www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2023; SP-12(11): 546-552

www.thepharmajournal.com
Received: 09-10-2023
Accepted: 13-11-2023

K Sakunthala Devi
Professor & Head, Department of AGB, College of Veterinary Science, Sri Venkateswara Veterinary University, Tirupati, Andhra Pradesh, India

Kutty Kumar
Assistant Professor, College of Veterinary Science, Sri Venkateswara Veterinary University, Tirupati, Andhra Pradesh, India

S Vani
Assistant Professor, Department of AGB, College of Veterinary Science, Sri Venkateswara Veterinary University, Tirupati, Andhra Pradesh, India

Corresponding Author:
K Sakunthala Devi
Professor & Head, Department of AGB, College of Veterinary Science, Sri Venkateswara Veterinary University, Tirupati, Andhra Pradesh, India

Author productivity on application of random regression models in animal breeding research through Lotka's law

K Sakunthala Devi, Kutty Kumar and S Vani

Abstract
The purpose of this paper is to know whether the author productivity pattern on usage of random regression models in animal breeding research adheres to Lotka's inverse square law of scientific productivity. Since the law was introduced, it has been tested in various fields of knowledge and results were varied. The data on 236 number of animal breeding research publications where random regression models used were downloaded from the PubMed database to know about the authorship productivity pattern and citations. Data were analyzed by using bibliometric indicators like publication and citation growth, co-authorship pattern, prolific authors and sankey diagram authors, title of research including country where research was carried out. Lotka's inverse square law was applied to assess authors' productivity pattern of animal breeding research that used random regression models and further Kolmogorov-Smirnov (K-S) goodness-of-fit test was applied for testing of observed and expected author productivity in the data. Inferences drawn for the set objectives in this study on authorship pattern, collaboration trend and authors' productivity pattern revealed that, the author "Misztal Ignacy" was found as most prolific author on usage of RRM in animal breeding research. The Chi-square test expected value of Lotka's law varied significantly from the observed value and K-S goodness-of-fit test revealed that this study does not adhere to Lotka's inverse square law of scientific productivity.

Keywords: RRM, animal breeding research, author productivity, Lotka's law

Introduction

Henderson, (1982) [6], Laird and Ware, (1982) [10] developed random regression models (RRM) for analysing test-day milk yield data in dairy cattle breeding. RRM are extension of repeatability models that allow researchers to analyse genetic variability over time without increasing model complexity (Lidauer *et al.*, 2003) [11] and have become common for genetic evaluation of longitudinal traits (Paneru *et al.*, 2021) [16]. Various researchers opined that compared to other statistical methods random regression models fit random genetic and environmental effects across time, which improves predicted breeding values in animal breeding research and are more parsimonious, computationally efficient and faster to reach convergence than other traditional models (Oliveira *et al.*, 2019) [15].

Indeed, areas of animal breeding research that have already utilised RRM includes analysing of various production traits, genotype by environmental interactions *etc.*, (Calus and Veerkamp, 2003) [4]. Several reviews on the use of RRM for analysis of test day records of dairy cattle have been given and research is still continuing on determining the order of fit for random regressions for additive genetic and permanent environmental effects that explains the most variations in the observations.

Lotka's law was developed by Alfred J. Lotka (1926) [12] which is also known as "the inverse square law of scientific productivity". It states that the number of authors publishing a certain number of articles is a fixed ratio to the number of authors publishing a single article. In calculating research productivity of authors, Lotka's law helps to identify patterns in the distribution of publications and outliers or deviations from these patterns on usage of RRM in animal breeding research. This law provides fundamental theoretical base for bibliometric studies involving authorships. The Law, as opposed to being a universal rule, is a methodical way of charting the productivity of authors across various fields of study. In any case, it opens up a broader window into deducing the authors' overall output trend in a given field.

Hence, the present study was planned with an idea to find out the authors' productivity pattern on usage of random regression models in animal breeding research with the following objectives through applying Lotka's law

- To verify the applicability of the Lotka's Inverse Square law on usage of RRM in animal breeding research.
- To apply Kolmogorov Smirnov (K-S) goodness-of- fit test for verifying validity of Lotka's Law.
- To identify the top prolific authors and Co-authorship pattern on usage of RRM in animal breeding research.
- Visualization of Sankey diagram through title, author, and country of origin.

