Growth of science in activated sludge modelling – a critical bibliometric review
Markus Ahnert and Peter Krebs

ABSTRACT
In this paper, the tool of bibliometric analysis is applied to the field of activated sludge modelling and its suitability as a first step of a literature analysis is assessed. The analysis is applied to the total dataset considered as well as a time-based classification. It can be shown that this tool is very well suited to filtering the relevant authors and publications, thus enabling a subsequent visual review. The methodology presented can also be applied to sub-disciplines or other subject areas. However, the sole use of the multiple statistical and visual tools is critically questioned. Thus, misinterpretations and apparent findings can result from structural problems in the data or parameters used. Not all of the metrics used are suitable for finding relevant publications, but rather for ranking the criteria studied. However, the latter represents the most widespread application of bibliometrics. From the analysis of the keywords, it could be deduced that there has been a temporal shift from fundamental model aspects to detailed questions such as the integration of sorption and adsorption processes or anaerobic digestion. The modelling of biological phosphorus removal has also surprisingly lost a great amount of importance in the scientific literature.

Key words | activated sludge modelling, bibliometric analysis, exponential growth, keyword analysis

HIGHLIGHTS
• First application of a bibliometric analysis in the field of activated sludge modelling with specific guidance for implementation.
• Demonstration that the tool can be used to analyse very clear trends in research based on time-dependent measures.
• Evidence of the importance of using such a tool in analysing the diversity of information available today in combination with professional expertise by users.

INTRODUCTION
Motivation
An essential step in deciding on the research direction as well as the development of emerging fields is the analysis of the existing literature from this thematic field. At the same time, an analysis of the existing literature can be used to identify research topics that have possibly already been exhaustingly studied and are therefore less eligible for funding. Gujer (2006) did this for the field of activated sludge modelling in a minimalistic and subjective review as he explained in his paper. He discovered from his review that this area of research obviously suffered from declining research interest because essential aspects had been researched. Therefore, the (limited) resources in water research should be shifted to other areas in order to solve the worldwide problems of water supply and sanitation. Other authors argue similarly. For example, Sin & Al (2021) have published a remarkable article on linking activated sludge modelling with artificial intelligence, machine
learning and biological molecular data because, in their opinion, the limits of conventional activated sludge models have been reached.

Own analyses comparable to the work of Gujer (2006) have shown that the original analysis results were probably only a time-limited sample and the conclusions are no longer valid in this way. Therefore, the question arose whether a literature analysis have to be conducted in such a way that misinterpretations can be avoided and further insights can be gained. For this purpose, the tool of bibliometric analysis was identified as suitable and useful and will be practically applied in this paper. The paper of Gujer acts as a recurrent theme for this investigation.

Introduction into bibliometrics

The number of published scientific articles is continuously increasing every year at a rising rate. Therefore, it is hardly possible to search in a subject area only by directly screening the available references. Tools are needed that can carry out a (semi-)automatic or systematised narrowing down and thus limit the scope of the direct search to such an extent that it is still feasible at all. The term ‘bibliometrics’, which is now used for this, goes back to Pritchard (1969) who places the origins of the statistical analysis of literature a few decades back, while the publication of the ‘Science Citation Index’ in 1964 and the development of ideas and concepts by Price (1965) also methodically and substantially backed this up with usable data.

In the field of water resources and especially wastewater treatment, this tool has been used only rarely and rather specifically. Moreover, its focus has leaned towards ranking countries, institutions or authors that are important in a particular thematic area (e.g. Chuang et al. 2011).

A positive example is Osman et al. (2020). In this review, bibliographic analyses are used in addition to specific technical research to identify current trends in the literature dealing with biogas production processes. Another example (Zheng et al. 2018) is more representative of the problems and challenges that can arise with such an analysis. For example, a very high information density in the form of information extracted from the literature data combined with some graphically sophisticated but possibly difficult-to-understand figures is not necessarily suitable for actually bringing clarity to the chosen task.

The extent to which the bibliometric analysis of a single journal (Wang et al. 2010) yields real insights may also be doubted. The insights for future research activities from this analysis were therefore rather limited.

Objective

The aim of the following investigations is first to continue and expand the bibliometric analysis that was originally carried out for the first time by Gujer (2006) for the field of activated sludge modelling. To prevent from ending in itself the Gujer-oriented analysis is expanded and the same database is used for a more sophisticated bibliometric analysis. It is expected, that at the same time, trends and developments are to be obtained retrospectively and prospectively from the data set. In this way, the usability and possible advantages and disadvantages of this type of analysis, especially in the field of water and environmental research, are to be worked out and thus an assessment of the ‘bibliometric analysis’ tool used is to be made. The objective is also to give helpful advices for adaptation of the methodology.

Typically, one would compare a new methodology with an already proven one. In the present case, this represents the purely manual review of the literature searched. Due to the large number of sources in the compiled database, this procedure is not possible in this way. Therefore, a methodological comparison must be dispensed with and only an evaluation of the findings from the bibliometric analysis as a new method can be derived.

The linkage of actually different topics by the parallel consideration of the Gujer evaluation as well as the general test of the bibliometric analysis seems therefore necessary despite possible increase of the complexity of the explanations.

MATERIALS AND METHODS

Database and tools

Selection of database

For all subsequent evaluations, a search was carried out in the Web of Science (http://isiknowledge.com) identical to the procedure according to Gujer (2006). The date of the query was 02.01.2021. Since the data set already contained documents stored with a publication date of 2021, these documents were not included in the evaluation.

In addition to the Web of Science (WoS), Scopus (https://www.scopus.com) is another database with citation data. Similar information is also available via Google Scholar (https://scholar.google.com/). While WoS and Scopus are chargeable, Google Scholar (GS) can be used free of charge. However, since a significant proportion of Google Scholar is also internet sources, websites, etc., this

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.191/895647/wst2021191.pdf by guest
was not used in the following analyses, as a purely literature-based investigation was to be conducted.

There are significant differences between the two databases WoS and Scopus (Zhu & Liu 2020), for example in the distribution and ranking of the countries of origin of the articles. Although differences in the published content could also be expected from this, Zhu & Liu (2020) were able to show that at least the main subject areas are similarly distributed and interconnected in a comparable way.

Aguillo (2012) examines the possibilities of using GS. The author cautions against the uncritical use of the data of this bibliography and citation index due to the lack of control mechanisms, but at the same time highlights the possibilities that can be expected in the future if appropriate quality improvements are made.

Larsen & von Ins (2010) have shown that the use of classic bibliographies will no longer cover actual research activities in the future. By analysing the development of publication activities over time in various citation indices, relevant conclusions were drawn for bibliometric analyses to be carried out in the future.

As the original analysis by Gujer (2006) was based only on the WoS data, a comparison was made with data from an identical search (see section on Bibliometric analysis of the entire data set) at Scopus to identify possible imbalances due to the limitation to only one database.

A comparison of the countries with the most publications (Figure 1 left) reveals a slight shift between the USA and China. While in WoS by far the most articles in the searched dataset come from China, the distribution in Scopus is somewhat broader at the top. In WoS, the top 5 countries cover 44% of the included datasets, in Scopus only 33%. Therefore, the regional diversity in Scopus appears to be somewhat larger. Also there are some differences in types of documents included. In Scopus a larger percentage of proceeding papers and a lower percentage of reviews is included (Figure 1 right).

