

# Under-reporting research relevant to local needs in the global south. Database biases in the representation of knowledge on rice

Ismael Rafols,<sup>1,2</sup> Tommaso Ciarli<sup>2</sup> and Diego Chavarro<sup>2</sup>

<sup>1</sup>*i.rafols@ingenio.upv.es*

Ingenio (CSIC-UPV), Universitat Politècnica de València, València, Spain,  
SPRU (Science and Technology Policy Research), University of Sussex, Brighton, UK,  
and Observatoire des Sciences et Techniques (HCERES-OST), Paris, France

<sup>2</sup>*t.ciarli@sussex.ac.uk, diego.chavarro@sussex.ac.uk*

SPRU, Science Policy Research Unit, University of Sussex, Brighton, UK

## Introduction

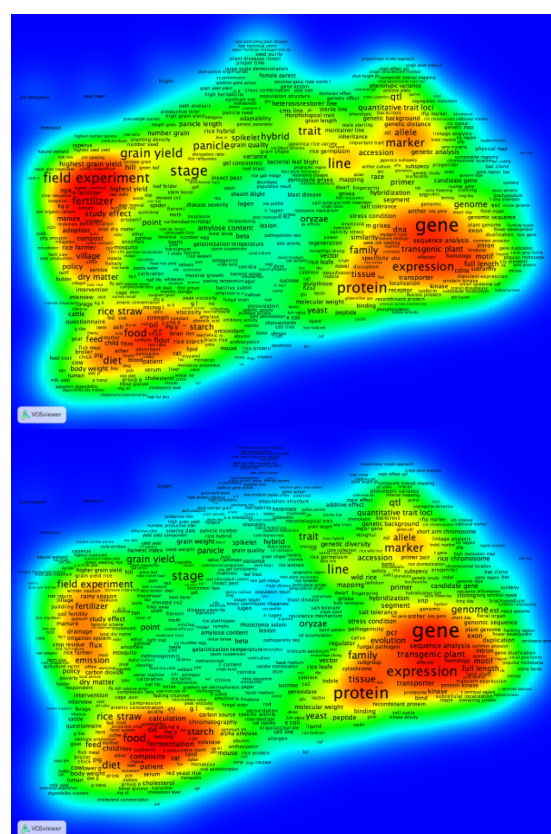
There is an increasing demand for science to help in addressing grand challenges or societal problems, such as tackling obesity, climate change or pandemics. In this context, it becomes important to understand what different sciences can offer to tackle these problems, and towards which directions scientific research should be developed. A useful starting point is to investigate what is the existing science supply, and which research options are better aligned to address grand challenges and societal demands (Sarewitz & Pielke, 2007). In order to map the science supply, we need a representation of the knowledge on research topics relevant for a problem.

Bibliometrics can provide very helpful tools for developing knowledge representations. However, these representations are highly dependent on the data and methods used. As a result, bibliometric tools or indicators often reproduce the biases in the data collection and treatment. For example, it has been shown that conventional bibliometric analyses are biased against non-English languages (Van Leeuwen et al., 2001), developing countries (Velho & Krige, 1986), applied science (Van Eck et al., 2013), the social sciences and humanities (Martin et al., 2010) and interdisciplinary research (Rafols et al., 2012). The aim of this paper is to investigate the biases introduced by available databases in the representation of research topics.

In a previous study on rice research, we showed that the bibliographic database CAB Abstracts (CABI) – which is focussed on agriculture and global health – has a larger coverage of rice research for most low income countries than Web of Science (WoS) or Scopus (Ciarli, Rafols & Llopis, 2014). For example, India has twice the number of publications in CABI on rice compared to Scopus and about 4 times those in WoS. In this study, we present evidence that shows that this unequal coverage distorts significantly the knowledge representation of rice research, globally and for different countries. Such bias may have

policy effects, in particular for a societal issue such as rice production.

As shown in Figure 1, we find that the journal coverage of the bibliometric databases WoS and Scopus under-represent some of the more application oriented topics (namely: i) production, productivity and plant nutrition (top left); ii) plant characteristics (top center); and iii) diseases, pests and plant protection (center).

