

## Review Article

Elżbieta Bielecka\* and Elżbieta Burek

# Spatial data quality and uncertainty publication patterns and trends by bibliometric analysis

<https://doi.org/10.1515/geo-2019-0018>

Received Oct 02, 2018; accepted Mar 05, 2019

**Abstract:** Using the literature review and quantitative analysis, the research on the quality and uncertainty of spatial data have been compared and analysed according to years of publication, authors, document types, WoS categories, and countries. The paper portrayed the development in the field, studied the state and evolution of the most productive and influential journals, conferences, and research institutions. The results showed that remote sensing, computer science, and geography relate mostly to data imperfection and assessment of its uncertainty. This relation is clearly translated into the most productive journals, and conferences proceedings. The top-ranked countries in this field are United States, China, and the United Kingdom.

**Keywords:** data imperfection, trust in spatial data, bibliometry, citation analysis, Lorentz curve, concentration

## 1 Introduction

### 1.1 Spatial data quality and uncertainty – overview of research development

The heterogeneity of the real world, technologies for data acquisition and processing, database management tools and platforms leads to a large amount of duplicated, inconsistent, ambiguous, and incomplete spatial data. Therefore spatial data quality and uncertainty (SDQ&U) is an increasingly important issue in geographical information science with thousands of publications in countless journals, conferences, and books. At the early stage of GIS development (from the 1960s until the mid-1980s), the imperfection of spatial data was mainly expressed as errors in geographical position and topology [1–3]. In the following years, the development of GIS technology and its use in

the decision-making process have fostered increasing interest in the quality and uncertainty research [4, 5]. Issues of quality and uncertainty of spatial data have become even more important because data imperfections propagated through spatial analysis affect the decision-making process. Recently, researchers have focused their scientific attention on uncertainty modelling and the final impact of data imperfection on the spatiotemporal analysis [6]. Moreover, since 2004 data delivered by volunteers (crowdsourcing data) have become the subject of extensive research. These studies have a broader context, analysing not only different types of errors but also the behaviour of volunteers expressed in form of their mapping activity in a given area *e.g.*: [7–12].

Although the concepts of uncertainty in spatial data and spatial data quality are similar, the standards have evolved only for the latter. The quality of data is understood as ‘degree to which a set of inherent characteristics fulfils requirements’ [13] with respect to the immanent attributes related to the geospatial nature of the data, like positional accuracy or spatial resolution. This definition was further adapted in ISO 19100 series for geographical information, *e.g.* ISO 19101-2:2002 Geographic information – Reference model [14], ISO 19157:2013 Data quality [15] and ISO/TS 19158:2012 Quality as assurance of data supply [16]. Uncertainty expressed as a ‘parameter, associated with the result of a measurement that characterises the dispersion of the values that could reasonably be attributed to the measured’ was defined in the International vocabulary of metrology – Basic and general concepts and associated terms (VIM) - ISO/IEC Guide 99:2007 [17]. However, the issue how to evaluate and describe data uncertainty still remains open.

Many international organisations include theory of quality and uncertainty of spatial data in their research programmes. The International Cartographic Association (ICA) research agenda focuses on visualization of data quality in general, and spatially varying quality in particular [20]. The International Federation of Surveyors (FIG) deals with the quality of data as part of the work of the Commission III Spatial Information Management in the field of establishing data-quality-standards relevant to

\*Corresponding Author: Elżbieta Bielecka: Military University of Technology Warsaw, Poland; Email: [elzbieta.bielecka@wat.edu.pl](mailto:elzbieta.bielecka@wat.edu.pl)  
Elżbieta Burek: Military University of Technology Warsaw, Poland

spatial information management in cooperation with international spatial data standard committees [21]. Great interest in data quality and uncertainty is also visible in the International Society of Photogrammetry and Remote Sensing (ISPRS) program, where several Commissions deal with this issue, including problems related to quality and uncertainty modelling as well as the semantic modelling and linking of ontologies. ISPRS-WII/4 working group has recently introduced the new term ‘trust in spatial data’ [6], which refers to the level of matching real-world phenomena with their abstraction in a database. Data quality and data usability are a priority in the Association of Geographic Information Laboratories in Europe (AGILE) scientific programme [22] with the number of papers on spatial data quality growing consistently [23]. Regardless of the conferences organized by the aforementioned associations (e.g. International Symposium on Spatial Data Quality (ISSDQ) or International Workshop on Spatial Data Quality), several national seminars and conferences on data quality issues take place every year. Furthermore, SDQ sessions are arranged within the framework of most conferences related to geographic information, spatial data infrastructure, data analysis, and management.

The steady rise of SDQ&U research results in a large and growing number of publications. Analysis of these publications is possible thanks to bibliometrics, *i.e.* a statistical analysis of publications.

## 1.2 Bibliometrics

The term ‘bibliométrie’ was coined by the Belgian businessman Paul Otlet in 1934 and defined as ‘the measurement of all aspects related to the publication and reading of books and documents’ [24]. Thirty-five years later, the term was popularised by Pritchard in the publication entitled “Statistical Bibliography or Bibliometrics” [25].

Although there are numerous bibliometric databases, the most commonly used ones are: Web of Science (WoS), Scopus and Google scholar [26–29]. Web of Science, owned by Clarivate Analytics (previously by Thomson Reuters), has been the foremost important source of analysing bibliometrics as it has a world-wide and multi-disciplinary coverage of refereed journals (approximately 10,000). As stated by Bar-Ilan [29], until 2004, the Science Citation Index Expanded (SCIE), the Social Science Citation Index (SSCI) and the Arts and Humanities Index databases, available together by using the WoS, were the only exhaustive sources of data citation. An alternative to WoS is Scopus developed by Elsevier. It indexes more

than 15,000 refereed journals, but its reference lists have been indexed consistently only since 1996. Both databases belong to commercial providers and require an access fee. They vary in terms of the covered period: WoS has been archiving publications since 1900, while Scopus – since 1966. WoS and Scopus have been dominant among the academic society, mainly through the annual release of the journal impact factor, an essential tool that evaluates the significance and scientific position of a given publication. On the other hand, Google Scholar, developed by Google Inc., is freely accessible. This database indexes most of scientific literature available on the Web, e.g.: books, preprints, journals, reports, books of abstracts, as well as files from digital archives. There is a debate in academic community on whether Google Scholar might become an alternative to the commercial citation databases as a source for evaluating studies, primarily due to the unknown quality of the resources indexed and overall policy [26–30].

Bibliometrics is perceived as an effective tool for analysing and monitoring scientific achievements and research trends in numerous disciplines of science, technology, and humanities. An essential component of bibliometrics is citation analysis, as it is used to indicate the impact of publications and expresses the significance of the obtained results for other, later studies. Garfield [32] enumerated fifteen reasons why scientists refer to existing publications. Among them giving credit to pioneers and peers, providing background for methodology, as well as correcting own work and the work of others are of utmost importance. However, many researchers criticised citation analysis and demonstrated its limitations. The main drawbacks, as found by Smith [33], are: (1) only a small part of what is read is cited; (2) not all the best publication in specific fields are easily accessible; (3) impact is not equivalent with quality, because papers could be referred to for reasons irrelevant to their merit; (4) bibliographic coupling and self-citations. Recently, the popularity of bibliometric mapping of science has increased [34]. It comprises a set of methods to create an overview and to depict the development of a given research field.

While many bibliometric studies have been carried out in disciplines dealing with spatial data including remote sensing [35, 36], GIS [37] and GISciences in general [38] the issues of spatial data quality and uncertainty have been broadly summarised only during a panel discussion that took place at the Sixth International Symposium on Spatial Data Quality (ISSDQ), held in Canada, in July 2009 [18].

This paper aims to provide a statistical overview of spatial data quality studies by bibliometric analysis of publications indexed in WoS from 1990 until the 23<sup>rd</sup> of May

2018. WoS was selected after an in depth literature review. This decision was supported by the following reasons: multidisciplinary *i.e.* wide and exhaustive indexing of the journals, books, and proceedings, the time span (since 1990), as well as a warranty of high quality of indexed publications [26, 29, 31, 39, 40].

We reveal the patterns and trends of scientific publications, geographical distribution, as well as the most productive journals and authors along with the most cited papers. The research output and impact in specific fields of research was also determined. The analysis provides several insights which may aid researchers, data providers, and public administration in understanding the development of the field. The next section (section 2) provides a description of methods and data used. Then the results are presented and discussed in sections 3 and 4. The paper ends with a brief concluding section.

