## Supplementary Material 1 — Comparison between predictive variables of common ragweed interference in soybean

### Objective

This supplementary material aims to compare the performance of two variables—plant density and shoot biomass of *Ambrosia artemisiifolia* (common ragweed)—as predictors of soybean yield interference. The goal is to justify the selection of plant density as the main variable in the segmented regression analyses of the main study, while also highlighting the complementary value of biomass in multivariate analyses.

### Method

Both variables were tested as predictors of soybean yield loss through segmented regression (the statistical procedure is described in the Materials and Methods section of the main article), with the objective of identifying critical interference thresholds (ψ). The fitted models were compared based on statistical criteria (adjusted $R^2$, significance, model stability) and agronomic considerations (field applicability and ease of measurement).

### Results and interpretation

Both variables were evaluated as predictors of interference using segmented regression to identify the critical infestation threshold of the weed. Plant density yielded a better statistical fit, with a higher adjusted $R^2$ (0.39), more expressive coefficients, and model significance ($p = 0.012$). Additionally, it represents a more practical and operational variable for field use, being easily measured by growers (Table S1).

On the other hand, although shoot biomass showed lower statistical performance ($R^2 = 0.30$), it was retained in the multivariate analyses due to its functional value, representing a more continuous and physiologically meaningful indicator of competitive pressure exerted by the weed. Biomass was used in the PCA, correlation matrix, multiple regression, and PLS to characterize the agronomic mechanisms of interference.

The final choice of plant density as the primary predictive variable in the segmented regression was thus based on statistical robustness, practical applicability, and alignment with agronomic management objectives.

### Supplementary Conclusion

The selection of plant density as the primary predictor in the segmented regression modeling was based on its statistical robustness, field applicability, and alignment with management objectives. Biomass, in turn, proved to be a valuable complementary variable for functionally characterizing weed interference, particularly in multivariate analyses (PCA, correlation, multiple regression, and PLS).

**Table S1** - Comparison between segmented regression models fitted using common ragweed plant density and shoot biomass as predictors of soybean yield loss.

| Criterion | Model with density | Model with biomass |
|---|---|---|
| Predictor variable | Plant density (plants $m^{-2}$) | Weed shoot biomass (g $m^{-2}$) |
| Measurement scale | Ordinal (experimental levels) | Continuous (g $m^{-2}$) |
| Breakpoint (ψ) | 1 plant $m^{-2}$ | 24.89 g $m^{-2}$ |
| Slope before ψ | +14.25 | +0.66 |
| Slope after ψ | ~ +2.90 (14.25 – 11.35) | ~ +0.05 (0.66 – 0.61) |
| Adjusted $R^2$ | 0.39 | 0.30 |
| Model significance | Significant ($p = 0.012$) | Marginal |
| Practical applicability | High — easily measured in the field | Moderate — requires collection and drying |
| Statistical sensitivity | High — stable model | Moderate — greater variability |
| Use in multivariate analyses | Not directly applied | Yes — used in PCA, correlations, and PLS |

## Supplementary Material 2 – Complementary analyses of agronomic interference mechanisms

This supplementary material compiles three complementary statistical approaches used to investigate the mechanisms by which *Ambrosia artemisiifolia* (common ragweed) interferes with soybean yield. The analyses include: (i) partial least squares regression (PLS), (ii) Random Forest modeling, and (iii) Pearson correlation matrix. The results reinforce the consistency of the multivariate patterns identified, indicating that the number of pods, final soybean plant density, thousand-seed weight, and weed biomass are the main indicators of yield loss.

