Journal of Stored Products Research 120 (2027) 103285

Contents lists available at ScienceDirect

Journal of Stored Products Research

journal homepage: www.elsevier.com/locate/jspr

A surveillance-based interpretation of commercial soybean reception reveals season-dominated organization across 16 harvest seasons

Maicon Marinho Vieira Araujo *, Paulo Otávio Aldaves dos Santos Guedes

Federal University of Mato Grosso (UFMT), Faculty of Agronomy and Animal Science (FAAZ), Storage Technology Center (NTA), Cuiabá, Mato Grosso, Brazil

ARTICLE INFO

Keywords:
Surveillance
Postharvest
Soybean
Storage
Climate
Archives

ABSTRACT

Postharvest science is built mainly on controlled experiments that resolve how management changes lots after intake. Commercial laboratories, meanwhile, continuously generate standardized physical, compositional and sanitary observations that, to our knowledge, are seldom treated as scientific evidence for population-level inference. We propose Surveillance-based Postharvest Science (SPS) as a potentially transferable interpretive framework whose first empirical application is evaluated here: reception records are read as a storage-entry state—an integrated analytical profile at the start of the postharvest chain—whose organization becomes visible when long-term archives are analysed as integrated observational systems. SPS rests on five principles (Figure 1) and four formal constructs: storage-entry state, year-conditioning, trait-specific seasonal signal and climatic readability. Using 2562 commercial soybean lots from a Mato Grosso reception network across 16 harvest seasons (2010/11–2025/26), seasonal structuring exceeded regional structuring for every trait (largest partial R^2_season for acidity, 0.415); trait-specific seasonal signal ranked acidity (Adj. R^2 = 0.461) and ether extract (0.263) above moisture (0.163); climate–quality associations were selective (mold–rainfall retained within a pre-specified FDR focus set) while domain-level adjusted RDA remained weak; and no monotonic 16-season drift was detected, with seasons instead grouping into compositional year types. Empirical inference is restricted to the monitored network; transferability of the framework remains to be evaluated independently.

1. Introduction

Commercial grain trade depends on a brief but consequential laboratory moment: reception. Before drying schedules are set and storage bins are assigned, lots are scored under harmonized protocols for physical, compositional and sanitary attributes that govern price, segregation and operational routing (Bartosik et al., 2023; Ziegler et al., 2021). In practice, each score is usually consumed alone—as a tolerance check—then discarded as a scientific object. That habit is efficient for commerce and insufficient for inference. The condition that begins the stored-grain pathway is jointly physical, chemical and sanitary; reading it one attribute at a time leaves the organizing structure of incoming quality invisible (Anukiruthika and Jayas, 2024; Magan et al., 2011).

A second habit shapes the literature. Postharvest knowledge is disproportionately produced under designed contrasts that ask how a treatment moves a lot after intake (Coradi et al., 2021; Cañizares et al., 2024; Lima et al., 2024; Souza et al., 2023). Those designs are indispensable for mechanism. They are also selective: they compress the very mixture of weather, genotype turnover, harvest pace, transport and pre-storage handling that defines commercial arrivals across seasons (Jaques et al., 2022; Toomer et al., 2023). The result is an evidentiary asymmetry. Laboratories already record standardized physical, compositional and sanitary observations at the entrance of the chain; to our knowledge, however, such commercial archives are seldom analysed as population-level observational systems spanning many seasons and many lots (Dubal et al., 2023; Bartosik et al., 2023).

This study proposes that routine commercial reception records should be interpreted not as independent quality measurements, but as the observable expression of an integrated storage-entry state whose organization becomes evident only through long-term analysis of large monitored archives.

We call the corresponding interpretive programme Surveillance-based Postharvest Science (SPS; Fig. 1). We propose SPS as a potentially transferable interpretive framework whose first empirical application is evaluated here; it is not a substitute for experimental postharvest research. Five commitments organize the programme: laboratories continuously generate standardized physical, compositional and sanitary observations; those observations jointly describe a

* Corresponding author.
E-mail addresses: maicon.araujo@ufmt.br, maiconmarinho@outlook.com (M.M.V. Araujo).

https://doi.org/10.1016/j.jspr.2026.103285
Received 6 August 2026; Received in revised form 3 October 2026; Accepted 24 October 2026
0022-474X/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
The Article Processing Charge for the publication of this research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (ROR identifier: 00x0ma614).

multidimensional condition at pathway entry; multi-season archives can disclose emergent organization at population scale; climate recovers only part of that organization; and management decisions should combine surveillance evidence with experimental knowledge rather than collapse either into the other.

Four constructs make the programme operable. The storage-entry state is the integrated physical, compositional, chemical and sanitary profile of a grain lot at reception—a multidimensional analytical construct inferred from co-measured attributes, not a latent biological entity observed directly and not equated with any single compliance score. Year-conditioning is the composite seasonal signal for which harvest season stands in an archive (measured climate plus unmeasured cultivar turnover, timing, logistics, within-lab change and population mix); season effect, year-conditioning and climate effect are therefore kept distinct. Trait-specific seasonal signal replaces any "information hierarchy" wording: traits are ranked by partial $R^2/\Delta R^2$_season and by stability under leave-one-season sensitivity, not by Shannon information, predictive information gain or decision value. Climatic readability is the degree of observational alignment between seasonal reception variation and harvest-window climate; it is an alignment diagnostic, not a validated prediction score (Magan et al., 2011; Coradi et al., 2021).

From these constructs follow three propositions. First, incoming commercial lots are expected to show structured organization at storage entry, with year-conditioning typically outweighing place-conditioning when harvest-season processes dominate. Second, attributes should rank unequally in the temporal structure they encode. Third, harvest-window climate should explain only part of the structure, because genotype, management, timing, drying and logistics also shape the lot before it is scored (Ziegler et al., 2021; Toomer et al., 2023).

Mato Grosso soybean intake supplies a demanding first empirical application rather than the definition of SPS. A large commercial production system, season-to-season climatic contrast and a consolidated single-laboratory surveillance stream yield a multi-year archive with protocol continuity and authentic operational heterogeneity (CONAB, 2025; USDA-FAS, 2025). Using 2562 lots across sixteen harvest seasons (2010/11–2025/26), we ask whether this monitored reception archive exhibits the temporal, spatial and climatic organization anticipated under SPS: whether season or region dominates structuring, how traits rank in seasonal signal, and how far harvest-window climate accounts for the pattern (climatic readability). The intended contribution is methodological and conceptual—an interpretive framework for commercial surveillance in postharvest science—with soybean as an initial empirical application, not an independent multi-system validation.

2. Materials and methods

2.1. Study design and reporting

This retrospective observational study examined whether a long-term commercial soybean archive exhibits the temporal, spatial and climatic organization anticipated by the proposed Surveillance-based Postharvest Science (SPS) framework. Investigators assigned no treatments. The target of inference was population-level organization of the observed reception profile, rather than the causal effect of a drying or storage treatment. Reporting was informed by the STROBE checklist for observational studies, particularly regarding study population, variable definition, missing data, potential bias and limits of inference; STROBE