## 2 Methods and Data

Data on papers addressing spatial data quality issues have been retrieved from the Web of Science (WoS) Core Collection with use of the online search application. The following query: 'spatial data' quality OR 'spatial data' uncertainty OR geodata quality OR geodata uncertainty OR 'geographic data' quality OR 'geographic data' uncertainty OR 'geospatial data' quality OR 'geospatial data' uncertainty was used to extract the relevant papers. Data were collected on the 23<sup>rd</sup> of May 2018. Phrases 'spatial data', 'geographical data', 'geospatial data', and 'geodata' were used to cover all cultural differences in defining data related to a geographic location *i.e.* location on the Earth. They are considered synonyms and are most often used interchangeably as stated by [41]. The WoS was searched through the general search interface, including such fields as: author(s), author identifier, title, abstract, keywords, keywords plus, publication name, document type, publication year, addresses, organization-enhanced, conference. The study consisted of two consecutive stages, namely: pre-processing and analyses. The pre-processing stage includes data cleaning and data sorting. Data cleaning comprises mainly checking authors with the same surname and different initials, *e.g.*: SHI W, who also published as: SHI WEI, SHI WZ, SHI WENZHONG, SHI WEN ZHONG, SHI W Z, or SHI WEN. Finally, to facilitate further analysis, publications of authors with different initials were merged. The cleaned data were then sorted by: years, authors, organisations, document types, times cited, and other bibliographic details. This enabled us to conduct further anal-

yses of SDQ&C publications patterns and trends. The analysis embraces: (1) general publication output and citation analysis from 1990 to 2018; (2) data screening to find some most cited papers and prominent authors, which allowed us to show key research topics as well as problems that remain unresolved in the field of SDQ&U studies; (3) focusing on research categories, journals and conferences to portray the scientific disciplines in which research on the spatial data quality is of utmost importance. Furthermore, exploration of titles, abstracts, and keywords showed some regional diversity in defining concepts related to spatial data imperfection. In particular, the following aspects were investigated to present the state and reveal the trends in research on spatial data quality and uncertainty.

1. Publication output and citation analysis - by using essential sciences indicators, *e.g.* total number of publication (TP), total number of citations (TC), average number of citation per publication (CPP) as defined in INCites Indicator Handbook [40].
2. Inequalities within subject categories, journals, conferences, and books expressed by the number of publications (TP), coefficient of dispersion ( $cv$ ), Lorentz curve, and Gini index.
3. Focus on the most productive journals (journal impact factor - IF, NP), authors (h index, NP, TC), and highly cited papers.
4. Productiveness of countries and institutions conveyed by the strength coefficient ( $s_{ij}$ ) and shown as a co-operation network.
5. Key words analysis, particularly co-occurrence between terms expressed by the strength coefficient ( $s_{ij}$ ) and the number of uses in the 5-year window.

Data were collected on the 23<sup>rd</sup> of May 2018. Phrases 'spatial data', 'geographical data', 'geospatial data', and 'geodata' were used to cover all cultural differences in defining data related to a geographic location *i.e.* location on the Earth. They are considered synonyms and are most often used interchangeably as stated by [41].

The coefficient of dispersion ( $cv$ ) is a measure used to quantify whether a set of observed occurrences are clustered or dispersed as compared to a standard statistical model. It is expressed as the variance ( $\sigma^2$ ) divided by the mean:

$$cv = \frac{\sigma^2}{\mu} \quad (1)$$

The coefficient of dispersion ( $cv$ ) is equal to 1 for a random or Poisson distribution, while  $cv > 1$  indicates under-dispersion or aggregation, and  $cv < 1$  shows over-dispersion or an even distribution.

The Lorenz curve is generally used to represent economic inequality, mainly income or wealth, although it could also be used to denote unequal distribution in any system, e.g.: the number of publication in journals, conference proceedings, or science categories. The Lorenz curve is a function of the cumulative proportion of ordered publication counts mapped onto the corresponding cumulative proportion of their number from 1990 till the 23<sup>rd</sup> of May 2018. The curve is expressed as:

$$L(y) = \frac{\int_0^y x dF(x)}{\mu} \quad (2)$$

where  $F(x)$  is the cumulative distribution function of ordered individuals and  $\mu$  is the average size.

The Gini index (Eq. 3), which is derived from the Lorenz Curve, was used as an indicator of output inequalities in WoS research categories, journals and conference proceedings. It ranges from 0 for perfect equality (the same number of publications are assigned to every subject category, journal or conference proceeding) to 1 for perfect inequalities.

$$G(x) = \frac{\sum_{i=1}^n (2i - n - 1) x'_i}{\mu n^2} \quad (3)$$

of  $n$  ordered individuals with  $x'_i$  the size of individual  $i$  and  $x'_1 < x'_2 < \dots < x'_n$ , where:  $n$  – number of WoS research categories (or journals or conference proceedings),  $x'_i$  – number of publications in a WoS research category (or journal or conference proceeding) in ascending order  $< x'_1 < x'_2 < \dots < x'_{n-1} < x'_n$  and  $\mu$  – mean value.

The co-coupling analysis of counties, institutions, authors and words based on the association of strength coefficient  $s_{ij}$ , which was described in detail in van Eck and Waltman [42]:

$$\begin{aligned} s_{ij} &= \frac{2mc_{ij}}{c_i c_j} \\ c_i &= \sum_{j=1} c_{ij} \\ m &= \frac{1}{2} \sum_i c_i \end{aligned} \quad (4)$$

where:  $c_{ij}$  – the number of links (e.g., co-occurrence links, co-citation, bibliographic coupling) between nodes  $i$  and  $j$  ( $c_{ij} = c_{ji} \geq 0$ );  $c_i$  – the number of links of node  $i$ ;  $m$  – the total number of links in the network.

The contributions of countries and institutions were investigated by the affiliation of authors'. No collaboration was noticed where the authors of a publication were affiliated at the same institution. National collaboration was

designated to authors from the same countries, but different institutions, while international collaboration was assigned to those who published with researchers from at least two countries.

Authors' key words and key words plus were used to analyse the co-wording. Key words plus were added by Thomson Reuters editors to highlight additional relevant but unnoticed key words that were not listed in a publication.

The networks of institutions, authors, and key words were elaborated via VOSviewer, the software tool developed by van Eck and Waltman [42, 43]. The importance of institutions, authors, and key words was expressed in form of the size of circles and labels. This importance was measured by *total link strength* ( $s_{ij}$ ), an attribute that indicates the total strength of a given institution/ author/ key word with other institutions/ authors/ key words, respectively. The colour designates the cluster, while the lines represent the links. The distance between a cluster and items belonging to the cluster shows the relatedness of the institutions/ authors / key words to each other [44].

## 3 Results

### 3.1 Publication outputs

The total number of publications (TP) on spatial data quality and uncertainty from 1990 till May 23 2018 equals 2,090, till the end of 2017 – 2,069. Only one paper related to SDQ&U was published in 1990, a year later – as many as 6 papers. During the last decade of the 20<sup>th</sup> century the total number of publication grew to 155, at the end of the first decade of the 21<sup>st</sup> century it rose 6-fold, and in the beginning of 2018 (23 May) 2.25-fold, and reached the value of 2,140. The annual increase in the global number of publications on SDQ&U could be described by the second degree polynomial (R square = 0.9485), with three obvious peaks in 2003, 2009-2010 and 2016, marked in dark blue in Figure 1. Research articles accounted for 59.4%, followed by

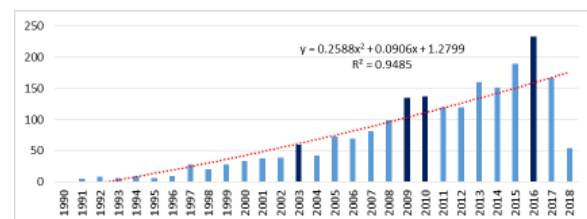


Figure 1: The increase in the annual number of publications since 1990 till May 2018

**Table 1:** Annual research outputs in spatial data quality and uncertainty

PY	TP	PC	PC/P	TC	CPP	AU	J	PPJ	CU
1990	1	17	10.8	0	0.00	1	1	1.0	1
1991	6	59	11.7	20	65.17	17	5	1.0	2
1992	14	65	12.1	34	33.79	11	5	1.6	1
1993	20	91	11.3	54	28.25	15	5	1.2	3
1994	30	100	12.0	84	29.83	23	9	1.1	5
1995	36	92	10.5	120	35.86	14	5	1.2	5
1996	45	113	12.5	165	30.29	19	10	0.9	6
1997	73	365	11.5	238	25.86	60	26	1.1	14
1998	93	197	14.6	331	23.10	53	21	1.0	7
1999	121	412	11.7	452	23.04	74	30	0.9	15
2000	155	313	10.2	607	21.42	94	28	1.2	15
2001	193	462	9.1	800	22.81	110	37	1.0	15
2002	232	502	9.7	1032	21.84	105	47	0.8	17
2003	292	765	11.8	1324	20.68	185	58	1.0	21
2004	334	481	11.5	1658	20.24	133	41	1.0	19
2005	407	858	12.8	2065	19.79	225	62	1.2	28
2006	476	670	12.9	2541	20.10	219	61	1.1	24
2007	557	741	12.2	3098	19.07	276	81	1.0	33
2008	656	1008	9.2	3754	18.34	291	88	1.1	34
2009	791	1578	14.7	4545	16.88	399	114	1.2	31
2010	928	2005	9.9	5473	16.55	391	117	1.2	33
2011	1048	1379	13.0	6521	15.84	405	117	1.0	39
2012	1168	1500	12.6	7689	15.68	413	114	1.1	40
2013	1328	1674	15.3	9017	14.62	578	152	1.1	45
2014	1479	1819	10.0	10496	13.89	557	148	1.0	52
2015	1669	2144	15.2	12165	12.79	706	168	1.1	51
2016	1902	2828	8.1	14067	11.47	809	189	1.2	57
2017	2069	1947	11.8	16136	10.60	673	149	1.1	48
2018	2140	555	17.0	18224	10.51	181	55	1.0	32

PY: publication year, TP: total number of publications, PC: page count, PC/P: page count per publication, TC: total cited references count, CPP: cited references per publication, AU: number of authors, J: number of journals (conferences), CU: number of countries, PPJ: average number of publications per journal.

proceedings papers (34.1%), book chapters (3.2%), and reviews (2.1%). The remaining 1.2% were editorial materials; book reviews, books (as many as 5), and meeting abstracts.

