Evaluating the Scientific Impact of XSEDE

Fugang Wang
Indiana University
Bloomington, Indiana

Gregor von Laszewski*
Indiana University
Bloomington, Indiana
laszewski@gmail.com

Timothy Whitson
Indiana University
Bloomington, Indiana

Geoffrey C. Fox
Indiana University
Bloomington, Indiana

Thomas R. Furlani
Center for Computational Research,
University at Buffalo, SUNY
Buffalo, New York

Robert L. DeLeon
Center for Computational Research,
University at Buffalo, SUNY
Buffalo, New York

Steven M. Gallo
Center for Computational Research,
University at Buffalo, SUNY
Buffalo, New York

ABSTRACT

We use the bibliometrics approach to evaluate the scientific impact of XSEDE. By utilizing publication data from various sources, e.g., ISI Web of Science and Microsoft Academic Graph, we calculate the impact metrics of XSEDE publications and show how they compare with non-XSEDE publication from the same field of study, or non-XSEDE *peers* from the same journal issue. We explain the dataset and data soruces involved and how we retrieved, cleaned, and curated millions of related publication entries. We then introduce the metrics we used for evaluation and comparison, and the methods used to calculate them. Detailed analysis results of Field Weighted Citation Impact (FWCI) and the peers comparison will be presented and discussed. We also explain how the same approaches could be used to evaluate publications from a similar organization or institute, to demonstrate the general applicability of the present evaluation approach providing impact even beyond XSEDE.

CCS CONCEPTS

• **Information systems → Information systems applications**; *Information retrieval*; Data management systems;

KEYWORDS

Scientific impact, bibliometrics, h-index, Technology Audit Service, XDMoD, XSEDE

ACM Reference Format:

Fugang Wang, Gregor von Laszewski, Timothy Whitson, Geoffrey C. Fox, Thomas R. Furlani, Robert L. DeLeon, and Steven M. Gallo. 2018. Evaluating the Scientific Impact of XSEDE. In *PEARC '18: Practice and Experience in Advanced Research Computing, July 22–26, 2018, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3219104.3219124

*Corresponding Author.

PEARC '18, July 22–26, 2018, Pittsburgh, PA, USA

ACM ISBN 978-1-4503-6446-1/18/07…$15.00
https://doi.org/10.1145/3219104.3219124

1 INTRODUCTION

To identify the impact of *scientific advancements enabled by enhanced cyberinfrastructure*, it is important to conduct a comprehensive analysis of achievements that can be attributed to the use of the advanced infrastructure, such as that provided by the Extreme Science and Discovery Environment (XSEDE) [5, 18].

We use the bibliometrics approach to evaluate the scientific impact of XSEDE. By acquiring related publication and citation data from multiple sources we calculate various metrics that show the impact of the publications and how they compare to their non-XSEDE peers that were published in the same journals, or in the same field of study. By processing millions of publication data entries we normalized the citation count by field of study. This essentially eliminates the problem that different fields of study have different publication characteristics. We introduced a novel [19] method to compare the target publications group with their peers published in the same publication venue to further show how the target publications group performs compared to their peers within the same publication venue.

2 RELATED WORKS

Bibliometrics based analysis has been the most commonly used method to evaluate the research impact of an individual, a research group, or even an organization. Publication count and citation count based metrics provide an effective way to show the quantity and quality, and the impact of scientific research activities. For instance, it was used to evaluate the quality of research in the United Kingdom [16, 17]. Most popular college/university rankings use citation based bibliometrics as an important factor to evaluate the quality of their research, e.g., the overall publication count and citation count in a certain year or a year range; the number of papers published in certain top journals; the number of highly cited papers; rating the citation count of a paper or its percentile ranking, etc.

Compute Canada, a virtual organization similar to XSEDE, also uses a bibliometrics based analysis to evaluate the impact of their research [1].

Some previous limited work studied the impact of TeraGrid [6], the early version of XSEDE, by analyzing the publications for one specific research allocation quarter, which involved a very limited number of researchers and publications. Our work is unique in that it provides a *comprehensive* analysis superior in data volume, and with novel analyses approach such as Field Weighted Citation Impact and journal publication-based peers comparison.

In addition to the more intuitive direct metrics of publication and citation count, some other derivative metrics such as h-index [13] and g-index [11] combines both publication and citation count to generate one metric. I10-index [2] in contrast measures only the count of those publications that received at least ten references by other publications. In our evaluation we calculated such metrics for various XSEDE research entities to show the impact and comparison of the entities on the same level, e.g., individual, project, research field of study, organization, etc. The results are presented on the XDMoD scientific impact portal [4].

Usage based metrics [7, 8] have also been proposed including metrics such as views and downloads, instead of the more formal citations of publications. However the applicability of this approach may be limited because the usage data may not be available from a publisher, or different publishers may have different criteria to measure the usage data. Thus, it would create an inconsistent comparison for papers published by different publishers. For this reason we did not present such metrics in this paper.

3 METHODOLOGY

We first introduce the methodology we use in the bibliometrics analysis to evaluate the scientific impact of XSEDE publications. This includes the specification of the dataset and data sources used in the analysis, the approaches to define and calculate the various metrics, and the information about the sophisticated software and service framework developed to facilitate this unique and comprehensive evaluation study on XSEDE.

