gosset: An R package for analysis and synthesis of ranking data in agricultural experimentation

## Kauê de Sousa1,2[\*], David Brown3,4, Jonathan Steinke2,5, Jacob van Etten2

1 Department of Agricultural Sciences, Inland Norway University of Applied Sciences, 2318 Hamar, Norway  
2 Digital Inclusion, Bioversity International, Parc Scientifique Agropolis II, 34397, Montpellier Cedex 5, France  
3 Laboratory of Geo-Information Science and Remote Sensing, Wageningen University & Research, Droevendaalsesteeg 3, 6708 PB, Wageningen, The Netherlands  
4 Digital Inclusion, Bioversity International, 30501, Turrialba, Costa Rica  
5 Humboldt University Berlin, Berlin, Germany

[\*] Correspondence should be addressed to: [kaue.desousa@inn.no](mailto:kaue.desousa@inn.no)

# Abstract

Appropriate data management and analysis are necessary to produce practical information from agricultural experimentation data. There is also an ongoing trend advocating for programmatic tools that supports reproducible workflows in scientific research. We developed the R package gosset, providing functionality to support analysis workflows with rank-based models, such as Plackett-Luce and Bradley-Terry. The gosset package facilitates data preparation, modelling and results presentation stages. We demonstrate the functionality of the package with a case of on-farm evaluations of common bean (*Phaseolus vulgaris* L.) genotypes in Nicaragua.

**Keywords**: Bradley-Terry, Plackett-Luce, data science, on-farm trials, participatory-research, tricot

# Required Metadata

# Current code version

*Code metadata*

| **Nr.** | **Code metadata description** | **Please fill in this column** |
| --- | --- | --- |
| C1 | Current code version | 0.4 |
| C2 | Permanent link to code/repository used for this code version | <https://github.com/AgrDataSci/gosset> |
| C3 | Code Ocean compute capsule |  |
| C4 | Legal Code License | MIT |
| C5 | Code versioning system used | Git |
| C6 | Software code languages, tools, and services used | R |
| C7 | Compilation requirements, operating environments & dependencies |  |
| C8 | If available Link to developer documentation/manual | <https://agrdatasci.github.io/gosset/> |
| C9 | Support email for questions | [kaue.desousa@inn.no](mailto:kaue.desousa@inn.no) |

# Motivation and significance

Participatory experimentation approaches have been increasingly applied in agricultural research [1]. While collecting data in ranking format is uncommon in general agricultural research settings, it is often collected in participatory experiments [2]. Recently developed approaches for on-farm experimentation, such as the tricot methodology are based on the collection of data in ranking format [3]. On the other hand, newly proposed approaches for synthesis of crop variety evaluation data largely depend on the analysis of ranking data [4]. The analysis of ranking data requires the use of appropriate statistical models such as Plackett-Luce [5,6] or Bradley-Terry [7]. Functionality for fitting Bradley-Terry and Plackett-Luce models are available in R with the packages BradleyTerry2 and PlackettLuce respectively [8,9]. However, extended functionality was required for the entire data science workflow, which usually includes: (1) Data preparation and cleaning, (2) modelling and validation, and (3) results presentation. For (1) gosset provides functions for converting and preparing data into ranking or pairwise format required by the packages PlackettLuce and BradleyTerry2 respectively. For (2), gosset provides functions for model selection and validation using cross-validation. In the case of (3), enhanced functionality for plotting model results is provided by the gosset package.

# Software description

The R package gosset provides functionality supporting the analysis workflows in agricultural experimentation, especially with rank-based approaches. The package is available in The Comprehensive R Archive Network (CRAN) and can be installed by executing install.packages("gosset"). The package is named in honor to William Sealy Gosset, known by the pen name ‘Student’, a pioneer of modern statistics in small sample experimental design and analysis [10].

## Software Architecture

The R package gosset is structured following the guidelines described in the manual for creating R add-on packages [11]. This structure basically consist of files DESCRTIPTION, LICENSE, NAMESPACE and NEWS, and directories data, dev, docs, inst, man, R, and vignettes. The package functions were developed following the S3 methods style [11] and are contained in the R sub-directory.