The average article length varied slightly with the coefficient of dispersion equal to  $cv = 0.172$ . A typical publication had approximately 12 pages ( $RMS=0.38$ ) and was written by 3 authors affiliated to 2 or 3 countries.

The publications on SDQ&U were cited 22,737 times by 20,352 articles. The rate of self-citations was just 4.4%. The average number of references per article (excluding self-citations) was 10.60, and it fluctuated moderately with the coefficient of dispersion  $cv = 0.53$ . As many as 781 publications (35.5%) were not cited, while 223 (which is 10%) were referred to only once. The percent of publica-

tions that were quoted at least 10 times slightly exceeds 25. Only 37 scientific papers were cited at least 100 times. The distinct growth in the number of publications and the citations shows the steady increase and communication in the SDQ&U research during the past three decades (Table 1). The correlation between the number of authors who conducted research on SDQ&U and the number of journals is significant, Pearson's  $r=0.989$ .

Moreover, a strong positive correlation was found between the annual total number of publications and the cumulative references count. On the other hand, cited references per publication are characterized by a moderate, negative correlation with the number of publications and the total count of references, and a strong negative corre-

**Table 2:** Correlation matrix

	time span	NP	TC	CPP
time span	1			
NP	0.93	1		
TC	0.89	0.99	1	
CPP	-0.81	-0.56	-0.52	1

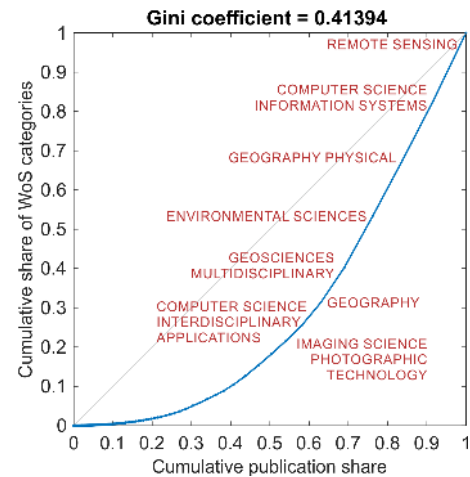
lation with the time that passed between the publication year and the year of our analysis (Table 2).

The first paper indexed in WoS, entitled “Generation of Project Base Maps Using US geodata for the National Water-Quality Assessment Program”, was written by Mladinich [45], and published in Technical Papers of 1990 ACSM-ASPRS Annual Convention, vol 2: Cartography. This article has not been cited yet. The first cited articles were published a year later, in 1991. Since May 2018 they were cited 313 times without self-citation, with an average number of 62.6 citations per item.

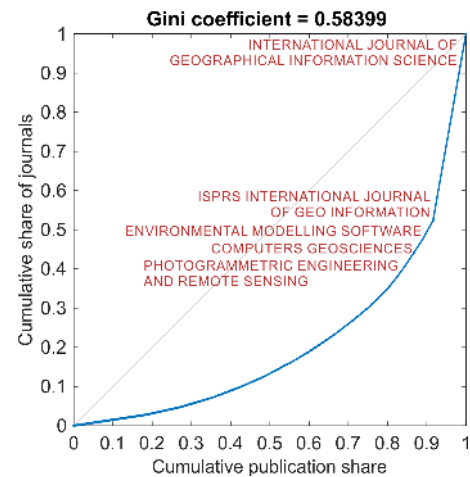
### 3.2 Subject categories, journals and conferences

Global research on spatial data quality and uncertainty spanned over 145 WoS research categories, which accounts for 41.4% of all research areas. For 27 categories, the number of publications was one, and for the next 14 only two publications. The median of the number of publications in the WoS categories equals 7, while the mean value - 32.15. 90% of publications on SDQ&U fall into WoS categories. This diversification is best reflected by Lorentz curve and Gini coefficient (Figure 2). The blue line shows the distribution of publication output among the WoS categories from 1990 till the 23<sup>rd</sup> of May 2018, while the thin, grey line at the 45° angle presents perfectly equal distribution of publications. The further away from the diagonal (e.g.: Geosciences Multidisciplinary, Geography, Computer Science Interdisciplinary Applications, and Imaging Science Photographic Technology), the more unequal is the distribution.

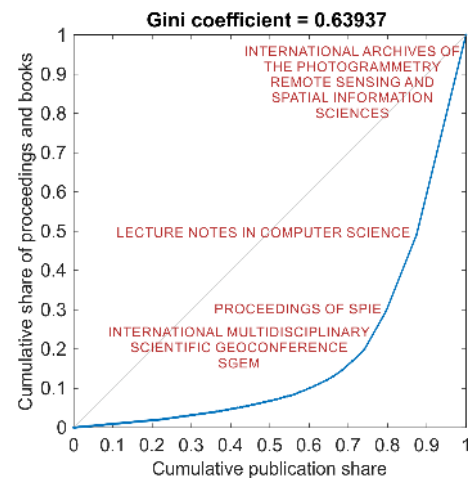
Remote Sensing accounted for 18.8% of all publications (402 papers), with an exponential increase in the number of publications in 5-year windows. The next most productive WoS category - Computer Science Information Systems was attributed to 17.5% papers, and until the year 2010 the number of these publications has grown steadily. In the second decade (2011-2018), the number of publications in this category fell by almost 50%. Geography physical and Geography comprised 10.9% and 17.0% output, re-



(a) WoS categories



(b) Journals



(c) Conferences proceedings

**Figure 2:** Statistical dispersion of SDQ&U outputs across subject categories, journals and proceedings

**Table 3:** Most productive journals in spatial data quality and uncertainty research

Journal	TP	% of all pub.in a journal	IF	IF 5-year (R)	TC	CPP	AU	AU/P	CU
International Journal of Geographical Information Science	75	4.08	2.502	2.545 (9)	1408	18.77	205	2.7	30
ISPRS International Journal of Geo-Information	24	2.34	1.502	1.672 (16)	127	5.29	79	3.3	14
Environmental Modelling Software	20	0.62	4.404	4.979 (3)	353	17.65	68	3.4	17
Computers Geosciences	20	0.41	2.533	2.818 (7)	433	21.65	74	3.7	16
Photogrammetric Engineering and Remote Sensing	18	0.31	2.493	2.512 (10)	638	35.44	63	3.5	10
Transactions in GIS	17	3.07	2.252	2.216 (11)	397	23.35	53	3.1	11
Geodetski Vestnik	14	1.86	0.234	0.290 (17)	19	1.36	24	1.7	2
ISPRS Journal of Photogrammetry and Remote Sensing	13	0.66	6.387	6.457 (1)	256	19.69	48	3.7	15
International Journal of Digital Earth	13	2.56	2.292	2.978 (6)	384	29.54	40	3.1	14
Science of the Total Environment	12	0.05	4.900	5.102 (2)	263	21.92	66	5.5	11
International Journal of Remote Sensing	11	0.10	1.724	1.986 (13)	119	10.82	31	2.8	7
Environmental Monitoring and Assessment	11	0.11	1.687	1.974 (14)	93	8.45	48	4.4	10
Environmetrics	11	0.76	1.532	1.84 (15)	199	18.09	25	2.3	7
Journal of Environmental Management	10	0.12	4.010	4.712 (4)	207	20.70	40	4.0	9
Applied Geography	10	0.42	2.687	3.401 (5)	280	28.00	30	3.0	8
Stochastic Environmental Research and Risk Assessment	10	0.59	2.629	2.722 (8)	67	6.70	45	4.5	12
Journal of the American Water Resources Association	10	0.40	1.717	2.158 (12)	305	30.50	32	3.2	5

TP: total number of publications on SDQ&U, TC: total cited references count, IF: 2016 ISI impact factor; CPP: cited references per publication, AU: number of authors, CU: countries; R-rank

spectively. In both categories, a gradual, linear increase in the number of publications was observed. On the other hand, the outputs denoted to Environmental Sciences and Geosciences Multidisciplinary categories grew gradually and covered respectively 16.5 and 13.5% of publications on SDQ&U.

SDQ&U research works were published in 832 SCI indexed journals in the years 1990-2018 (May). There is a significant concentration of articles in 17 journals, which take

98% of all publications on SDQ&U (Table 3). This large dispersion is visible on the Lorenz curve (Figure 2b), and also confirmed by a high value of the coefficient of dispersion  $cv = 1.82$ .