A direct comparison of the number of papers in individual journals according to the data source used is shown in Figure 2 (left). With two relevant exceptions (Journal of the Water Pollution Control Federation, Environmental Science and Technology), the WoS always contains more articles per journal. Two major differences can be derived from the temporal development (Figure 2 right). Scopus obviously contains more older documents from before 1990, while WoS contains more articles from 2016 onwards. At least in the case of the journal of WPCF, which was published under this name from 1960 to 1989, a connection can be made.

Based on this analysis, a certain difference between the data sets from WoS and Scopus is evident in the subject area investigated. However, since the publication of Gujer (2006) on which this article is based, relies on the WoS data sets, therefore WoS was also used as the only data source out for the subsequent evaluations. Since the differences found between the two databases do exist, but are not considered significant for the thematic field this seems to be acceptable. In independent future investigations the combination of data from both databases seems to be evident.

Development of the number of data sets in the Web of Science

In order to classify the data sets used in the entirety of the WoS database, relevant information is required, which could not be researched directly. Therefore, sources that had direct access were used. What is needed are data on the number of papers published each year.

A source for this figure can be found in Fortunato et al. (2018). This contains a chart of the papers recorded in the WoS since 1900, which has been digitised for own comparative purposes. The authors show the already known exponential trend in the development of the number of publications. Furthermore, they contain studies on the development of the number of authors per publication. This is demonstrably increasing (exponentially), especially in the engineering sciences. However, this trend is certainly only a snapshot. The exponential increase observed is not a
development of recent years. Price (1986) has already shown that this also applies to the number of journals from 1650 until today. Bornmann & Mutz (2015) also provide analyses of the development of general publication activity, essentially also based on WoS data, supplemented by other internal data sources that are not publicly available. However, there is no critical discussion of the data source used. Data from diagrams were also digitised from this source for the authors’ own use. Interestingly, Bornmann & Mutz (2015) have segmented the exponential growth in publication activity into three time periods from 1650 to 2012. According to this, the strongest growth can be seen since about 1925–1950. Furthermore, data on the number of documents in WoS were extracted from Patience et al. (2013). The data thus obtained from various sources can be used to calculate the doubling time as a suitable comparative variable of the rate of growth in publication activity.

Bibliometric analysis of the entire data set

For the bibliometric analysis, the software biblioshiny was used, which, based on the R package bibliometrix (Aria & Cuccurullo 2017), can perform various statistical evaluation procedures with databases of different bibliographies such as WoS. The browser interface biblioshiny makes it possible to use it even without knowledge of R. Due to its ease of use and the large number of evaluation options, this tool was used for the following statistical analyses.

In addition, there are a number of freely available software tools for conducting bibliometric analyses. Table 1 shows some examples with the frequency of citation in various data sources.

CiteSpace and VOSViewer in particular are very much oriented towards visual analysis, which may explain their widespread use.

For the analysis, the same search terms activated AND sludge AND model (abbreviated as AuSuM in the following) as in Gujer (2006) were used to search the WoS Science Citation Index expanded up to and including 2020. It is easy to understand that for a search for publications on activated sludge modelling the selection of search terms is too general and other reasonable possibilities would have resulted in a better thematic limitation. However, due to the direct relation to the Gujer search, this approach was intentionally chosen.

The data were exported to and prepared for analysis with biblioshiny. Some simple analyses can be carried out directly in the WoS. For this purpose, the assignment of the extracted articles to the WoS main categories was determined and compared in additional searches. These are:

- Environmental Sciences
- Water Resources
- Environmental Engineering
- Chemical Engineering

Since articles can also be assigned to several categories, an analysis of totals is subject to error. For these categories, the total number of papers per year was determined and applied to the total number of the year in the AuSuM search. This was calculated cumulatively in absolute and
relative terms and presented in relation to the maximum proportion of papers in this category.

For the bibliometric analysis, existing measures were used or combined to form own measures (see Table 2). The measures were selected for two main reasons: Either it is obvious that this measure can be useful in the analysis or they are measures that have been used in many other publications for bibliometric analysis and thus can be used comparatively to the findings from other studies.

In the analysis, a special focus was given to links within the database (local), as it was expected that these could then also be closely related to it thematically.

The analysis showed that there are obviously inaccuracies and errors, especially in the naming of authors. This has a great influence on all author-relevant evaluations. It also has a negative impact on the generation of collaboration or citation networks. On the one hand, a manual correction is extremely time-consuming due to the amount of data and, on the other hand, it is not possible to reliably assign which authors with the same name are really the same persons. On the other hand, the error is of course only significant for authors with high relevance, so that correcting a few people might be sufficient. This was not completely done in the context of this analysis, as the focus was not on the authors but rather on thematic aspects.

In addition to the data of descriptive statistics, journal-based and author-based measures are of interest. However, the absolute number of publications is not really a suitable parameter for the significance of a journal or author; much more interesting is the number of citations that journals/authors receive for their papers. In order to identify relevant journals or papers even with a relatively low number of publications, a relative impact measure is a suitable parameter.

Indices based on publications are very widespread and widely used to assess the publication activity and importance within an author’s field. The best known is the Hirsch or h-index, which as m-quotient contains a reference to the time period of the publications (Hirsch 2005). The g-index was developed to take highly cited papers more

| Bibliometric measure | Description | Remark |
|----------------------|-------------|--------|
| Descriptive statistics | main features of dataset | |
| journal-based measures: | | |
| Most relevant sources | number of papers per journal (simple measure of publication mass) | combination of both for an impact measure; the impact of a journal is higher when the ratio of no. of citations per no. of papers is higher |
| Most local cited sources | number of citations within dataset per journal (measure of relevance) | |
| Source dynamics | comparing variations in percentage of papers per journals per year | use of logarithmic presentation for better interpretation |
| author-based measures: | | |
| most relevant authors | number of papers per author | combination of both for an impact measure analogue to journal impact measure as author impact |
| most local cited authors | number of citations within dataset per author | |
| author local impact | different indices (h, m, g) and number of total citations | analysis of correlation between author impact and indices |
| document-based measures | | |
| most global cited documents | citations from inside and outside of database | possibly thematically inappropriate documents |
| most local cited documents | citations only from inside of database | no consideration of missing documents within the database |
| most cited references | citations inside of database of papers from in- and outside of database | |
into account (Egghe 2006). The use of such indices is very widespread, but also highly controversial and the subject of many scientific discussions. It is repeatedly questioned whether such measures are necessary beyond the standard bibliometric measures (e.g. Bornmann et al. 2009), since empirical correlations between indices and standard bibliometric measures can often be found (e.g. Gaster & Gaster 2012).

Various measures are also available for the analysis of the most important documents. Different influencing factors from inside and outside the database have to be taken into account, as will be explained in the presentation of the results.

In addition to the measures listed with regard to authors or documents, a wide variety of evaluations can be carried out with the help of bibliometric software with relation to conceptual, intellectual and social structure. Due to the complexity of the relationships presented, the evaluation is usually carried out visually using various graphical tools. The use of these tools can be helpful in identifying clusters and connections with regard to authorship, thematic references, participating institutions, grouping of keywords or the scientific contribution from different countries. After intensive own tests with these tools, it can be stated that they should primarily be used as a supplement to conventional measures.

