FACTORS ASSOCIATED WITH CROP YIELD: A CROSS-COUNTRY CORRELATION ANALYSIS

Keldibekov Jahongir[1], Khasanov Dilmurod[2]

[1]Teacher, School 4, Zafarobod Disctrict, Jizzakh, Uzbekistan
[2]Jizzakh branch of the National University of Uzbekistan named after Mirzo Ulugbek, Jizzakh, Uzbekistan
E-mail: jahongirkeldibekov9@gmail.com, tatusf2015@gmail.com

K E Y W O R D S

crop yield, temperature, rainfall, pesticides, Pearson correlation, open agricultural data

A B S T R A C T

Crop productivity responds to climate, water and nutrient availability, soil conditions, crop genetics and biotic pressure. This study distinguishes these agronomic mechanisms from the associations measurable in a public country-level crop-yield dataset. An open Zenodo compilation of FAO yield and pesticide data with rainfall and temperature variables was audited and condensed from 25,932 source rows to 13,130 unique country–year–crop observations for 1990–2013 (excluding 2003). Pearson and Spearman correlations were calculated separately for wheat, maize, rice and soybeans; within-country temporal correlations were used as a sensitivity check. The strongest pooled absolute Pearson association was temperature for maize ($r = -0.481$), whereas pesticide totals were positively correlated with soybean yield ($r = 0.363$). After removing country means, the maize temperature correlation changed to $r = 0.208$. Rainfall was fixed within each country in the supplied dataset, and pesticide quantities were country totals rather than crop application rates. These findings describe associations and data limitations; they do not estimate causal or field-level effects.

1. Introduction

Yield, expressed as harvested product per unit area, is shaped by interacting environmental and management processes. Temperature influences phenology, respiration and grain filling; water stress restricts photosynthesis and nutrient transport; nutrient deficits and adverse soil properties limit root growth. The same rainfall total can be beneficial under one soil or growth stage and harmful under another. Analyses of observed crop yields therefore require the crop, place, period and measurement scale to be specified [1], [2].

Research commonly studies temperature and precipitation because they are observed over large areas; agronomic experiments additionally measure irrigation, nitrogen, phosphorus, soil texture, pH, planting density, cultivar and pest pressure. Irrigation can moderate losses associated with higher temperatures [1], while precise water and fertilizer management depends on soil and crop-specific requirements [3]. Simple correlations are useful for screening patterns, but geographic differences, shared time trends and unmeasured inputs can change their sign or magnitude.

The objective here is to classify the principal mechanisms affecting crop productivity, examine which of them are actually recorded in one reproducible open dataset, and quantify yield correlations by crop. The analysis retains the user-specified subject while interpreting "directly affecting" as an agronomic mechanism, not a causal conclusion from observational correlations.

2. Materials and Methods

2.1. Factor selection and dataset

The analysis uses the open dataset deposited by Laktionov, Semenov and Diachenko [4]. Its

(cc) BY

documentation identifies FAO as the source of crop yields and pesticide totals, and a separate public crop-yield collection as the source of rainfall and temperature. Yield is reported in hectograms per hectare (hg ha^{-1}); 10,000 hg ha^{-1} equals 1 t ha^{-1}. Each record names a country, year and crop, and supplies annual rainfall (mm), pesticide use (tonnes), temperature (°C) and yield. The four selected crops are wheat, maize, paddy rice and soybeans.

Table 1 distinguishes mechanisms typically investigated in agronomy from the variables the dataset can support. Pesticide mass is a country-wide total and cannot stand for pest incidence, efficacy or crop-specific treatment intensity. Fertilizer rates, irrigation, soil attributes and cultivar are absent, so they are reviewed conceptually but excluded from the numeric correlations.

Table 1.

Agronomic factor groups and their representation in the dataset

Factor group	Plausible yield pathway	Available variable
Climate	Heat, photosynthesis and seasonal water supply	Temperature; long-term rainfall
Soil and nutrients	Root conditions and N–P–K uptake	Not available
Water management	Irrigation timing and water deficit	Not available
Protection and biology	Pest losses, disease and treatment	National pesticide total only
Genetics and husbandry	Cultivar, sowing date and plant density	Crop category only

2.2. Data audit and unit of analysis

Inspection found 25,932 source rows but only 13,130 distinct country–year–crop groups. In 3,690 groups, the same yield was repeated beside different temperature values, up to 19 rows per group. As the deposited file does not identify the repeated temperature measurements, their arithmetic mean was used as a pragmatic summary; one yield observation was retained per group. No group contained conflicting yield values and no analysed field was missing. The resulting panel spans 101 countries, ten crop categories and 23 years between 1990 and 2013; 2003 is absent. Rainfall has exactly one value per country across the entire period. Such rainfall values represent between-country differences in this file, not year-to-year rainfall fluctuations.

