

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
BACHARELADO EM ESTATÍSTICA

Wanessa Bastos de Oliveira Carvalho

**Calibração do modelo de simulação de cultura DSSAT: uma abordagem
através do otimizador bayesiano**

Goiânia
2025



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TRABALHO DE CONCLUSÃO DE CURSO DE GRADUAÇÃO NO REPOSITÓRIO INSTITUCIONAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio do Repositório Institucional (RI/UFG), regulamentado pela Resolução CEPEC no 1240/2014, sem ressarcimento dos direitos autorais, de acordo com a Lei no 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo dos Trabalhos de Conclusão dos Cursos de Graduação disponibilizado no RI/UFG é de responsabilidade exclusiva dos autores. Ao encaminhar(em) o produto final, o(s) autor(a)(es)(as) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do Trabalho de Conclusão de Curso de Graduação (TCCG)

Nome(s) completo(s) do(a)(s) autor(a)(es)(as): Wanessa Bastos de Oliveira Carvalho.

Título do trabalho: Calibração do modelo de simulação de cultura DSSAT: uma abordagem através do otimizador bayesiano.

2. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador) Concorda com a liberação total do documento [x] SIM [] NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante: a) consulta ao(à)(s) autor(a)(es)(as) e ao(à) orientador(a); b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo do TCCG. O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro.

Obs.: Este termo deve ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **David Henriques Da Matta, Professor do Magistério Superior**, em 08/12/2025, às 09:38, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Wanessa Bastos De Oliveira Carvalho, Discente**, em 08/12/2025, às 16:38, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5815870** e o código CRC **01CAF213**.

Referência: Processo nº 23070.059692/2025-31

SEI nº 5815870

Wanessa Bastos de Oliveira Carvalho

**Calibração do modelo de simulação de cultura DSSAT: uma abordagem
através do otimizador bayesiano**

Projeto de trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Estatística da Universidade Federal de Goiás para aprovação no componente curricular TCC, como parte das exigências para a obtenção do título de bacharel em Estatística.

Orientador: David Henriques da Matta

Goiânia
2025

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Carvalho, Wanessa Bastos de Oliveira

Calibração do modelo de simulação de cultura DSSAT [manuscrito] : uma abordagem através do otimizador bayesiano / Wanessa Bastos de Oliveira Carvalho. - 2025.

54, liv f.: il.

Orientador: Prof. Dr. David Henriques da Matta.

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Goiás, Instituto de Matemática e Estatística (IME), Estatística, Goiânia, 2025.

Bibliografia.

Inclui siglas, abreviaturas, símbolos, lista de figuras, lista de tabelas.

1. Aprendizado de máquina. 2. Fenologia. 3. Processo gaussiano. 4. Probabilidade. 5. Função de aquisição. I. Matta, David Henriques da, orient. II. Título.

CDU 519.22



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

ATA DE DEFESA DE TRABALHO DE CONCLUSÃO DE CURSO

Aos vinte e seis dias do mês de novembro do ano de 2025 iniciou-se a sessão pública de defesa do Trabalho de Conclusão de Curso (TCC) intitulado “Calibração do modelo de simulação de cultura DSSAT: uma abordagem através do otimizador bayesiano”, de autoria de Wanessa Bastos de Oliveira Carvalho, do curso de Estatística, do Instituto de Matemática e Estatística da UFG. Os trabalhos foram instalados pelo Prof. Dr. David Henriques da Matta com a participação dos demais membros da Banca Examinadora: Renato Rodrigues Silva (IME/UFG), Márcio Augusto Ferreira Rodrigues (IME/UFG) e Ludmilla Ferreira Justino (Embrapa/GO). Após a apresentação, a banca examinadora realizou a arguição da estudante. Posteriormente, de forma reservada, a Banca Examinadora atribuiu a nota final de 8,3, condicionando a aprovação do TCC à entrega da versão final corrigida, conforme as alterações solicitadas, dentro do prazo estipulado.

Proclamados os resultados, os trabalhos foram encerrados e, para constar, lavrou-se a presente ata que segue assinada pelos Membros da Banca Examinadora.



Documento assinado eletronicamente por **Marcio Augusto Ferreira Rodrigues, Professor do Magistério Superior**, em 26/11/2025, às 22:06, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Renato Rodrigues Silva, Professor do Magistério Superior**, em 27/11/2025, às 14:32, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ludmilla Ferreira Justino, Usuário Externo**, em 27/11/2025, às 15:50, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **David Henriques Da Matta, Professor do Magistério Superior**, em 08/12/2025, às 16:12, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5793982** e o código CRC **EE22A20B**.

Por isso não tema, pois estou contigo; não tenha medo, pois sou o seu Deus. Eu o fortalecerei e o ajudarei; eu o segurarei com a minha mão direita vitoriosa. (Isaías 41:10)

AGRADECIMENTOS

Agradeço primeiramente a Deus que esteve comigo e me ajudou, não só no desenvolvimento desse trabalho de conclusão de curso, mas em todos os momentos da graduação. Agradeço ao meu marido André Luiz Carvalho dos Santos e minha mãe Ana Elpídia Bastos, que com teu seu apoio, se tornaram parte fundamental na finalização deste curso de graduação.

Agradeço ao meu orientador Professor Dr. David Henriques da Matta que me desafiou e ensinou durante todo o projeto. O doutorando Gutemberg Resende Honório Filho que se dedicou e esforçou a me ajudar, assim como o Dr. Alexandre Bryan Heinemann e a Dra. Ludmilla Ferreira Justino.

RESUMO

O presente trabalho buscou mostrar que, através da aplicação da técnica de Otimização Bayesiana (BO), foi possível calibrar os parâmetros em um modelo de simulação de culturas como o DSSAT (*Decision Support System for Agrotechnology Transfer*) de forma eficiente e computacional, sendo uma técnica estatística e de Aprendizagem de Máquina que encontrou de forma robusta os parâmetros do DSSAT através de otimização. Neste trabalho a calibração foi focada em parâmetros do submodelo fenológico CROPGRO-*Drybean* para a cultivar de feijão-comum BRS *Esplendor*, de modo a avaliar e demonstrar a eficiência da técnica de maneira isolada, mantendo todos os demais parâmetros constantes e aplicando a técnica de BO apenas nos parâmetros fenológicos do DSSAT. A utilização da técnica de BO foi feita a partir do software R na função *ParBayesianOptimization* (PBO) e buscou estimar um conjunto de parâmetros cujo objetivo foi minimizar a Função Objetivo, que, neste trabalho, foi dada pelo MAPE (médias dos erros percentuais absolutos). Foi uma técnica adequada quando a otimização era computacionalmente cara e, além disso, destacou-se pela forma inteligente de encontrar os pontos candidatos que compuseram o resultado da otimização, utilizando uma Função de Aquisição que foi alvo de estudo neste trabalho. Os resultados, portanto, visaram mostrar a eficiência da calibração via BO quando comparada com a calibração manual feita por um especialista na área agrônômica e também com a calibração por estimativa separada e simultânea propostas no Protocolo de Calibração para Modelos de Cultura de Solo. Os resultados observaram os box-plots dos resultados obtidos por cada metodologia a fim de identificar a eficiência das calibrações e, desse modo, a otimização bayesiana revelou-se uma técnica adequada à estimação dos hiperparâmetros fenológicos da cultivar BRS *Esplendor* quando comparada com protocolos agrônômicos e calibração manual, pois os resultados foram capazes de otimizar os valores dos hiperparâmetros reduzindo o MAPE (Erro Médio Absoluto Percentual). Isso mostrou que a técnica pôde ser expandida para demais grupos, além da fenologia do feijão-comum.

Palavras-chave: Aprendizado de Máquina. Fenologia. Processo Gaussiano. Probabilidade. Função de Aquisição.

ABSTRACT

The present work sought to show that, through the application of the Bayesian Optimization (BO) technique, it was possible to efficiently and computationally calibrate the parameters of a crop simulation model such as DSSAT (Decision Support System for Agrotechnology Transfer), with BO being a statistical and Machine Learning technique that robustly identified DSSAT parameters through optimization. In this study, the calibration was focused on parameters of the phenological submodel CROPGRO-Drybean for the common bean BRS Esplendor cultivar, in order to evaluate and demonstrate the effectiveness of the technique in isolation, keeping all other parameters constant and applying BO only to the phenological parameters of DSSAT. The use of the BO technique was carried out in the R software through the *ParBayesianOptimization* (PBO) function, which sought to estimate a set of parameters whose objective was to minimize the Objective Function, which, in this study, was given by the MAPE (Mean Absolute Percentage Error). It was an appropriate technique when optimization was computationally expensive and, moreover, it stood out for its intelligent way of selecting candidate points that composed the optimization result, using an Acquisition Function, which was the subject of analysis in this work. The results therefore aimed to show the efficiency of BO-based calibration when compared with manual calibration performed by an agronomic expert and with the separate and simultaneous estimation approaches proposed in the Calibration Protocol for Soil-Crop Models. The results examined the boxplots of the outputs obtained by each methodology in order to identify the efficiency of the calibrations, and, in this way, Bayesian Optimization proved to be an adequate technique for estimating the phenological hyperparameters of the BRS Esplendor cultivar when compared with agronomic protocols and manual calibration, as the results were able to optimize the hyperparameter values by reducing the MAPE (Mean Absolute Percentage Error). This showed that the technique could be expanded to other groups beyond the phenology of common bean.

Key-words: Machine Learning. Phenology. Gaussian Process. Probability. Acquisition Function.

LISTA DE FIGURAS

Figura 1: Estádios do desenvolvimento da planta

Figura 2: Fase de inicialização da função PBO

Figura 3: Esquema gráfico do processo de Amostragem por Hipercubo Latino

Figura 4: Processo da função de aquisição na função PBO

Figura 5: Gráfico da exploração versus exploração em cada função de aquisição

Figura 6: Resumo das funções de aquisição

Figura 7: Dados de inicialização da função optim

Figura 8: Processo de otimização pela função optim

Figura 9: Resultado APE comparativo da saída ADAP

Figura 10: Resultado APE comparativo da saída MDAP

Figura 11: Resultado APE comparativo da saída PD1P

Figura 12: Resultado APE comparativo da saída PDFP

Figura 13: Todos os resultados APE dos hiperparâmetros simulados

Figura 14: Resultado médio das comparações - MAPE

LISTA DE TABELAS

Tabela 1: Descrição dos parâmetros fenológicos.

Tabela 2: Dados de saída do DSSAT.

Tabela 3: Limites em dias fototérmicos dos hiperparâmetros do modelo.

Tabela 4: Combinações da BO que superaram ou são equivalentes ao Step 6

Tabela 5: Quantidade de pontos iniciais e iterações que tornam o BO mais eficaz

Tabela 6: Resultados da PBO para os melhores resultados

Tabela 7: Limites versus valores calibrados via BO

LISTA DE ABREVIATURAS E SIGLAS

UFG	Universidade Federal de Goiás
IME	Instituto de Matemática e Estatística
CSM	<i>Cropping System Model</i>
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
DSSAT	Sistema de Suporte à Decisão para Transferência de Agrotecnologia
CROPGRO	Modelo de crescimento das culturas
EI	Melhoria Esperada
EIPS	Melhoria Esperada por segundo
UCB	Limite Superior de Confiança
POI	Probabilidade de Melhora
GP	Processo Gaussiano
EGO	Efficient Global Optimization
EMFL	Tempo entre a emergência da planta e o aparecimento da primeira flor
FLSH	Tempo entre a primeira flor e a primeira vagem
FLDS	Tempo entre a primeira flor e a primeira semente
SDPM	Tempo entre a primeira semente e a maturidade fisiológica
LHS	Amostragem por Hipercubo Latino
PBO	<i>ParBayesianOptimization</i>
BO	Otimização Bayesiana
RMSE	Raiz do Erro Quadrático Médio
MAPE	Erro Percentual Absoluto Médio
MQO	Mínimos Quadrados Ordinários
MQP	Mínimos Quadrados Ponderados
ADAP	Dias após o plantio até o florescimento
MDAP	Dias após o plantio até a maturação
PD1P	Dias após o plantio até a fase 1 do desenvolvimento das vagens
PDFP	Fase final do desenvolvimento das vagens