Materials and Methods

The information for this study was retrieved on December 11, 2022, from the PubMed database https://pubmed.ncbi.nlm.nih.gov. The time period covered under this study was 13 years (January 1, 2010 to December 11, 2022). Out of 276 publications retrieved from PubMed core collections 14.49 percent were exempted and as per Prism, 2020 remaining 85.51 percent of publications were utilized for further analysis. Flow chart of documents considered in the research was shown in Figure 1. All the odds and bibliographic information for 236 research publications written by 1004 authors published in 62 sources was gathered for this study by searching with the title "Random Regression Model utilized in Animal Breeding Research" at the PubMed core collection. The purpose of this analysis on usage of random regression models in animal breeding research is to verify Lotka's law of scientific production following Pao, (1985) [17] guidelines. On top of that, the data was examined with the help of 'R' programming language (Aria and Cuccurullo, 2017) and the VOS Viewer software (Van Eck and Waltman, 2010). Among 236 publications used, proper citations were given for each document and then performed statistical analysis at 1 percent level of significance to verify the applicability of Lotka's law.

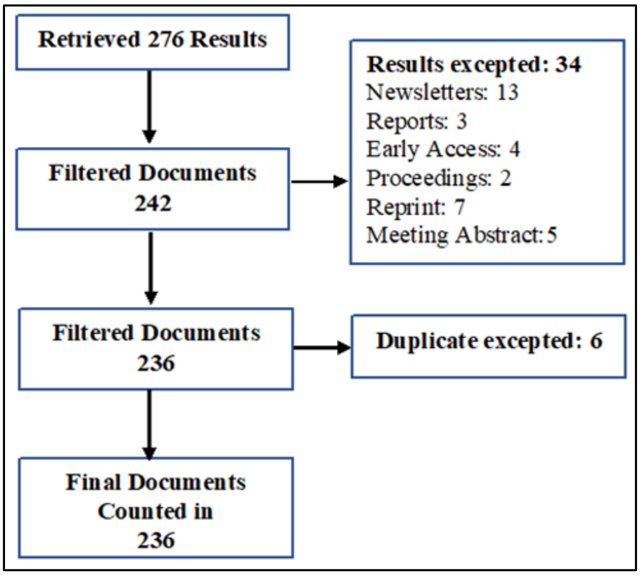

Fig 1: Flow Chart of Documents Considered on Research

Results and Discussion

As shown in Figure 2, the use of the Random Regression Model in the study of animal breeding research had been expanded worldwide during the past 13 years. The average age of the documents since publication was found to be 6.47 years, with a growth rate of 1.19 percent per year and the highest number of citations were observed for the documents published in the year 2010 (15.38%) followed by the year 2011 (14.10%). Highest number of publications were recorded in the year 2021 (13.14%) followed by 2018 (9.75%).

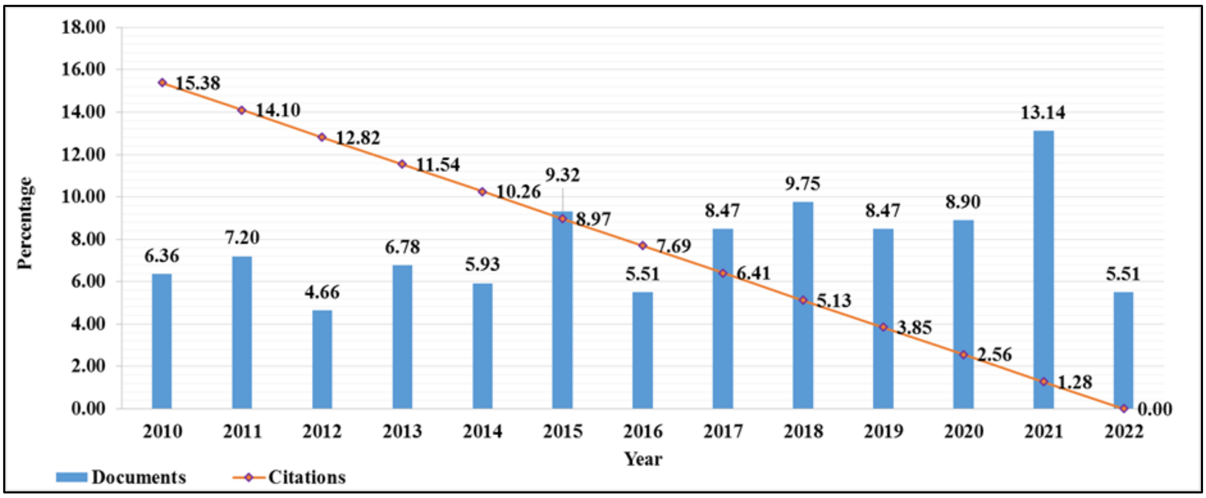

Fig 2: Year wise Publications along with Citation

Lotka's Inverse Square Law determining the value of 'n,' 'c' and 'C V.'

