Journal of Informetrics 12 (2018) 31–41

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Double rank analysis for research assessment

Alonso Rodríguez-Navarro [a,*], Ricardo Brito [b]

[a] Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid, Campus de Montegancedo, 28223-Pozuelo de Alarcón, Madrid, Spain
[b] Departamento de Física Aplicada I and GISC, Universidad Complutense de Madrid, 28040, Madrid, Spain

ARTICLE INFO

Article history:
Received 7 July 2017
Received in revised form 18 October 2017
Accepted 20 November 2017
Available online 27 November 2017

Keywords:
Research assessment
Bibliometric analysis
Citation distribution
Lognormal distribution
Double-rank analysis

ABSTRACT

Reliable methods for the assessment of research success are still in discussion. One method, which uses the likelihood of publishing very highly cited papers, has been validated in terms of Nobel prizes garnered. However, this method cannot be applied widely because it uses the fraction of publications in the upper tail of citation distribution that follows a power law, which includes a low number of publications in most countries and institutions. To achieve the same purpose without restrictions, we have developed the double rank analysis, in which publications that have a low number of citations are also included. By ranking publications by their number of citations from highest to lowest, publications from institutions or countries have two ranking numbers: one for their internal and another one for world positions; the internal ranking number can be expressed as a function of the world ranking number. In log–log double rank plots, a large number of publications fit a straight line; extrapolation allows estimating the likelihood of publishing the highest cited publication. The straight line derives from a power law behavior of the double rank that occurs because citations follow lognormal distributions with values of μ and σ that vary within narrow limits.

1. Introduction

Research assessment is the cornerstone of research policy. States and private companies invest large amounts of funds and other resources in scientific and technological research; therefore, as in any other productive system, research must be analyzed in terms of productivity and cost efficiency (Garfield & Welljams-Dorof, 1992) for the sake of taxpayers and shareholders. However, in contrast to the case for other productive systems, this simple idea hides a complex problem because for a long time there has not been complete agreement about the procedure for measuring research performance. Thus, based on the numbers of papers or of citations, a profusion of indicators have been proposed (van Noorden, 2010; historical studies in Delanghe, Sloan, & Muldur, 2011; Godin, 2006; Leydesdorff, 2005) that define research success "operationally" as simply amounting to the score of the proposed index (Harnad, 2009).

Fundamental to this issue the question arises of whether the assessment of the success of basic research in countries and institutions is better represented by the total number of published papers or by only the number of papers that were more influential and received a high number of citations. The origin of this question is partially conceptual, as it depends on whether a Kuhnian view of scientometrics is accepted (Andras, 2011; Martin & Irvine, 1983; Rodríguez-Navarro, 2012), but a wrong answer has real consequences. A good example is the European paradox and the notion of the excellence of European research that has led the research policy of the European Union astray for 20 years. By wrongly identifying excellence with

* Corresponding author.
E-mail addresses: alonso.rodriguez@upm.es (A. Rodríguez-Navarro), brito@ucm.es (R. Brito).

https://doi.org/10.1016/j.joi.2017.11.004
1751-1577/

the total number of publications, research policy has focused on the transfer of knowledge to the manufacturing sector when the real problem has been insufficient knowledge generation (Bonaccorsi, 2007; Dosi, Llerena, & Labini, 2006; Herranz & Ruiz-Castillo, 2013; Rodríguez-Navarro, 2016; Rodriguez-Navarro & Narin, 2017).

Counting publications with a high number of citations seems a simple dichotomous procedure to estimate the size of research output that is very influential. In another approach, the citation range can be divided into categories, considering that all publications in the same category have similar merit, and comparing the share of the publications in each category (Albarrán, Herrero, Ruiz-Castillo, & Villar, 2017). Percentile-based categories (Bornmann & Mutz, 2011; Bornmann, Leydesdorff, & Mutz, 2013; Waltman & Schreiber, 2013) can be used in both approaches. In dichotomous procedures, percentile indicators count the number or percentage of articles from a country or institution that belong to the top-x% of all cited papers in the world and that therefore exceed a certain citation level. At low citation levels, which apply to high percentiles (e.g., top-50%), the method provides results that are not very different from counting all publications, but this does not occur for small percentiles (e.g. the top-1.0 or top-0.1%), which implies highly cited papers. As mentioned above, the percentile dichotomous method is statistically robust but ambiguous without further definitions, because when two countries or institutions are compared, the research performance ratio that results varies depending on the percentile used (Rodríguez-Navarro, 2012, 2016). Furthermore, even if the smallest percentile that can be reliably applied to most countries and institutions, the top-1%, is used, the citation level of most publications in this percentile is not very high and the results do not correlate with the number of Nobel Prizes garnered by countries and institutions (Rodríguez-Navarro, 2011b).

Previous research has demonstrated that the capacity to publish highly cited papers reflects the capacity of the research actors to make important discoveries or to promote important advancements in science; this approach has been validated by correlation with the number of Nobel Prizes garnered (Rodríguez-Navarro, 2011a, 2016). Unfortunately, this correlation is satisfied when the citation level is very high, which gives rise to another problem because the number of publications with such a high citation level is too low to be counted reliably. As an alternative this number can be calculated instead of counted by using the function that describes the upper tail of the citation distribution. In most of the cases studied the data in this tail fit a power law with exponential cut-off or a lognormal distribution (Brzezinski, 2015; Katz, 2016; Price, 1976; Ruiz-Castillo, 2012); a power law function has been used for this calculation purpose with very active research actors (Rodríguez-Navarro, 2016). However, unfortunately, very few countries and institutions can be evaluated using this method because the proportion of publications that can be treated as a power law in the upper tail is normally low. Although across research areas, the average proportion is 2% of all articles, the percentage is much lower in many cases (Brzezinski, 2015; Ruiz-Castillo, 2012), which implies that in many institutions and countries the tail of the distribution that can be used for the evaluation is practically non-existent.

Taking into consideration the issues raised in this brief discussion, it seems that a convenient method of research assessment should allowed to calculate the capacity of the research actors to publish highly cited papers but that this calculation should be made by using the total number or a large proportion of their publications. With such a method, even not very active countries or institutions could be evaluated by their capacity to publish very highly cited papers. Pursuing this goal we describe here the double rank analysis, a new method that reveals the structure of the citation counts of the papers published by a country and institution in relation with the papers of the world.