LISTA DE SÍMBOLOS

σ	Incerteza do Processo Gaussiano
μ	Média do Processo Gaussiano
%	Porcentagem
ω	Peso do Método de Mínimos Quadrados Ponderados
θ	Parâmetro a ser estimado na inferência bayesiana
<i>MAPE</i>	Erro Percentual Absoluto Médio
A_i	Valor real da métrica no período i,
F_i	Valor previsto pelo modelo no período i
<i>EI</i>	Melhoria Esperada
<i>EIPS</i>	Melhoria Esperada por segundo
<i>UCB</i>	Limite Superior de Confiança
<i>POI</i>	Probabilidade de Melhora
X	Matriz de entrada normalizada
Z	Escore normalizados
V_o	Valor original do dado de entrada
S_o	Dado original do vetor escore
y_{MAX}	Resultado máximo no vetor escore
<i>eps</i>	Parâmetro de entrada para controle de EI, EIPS e POI
$pnorm(z)$	Probabilidade normal padrão no ponto Z
$dnorm(z)$	Densidade da normal padrão até o ponto Z
$P(A B)$	Probabilidade condicional de A dado B
$P(A \cap B)$	Probabilidade conjunta de A e B
$P(B)$	Probabilidade do evento B
$\pi(\theta X)$	Distribuição a posteriori
$L(X \theta)$	Função de verossimilhança
$\pi(\theta)$	Priori de θ
$\pi(X)$	Preditiva de X
SQE_{Grupo}	Soma dos Quadrados dos Erros dos Grupos

SUMÁRIO

INTRODUÇÃO	15
1 FUNDAMENTAÇÃO TEÓRICA	17
1.1 Fundamentos da Otimização Bayesiana.....	17
1.2 Calibração de Modelos Agrícolas e Otimização.....	18
1.3 O Sistema DSSAT e o Modelo CROPGRO	19
1.4 Introdução à fenologia.....	20
1.5. Funcionamento do código.....	22
1.5.1 Fase de Inicialização	23
1.5.2 Fase de Amostragem e Avaliação	27
1.5.3 Fase de Ajuste do Processo Gaussiano	28
1.5.4 Fase da Função de Aquisição e Iteração.....	30
1.5.4.1 UCB (Upper Confidence Bound).....	32
1.5.5 Fase Final	38
1.6 Inferência Bayesiana e o Otimizador Bayesiano.....	38
2 METODOLOGIA	41
3 RESULTADOS	43
Seção 3.1: ADAP	43
Seção 3.2: MDAP.....	44
Seção 3.3: PD1P.....	45
Seção 3.4: PDFP	46
Seção 3.5: Resultados Gerais	47
CONCLUSÃO	52
REFERÊNCIAS	53

INTRODUÇÃO

Modelos mecanicistas, como o Sistema de Apoio à Decisão para Transferência de Agrotecnologia (*Decision Support System for Agrotechnology Transfer - DSSAT*), são amplamente utilizados na agricultura para simular o crescimento, o desenvolvimento e a produtividade das culturas (Jones et al., 2003). Podem ser utilizados para simular o desenvolvimento do feijão-comum (*Phaseolus vulgaris L.*), por exemplo, que compõe a base da segurança alimentar brasileira e é uma das principais fontes de proteína vegetal no mundo (Silva; Wander, 2013). O feijão-comum é comumente consumido em países menos desenvolvidos, pois ajuda no combate à fome e à desnutrição. Além disso, o Brasil se destaca como um dos maiores produtores mundiais de feijão-comum (Faostat, 2019).

O DSSAT é uma plataforma que simula o desenvolvimento das plantas (Jones et al., 2003) de forma matemática e incorpora aspectos biológicos e físicos que ocorrem na natureza para que essa simulação seja possível. O modelo CROPGRO-Drybean, implementado no DSSAT, tem sido extensivamente aplicado para simular aspectos do cultivo de feijão-comum, assim como caracterizar cultivares e oferecer uma análise de riscos climáticos (Heinemann et al., 2017; Oliveira et al., 2012). A previsão dos estádios fenológicos é fundamental para a determinação do impacto das mudanças climáticas nos cultivos de forma a garantir a segurança alimentar em períodos de crise climática, pois permite antecipar eventos críticos no ciclo da cultura (Heinemann et al., 2017).

Apesar de o DSSAT ser uma plataforma robusta, ele requer a aplicação precisa de um processo de calibração acurado. O DSSAT possui um número grande de parâmetros que se relacionam de forma complexa, tornando o modelo de difícil modelagem (Hoogenboom et al., 2012). Os métodos utilizados até o momento são baseados em “tentativa e erro”, em que o pesquisador muda os valores dos parâmetros manualmente com base em suas experiências (Van Oort; Van Ittersum, 2013). Desse modo, torna-se necessário o desenvolvimento de um modelo matemático que promova uma calibração mais robusta, sistemática e automatizada (Wallach et al., 2024).

Uma alternativa para a calibração (processo sistemático de ajuste ou ainda estimação) dos parâmetros do DSSAT é a Otimização Bayesiana, uma ferramenta estatística que correlaciona aprendizado de máquina com maximização de uma função complexa e de alto custo computacional (Snoek et al., 2012). Essa abordagem busca encontrar os valores dos parâmetros que maximizem uma função objetivo, equilibrando exploração e exploração de pontos a partir de uma função de aquisição otimizadora (Jones et al., 2003). Com base no exposto, o objetivo deste trabalho foi avaliar o desempenho do otimizador bayesiano na calibração dos parâmetros fenológicos do modelo CROPGRO-Drybean para a previsão das datas fenológicas.

1 FUNDAMENTAÇÃO TEÓRICA

1.1 Fundamentos da Otimização Bayesiana

A Otimização Bayesiana é uma função estatística que busca encontrar o ponto ótimo global da forma mais eficiente, com o menor esforço possível. Nessa área, se destaca o trabalho de Jones et al. (2003), que apresentou o *Efficient Global Optimization* (EGO), que popularizou o uso do Processo Gaussiano e introduziu pela primeira vez a ideia da função de aquisição Expected Improvement (EI). A EI equaciona a ideia de Exploração (aproveitamento) versus a Exploração (busca), que é utilizada e aperfeiçoada em algumas funções que são a base da otimização bayesiana e, portanto, da função PBO (função principal deste trabalho).

Snoek et al. (2012) aperfeiçoaram a otimização bayesiana, acrescentando a ela o caráter de aprendizado de máquina. Assim, a técnica passou a permitir a calibração dos hiperparâmetros e utilizar a ideia de regressão. Os autores demonstraram que a otimização bayesiana não apenas reduz o custo computacional, mas também obtém excelentes resultados. A otimização bayesiana leva esse nome por conta de seu caráter estatístico Bayesiano, no qual:

- Distribuição a priori: o processo gaussiano (GP) inicial é usado como uma distribuição a priori, pois nela há apenas uma suposição de como é o comportamento da função.
- Distribuição a posteriori: o modelo absorve esses novos pontos, atualizando a distribuição a priori (GP) e tornando o modelo mais robusto. Portanto, a distribuição a posteriori corresponde ao processo gaussiano ajustado.
- Função de atualização: a função de aquisição é utilizada para atualizar a função a priori, guiando o processo de forma inteligente em busca de pontos promissores através da otimização e da maximização da utilidade. Os pontos promissores são absorvidos pelos resultados, e assim o Processo Gaussiano é ajustado.

Desse modo, pode-se concluir que a Otimização Bayesiana é uma boa opção para a calibração dos parâmetros DSSAT, os quais são muito caros

computacionalmente e requerem um bom acerto, devido à sua importância para a confiabilidade das simulações (Snoek et al., 2012; Wallach et al., 2024). O uso da OB, portanto, contribui significativamente para a otimização e a melhoria das previsões dos sistemas agrícolas (Hoogenboom et al., 2012).

1.2 Calibração de Modelos Agrícolas e Otimização

Cultivar é o termo agrônômico que denota um grupo de plantas com características desejáveis. Cada cultivar possui atributos próprios, tanto morfológicos (tamanho, coloração, etc.) quanto fisiológicos e agrônômicos (resistência a doenças, produtividade, adaptação ao solo ou clima, etc.) (Storck, 1993; Faria et al., 2009). Essas diferenças tornam necessária a calibração do modelo de simulação que as utilizará, pois, se o modelo matemático que propõe simular todo o processo dessa planta não possuir tal calibração, torna-se inadequado para prever com eficácia o processo (Boote et al., 1996; Andrade et al., 2009).

Enquanto a cultivar trata da variedade agrícola que é plantada (testada e nomeada), o genótipo é o conjunto de informações genéticas de uma planta. Toda cultivar tem um genótipo, ou um conjunto deles, mas nem todo genótipo é um cultivar. Os valores dos hiperparâmetros que são encontrados na otimização bayesiana desempenham o papel de calibrar o modelo para o determinado genótipo de feijão-comum, que nesse trabalho será o genótipo BRS *Esplendor*. A calibração de modelos é discutida por Van Oort e Van Ittersum (2013). Os autores apresentam técnicas estatísticas e mostram que a calibração é essencial para que o modelo seja confiável, ou seja, impacta diretamente a confiabilidade das previsões, principalmente quando tais previsões avaliam as mudanças climáticas.

Os métodos historicamente utilizados para estimar os parâmetros do DSSAT são frequentemente baseados em tentativa e erro (trial-and-error), nos quais o pesquisador insere valores de forma manual com base em sua experiência e conhecimento prévio do cultivo (Hoogenboom et al., 2012). No entanto, essa abordagem é dispendiosa, subjetiva e não garante a convergência para o ótimo global, impactando a confiabilidade das previsões (Van Oort; Van Ittersum, 2013).

Esta limitação justifica a necessidade de métodos automatizados e estatisticamente robustos, como a Otimização Bayesiana (Wallach et al., 2024).

1.3 O Sistema DSSAT e o Modelo CROPGRO

O DSSAT é uma das plataformas de simulação agrícola mais difundidas pelo mundo. Ele possui uma série de modelos de culturas, módulos de solo, clima e manejo, permitindo a análise de cenários agrícolas (Hoogenboom et al., 2012). O sistema foi inicialmente desenvolvido para leguminosas, como o feijão.

Hoogenboom et al. (2012) exploram os estudos a respeito do DSSAT e diz que ele não é um único programa, mas sim um sistema com design modular e essa modularidade permite a criação do Cropping System Model (CSM) que oferece componentes separados para lidar com o solo, clima e a cultura. A partir do CSM é possível especializar um estudo considerando apenas um submodelo do DSSAT, e isso é fundamental para este trabalho, pois o foco desse estudo é apenas no submodelo fenológico do CROPGRO (Crop Growth Model) – Drybean, modelo de crescimento das culturas do feijão-comum.

Segundo Boote et al. (1998), o CROPGRO é um modelo processual baseado em equações que descrevem os principais processos fisiológicos do crescimento vegetal, como fotossíntese, balanço hídrico, partição de biomassa e desenvolvimento fenológico. A forma de modelar do CROPGRO utiliza parâmetros fisiológicos e ambientais que conferem ao modelo uma flexibilidade ao simular diversos tipos de cultivares e diversas condições de cultivo.

Mais uma vez faz-se necessário apontar que a calibração dos parâmetros é fundamental para mitigar os desvios inseridos na tentativa de calibração manual. Os valores dos parâmetros podem interferir significativamente no resultado da simulação, tornando essa plataforma muito sensível a erros. Hoogenboom et al. (2012) explicam que a estimação dos parâmetros é geralmente feita a partir de medições de ensaios de campo, tornando válida a necessidade de desenvolver métodos mais eficientes para essa tarefa. Portanto, a utilização de uma metodologia

eficiente de calibrar o modelo como a Otimização Bayesiana, tornam os resultados do DSSAT mais confiáveis.

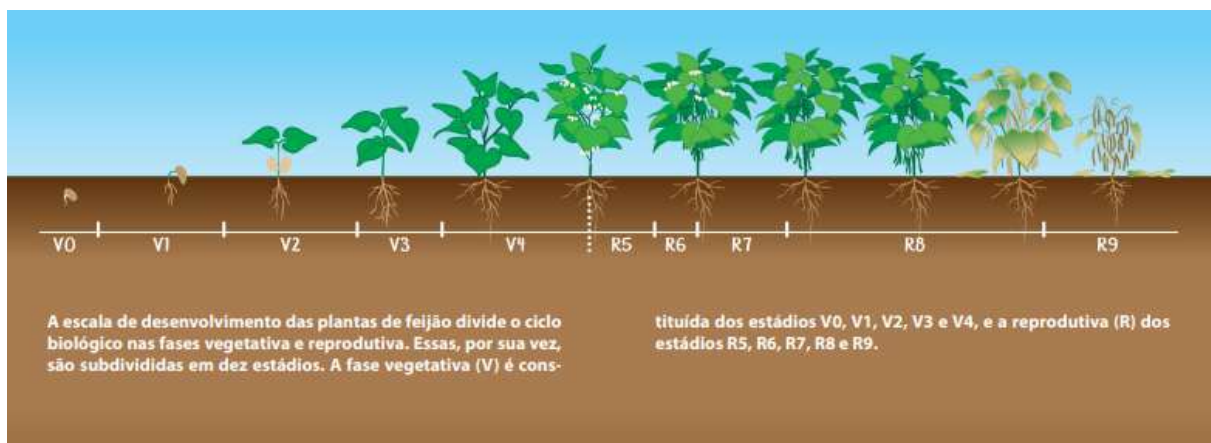
1.4 Introdução à fenologia

"Fenologia refere-se ao estudo dos fenômenos periódicos dos seres vivos e suas relações com as condições do ambiente e a correlação com os aspectos morfológicos" (Oliveira et al., 2018, p. 10). Ela possibilita a observação de características morfológicas e o respectivo tempo associado às características. Portanto, a fenologia representa o desenvolvimento da planta ao longo do tempo e permite a diferenciação das etapas do ciclo:

- Germinação;
- Emergência;
- Desenvolvimento vegetativo;
- Florescimento;
- Frutificação;
- Formação das sementes;
- Maturação.