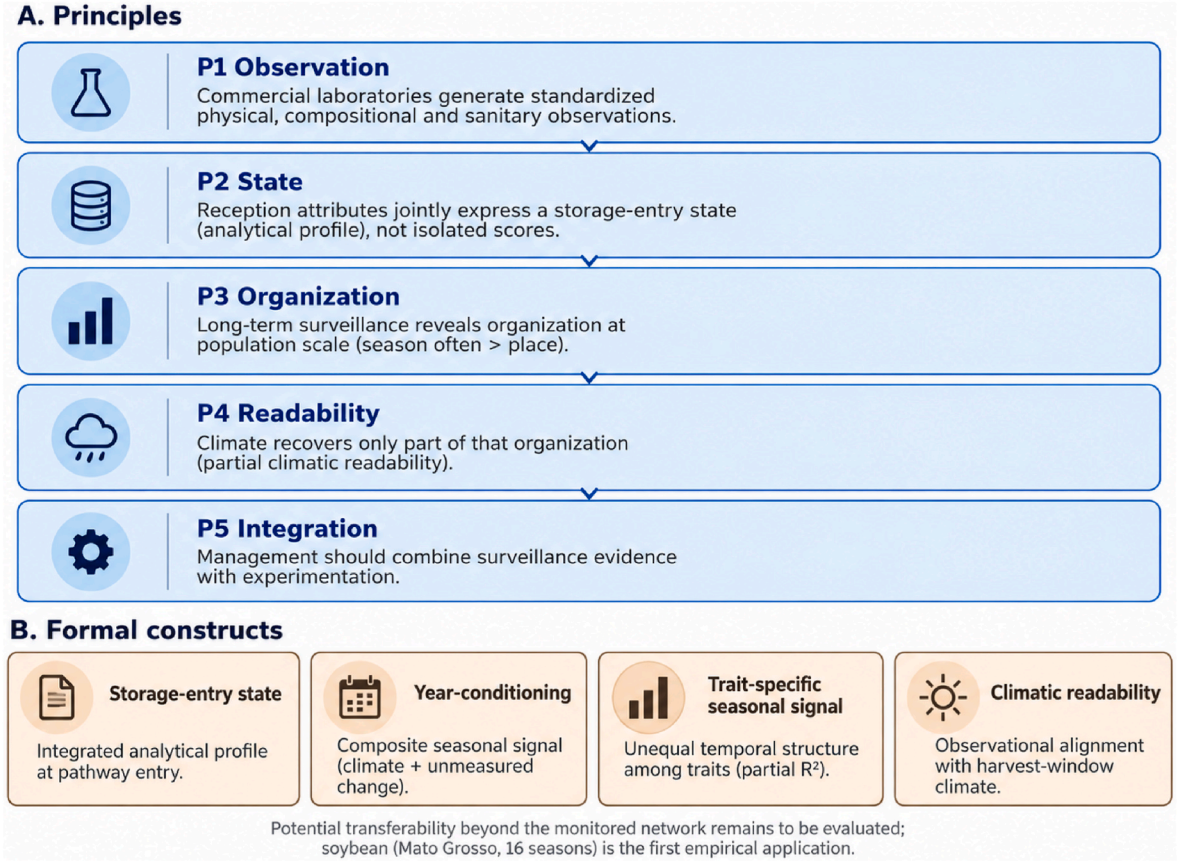


Fig. 1. Surveillance-based Postharvest Science (SPS). (A) Five principles linking commercial laboratory archives to complementary postharvest evidence. (B) Formal constructs: storage-entry state, year-conditioning, trait-specific seasonal signal and climatic readability. Network-specific coefficients are not transferable. Potential transferability of the conceptual framework beyond the monitored soybean reception network remains to be evaluated independently. Soybean (Mato Grosso, 16 seasons) is the first empirical application.

was used as guidance, not as a claim of complete checklist compliance for every item.

2.2. Commercial surveillance network and sample inclusion

The analytical unit was one commercial soybean lot represented by a traced composite sample (~1 kg). All laboratory determinations used in this study were performed at the Núcleo de Tecnologia em Armazenagem (NTA), Federal University of Mato Grosso (UFMT). Records entered the archive through routine commercial laboratory demand within this NTA surveillance stream (a large-scale monitored population), not through probability sampling of all Mato Grosso production. Inclusion required a harvest-season label, a network macro-region (North, East, South or West), a municipality name and at least one of the six reception traits analysed below. Duplicate records were removed using available traceability fields before curation.

2.3. Operational definition of the reception profile

The conceptual construct under SPS is the storage-entry state (analytical profile, not a latent entity). Operationally, the reception profile was the six-trait vector of moisture (%), test weight (kg hL^{-1}), crude protein (%), ether extract (%), acidity (g 100 g^{-1}) and moldy grains (%) recorded in the curated analytical dataset (n = 2562 lots; 16 harvest seasons, 2010/11–2025/26; 93 municipalities).

Samples were obtained in the commercial reception pathway before long-term storage. Depending on routine laboratory workflow, samples could undergo cleaning for analytical preparation. Commercial drying was not experimentally standardized across lots or seasons; when drying occurred before final laboratory recording, it formed part of the observed operational pathway rather than a controlled treatment. Consequently, the reception profile represents the condition available to dump-pit and laboratory decision-making at the storage-entry window, and may already incorporate routine pre-storage operations. It should not be interpreted as a purely field-inherited vector free of all post-harvest intervention.

2.4. Laboratory determinations and interannual comparability

All determinations for the soybean archive analysed here were performed at NTA/UFMT, within the same commercial soybean-quality surveillance line documented by Andrade (2016) for Mato Grosso harvests. Physical and sanitary classification followed the Brazilian official soybean grading standard (Portaria SDA no. 532/2022; MAPA, 2022). Moldy grains were scored visually as the mass percentage of grains with fungal colonization or mold evidence under that standard and retained as the sanitary coordinate with continuous seasonal coverage. Moisture was determined with a G810 analyser and confirmed by oven drying at 105 °C for 24 h (wet basis) according to laboratory routine. Test weight was measured with hectolitre equipment and expressed as kg hL^{-1}. Crude protein followed Kjeldahl (AOAC Official Method, 2001.11) and ether extract Soxhlet extraction (AOAC Official Method 920.39). Near-infrared spectroscopy was used under calibrations referenced to wet chemistry (AOAC International, 2023a, 2023b); season-specific NIR validation statistics are not re-reported here beyond that calibration lineage. Acidity (lipid acidity) was determined by alkaline titration of free fatty acids following AOAC Official Method 940.28 and expressed as g oleic-acid equivalent per 100 g (AOAC International, 2023c). Complementary sanitary assays, when recorded, included blotter tests and additional visual identification under standard laboratory methodology, but were not forced into the primary reception profile. All assays were performed in this single NTA laboratory stream; residual season effects may still partly reflect within-lab changes in client mix, equipment generation or workflow intensity across the 16 seasons.

2.5. Climate data and seasonal aggregation

Harvest-window climate (September–March) was summarized from INMET stations (INMET, 2026) as seasonal means: rainfall, mean and maximum temperature, mean relative humidity, rainy days, days >20 mm and >50 mm, fungal-risk hours, hours >35 °C, maximum daily rainfall and global radiation. The September–March envelope spans the main Mato Grosso soybean sowing-to-harvest window and was chosen a priori as a statewide seasonal exposure; it is not cultivar- or municipality-specific phenology. Fungal-risk hours were defined as cumulative hours that simultaneously met temperature and humidity conditions favourable to fungal development in the climate-processing pipeline. Station counts were approximately 39 in most seasons and 12 in 2025/26 after coverage quality control. Primary climate–quality analyses used these statewide seasonal envelopes. A secondary ecological layer reused previously computed lot–nearest-station associations as a reduced-scale sensitivity (not a causal exposure model). Neither layer substitutes laboratory measurement of the reception profile.

2.6. Data curation, missingness and quality control

Season, region and units were harmonized. No global imputation was applied. Trait-wise missingness in the curated dataset was: moisture 36, test weight 53, protein 37, ether extract 36, acidity 91 and moldy grains 82 (of 2562). Range filters for inferential layers are reported in Supplementary Table S1. Test weight outside 50–90 kg hL^{-1} was excluded (nine records), separating impossible entries (including five zeros and one value of 22,141.273 kg hL^{-1}) from the retained operational range. Additional domain exclusions removed one moisture, two protein, one ether-extract and two acidity records. Coverage analyses used all curated lots; confirmatory models used trait-wise complete cases after filtering (moisture 2525; test weight 2500; protein 2523; ether extract 2525; acidity 2469; moldy grains 2480); multivariate analyses required all six coordinates simultaneously (n = 2243).

