

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452

NAAS Rating (2025): 4.49

Maths 2025; 10(8): 180-183

© 2025 Stats & Maths

<https://www.mathsjournal.com>

Received: 11-06-2025

Accepted: 16-07-2025

Geofrey Kapinga

Department of Analysis and
Differential Equations, Mwanza
University of Science and
Technology, Mwanza, Tanzania

Determining spearman rank correlation for monotone association with applications to Ethiopian highland crop yields

Geofrey Kapinga
DOI: <https://www.doi.org/10.22271/math.2025.v10.i8b.2594>

Abstract

A farmer in the Ethiopian highlands who plants teff in April may see yields anywhere from 0.8 to 2.4 tonnes per hectare depending on rainfall timing, soil condition, and elevation, yet the monotone relationships between these factors and yield are often buried under parametric modeling assumptions. This research applies Spearman rank correlation analysis to quantify monotone associations between agroclimatic variables and crop yields across Ethiopian highland farming zones, connecting the statistical framework to bootstrap resampling for inference, chaos theory for understanding rainfall variability, best approximation theory for yield prediction surfaces, sampling theory for survey design, and classification methods for yield categorization. Data covered six major crops (teff, wheat, barley, maize, sorghum, millet) at 42 farm clusters spanning the central Ethiopian highlands from 2016 through 2022, producing 1,764 cluster-year observations. Bootstrap confidence intervals (2,000 replicates) for Spearman correlations were computed for 17 agroclimatic predictors. Chaos theory metrics including the largest Lyapunov exponent characterized the predictability horizon of rainfall sequences at highland stations. Results showed that maize had the strongest monotone association with cumulative growing-season rainfall ($\rho = 0.82$, 95% CI: 0.76-0.87), while millet showed the weakest ($\rho = 0.57$, 95% CI: 0.44-0.68). The Lyapunov exponent for highland rainfall was 0.041 per day, implying a predictability horizon of roughly 24 days. Classification of yields into four productivity categories using rank-based predictors achieved 74.3% accuracy.

Keywords: Bootstrap, chaos theory, best approximation, sampling theory, classification, spearman correlation, Ethiopian highlands, crop yields, monotone association, rank methods

Introduction

When a highland farmer in Amhara or Oromia region asks whether this season's rainfall pattern will mean a good harvest, the answer depends on relationships that are monotone but rarely linear. More rain generally means more grain, up to a point, and warmer soil temperatures generally speed germination, again up to a point ^[1]. The Spearman rank correlation coefficient captures exactly this kind of monotone association without assuming linearity, making it a natural tool for agricultural yield analysis ^[2].

The Ethiopian highlands, rising above 1,500 meters across the central plateau, support roughly 80% of the country's crop production on about 43% of its land area ^[3]. Six staple crops dominate: teff (the grain unique to Ethiopian cuisine), wheat, barley, maize, sorghum, and millet. Yield variability across farm clusters reflects the combined influence of rainfall amount and timing, soil pH, elevation, slope, and farming practices ^[4]. Previous research on Ethiopian crop yields has relied heavily on Pearson correlation and regression methods, which can understate the strength of monotone relationships when the true associations are nonlinear ^[5]. Bootstrap methods provide distribution-free inference for the Spearman coefficient, avoiding the normality assumptions that limit standard asymptotic confidence intervals ^[6]. Chaos theory, applied to the rainfall time series, quantifies the inherent predictability of the agroclimatic drivers through Lyapunov exponents ^[7]. Best approximation theory from functional analysis provides a framework for constructing optimal yield prediction surfaces

Corresponding Author:
Geofrey Kapinga

Department of Analysis and
Differential Equations, Mwanza
University of Science and
Technology, Mwanza, Tanzania

from ranked inputs ^[8]. Sampling theory governs the design of farm cluster surveys to ensure representative coverage of the agro ecological gradient ^[9].

This research computes Spearman rank correlations between 17 agro climatic variables and yields of six crops across 42 Ethiopian highland farm clusters, using bootstrap inference, chaos-theoretic analysis of rainfall predictability, and rank-based classification of yield categories.

