



**UNIVERSIDADE FEDERAL DE GOIÁS  
ESCOLA DE AGRONOMIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E  
MELHORAMENTO DE PLANTAS**

**AMBIENTIPAGEM NO MELHORAMENTO DE  
PLANTAS E PLATAFORMA 3,5K DE  
COVARIÁVEIS AMBIENTÍPICAS**

**LUCAS LEMES DE SOUZA PEIXOTO**

Orientador:

**Prof. Dr. Rafael Tassinari Resende**

Coorientador:

**Prof. Dr. Gustavo Eduardo Marcatti**



UNIVERSIDADE FEDERAL DE GOIÁS  
ESCOLA DE AGRONOMIA

## TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

### E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 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 das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) 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 material bibliográfico

Dissertação     Tese     Outro\*: \_\_\_\_\_

\*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

**Exemplos:** Estudo de caso ou Revisão sistemática ou outros formatos.

#### 2. Nome completo do autor

Lucas Lemes de Souza Peixoto

#### 3. Título do trabalho

Ambientipagem no Melhoramento de Plantas e Plataforma 3,5K de Covariáveis Ambientípicas

#### 4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento  SIM     NÃO<sup>1</sup>

[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) autor(a) e ao(à) orientador(a);
- b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação. 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;
- Publicação da dissertação/tese em livro.

**Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.**



Documento assinado eletronicamente por **Rafael Tassinari Resende, Professor do Magistério Superior**, em 03/10/2025, às 14:13, 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 **Lucas Lemes De Souza Peixoto, Discente**, em 06/10/2025, às 23:47, 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](https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **5697065** e o código CRC **CAB879BE**.

---

Referência: Processo nº 23070.027075/2025-76

SEI nº 5697065

**LUCAS LEMES DE SOUZA PEIXOTO**

**AMBIENTIPAGEM NO MELHORAMENTO DE PLANTAS E  
PLATAFORMA 3,5K DE COVARIÁVEIS AMBIENTÍPICAS**

Dissertação apresentada ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas, da Escola de Agronomia, da Universidade Federal de Goiás, como exigência parcial para obtenção do título de Mestre em Genética e Melhoramento de Plantas.  
Área de concentração: Genética e Melhoramento de Plantas.

Orientador:

**Prof. Dr. Rafael Tassinari Resende**

Coorientador:

**Prof. Dr. Gustavo Eduardo Marcatti**

Goiânia, GO – Brasil

Agosto - 2025

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

Peixoto, Lucas Lemes de Souza  
AMBIENTIPAGEM NO MELHORAMENTO DE PLANTAS E  
PLATAFORMA 3,5K DE COVARIÁVEIS AMBIENTÍPICAS  
[manuscrito] / Lucas Lemes de Souza Peixoto. - 2025.  
LXXXIV, 84 f.: il.

Orientador: Prof. Dr. Rafael Tassinari Resende; co-orientador Dr. Gustavo Eduardo Marcatti.

Dissertação (Mestrado) - Universidade Federal de Goiás, Escola de Agronomia (EA), Programa de Pós-graduação em Genética e Melhoramento de Plantas, Goiânia, 2025.

Bibliografia.

Inclui siglas, mapas, abreviaturas, símbolos, gráfico, tabelas.

1. Tipagem Ambiental. 2. Ambientômica. 3. Soilgrids. 4. NASAPOW. 5. WORLDCLIM. I. Resende, Rafael Tassinari, orient. II. Título.

CDU 633



UNIVERSIDADE FEDERAL DE GOIÁS

ESCOLA DE AGRONOMIA

### ATA DE DEFESA DE DISSERTAÇÃO

Ata nº 137 da sessão de Defesa de Dissertação de Lucas Lemes de Souza Peixoto, que confere o título de Mestre em Genética e Melhoramento de Plantas, na área de concentração em Genética e Melhoramento de Plantas.

Aos seis dias do junho de dois mil e vinte e cinco, a partir das quatorze horas, on-line, realizou-se a sessão pública de Defesa de Dissertação intitulada "**Ambientipagem no Melhoramento de Plantas e Plataforma 3,5K de Covariáveis Ambientílicas**". Os trabalhos foram instalados pelo Orientador, Professor Rafael Tassinari Resende (PPGGMP/UFG), com a participação dos demais membros da Banca Examinadora: Doutor Renato Domiciano Silva Rosado (PPGGMP/UFG), membro titular interno e Doutor Gerson dos Santos Lisboa (FCT/UFG), membro titular externo. Durante a arguição os membros da banca fizeram sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido o candidato **Aprovado** pelos seus membros. Proclamados os resultados pela Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora.

#### TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **Rafael Tassinari Resende, Professor do Magistério Superior**, em 09/06/2025, às 10:57, 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 **Gerson Dos Santos Lisboa, Professor do Magistério Superior**, em 09/06/2025, às 12:39, 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 Domiciano Silva Rosado, Usuário Externo**, em 11/06/2025, às 23:10, 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](https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **5386569** e o código CRC **5FBDE58E**.

Referência: Processo nº 23070.027075/2025-76

SEI nº 5386569

*“Aos que preferem edificar em silêncio, como Marco Vipsânio Agripa, cuja grandeza está nas fundações invisíveis, mas indispensáveis.*

*Iis qui tacite aedificare malunt, sicut Marcus Vipsanius Agrippa, cuius maiestas in fundamentis latentibus, sed necessariis, innititur.”*

Em gratidão e reconhecimento, dedico esta obra à minha mãe, **Meire**, ao meu tio **Edson** e ao meu orientador e amigo, **Rafael**.

## AGRADECIMENTOS

*À minha mãe, Meire Jane e ao Duda por todo o carinho e ensinamentos que propiciaram a minha formação, sempre permitindo e influenciando positivamente a continuidade dos meus estudos e meu desenvolvimento profissional.*

*Ao meu Tio Edson, por todo o aconselhamento, bem como por ser essa inspiração.*

*Ao meu orientador, Prof. Dr. Rafael Tassinari Resende, por todo o incentivo, ensinamentos e orientação ao longo da minha trajetória, tanto acadêmica como pessoal.*

*Ao meu coorientador, Dr. Gustavo Eduardo Marcatti, pelos ensinamentos e experiência de vida compartilhada.*

*Aos amigos do Programa de Genética e Melhoramento de Plantas da UFG, pelo conhecimento compartilhado durante esta jornada, bem como pela ajuda e direcionamentos fundamentais ao desenvolvimento deste trabalho.*

*À Universidade Federal de Goiás por toda infraestrutura e apoio institucional indispensáveis à realização deste trabalho.*

*Aos membros da minha banca de qualificação, Prof. Dr. Alexandre Siqueira Guedes Coelho e Prof. Dr. Alisson Neves Harmyns Moreira, pela disponibilidade, pelas valiosas contribuições e pelo tempo dedicado à leitura e avaliação deste trabalho, que certamente enriqueceram e aprimoraram a pesquisa desenvolvida.*

*Aos Prof. Dr. Lee Hickey, Dra. Cibele H. Amaral e Prof. Dr. Yunbi Xu, pela colaboração e contribuição científica por meio do artigo publicado, que serviu como base para o desenvolvimento desta pesquisa.*

*À CAPES, pelo financiamento e concessão da bolsa de estudos, que possibilitou minha dedicação integral à pesquisa.*

**Muito obrigado!**

## SUMÁRIO

|   |    |
|---|----|
| RESUMO.....   | 5  |
| ABSTRACT.....   | 6  |
| 1 INTRODUÇÃO.....   | 7  |
| 2 AMBIENTÔMICA VIA SATÉLITES E SUAS APLICAÇÕES NO<br>MELHORAMENTO DE PLANTAS.....               | 9  |
| 2.1 INTRODUÇÃO.....   | 10 |
| 2.2 INTEGRAÇÃO DA POPULAÇÃO - ALVO DE AMBIENTES NO<br>CONTEXTO DE AMBIENTÔMICA.....             | 13 |
| 2.3 APLICANDO A AMBIENTÔMICA AO MELHORAMENTO: A<br>NECESSIDADE DE AMBIEN TIPAGEM EXAUSTIVA..... | 15 |
| 2.4 SISTEMAS DE SATÉLITES E SENSORIAMENTO REMOTO PARA<br>AMBIENTIPAGEM .....                    | 20 |
| 2.4.1 SATÉLITES E SUA CLASSIFICAÇÃO.....  | 20 |
| 2.4.2 SENSORES DE SATÉLITE PARA AMBIENTIPAGEM.....  | 22 |
| 2.4.2.1 SENSORES ÓPTICOS.....   | 22 |
| 2.4.2.2 SENSORES DE RADAR.....  | 25 |
| 2.4.2.3 SENSORES LIDAR.....   | 26 |
| 2.4.2.4 SENSORES METEOROLÓGICOS.....  | 27 |
| 2.4.3 PRODUTOS DE SENSORIAMENTO REMOTO PARA AMBIENTÔMICA.....                                   | 28 |
| 2.5 PLATAFORMAS DE GESTÃO DE INFORMAÇÕES AMBIENTÔMICAS E CIBER<br>INFRAESTRUTURA.....           | 31 |
| 2.6 AMBIENTÔMICA ASSISTIDA POR INTELIGÊNCIA ARTIFICIAL (IA).....                                | 34 |
| 2.7 AMBIENTÔMICA PARA O MELHORAMENTO DE CULTURAS.....   | 36 |
| 2.7.1 AMBIENTÔMICA PARA O MELHORAMENTO PREDITIVO.....   | 38 |
| 2.7.2 AMBIENTÔMICA EM ESCALA DE CAMPO E OTIMIZAÇÃO DE PROGRAMAS<br>DE MELHORAMENTO.....         | 40 |
| 2.7.3 DADOS TEMPORAIS E ADAPTAÇÃO ÀS MUDANÇAS CLIMÁTICAS.....                                   | 41 |
| 2.8 CONSIDERAÇÕES FINAIS.....   | 43 |
| 2.9 REFERÊNCIAS.....  | 44 |
| 3 DESENVOLVIMENTO DE PIPELINE INTEGRADO DE AMBIENTIPAGEM 3,5K<br>PARA MODELAGEM G×A.....        | 57 |
| 3.1 INTRODUÇÃO.....   | 58 |
| 3.2 MATERIAIS E MÉTODOS.....  | 60 |
| 3.2.1 DETALHAMENTO TÉCNICO POR PLATAFORMA.....  | 63 |
| 3.2.2 ANÁLISE ESTATÍSTICA E EXPLORATÓRIA DAS COVARIÁVEIS<br>AMBIENTAIS.....                     | 64 |
| 3.3 RESULTADOS.....   | 65 |

|     |                           |    |
|-----|---------------------------|----|
| 3.4 | DISCUSSÃO.....            | 73 |
| 3.5 | CONCLUSÃO.....            | 75 |
| 3.6 | REFERÊNCIAS.....          | 76 |
| 4   | CONSIDERAÇÕES FINAIS..... | 78 |

## RESUMO

PEIXOTO, L. L. S. **AMBIENTIPAGEM NO MELHORAMENTO DE PLANTAS E PLATAFORMA 3,5K DE COVARIÁVEIS AMBIENTÍPICAS**. Dissertação (Mestrado em Genética e Melhoramento de Plantas) — Escola de Agronomia, Universidade Federal de Goiás, Goiânia, 2025.<sup>1</sup>

A ambientipagem e a ambientômica têm ganhado destaque por permitirem a caracterização detalhada dos sistemas agrícolas com base em informações multiescalares sobre condições ecológicas. Este trabalho tem como objetivo integrar dados ambientais provenientes de plataformas públicas (NASA POWER, SoilGrids e WorldClim) por meio de um *pipeline* automatizado e escalável, com vistas à construção de malhas espaciais contínuas de covariáveis do espaço agrícola e análise territorial. Inicialmente, apresenta-se a revisão bibliográfica em formato de capítulo, sendo uma versão em português para o artigo publicado na revista *Molecular Plant*, que discute os fundamentos conceituais da ambientômica, as potencialidades do sensoriamento remoto, os desafios para integração dos dados e as perspectivas futuras da área. Em seguida, se desenvolve em um *pipeline* constituído em linguagem Python capaz de realizar a aquisição automatizada, padronização, interpolação espacial e integração de múltiplas fontes de dados ambientais. A partir da matriz gerada, foram aplicadas Análises de Componentes Principais (PCA), destacando gradientes ecológicos relevantes para a recomendação de genótipos. A proposta metodológica permitiu organizar 3.507 camadas ambientais e gerar um conjunto de dados georreferenciado com extração de covariáveis por pixel, adequado para o uso em análises com enfoque em interação genótipos  $\times$  ambientes (G $\times$ A). A estrutura matricial resultante oferece uma interface acessível para melhoristas, mesmo sem domínio avançado em Sistema de Informações Geográficas (SIG). Assim, esta dissertação oferece uma contribuição conceitual e prática para o avanço da ambientipagem no contexto do melhoramento genético, promovendo a integração entre estatística, geoinformação e plataformas de ambientipagem para aplicabilidade dentro dos programas de melhoramento de plantas.

Palavras-chave: Tipagem Ambiental; Ambientômica; Soilgrids; NASAPOWER; WORLDCLIM

---

<sup>1</sup> Orientador: Prof. Dr. Rafael Tassinari Resende. Universidade Federal de Goiás.

<sup>1</sup> Coorientador: Prof. Dr. Gustavo Eduardo Marcatti. Universidade Federal de São João Del-Rei

## ABSTRACT

PEIXOTO L. L. S. **DEVELOPMENT OF AN INTEGRATED 3,5K ENVIROTYPING PIPELINE FOR A G×E MODELING** Dissertation (Master in Genetics and Plant Breeding) – Escola de Agronomia. Federal University of Goiás, Goiânia, 2025.<sup>1</sup>

Envirotyping and enviromics have gained prominence for enabling the detailed characterization of agricultural systems based on multiscale information about ecological conditions. This work aimed to integrate environmental data from public platforms (NASA POWER, SoilGrids, and WorldClim) through an automated and scalable pipeline, targeting the construction of continuous spatial grids of agricultural-space covariates and territorial analysis. Initially, a literature review is presented in chapter format, as a Portuguese version of the article published in the journal *Molecular Plant*, which discusses the conceptual foundations of enviromics, the potential of remote sensing, the challenges of data integration, and future perspectives in the field. Subsequently, a pipeline developed in Python was implemented, capable of performing automated acquisition, standardization, spatial interpolation, and integration of multiple sources of environmental data. From the generated matrix, Principal Component Analyses (PCA) were applied, highlighting ecological gradients relevant for genotype recommendation. The proposed methodology enabled the organization of 3,507 environmental layers and the generation of a georeferenced dataset with pixel-wise extraction of covariates, suitable for analyses focused on genotypes × environment (G×E) interaction. The resulting matrix structure provides an accessible interface for breeders, even without advanced knowledge in Geographic Information Systems (GIS). Thus, this dissertation offers both conceptual and practical contributions to the advancement of envirotyping in the context of genetic improvement, promoting the integration of statistics, geoinformation, and envirotyping platforms for practical application within plant breeding programs.

Keywords: Envirotyping; Enviromics; SoilGrids; NASA POWER; WorldClim

---

<sup>1</sup> Advisor: Prof. Dr. Rafael Tassinari Resende. Universidade Federal de Goiás.

<sup>1</sup> Co-advisors: Dr. Gustavo Eduardo Marcatti. Universidade Federal de São João del-Rei

# 1 INTRODUÇÃO

A produção agrícola enfrenta desafios cada vez mais complexos diante das mudanças climáticas e da intensificação das pressões ambientais sobre os ecossistemas cultivados. A variabilidade espacial e temporal de fatores como temperatura, radiação, umidade e características do solo influenciam diretamente o desempenho fenotípico das plantas, afetando a estabilidade produtiva de cultivares. Nesse cenário, compreender e incorporar a influência do ambiente nos programas de melhoramento genético vegetal torna-se imprescindível. A interação genótipos  $\times$  ambientes ( $G \times A$ ) é central para explicar parte dessa variabilidade, e estratégias capazes de caracterizar ambientes de maneira exaustiva podem otimizar a seleção de genótipos mais adaptados a diferentes condições de cultivo (van Eeuwijk et al., 2016; Resende et al., 2021). O avanço de ferramentas computacionais e o acesso ampliado a dados ambientais provenientes de sensores remotos impulsionam a emergência de abordagens como a ambientômica, que permite integrar múltiplas camadas de dados ambientais ao processo de melhoramento.