2.7. Statistical estimands and analysis plan

The storage-entry state is a conceptual construct. The statistical estimands were: (i) trait-specific variation associated with harvest season and macro-region; (ii) the incremental seasonal signal of each trait; (iii) season-level associations between climate descriptors and trait means; (iv) multivariate organization and variance partitioning of the six-trait profile; and (v) monotonic drift across the 16 seasonal means. Analyses were organized to evaluate the three SPS propositions without treating SPS as already validated.

2.7.1. Temporal and regional structuring (P1)

Primary trait-wise models were linear mixed models

$$y = \text{season} + \text{region} + (1 \mid \text{municipality})$$

with categorical season (16 levels) and region (4 levels). Municipality random intercepts addressed spatial dependence among lots. When mixed models are summarized alongside OLS, municipality-clustered sandwich standard errors were used as a companion specification. Partial R^2 for season (given region) and for region (given season), together with $\Delta R^2_season = R^2_season + region - R^2_region$, quantified relative structuring; comparisons emphasize effect-size metrics rather than p-values alone. Season × region interaction was tested by nested comparison of additive versus interactive fixed-effect models (all season–region cells had $n \geq 6$). Leave-one-season-out checks summarized (i) stability of Adj. $R^2/\Delta R^2_season$ and (ii) region-mean predictive RMSE for held-out seasons. For moldy grains (≈17% exact zeros), a two-part sensitivity—linear-probability presence model and severity among positives—supplemented an asinh-transformed model. Residual skewness, Jarque–Bera tests, Levene heteroscedasticity across seasons and

Cook's-distance outliers were inspected diagnostically.

2.7.2. Trait-specific seasonal signal (P2)

Traits were ranked by Adj. R^2 of the two-factor models, by ΔR^2_season and by partial R^2_season, then inspected through seasonal means with 95% confidence intervals and leave-one-season stability. Under SPS this ranking is the trait-specific seasonal signal construct (not Shannon information or decision value).

2.7.3. Climate–quality associations (P3)

Climate predictors were grouped a priori into four domains (wetness, heat, radiation, extreme rainfall) and scored as within-domain standardized means. Domain-constrained RDA reports raw and adjusted R^2 with a global permutation test and leave-one-season R^2. Spearman correlations linked seasonal climate descriptors to seasonal trait means ($n = 16$). A pre-specified focus set (mold–rainfall/RH; oil–radiation; protein–temperature; moisture–rainfall/RH; acidity–RH/temperature) is reported with bootstrap percentile intervals and FDR within the focus set; intense-rainfall vs test-weight pairs are labelled exploratory. FDR across the full climate × trait matrix (72 tests) is also reported. Climatic readability is therefore an observational alignment diagnostic, not a validated predictive score.

2.7.4. Multivariate ordination and variance partitioning

PCA on z-scored complete cases provided an exploratory ordination of the six-trait reception profile (Jolliffe and Cadima, 2016), with bootstrap loading intervals and sensitivity runs excluding 2010/11 and excluding moldy grains. PERMANOVA-style centroid partitions on Euclidean distances of the standardized profile, plus season × region cell models, partitioned multivariate variation; PERMDISP tested homogeneity of multivariate dispersions. Complete-case versus curated-archive composition was compared (χ^2 on season membership) as a coverage check.

2.7.5. Trend analysis and season typology

Mann–Kendall tests with Sen slopes (and lag-1 prewhitened sensitivity) on one mean per season assessed monotonic change (Mann, 1945; Kendall, 1975; Sen, 1968). Pettitt change-point tests and a k-means typology of the 16 seasonal z-mean profiles ($k = 3$) described year-type structure rather than linear drift alone. With $n = 16$, the trend battery remains interpretive.

2.8. Limits of inference, software and data governance

Empirical inference is restricted to the monitored soybean reception network. Transferability of the SPS conceptual framework to other commodities and regions remains to be evaluated independently. Because cultivar identity was inconsistently recorded and strongly unbalanced across seasons (256 labels), it was not included as a predictor; estimated season effects may partly reflect cultivar turnover and other unmeasured temporal changes within the single NTA stream (equipment generation, client profile and commercial logistics).

Analyses used Python (pandas, NumPy, SciPy, statsmodels; McKinney, 2010; Harris et al., 2020; Virtanen et al., 2020; Seabold and Perktold, 2010). $\alpha = 0.05$ unless stated. Lot-level records were analysed after removal of producer/company identifiers for secondary research use under institutional laboratory-record policies; de-identified outputs may be shared by the corresponding author upon reasonable request.

3. Results

3.1. Sampling frame

The curated analytical dataset comprised 2562 lots (North 805, South 667, East 620, West 470) across 16 seasons and 93 municipalities (Fig. 2). Annual lot counts ranged from 90 (2013/14) to 233 (2014/15 and 2020/21); municipal coverage peaked at 45 in 2014/15 (Table 1). Trait-wise confirmatory n after filtering remained high (moisture 2525; test weight 2500; protein 2523; ether extract 2525; acidity 2469; moldy grains 2480; six-trait complete cases 2243). Range exclusions are summarized in Supplementary Table S1.

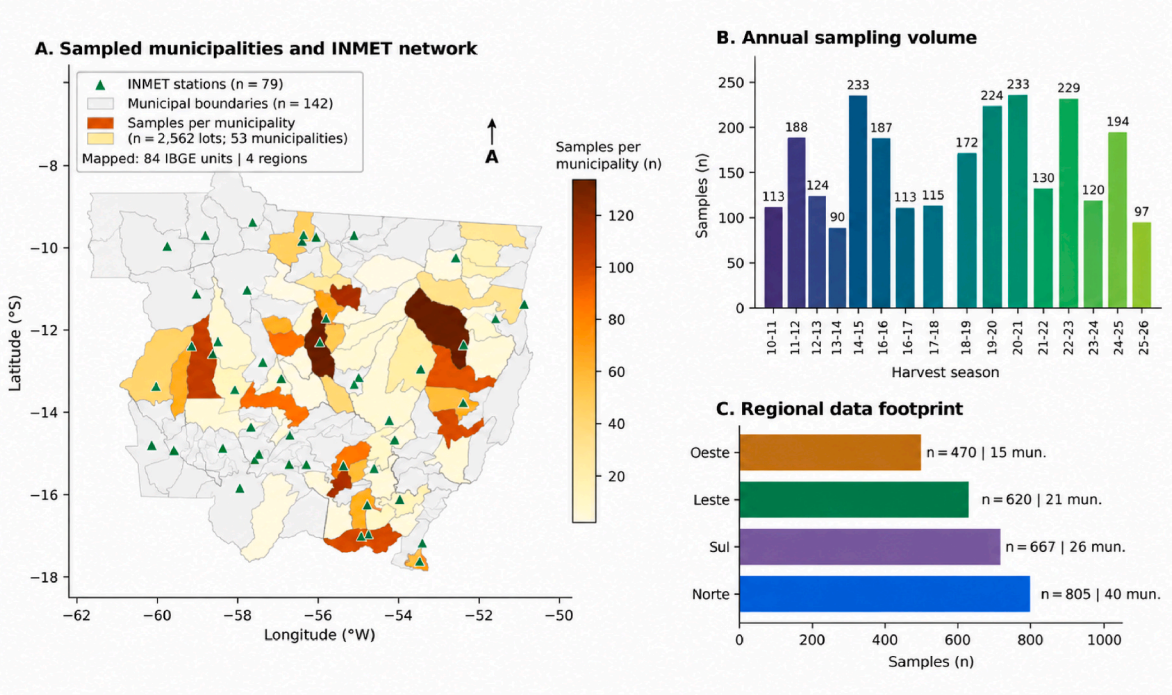


Fig. 2. Sampling frame for storage-entry soybean quality in Mato Grosso (2010/11–2025/26). (A) Municipal intensity and INMET stations. (B) Annual lot counts. (C) Macro-regional footprint. Map lines delineate study areas and do not necessarily depict accepted national boundaries.

Table 1
Sampling frame and seasonal climate envelopes.

