Adaptability and Stability of Coffea Canephora to Dynamic Environments Using Bayesian Approach

Fabio Luiz Partelli

Universidade Federal do Espírito Santo, Centro Universitário Norte do Espírito Santo

Flavia Alves Silva (✉ flavia_uems@hotmail.com)

Universidade Estadual do Norte Fluminense Darcy Ribeiro

André Monzoli Covre

Universidade Federal do Espírito Santo, Centro Universitário Norte do Espírito Santo

Gleison Olios

Universidade Federal do Espírito Santo, Centro Universitário Norte do Espírito Santo

Caio Cezar Guedes Correa

Universidade Federal do Espírito Santo

Alexandre Pio Viana

Universidade Estadual do Norte Fluminense Darcy Ribeiro

Research Article

Keywords: prior, bi-segmented regression, GxE interaction, Conilon

Posted Date: March 3rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1370632/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Adaptability and stability of *Coffea canephora* to dynamic environments using Bayesian approach

Fabio Luiz Partelli¹, Flavia Alves da Silva^{2*}, André Monzoli Covre¹, Gleison Oliosi¹, Caio Cezar Guedes Correa³, Alexandre Pio Viana²

¹Universidade Federal do Espírito Santo, Centro Universitário Norte do Espírito Santo, 29932-540, São Mateus, Espírito Santo, Brasil.

²Universidade Estadual do Norte Fluminense Darcy Ribeiro, Laboratório de Melhoramento Genético Vegetal, 28013-602, Campos dos Goytacazes, Rio de Janeiro, Brasil.

³Universidade Federal do Espírito Santo, Centro de Ciências Agrárias e Engenharias, 29500-000, Alegre, Espírito Santo, Brasil.

Corresponding author: Flavia Alves da Silva; e-mail: flavia_uems@hotmail.com

Abstract

The aim of the work was to use the Bayesian approach, modeling the interaction of coffee genotypes with the environment, using a bisegmented regression to identify stable and adapted genotypes. A group of 43 promising genotypes of *Coffea canephora* was chosen. The genotypes were arranged in a randomized block design with three replications of seven plants each. The experimental plot was harvested four years in the study period, according to the maturation cycle of each genotype. The proposed Bayesian methodology was implemented in the free program R using rstanarm and coda. After fine adjustments in the approach, it was possible to make inferences about the significant GxE interaction, and to discriminate the coffee genotypes regarding production, adaptability and stability.

Keywords: prior, bi-segmented regression, GxE interaction, Conilon.

20 Introduction

21 Worldwide, around 174 million bags of coffee are produced every year. Of this total output, 59.8% is Arabica
 22 (*Coffea arabica*) and 40.2% Robusta/Conilon (*Coffea canephora*) ¹. The entire coffee chain accounts for an annual
 23 revenue of around 172,000 million USD, and Brazil for approximately 32% of the global production ².

24 Conilon coffee is commonly vegetatively propagated, which warrants a uniform crop development and better
 25 fruit quality, with high genetic variability between cultivars. These genotypes differ in their interaction with the
 26 environment, making breeding a challenge ³⁻⁵.

27 The ultimate goal in plant breeding programs is to recommend cultivars for commercial use. To do it is necessary
 28 to model the behavior of genotypes across environments, to get greater confidence and better based-decision about the
 29 recommendations of superior cultivars. This modeling of Genotype x Environment interaction (GE) can be defined as the
 30 differential response of genotypes to environmental variation. Basically, it is necessary to discriminate cultivars as to the
 31 most stable ones, that is, those with less significant GE, or those more adapted to specific environments as more responsive
 32 to a good environment ⁶.

33 The identification of the causal factors of GE and the quantification of these on the residual variation, not
 34 explained by the main effects, are key elements for the understanding of the phenotypic adaptation. In breeding programs
 35 there is a need for decision making, such as selecting strains in multi-environmental trials. Analytical methods linked to
 36 multi-environmental characterization protocols allow useful inferences to decision making, such as the identification of
 37 environmental factors that determine the GE interaction, and mainly the adaptability and stability of genotypes. The
 38 objective is to identify individuals with predictable behavior and who are responsive to environmental variations, under
 39 specific or broad conditions ⁷⁻⁹.

40 There are currently several methodologies available in studies of adaptability and stability, varying in terms of
 41 the concepts employed and statistical principles. All are based on the existence of the GE interaction and the choice
 42 depends on the experimental data, number of available environments, required precision and type of information desired
 43 ¹⁰.

44 The simplest method is based on unfolding the sum of squares of the effects of environments from a joint analysis
 45 on effects of environments within each genotype. Other methods may use some different abstractions of this as the
 46 arithmetic mean of the components of variance ¹¹, and decomposition of the sum of squares in to isolated genotypes ¹².
 47 Later, methods based on linear regression were proposed ^{13,14}, where for each genotype the dependent variable regresses
 48 according to an environmental index defined as the average of all genotypes in the environment. Subsequently, a
 49 correction was proposed due to the implicit errors in the measurements of the genotypes themselves, where the
 50 independent variable was associated with errors, invalidating the estimator ¹⁵. Subsequently it was proposed to evaluate
 51 the genotypes in favorable and unfavorable environments to identify the ideotype, using a double regression ¹⁶.

52 With this approach, there is a principle that the ideotype would be the one with high productive capacity, high
 53 stability, little sensitive to the adverse conditions of unfavorable environments, but capable of satisfactorily responding
 54 to the improvement of the environment. Taking a single regression coefficient in the conventional methodology
 55 considering all environments, this ideotype would be discarded because having a double slope, its deviations, which
 56 should be examined in different environments, could be relatively high in relation to the estimated line. However, this
 57 method is impractical when there is a small number of environments in the subgroups. It was then proposed to use a single
 58 regression again, but bisegmented ¹⁷. This method was later extended to a simpler operation and with statistical properties
 59 more suited to the purpose of improvement ¹⁸.