The International Journal of Geographical Information Science published the most, *i.e.* as many as 75 articles on SDQ&U, which constitutes 3.6% of the total output. It is followed by ISPRS International Journal of Geo-Information (24 papers, 1.5%), Environmental Mod-

**Table 4:** Top 5 conferences proceedings (books) with papers (chapters) on SDQ&U

Titles	TP	% of all papers	TC	CPP	AU	CU
International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences	87	4,163	55	0,63	279	38
Lecture Notes in Computer Science	54	2,584	273	5,06	148	24
Proceedings of SPIE	38	1,818	8	0,21	139	13
International Multidisciplinary Scientific Geoconference SGEM	21	1,005	8	0,38	55	6
IEEE International Symposium on Geoscience and Remote Sensing IGARSS	19	0,909	6	0,32	64	10

TP: total number of publications in the field, TC: total cited references count, CPP: cited references per publication, AU: number of authors, CU: countries

eling Software (20, 0.96%), and Computers Geosciences (20, 0.96%). However, SDQ&U related articles are a small fraction of all papers, assuming the highest percentage (4.08%) for International Journal of Geographical Information Science and 3.07% for Transaction in GIS respectively (Table 3. column 3). On average, the percentage of papers on SDQ&U in the aforementioned journals slightly exceeds 1%. The most influential journals that published research papers on data quality were ISPRS Journal of Photogrammetry and Remote Sensing as well as Science of the Total Environment, with the 5-year impact factor of 6.457, and 5.102 accordingly. At the same time, Photogrammetric Engineering and Remote Sensing and International Journal of Digital Earth provide the highly cited articles with CPP equals 35.44 and 29.54 respectively. It is worth noting that publications related to crowdsourcing data, mainly OpenStreetMap, are most frequently cited.

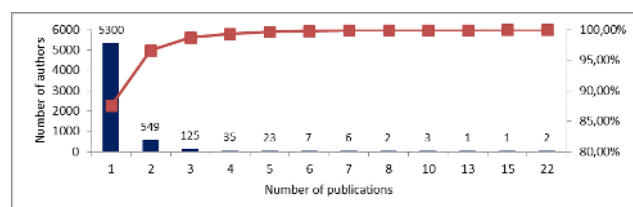
Articles published in 17 top journals were cited 18.70 times on the average, achieving a considerable influence on further research. The most frequently cited paper with an average annual number of citations amounting to 32.89 is entitled “Crowdsourcing geographic information for disaster response: a research frontier”, and it was published in 2010 in International Journal of Digital Earth. The article was written by Goodchild, M. F. and Glennon, J. A. from University of California Santa Barbara, USA, and since 21 May 2018 it has been cited 289 times.

Research works on SDQ&U were presented at 448 international conferences, with the average number of papers equal to 2.56. The dispersion of the number of papers in conference proceedings measured by Gini coefficient equals 0.64 (Figure 2c) and is lower than WoS categories and journals (Gini – 0.56 and 0.41 respectively). The five most productive conference proceedings are presented in Table 4.

The highly cited conference paper “Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent mode” written by Song Sen, Liu Liang; Edwards Scott V. and Wu Shaoyuan was published in 2012 in Proceedings of the National Academy of Sciences of the United States of America. The paper was cited 179 times, while the average number of citations per year equals 26.42. It is a result of co-operation between American and Chinese universities.

### 3.3 The most influential authors

Eight out of 6,053 authors were the most productive and they have made outstanding achievement on SDQ&U research. Among these authors Shi W.Z. from Hong Kong Polytechnic University and Goodchild M.F. from University of California published the largest number (25 and 22) articles, while Gelfand A.E. received the highest citation rate (28.88). Over 87.5% researchers have published just one article, which indicates that quality issues were treated as part of larger projects. The most productive authors in the field have conducted their research at universities in the USA (4 people), Europe (3 people: Netherlands, Spain, Germany), Canada (1), China (1), and Tunisia (1) (see Table 5).

**Figure 3:** Number of publications by authors



**Table 5:** The 10 most productive authors 1990 - 2018 (23 May) ordered by the number of publications related to spatial data quality

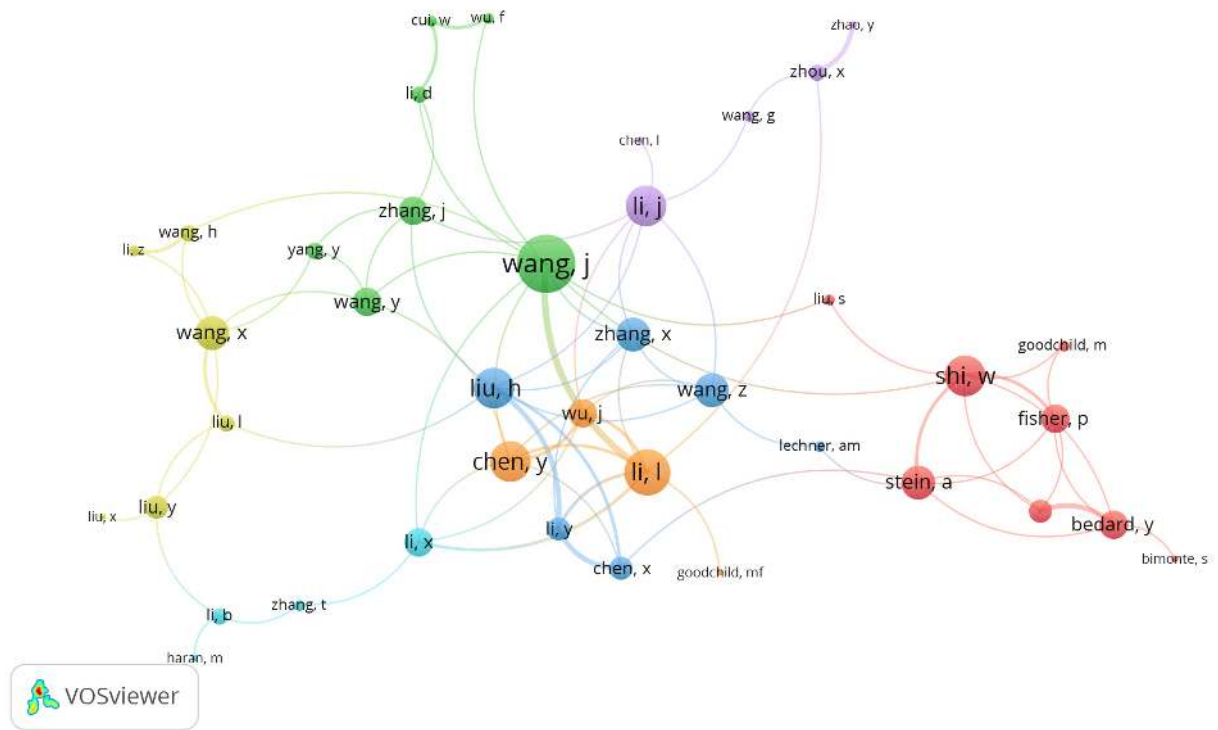
Authors	Institution/ Country	TP	%ISI (R)	h-index (R)	TC	% STC (R)	CPP (R)	Main research area
Shi W.Z. <sup>1</sup>	Hong Kong Polytechnic University, China	25	15.11 (5)	28 (4)	59	3.61 (7)	5.93 (6)	Remote sensing, imaging science photographic technology
Goodchild M.F. <sup>1</sup>	Univ Calif Santa Barbara, USA	22	10.58 (7)	39 (2)	538	10.38 (4)	24.45 (2)	Geography, geography physical
Stein A.	Univ Twente, Nederlands	13	4.47 (9)	37 (3)	148	3.17 (8)	11.38 (5)	Remote sensing, environmental geosciences
Beaubouef T.	Se Louisiana Univ, USA	10	43.48 (1)	6 (8)	45	18.07 (3)	4.50 (7)	Computer science, artificial intelligence
Petry F.E.	Se Louisiana Univ, USA	10	8.85 (8)	18 (6)	20	1.68 (10)	2.00 (9)	Computer science artificial intelligence
Devillers R	Mem Univ Newfound- land, Canada	8	16.67 (4)	11 (7)	116	25.72 (5)	14.50 (4)	Environmental sciences ecology, computer science
Gelfand AE	Duke Univ, USA	8	3.67 (10)	45 (1)	231	2.04 (9)	28.88 (1)	Statistics probability
Ariza- Lopez FJ	Univ Jaen, Spain	7	23.33 (3)	5 (9)	19	29.69 (1)	2.13 (8)	Remote sensing, geology, physical geography
Faiz S	LTSIRS Lab, University of Tunis, Tunisia	7	24.14 (2)	3 (10)	3	10.00 (5)	1.03 (10)	Computer sciences, engineering, remote sensing
Hofle B	Heidelberg Univ, Germany	7	11.29 (6)	19 (5)	43	3.65 (6)	19.31 (3)	Remote sensing, geography physical

TP – total publication related to SDQ&U, %ISIP – % of all ISI indexed publication; h-index – Hirsch index; TC – total citations of SDQ&U papers; % STC - % sum of all times ISI publications; CPP – total citations per all ISI publications; R – rank

<sup>1)</sup> publications of Goodchild M.F. and Goodchild M. as well as Shi W. and Shi W.Z. were put together; both authors have published their research using different initials of their middle names. Merging was done after an in-depth analysis of their affiliations.

