions. It probably means that human resources and good ideas provide a better guarantee of success than local economic activities.

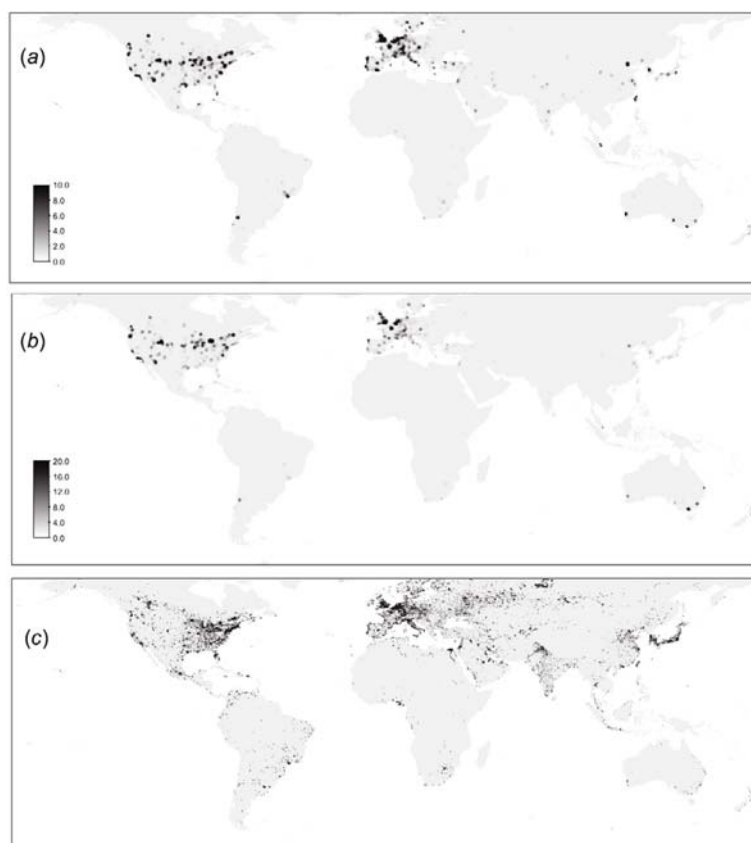


Figure 6. World maps of bibliometric parameters for geostatistics estimated using a sample of 4000 articles: (a) density of published research articles generated using the isotropic Gaussian kernel with a standard deviation of 0.5 arcdegrees; (b) the same but weighted using the CRs for each article; (c) night light image of the world for the period 1992–2002. Based on results in January 2008

Discussion

By conducting the analyses described in this paper, we gained important insights that are discussed in this section. First, we discovered that it is extremely hard to define a field by specifying only few keywords. The publications today are increasingly inter/multi-disciplinary and it is often hard to identify a crisp boundary and define the domain of a scientific field. One solution to this problem is to first look for all possible publications containing ANY of the keywords, then by relevance and then take a limited sample of items. This is certainly better than if only one or just a few keywords are used, because there is always a danger of being biased. However, a problem with

taking into account all possible publications where ANY of the words can be found is that their number can easily exceed 10^5 , which usually makes it rather difficult to analyze.

In the case of journal articles, many review articles can easily fall under several research fields. Review articles also get cited more frequently, which always puts them on the top of the most cited lists. The same happens with pure mathematical/statistical books that cover numerous fundamental topics: e.g. the book of Venables, W. N. and Ripley, B. D. "*Modern Applied Statistics with S*" originally appeared as the second on our Top 25 list. However, we eventually decided to remove it because <30% of its content can be really connected with analysis of spatial phenomena. Here we also need to mention that, while querying books, we had to exclude all general software manuals, glossaries and encyclopedias because they are much more likely to be listed as bestsellers for multiple terms. In that sense, it would have helped us considerably if the research groups themselves had assigned their publications to a specific field or domain.

Importing and analyzing library items is not a straightforward exercise. According to the author of PoP, Anne-Wil Harzing, citations per paper (and related measures) are rather sensitive to GS errors as the results for most authors display far more papers (e.g. because of inaccurately referenced duplicates) than they have actually written [HARZING, 2008]. We would suggest filtering the PoP results first before any further analysis is performed (see also [BAR-ILAN, 2008]). In addition, GS's coverage is not equally good in all areas of science, which can lead to biased estimates for interdisciplinary topics/fields [HARZING, 2008]. It seems that the most sensible thing to do is to start the search using PoP, filter out duplicates and purely 'grey' sources, but then also cross-check the results with the commercial databases WoS and/or SCOPUS. Note that matching GS and WoS/SCOPUS items is also not an easy task. For some reasons, Google trims long titles, which might be a problem especially if the first words/letters of the title are trimmed. We nevertheless foresee that it is probably only a matter of time until GS is accepted by research organizations as the most accurate and most accessible tool for assessing the impact of authors [JACSÓ, 2005; NORUZI, 2005].