More recently, other methods have also been proposed as Additive Multiplicative Models Interaction – AMMI¹⁹, GGE Biplot²⁰, Extended Centroid Method^{21,22}, non-parametric analysis²³. Although there are many methods available, these procedures have limitations in dealing with unbalanced data, non-orthogonal experiments (incomplete blocks), and heterogeneity of variance between various environments.

Some new approaches help to circumvent these problems, including Intelligence Artificial^{24,25} and based on Bayesian approaches^{26–28}. These approaches are particularly useful to avoid false positives when few environments ($n < 10$) are evaluated. Both approaches can be combined with other techniques (regression), but the Bayesian approaches allows to incorporate additional information about the parameters through prior distributions incorporated to model with a probability distribution. In practice, this information can be obtained using information from previous studies²⁹, improving the model goodness of fitting³⁰. Plant breeders can now gain new insights by leveraging previously published studies, and also usefully leverage information from previous experiments within programs. Other reviews may provide more relevant information on the use of the Bayesian approach^{31,32}.

In coffee plants, as well as other perennial fruit trees worldwide appreciated, situations such as those described above are commonly encountered, such as unbalanced, non-orthogonal data and few data available. Allied to this, is the time needed to obtain data on perennial plants, which make it very difficult to carry out large experiments in different locations. So, we use an approach not yet reported for coffee. The objective was to use the Bayesian approach by modeling the interaction of coffee genotypes with the environment using a bisegmented regression to identify stable and adapted genotypes. We also provide a workflow for implementation in free and open-source R software and prior estimates for breeders worldwide.

Methodology

Plant Material

A group of 43 promising *C. canephora* genotypes was chosen, most of which had been selected by coffee farmers in the State of Espírito Santo – Brazil (Supplementary Table 1). These 42 cutting and one seed-propagated genotypes were propagated again by cuttings and planted in an experimental plot for selection for high yield potential and agronomic traits of interest. The experiment was planted in April 2014, at a spacing of 3.5 m x 1.0 m, totaling 2,857 plants ha⁻¹, in the county of Itabela, Bahia, Brazil (lat 16° 36' 52.00" S, long 39° 30' 33.00" W, alt 140 m asl). According to Köppen's classification, the regional climate is Aw tropical, with hot humid summers and dry winters³³, mean minimum temperatures of over 15 °C (July and/or August) and mean maxima of over 35 °C in January and/or February in some years (Supplementary Figure 1).

The genotypes were arranged in a randomized block design with three replications of seven plants each. The number of orthotropic branches was controlled by pruning, to maintain around 12,000 - 15,000 stalks per hectare. In all experimental years, mechanical and chemical weeding was performed at least once and insecticides and fungicides were applied. The experimental area was drip-irrigated in all evaluation years. Fertilizers (N, P₂O₅ and K₂O) were applied according to the plant requirements and phenological stages (600, 100 and 400 kg ha⁻¹ year⁻¹, respectively).

The experimental plot was harvested four times in the study period (2016, 2017, 2018 and 2019), according to the maturation cycle of each genotype. The plots were harvested separately for each genotype, measuring the production in liters per plot. Later, the individual yield of each genotype was converted into bags of green coffee per hectare, based on two harvests and extrapolated to the others and, based on the plant spacing, yield per hectare was calculated.

Statistical Analyses

The data were subjected to analysis of variance in each environment, and the homogeneity of residual variance was later verified using Hartley's test. Posteriorly was performed a joint variance analysis, by model:

$$Y_{ijk} = \mu + r/e_{k(j)} + e_j + g_i + ge_{ij} + \varepsilon_{ijk} \quad (1)$$

where Y_{ijk} is the phenotypic mean; μ is the general mean; $r/e_{k(j)}$ is the effect of the k^{th} repetition in the j^{th} environment; g_i is the fixed effect of the i^{th} genotype; e_j is the effect of the j^{th} environment Normally Independently Distributed (NID)($0, \sigma_e^2$); ge_{ij} is the effect of the interaction of the i^{th} genotype in the j^{th} environment NID ($0, \sigma_{ge}^2$); and ε_{ijk} is the experimental error NID ($0, \sigma_\varepsilon^2$).

Bayesian Inference

The bisegmented regression model used is:

$$Y_{ij} = \beta_{i0} + \beta_{i1}I_j + \beta_{i2}T(I_j) + e_{ij} \quad (2)$$

where Y_{ij} is the response of genotype i in environment j , β_{i0} is the mean response of genotype i , β_{i1} is the slope under the first regime (the linear regression coefficient related to the unfavorable environments), and β_{i2} represents the change in slope from the first to the second regime ($\beta_{i1} + \beta_{i2}$ is the slope after the change-point, that is, the linear response to the favorable environments). Further, I_j is the coded environmental index, $T(I_j) = 0$ if $I_j > 0$, or $T(I_j) = I_j - \bar{I}_+$ if $I_j > 0$ and $\bar{I}_+$ is mean of the coded environmental index considering only environments with positive indexes and e_{ij} is the error term, NID ($0, \sigma^2$).