Papers related to spatial data quality had the highest percentage share in the publishing output of Beaubouef T. (from Se Louisiana Univ, United States). 10 out of 23 publications indexed in ISI (43%) addressed the SDQ&U issue. For comparison, with Goodchild M.F., who published the most on spatial data quality, this percentage was just over 10% (see Table 5). The most productive authors conducted their research in Remote Sensing, Geography physical, and Computer Science.

Only 40 scholars grouped in 7 clusters have worked on spatial data quality in broader national or international networks, which is reflected in the number of common publications equal or greater than 5. The collaborations network of these authors is presented in Figure 4. The circle size shows the authors' co-operations measured by the number of links between other scholars, while the thickness of lines demonstrates the co-operation strength expressed by the strength coefficient  $s_{ij}$ . The distance be-



**Figure 4:** Network of scholars' co-authorship in the field of spatial data quality

tween the authors indicates the relatedness of collaboration.

The most collaborative authors are: Wang J. ( $s_{ij} = 13$ ), Li L. ( $s_{ij} = 13$ ), Liu H. ( $s_{ij} = 10$ ), Shi W. ( $s_{ij} = 9$ ), Chen Y. ( $s_{ij} = 7$ ), and Stein A. ( $s_{ij} = 7$ ). However, only the authors grouped in the red cluster represented extensive international co-operation (China, the Netherlands, Canada, USA, and the United Kingdom). The quantity of multi-author papers roughly corresponds to the content, accounting for only 14.6% of all publications in the field.

### 3.4 Institutions and countries productiveness and geographic distribution

There were 1,928 research institutions that participated in spatial data quality research, out of which 234 conducted neither national nor international cooperation in the field of spatial data quality. Scholars from the Chinese Academy of Agricultural Sciences and Wuhan University (China) published the most papers related to research on spatial data quality. They are followed by two Australian universities: of Melbourne and Queensland. The

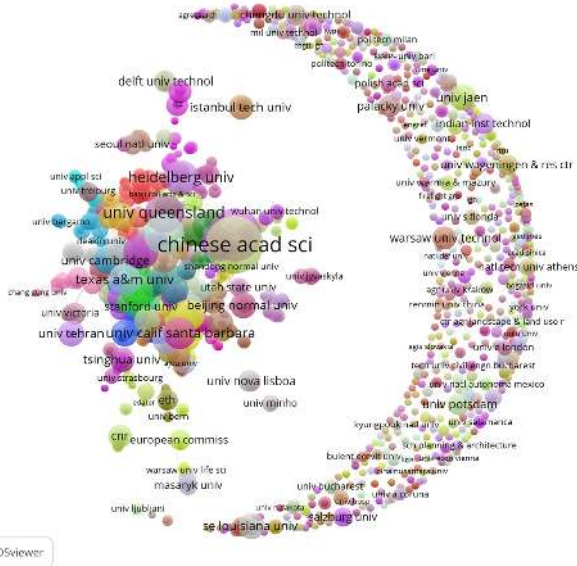
Hong Kong Polytech Univ. was ranked in the fifth position. The 50 top scientific institutions conducting research in the field were dominated by 16 universities from the United States. The top ones included five universities from China (including Singapore and Hong Kong) and 16 from Europe, e.g.: universities from the UK, Germany, Italy, Netherlands, and France as well as international organisations like the International Union for Conservation of Nature (IUCN), located in Switzerland. The co-operation network is presented in Figure 5. The size of the circle and the label are determined by the potency of an institution. The distance between two institutions indicates their relationship in terms of co-citation links: the closer two institutions are the stronger the co-operations. Collaboration in the SDQ&U research is frequent, and is dominated by national co-operations. The most internationally collaborative countries are: USA-Canada-Australia-UK-Netherlands. China, Germany, France, Italy, and India have started broad international co-operation 2010 onwards. Five years later, Eastern European countries such as the Czech Rep., Poland, and Romania also joined international cooperation in the field of spatial data quality research.

Research on spatial data quality and uncertainty has been conducted in 95 countries from all continents (Fig-

**Table 6:** The 10 most productive countries in spatial data quality and uncertainty research during 1990-2018 (May)

Countries	TP	TPj	TPcp	TC	CPP (R)	R&D	TP GISs	TP RS	TP AI
USA	561	438	123	10372	18.62 (1)	405.3	3804	23914	4144
China	312	143	169	1433	4.59 (10)	337.5	1938	8989	2392
UK	189	143	29	2752	14.56 (5)	21.1	915	4594	830
Germany	167	112	48	2462	14.74 (4)	26.8	713	4918	590
Australia	154	100	24	2739	17.78 (2)	38.4	509	2442	1718
Canada	94	73	21	1058	11.25 (7)	24.3	699	4344	823
Italy	89	63	28	1120	12.58 (6)	19	377	4361	797
France	89	58	22	763	10.17 (9)	42.2	287	4509	869
The Netherlands	79	54	18	1244	15.74 (3)	10.8	522	2247	304
India	60	34	24	394	6.57(9)	36.1	183	3195	643

TP - total publications in the field; TPj – total articles in JRC indexed journals; TPcp – total papers in WoS indexed conference proceedings; TC – total citation counts in the field; CPP – average citations per publication; R – rank; R&D – investments in research and development in billions; TP GISs – total publications in GIScience journals [38]; TP RS – total publication in remote sensing [35, 36]; TP AI – total publications in artificial intelligence [46].

**Figure 5:** Network of institutional co-operation

ure 6). The first rank belongs to the USA (557 publications), followed by: China, Germany, the UK, Australia, Canada, Italy, the Netherlands, France, and India (Table 6). Poland, a Central European developing country, was ranked as the twelfth country (88.2%), after Spain (89.3%). Norway was the most influential country, if average citation per publication is considered (CPP=45.14). The CPP was 21.26 and 18.31 for the United States and the United Kingdom respectively. Chinese publications are relatively the least cited (CPP=4.73), which shows that they have had low impact on world science so far (Figure 6).

On the other hand, scholars from 19 countries have published just one publication, 13 of which have not been

referenced till May 2018. Nine countries with two publications were ranked in the 67<sup>th</sup> position (21.2 percentile of the output).

In general, a linear growth in the number of publications was observed in the 5-year window starting in 1990 (Figure 7). The exceptions are China, Canada and the Netherlands, where after 2009 a slight decrease in the number of publications was visible.

The analysis of previous bibliometric studies concerning publications in GIScience journals (after [38]), remote sensing (after [35, 36]), and artificial intelligence (after [46]) indicates a very strong correlation between these research areas and spatial data quality ( $R^2=0.98$  and  $R^2$  adjusted amounts 0.96). Moreover, the number of publications is strongly positively correlated with country investment in research and development: the Pearson'  $r$  coefficient equals 0.92. This dependence is also underlined by journals and conference proceedings in which scientists mainly publish (see Figure 3, Tables 3 and 4).

### 3.5 Co-words analysis

The word that dominated among 8,327 key words used by authors and key words plus in the research output related to spatial data quality and uncertainty was GIS (Table 7). It was used 312 times and strongly co-coupled ( $s_{ij} = 218$ ) with the other frequently used key words (Figure 8). In relation to papers on SDQ&U, GIS was mainly associated with computing environment as well as data and databases management, much less with analyses and applications. Uncertainty was the second most frequently used word. It appeared 167 times, out of which 120 times in the pub-