Articles published in Chemical Abstracts between 1907 and 1916 were analyzed by Alfred J. Lotka, (1926) [12], while conducting research in the field of author's scientific output. According to the Authorship Decay Law, "The probability of authors making a single publication is essentially 60 percent," and the number of authors making 'n' contributions is approximately $1/n^2$ of those contributing one. *i.e.*, 60 percent of authors make a single contribution to the field, whereas 15 percent produce two copies ($1/2^2 * 60$), 7 percent disperse three copies ($1/3^2 * 60$) and so on. The results of the calculations used to determine the values of n, C and the critical value which are used to assess the data set's suitability are shown in Table 1.

Lotka's Law articulated in the resulting equation as $x^n * y = c$

Where 'x' refers to the number of articles published (1, 2, 3, 4….); 'y' refers to the number of authors making 'x' contributions as $x^n y$=constant; 'n' is an exponent that is constant for a given set of data; and 'c' is constant.

Pao (1985) [17] stated that the value of 'n' differs based on the data set. When n=2 used for a given data set, then it is termed as "Inverse-square law".

The formulae used are as follows

$$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2}$$

$$c = \frac{1}{\sum_1^{p-1}\frac{1}{x^n} + \frac{1}{(n-1)(p^{n-1})} + \frac{1}{2p^n} + \frac{n}{24(p-1)^{n+1}}} = 0.821834$$

$$\text{Critical Value (C V)} = \frac{1.63}{\sqrt{\sum y_x + \sqrt{\frac{\sum y_x}{10}}}} = 0.051188$$

Kolmogorov-Smirnov (K-S) goodness-of-fit test

Pao, (1986) [18] recommended non-parametric Kolmogorov-Smirnov (K-S) goodness-of-fit test to calculate the maximum deviation as D_{max} = (Fo-Fe) to examine whether the observed frequency pattern on productivity of the author confirmed to the predicted frequency pattern.

Where Fe = Expected author's cumulative frequency.

Fo = Observed author's cumulative frequency of a sample of ‘n’ observations.

At 1% level of significance, the K-S statistic is equivalent to $1.63/\sqrt{\sum yx}$

In Lotka's Law, the deviation represents the difference between the observed number of authors who have published a certain number of papers, and the expected number of authors if the distribution of publications followed a power law. The critical value is a threshold at which the deviation between the observed and expected number of authors becomes statistically significant. As shown in Table 2, this study revealed the value of deviation (D) from the author productivity on usage of RRM in animal breeding research was to be 0.18086 and the K-S value was 0.051188.

The value of D was greater than the critical value and therefore K-S test did not support the applicability of Lotka’s law. This indicates that the distribution deviates from a power law which may be due to abridgement of collaboration in planning and publishing research findings.

The author productivity fit for Lotka’s law is depicted graphically in Figure 3, considering the number of documents and percentage of authors. Inverse square law relationship revealed a “long tail” shaped curve indicating that there was a negative correlation between the number of articles written by an author and proportion of author's relative productivity in the field.