Depending on the software used, the preparation and generation is dependent on a wide variety of parameters and specifications, so that direct comparability and verification of the results of different tools is hardly possible. The interpretation of the representations is also very complex. Through different colours, line thicknesses, symbol and font sizes or distances, several properties or characteristics can be represented at the same time. The possible dominance of individual variables in this representation can make interpretation difficult. However, these visualisation tools have proven to be suitable in our own analysis to work out certain clustering. This has been done, on the one hand, for the keywords contained in the articles and, on the other hand, with regard to the co-citation of cited references.

**Analysis of time-related developments**

In addition to absolute numbers of publications or citations, an additional consideration of the temporal component can provide valuable information on developments and trends. Therefore, an analysis of temporal developments is carried out on the following measures:

- Change in the proportion of document types.
- Collaboration index: the Collaboration Index (CI) is calculated as total authors of multi-authored articles divided by total multi-authored articles (Elango & Rajendran 2012; Koseoglu 2016).
- Average citations per documents and average citations per year per documents.
- Most locally cited sources: evaluation by rankings of the top 20 for each year, analysis of the rankings over time allows conclusions to be drawn about the impact of the journals.
- Most locally cited references.
- Country scientific production and most cited countries: development and presentation of a country impact, i.e. the change in the significance of the scientific contribution from different regions of the world.

**Keyword development within the examined topic area**

Due to the exponential increase in the number of articles, it is of interest for the analysis of keyword development whether an effect on the keywords used can also be determined. Therefore, an evaluation of the temporal development of the papers contained in the data set of the AuSuM search was compared with the various existing keywords. The correlation can be described very well with a linear function (Figure 3). This allows the interpretation that the number of keywords used behaves very similarly to the development of the published articles and thus there are no time-variable effects on the number or use of keywords.

This ensures that an evaluation of the development of keywords over time provides plausible results through a relative representation in relation to the number of publications.

In the period 1990–2020, 18,143 different keywords were used in the articles of the database of the topic area (total number of keywords 81,765). Their frequency was evaluated per year and related to the total number of different keywords per year. This determined the relative share of this keyword. Then, for the 200 keywords with the largest

![Figure 3](http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.191/895647/wst2021191.pdf)
RESULTS AND DISCUSSION

Bibliometric analysis of the entire database

The database used contains 7,134 documents from 649 sources. On average, the documents are cited 26.7 times or 2.4 times per year. There are essentially three document types (see Table 3), which differ in their statistical characteristics. For example, reviews typically contain significantly more references and are also cited significantly more due to the collected content. The distribution of document types is similar to the results of Zandi et al. (2019) (Analysis of the use of membrane application in industrial biowaste treatment) and Davarazar et al. (2020) (Treatment technologies for pharmaceutical effluent), who determined a share of 80% for articles, 18% (Zandi et al. 2019)/9% (Davarazar et al. 2020) for proceeding papers and 1/4% for reviews respectively. Scientific collaboration played an extremely important role in the preparation of the documents. Only 1.6% of the documents were written by single authors.

Since the initial motivation for this article was to validate the statement made in the article by Gujer (2006), the first step is to compare the data obtained with regard to the total number of articles in the WoS database used and the topic-specific search for AuSuM. Figure 4(a) shows the exponential development of the articles in the database from various sources and provides the corresponding regression equations and coefficients of determination. What is noticeable here is the increased number of publications in the Bornmann & Mutz (2015) source, which can be explained by the integration of further data sources. The other sources (Fortunato et al. 2018; Patience et al. 2019, own WoS search) agree very well.

The development of the number of articles from the Gujer search (Figure 4(b)) in the original article and in the current analysis show certain deviations. Obviously, journals were subsequently included in the WoS data source, so that the number of available articles has increased. However, the influence is only slight. In both cases, an increasing trend can be observed, which also shows an exponential trend.

In order to eliminate this exponential trend, the articles from the AuSuM search were put into perspective with the corresponding annual data from the total WoS database (Figure 4(d)). From this, the beginning of publication activity in the subject area of ‘activated sludge modelling’ can be identified very well from the mid/late 1980s. From the mid-1990s onwards, articles from this topic area have a largely stable share of 0.02% of the total number of articles in the WoS. This refutes the conclusion of Gujer (2006) that there is a decline in scientific activity in this subject area.

This can also be documented by the doubling times that can be calculated from the exponential functions (Figure 4(c)). These range from 13.6 to 25.8 years, depending on the database used. Interestingly, this fits very well with an analysis of the number of scientific journals by Price (1963). This includes the existing journals in the period 1700–1950. The doubling time of new publications is 17.8 years in the area of scientific articles.

The doubling time of the number of articles in the field of ‘activated sludge modelling’ of 2.6 (Gujer period) and 5.7 years respectively shows that this is a relatively young field with intensive publication activity. The increase in the doubling time from 2.6 to 5.7 years again indicates the beginning of a slowdown in growth – the subject area is thus slowly entering the scope of more established research areas. Larsen & von Ins (2010) come to similar results and effects for ‘young’ and ‘old’ subject areas in a detailed analysis of various subject-specific literature databases.

At this point, it should be noted that due to the strong exponential character of the growth in publications,
comparisons of absolute numbers of publications or citations can very quickly lead to misinterpretations or that effects in older periods are not recognised due to the lower numbers at the time. This is exemplified by evaluations conducted primarily with absolute numbers, e.g. by Zheng et al. (2015). Therefore, the use of relative parameters seems to make more sense.

The results of the evaluation according to the four WoS categories are shown in Figure 5. On the left, the stacked share of the total number of publications in WoS is shown cumulatively. If one compares this representation with Figure 4(c), one can obviously assume an increase in the allocation to other categories from 2010 onwards.

If the analysis of the occurrence of the time-based maximum is considered (Figure 5 right), it can be seen that obviously new categories are occupied by the subject area (chemical engineering), which suggests an increased use in practically oriented areas.

Figure 4 | (a) growth of papers in WoS database from different sources; (b) growth of papers in AuSuM search in WoS from Gujer (2006) and actual data; (c) doubling times of publications in different sources; (d) relative part of AuSuM data in WoS database per year. Please refer to the online version of this paper to see this figure in color: http://dx.doi.org/10.2166/wst.2021.191.

| source               | time span   | doubling time [a] |
|----------------------|-------------|-------------------|
| Patience et al. 2019 | 1989-2017   | 17.3              |
| Fortunato et al. 2018| 1900-2013   | 14.1              |
| Bornmann and Mutz 2015| 1980-2012   | 25.8              |
| own WoS search       | 1900-2020   | 13.6              |
| Price 1986 (no. of journals) | 1700-1950 | 17.8              |
| Gujer 2006           | 1985-2004   | 2.6               |
| own WoS search AuSuM | 1960-2020   | 5.7               |

Figure 5 | Cumulated percentage of papers per year in WoS search by main categories (left absolute, right relative values). Please refer to the online version of this paper to see this figure in color: http://dx.doi.org/10.2166/wst.2021.191.
Journal-based measures

These evaluations show which journals contain a particularly large number of articles or whose articles were cited particularly frequently. Neither of these alone is very meaningful; the absolute number of articles says nothing about the quality. However, if the ratio of citations of this journal to the articles published in it is particularly high, a high significance can be concluded. Obviously, the articles are particularly important for authors. Therefore, this is shown for the 30 journals with the most publications as well as 36 other relevant ones in Figure 6 left. The resulting factor of about 27 citations per document is already known from the descriptive statistics. There is a large scatter around this equalising line. Some journals have an above-average number of citations per document, others a very below-average number. Among the two journals with the largest number of publications and citations, Water Research is obviously the journal with the greater impact compared to Water Science and Technology, despite a somewhat lower number of publications.

