Scientometrics (2012) 91:269–275
DOI 10.1007/s11192-011-0516-5

The quality-quantity-quasity and energy-exergy-entropy exegesis of expected value calculation of citation performance

Gangan Prathap

Received: 10 August 2011 / Published online: 20 September 2011

Abstract Quantitative assessment of information production processes requires the definition of a robust citation performance indicator. This is particularly so where there is a need to introduce a normalization mechanism for correcting for quality across field and disciplines. In this paper, we offer insights from the "thermodynamic" approach in terms of quality, quantity and quasity and energy, exergy and entropy to show how the recently introduced expected value measure can be rationalized and improved. The normalized energy indicator E is proposed as a suitable single number scalar indicator of a scientist's or group's performance (i.e. as a multiplicative product of quality and quantity), when complete bibliometric information is available.

Keywords Bibliometrics · Normalization · Performance Indicators · Crown Indicator · Quality · Quantity · Quasity · Energy · Exergy · Entropy · Expected Value · Citation Performance

Introduction

The holy grail of quantitative assessment of information production processes has been to find a single meaningful indicator from the quantity data (the number of papers P published), or from what we have recently called the quasity term (citations C received) and a suitable quality measure (say, impact, defined as the ratio of citations to publications, i.e. $i = C/P$). The neologism, quasity, needs some explanation. Table 1 draws on some metaphors and analogies from classical mechanics, electrical engineering and bibliometrics, to show that for a bibliometrics-thermodynamics consilience, a quasity term must be introduced so that mathematically and metaphysically, quasity = quantity × quality, and energy (or exergy) = quality × quasity.

G. Prathap (✉)
CSIR National Institute of Science Communication and Information Resources,
New Delhi 110 012, India
e-mail: gp@cmmacs.ernet.in

Table 1 The metaphors or analogies drawing the bibliometrics-thermodynamic consilience

Analogies	Quality	Quantity	Quasity	Energy/Exergy
Mechanical	Velocity	Mass	Momentum	Energy
Electrical	Current	Resistance	Voltage	Power
Scientometric	Impact	Papers	Citations	Energy/Exergy

The Quality-Quantity-Quasity paradigm leads through the "thermodynamic" analogy to the trinity of terms we denote by Energy-Exergy-Entropy (Prathap 2011a, b). In this way, it is possible to describe research performance through a single scalar indicator, the exergy term $X = (C/P)\cdot C$, when complete bibliometric information is not available and only P and C are known for any portfolio of work. The presumption here is that all papers, irrespective of journal or field in which it has appeared are weighted equally.

However, publication and citation practices vary significantly from field to field (or from journal to journal). The comparison of quality indicators across disciplines becomes a very difficult task, and normalization procedures have to be introduced to give a weighted quality term.

The first standard to emerge in evaluative bibliometric practice for the normalization mechanism was the so called crown indicator (originally introduced by the Centre for Science and Technology Studies at Leiden, and hence known as the CWTS approach). The "crown indicator" is a variation of Schubert and Braun's (1986) RCR = MOCR/MECR.

This can be thought of as a "add-divide" approach as the sequence of operations requires citations received by the papers and the citation norms for the papers (depending on the journal/field, document type, and/or vintage) to be separately added and then divided to get what is essentially an arithmetic average even if the underlying distribution is a very complex one. This was challenged by Opthof and Leydesdorff (2010), who proposed an alternative divide-add approach. In response to this, a new crown indicator was introduced by CWTS: the mean normalized citation score (MNCS) (Waltman et al. 2011b). Bornmann and Mutz (2011) summed this up very neatly thus: The previous crown indicator calculated the average citation rate over all papers in a publication set and then field-normalized the citation rate; the new crown indicator field-normalized each paper's citation impact in a paper set before an average value (harmonic average) over the field-normalized impact values was calculated. However, both old and new indicators suffer from the weakness that all the operations are based on arithmetic averages of ratios or ratios of arithmetic averages (Bornmann and Mutz 2011). As citation data is highly skewed, this will not lead to robust measures. Instead, Bornmann and Mutz (2011) extend an earlier idea (Bornmann 2010) to calculate a single number measure for the citation impact that is not based on the arithmetic average but uses reference distributions based on the calculation of percentiles. This allows each paper in an oeuvre to be field-normalized with a matching reference standard (the norm). The reference standard (or norm) for a paper is obtained by collecting all papers published in the same year, with the same document type and belonging to the same field (defined by a discipline-oriented database) and these are then categorized into what are called impact (quality) classes. They use six percentile impact classes: 99th – top 1%, 95th, 90th, 75th, 50th, and < 50th – bottom 50% (following the approach of the National Science Board 2010) (see here also Bornmann, de Moya-Anegón, and Leydesdorff 2010). Through the use of the citation limits that define these classes the paper in question can be assigned to one of the six citation impact (quality) classes and weights are assigned accordingly. This procedure is repeated for all papers in the oeuvre. Bornmann and Mutz (2011) also point out that nothing prevents the use of the actual percentile value for each paper (this then actually serves as the quality weight) instead of the six-point percentile impact class proposed except for the extra effort involved. An expected value (EV) is then proposed but as we shall show later, this is a overall quality proxy and not a proxy for total performance. Leydesdorff et al. (2011) make the same observation.