The modified bisegmented regression model³⁴ showed can be performed by the Bayesian approach^{6,28}, where the likelihood function and the joint posterior distribution can be founded. In summary, assuming $e_{ij}|I\sigma_{ie}^2 \sim N(0, I\sigma_{ie}^2)$, each observation y_{ij} has a distribution $y_{ij} \sim N(\beta_{i0} + \beta_{i1}I_j + \beta_{i2}T(I_j); I\sigma_{ie}^2)$, and the likelihood function for each genotype is given by:

$$L_i(\beta_{i0}, \beta_{i1}, \beta_{i2}, \sigma_{ie}^2, y_{ij}) = \frac{1}{(\sqrt{2\pi\sigma_{ie}^2})^a} \exp \left\{ -\frac{1}{2\sigma_{ie}^2} \sum_{I_{j \leq 0}} [y_{ij} - \beta_{i0} - \beta_{i1}I_j]^2 - \frac{1}{2\sigma_{ie}^2} \sum_{I_{j > 0}} [y_{ij} - \beta_{i0} - \beta_{i1}I_j - \beta_{i2}(I_j - \bar{I}_+)]^2 \right\} \quad (3)$$

The prior distributions for the parameters ($\beta_{i0}, \beta_{i1}, \beta_{i2}, \sigma_{ie}^2$) are given by:

$$\beta_{i0} | \mu_{\beta_{i0}}, \sigma_{\beta_{i0}}^2 \sim N(\mu_{\beta_{i0}}, \sigma_{\beta_{i0}}^2) \quad (4)$$

$$\beta_{i1} | \mu_{\beta_{i1}}, \sigma_{\beta_{i1}}^2 \sim N(\mu_{\beta_{i1}}, \sigma_{\beta_{i1}}^2) \quad (5)$$

$$\beta_{i2} | \mu_{\beta_{i2}}, \sigma_{\beta_{i2}}^2 \sim N(\mu_{\beta_{i2}}, \sigma_{\beta_{i2}}^2) \quad (6)$$

$$\frac{1}{\sigma_{ie}^2} = \tau_{ie} | \alpha_1, \beta_1 \sim \text{Gamma}(\alpha_i, \beta_i) \quad (7)$$

Where $\mu_{i0}, \sigma_{\beta_{i0}}^2, \mu_{i1}, \sigma_{\beta_{i1}}^2, \mu_{i2}, \sigma_{\beta_{i2}}^2$, and α_i, β_i are the known parameters. This last prior distribution is the Gamma distribution with mean and variance equal to $\frac{\alpha}{\beta}$ e $\frac{\alpha}{\beta^2}$, respectively. Additionally, σ_{ie}^2 the precision is equal to $\frac{1}{\sigma_{ie}^2}$.

The joint posterior distribution is proportional to the product of the likelihood function (eq 3) and the prior distributions (eqs 4-7).

$$P(\beta_{i0}, \beta_{i1}, \beta_{i2}, \tau_{ie} = \frac{1}{\sigma_{ie}^2} | y_{ij}) \propto \frac{1}{(\sqrt{2\pi\sigma_{ie}^2})^a} \exp \left\{ -\frac{1}{2\sigma_{ie}^2} \sum_{I_{j \leq 0}} [y_{ij} - \beta_{i0} - \beta_{i1}I_j]^2 - \frac{1}{2\sigma_{ie}^2} \sum_{I_{j > 0}} [y_{ij} - \beta_{i0} - \beta_{i1}I_j - \beta_{i2}(I_j - \bar{I}_+)]^2 \right\} - \frac{1}{\sqrt{2\pi\sigma_{\beta_{i0}}^2}} \exp \left[-\frac{1}{2\sigma_{\beta_{i0}}^2} (\beta_{i0} - \mu_{\beta_{i0}})^2 \right] * \frac{1}{\sqrt{2\pi\sigma_{\beta_{i1}}^2}} \exp \left[-\frac{1}{2\sigma_{\beta_{i1}}^2} (\beta_{i1} - \mu_{\beta_{i1}})^2 \right] * \frac{1}{\sqrt{2\pi\sigma_{\beta_{i2}}^2}} \exp \left[-\frac{1}{2\sigma_{\beta_{i2}}^2} (\beta_{i2} - \mu_{\beta_{i2}})^2 \right] * \frac{\beta_i^{\alpha_1} \tau_{ie}^{\alpha_1-1} e^{-\beta_i \tau_{ie}}}{\Gamma(\alpha_1)} \quad (8)$$

To make inferences regarding the parameters in eq 8, the Markov chain Monte Carlo (MCMC) was used to obtain the posterior marginal distributions for each parameter. The marginal distribution samples of the stability parameter, σ_{di}^2 , were obtained indirectly. This parameter is a function of σ_{ie}^2 according to the following expression: $\hat{\sigma}_{di}^2 = \hat{\sigma}_{ie}^2 - \frac{MSR}{r}$, where MSR is the residual mean square obtained from the variance analysis and r is the number of repetitions of the experiment.

We adopted four MCMC chains considering 100,000 iterations (nitt) each of the Gibbs sampler algorithm, after carrying out some convergence tests of the chains with a visual inspection of the behavior of the parameter estimates, starting at 10,000 iterations and increasing until stability is obtained. We set the burn-in to 10,000 iterations and thinned every four iterations (thin = 1:4). In each chain, we analyzed the posterior mean, standard deviation, 95% credibility intervals, and convergence criterion statistics according to ^{35,36}. The methodology was implemented in software R ³⁷, and the joint distribution samples were obtained using the rstanarm::stan_glm package::function ³⁸. The MCMC chain convergence at 100,000 iterations was accessed by Geweke's criterion, according to the recommendations of ³⁹, using the coda::gewekediag package::function ⁴⁰.

Prior's distributions

We left from a minimally informative prior distributions as tested in the works described in the literature ⁴¹, which were represented by distributions with large variances:

$$\beta_{i0} | \mu_{\beta_{i0}}, \sigma_{\beta_{i0}}^2 \sim N(\mu_{\beta_{i0}} = 0, \sigma_{\beta_{i0}}^2 = 100000) \quad (9)$$

$$\beta_{i1} | \mu_{\beta_{i1}}, \sigma_{\beta_{i1}}^2 \sim N(\mu_{\beta_{i1}} = 0, \sigma_{\beta_{i1}}^2 = 100000) \quad (10)$$

$$\beta_{i2} | \mu_{\beta_{i2}}, \sigma_{\beta_{i2}}^2 \sim N(\mu_{\beta_{i2}} = 0, \sigma_{\beta_{i2}}^2 = 100000) \quad (11)$$

$$\tau_{ie} | \alpha_i, \beta_i \sim \text{Gamma}(\alpha_i = 0.001, \beta_i = 0.001) \quad (12)$$

Later we also tested another approach used in the literature ^{28,42-44}, that was characterized by the estimates obtained from the frequentist analysis of the bi-segmented model, used as information to define the hyperparameters. But we take advantage of the biggest advantage of the Bayesian approach, which is to be able to incorporate information a priori from previous data, and we use the data described in our previous works ^{4,45-47} to make inferences about hyperparameters in distributions.