A primeira parte da dissertação, apresentada sob a forma de artigo publicado na revista *Molecular Plant*, discute o campo da ambientômica sob uma perspectiva conceitual abrangente, com ênfase na caracterização ambiental por meio de dados orbitais, destacando seu papel no avanço do melhoramento genético. É apresentado as principais fontes e plataformas de aquisição de dados ambientais (como NASA POWER, SoilGrids, MODIS e WorldClim), discutindo suas resoluções e aplicações em diferentes escalas de análise. A ambientômica é discutida como uma terceira abordagem ômica, ao lado da genômica e da fenômica, capaz de alimentar modelos preditivos que consideram a complexa interação  $G \times A \times M$  (genótipos  $\times$  ambientes  $\times$  manejos). São exploradas as potencialidades de sensores ópticos, radar, LiDAR e dados meteorológicos para a caracterização do ambiente em larga escala, e defendida a necessidade de metodologias integrativas, infraestrutura cibernética robusta e colaboração multidisciplinar para o avanço da área (Resende et al., 2024).

Mediante a necessidade de integração das diferentes plataformas públicas, que disponibilizam informações do ambiente é apresentado um *pipeline* automatizado de ambientipagem, voltado à aquisição, padronização e integração de dados ambientais em larga escala. Esse *pipeline* foi desenvolvido em Python e integra diferentes bibliotecas geoespaciais, como Xarray, Rioxarray e GeoPandas, otimizando a manipulação e o processamento de arquivos raster. As plataformas NASAPOWER, 11 SoilGrids e WorldClim são utilizadas como fontes de dados ambientais, compondo uma malha espacial contínua de covariáveis ambientais por pixel. A partir dessa base, foi possível aplicar análises estatísticas multivariadas como a Análise de Componentes Principais (PCA), com vistas à estratificação ambiental, interpretação de gradientes e suporte à recomendação de genótipos. A proposta tem como diferencial a disponibilização de dados já processados em estruturas matriciais familiares a realidade operacional de melhoristas, mesmo aqueles sem experiência em Sistemas de Informação geográfica (SIG).

Dessa forma, esta dissertação está estruturada para combinar a fundamentação teórica e conceitual da ambientômica com o desenvolvimento prático de ferramentas computacionais aplicáveis ao melhoramento vegetal. Ao integrar os aportes de plataformas de dados ambientais com métodos analíticos e automatização de processos, o trabalho contribui para simplificar um dos principais gargalos à adoção da ambientipagem: a complexidade técnica na obtenção e análise dos dados de diferentes plataformas. A convergência do trabalho aqui apresentado reforça a necessidade de soluções que aliem robustez científica e aplicabilidade operacional, possibilitando que programas de melhoramento possam incorporar a variabilidade ambiental em suas estratégias de seleção de maneira mais prática.

## 2 AMBIENTÔMICA VIA SATÉLITES E SUAS APLICAÇÕES NO MELHORAMENTO DE PLANTAS

### RESUMO

Ambientômica refere-se à caracterização de micro e macroambientes com base em grandes conjuntos de dados ambientais. Ao fornecer recomendações genótípicas com extrapolação preditiva em nível específico de local, a ambientômica pode subsidiar decisões no melhoramento de plantas em condições variadas e antecipar a produtividade diante das mudanças climáticas. A integração baseada em ambientômica de estatística, ambientipagem (isto é, determinação de fatores ambientais) e sensoriamento remoto pode ajudar a desvendar a complexa interação entre genética, ambiente e manejo. Para apoiar esse objetivo, a ambientipagem exaustiva, com a geração de perfis ambientais precisos, poderá melhorar significativamente as previsões de desempenho genotípico e o ganho genético em culturas agrícolas. Atualmente, plataformas de gestão informacional já agregam diversos conjuntos de dados ambientais obtidos por meio de sensores ópticos, térmicos, radar e detecção de alcance por luz (LiDAR), os quais capturam informações detalhadas sobre a vegetação, a estrutura da superfície e o relevo. Essa abundância de informações, combinada com dados climáticos de acesso livre, impulsiona pesquisas inovadoras em ambientômica. Embora a ambientômica detenha um imenso potencial para o melhoramento genético, ainda persistem alguns desafios, como: (1) a necessidade de métodos integradoras para coletar sistematicamente dados de campo que possam ser escalonados e ampliados para observações em larga escala via satélite; (2) o desenvolvimento de modelos de inteligência artificial de ponta para integração, simulação e predição de dados; (3) a criação de ciberinfraestruturas capazes de processar grandes volumes de dados em múltiplas escalas, oferecendo interfaces intuitivas para entrega de previsões a usuários finais; e (4) a promoção da colaboração e do compartilhamento de dados entre agricultores, melhoristas, fisiologistas, especialistas em geoinformação e programadores vinculados a instituições de pesquisa. Superar esses desafios é essencial para aproveitar plenamente o potencial dos grandes volumes de dados capturados por satélites, com vistas a transformar a agricultura e o melhoramento genético no século XXI por meio da ambientômica.

Palavras-chave: ambientipagem, melhoramento de precisão, interações genótipo–ambiente, sensoriamento remoto, modelos preditivos, informação ambientômica.

# SATELLITE-ENABLED ENVIROMICS AND ITS APPLICATIONS IN PLANT BREEDING

## ABSTRACT

Enviromics refers to the characterization of micro- and macroenvironments based on large-scale environmental datasets. By providing genotypic recommendations with predictive extrapolation at a site-specific level, enviromics could inform plant breeding decisions across varying conditions and anticipate productivity in a changing climate. Enviromics-based integration of statistics, envirotyping (i.e., determining environmental factors), and remote sensing could help unravel the complex interplay of genetics, environment, and management. To support this goal, exhaustive envirotyping to generate precise environmental profiles would significantly improve predictions of genotype performance and genetic gain in crops. Already, informatics management platforms aggregate diverse environmental datasets obtained using optical, thermal, radar, and light detection and ranging (LiDAR) sensors that capture detailed information about vegetation, surface structure, and terrain. This wealth of information, coupled with freely available climate data, fuels innovative enviromics research. While enviromics holds immense potential for breeding, a few obstacles remain, such as the need for (1) integrative methodologies to systematically collect field data to scale and expand observations across the landscape with satellite data; (2) state-of-the-art AI models for data integration, simulation, and prediction; (3) cyberinfrastructure for processing big data across scales and providing seamless interfaces to deliver forecasts to stakeholders; and (4) collaboration and data sharing among farmers, breeders, physiologists, geoinformatics experts, and programmers across research institutions. Overcoming these challenges is essential for leveraging the full potential of big data captured by satellites to transform 21st century agriculture and crop improvement through enviromics.

Keywords: envirotyping, precision breeding, genotype–environment interactions, remote sensing, predictive models, enviromic information

## 2.1 INTRODUÇÃO

As alterações no clima global e seus consequentes impactos sobre as paisagens agrícolas despertam sérias preocupações em relação à segurança alimentar. O conceito de ambientômica, que aborda a miríade de variáveis ambientais que influenciam o crescimento e o desenvolvimento das plantas, oferece *insights* fundamentais. Estes elucidam como o aprimoramento do melhoramento de culturas pode ser alcançado por meio da decodificação das complexas interações entre genótipos, ambientes e manejos ( $G \times A \times M$ ), visando à adaptação eficaz às mudanças climáticas. Fatores terrestres e atmosféricos influenciam diretamente o crescimento, o desenvolvimento e a produtividade das culturas (Cooper et al., 2014; Xu 2010, 2016), e genótipos que se destacam em uma região podem apresentar baixo desempenho em outra, em razão das condições ambientais específicas (Xu,

2010). Além da compreensão dessas interações  $G \times A$ , o conhecimento das características variáveis do ambiente (Piepho & Blancon, 2023) pode subsidiar predições genômicas direcionadas, estratégias de melhoramento genético e práticas de manejo que impulsionem a produtividade agrícola.

Para avaliar as interações  $G \times A$  e  $G \times A \times M$ , genótipos replicados são geralmente cultivados em ambientes caracterizados por combinações específicas de locais e anos. Os dados resultantes podem ser analisados por meio de abordagens como métodos de adaptabilidade e estabilidade (van Eeuwijk et al., 2016; Crossa et al., 2022), ou por técnicas gráficas, como o biplot GGA (genótipo + interação genótipo-ambiente) e o modelo de efeitos principais aditivos e interação multiplicativa (AMMI), para mapear as respostas de diferentes genótipos a ambientes variados (Neisse et al., 2018; Olivoto et al., 2019). Além disso, os ensaios em múltiplos ambientes (MET – *Multi-Environment Trials*), que utilizam modelos estatísticos para analisar estruturas de variância e covariância entre ambientes, pode detalhar o desempenho genotípico sob condições específicas de cultivo e manejo (Malosetti et al., 2013). A modelagem fatorial analítica também tem ganhado popularidade recentemente, pois acelera a convergência dos modelos na análise de grandes conjuntos de dados com múltiplos ambientes (Krause et al., 2020; Smith et al., 2021). Destaca-se ainda que modelos de crescimento de culturas possibilitam simulações virtuais do comportamento genotípico sob cenários hipotéticos; a integração de informações genéticas e ambientais por meio de modelos mecanísticos ecofisiológicos pode aumentar a precisão dessas simulações (Bustos-Korts et al., 2019; Rincent et al., 2019).

A ambientômica, como uma das ômicas, representa uma abordagem relativamente distinta para avaliar as interações  $G \times A$  e  $G \times A \times M$  com base em dados ambientais (Resende et al., 2021). O termo ambientômica surgiu inicialmente na literatura psiquiátrica, em meados da década de 1990 (Anthony et al., 1995). No contexto do melhoramento genético de plantas, foi mencionado pela primeira vez por Xu (2016), e sua exploração mais aprofundada começou com um *preprint* no bioRxiv em 2019, culminando na publicação de Resende et al. (2021). Desde então, a ambientômica tem ganhado rapidamente popularidade na comunidade de melhoristas, destacando-se em pesquisas conduzidas por diversos grupos (Costa-Neto et al., 2021a; Cooper & Messina, 2021; Crossa et al., 2021; Resende et al., 2022). Um ambientoma, assim como o genoma e o fenoma, corresponde a um conjunto de ambiótipos representado por todos os fatores ambientais que afetam o crescimento e o desenvolvimento de um organismo, abrangendo variáveis

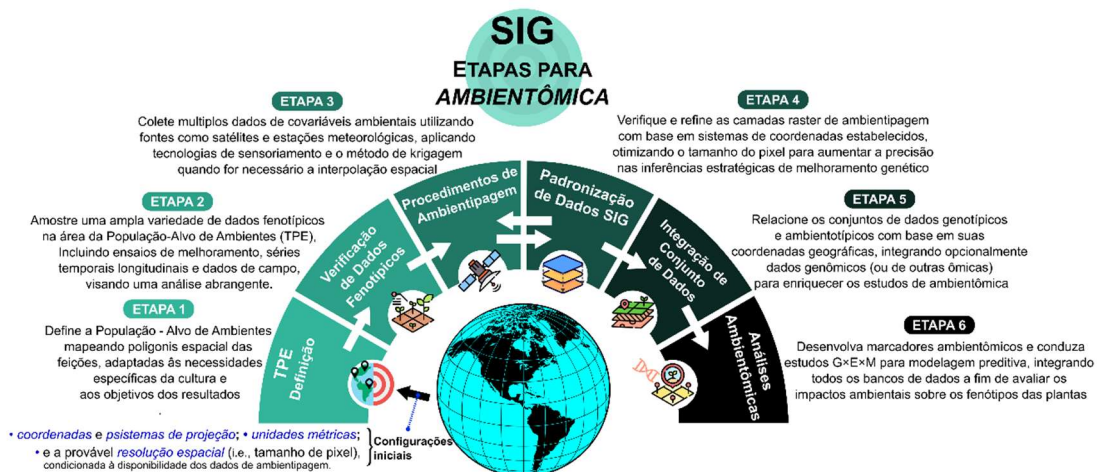
climáticas e da paisagem (Xu, 2016; Costa-Neto & Fritsche-Neto, 2021; Resende et al., 2021). A origem do termo ambiótipo é atribuída a Patten (1991), tendo sido retomada por Beckers et al. (2009) em estudos genéticos com camundongos. Já o conceito de ambientipagem foi proposto por Xu em 2010 e publicado formalmente em 2016 (Xu et al., 2022), com o objetivo de descrever a coleta de dados ambientais para a caracterização de ambientes, conforme também discutido por Cooper et al. (2014). No caso das culturas agrícolas, a abordagem ambientômica prioriza a análise espacial de dados, integrando informações experimentais e de campo (*on-farm*) para validação precisa de modelos em diferentes escalas. Se a ambientipagem for considerada a terceira tecnologia de tipagem (Xu, 2016), ao lado da genotipagem e da fenotipagem, então a ambientômica representa uma terceira abordagem ômica, juntamente com a genômica e a fenômica (Resende et al., 2021).

Sensores de satélite são uma fonte de dados indispensável para a enviromics. Por exemplo, satélites de observação da Terra, como o Landsat, MODIS (*Moderate Resolution Imaging Spectroradiometer*) e Sentinel, monitoram fenômenos de superfície, incluindo clima, composição da vegetação, uso da terra e poluição do ar ou da água (Zhao et al., 2022). Posicionados em órbitas mais altas, os satélites meteorológicos GOES (*Geostationary Operational Environmental Satellites*) e Meteosat fornecem dados sobre condições meteorológicas e climáticas, captando parâmetros como luz, temperatura, umidade e velocidade do vento (Krinitskiy et al., 2023). Além disso, ferramentas de reanálise, como o *Modern-Era Retrospective Analysis for Research and Applications, Version 2* (MERRA-2), utilizam observações globais (Reichle et al., 2017) para calibração, determinação de incertezas e avaliação de produtos de dados, conforme destacado por Brönnimann et al. (2018).

Os dados atualmente derivados de satélites oferecem uma visão detalhada de sítios geográficos, permitindo uma análise mais abrangente da influência do ambiente sobre o desempenho genotípico. Nessa perspectiva, examinamos a relação entre os dados ambientotipos modernos e seus efeitos sobre a ambientômica e o melhoramento de plantas, com ênfase especial nas tecnologias espaciais/satelitais que facilitam a compreensão das interações entre genótipos de culturas e ambiótipes, destacando o papel da ambientômica como fornecedora de informações em escala ômica. Nosso objetivo é atualizar a base de conhecimento, abordando a lacuna de revisões abrangentes diante do rápido avanço tecnológico e das inovações em análise de dados, com vistas a aumentar a precisão e a eficiência no manejo e melhoramento de culturas frente às mudanças climáticas.

## 2.2 INTEGRAÇÃO DA POPULAÇÃO - ALVO DE AMBIENTES NO CONTEXTO DE AMBIENTÔMICA

A etapa inicial na gestão de estruturas de ambienteômica aplicadas a estudos com culturas envolve a identificação da população-alvo de ambientes (TPE – *Target Population of Environments*) (Figura 1, etapa 1). A TPE representa a composição e a frequência dos tipos ambientais dentro de uma região de interesse para programas de melhoramento de plantas, incluindo a variedade de ambientes nos quais os genótipos candidatos serão avaliados quanto ao desempenho sob diferentes condições de crescimento (Cooper et al., 2014; Chenu, 2015). Com a disponibilização de dados de GPS para uso civil, tornou-se possível determinar a variabilidade de fatores ambientais nas escalas regional, da propriedade e da parcela. No contexto da ambienteômica, o foco não está em um único local experimental, como ocorre em abordagens MET (*Multi-Environment Trials*), mas sim na totalidade da TPE, que é concebida como uma forma geoespacial virtual representando diretamente o ambiente real no campo.



**Figura 1.** Etapas essenciais em SIG para a ambienteômica visando o aprimoramento do melhoramento de precisão por meio da integração de dados fenotípicos e ambientais. Este fluxo de trabalho tem início com o estabelecimento da população-alvo de ambientes (TPE) e culmina na integração detalhada de dados para análise em enviromics. Inclui etapas-chave como a validação dos dados fenotípicos, a coleta estratégica de dados de envirotipagem via sensoriamento remoto, a padronização criteriosa dos dados geoespaciais e a análise abrangente dos dados envirotípicos. Essa abordagem sistemática é projetada para refinar a modelagem preditiva e otimizar os resultados agrícolas ao avaliar e explorar a interação entre genótipo, ambiente e práticas de manejo. Adaptado de Resende et al. (2024).