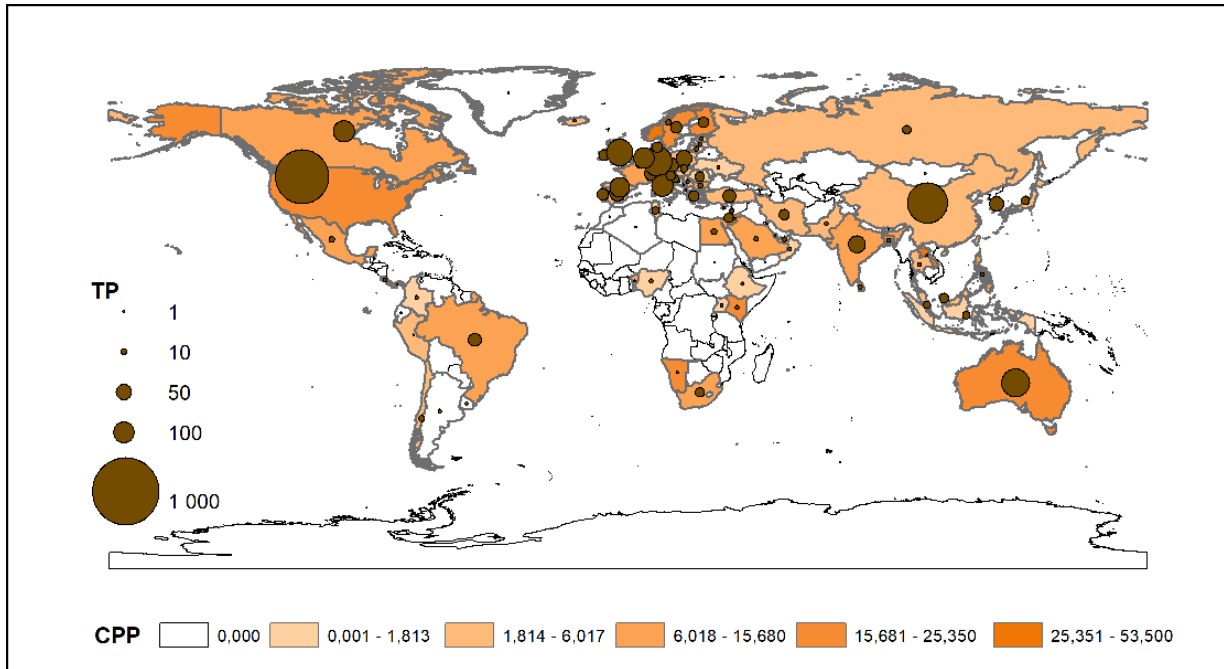


Figure 6: Geographical distribution of countries involved in SDQ&U research

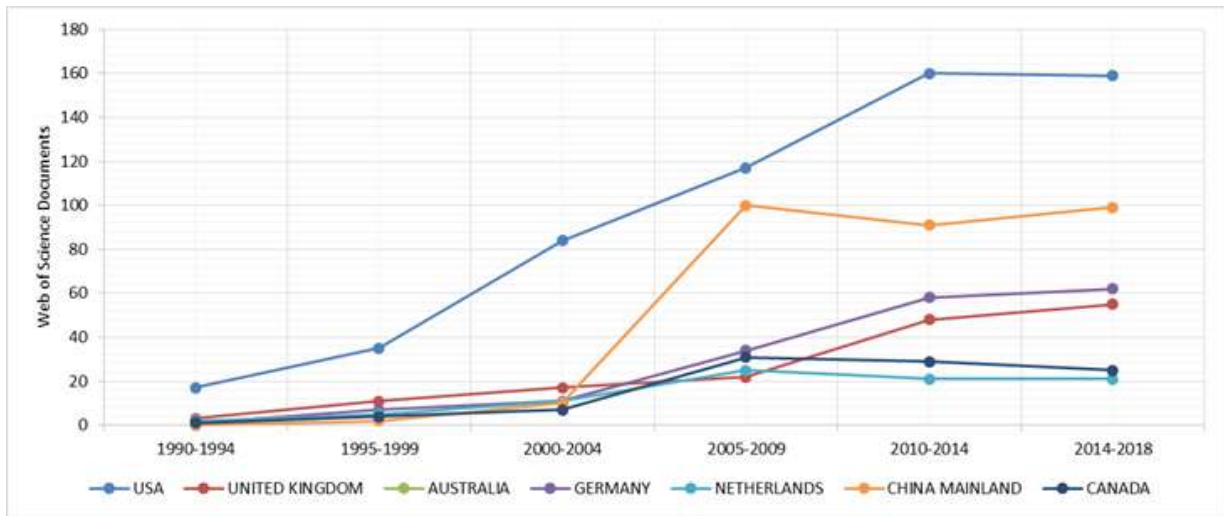


Figure 7: Temporal trend in the output of the field, by the selected country

lication titles (GIS only 68). Uncertainty was mainly collocated with spatial data, information, modelling, assessment, management, and visualization. The term has been used more frequently since 2005, when spatial data users became more aware of the high influence of data uncertainty on the results of spatial analysis. Quality as a key word occurred 116 times (but only 75 times in the titles). Its usage is strongly connected with growing availability of volunteered geographic information, mainly OpenStreetMap (Figure 8 red cluster).

Key words analysis revealed that they could be clustered in 4 groups. The first group contains such frequently used concepts related to data imperfection as: spatial data quality, completeness, geometry, uncertainty assessment, trust, topological relations. It includes mainly papers on the basic problem of imperfection of spatial data in relation to established standards (national or international) or reference, more reliable data. The second group is focused on uncertainty related to any type of data processing. The concepts underline both raster and vector data analysing, as well as methods of simulation, interpolation,

**Table 7:** Key words used more than 50 times

Keyword	Total occurrences	$s_{ij}$	1990-1994	1995-1999	2000-2004	2005-2009	2010-2014	2015-2018
GIS	297	2419	3	6	22	95	83	90
Uncertainty	167	1575	1	9	8	39	66	47
Spatial data	158	1263	5	4	11	37	50	52
Quality	116	1078	0	0	9	15	33	63
Model	108	973	0	6	10	24	34	35
Management	96	1069	0	1	11	19	34	32
Classification	73	698	1	6	5	14	28	19
Information	72	667	0	3	2	14	30	25
Spatial data quality	67	483	0	0	2	12	27	26
OpenStreetMap	55	88	0	0	0	0	14	42
Scale	52	80	1	1	6	8	16	20
Volunteered Geographic Information	52	67	0	0	0	0	7	45
Remote Sensing	50	43	0	1	4	8	20	17

**Table 8:** The most frequent key-words

Concepts group	Key-words ( $s_{ij}$ )
Data imperfection	Spatial data quality (483), completeness (154), uncertainty assessment (50), trust (47), topological relations (41), quality information (27), consistency (23), confidence (20), spatial data uncertainty (21), precision (18), model uncertainty (14), positional uncertainty (14), inconsistency (9), geometry (7), high quality (7), correction (5)
Data processing and analysis	Simulation (389), interpolation (295), kriging (126), modelling (101), data mining (91), spatial statistics (89), quality control (76), fuzzy set (71), generalization (31), 3D modelling (24), sensitivity analysis (28), decision tree (22), spatial pattern (14)
Data storage management and delivery	GIS (2419), spatial data (1263), access (116), web services (87), spatial database (64), information system (21), twitter (20), web GIS (17), GIS database (15), geospatial web services (10), spatial spatial query (6), large dataset (5)
Applications	Agriculture (erosion, livestock) (152), geology (65), topography (70), environmental monitoring and management (air, forest, land use/land cover, soil, water, contamination) (44), visualisation (30)

kriging, modeling, data mining, fuzzy set theory, spatial statistics, and interpolation. The third key words group comprises uncertainty co-located with data storage, management, and delivery; while the fourth - applications for which imperfection is pertinent, like: agriculture, forest, land cover/land use, and geology. The key words groups ordered by total link strength ( $s_{ij}$ ) coefficient are shown in Table 8.

The co-occurrence of relationships among key words used at least 10 times is shown in Figure 8. The colour indicates the period when the words were mainly used, while the size is proportional to the occurrence frequency. The lines depict the relationships between the words: the more lines the stronger the connection between the words.

## 4 Discussion

Although the study of spatial data imperfection started to appear in the mid-1980s with the widespread availability of GIS, the first WoS indexed paper was published in 1990. In the past years the perception of some issues of SDQ&U has changed, which was described exhaustively by Devillers *et al.* [18] and Lowel [47]. The trends noticed by GIS researchers [5, 6, 18, 23, 47, 48] have also been revealed in this study. They could be summarized as paradigm shift from location uncertainties of geographical features and phenomena, through fitness for use data evaluation, to uncertainty in decision making. Similar tendencies were ob-

