

## Assessing coffee yield: predictive modeling based on phenological sensitivity to climate variability

Marcus Vinicius Oliveira Noronha<sup>203</sup>, João Paulo Silva<sup>204</sup>, Marcos Augusto Plachi<sup>205</sup>, Douglas Martins de Santana<sup>206</sup>, Marco Antonio Tavares Rodrigues<sup>207</sup>, Priscilla Rocha Silva Fagundes<sup>208</sup>, Luciana Alvim Santos Romani<sup>209</sup>, Silvia Maria Fonseca Silveira Massruhá<sup>210</sup>, Durval Dourado Neto<sup>211</sup>

### Abstract

This study aimed to validate the agro-meteorological model developed by Santos and Camargo (2006) for predicting Arabica coffee yield, using 14 years of data (2011–2024) from Agrotechnological District (DAT) Caconde and the Volcanic Region of Poços de Caldas. The model, based on climatic variables and phenological sensitivity coefficients (Ky), was tested under 16 scenarios combining genotypes, productivity ranges, and two spatial levels: plot and regional. Results indicated better performance at the regional scale, with lower errors (MAE, RMSE) and high R<sup>2</sup> values (> 0.9), especially in areas with homogeneous productivity. In contrast, plot-level scenarios, such as those for the Bourbon cultivar, showed high variability and lower predictive accuracy. Water deficit was identified as the main factor associated with yield losses. The findings highlight the model's potential for strategic planning at broader spatial scales and the need for local calibration to improve accuracy in heterogeneous environments.

**Keywords:** Agricultural productivity; Agro-meteorological models; Climate variability.

---

<sup>203</sup> Crop Science, Esalq/USP. ORCID: 0009-0004-7935-6380. E-mail: marcus.noronha@usp.br.

<sup>204</sup> Feagri, Unicamp. ORCID: 0000-0002-0695-9111. E-mail: jp\_e@hotmail.com.

<sup>205</sup> Agronomist, Bela Vista da Fumaça farm. E-mail: fbelavista.bv@gmail.com.

<sup>206</sup> Crop Science, Esalq/USP. ORCID: 0000-0002-5158-745X. E-mail: douglas.martinsantana@usp.com.

<sup>207</sup> Crop Science, Esalq/USP. ORCID: 0009-0000-3197-7234. E-mail: marco.a.t.rodrigues@usp.br.

<sup>208</sup> Institute of Agricultural Economics (IEA). ID Lattes: 8466887937334808. E-mail: priscilla@iea.sp.gov.br.

<sup>209</sup> Embrapa Digital Agriculture. ORCID: 0000-0002-7386-3515. E-mail: luciana.romani@embrapa.br.

<sup>210</sup> Embrapa's headquarters. ORCID: 0000-0002-8055-0680. E-mail: silvia.massruha@embrapa.br.

<sup>211</sup> Crop Science, Esalq /USP. ORCID: 0000-0002-8452-8288. E-mail: ddourado@usp.br.

## 1. Introduction

Agricultural modeling has significantly contributed to crop yield forecasting, strategic decision-making, and the development of public policies. In coffee cultivation, although research is less extensive compared to annual crops, the complexity of the coffee plant's phenological cycle presents challenges for building accurate predictive models. This cycle includes phases that are highly sensitive to edaphoclimatic conditions, such as bud formation and induction, flowering, fruit expansion, and grain filling, which directly affect crop yield.

The agro-meteorological model proposed by Santos and Camargo (2006) links climatic variables to phenological stages, offering a valuable tool for estimating yield losses and adjusting management strategies. Recent studies have introduced improvements to the model, including physiological variables, CO<sub>2</sub> fertilization, and the use of diverse meteorological data sources (Verhage et al., 2017; Valeriano et al., 2018; Freitas et al., 2025). Despite advances in machine learning, empirical modeling remains widely used in coffee research.

Therefore, the present study aimed to validate the model developed by Santos and Camargo (2006), using two combinations of water sensitivity coefficients, under different scenarios defined by genotype, productivity class, and spatial scale (plot and regional levels), in the DAT Caconde and the Vulcanic Region of Poços de Caldas.

## 2. Methods

This study analyzed 14 years of Arabica coffee yield data (2011–2024) from local plots at Boa Vista da Fumaça farm (Caconde, SP) and regional data from the Vulcanic Region of Poços de Caldas. Climatic variables (precipitation and temperatures) were obtained from INMET (2009–2024), and the region is classified as Cwa (Köppen), with altitude and terrain favorable for coffee cultivation.

The agro-meteorological model by Santos and Camargo (2006) was used to estimate yield losses due to water deficit (fWD), cold (fCold), and heat stress (fHeat). Calculations included phenological sensitivity coefficients (Ky), temperature penalty functions, and an adjustment factor based on the previous year's productivity (fYaa).