To describe the double rank analysis, this article has two parts. In the first part (Section 2) we describe the characteristics of the double rank plots constructed with empirical data from two quiescent and two hot research areas. These double rank plots showed power law behavior. In the second part (Section 3) we investigated the mathematical reasons for this behavior. For this purpose we studied the double rank plots generated from simulated citation distributions that follow lognormal distributions.

2. Empirical double ranks

To perform the double rank analysis for the publications of a country or institution in a research field, we first construct two citation-rank plots, one for the country or institution and another for the world, by ranking the publications from the highest to the lowest number of citations. Because the publications of the country or institutions are in the two plots, they have two ranking numbers and the internal ranking number can be expressed as a function of the world ranking number. This function indicates how relevant are the publications of the selected country compared with the publications of the world. For instance, highly competitive institutions will typically get low numbers in the world ranking number, while low-performing institutions will concentrate many publications at the end of the world list, showing very high world ranking numbers. Double rank analysis can be interpreted as a Zipf's plot (Newman, 2005) in which the world ranking number substitutes for the number of citations.

The rest of this section applies the double rank analysis to several countries and research areas in order to reveal the underlying structure of citation counts. In particular, we will study two research areas, "plant sciences," and "physiology", and two research topics, "graphene" and "photovoltaic cells."

2.1. Citation counts

Citation counts were retrieved from the Thomson Reuters Web of Science (WoS) and its "Advanced Search" feature. To retrieve the corresponding publications from the WoS, we used the tags for the research area (SU=), topic (TS=), and year

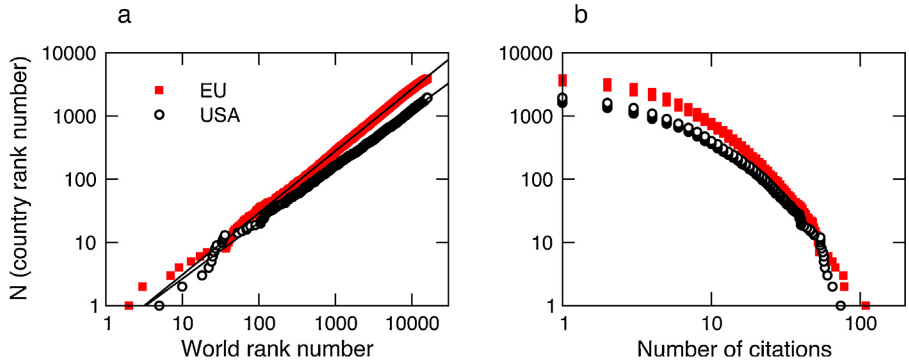

Fig. 1. Double-rank (a) and rank/citation (b) plots of domestic articles from the USA and EU published in 2012 in the WoS research area of 'plant sciences'.

of publication (PY=). We also studied certain institutions; for this purpose we used the organization-enhanced tag (OG=). In all cases we counted only "articles." The publication year was 2012 and we used a citation window of two years either 2014–2015 or 2015–2016 or a three-year window, 2014–2016. As explained below, for country or institution publications, we sometimes counted only domestic articles.

It is worth noting that a two-year citation window might be too narrow for evaluation purposes but not for our purposes; we used it because in the two hot topics that we investigated the number of publications is increasing very rapidly. For example, in a five-year window (2011–2015) the number of world publications on graphene increased almost four times (Rodriguez-Navarro & Narin, 2017). This rapid increase implies that the probability of citation of a paper published in 2009 is four times higher in the last than in the first year of the citation window. This does not occur in many other research areas; to avoid these enormous differences between areas we used a two- or three-year counting window in all cases. For evaluating purposes, domestic articles might not reflect the whole research capacity of a country but this is irrelevant for the purpose of our study.

2.2. Construction of the double rank plot

To construct the double rank plots we sorted the articles in decreasing order of the number of citations, numbering them from 1, and plotted their ordinal numbers as a function of their number of citations (Newman, 2005). We omitted articles without citations but this omission did not affect to our study. We show below that, in several cases, even articles with two or three citations deviate from the dominant trend of the double rank plot and cannot be used for the predictive purposes of the double rank plot. When several articles had the same number of citations, which occurred repeatedly at low citation levels, we ordered them by their publication date. Once the world and country articles had been ordered, each country article had one ranking number in each series of articles and we constructed the double rank plots by plotting the country ordinal number as a function of its world ordinal number.

2.3. Results of the double rank analysis

As a first step we studied the double rank plot in two quiescent research areas, "plant sciences" and "physiology" (Web of Science), which are characterized by a much lower number of citations than areas that include hot topics (graphene, lithium batteries, photovoltaic cells, stem cells, etc.). In "physiology" we compared the USA and the EU, and in "plant sciences" we made the same comparison and also compared three countries that are supposed to be very different in competitiveness but publish a similar number of articles: Brazil, Germany, and Spain.

Using log-transformed data, the double rank plots for the USA and EU followed straight lines in both research areas (Figs. 1 and 2a). Remarkably, only some of the first points (5–30), which correspond to the most cited articles, were noisy and showed notable deviations with reference to the straight line fitted to the rest of the points. Such highly cited, cutting edge research publications are intrinsically subject to heavy fluctuations, as they are top papers containing seminal contributions. The omission of these points has little relevance because the total number of points that could be used for fitting the straight line was very high: 3918 and 1959 in "plant sciences," and 2113 and 2423 in "physiology" for the EU and the USA, respectively. Thus, these points could be omitted from the fitting without any significant effect on further calculations.