Season	N lots	Municipalities	Cultivars	Norte	Leste	Sul	Oeste	Rainfall (mm)	Mean T (°C)	Mean RH (%)	Rainy days	Fungal-risk hours	Hours >35 °C	Radiation (MJ m^{-2})
2010-2011	113	17	21	32	30	26	25	1261	25.57	74.75	88	2048	103.6	4786
2011-2012	188	26	49	33	29	91	35	1250	25.45	75.48	110.9	2282	72.21	3870
2012-2013	124	31	9	38	21	31	34	1393	25.76	75.9	118.5	2429	106.4	3721
2013-2014	90	21	47	32	20	11	27	1398	25.25	77.44	120.2	2464	45.15	3663
2014-2015	233	45	1	95	49	59	30	1157	25.67	75.98	95.73	2298	80.53	3533
2015-2016	187	28	48	66	31	58	32	952	26.68	73.82	93.42	2107	168.5	3612
2016-2017	113	39	61	50	29	22	12	1258	25.41	77.08	107.2	2405	55.79	3500
2017-2018	115	22	25	42	36	19	18	1275	25.87	75.24	105.2	2206	103.8	3350
2018-2019	172	23	25	51	65	26	30	1321	25.71	76.76	103.1	1899	41.69	3485
2019-2020	224	25	27	78	35	89	22	1176	26.21	74.48	88.13	2028	123.5	3752
2020-2021	233	25	23	69	84	58	22	874.4	26.72	68.56	47.1	710.5	160.1	3513
2021-2022	130	22	23	28	32	43	27	508.1	26.65	73.31	23.64	286.8	25.34	2863
2022-2023	229	26	27	58	71	45	55	1244	25.34	73.94	94.28	1803	65.18	3681
2023-2024	120	18	20	33	33	48	6	1048	27.02	73.28	87.09	1766	202.4	4038
2024-2025	194	20	23	53	41	28	72	1144	26.73	70.92	74.55	1092	119.8	3397
2025-2026	97	20	22	47	14	13	23	1090	25.88	72	102.8	1931	107.4	3758

3.2. P1 — Seasonal versus regional structuring

In season + region models, season structured every trait more strongly than region (Table 2). Season-only R^2 exceeded region-only R^2 throughout: acidity 0.420 vs 0.086; ether extract 0.267 vs 0.004; protein 0.141 vs 0.002; moisture 0.121 vs 0.060; moldy grains 0.062 vs 0.019; test weight 0.039 vs 0.029. Partial R^2_season (season given region) mirrored that order (acidity 0.415; ether extract 0.265; protein 0.142; moisture 0.115; mold 0.061; test weight 0.040). Municipality random-intercept mixed models converged for all traits; ICCs were modest for composition (protein 0.044; oil 0.098) and higher for acidity (0.214) and especially mold (0.512), indicating substantial municipal clustering for the sanitary coordinate. Municipality-clustered tests left season significant for every trait. Season × region interaction raised R^2 only modestly ($\Delta R^2 \approx 0.02$–0.06) and did not overturn season > region. Leave-one-season predictive checks showed region means only weakly beating a global mean for most traits, consistent with year-conditioning dominating place. PERMANOVA-style partitions on complete cases recovered the same multivariate ordering (season $R^2 = 0.188$; region $R^2 = 0.033$; additive ≈ 0.211; cell interaction $\Delta R^2 \approx 0.040$; Table 4). PERMDISP indicated heterogeneous seasonal dispersions ($p = 0.005$) and marginal regional dispersion differences ($p = 0.060$). For moldy grains (16.8% exact zeros), a two-part sensitivity showed stronger seasonal structure in presence (ΔR^2_season = 0.157) than in severity among positives (0.047); asinh Adj. $R^2 = 0.123$. Within this monitored network, P1 was supported as stronger seasonal than regional structuring—not as independent validation of SPS as a whole (see Table 5).

3.3. P2 — Trait-specific seasonal signal

Adjusted R^2 from the two-factor models ranked traits as: acidity 0.461 (RMSE = 0.458) > ether extract 0.263 (1.078) > moisture 0.163 (1.529) > crude protein 0.137 (1.581) > moldy grains 0.072 (0.832) > test weight 0.060 (1.874) (Table 2). The same order was recovered by ΔR^2_season and by partial R^2_season. Leave-one-season refits kept that ranking stable (mean Adj. R^2 across held-out seasons tracked the full-sample order). Seasonal trajectories showed the practical content of that ranking (Fig. 3). Acidity spanned 0.36 g 100 g^{-1} (2011/12) to 1.87 g 100 g^{-1} (2010/11); ether extract 19.28% (2015/16) to 21.55% (2010/11); moisture 11.48% (2011/12) to 13.94% (2010/11). The 2010/11 season combined high moisture, high acidity and high mold; 2011/12 combined the opposite pole. Test weight stayed near 68–69 kg hL^{-1} and contributed little seasonal structure.

Regional means modified water status, acidity and mold more than composition (Fig. 4): South was driest (11.64% moisture); North

Table 2
Seasonal versus regional structuring and trait-specific seasonal signal.

trait	n	Adj_R2_full	R2_season	R2_region	delta_R2_season	partial_R2_season	partial_R2_region	mixed_icc	delta_R2_interaction
Acidity	2469	0.4614	0.4204	0.08612	0.3792	0.415	0.0776	0.2135	0.03621
Ether extract	2525	0.263	0.2672	0.004389	0.2639	0.265	0.001469	0.09809	0.03177
Crude protein	2523	0.1367	0.1414	0.001507	0.1413	0.1415	0.00167	0.04411	0.01815
Moisture	2525	0.1627	0.1206	0.06048	0.1081	0.1151	0.05467	0.06691	0.0612
Moldy grains	2480	0.07153	0.06229	0.01857	0.0597	0.06083	0.01705	0.5123	0.03568
Test weight	2500	0.06038	0.03891	0.02871	0.03844	0.03958	0.02939	0.108	0.03859

Table 3
Pre-specified climate–quality focus associations.

Quality	Climate	set	Spearman_rho	Spearman_p	boot_lo	boot_hi	Spearman_q_prespec	sig_FDR_prespec
Moldy grains	rainfall_mm	prespecified	0.6853	0.003392	0.2474	0.902	0.03052	True
Moldy grains	mean_rh_pct	prespecified	0.5647	0.02266	0.1015	0.8691	0.07609	False
Ether extract	global_radiation_mj_m2	prespecified	0.5559	0.02536	0.1337	0.8036	0.07609	False
Crude protein	mean_temp_c	prespecified	−0.3059	0.2493	−0.7106	0.2151	0.3205	False
Crude protein	max_temp_c	prespecified	−0.4412	0.08715	−0.8107	0.06699	0.1961	False
Moisture	rainfall_mm	prespecified	−0.04118	0.8797	−0.5241	0.5654	0.8797	False
Moisture	mean_rh_pct	prespecified	−0.1206	0.6564	−0.5638	0.3823	0.7385	False
Acidity	mean_rh_pct	prespecified	0.3353	0.2043	−0.1512	0.7471	0.3205	False
Acidity	mean_temp_c	prespecified	−0.3147	0.2352	−0.7157	0.2081	0.3205	False
Test weight	days_rain_gt50mm	exploratory	0.6382	0.007801	0.0629	0.9075	nan	nan
Test weight	max_daily_rainfall_mm	exploratory	0.7412	0.001018	0.4155	0.9207	nan	nan

Table 4
Multivariate diagnostics and domain RDA summary.

metric	value
PCA_PC1_pct	25.23
PCA_PC2_pct	19.63
PCA_cum_pc12	44.86
PCA_drop2010_cum_pc12	45.55
PCA_without_mold_cum_pc12	50.95
PERMANOVA_R2_season	0.1878
PERMANOVA_R2_region	0.0335
PERMANOVA_R2_additive	0.2109
PERMANOVA_deltaR2_interaction	0.04027
PERMDISP_season_p	0.005
PERMDISP_region_p	0.06
RDA_domains_R2	0.2753
RDA_domains_R2_adj	0.01183
RDA_domains_p_perm	0.404
completeness_chi2_p	2.431e-36

combined higher moisture (12.60%), acidity (1.62 g 100 g^{-1}) and mold (0.82%); protein and oil remained nearly flat across regions (~37.4–37.6% and ~20.2–20.4%). Under the operational metric used here, P2 was supported as unequal trait-specific seasonal signal, with acidity and oil at the top.