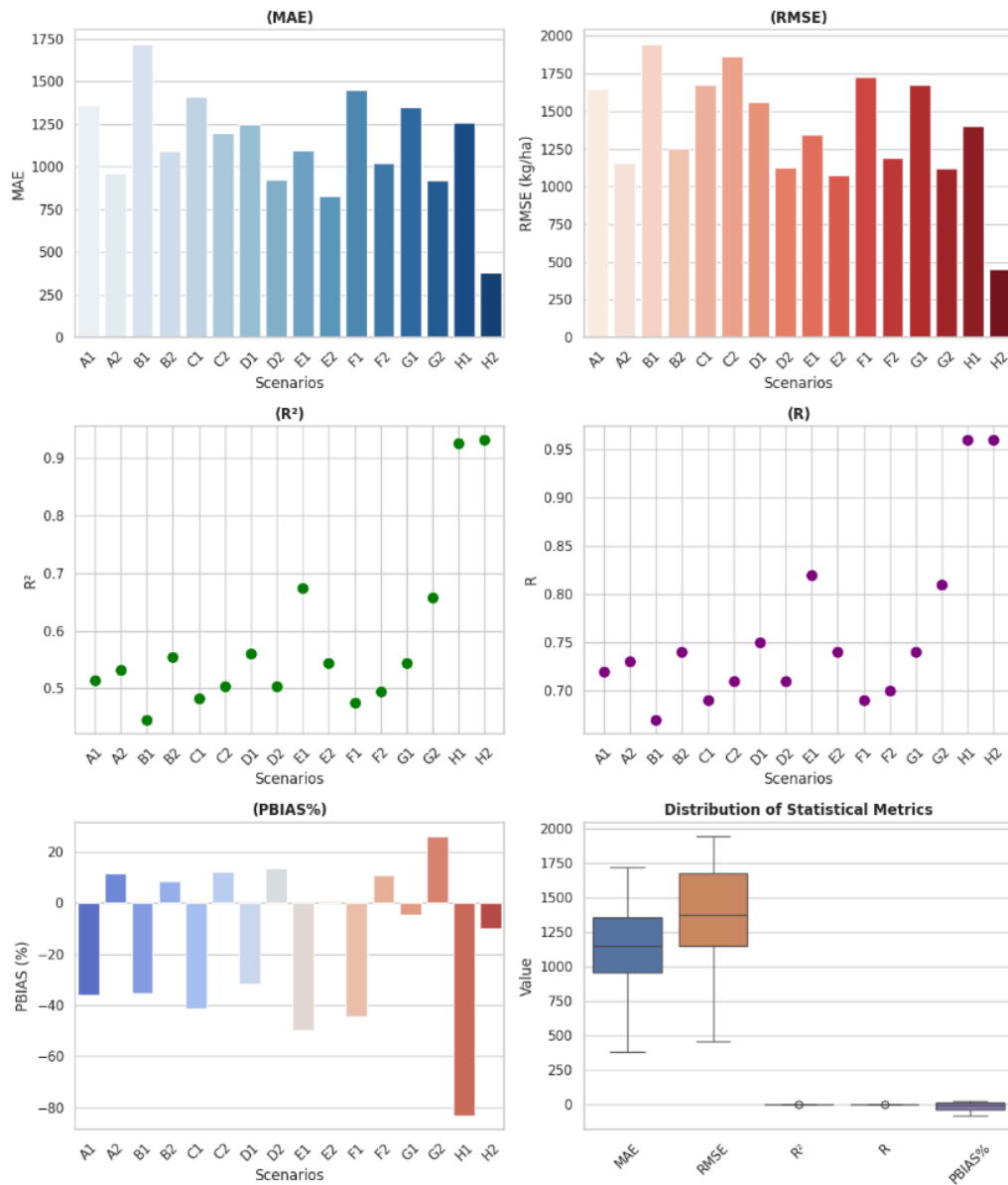
Sixteen validation scenarios were tested at two spatial levels (plot and regional), combining genotypes, productivity classes, and geographic scales. Model performance

was evaluated using MAE, RMSE,  $R^2$ , R, and PBIAS%, after outlier removal based on the interquartile range.

### 3. Results and Discussion

The results demonstrated substantial variations in model performance across the different scenarios evaluated, reflecting the influence of spatial scale, genetic material, and productivity classes on the model's ability to predict coffee yield (Figure 1). The analysis of statistical metrics (MAE, RMSE,  $R^2$ , PBIAS%, and Pearson correlation) showed that regional-level scenarios outperformed plot-level scenarios. This difference is associated with greater productivity homogeneity in regional areas, which tends to smooth out the effects of local variability such as management practices, biennial yield fluctuations, and edaphoclimatic differences (Pereira et al., 2008).