Furthermore, some particularly positive (e.g. Applied and Environmental Microbiology, Journal of Water Pollution Control Federation) and negative examples (e.g. Water Environment Research, Environmental Technology, Desalination and Water Treatment) are noticeable.

Especially in the case of journals with few articles in the database, the influence of papers with a high number of citations can very quickly lead to a high ratio value, as can be seen in the example of Journal of Water Pollution Control Federation. Although this journal has been discontinued or renamed in 1989, it obviously contains some significant articles that are cited very often and continuously.

Source dynamics

Important and very interesting for the analysis of a topic are the main journals in which the articles are published. Due to the sometimes very long periods of time from which the analysed data originate, a temporal component is also relevant in order to identify journals that have published current content on the searched topic. This is shown for the top-ranked 10 journals (number of publications) in Figure 6 on the right. For better comparability, a semi-logarithmic representation is used. Similar to the development of publication in ‘old’ disciplines, longer-established journals are obviously less in demand and new journals initially show increased publication activity. This can be seen very clearly in the journals shown in dashed lines. However, we can only speculate about the reasons.

The combination of the two measures shown in Figure 6 left can therefore be used very well to identify relevant sources in the subject area under investigation. It should be noted, however, that the proportion of articles in these sources that are relevant to the topic under investigation may well be small. Therefore, it does not seem advisable to look through sources with a high impact in their entirety, but to examine documents from these sources in detail as a priority in a literature search.

Author impact measures

A large number of measures can be used to analyse the significance of individual authors. Their mutual relationships and correlations are shown in Figure 7. It can be empirically shown that there is a linear relationship between the h-index and the g-index (Figure 7(a)). There is a mathematical relationship between the h-index and the total number of citations (Figure 7(b)) with a power function. It is therefore...
possible to work with the standard bibliographic measures SBM instead of the various indices.

Measures based purely on the number of publications obviously have only a minor significance for the impact in the subject area under investigation. This should always be considered in connection with citations. Therefore, the authors’ publication activity was compared with the number of citations received (Figure 7(c)). The author Mogens Henze has a very high number of local citations (i.e. within the subject area). The lines indicate the average number of citations per publication – once with and once without Mogens Henze’s data set. All points above the lines represent authors with an above-average number of citations per publication. In this representation, authors with very many publications generally appear in the below-average range, although the total number of citations received is of course very high.

At this point it must be mentioned that the results for the authors Vanrolleghem and Van Loosdrecht had to be corrected because the database contained different spellings.
For other authors with a comparatively large number of publications, no such errors could be found during review, but further errors cannot be excluded and are to be expected. For the two authors concerned, this has a direct influence on the internal calculation of the indices used, which could not be corrected. See also further explanations in the chapter Limitations.

Depending on the criteria used (h/g/m-index, number of papers, total citations), the authors are ranked differently in the ranking lists. The various comparisons in the evaluations of Figure 7(c)–7(f) clearly show that there are sometimes enormous differences in the ranking of individual authors. If it is assumed that the work of authors with a high number of citations is considered significant for the respective subject area, then only the different number of publications per author must be taken into account in order to eliminate this influencing factor and not overlook authors with few but significant publications. Therefore, the author impact as a ratio of citations received to the number of publications seems to be a very suitable measure. At the same time, only easily available SBMs need to be used for the calculation.

Table 4 shows the 10 authors with the highest author impact and the other 10 authors with the largest number of local citations, sorted by author impact. It is noticeable that the highest values for author impact are achieved by authors with a rather medium number of publications.

Thus, author impact as a relative measure appears to be very well suited for identifying important authors in the subject area under investigation and for giving priority to their frequently cited articles in a subsequent literature review.

Document-based measures

The 20 most globally cited documents contain 6 reviews. 4 articles deal with the basics and properties of biological phosphorous elimination, three articles with the removal of dyes from wastewater and two with membrane fouling.

Among these 20 articles, only two documents (Smolders et al. 1994; Gujer et al. 1999) can be directly assigned to the original topic of activated sludge modelling. This shows the limited suitability of the measure ‘global citations’.

When analysing the most locally cited documents, it is obvious that these are for the most part very relevant contributions to the subject area under investigation. Therefore, this type of search should be carried out in every bibliometric search. However, the analysis of the most locally cited documents has a methodological weakness. If relevant articles are not included in the database because they are not represented by the search words (for whatever reason), they will not be recognised as significant documents in this evaluation either. Therefore, instead of using the most locally cited documents, it is recommended to analyse the most cited references. This analysis inevitably includes papers that are only partially related to the actual topic (such as the citation of guidelines for analytical procedures). However, papers that are relevant to the topic area but are not included in the search are also taken into account. This was the case for some obviously relevant papers in the present evaluation, such as ASM1 (Henze et al. 1987) or the clarifier model by Takacs et al. (1991).

Structural analysis

As already explained in the method description, the interpretation of the results of a structural analysis is a great challenge. For the examined data set, all measures available in the software tools bibliometrix and VOSViewer were applied for structural analysis and the results obtained were evaluated. The following main findings emerged:

- Author-related networks:
  Due to inaccurate naming, some authors appear more than once in the networks. This questions the entire statement of such representations. Correction of errors in advance is therefore indispensable for this type of evaluation.

- Organisation-related networks:
  The previously undiscussed classification according to institutions of the publishing authors can be used as a supplement, but the achievable gain in knowledge in addition to the described evaluations is rather low.

- Country-related networks:
  These are basically dominated by the large scientific nations of the USA and China. The links and distances between them can be helpful in classifying networks spatially. Direct concrete findings for a subject-specific analysis are rather difficult to obtain.

- Source-related networks:
  Due to the dominance of the two journals Water Research and Water Science and Technology in the subject area studied, a network analysis is difficult. Other source-related measures presented here are more useful for this purpose.

- Document-related networks (citations, co-citations, keywords, co-occurrence):
  These networks can sometimes provide valuable supplements to other measures. Thematic connections and temporal developments can be identified.
| Authors       | HENZE M | GUJER W | MINOT | COMEAU Y | SPANJERSH | BATSTONE DJ | WENTZEL MC | OEHMENA | SIN G | RITTMANN BE |
|---------------|---------|---------|-------|----------|-----------|------------|------------|---------|-------|-------------|
| Articles      | 42      | 48      | 31    | 21       | 29        | 27         | 51         | 36      | 34    | 22          |
| Total Citations | 3,178   | 3,604   | 3,433 | 377      | 845       | 1,139      | 2,137      | 2,335   | 973   | 616         |
| Local Citations | 2,743   | 763     | 446   | 282      | 367       | 340        | 622        | 399     | 371   | 232         |
| Author impact | 65.3    | 15.9    | 14.4  | 13.4     | 12.7      | 12.6       | 12.2       | 11.1    | 10.9  | 10.5        |
| h-index       | 25      | 26      | 22    | 10       | 14        | 18         | 23         | 22      | 17    | 13          |
| g-index       | 42      | 48      | 31    | 19       | 29        | 27         | 46         | 36      | 31    | 22          |
| m-index       | 0.68    | 0.55    | 0.73  | 0.48     | 0.45      | 0.95       | 0.70       | 1.22    | 0.94  | 0.46        |