We mentioned above that the upper tail of the distribution of citations can be fitted to power law with exponential cut-off or lognormal functions (Brzezinski, 2015; Katz, 2016; Price, 1976; Ruiz-Castillo, 2012) and that a power law function can be used to predict the frequency of highly cited publications in the USA and EU (Rodríguez-Navarro, 2016). However, in comparison with the straight-line fittings in the double rank plots, fitting straight lines to the upper tails of the rank/citation plots of the log-transformed data was much more uncertain (Figs. 1 and 2b). In the case of "plant sciences" reasonable fittings were not achievable; in the case of "physiology," the power law tail might include one hundred points, but significantly different straight lines were obtained depending on the number of points that were fitted. Despite these problems, the

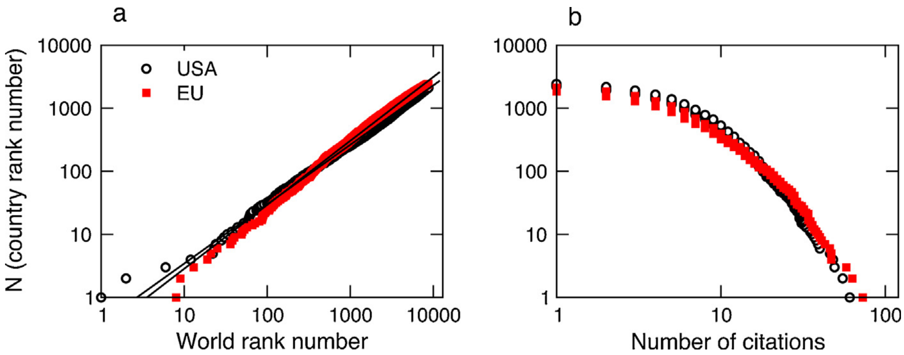

Fig. 2. Double-rank (a) and rank/citation (b) plots of domestic articles from the USA and EU published in 2012 in the WoS research area of 'physiology'.

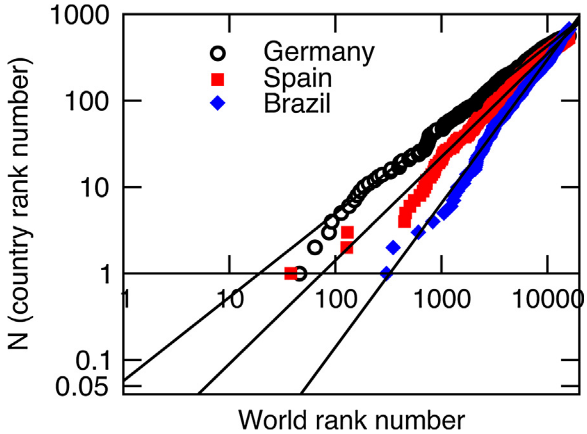

Fig. 3. Double-rank plot of domestic articles from Germany, Spain and Brazil published in 2012 in the WoS research area of 'plant sciences'. Extending the regression lines to cut the y-axis is a simple approach to estimate the likelihood of publishing the most highly cited publication. Total extension is shown only for Germany. A more precise mathematical calculation is given in the text.

shape of the tails revealed minor differences between the EU and the USA in research performance in these research fields, which was consistent with the results of the double rank analysis.

The double rank plots of three countries that publish similar number of papers in "plant sciences" but are supposed to be very different in research efficiency: Brazil, Germany, and Spain (Fig. 3), also followed straight lines, excluding very few points corresponding to the most cited publications. The slope of the lines indicated the research efficiency of the countries, lower slopes for higher efficiencies.

In preliminary results of this study (not shown) we found that the rank/citation log–log plots in hot topics—graphene and photovoltaic cells—were noisy and, in most cases, the data points visibly deviated from straight lines, which do not occur in other research areas (Rodríguez-Navarro, 2016). Consequently, in graphene and photovoltaic cells, the absence of a function that describes the tail made it impossible to estimate the frequency of publishing a very highly cited paper. In contrast with these results, the double rank plots of the log-transformed data could be fitted reasonably well by straight lines, and included hundreds of points. Fig. 4 shows the double rank plots of domestic articles in 2012 from the EU and the USA on photovoltaic cells, and from the USA, South Korea, and Germany on graphene. These results show that the double rank plots in hot topics were more complex and noisy than those in the two quiescent research areas studied above, but that straight lines could be fitted if the articles with a higher number of citations were omitted. The statistical noise of these points was demonstrated by plotting together the articles on graphene from the USA in two adjacent years (Fig. 5). It is reasonably to assume that the world competitiveness of the USA in two adjacent years should be very similar and, in fact, the two double rank plots were coincident if the first 30 points were not considered. It is worth stressing again that eliminating some noisy points of highly cited articles in double rank plots does not substantially decrease the accuracy of fitting because the number of data points that can be fitted to a straight line is very high.

Divergent straight lines of the log–log double rank plots corresponding to different countries indicate that a single index for research assessment is not possible without defining the citation level that is required. As a demonstrative exercise of assessment we can estimate the likelihood that a country publishes the most cited paper, ranking in the world's first position. This likelihood corresponds to the value of N at the intersection of the fitted straight line and the vertical axis in the log–log double rank plot (this is shown in Fig. 3) and, more accurately, by fitting a power law to the data before they were log transformed. In "plant sciences," these likelihoods amounted to 0.32 for the EU and 0.34 for the USA (omitting the first 20

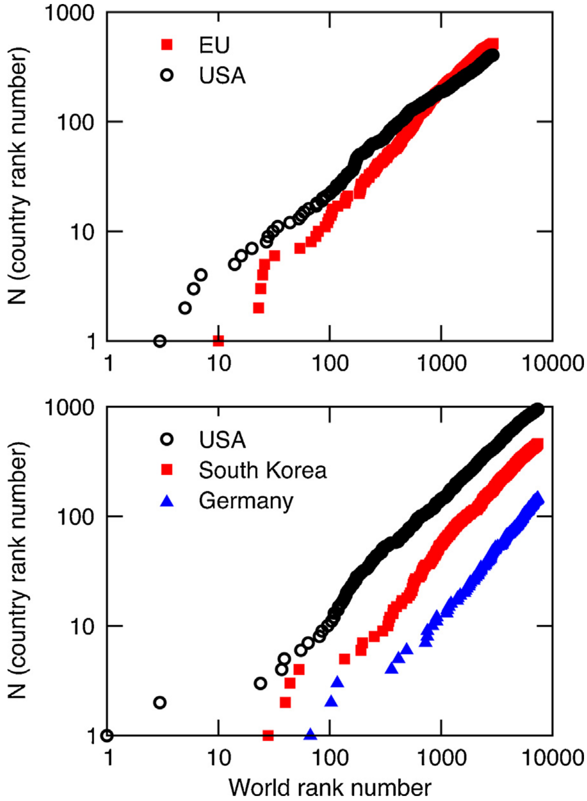

Fig. 4. Double-rank plot for domestic publications from the USA and EU in 2012 in 'photovoltaic cells' (upper panel), and from the USA, South Korea and Germany in graphene (lower panel).