Após a identificação da TPE (população-alvo de ambientes), os dados fenotípicos devem ser analisados em diferentes contextos ambientais (Figura 1, etapa 2). Conduzir um estudo de ambientômica - ou mesmo um estudo abrangente de interações  $G \times E \times M$  - com apenas alguns ensaios representativos em ambientes variados é análogo à realização de um estudo genômico utilizando poucas plantas individuais ou apenas alguns marcadores de polimorfismo de nucleotídeo único (SNPs) (Resende et al., 2021). A gama ambiental deve gerar variação ambientipagem suficiente para viabilizar a análise em ambientômica. Isso envolve a ampliação da coleta de dados para incluir estudos longitudinais e dados coletados diretamente no campo (*on-farm*), oferecendo informações abrangentes sobre o desempenho genotípico em diferentes condições ambientais. Khosla (2023) demonstrou o uso de amostragem em grade e tecnologias com sensores para a coleta de dados ambiótipos juntamente aos dados de desempenho fenotípico de culturas. Sensores de solo aprimorados podem ser utilizados para medições *in situ* rápidas, confiáveis e de baixo custo em escala de parcela. Por meio de experimentos de campo e de laboratório, são coletadas camadas de dados tanto das culturas quanto do ambiente, fornecendo informações sobre as condições genotípicas dentro da TPE. Ao integrar essas diversas fontes de dados, a análise ganha em profundidade e confiabilidade, conectando as definições ambientais teóricas a resultados agrícolas concretos e aplicáveis no mundo real.

Até o momento, a ambientômica tem sido aplicada ao melhoramento de plantas na etapa de seleção e recomendação de cultivares, tendo como foco a população-alvo de ambientes (TPE). Por exemplo, conceitos de ambientômica foram recentemente utilizados para quantificar os efeitos do clima na adaptação de germoplasmas elite de feijão-comum (*Phaseolus vulgaris*) no Brasil, resultando na identificação de limites climáticos e fases de desenvolvimento críticas para cada cenário de produção, orientando a seleção de variedades adaptadas às mudanças climáticas (Heinemann et al., 2022). Em última instância, a ambientipagem precisa ser escalonada para parcelas experimentais pequenas ou até mesmo para plantas individuais, a fim de alcançar o mesmo nível de resolução proporcionado atualmente pela genotipagem e fenotipagem (Xu, 2016). Para alcançar esse objetivo, são necessários três tipos de avanços técnicos e científicos. Primeiro, a ambientipagem de todos os fatores ambientais deve ser viável, provavelmente por meio do acoplamento de sensores equipados em satélites com sensores e sondas terrestres. Segundo, todos os sensores e sondas devem ser acessíveis e possuir resolução, capacidade de processamento e eficiência suficientemente altas para permitir a ambientipagem em nível de planta individual. Terceiro,

são necessárias ferramentas de gerenciamento de informações ambientômicas que estejam equipadas com sistemas de computação robustos e modelagem preditiva assistida por inteligência artificial.

### **2.3 APLICANDO A AMBIENTÔMICA AO MELHORAMENTO: A NECESSIDADE DE AMBIENTIPAGEM EXAUSTIVA**

O termo ambiente é utilizado de diferentes maneiras. Ele pode se referir a: (1) condições naturais que afetam a existência humana, conforme discutido por cientistas ambientais; (2) condições sociais e culturais que moldam a vida individual ou comunitária, conceito pertinente às ciências sociais; (3) sítios geográficos, que apresentam variações anuais; e (4) fatores físicos, químicos, bióticos e abióticos que influenciam o crescimento e o desenvolvimento de um organismo, os quais, em conjunto, constituem um ambiótipo. No contexto da ambientômica, favorecem-se os dois últimos usos, com ênfase na coleta de dados ambientais para locais georreferenciados e na análise de sua variação espacial e temporal. Em particular, considerar as camadas complexas dos ambientes como ambiótipos estabelece a base para a análise criteriosa de modelos preditivos no melhoramento de plantas, nos quais as relações entre fatores ambientais e o desempenho genotípico devem ser analisadas com precisão e profundidade (Figura 1, etapa 3).

Sob a perspectiva da genética quantitativa, o ambiente é um dos dois termos utilizados para explicar a variação fenotípica. O ambientoma é definido como o conjunto completo de condições externas que afetam o desempenho fenotípico (Costa-Neto & Fritsche-Neto, 2021). As abordagens de ambientômica identificam padrões ocultos por meio do próprio ambiótipo e de sua interação com o genótipo, o que é essencial para compreender e aprimorar culturas agrícolas. Em comparação com o genótipo — que é composto por diversos genes responsáveis pela determinação do fenótipo —, o ambiótipo envolve numerosos fatores ambientais, cada um com efeitos distintos sobre o fenótipo (Cooper e Messina, 2021). Alguns desses fatores podem ter efeitos marcantes e ser amplamente previsíveis, como o fotoperíodo (duração do dia e da noite), os padrões de temperatura, a precipitação anual e sazonal, as propriedades do solo e determinados estresses abióticos específicos.

Muito se discute sobre modelos explicativos versus preditivos (por exemplo, em um fórum liderado por Leo Breiman, desenvolvedor do *Random Forest*; Breiman, 2001).

Enquanto os modelos explicativos têm como objetivo compreender as relações causais entre variáveis, os modelos preditivos buscam prever resultados futuros com base em dados passados ou atuais (Shmueli, 2010). Ambas as abordagens — predição e explicação — são importantes para a ambientômica, oferecendo contribuições distintas que, de forma sinérgica, podem potencializar o melhoramento de culturas, tornando-se, portanto, indispensáveis para o avanço da área (Costa-Neto et al., 2023). Entretanto, é importante lembrar que, quando duas variáveis parecem estar relacionadas, assumir que uma delas pode prever com precisão a outra pode levar a inferências populacionais equivocadas, pois aquilo que parece eficaz em um conjunto de dados amostrais pode não se sustentar universalmente. Testar essas suposições é especialmente crucial no melhoramento genético de plantas, em que modelos preditivos são preferidos em relação aos explicativos para identificar genótipos com fenótipos desejáveis, como alto rendimento e resistência a doenças.

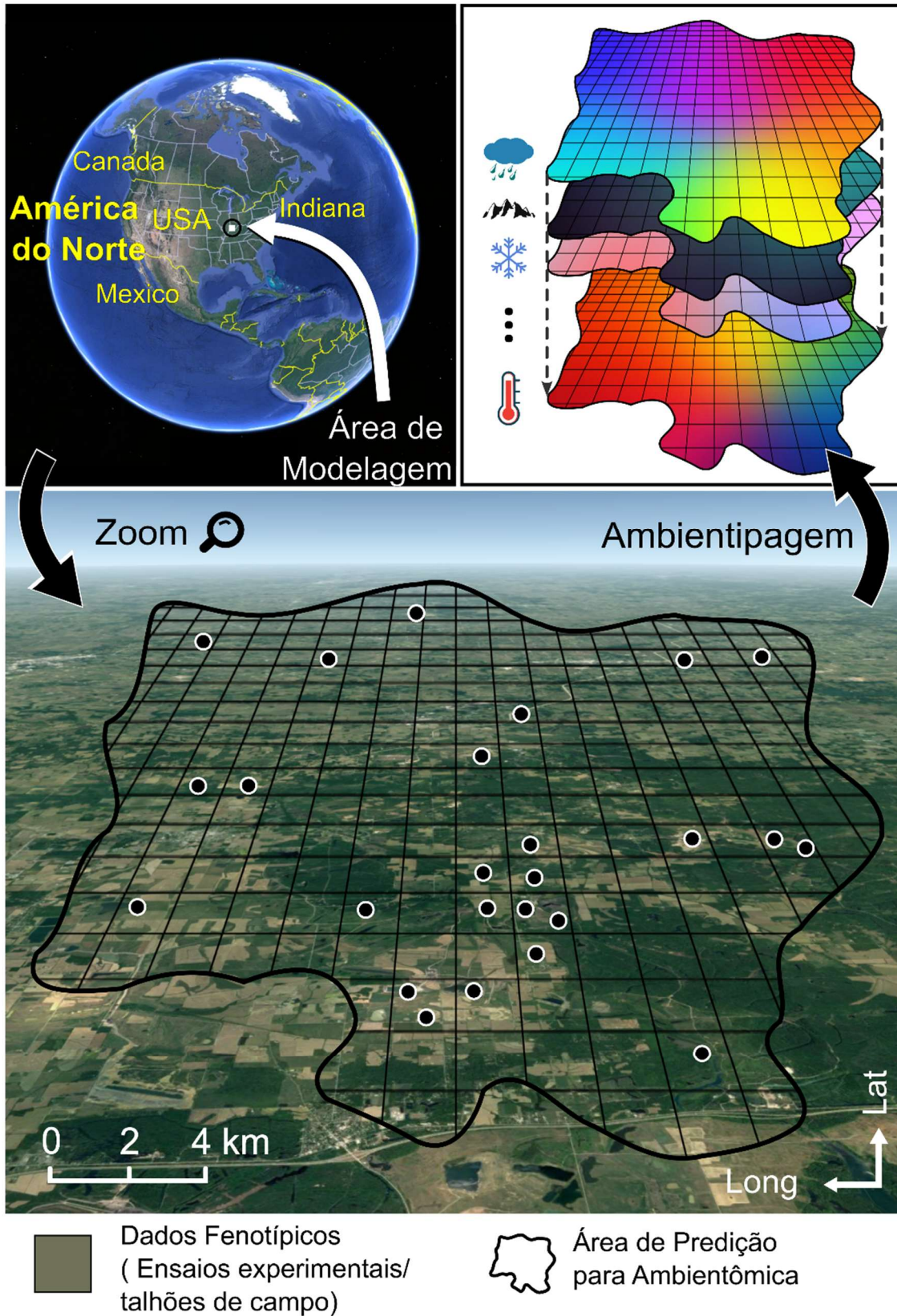
O objetivo do melhoramento de plantas é identificar genótipos com desempenho fenotípico desejável, e os modelos preditivos podem facilitar essa identificação. Por exemplo, a seleção genômica utiliza marcadores genéticos para construir um modelo de predição do mérito genético dos candidatos selecionados em relação a características complexas (Resende, 2024). A seleção genômica apresenta vantagens em relação à seleção assistida por marcadores baseada em QTLs (*locus* de característica quantitativa), especialmente para características complexas controladas por muitos genes de pequeno efeito (Budhlakoti et al., 2022). Para dar suporte à predição genômica, é necessário o uso de genotipagem de alta densidade, já que a adição de mais dados de marcadores pode revelar melhor a variação genética associada à característica de interesse (Sousa et al., 2019). De forma semelhante, informações ambientais mais detalhadas, quando combinadas com dados genômicos (e/ou outras abordagens ômicas), podem aperfeiçoar os resultados preditivos no contexto da ambientômica (Araújo et al., 2024; Callister et al., 2024). O valor de dados exaustivos é exemplificado pelos estudos de Millet et al. (2019), que analisaram o desempenho do milho (*Zea mays*) em diferentes ambientes europeus, e de Li et al. (2021), que integraram fatores ambientais em estudos de associação genômica ampla (GWAS) para culturas como o trigo (*Triticum aestivum*), o milho e a aveia (*Avena sativa*).

Diversas abordagens podem ser utilizadas para caracterizar ambiente. Por exemplo, o caminho do simulador de sistemas de produção agrícola (APSIM), que categoriza ambientes para modelagem agrícola, utiliza dados observados e simulações de modelos para definir tipos distintos de ambientes com base em fatores como clima, solo e

práticas de manejo (Holzworth et al., 2018). Métodos estatísticos para caracterização de tipos de ambiente, como o iclass, baseiam-se na análise de conjuntos de dados observacionais, como a produtividade de culturas, para identificar grupos ou clusters de ambientes. Esses métodos têm como objetivo minimizar a interação cruzada  $G \times A$  na classificação dos tipos de ambiente para fins de pesquisa ou melhoramento genético (Smith et al., 2021).

A ambientômica pode aproveitar tanto a modelagem agrícola quanto os modelos estatísticos, integrando informações multidimensionais provenientes da ambientipagem e da genotipagem por meio do uso de modelos baseados em kernel ou de regressão aleatória (Jarquín et al., 2014; Costa-Neto et al., 2021b; Resende et al., 2021). Li et al. (2022) demonstraram que os métodos utilizados em genômica e fenômica também são eficazes na ambientômica, aplicando essas abordagens para prever o impacto das condições climáticas sobre o desempenho do trigo. É importante destacar que o ambiótipo pode ser considerado um fator independente que influencia significativamente o fenótipo, sua predição e sua seleção — e não apenas como um cofator nas interações  $G \times A$ .

A ambientômica integra diversos métodos  $G \times A \times M$  dentro de sua estrutura, embora seja fundamentalmente distinta de modelos baseados na ecofisiologia, como os modelos de crescimento de culturas ou aqueles derivados exclusivamente de ensaios em múltiplos ambientes (METs). Para viabilizar a análise, é necessário obter dados suficientes de genômica - ambientômica - fenômica ( $G - A - P$ ), o que pode ser alcançado por meio da incorporação de conjuntos de dados desbalanceados aos modelos. De fato, para modelos preditivos, conjuntos de dados desbalanceados e extensos são muito mais vantajosos do que dados escassos e balanceados, nos quais todos os genótipos estão presentes em todos os ensaios (Resende et al., 2021). Estratégias estatísticas podem lidar de forma eficaz com predições genéticas, além de estimar os componentes de variância genética e residual com base em dados desbalanceados de ensaios (Schmidt et al., 2019; Dias et al., 2020). Os procedimentos de ambientipagem se inspiram na decomposição das interações  $G \times A$ , a partir de dados obtidos por ambientipagem exaustiva ou de alta capacidade de *throughput* (Cooper et al., 2014; Xu, 2016). A Figura 2 apresenta uma estrutura hipotética de dados de ambientipagem que poderia apoiar a seleção em escala de ambientoma. A área mostrada está localizada em Indiana, EUA, onde diferentes camadas de dados de ambientipagem podem ser coletadas. Embora este estudo de caso hipotético foque em uma área específica, os conceitos se aplicam a regiões amplas, desde países individuais até continentes inteiros – ou mesmo regiões intercontinentais – com a TPE ajustada aos interesses específicos.



**Figura 2.** Um exemplo de área de predição localizada na América do Norte. São apresentados vinte e cinco pontos de amostragem contendo dados fenotípicos (ensaio experimentais ou realizados em campo). A área de predição está localizada em Indiana, EUA. Diversas camadas de dados de ambientipagem para a área também são representadas. Adaptado de Resende et al. (2024).

Para integrar dados de ambientipagem provenientes de fontes diversas, a cartografia e a geodésia utilizam datums, ou quadros de referência, compostos por um sistema de coordenadas e de referência para representar a superfície terrestre (uma etapa importante da etapa 4 da Figura 1). Os elipsoides e os parâmetros de transformação de datums globais, como WGS84, NAD83, ED50, SIRGAS e ITRF, cada um otimizado para regiões específicas, garantem o referenciamento preciso na sobreposição de dados e em mapas. Embora o WGS84 seja o padrão global, datums regionais continuam a ser utilizados em aplicações locais, especialmente em mapeamentos de alta precisão e estudos geodésicos, sendo que a integração global tem sido favorecida pelos avanços tecnológicos. O conceito de cubo de dados (*data cube*), que utiliza técnicas de geocodificação e correção de imagens, permite o tratamento de diferentes dados geoespaciais com ferramentas como *gdal* e *xcube* para análises multivariadas. Destaca-se que a biblioteca *gdal* possibilita a construção e o processamento sob demanda de cubos de dados a partir de coleções de imagens de satélite (Appel e Pebesma, 2019). Além disso, o pacote Python CGC facilita a co-oclusterização e tri-clusterização de cubos geoespaciais para identificação de padrões nas dimensões espacial, temporal e temática (Nattino et al., 2022).

A aquisição de uma miríade de pontos de dados sobre atributos terrestres e características da superfície, como os padrões de crescimento de culturas agrícolas em nível de propriedade rural, tornou-se viável, especialmente quando se consideram dados temporais, segmentados de acordo com a acessibilidade da plataforma de monitoramento ambiental e integrados às normas climatológicas (Uthes et al., 2020). Importante destacar que a incorporação das normas climatológicas proporciona uma representação mais robusta de médias relevantes para pontos geográficos específicos. Essa abordagem facilita a detecção de desvios climáticos, como alterações incomuns de temperatura ou padrões de precipitação, que justificam investigações mais aprofundadas.