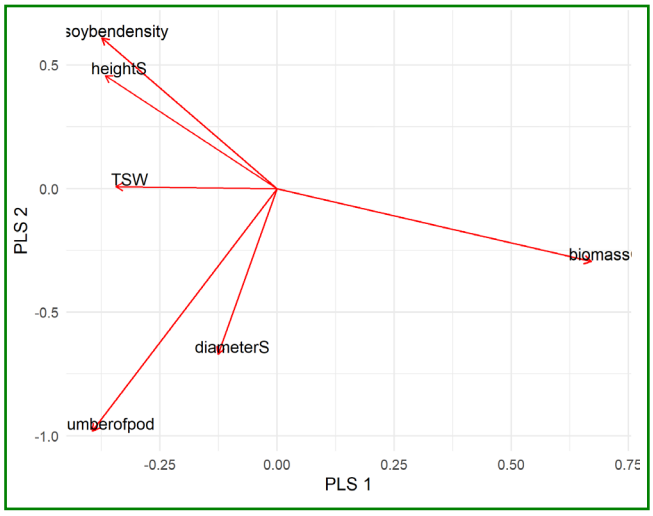


**Figure S1** - Biplot of the partial least squares (PLS) regression based on agronomic variables of soybean and weed attributes. Arrows represent the loadings of each variable in the two principal components. The first two components explained 52% of the variation in soybean yield reduction. Variables with longer arrows contributed more strongly to the latent structure of interference. TSW (weight of a thousand seeds); BiomassC (common ragweed biomass); DiameterS (soybean diameter); Numberfood (number of pods); HeightS (soybean height); Soybendensity (population density)

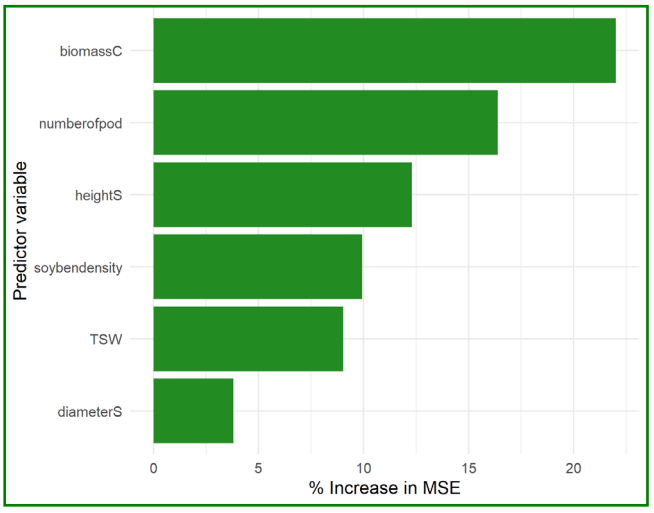


**Figure S2** - Relative importance of predictor variables in the Random Forest model explaining soybean yield reduction. The model identified the number of pods per plant (numberofpod), soybean plant density (soybendensity), thousand seed weight (TSW), and weed biomass (biomassC) as the most relevant predictors. Model performance: $R^2$ = 0.39; RMSE ≈ 14.2%.

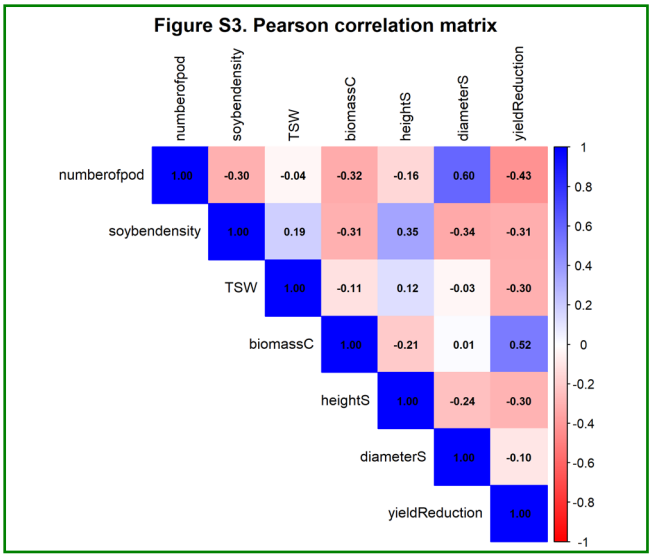


**Figure S3** - Pearson correlation matrix among soybean agronomic traits and weed attributes. Red tones indicate negative correlations; blue tones indicate positive correlations. The variable yieldReduction was negatively correlated with productive components of soybean and positively correlated with weed biomass and diameter. TSW (weight of a thousand seeds); BiomassC (common ragweed biomass); DiameterS (soybean diameter); Numberfood (number of pods); HeightS (soybean height); Soybendensity (population density); YeldReduction (reduced productivity).