The model (hyperparameters) was chosen considering the one that presented more accurate metrics in relation to the chosen prior. The priori choice was justified as suggested by ^{48,49}, with the modification of the alpha.mu and alpha. Vex tension parameters for variances and values to means, based on the lowest Deviance Information Criterion (DIC) value among the models, where $DIC = D(\hat{\theta}) - 2p_D$. Here, $D(\hat{\theta})$ is a point estimate of the deviance obtained by replacing the parameters with their posterior mean estimates in the likelihood function and $2p_D$ is given by the effective number of parameters in the models. A lack of increment in the informative capacity of the priori was considered when $|\Delta| < 2$ between the DIC of the models ⁵⁰. To assist priori decision making, the predictive capacity of the model was also considered, in a cross-validation, with ten folds (90 % training, and 10 % validation). Thus, we considered the mean correlation among the response of genotype i in environment j and phenotypic mean those observed from the folds of the cross-validation.

Results and discussion

The individual phenotypic value, when evaluated in an environment, is the result of the action of the genotypic effect under the influence of the environment to which it is subjected. However, when assessing the same individual in various environments, an additional component that influences its phenotypic value often appears, which is called the interaction between genotypic and environmental effects (GE). This interaction quantifies the different behavior of

genotypes in the face of environmental variations. The GE interaction is one of the biggest problems of breeding programs of any kind, both in the selection and selection phase and in the genotype recommendation stage. In the coffee genotypes evaluated in this work, it was possible to observe that there was a significant GE interaction ($P < 0.001$) for the production of bags of green beans per hectare (Table 1).

TABLE 1 HERE

The causes of the GxE interaction can be attributed to physiological, adaptive factors, related to the measurement scale of variables, among others. Which set of mechanisms allows a genotype to respond better to an environment is a question still at the frontier of science, usually trying to be answered involving molecular techniques, where the Bayesian approach is also gaining prominence. Here, we use this approach only to identify these genotypes and their response with less bias and variance. The breeder, in addition to identifying an approach that can bring better results to his pipeline, needs to adjust his model to the data. In this work, as described in the methodology, we chose a priori from frequentist approaches (Figure 1) to obtain, with the Bayesian approach, estimates with shorter credibility intervals.

Figure 1 Difference between coefficient estimates, credibility intervals (Bayesian approach, left), and confidence intervals (frequentist approach, right). Coefficients estimated by the Bayesian approach were obtained after fit with the function `rstanarm::stam_glm` using priors from frequentist, and by the frequentist approach using the function `stats::glm` on previous experiments.

Note that in this example of any genotype used to illustrate (Figure 1), not only were the credibility intervals much smaller compared to the confidence intervals, but the estimates of some parameters were changed/corrected. This reduction in the limits of the credibility interval very well described when it was proposed^{51,52} has also been evidenced by other studies using this approach with similar models for Stability and Adaptability^{28,41,43,44,53}, but not for coffee cultivation. For example, this genotype would have its negative estimate for the parameter referring to the unfavorable environment, whereas in the Bayesian approach, its estimate was close to zero (probably positive) and the error estimate much lower (parameter related to stability). This in itself is an interesting result, but it is particularly useful when the

breeder is dealing with dynamic environments, where the variance of genotypes can generate wide confidence intervals in the frequentist approach, as highlighted in ⁵⁴. This was just an illustration that we believe will help breeders when choosing their models. The results specific to the genotypes studied here will be discussed throughout the text below.

We would also like to highlight the process of choosing the number of iterations of the MCMC chains. As this step may require additional time to adjust the model, due to the computational power required for so many calculations, we illustrate here an example of chains from our starting point and the final number of iterations until stability in the chains is obtained (Figure 2).

Figure 2 Visual inspection of chain convergence in four independent chains to take posterior mean of each coefficient in a Bayesian Bi-segmented Regression fitted model. Left intercept coefficient from the model shows problems in convergence using only 10 thousand iterations. In right, the same coefficient in chains of 100 thousand iterations. In both cases, burn-in was disregarded.

It is possible to observe that in chains with only 10,000 iterations, at times some of the four chains fail to estimate for the longest time the same value for the parameter. This is a problem, because with these chains, we seek precisely the marginal distribution, and this stability could be interpreted as a point estimate, which is the opposite. This could also mean an overfit of the model, limiting inferences about culture to the data used here. Better discussions of this can be found at ^{55,56}.

After refining the model, the objective was to access the interactions between genotypes and the environment, since significant interaction between these factors had already been detected (Table 1). The edaphoclimatic conditions of each environment (E) (Supplementary Figure 1), which presents differences in climatic effects year by year, such as rainfall and temperature, can be the mainly font of alterations in the environment factor that affect genotypes responses. The extreme cases are the temperatures higher than 37 °C, that can be tolerate by *C. canephora*, using a molecular mechanism of maintenance of photoprotection and antioxidant ^{57,58} reducing grain weight and yield ⁵⁹. Mean temperatures lower than 17 °C and higher than 31 °C, also affect the growth of *C. canephora* ^{60,61}. These dynamic environments are a challenge, because although the criteria for testing cultivation value and use in the process of launching new cultivars generally consider some environments, but not all are tested ⁶². This puts the farmer in a situation where he depends on an often-inaccurate recommendation for his specific region.