Table 2 gives five actual wheat observations after aggregation. Values are displayed in the original yield unit; for example, the Albanian 1990 yield of 30,197 hg ha^{-1} corresponds to 3.0197 t ha^{-1}. The temperature for Brazil in 2010 is the mean of nine source values, rather than one of the individual entries.

Table 2.

Illustrative records from the analysed wheat panel

Country	Year	Rain (mm)	Temp (°C)	Pesticides (t)	Yield (hg ha^{-1})
Albania	1990	1,485	16.37	121.00	30,197
Australia	2000	534	16.67	33,475.00	18,209
Brazil	2010	1,761	23.64	342,580.00	28,288
India	2011	1,083	25.82	55,540.00	29,886
France	2013	867	11.01	66,497.29	72,656

(cc) BY

2.3. Correlation procedure

Pearson's coefficient measures linear association between each numeric factor X and yield Y within a crop. Spearman's coefficient applies the same idea to ranked values, indicating whether a monotone relationship persists despite skew or outliers. To contrast geographic and temporal patterns, temperature and pesticide values and yields were separately centered on their country-specific mean, then correlated across available years. The analysis is descriptive; no formal independent-observation significance claim is made because records within countries share climate and management context.

$$r_{xy} = \Sigma_i (x_i - \bar{x})(y_i - \bar{y}) / \sqrt{[\Sigma_i (x_i - \bar{x})^2 \Sigma_i (y_i - \bar{y})^2]}$$

The bar denotes the sample mean and i indexes country–year observations of one crop. Country centering replaces x_i and y_i by their deviations from the respective country means. This eliminates all rainfall variation in the deposited data, so a within-country rainfall coefficient is undefined.

3. Results and Discussion

3.1. Crop-specific pooled associations

Table 3 reports correlations from unique country–year–crop observations. The sample sizes are 1,810 for wheat, 2,028 for maize, 1,502 for paddy rice and 1,242 for soybeans. Temperature has the largest absolute pooled Pearson coefficient among the three measured variables for wheat, maize and rice. Pesticide totals show the largest coefficient for soybeans. These are rankings of observed associations within this dataset, not rankings of agronomic causal importance.

Table 3.

Pooled correlations with crop yield by factor

Crop	n	Rain r	Temp r	Pest r	Rain ρ	Temp ρ	Pest ρ
Wheat	1810	-0.049	-0.431	+0.052	+0.006	-0.449	+0.262
Maize	2028	-0.131	-0.481	+0.194	-0.077	-0.505	+0.520
Paddy rice	1502	-0.121	-0.307	+0.277	-0.064	-0.323	+0.557
Soybeans	1242	-0.011	-0.257	+0.363	+0.008	-0.258	+0.545

The maize temperature coefficient is −0.481, yet the rice coefficient is −0.307: crop-specific responses and country composition both matter. Rainfall Pearson coefficients are small and negative in this compiled file, including −0.131 for maize. This must not be read as evidence that reducing rainfall raises yield: the rainfall column is constant over time for every country and mixes agroclimatic, economic and geographic differences. Pesticide totals are more strongly associated with soybean yield (r = 0.363) than with wheat yield (r = 0.052), but a national tonnage is heavily influenced by national agricultural scale. The gap between Pearson and Spearman results, especially for maize pesticides (0.194 versus 0.520), suggests nonlinearity, skew and/or outlier sensitivity.

3.2. Within-country sensitivity and interpretation

Table 4 compares pooled correlations with correlations after country means are removed. The change in temperature sign is conspicuous: maize shifts from −0.481 to +0.208, wheat from −0.431 to +0.144, and rice from −0.307 to +0.220. The within-country statistic still mixes secular yield improvement, changes in management and temperature variability; it is not a causal temperature response. The soybean temperature relation becomes weak (+0.023). For pesticides, the within-country correlations remain positive but are based on national totals, not dose applied to the selected crop.

(cc) BY

Table 4.