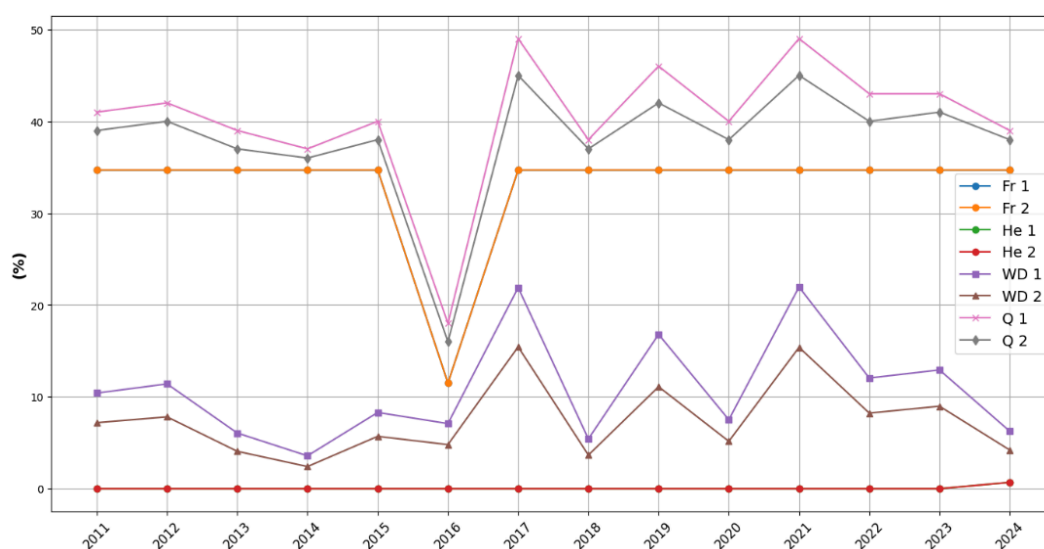
MAE and RMSE values varied considerably across the evaluated scenarios. In specific plots, such as scenario B1 (Bourbon cultivar), the model showed high error rates, with a MAE of 1717.5 kg ha<sup>-1</sup> and RMSE of 1943 kg ha<sup>-1</sup>, indicating substantial deviation between estimated and observed values. In contrast, scenario H2 (regional level) recorded the lowest absolute error values (MAE = 380 kg ha<sup>-1</sup>; RMSE = 453 kg ha<sup>-1</sup>), reflecting greater accuracy. The analysis of the coefficient of determination further reinforced this trend: while some regional scenarios, such as H1 and H2, achieved  $R^2$  values above 0.9, most scenarios scored below 0.6, and scenarios like B1, C1, and F1 showed  $R^2$  values below 0.5, indicating limited explanatory power.



**Figure 1.** Model performance statistics across different scenarios: MAE, RMSE ( $\text{kg ha}^{-1}$ ),  $R^2$ , correlation (R), PBIAS (%), and overall distribution of statistical metrics. The evaluated scenarios include total dataset (A), by cultivar (B: Bourbon, C: Catuaí IAC 144, D: Mundo Novo 376/4), by productivity range (E:  $Y < 1200$ , F:  $1200 \leq Y \leq 1800$ , G:  $Y > 1800 \text{ kg ha}^{-1}$ ), and by the Vulcanic Region (H), at the plot level (1) and regional level (2).

The correlations between observed (Yob) and estimated (Yes) values followed a similar pattern, with coefficients exceeding 0.9 in the best-performing scenarios and falling below 0.7 in those with lower predictive capacity. PBIAS% values also exhibited a wide range, from moderate underestimations to extreme overestimations. Scenario H1, for instance, recorded a PBIAS of -83%, while scenarios such as B2, E2, G1, and H2 showed values within the satisfactory range ( $< \pm 10\%$ ) according to Wu et al. (2019). The weakest performances were observed in plot-level scenarios, underscoring the model's limitations in capturing microenvironmental variability.

Temporal analysis of yield losses revealed that the 2017 and 2021 growing seasons experienced the greatest total losses, largely associated with water deficit (WD) conditions (Figure 2). At the plot level, WD-related losses exceeded 20%, while at the regional level they ranged between 5% and 15%, confirming the greater vulnerability of smaller production units. Losses attributed to high temperatures (He) were nearly negligible throughout the time series, except for a slight increase in 2024 (0.8%), linked to ten-day periods with average temperatures above 24°C (Pereira et al., 2008). Conversely, frost-related losses (Fr) were consistently high, except in 2016, a year with higher minimum temperatures. These losses reflect the crop's sensitivity to minimum temperatures below 2°C, which can cause severe and irreversible damage to plant tissues (Santos; Camargo, 2006; Bussoni et al., 2022).