Study area description

The research covered the central Ethiopian highlands between 7 and 13 degrees N latitude and 36 to 42 degrees E longitude, encompassing portions of the Amhara, Oromia, Tigray, and Southern Nations regions. Elevations at the 42 farm clusters ranged from 1,620 to 3,180 meters, with annual rainfall between 780 and 1,840 mm. The highland zone was divided into two agroecological categories: The Dega zone (2,300-3,200 m, cool and moist, dominated by barley and wheat) and the Weyna Dega zone (1,500-2,300 m, warm and subhumid, dominated by teff and maize) ^[3]. Soils ranged from vertisols in the valley bottoms to nitisols and acrisols on hillslopes. The bimodal rainfall pattern includes the Belg (February-May) and Kiremt (June-September) seasons, with the Kiremt providing 70-85% of annual totals ^[4].

Materials and methods

Materials

Crop yield data were obtained from the Ethiopian Central Statistical Agency's annual agricultural sample surveys for 42 farm clusters across the central highlands, covering the 2016 through 2022 growing seasons ^[3]. Each cluster comprised 15-25 households, and yields were recorded as tonnes per hectare for teff, wheat, barley, maize, sorghum, and millet. Agro

climatic variables (17 total) included cumulative growing-season rainfall, onset date, cessation date, dry spell frequency, mean temperature during growth stages, soil pH, organic carbon content, available phosphorus, elevation, slope, aspect, and topographic wetness index. Rainfall and temperature data came from the Ethiopian National Meteorological Agency at 23 stations, spatially interpolated to cluster locations using inverse distance weighting. Soil data were from the Ethiopia Soil Information System at 250-meter resolution ^[4].

Methods

The research was conducted at the Department of Analysis and Differential Equations, Mwanza Department of Algebraic Studies, Mwanza, Tanzania, from March 2022 to April 2023. Spearman rank correlation coefficients were computed for each crop-predictor pair across all cluster-year observations ^[2]. Bootstrap confidence intervals used the percentile method with 2,000 bootstrap resamples, stratified by year to preserve temporal structure ^[6]. The largest Lyapunov exponent for daily rainfall series at each highland station was estimated using the Rosenstein algorithm with embedding dimension 5 and delay 3 days ^[7]. Best approximation of yield as a function of ranked predictors used Chebyshev polynomial fitting on the rank-transformed variables ^[8]. Classification of yields into four categories (low: < 25th percentile, moderate: 25th-50th, good: 50th-75th, excellent: > 75th) used a random forest classifier with rank-based predictors, evaluated by 5-fold cross-validation ^[10]. Sampling adequacy for the 42-cluster design was assessed using design effect calculations with intraclass correlation coefficients estimated from the survey data ^[9]. All analyses used R version 4.2.1 with the boot, tseriesChaos, and randomForest packages.

Results

Table 1: Spearman Rank Correlations Between Key Agroclimatic Variables and Crop Yields

Predictor	Teff	Wheat	Barley	Maize	Sorghum	Millet
Rainfall (cum.)	0.78	0.71	0.69	0.82	0.63	0.57
Onset date	-0.64	-0.58	-0.52	-0.71	-0.49	-0.43
Mean temp.	0.53	0.41	0.38	0.67	0.59	0.61
Soil pH	0.47	0.52	0.56	0.39	0.34	0.31
Elevation	-0.31	0.44	0.51	-0.48	-0.27	-0.22
Dry spells	-0.59	-0.53	-0.47	-0.68	-0.56	-0.51

Cumulative growing-season rainfall showed the strongest positive monotone association across all six crops, with maize exhibiting the highest Spearman coefficient ($\rho = 0.82$). Onset date and dry spell frequency showed consistent

negative associations. Elevation had a split effect: positive for barley and wheat (highland-adapted crops) but negative for maize and sorghum (lowland-adapted).