A fenologia do feijoeiro possui duas fases distintas e com denominações próprias, a saber: fase vegetativa (V) e fase reprodutiva (R). Na fase vegetativa, a forma de identificar o desenvolvimento é dada pelos seguintes estádios: V0, V1, V2, V3 e V4. Já a fase reprodutiva, é dada pelos estádios R5, R6, R7, R8 e R9 (Oliveira et al., 2018), como ilustrado na Figura 1. Em cada estádio, é necessário considerar a influência de todos os fatores que direta ou indiretamente afetam o comportamento da cultura (Oliveira et al., 2018; Boote et al., 1996).

Figura 1: Estádios do desenvolvimento da planta



Fonte: OLIVEIRA et al., 2018.

Neste trabalho, o modelo CROPGRO-Drybean foi selecionado para simular o desenvolvimento da cultivar BRS *Esplendor*, conforme será detalhado posteriormente. Segundo Oliveira et al. (2018), a identificação do estágio de desenvolvimento da planta é fundamental para a correta tomada de decisão. Além disso, a previsão correta das datas de cada fase é imprescindível para as ações de manejo, como adubação e irrigação (Boote et al., 1996; Oliveira et al., 2018). Inicialmente, serão utilizados apenas os hiperparâmetros fenológicos descritos na Tabela 1 como base para a otimização, deixando a expansão do escopo para incorporar todos os parâmetros do DSSAT como sugestão para estudos futuros.

Tabela 1 - Descrição dos parâmetros fenológicos.

EM-FL	Tempo entre a emergência da planta e o aparecimento da primeira flor
FL-SH	Tempo entre a primeira flor e a primeira vagem
FL-SD	Tempo entre a primeira flor e a primeira semente
SD-PM	Tempo entre a primeira semente e a maturidade fisiológica

Fonte: Autor

Como resultado da Otimização Bayesiana (BO), os dados de saída também corresponderão a datas de importância fenológica: ADAP, MDAP, PD1P e PDFP, que estão descritas na Tabela 2. Para o pleno funcionamento do sistema DSSAT,

todos os hiperparâmetros deverão ser introduzidos na simulação, porém aqueles não sujeitos à otimização fenológica serão mantidos constantes.

Tabela 2: Dados de saída do DSSAT.

ADAP	Data do florescimento ou início da fase reprodutiva
MDAP	Data de maturidade fisiológica da planta – fim do ciclo
PD1P	Dias após o plantio até a fase 1 do desenvolvimento das vagens
PDFP	Data da formação da primeira semente

Fonte: Autor

O feijão-comum é um dos principais ingredientes na mesa do brasileiro. Ele constitui a base da segurança alimentar e uma das principais fontes de proteína vegetal no mundo. É amplamente utilizado em países menos desenvolvidos, pois ajuda no combate à fome e à desnutrição. Cargnin (2010) esclarece que o Brasil é um dos maiores produtores de feijão. Contudo, a produção nacional não é capaz de suprir toda a demanda interna do produto, resultando na importação de aproximadamente 100 mil toneladas de feijão-comum todos os anos. Com o objetivo de reduzir o nível de importação e fazer frente à demanda de feijão-comum, o programa de melhoramento genético lançou a cultivar BRS Esplendor no Distrito Federal, visto que esta cultivar possui alto potencial produtivo e uma resistência maior a fungos.

1.5. Funcionamento do código

Durante o uso do DSSAT, o pesquisador necessita imputar os valores dos parâmetros corretamente, de forma que a previsão seja a mais precisa possível. Com os anos de experiência, é possível que a intuição do pesquisador seja apurada o suficiente para operar uma boa calibração do modelo. Contudo, esse método se baseia na ideia de tentativa e erro (Van Oort; Van Ittersum, 2013). A Otimização Bayesiana (BO) é uma estratégia que, além de calibrar o modelo matematicamente,

transforma um problema de alto custo computacional em um processo de aprendizado e de otimização inteligente via função de aquisição.

A Otimização Bayesiana (BO) proposta neste trabalho foi implementada com base no software R e utiliza a função `bayesOpt` do pacote *ParBayesianOptimization*. O objetivo principal dessa função é maximizar uma função objetivo ajustando um Processo Gaussiano (GP) e, em seguida, proceder com uma função de aquisição que irá nortear o próximo ponto promissor (Snoek et al., 2012).

Outras técnicas como o Algoritmo Genético (AG), por exemplo, retornam uma otimização dos parâmetros. Contudo, o Algoritmo Genético conduz uma busca cega na população, enquanto o Otimizador Bayesiano conduz uma busca inteligente, além de funcionar muito bem com pequenas amostras (Snoek et al., 2012). Portanto, o resultado final da Otimização Bayesiana encontra a combinação de coeficientes que permite ao DSSAT simular o ciclo da cultivar BRS *Esplendor* com a maior precisão possível, evitando que vieses humanos sejam adicionados ao processo, sendo, assim, uma alternativa robusta ao processo manual de tentativa e erro (Wallach et al., 2024).

1.5.1 Fase de Inicialização

A função PBO (*ParBayesianOptimization*) admite como entrada valores que irão conduzir o processo ao melhor resultado possível, dadas as características dos dados analisados. A Função Objetivo (FUN) é aquela que norteia o processo, pois é a partir dela que é possível saber se a função de aquisição encontrou pontos ótimos que superaram os pontos iniciais já avaliados pelo algoritmo.

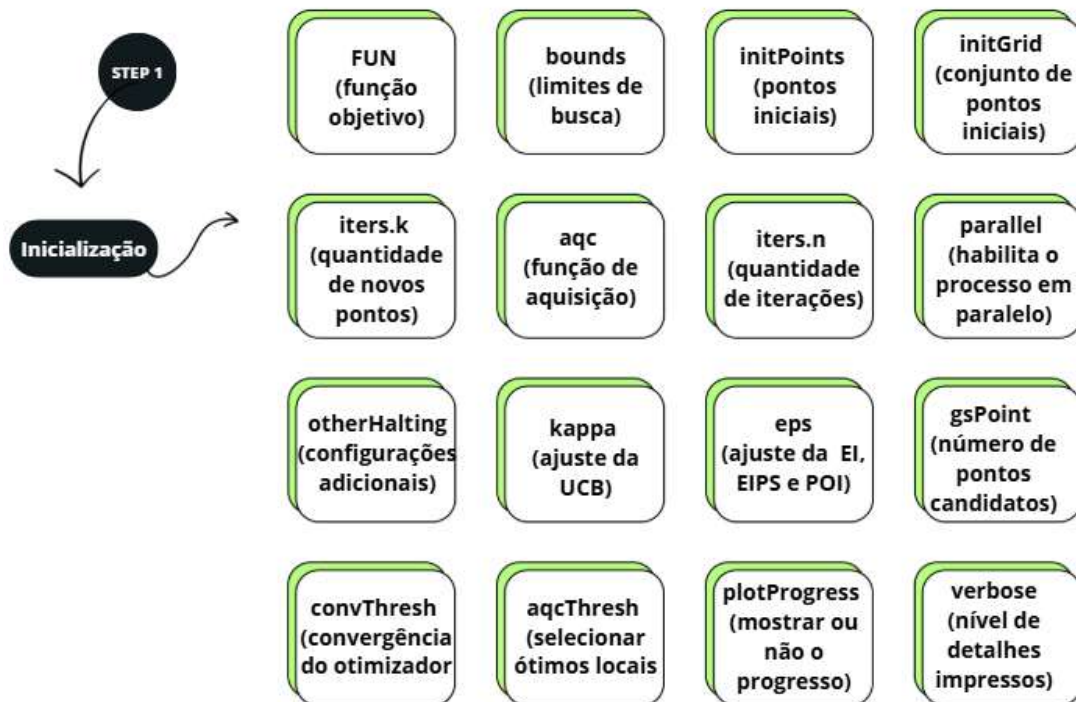
A função utilizada como ponto de calibração foi a minimização do MAPE (*Mean Absolute Percentage Error*), cuja forma analítica é dada pela Equação (1). Esta função retorna o erro percentual relativo, cujo suporte está entre 0 e 100 (Chai; Drapper, 2010).

$$MAPE^* = \frac{100}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right|, \quad (1)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right|. \quad (2)$$

Neste estudo foi utilizada a fórmula (2), sendo n o número total de observações, A_i o valor real da métrica no período i , F_i o valor previsto pelo modelo no mesmo período. Assim, $|A_i - F_i|$ calcula o erro absoluto (diferença) entre os valores reais e previstos e $\left| \frac{A_i - F_i}{A_i} \right|$ calcula o erro percentual absoluto (APE) de cada observação. Para proceder com a média basta somar os erros percentuais absolutos e dividi-los pelo número de observações n . Esse resultado possui um suporte entre 0 e 1 e foi a medida relativa que balizou a função objetivo a encontrar o menor MAPE para proceder com a calibração do modelo (Chai; Draper, 2010).

Figura 2: Fase de inicialização da função PBO



Fonte: Autor

Na sequência, e ainda com referência à Figura 2, os limites de busca (bounds) determinam o universo de valores que serão explorados dentro de cada

parâmetro. Definiu-se uma lista com os limites inferiores e superiores de cada hiperparâmetro. Dado o escopo de estudo deste trabalho, os limites foram definidos com base em conhecimentos prévios (teóricos e experiência profissional) sobre o feijão-comum, por profissionais da área agrônômica.

O algoritmo somente utiliza os valores que estiverem dentro desses limites, sendo imprescindível que o valor ótimo esteja presente no intervalo. Há implicações negativas para o resultado da otimização caso os limites de busca sejam negligenciados (Jones et al., 1998). Limites muito estreitos podem não conter o valor ótimo, enquanto aqueles com limites muito largos resultam na falta de eficiência e elevado custo computacional, pois haverá uma maior exploração da área. Os limites utilizados neste trabalho são apresentados na Tabela 3.

Tabela 3: Limites em dias fototérmicos dos hiperparâmetros do modelo.

Hiperparâmetros	Valores em dias fototérmicos
EMFL	15° a 55°
FLSH	5° a 20°
FLSD	5° a 20°
SDPM	15° a 40°

Fonte: Autor

Na sequência, o usuário opta pela entrada da quantidade de pontos iniciais (initPoints) ou pela disponibilização de um data-frame (quadro de dados) dos pontos que ele entende que geram maior celeridade ao processo (initGrid). A inicialização com o initGrid nada mais é do que o uso direto de tais pontos para a avaliação e para o ajuste do Processo Gaussiano (compreendido em seções posteriores). Como são mutuamente exclusivas, o não uso do initGrid requer que o algoritmo gere o número de pontos por amostragem (initPoints) para, só então, realizar a avaliação e o ajuste do processo.

Ao final de cada iteração, a função de aquisição determina os pontos que otimizam a utilidade. Essa quantidade pode ser determinada pelo usuário em um número inteiro (iters.k) ou por padrão (default) com iters.k = 1. O parâmetro acq

(Figura 2) é o parâmetro que define qual função de aquisição será utilizada no código e será melhor detalhada a seguir.

Outros dois parâmetros que controlam a função de aquisição e podem ser definidos pelo usuário são o kappa e o eps. Enquanto o kappa é utilizado na função de aquisição UCB (*Upper Confidence Bound*), o eps é utilizado em EI (*Expected Improvement*), EIPS e POI, mas ambos servem para o mesmo propósito: equilibrar a exploração e a exploração da função de aquisição. O kappa controla o peso dado à incerteza do modelo (desvio padrão do Processo Gaussiano), logo, quanto maior o kappa, maior a exploração em detrimento à exploração. Já o eps define uma margem de melhoria; quanto maior seu valor, mais conservador se torna o algoritmo, incentivando a exploração contra a exploração.

O gsPoints é um parâmetro que também faz parte da fase de aquisição e pode ser definido pelo usuário, como mostra a Figura 2 e será definido posteriormente. O parâmetro parallel habilita a paralelização do processo, que por padrão estará desabilitada. O parallel define como a Função Objetivo (FUN) será executada, ou seja, permite que o programa utilize os núcleos de processamento computacional disponíveis para executar $iters.k$ simulações simultaneamente, reduzindo o tempo total para a finalização do processo.