**Figure 2** Estimated losses (%) due to frost (Fr), high temperatures (He), water deficit (WD), and total losses (Q) at the plot level (1) and regional level (2).

In the cultivar-specific scenarios (B, C, D), the model did not perform satisfactorily for any of the three varieties, with the Bourbon cultivar (B1) standing out negatively due to high error dispersion and an  $R^2$  of approximately 0.44. On the other hand, grouping by productivity ranges (scenarios E, F, and G) resulted in better performance, particularly in scenario E2 (productivity < 1200 kg ha<sup>-1</sup>, regional level), which showed a near-zero bias and low variability in errors. This pattern suggests that homogeneous productivity ranges favor model performance by reducing statistical noise associated with diverse cropping systems.

Previous studies support these findings. Verhage et al. (2017) reported an  $R^2$  of 0.96 and errors around  $\pm 200$  kg ha<sup>-1</sup> after refining sensitivity coefficients and including variables such as CO<sub>2</sub> enrichment. Freitas et al. (2025) highlighted that incorporating root growth curves, leaf area index, and crop coefficients contributed to error reduction, although  $R^2$  values remained between 0.55 and 0.65. Similarly, Valeriano et al. (2018) demonstrated the model's viability using different meteorological data sources (ECMWF, NASA, and local stations), with error rates comparable to those observed in scenario H2, reinforcing the model's applicability in data-scarce environments.

Ultimately, the integrated analysis of performance metrics shows that the model's effectiveness is directly dependent on the scale of application, data homogeneity, and proper parameterization of sensitivity coefficients (Ky). The higher efficiency observed at the regional level highlights the potential of this modeling approach as a strategic management tool in coffee production, provided it is adequately tailored to local conditions. The heterogeneity in performance across scenarios underscores the importance of calibration by region, cultivar, or productivity class as an essential strategy to enhance predictive reliability, particularly in the face of climate change challenges.

#### **4. Conclusion**

The model demonstrated good predictive capacity when applied to regional-level data using Ky series calibrated at the regional scale. However, when applied to local-level data, the model did not yield satisfactory results with the parameter combinations used.

#### **Acknowledgements**

The authors would like to thank Fapesp (Process 2022/09319-9 [2023/11113/2]) for the funding.

## References

BUSSONI, C. V. A.; MOREIRA, D. S.; MACHADO, J. P. Avaliação do modelo WRF para aplicação de um índice de previsão de geada na Região Sul do Brasil. **Revista Brasileira de Meteorologia**, v. 37, n. 3, p. 279-287, July/Sept. 2022. DOI: <https://doi.org/10.1590/0102-77863730084>.

FREITAS, C. H. de; COELHO, R. D.; COSTA, J. de O.; SENTELHAS, P. C. Equationing arabica coffee: Adaptation, calibration, and application of an agrometeorological model for yield estimation. **Agricultural Systems**, v. 223, 104181, Feb. 2025. DOI: <https://doi.org/10.1016/j.agsy.2024.104181>.

PEREIRA, A. R.; CAMARGO, A. P. de; CAMARGO, M. B. P. de. **Agrometeorologia de cafezais no Brasil**. Campinas: Instituto Agronômico, 2008.

SANTOS, M. A. dos; CAMARGO, M. B. P. de. Parametrização de modelo agrometeorológico de estimativa de produtividade do cafeeiro nas condições do Estado de São Paulo. **Bragantia**, v. 65, n. 1, p. 173-183, 2006. DOI: <https://doi.org/10.1590/S0006-87052006000100022>.

VALERIANO, T. T. B.; ROLIM, G. de S.; APARECIDO, L. E. de O.; MORAES, J. R. da S. C. de. Estimation of coffee yield from gridded weather data. **Agronomy Journal**, v.110, n. 6, p. 2462-2477, 2018. DOI: <https://doi.org/10.2134/agronj2017.11.0649>.

VERHAGE, F. Y. F.; ANTEN, N. P. R.; SENTELHAS, P. C. Carbon dioxide fertilization offsets negative impacts of climate change on Arabica coffee yield in Brazil. **Climatic Change**, v. 144, p. 671-685, 2017. DOI: <https://doi.org/10.1007/s10584-017-2068-z>.

WU, J.; CHEN, X.; YU, Z.; YAO, H.; LI, W.; ZHANG, D. Assessing the impact of human regulations on hydrological drought development and recovery based on a 'simulated-observed' comparison of the SWAT model. **Journal of Hydrology**, v. 577, 123990, Oct. 2019. DOI: <https://doi.org/10.1016/j.jhydrol.2019.123990>.