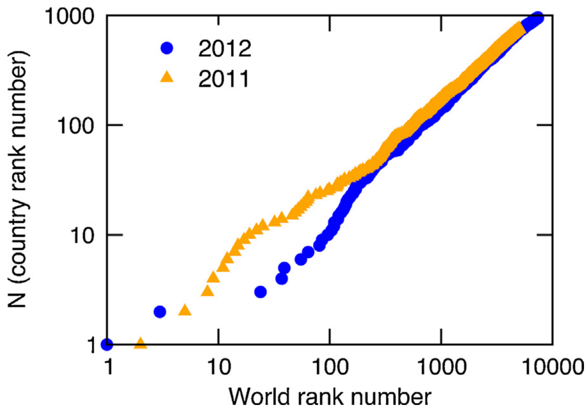

Fig. 5. Double-rank plot of domestic articles from the USA published in 2012 and 2011 in the WoS research topic of 'graphene'.

points). In "physiology" the likelihoods amounted to 0.38 for the EU and 0.33 for the USA (omitting the first 20 points). The likelihoods of publishing the most cited article in the world in plant sciences for Brazil, Germany and Spain were: $9{\cdot}10^{-5}$, 0.06 (shown in Fig. 3), and 0.007, respectively (omitting the first 30 points).

2.4. Applicability of the double rank method

The finding that log–log double rank plots are straight lines for most of the highly and lowly cited publications indicates that the non-transformed data probably fit a power law. It also strongly suggests that research systems are complex systems that behave as single operating units in which performance can be analytically described and calculated at any citation level. Because in most cases the double rank plots are not parallel lines, the research performance ratio between two or more countries or institutions depends on the citation level at which it is estimated. This has been previously described for the upper tail (Rodríguez-Navarro, 2016). If the citation level of the most highly cited paper of the year is selected the likelihood of publishing this paper would correspond to the value of N at the intersection of the fitted straight line and the vertical axis in the log–log double rank plot (Figs. 1–5). However, this is only an example; specific indices at particular citation levels should be defined by the evaluating institutions.

Log-log double rank plots that followed straight lines were found in quiescent (Figs. 1–3) and actively progressing (Fig. 4) research fields, and in countries of different competitiveness (Figs. 3 and 4). Although the first few data points showed notable statistical noise, and a certain number of publications in the upper fragment of the plot showed an appreciable deviation, the number of points that could be fitted to a straight line was very high, and included many publications that were not highly cited. This discussion is valid for a power law function in the original data, as straight lines found in the log-transformed data are mathematically equivalent to a power law in the original, non-transformed data.

To estimate the likelihood of publishing a very highly cited paper using only the upper tail and the power law method, it was necessary to fix a certain level of citations for each year and each field of research (Rodríguez-Navarro, 2016). Therefore, comparisons and statistical averages of the results of several years and fields present some difficulties. In the double rank analysis the approach is different because it allows to estimate the likelihood of a certain position among all world publications (e.g., the first). Therefore, for any year and in any discipline the results are comparable.

In principle, the double rank analysis excludes the requirement of normalization for different citation practices (Glänzel & Moed, 2013; Ioannidis, Boyack, & Wouters, 2016; Ruiz-Castillo & Waltman, 2015), because the research field or topic is the same for both the world and the country or institution publications that are being studied. Although field normalization is a complex problem with no perfect solution (Ruiz-Castillo & Waltman, 2015 and references therein), the different levels of citations of different topics in the same research field are expected to have a more comprehensive treatment in the double rank analysis than in the share of highly cited publications. For example, the share of the top 1% of cited publications in "materials science" would reflect the research on graphene and other nanomaterials, but steel, whose publications are not cited as often as those for the other materials, would be almost excluded from the selected publications and, consequently, steel would remain almost non-existent in the evaluation. The double rank analysis is more likely to provide a reasonably solution to this issue, because it uses most of the papers in a certain research field and not only the publications in the high-citation tail.

To give a proper statistical significance the double rank analysis described here, each publication, either domestic or multinational, should be counted only once. Under this restriction, further research is needed to determine the counting method of multinational publications. In principle, fractional or multiplicative methods (Perianes-Rodriguez & Ruiz-Castillo, 2015; Waltman & van-Eck, 2015) cannot be directly applied to the double rank analysis. Independently of the treatment given to multinational publications, it is worth noting, however, that fractional or multiplicative counting are mathematical solutions for a non-mathematical problem (Rodríguez-Navarro, 2012; Zanotto, Haeffner, & Guimaraes, 2016). Thus, collaborations between countries with high and low levels of research performance increase the citation level of the low-level country more than collaborations between two low-level countries; despite this, fractional or multiplicative counting are applied identically in the two cases.

3. Mathematical modeling

The empirical findings described in the previous section suggested that there is a mathematical law governing citation distributions and the structure of the double rank plots. The existence of this mathematical law indicates that the citations received by the publications of countries and institutions behave as a complex system (Section 2.4). In order to study the properties of this system, we carried out a mathematical analysis by means of simulating citation counts. This simulation approach was also aimed to generalize the empirical findings eliminating fluctuations, such as those shown in Fig. 5.

3.1. Citation simulated lognormal series

For the numerical study of the double rank, we generated random numbers simulating the distribution of citation. As several studies suggest the distribution of citations obey a lognormal distribution (Evans, Hopkins, & Kaube, 2012; Redner, 2005; Radicchi, Fortunato, & Castellano, 2008; Stringer, Sales-Pardo, & Amaral, 2010; Thelwall & Wilson, 2014a, 2014b). It is worth noting that this lognormal distribution applies to the total number of publications while the power law mentioned above (Section 1) applies only to the upper tail. The procedure used here to generate lognormal random numbers follows the strategy described in(Press, Flannery, Teulosky, & Vetterling, 1989, chapter 7). In short, we used the Box-Muller technique to generate random numbers with a normal distribution (with zero average and unit variance). These numbers were then transformed to have an average μ and deviation σ, then we took their exponentials to get the desired lognormal distribution. Once all the random numbers had been generated, they were sorted in decreasing order. In order to avoid the intrinsic fluctuations inherent to the generation of random numbers, we generated 10,000 lognormal distributions for each parameter set, and averaged them to obtain the final data points (Thelwall, 2016). These lists were used for the double rank analysis described later.

