Collaborative characteristics and networks of national, institutional and individual contributors using highly cited articles in environmental engineering in Science Citation Index Expanded

Hui-Zhen Fu and Yuh-Shan Ho*

Collaboration has strongly intensified recently and its characteristics are attractive and multidimensional. This study explored the trends, characteristics and citation impact of collaboration as well as collaboration networks for the macrolevel of country, the mesolevel of institution and the microlevel of author, using highly cited articles in environmental engineering (EE) category of the Science Citation Index Expanded. A data visualization and manipulation software, Gephi 0.9.1 was used to visualize the relationships of collaboration among authors, institutions and countries. Although collaboration in general is known to be beneficial in many research areas, collaboration is not beneficial for highly cited research in the EE field. Single author, single country and single institution articles were more visible than articles involving international collaboration, inter-institutional collaboration and individual collaboration. There were certain group collaborations, which played an important role in publishing highly cited articles in the EE field.

Keywords: Collaboration, co-authorship, environmental engineering, highly cited articles, Y-index.

In the early 1990s, it was reported that internationally collaborative publications received many more citations than single country publications and single institution publications1. Collaboration has been strongly intensified in recent years owing to rapid development in scientific communication2. Collaboration also leads to higher citation impact in almost all science fields3–5. Moreover, as the number of countries per paper increases, papers were more inclined to receive more citations than the ones with less countries affiliated6. Recently, certain countries had different performances on the citation impact of collaboration. There was null benefit resulting from collaboration with Iran7. International collaboration did not increase scientific impact for American scientists publishing in Nature and Science8. Citation was a direct indicator of high impact or visibility in the scientific community, and therefore most cited articles were important for investigation9. It was accepted that highly cited articles were associated with high quality research10, and the most cited articles were also confirmed to be the most important ones by a vast majority of the elite group11. Therefore, this study we chose highly cited articles in a certain field as the object of study, and attempted to explore collaborative characteristics and patterns for the macrolevel of country, the mesolevel of institution and the microlevel of author.

Due to the multidisciplinary, interdisciplinary and increasing attention features, environmental engineering (EE) in Science Citation Index Expanded (SCI-EXPANDED) was selected as an example, including ‘resources that discuss the effects of human beings on environment and the development of controls to minimize environmental degradation’12. Relevant topics in this category include ‘water and air pollution control, hazardous waste management, land reclamation, pollution prevention, bioremediation, incineration, management of sludge problems, landfill and waste repository design and construction, facility decommissioning, and environmental policy and compliance’12. The original publication in the EE category could be dated back to 1967 by Water Research with a history of more than 40 years13. There were more than 1.7 million publications in the EE category during the past 46 years, based on SCI-EXPANDED (data updated on 7 November 2014). Highly cited articles...
in the EE category were retrieved for investigation, to characterize the citation impact and collaboration networks of the countries, institutions and authors.

To characterize the collaborative patterns and compare the citation impact of highly cited articles for different levels in EE, some newly developed indicators were employed. Although the indicator of total citations (TC) of a paper has been traditionally applied to evaluate research\textsuperscript{11}, and has been employed to select highly cited articles in various fields of medicine\textsuperscript{12}, major depression\textsuperscript{13}, biomarkers\textsuperscript{14} and radiologic field\textsuperscript{15}, the indicator of total citations could not be re-examined due to the regular update of Web of Science (WoS)\textsuperscript{16}. TC\textsubscript{year} which was defined as the total citations of a paper cited from its publication to a specific year\textsuperscript{17,18}, was introduced to overcome the limitations of the traditional citation indicator\textsuperscript{19}. The citations per publication, dividing total citations of a group by its number of publications, was employed\textsuperscript{20}. Correspondingly, TC\textsubscript{year} was used to replace the traditional dominator of total citations and produced a new indicator, CPP\textsubscript{year}, which could also assure its repeatability. In an era of increasing number of authors per article\textsuperscript{21}, Y-index \((j, h)\) relating to first authorship and corresponding authorship was proposed to characterize the contribution by authors, institutions and countries of highly cited articles\textsuperscript{22}. Recently, to reveal collaborative patterns, increasing attention was attached to the visualization of collaboration networks. Gephi is an interactive visualization and exploration platform for all kinds of networks\textsuperscript{23}. Gephi is a recently developed software, but it has already been used to generate a global map of science based on citation relations among journals\textsuperscript{24}, co-authorship\textsuperscript{25} and relationship among institutes\textsuperscript{26}.

Generally, newly developed indicators including TC\textsubscript{year}, CPP\textsubscript{year} and Y-index, and software Gephi 0.9.1 were employed to obtain accurate, rich and fresh information of collaboration characteristics of highly cited articles in WoS category of environmental engineering (EE). This study not only revealed active countries, institutions and authors in EE, but also identified trends, citation impact, and collaboration networks – international collaboration, inter-institutional collaboration and co-authorship to help scientific workers seek experience and collaboration. Moreover, the collaboration patterns for the macrolevel of country, the mesolevel of institution and the microlevel of author were explored, providing detailed characteristics of collaboration by highly cited articles.

**Methodology**

**Data collection**

Data was collected from the SCI-EXPANDED database of WoS Core Collection from ThomsonReuters (updated on 5 February 2015). The science edition Journal Citation Reports (JCR) of 2013 reported 8539 journals indexed in 176 categories in WoS. There were 46 journals listed in WoS category of EE in 2013 JCR. The schematic for searching process of highly cited articles is shown in Figure 1. There were 38,901,715 documents from 1967 to 2013 found in SCI-EXPANDED. Results were then refined by the WoS category of EE (172,621 documents). A filter of \(TC\textsubscript{2013} \geq 100\) extracted 3658 highly cited documents\textsuperscript{20,28}, while \(TC\textsubscript{2013}\) presents the total citations since publication to the end of 2013. After the filtration of \(TC\textsubscript{2013}\), the filtered publications included articles (3304 articles; 90% of 3658 documents), reviews (303; 8.3%), proceedings papers (151; 4.1%), notes (40; 1.1%), letters (7; 0.19%), editorial materials (3; 0.082%), and one for addition correction. In the EE category, the 3,304 articles (2.1% of 172,621 documents) having \(TC\textsubscript{2013} \geq 100\) were retrieved as highly cited articles for further analysis.