An alternative to mitigate the influences of the GxE interaction is to recommend/use genotypes with wide adaptability and good stability. Obviously, the most productive genotypes are preferred, aggregated with the characteristics already mentioned. Some coffee genotypes were studied to tolerance characteristics related to environment ^{45,61,63–65}. These studies have shown that they have genetic variability allowing margin to identify more stable genotypes, and that they respond well to favorable environments, or even are able to tolerate more extreme conditions.

We gathered information about the average production of green grain bags per hectare and the $\hat{\sigma}_{di}^2$ parameter, together with the coefficient of determination of the model (R^2) for *C. canephora* genotypes in order to highlight the most productive genotypes, which may have a genetic basis that allows more stability (Figure 3). Where, easily a farmer could choose a genotype that would meet the desired level of production, which had good stability, with some level of confidence in the information.

It was possible to observe that the production average in the evaluated genotypes is between 96.8 and 101.1 bags·ha⁻¹, being possible to choose several genotypes with green grain bags yield above the average. However, when observing the stability of these genotypes that have production mean above the overall mean (Figure 3, in the right half), it is noticed that only eight of these genotypes have good stability (represented by green bars). Thus, the genotypes Z21 (136.4 bags·ha⁻¹), Peneirão (140.1 bags·ha⁻¹) and Ouro negro 2 (138.1 bags·ha⁻¹), ranked from the most stable to the least stable, can be chosen by farmers as the most stable and most productive (more than one standard deviation above the overall mean).

When observing the model determination coefficients (annotations above bars in the right side of Figure 3), we found that the adjusted model has a good ability to describe our data for most genotypes. This may be an indication that the information regarding the estimates of the model parameters may have a low bias when considered by farmers for decision making and choice of genotypes, with a majority of R^2 higher than 0.90. Only four genotypes showed R^2 below 0.9 (700, CH1, Verdim D and Clementino), but since none showed good stability, nor were they among the most productive, they were not recommended in any way. With the exception of the Verdim D genotype, which showed good adaptability, but still was not recommended by good production, as it had low stability, with other genotypes that met the same requirements.

250

251

252

253

Figure 3 $\hat{\beta}_0$ (harvest average, at left) and $\hat{\sigma}^2_{di}$ (stability in the bars, with values of R^2 of model in each bar, at right) estimates in Z-Score Normalization for *C. canephora* genotypes. These parameters were taken from a bisegmented Bayesian regression, modeled for describe adaptability and stability. In green, values referring to the estimate of the parameter above the average, and in red, estimates below the general average.

The LB1 genotype was the most productive (143.4 bags·ha⁻¹), but showed low stability. But the farmer can still consider his adaptability, how responsive he can be to a favorable environment. The adaptability estimates according to the model used were plotted to show the genotypes that responded positively to the favorable environment (Figure 4).

Figure 4 The slope between β_{i1} and β_{i2} , the adaptability to the favorable environment for *C. canephora* genotypes. These parameters were taken from a bisegmented Bayesian regression, modeled for describe adaptability and stability. In green, straight referring to positive response to favorable environment, and in red, straight of genotypes that do not respond to the favorable environment. The slope makes reference to how much a genotype can respond to the favorable environment.

It is possible to observe that the LB1 genotype is the twelfth among the genotypes that best respond to the favorable environment. Now, the farmer must decide whether the difference between his production and the production between the genotypes that most respond to a favorable environment compensates for his low stability. The Z21 genotype is the one that most responds to the favorable environment (Figure 4, the straight with the highest slope in β_{i2}). However, the difference between the production of genotypes LB1 and Z21 is only ~ 5 %. So, it doesn't pay to get the most productive genotype, if there is another among the top5 that responds much more to the favorable environment. However, it is necessary to note that the Z21 genotype is also the genotype whose production is severely impaired when subjected to an unfavorable environment (Figure 4, the line with the lowest point in β_{i1}). This would be a recommended genotype for producers with a high level of technology and who will need more inputs for their environment.

Some genotypes showed low adaptability to the favorable environment (red lines on the slope between β_{i1} and β_{i2}), and although some of these genotypes may have a production above the general average, such as Bamburral (108.4 bags·ha⁻¹), it may not be interesting for the farmer to use this genotype, because, for example, if the precipitation is greater than expected, the genotype may not take advantage of this extra resource. On the other hand, there are other genotypes that also have an average production above the general average, with stability and that do not respond as well to the favorable environment, but can capture part of this type of resource.

Interestingly, it is possible to observe in practice a breeding program corroborating our data. For example, we do not recommend the Bamburral because it does not meet all characteristics, but this genotype has some interesting characteristics. It was used among five others to donate genetic material to compose one of the last cultivars launched, called “Tributun”, registered as number 37808 by the National Registry of Cultivars (Registro Nacional de Cultivares, RNC) by the Brazilian Ministry of Agriculture, Livestock and Food Supply (Ministério da Agricultura, Pecuária e Abastecimento, MAPA). This breeding process uses clones discovered by the farmers themselves, evaluated in the north of Espírito Santo state, Brazil ⁶².

Among the stable genotypes, with good production in the unfavorable environment and which are also able to respond to improvements in the environment are cultivars 18 (105.7 bags·ha⁻¹), Z18 (108.9 bags·ha⁻¹), and mainly Ouro negro 2 (138.1 bags·ha⁻¹) and Peneirão (140.1 bags·ha⁻¹), which are among the most productive genotypes among those evaluated. These genotypes can serve well smaller, less technician farmers and also family agriculture, these farmers competitive production with the market and greater profitability.

Conclusion

It was possible to use previous information on coffee genotypes as prior information on parameter distributions of an Adaptability and Stability model, which allowed obtaining shorter credibility intervals and good evidence of low bias in the model by the determination coefficient. After fine adjustments in the approach, it was possible to make inferences about the significant GxE interaction, and to discriminate the coffee genotypes regarding production, adaptability and stability. This is still a new approach for perennials, and since it allows more accurate estimates it can be advantageous when planning breeding programs.