Os parâmetros verbose e plotProgress imprimem os resultados de saída e os resultados do progresso do processo. O parâmetro otherHalting admite novos critérios de parada adicionais para o algoritmo. Os mais comuns são:

- timeLimit: permite ao usuário determinar um tempo máximo de execução para a otimização.
- minUtility: permite ao usuário determinar um valor mínimo para a função de aquisição. Isso ajuda a otimização ser mais célere, pois uma vez que o valor máximo da otimização é menor que esse critério, então o algoritmo já convergiu de uma forma eficiente e não precisa continuar o processo iterativo.

O parâmetro convThresh (*Convergence Threshold*) monitora o valor da função de aquisição. O processo de otimização cessa se a função de aquisição ficar abaixo de um limite definido pelo usuário. Esse parâmetro busca parar o processo caso o próximo ponto não trouxer uma melhoria significativa. Já o parâmetro

acqThresh (Acquisition Threshold) também determina um critério de parada, porém aqui a função verifica o valor máximo da utilidade encontrado pela função de aquisição. Se esse valor for menor que o determinado pelo usuário, então a otimização parará. Aquele observa a convergência do processo de otimização, enquanto esse se concentra no potencial do próximo passo a ser dado.

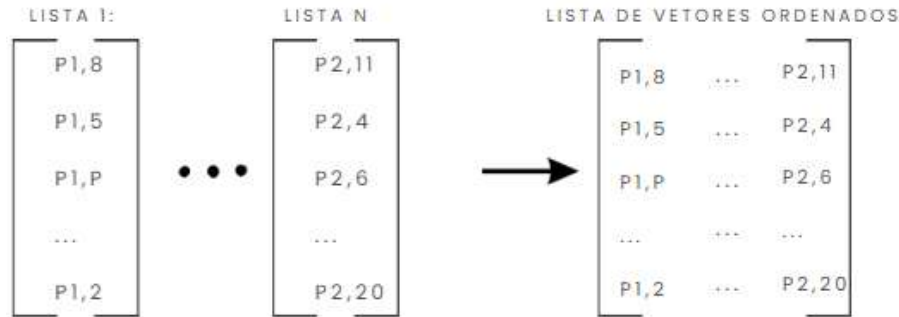
Em suma, esses são os parâmetros de entrada que, se combinados de forma eficiente, resultam em uma otimização rápida e com boa convergência. Isso evita desperdício computacional e esforço desnecessário caso o usuário tenha um bom conhecimento sobre a questão proposta.

1.5.2 Fase de Amostragem e Avaliação

Nessa etapa do algoritmo, a quantidade de pontos iniciais (*initPoints*) é utilizada como valor chave, portanto, por simplificação, a chamaremos de *P*. Nesse trabalho é utilizado o *initPoint* ao invés de *initGride*. Após a inicialização com os valores que o usuário entende serem melhores para a otimização, o algoritmo divide cada parâmetro em *P* partes de forma uniforme respeitando os limites de cada um, formando *P* partições por parâmetro.

Sequencialmente, é selecionado um ponto de forma aleatória dentro de cada partição e esses pontos farão parte de uma lista criada para cada parâmetro. As posições dos pontos dentro das listas são reordenadas de forma aleatória. Uma nova lista de vetores ordenados é criada combinando as *i*-ésimas posições de cada lista como mostra a Figura 3.

Figura 3: Esquema gráfico do processo de Amostragem por Hipercubo Latino



Fonte: Autor

Na Figura 3, há a disposição de N parâmetros e P pontos reordenados, gerando, portanto, P pontos de vetores ordenados. Todo esse processo recebe o nome de Amostragem por Hipercubo Latino ou ainda LHS do inglês *Latin Hypercube Sampling*. A LHS é utilizada para garantir que a otimização ocorra com um espaço bem distribuído, evitando gasto com procuras aleatórias.

Com os P pontos disponíveis, o programa é capaz de gerar uma simulação DSSAT para cada vetor ordenado, obtendo como resultado os valores negativos do MAPE para cada um e os armazenando em um vetor escore. Essa etapa é denominada avaliação dos pontos. Os valores de mínimo e/ou máximo do vetor escore são importantes para a utilização de algumas funções de aquisição, assim como o vetor escore será utilizado para o ajuste do processo gaussiano.

1.5.3 Fase de Ajuste do Processo Gaussiano

Com o vetor escore abastecido com os P valores gerados na etapa de avaliação, é ajustado um processo gaussiano. Antes do processo ser ajustado, os valores dos parâmetros (dados de entrada) e os resultados (escore ou dados de saída) são normalizados de forma que o suporte da função esteja entre 0 e 1. Os dados são normalizados utilizando a técnica min-max, também conhecida como reescalonamento linear.

$$\frac{\text{Valor normalizado}}{\text{Dados de entrada}} = \frac{V_o - \min \{bound\}}{\max \{bound\} - \min \{bound\}}, \quad (3)$$

$$\frac{\text{Escore normalizado}}{\text{Dados de saída}} = \frac{S_o - \min \{vetor\}}{\max \{vetor\} - \min \{vetor\}}. \quad (4)$$

Das equações (3) e (4), tem-se que V_o é o valor original do dado de entrada, $\min\{bound\}$ é o valor do limite inferior definido no bound desse parâmetro, assim como $\max\{bound\}$ é o seu limite superior. De forma análoga, S_o é o dado original do vetor escore, $\min\{vetor\}$ é o valor do menor MAPE do vetor escore, assim como $\max\{vetor\}$ é o seu maior valor.

A normalização é crucial para ajustar valores com escalas distintas em uma mesma função e possibilita a aprendizagem do modelo com a relação de dados de entrada com os respectivos dados de saída. O ajuste do modelo é feito com os dados normalizados e não constitui um algoritmo de otimização, mas sim de aprendizado de máquina.

A PBO utiliza a função de Kringin (K_m) derivada do pacote DiceKriging, onde X é a matrix de entrada normalizada e Z são os escores normalizados. A função K_m é uma técnica de interpolação, pois pega o conjunto de dados e usa para prever valores em outros pontos, funcionando como um Modelo de Regressão Preditiva. A aprendizagem de máquina no algoritmo (Processo Gaussiano) utiliza a função K_m , que aprende com os resultados obtidos no modelo e generaliza para os pontos que ainda não foram avaliados, modelando a covariância entre os dados por meio da função de covariância (kernel). A K_m supõe que os dados estão correlacionados, de modo que valores próximos em X geram resultados próximos em Z .

A estimação interna consiste em encontrar os melhores valores para os hiperparâmetros do kernel, que são: parâmetro de escala e efeito pepita (*nugget effect*). O parâmetro de escala estuda a distância de correlação (length-scale ou rugosidade), uma medida que a função KM utiliza para saber quão rugosa é a função objetivo, pois pontos próximos no espaço de busca possuem valores de escore mais correlacionados que aqueles que estão distantes. Por outro lado, o efeito pepita é aquele que adiciona ruído ao modelo. Ou seja, adiciona uma

aleatoriedade ao modelo que não pode ser explicada pela correlação entre os pontos, evitando um overfitting no modelo. Tais hiperparâmetros do kernel são estimados utilizando a Máxima Verossimilhança.

A rugosidade da função mostra o quão suave ou não é a função objetivo. Funções suaves tendem a um comportamento em que pequenas mudanças nos parâmetros resultam em pequenas mudanças no escore. Nesse caso, o valor da length-scale é alto, pois a correlação se estende a pontos mais distantes. Contrariamente, curvas mais rugosas possuem length-scale baixo e pequenas mudanças nos parâmetros geram grandes distúrbios no escore, pois os pontos só são correlacionados quando então bem pertos um dos outros no campo de busca. Essa informação é obtida fazendo-se a Máxima Verossimilhança e é importante para saber a precisão das previsões futuras. Em seguida, o algoritmo trabalha o modelo do Processo Gaussiano ajustado (GP) com valor esperado (μ) e incerteza (σ^2).

O GP é um processo de aprendizado de máquina supervisionado para modelos de regressão, que utiliza como otimizador a Máxima Verossimilhança para encontrar os melhores hiperparâmetros, conforme apresentado anteriormente. Ele aprende como é a relação entre os dados de entrada e de saída, com o objetivo de prever os resultados, o valor esperado e a incerteza para pontos que não foram avaliados. No entanto, por si só, ele não é capaz de otimizar o modelo, pois a fase da função de aquisição é que determinará os pontos que a GP utilizará como teste, dado que já aprendeu com os dados da fase inicial.

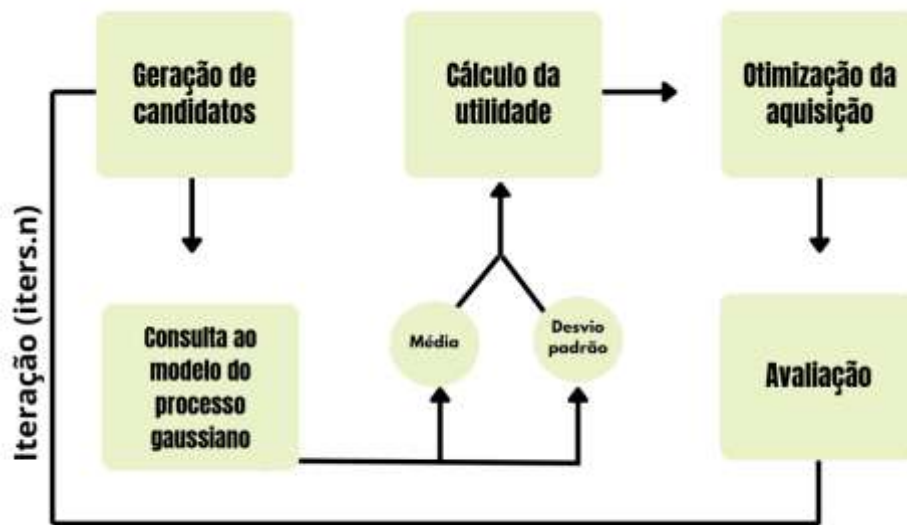
1.5.4 Fase da Função de Aquisição e Iteração

A otimização do código é a parte principal e que confere celeridade a todo o processo. A função de aquisição é a estratégia utilizada para guiar o algoritmo rumo a uma convergência mais rápida e com menor gasto computacional, pois, ao invés de testar todos os pontos possíveis, somente testará os pontos promissores definidos nessa fase.

A função de aquisição é dependente das informações do Processo Gaussiano para definir em quais regiões os pontos são considerados mais promissores. Ela

balanceia dois objetivos diferentes: exploração e a exploração. A exploração norteia o código de modo que os pontos já observados são mais prováveis de conter o valor ótimo. Já a exploração segue a linha contrária, pois prioriza a busca em regiões que ainda não foram exploradas (novos pontos observados). Quando da função de aquisição, a função PBO segue o processo descrito na figura 4.

Figura 4: Processo da função de aquisição na função *ParBayesianOptimization*.



Fonte: Autor

A geração de candidatos é feita da mesma forma que a fase inicial para a geração dos pontos iniciais, através da Amostragem por Hipercubo Latino. Ao invés da função `initPoints`, a função que controla a quantidade de pontos é a `gsPoints` e cria uma base para a busca global. Como já foi dito anteriormente, o `gsPoints` considera como padrão o valor de aproximadamente 10.000 pontos, mais especificamente $\max(100, \text{length}(\text{bounds})^3)$, garantindo que a busca inicial cubra todo o espaço.

Para cada um dos pontos candidatos, a função de aquisição (definida pelo usuário) obtêm os valores da média (μ) e incerteza (σ) do Processo Gaussiano. Nessa fase, há a utilização do aprendizado de máquina no algoritmo, pois os dados históricos obtidos na fase de inicialização são usados como dados de treino do modelo. Na fase de aquisição, o modelo já aprendeu e realiza, de forma mais

eficiente, a inferência (previsão) dos valores esperados e do desvio padrão. Em seguida, serão apresentadas as características de cada tipo de função de aquisição.

1.5.4.1 UCB (Upper Confidence Bound)

A UCB, que na tradução direta significa Limite Superior de Confiança, escolhe pontos candidatos com maior potencial dada a combinação entre previsão e incerteza com a seguinte fórmula (5).

$$UCB = \mu_{GP_{pred}} + \kappa(\sigma_{GP_{pred}}) \quad (5)$$

Quanto maior a média do Processo Gaussiano predito para o ponto candidato ($\mu_{GP_{pred}}$), maior a preferência pela exploração e por permanecer nas regiões próximas aos pontos observados. Porém, quanto maior o desvio padrão (incerteza) do Processo Gaussiano predito ($\sigma_{GP_{pred}}$) ou o valor de kappa, maior a exploração de regiões mais distantes dos pontos observados. Assim, pode-se notar que quando o modelo possui uma alta variabilidade em torno da média, ele buscará compensar a incerteza de encontrar o ponto ótimo na localidade original, procurando por regiões mais distantes.