To reduce the error from the database, the data from WoS was checked and reclassified. Publications affiliated to England, Scotland, Northern Ireland and Wales were included in the title of United Kingdom (UK). Publications from Federal Republic of Germany (Fed Rep Ger) were grouped under Germany\textsuperscript{29}. Similarly, articles from Czechoslovakia, Yugoslavia and USSR were checked and assigned to Czech Republic, Croatia and Russia respectively\textsuperscript{29}. To enlarge the sample with author information, the author of single author articles is regarded as both first author and corresponding author.

**Indicators**

The indicators calculated on highly cited articles and abbreviations are given in Table 1. Production-related indicators (including number of different kinds of publications) and scientific impact-related indicators (including number of total citations, citations per publication) were mainly relied upon and some other indicators (including Y-index and ranking) were also employed. Y-index \((j, h)\) relating to number of first author publications (FP) and corresponding author publications (RP) was applied to estimate the performance of authors\textsuperscript{30}. Concretely, \(TC\textsubscript{2013}\) and CPP\textsubscript{2013} having the advantage of repeatability were used to characterize the citation impact of collaboration. All authors affiliations determined the corresponding contributing institutions and countries. Five types have been classified based on the country and institution\textsuperscript{32}: (1) ‘single country/institution article’, if the articles were affiliated to the same country/institution; ‘single author article’, if the articles were affiliated to only one author; (2) ‘multi-country article/multi-institution’, if the articles were co-authored by researchers from more than one country/institution; ‘multi-author article’, if the articles were affiliated to more than one author; (3) ‘first author article’, if the first author’s address was from ‘A’ country or institution for analysis, and therefore this publication was labelled as ‘A’ country/institution’s first author article; (4) ‘corresponding author article’, if the corresponding...
Figure 1. Schematic for searching the articles in environmental engineering category.

Table 1. Introduction of the bibliometric indicators and abbreviations used in subsequent analysis

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Denotation</th>
<th>Definition</th>
<th>Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of publication</td>
<td>TP</td>
<td>Number of total number of highly cited articles in the EE field produced by the analysed unit (country, institution, or author) during the analysed time span.</td>
<td>Counting the number of publications</td>
</tr>
<tr>
<td></td>
<td>SP</td>
<td>Number of single country/institution/author articles in the EE field produced by the analysed unit (country, institution, or author) during the analysed time span.</td>
<td>Counting the number of publications</td>
</tr>
<tr>
<td></td>
<td>MP</td>
<td>Number of multi-country/multi-institution/multi-author articles in the EE field produced by the analysed unit (country, institution, or author) during the analysed time span.</td>
<td>Counting the number of publications</td>
</tr>
<tr>
<td></td>
<td>FP</td>
<td>Number of first author articles in the EE field produced by the analysed unit (country, institution, or author) during the analysed time span.</td>
<td>Counting the number of publications</td>
</tr>
<tr>
<td></td>
<td>RP</td>
<td>Number of ‘corresponding author articles’ in the EE field produced by the analysed unit (country, institution, or author) during the analysed time span.</td>
<td>Count the number of publications</td>
</tr>
<tr>
<td>Number of citations</td>
<td>TC2013</td>
<td>The total number of citations to highly cited articles in the EE field from its publication year to 2013.</td>
<td>Web of Science</td>
</tr>
<tr>
<td>Citations per publication</td>
<td>CPP2013</td>
<td>Number of total citations per highly cited articles.</td>
<td>TC2013/TP</td>
</tr>
<tr>
<td>(CPP2013)</td>
<td>CPP_sp</td>
<td>Number of citations per single author, institution, or country articles.</td>
<td>TC2013/SP</td>
</tr>
<tr>
<td></td>
<td>CPP_mp</td>
<td>Number of citations per multi-country, multi-institution, or multi-author articles.</td>
<td>TC2013/MP</td>
</tr>
<tr>
<td></td>
<td>CPP_fp</td>
<td>Number of citations per first author articles.</td>
<td>TC2013/FP</td>
</tr>
<tr>
<td></td>
<td>CPP_rp</td>
<td>Number of citations per corresponding author articles.</td>
<td>TC2013/RP</td>
</tr>
<tr>
<td>Ranking</td>
<td>RTP</td>
<td>Ranking of number of total highly cited articles.</td>
<td>Sequencing by TP</td>
</tr>
<tr>
<td></td>
<td>RSP</td>
<td>Ranking of number of single author, institution, or country articles.</td>
<td>Sequencing by SP</td>
</tr>
<tr>
<td></td>
<td>RPM</td>
<td>Ranking of number of multi-country, multi-institution, or multi-author articles.</td>
<td>Sequencing by MP</td>
</tr>
<tr>
<td></td>
<td>RFP</td>
<td>Ranking of number of first author articles.</td>
<td>Sequencing by FP</td>
</tr>
<tr>
<td></td>
<td>RRP</td>
<td>Rank of number of corresponding author articles.</td>
<td>Sequencing by RP</td>
</tr>
<tr>
<td>Y-index (j, h)</td>
<td>j</td>
<td>j is publication performance, which is a constant related to publication quantity.</td>
<td>j = FP + RP</td>
</tr>
<tr>
<td></td>
<td>h</td>
<td>h is publication characteristics, which can describe the proportion of RP to FP.</td>
<td>h = tan⁻¹ (RP/FP)</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>N/A</td>
<td>not available</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCI-EXPANDED</td>
<td>Science Citation Index Expanded.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EE</td>
<td>Environmental Engineering.</td>
<td></td>
</tr>
</tbody>
</table>
author’s address was from ‘B’ country or institution, and therefore this publication was labelled as ‘B’ country/institution’s corresponding author article.