## Software Functionalities

### Data management and preparation

* rank\_binomial transforms a ranking object into a binary comparisons, as required by package BradleyTerry2 [8].
* rank\_numeric converts numeric values into rankings. The parameter ascending = indicates if the rankings should be made considering the numeric values in ascending order. The default is asceding = FALSE. This function is useful when the data were collected as numerical observations, for instance, in a experiment measuring crop yield.
* rank\_tricot transforms data in tricot format into PlackettLuce rankings.

### Modelling

* AIC computes the Akaike Information Criterion [12]] for a Bradley-Terry model fitted with BradleyTerry2 [8] or a Plackett-Luce model fitted with [9].
* btpermute deviance-based forward variable selection [13] procedure for Bradley-Terry models.
* crossvalidation performs k-fold cross-validation, where k could be specified by the user.The default is 10-fold. Folds can be provided as a vector for a custom cross-validation, such as blocked cross-validation.
* forward executes forward variable selection with cross-validation.
* kendallTau computes the Kendall-tau rank correlation coefficient between two rankings [14].
* kendallW computes Kendall’s W (coefficient of concordance) among observed rankings and those predicted by the Plackett-Luce model [15].
* pseudoR2 computes goodness-of-fit metrics, such as McFadden’s pseudo-R2 [16].

### Visualization and results presentation

* compare is a visualization approach to compare measures from two different methods [17].
* plot provides an alternative S3 method to plot.pltree() method implemented by the PlackettLuce package.
* regret computes the regret coefficients, the loss under the worst possible outcome. A common heuristic in risk assessment strategy [18].
* reliability computes the probability of a set of items to outperform a reference item. A common heuristic in plant breeding [19].
* worth\_bar creates a bar plot of the estimated *worth* for each evaluated item.
* worth\_map creates a heatmap plot of the estimated log-worth for all items considering each of the evaluated traits.

# Illustrative example

To demonstrate the functionality of the gosset package, we use the nicabean dataset, which was generated with decentralized on-farm trials of common bean (*Phaseolus vulgaris* L.) varieties in Nicaragua over five seasons (between 2015 and 2016). Following the tricot approach [3], farmers were asked to test in their farms three varieties of common bean. The varieties were randomly assigned as incomplete blocks of size three (out of 10 varieties). The farmers assessed which of the three varieties had the best and worst performance in eight traits (vigor, architecture, resistance to pests, resistance to diseases, tolerance to drought, yield, marketability, taste). The farmers also provided their overall appreciation about the varieties, i.e., which variety had the best and the worst performance based on the overall performance considering all the traits. To analyze the data, we use the Plackett-Luce model implemented in the R package *PlackettLuce* [9].

The nicabean is a list with two data frames. The first, trial, contains the trial data with farmers’ evaluations, ranked from 1 to 3, with 1 being the higher ranked variety and 3 the lowest ranked variety for the given trait and incomplete block. The rankings in this dataset were previously transformed from tricot rankings (where participants indicate best and worst) to ordinal rankings using the function rank\_tricot(). The second data frame, covar, contains the covariates associated to the on-farm trial plots and farmers. This example will require the packages PlackettLuce [9], climatrends [20], and ggplot2 [21].

library("gosset")  
library("PlackettLuce")  
library("climatrends")  
library("ggplot2")  
  
data("nicabean", package = "gosset")  
load("nicabean\_temp\_ind.rda")  
  
dat <- nicabean$trial  
  
covar <- nicabean$covar  
  
traits <- unique(dat$trait)

To start the analysis of the data, we transform the rankings into Plackett-Luce rankings using the function rank\_numeric(). We run iteratively over the traits adding the rankings to a list called R. Since the varieties are ranked in an ascending order, with 1 being the higher ranked and 3 the lower ranked, we use the argument asceding = TRUE to indicate which order should be used.