The simulation of a double rank plot requires two series of simulated citations one for the world and another for the country or institution. For this purpose, some discussion is required about the simulation of the world distribution of citations. It can be considered that the world distribution of citations arises from the assembly, or combination, of many lognormal distributions of citations that correspond to the publications of many individual research institutions. Therefore, the world production could be generated by merging many lognormal distributions—there might be more than 1000 research institutions that significantly contribute to the world production. Such an assembly would allow us to construct the double rank plot, as every paper of a simulated institution would also be in the simulated world series. While this approach is probably the best, it requires a very large number of lognormal distributions to be combined. We therefore took a different approach. We generated a large series of data points which simulated the world publication list, and small series, which simulated the publication list of countries/research institution, each one with its own set of parameters μ and σ. Then we combine the large series with individual small series of data points ($\leq$0.3% of the number of data points of the large series). This combination of the large and small series did not significantly affect to the μ and σ values of the large series (they changed by less than 0.2%). The number of data points of the simulated World Series was fixed at 150,000 and the number for the institution series varied from 500 to 100.

Next, the continuous lognormal distributions were discretized to create a discrete lognormal distribution of citations. In order to avoid zero citations, continuous data points lower than 0.5 were omitted. After these operations the μ and σ values of the resulting data points did not differ appreciably from the original μ and σ values. A more complex issue was to decide whether to combine the series before or after discretizing the continuous series. If they were combined after discretizing, the double rank plots progressed in steps instead of in a continuous way, which might affect the fitting of the double rank plot. These discontinuities occurred because without an additional criterion, after ordering the combined data points of two series, those that have the same value in the two series cannot be intercalated. For example, in our simulation, the data points amounting to 10 simulated citations were 4995 in the world's series and 18 in series s1; after combining and ordering, the 18 points of series s1 either precede or go after the 4995 data points of the world series. This did not occur with real citations because in this case the publication date was an additional criterion for ordering the data points. Therefore, the best options were to combine the data points of the series before discretization, or to combine the continuous data points rounded to three decimals, which was the solution that we used. Obviously, the double rank plot implies discrete data.

3.2. Parameters of generated lognormal series

The properties of double rank plots obtained from combinations of data points that follow lognormal distributions are not universal but depend on the parameters of the combined series. Therefore, we simulated lognormal distributions with parameters that were similar to those found in actual citation counts. Although the μ and σ values of lognormal citation distributions have been reported (Evans et al., 2012; Golosovsky & Solomon, 2012) we performed our own calculations using a two- or three-year window (Section 2.1) in countries and institutions that represent the highest level of scientific performance and a reasonable minimum, but excluding cases showing very low scientific performance. The values of μ and σ were obtained by using a maximum likelihood method (Pawitan, 2013). An alternative estimation of these parameters comes from a Levenberg-Marquardt Algorithm (Press et al., 1989, chapter 14), used to solve nonlinear least squares problems. The parameters fitted by both methods are equal within the statistical error, particularly for the parameter μ (where they are equal to within a few percentage points), although σ shows slightly higher variability. The maximum values of μ and σ were calculated for the USA and the most competitive institutions in the USA. For the minimum values of μ and σ we selected Brazil and Brazilian institutions because Brazil is still developing an efficient research system.

The world series that we describe in this section simulates the research area of chemistry (SU = chemistry in the Advanced Search feature of the Web of Sciences; see parameters in Table 1), counting citations in a two-year window. We selected this research area because the μ of the distribution of the real data is intermediate between the quiescent and hot areas studied in the first part of this article (Section 2).

We also determined the μ and σ values of citation distributions in hot topics and quiescent scientific areas for the world and for very competitive and much less competitive research institutions. In this search (Table 1; other characteristics to be described elsewhere) we found that differences between research areas were notable (e.g., world μ = 2.5 and 1.4 for Li-batteries and plant sciences, respectively), and that the differences were even higher between the most competitive institution in the most cited area and institutions in countries currently developing a research system in the least cited area (e.g., μ = 3.1 for MIT in lithium batteries versus μ = 1.0 for the Universidade de Sao Paulo in plant sciences). However, in the same research area the differences between countries and institutions were smaller (e.g., in chemistry μ varied from 2.4 to 1.3). The values of μ and σ were highly correlated (Pearson correlation coefficient = 0.85, $p < 10^{-6}$, n = 36) and on average μ was 1.8 times higher than σ. In heterogeneous areas such as chemistry this ratio was higher and slightly more variable (from μ = 2.4 and σ = 1.1 to μ = 1.3 and σ = 0.9) than in homogenous areas such as plant sciences (μ = 1.7 and σ = 1.0 versus μ = 1.0 and σ = 0.8).

As a result of this preliminary study we selected the following μ and σ values: for the world series we took μ = 1.7 and σ = 1.0, and for the institutions we took μ ranging from 2.4, which is the figure for MIT, to 1.5, as well as some intermediate parameters (Table 1 summarizes the parameters for these series).

3.3. Simulation results

Attending to these considerations, we generated a world series and eight country/institution series with the parameters recorded in Table 2. To further validate our simulations we compared our data with real citation distributions. For this purpose we counted the number of simulated articles and the corresponding citations in three broad classes as described

Table 1
Calculated parameters of lognormal distributions in different scientific fields. "Articles" published in 2012.