Fruchterman Rheingold layout algorithm

Gephi’s last version Gephi 0.9.1, launched in February 2016 (https://gephi.org/), was employed to analyse the collaboration networks of authors, institutions and countries in this study. It is capable of transforming the network into a map and Fruchterman Rheingold is its default layout algorithm31. Fruchterman and Rheingold assume that only neighbouring vertices attract each other and all vertices repel each other32.\(k\) is the optimal distance between vertices

\[
k = C \sqrt{\frac{\text{Area}}{\text{Number of vertices}}},
\]

where constant \(C\) is found experimentally. The vertices are uniformly distributed in the frame and \(k\) is the radius of the empty area around a vertex. The attractive force and the repulsive force are similar to spring force and electrical force respectively (https://github.com/gephi/gephi/wiki/Fruchterman-Reingold).

Attractive force: \(f_a(d) = \frac{d^2}{2k}\),

Replusive force: \(f_r(d) = -k^2/d\).

where \(d\) is the distance between two vertices32. The ideal distance between vertices \(k\) is where the attractive and repulsive force exactly cancel each other out. The produced graph could be described by \(G = (N, E)\), which consists of a set of \(N\) nodes and a set of \(E\) edges. In Gephi, a node could identify the research ID and label of a particular item, while a pair of nodes form an edge32.

Results and discussion

Overall trends of international collaboration, institutional collaboration and co-authorship

To understand the trends of international collaboration, institutional collaboration and co-authorship of the EE field, the percentage of multi-country, multi-institution, and multi-author articles and the total number of highly cited articles during 1967–2011 are displayed in Figure 2. There were 3044 highly cited articles published during 1967–2011. The number of articles kept increasing from 59 in 1967 to 10,798 articles in 2013. The average of percentage of multi-author articles was the greatest (86%), followed by multi-institution articles (35%) and multi-country articles (14%). Multi-author articles were published during 1967–2011, but there were no multi-institution articles before 1973 and no multi-country articles before 1976. The percentage of multi-author articles fluctuated around 85%, while there were increasing trends for multi-institution and multi-country articles during 1967–2011. This meant that the collaboration among institutions and countries has been strengthened in the past decades.

Characteristics and citation impact of international collaboration

To explore the research role and collaboration of contributors, five indicators including total number, independent, collaborative, first author and corresponding author articles were used together for countries and institutions33,34. There were 72 countries/territories contributing to these 3258 highly cited articles with author information from the WoS Core Collection. The 20 most productive countries are listed with number of publications and citations per publication in Table 2. According to all these five indicators, USA won the first place. Researchers from USA contributed to 1624 articles (50% of 3258 highly cited articles), followed distantly by Canada with 273 articles and UK where researchers published 214 articles. It is common that USA took absolute predominance of highly cited articles in various fields, which was paralleled by reports of most highly cited articles in EE (66%)35, adsorption research (61%)19, wetland field (72%)36, socio-ecological research (56%)37 and chemical engineering (49%)23. Although China has become the second since 2006 in scientific production and is closing up with USA38, USA’s leading place is secure in the scientific world according to the research performance of highly cited articles in the near future.

Figure 2. Trends of international, inter-institutional and interpersonal collaboration.
articles (CPPMP = 184) on average, which meant that citations per publication of multi-country collaboration was higher than the citations per publication of single country articles (CPPSP = 193) were added to the corresponding author list for publication. The first author and corresponding author are important authorship, providing detailed information of the characteristics of collaboration. A survey revealed that the greatest contributor of work to a study is the first author, and as for the initial conception and supervision, the greatest contributor is the corresponding author. The home base of a study or the origin of the paper might be observed by the address of the corresponding author. USA still took the lead in international research collaboration. To be specific, larger the national scientific size, usually, lesser the international collaboration. This phenomenon was not common since various studies revealed that international collaboration played an important role in improving citation impact, such as academic publishing of China and Germany in physics, environmental assessment-related research, clinical reproductive medicine research, and eight scientific journals: Nature, Science, Circulation, Blood, Proceedings of the National Academy of Sciences of the United States of America, Journal of the American Chemical Society, Physical Review Letters and Astrophysics Journal. Among the top countries, international collaboration led to higher impact only for Canada, Netherlands, Spain and Belgium. Taiwan had the highest citations per multi-country article (CPPMP = 251), while the highest citations per single country article (CPPSP = 231) was found in Switzerland. Generally, international collaboration was not beneficial for highly cited articles in the EE field. The average percentage of international collaboration in EE during 2001–2011 was 28%, less than the global internationally collaborative Science and Engineering papers percentage of 32% from 2001 to 2013 (ref. 23). Most countries had more single-country articles, while six European countries including UK, Germany, France, Denmark, Belgium and Norway had more multi-country articles than single-country articles. The European countries with smaller size pursued international collaboration more intensively than USA. The national scientific size was negatively correlated with the amount of international research collaboration. To be specific, larger the national scientific size, usually, lesser the international collaboration. It is noticeable that the citations per publication of multi-country articles (CPPMP = 184) on average, which meant that articles with international collaboration attracted less citations than articles without international collaboration. This phenomenon was not common since various studies revealed that international collaboration played an important role in improving citation impact, such as academic publishing of China and Germany in physics, environmental assessment-related research, clinical reproductive medicine research, and eight scientific journals: Nature, Science, Circulation, Blood, Proceedings of the National Academy of Sciences of the United States of America, Journal of the American Chemical Society, Physical Review Letters and Astrophysics Journal. Among the top countries, international collaboration led to higher impact only for Canada, Netherlands, Spain and Belgium. Taiwan had the highest citations per multi-country article (CPPMP = 251), while the highest citations per single country article (CPPSP = 231) was found in Switzerland. Generally, international collaboration was not beneficial for highly cited articles in the EE field. In the EE category, no journals used alphabetical order for the authors list for publications. The first author and corresponding author are important authorship, providing detailed information of the characteristics of collaboration. A survey revealed that the greatest contributor of work to a study is the first author, and as for the initial conception and supervision, the greatest contributor is the corresponding author. The home base of a study or the origin of the paper might be observed by the address of the corresponding author. USA still took the lead in terms of first author articles and corresponding author articles. Except USA, Switzerland, Germany, Denmark, and Norway, 15 countries of the top 20 countries (75%) had higher CPPs for corresponding author articles than first author articles (CPPsp > CPPfp). This meant that articles that contributed to the initial conception and supervision by scientists in top countries, were inclined to obtain more citations. That in turn resulted in high research level of the top countries.