Alguns fatores ambientais são previsíveis, pois são amplamente determinados pela longitude, latitude e altitude, enquanto outros são imprevisíveis devido a variáveis aleatórias, como mudanças nas condições climáticas. Mesmo para os climatologistas mais experientes, prever eventos climáticos anômalos é uma tarefa desafiadora (Brady & Spring, 2021), o que dificulta os esforços para identificar cultivares adequadas ao cultivo diante da variação climática. Tais variações nem sempre resultam em eventos concretos. Dados de satélite oferecem informações temporais, ou seja, medições em intervalos determinados (ver

Tabela Suplementar 1; discutida em detalhe abaixo). Algumas medições são influenciadas por fatores atmosféricos, incluindo (mas não se limitando a) formações de nuvens, poluentes, raios e diversas formas de radiação, como raios solares e raios cósmicos. É importante ressaltar que, embora a ambientômica exija menor granularidade em comparação com a fenômica, demanda atenção aos detalhes em nível de pixel, inclusive nos níveis de parcelas experimentais, blocos e até plantas individuais (Xu, 2016; Xu et al., 2022). Além disso, o conceito de “pan-ambientômica” abrange a ambientômica em múltiplas dimensões, incluindo tempo, espaço, diferentes localidades e estágios de desenvolvimento. Para discussões adicionais sobre as relações complexas entre ambiótipo, ambientipagem, ambientoma, ambientômica e pan-ambientômica, ver Crossa et al. (2021) e Guo & Li (2023).

Uma variedade de fontes fornece dados para estudos de ambientômica. Estações meteorológicas oferecem informações sobre parâmetros climáticos como temperatura, umidade, precipitação e velocidade do vento. Estações hidrológicas monitoram dados relacionados aos recursos hídricos, como os níveis de rios e lagos. Sensores *in situ* e redes de sensores coletam dados em pontos ambientais específicos, muitas vezes associados à Internet das Coisas (IoT). Drones e veículos aéreos não tripulados (VANTs/UAVs) podem realizar coletas de dados em alta resolução em áreas de difícil acesso. O *crowdsourcing*, que envolve a colaboração pública, também pode ser utilizado para coletar e disponibilizar informações. Dispositivos móveis podem captar dados geoespaciais em tempo real. Além disso, coleções de dados históricos fornecem informações sobre padrões e tendências ao longo do tempo, como registros climáticos antigos, mapas e documentos.

## **2.4 SISTEMAS DE SATÉLITES E SENSORIAMENTO REMOTO PARA AMBIENTIPAGEM**

### **2.4.1 SATÉLITES E SUA CLASSIFICAÇÃO**

Cientistas de plantas possuem amplo domínio sobre tecnologias avançadas de genotipagem, bioinformática e, cada vez mais, sobre fenotipagem de alto desempenho. Tanto VANTs (UAVs) quanto satélites podem oferecer imagens de alta resolução, com grande potencial para estimativas precisas em parcelas experimentais de melhoramento de diferentes tamanhos, possibilitando uma fenotipagem padronizada e econômica em

programas de melhoramento (Pinto et al., 2023). É importante destacar que os dados de ambientômica podem, em alguns casos, intersectar-se com a fenômica aérea, capturando a influência combinada de todos os fatores ambientais, mesmo diante de pequenas variações. Um estudo recente que integrou dados de VANTs e do satélite Sentinel-2 obteve avanços significativos na predição da produtividade da cana-de-açúcar (*Saccharum officinarum*), oferecendo uma solução de baixo custo para a gestão de produtividade (Som-ard et al., 2024). Utilizando dados multiespectrais e de micro-ondas do satélite ALOS (Advanced Land Observing Satellite), especificamente dos sensores AVNIR-2 e PALSAR, Domingues et al. (2023) demonstraram a eficácia de redes neurais artificiais para estimar com precisão o volume de madeira em uma plantação comercial de eucalipto no Brasil. No entanto, poucos estudos têm se concentrado na integração entre o melhoramento genético e a ambientipagem utilizando satélites (Xu et al., 2022).

Os satélites são verdadeiras maravilhas tecnológicas projetadas para missões específicas e com características únicas, variando desde os pequenos CubeSats (satélites modulares em forma de cubo, com apenas alguns centímetros, utilizados em missões espaciais de baixo custo) até grandes satélites que pesam várias toneladas (Levchenko et al., 2018). A escolha do tipo de satélite depende da carga útil (*payload*), das restrições de lançamento e das aplicações-alvo, que podem variar desde o monitoramento da superfície terrestre até comunicações globais. Satélites de observação da Terra operam frequentemente em órbitas baixas, a fim de obter imagens de alta resolução, enquanto satélites de comunicação e navegação operam em órbitas mais altas, como as geoestacionárias (Zhao et al., 2022). A estabilidade e orientação dos satélites são asseguradas por sistemas de controle de atitude, que podem variar: alguns utilizam giroscópios e rodas de reação, enquanto outros recorrem a sistemas mais sofisticados, combinando rodas de reação e ímãs.

Os satélites podem ser agrupados com base em sua finalidade. Satélites astronômicos são utilizados para observar o espaço e corpos celestes (por exemplo, os renomados Hubble e James Webb). Satélites de comunicação transmitem sinais de rádio, televisão, telefonia e internet. Satélites de observação da Terra monitoram recursos terrestres, como vegetação, solo, água e clima. Satélites meteorológicos são usados para previsão do tempo e coleta de dados climáticos. Satélites militares são empregados para fins de defesa, espionagem, navegação e comunicação. Por fim, as estações espaciais abrigam astronautas e experimentos científicos no espaço (Jakobsen et al., 2022). Os satélites mais adequados para a ambientômica são os satélites de observação da Terra e os satélites meteorológicos.

Esses podem ser colocados em órbitas baixas, para monitorar a superfície terrestre e coletar dados científicos (Levchenko et al., 2018), ou em órbitas geostacionárias (mantendo-se em posição fixa em relação à Terra, a grandes altitudes), para monitorar condições climáticas e prever tempestades (Krinitskiy et al., 2023). Os satélites de observação da Terra também variam conforme o tipo de sensor: óptico, termal, radar e LiDAR (*Light Detection and Ranging*; Figura 3A). Esses sensores e suas aplicações na ambientipagem são discutidos em detalhe adiante.

## **2.4.2 SENSORES DE SATÉLITE PARA AMBIENTIPAGEM**

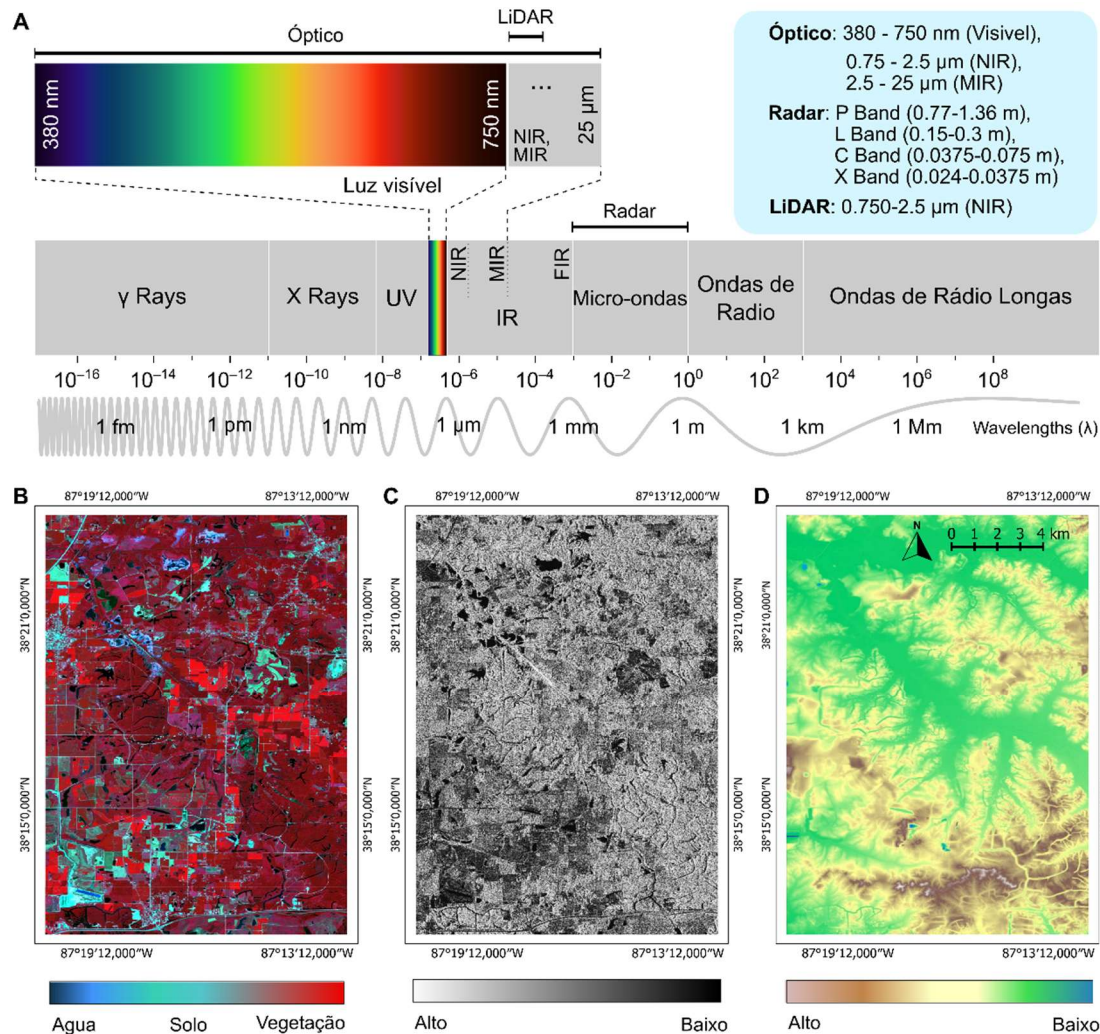
### **2.4.2.1 SENSORES ÓPTICOS**

Sensores ópticos capturam imagens da superfície terrestre utilizando o espectro eletromagnético visível, infravermelho próximo e infravermelho de ondas curtas, permitindo uma análise detalhada por pixel de uma área agrícola. Ao monitorar variáveis estratégicas, como índice foliar, teor de umidade do solo e estado fitossanitário das plantas, os melhoristas obtêm informações relevantes sobre o desenvolvimento das culturas, podendo detectar sinais precoces de estresses bióticos ou abióticos. Ao fornecerem visões aéreas contínuas e precisas, esses sensores ópticos possibilitam a adoção de práticas agrícolas mais eficientes, aumentando a produtividade das lavouras e favorecendo a tomada de decisões embasadas na gestão das áreas cultivadas. Esses sensores são frequentemente utilizados em drones, e a abordagem multiespectral, que explora diversas bandas espectrais, facilita seu uso em experimentos de fenotipagem voltados ao melhoramento genético.

As bandas espectrais dos sensores ópticos abrangem comprimentos de onda úteis para aplicações agrícolas, fornecendo dados ambiótipos valiosos para a ambientômica (Figuras 3A e 3B). A banda azul (450–495 nm) é útil para detectar vegetação saudável e medir nutrientes e clorofila nas plantas. A banda verde (500–575 nm) é utilizada para avaliar a saúde vegetal, permitindo identificar áreas com maior densidade de vegetação. Já a banda verde 1 (500–550 nm) é empregada para analisar a saúde foliar e avaliar as respostas das plantas ao estresse (Yang et al., 2022).

Em alguns casos, o comportamento da vegetação atua como um indicador indireto da qualidade ambiental, sendo que plantas saudáveis e vigorosas indicam ambientes

favoráveis ao desenvolvimento das culturas — conceito conhecido como caracterização baseada em plantas (*plant-based characterization*) (Skovsgaard & Vanclay, 2008). Diferentes bandas espectrais dentro do espectro eletromagnético são fundamentais para a mensuração indireta de atributos das plantas. A região do infravermelho próximo (NIR - 750 -1300 nm) é utilizada para avaliar a saúde das culturas, com base na reflexão das estruturas foliares e do dossel (Kokaly et al., 2003), enquanto as bandas vermelhas (600 - 750 nm) são usadas para estimar os níveis de clorofila, servindo de base para a análise da cobertura vegetal e da produtividade (Venancio et al., 2019). As bandas da borda do vermelho (*red-edge*, que fazem a transição entre o vermelho e o NIR) são empregadas para detectar alterações sutis na vegetação relacionadas ao crescimento e ao estresse. A eficácia dessas bandas na identificação de características das plantas depende do tipo de cultura, dos níveis de estresse e de outras condições ambientais. Índices de vegetação derivados dessas bandas permitem quantificar os impactos ambientais sobre a vegetação: o Índice de Vegetação por Diferença Normalizada (NDVI) avalia a saúde das plantas contrastando a luz NIR e vermelha; O Índice de Vegetação Ajustado ao Solo (SAVI) corrige o NDVI para o fundo do solo, aumentando a precisão em áreas com vegetação esparsa; o Índice de Vegetação Melhorado (EVI) aprimora o NDVI ao corrigir os efeitos da atmosfera e do solo, permitindo detecção mais sensível de mudanças; O Índice de Diferença Normalizada da Água (NDWI) foca no conteúdo de água, sendo útil em estudos hidrológicos (Silva et al., 2020). Em conjunto, esses índices e bandas espectrais fornecem informações abrangentes sobre a qualidade ambiental e a vegetação, demonstrando que a eficácia da análise espectral é condicional ao contexto agrícola específico.



**Figura 3. Imagens capturadas por três diferentes tipos de sensores.** (A) Faixa do espectro coberta por diferentes tecnologias de sensores; são mostrados comprimentos de onda que vão dos raios gama até ondas de rádio longas, juntamente com bandas específicas utilizadas por sensores ópticos, de radar e LiDAR para aquisição de dados de ambientipagem; sensores térmicos cobrem o infravermelho próximo (NIR), infravermelho médio (MIR) e infravermelho distante (FIR) (essa categorização não é padronizada e pode variar entre fontes). (B) Sensor óptico, mostrando uma imagem em cores falsas NIR-R-G do satélite Sentinel-2 com resolução de 10 metros. (C) Sinal de retroespalhamento (micro-ondas) obtido com o sensor Sentinel-1 de radar de abertura sintética (SAR), com resolução aproximada de 10 metros. (D) Topografia obtida por tecnologia LiDAR aerotransportada, com espaçamento médio entre pontos de 1,5 metro, proveniente do IndianaMap (IndianaMap 2011). UV: ultravioleta. Adaptado de Resende et al. (2024).

Sensores ópticos também podem ser utilizados em avaliações da qualidade ambiental, incorporando características da paisagem com base terrestre ("*Earth-based*") (Skovsgaard & Vanclay, 2008). As bandas do infravermelho de ondas curtas (SWIR) são úteis para identificar composições do solo e detectar variações de umidade tanto no solo quanto no dossel vegetal, as bandas SWIR-1, em conjunto com as bandas NIR, são especialmente valiosas para a estimativa do teor de umidade do solo e para o monitoramento da saúde das culturas sob diferentes estados de hidratação (Yue et al., 2019). As bandas

ultravioleta-azul e amarela, embora menos utilizadas, contribuem para pesquisas sobre dispersão de aerossóis atmosféricos (Bautista et al., 2022). Sensores hiperespectrais apresentam alta resolução espectral, cobrindo comprimentos de onda que vão do visível ao NIR e até mesmo ao espectro SWIR, oferecendo dados detalhados por pixel em diversas bandas estreitas. Rizzo et al. (2023) construíram um mapa global de alta definição da cor do solo, com precisão de 30 metros, com base em mais de três décadas de dados dos satélites Landsat e em medições espectrais em solo, estabelecendo uma base sólida para o monitoramento e manejo dos recursos edáficos no futuro. Esses sensores, com resolução espacial limitada a 30 metros, podem ser utilizados para coletar uma grande quantidade de dados ambiótipos, como tipos de solo e níveis de umidade, promovendo avanços significativos na pesquisa ambiental e agrícola (Yue et al., 2019).

#### **2.4.2.2 SENSORES DE RADAR**

Diferentemente dos sensores ópticos, os sensores de radar emitem micro-ondas e detectam a radiação refletida. Esses sensores possuem capacidade de visão noturna, pois conseguem penetrar através das nuvens, e fornecem imagens com resolução espacial mais baixa, sendo, portanto, bem adequados para condições climáticas adversas e para o mapeamento tridimensional (3D) da superfície terrestre (Wang et al., 2018). Por exemplo, a imagem produzida por sensores de radar mostrada na Figura 3C demonstra o fenômeno do retroespalhamento (*backscattering*), ou seja, o retorno da energia de micro-ondas emitida pelo sensor SAR (Radar de Abertura Sintética) do satélite Sentinel-1, que opera na banda C, de volta ao próprio sensor. Esse retorno revela detalhes sobre a topografia, estrutura e umidade da superfície, com base na variação do sinal. Dessa forma, os sensores de radar podem facilitar a ambientipagem ao fornecer informações sobre as condições do solo, como vegetação, umidade do solo e disponibilidade hídrica. Eles também podem reforçar o monitoramento da saúde das plantas, ao detectar mudanças associadas a doenças e estresse hídrico (Emmerik et al., 2017). Os dados desses sensores podem ser utilizados para analisar características fenotípicas associadas ao desempenho das plantas em diferentes ambientes (Al-Turjman, 2019).