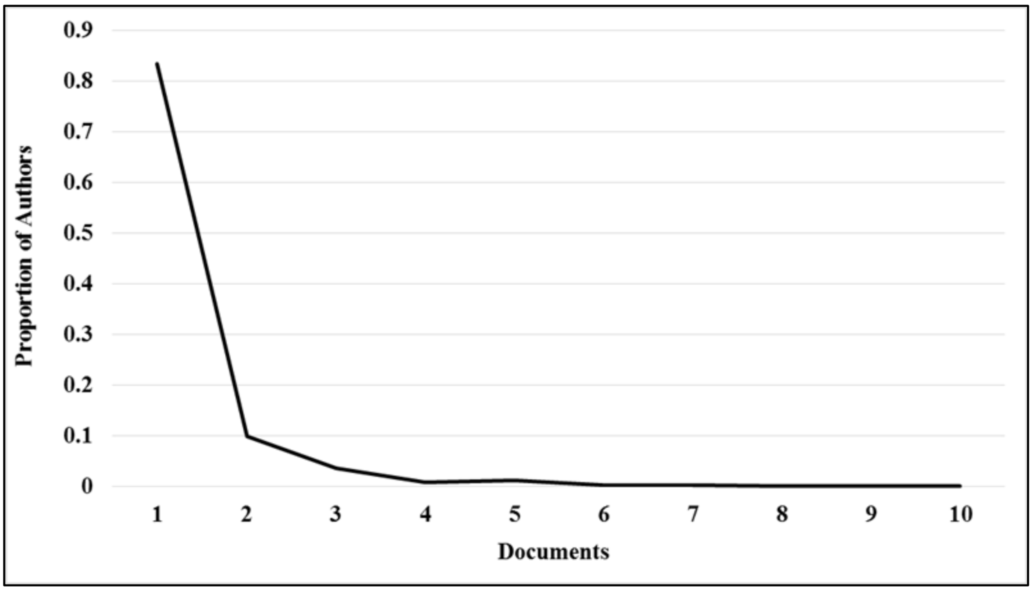

Fig 3: Author Productivity Fit for Lotka’s Law

Table 1: Lotka’s Law (n-value) output on author productivity of random regression model used in animal breeding research

S. No	No of Articles (*x*)	No of Authors Observed (*y*)	Percentage of Authors	Total no of Contributions	Log (*x*)	Log (*y*)	Log x^2	*Log x * Log y*	x^n	$1/x^n$
1	1	837	83.37	837	0.0000	2.9227	0.0000	0.0000	1.0000	1.0000
2	2	100	9.96	200	0.3010	2.0000	0.0906	0.6021	7.6206	0.1312
3	3	36	3.59	108	0.4771	1.5563	0.2276	0.7425	24.9987	0.0400
4	4	8	0.80	32	0.6021	0.9031	0.3625	0.5437	58.0732	0.0172
5	5	12	1.20	60	0.6990	1.0792	0.4886	0.7543	111.6638	0.0090
6	6	4	0.40	24	0.7782	0.6021	0.6055	0.4685	190.5045	0.0052
7	7	4	0.40	28	0.8451	0.6021	0.7142	0.5088	299.2628	0.0033
8	8	1	0.10	8	0.9031	0.0000	0.8156	0.0000	442.5510	0.0023
9	9	1	0.10	9	0.9542	0.0000	0.9106	0.0000	624.9354	0.0016
10	11	1	0.10	11	1.0414	0.0000	1.0845	0.0000	1125.0619	0.0009
		1004	100.00	1317	6.6012	9.6654	5.2997	3.6199	2885.6719	1.2107

Table 2: Kolmogorov-Smirnov Test (Goodness – of - Fit)

S. No	No of Articles (x)	Authors (y_x)	Observed Author: Relative Frequency ($y/\sum y_x$)	Observed Author: Cumulative Frequency (Fo)	Expected Authors: Relative Frequency $C^*(1/x^n)$	Expected Authors: Cumulative Frequency % (Fe)	Deviation (Fo-Fe)
1	1	837	0.83367	0.83367	0.82594	0.82594	0.00772
2	2	100	0.09960	0.93327	0.00000	0.82594	0.10733
3	3	36	0.03586	0.96912	0.00000	0.82594	0.14318
4	4	8	0.00797	0.97709	0.00000	0.82594	0.15115
5	5	12	0.01195	0.98904	0.00000	0.82594	0.16310
6	6	4	0.00398	0.99303	0.00000	0.82594	0.16709
7	7	4	0.00398	0.99701	0.00979	0.81615	0.18086
8	8	1	0.00100	0.99801	0.00662	0.82277	0.17524
9	9	1	0.00100	0.99900	0.00000	0.82277	0.17623
10	11	1	0.00100	1.00000	0.00000	0.82277	0.17723
		1004	1.00000				

For comparison purposes, leading researches on usage of RRM in ABR was depicted in Figure 4, which shows that Misztal Ignacy from Department of Animal and Dairy Science, University of Georgia, Athens published 11 papers, during the year 2010 to 2021 (2010-3; 2015-1; 2016-2; 2019-1; 2021-4); followed by Jamrozik Janusz, Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Canada published 9 papers (2010-2; 2011-1; 2013-2; 2018-2; 2019-2); and Masuda Yutaka, Department of Animal and Dairy Science, University of Georgia, Athens published 8 papers (2011-1; 2015-1; 2016-2; 2017-1; 2019-1; 2020-1; 2021-1). It was observed that a common methodological approach was utilized by these authors in all the documents while using random regression models in animal breeding research.