### Characteristics and citation impact of international collaboration

Among the total articles, 1566 articles (48%) were contributed by inter-institutional collaboration and 1690 articles (52%) were published by single institutions. Table 3 shows the 20 most productive institutions whose authors published no less than 30 highly cited articles. Among these 20 institutions, 13 institutions were located in USA, two in Canada and Switzerland respectively, and one each in China, Spain and UK. The Environmental

<table>
<thead>
<tr>
<th>Country</th>
<th>R TP (TP)</th>
<th>R MP (MP)</th>
<th>R SP (SP)</th>
<th>CPP MP</th>
<th>CPP SP</th>
<th>R FP (FP)</th>
<th>R RP (RP)</th>
<th>CPP FP</th>
<th>CPP RP</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>1 (1624)</td>
<td>1 (294)</td>
<td>1 (1330)</td>
<td>187</td>
<td>199†</td>
<td>1 (1503)</td>
<td>1 (1173)</td>
<td>199†</td>
<td>198</td>
</tr>
<tr>
<td>Canada</td>
<td>2 (273)</td>
<td>2 (126)</td>
<td>2 (147)</td>
<td>184†</td>
<td>176</td>
<td>2 (211)</td>
<td>2 (168)</td>
<td>177</td>
<td>181†</td>
</tr>
<tr>
<td>UK</td>
<td>3 (214)</td>
<td>3 (116)</td>
<td>5 (98)</td>
<td>179</td>
<td>214†</td>
<td>3 (148)</td>
<td>4 (131)†</td>
<td>201</td>
<td>204†</td>
</tr>
<tr>
<td>Switzerland</td>
<td>4 (174)</td>
<td>5 (69)</td>
<td>4 (105)†</td>
<td>207</td>
<td>231†</td>
<td>4 (140)</td>
<td>5 (110)†</td>
<td>221†</td>
<td>220</td>
</tr>
<tr>
<td>China</td>
<td>5 (173)</td>
<td>6 (66)‡</td>
<td>3 (107)†</td>
<td>185</td>
<td>207†</td>
<td>5 (137)</td>
<td>3 (135)†</td>
<td>203</td>
<td>204†</td>
</tr>
<tr>
<td>Germany</td>
<td>6 (149)‡</td>
<td>4 (84)‡</td>
<td>9 (65)‡</td>
<td>186</td>
<td>204‡</td>
<td>8 (89)‡</td>
<td>8 (85)‡</td>
<td>196‡</td>
<td>194</td>
</tr>
<tr>
<td>Japan</td>
<td>7 (145)</td>
<td>8 (61)‡</td>
<td>6 (84)‡</td>
<td>173</td>
<td>177‡</td>
<td>6 (102)‡</td>
<td>6 (88)‡</td>
<td>171</td>
<td>171‡</td>
</tr>
<tr>
<td>Netherlands</td>
<td>8 (135)</td>
<td>9 (56)‡</td>
<td>7 (79)‡</td>
<td>194‡</td>
<td>164‡</td>
<td>6 (102)‡</td>
<td>6 (88)‡</td>
<td>174‡</td>
<td>177‡</td>
</tr>
<tr>
<td>France</td>
<td>9 (113)</td>
<td>6 (66)‡</td>
<td>13 (47)‡</td>
<td>179*</td>
<td>187*</td>
<td>12 (69)*</td>
<td>12 (60)*</td>
<td>187*</td>
<td>191‡</td>
</tr>
<tr>
<td>Sweden</td>
<td>10 (105)</td>
<td>10 (52)</td>
<td>11 (53)‡</td>
<td>169</td>
<td>169</td>
<td>11 (72)‡</td>
<td>10 (69)</td>
<td>177‡</td>
<td>179‡</td>
</tr>
<tr>
<td>Spain</td>
<td>11 (100)</td>
<td>11 (42)‡</td>
<td>10 (58)‡</td>
<td>189‡</td>
<td>161‡</td>
<td>9 (77)‡</td>
<td>11 (68)</td>
<td>166‡</td>
<td>172‡</td>
</tr>
<tr>
<td>India</td>
<td>12 (89)</td>
<td>18 (21)‡</td>
<td>8 (68)‡</td>
<td>177</td>
<td>187‡</td>
<td>10 (75)‡</td>
<td>9 (73)‡</td>
<td>188‡</td>
<td>188‡</td>
</tr>
<tr>
<td>Australia</td>
<td>13 (80)</td>
<td>13 (32)</td>
<td>12 (48)†</td>
<td>204†</td>
<td>197‡</td>
<td>13 (58)‡</td>
<td>13 (56)</td>
<td>191‡</td>
<td>192‡</td>
</tr>
<tr>
<td>Italy</td>
<td>14 (69)</td>
<td>14 (31)</td>
<td>15 (38)‡</td>
<td>179‡</td>
<td>178‡</td>
<td>14 (51)‡</td>
<td>14 (48)‡</td>
<td>170‡</td>
<td>171‡</td>
</tr>
<tr>
<td>Denmark</td>
<td>15 (63)</td>
<td>12 (37)†</td>
<td>18 (26)‡</td>
<td>200†</td>
<td>176‡</td>
<td>17 (43)‡</td>
<td>18 (39)‡</td>
<td>182‡</td>
<td>170‡</td>
</tr>
<tr>
<td>South Korea</td>
<td>15 (63)</td>
<td>15 (28)</td>
<td>16 (35)‡</td>
<td>178‡</td>
<td>169‡</td>
<td>15 (46)‡</td>
<td>15 (45)‡</td>
<td>171‡</td>
<td>172‡</td>
</tr>
<tr>
<td>Taiwan</td>
<td>17 (55)‡</td>
<td>20 (13)‡</td>
<td>14 (42)‡</td>
<td>251‡</td>
<td>186‡</td>
<td>15 (46)‡</td>
<td>16 (44)‡</td>
<td>195‡</td>
<td>195‡</td>
</tr>
<tr>
<td>Turkey</td>
<td>18 (51)‡</td>
<td>19 (17)‡</td>
<td>17 (34)‡</td>
<td>168‡</td>
<td>164‡</td>
<td>17 (43)‡</td>
<td>17 (43)‡</td>
<td>158‡</td>
<td>158‡</td>
</tr>
<tr>
<td>Belgium</td>
<td>19 (41)‡</td>
<td>16 (24)‡</td>
<td>20 (17)‡</td>
<td>190‡</td>
<td>164‡</td>
<td>19 (25)§</td>
<td>20 (23)‡</td>
<td>170‡</td>
<td>172‡</td>
</tr>
<tr>
<td>Norway</td>
<td>20 (33)‡</td>
<td>16 (24)‡</td>
<td>23 (9)‡</td>
<td>161‡</td>
<td>170‡</td>
<td>21 (19)‡</td>
<td>21 (18)‡</td>
<td>164‡</td>
<td>162‡</td>
</tr>
</tbody>
</table>