Books are particularly hard to evaluate especially if we are dealing with edited books where each chapter is referred to separately. Neither GS nor SCOPUS perceive specific book chapters as belonging to the same book, which means that we had to sum the citation statistics ourselves. This leads to another difficulty – should the book chapters be treated as research articles? It seems that books are increasingly produced as a compilation of research articles and should therefore be treated as equivalent to articles published in a journal. In that sense, GS seems to be the least sensitive to this problem because in GS any type of publication that is cited will appear on the list.

At the moment, both SCOPUS and WoS follow a principle of assigning the credit for citations equally to all authors. This does not seem to be fair and we assume that some decay function should be used to attach weights to authors based on their

importance: e.g. the first authors would get a weight of 0.45, second 0.3, third 0.25, etc. This is something that should definitively be considered in future investigations.

Another difficulty we discovered with all bibliometric databases is the problem of duplicate authors and their exact spatial location. Many authors have the same names (even with two middle letters!), many change locations, some change names, etc. To account for this, SCOPUS Author Identifier functionality distinguishes between these names by assigning each author a unique number. For example, the principal author of this article has two identifiers:

AU-ID("Hengl, Tomislav" 6602864740)

OR AU-ID("Hengl, Tomislav" 13409408400)

where ID 6602864740 refers to the author who has worked at the ITC, Enschede, the Netherlands and ID 13409408400 to the period when the author was an employee of the JRC Ispra in Italy. SCOPUS further allows users to group the same authors with multiple identifiers. WoS does not make this distinction but only keeps the most recent address in the database. On the other hand, an advantage of WoS is that it allows users to zoom into research articles and observe how the citations change for each article.

Unique identifiers such as Digital Object Identifier (DOI) and ISBN for books will be increasingly important – in order to reduce duplication and errors. Similar universal identifiers could also be introduced for authors and other library items such as computer programs, maps and the like. Finally, bibliometrics would benefit significantly if a unique web database were created in which anybody was able to check the correct reference of all library items in the world at any time. This would certainly contribute to the democratization of science and soften the monopolies some scientific information indexing companies currently have. Although Google seems to be closest to reaching this goal, only truly non-profit international associations can serve such purposes. Currently, the best example of such an organization is the Online Computer Library Center (www.oclc.org).

Finally we should mention that spatial display and analysis of bibliometric data has many benefits. Once library items are matched with the address of the corresponding author, various variables can be displayed spatially and spatial patterns explored – density of publications, spatio-temporal trends and geographical distribution of citation statistics. Unfortunately, finding latitude/longitude for many items in our exercise required a lot of manual work. Google Earth seems to be the perfect solution to bring this gap and match the authors' addresses and geographical locations (it will often find correct location – as long as the city name, country and/or postcode are provided – even if the city name is misspelled or incomplete). In that sense, Google Scholar could be improved if the programmers would also include spatial reference, in the PoP software.

Conclusions

The results of this investigation show that geostatistics is an active scientific field in which the number of publications is still increasing. According to GS, the number of citations per publication has reached a steady state, while the authors' h -index seems to have been declining since 1999. We assumed that this is happening because there are an increasing number of younger authors with a smaller h -index, but we do not have the authors' age in our database and so were not able to test this assumption.

The geostatistics books receive equal citations per item as journal articles. The CR can be then used to depict the most cited items on average. The four most cited geostatistics books are "*Statistics for Spatial Data*" (215 citations per year), "*An Introduction to Applied Geostatistics*" (100 citations per year), "*GSLIB: Geostatistical Software Library and User's Guide*" (96 citations per year) and "*Geostatistics for Natural Resources Evaluation*" (83 citations per year). The four most cited journal articles are "Improved global sea-surface temperature analyses using optimum interpolation" (104 citations per year), "Spatial autocorrelation: Trouble or new paradigm?" (48 citations per year), "A statistical-topographic model for mapping climatological precipitation over mountainous terrain" (42 citations per year) and "Spatial pattern and ecological analysis" (39 citations per year). Although most of the library items in the Top 25 list are over 7–8 years old, there are many more recent publications that compete with the publications at the top of the list.

In order to allow a fair comparison between the authors/institutes, more sophisticated indices need to be developed that take into account authors' (scientific) age and the characteristics of a research field. Currently, the citation rate (CR) seems to be more appropriate for a fair evaluation of authors than the number of publications or the authors' h -index. For geostatistics, we discovered that the individual impact of an author (citations per year per article) is correlated with author's number of publications. However, the number of publications explains only 60% of the variation in citations and this number progressively decreases for author with a lower number of publications. The problem with the h -index is that it is a function of the scientific age of an author and it can be much smaller for relatively young authors (who might be writing articles that are just as good as those written by senior researchers). The results of this research have shown that it is probably more useful to evaluate individual publications, extract the hottest articles and then assign credits to authors and home institutes than to focus on authors' h -indices, which can differ significantly from web service to web service.