O uso da UCB é mais recomendado quando o espaço de busca é muito grande ou complexo, com muitos picos potenciais, e deseja garantir que o algoritmo explore todas as regiões para evitar ficar preso em um ótimo local. Vale ressaltar que o valor de kappa é introduzido pelo usuário, mas tem como padrão o valor de 2.576, que é o valor do quantil da normal padrão que retorna um nível de confiança de 99%.

1.5.4.2 EI (Expected Improvement)

A EI ou Melhoria Esperada corresponde ao tipo de função de aquisição que escolhe o ponto que possui uma maior probabilidade de superar o melhor resultado já obtido.

$$EI = (\mu_{GP_{pred}} - y_{MAX} - eps) * (pnorm(z) + \sigma_{GP_{pred}} * dnomr(z)) \quad (6)$$

$$z = \frac{\mu_{GP_{pred}} - y_{MAX}}{\sigma_{GP_{pred}}} \quad (7)$$

Pela fórmula (6), pode-se observar que, além de considerar o valor da esperança do processo gaussiano predito para o candidato ($\mu_{GP_{pred}}$) e do valor desvio padrão do processo gaussiano ($\sigma_{GP_{pred}}$), o algoritmo também considera o valor de eps, que por padrão é zero, o y_{MAX} que é o melhor valor encontrado dentro do vetor escore até o momento, $pnorm$ ou $\Phi(z)$, que calcula a probabilidade do evento z ocorrer, e a densidade da normal padrão, $dnomr$ ou $\phi(z)$, que ajuda a calcular o valor esperado da melhoria. Aqui, z é o valor padronizado do ponto candidato.

O primeiro termo da fórmula capta a melhoria potencial ($\mu_{GP_{pred}} - y_{MAX} - eps$) do ponto candidato e o quanto ele supera o melhor ponto já encontrado. Já o segundo termo capta a probabilidade e o valor esperado dessa melhoria ($pnorm(z) + \sigma_{GP_{pred}} * dnomr(z)$). Portanto, a função EI prioriza a exploração, onde os pontos buscados estarão sempre próximos dos pontos já observados anteriormente e também a melhoria desses pontos. Enquanto a UCB se contenta em encontrar um ponto ótimo em uma grande região de busca, o EI busca encontrar um ponto cada vez melhor dentro de um campo de busca reduzido. Se o objetivo do usuário é a exploração, então a EI é uma boa solução inicial, pois ela foca em refinar os resultados existentes.

1.5.4.3 EIPS (Expected Improvement Per Second)

A EIPS ou Melhoria Esperado por Segundo é uma atualização da EI em que a função agora é penalizada pelo tempo.

$$EIPS = \frac{(\mu_{GP_{pred}} - y_{MAX} - eps) * (pnorm(z) + \sigma_{GP_{pred}} * dnomr(z))}{\mu_{GPE}} \quad (8)$$

A previsão do tempo de execução esperado (μ_{GPE}) é feita por timeGP, que é um modelo do Processo Gaussiano. Então, além de preferir exploração a exploração, ela também é mais eficiente com relação ao tempo. Portanto, visa encontrar o melhor ponto possível o mais rápido possível. A EIPS é fundamental para problemas onde o tempo de execução da função objetivo não é constante entre os pontos, permitindo que a otimização seja não apenas eficaz, mas também eficiente em relação ao tempo.

1.5.4.4 POI (Probability Of Improvement)

A POI ou Probabilidade de Melhoria é uma simplificação do EI e apenas calcula a probabilidade de um ponto candidato superar o melhor ponto atual. De todas as funções de aquisição, a POI é a mais conservadora, sendo, portanto, a função que mais explora e menos explora novas áreas.

$$POI = pnorm(z) \quad (9)$$

$$P(f(x) > y_{MAX} + eps) \quad (10)$$

$$z = \frac{\mu_{GP_{pred}} - y_{MAX} - eps}{\sigma_{GP_{pred}}} \quad (11)$$

Note que, pela figura 5, apenas a função UCB prefere a exploração ao invés da exploração, ou seja, ela busca pontos mais distantes daqueles já observados na fase inicial, enquanto que as demais funções são mais conservadoras e preferem áreas de buscas mais próximas dos pontos já conhecidos. As funções EI, EIPS e POI assumem que o ótimo está próximo aos pontos já observados e o seu intento, portanto, é a melhora da estimativa. Dentre elas, a POI busca por exploração de forma mais incisiva que as demais.

Figura 5: Gráfico da exploração versus exploração em cada função de aquisição.



Fonte: Autor

A Figura 6 resume as funções de aquisição de forma simplificada e visual, mostrando que EI, EIPS e POI estão na zona de exploração, enquanto que a UCB está na zona de exploração. A EI é a função central, da qual derivam a EIPS e a POI.

Figura 6: resumo das funções de aquisição



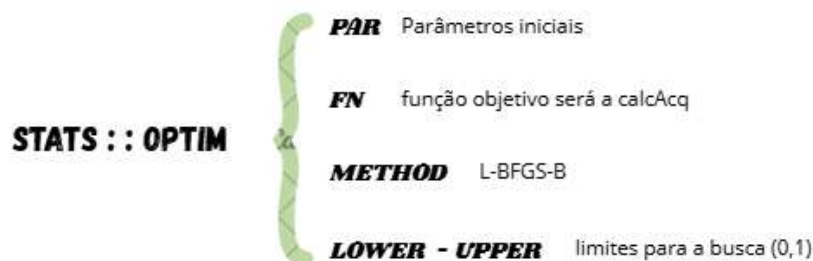
Fonte: Autor

1.5.4.5 Otimização da função de aquisição e maximização da utilidade

Dentro da função de aquisição ainda há um ponto crucial para a otimização do algoritmo. O refinamento da busca por um ótimo da função de aquisição é feito a partir da utilidade. Portanto, o objetivo é maximizar o valor da utilidade e encontrar os pontos mais promissores dentre todos.

Como já visto anteriormente, os pontos utilizados dentro da função de aquisição foram normalizados e possuem suporte entre 0 e 1. A função PBO possui um comando chamado de `stats::optim`, e cada ponto candidato normalizado passa por essa nova função, que retorna o ponto encontrado que maximiza localmente a função de aquisição juntamente com o seu valor de utilidade. Dentro da função `stats::optim`, as principais configurações para a perfeita execução do programa estão dispostas na Figura 7.

Figura 7: dados de inicialização da função `optim`

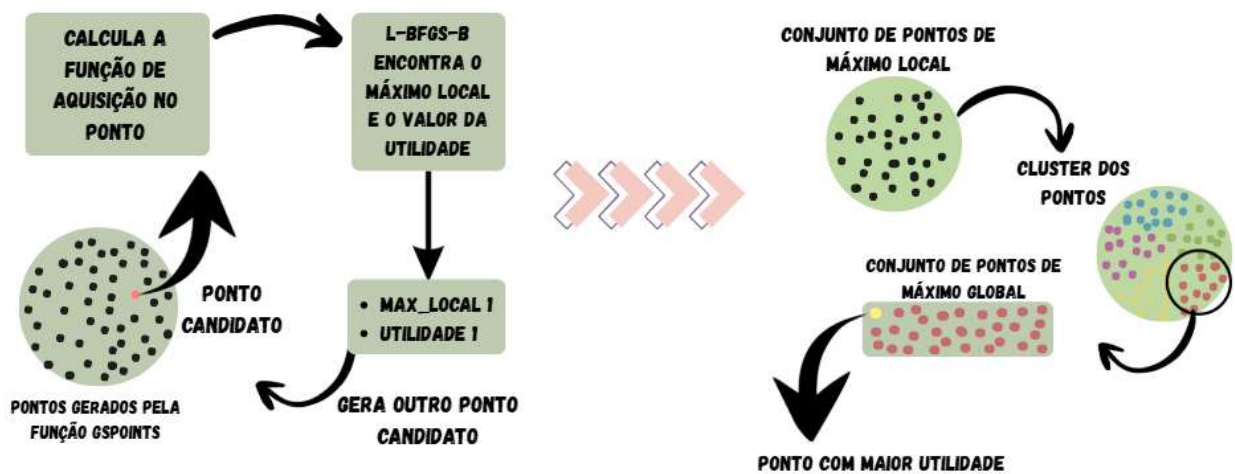


Fonte: Autor

Da função `gsPoints` são gerados milhares de pontos candidatos, como visto na seção anterior, e o parâmetro inicial (`PAR`) utiliza cada ponto candidato como ponto de partida. A função a ser otimizada (`FN`) é chamada de “`calcAcq`” e calcula a função de aquisição para esse ponto específico. O L-BFGS-B é um método de otimização numérico que serve para encontrar mínimos locais. Como o objetivo da PBO é encontrar o máximo, o método controla essa diferença utilizando o parâmetro “`control = list(fnscale = -1)`”, assim o resultado sairá com um valor invertido. Já os limites de busca (`lower` e `Upper`) correspondem ao espaço de busca dos parâmetros.

O resultado encontra o ponto que maximiza localmente a função de aquisição e o valor da utilidade. Ao final de todo o processo os resultados são coletados e agrupados (“dbscan”) para encontrar os máximos locais únicos. O ponto escolhido é aquele que, dentre os pontos de máximo local, possui a maior utilidade, como mostra a figura 8.

Figura 8: processo de otimização pela função optim.



Fonte: Autor

Os cálculos da utilidade, do ótimo local e do ótimo global são dependentes do tipo de função de aquisição que o código utiliza, pois são fundamentais na função “calAcq”. Após a determinação do ponto de maior utilidade dentro do conjunto de pontos que maximizam globalmente a função de aquisição, o algoritmo escolhe os $iters.k$ pontos que farão parte do resultado final da rodada.

Por fim, a fase de avaliação consiste em fazer a simulação DSSAT (usar a função objetivo) em cada um dos novos $iters.k$ pontos que serão agregados ao histórico score (resultados). Portanto, ao final de uma iteração, o tamanho do vetor score será de $initPoints + iters.k$, assim como que, ao final do processo com todas as iterações, o tamanho do vetor score será $initPoints + iters.k * iters.n$.

1.5.5 Fase Final

Ao final de todas as iterações, o programa retorna os seguintes resultados (dispostos de forma geral e sem considerar cada caso):

- O histórico de todas as simulações (scoreSummary);
- O modelo de Aprendizado de Máquina (GauProList);
- Os parâmetros de controle (aptPars);
- Os critérios de parada, quando houver (stopStatus);

O histórico de todas as simulações é uma tabela de dado como registro de cada simulação em cada época (iteração) e até mesmo da fase inicial. Em cada linha ele contém:

- Score: resultado da função objetivo (MAPE);
- Epoch: o número da iteração ou época;
- Iteration: a ordem da avaliação dentro da iteração;
- Elapsed: tempo de execução da simulação, quando utilizado a EIPS.

O modelo de aprendizado de máquina contém o modelo do Processo Gaussiano ajustado por último e que resultou no vetor score (scoreGP) e também o tempo de execução que a simulação utilizou no caso da função de aquisição EIPS (timeGP). Os parâmetros de controle contém todos os parâmetros que foram utilizados na otimização: limites, acq (função de aquisição), valores de kapps, eps, gsPoints, iters.n, iters.k. Também há as funções auxiliares que devolvem os parâmetros de entrada que geraram o melhor resultado (getBestPars(optObj) e o valor do melhor resultado alcançado.

1.6 Inferência Bayesiana e o Otimizador Bayesiano

A Otimização Bayesiana (BO) herda esse nome da lógica de Inferência Bayesiana, que utiliza o Teorema de Bayes para encontrar uma distribuição a posteriori, dado que se conhece a função de verossimilhança dos dados e a distribuição a priori do parâmetro dos dados (Gelman et al., 2013; De Finett, 1974).

O Teorema de Bayes expressa a probabilidade de um evento qualquer A ocorrer, dado que outro evento B ocorreu, seguindo a seguinte forma analítica:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}, \quad (12)$$

Onde:

- $P(A|B)$ é a probabilidade a posteriori (a crença atualizada).
- $P(B|A)$ é a função de verossimilhança (a evidência dos dados).
- $P(A)$ é a probabilidade a priori (a crença inicial).
- $P(B)$ é a probabilidade marginal (constante de normalização).

A Inferência Bayesiana demonstra que, se X é uma variável aleatória cujo seu parâmetro θ também é assumido como sendo uma variável aleatória, então, a sua forma é dada por:

$$\pi(\theta|X) = \frac{L(X|\theta)\pi(\theta)}{\pi(X)}, \quad (13)$$

$$\pi(X) = \int_{\Theta} L(X|\theta)\pi(\theta) d\theta. \quad (14)$$

Onde $L(X|\theta)$ é a função de verossimilhança de X , $\pi(\theta)$ a função a priori de θ , Θ o espaço paramétrico que define o suporte de θ , $\pi(X)$ a preditiva e, por fim, $\pi(\theta|X)$ a posteriori. Similarmente, pode-se considerar que a menos de uma constante, a posteriori também pode ser escrita como:

$$\pi(\theta|X) \propto L(X|\theta)\pi(\theta) \quad (15)$$

O objetivo dessa função é encontrar a distribuição a posteriori, levando em consideração a função de verossimilhança e a distribuição a priori de θ . É uma forma de atualizar o conhecimento sobre uma forma incerta, usando os dados observados e a ideia de probabilidade (De Finetti, 1974).