**Table 2.** Characteristics of top 20 countries with ranks, number of articles, and citations per publication
The US Environmental Protection Agency took the first place by 97 articles in total and 78 inter-institutionally collaborative articles, but the second place by the Geological Survey of USA had the most of 59 corresponding author articles and six single author articles. The California Institute of Technology (Caltech) ranked third in terms of total articles, but had the most independent articles and first author articles. It is worthwhile that 4 of the top 20 institutions (US Environmental Protection Agency, US Geological Survey, Environment Canada and Consejo Superior de Investigaciones Científicas) were national government institutions, rather than universities. The US Environmental Protection Agency published two articles with TC_{2013} = 1000 in *Water Research* and *Environmental Science and Technology* respectively. One article concerned the sorption of hydrophobic compounds on natural sediments\(^5\), while the other reported a perspective view of Environmental Protection Agency’s priority pollutants\(^5\). The US Geological Survey published the most cited article with TC_{2013} = 2934, providing the first nationwide reconnaissance of organic wastewater contaminants of USA\(^7\). The three most cited articles of Environment Canada in *Environmental Science and Technology* with TC_{2013} > 200 focused on organochlorine contaminants in Arctic marine food chains\(^48,49\) and long-chain perfluorinated acids in biota from the Canadian Arctic\(^50\). Although the Chinese Academy of Sciences ranked fifth, it is noticed that there were not more than 100 branches distributed in various cities of China\(^5\).

With respect to all highly cited articles, the citations per publication of single institution articles (CPP_{SP} = 195) was higher than the citations per publication of multi-country articles (CPP_{MP} = 186) on average. In terms of the top three institutions, US Environmental Protection Agency and US Geological Survey had higher CPP_{SP}, and Caltech had higher CPP_{MP} (Table 3). This meant that articles without inter-institutional collaboration obtained more citations than articles with it. This performance provided another evidence that collaboration involved in highly cited articles did not help improve citation impact in EE. The articles published only by Swiss Federal Institute of Technology in Switzerland had the highest number of citations (CPP_{SP} = 339), while the articles published by Swiss Federal Institute of Technology in Switzerland had the highest number of citations (CPP_{SP} = 195) and the number of citations (CPP_{MP} = 170). This could be understood that Switzerland had a strong research ability.

In terms of first authorship and corresponding authorship, the US Geological Survey published the most corresponding author articles, while Caltech contributed the most first author articles. Only four institutions including Caltech, Stanford University, University of Wisconsin and Swiss Federal Institute of Technology had higher CPPs for first author articles than corresponding author articles (CPP_{SP} > CPP_{RP}). Corresponding author articles from the other 16 institutions had higher CPPs than their first author articles (CPP_{SP} > CPP_{RP}). This means that, if the top institutions contribute more initial conception and supervision articles, the articles could get greater number of citations. This phenomenon was similar with the rest of top countries.

**Characteristics and citation impact of co-authorship**

Although the authorship list can be created based on contribution, alphabetical order, or reverse seniority, ordering.
by contribution was the most often used approach\textsuperscript{52}. Y-index based on first author articles and corresponding author articles has been used to evaluate the characteristics of authors, institutions and countries for highly cited articles in chemical engineering research\textsuperscript{23} independent research of China\textsuperscript{1}, social work field\textsuperscript{35} and adsorption research\textsuperscript{54}. This indicator is more discriminative for authors than institutions and countries\textsuperscript{23,54}. A total of 584 authors (8.0\%) published more corresponding author articles than first author articles (8.0\%) had the same corresponding author articles as first author articles ($h > 0.7854$); 1266 authors (18\%) had the same corresponding author articles as first author articles ($h = 0.7854$); 775 authors (11\%) had more first author articles than corresponding author articles; and 4582 (64\%) authors did not have first and corresponding author articles. Top 26 authors with high value of the $j$ parameter of Y-index ($j \geq 10$) are displayed in Figure 3. Distance of point from the origin of coordinate directly presents $j$ parameter of Y-index of a certain author. The parameter $h$ differentiates the nature of the leadership role: first or corresponding author role. The highest values of $j$ parameter of Y-index of three researchers were V. K. Gupta ($j = 31; h = 0.8176$), B. E. Logan (30; 1326) and K. Kaman (21; 1.190). These top three authors were above the line $y = x$ ($h = 0.7854$) which meant they had more corresponding author articles than first author articles. Y-index is useful especially for the authors with the same production. For example, the $j$ of six authors including W. Choi, J. C. Crittenden, S. J. Hug, C. P. Huang, P. L. McCarty and B. G. Oliver were all the same of 10. However, $h$ of Choi was 1.326, $h$ of Hug was 1.166; $h$ of Crittenden, McCarty and Huang were the same of 0.9828; and $h$ of Oliver was 0.7854. Choi had greater proportion of corresponding author articles to first author articles than Crittenden, Huang, Hug, McCarty, and Oliver. In addition, J. J. Schauer and Y. S. Ho were the only top authors who published more first author articles than corresponding author articles. Ho published two articles related to pseudo-second order kinetic model in 1998 and 2000 with TC$_{2013} > 1000$ (refs 55 and 56). Furthermore, Ho’s article relating to the kinetic model ranked first in annual citations in the WoS category of chemical engineering since 2008 (ref. 23). It has also been reported that the top most authors contributing to top cited articles were inclined to be assigned as the corresponding authors\textsuperscript{54}. Among the top 26 authors, 15 authors had higher CPP$_{FP}$ than CPP$_{RP}$; eight authors had lower CPP$_{FP}$ than CPP$_{RP}$ and three authors had the same CPP$_{FP}$ and CPP$_{RP}$. Like the results of institutions and countries, the top authors contributed more to the initial conception and supervision of highly cited articles, but the relationship between authorship of the top author and citations impact was not obvious.