Any researcher or research organization can be successfully evaluated nowadays using web services such as Web of Science, SCOPUS, Google Scholar or something similar. Objective measures can be used to depict the most influential authors/publications and research institutes/organizations and distinguish them from the hyper-productive authors that nobody reads. If the library items are linked to

geographical location, such data can also be used to generate scientific productivity and excellence maps. The commercial scientific indexing companies could enhance their service if they would consider assigning the geographical location to library items to allow spatial exploration and analysis of bibliometric indices.

*

The authors would like to thank the editors of the journal *Scientometrics* for allowing us to retain the original title of the article. We would like to apologize to numerous colleagues whose work may have been omitted or unintentionally ignored in this study. Our intention is to periodically update the Top 25 lists of the articles, books and authors via the contact author's website, so that the most accurate information is presented.

References

- BAR-ILAN, J. (2008), Which h-index? – A comparison of WoS, Scopus and Google Scholar, *Scientometrics*, 74 (2) : 257–271.
- BATISTA, P. D., CAMPITELI, M. G., KINOCHI, O., MARTINEZ A. S. (2006), Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68 (1) : 179–189.
- BAUER, K., BAKKALBASI, N. (2005), An examination of citation counts in a new scholarly communication environment. *D-Lib Magazine*, 11 (9) : 1–4.
- BIHUI, J. (2007), The AR-index: complementing the h-index. *ISSI Newsletter*, 3 (1) : 1–6.
- CRESSIE, N. (1990), The origins of kriging. *Mathematical Geology*, 22 (3) : 239–252.
- CRESSIE, N. A. C. (1993), *Statistics for Spatial Data*, revised edition. John Wiley & Sons, New York.
- DIGGLE, P. J. (2003), *Statistical Analysis of Spatial Point Patterns*. A Hodder Arnold Publication.
- DOLL, C. N. H., MULLER, J. P., MORLEY, J. G. (2006) Mapping regional economic activity from night-time light satellite imagery. *Ecological Economics*, 57 (1) : 75–92.
- EASTMAN, J. R., FULK, M. (1993), Long sequence time series evaluation using standardized principal components. *Photogrammetric Engineering and Remote Sensing*, 59 (8) : 1307–1312.
- GANDIN, L. S. (1963), *Objective Analysis of Meteorological Fields. translated from Russian in 1965 by Israel Program for Scientific Translations*, Jerusalem, Gidrometeorologicheskoe Izdatel'stvo (GIMIZ), Leningrad.
- GILES, J. (2005), Start your engines. *Nature*, 438 (1) : 554–555.
- HARZING, A. W., VAN DER WAL, R. (2008), Google Scholar: the democratization of citation analysis? *Ethics in Science and Environmental Politics*, in press.
- ISAAKS, E. H., SRIVASTAVA, R. M. (1989), *Applied Geostatistics*. Oxford University Press, New York.
- JACSO, P. (2005), Google Scholar: the pros and the cons. *Online Information Review*, 29 (2) : 208–214.
- JOURNEL, A. G. (1986), Mining geostatistics. *Mathematical Geology*, 18 : 119–140.
- JOURNEL, A. G., HUIJBREGTS, C. J. (1978), *Mining Geostatistics*. Academic Press, London
- KOUSHA, K., THELWALL, M., (2007), Google Scholar citations and Google Web/Url citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*, 5 (7) : 1055–1065.
- LASAPONARA, R. (2006), On the use of principal component analysis (PCA) for evaluating interannual vegetation anomalies from SPOT/VEGETATION NDVI temporal series. *Ecological Modelling*, 194 (4) : 429–434.
- MATHERON, G. (1962), *Traité de géostatistique appliquée, Mémoires du Bureau de Recherches Géologiques et Minières*, Vol 14. Editions Technip, Paris.
- MEHO, L. I., YANG, K. (2007), A new era in citation and bibliometric analyses: Web of Science, Scopus, and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58 : 1–21.

- MINASNY, B., HARTEMINK, A. E., MCBRATNEY, A. (2007), Soil science and the h index. *Scientometrics*, 73 (3) : 257–264.
- NORUZI, A. (2005), Google Scholar: The new generation of citation indexes. *LIBRI*, 55 (4) : 170–180.
- PIWOWAR, J. M., MILLWARD, A. A. (2002), Multitemporal change analysis of multispectral imagery using principal components analysis. In: *Geoscience and Remote Sensing Symposium IGARSS '02*, vol 3, IEEE International, pp. 1851–1853.
- ROEDIGER, H. L. (2006), The h index in Science: A new measure of scholarly contribution. *The Academic Observer* 19 : 1–4.
- STEIN, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*. Series in Statistics, Springer, New York.
- WEBSTER, R., OLIVER, M. A. (2007), *Geostatistics for Environmental Scientists*. Statistics in Practice, Wiley, Chichester.
- YOUDEN, W. J. (1951), *Statistical Methods for Chemists*. John Wiley & Sons, New York.
- ZHOU, F., HUAI-CHENG, G., YUN-SHAN, H., CHAO-ZHONG, W. (2007), Scientometric analysis of geostatistics using multivariate methods. *Scientometrics*, 73 : 265–279.