3.4. P3 — Partial climatic readability

Harvest-window climate varied widely among seasons (Fig. 5; Table 1): rainfall 508 mm (2021/22) to 1398 mm (2013/14); fungal-risk hours 287 to 2464; mean RH 68.6–77.4%; mean temperature 25.25–27.01 °C. The 2012/13–2013/14 window was wet/humid; 2021/22 was dry; 2023/24 peaked in hours >35 °C.

Season-level climate–quality correlations were selective (Fig. 6; Table 3). Within the pre-specified focus set, mold versus rainfall retained FDR $q = 0.031$ ($\rho = 0.685$; bootstrap 95% interval 0.25–0.90); mold–RH and oil–radiation were nominally positive but did not retain focus-set FDR. Intense-rainfall metrics versus test weight remained exploratory ($\rho = 0.64$–0.74). Under FDR control of the full climate × trait matrix (72 tests; $n = 16$), no pair retained $q < 0.05$. Acidity—the strongest temporal encoder under P2—lacked stable climate links among the retained indicators. A four-domain RDA (wetness, heat, radiation, extreme rainfall) yielded constrained $R^2 = 0.275$ but adjusted $R^2 \approx 0.012$ and a nonsignificant global permutation test ($p = 0.40$; Table 4). Lot–nearest-station correlations (secondary layer) similarly highlighted moisture/acidity/mold associations with temperature and wetness descriptors without elevating climate to a surrogate state. P3 is therefore supported only as partial observational readability: selected climate–trait links appear, but multivariate climate alignment is not robust at $n = 16$.

3.5. Drift check — oscillating baseline and year types

Mann–Kendall tests on 16 seasonal means were non-significant at $\alpha = 0.05$ for all traits; lag-1 prewhitening did not change that conclusion. Acidity ($\tau = -0.350$, $p = 0.065$; Sen slope −0.030 g 100 g^{-1} $season^{-1}$, 95% CI overlapping zero) and moldy grains ($\tau = -0.367$, $p = 0.053$) only approached decline. Pettitt change-point tests were likewise non-significant for every trait. A k-means typology of the 16 z-mean reception profiles ($k = 3$) separated (i) the 2010/11 high-moisture/high-oil/high-acidity/high-mold pole, (ii) high-protein/high-oil seasons and (iii) low-oil/low-protein seasons—an oscillating year-type structure rather than a monotonic 16-season drift.

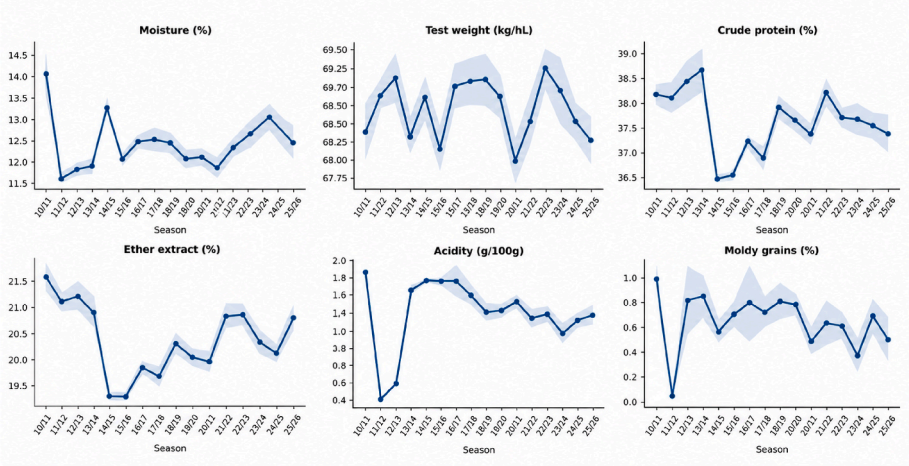


Fig. 3. Seasonal trajectories of quality traits (±95% CI) across 16 harvest seasons.

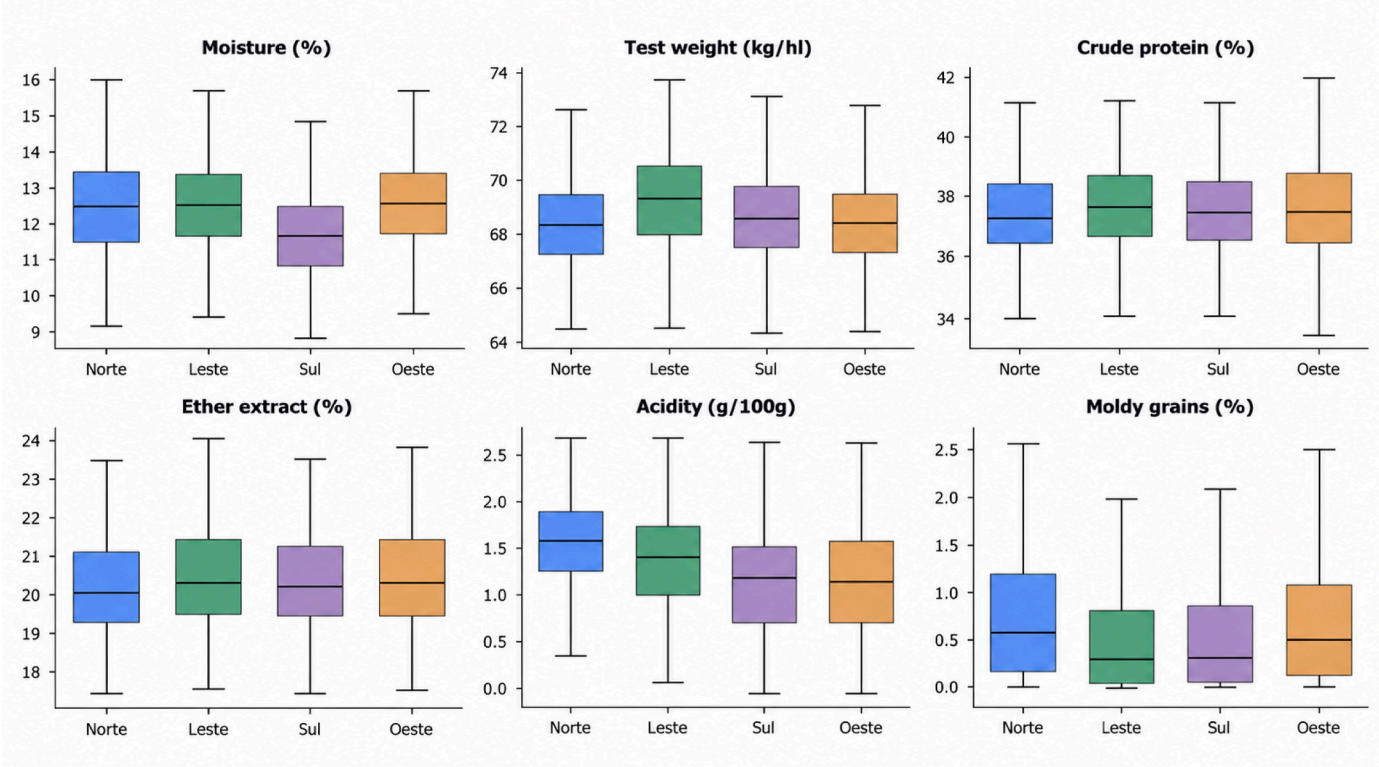


Fig. 4. Regional modifiers of storage-entry quality (North, East, South, West).