3.1 Dataset and data sources

Several data sets and sources are involved in this study which includes XSEDE publications, Microsoft Academic Graph (MAG), and Web of Science Data. For all data we used the same time period between 2005 and 2016, which is the same as for the XSEDE publications. We describe the basic features of the datasets next.

XSEDE publications. This dataset includes publications from XSEDE as well as from TeraGrid. This data is collected from two sources. One is from the user-submitted data from the XSEDE user portal; another is from the past TeraGrid/XSEDE project reports submitted to NSF. For the latter, we extracted the publications appendix from the reports and then parsed and curated with significant effort the publication records text, before putting them into a structured database. There were over 20 thousand raw entries.

Microsoft Academic Graph (MAG). This dataset was retrieved with the API provided by Microsoft. The data was then curated and cleaned and put into a MongoDB database. This dataset has about 58 million entries.

ISI Web of Science (WoS). For the publication venues with at least 10 XSEDE papers appearing in them, we retrieved all the publications published in them to facilitate the peers comparison study. This dataset has about 2 million entries.

3.2 Field Weighted Citation Impact Analysis

The field Weighted Citation Impact (FWCI) metric is proposed as one of the snowball metrics [10]. It calculates the average citation count of a target group of publications based on their field of science, and then compare that with the average citation count of the whole field of science in the same time period. The result is a ratio

$$FWCI = avg(CC_{group})/avg(CC_{field})$$

A *FWCI* value greater than 1 indicates that the pertinent publication group had more citations than the expected value of the field of science, while a value less than 1 indicates that the average citation count that the group received was less than the expected value for the applicable field of science.

In this study we introduced the following process to calculate the *FWCI* values for the XSEDE publications.

(1) Query every raw XSEDE publication by title against the MAG data set, and verify the matching ones by checking other properties such as published year. After this process we identify the verified matching records in the MAG for all valid XSEDE publications. During this process we use elasticsearch [12] to improve both the accuracy and the performance of the query. This is important because of the size of the dataset.

(2) For each of the 58 million MAG data records, we use the assigned field of study values, along with other related data form MAG, to *trace* upward to the top levels of the hierarchical fields. This process narrowed down the 30k different assigned fields of study to 19 overall top level *fields of study* as defined in the MAG dataset. One thing to note is that each publication is assigned to multiple science fields in the original publication records, and the final top level science field category of a publication may not be unique either. However as a lot of research publications are themselves multi-disciplinary we think that such results are valid and acceptable.
In the following analysis we counted a publication in all the top level science fields that we found following this tracing process.

(3) Once we have each and every publications in the MAG dataset, we can calculate the average citation count by each top level field, for all the MAG publications and XSEDE publications respectively. Following that we can calculate the ratio to get the FWCI values.

3.3 Metric for Journal Publication-based Peer Comparison

An importnat achievement is our novel and sophisticated *Journal Publication-based Peer Comparison (JPPC)* metric as discussed in [19]. We used the following process to obtain the data needed for the analysis.

(1) We start this analysis by querying all XSEDE publications against a third party data source - WoS [3]. The XSEDE data, as explained before, contains the publication entries extracted from past TeraGrid/XSEDE reports to NSF, and the publication data from the XSEDE user portal. Both are user-submitted data or compiled from user-submitted data, thus this query and verification process is needed to ensure the quality and accuracy of the dataset. This resulted in about nine thousand verified publications at the time of the study.
(2) From this verified publications list, we find the subset of all publication venues with at least 10 XSEDE publications. For each of the publications published in these venues we retrieve from WoS the extended metadata to get the exact volume and issue number of the publication venue. The reasons why we chose a threshold value of 10 to identify a publication venue subset are:
  (a) This ensures the statistical significance of the analysis results.
  (b) This eases the data retrieval work substantially. While we have ~1400 distinct publication venues identified from all the verified XSEDE publications, the subset when we use 10 as the minimum number of publications appearing in the venue was reduced to ~120 publication venues.
  (c) Using this criterion, the number of XSEDE publications in the peers comparison was about five thousand, or about 56% of all the verified ones. This represents a good portion of all the data.
(3) For all ~120 publication venues, we retrieved all the publications data published in them during the same time period as the TeraGrid/XSEDE publications (2005-2016).

Based on this data, we can establish suitable comparison peer groups, which are based each on a single journal issue (or journal volume when no issue data available for some publications) that an XSEDE publication appeared in. For each comparison peer group, we rank the citation count of each publication (including the XSEDE ones and the peers). The calculated percentile ranking values serve as the basis of the peer comparison study. The comparison is between publications that we identified as XSEDE papers and those that were not.
To apply the percentile ranking to the field of science of XSEDE publications among the journal issues where each publication was published, we aggregate them based on the project field of science data obtained from the XSEDE central database (XDcDB). These XSEDE fields of science are self-reported by the researchers. We then calculate the average and median percentile rank for each field of science (FOS).