R <- vector(mode = "list", length = length(traits))  
  
for (i in seq\_along(traits)) {  
   
 dat\_i <- subset(dat, dat$trait == traits[i])  
   
 R[[i]] <- rank\_numeric(data = dat\_i,  
 items = "item",  
 input = "rank",   
 id = "id",   
 ascending = TRUE)  
}

Then, using the function kendallTau() we can assess the Kendall tau () coefficient [14]. This approach can be used, for example, to assess the drivers of farmers choices or to prioritize traits to be tested in a next stage of tricot trials (e.g. a lite version of tricot with no more than 4 traits to assess). We use the overall appreciation as the reference trait, and compare the Kendall tau with the other 8 traits.

baseline <- which(grepl("OverallAppreciation", traits))  
  
kendall <- lapply(R[-baseline], function(X){  
 kendallTau(x = X, y = R[[baseline]])  
})  
  
kendall <- do.call("rbind", kendall)  
  
kendall$trait <- traits[-baseline]

The kendall correlation shows that farmers prioritized the traits yield ( = 0.749), taste ( = 0.653) and marketability ( = 0.639) when assessing overall appreciation.

Kendall tau correlation between ‘overall performance’ and the other traits assessed in the Nicaragua bean on-farm trials.

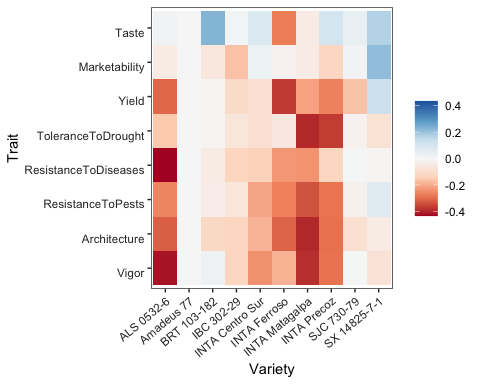
| trait | kendallTau |
| --- | --- |
| Vigor | 0.439 |
| Architecture | 0.393 |
| ResistanceToPests | 0.463 |
| ResistanceToDiseases | 0.449 |
| ToleranceToDrought | 0.411 |
| Yield | 0.749 |
| Marketability | 0.639 |
| Taste | 0.653 |

Then, for each trait, we fit a Plackett-Luce model using the function PlackettLuce() from the package of the same name. This will allow us to continue the analysis of the trial data using the other functions in the package gosset.

mod <- lapply(R, PlackettLuce)

The worth\_map() function can be used to visually assess item performance based on different characteristics. The values represented in a worth\_map are log-worth estimates. From the breeder or product developer perspective the function worth\_map() offers a visualization tool to help in identifying item performance based on different characteristics.

worth\_map(mod[-baseline],  
 labels = traits[-baseline],   
 ref = "Amadeus 77") +  
 labs(x = "Variety",  
 y = "Trait")



Trait performance (log-worth) of bean varieties in Niragua. Variety ‘Amadeus’ is set as reference (log-worth = 0). Blue values indicate a superior performance of varieties for a given trait, compared to the reference. Red values indicate a variety with weak performance for the given trait, compared to the reference.

To consider the effect of climate factors on variety performance, we used climatic covariates to fit a Plackett-Luce tree. For brevity, we used the variables maximum night time temperature (maxNT) and maximum day time temperature, but more variables can be used in a Plackett-Luce tree. Temperature data were obtained from the AgERA5 dataset, using the R package ag5Tools [22]. The temperature data downloaded from AgERA5 were used for computing the climatic variables with the package *climatrends* [20]. In this case, we only use the rankings corresponding to the overall performance of varieties.

# temp <- temperature(covar[, c("longitude","latitude")],   
# day.one = covar[, "planting\_date"],  
# span = 80)  
#   
# head(temp[, 1:2])

To be coupled with climatic covariates, the rankings should be made formatted as grouped rankings. For this we used the function rank\_numeric with the parameter group = TRUE.