The most productive and stable genotypes have an average production 2.5 times higher than the less productive genotypes and with low stability. The Z21 genotype is recommended to compose part of selected genetic material for highly technical farmers, as it responds very well to the favorable environment, being one of the most productive and with excellent stability. While still maintaining high production and stability, but with lower adaptability, the Ouro Preto 2 and Peneirão genotypes can add to the cultivated clones serving small farmers with lower technology level, or environments that will not offer many resources for the plants.

References

1. United States Department of Agriculture. Coffee: World Markets and Trade. (2019).
2. National Food Supply Company. Brazilian coffee production statistics. 1–62 (2019).

- 310 3. Partelli, F. L. *et al.* Root system distribution and yield of “Conilon” coffee propagated by seeds or
311 cuttings. *Pesquisa Agropecuaria Brasileira* **49**, 349–355 (2014).
- 312 4. Partelli, F. L. *et al.* Tributun: a coffee cultivar developed in partnership with farmers. *Crop*
313 *Breeding and Applied Biotechnology* **20**, e30002025 (2020).
- 314 5. Silva, L. O. E. *et al.* Root Trait Variability in *Coffea canephora* Genotypes and Its Relation to Plant
315 Height and Crop Yield. *Agronomy* **10**, 1394 (2020).
- 316 6. Oliveira, T. R. A. *et al.* Adaptability and stability evaluation of maize hybrids using Bayesian
317 segmented regression models. *PLOS ONE* **15**, e0236571 (2020).
- 318 7. Teodoro, P. E. *et al.* Adaptability and stability of cotton genotypes regarding fiber yield and quality
319 traits. *Crop Science* **59**, 518–524 (2019).
- 320 8. Mohammadi, R., Sadeghzadeh, B., Ahmadi, M. M. & Amri, A. Biological interpretation of
321 genotype $\times$ environment interaction in rainfed durum wheat. *Cereal Research*
322 *Communications* 1–8 (2020) doi:htdoi.org/10.1007/s42976-020-00056-7.
- 323 9. Krause, M. D. *et al.* Boosting predictive ability of tropical maize hybrids via genotype-by-
324 environment interaction under multivariate GBLUP models. *Embrapa Milho e Sorgo-Artigo em*
325 *periódico indexado (ALICE)* (2020).
- 326 10. Nascimento, M. *et al.* Bayesian segmented regression model for adaptability and stability
327 evaluation of cotton genotypes. *Euphytica* **216**, 30 (2020).
- 328 11. Plaisted, R. L. & Peterson, L. C. A technique for evaluating the ability of selections to yield
329 consistently in different locations or seasons. *American Potato Journal* **36**, 381–385 (1959).
- 330 12. Wricke, G. ERFASSUNG DER WECHSELWIRKUNG ZWISCHEN GENOTYP UND
331 UMWELT BEI QUANTITATIVEN EIGENSCHAFTEN. *Zeitschrift für pflanzenzüchtung* **53**, 266
332 (1965).
- 333 13. Finlay, K. W. & Wilkinson, G. N. The analysis of adaptation in a plant-breeding programme.
334 *Australian journal of agricultural research* **14**, 742–754 (1963).
- 335 14. Eberhart, S. A. & Russell, W. A. Stability parameters for comparing varieties 1. *Crop science* **6**,
336 36–40 (1966).
- 337 15. Tai, G. C. C. Genotypic stability analysis and its application to potato regional trials. *Crop science*
338 **11**, 184–190 (1971).
- 339 16. Verma, M. M., Chahal, G. S. & Murty, B. R. Limitations of conventional regression analysis a
340 proposed modification. *Theoretical and Applied Genetics* **53**, 89–91 (1978).
- 341 17. Silva, J. G. C. da & Barreto, J. N. Aplicação da regressão linear segmentada em estudos da interação
342 genótipo $\times$ ambiente. *Simpósio de Estatística Aplicada à Experimentação Agronômica* **1**, 49–50
343 (1985).
- 344 18. Cruz, C. D., Torres, R. A. de & Vencovsky, R. An alternative approach to the stability analysis
345 proposed by Silva and Barreto. *Revista Brasileira de Genética* **12**, 567–580 (1989).
- 346 19. Gauch Jr, H. G. *Statistical analysis of regional yield trials: AMMI analysis of factorial designs*.
347 (Elsevier Science Publishers, 1992).
- 348 20. Yan, W., Hunt, L. A., Sheng, Q. & Szlavnick, Z. Cultivar evaluation and mega-environment
349 investigation based on the GGE biplot. *Crop Science* **40**, 597–605 (2000).