3.4 Software Architecture Supporting the Study

We enhanced our previously developed software framework [20] to support the study, which includes data acquisition, cleanup, processing and presentation. The framework is based on a distributed set of software services. This service-oriented framework is integrated as part of a layered architecture consisting of components for:

- A data layer that retrieves publication and citation data from external sources. This includes data from the ISI Web of Science; Microsoft Academic Graph; Google Scholar, and very importantly the NSF award database.
- Business logic layer that deals with:
  - parsing and processing while correlating data from various databases and services, such as the XSEDE central database (XDcDB).
  - a metrics generation and analysis system for different aggregation levels – users, projects, organization, field of science.
- a presentation layer using a lightweight portal in addition to exposing some data via a RESTful API [4].

Due to the use of the Software as a Service (SaaS) approach, our framework is *expandable* as we are able to integrate new services and data resources as required. Hence our framework can be adapted to other resource providers as demonstrated in [19]. Obviously, adaptation could mean that we have to change the bibliometric data, which could mean that we need to integrate new data sources and curation services spending significant effort to integrate such data.

3.4.1 Service Integration into XSEDE and XDMoD. Our current framework for XSEDE includes services that are motivated by our initial findings from the XSEDE bibliometric data. A RESTful service is integrated into the *XSEDE User Portal* as part of the publication discovery service.
The various impact metrics of different levels of XSEDE entities - person, project, organization, field of study - as well as part of the analyses are available on the *XDMoD scientific impact portal* [4].

4 RESULTS AND DISCUSSIONS

In this paper, we discuss results specifically targeting the analysis of data related to XSEDE.

4.1 Field Weighted Citation Impact Metrics

First we show the calculated FWCI values in Figure 1. The plot lists the FWCI for the top-level fields of science as defined from the MAG data. Each data point also has the number of XSEDE publications as well as the number of all publications in that field. The red vertical line indicates the point at which FWCI=1. Figure 1 shows all fields but one (political science, with only 3 publications) that had FWCI values greater than 1, with the majority fields having much higher values.
In Figure 2 we display the same data but sort it based on the number of XSEDE publications in the field. This emphasizes the FWCI for the fields that the majority of the XSEDE publications fell within.
Figure 3 compares the expected citation count (denoted by mean_all) based on all publications in a given field of science with the actual average citation count (denoted by mean_xd) for XSEDE publications in each field of science, while including the FWCI values at the same time (line chart with dots). The plot was sorted based on the number of XSEDE publications in the field. This again indicates that XSEDE publications received much higher citation counts and implies a higher scientific impact than their non-XSEDE peers.
In Figure 4 we show the extra citations XSEDE publications receive for each field, compared to the expected overall field of science value. Figure 4 also indicates how much impact that access to XSEDE resources has on each individual field of science.

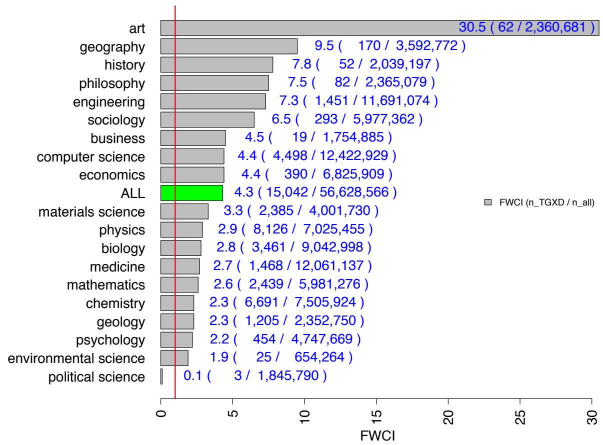

Figure 1: Field Weighted Citation Impact (FWCI) by Field sorted by FWCI.

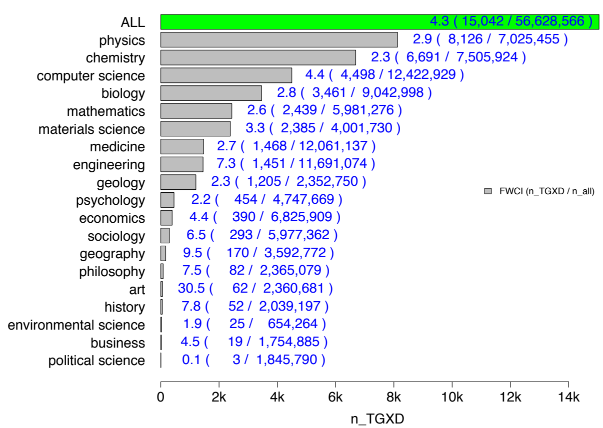

Figure 2: FWCI by Field sorted by Publication Count of the Field.

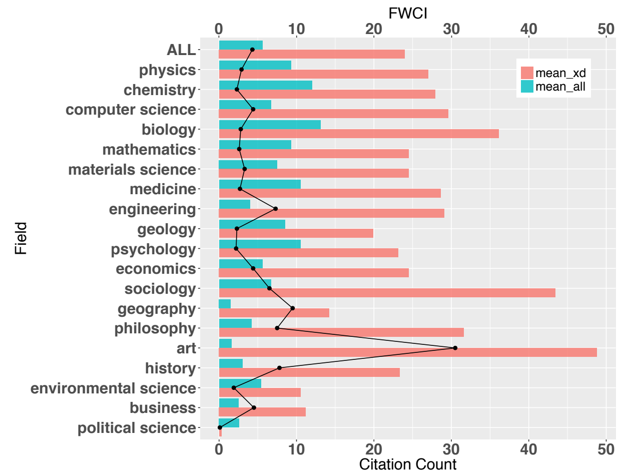

Figure 3: FWCI with Expected Citation Count and Actual Citation Count from XD Publications.