Research field or topic	Research actor	Counting window	N	μ	σ
Plant Sciences	World	2014–2015	1750	1.37	0.96
Plant Sciences	University Wageningen	2014–2015	190	1.69	0.97
Plant Sciences	EU, domestic	2014–2015	4434	1.50	0.91
Plant Sciences	USA, domestic	2014–2015	2321	1.43	0.98
Plant Sciences	Spain, domestic	2014–2015	643	1.38	0.85
Plant Sciences	Germany, domestic	2014–2015	679	1.38	0.92
Plant Sciences	South Korea, domestic	2014–2015	415	1.31	0.88
Plant Sciences	Brazil, domestic	2014–2015	971	0.97	0.78
Physiology	World	2014–2015	9829	1.48	0.86
Physiology	USA, domestic	2014–2015	2601	1.59	0.82
Physiology	EU, domestic	2014–2015	2351	1.48	0.87
Graphene	World	2014–2015	7901	2.29	1.20
Graphene	Singapore, domestic	2014–2015	194	2.54	1.21
Graphene	USA, domestic	2014–2015	1002	2.45	1.16
Graphene	Germany, domestic	2014–2015	152	2.29	1.12
Graphene	South Korea, domestic	2014–2015	454	2.28	1.12
Li-batteries	World	2014–2016	3902	2.50	1.19
Li-batteries	USA, domestic	2014–2016	516	2.85	1.13
Li-batteries	Germany, domestic	2014–2016	120	2.32	1.02
Li-batteries	South Korea, domestic	2014–2016	330	2.23	1.25
Li-batteries	EU, domestic	2014–2016	404	2.10	1.11
Photovoltaic cells	World	2014–2016	3202	2.19	1.14
Photovoltaic cells	USA	2014–2016	424	2.51	1.21
Photovoltaic cells	Germany	2014–2016	106	2.26	0.94
Photovoltaic cells	South Korea	2014–2016	351	1.75	1.07
Chemistry	World	2015–2016	151163	1.74	0.91
Chemistry	MIT	2015–2016	665	2.36	1.13
Chemistry	Chinese Acad. Sci.	2015–2016	6910	1.86	1.12
Chemistry	CNRS	2015–2016	4527	1.65	0.98
Chemistry	Universi. de Sao Paulo	2015–2016	618	1.32	0.89
Engineering	World	2015–2016	148075	1.49	0.99
Engineering	MIT	2015–2016	603	1.78	1.08
Engineering	KU Leuven	2015–2016	504	1.59	1.10
Engineering	KAIST (South Korea)	2015–2016	640	1.35	0.98

Table 2
Parameters of the lognormal distributions that were combined to perform the double rank analysis. The world series was combined with de s1-s8 series individually. L1 indicates the likelihood of publishing the publication ranking in the world first position: the point at which the double rank power-law plot cuts the vertical axis (see Fig. 6).

Series	lognormal			Double rank
	N	μ	σ	L1
World	150.000	1.7	1.0	–
s1	500	2.4	1.1	0.36
s2	500	2.1	1.1	0.13
s3	200	2.1	1.1	0.061
s4	500	1.7	1.0	0.0035
s5	300	1.7	1.0	0.0022
s6	100	1.7	1.0	0.00081
s7	500	1.5	0.9	$9.5 \cdot 10^{-5}$
s8	200	1.5	0.9	$4.5 \cdot 10^{-5}$

previously (Ruiz-Castillo & Waltman, 2015): "(i) articles with low impact, or a number of citations less than or equal to m_1; (ii) articles with a fair impact, or citations greater than m_1 and less than or equal to m_2; (iii) articles with a remarkable or outstanding citation impact above m_2. In this classification system: "m_1 = mean citation of the cluster citation distribution, and m_2 = mean citation for articles with a number of citations above m_1." The results recorded in Table 3 show that the percentages of simulated articles and citations in each class of our simulations—approximately 70/20/10 for articles and 30/30/40 for citations—are remarkable coincident with previous results (Albarrán, Perianes-Rodríguez, & Ruiz-Castillo, 2015; Li, Radicchi, Castellano, & Ruiz-Castillo, 2013; Ruiz-Castillo & Waltman, 2015). This coincidence demonstrates that our simulations reproduce the citation distributions across scientific fields, which show high similarity between them (Ruiz-Castillo & Waltman, 2015 and references therein).

Once citations lists were created, we proceeded to construct double rank plots from simulated series. Fig. 6 shows the simulated double rank plots for series s1, s2, s5, and s7 described in Table 2. As can be expected from the log–log transformed data shown in this figure, large stretches of the double rank plots engendered by combining each small series with the simulated world series could be fitted to power laws, in agreement with real data. In all the series, however, the first 5–10

Table 3
The skewness of simulated data. Distribution of simulated papers and citations in three categories attending to the number of simulated citations (Table 2).

Series	Percentage of articles in category			Percentage of citations in category		
	1	2	3	1	2	3
World	70.7	19.9	9.4	32.7	29.9	37.4
s1	71.0	20.6	8.4	29.5	31.1	39.4
s2	72.1	19.6	8.3	30.3	30.5	39.2
s3	73.6	20.2	6.2	30.3	30.7	38.9
s4	70.5	19.8	9.7	32.5	30.0	37.5
s5	71.1	19.7	9.2	32.8	29.9	37.2
s6	70.4	19.4	10.2	33.1	28.6	38.3
s7	65.9	23.1	11.0	31.3	34.5	34.1
s8	66.0	22.8	11.1	31.4	31.8	36.8

Category 1 = simulated publications with a low number of citations, below m_1. Category 2 = simulated publications with a fair number of citations, above m_1 and below m_2. Category 3 = simulated publications with a large number of citations, above m_2, where m_1 = mean number of citations of simulated publications, m_2 = mean number of citation of simulated publications with a number of citations above m_1 (Ruiz-Castillo & Waltman, 2015).

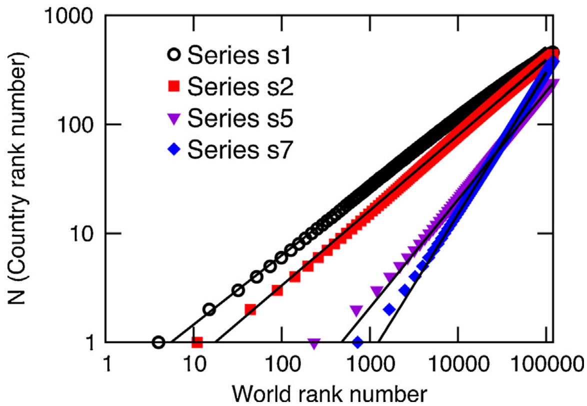

Fig. 6. Double-rank plot of simulated data for series s1, s2, s5, and s7 with parameters cited in Table 2.

points in the lower part of the plots clearly deviated from power laws. Because fluctuations are not present in our simulations, such deviations appear to be intrinsic. The middle part obeyed a power law, with a well-defined exponent. In the upper part of the plots there was a certain deviation from power laws, but the number of data points affected was low except in those series in which the μ of the series was very different from the μ of the simulated world series (e.g., s1 and s2; Table 2). Even in these cases the number of points that could be fitted to a power law was a large proportion of the data points of the small series. The exponent of the power law was quite robust. For instance the exponent obtained by fitting the worst case, s1 in Fig. 6 changes from b = 0.599, when fitting the whole curve, to b = 0.626, when fitting the central 150 points (a difference of less than 4%). For all the other cases, like s2 and s7 when μ and σ were closer to the world parameters, the fitting was much better. Therefore we conclude that the double rank plot can be fitted to a power law.