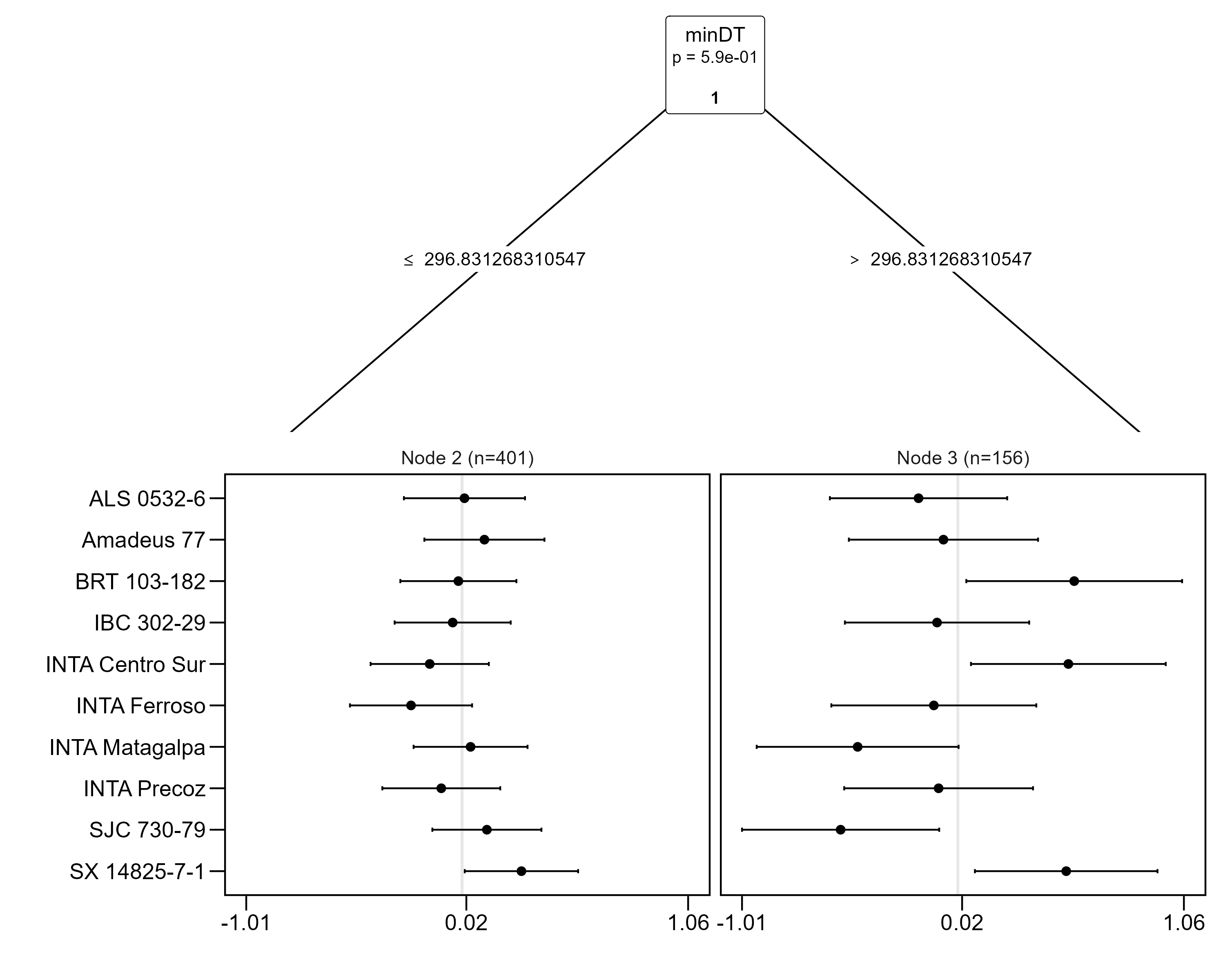
R <- vector(mode = "list", length = length(traits))  
  
for (i in seq\_along(traits)) {  
   
 dat\_i <- subset(dat, dat$trait == traits[i])  
   
 R[[i]] <- rank\_numeric(data = dat\_i,  
 items = "item",  
 input = "rank",   
 id = "id",   
 ascending = TRUE,  
 group = TRUE)  
}

Now we can fit the Plackett-Luce tree with climate covariates. Notice that we used the 9th element in the list of rankings as we are only using the overall appreciation ranking.