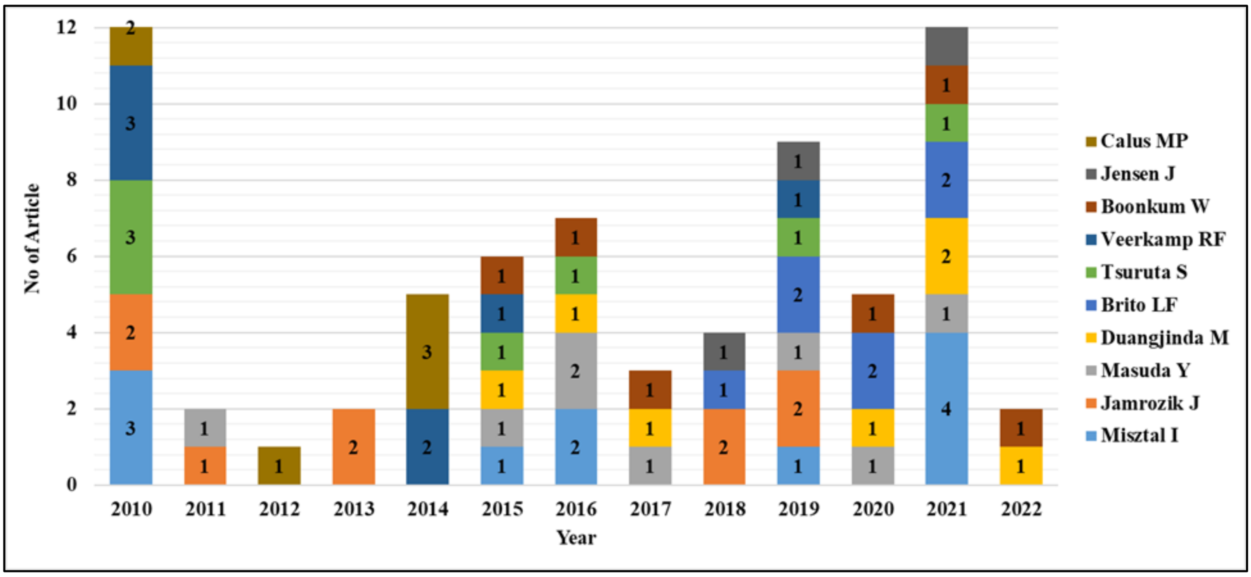

Fig 4: Leading researchers on usage of RRM in Animal Breeding Research

Co-authorship network depicting the relationship between authors as revealed by their article publications on random regression model usage in animal breeding research was shown in Figure 5. One thousand four highly productive authors wrote 236 papers, with an average of 1 author per paper with a range of 1 to 25. A total of 13 clusters with 146 connections among them were observed, of which Cluster 1 (Red) had the most author collaboration with 16 links with a total strength of 17 documents and two authors (Tyrisevä *et al.*, (2011) [22], Biotechnology and Food Research, Biometrical Genetics, MTT Agrifood Research, Finland and Fediaevsky *et al.*, (2010) [5], National Research Institute for Agriculture, Food and Environment (INRAE) France). Cluster 2 (Green) contained two documents, 16 links with a total strength of 14 authors (Hong *et al.*, (2022) [7], College of Animal Science and National Engineering Research Center for Breeding Swine Industry, South China Agricultural University, Guangzhou, China and Kang *et al.*, (2017) [8], National Engineering Laboratory for Animal Breeding, Key Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture; College of Animal Science and Technology, China Agricultural University, Beijing, China). Whereas cluster 3 (Blue) has four documents and 16 links with a total strength of 17 authors.

Cluster 4 (Yellow), 5 (Purple), 6 (Flourescent blue), 7 (Light brown), 8 (Dark brown), 9 (Pink) revealed moderate author collaboration (Figure 5). Whereas, cluster 10 (Light rose), 11 (Green), 12 (Blue) and 13 (Orange) revealed meagre author collaboration.

Wilson, (2022) [24] opined that to enhance the quality of research work Scientists need to exchange new ideas and discuss the research plans and also stated that Internationalization of research provides access to varied skills, faculties, networking and International awareness on research for the benefit of the society.