A BO nada mais é que a aplicação sequencial da Inferência Bayesiana, em que a distribuição a priori é dada pelo Processo Gaussiano (não ajustado); a verossimilhança é assumida pelo vetor escore (vetor de resultados) incrementado a cada nova iteração e, por fim, a posteriori será o processo gaussiano ajustado, cuja função atualização é a função de aquisição (Jones et al., 1998; Gelman et al., 2013). Por isso, a Otimização Bayesiana leva esse nome, em homenagem à Inferência Bayesiana.

2 METODOLOGIA

Os resultados obtidos a partir da simulação DSSAT, utilizando a Otimização Bayesiana através da função PBO, foram utilizados em comparação com os resultados propostos no Protocolo de Calibração para Modelos de Cultura de Solo (*A calibration protocol for soil-crop models*) e em comparação com a calibração manual de um especialista da Embrapa Arroz e Feijão. Os resultados têm como objetivo avaliar a eficiência da busca informada probabilística (Otimização Bayesiana) contra a estimação estatística estruturada e sequencial (Protocolo) são superiores, formando uma alternativa robusta e computacional para a calibração dos hiperparâmetros do sistema DSSAT. Além do Protocolo, há resultados de uma calibração manual (tentativa e erro) feita por um agrônomo experiente na área chamada de Calibração Manual ou ainda *Default Value*.

O Protocolo possui duas formas de calibração: *Step 6* (estimação separada por grupo de variáveis) e o *Step 7* (estimação simultânea). Na estimativa separada (*Step 6*) os parâmetros são estimados para cada grupo de variáveis (fenologia, biomassa, etc.) separadamente. Cada grupo é ajustado utilizando MQO (Método dos Mínimos Quadrados Ordinários) e é avaliado segundo o AICc (Critério de Informação de Akaike Corrigido). No *Step 6*, os grupos candidatos são testados um a um, de forma muito semelhante ao procedimento do método *Stepwise Forward*, ou seja, se a adição de um grupo aumenta o valor do AICc, então o grupo permanecerá na listagem final de variáveis a serem estimadas. O objetivo desse procedimento é fornecer um grupo que realmente necessita ser estimado por MQO evitando a super parametrização.

Na estimativa simultânea (*Step 7*), todos os parâmetros selecionados são ajustados simultaneamente via Método dos Mínimos Quadrados Ponderados (MQP), sendo que cada grupo de variáveis receberá um peso (ω) proporcional ao erro residual obtido no *Step 6* e é calculado da seguinte forma:

$$\omega = \frac{SQE_{Grupo}}{n - p} \quad (16)$$

Onde 'n' é o número de observações, 'p' o número de grupos estimados no *Step 6* e SQE_{grupo} é a soma dos quadrados dos erros do grupo. O objetivo do *Step 7* é minimizar a soma ponderada dos erros quadrados em todos os grupos, considerando a interdependência entre os grupos.

Além disso, Otimização Bayesiana faz a busca de parâmetros simultaneamente, onde a função de aquisição já considera o erro total do modelo para todos os dados combinados, sem a necessidade das etapas sequenciais como a do protocolo. Sabe-se que a simulação DSSAT retorna as datas ADAP, MDAP, PD1P e PDFP e elas serão simuladas para as 4 formas de calibração apresentadas anteriormente e elencadas a seguir. Posteriormente, serão abordados os resultados obtidos para cada uma dessas datas.

- Calibração via Otimização Bayesiana alternando *initPoints* e *iters.n*;
- Calibração Manual – *Default Value*;
- Calibração Protocolo – *Step 6*;
- Calibração Protocolo – *Step 7*.

3 RESULTADOS

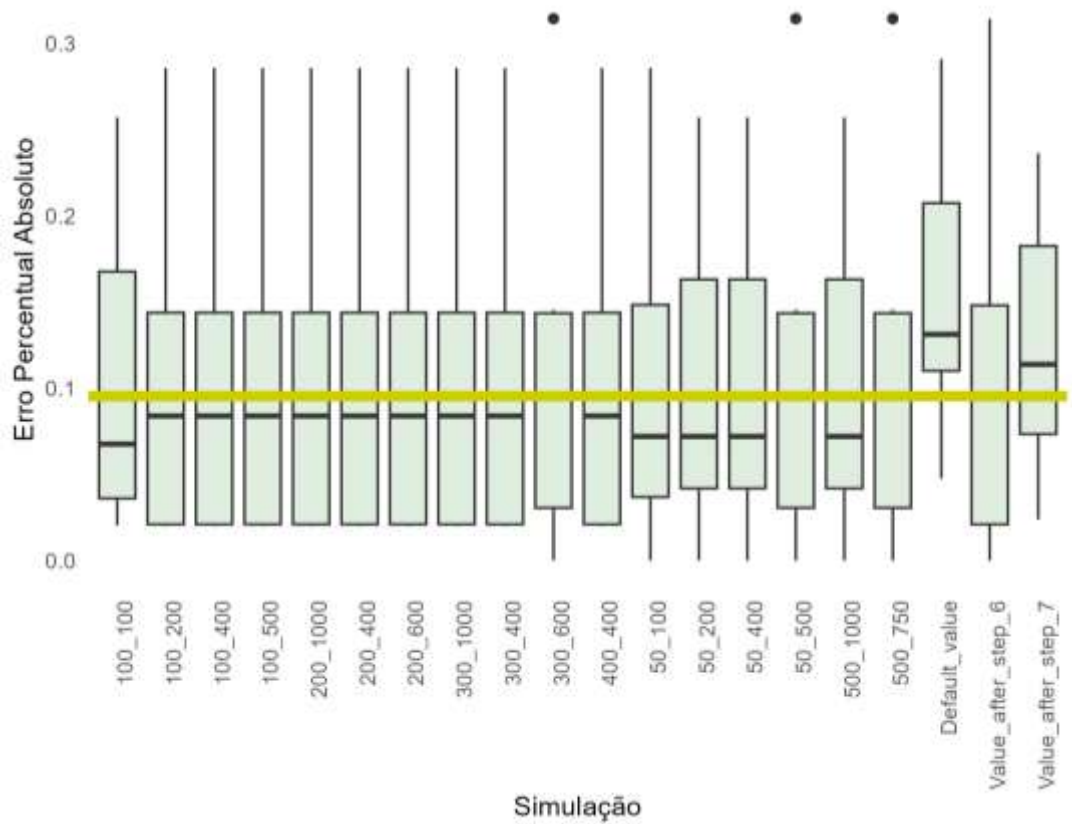
Seção 3.1: ADAP

Os resultados obtidos na comparação do protocolo com a BO, na data do florescimento (início da fase reprodutiva - ADAP), são evidenciados na Figura 9. No eixo das abcissas encontram-se as diferentes simulações, sendo que as três últimas correspondem ao protocolo e à calibração manual. No eixo das ordenadas há os valores do erro absoluto. A apresentação dos resultados da BO segue a seguinte lógica: X_Y, no qual X é a quantidade de pontos iniciais e Y, a quantidade de iterações. Os gráficos Box-Plots foram gerados utilizando a métrica do APE (Erro Percentual Absoluto), com suporte em 0 e 1 definido na Fase de Inicialização.

O objetivo na visualização dos Box-Plots é encontrar a mediana mais baixa, ou seja, a mediana mais próxima a zero. Assim sendo, a BO apresenta resultados superiores ou similares às demais calibrações em todas as combinações. As combinações 300_600, 50_500 e 500_750 assemelham-se aos do protocolo na calibração por estimativa separada (*step* 6), porém, nesses três pontos há a presença de outliers que demonstram que há erros calculados fora do padrão, com valores de aproximadamente 30%. O valor mais discrepante, com o pior resultado, foi aquele obtido pelo processo de calibração manual (*Default_Value*). Mas no geral, a BO obteve os melhores resultados para ADAP.

Como os resultados do *Step* 6 foram sempre superiores aos dos *Step* 7 e do *Default Value* (calibração manual), foi traçada uma reta (reta amarela nas imagens abaixo) na linha da mediana do *Step* 6, com o objetivo de facilitar a comparação visual entre os resultados da BO e dos resultados do *Step* 6. A comparação entre os resultados fica mais clara, pois, uma vez que a BO é superior em qualidade ao *Step* 6, também será superior aos demais resultados de comparação descritos acima.

Figura 9: Resultado APE comparativo da saída ADAP



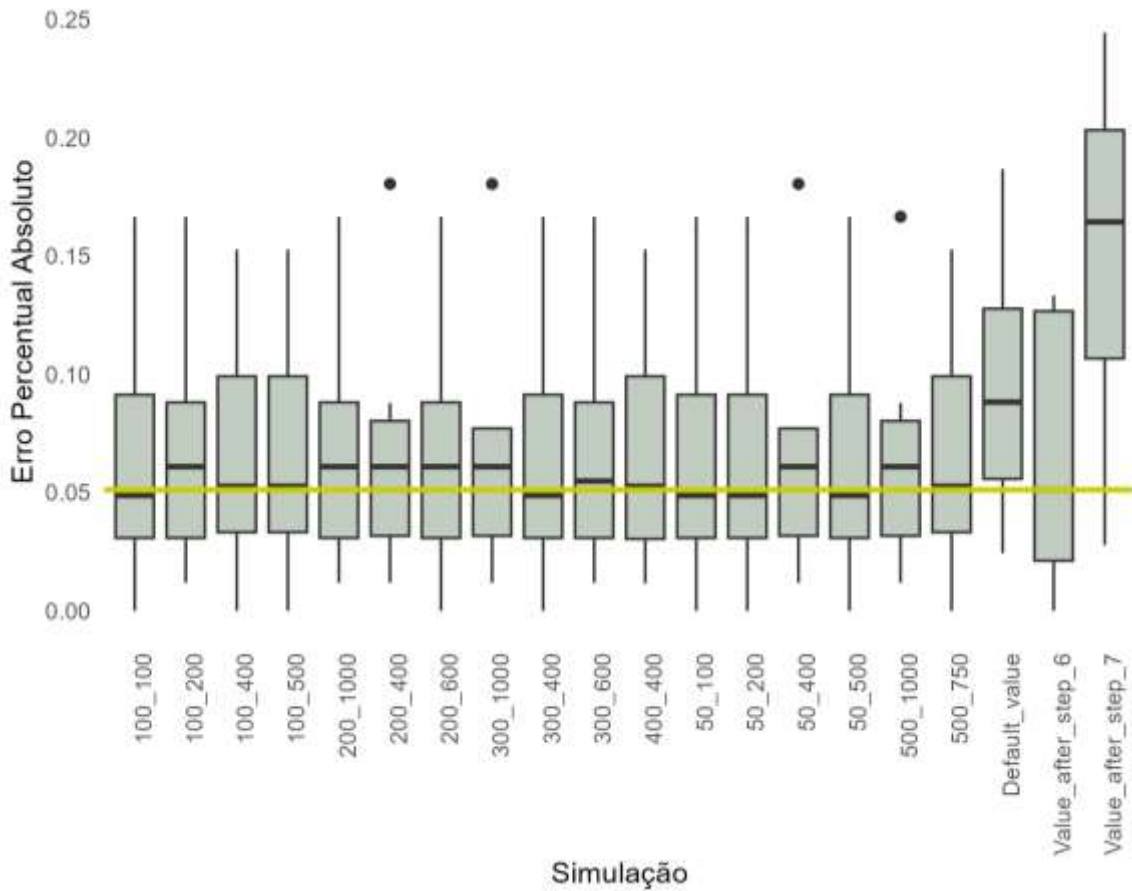
Fonte: Autor

Seção 3.2: MDAP

Os resultados obtidos na comparação da calibração via protocolo e manual com a BO, na data de maturidade fisiológica da planta (MDAP), são evidenciados na Figura 10. O pior resultado é dado pela calibração por estimativa simultânea no *step* 7. Já o protocolo no *step* 6 (calibração por estimativa separada) possui resultados muito próximos aos da BO.

A calibração por BO gera bons resultados, mas em muitos pontos, seus resultados são inferiores aos do *Step* 6. Os pontos outliers encontrados em 200_400, 300_1000, 50_400 e 500_1000 evidenciam que essas combinações levam a piores resultados na BO, uma vez que produz erros fora da padronização (em torno de 18%).