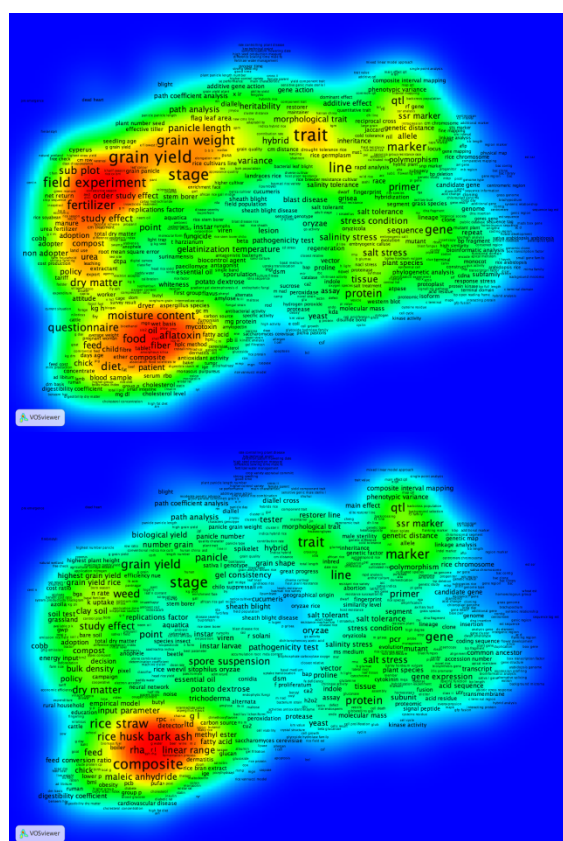


**Figure 1. Publication density for rice research in CABI (top) and in WoS (bottom). The top left and top right areas under-report in WoS are related to production and seed characteristics.**

Given that these are issues relevant to small farmers, producing for the local market, and with no access to the seeds developed with molecular biology techniques (GM – bottom left), we pose the

question whether the inadvertent effect of the biases in the dominant database is to under-represent, the type of research that has most chances of being relevant for improving their wellbeing, without introducing the use of the highly contested GM seeds.

Figure 2 illustrates that under-representation of research on production, pest and seed characteristics is particularly acute in some countries with molecular biology research (related to GM), but with a focus on research to address food security and local farming needs (in this case Iran). Rice research in these countries tends to be more focused on increasing crop yield, precisely the topic under-represented in WoS and Scopus.



**Figure 2. Publication density for rice research in Iran for CABI (top) and WoS (bottom).**

## Conclusions

Since knowledge representation can play a significant role in framing research strategies, policy and technological development, in this ignite talk we want to draw attention to the topic bias in the dominant bibliometric databases. From a technical point of view, few bibliometric and science policy experts will be surprised to hear that WoS and Scopus, are under-representing low income countries and more applied research. Given these results, we pose the question whether such conceptual biases may result in strategies that do not take into account knowledge and techniques which may be developed in closer connection to

farmers and consumers local needs. This study does not answer this question, but it shows that it is a meaningful and important issue for bibliometrics to address: bibliometric exercise that use dominant databases may have a negative effect on policies relevant to important social issues, particularly in developing countries.

## Information on methods and data

Publications on rice for the period 2003-2012 were downloaded from the WoS (including SCI-Expanded, SSCI, A&HCI, CPCI-S i CPCI-SSH) searching “rice” or “oryza” in the field “topic”. Scopus records were downloaded searching in title, abstract or keywords, i.e. TIT-ABS-KEY (“rice” OR “oryza”). Similarly, documents with “rice” or “oryza” were searched in title and abstract of the database CAB Abstracts. The records of the different databases were matched with multiple matching algorithms. The analysis was carried out using Vantage Point, the statistical package R and the visualisation programme VOSviewer.

## Acknowledgments

We acknowledge support from the EU (Marie Curie Integration fellowship to IR), the UK ESRC (RES-360-25-0076) and the US NSF (Award #1064146). The findings and observations contained in this paper are those of the authors and do not necessarily reflect the views of the funders.

## References

- Ciarli, T., et al. (2014). The under-representation of developing countries in the main bibliometric databases. *Proceedings of the S&T Indicators Conference* (97–105). Leiden.
- Martin, B. R., Tang, P., Morgan, M., & al. (2010). *Towards a Bibliometric Database for the Social Sciences and Humanities – A European Scoping Project* (A report for DFG, ESRC, AHRC, NWO, ANR and ESF). Brighton, UK: SPRU.
- Rafols, I., et al. (2012). How journal rankings can suppress interdisciplinarity. *Research Policy*, 41(7), 1262–1282.
- Sarewitz, D., & Pielke, R. A. (2007). The neglected heart of science policy: reconciling supply of and demand for science. *Environmental Science & Policy*, 10(1), 5–16.
- Van Eck, N. J. et al. (2013). Citation Analysis May Underestimate the Impact of Clinical Research as Compared to Basic Research. *PLoS ONE*, 8(4), e62395. doi:10.1371/journal.pone.0062395
- Van Leeuwen, T. N., et al. (2001). Language biases in the coverage of the Science Citation Index. *Scientometrics*, 51(1), 335-346.
- Velho, L., & Krige, J. (1984). Publication and citation practices of Brazilian agricultural scientists. *Social Studies of Science*, 14(1), 45-62.