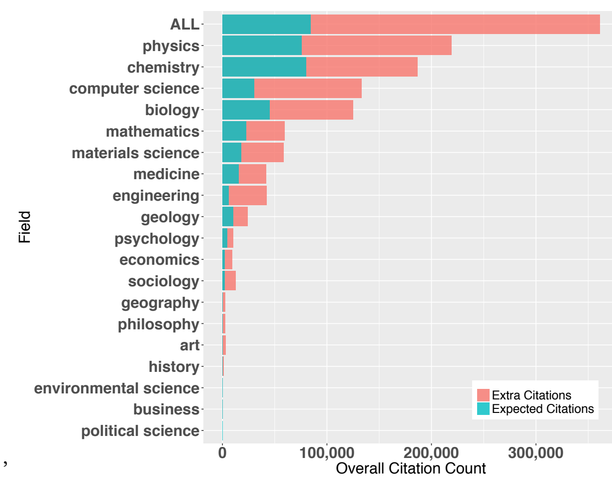

Figure 4: Extra Citation Count Achieved by XD Publications.

The availability of all the publications for each field makes it possible to calculate other interesting statistics, in addition to the above presented FWCI results. In Table 1 we display for each field the highly cited XSEDE papers (defined as top 1% and top 5% in citation count in that field) and the percentage of how many XSEDE publications fall into each category. The results show that for most fields a higher than expected percentage of XSEDE publications fall into the highly cited papers categories. E.g., when we consider all the publications and fields together, 4.8% XSEDE publications were in the top 1% highly cited group while 22.5% were in the top 5% highly cited group.

4.2 XSEDE Peer Data Analysis

Now we present a number of graphs and tables that show the results from the peer comparison study. Figure 5 shows the average percentile rank of XSEDE publications grouped by each publication venue. Figure 6 shows the same publication data but presents the median percentile rank values.

When we aggregate the results by fields of study instead of by individual journal, we get the results shown in Figure 7.

These plots show that for majority of the publications venues, or fields of science, XSEDE publications have a higher percentile ranking based on citation count.

When we consider the overall comparison results, Figure 8 shows the distribution of the XSEDE publication's percentile rank in each 10% increment group. Values above 50% indicate that the XSEDE publications are cited at a higher rate than their non-XSEDE peers. Again the result show the distribution skewed to the higher end, which means that XSEDE publications are cited more frequently than their non-XSEDE peers.

Figure 9 shows the empirical cumulative distribution of the percentile ranks compared to that of the peers group. The XSEDE publication curve is entirely to the right of the overall publication curve which is another indication that the XSEDE publications have a higher impact. Figure 10 shows the kernel density of the distributions of XSEDE publications' percentile ranking and that of

Table 1: Highly Cited Papers Statistics (in top 1% and 5%)

Field	# in top 1%	% in top 1%	# in top 5%	% in top 5%	# per 100,000	# XSEDE pubs
ALL	727	4.8	3380	22.5	26.6	15042
physics	292	3.6	1204	14.8	115.7	8126
chemistry	177	2.6	782	11.7	89.1	6691
computer science	223	5.0	1037	23.1	36.2	4498
biology	102	2.9	453	13.1	38.3	3461
mathematics	68	2.8	351	14.4	40.8	2439
materials science	108	4.5	446	18.7	59.6	2385
medicine	42	2.9	213	14.5	12.2	1468
engineering	111	7.6	414	28.5	12.4	1451
geology	33	2.7	183	15.2	51.2	1205
psychology	11	2.4	53	11.7	9.6	454
economics	26	6.7	101	25.9	5.7	390
sociology	18	6.1	63	21.5	4.9	293
geography	19	11.2	87	51.2	4.7	170
philosophy	11	13.4	31	37.8	3.5	82
art	15	24.2	39	62.9	2.6	62
history	6	11.5	19	36.5	2.6	52
environmental science	1	4.0	3	12.0	3.8	25
business	1	5.3	6	31.6	1.1	19
political science	0	0.0	0	0.0	0.2	3