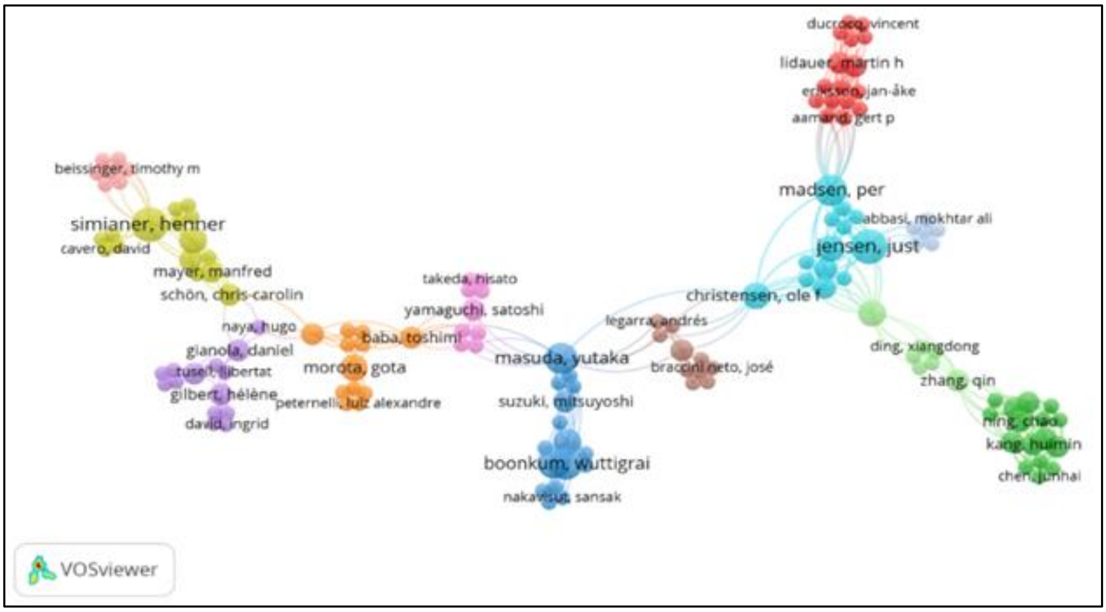

Fig 5: Co-Authorship in RRM in Animal Breeding Research

Sankey diagrams are employed in order to graphically depict the exchange of data between various nodes. The magnitude of the flow was represented by the width of the lines in the diagram, and the direction of the flow is as shown in Figure 6. The left and right sides of the Sankey diagram represent "Title" and "Country," while the element in the middle refers to the "Author". Elements progress from titles to author and then to geographical locations. Thickness of line indicates the amount of flow of scientific information being transmitted from the left to right. According to this study, the word "Regression" was the most popular title used in animal breeding research studies published in the past, followed by the titles Genomic, genetic, dairy and cattle. The terms "estimation," "yield," "milk," "traits" and "parameters" were used with less frequency but may be associated with developing themes in this area. Most of these titles are given by Jamrozik, Duangjinda, Brito, Miszatal and Masuda. This analysis revealed the significant contribution of different authors working on different topics with most commonly used titles in the area of Animal Breeding Research. Thailand, Canada, Brazil, Denmark and U.S.A. were found to be the most prominent Nations that used random regression model in the field of animal breeding research.