In this paper, we combine insights from the "thermodynamic" approach in terms of quality, quantity and quasity and energy, exergy and entropy to show how the recently introduced expected value (EV) measure can be rationalized and improved. The normalized energy indicator E is proposed as a robust single number scalar indicator of a scientist's or group's performance (i.e. as a multiplicative product of quality and quantity), when complete bibliometric information is available.

The thermodynamics of normalization

The normalization mechanisms for measuring "quality" can be explained following the treatment of the subject by Waltman et al. (2011a, b) Quality, when defined as an average, where complete bibliometric information is not available, is the ratio of total citations (C) to the number of papers published (P). As publication and citation practices vary significantly from field to field, it is necessary to derive a "weighted" or "normalized" quality value. For a given portfolio of publications, we count the number of citations each paper has received. To obtain the normalized "quality" value for field or journal, we must take notice of the expected number of citations each publication could have received in the journal, or field, to which it belonged. This is done using the average number of citations of all publications of the same document type (i.e., article, letter, or review) published in the same journal, or field, in the same year. The CWTS crown indicator, as originally proposed, is obtained as the ratio of the sum of the actual number of citations of all publications to the sum of the expected number of citations of all publications (i.e., first **add**, then **divide**).

An alternative, and arguably better procedure for normalization, is also possible. One can first take for each publication, the ratio of its actual number of citations and its expected number of citations (first **divide**), and then **add** these ratios and finally divide by the total number of publications to obtain the average of the ratios; hence the **divide-add** procedure. Such an indicator was first introduced by Lundberg (2007), as the item-oriented field-normalized citation score average. Recent advocacy by Opthof and Leydesdorff (2010) in favour of this divide-add normalization strategy has led to a vigorous debate, see Van Raan et al. (2010), Bornmann (2010), Bornmann and Mutz (2011), Gingras and Lariviere (2011), Leydesdorff and Opthof (2010, 2011), Moed (2010), and Spaan (2010). Indicators that rely on this divide-add normalization mechanism are used by various institutes, e.g., Karolinska Institute in Sweden (Rehn and Kronman 2008), Science-Metrix in the US and Canada (e.g., Campbell et al. 2008 p. 12), etc.

Note that the foregoing discussion is only about how the quality of a publication should be quantified given the number of citations it has earned (i.e., the impact dimension), and the need for normalizing this depending on the field or journal in which it has appeared. The expected value (EV) measure finally arrived at is also a proxy for quality. The quantity and quasity (or productivity) dimensions have been addressed only recently (Leydesdorff et al. 2011 and Leydesdorff and Bornmann 2011). Recent work (Prathap 2011a, b) has also indicated that energy-like indicators are robust markers of performance as a multiplicative product of quality and quantity. The purpose of this paper is to show that such energy and exergy indicators can be derived for field or journal normalization as well, using the normalized quality values, and the appropriate quantity values, and this is demonstrated below.

The "normalized" energy and exergy indicators using the expected value (EV) approach