In an example of our study we used the double rank plot to estimate the likelihood that a certain institution will publish the most highly cited paper of the year (the paper that ranks number one in the ordinate axis of the double rank plot). For this purpose, the deviation of the lower data points of the double rank plot from a power law (Fig. 6 and 7) raised the question of whether these data points could be ignored. To answer this question we used series s4, s5, and s6. The μ and σ values of these series were equal to those of the simulated world and combined series. The results of the double rank plot are presented in Fig. 7. As can be seen, they follow parallel straight lines, except in the lower part of the citation counts. Because of the coincidence of the parameters, the exponent of the power law is simply equal to 1, and the probability that the highest number of the combined series belong to any of these series is equal to the ratio between the number of data points of each small series and the number of the large series: 3.3 10^{-3}, 2 10^{-3}, and 6.6 10^{-4} for s4, s5, and s6, respectively, which are similar to the likelihoods calculated from the fitted power laws: 3.5 10^{-3}, 2.2 10^{-3}, and 8.1 10^{-4} (Table 2), which were obtained by ignoring the lower points that deviate from the power law. Therefore, the lower points that deviate from the power law should be omitted in double rank calculations of the likelihood of publishing the paper ranking first among world publications.

3.4. Mathematical validation of the double rank analysis

Limiting the simulations to the mathematical parameters prevailing for real lognormal distributions of citations (Table 2), the simulated articles and corresponding citations (Table 3) reproduce the skewed characteristics of the citation distributions

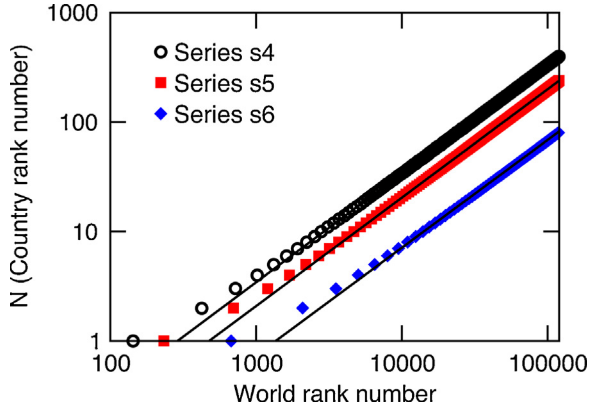

Fig. 7. Double-rank plot of simulated data for series s4, s5, and s6, with $N = 500$, 300 and 100, respectively, but otherwise equal values of μ and σ.

that have been shown for real scientific articles (Albarrán et al., 2015; Li et al., 2013; Ruiz-Castillo & Waltman, 2015). In these simulations, our study reveals that large parts of the log–log double rank plots follow straight lines, in excellent agreement with real citation data (Fig. 1–5). Some of the first points showed visible deviations in all cases and some of the last data points also deviated when the μ value of the simulated institution was much higher than that corresponding to the world simulation (Fig. 6). However, even in the cases with the highest differences in μ values, a large number of data points could be used to fit the power law. These coincidences between the empirical and simulation analyses validate the general use of the double rank analysis for research assessment.

4. Conclusions and perspectives

A research assessment of countries and institutions based on their capacity to publish very highly cited papers seems to be a reasonable approach to solving the complex problem that this type of assessment creates. The major advantage of this method is that it can be validated in terms of the number of Nobel Prize awards (Rodríguez-Navarro, 2011a, 2016), but it presents a great limitation because it cannot be widely applied. As already described, the limitation arises because the proportion of publications that can be fitted to the power law is very low in many cases.

To solve this problem we describe here a double rank analysis for the publications of countries and institutions. However, research evaluation is a complex problem and extensive application of the double rank method for evaluation purposes requires the previous solution of several technical and mathematical details that are not sufficiently treated in this study. We are working to fill this gap.

We have shown how the double rank method can be applied to rank the research of institutions or countries by performing a comparison of the citations of papers produced by the institution with those of the rest of the world. Exactly the same method can be applied to rank institutions or universities within a country, by replacing the comparison with the world by a comparison with the country, allowing a national ranking to be created. However, it would be unsuitable to use this method as a personal indicator, as the number of papers needs to be high. In that respect, we remind readers that in this paper we use as few as 100 simulated publications, but they are averaged over 10,000 realizations, and that averaging eliminates the intrinsic noise of citations. It is likely that for a researcher with, say, 100 papers, the double rank plot will be very noisy and could give a wrong understanding of the importance of the research produced by that individual. Furthermore, it must be pointed out that the number of citations correlates with the importance of a publication but does not measure its importance.

Moreover, the method can be applied to any other social or scientific assessment where a ranking exists. We can think, for example, of the distribution of incomes of individuals in a country and a comparison with the world incomes, or the success of a health care system if a "success indicator" is defined, or company activity ranked by its annual turnover. These are only few examples of applicability of the double rank method. The function of the double rank plot must certainly be studied in each case.

Author contributions

Alonso Rodríguez-Navarro: Conceived and designed the analysis, Performed the analysis, Wrote the paper.

Ricardo Brito: Conceived and designed the analysis, Contributed data or analysis tools, Performed the analysis, Wrote the paper.

Acknowledgements

We thank Javier Ruiz-Castillo for his helpful comments on an earlier version of this article. This work was supported by the Spanish Ministerio de Economía y Competitividad, grant number FIS2014-52486-R.