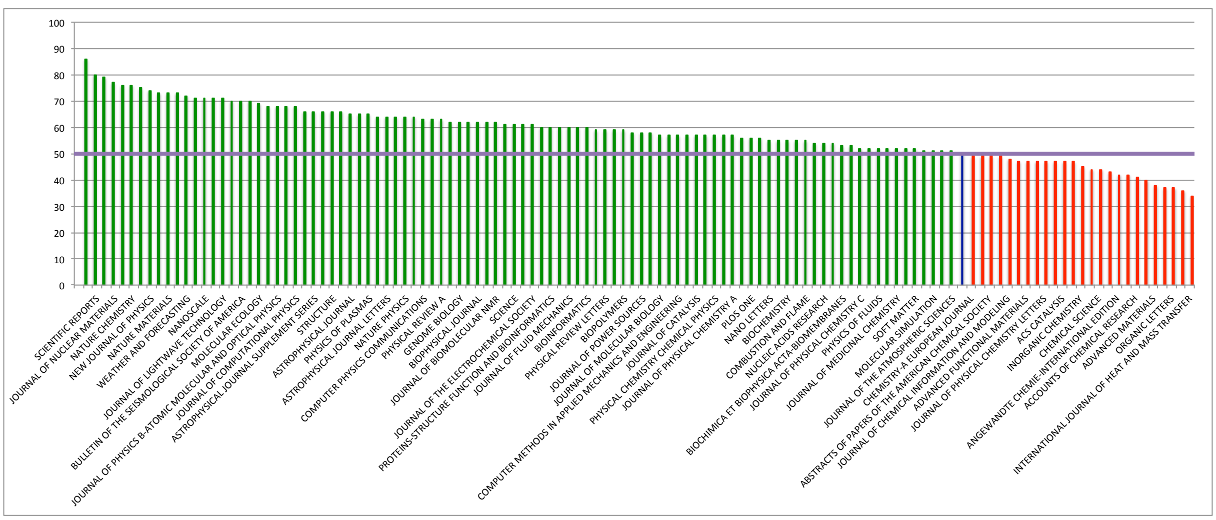

Figure 5: Average percentile ranking of XD publications by journal (by ISI)

peers'. As expected, the non-XSEDE peer publications are evenly distributed by percentile ranks with the spike at 50% mostly coming from more recently published journal issues where most publications were not yet cited. The XSEDE publications are weighted to the higher percentile ranks side again. This again shows that XSEDE publications tend to be more highly cited compared to their peers published in the same journal issue.

Table 2 lists the average and median rankings and citations received by XSEDE and non-XSEDE peer publication groups.

We used several non-parametric statistical tests to decide whether the XSEDE and non-XSEDE population distributions are identical without assuming that they follow a normal distribution. We used the Mann-Whitney-Wilcoxon test [15], Mood's median test [9], and Kruskal-Wallis test [14]. The results are as the following.

Table 2: Basic statistics of XSEDE publications group and peers group

	Number of	Rank		Citations	
	Publications	Average	Median	Average	Median
XD	5078	59	63	28	12
Peers	356464	49	49	15	5

Wilcox test for citation count

- W = 1160300000, p-value < 2.2e-16. Alternative hypothesis: true location shift is not equal to 0

Wilcox test for percentile ranking

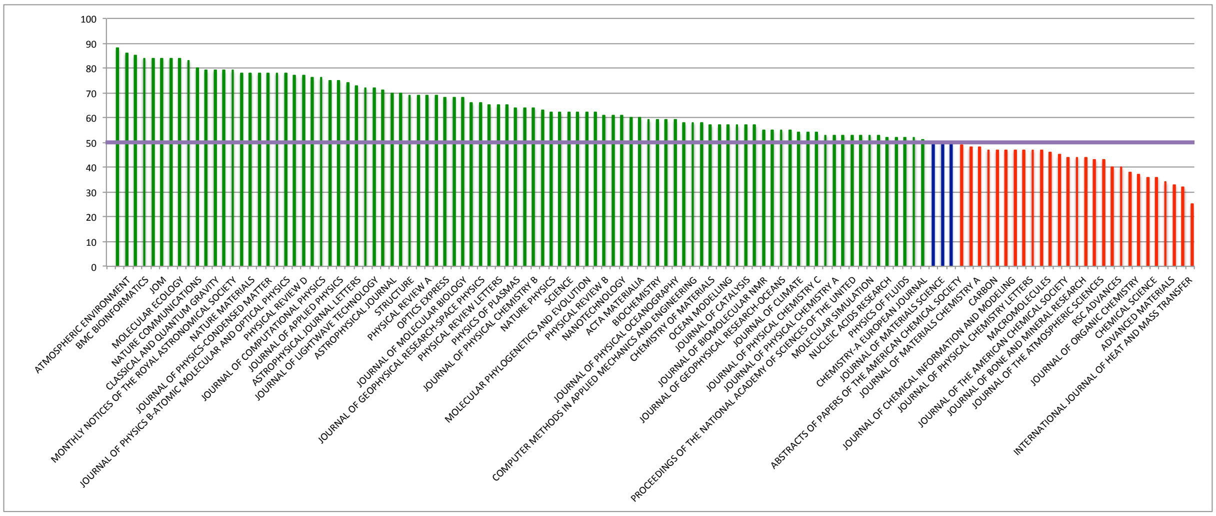

Figure 6: Median percentile ranking of XD publications by journal (by ISI)