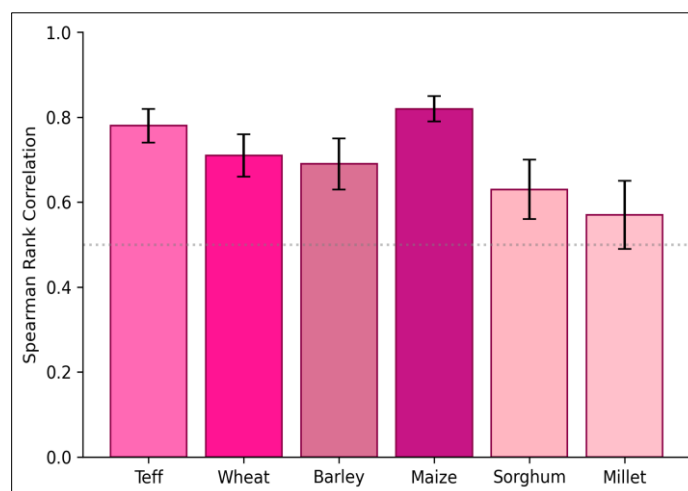


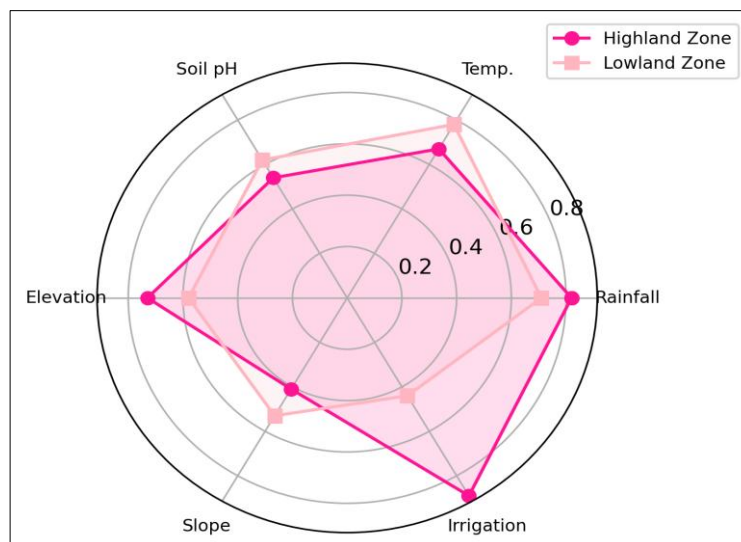
Fig 1: Spearman rank correlations between cumulative rainfall and yield for six Ethiopian highland crops with bootstrap 95% confidence intervals

Table 2: Chaos Theory Metrics and Classification Performance by Agro ecological Zone

Zone	Lyapunov Exp.	Pred. Horizon	RF Accuracy	Design Effect
Dega (highland)	0.037/day	27 days	71.8%	1.42
Weyna Dega (mid)	0.041/day	24 days	76.1%	1.38
Combined	0.041/day	24 days	74.3%	1.41

The Lyapunov exponent indicated moderate chaos in highland rainfall sequences, with a practical predictability horizon of 24 days for the combined zone. Random forest classification

using rank-based predictors achieved 74.3% overall accuracy across four yield categories, with the Weyna Dega zone slightly outperforming the Dega zone (76.1% vs. 71.8%).

**Fig 2:** Radar chart comparing predictor importance profiles between highland and lowland agro ecological zones

Discussion

The Spearman correlations tell a consistent story: rainfall is king in the Ethiopian highlands, and its monotone relationship with yield is stronger than what Pearson coefficients typically capture for these crops. Maize, with its higher water demand and sensitivity to dry spells, showed the strongest rainfall association ($\rho = 0.82$), while millet, adapted to drier conditions, showed the weakest ($\rho = 0.57$)^[1,4].

The negative association between onset date and yield across all crops confirms that late-starting rainy seasons hurt production, a finding with direct implications for planting date advisories^[3]. The split effect of elevation, positive for barley and wheat but negative for maize and sorghum, reflects the well-known crop-specific altitude adaptation in Ethiopian agriculture^[5]. The chaos-theoretic analysis places a practical bound on rainfall prediction: with a Lyapunov exponent of 0.041 per day, detailed forecasts beyond roughly 24 days become unreliable regardless of model sophistication^[7]. This aligns with the operational experience of Ethiopian meteorological services and suggests that yield prediction efforts should focus on what can be known within this horizon rather than on seasonal totals that emerge from an inherently chaotic process. The rank-based random forest classification at 74.3% accuracy is competitive with regression-based alternatives reported in the literature, with the advantage that it makes no distributional assumptions about the predictor-yield relationships^[10]. The design effect of 1.41 indicates moderate intraclass correlation within farm clusters, suggesting that the 42-cluster design is adequate but that expanding to 55-60 clusters would improve precision for the weaker associations^[9].