R <- R[[9]]  
  
pld <- cbind(R, nicabean\_temp\_ind)  
  
pl <- pltree(R ~ .,   
 alpha = .6,  
 data = pld)

The following is an example of the plot made with the function plot in the gosset package.

plot(pl)



“Plackett-Luce tree plot with gosset function *plot*”

Table 3 shows the results of reliability for the evaluated common bean varieties in each of the resulting nodes of the Plackett-Luce tree.

# Impact

Reproducible and efficient workflows are fundamental in scientific research (Lowndes et al. 2017). The gosset package provides functionality that was not previously available from other R packages and which enabled scientific studies based on the analysis of ranking data. This functionality enables making the entire workflow to be reproducible and more efficient. The utility of the gosset package has been demonstrated by enabling studies based on the analysis of ranking data. For instance, [23], [24] and [25] applied the Plackett-Luce model in combination with recursive partitioning [9,26]. In these studies, the gosset package supported data preparation, model validation and results presentation tasks.

# Conclusions

We described the functionality of the R package gosset to support the synthesis and analysis of ranking data. The package provide functions not available in existing R packages for analyzing ranking data. We provided an illustrative example covering the main functionality across the stages involved in the analysis workflow.

# Conflict of Interest

No conflict of interest exists: We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

# Acknowledgements

We acknowledge Olga Spellman (Science Writing Service of the Alliance of Bioversity International and CIAT) for English editing of this manuscript

# References

1. De Roo, N.; Andersson, J.A.; Krupnik, T.J. On-Farm Trials for Development Impact? The Organisation of Research and the Scaling of Agricultural Technologies. *Experimental Agriculture* **2019**, *55*, 163–184.

2. Coe, R. Analyzing ranking and rating data from participatory on-farm trials. In *Quantitative Analysis of Data from Participatory Methods in Plant Breeding*; Bellon, M.R., J. Reeves, Eds.; CIMMYT: Mexico City, Mexico, 2002; pp. 44–65.

3. van Etten, J.; Beza, E.; Calderer, L.; Van Duijvendijk, K.; Fadda, C.; Fantahun, B.; Kidane, Y.G.; Van De Gevel, J.; Gupta, A.; Mengistu, D.K.; et al. First Experiences with a Novel Farmer Citizen Science Approach: Crowdsourcing Participatory Variety Selection Through on-Farm Triadic Comparisons of Technologies (Tricot). *Experimental Agriculture* **2019**, *55*, 275–296, doi:[10.1017/S0014479716000739](https://doi.org/10.1017/S0014479716000739).

4. Brown, D.; Van den Bergh, I.; de Bruin, S.; Machida, L.; van Etten, J. Data Synthesis for Crop Variety Evaluation. A Review. *Agronomy for sustainable development* **2020**, *40*, 1–20.

5. Luce, R.D. *Individual Choice Behavior*; Courier Corporation, 1959; p. 153;

6. Plackett, R.L. The Analysis of Permutations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **1975**, *24*, 193–202, doi:[10.2307/2346567](https://doi.org/10.2307/2346567).

7. Bradley, R.A.; Terry, M.E. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* **1952**, *39*, 324–345.

8. Turner, H.; Firth, D. [Bradley-Terry Models in R: The BradleyTerry2 Package](https://www.jstatsoft.org/v48/i09/). *Journal of Statistical Software* **2012**, *48*, 1–21.

9. Turner, H.L.; van Etten, J.; Firth, D.; Kosmidis, I. Modelling rankings in R: the PlackettLuce package. *Computational Statistics* **2020**, doi:[10.1007/s00180-020-00959-3](https://doi.org/10.1007/s00180-020-00959-3).

10. Ziliak, S.T. How Large Are Your G-Values? Try Gosset’s Guinnessometrics When a Little “p” Is Not Enough. *The American Statistician* **2019**, *73*, 281–290, doi:[10.1080/00031305.2018.1514325](https://doi.org/10.1080/00031305.2018.1514325).