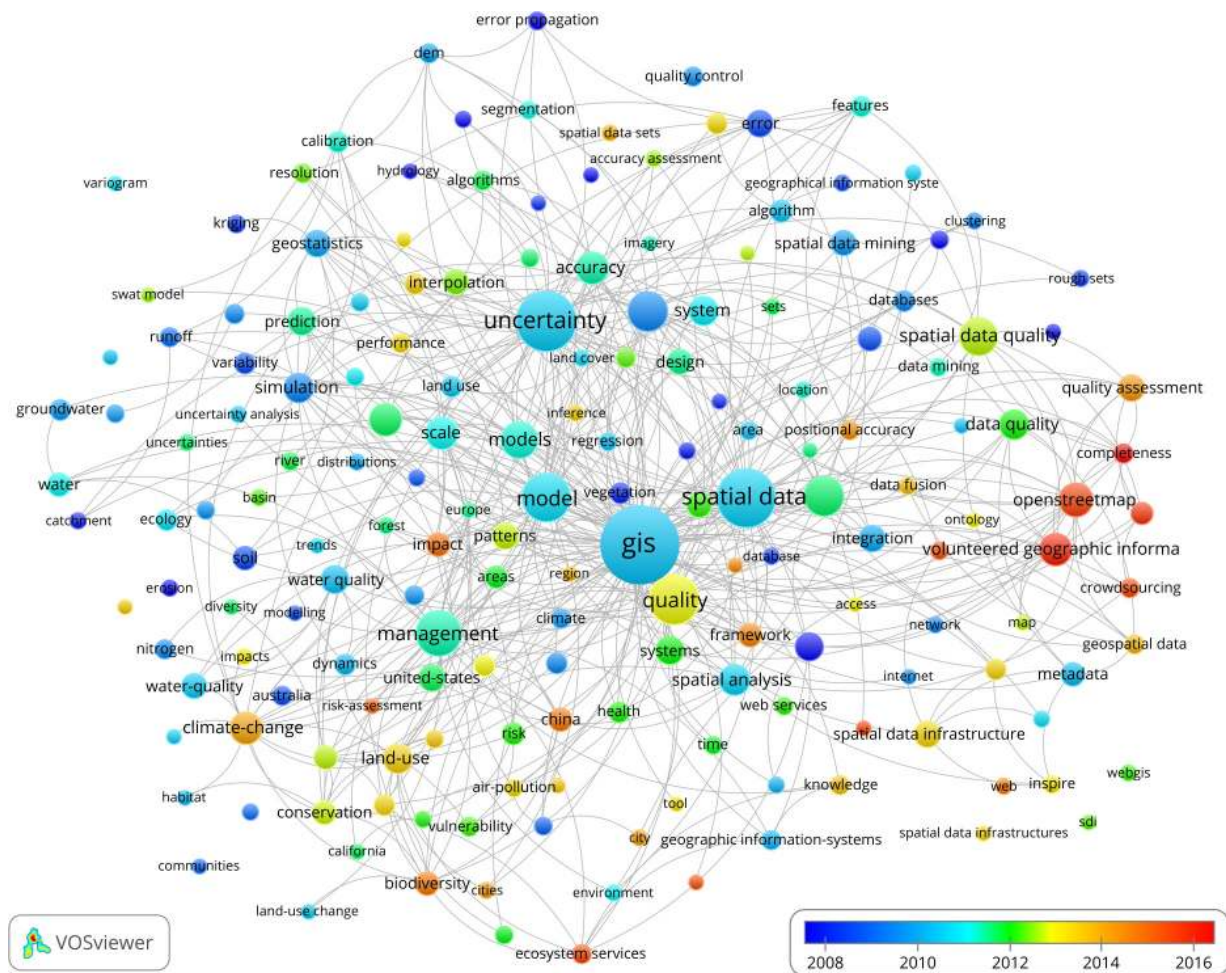


Figure 8: Key-words co-network in 1990-2018

served by [35–37] while analysing the global scientific production in the fields of remote sensing or GIS. They have been highlighted, although not very clearly, by Tong *et al.* [6].

The broader summary of SDQ&U research was discussed by Devillers *et al.* [18], who focused particularly on the main achievements, failures, and opportunities for future research. More recent studies, from 2000 to 2014, regarding the uncertainty of spatial information, were presented by Tong *et al.* [6]. However, the purpose and scope, expressed by the research query, as well as the period covered by in their research are clearly different from those presented in this paper. They focused on uncertainty of spatial information and spatial analysis as well as the contribution and position of China in the field. On the other hand, the presented study concerns only research on SDQ&U without pre-defined countries, regions or institutions of special importance.

Some limitation of the research on data imperfection results from semantic inconsistency. Devillers *et al.* [18]

and Fisher [19] noticed that use of the terms ‘data quality’ and ‘data uncertainty’ varies largely between countries and communities. This has led to the confusion of readers because the same concepts are named differently, or various concepts are labelled identically. In general, uncertainty describes ambiguous or unknown data, whereas quality underlines the degree of data compliance with the standards used for data acquisition. The problem of semantic plasticity was emphasized by [12, 18, 19, 41, 49] and is clearly noticeable in the previous bibliometric analysis [6, 26, 38].

The analysis of keyword temporal trajectory for WoS indexed publication on SDQ&U shows more often used terms (see Table 7). Some of them, such as: uncertainty, quality, remote sensing, model, and classification are also perceived as most important in global publication analysis on remote sensing [35, 36], artificial intelligence [46], and GIS [37].

The volume of research on SDQ&U is relatively small, especially when compared with papers related to GIS [37],

GIScience [38] or RS [35, 36]. Nevertheless, it constitutes an important fraction of these studies. Increase in the SDQ&U volume in the years 2008 and 2009 (see Figure 1) is connected with the growing importance of semantics [18] especially when comparing official (governmental) and crowdsourcing data [12, 48]. Research on data quality is quite elitist, where the small group of scientists has a strong influence on the development of this field (see Table 5), which, as stated by [27], is also the case in many other disciplines (e.g.: cartography, architecture and urban planning, sociology).

The average number of authors per paper equals 2.90. This confirms Biljecki's statement that single-authored papers related to GISciences are 'falling out of fashion' [38]. Seven journals classified as GISciences journal by [38] are also the most productive journals in the field of spatial data quality and uncertainty. They are: International Journal of Geographical Information Science, ISPRS International Journal of Geo-information, Computer Geosciences, Photogrammetric Engineering Remote Sensing, Transaction in GIS, ISPRS Journal of Photogrammetry and Remote Sensing, and International Journal of Digital Earth. However the percent of publications on SDQ&U in these journals does not exceed 4.5, and on average equals 2.

Six leaders in the theoretical and practical data quality research in China are mentioned by Tong *et al.* [6]. Two of them, namely prof. Jinfeng Wahn and prof. Wenzhong Shi, are clearly visible in the network of scholars' co-operation in the analysed paper sample (see Figure 4, green and red clusters). Jinfeng Wahn is very active in data imperfection related to raster data obtained from remote sensing sensors. He established the unbiased estimation theory for heterogeneous land surface. Wenzhong Shi broadly cooperates with many universities. His research is focused mainly on theoretical models of positional and attribute spatial data uncertainty [6].

The collaboration pattern of countries and universities in SQG&U research, dominated by USA and China, is very similar in many disciplines, especially in GISciences [38], GIS [37] remote sensing [35, 36], artificial intelligence [46], and many other disciplines and is dominated by national co-operations. However, the top 10 countries also broadly cooperated with numerous low ranked countries, like: Brazil, the Czech Republic, Poland, Finland, Belgium, Iran, Greece, and Portugal.

The analysed papers, *i.e.* those that met the criteria described in the Data and Methods section, mainly cover methodological and basic research, therefore they do not comprise all data quality issues published in the GIS [37], Remote Sensing [35, 36] journals, and even those analysed by Tong *et al.* in the paper 'Uncertainty of Spatial Informa-

tion and Spatial Analysis' [6]. Moreover, the obtained results could differ significantly after the analysis of publications indexed by the Scopus and Google Scholar citation databases. This has been emphasized by numerous scientists, *e.g.* [28–31, 50]. Flagas *et al.* [28] noticed that in medical studies Scopus offers about 20% more coverage than Web of Science, whereas Google Scholar delivers results of inconsistent accuracy. Similar dependencies were observed in computer science [51] and earth sciences [26, 27]. This justifies the decision to extend our further research into scientific publications output on SDQ&U including Scopus and Google Scholar databases.

## 5 Conclusions

The paper provided a comprehensive and longitudinal survey study on spatial data quality and uncertainty research production in the last 30 years. Its main findings demonstrate that spatial data quality is an essentially multidimensional and multifaceted issue. This is confirmed by a wide range of scientific journals, where papers on spatial data quality are published, and thousands of scientists, working in academia or scientific institution located in dozens of countries. Models of data imperfection, quality validation, and assurance as well as quality indices (*i.e.*: completeness, consistency, and positional accuracy) are the main research topics and frontier exploration subjects since the 1990s. This means that they have been on the scientific agenda for 3 decades already. However, research on spatial data quality and uncertainty has been evolving and it has become more collaborative and more competitive.

Remote sensing and geography were the main research subject categories where quality and uncertainty are of utmost importance at all stages of research, starting from data acquisition and data processing, to information retrieval by many analytical techniques.

Although the publications on SDQ&U constituted only a small fraction of the total research output of authors (on average 16%) or journals (1.1%) they had a significant influence on the authors' potency, which is clearly demonstrated by the CPP analysis.

The research showed that research on spatial data quality and uncertainty is dominated by a few countries, institutions, and scholars, who produce more than 90% of the volume. The most influential countries are USA, China and UK, and it is quite unsurprising, considering their productiveness in remote sensing, artificial intelligence, and GIS research as well as investments on research and development.

Key words analysis, which provides important information about research trends, revealed that the term ‘quality’ was used in the context of data evaluation (e.g. completeness and positional accuracy), mainly VGI data. Uncertainty, however, was associated with remote sensing and spatial analysis in many environmental applications. Moreover, the research analyzed the relatedness between scale and uncertainty

**Acknowledgement:** The research was founded by the statutory project conducted at the Military University of Technology in Warsaw, Faculty of Civil Engineering and Geodesy No PBS/333/2017.