| Authors       | TAKACSI | LIU Y | EKAMA GA | JEPSSON U | DAIGGER GT | Gernaey KV | ORHON D | NI BJ | VANROLLEG-HEMPA | VAN LOOS-DRECHTMCM |
|---------------|---------|-------|----------|------------|-------------|------------|---------|-------|-----------------|--------------------|
| Articles      | 39      | 45    | 62       | 37         | 35          | 45         | 137     | 94    | 145             | 119                |
| Total Citations | 913     | 1,653 | 2,036    | 1,277      | 1,144       | 1,714      | 3,193   | 3,089 | 5,221           | 8,988              |
| Local Citations | 387     | 422   | 538      | 296        | 247         | 253        | 748     | 483   | 619             | 358                |
| Author impact | 9.9     | 9.4   | 8.7      | 8.0        | 7.1         | 5.6        | 5.5     | 5.1   | 4.3             | 3.0                |
| h-index       | 17      | 20    | 26       | 20         | 20          | 20         | 32      | 32    | 36              | 43                 |
| g-index       | 29      | 40    | 43       | 35         | 33          | 41         | 48      | 52    | 58              | 77                 |
| m-index       | 0.46    | 0.77  | 0.62     | 0.69       | 0.42        | 1.11       | 0.97    | 2.13  | 1.24            | 1.72               |
Due to the suitability of document-related measures for visualisation in networks, two examples were selected for further visualisation (Figure 8). The following features of the visualisation are used to code the properties in the data examined:

- Item size and font size: weight of measure
- Colour: cluster membership
- Lines: links between different elements; line thickness represents strength of link
- Distance: relatedness of connection

The network shown on the left of Figure 8 shows co-occurrence of papers. The right side shows the co-citation network of used keywords. A co-citation exists when a certain paper cites both connected papers. Depending on the number of such co-citations the item size increases in the network.

Co-occurrence exist when two certain keywords are both used in the same document. This makes it possible to analyse the proximity of keywords in terms of content and the thematic connection between them. This is particularly helpful when the analysis concerns an unknown subject area, because then knowledge and experience about related topics, concepts, methods or terms are usually lacking.

Analysis of the clusters of the co-citation network (Figure 8 left):

- Red: This represents the core cluster of activated sludge modelling, represented by Gujer et al. (1999) and Ekama et al. (1986). Due to its central position in the network, there is a close connection to dark yellow and blue/cyan.
- Green: This cluster very clearly represents publications on biological P elimination (Smolders et al. 1994). The clear separation and spatial distance to all other clusters is obvious, but there is also still a certain closeness as well as multiple links to the core ASM cluster (red).
- Dark yellow: This cluster bundles publications on practical and technical aspects of ASM use (main paper Henze et al. (2000), Takacs et al. (1991), Copp (2002)). The listed articles by Henze (compilation of ASM1-3) and Takacs (secondary clarifier model) represent ‘standard citations’ in wastewater treatment plant modelling. However, the practice of citing Henze et al. (2000) instead of the original papers on ASM 1-3 clustered therein seems questionable.
- Purple: This cluster contains various publications on modelling sorption processes with respect to dyes (Ho & McKay 1999) as well as papers related to pharmaceuticals and micro pollutants (Joss et al. 2006). Although both topics are classified in one cluster, the spatial distance shows certain differences and separations. Furthermore, the lack of links to other clusters and directly to each other is conspicuous. By modifying the settings when creating the network, a further delimitation of these two topics could possibly have been achieved.
- Cyan: In many publications on the practical simulation of wastewater treatment plants using operational data, the analytical guidelines used are cited. This is represented in this cluster by the sources of Dubois et al. (1956) and various versions of the procedures of the APHA (American public health association). This can also be deduced from the spatial proximity to red and dark yellow. The partial mixing with blue markers is noteworthy. Obviously, the algorithm has problems with a clean distinction with the selected settings.
- Blue: This cluster represents thematically membranespecific publications (Le-Clech et al. 2006). Here, too, the spatial distance to the red and yellow clusters is recognisable; nevertheless, a variety of links are present.
- Orange: This cluster is a special case, characterised more by its heterogeneity and distribution around the red and yellow clusters. These are predominantly practically oriented works (e.g. Stare et al. 2007). These have multiple connections and thus also proximity in terms of content to the two main clusters and represent a certain transition as well as links to the other clusters, because obviously the thematic scope is very high or heterogeneous. As a result, a concrete allocation and spatial classification, including a corresponding distance, appears rather difficult. Nevertheless, a small orange cluster outside the yellow cluster is noticeable.

From the co-occurrence network (Figure 8 right) of all keywords (author keywords and keywords plus in WoS), the content relationships of the clusters can be derived very well, as already summarised in the illustration. With this type of representation, it can be helpful to exclude the most frequently used keywords from the network generation in order to minimise dominance. In this case, this was done by hiding the keyword ‘activated sludge’. This also makes individual keywords more recognisable. Although the clusters are very clearly identifiable, some unusual classifications are noticeable. For example, the markers for ‘optimization’ (red) and ‘waste water’ (green) are not located in the actual cluster area.

This illustration represents a network for the entire period of the examined dataset. It is therefore not suitable for deriving time-related developments. A possibility of
network use taking into account a time reference is explained in the next section.

**Analysis of time-related developments**

The analysis of the total data set described so far is only a retrospective view and of limited use for deriving possible developments and future trends. Therefore, this chapter makes detailed use of time-related criteria. The resulting representations are summarised in Figure 9.

The time-related changes within the document types are shown in Figure 9(a). While the reviews show a largely constant but rather small share, a strong temporal dynamic can be seen in the share of proceeding papers. While at the beginning they accounted for about 30–40% of the documents, this has decreased since 2002 to currently only a negligible share. This is due to the initiatives and guidelines in the past to change the publication practice towards correctly reviewed articles, because the review process for conference submissions does not correspond to the practice usually applied to journals. This does not mean, however, that fewer documents have emerged from conference submissions. Many of the published documents are based on conference publications, but were subsequently subjected to another review process.

Another very interesting development can be derived from the Collaboration Index CI (Figure 9(b)). Between 1990 and 2020, this index almost doubled. This means that the number of authors per paper has almost doubled as well. This can be interpreted as the increasing complexity of scientific research, but also as a change in publication practice in order to help as many people as possible from one’s own working environment to increase the number of their publications. Neither of these assumptions can be directly tested with this form of data analysis.
The development of citations per document is shown in Figure 9(c). On the left axis, this is shown as an absolute value, and on the right axis, it is shown by year. It can be seen that publications have an average of 30–50 citations up to around 2005, while this then decreases to 0 for publications published thereafter. This is to be expected, since young publications cannot have been cited so often yet. At some point, a kind of saturation will occur or newer publications will be available that are cited with priority. For the evaluated dataset, it can be estimated that citations can be expected up to about 10 years after publication of a document, after which this decreases strongly. To eliminate this temporal aspect, the number of citations was related to the year and displayed on the right axis. This results in an average citation rate of 2.5–3 citations/document/year for ‘younger’ publications (from about 2000–2005).

The importance of individual sources in their total as well as with a time-based component has already been...
explained in a previous chapter and the results are shown in Figure 6 (right). Figure 9(d) shows the annual ranking of five sources that are included in the top 20 locally cited sources over the entire data period of 31 years. This shows that Water Research and Water Science and Technology are continuously in the top positions, while other journals have either declined in importance or gained in influence. It should be noted that especially in the case of ‘smaller’ journals, their influence can fluctuate greatly with the personal activity of individual authors or teams of authors who primarily publish in them. An example of this is the journal Water SA with a strong reference to South African authors and their strong activity in the subject area, primarily in the 1980s and 1990s.