Figura 10: Resultado APE comparativo da saída MDAP

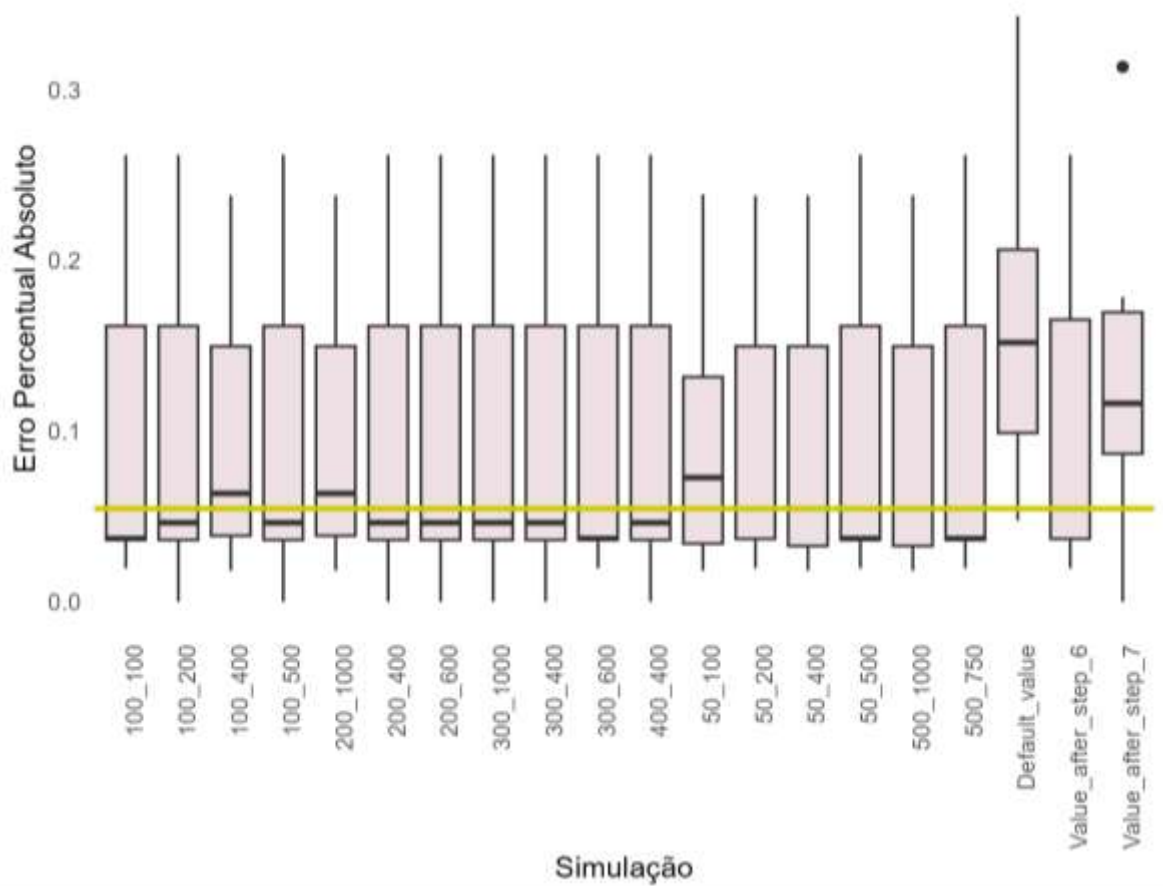


Fonte: Autor

Seção 3.3: PD1P

Os resultados obtidos na comparação da calibração via protocolo e manual com a BO, na data intermediária que marca o início da formação das vagens (PD1P), são evidenciados na Figura 11. O pior resultado é dado pela calibração manual (*Default Value*), seguido da calibração via estimativa simultânea (*Step 7*), que é a única a apresentar outlier. Entretanto, como a calibração via *Step 7* já se mostra inferior na maioria dos resultados, esse ponto outlier não receberá tanta importância. A calibração via estimação separada (*step 6*) do protocolo gera bons resultados, porém, a BO o supera na maioria dos pontos, sendo, nesse resultado, a melhor forma de calibrar os parâmetros do DSSAT.

Figura 11: Resultado APE comparativo da saída PD1P



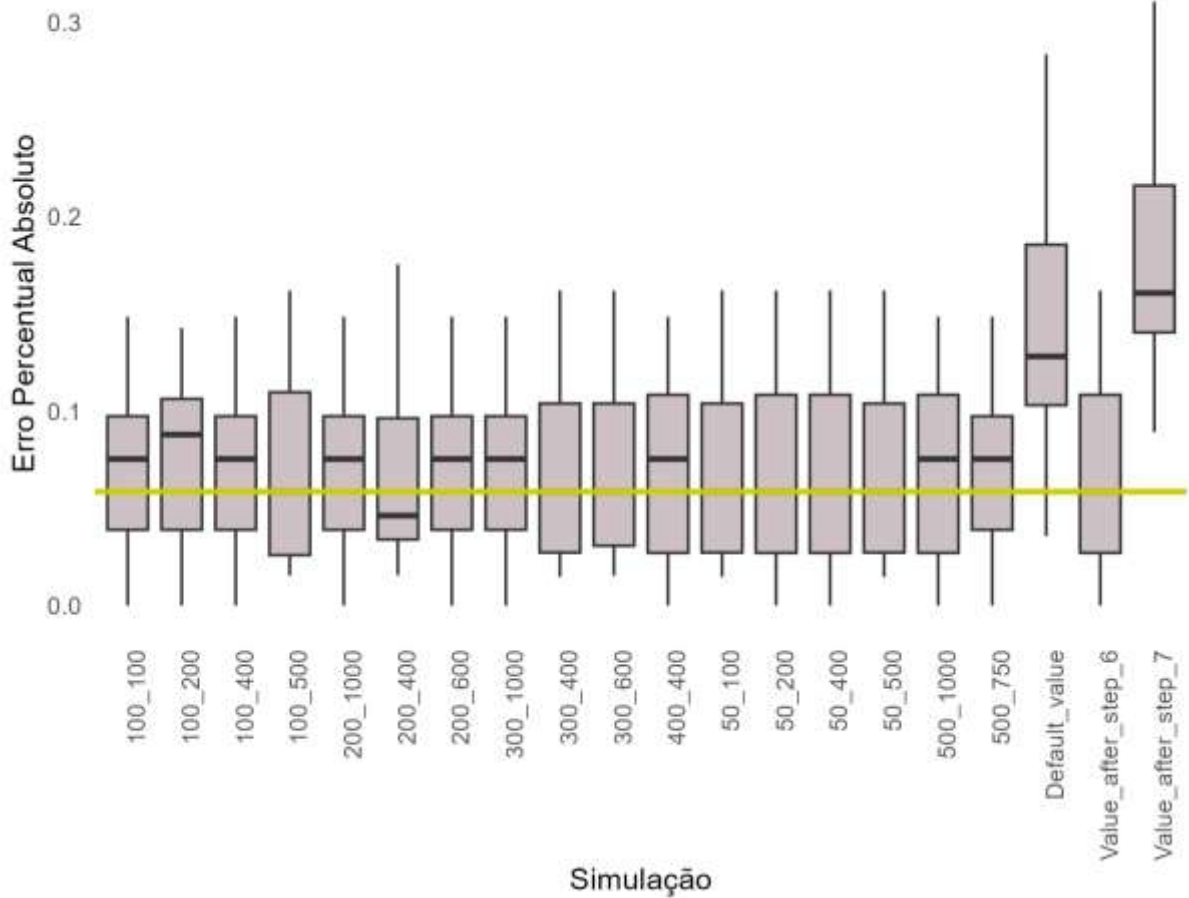
Fonte: Autor

Seção 3.4: PDFP

Os resultados obtidos na comparação da calibração via protocolo e manual com a BO, na data que marca a formação da primeira semente são evidenciados na Figura 12. Os piores resultados foram encontrados na calibração via estimativa simultânea (*Step 7*) e na calibração manual (*Default Value*).

A calibração via estimativa separada - *Step 6* - possui um bom resultado e tem mediana semelhante à calibração via BO em diversos pontos, porém, na maioria das ocorrências supera os resultados da BO. Sendo assim, é importante analisar os resultados conjuntamente, pois, em quase todas as datas, a Otimização Bayesiana tem se mostrado um método de calibração bastante preciso e eficiente, embora, nesse último resultado, a calibração via *Step 6* tenha se mostrado superior.

Figura 12: Resultado APE comparativo da saída PDFFP



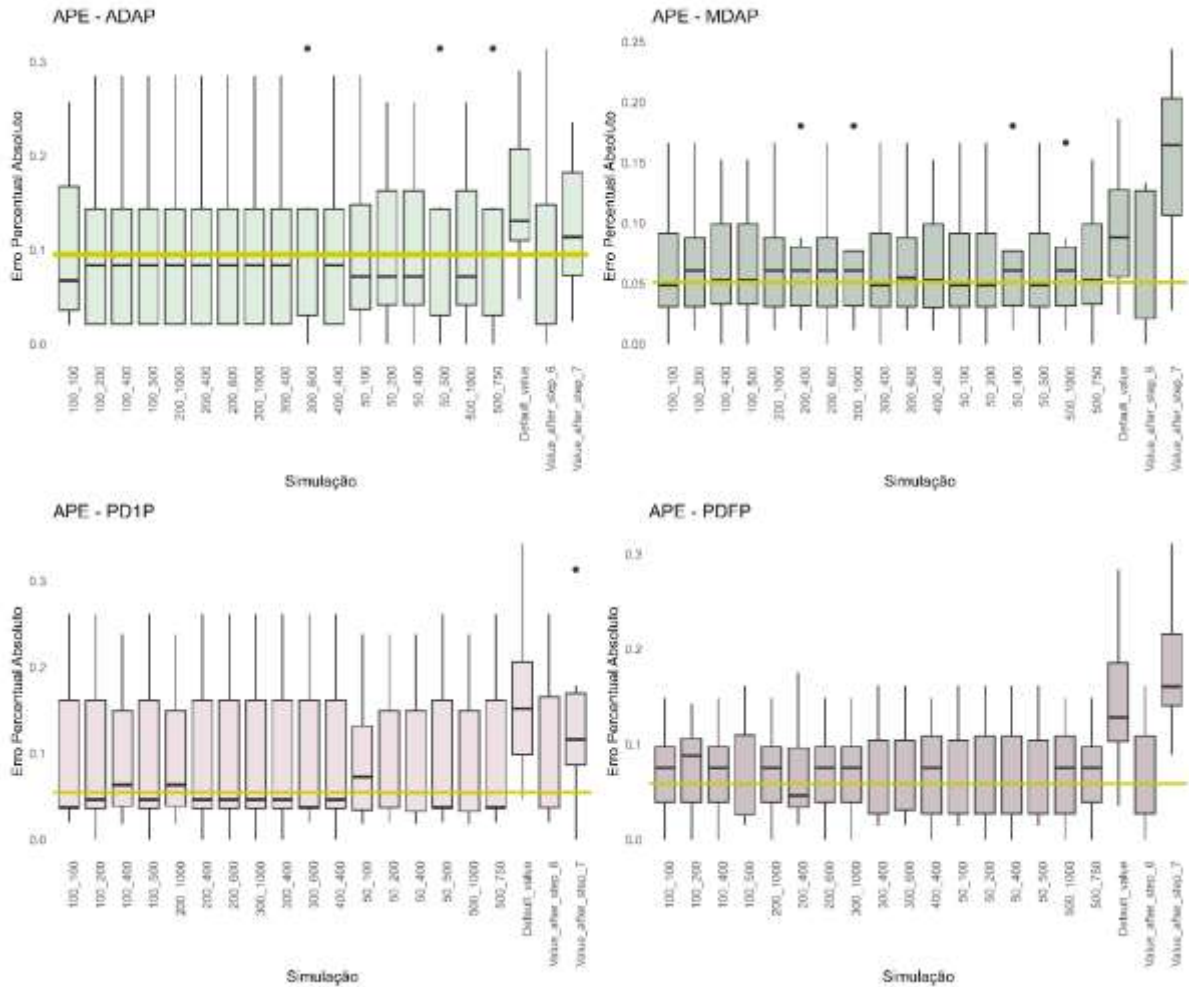
Fonte: Autor

Seção 3.5: Resultados Gerais

De modo geral, os piores resultados sempre contavam com a presença da calibração por estimativa simultânea (*Step 7*) e da calibração manual (*Default Value*). Já a calibração via otimização bayesiana teve bons resultados quando comparada com as demais calibrações na maioria dos casos.

Porém, em alguns momentos, ela era superada pela estimativa separada – *Step 6*, então, como já dito anteriormente, será necessária a análise mais completa e conjunta dos resultados dos hiperparâmetros do subgrupo fenológico.