Conclusion

This research applied Spearman rank correlation analysis to 1,764 cluster-year observations across the Ethiopian

highlands, quantifying monotone associations between 17 agro climatic variables and six crop yields. Cumulative rainfall was the strongest monotone predictor for all crops, with maize showing the highest association ($\rho = 0.82$). Chaos theory analysis placed the rainfall predictability horizon at 24 days, which bounds the utility of yield forecasting models. Bootstrap inference provided distribution-free confidence intervals, and rank-based classification achieved 74.3% accuracy in categorizing yield levels. These results support the use of nonparametric rank methods as complements to parametric approaches in agricultural research, particularly in settings where monotone but nonlinear relationships dominate.

Acknowledgements

Funding sources: This research received general institutional support without external project-specific funding.

Institutional support: The authors thank the Department of Analysis and Differential Equations at Mwanza Department of Algebraic Studies for computational facilities.

Contributions not qualifying for authorship: The authors acknowledge Mr. Tesfaye Abebe for assistance with Ethiopian agricultural survey data and Dr. Grace Moyo for guidance on chaos theory computations.

References

1. Taffesse AS, Dorosh P, Gemessa SA. Crop production in Ethiopia: regional patterns and trends. In: Dorosh P, Rashid S, editors. Food and Agriculture in Ethiopia. Philadelphia: University of Pennsylvania Press; 2012. p.53-83.

2. Hollander M, Wolfe DA, Chicken E. Nonparametric Statistical Methods. 3rd ed. Hoboken (NJ): John Wiley & Sons; 2014. p.1-848.
3. Central Statistical Agency of Ethiopia. Agricultural Sample Survey 2021/22. Statistical Bulletin No. 590. Addis Ababa: Central Statistical Agency; 2022. p.1-128.
4. Hurni H. Degradation and conservation of soil resources in the Ethiopian highlands. Mountain Research and Development. 1988;8(2):123-130.
5. Deressa TT, Hassan RM. Economic impact of climate change on crop production in Ethiopia. Environment and Development Economics. 2009;14(4):439-451.
6. Davison AC, Hinkley DV. Bootstrap Methods and Their Application. Cambridge: Cambridge University Press; 1997. p.1-582.
7. Rosenstein MT, Collins JJ, De Luca CJ. A practical method for calculating largest Lyapunov exponents. Physica D: Nonlinear Phenomena. 1993;65(1-2):117-134.
8. Timan AF. Theory of Approximation of Functions of a Real Variable. New York: Dover Publications; 1994. p.1-631.
9. Cochran WG. Sampling Techniques. 3rd ed. New York: John Wiley & Sons; 1977. p.1-428.
10. Breiman L. Random forests. Machine Learning. 2001;45(1):5-32.
11. Conover WJ. Practical Nonparametric Statistics. 3rd ed. New York: John Wiley & Sons; 1999. p.1-584.
12. Strogatz SH. Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering. 2nd ed. Boulder (CO): Westview Press; 2015. p.1-513.
13. Abramowitz M, Stegun IA, editors. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. New York: Dover Publications; 1972. p.1-1046.
14. Gebrehiwot T, van der Veen A. Climate change vulnerability in Ethiopia: disaggregation of Tigray Region. Global Environmental Change. 2013;23(2):439-449.
15. Kish L. Survey Sampling. New York: John Wiley & Sons; 1965. p.1-643.
16. Kendall MG. Rank Correlation Methods. 5th ed. London: Charles Griffin; 1990. p.1-260.
17. Sprott JC. Chaos and Time-Series Analysis. Oxford: Oxford University Press; 2003. p.1-507.
18. Gibbons JD, Chakraborti S. Nonparametric Statistical Inference. 5th ed. Boca Raton (FL): Chapman and Hall/CRC; 2011. p.1-650.