- 350 21. Rocha, R. B., Muro-Abad, J. I., Araújo, E. F. & Cruz, C. D. Avaliação do método centróide para
351 estudo de adaptabilidade ao ambiente de clones de *Eucalyptus grandis*. *Ciência Florestal* **15**, 255–
352 266 (2005).
- 353 22. Nascimento, M. *et al.* Alteração no método centroide de avaliação da adaptabilidade genotípica.
354 *Pesquisa Agropecuária Brasileira* **44**, 263–269 (2009).
- 355 23. Lin, C.-S. & Binns, M. R. A superiority measure of cultivar performance for cultivar× location
356 data. *Canadian journal of plant science* **68**, 193–198 (1988).
- 357 24. Nascimento, M., Nascimento, A. C. C. & Barroso, L. M. A. RNA - Aplicação em Estudos de
358 Adaptabilidade e Estabilidade Fenotípica. in *Inteligência computacional aplicada ao*
359 *melhoramento genético* 278–291 (2018).
- 360 25. Teodoro, P. E. *et al.* Redes neurais artificiais para identificar genótipos de feijão-caupi
361 semiprostrado com alta adaptabilidade e estabilidade fenotípicas. *Pesquisa Agropecuária*
362 *Brasileira* **50**, 1054–1060 (2015).
- 363 26. Nascimento, M. *et al.* Abordagem bayesiana para avaliação da adaptabilidade e estabilidade de
364 genótipos de alfaça. *Pesquisa agropecuária brasileira* **46**, 26–32 (2011).
- 365 27. Couto, M. F. *et al.* Eberhart and Russel's Bayesian Method in the selection of popcorn cultivars.
366 *Crop Science* **55**, 571–577 (2015).
- 367 28. Nascimento, M. *et al.* Bayesian segmented regression model for adaptability and stability
368 evaluation of cotton genotypes. *Euphytica* **216**, 30 (2020).
- 369 29. Johnson, S. R., Tomlinson, G. A., Hawker, G. A., Granton, J. T. & Feldman, B. M. Methods to
370 elicit beliefs for Bayesian priors: a systematic review. *Journal of clinical epidemiology* **63**, 355–
371 369 (2010).
- 372 30. Lemaire, L., Jay, F., Lee, I., Csilléry, K. & Blum, M. G. B. Goodness-of-fit statistics for
373 approximate Bayesian computation. *arXiv preprint arXiv:1601.04096* (2016).
- 374 31. Beaumont, M. A. & Rannala, B. The bayesian revolution in genetics. *Nature Reviews Genetics*
375 (2004) doi:10.1038/nrg1318.
- 376 32. Xavier, A., Muir, W. M., Craig, B. & Rainey, K. M. Walking through the statistical black boxes of
377 plant breeding. *Theoretical and Applied Genetics* (2016) doi:10.1007/s00122-016-2750-y.
- 378 33. Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. L. M. & Sparovek, G. Köppen's climate
379 classification map for Brazil. *Meteorologische Zeitschrift* **22**, 711–728 (2013).
- 380 34. Cruz, C. D., Torres, R. A. de & Vencovsky, R. An alternative approach to the stability analysis
381 proposed by Silva and Barreto. *Revista Brasileira de Genética* **12**, 567–580 (1989).
- 382 35. Geweke, J. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior
383 Moments. *Bayesian Statistics* 1–30 (1992) doi:https://doi.org/10.21034/sr.148.
- 384 36. Raftery, A. L. & Lewis, S. One long run with diagnostics: implementation strategies for Markov
385 chain Monte Carlo. *Statistical Science* **7**, 475 (1992).
- 386 37. R Core Team. R: A Language and Environment for Statistical Computing. (2020).
- 387 38. Goodrich, B., Gabry, J. & Brilleman, S. rstanarm: Bayesian Applied Regression Modeling via Stan.
388 *R package version* (2019).

- 389 39. Cowles, M. K. & Carlin, B. P. Markov chain Monte Carlo convergence diagnostics: a comparative
390 review. *Journal of the American Statistical Association* **91**, 883–904 (1996).
- 391 40. Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: convergence diagnosis and output analysis
392 for MCMC. *R news* **6**, 7–11 (2006).
- 393 41. Oliveira, T. R. A. *et al.* Adaptability and stability evaluation of maize hybrids using Bayesian
394 segmented regression models. *PLOS ONE* **15**, e0236571 (2020).
- 395 42. Teodoro, P. E., Nascimento, M., Torres, F. E., Barroso, L. M. A. & Sagrilo, E. Perspectiva
396 bayesiana na seleção de genótipos de feijão-caupi em ensaios de valor de cultivo e uso. *Pesquisa*
397 *agropecuária brasileira* **50**, 878–885 (2015).
- 398 43. Nascimento, M. *et al.* Abordagem bayesiana para avaliação da adaptabilidade e estabilidade de
399 genótipos de alfafa. *Pesquisa agropecuária brasileira* **46**, 26–32 (2011).
- 400 44. Couto, M. F. *et al.* Eberhart and Russel’s Bayesian Method in the selection of popcorn cultivars.
401 *Crop Science* **55**, 571–577 (2015).
- 402 45. Martins, M. Q. *et al.* Adaptability and stability of *Coffea canephora* genotypes cultivated at high
403 altitude and subjected to low temperature during the winter. *Scientia Horticulturae* **252**, 238–242
404 (2019).
- 405 46. Partelli, F. L. *et al.* Microclimate and development of “Conilon” coffee intercropped with rubber
406 trees. *Pesquisa Agropecuária Brasileira* **49**, 872–881 (2014).
- 407 47. Araújo, A. V., Partelli, F. L., Oliosi, G. & Pezzopane, J. R. M. Microclimate, development and
408 productivity of robusta coffee shaded by rubber trees and at full sun. *Revista Ciência Agrônômica*
409 **47**, 700–709 (2016).
- 410 48. Hadfield, J. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm
411 R package. *Journal of Statistical Software* **33**, 1–22 (2010).
- 412 49. Hadfield, J. Package ‘MCMCglmm.’ (2019).
- 413 50. Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A. Bayesian measures of model
414 complexity and fit. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **64**,
415 583–639 (2002).
- 416 51. Resende, M. D. V. Inferencia Bayesiana e simulação estocástica (amostragem de Gibbs) na
417 estimação de componentes de variância e de valores genéticos em plantas perenes. *Embrapa*
418 *Florestas-Docmentos (INFOTECA-E)* (2000).
- 419 52. Gamerman, D. & dos Santos Migon, H. *Inferência estatística: uma abordagem integrada*. (Instituto
420 de Matemática, Universidade Federal do Rio de Janeiro, 1993).
- 421 53. Teodoro, P. E., Nascimento, M., Torres, F. E., Barroso, L. M. A. & Sagrilo, E. Perspectiva
422 bayesiana na seleção de genótipos de feijão-caupi em ensaios de valor de cultivo e uso. *Pesquisa*
423 *agropecuária brasileira* **50**, 878–885 (2015).
- 424 54. Bernardo Júnior, L. A. Y. *et al.* AMMI Bayesian Models to Study Stability and Adaptability in
425 Maize. *Agronomy Journal* **110**, 1765–1776 (2018).
- 426 55. Papastamoulis, P. Overfitting Bayesian mixtures of factor analyzers with an unknown number of
427 components. *Computational Statistics & Data Analysis* **124**, 220–234 (2018).