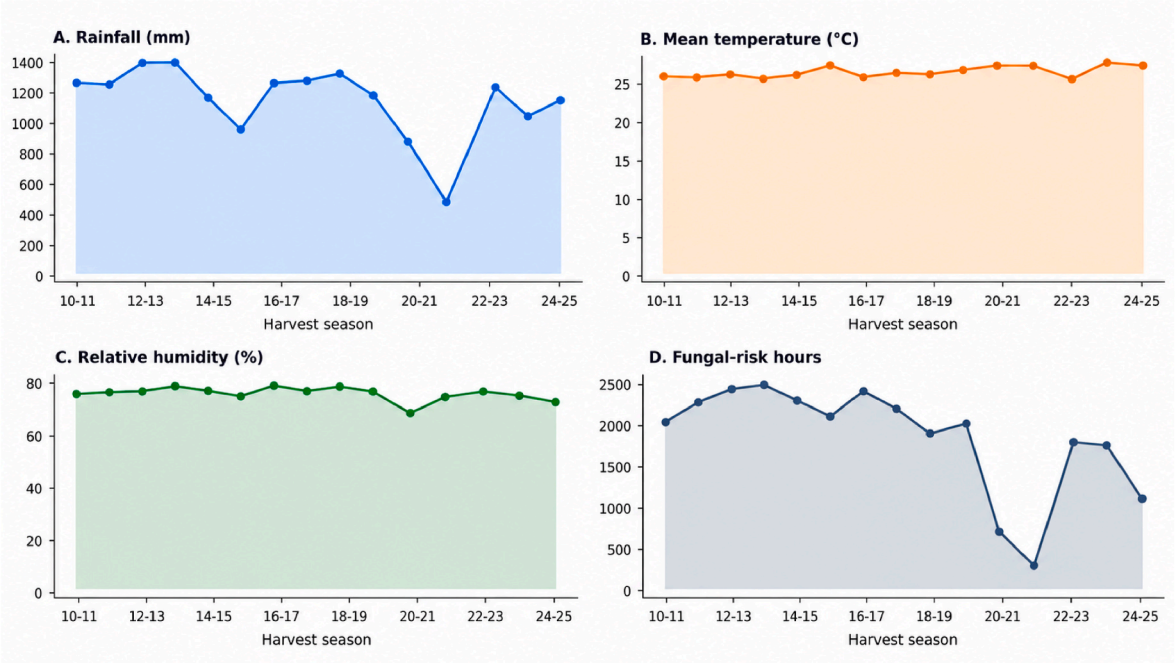


Fig. 5. Harvest-window climate envelopes used as readability candidates.

3.6. *Multivariate structure of the reception profile*

In complete cases (n = 2243), PCA provided an exploratory ordination of the standardized six-trait profile: PC1 (25.2%) opposed ether extract (loading 0.54) and protein (0.45) to acidity (−0.58) and moisture (−0.39); PC2 (19.6%) opposed mold (0.67) to test weight (−0.57); cumulative PC1+PC2 = 44.9% (Fig. 7; Table 4). Bootstrap loading intervals and leave-one-season PCA kept the PC1–PC2 geometry recognizable; excluding 2010/11 left cum PC1+PC2 ≈ 45.6%, and dropping moldy grains raised it to ≈ 51.0%. Season membership differed between complete-case and incomplete lots (χ^2 $p < 0.001$), so multivariate inference is conditioned on the complete-case filter. Regional centroids remained near the origin relative to the lot cloud, consistent with weaker place than seasonal structuring. The joint pattern is compatible

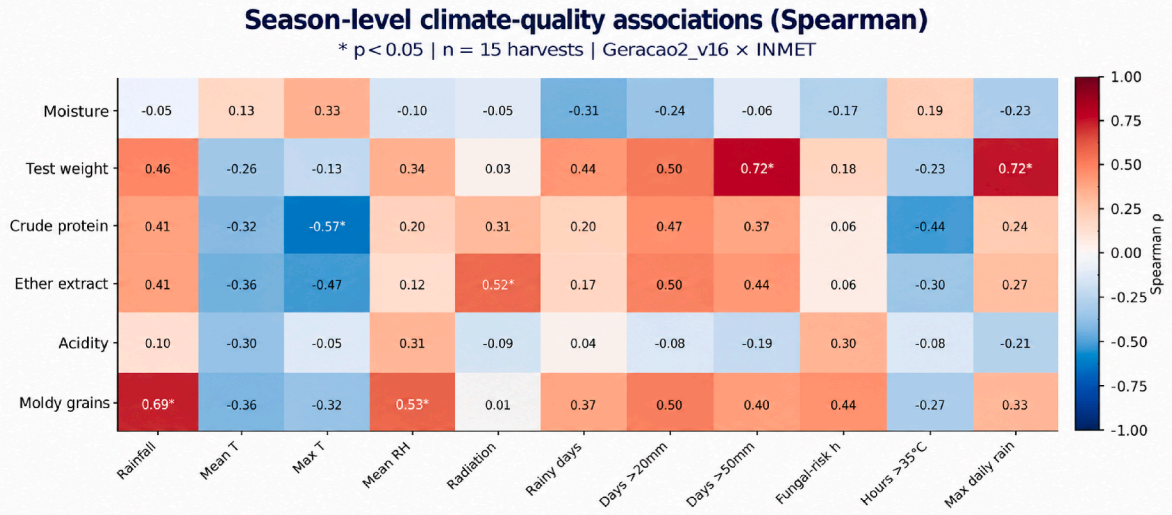


Fig. 6. Partial climatic readability: season-level Spearman associations (asterisks: $p < 0.05$).

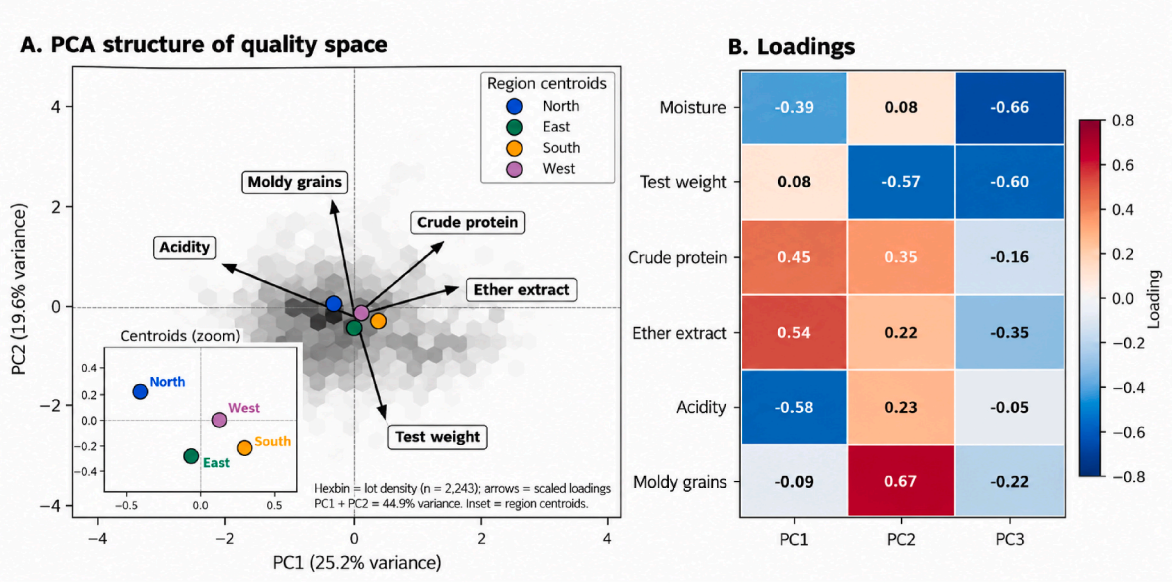


Fig. 7. Multivariate structure of the reception profile. (A) PCA biplot with density and centroid inset. (B) Loadings on PC1–PC3.

with reading the six traits as a coordinated reception profile, without treating PCA as ontological proof of the storage-entry construct.