The development of the most cited references over time is a very valuable method of analysis for changing interests within the subject area. Figure 9(d) shows the 10 most cited sources with the cumulative number of citations. The activity can be deduced from the increase. There are some documents that show a very continuous increase, while others stagnate. In addition, new documents are being added.

Kappeler & Gujer (1992) and Ekama et al. (1986) deal with modelling principles. These have apparently not been used for publications since 2007/2008. Either there are newer sources to cite or the topic no longer plays a role in current publication activity.

The situation is different with the publications by Lowry et al. (1951) and Dubois et al. (1956). Both have been cited continuously since about 2009/2010. However, this is not due to the late publication date of these documents, as might first be assumed when ‘new’ articles appear in such an evaluation. These papers from the 1950s contain basic principles for determining the composition of organic substrates. Due to the increased activities in modelling the anaerobic processes in sludge treatment in recent years (see also the following analysis of the temporal change of the keywords), there is a need for such a characterisation of the sludges used. The decreasing interest in modelling biological P elimination can also be seen from the decreasing citation frequency, as exemplified by the article by Smolders et al. (1994).

The last time-dependent evaluation was the analysis of the influence of the countries of the authors’ institutions. This is shown for 16 countries in Figure 9(f). Here, too, a relative representation was made in the form of the citations received on the total number of publications published in the form of a country impact. Only some data are shown with different colours in Figure 9(f) that were discussed in the following, while all other only show their more or less constant and minor relevance. Over the course of the 31 years under consideration, institutions from the Netherlands or USA have obviously lost a massive amount of impact with their publications in the subject area, while Switzerland remains at a constantly high level. Countries like Australia or Turkey and with minor characteristic also UK obviously intensified their scientific activities in this field in the 1990s and first had to establish themselves in the scientific community. It is also interesting to note that almost all countries except Switzerland, Denmark and Canada (at certain different levels) suffer from a declining interest in published papers, which is reflected in a continuously decreasing country impact. This could be interpreted by the current scientific practice of increased publication activity as the main measure for scientific success. It should be mentioned that this country-specific evaluation is subject to the disadvantage that only the countries of the institutions of the main authors are taken into account. Many international author teams are thus not represented properly.

Keyword development within the scientific field under investigation

The analysis of the keywords used and their temporal dynamics is only possible in an automated way due to the volume of data. By limiting the selected keywords to be examined manually with the described methodology, a detailed and technically based evaluation is then possible.

In total, 87 of the 250 keywords examined with the most frequent use in the individual years showed a time-dynamic behaviour. They were classified into the following groups:

- Increase in the use of this keyword (with subcategory ‘strong increase’)
- Decrease in the use of the keyword
- Temporarily increased use with subsequent decrease

The classification and the corresponding keywords with the numbers are listed in Table 5.

The keywords listed can in turn be grouped into thematic clusters and an evaluation can be carried out from the classification for this thematic field (abbreviations of the classification according to Table 5):

- (Biological) phosphorous removal: ia = 1; da = 7; ih = 7

Due to the large number of identified keywords with temporal dynamics from this field, there is a very strong incidence that this topic is obviously suffering from a very strong decline in interest.
In contrast to phosphorus elimination, this topic is gaining enormous growth and is obviously currently very intensive in research with various sub-aspects.

- Adsorption: ia = 6
This topic also shows a strong to very strong increasing tendency with regard to the use of activated carbon as well as the adsorptive purification of industrial wastewater (dyes).

- ASM fundamentals: da = 7
Keywords that can be associated with the model fundamentals show a continuously decreasing use in their range. This can be reconciled with the argumentation of Gujer (2006), according to which the fundamental topic has obviously been researched enough, but topics derived from it are now emerging as focal points, both in application-oriented research and in theoretical considerations of special topics.

In addition, temporal dynamics (+/-) can be derived from the keywords for the following topics:

- Pharmaceuticals, micro pollutants
- Granular sludge
- Biosorption
- Membrane technology
- Sludge settling properties

The analysis of the temporal behaviour of the keywords shows that only a retrospective evaluation is possible or, in the case of newly appearing topics, a critical mass of publications must first be reached in order to be detected with
this type of analysis. Then, however, it is already an established field of research.

The tools for network generation from bibliometrics can also be used to analyse the temporal development of time-dynamic developments in the thematic details studied. An example is shown in Figure 10 for the development of the keywords. For this purpose, an overlay with a time reference can be assigned to the markers. However, it can be seen from the automatically generated time axis that aggregations are obviously being carried out or that an incorrect weighting is being applied due to the exponentially increasing number of publications and thus also of keywords used.

Nevertheless, a topic-specific temporal clustering is clearly recognisable. While the ASM fundamentals and phosphorus-related keywords tend to appear in shades of blue in the early years, there is a shift towards the topics of biosorption, adsorption and anaerobic digestion as time goes on. This basically corresponds to the above findings from the temporal keyword analysis. However, it should also be noted here that these visually based evaluations should rather be used as a supplement to a detailed data analysis.

Another interesting possibility for analysing the temporal change in the importance of certain procedures, methods or keywords can be carried out via so-called s-curves of the number of articles per time (Mao et al. 2020). This determines the occurrence of a turning point in the increase of the number of the feature under consideration towards a stagnant development.

**CRITICAL COMMENTS ON THE METHODOLOGY**

During the work on this bibliometric analysis, various questions, problems, limitations and open issues arose, which will be mentioned below.

First of all, the selection of search words by Gujer (2006) should be discussed. The aim of the original publication was certainly to analyse the subject area of ‘activated sludge modelling’. This is too broad with the search term AuSuM actually used. A current search for the following search term (‘activated sludge model’ or ‘activated sludge modelling’ or ‘activated sludge modelling’) resulted in 836 (WoS) or 1,147 (Scopus) references for the period up to 2021. This is a magnitude lower than the search with the original search terms of Gujer (2006). From technical reasons it is also possible to use more complex search terms with use...
of logical operators to specify it more in detail. As already mentioned, however, this was also chosen identically as a basis for comparison in the present work.

Furthermore, the use of a large number of existing bibliometric measures must be critically questioned. Which statements and findings can really be obtained from a certain measure? It is recommended to work with the SBM instead of various indices, because these figures are more directly related to publication and citation activity (source on SBM). The comparison of various indices has shown that, for example, authors are given very different meanings as a result. However, it is precisely such indices that gain great popularity and are widely used in the evaluation of scientific productivity.

The biggest problem with data analysis based on authors has turned out to be the incorrect spelling of names. Mostly, these arise on the citation side through inaccurate use of first names, but also through different spellings of complex surnames. It can also not be excluded that problems arise from arbitrary abbreviation of first names due to the recording practice at the individual sources. Name changes due to marriage are also not taken into account. It is therefore recommended that authors only be referenced via unique identification numbers (e.g. ResearcherID, OrcidID), as is now good practice for articles via the digital object identifier (DOI). Subsequent correction of these errors is no longer possible, so that all author- and co-author-based evaluations should be critically examined. Since it seems difficult to impossible to do this retrospectively for all available authors as well, author-based analysis should always be used with very great caution and not as an essential element of a research.