Figura 13: Todos os resultados APE dos hiperparâmetros simulados



Fonte: Autor

Considerando a Figura 13 e o APE (Erro Percentual Absoluto), é possível encontrar as combinações que geram melhores resultados que aqueles obtidos pela estimativa separada *Step 6* (Tabela 4). Os valores em negrito da Tabela 4 mostram as combinações presentes em todas as datas de saída que apresentam resultados superiores ao do *Step 6*, dados esses valores, foi construída a Tabela 5.

Com base nas combinações da Tabela 5, e considerando apenas as combinações mais rápidas (50_200 e 50_500), os resultados obtidos a partir da BO, de acordo com a função PBO proposta neste trabalho, são apresentadas na Tabela 6.

Tabela 4: combinações da BO que superam ou são equivalentes ao *Step 6*

ADAP	MDAP	PD1P	PDFP
	100_100	100_100	200_400
	300_400	100_200	100_500
	50_100	100_500	200_400
	50_200	200_400	300_400
	50_500	200_500	300_600
Todas as combinações são superiores ou equivalentes		300_1000	50_100
		300_400	50_200
		300_600	50_400
		400_400	50_500
		50_200	
		50_400	
		50_500	
		500_1000	
		500_750	

Fonte: Autor

Tabela 5: Quantidade de pontos iniciais e iterações que tornam o BO mais eficaz.

Quantidade de pontos iniciais	Quantidade de iterações finais
50	200
50	500
300	400

Fonte: Autor

Tabela 6: resultados da PBO para os melhores resultados

Combinação	Época	iteração	EMFL	FLSH	FLSD	SDPM	gpUtility	Escore
50_200	89	139	33,9601	7,2258	12,5872	16,1679	0,0123	-0,0827
50_500	105	155	33,8574	5,4082	12,2513	18,3871	0,0044	-0,0822

Fonte: Autor

A época retratada na Tabela 6 representa o ciclo de treinamento do Modelo de Aprendizado de Máquina (Processo Gaussiano), e corresponde ao estágio da otimização. Já a Iteração indica o número de passos dentro da BO em que o resultado (Escore) foi encontrado. Portanto, na primeira linha, por exemplo, encontra-se o número de 200 iterações como limitante da busca, e na 139ª iteração, o resultado *Escore* = $-0,0827$ foi encontrado. Isso mostra que o processo obteve

um gasto computacional maior que o necessário. Portanto, pode-se pensar em ajustar o parâmetro `convThresh` ou `acqThresh` para que a otimização pare quando encontrar um escore mínimo que não acarrete em maiores acréscimos posteriores.

Tabela 7: Limites versus valores calibrados via BO

Hiperparâmetros	Valores em graus/dia	Valor via BO (50_200)	Valor via BO (50_500)	Valor via BO (300_400)
EMFL	15° a 55°	33,96°	33,86°	33,95°
FLSH	5° a 20°	7,23°	5,41°	5,13°
FLSD	5° a 20°	12,58°	12,25°	13,39°
SDPM	15° a 40°	16,17°	18,38°	15,43

Fonte: Autor

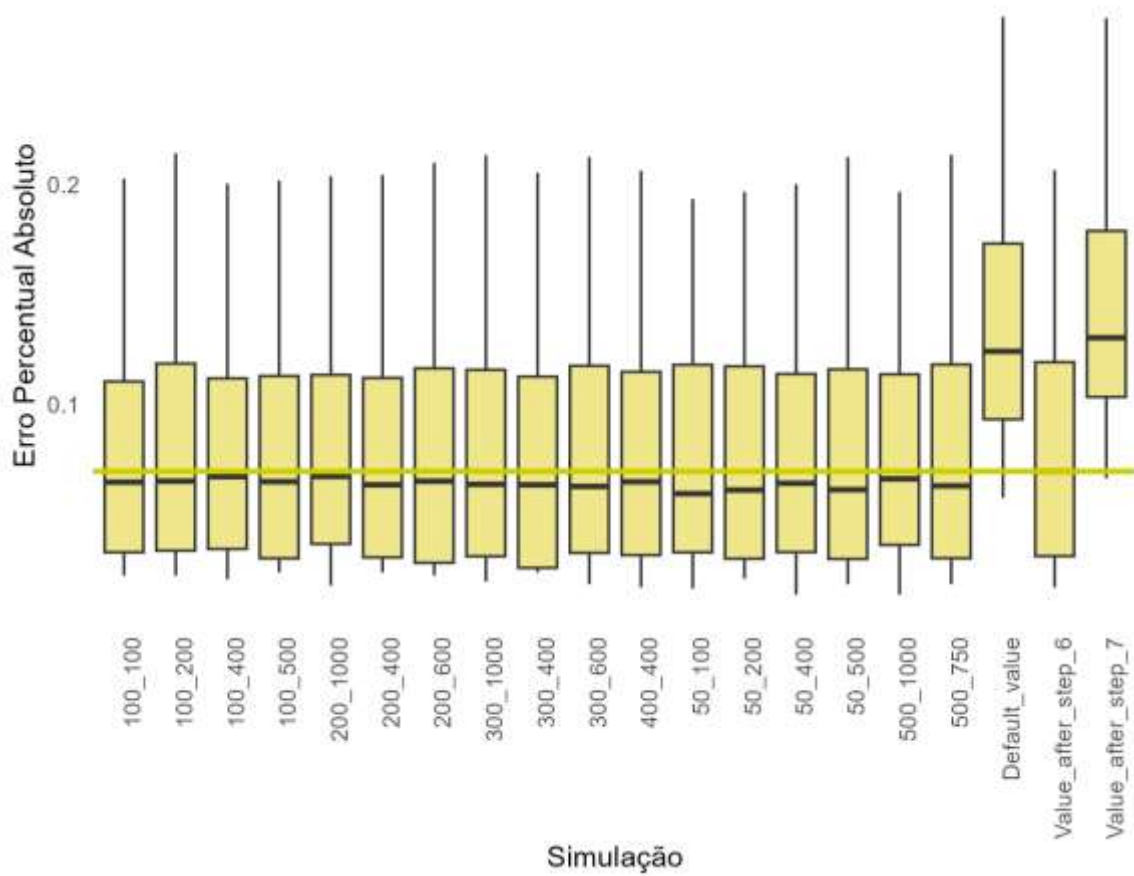
Os valores calibrados para os hiperparâmetros se encontram na Tabela 7 e demonstram que os resultados das calibrações estão bem centradas, com exceção do SDPM. Para esse caso, o melhor seria modificar o intervalo de SDPM de 15° a 40° para 10° a 35° e observar se a nova calibração ainda permanece em aproximadamente 15°, pois a função PBO só irá permitir valores de busca dentro dos limites iniciais. O “gpUtility” é o valor da utilidade calculada dentro da função de aquisição pelo Processo Gaussiano. Dentro do conjunto de valores que oferecem um máximo global, o valor escolhido será aquele com maior utilidade, pois considerará que o valor é mais promissor, nesse caso, 50 pontos iniciais e 200 iterações.

Finalmente, o valor “escore” é o resultado da função objetivo MAPE e ele resulta no valor que minimize o MAPE ou que maximize o -MAPE. Logo, o melhor valor encontrado na Tabela 6 é o *score* = -0,0822146, pois é o mais próximo de zero. A combinação associada ao menor escore é a de 50 pontos iniciais e 500 iterações, de forma que os valores em destaque na Tabela 7 são considerados os valores dos hiperparâmetros com melhor calibração via Otimização Bayesiana.

Contudo, essa não é a única forma de encontrar combinações de número de pontos iniciais e número de iterações. Observa-se na Figura 14, que em média, o

resultado da BO é sempre superior ao de todas as calibrações disponíveis, com destaque para as combinações 100_500 e 300_400. Nesses casos, o método de Otimização Bayesiana se mostra superior a todos os demais quando considerada a média.

Figura 14: Resultado médio das comparações - MAPE



Fonte: Autor

CONCLUSÃO

O presente trabalho buscou explorar a Otimização Bayesiana na calibração de hiperparâmetros do submodelo fenológico CROPGRO-*Drybean*, aplicado ao cultivar de feijão-comum BRS *Esplendor*, e obteve excelentes resultados em comparação com as calibrações propostas no Protocolo de Calibração para Modelos de Cultura de Solo e na calibração manual realizada por um especialista da Embrapa Arroz e Feijão, na região de Goiás.

Em média, a Otimização Bayesiana provou ser mais eficiente que a calibração manual, chamada neste trabalho de *Default Value*, que a calibração por estimação separada (*Step 6*) e pela estimação simultânea (*Step 7*). Isso mostra que ela é uma alternativa eficiente para a calibração dos hiperparâmetros do submodelo do DSSAT. Dessa forma, torna-se evidente que a simulação das datas de desenvolvimento fenológico é mais precisa quando se utiliza o método de calibração por Otimização Bayesiana.

O Processo Gaussiano e a estrutura de Aprendizado de Máquina associada a uma função de aquisição nortearam todo o processo de forma eficaz e contribuiu para o sucesso da calibração, tornando o processo mais inteligente e célere que outros métodos de calibração já consolidados no âmbito agrônomo.

REFERÊNCIAS

BOOTE, K. J. et al. The CROPGRO model for grain legumes. In: SINGH, U. et al. (Org.). Understanding Options for Agricultural Production. Dordrecht: Springer, 1998. p. 99-128.

CARGNIN, A.; ALBRECHT, J. C. BRS Esplendor: nova cultivar de feijoeiro comum do grupo comercial preto para o Distrito Federal. Planaltina, DF: Embrapa Cerrados, 2010. 3 p. (Embrapa Cerrados. Comunicado Técnico, 168).

CHAI, T.; DRAPER, N. The relationship between the coefficient of determination and the mean absolute percentage error (MAPE). Journal of Agricultural, Biological and Environmental Statistics, v. 15, n. 4, p. 539–558, 2010.

DE FINETTI, B. Theory of probability: a critical introductory treatment. Chichester: Wiley, 1974.

FAOSTAT. Food and Agriculture Organization of the United Nations. Statistics Division. 2019. Disponível em: <https://www.fao.org/faostat/en/#data/QC/visualize>. Acesso em: 29 de novembro de 2025.

FARIA, A.P.; MODA-CIRINO, V.; BURATTO, J.S.; SILVA, C.F.B. da; DESTRO, D. Interação genótipo x ambiente na produtividade de grãos de linhagens e cultivares de feijão. Acta Scientiarum. Agronomy. Maringá, v. 31, n. 4, p. 579-585, 2009.

GELMAN, A. et al. Bayesian data analysis. 3rd ed. Boca Raton: CRC Press, 2013.

HOOGENBOOM, G. et al. The Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.5 [CD-ROM]. University of Florida, Gainesville, FL, 2012.

JONES, D. R.; SCHONLAU, M.; WELCH, W. J. Efficient Global Optimization of Expensive Black-Box Functions. Journal of Global Optimization, v. 13, n. 4, p. 455–492, 1998.

JONES, J.W.; HOOGENBOOM, G.; PORTER, C.H.; BOOTE, K.J.; BATCHELOR, W.D.; HUNT, L.A.; WILKENS, P.W.; SINGH, U.; GIJSMAN, A.J.; RITCHIE, J.T. The

DSSAT cropping system model. *European Journal of Agronomy*, v. 18, p. 235-265, 2003.

OLIVEIRA, L. F. C. de et al. *Conhecendo a fenologia do feijoeiro e seus aspectos fitotécnicos*. 2. ed. rev. e ampl. Brasília, DF: Embrapa, 2018.

SILVA, O. F.; WANDER, A. E. *O Feijão-Comum no Brasil: Passado, Presente e Futuro*. Embrapa Arroz e Feijão, 2013.

SNOEK, J.; LAROCHELLE, H.; ADAMS, R. P. Practical Bayesian Optimization of Machine Learning Algorithms. In: *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe, Nevada, 3-6 dez. 2012. p. 2951-2959.

VAN OORT, P. A. J.; VAN ITTERSUM, M. K. Calibrating crop growth models: An overview of methods and their application. *Journal of Agricultural Science*, v. 151, n. 2, p. 219–235, 2013.

WALLACH, D. et al. A calibration protocol for soil-crop models. *Environmental Modelling and Software*, v. 180, p. 106147, 2024. Disponível em: <https://doi.org/10.1016/j.envsoft.2024.106147>. Acesso em: 16 de outubro de 2025.

WANG, J. An Intuitive Tutorial to Gaussian Processes Regression. [S. l.: s. n.], 2021. Disponível em: http://www.cs.queensu.ca/~jjewang/teaching/CISC881/CISC881_tutorial.pdf. Acesso em: 25 ago. 2025.

WILSON, S. *ParBayesianOptimization: Parallel Bayesian Optimization of Hyperparameters*. R package version 1.2.6. CRAN, 2022. Disponível em: <https://cran.r-project.org/web/packages/ParBayesianOptimization/ParBayesianOptimization.pdf>. Acesso em: 25 ago. 2025.