O mapeamento de culturas é outra aplicação importante, pois os sensores de radar podem auxiliar na identificação das distribuições espaciais de diferentes culturas em grandes áreas agrícolas, contribuindo para o planejamento de experimentos de

melhoramento genético vegetal. Utilizando dados de um satélite de imagem radar em banda C com polarização dual e um classificador do tipo QUEST (Quick, Unbiased, Efficient Statistical Tree), Mishra et al. (2017) mapearam a distribuição espacial das áreas de cultivo de arroz (*Oryza sativa*), alcançando uma precisão notável de 88,6%. Essa abordagem demonstra-se como uma ferramenta eficaz para o mapeamento de lavouras de arroz e pode, potencialmente, melhorar a previsão de cadeias de suprimento agrícolas.

#### 2.4.2.3 SENSORES LIDAR

Os sensores LiDAR são sensores ativos de sensoriamento remoto que emitem feixes de laser nos comprimentos de onda do verde e do infravermelho próximo (NIR). Esses sensores são capazes de modelar com alta precisão a superfície terrestre (Figura 3A). Satélites equipados com sensores LiDAR emitem pulsos de laser em direção à superfície da Terra e medem o tempo que o pulso leva para retornar ao sensor. Essas informações são utilizadas para calcular a distância entre o satélite e o ponto da superfície que refletiu o pulso. Ao combinar múltiplas medições de diferentes emissões, os satélites LiDAR fornecem informações detalhadas em 3D da superfície, incluindo as alturas do terreno e da vegetação. Modelos ainda mais detalhados podem ser gerados por meio de nuvens de pontos obtidas por sistemas LiDAR aéreos ou terrestres. Dois sistemas LiDAR em órbita, o IceSat-2 da NASA e o GEDI (*Global Ecosystem Dynamics Investigation*), produzem dados na forma de contagem de fótons e amostras de forma de onda da superfície da Terra, que requerem interpolação para a criação de raster. Modelos digitais de elevação de qualidade também podem ser derivados de tecnologias de radar orbital, como o SRTM (*Shuttle Radar Topography Mission*) e o ALOS-PALSAR (*Phased Array type L-band Synthetic Aperture Radar*).

O *Airborne Laser Terrain Mapper* é um sensor aerotransportado que realiza mapeamento topográfico de alta resolução e modelagem do relevo, fornecendo dados precisos. O sensor de terras, vegetação e gelo instalado no satélite ICESat mede a altura da superfície terrestre, da vegetação e do gelo, permitindo estudos sobre mudanças climáticas e o monitoramento ambiental. O GEDI, acoplado ao satélite ICESat-2, mapeia a estrutura vertical das florestas, facilitando o estudo da dinâmica dos ecossistemas terrestres (Alvites et al., 2022). O sofisticado sensor ATLAS LiDAR, também integrado ao ICESat-2, é utilizado para medir a altura do gelo polar e da superfície terrestre, contribuindo para pesquisas climáticas e o monitoramento ambiental.

A aquisição de dados LiDAR é dispendiosa, principalmente devido ao alto custo dos sensores e às despesas associadas à sua integração em dispositivos aerotransportados. Além disso, os arquivos gerados são volumosos. Ainda assim, os sensores LiDAR desempenham diversas funções na agricultura. A Figura 3D ilustra o uso de dados LiDAR aerotransportados para modelar a topografia da área apresentada na Figura 2. Os dados foram obtidos por meio da plataforma gratuita IndianaMap Framework LiDAR (IndianaMap, 2011) e são apresentados com fins comparativos e para a exploração das funcionalidades do LiDAR. Sensores LiDAR podem mapear a topografia de áreas agrícolas, aprimorando a precisão no planejamento do uso do solo e da irrigação (Debnath et al., 2023). Além disso, os dados LiDAR fornecem informações sobre a estrutura e altura das culturas, facilitando a avaliação da saúde das plantas e a detecção precoce de problemas, como estresse hídrico ou doenças. Quando instalados em satélites, os sensores LiDAR podem ser usados para monitoramento em larga escala, cobrindo rapidamente extensos territórios — um recurso valioso na agricultura moderna, que exige gestão inteligente de recursos e monitoramento contínuo para aumentar a produtividade. Já quando montados em VANTs (UAVs), os sensores LiDAR fornecem medições geométricas especialmente vantajosas para a captura da altura das plantas e a modelagem da biomassa em nível de parcela, em programas de melhoramento genético. No entanto, ainda não se sabe ao certo se o mesmo nível de resolução pode ser alcançado com sensores LiDAR baseados em satélites.

#### **2.4.2.4 SENSORES METEOROLÓGICOS**

Satélites meteorológicos equipados com um conjunto específico de sensores monitoram as condições atmosféricas. Câmeras infravermelhas térmicas de alta resolução medem a temperatura da superfície terrestre, auxiliando em estudos climáticos e na detecção de variações térmicas. As câmeras térmicas geralmente apresentam baixa resolução espacial e espectral. Uma exceção notável é a missão ECOSTRESS da NASA, que possui capacidade espacial de 70 metros. Os produtos de nível inferior da missão ECOSTRESS incluem informações sobre a eficiência do uso da água e o índice de estresse por evaporação (*evaporative stress index*) (Fisher et al., 2020), que contribuem para elucidar a capacidade de diferentes genótipos lidarem com condições mais secas.

Radiômetros medem a radiação solar e infravermelha, fornecendo informações importantes sobre o clima e a energia solar incidente na Terra. Radiômetros de micro-ondas medem a temperatura da superfície do mar e outras características oceânicas, contribuindo para estudos oceanográficos e para a previsão de fenômenos marinhos. Sondas atmosféricas são utilizadas para medir temperatura, umidade e pressão em diferentes altitudes, oferecendo informações sobre os processos atmosféricos e facilitando a previsão do tempo. Satélites meteorológicos também incluem imageadores com alta resolução temporal, permitindo imagens detalhadas da cobertura de nuvens e dos padrões climáticos, fornecendo dados essenciais para a previsão e o monitoramento de eventos climáticos extremos.

Os sensores de superfície e sensores meteorológicos se complementam, facilitando o estudo e o monitoramento dos ambientes terrestres e atmosféricos. O uso combinado de tecnologias avançadas, como câmeras ópticas, sensores multiespectrais e hiperespectrais, SAR, radiômetros e sensores atmosféricos, proporciona uma visão ampla e precisa da Terra e de seus fenômenos climáticos. Colaborações contínuas nessas áreas contribuirão tanto para a pesquisa básica quanto para a aplicação de dados de satélite na agricultura e no melhoramento de plantas, promovendo uma compreensão mais profunda das interações entre genótipos e condições agroclimáticas e apoiando a tomada de decisão em setores relacionados à ambientômica, como agricultura, ciência do clima, meteorologia e políticas ambientais.

### **2.4.3 PRODUTOS DE SENSORIAMENTO REMOTO PARA AMBIENTÔMICA**

O panorama atual do sensoriamento remoto oferece uma ampla variedade de produtos, cada um desenvolvido para aplicações específicas (Khanal et al., 2020; Lechner et al., 2020; Weiss et al., 2020). Esses produtos (detalhados na Tabela Suplementar 1) são provenientes de missões de monitoramento terrestre e, frequentemente, envolvem múltiplos satélites. Seus diversos sensores capturam “cenas” - imagens ou conjuntos de dados que representam áreas da superfície terrestre em momentos específicos - fundamentais para a ambientipagem. Essas cenas são apresentadas em formato raster, que incorpora coordenadas geográficas ou métricas. A cobertura espacial e os intervalos de captação de uma cena geralmente são definidos durante o planejamento da missão, embora alguns produtos ofereçam flexibilidade para customização. As cenas podem incluir múltiplas bandas espectrais, permitindo captar variações temporais e espectrais, o que enriquece a

dimensionalidade dos dados. A escolha de um produto de sensoriamento remoto para aplicações em ambientômica requer o entendimento dos tipos de resolução (espacial, temporal, espectral e radiométrica), os quais determinam a adequação do produto para finalidades específicas (Jensen, 2009; Khanal et al., 2020).

A resolução espacial determina o quão bem o menor objeto identificável em uma cena de sensoriamento remoto pode ser distinguido. Essa resolução está intimamente ligada ao tamanho do pixel, sendo que pixels menores oferecem maior nível de detalhe. Uma alta resolução espacial permite detectar características específicas, como deficiência hídrica em determinadas culturas, sendo ideal para a análise de áreas reduzidas, como unidades experimentais de amostragem (Jensen, 2009; Khanal et al., 2020). No entanto, produtos com alta resolução espacial (pixel < 5 m) geralmente implicam maiores custos, tanto financeiros quanto computacionais, exigindo mais recursos para o processamento dos dados e, frequentemente, cobrindo áreas menores por cena. A seleção da resolução espacial em um programa de melhoramento genético é influenciada por fatores como o tipo de cultura, escala de produção, fonte dos dados e estágio do ciclo de melhoramento. Por exemplo, o espaço ocupado por populações de plantas pode variar significativamente, sendo que populações em gerações iniciais ocupam menos área em comparação com gerações mais avançadas (por exemplo, F5 ou F6; Chaves e de Miranda Filho, 1992).

A resolução temporal (frequência de captura de imagens no mesmo local) afeta a capacidade de monitorar mudanças ao longo do tempo. Uma resolução temporal mais alta significa intervalos mais curtos entre as capturas, como exemplificado pelo ciclo de 16 dias da missão Landsat (Jensen, 2009). No entanto, o uso prático das imagens pode ser comprometido por condições atmosféricas, especialmente cobertura de nuvens, que pode impedir a coleta de dados — sobretudo na estação chuvosa, quando as plantas estão em fases críticas de desenvolvimento. Esse desafio é particularmente evidente para sensores ópticos. A escolha da resolução temporal, portanto, depende de necessidades específicas, como o ciclo de vida da cultura, o nível de detalhe desejado ao longo desse ciclo e os estágios fenológicos de interesse (Yang et al., 2022). Uma alta resolução temporal aumenta a probabilidade de obtenção de imagens utilizáveis e livres de nuvens em períodos críticos, como durante o pico de crescimento vegetativo, o que é essencial para análises precisas e para a tomada de decisão fundamentada (Lechner et al., 2020).

A resolução espectral é determinada pela capacidade do sensor de distinguir diferentes comprimentos de onda ao longo do espectro eletromagnético (Jensen, 2009). As

bandas do infravermelho próximo (NIR) e vermelha são essenciais para o cálculo de índices de vegetação (como o NDVI), que auxiliam na avaliação do vigor das plantas e dos níveis de estresse. As bandas azul e verde contribuem para imagens em cores reais (RGB: Vermelho, Verde e Azul) e para índices relacionados à vegetação e ao solo. As bandas do infravermelho médio ajudam na avaliação do conteúdo hídrico das plantas, enquanto as bandas termais são fundamentais para medir a temperatura da superfície terrestre. Um produto de sensoriamento remoto ideal combina bandas visíveis (RGB) e NIR, permitindo estudos abrangentes sobre vegetação e avaliações ambientais (Silva et al., 2020; Voitik et al., 2023). As bandas do infravermelho médio e termais são especialmente úteis para estudos sobre resistência à seca. Os avanços tecnológicos levaram ao desenvolvimento de sensores multiespectrais, que oferecem ampla cobertura espectral com número limitado de bandas, e de sensores hiperespectrais, que proporcionam alta resolução espectral com mais de 100 bandas. Embora os sensores hiperespectrais ainda não sejam amplamente utilizados, seu potencial para identificar características específicas das plantas ou condições ambientais é significativo, prometendo avanços futuros no monitoramento preciso de plantas e do ambiente (Terentev et al., 2022).

Embora as resoluções espacial, temporal e espectral sejam fatores primários na escolha de produtos de sensoriamento remoto para ambientômica, outros fatores também são relevantes, como o custo dos produtos, especialmente entre opções gratuitas e pagas. Produtos com alta resolução espacial (<5 m) e alta resolução temporal (<5 dias) raramente são gratuitos, sendo que os produtos com alta resolução temporal geralmente apresentam resolução espacial moderada (>30 m). No entanto, para muitas aplicações em ambientômica, dados com resolução ultra-alta podem não ser necessários, e é possível aproveitar dados obtidos para outros fins, como planejamento de infraestrutura ou modelagem de previsão de safras. A evolução do poder computacional e das ferramentas de gerenciamento de dados tende a tornar produtos de alta resolução mais acessíveis no futuro. Além disso, o *status* operacional das missões satelitais deve garantir a viabilidade de longo prazo dos métodos de ambientômica tecnologias de sensores, mesmo com a descontinuação de alguns sensores, costumam ser substituídas, garantindo a disponibilidade contínua de dados. O uso de uma ampla variedade de fontes de sensoriamento remoto aumenta a precisão das análises ambientais e reduz a dependência de fontes únicas de dados. Por fim, o tipo de sensor — seja óptico de altíssima resolução, óptico convencional, radar, LiDAR ou meteorológico — representa uma decisão fundamental no sensoriamento remoto, já que cada um oferece

vantagens específicas para determinadas aplicações, conforme resumido na Tabela Suplementar 1.

Escolher a melhor ferramenta de sensoriamento remoto para a ambientômica requer a compreensão de como os fatores ambientais influenciam a capacidade produtiva de um genótipo. Isso pode ser avaliado por meio de estratégias baseadas no ambiente (“*Earth-based*”) e nas plantas (“*Plant-based*”), conforme descrito por Skovsgaard & Vanclay (2008). As avaliações *Earth-based* se concentram em características físicas, como o clima, a topografia e o solo. Sensores meteorológicos coletam dados climáticos (por exemplo, precipitação e temperatura), enquanto sensores de radar e LiDAR são valiosos para modelagem topográfica, oferecendo informações sobre atributos do terreno e tipos de solo. Já as avaliações *Plant-based* utilizam sensores ópticos para avaliar características relacionadas às culturas, correlacionando bandas específicas do espectro óptico e índices derivados com a produção agrícola e fatores de estresse (como doenças e déficit hídrico; Tomar et al., 2014; Khanal et al., 2020; Voitik et al., 2023).

## **2.5 PLATAFORMAS DE GESTÃO DE INFORMAÇÕES ENVIROMICAS E CIBERINFRAESTRUTURA**

O processo de extração, transformação e carregamento (ETL) também pode ser utilizado para lidar com a diversidade e complexidade dos dados ambientais (Aydinoglu, 2016). A gestão desses dados pode ser dividida em três etapas principais: Extração (representada na Figura 1 pelos passos 1 - 2): consiste na coleta de informações provenientes de diversas fontes, como satélites e sensores (por exemplo, MODIS e Sentinel-2), estações meteorológicas e repositórios climáticos (como MERRA-2 e ERA5), abrangendo dados sobre clima, radiação, solo, vegetação e topografia. Transformação (Figura 1, passos 3 - 5): os dados são processados e preparados para análises avançadas por meio de procedimentos como limpeza, normalização, integração, extração de características e cálculos de índices relevantes. Carregamento (Figura 1, passo 6): os dados transformados são inseridos em um ambiente analítico apropriado, como um Sistema de Informações Geográficas (SIG) ou plataforma de dados ambientais, fornecendo uma base sólida para análises mais detalhadas. Nesta seção, abordamos a etapa de extração.

Dados ambientais para a ambientômica podem ser extraídos de plataformas (ou repositórios) que agregam conjuntos de dados e sistemas de análise, reunindo informações abrangentes e retrospectivas sobre o ambiente e a superfície terrestre. Plataformas de gestão da informação integram uma variedade de fontes de dados, como observações por satélite, medições de superfície e dados de modelos climáticos, para fornecer informações ambíóticas detalhadas, como variáveis meteorológicas e dados sobre o clima global e as mudanças climáticas em escala regional ou global. Grande parte desses dados, como os provenientes das missões EOS, LANDSAT, Sentinel, ALOS, GEDI e RADARSAT, pode ser facilmente adquirida por meio da plataforma Google Earth Engine (Velasco-Montoya et al., 2023). Além disso, códigos de SIG facilitam a colaboração entre projetos de *software*. A organização *Open Geospatial Solutions* no GitHub hospeda projetos de código aberto desenvolvidos e mantidos por uma comunidade de especialistas em software geoespacial, os quais são gratuitos para uso e modificação, sob licença do *Massachusetts Institute of Technology* (MIT). Para mais informações, consulte a página Open Geospatial Solutions no GitHub, mantida por Wu, Aybar e Brown.