56. van Havre, Z., White, N., Rousseau, J. & Mengersen, K. Overfitting Bayesian Mixture Models with an Unknown Number of Components. *PloS ONE* **10**, e0131739 (2015).
57. Martins, M. Q. *et al.* Protective response mechanisms to heat stress in interaction with high [CO₂] conditions in coffea spp. *Frontiers in Plant Science* **7**, (2016).
58. Rodrigues, W. P. *et al.* Long-term elevated air [CO₂] strengthens photosynthetic functioning and mitigates the impact of supra-optimal temperatures in tropical Coffea arabica and C. canephora species. *Global change biology* **22**, 415–31 (2016).
59. Ramalho, J. C. *et al.* Stress cross-response of the antioxidative system promoted by superimposed drought and cold conditions in Coffea spp. *PLoS ONE* **13**, (2018).
60. Partelli, F. L., Vieira, H. D., Silva, M. G. & Ramalho, J. C. Seasonal vegetative growth of different age branches of conilon coffee tree. *Semina: Ciências Agrárias* **31**, 619–626 (2010).
61. Covre, A. M., Partelli, F. L., Bonomo, R., Braun, H. & Ronchi, C. P. Vegetative growth of Conilon coffee plants under two water conditions in the Atlantic region of Bahia State, Brazil. *Acta Scientiarum. Agronomy* **38**, 535 (2016).
62. Partelli, F. L. *et al.* Tributun: a coffee cultivar developed in partnership with farmers. *Crop Breeding and Applied Biotechnology* **20**, e30002025 (2020).
63. Gomes, W. R. *et al.* Genetic diversity of standard leaf nutrients in Coffea canephora genotypes during phenological phases. *Genetics and molecular research : GMR* **15**, (2016).
64. Dubberstein, D., Partelli, F. L., Machado Dias, J. R. & Espindula, M. C. Influence of fertilization on coffee trees vegetative growth in west south Amazon. *Coffee Science* **12**, 50–59 (2017).
65. Giles, J. A. D. *et al.* Divergence and genetic parameters between coffea sp. genotypes based in foliar morpho-anatomical traits. *Scientia Horticulturae* **245**, 231–236 (2019).

Acknowledgements

The authors are indebted to the pioneer breeders, i.e., the coffee farmers, who made the first steps in the process of selection of most of the superior genotypes available nowadays. Therefore, the traditional names of the clones were maintained as used among coffee growers. We further acknowledge the support and funds of the Federal University of Espírito Santo (UFES), the Foundation for Research and Innovation Support of Espírito Santo (FAPES), the Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Author contributions

Fabio Luiz Partelli: conceptualization; Flavia Alves da Silva and Caio Cezar Guedes Corrêa: methodology, formal analysis and figures; Flavia Alves da Silva: writing original draft preparation; André Monzoli Covre and Gleison Oliosio: investigation, data curation; Alexandre Pio Viana and Fabio Luiz Partelli: supervision, writing, reviewing and editing.

Data availability

The full phenotypic information, breeding values, scripts and chains generated used in this study, have been submitted at the Open Science Framework and was awarded the public DOI identifier: <https://doi.org/10.17605/OSF.IO/D8T2R>.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Figure legends

Figure 1 Difference between coefficient estimates, credibility intervals (Bayesian approach, left), and confidence intervals (frequentist approach, right). Coefficients estimated by the Bayesian approach were obtained after fit with the function `rstanarm::stam_glm` using priors from frequentist, and by the frequentist approach using the function `stats::glm` on previous experiments.

Figure 2 Visual inspection of chain convergence in four independent chains to take posterior mean of each coefficient in a Bayesian Bi-segmented Regression fitted model. Left intercept coefficient from the model shows problems in convergence using only 10 thousand iterations. In right, the same coefficient in chains of 100 thousand iterations. In both cases, burn-in was disregarded.

Figure 3 $\hat{\beta}_0$ (harvest average, at left) and $\hat{\sigma}_{di}^2$ (stability in the bars, with values of R^2 of model in each bar, at right) estimates in Z-Score Normalization for *C. canephora* genotypes. These parameters were taken from a bisegmented Bayesian regression, modeled for describe adaptability and stability. In green, values referring to the estimate of the parameter above the average, and in red, estimates below the general average.

Figure 4 The slope between $\hat{\beta}_{i1}$ and $\hat{\beta}_{i2}$, the adaptability to the favorable environment for *C. canephora* genotypes. These parameters were taken from a bisegmented Bayesian regression, modeled for describe adaptability and stability. In green, straight referring to positive response to favorable environment, and in red, straight of genotypes that do not respond to the favorable environment. The slope makes reference to how much a genotype can respond to the favorable environment.

Tables

Table 1 Mean squares from joined analysis of variance for yield of 43 coffee genotypes assessed in four environments. Harvest 2016, 2017, 2018 and 2019.

Factor	DF	MS	F value	Pr(>F)
Environments (E)	3	153585	267.216	2E-16***
Blocks	2	546	0.95	0.388
Genotypes (G)	42	6674	11.612	2E-16***
Blocks/Environments	6	790	1.375	0.224
Genotypes x Environments (GxE)	126	1448	2.519	1.7E-11***

	Residuals	336	575
493	*** Significant at 0.001 probability levels by F test.		
494			

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarymaterial.docx](#)