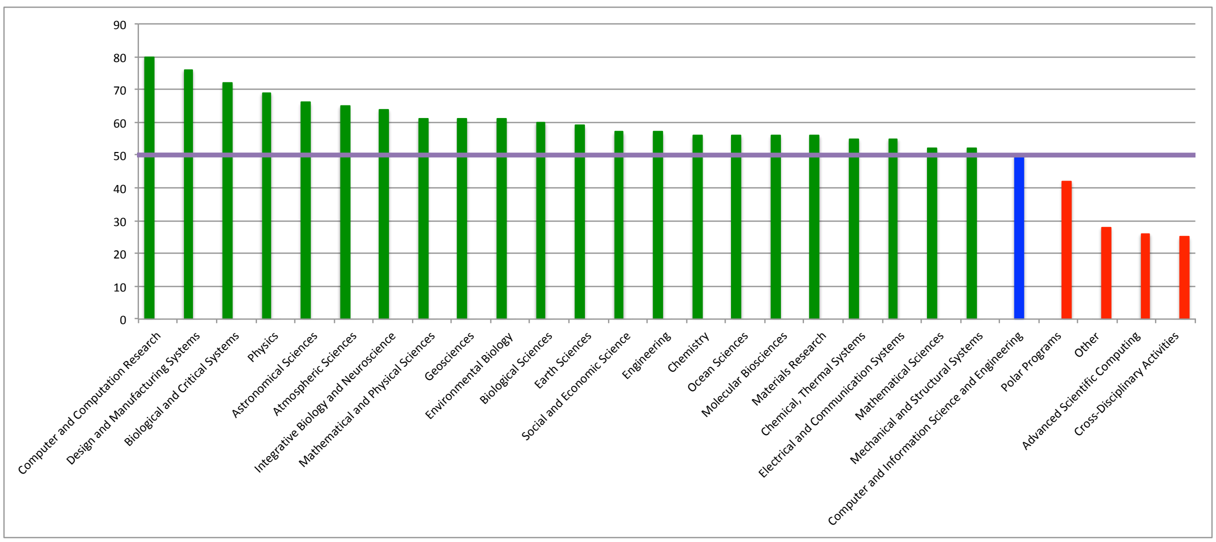

Figure 7: Average percentile ranking of XD publications by Field of Study (by ISI)

- W = 1090700000, p-value < 2.2e-16. Alternative hypothesis: true location shift is not equal to 0

Mood's median test for citation count

- p-value = 3.299883e-172

Mood's median test for percentile ranking

- p-value = 8.83052e-71

Kruskal-Wallis Test for citation count

- Kruskal-Wallis chi-squared = 1207.6, df = 1, p-value < 2.2e-16

Kruskal-Wallis Test for percentile ranking

- Kruskal-Wallis chi-squared = 632.35, df = 1, p-value < 2.2e-16

All of these results strongly indicate that the differences that we see between the XSEDE and the non-XSEDE publication metrics are statistically significant.

We also performed a T-test to test the citation count differences and percentile rank differences. Even though the distribution of the citation count of the XSEDE publication group and the peers group are not necessarily normally distributed, due to the central limit theorem, when the sample size is large enough, it is rational to use the T-test to not only test if there is a statistical difference between the two groups, as having been shown by the several previous tests, but also to quantify the difference between the means. The t-test results for both citation count and percentile ranking are given below.

T-test for citation count (Welch Two sample t-test)

- T=9.8328, df=5105.5, p-value< 2.2e-16, 95% confidence interval: $[10.90, 16.32]$

T-test for ranking (Welch Two sample t-test)

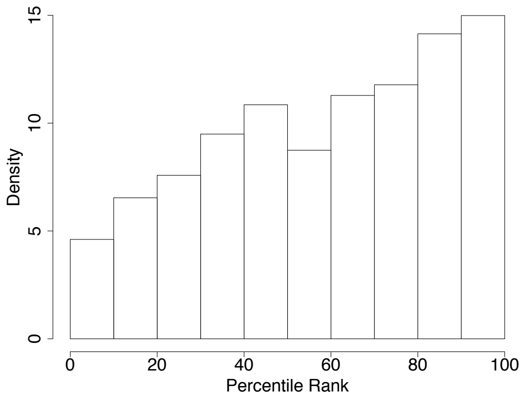

Figure 8: Histogram of Percentile Ranking

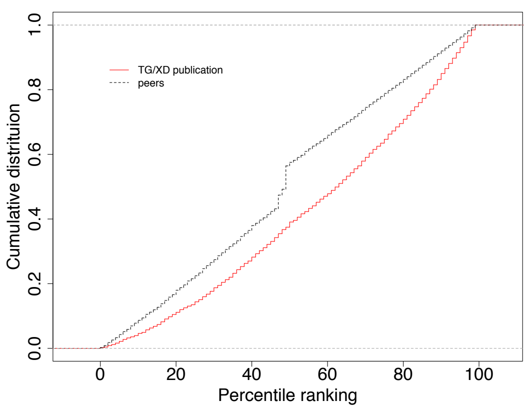

Figure 9: Empirical Cumulative Distribution of Percentile Ranks

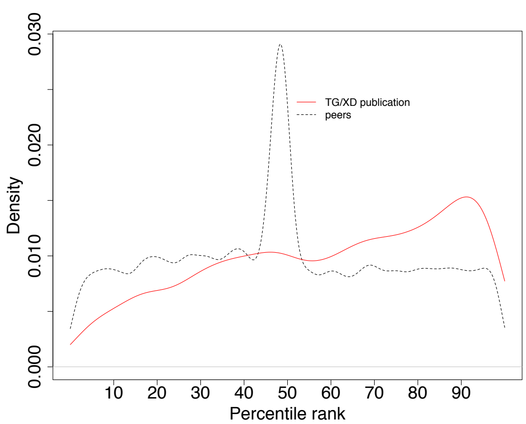

Figure 10: Kernel Density of the distributions of XSEDE publications' percentile ranking and that of peers'

- T=25.412, df=5105.5, p-value<2.2e-16, 95% confidence interval: [9.07, 10.59]

The results show that the XSEDE group has a statistically higher citation ranking and a statistically higher mean citation rate than the non-XSEDE peer group.