O MERRA-2 é um conjunto de dados climáticos de longo prazo que oferece uma avaliação retrospectiva e abrangente das condições atmosféricas passadas, com informações detalhadas sobre variáveis meteorológicas como temperatura, umidade, vento e pressão atmosférica (Gelaro et al., 2017). Da mesma forma, o ERA5, desenvolvido pelo Centro Europeu de Previsão do Tempo de Médio Prazo (ECMWF), fornece dados de reanálise climática global de alta resolução, cobrindo um longo período e permitindo análises climáticas em escala global (Hersbach et al., 2020).

O Reanalysis NCEP/NCAR, uma colaboração entre o *National Centers for Environmental Prediction* (NCEP) e o *National Center for Atmospheric Research* (NCAR), disponibiliza um conjunto de dados consistente e abrangente com décadas de informações climáticas. O Serviço de Mudança Climática Copernicus (C3S), parte do programa Copernicus da União Europeia, oferece dados climáticos fundamentais, incluindo observações por satélite, dados de superfície e resultados de modelos climáticos, para análise das mudanças climáticas e seus impactos. O MODIS, sensor satelital a bordo das missões Terra e Aqua da NASA, é uma ferramenta para monitoramento ambiental e climático que adquire dados espectrais e imagens de alta resolução temporal da superfície terrestre, oferecendo uma visão detalhada e abrangente das condições do planeta. Essas plataformas são fundamentais para pesquisas climáticas, estudos sobre mudanças ambientais, previsões

meteorológicas e análises retrospectivas, fornecendo dados que contribuem para uma melhor compreensão do clima global e de sua complexidade ao longo do tempo.

Diversos pacotes em R e Python podem ser utilizados para coletar dados de diferentes fontes para modelagem ambiótica. A biblioteca Requests do Python (Reitz, 2024) realiza requisições HTTP a serviços *web* ou APIs, incluindo dados ambientais. A biblioteca pyModis (Delucchi e Neteler, 2013) permite extrair dados do satélite MODIS, e o pacote Sentinelsat extrai dados dos satélites Sentinel. O Geopy possibilita a extração de informações geográficas, enquanto bibliotecas de raspagem da *web* como BeautifulSoup e Scrapy facilitam a coleta de dados diretamente de sites (Kouzis-Loukas, 2016). O pacote Pyproj auxilia na extração e transformação de coordenadas geoespaciais. Para mais detalhes sobre esses e outros pacotes Python para análise SIG, consulte Westra (2016).

O pacote nasapower do R (Sparks, 2018) fornece acesso aos dados da NASA POWER, permitindo a extração de informações climáticas e meteorológicas. Os pacotes raster, sf e terra são ferramentas robustas para a extração e o gerenciamento de dados geoespaciais do tipo raster (Hijmans et al., 2022). O pacote stars pode ser utilizado para extrair dados espaço-temporais (Pebesma e Bivand, 2023), enquanto o pacote sen2r permite a extração de dados dos satélites Sentinel-2 (Ranghetti et al., 2020). Alguns pacotes também auxiliam no download de dados, tornando-os facilmente acessíveis, como o EnvRtype (Costa-Neto et al., 2021a) e o SoilType (Fritsche-Neto, 2023).

Algumas plataformas de gerenciamento de dados oferecem recursos e conjuntos de dados que podem ser estrategicamente utilizados para obter informações de ambientipagem. Utilizando classificações climáticas detalhadas descritas por Köppen–Geiger (Cui et al., 2021) e Ecorregiões (Dinerstein et al., 2017), é possível categorizar ambientes de cultivo com base em condições climáticas específicas. Os mapas de propriedades do solo do SoilGrids (Poggio et al., 2021) fornecem informações sobre as propriedades físicas e químicas de múltiplas camadas do solo. Dados sobre energia solar do nasapower (Sparks, 2018), bem como informações sobre vento, precipitação, umidade sazonal, sazonalidade climática e até eventos extremos via WorldClim1 (Hijmans et al., 2005) e WorldClim2 (Fick e Hijmans, 2017), podem ser usados para modelar as complexas relações que afetam o crescimento e desenvolvimento das plantas. A *Environmental Data Initiative* (Gries et al., 2023) fornece acesso a uma grande variedade de dados ambióticos, enriquecendo os estudos em ambientômica com informações sobre as condições ambientais passadas e atuais. O NASA GeneLab (Berrios et al., 2021) também contribui para a coleta e

organização de dados ômicos, disponibilizando esses dados oriundos de missões espaciais e experimentos análogos, promovendo descobertas científicas e esclarecendo os efeitos dos ambientes espaciais sobre a biologia. Todos esses recursos capacitam os pesquisadores a considerar uma infinidade de fatores ambientais, aprimorando nossa compreensão das interações genótipos  $\times$  ambientes ( $G \times A$ ) e facilitando a seleção de plantas para uma agricultura sustentável e para a conservação da biodiversidade.

Apesar da disponibilidade de numerosos repositórios de dados de satélite, o processamento e a análise de grandes volumes de dados ainda exigem considerável poder computacional e conhecimento especializado em ciência de dados da Terra. A ciberinfraestrutura necessária para essas atividades inclui computação de alto desempenho e serviços em nuvem, como *Amazon Web Services*, *Planetary Computer/Azure* e *Google Earth Engine*. A expectativa de aumento no uso de ferramentas web, como aplicações do Google Earth Engine e painéis R Shiny, na próxima década, ressalta a crescente necessidade de ferramentas avançadas que tornem a análise de dados mais acessível. Nesta era da análise de big data, a capacidade de aproveitar produtos de ambientipagem com finalidades específicas permitirá que pesquisadores se engajem ativamente na revolução dos dados abertos de satélite, facilitando a tomada de decisões informadas nas áreas ambiental e agrícola (Khanal et al., 2020; Vance et al., 2024).

## **2.6 AMBIENTÔMICA ASSISTIDA POR INTELIGÊNCIA ARTIFICIAL (IA)**

Como discutido anteriormente, dados de diversas fontes estão sendo combinados de forma integrada graças a plataformas que oferecem acesso uniforme aos dados, soluções de armazenamento, interfaces baseadas em aplicações e middleware, facilitando a fusão dos dados genotípicos - ambiótipos - fenotípicos ( $G-A-P$ ) em redes de conhecimento abrangentes (Lund, 2020). Essas ferramentas utilizam algoritmos de mineração de dados para integrar diferentes fluxos de dados (Marsh et al., 2021). Técnicas como concatenação, transformação e integração baseada em modelos possibilitam a fusão eficaz de conjuntos de dados (Picard et al., 2021). Conjuntos de dados multiômicos tornam esse processo desafiador devido aos seus diferentes formatos, escalas e dimensões, além de frequentemente serem ruidosos, esparsos e coletados em condições distintas. Para lidar com essas complexidades, são necessários padrões internacionais e algoritmos de aprendizagem profunda (*deep*

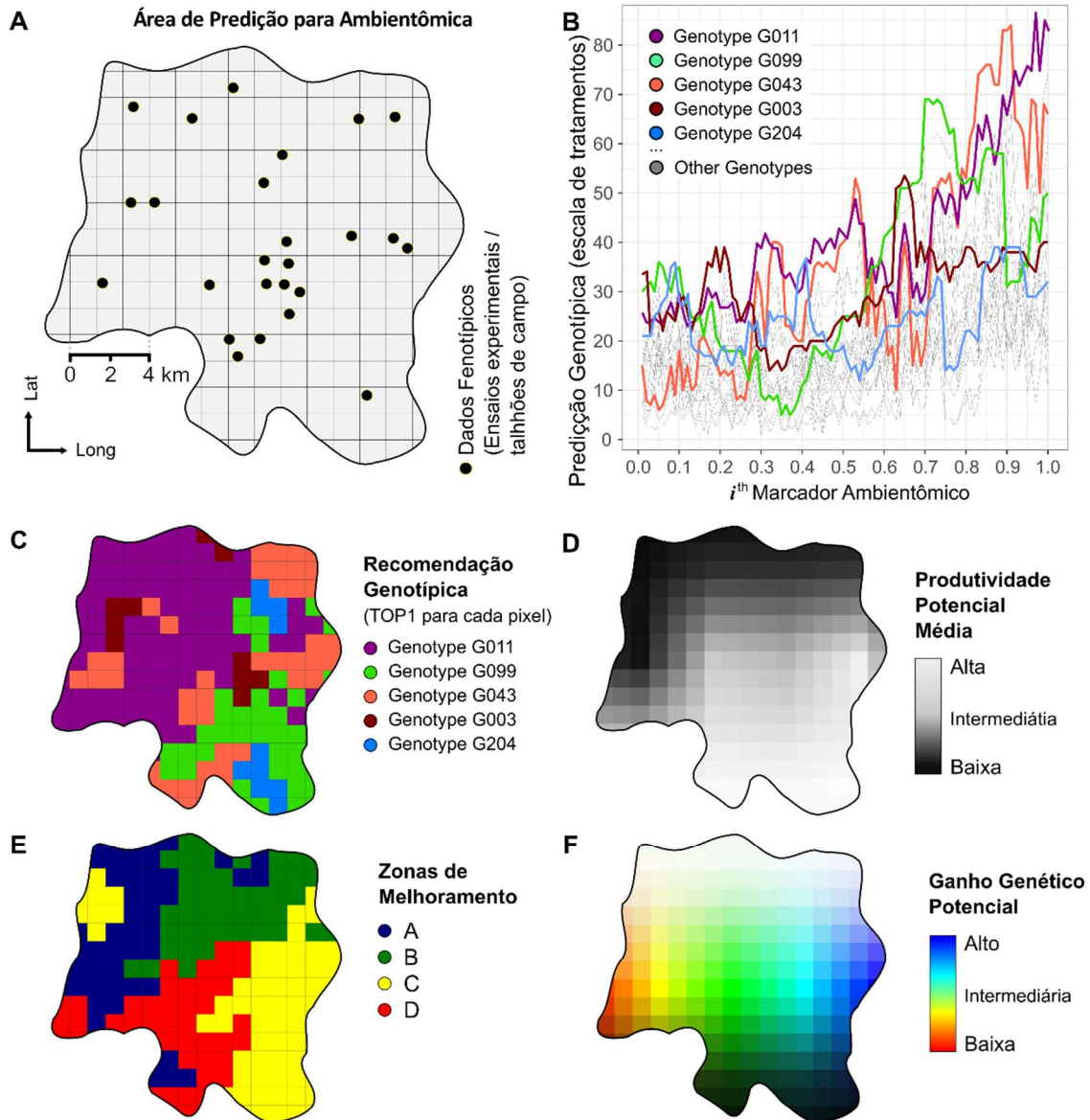
*learning*) para manejar padrões não lineares e facilitar a integração dos dados (Selby et al., 2019; Montesinos-López et al., 2022). A IA surgiu como uma estratégia eficaz para superar as complexidades dos conjuntos de dados de ambientômica, melhorando as previsões no melhoramento genético vegetal por meio da integração dos dados G - A - P. Isso envolve considerar as nuances estruturais dos dados de melhoramento e aplicar métodos estatísticos para otimizar as previsões e os processos de tomada de decisão no melhoramento de plantas (Xu et al., 2022).

Além dos métodos tradicionais de predição (como modelos mistos e/ou bayesianos), os métodos de IA e aprendizado de máquina (*machine learning – ML*) podem auxiliar a ambientômica ao aprimorar as previsões genotípicas e as recomendações (Resende et al., 2021; Costa-Neto et al., 2023). Tanto a IA quanto o ML estão provocando uma mudança de paradigma no geoprocessamento (SIG) e no melhoramento genético. No geoprocessamento, a IA utiliza algoritmos avançados e modelos de ML para extrair conhecimento a partir de dados espaciais, possibilitando análises mais precisas e decisões mais informadas em áreas como monitoramento ambiental e gestão de desastres. A IA pode melhorar a precisão do mapeamento, o reconhecimento de padrões espaciais e a modelagem preditiva, revolucionando a forma como compreendemos e interagimos com as informações geográficas (Khan et al., 2022; Montesinos-López et al., 2022). Técnicas emergentes de IA são úteis para identificar e selecionar características desejáveis em conjuntos de dados genéticos. Seu potencial para melhorar a seleção de genótipos, otimizar a produtividade de culturas e ampliar a adaptabilidade ao clima é uma área de pesquisa ativa (Hayes et al., 2023; Negus et al., 2024).

A sinergia transformadora entre Geoprocessamento + IA (GeoAI) (Song et al., 2023) e o melhoramento genético vegetal tem o potencial de enfrentar desafios globais urgentes relacionados à segurança alimentar e à gestão sustentável do uso da terra. Redes neurais artificiais e outras técnicas de IA têm sido aplicadas com sucesso para diversos fins nas geociências e na engenharia geotécnica (Noack et al., 2014; Kim et al., 2019; Samui, 2020). Além disso, diversos estudos demonstraram o desempenho superior dos algoritmos XGBoost e Random Forest na predição de propriedades geológicas (Naghbi et al., 2020; Zhang et al., 2021). Revisões notáveis de Negus et al. (2024) e Khan et al. (2022) destacam as oportunidades de explorar a IA no melhoramento genético vegetal, o que pode transformar o aprimoramento de culturas e levar a avanços significativos na agricultura.

## 2.7 AMBIENTÔMICA PARA O MELHORAMENTO DE CULTURAS

Após a identificação dos alvos ambientais, coleta de dados fenotípicos e aquisição dos dados ambientais, a próxima etapa da ambientômica consiste na fusão dos conjuntos de dados (Figura 1, etapa 5). Essa etapa combina as informações ambientais com as observações fenotípicas e, possivelmente, dados genômicos ou de outras abordagens ômicas. A principal questão é garantir que as coordenadas geográficas dos locais com dados fenotípicos utilizem o mesmo sistema de coordenadas (ou seja, o mesmo *Datum*) dos dados de ambientipagem. Essa tarefa assegura uma avaliação aninhada que minimiza o ruído analítico e esclarece as interações entre fatores, estabelecendo as bases para análises abrangentes em ambientômica. A integração de dados multidimensionais, tecnologias de big data e inteligência artificial possibilita o desenvolvimento de um esquema de melhoramento inteligente e integrado G–A–P (Genótipo – Ambiente – Fenótipo), conduzindo a previsões fenotípicas mais precisas e maiores ganhos genéticos por meio de plataformas integrativas de melhoramento e iniciativas de código aberto (Xu et al., 2022; Vance et al., 2024). Ainda assim, as interações  $G \times A \times M$  (Genótipos  $\times$  Ambientes  $\times$  Manejos) representam desafios para a otimização dos ganhos genéticos e da produtividade das culturas. Impulsionado pelos avanços nas tecnologias genéticas, genômicas e de sensoriamento remoto, esse cenário pode levar ao surgimento da ambientômica para a predição de caracteres complexos (Cooper e Messina, 2021).



**Figura 4. Uma análise hipotética de ambientômica.** Esta figura serve como um recurso para decisões locais de melhoramento genético baseadas em informações. **(A)** Área de predição para ambientômica ou a População Alvo de Ambientes (TPE). **(B)** Resultado de um modelo de regressão aleatória infinitesimal baseado em ambientômica. **(C)** Recomendações genotípicas: classificação completa de todos os materiais genéticos avaliados, com extrapolação preditiva para todos os pixels da área. **(D)** Ganhos genéticos potenciais: ambientes ideais para a realização de experimentos de melhoramento. **(E)** Zonas de melhoramento: polígonos geográficos que minimizam as interações genótipo  $\times$  ambiente ( $G \times A$ ). **(F)** Produtividade potencial: projeção da produtividade para toda a área ou para outros caracteres fenotípicos desejados. Os resultados se aplicam tanto à média da população quanto aos materiais genéticos selecionados ou recomendados. Adaptado de Resende et al. (2024).

A ambientômica também pode se beneficiar da integração de dados ecofisiológicos, oferecendo soluções para uma agricultura inteligente frente ao clima, práticas de campo mais econômicas e cenários futuros para o melhoramento de plantas (Costa-Neto et al., 2021b). A incorporação de conceitos probabilísticos de modelos Bayesianos aprimora ainda mais os processos de recomendação de cultivares em ensaios

multiambientes, ampliando nossa compreensão das interações  $G \times A$  (Genótipos  $\times$  Ambientes) (Dias et al., 2020). Destaca-se que países como Austrália e outras nações participantes da Rede Internacional de Melhoramento do Trigo (*International Wheat Improvement Network*) obtiveram ganhos genéticos substanciais ao combinar o sensoriamento remoto com a seleção genômica em estratégias de cruzamento parental e seleção de progênes, conforme evidenciado na revisão abrangente de Chen et al. (2022), fornecendo plataformas valiosas para pesquisas em andamento e o aprimoramento das metodologias de melhoramento.