Prathap (2011a) proposed a "thermodynamic" analogy and introduced what are called the energy, exergy and entropy terms associated with a bibliometric sequence. This can be displayed as time series (variation over time), or in event terms (variation as papers are published) and also in the form of phase diagrams (energy–exergy–entropy representations). Exergy is the most meaningful single number scalar indicator of a scientist's performance (i.e. as a multiplicative product of quality and quantity), when only the total number of papers P and the total number of citations C in the portfolio is known, and not the actual bibliometric sequence of citations of all papers, c_k, for $k = 1$ to P. Thus, in moving from a system which has k degrees of freedom to one which has only one degree of freedom (at the oeuvre level), one can expect the loss of information to be reflected as a shortfall between energy and exergy. Exergy, which is an energy like term, is computed from the formula, $X = iC$, where i is a measure of quality, expressed as the ratio of total citations C to total papers published P. Note that as originally proposed, it is an unweighted term, and does not take into account field differences. However, there is no difficulty in bringing in field-normalized or journal-normalized impact terms and we shall demonstrate how this is done through empirical examples below. If complete bibliometric data is available, i.e., the entire sequence c_k, for $k = 1$ to P, is known, then the energy $E = \sum c_k^2$ is a measure of effort expended. The shortfall between E and X can be taken as an entropy measure (Prathap 2011a). These metaphors are useful in drawing the bibliometric-thermodynamic consilience (see Table 1).

We shall use as an empirical basis, data from Bornmann and Mutz (2011). Although a critique of this scheme has since appeared (Leydesdorff, Bornmann, Mutz and Opthof 2011), our purpose here is solely to start with that data and re-interpret Table 1 of Bornmann and Mutz (2011) using the Quality-Quantity-Quasity and Energy-Exergy-Entropy paradigms. Papers from three research groups are categorized into six percentile impact classes (99th – top 1%, 95th, 90th, 75th, 50th, and < 50th – bottom 50%). Weights are assigned in the manner shown (see Table 2): A paper in the 99th percentile class is valued at six times that in the < 50 percentile class, etc. Unlike, Bornmann and Mutz (2011), we do not use frequency or probability arguments but instead define the quasity term as quasity = quantity × quality. The number of papers in each class is taken as the quantity term and is weighted by the corresponding quality value to get the quasity term. As we shall see, for each class, this is an "expected value" that takes into account the effects of size and quality. The total quasity divided by the total quantity (both are presumed to have additive properties akin to conservation of mass and momentum respectively) will then yield the same expected value (EV) that was proposed in Bornmann and Mutz (2011). Note that the EV thus calculated is a quality value and can take a range from 1.0 (all papers belong to the bottom 50%) to 6.0 (all papers belong to the top 1%). Thus a research group that has all its 10 papers in the top 1% and another that has all its 100 papers in the same percentile class have the same EV. To distinguish the two groups for total performance in a way that takes both quality and quantity into account, one must use either

Table 2 Data taken from the example from Bornmann and Mutz (2011) but now re-interpreted using the Quality-Quantity-Quasity and Energy-Exergy-Entropy paradigms

Percentile impact class	Norm	Research group 1				Research group 2				Research group 3			
	Numbering of classes (X)			"EV"				"EV"				"EV"	
	Quality	Quality	Quantity	Quasity	Energy	Quality	Quantity	Quasity	Energy	Quality	Quantity	Quasity	Energy
<50	1	1	43	43	43	1	17	17	17	1	68	68	68
50th	2	2	22	44	88	2	28	56	112	2	43	86	172
75th	3	3	33	99	297	3	19	57	171	3	27	81	243
90th	4	4	21	84	336	4	39	156	624	4	36	144	576
95th	5	5	23	115	575	5	22	110	550	5	40	200	1000
99th	6	6	14	84	504	6	45	270	1620	6	36	216	1296
Total			156	469	1843		170	666	3094		250	795	3355
Oeuvre Statistics		EV	Quantity	Quasity	Exergy	EV	Quantity	Quasity	Exergy	EV	Quantity	Quasity	Exergy
		3.01	156	469	1410.01	3.92	170	666	2609.15	3.18	250	795	2528.10

the total quasity score, or sum up the total energies to give the energy total *E*. Note that this is now based on the normalization using percentile classes. If an individual or research group has 100 papers in class 1 and another has the same number in class 6, then their respective *E* will differ by a factor of 36!.

Conclusions

The quantitative assessment of information production processes for performance requires that the correct multiplicative product of quality and quantity be made. Recent "thermodynamic" insights (Prathap 2011a, b) have led to the understanding that this can be done by introducing what are called the energy and exergy terms. However, due to the fact that citation practices vary significantly from journal to journal or from field to field, it is not a simple matter to come up with a unique definition of normalized quality. Bornmann and Mutz (2011) have proposed a field-normalization procedure against a matching reference standard (the norm) using weights for quality that vary according to six pre-defined percentile impact classes. In this paper, we offer insights from the "thermodynamic" approach in terms of Quality-Quantity-Quasity and Energy-Exergy-Entropy to show how these results can be re-interpreted. In general cases, the normalized energy indicator *E* is proposed as a single number scalar indicator of a scientist's or research groups performance (i.e. as a multiplicative product of quality and quantity), when complete bibliometric information is available. In this regard, it is meaningful to mention here that more recently, Leydesdorff and Bornmann (2011) have further developed their ideas to yield an alternative scalar sum: the Integrated Impact Indicator (I3). This is different from Bornmann & Mutz's EV (which is a mean percentile rank, and therefore a quality proxy as shown in this paper). It can also be seen that the indicator I3 (which is now based on an integration strategy and not the mean percentile rank) and the Energy (or Exergy at oeuvre level) assume that total impact is a scalar sum.