Another significant issue is the question of the data source used. According to Larsen & von Ins (2010), future analyses should include several citation indices. This demand is fully agreed with. For the present work, only the WoS data were used for reasons of comparison with Gujer (2006). From this follows the need to have appropriately combinable data interfaces in the various sources in order to generate an overall database for a bibliometric analysis. However, due to the inaccuracies already contained in the existing individual databases, it is suspected that combining data from multiple sources will increase these errors.

The automated analysis of the temporal keyword development cannot concretely filter out special developments and new research directions. For example, the combination of ASM and CFD (Laurent et al. 2014), consideration of individual organisms (Gujer 2002; Meister & Rauch 2016) or the combination of ASM with microbiological data using machine learning and artificial intelligence (Sin & Al 2021).

Especially the last cited concept seems to be very interesting and fits perfectly into the current research activities focusing on microbiological aspects. However, it requires a completely new level of technical prerequisites and massively increases the general complexity.

New specializations in the field of activated sludge modelling have been developed in the past, promising a new level of knowledge. Models became more complex, new linkages into other specialties were created. At the same time, there is still a significant need for simple and easy to understand tools especially when used in practice or education. This range will always remain in the field of activated sludge modelling (as certainly in many other research areas). As an example the problem of complexity vs. simplicity is also intense discussed by Glover et al. (2006), although they work in the field of CFD and ASM combination.

If nothing else, this was very impressively demonstrated by the activated sludge models themselves. The ASM2 or ASM2D is much more complex than the ASM1, but is used much less frequently in practical applications than the ASM3 developed in parallel, which in turn contains fewer processes and fractions.

Therefore, any further development is of course to be welcomed. However, regular verification of new developments appears to be just as important. For this purpose, a bibliometric analysis can be used very well.

Many of the available tools for bibliometrics focus on generating maps and networks to visualise relationships. Due to the complexity of the data, this can be a very useful tool. From this author's own application, however, it can be stated that these visualisations are often very difficult to interpret. In any case, with the available setting parameters, a statement that is as clearly comprehensible as possible should be achieved. Therefore, it is recommended to use such visualisations in addition to the analysis process. This can be done in a mutual application of concrete bibliometric measures in order to secure or verify the findings in each case.

RECOMMENDATIONS FOR CONDUCTING A BIBLIOMETRIC ANALYSIS

In summary, the following are recommended or suggested for conducting such analyses:

- Combining sources from all available databases.
- Accounting for exponential growth by developing relative measures and representations.
- Focusing on internal links provides more important insights than as external references (e.g., citations, co-citations).
- A bibliometric analysis should only be the first (technical) step. It should be followed by a direct analysis including expert knowledge. Therefore, the methodology is rather unsuitable for unknown topics and in this case a person should be integrated into the research who has a suitable professional background.
- Individual journals, authors or institutions are only of limited use as indicators for important or relevant papers. In the case of authors, there is the problem of name inconsistencies, and in the case of journals, the publication policy can also be an influencing factor. For institutions, the general size or the number of employees has a direct effect.
- The analysis of keywords as well as their development over time has proven to be very informative. However, the available software tools can only be used for this purpose to a limited extent and separate data preparation is necessary.

Special importance is attached to so-called standard citations and standard keywords, in our case e.g. the publications on the ASM. These occur in a very large number, so that other effects may be underrepresented in the analyses. This is highly true for the visual representations of the networks.

CONCLUSIONS

The motivation and starting point for preparing this publication was the, in some ways, provocative publication by Gujer (2006) with the question of whether scientific resources should be used in other areas of water research instead of activated sludge modelling, because this topic has obviously already been sufficiently researched. At the present time, it can be stated that Gujer’s conclusions were not correct. The topic is still being intensively researched and many questions are still open or were newly developed.

With the present analysis, however, it could be shown that a further development and deepening within the topic area can be objectively proven through the use of bibliometric methods, but only as a first part of a more complex analysis. At the same time, the application of the diverse possibilities of bibliometric software tools raises the question of the significance and the evaluation and interpretation of the results.

A bibliometric analysis is very well suited to identify journals and documents with high impact. However, this should never be determined by absolute numbers, but should be represented by corresponding relative parameters. The analysis of author-based measures should be of secondary importance, because errors currently still have too great an influence. In general it is questionable if rankings of journals, authors or institutions are good indicators for high-quality papers.

Furthermore, in publication practice and bibliometric analysis, attention should be paid to special publication practices. Summarising and republishing essential documents can lead to problems in subsequent bibliometric analyses if both original publications and replications are cited in a different form. This can be demonstrated very well in the field under study with the ASM publications and their republication as IWA report (Henze et al. 2000).

Specifically, the analysis of the keywords used appears to be a powerful tool when related to a temporal basis. Thus, changes and developments in the publication orientation can be discovered or verified. However, due to the publication process, a time lag of several months up to one year has to be considered. Therefore it can only be used in a retrospective way. This type of analysis of the available scientific literature is suited in general to exploring subject areas outside the discipline, e.g. in the search for new fields of research or possible links between different scientific disciplines. In a first step, a comprehensive amount of information can be screened and systematised in order to then selectively analyse the essential sources in terms of content. In an age of increasing information abundance, it will therefore be difficult to conduct meaningful literature analyses without the tool of bibliometric analysis. The last step of such an exploration has to be done by use of expert knowledge, an automatic analysis will surely fail.

Therefore, it seems absolute necessary to use all available databases to compensate for differences in content coverage. In doing so, utmost care is required when combining information from different sources regarding correct naming and combination of data fields from different databases. In any case, the high dynamics due to the exponential growth of new sources must be taken into account in the analyses in order to avoid misinterpretations. This is not yet been included in many available bibliometric measures, so there is still a need for development here.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories.
REFERENCES

Aguillo, I. F. 2012 Is google scholar useful for bibliometrics? a webometric analysis. Scientometrics 91 (2), 343–351.

Aria, M. & Cuccurullo, C. 2017 Bibliometrix: an R-tool for comprehensive science mapping analysis. Journal of Informetrics 11 (4), 959–975.

Bornmann, L., Mutz, R. & Daniel, H.-D. 2009 Do we need the h-index and its variants in addition to standard bibliometric measures? Journal of the American Society for Information Science and Technology 60 (6), 1286–1289.

Bornmann, L. & Mutz, R. 2015 Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. Journal of the Association for Information Science and Technology 66 (11), 2215–2222.

Chen, C. 2004 Searching for intellectual turning points: progressive knowledge domain visualization. Proceedings of the National Academy of Sciences 101 (suppl 1), 5303–5310.

Chuang, K. Y., Wang, M. H. & Ho, Y. S. 2021 High-impact papers presented in the subject category of water resources in the essential science indicators database of the institute for scientific information. Scientometrics 87 (3), 551–562.

Copp, J. B. 2002 The Cost Simulation Benchmark: Description and Simulator Manual: A Product of Cost Action 624 and Cost Action 682. Office for Official Publications of the European Communities, Luxembourg. https://op.europa.eu/en/publication-detail/-/publication/8448ef88-37dd-4d1a-823f-3143e7902429.

Davarazar, M., Mostafaei, A., Jahanianfard, D., Davarazar, P., Ghiasi, S. A. B., Gorchich, M., Nemati, B., Kamali, M. & Aminabhavi, T. M. 2020 Treatment technologies for pharmaceutical effluents: a scientometric study. Journal of Environmental Management 254, 11.

Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. T. & Smith, F. 1956 Colorimetric method for determination of sugars and related substances. Analytical Chemistry 28 (3), 350–356.