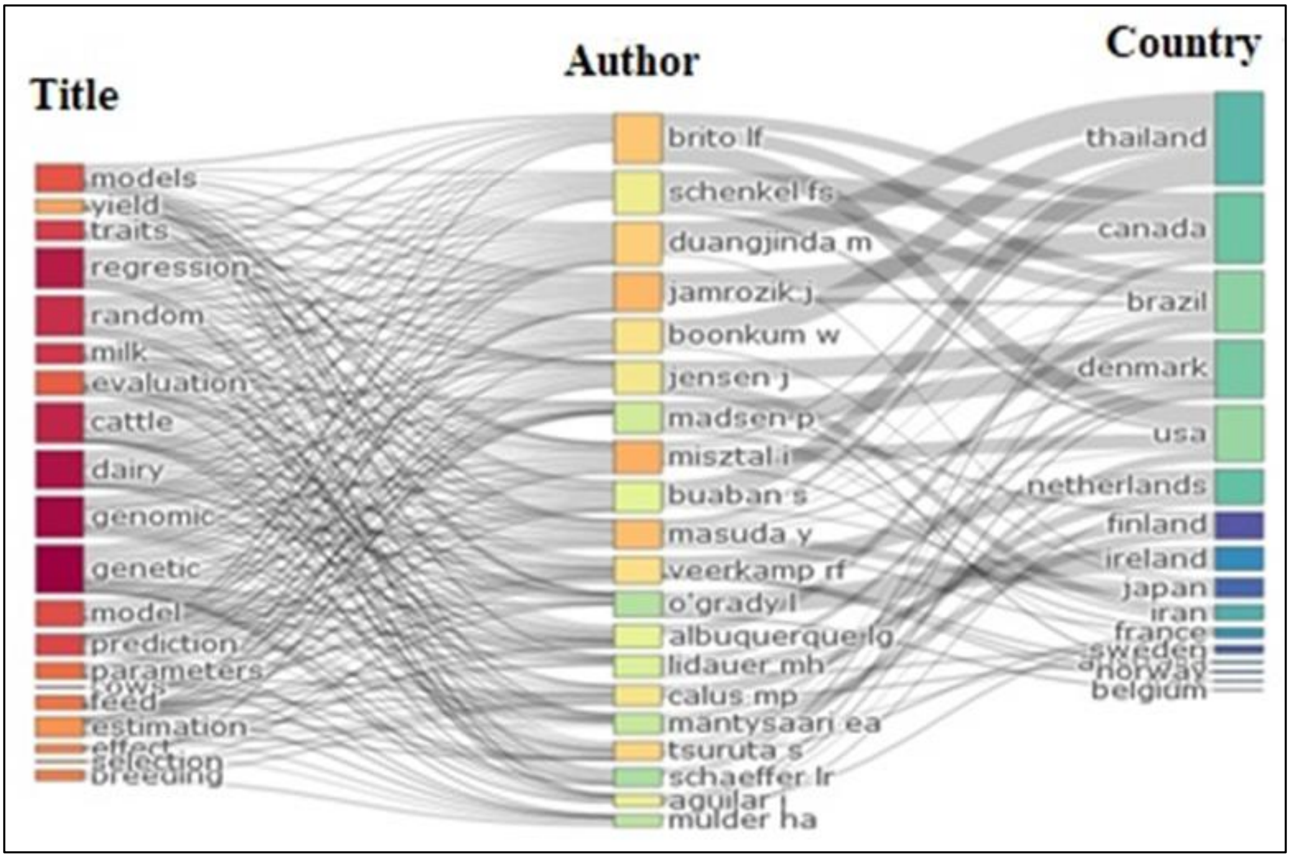

Fig 6: Sankey Diagram for Title, Author and Country

This study investigated the outputs of animal breeding research that used random regression models during the period from 2010 to 2022. The results showed that the highest number of publications were in the year 2021, followed by 2018. The majority of the research outputs were published as research articles. Thailand, Canada, Brazil, Denmark and U.S.A. were the most contributed countries to carry the research on usage of random regression models in animal breeding research. Single author contributed a greater number of articles in the relevant topic and the applicability of Lotka's law stated that the D value (0.18086) is greater than that of critical value (0.051188) indicating that the data did not fit the Lotka's law and designated a shift from the distribution of authors' productivity. This could suggest that the data set may not be sufficient representative of the population under this studied or may be due to the assumptions underlying Lotka's law did not hold in this instance. The source of the discrepancy would require additional research. The inferences found out from this study would be a baseline for future study on usage of RRM in animal breeding research.

Conclusion

The Chi-square test expected value of Lotka's law varied significantly from the observed value and K-S goodness-of-fit test revealed that this study does not adhere to Lotka's inverse square law of scientific productivity. The inferences found out from this study would be a baseline for future study on usage of RRM in animal breeding research.

Acknowledgment

The PubMed Core Collection database and the R programming language with VOS Viewer Open Source software, which were both used for this statistical analysis to visualize the results of this study, are both well appreciated by the authors.