### 2.7.1 AMBIENTÔMICA PARA O MELHORAMENTO PREDITIVO

Por fim, nossa discussão se volta para a etapa 6 do processo de ambientômica, conforme representado na análise de ambientômica (Figura 1), cujos resultados hipotéticos de Resende et al. (2021) são ilustrados na Figura 4A. Um dos modelos utilizados em ambientômica é a regressão aleatória (*random regression*), na qual pulsos horizontais ao longo do gradiente de marcadores ambientômico (eixo x) representam um novo ensaio predito (Figura 4B), ou seja, um experimento 100% virtual. Esse modelo prevê o comportamento de cada material genético e sua ordenação, com extrapolação preditiva em nível específico de localidade, cobrindo todos os pixels da área (Figura 4C). A produtividade potencial média de cada material genético selecionado/recomendado pode ser estimada, conforme mostrado na Figura 4D. Por exemplo, ao integrar dados climáticos e geográficos, foi possível selecionar genótipos ideais de eucalipto para uma ampla região, adequando a escolha de cultivares clonais para maximizar o volume de madeira em diferentes idades de plantio. Essa análise evidenciou o poder da estratificação ambiental inovadora para otimizar a produtividade (Marcatti et al., 2017).

Comparado aos genótipos — nos quais variedades endogâmicas ou híbridas podem, em geral, ser replicadas, ou genótipos de locus único ou haplótipos multilocus podem ser replicados por grupos de indivíduos —, os ambiótipos para determinados fatores ambientais são considerados, em geral, replicáveis para um local específico de ambientipagem, conforme determinado por longitude, latitude e altitude. Fatores ambientais principais (ambiótipos), como a duração sazonal do dia/noite, variações de temperatura e ambientes manejados, são geralmente consistentes e amplamente previsíveis, enquanto os fatores ambientais secundários são em grande parte imprevisíveis. A acurácia preditiva do

desempenho fenotípico é determinada por combinações complexas de genótipos, ambiótipos e suas interações (Araújo et al., 2024). Portanto, apenas os fatores ambientais principais podem ser usados com confiança para classificação e predição. No entanto, de forma semelhante aos índices de seleção utilizados na avaliação de fenótipos quantitativos, índices ambientômicos podem ser construídos para cada localidade específica de ambientipagem, utilizando informações extraídas de todos os fatores ambientais, com base em suas contribuições individuais (infinitesimais) para a variação total ambientômica e em seus relacionamentos ambiótipos (Costa-Neto et al., 2021b; Li et al., 2021; Resende et al., 2021; Piepho, 2022; Xu et al., 2022).

Avançando ainda mais, se uma interação  $G \times A$  estiver significativamente associada a grupos de ambientes dentro da TPE (*Target Population of Environments*), maiores ganhos genéticos podem ser alcançados por meio da reorganização dos experimentos em "mega-ambientes" (Crespo-Herrera et al., 2021; Krause et al., 2022). Isso leva à discussão sobre "zonas de melhoramento" versus mega-ambientes (Gauch & Zobel, 1997), em que zonas de melhoramento referem-se à reagregação de pixels que minimiza a interação  $G \times A$ , e não a agrupamentos de experimentos. Em outras palavras, zonas de melhoramento são polígonos geográficos que minimizam a interação  $G \times A$  (Figura 4E e vide Callister et al., 2024). A partição dos ambientes (regiões geográficas) em subgrupos homogêneos é uma prática de longa data (DeLacy & Cooper, 1990; Ouyang et al., 1995). Contudo, nesta era da ambientômica, temos disponíveis dados ambientais muito mais diversos e completos do que nunca para a formação desses subgrupos, por meio da classificação ou agrupamento dos locais experimentais ambientais com base em todas as informações disponíveis. Com base em níveis de similaridade ambientômica ou índices, zonas de melhoramento, estações experimentais e locais de ensaios MET (*multi-environment trials*) podem ser definidos e otimizados. Quanto mais informação ambientômica for utilizada, melhores serão as estratégias que podem ser desenvolvidas para otimizar os programas e fluxos de melhoramento genético.

É possível alcançar ganhos potenciais de seleção em nível de pixel em uma área. Os ambientes ótimos para implementar experimentos de melhoramento estão representados na Figura 4F, e essa informação pode ser especialmente útil para a alocação de novos experimentos em próximas safras. Afinal, locais com maior potencial de ganho tendem a fornecer seleções genéticas mais precisas. De acordo com Fernandes-Filho et al. (2023), a inclusão de informações ambientais na forma de ambientômica em modelos de predição

genômica para avaliar o desempenho de genótipos em experimentos de melhoramento da alfafa (*Medicago sativa*) em múltiplas colheitas resultou em aumento da variância genética, redução da variância do erro e melhoria da capacidade preditiva, especialmente em relação à adaptabilidade e persistência das famílias genéticas avaliadas.

### **2.7.2 AMBIENTÔMICA EM ESCALA DE CAMPO E OTIMIZAÇÃO DE PROGRAMAS DE MELHORAMENTO**

Atualmente, é difícil prever os efeitos da genética e das práticas de manejo no desempenho das culturas em um ambiente específico, especialmente em escalas regionais a globais. A modelagem de culturas em múltiplas escalas permitirá projetar sistemas do gene à fazenda voltados para uma produção agrícola resiliente e sustentável frente às mudanças climáticas. Esse tipo de modelagem pode ser aprimorado por meio da representação detalhada de características das culturas, da interface entre modelos de crescimento de plantas e modelos em larga escala, do aperfeiçoamento da representação das respostas fisiológicas às mudanças climáticas e práticas de manejo, da redução de lacunas nos dados e do aproveitamento de dados provenientes de múltiplas fontes (Peng et al., 2020). Os modelos de crescimento de culturas oferecem uma abordagem para prever a produtividade agrícola em cenários  $G \times A \times M$  (Genótipos  $\times$  Ambientes  $\times$  Manejos), possibilitando o desenho e a testagem rápida de estratégias inovadoras de melhoramento de plantas com base em uma compreensão integrada dessas interações. Isso cria oportunidades para identificar e implementar caminhos para o aumento da produtividade, por meio da integração de ganhos genéticos provenientes do melhoramento com estratégias de manejo agrícola (Cooper et al., 2020).

Em nível de campo, Guan et al. (2023) propuseram uma estrutura escalável para quantificar os impactos de carbono em áreas agrícolas, mensurando as emissões de gases de efeito estufa, como  $N_2O$  e  $CH_4$ , além de alterações no estoque de carbono do solo. Uma solução do tipo “sistema de sistemas” foi sugerida, baseada na integração de diversas abordagens (por exemplo, observações variadas, dados de sensores/in situ e modelagem). Essa abordagem é composta por cinco componentes: Coleta escalável de dados reais e sensoriamento multiescalar das condições ambientais (A), de manejo (M) e das culturas no nível local; Modelagem avançada para sustentar a quantificação; Integração sistemática entre modelos e dados no nível das propriedades agrícolas; Alta eficiência computacional e

uso de IA, permitindo escalabilidade para milhões de áreas agrícolas individuais a baixo custo; Sistemas de validação robustos e em múltiplos níveis, assegurando a fidelidade da solução e sua escalabilidade real. Essa solução proposta deve ser generalizada e utilizada para quantificar outras medições ambientômicas no contexto do melhoramento vegetal.

Plataformas como *The Climate Corporation* e *Earth Observing System Data Analytics* (EOSDA) *Crop Monitoring* já oferecem suporte significativo ao setor agrícola e ao melhoramento genético, fornecendo vastos conjuntos de dados — incluindo dados meteorológicos de mais de 2,5 milhões de locais e 150 bilhões de observações de solo, o que resulta em 10 trilhões de pontos de simulação climática. Esses dados permitem o monitoramento detalhado da umidade do solo, a criação de zonas para uso eficiente de recursos e a avaliação da saúde da vegetação, impulsionando avanços genéticos e promovendo práticas agrícolas sustentáveis. Em especial, a EOSDA se destaca na oferta de dados críticos de solo e vegetação, por meio de insights gerados por satélites. A plataforma integra imagens de satélite com dados de alta resolução coletados por drones (UAVs) e veículos terrestres, aumentando a precisão dos dados ambientômicos para locais específicos. Essa abordagem - detalhada em <https://eos.com/blog/how-precision-farming-fights-climate-change/> - apoia a seleção de locais ideais de ensaios com base em similaridades ambientais, acelera o melhoramento por meio de ambientes controlados e fortalece as estratégias de seleção. Notavelmente, a adoção de métodos de predição genômica-ambientômica, conforme discutido por Xu et al. (2022), pode oferecer maior acurácia do que as predições genômicas tradicionais, aumentando significativamente a eficiência do melhoramento e os ganhos genéticos.

### **2.7.3 DADOS TEMPORAIS E ADAPTAÇÃO ÀS MUDANÇAS CLIMÁTICAS**

Os dados de satélite oferecem informações temporais, permitindo medições dinâmicas em diversos momentos no tempo, o que amplia a compreensão sobre o período de crescimento das culturas e as similaridades entre diferentes locais. No entanto, eventos atmosféricos, como nuvens, poluição, relâmpagos e radiações solar/cósmica, podem afetar essas medições. Rustowicz (2017) investigou o uso de imagens de satélite em séries temporais e técnicas de aprendizado de máquina para a classificação de culturas. Avançando nessa direção, Pazúr et al. (2021) destacaram o valor dos sensores de satélite com alta resolução temporal para estudos de ecologia da paisagem, demonstrando que a inclusão de

informações temporais melhora a acurácia do mapeamento da paisagem e a identificação de elementos importantes da paisagem.

Explorando o impacto das mudanças climáticas na agricultura, o estudo de Rezaei et al. (2023) investiga os efeitos de temperaturas mais elevadas, níveis aumentados de CO<sub>2</sub> e alterações na disponibilidade hídrica sobre a produtividade das culturas agrícolas. A pesquisa revela respostas variadas entre culturas C3 e C4 diante de estresse hídrico e altos níveis de CO<sub>2</sub>, apontando um aumento potencial na variabilidade dos rendimentos das plantas. Em especial, culturas em latitudes mais baixas podem sofrer reduções severas na produtividade, com perdas estimadas entre 7% e 23% na ausência de medidas adaptativas. O estudo ressalta a importância de uma abordagem multidisciplinar, recomendando a combinação de avaliações biofísicas da produtividade com análises econômicas e ambientais, para melhor compreensão das interações complexas entre fatores como: perda de nitrogênio, mudanças na matéria orgânica do solo, e qualidade nutricional das culturas (Ciscar et al., 2019). Ao empregar ambientipagem e previsões climáticas precisas, o estudo sugere estratégias para enfrentar a diversidade de desafios das mudanças climáticas, defendendo o desenvolvimento de métodos agrícolas inovadores para manter a produtividade frente às mudanças ambientais. A integração de dados de ensaios, melhoramento genético e dados de campo com anotações genotípicas, por meio de estruturas de big data, permite esclarecer as interações  $G \times A \times M$  (Genótipos  $\times$  Ambientes  $\times$  Manejos), ajudando a lidar com os desafios impostos pelas mudanças climáticas. Tal integração, aliada ao uso de materiais genéticos diversos em ambientes variados, representa um caminho promissor para a melhoria do desempenho das cultivares. A caracterização dos ambientes agrícolas é fundamental para o aprimoramento do melhoramento genético e da seleção de germoplasma adaptada ao TPE (*Target Population of Environments*), conforme apontado por Chenu (2015). Saltz et al. (2018) exploraram a variabilidade das interações  $G \times A$ , defendendo estudos sobre suas bases biológicas por meio da análise de características associadas ao desempenho atual das plantas, que podem servir como indicadores de adaptação a condições futuras. Essa abordagem de continuidade temporal é essencial para prever as mudanças necessárias nos genótipos visando a adaptação ao clima em transformação. Assim, desenvolver estratégias de cultivo que incorporem essas interações e considerem os impactos reais das mudanças climáticas é essencial para manter a produtividade agrícola (Cooper et al., 2021). De forma notável, uma estrutura tecnológica G

× M foi proposta para adaptar-se às mudanças climáticas e garantir a segurança alimentar, conforme Messina & Cooper (2022).

## 2.8 CONSIDERAÇÕES FINAIS

A ambientômica emergiu como uma abordagem poderosa para o aprimoramento do melhoramento de plantas, ao possibilitar a integração de informações multidimensionais oriundas de dados de sensoriamento remoto via satélite. A utilização conjunta de estatísticas tradicionais, big data e inteligência artificial (IA) com múltiplos conjuntos de dados ambientais — muitos deles derivados de sensores orbitais — pode fornecer uma visão precisa das interações  $G \times A \times M$  (Genótipos  $\times$  Ambientes  $\times$  Manejos). Essa integração viabiliza o desenvolvimento de esquemas inteligentes e integrados de melhoramento  $G - A - F$  (Genótipos – Ambientes – Fenótipos), permitindo predições fenotípicas mais precisas e maiores ganhos genéticos. Isso se torna possível ao se prever ensaios completamente virtuais, que replicam fielmente a realidade, eliminando a necessidade de experimentos físicos e, assim, reduzindo custos operacionais. A Ambientômica oferece uma dimensão complementar à Genômica e à Fenômica, representando um caminho promissor e inovador para o avanço sustentável da ciência das culturas agrícolas, conectando o conhecimento científico à realidade de campo e contribuindo para uma produção agrícola resiliente diante das mudanças climáticas.

Com a Ambientômica, as possibilidades são vastas. Recomendações genotípicas com extrapolação preditiva para todos os pixels de uma área capacitam melhoristas a tomarem decisões localizadas com base em informações detalhadas sobre o comportamento dos materiais genéticos sob diferentes condições de cultivo. Além disso, o conceito de zonas de melhoramento, ou seja, polígonos geográficos que minimizam as interações  $G \times A$  (e em alguns casos  $G \times E \times M$ ), representa uma abordagem inovadora para otimizar a produtividade em larga escala. A projeção da produtividade potencial de toda uma área, bem como de outros caracteres fenotípicos de interesse, demonstra a utilidade da ambientômica para obter *insights* preditivos da produtividade sob diferentes cenários climáticos. Esses achados podem subsidiar a seleção de variedades mais adaptadas a ambientes locais específicos, contribuindo para reduções de custos em campo tanto no presente quanto no futuro.

Contudo, muitos programas de melhoramento enfrentam uma lacuna significativa por não conseguirem traduzir dados derivados de satélites em conhecimento prático para orientar decisões. Para enfrentar esse desafio, é necessário um esforço direcionado para capacitar profissionais do melhoramento de plantas com conhecimentos sobre processamento de dados, métodos analíticos e o uso de tecnologias avançadas, como IA e modelagem estatística. Ao fortalecer essas competências fundamentais, os programas de melhoramento poderão liberar todo o potencial da Ambientômica, utilizando os dados de satélite como ferramenta-chave no avanço da agricultura sustentável.

Apesar dos desafios persistentes — como a necessidade de dados satelitais gratuitos e de alta resolução, e da integração eficiente com dados genômicos e fenômicos, é evidente que a ambientômica oferece uma nova oportunidade para aumentar a produtividade e a sustentabilidade agrícola. Existe também a demanda por integrar dados satelitais de baixa resolução com imagens de alta resolução de UAVs, por meio de imputação ou IA, reduzindo os custos associados à obtenção de imagens de alta qualidade em larga escala. Além disso, modelos avançados de IA podem melhorar várias etapas da ambientômica, desde aumento e fusão de dados até previsão com modelos complexos. A colaboração entre agrônomos, fisiologistas, geneticistas, especialistas em geoinformação e programadores, em diferentes instituições de pesquisa, é essencial para avançar nesse campo e aproveitar plenamente o potencial dos dados de envirotipagem para transformar a agricultura do século XXI.

## 2.9 REFERÊNCIAS

AL-TURJMAN, F. The road towards plant phenotyping via WSNs: An overview. **Computers and Electronics in Agriculture**, v. 161, p. 4-13, 2019. <https://doi.org/10.1016/J.COMPAG.2018.09.018>

ALVITES, C.; MARCHETTI, M.; LASSERRE, B.; SANTOPUOLI, G. LiDAR as a tool for assessing timber assortments: A systematic literature review. **Remote Sensing**, v. 14, n. 18, p. 4466, 2022.

ANTHONY, J. C.; EATON, W. W.; HENDERSON, A. S. Looking to the future in psychiatric epidemiology. **Epidemiologic reviews**, v. 17, n. 1, p. 240-242, 1995.