References

Bornmann, L. (2010). Towards an ideal method of measuring research performance: Some comments to the Opthof and Leydesdorff (2010) paper. *Journal of Informetrics, 4*(3), 441–443.

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*(1), 228–230.

Bornmann, L., de Moya-Anegón, F., & Leydesdorff, L. (2010). Do scientific advancements lean on the shoulders of giants? A bibliometric investigation of the Ortega hypothesis. *PLoS One, 5*(10), e11344.

Campbell, D., Archambault, E., & Côte', G. (2008). *Benchmarking of Canadian Genomics—1996–2007*. http://www.science-metrix.com/pdf/SM_Benchmarking_Genomics_Canada.pdf.

Gingras, Y., & Lariviere, V. (2011). There are neither ''king'' nor ''crown'' in scientometrics: Comments on a supposed ''alternative'' method of normalization. *Journal of Informetrics, 5*(1), 226–227.

Leydesdorff, L., & Bornmann, L. (2011). Integrated impact indicators (I3) compared with Impact Factors (IFs): An alternative design with policy implications. *Journal of the American Society for Information Science and Technology*. doi: 10.1002/asi.21609.

Leydesdorff, L., & Opthof, T. (2010). Normalization at the field level: Fractional counting of citations. *Journal of Informetrics, 4*(4), 644–646.

Leydesdorff, L., & Opthof, T. (2011). Remaining problems with the ''new crown indicator'' (MNCS) of the CWTS. *Journal of Informetrics, 5*(1), 224–225.

Leydesdorff, L., Bornmann, L., Mutz, R., & Opthof, T. (2011). Turning the tables in citation analysis one more time: Principles for comparing sets of documents. *Journal of the American Society for Information Science and Technology, 62*(7), 1370–1381.

Lundberg, J. (2007). Lifting the crown—citation z-score. *Journal of Informetrics, 1*(2), 145–154.
Moed, H. F. (2010). CWTS crown indicator measures citation impact of a research group's publication oeuvre. *Journal of Informetrics, 4*(3), 436–438.
National Science Board. (2010). Science and engineering indicators 2010, appendix tables. Arlington, VA, USA: National Science Foundation (NSB 10-01).
Opthof, T., & Leydesdorff, L. (2010). Caveats for the journal and field normalizations in the CWTS (''Leiden'') evaluations of research performance. *Journal of Informetrics, 4*(3), 423–430.
Prathap, G. (2011a). The Energy–Exergy–Entropy (or EEE) sequences in bibliometric assessment. *Scientometrics, 87*, 515–524.
Prathap, G. (2011b). Quasity, when quantity has a quality all of its own—toward a theory of performance. *Scientometrics, 88*, 555–562.
Rehn, C., & Kronman, U. (2008). *Bibliometric handbook for Karolinska Institutet*. http://ki.se/content/1/c6/01/79/31/bibliometric_handbook_karolinska_institutet_v_1.05.pdf.
Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics, 9*(5), 281–291.
Spaan, J. A. E. (2010). The danger of pseudoscience in Informetrics. *Journal of Informetrics, 4*(3), 439–440.
Van Raan, A. F. J., Van Leeuwen, T. N., Visser, M. S., Van Eck, N. J., & Waltman, L. (2010). Rivals for the crown: Reply to Opthof and Leydesdorff. *Journal of Informetrics, 4*(3), 431–435.
Waltman, L., Van Eck, N. J., Van Leeuwen, T. N., Visser, M. S., & Van Raan, A. F. J. (2011a). Towards a new crown indicator: an empirical analysis. *Scientometrics*, doi: 10.1007/s11192-011-0354-5.
Waltman, L., Van Eck, N. J., Van Leeuwen, T. N., Visser, M. S., & Van Raan, A. F. J. (2011b). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics, 5*(1), 37–47.