## References

- [1] Goodchild M.F., Stepping over the line: Technological constraints and the new cartography. *The American Cartographer*, 1988, 15, 311–20
- [2] Burrough P.A., Principles of Geographical information systems for land resources assessment (monographs on soil and resources survey). Oxford University Press, New York, 1986
- [3] MacEachren A.M., Visualizing uncertain information. *Cartographic Perspectives*, 1992, 13, 10–19
- [4] Goodchild M.F., Geographical Information Science. *Int J Geogr Inf Syst.*, 1992, 6(1), 31-45
- [5] Devillers R., Goodchild H., (eds). *Spatial Data Quality: From Process to Decisions*. Boca Raton, FL, CRC Press, 2009.
- [6] Tong X., Xie H., Liu S., Jin Y., Shi W., Wang J., Pei T., GeY., Zhu, Ch., Uncertainty of Spatial Information and Spatial Analysis. In: *The Geographical Sciences During 1986—2015*, Springer Geography. Springer, Singapore, 2016, 511-522
- [7] Goodchild M.F., Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. *International Journal of Spatial Data Infrastructures Research*, 2007, 2, 24-32
- [8] Flanagan A.J., Metzger M.J., The credibility of volunteered geographic information. *GeoJournal*, 2008, 72, 137–148
- [9] Haklay M., Weber P., *OpenStreetMap: User-Generated Street Map*. IEEE Pervasive Computing, 2008, 7(4): 12-18
- [10] Zielstra D., Zipf A., A comparative study of proprietary geodata and volunteered geographic information for Germany. *Proceedings of 13th AGILE International Conference on Geographic Information Science*, Guimarães (PRT), 4 Jun 2010, 1-15
- [11] Bégin D., Devillers R., Roche S., Assessing volunteered geographic information (VGI) quality based on contributors’ mapping behaviours. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2013, XL-2/W1, 149-154
- [12] Nowak Da Costa J., Novel Tool to Examine Polygon Features Completeness Based on a Comparative Study of VGI Data and Official Polish Building Datasets. *Geodetski vestnik*, 2016, 60(3), 495-508
- [13] ISO 9000:2005 Quality management systems – Fundamentals and vocabulary (revised by ISO 9000:2015), 2005
- [14] ISO 19101-1:2002 Geographic information (revised by ISO 19101:2014), 2014
- [15] ISO 19157:2013 Geographic information - Data quality, 2013
- [16] ISO/TS 19158:2012 Geographic information - Quality as assurance of data, 2012.
- [17] ISO/IEC Guide 99:2007 International vocabulary of metrology – Basic and general concepts and associated terms (VIM), 2007
- [18] Devillers R., Bedard Y., Fisher P., Stein A., Chrisman N., Shi W., Thirty years of research on spatial data quality: achievements, failures, and opportunities. *Transaction in GIS*, 2010, 14(4), 387-400
- [19] Fisher P.F., Data quality and uncertainty: Ships passing in the night! In Shi W, Goodchild M.F., Fisher P.F. (Ed.) *Proceedings of the Second International Symposium on Spatial Data Quality*. Hong Kong, Hong Kong Polytechnic University, 2003, 17–22
- [20] Virrantaus K., Fairbairn D., Kraak M-J., ICA Research Agenda on Cartography and GI Science. *The Cartographic Journal*, 2009, 46(2), 63-75
- [21] FIG., <http://www.fig.net/organisation/comm/3/index.asp> (retrieved on 18 April 2018)
- [22] Craglia M., Gould M., Kuhn W., Toppen, F., *The Agile Research Agenda*. 7th ECGI&GIS Workshop, 2001, <http://www.lmu.jrc.it/Workshops/7ec-gis/papers/html/agile/agile.htm>
- [23] Di Donato P., Salvemini, M., An analysis of AGILE conferences’ papers: a snapshot of the GI&GIS research in Europe. In: Fullerton K. (Ed.), *11 th EC GI & GIS Workshop ESDI: Setting the Framework*. Sardinia, June 2005, 1-6
- [24] Rousseau R., *Library Science: Forgotten Founder of Bibliometrics*. *Nature*, 2014, 510-218
- [25] Pritchard A., *Statistical Bibliography or Bibliometrics?* *Journal of Documentation*, 1969, 25(4), 348-349
- [26] Siłka P., Śleszyński P. Jaworska B., Citation Analysis Of Polish Academy Of Sciences Committee Members in Google Scholar. *Zagadnienia Naukoznawstwa*, 2016, 52 (4), 529-560 (in Polish with English summary)
- [27] Śleszyński P., Citations and impact of the Polish geographical centers by Google Scholar. *Przegląd Geograficzny*, 2013, 85(4), 599-628 (in Polish with English summary)
- [28] Flagas M., Pitsouni E.I., Malietzis G.A., Pappas G., Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *The FASEB Journal*, 2008, 22, 338-342, doi: 10.1096/fj.07-9492LSF
- [29] Bar-Ilan J., Informetrics at the beginning of the 21st century—A review. *Journal of Informetrics*, 2008, 2, 1–52
- [30] Halevi G., Moed H., Bar-Ilan J., Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the Literature. *Journal of Informetrics*, 2017, 11(3), 823-834, doi: 10.1016/j.joi.2017.06.005
- [31] Bar-Ilan J., Citations to the “Introduction to informetrics” indexed by WOS, Scopus and Google Scholar. *Scientometrics*, 2010, 82(3), 495–506 doi: 10.1007/s11192-010-0185-9
- [32] Garfield E., *Citation Indexing, Its Theory and Applications in Science, Technology and Humanities*. New York, Wiley, 1979
- [33] Smith L.C., Citation analysis. *Library Trends*, summer 1981, 83-105
- [34] Groneberg D.A., Biomedical research in Wrocław: A combined Density-Equalizing mapping and scientometric analysis. *Arch. Immunol. Ther. Exp.*, 2018 66, 1, doi: <https://doi.org/10.1007/s00005-017-0502-6>



- [35] Zhang H., Huang M., Qing X., Li G., Tian Ch., Bibliometric Analysis of Global Remote Sensing Research during 2010–2015. *ISPRS Int. J. Geo-Inf.* 2017, 6(11), 332, doi: <https://doi.org/10.3390/ijgi6110332>
- [36] Zhuang Y., Liu X., Nguyen T., He Q., Hong S., Global remote sensing research trends during 1991–2010: a bibliometric analysis. *Scientometrics*, 2013, 96 (1), 203–219, doi:10.1007/s11192-012-0918-z
- [37] Tian Y., Wen C., Hong S., Global scientific production on GIS research by bibliometric analysis from 1997 to 2006. *Journal of Informetrics*, 2008, 2(1), 65–74. doi:10.1016/j.joi.2007.10.001
- [38] Biljecki F., A scientometric analysis of selected GIScience journals. *International Journal of Geographical Information Science*, 2016, 30(7), 1302-1335, doi: 10.1080/13658816.2015.1130831
- [39] Clarywate Analytics, 15 February 2019, <https://clarivate.com/products/web-of-science/web-science-form/web-science-core-collection/>
- [40] INCites, INCites Indicator Handbook. Thompson Reuters, 2014
- [41] Koch W.G., Lexical knowledge sources for cartography and GIS – development, current status and outlook. *Geodesy and Cartography*, 2016, 65(2), doi: <https://doi.org/10.1515/geocart-2016-0014>
- [42] van Eck, N.J. Waltman L., How to normalize co-occurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, 2009, 60(8), 1635-1651.
- [43] van Eck N.J., Waltman L., Visualizing bibliometric network. In Y. Ding, R. Rousseau, and D. Wolfram (eds.) *Measuring scholarly impact: Methods and practice*. Springer, 2014, 285-320
- [44] van Eck N.J. Waltman L., VOSviewer Manual, version 1.6.8. Universiteit Leiden, CWTS, 27 April 2018
- [45] Mladinich C.S., Generation of Project Base Maps Using US Geodata for the National Water-Quality Assessment Program. *Technical Papers - 1990 ACSM-ASPRS Annual Convention*, vol 2: CARTOGRAPHY, 1990, 115-131
- [46] Niu J., Tang W., Xu F., Zhou X., Song Y., Global Research on Artificial Intelligence from 1990–2014: Spatially-Explicit Bibliometric Analysis. *ISPRS Int. J. Geo-Inf.* 2016, 5(66), doi:10.3390/ijgi5050066
- [47] Lowell K., Why aren't we making better use of uncertainty information in decision making? *Proceedings of 6th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Portland, Maine, USA, 2004. <https://www.spatial-accuracy.org/system/files/Lowell2004accuracy.pdf>
- [48] Bielecka E., Geographical data sets fitness of use evaluation. *Geodetski Vestnik*, 2015, 59(2), 335-348.
- [49] Bielecka E., Leszczynska M., Usability of OpenStreetMap forest data. *SYLWAN*, 2018, CLXII (6), 460-468
- [50] Cavacini A., What is the best database for computer science journal articles? *Scientometrics*, 2015, 102(3), 2059-2071, doi: 10.1007/s11192-014-1506-1
- [51] Martín-Martín A., Orduna-Malea E., Delgado López-Cózar E., Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: a multidisciplinary comparison. *Scientometrics*, 2018, 116 (3), 2175–2188, doi:10.1007/s11192-018-2820-9