4. Discussion

4.1. Storage-entry state as an analytical construct

The storage-entry state is the conceptual centre of SPS. It is defined here as an integrated reception profile—an analytical multidimensional construct—not as a latent biological entity observed directly. In this soybean application, six co-measured attributes formed a coherent joint object: PC1 opposed oil/protein to acidity/moisture and PC2 opposed mold to test weight (Fig. 7). That geometry is what the formal definition anticipates—a coordinated physical–compositional–chemical–sanitary profile at pathway entry, not a checklist of independent tolerances (Anukiruthika and Jayas, 2024; Magan et al., 2011). Checklist reading remains necessary for trade. It cannot, by itself, reveal emergent organization across seasons. Treating reception as a state changes the scientific question from "did each attribute pass?" to "what profile begins the postharvest chain, and how is that profile organized in commercial populations?"

4.2. Commercial surveillance as an observational evidence layer

SPS Principle 1–3 reposition commercial laboratories as continuous generators of standardized physical, compositional and sanitary observations. Experiments resolve post-intake mechanisms under designed contrasts (Coradi et al., 2021; Cañizares et al., 2024; Lima et al., 2024). Surveillance can resolve large-scale pattern across years within a monitored network. Neither layer replaces the other. The soybean archive shows what becomes visible only in the surveillance layer: year-conditioning stronger than place-conditioning, unequal temporal structuring among traits, oscillating rather than monotonically drifting baselines, and joint multivariate structure. Those are not findings an intake experiment on a few lots is built to produce. They are findings an observational system can produce when interpreted as SPS rather than as scattered quality paperwork (Bartosik et al., 2023; Ziegler et al., 2021).

4.3. Trait-specific seasonal signal

Surveillance variables differ in explained temporal structure (Adj. R^2, ΔR^2_season and partial R^2_season), which is the trait-specific seasonal signal construct—not Shannon information or decision value. In this network, acidity and ether extract carried more seasonal explained variation than moisture. Moisture remains indispensable for storability physics and for contractual single-trait decisions (Ziegler et al., 2021), yet it was an incomplete encoder of year type in this archive. The ranking therefore warns against moisture-only summaries of the reception profile without denying moisture's role (Souza et al., 2023; Lima et al., 2024; Jordan et al., 2025). Other crops may reorder traits; what may travel—subject to independent tests—is the claim that traits differ in temporal signal, not that Mato Grosso ranks are universal.

The seasonal trajectories of crude protein, ether extract and acidity also deserve interpretation from a compositional perspective. Protein and oil are the principal value-bearing constituents of soybean, but their final concentrations are not fixed properties of the crop: they reflect genetic background, environmental conditions and genotype × environment interactions acting during seed development and filling (Rotundo and Westgate, 2009; Assefa et al., 2019; Liu et al., 2022). Environmental conditions, particularly temperature and water availability during seed development, can modify the accumulation and final concentration of major seed constituents, and protein and oil do not necessarily respond in parallel to these environmental drivers (Rotundo and Westgate, 2009; Assefa et al., 2019). Consequently, interannual changes in these traits can generate commercially meaningful compositional differences among harvests without implying progressive deterioration of the grain itself. This interpretation is particularly relevant here because cultivar identity was not sufficiently balanced across seasons to separate genetic turnover from environmental and other components of year-conditioning.

Acidity requires a different chemical interpretation. Unlike protein and ether extract, lipid acidity reflects the free-fatty-acid fraction and is therefore related to lipid hydrolysis and the condition of the lipid matrix rather than simply to the amount of oil present. The multivariate opposition observed here—protein and ether extract toward one side of PC1 and acidity and moisture toward the other—should therefore be interpreted as a population-level compositional pattern at reception, not as evidence that one group of traits chemically caused the other. This distinction is especially important because the archive represents grain at the storage-entry window and may already incorporate heterogeneous pre-reception conditions and routine handling. Accordingly, the pronounced seasonal signal of acidity may provide information complementary to bulk protein and oil measurements, while its underlying biochemical drivers require controlled investigation.

These compositional differences also have practical implications. Protein and oil concentrations are central attributes of soybean commercial and processing value, and environmental variation in seed composition can generate differences relevant to protein meal and crude oil yield (Rotundo and Westgate, 2009; Assefa et al., 2019). Long-term surveillance may consequently help processors and commercial laboratories recognize harvests with systematically different raw-material profiles and support decisions on characterization, segregation and processing allocation. Importantly, however, the trajectories in Fig. 3 should not be interpreted as evidence of a sustained decline in soybean composition. Formal trend tests detected no significant monotonic change across the 16 seasonal means, and the archive instead resolved oscillating compositional year types. The surveillance signal is therefore one of interannual compositional reorganization rather than progressive chemical deterioration.

4.4. Partial climatic readability

Climatic readability is defined here as the degree of observational alignment between seasonal reception variation and harvest-window climate, not as validated prediction. In this application, selected climate–trait links suggested partial readability (mold with wetness; oil with radiation; intense-rainfall metrics with test weight), but full-matrix FDR and adjusted domain-level RDA did not support a stable multivariate climate solution at $n = 16$. Acidity—the strongest temporal encoder under P2—lacked stable climate links (Magan et al., 2011; Jaques et al., 2022; Dubal et al., 2023). That dissociation remains theoretically important: if the strongest encoder of year-conditioning is not climate-legible at seasonal scale, climate triage cannot replace laboratory measurement of the state (Coradi et al., 2021; Toomer et al., 2023). SPS Principle 4 therefore treats climate as a readability candidate, not a surrogate state.

4.5. Toward surveillance-based postharvest science

SPS is proposed as a potentially transferable interpretive framework (Fig. 1), not as a programme already validated beyond this network. Empirical inference remains restricted to the monitored Mato Grosso soybean reception archive. What may travel—subject to independent evaluation—are the constructs and questions (storage-entry profile as analytical construct; year-conditioning as a partially unidentified pre-reception bundle; trait-specific seasonal signal; exploratory climatic readability), not the numerical ranks or R^2 values. A compact validation matrix mapping SPS principles to evidence layers and analyses is given in Section 4.6. A broader multi-commodity agenda is left for subsequent work. Predictive algorithms, if used, should remain subordinated to this interpretive layer rather than replace it (Andre et al., 2022; Lima et al., 2024).

4.6. SPS evidence matrix for this application

5. Conclusions

Commercial laboratory archives can be read as structured observational systems for postharvest science when reception traits are treated as a coordinated analytical profile rather than as isolated measurements. Surveillance-based Postharvest Science is offered as a potentially transferable interpretive framework for that reading; the Mato Grosso soybean series (16 seasons; 2562 lots) provides an initial empirical application, not an independent multi-system validation. Within this monitored network, the reception profile showed stronger seasonal than regional structuring, unequal trait-specific seasonal signals, only partial climatic readability and an oscillating year-type baseline rather than monotonic drift. The lasting claim is the interpretive approach and its constructs; transferability beyond this archive requires independent evaluation.

Table 5
SPS evidence matrix for this application.

Principle	Evidence in this archive	Analysis
Observation	Single-lab NTA protocol continuity; curated coverage across 16 seasons	Sampling frame; missingness audit; Supplementary Table S1
State	Coordinated six-trait profile (not a latent entity)	PCA with bootstrap/ sensitivity; complete-case checks
Organization	Season > region structuring; interaction modest; seasonal dispersions heterogeneous	Mixed models/partial R^2; PERMANOVA-style partitions; PERMDISP
Readability	Selective climate–trait links; weak adjusted multivariate climate fit	Domain RDA; pre-specified Spearman + FDR; nearest-station sensitivity
Integration	Operational reading without replacing experiment	Discussion limits; future multi-commodity agenda

CRediT authorship contribution statement

Maicon Marinho Vieira Araujo: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Paulo Otávio Aldaves dos Santos Guedes:** Writing – original draft, Validation, Software, Methodology, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author(s) used Cursor (Composer) and ChatGPT (OpenAI) to assist with manuscript organization, language editing, formatting, and revision support. After using these tools/services, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Maicon Marinho Vieira Araujo reports administrative support, article publishing charges, and equipment, drugs, or supplies were provided by Federal University of Mato Grosso, Faculdade de Agronomia e Zootecnia. Reports a relationship with that includes:. Has patent pending to. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the Núcleo de Tecnologia em Armazenagem (NTA), Universidade Federal de Mato Grosso (UFMT), and collaborating producers and technical staff in the surveillance network.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jspr.2026.103285.