The single author article was more theoretical and had greater share of the initial idea\textsuperscript{57}, and has been used for evaluation of authors, institutions and countries\textsuperscript{58}. Detailed information of the top 26 authors is listed in Table 4. G. R. Cass from Caltech had no first author articles and no single-author articles. B. E. Logan from Pennsylvania State University in USA published 30 highly cited articles in total and had the most number of corresponding author articles. Although V. K. Gupta from Indian Institute of Technology in India ranked 11th in total highly cited articles, Gupta had the most first author articles (15 articles). A total of 7895 authors (97\%) had no single-author articles, while only 3\% of authors had single-author articles. However, among the top 26 authors in Table 4, 11 authors (42\%) had single-author articles. The percentage of top authors having single-author articles was much higher than the average. H. Brix published the most of three single-author articles, with the CPP$_{SP} = 258$, and four authors D. Mackay, D. Grosjean, Y. S. Ho, B. G. Oliver had two single-author articles. The citations per publication of single-author articles (CPP$_{SP} = 177$) were a little higher than the citations per publication of multi-author articles (CPP$_{MA} = 174$) for all highly cited articles. Similar to the results of international collaboration, and intra-institutional collaboration, inter-personal collaboration also did not help enhance impact of highly cited articles in the EE field.

\textbf{Collaboration network of multi-country, multi-institution, multi-author articles}

To identify in-depth collaboration relationship of countries, institutions and authors, Fruchterman and Rheingold algorithm in Gephi 0.9.1 was employed to characterize the collaboration network, with the same

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{Top 26 authors with Y-index ($j \geq 10$).}
\end{figure}
ties of collaborating national pairs. A node which had
edges, suggesting that these 70 institutions had 377 varie-
tries, institutions, or authors, and the thickness of the line
there is a collaboration relationship between two coun-
tries, institutions, or authors were mapped by Gephi. For a total of 70 coun-
tinstitutions is illustrated in Figure 5. To be more focused,
condition of area = 10,000, gravity = 10, and speed = 1.0.
According to $G = (N, E)$, $N$ is the number of nodes
(analysed units) which are involved in the analysed
network of all countries, top 20 institutions and top 26
articles. About a half of collaborating pairs (198 of 377
pairs) had only one article. The reason why other coun-
tries did not form an apparent research network among
themselves could be possibly attributed to the small quan-
ty of publications by them.

Institutional collaboration network of the top 20 insti-
tutions is illustrated in Figure 5. To be more focused,
only the names of more active nodes (top 20 institutions)
are shown here. It comprises 547 nodes and 900 undi-
rected weighted edges. A node represents an institution,
the size of which depends on the weight. Edge thickness
among nodes is determined by the number of collabora-
tive publications between them. Similar to the results of
countries, the network was dominated by institutions
from USA. The most frequently collaborating institutional
pair was Caltech and Oregon State University with 24
articles, and the second pair was University of Toronto
and Environment (Canada) with 15 articles. A total of
860 collaborating institutional pairs (76% of 900 pairs)

Table 4. Characteristics of top 26 authors with the highest $Y$-index ($j, h$)