11. Team, R.C. [Writing r Extensions](https://cran.r-project.org/doc/manuals/r-release/R-exts.html) 2022.

12. Akaike, H. A New Look at the Statistical Model Identification. *IEEE transactions on automatic control* **1974**, *19*, 716–723.

13. Lysen, S. Permuted Inclusion Criterion: A Variable Selection Technique. *Publicly accessible Penn Dissertations* **2009**, 28.

14. Kendall, M.G. A new measure of rank correlation. *Biometrika* **1938**, *30*, 81–93, doi:[10.1093/biomet/30.1-2.81](https://doi.org/10.1093/biomet/30.1-2.81).

15. Kendall, M.G.; Smith, B.B. [The Problem of m Rankings](http://www.jstor.org/stable/2235668). *The Annals of Mathematical Statistics* **1939**, *10*, 275–287.

16. McFadden, D. et al. Conditional Logit Analysis of Qualitative Choice Behavior. **1973**.

17. Bland, J.M.; Altman, DouglasG. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *The Lancet* **1986**, *327*, 307–310, doi:[10.1016/s0140-6736(86)90837-8](https://doi.org/10.1016/s0140-6736(86)90837-8).

18. Loomes, G.; Sugden, R. Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. *The Economic Journal* **1982**, *92*, 805–824, doi:[10.2307/2232669](https://doi.org/10.2307/2232669).

19. Eskridge, K.M.; Mumm, R.F. Choosing Plant Cultivars Based on the Probability of Outperforming a Check. *Theoretical and Applied Genetics* **1992**, *84*, 494–500, doi:[10.1007/BF00229512](https://doi.org/10.1007/BF00229512).

20. de Sousa, K.; van Etten, J.; Solberg, S.Ø. [*climatrends: Climate Variability Indices for Ecological Modelling*](https://CRAN.R-project.org/package=climatrends); 2020;

21. Wickham, H. [*ggplot2: Elegant Graphics for Data Analysis*](https://ggplot2.tidyverse.org); Springer-Verlag New York, 2016; ISBN 978-3-319-24277-4.

22. Brown, D.; de Sousa, K. [*ag5Tools: Toolbox for Downloading and Extracting Copernicus AgERA5 Data*](https://agrdatasci.github.io/ag5Tools/); 2022;

23. van Etten, J.; de Sousa, K.; Aguilar, A.; Barrios, M.; Coto, A.; Dell’Acqua, M.; Fadda, C.; Gebrehawaryat, Y.; van de Gevel, J.; Gupta, A.; et al. Crop variety management for climate adaptation supported by citizen science. *Proceedings of the National Academy of Sciences* **2019**, *116*, 4194–4199, doi:[10.1073/pnas.1813720116](https://doi.org/10.1073/pnas.1813720116).

24. Moyo, M.; Ssali, R.; Namanda, S.; Nakitto, M.; Dery, E.K.; Akansake, D.; Adjebeng-Danquah, J.; van Etten, J.; de Sousa, K.; Lindqvist-Kreuze, H.; et al. Consumer Preference Testing of Boiled Sweetpotato Using Crowdsourced Citizen Science in Ghana and Uganda. *Frontiers in Sustainable Food Systems* **2021**, *5*, 6, doi:[10.3389/fsufs.2021.620363](https://doi.org/10.3389/fsufs.2021.620363).

25. de Sousa, K.; van Etten, J.; Poland, J.; Fadda, C.; Jannink, J.-L.; Kidane, Y.G.; Lakew, B.F.; Mengistu, D.K.; Pè, M.E.; Solberg, S.Ø.; et al. Data-Driven Decentralized Breeding Increases Prediction Accuracy in a Challenging Crop Production Environment. *Communications biology* **2021**, *4*, 1–9.

26. Zeileis, A.; Hothorn, T.; Hornik, K. Model-Based Recursive Partitioning. *Journal of Computational and Graphical Statistics* **2008**, *17*, 492–514, doi:[10.1198/106186008X319331](https://doi.org/10.1198/106186008X319331).