Data availability

Data will be made available on request.

References

Andrade, P.J., 2016. Qualidade De Grãos De Soja Produzidos Em Mato Grosso Entre 2006 E 2016. Tese (Doutorado Em Agricultura Tropical), MT. Universidade Federal de Mato Grosso, Faculdade de Agronomia e Zootecnia, p. 84.

Andre, G.S., Coradi, P.C., Teodoro, L.P.R., Teodoro, P.E., 2022. Predicting the quality of soybean seeds stored in different environments and packaging using machine learning. Sci. Rep. 12, 8793. https://doi.org/10.1038/s41598-022-12863-5.

Anukiruthika, T., Jayas, D.S., 2024. Mathematical modeling for management of stored-grain ecosystems: approaches, opportunities, and research needs. J. Stored Prod. Res. 106, 102304. https://doi.org/10.1016/j.jspr.2024.102304.

AOAC International, 2023a. AOAC official method 2001.11. Protein (Crude) in animal feed, forage (plant tissue), grain, and oilseeds. Official Methods of Analysis of AOAC INTERNATIONAL, twenty-second ed. Oxford University Press.

AOAC International, 2023b. AOAC official method 920.39. Fat (Crude) or ether extract in animal feed. Official Methods of Analysis of AOAC INTERNATIONAL, twenty-second ed. Oxford University Press.

AOAC International, 2023c. AOAC official method 940.28. Fatty acids (Free) in crude and refined oils. Official Methods of Analysis of AOAC INTERNATIONAL, twenty-second ed. Oxford University Press.

Assefa, Y., Purcell, L.C., Salmeron, M., Naeve, S., Casteel, S.N., Kovács, P., et al., 2019. Assessing variation in US soybean seed composition (protein and oil). Front. Plant Sci. 10, 298. https://doi.org/10.3389/fpls.2019.00298.

Bartosik, R., Rodrigues, A., Alvarez, A., 2023. Grain storage management: current knowledge and research challenges. Cereal Chem. 100, 14–29. https://doi.org/10.1002/cche.10636.

Cañizares, L.C.C., Gaioso, C.A., Timm, N.S., Meza, S.L.R., Ramos, A.H., Oliveira, M., et al., 2024. Influence of broken kernels content on soybean quality during storage. Grain Oil Sci. Technol. 7, 105–112. https://doi.org/10.1016/j.gaost.2024.03.002.

CONAB (Companhia Nacional de Abastecimento), 2025. Acompanhamento Da Safra Brasileira De Grãos: Safra 2024/25, 12th Survey. CONAB, Brasília, DF.

Coradi, P.C., Maldaner, L., Ferreira, E.G.M., Puchetti, N., Lemes, E.M., Teodoro, P.E., 2021. Effect of drying temperature and air flow on the quality of seeds and grains of soybean cultivars. J. Stored Prod. Res. 92, 101797. https://doi.org/10.1016/j.jspr.2021.101797.

Dubal, I.T.P., Coradi, P.C., Bilhalva, N.S., Biduski, B., Lutz, E., Mallmann, C.A., et al., 2023. Monitoring of carbon dioxide and equilibrium moisture content for early detection of physicochemical and morphological changes in soybeans stored in vertical silos. Food Chem. 436, 137721. https://doi.org/10.1016/j.foodchem.2023.137721.

Harris, C.R., Millman, K.J., van der Walt, S.J., et al., 2020. Array programming with NumPy. Nature 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2.

INMET (Instituto Nacional de Meteorologia), 2026. Banco De Dados Meteorológicos (BDMEP). INMET, Brasília, DF. https://bdmep.inmet.gov.br/. (Accessed 6 August 2026).

Jaques, L.B.A., Coradi, P.C., Rodrigues, H.E., Dubal, I.T.P., Padia, C.L., Lima, R.E., Souza, G.A.C., 2022. Post-harvesting of soybean seeds – engineering, processes technologies, and seed quality: a review. Int. Agrophys. 36, 59–81.

Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. Philos. Trans. R. Soc. A 374, 20150202. https://doi.org/10.1098/rsta.2015.0202.

Jordan, J.P., Christ, D., Jordan, R.A., Motomiya, A.V.A., Santos, R.C., Oliveira, F.C., Silva, L.B., Coelho, S.R.M., 2025. Quality of soybeans stored under controlled temperature and relative humidity conditions with heat pump technology: a comparison with conventional bag storage. Cienc. E Agrotecnol 49, e011525. https://doi.org/10.1590/1413-7054202549011525.

Kendall, M.G., 1975. Rank Correlation Methods, fourth ed. Charles Griffin, London.

Lima, R.E., Coradi, P.C., Rodrigues, D.M., Teodoro, P.E., Teodoro, L.P.R., Oliveira, D.P., 2024. Monitoring and predicting the quality of soybeans for different drying and storage technologies on a real scale using sensors and machine learning models. J. Stored Prod. Res. 108, 102386. https://doi.org/10.1016/j.jspr.2024.102386.

Liu, A., Cheng, S.S., Yung, W.S., Li, M.W., Lam, H.M., 2022. Genetic regulations of the oil and protein contents in soybean seeds and strategies for improvement. Adv. Bot. Res. 102, 259–293. https://doi.org/10.1016/bs.abr.2022.03.002.

Magan, N., Medina, A., Aldred, D., 2011. Possible climate-change effects on mycotoxin contamination of food crops pre- and postharvest. Plant Pathol. 60, 150–163. https://doi.org/10.1111/j.1365-3059.2010.02412.x.

Mann, H.B., 1945. Nonparametric tests against trend. Econometrica 13, 245–259.

MAPA, 2022. Portaria SDA no. 532, de 14 de fevereiro de 2022. Regulamento Técnico Definindo O Padrão Oficial De Classificação Da Soja E De Seus Subprodutos.

McKinney, W., 2010. Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference, pp. 56–61.

Rotundo, J.L., Westgate, M.E., 2009. Meta-analysis of environmental effects on soybean seed composition. Field Crops Res. 110, 147–156. https://doi.org/10.1016/j.fcr.2008.07.012.

Seabold, S., Perktold, J., 2010. Statsmodels: econometric and statistical modeling with python. Proceedings of the 9th Python in Science Conference, pp. 92–96.

Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau. J. Am. Stat. Assoc. 63, 1379–1389.

Souza, D.G., Resende, O., Zuchi, J., Ferreira, P.V.A., Mabasso, G.A., 2023. CO2 monitoring associated with the quality of soybean grains stored in different moisture contents. Food Sci. Technol. 43. https://doi.org/10.5327/fst.29023.

Toomer, O.T., Oviedo-Rondón, E.O., Ali, M., Patiño, D., Joseph, M., Frinsko, M., et al., 2023. Current agronomic practices, harvest and post-harvest processing of soybeans (*Glycine max*)—A review. Agronomy 13, 427. https://doi.org/10.3390/agronomy13020427.

USDA-FAS, 2025. Oilseeds: World Markets and Trade. United States Department of Agriculture, Foreign Agricultural Service, Washington, DC.

Virtanen, P., Gommers, R., Oliphant, T.E., et al., 2020. SciPy 1.0: fundamental algorithms for scientific computing in python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

Ziegler, V., Paraginski, R.T., Ferreira, C.D., 2021. Grain storage systems and effects of moisture, temperature and time on grain quality – a review. J. Stored Prod. Res. 91, 101770. https://doi.org/10.1016/j.jspr.2021.101770.