References

Albarrán, P., Perianes-Rodríguez, A., & Ruiz-Castillo, J. (2015). Differences in citation impact across countries. *Journal of the Association for Information Science and Technology*, *66*, 512–525.
Albarrán, P., Herrero, C., Ruiz-Castillo, J., & Villar, A. (2017). The Herrero-Villar approach to citation impact. *Journal of Informetrics*, *11*, 625–640.
Andras, P. (2011). Research: Metrics, quality, and management implications. *Research Evaluation*, *20*, 90–106.
Bonaccorsi, A. (2007). Explaining poor performance of European science: Institutions versus policies. *Science and Public Policy*, *34*, 303–316.
Bornmann, L., & Mutz, R. (2011). Further steps towards an ideal method of measuring citation performance: The avoidance of citation (ratio) averages in field-normalization. *Journal of Informetrics*, *5*, 228–230.
Bornmann, L., Leydesdorff, L., & Mutz, R. (2013). The use of percentile rank classes in the analysis of bibliometric data: Opportunities and limits. *Journal of Informetrics*, *7*, 158–165.
Brzezinski, M. (2015). Power laws in citation distribution: Evidence from Scopus. *Scientometrics*, *103*, 213–228.
Delanghe, H., Sloan, B., & Muldur, U. (2011). European research policy and bibliometric indicators. *Scientometrics*, *87*, 389–398.
Dosi, G., Llerena, P., & Labini, M. S. (2006). The relationships between science, technologies and their industrial exploitation: An illustration through the myths and realities of the so-called 'European Paradox'. *Research Policy*, *35*, 1450–1464.
Evans, T. S., Hopkins, N., & Kaube, B. S. (2012). Universality of performance indicators based on citation and reference counts. *Scientometrics*, *93*, 473–495.
Garfield, E., & Welljams-Dorof, A. (1992). Citation data: Their use as quantitative indicators for science and technology evaluations and policy-making. *Science and Public Policy*, *19*, 321–327.
Glänzel, W., & Moed, H. F. (2013). Opinion paper: Thoughts and facts on bibliometric indicators. *Scientometrics*, *96*, 381–394.
Godin, B. (2006). On the origins of bibliometrics. *Scientometrics*, *68*, 109–133.
Golosovsky, M., & Solomon, S. (2012). Runaway events dominate the heavy tail of citation distribution. *The European Physical Journal Special Topics*, *205*, 303–311.
Harnad, S. (2009). Open access scientometrics and the UK research assessment exercise. *Scientometrics*, *79*, 147–156.
Herranz, N., & Ruiz-Castillo, J. (2013). The end of the european paradox. *Scientometrics*, *95*, 453–464.
Ioannidis, J. P. A., Boyack, K., & Wouters, P. F. (2016). Citation metrics: A primer on how (not) to normalize. *PLoS Biology*, *14*(9), e1002542.
Katz, J. S. (2016). What is a complex innovation system? *PUBLIC LIBRARY OF SCIENCE*, *11*(6), e0156150.
Leydesdorff, L. (2005). Evaluation of research and evolution of science indicators. *Current Science*, *89*, 1510–1517.
Li, Y., Radicchi, F., Castellano, C., & Ruiz-Castillo, J. (2013). Quantitative evaluation of alternative field normalization procedures. *Journal of Informetrics*, *7*, 746–755.
Martin, B. R., & Irvine, J. (1983). Assessing basic research. Some partial indicators of scientiic progress in ratio astronomy. *Research Policy*, *12*, 61–90.
Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Conterporary Physics*, *46*, 323–351.
Pawitan, Y. (2013). *All likelihood: Statistical modelling and interference using likelihood*. Oxford United Kingdom: Oxford University Press.
Perianes-Rodriguez, A., & Ruiz-Castillo, J. (2015). Multiplicative versus fractional counting methods for co-authored publications. The case of the 500 universities in the Leiden Ranking. *Journal of Informetrics*, *9*, 974–989.
Press, W. H., Flannery, B. P., Teulosky, S. A., & Vetterling, W. T. (1989). *Numerical recipies, fortran version*. Cambridge United Kingdom: Cambridge University Press.
Price, D. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, *27*, 292–306.
Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 17268–17272.
Redner, S. (2005). Citation statistics from 110 years of physical review. *Physics Today*, *58*, 49–54.
Rodríguez-Navarro, A. (2011a). A simple index for the high-citation tail of citation distribution to quantify research performance in countries and institutions. *PUBLIC LIBRARY OF SCIENCE*, *6*(5), e20510.
Rodríguez-Navarro, A. (2011b). Measuring research excellence: Number of Nobel Prize achievements versus conventional bibliometric indicators. *Journal of Documentation*, *67*, 582–600.
Rodríguez-Navarro, A. (2012). Counting highly cited papers for university research assessment: conceptual and technical issues. *PUBLIC LIBRARY OF SCIENCE*, *7*(10), e47210.
Rodríguez-Navarro, A. (2016). Research assessment based on infrequent achievements: A comparison of the United States and Europe in terms of highly cited papers and Noble Prizes. *Journal of the Association for Information Science and Technology*, *67*, 731–740.
Rodriguez-Navarro, A., & Narin, F. (2017). European paradox or delusion-Are European science and economy outdated? *Science and Public Policy*, http://dx.doi.org/10.1093/scipol/scx021
Ruiz-Castillo, J., & Waltman, L. (2015). Field-normalized citation impact indicators using algorithmically constructed classification systems of science. *Journal of Informetrics*, *9*, 102–117.
Ruiz-Castillo, J. (2012). The evaluation of citation distribution. *SERIEs*, *3*, 291–310.
Stringer, M. J., Sales-Pardo, M., & Amaral, L. A. N. (2010). Statistical validation of a global model for the distribution of the ultimate number of citations accrued by papers published in a scientific journal. *Journal of the American Society for Information Science*, *61*, 1377–1385.
Thelwall, M., & Wilson, P. (2014a). Distributions for cited articles from individual subjects and years. *Journal of Informetrics*, *8*, 824–839.
Thelwall, M., & Wilson, P. (2014b). Regression for citation data: An evaluation of different methods. *Journal of Informetrics*, *8*, 963–971.
Thelwall, M. (2016). The precision of the aritmetic mean, geometric mean and percentiles for citation data: An experimental simulation modelling approach. *Journal of Informetrics*, *10*, 110–123.
Waltman, L., & Schreiber, M. (2013). On the calculation of percentile-based bibliometric indicators. *Journal of the American Society for Information Science and Technology*, *64*, 372–379.
Waltman, L., & van-Eck, N. J. (2015). Field-normalized citation impact indicators and the choice of an appropriate counting mehod. *Journal of Informetrics*, *9*, 872–894.
Zanotto, S. R., Haeffner, C., & Guimaraes, J. A. (2016). Unbalanced international collaboration affects adversely the usefulness of countries' scientific output as well as their technological and social impact. *Scientometrics*, *109*, 1789–1814.
van Noorden, R. (2010). A profusion of measures. *Nature*, *465*, 864–866.

Alonso Rodríguez-Navarro is Emeritus Professor in the Center of Plant Biotechnology and Genomic, Polytechnic University of Madrid.

Ricardo Brito is Full Professor at the Department of Applied Physics, Complutense University of Madrid.