<table>
<thead>
<tr>
<th>Author</th>
<th>$R_{TP}$ (TP)</th>
<th>$R_{FP}$ (FP)</th>
<th>$R_{RP}$ (RP)</th>
<th>$R_{SP}$ (SP)</th>
<th>CPP$_{TP}$</th>
<th>CPP$_{FP}$</th>
<th>CPP$_{RP}$</th>
<th>CPP$_{SP}$</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>V. K. Gupta</td>
<td>11 (19)</td>
<td>1 (15)</td>
<td>2 (16)</td>
<td>N/A</td>
<td>197↑</td>
<td>194↑</td>
<td>189↑</td>
<td>N/A</td>
<td>Indian Institute of Technology, India</td>
</tr>
<tr>
<td>B. E. Logan</td>
<td>2 (30)</td>
<td>11 (6)</td>
<td>1 (24)</td>
<td>18 (1)</td>
<td>187</td>
<td>262↑</td>
<td>243↑</td>
<td>136</td>
<td>Pennsylvania State University, USA</td>
</tr>
<tr>
<td>K. Kannan</td>
<td>5 (26)</td>
<td>5 (8)</td>
<td>3 (15)</td>
<td>N/A</td>
<td>222↑</td>
<td>197↑</td>
<td>214↑</td>
<td>N/A</td>
<td>SUNY Albany, USA</td>
</tr>
<tr>
<td>C. T. Chiou</td>
<td>18 (16)</td>
<td>2 (11)</td>
<td>6 (10)</td>
<td>18 (1)</td>
<td>358↑</td>
<td>346↑</td>
<td>326↑</td>
<td>246</td>
<td>U.S. Geological Survey, USA</td>
</tr>
<tr>
<td>D. Mackay</td>
<td>3 (27)</td>
<td>3 (10)</td>
<td>6 (10)</td>
<td>3 (2)</td>
<td>240</td>
<td>242↑</td>
<td>194↑</td>
<td>419↑</td>
<td>University of Toronto, Canada</td>
</tr>
<tr>
<td>M. Elimelech</td>
<td>5 (26)</td>
<td>126 (2)</td>
<td>4 (14)</td>
<td>N/A</td>
<td>205↑</td>
<td>184↑</td>
<td>181↑</td>
<td>N/A</td>
<td>Yale University, USA</td>
</tr>
<tr>
<td>S. H. Lin</td>
<td>65 (8)</td>
<td>5 (8)</td>
<td>12 (8)</td>
<td>N/A</td>
<td>185</td>
<td>185↑</td>
<td>185↑</td>
<td>N/A</td>
<td>Yuan Ze University, Taiwan</td>
</tr>
<tr>
<td>T. A. Ternes</td>
<td>26 (13)</td>
<td>11 (6)</td>
<td>10 (9)</td>
<td>18 (1)</td>
<td>493↑</td>
<td>395↑</td>
<td>243↑</td>
<td>1,481↑</td>
<td>Federal Institute of Hydrology, Germany</td>
</tr>
<tr>
<td>J. J. Pignatel</td>
<td>18 (16)</td>
<td>15 (5)</td>
<td>6 (10)</td>
<td>18 (1)</td>
<td>274↑</td>
<td>209↑</td>
<td>232↑</td>
<td>74↑</td>
<td>Connectic Agricultural Experiment Station, USA</td>
</tr>
<tr>
<td>G. R. Cass</td>
<td>1 (38)</td>
<td>4 (14)</td>
<td>N/A</td>
<td>N/A</td>
<td>253↑</td>
<td>247↑</td>
<td>N/A</td>
<td>California Institute of Technology, USA</td>
<td></td>
</tr>
<tr>
<td>D. Grosjean</td>
<td>47 (9)</td>
<td>10 (7)</td>
<td>17 (7)</td>
<td>3 (2)</td>
<td>154</td>
<td>154↑</td>
<td>137↑</td>
<td>199↑</td>
<td>DGA INC, USA</td>
</tr>
<tr>
<td>J. J. Schauer</td>
<td>21 (15)</td>
<td>5 (8)</td>
<td>37 (5)</td>
<td>N/A</td>
<td>316↑</td>
<td>227↑</td>
<td>268↑</td>
<td>N/A</td>
<td>CALTECH, USA; Univ Wisconsin, USA</td>
</tr>
<tr>
<td>G. Mckay</td>
<td>21 (15)</td>
<td>59 (3)</td>
<td>6 (10)</td>
<td>N/A</td>
<td>189</td>
<td>443↑</td>
<td>380↑</td>
<td>N/A</td>
<td>Hong Kong University Science &amp; Technology, China</td>
</tr>
<tr>
<td>Y. S. Ho</td>
<td>39 (10)</td>
<td>3 (10)</td>
<td>54 (4)</td>
<td>3 (2)</td>
<td>502↑</td>
<td>352↑</td>
<td>543↑</td>
<td>340</td>
<td>Hong Kong University Science &amp; Technol, China; Peking University, China</td>
</tr>
<tr>
<td>T. Harner</td>
<td>14 (18)</td>
<td>15 (5)</td>
<td>10 (9)</td>
<td>N/A</td>
<td>173</td>
<td>191↑</td>
<td>174↑</td>
<td>N/A</td>
<td>University of Toronto, Canada</td>
</tr>
<tr>
<td>H. R. Buser</td>
<td>39 (10)</td>
<td>11 (6)</td>
<td>25 (6)</td>
<td>3 (2)</td>
<td>209</td>
<td>209↑</td>
<td>222↑</td>
<td>184↑</td>
<td>Swiss Federal Research Station, Switzerland</td>
</tr>
<tr>
<td>J. G. Yu</td>
<td>47 (9)</td>
<td>15 (5)</td>
<td>17 (7)</td>
<td>N/A</td>
<td>268↑</td>
<td>244↑</td>
<td>271↑</td>
<td>N/A</td>
<td>Wuhan University of Technology, China</td>
</tr>
<tr>
<td>F. Wania</td>
<td>47 (9)</td>
<td>15 (5)</td>
<td>17 (7)</td>
<td>18 (1)</td>
<td>229↑</td>
<td>202↑</td>
<td>203↑</td>
<td>137↑</td>
<td>University of Toronto, Canada</td>
</tr>
<tr>
<td>B. H. Hameed</td>
<td>65 (8)</td>
<td>28 (4)</td>
<td>12 (8)</td>
<td>N/A</td>
<td>187</td>
<td>189↑</td>
<td>189↑</td>
<td>N/A</td>
<td>Universiti Sains Malaysia, Malaysia</td>
</tr>
<tr>
<td>H. Brix</td>
<td>110 (6)</td>
<td>15 (5)</td>
<td>25 (6)</td>
<td>1 (3)</td>
<td>211↑</td>
<td>200↑</td>
<td>142↑</td>
<td>258↑</td>
<td>Aarhus University, Denmark</td>
</tr>
<tr>
<td>P. L. Mccarty</td>
<td>35 (11)</td>
<td>28 (4)</td>
<td>25 (6)</td>
<td>N/A</td>
<td>158↑</td>
<td>156↑</td>
<td>199↑</td>
<td>N/A</td>
<td>Stanford University, USA</td>
</tr>
<tr>
<td>W. Choi</td>
<td>47 (9)</td>
<td>126 (2)</td>
<td>12 (8)</td>
<td>N/A</td>
<td>129</td>
<td>150↑</td>
<td>163↑</td>
<td>N/A</td>
<td>Pohang University of Science and Technology, South Korea</td>
</tr>
<tr>
<td>J. C. Crittenden</td>
<td>65 (8)</td>
<td>28 (4)</td>
<td>25 (6)</td>
<td>N/A</td>
<td>143</td>
<td>149↑</td>
<td>141↑</td>
<td>N/A</td>
<td>Arizona State University, USA</td>
</tr>
<tr>
<td>C. P. Huang</td>
<td>30 (12)</td>
<td>11 (6)</td>
<td>25 (6)</td>
<td>N/A</td>
<td>161↑</td>
<td>154↑</td>
<td>164↑</td>
<td>N/A</td>
<td>University of Delaware, USA</td>
</tr>
<tr>
<td>S. J. Hug</td>
<td>65 (8)</td>
<td>59 (3)</td>
<td>17 (7)</td>
<td>N/A</td>
<td>176↑</td>
<td>163↑</td>
<td>168↑</td>
<td>N/A</td>
<td>EAWAG, Switzerland</td>
</tr>
<tr>
<td>B. G. Oliver</td>
<td>65 (8)</td>
<td>5 (8)</td>
<td>37 (5)</td>
<td>3 (2)</td>
<td>147↑</td>
<td>120↑</td>
<td>157↑</td>
<td>115↑</td>
<td>Environment Canada, Canada</td>
</tr>
</tbody>
</table>