Pooled and country-centered Pearson correlations

Crop	Temp pooled	Temp centered	Pest pooled	Pest centered
Wheat	-0.431	+0.144	+0.052	+0.129
Maize	-0.481	+0.208	+0.194	+0.170
Paddy rice	-0.307	+0.220	+0.277	+0.220
Soybeans	-0.257	+0.023	+0.363	+0.147

The contrast illustrates why a single global coefficient cannot identify a directly acting factor. Mechanistic studies find that temperature effects interact with water availability and differ by crop and location [2]. Field experiments, with soil, irrigation and nutrient inputs observed at plot scale, are needed to estimate responses to interventions. The present file has no field-level soil moisture, growth-stage weather, fertilizer application, crop-level pesticide use or pest incidence. Its repeated temperature rows also lack an explicit sampling definition. These limitations constrain both statistical adjustment and interpretation.

4. Conclusion

Temperature and water supply, soil and nutrient status, irrigation, pest and disease pressure, and cultivar and management practices are the main agronomic factor groups affecting crop productivity. In the analysed open panel, temperature, long-term rainfall and national pesticide totals were the only numeric proxies among these groups. After correcting duplicated yield observations, crop-specific pooled correlations showed the largest measured temperature association for maize ($r = -0.481$) and the largest pesticide association for soybeans ($r = +0.363$). The reversal of several temperature correlations after country centering demonstrates substantial dependence on the analysis scale. Future work should join crop-season weather with plot-level irrigation, soil tests, fertilizer application, variety and pest measurements, then estimate nonlinear and interaction effects under an appropriate study design. The present coefficients are a reproducible exploratory description, not proof that a factor directly caused a change in yield.

Data availability

The public source file df_yield_merged.csv and its component files are available from the cited Zenodo record [4]. The analysis used the published merged CSV, grouped by country, year and crop, with arithmetic mean temperature and one non-conflicting yield per group. All reported coefficients are reproducible from these rules.

References

1. P. Agnolucci, C. Rapti, P. Alexander, V. De Lipsis, R. A. Holland, F. Eigenbrod, and P. Ekins, “Impacts of rising temperatures and farm management practices on global yields of 18 crops,” Nature Food, vol. 1, pp. 562–571, 2020, doi: 10.1038/s43016-020-00148-x.
2. M. Matiu, D. P. Ankerst, and A. Menzel, “Interactions between temperature and drought in global and regional crop yield variability during 1961–2014,” PLOS ONE, vol. 12, no. 5, e0178339, 2017, doi: 10.1371/journal.pone.0178339.
3. Y. Xing and X. Wang, “Precise application of water and fertilizer to crops: challenges and opportunities,” Frontiers in Plant Science, vol. 15, 1444560, 2024, doi: 10.3389/fpls.2024.1444560.
4. I. Laktionov, S. Semenov, and G. Diachenko, “Dataset for Crop Yield Prediction,” Zenodo, version 1, 2026, doi: 10.5281/zenodo.18761663.
5. N. Raximov, D. Khasanov, "THE IMPORTANCE OF FORECASTING CROP YIELDS AND REGIONAL CROP DISTRIBUTION BASED ON AGRO-CLIMATIC CHARACTERISTICS AND MARKET DYNAMICS," Raqamli Transformatsiya va Sun’iy Intellekt, 4(5), 1-5, 2026. https://dtai.uz/articles/the-importance-of-forecasting-crop-yields-and-

(cc) BY

regional-crop-distribution-based-on-agro-climatic-characteristics-and-market-dynamics/

6. D. Khasanov, N. Raximov, N. Raximov, "PREDICTION THE YIELD OF GRAIN CROPS USING BASIC MACHINE LEARNING ALGORITHMS," Raqamli Transformatsiya va Sun'iy Intellekt, 2(5), 40-46, 2024. https://dtai.uz/articles/prediction-the-yield-of-grain-crops-using-basic-machine-learning-algorithms/
7. D. Khasanov, B. Daminova and M. Tojiyev, "The Impact of Normalization on Regression-Based Crop Yield Prediction: Accuracy and Efficiency Analysis," *2026 International Russian Smart Industry Conference (SmartIndustryCon)*, Sochi, Russian Federation, 2026, pp. 63-67, doi: 10.1109/SmartIndustryCon68821.2026.11493015.

(cc) BY *dtai.uz*