4.2.1 Journal peer comparison based on MAG data. Although we first integrated the MAG data in order to evaluate the field weighed impact of the XSEDE publications, we can follow the same approach as was done using the WoS data to conduct a similar peer comparison study. Figure 11 and Figure 12 show the average and median percentile ranking for XSEDE publications by each journal using MAG data. The overall results are pretty similar to what we got from the study with the WoS data.

5 CONCLUSION

We evaluated the scientific impact of XSEDE by examining the publications that were enabled by having access to the XSEDE resources. By curating the XSEDE publication data including cleansing, verifying and correlating the various data sources, we obtained a substantial very valuable dataset with which to compare and evaluate the scientific impact of XSEDE itself. While using two distinct analyses - *Field Weighted Citation Impact* analysis, and another novel *Journal Publications-based Peer Comparison* study, we found that XSEDE publications tend to be cited more than non-XSEDE publications. Various statistical tests show the results are statistically significant. The results from this study could potentially be used to inform to the XSEDE leadership team and the funding agency about the management of the facility, for example, to provide useful information to the resource allocation committee during proposal selection and approval. While the present study dealt exclusively with XSEDE data, the approaches and methods developed can be applied to evaluate publication data from a variety of different facilities or groups. In fact we have done similar analyses for NCAR, BlueWaters, and Bridges using the developed methodology and software framework.

6 ACKNOWLEDGMENTS

This work is part of the XSEDE Metrics Service (XMS) project sponsored by NSF under grant number OCI-1025159. Lessons learned from FutureGrid have significantly influenced this work. Gathering publications was first pioneered by FutureGrid, influencing the development in the XSEDE portal. We would like to thank Matt Hanlon and Maytal Dahan for their efforts to integrate part of the services into the XSEDE portal.

REFERENCES

[1] Compute Canada Report: Increasing Canadian Research Impact. Web Page. URL: https://www.computecanada.ca/compute-canada-report-increasing-canadian-research-impact/.
[2] i-10 index | google scholar citations open to all. URL: http://googlescholar.blogspot.com/2011/11/google-scholar-citations-open-to-all.html.
[3] ISI Web of Science. Web Page. URL: http://wokinfo.com/.
[4] XDMoD Scientific Impact Portal for XSEDE. Web Page. URL: https://sciimp.ccr.xdmod.org/xdportalpub/overview/.
[5] XSEDE. Web Page. URL: https://www.xsede.org/.
[6] J. Bollen, G. Fox, and P. R. Singhal. How and where the TeraGrid supercomputing infrastructure benefits science. *Journal of Informetrics*, 5(1):114–121, 2011.
[7] J. Bollen, M. A. Rodriguez, and H. Van de Sompel. MESUR: Usage-based Metrics of Scholarly Impact. In *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '07, pages 474–474, New York, NY, USA, 2007. ACM. URL: http://doi.acm.org/10.1145/1255175.1255273, doi:10.1145/1255175.1255273.

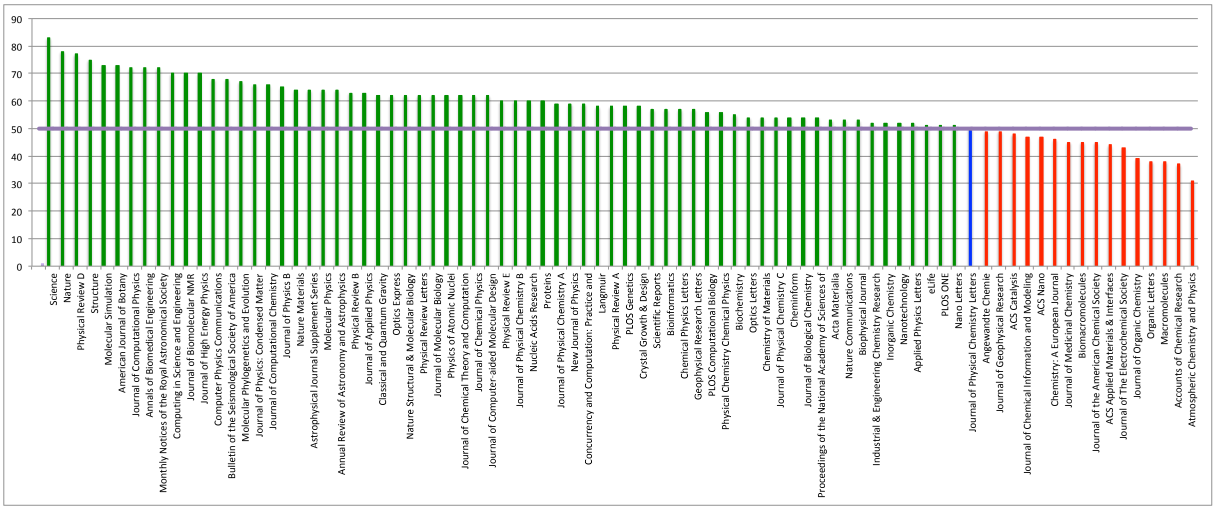

Figure 11: Average percentile ranking of XD publications by journal (by MS)