According to $G = (N, E)$, $N$ is the number of nodes
(analysed units) which are involved in the analysed
network of all countries, the network was dominated by institutions
among nodes is determined by the number of collabora-
tive publications between them. Similar to the results of
countries, the network was dominated by institutions
from USA. The most frequently collaborating institutional
pair was Caltech and Oregon State University with 24
articles, and the second pair was University of Toronto
and Environment (Canada) with 15 articles. A total of
860 collaborating institutional pairs (76% of 900 pairs)
more international collaborative articles had a larger size,
while more frequently collaborating national pairs were
connected by a thicker edge. It is noticeable that USA
was the centre of the global collaboration network. The
collaborating pair of USA and Canada was the strongest
with 65 articles, followed by the pair of USA and UK
with 37 articles, and the pair of USA and China with 28
articles. About a half of collaborating pairs (198 of 377
pairs) had only one article. The reason why other coun-
tries did not form an apparent research network among
themselves could be possibly attributed to the small quan-
ty of publications by them.

Institutional collaboration network of the top 20 institu-
tions is illustrated in Figure 5. To be more focused,
only the names of more active nodes (top 20 institutions)
are shown here. It comprises 547 nodes and 900 undi-
rected weighted edges. A node represents an institution,
the size of which depends on the weight. Edge thickness
among nodes is determined by the number of collabora-
tive publications between them. Similar to the results of
countries, the network was dominated by institutions
from USA. The most frequently collaborating institutional
pair was Caltech and Oregon State University with 24
articles, and the second pair was University of Toronto
and Environment (Canada) with 15 articles. A total of
860 collaborating institutional pairs (76% of 900 pairs)
Figure 4. International collaboration network of 70 countries by Fruchterman and Rheingold algorithm.

Figure 5. Inter-institutional collaboration network of top 20 institutions by Fruchterman and Rheingold algorithm.
had only one article, much higher than the collaborating national pairs. The research institutions in one country tended to collaborate with domestic organizations.

Co-author network of the top 26 authors is displayed in Figure 6. Each node represents individual authors, and links among nodes (edge) represent number of co-authored publications. The structure of the co-author network can be broadly divided into several independent parts. The largest part was dominated by the international group work in USA and Canada, including T. Harner in Canada, K. Kannan in USA, D. Mackay in Canada and F. Wania in Canada, which could partly explain the strongest collaborating bond of USA and Canada. The second largest part comprised J. J. Schauer in Caltech, G. R. Cass in Caltech and D. Grosjean in USA. Obviously, these two parts had a group of relatively fixed collaborators respectively. More than half of the top authors (14 authors) were involved in five parts with at least two top authors, indicating that well-performed team collaboration made great contribution to highly cited articles in the EE field.

Under the same condition of the Fruchterman Rheingold layout, the three figures of international, inter-institutional and co-author collaboration network looked very different from each other. The layout of international collaboration network was the most concentrated, and could be regarded as a whole. It is easy to tell the centre of the network, the USA. The layout of inter-institutional collaboration network was less compact. More than one centres could be identified and there were a number of connections between different centres. More connections were found in the same country. The co-author network layout was the most dispersed, having several independent parts, each of which contained abundant internal connections. Only few connections between these parts could be found. Certain international group collaborations played an important role in international collaboration.

Conclusion

The application of quantitative techniques revealed active contributors and useful information for scientific researchers in EE, and improved current understanding of collaboration patterns and characteristics of highly cited articles.
(1) Multi-country and multi-institution articles have increased during the study period, and reached around 27% and 57% in 2007 respectively. The percentage of multi-author articles stayed relatively stable.

(2) USA contributed a half to highly cited articles in EE, followed by Canada and the UK. European countries involved more international collaboration than other top countries. The more the top initial conception and supervision countries contributed, higher citations the articles would receive. Furthermore, international collaboration was not beneficial for highly cited articles.

(3) Unlike other fields usually led by universities, two US government institutions, Environmental Protection Agency and Geological Survey dominated, partly due to the character of public welfare of EE. Inter-institutional collaboration also did not help enhance the impact of highly cited articles. Similarly, if top institutions contributed more to initial conception and supervision of the articles, the articles would obtain greater citations.

(4) Y-index revealed the most contributing 26 authors, and top authors, contributed more to the initial conception and supervision of highly cited articles. Consequently, co-authorship did not play an important role in improving citation impact.

(5) Fruchterman and Rheingold algorithm in Gephi 0.9.1, successfully provided collaboration networks of countries, institutions and authors. International collaboration network layout was the most compact having several rings, with USA as the only global centre. Inter-institutional collaboration network layout was less dense, with more than one centre and relevant connections. Co-author network layout was the least dense, with several independent parts, which could be partly explained by various groups.

Excellent group work made great contributions to highly cited articles, and their experiences could be spread, learned and imitated by other scientific researchers in EE. Highly cited works did not seem to be beneficial from collaboration at the macrolevel of country, the mesolevel of institution and the microlevel of author. In order to improve the citation impact of works, environmental engineering researchers should pay more attention to initial conception of scientific work rather than collaboration. Although this study is especially related only to the field of EE, the results provide an interesting contribution to the discussion on the relationship between collaboration and citation impact. Further studies could help to gain deeper insights into collaboration patterns of highly cited publications.


29. Ho, Y. S., The top-cited research works in the Science Citation Index Expanded. *Scientometrics,* 2013, 94, 1297–1312.


53. Ho, Y. S., Classic articles on social work field in *Social Science Citation Index*: A bibliometric analysis. *Scientometrics,* 2014, 98, 137–155.


Received 2 January 2017; revised accepted 27 October 2017
doi: 10.18520/cs/v115/i3/410-421