Eghe, L. 2006 Theory and practise of the g-index. Scientometrics 69 (1), 131–152.

Ekama, G. A., Dold, P. L. & Marais, G. V. R. 1986 Procedures for determining influent COD fractions and the maximum specific growth rate of heterotrophs in activated sludge systems. Water Science and Technology 18 (6), 91–114.

Elango, B. & Rajendran, P. 2012 Authorship trends and collaboration pattern in the marine sciences literature: a scientometric study. International Journal of Information Dissemination and Technology 2 (3), 166. http://www.ijidt.com/index.php/ijidt/article/view/91.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D. & Barabási, A.-L. 2018 Science of science. Science 359 (6379), eaao0185.

Gaster, N. & Gaster, M. 2012 A critical assessment of the h-index. BioEssays 34 (10), 830–832.

Glover, G. C., Printemps, C., Essemiani, K. & Meinhold, J. 2006 Modelling of wastewater treatment plants – how far shall we go with sophisticated modelling tools? Water Science and Technology 55 (3), 79–89.

Gujer, W. 2002 Microscopic versus macroscopic biomass models in activated sludge systems. Water Science and Technology 45 (6), 1–11.

Gujer, W. 2006 Activated sludge modelling: past, present and future. Water Science and Technology 53 (3), 111–119.

Gujer, W., Henze, M., Mino, T. & van Loosdrecht, M. 1999 Activated sludge model No. 3. Water Science and Technology 39 (1), 183–193.

Harzing, A. W. 2007 Publish or Perish. Available from: https://harzing.com/resources/publish-or-perish.

Henze, M., Grady, C. P. L., Gujer, W., Marais, G. V. R. & Matsuo, T. 1987 A general model for single-sludge wastewater treatment systems. Water Research 21 (5), 505–515.

Henze, M., Gujer, W., Mino, T. & van Loosdrecht, M. 2000 Activated Sludge Models ASM1, ASM2, ASM2d and ASM3. IWA Scientific and Technical Report No. 9, IWA Publishing, London, UK.

Hirsch, J. E. 2005 An index to quantify an individual’s scientific research output. Proceedings of the National Academy of Sciences 102 (46), 16569–16572.

Ho, Y. S. & McKay, G. 1999 Pseudo-second order model for sorption processes. Process Biochemistry 34 (5), 451–465.

Joss, A., Zabczynski, S., Göbel, A., Hoffmann, B., Löfler, D., McArdell, C. S., Ternes, T. A., Thomsen, A. & Siegrist, H. 2006 Biological degradation of pharmaceuticals in municipal wastewater treatment: proposing a classification scheme. Water Research 40 (8), 1686–1696.

Kappeler, J. & Gujer, W. 1992 Estimation of kinetic parameters of heterotrophic biomass under aerobic conditions and characterization of wastewater for activated sludge modelling. Water Science and Technology 25 (6), 125–139.

Koseoglu, M. A. 2016 Mapping the institutional collaboration network of strategic management research: 1980–2014. Scientometrics 109 (1), 203–226.

Larsen, P. O. & von Ins, M. 2010 The rate of growth in scientific publication and the decline in coverage provided by science citation index. Scientometrics 84 (3), 575–603.

Laurent, J., Samstag, R. W., Ducoste, J. M., Griborio, A., Nopens, I., Batstone, D. J., Wicks, J. D., Saunders, S. & Potier, O. 2014 A protocol for the use of computational fluid dynamics as a supportive tool for wastewater treatment plant modelling. Water Science and Technology 70 (10), 1575–1584.

Le-Clech, P., Chen, V. & Fané, T. A. G. 2006 Fouling in membrane bioreactors used in wastewater treatment. Journal of Membrane Science 284 (1), 17–53.

Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. 1951 Protein measurement with the Folin phenol reagent. Journal of Biological Chemistry 195 (1), 265–275.

Mao, G., Hu, H., Liu, X., Crittenden, J. & Huang, N. 2020 A bibliometric analysis of industrial wastewater treatments from 1998 to 2019. Environmental Pollution 275, 115785.

Meister, M. & Rauch, W. 2016 Wastewater treatment modelling with smoothed particle hydrodynamics. Environmental Modelling & Software 75, 206–211.

Osman, A. I., Deka, T. J., Baruah, D. C. & Rooney, D. W. 2020 Critical challenges in biohydrogen production processes from the...
organic feedstocks. *Biomass Conversion and Biorefinery*.
https://doi.org/10.1007/s13399-020-00965-x.

Patience, G. S., Pounis, G., Patience, P. A. & Boffito, D. C. 2019 In: *Analysis in Nutrition Research* (G. Pounis, ed.). Academic Press, pp. 197–230.

Price, D. J. d. S. 1965 Networks of scientific papers. *Science* 149 (3683), 510–515.

Price, D. J. d. S. 1986 *Little Science, Big Science … and Beyond*. Columbia University Press, New York.

Pritchard, A. 1969 Statistical bibliography or bibliometrics. *Journal of Documentation* 25 (4), 348–349.

Sin, G. & Al, R. 2021 Activated sludge models at the crossroad of artificial intelligence – A perspective on advancing process modeling. *npj Clean Water* 4 (1), 16.

Smolders, G. J. F., van der Meij, J., van Loosdrecht, M. C. M. & Heijnen, J. J. 1994 Model of the anaerobic metabolism of the biological phosphorus removal process: stoichiometry and pH influence. *Biotechnology and Bioengineering* 43 (6), 461–470.

Stare, A., Vrečko, D., Hvala, N. & Strmčnik, S. 2007 Comparison of control strategies for nitrogen removal in an activated sludge process in terms of operating costs: a simulation study. *Water Research* 41 (9), 2004–2014.

Takacs, I., Patry, G. G. & Nolasco, D. 1994 A dynamic-model of the clarification thickening process. *Water Research* 25 (10), 1263–1271.

van Eck, N. J. & Waltman, L. 2014 *Citnetexplorer: a new software tool for analyzing and visualizing citation networks*. *Journal of Informetrics* 8 (4), 802–823.

van Eck, N. J. & Waltman, L. 2017 Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics* 111 (2), 1053–1070.

Wang, M. H., Yu, T. C. & Ho, Y. S. 2010 A bibliometric analysis of the performance of water research. *Scientometrics* 84 (3), 813–820.

Zandi, S., Nemati, B., Jahanianfard, D., Davarazar, M., Sheikhnajad, Y., Mostafaie, A., Kamali, M. & Aminabhavi, T. M. 2019 Industrial biowastes treatment using membrane bioreactors (MBRs) - a scientometric study. *Journal of Environmental Management* 247, 462–473.

Zheng, T. L., Wang, J., Wang, Q. H., Nie, C. H., Smale, N., Shi, Z. N. & Wang, X. N. 2015 A bibliometric analysis of industrial wastewater research: current trends and future prospects. *Scientometrics* 105 (2), 863–882.

Zheng, T. L., Li, P. Y., Wu, W. J., Liu, J. G., Shi, Z. N., Guo, X. S. & Liu, J. X. 2018 State of the art on granular sludge by using bibliometric analysis. *Applied Microbiology and Biotechnology* 102 (8), 3453–3473.

Zhu, J. W. & Liu, W. S. 2020 A tale of two databases: the use of Web of Science and Scopus in academic papers. *Scientometrics* 123 (1), 321–335.

First received 7 April 2021; accepted in revised form 6 May 2021. Available online 20 May 2021.