Conflict of interests

The authors have declared no conflict of interests exist

References

1. Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics. 2017;11(4):959-975. https://doi.org/10.1016/j.joi.2017.08.007
2. Bansal S, Mahendiratta S, Kumar S, Sarma P, Prakash A, Medhi B. Collaborative research in modern era: Need and challenges. Indian Journal of Pharmacology. 2019;51(3):137. https://doi.org/10.4103/ijp.IJP_394_19
3. Buaban S, Kuchida K, Suzuki M, Masuda Y, Boonkum W, Duangjinda M. Genetic analysis of the rates of conception using a longitudinal threshold model with random regression in dairy crossbreeding within a tropical environment: Genetic Analysis of Conception Rates in Dairy. Animal Science Journal. 2016;87(8):961-971. https://doi.org/10.1111/asj.12521
4. Calus MPL, Veerkamp RF. Estimation of Environmental Sensitivity of Genetic Merit for Milk Production Traits Using a Random Regression Model. Journal of Dairy Science. 2003;86(11):3756-3764. https://doi.org/10.3168/jds.S0022-0302(03)73982-4
5. Fediaevsky A, Calavas D, Gasqui P, Moazami-Goudarzi K, Laurent P, Arsac JN, *et al.* Quantitative estimation of genetic risk for atypical scrapie in French sheep and potential consequences of the current breeding programme for resistance to scrapie on the risk of atypical scrapie. Genetics Selection Evolution. 2010;42(1):14. https://doi.org/10.1186/1297-9686-42-14
6. Henderson CR. Analysis of Covariance in the Mixed Model: Higher-Level, Nonhomogeneous, and Random Regressions. Biometrics. 1982;38(3):623. https://doi.org/10.2307/2530044
7. Hong Y, Yan L, He X, Wu D, Ye J, Cai G, *et al.* Estimates of Variance Components and Heritability Using Random Regression Models for Semen Traits in Boars. Frontiers in Genetics. 2022;13:805651. https://doi.org/10.3389/fgene.2022.805651
8. Kang H, Zhou L, Mrode R, Zhang Q, Liu JF. Incorporating the single-step strategy into a random regression model to enhance genomic prediction of longitudinal traits. Heredity. 2017;119(6):459-467. https://doi.org/10.1038/hdy.2016.91
9. Kirkpatrick M, Lofsvold D, Bulmer M. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics. 1990;124(4):979-993. https://doi.org/10.1093/genetics/124.4.979
10. Laird NM, Ware JH. Random-Effects Models for Longitudinal Data. Biometrics. 1982;38(4):963. https://doi.org/10.2307/2529876
11. Lidauer M, Mäntysaari EA, Strandén I. Comparison of test-day models for genetic evaluation of production traits in dairy cattle. Livestock Production Science. 2003;79(1):73-86. https://doi.org/10.1016/S0301-6226(02)00142-2
12. Lotka AJ. The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences. 1926;16(12):317-323. https://www.jstor.org/stable/24529203
13. Masuda Y, Aguilar I, Tsuruta S, Misztal I. Technical note: Acceleration of sparse operations for average-information REML analyses with supernodal methods and sparse-storage refinements 1, 2. Journal of Animal Science. 2015;93(10):4670-4674. https://doi.org/10.2527/jas.2015-9395
14. Naranchuluum G, Ohmiya H, Masuda Y, Hagiya K, Suzuki M. Selecting the desirable method for predicting 305-day lactation yields in Mongolia: Prediction of lactation in Mongolia. Animal Science Journal. 2011;82(3):383-389. https://doi.org/10.1111/j.1740-0929.2010.00860.x
15. Oliveira HR, Lourenco Dal, Masuda Y, Misztal I, Tsuruta S, Jamrozik J, *et al.* Application of single-step genomic evaluation using multiple-trait random regression test-day models in dairy cattle. Journal of Dairy Science. 2019;102(3):2365-2377. https://doi.org/10.3168/jds.2018-15466
16. Paneru U, Dhungana KP, Kanu S, Sharma P. Genetic and Non-Genetic Factors on Productive and Reproductive Performance of Indigenous Buffalo. Journal of Nepal Agricultural Research Council. 2021;7:83-91. https://doi.org/10.3126/jnarc.v7i1.36925
17. Pao ML. Lotka's law: A testing procedure. Information Processing & Management. 1985;21(4):305–320. https://doi.org/10.1016/0306-4573(85)90055-X
18. Pao ML. An empirical examination of Lotka's law. Journal of the American Society for Information Science. 1986;37(1):26-33. https://doi.org/10.1002/asi.4630370105
19. Schaeffer LR. Application of random regression models

in animal breeding. Livestock Production Science. 2004;86(1-3):35-45. https://doi.org/10.1016/S0301-6226(03)00151-9
20. Singh A, Rana JS. Prediction of 30S-day milk yield based on test- day values in Murrah buffaloes. The Indian Journal of Animal Sciences. 2008;78(10):Art. 10. https://epubs.icar.org.in/index.php/IJAnS/article/view/4898
21. Thepparat M, Boonkum W, Duangjinda M, Tumwasorn S, Nakavisut S, Thongchumroon T. Genetic evaluation using random regression models with different covariance functions for test-day milk yield in an admixture population of Thailand goats: Goat Genetic Evaluation. Animal Science Journal. 2015;86(7):655-660. https://doi.org/10.1111/asj.12349
22. Tyrisevä AM, Meyer K, Fikse WF, Ducrocq V, Jakobsen J, Lidauer MH, *et al.* Principal component approach in variance component estimation for international sire evaluation. Genetics Selection Evolution. 2011;43(1):21. https://doi.org/10.1186/1297-9686-43-21
23. Van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometrics mapping. Scientometrics. 2010;84(2):523-538. https://doi.org/10.1007/s11192-009-0146-3
24. Wilson. Why International Scientific Collaboration Is Important; c2022. https://collaboratory.ist/international-scientific-collaboration