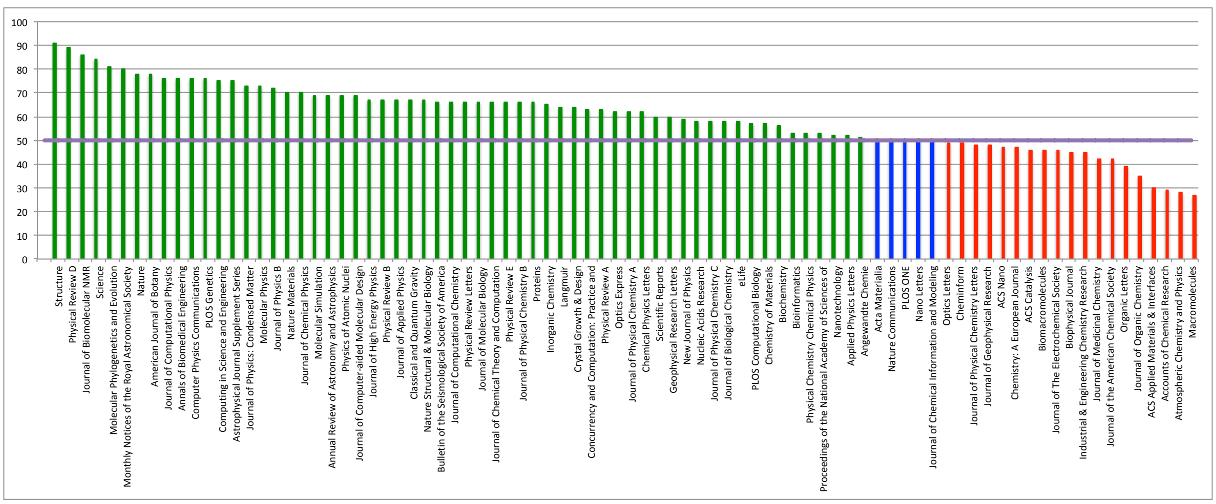

Figure 12: Median percentile ranking of XD publications by journal (by MS)

[8] J. Bollen, H. Van de Sompel, and M. A. Rodriguez. Towards Usage-based Impact Metrics: First Results from the Mesur Project. In *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '08, pages 231–240, New York, NY, USA, 2008. ACM. URL: http://doi.acm.org/10.1145/1378889.1378928, doi:10.1145/1378889.1378928.

[9] G. W. Brown, A. M. Mood, et al. On median tests for linear hypotheses. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. The Regents of the University of California, 1951.

[10] L. Colledge. Snowball metrics recipe book. 2014, 2014.

[11] L. Egghe. Theory and practise of the g-index. *Scientometrics*, 69(1):131–152, 2006.

[12] C. Gormley and Z. Tong. *Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine.* " O'Reilly Media, Inc.", 2015.

[13] J. E. Hirsch. An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences of the United States of America*, 102(46):16569–16572, 2005.

[14] W. H. Kruskal and W. A. Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.

[15] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60, 1947.

[16] T. Penfield, M. J. Baker, R. Scoble, and M. C. Wykes. Assessment, evaluations, and definitions of research impact: A review. *Research Evaluation*, 23(1):21–32, 2014.

[17] P. Thomas and D. Watkins. Institutional research rankings via bibliometric analysis and direct peer review: A comparative case study with policy implications. *Scientometrics*, 41(3):335–355, 1998.

[18] J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. Peterson, R. Roskies, J. Scott, and N. Wilkins-Diehr. XSEDE: Accelerating Scientific Discovery. *Computing in Science Engineering*, 16(5):62–74, Sept 2014. doi:10.1109/MCSE.2014.80.

[19] G. von Laszewski, F. Wang, G. C. Fox, D. L. Hart, T. R. Furlani, R. L. DeLeon, and S. M. Gallo. Peer comparison of xsede and ncar publication data. In *2015 IEEE International Conference on Cluster Computing*, pages 531–532, Sept 2015. doi:10.1109/CLUSTER.2015.98.

[20] F. Wang, G. von Laszewski, G. C. Fox, T. R. Furlani, R. L. DeLeon, and S. M. Gallo. Towards a scientific impact measuring framework for large computing facilities - a case study on xsede. In *Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment*, XSEDE '14, pages 25:1–25:8, New York, NY, USA, 2014. ACM. URL: http://doi.acm.org/10.1145/2616498.2616507, doi:10.1145/2616498.2616507.