APPEL, M.; PEBESMA, E. On-demand processing of data cubes from satellite image collections with the gdalcubes library. **Data**, v. 4, n. 3, p. 92, 2019.

ARAÚJO, M. S.; CHAVES, S. F. S.; DIAS, L. A. S.; FERREIRA, F. M.; PEREIRA, G. R.; BEZERRA, A. R. G.; ALVES, R. S.; HEINEMANN, A. B.; BRESEGHELLO, F.;

CARNEIRO, P. C. S.; KRAUSE, M. D.; COSTA-NETO, G. M. F.; DIAS, K. O. G. GIS-FA: an approach to integrating thematic maps, factor-analytic, and envirotyping for cultivar targeting. **Theoretical and Applied Genetics**, v. 137, n. 4, p. 80, 2024.

AYDINOGLU, A. C. Modelling, encoding and transforming of open geographic data to examine interoperability between GIS applications. **Geocarto international**, v. 31, n. 4, p. 446-461, 2016.

SAN BAUTISTA, A.; FITA, D.; FRANCH, B.; CASTIÑEIRA-IBÁÑEZ, S.; ARIZO, P.; SÁNCHEZ-TORRES, M. J.; BECKER-RESHEF, I.; URIS, A.; RUBIO, C. Crop monitoring strategy based on remote sensing data (Sentinel-2 and planet), study case in a rice field after applying glycinebetaine. **Agronomy**, v. 12, n. 3, p. 708, 2022.

BECKERS, J.; WURST, W.; DE ANGELIS, M. H. Towards better mouse models: enhanced genotypes, systemic phenotyping and envirotype modelling. **Nature Reviews Genetics**, v. 10, n. 6, p. 371-380, 2009.

BERRIOS, D. C.; GALAZKA, J.; GRIGOREV, K.; GEBRE, S.; COSTES, S. V. NASA GeneLab: interfaces for the exploration of space omics data. **Nucleic acids research**, v. 49, n. D1, p. D1515-D1522, 2021.

BRADY, R.; SPRING, A. Climpred: Verification of weather and climate forecasts. **The Journal of Open Source Software**, v. 6, n. 59, 2021.

BREIMAN, L. Statistical modeling: The two cultures (with comments and a rejoinder by the author). **Statistical science**, v. 16, n. 3, p. 199-231, 2001.

BRÖNNIMANN, S.; ALLAN, R.; ATKINSON, C.; BUIZZA, R.; BULYGINA, O.; DAHLGREN, P.; DEE, D.; DUNN, R.; GOMES, P.; JOHN, V.O.; JOURDAIN, SYLVIE; HAIMBERGER, LEOPOLD; HERSBACH, HANS; KENNEDY, JOHN; POLI, P.; PULLIAINEN, J.; RAYNER, N.; SAUNDERS, R.; SCHULZ, J.; STERIN, A.; STICKLER, A.; TITCHNER, H.; VALENTE, M. A.; VENTURA, C.; WILKINSON, C. Observations for reanalyses. **Bulletin of the American Meteorological Society**, v. 99, n. 9, p. 1851-1866, 2018.

BUDHLAKOTI, N.; KUSHWAHA, A. K.; RAI, A.; CHATURVEDI, K.K.; KUMAR, A.; PRADHAN, A. K.; KUMAR, U.; KUMAR, R. R.; JULIANA, P.; MISHRA, D. C.; KUMAR, S. Genomic selection: A tool for accelerating the efficiency of molecular breeding for development of climate-resilient crops. **Frontiers in Genetics**, v. 13, p. 832153, 2022.

BUSTOS-KORTS, D.; MALOSETTI, M.; CHENU, K.; CHAPMAN, S.; BOER, M. P.; ZHENG, B.; VAN EEUWIJK, F.A. From QTLs to adaptation landscapes: using genotype-to-phenotype models to characterize G × E over time. **Frontiers in Plant Science**, v. 10, p. 1-23, 2019.

CALLISTER, A. N.; COSTA-NETO, G.; BRADSHAW, B. P.; ELMS, S.; CROSSA, J.; BRAWNER, J. T. Enviromic prediction enables the characterization and mapping of

Eucalyptus globulus Labill breeding zones. **Tree Genetics & Genomes**, v. 20, n. 1, p. 3, 2024.

CHAVES, L. J.; MIRANDA FILHO, J. B. Plot size for progeny selection in maize (*Zea mays* L.). **Theoretical and Applied Genetics**, v. 84, n. 7, p. 963-970, 1992.

CHEN, C. J.; RUTKOSKI, J.; SCHNABLE, J. C.; MURRAY, S. C.; WANG, L.; JIN, L. W. X.; STICH, B.; CROSSA, J.; HAYES, B. J.; ZHANG, Z. Role of the genomics-phenomics-agronomy paradigm in plant breeding. **Plant breeding reviews**, v. 46, p. 627-673, 2022.

CHENU, K. Characterizing the crop environment-nature, significance and applications. **Crop physiology**, p. 321-348, 2015.

COOPER, M.; MESSINA, C. D.; PODLICH, D.; TOTIR, L. R.; BAUMGARTEN, A.; HAUSMANN, N. J.; WRIGHT, D.; GRAHAM, G. Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. **Crop and Pasture Science**, v. 65, n. 4, p. 311-336, 2014.

CISCAR, J. C.; RISING, J.; KOPP, R. E.; FEYEN, L. Assessing future climate change impacts in the EU and the USA: insights and lessons from two continental-scale projects. **Environmental Research Letters**, v. 14, n. 8, p. 084010, 2019.

COOPER, M.; MESSINA, C. D. Can we harness “enviromics” to accelerate crop improvement by integrating breeding and agronomy? **Frontiers in Plant Science**, v. 12, p. 735143, 2021.

COOPER, M.; TANG, T.; GHO, C.; HART, T.; HAMMER, G.; MESSINA, C. Integrating genetic gain and gap analysis to predict improvements in crop productivity. **Crop Science**, v. 60, n. 2, p. 582-604, 2020.

COOPER, M.; VOSS-FELS, K. P.; MESSINA, C. D.; TANG, T.; HAMMER, G. L. Tackling  $G \times E \times M$  interactions to close on-farm yield-gaps: creating novel pathways for crop improvement by predicting contributions of genetics and management to crop productivity. **Theoretical and Applied Genetics**, v. 134, n. 6, p. 1625-1644, 2021.

COSTA-NETO, G.; CRESPO-HERRERA, L.; FRADGLEY, N.; GARDNER, K.; BENTLEY, A. R.; DREISIGACKER, S.; FRITSCHÉ-NETO, R.; MONTESINOS-LÓPEZ, O. A.; CROSSA, J. Envirome-wide associations enhance multi-year genome-based prediction of historical wheat breeding data. **G3**, v. 13, n. 2, p. jkac313, 2023.

COSTA-NETO, G.; CROSSA, J.; FRITSCHÉ-NETO, R. Enviromic assembly increases accuracy and reduces costs of the genomic prediction for yield plasticity in maize. **Frontiers in Plant Science**, v. 12, p. 717552, 2021.

COSTA-NETO, G.; FRITSCHÉ-NETO, R. Enviromics: bridging different sources of data, building one framework. **Crop breeding and applied biotechnology**, v. 21, e393521S12, 2021.

COSTA-NETO, G.; GALLI, G.; CARVALHO, H.F.; CROSSA, J.; FRITSCHÉ-NETO, R. EnvRtype: a software to interplay enviromics and quantitative genomics in agriculture. **G3**, v. 11, n. 4, p. jkab040, 2021.

CRESPO-HERRERA, L. A.; CROSSA, J.; HUERTA-ESPINO, J.; MONDAL, S.; VELU, G.; JULIANA, P.; VARGAS, M.; PÉREZ-RODRÍGUEZ, P.; JOSHI, A. K.; BRAUN, H. J.; SINGH, R. P. Target population of environments for wheat breeding in India: definition, prediction and genetic gains. **Frontiers in Plant Science**, v. 12, p. 638520, 2021.

CROSSA, J.; FRITSCHÉ-NETO, R.; MONTESINOS-LOPEZ, O. A.; COSTA-NETO, G.; DREISIGACKER, S.; MONTESINOS-LOPEZ, A.; BENTLEY, A. R. The modern plant breeding triangle: optimizing the use of genomics, phenomics, and enviromics data. **Frontiers in plant science**, v. 12, p. 651480, 2021.

CROSSA, J.; MONTESINOS-LÓPEZ, O. A.; PÉREZ-RODRÍGUEZ, P.; COSTA-NETO, G.; FRITSCHÉ-NETO, R.; ORTIZ, R.; MARTINI, J. W. R.; LILLEMOM, M.; MONTESINOS-LÓPEZ, A.; JARQUÍN, D.; BRESEGHELLO, F.; CUEVAS, J.; RINCENT, R. Genome and environment based prediction models and methods of complex traits incorporating genotype  $\times$  environment interaction. **Genomic prediction of complex traits: Methods and protocols**, p. 245-283, 2022.

CUI, D.; LIANG, S.; WANG, D.; LIU, Z. A 1 km global dataset of historical (1979–2013) and future (2020–2100) Köppen–Geiger climate classification and bioclimatic variables. **Earth Syst. Sci. Data**, 13, 5087-5114. 2021.

DEBNATH, S.; PAUL, M.; DEBNATH, T. Applications of LiDAR in agriculture and future research directions. **Journal of Imaging**, v. 9, n. 3, p. 57, 2023.

DELACY, I. H.; COOPER, M. Pattern analysis for the analysis of regional variety trials. *In: Genotype-by-environment Interaction and Plant Breeding*, M.S. Kang, (ed). Louisiana State University. pp. 301-334, 1990.

DELUCCHI, L.; NETELER, M. G. **pyModis**: the Python library for MODIS data. *In: FOSS4G 2013*, Nottingham. 2013.

DIAS, K. O. G.; PIEPHO, H. P.; GUIMARÃES, L. J. M.; GUIMARÃES, P. E. O.; PARENTONI, S. N.; PINTO, M. O.; NODA, R. W.; MAGALHÃES, J. V.; GUIMARÃES, C. T.; GARCIA, A. A. F.; PASTINA, M. M. Novel strategies for genomic prediction of untested single-cross maize hybrids using unbalanced historical data. **Theoretical and Applied Genetics**, v. 133, n. 2, p. 443-455, 2020.

DINERSTEIN, E.; OLSON, D.; JOSHI, A.; VYNNE, C.; BURGESS, N. D.; WIKRAMANAYAKE, E.; HAHN, N.; PALMINTERI, S.; HEDAO, P.; NOSS, R., ET AL. An ecoregion-based approach to protecting half the terrestrial realm. **BioScience**, v. 67, n. 6, p. 534-545, 2017.

DOMINGUES, G. F.; SOUZA, G. S. A.; BARBOSA, R. A.; MARCATTI, G. E.; LORENZON, A. S.; LEITE, H.G. Estimated volume of eucalyptus plantations through ALOS satellite images. **TreeDimensional**, v. 11, p. e2023025, 2023.

VAN EMMERIK, T.; STEELE-DUNNE, S.; PAGET, A.; OLIVEIRA, R. S.; BITTENCOURT, P. R. L.; BARROS, F. D. V.; VAN DE GIESEN, N. Water stress detection in the Amazon using radar. **Geophysical Research Letters**, v. 44, n. 13, p. 6841-6849, 2017.

FERNADES FILHO, C. C.; ANDRADE, M. H. M. L.; NUNES, J. A. R.; JARQUIN, D. H.; RIOS, E. F. Genomic prediction for complex traits across multiples harvests in alfalfa (*Medicago sativa* L.) is enhanced by enviromics. **The Plant Genome**, v. 16, n. 2, e20306, 2023.

FICK, S. E.; HIJMANS, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. **International journal of climatology**, v. 37, n. 12, p. 4302-4315, 2017.

FISHER, J. B.; LEE, B.; PURDY, A. J.; HALVERSON, G. H.; DOHLEN, M. B.; CAWSE-NICHOLSON, K.; WANG, A.; ANDERSON, R. G.; ARAGON, B.; ARAIN, M. A. et al. ECOSTRESS: NASA's next generation mission to measure evapotranspiration from the international space station. **Water Resources Research**, v. 56, n. 4, e2019WR026058, 2020.

FRITSCHÉ-NETO, R. SoilType: An R package to interplay soil characterization in plant science. **Agronomy Journal**, v. 116, n. 3, p. 848-854, 2024.

GAUCH, H.G.; ZOBEL, R.W. Identifying mega-environments and targeting genotypes. **Crop science**, v. 37, n. 2, p. 311-326, 1997.

GELARO, R.; MCCARTY, W.; SUÁREZ, M. J.; TODLING, R.; MOLOD, A.; TAKACS, L.; RANGLES, C.; DARMENOV, A.; BOSILOVICH, M. G.; REICHLE, R., et al. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). **Journal of climate**, v. 30, n. 14, p. 5419-5454, 2017.

GRIES, C.; HANSON, P. C.; O'BRIEN, M.; SERVILLA, M.; VANDERBILT, K.; WAIDE, R. The environmental data Initiative: Connecting the past to the future through data reuse. **Ecology and Evolution**, v. 13, n. 1, e9592, 2023.

GUAN, K.; JIN, Z.; PENG, B.; TANG, J.; DELUCIA, E. H.; WEST, P. C.; JIANG, C.; WANG, S.; KIM, T.; ZHOU, W., et al. A scalable framework for quantifying field-level agricultural carbon outcomes. **Earth-Science Reviews**, v. 243, p. 104462, 2023.

GUO, T.; LI, X. Machine learning for predicting phenotype from genotype and environment. **Current Opinion in Biotechnology**, v. 79, p. 102853, 2023.

HAYES, B. J.; CHEN, C.; POWELL, O.; DINGLASAN, E.; VILLIERS, K.; KEMPER, K. E.; HICKEY, L. T. Advancing artificial intelligence to help feed the world. **Nature biotechnology**, v. 41, n. 9, p. 1188-1189, 2023.

HEINEMANN, A. B.; COSTA-NETO, G.; FRITSCHÉ-NETO, R.; MATTA, D. H.; FERNANDES, I. K. Enviromic prediction is useful to define the limits of climate adaptation: a case study of common bean in Brazil. **Field Crops Research**, v. 286, p. 108628, 2022.

HERSBACH, H.; BELL, B.; BERRISFORD, P.; HIRAHARA, S.; HORÁNYI, A.; MUÑOZ-SABATER, J.; NICOLAS, J.; PEUBEY, C.; RADU, R.; SCHEPERS, D.; ET AL. The ERA5 global reanalysis. **Quarterly journal of the royal meteorological society**, v. 146, n. 730, p. 1999-2049, 2020.

HIJMANS, R. J.; BIVAND, R.; FORNER, K.; OOMS, J.; PEBESMA, E.; SUMNER, M. D. Package 'terra'. **Maintainer: Vienna, Austria**, v. 384, 2022.

HIJMANS, R. J.; CAMERON, S. E.; PARRA, J. L.; JONES, P. G.; JARVIS, A. Very high resolution interpolated climate surfaces for global land areas. **International Journal of Climatology: A Journal of the Royal Meteorological Society**, v. 25, n. 15, p. 1965-1978, 2005.

HOLZWORTH, D.; HUTH, N. I.; FAINGES, J.; BROWN, H.; ZURCHER, E.; CICHOTA, R.; VERRALL, S.; HERRMANN, N. I.; ZHENG, B.; SNOW, V. APSIM Next Generation: Overcoming challenges in modernising a farming systems model. **Environmental Modelling & Software**, v. 103, p. 43-51, 2018.

INDIANAMAP. **IndianaMap Framework Lidar**. Indiana Statewide Imagery and LiDAR Program. Distributed by OpenTopography, 2011.

JAKOBSEN, P.; FERRUIT, P.; OLIVEIRA, C. A.; ARRIBAS, S.; BAGNASCO, G.; BARHO, R.; BECK, T. L.; BIRKMANN, S.; BOKER, T.; BUNKER, A. J.; et al. The near-infrared spectrograph (nirspec) on the james webb space telescope-i. overview of the instrument and its capabilities. **Astronomy & Astrophysics**, v. 661, p. A80, 2022.

JARQUÍN, D.; CROSSA, J.; LACAZE, X.; DU CHEYRON, P.; DAUCOURT, J.; LORGEUX, J.; PIRAUX, F.; GUERREIRO, L.; PÉREZ, P.; CALUS, M.; et al. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. **Theoretical and applied genetics**, v. 127, n. 3, p. 595-607, 2014.

JENSEN, J. R. **Remote sensing of the environment: An earth resource perspective 2/e**. Pearson Education India, 2009.

KARWOWSKA, K.; WIERZBICKI, D. Using super-resolution algorithms for small satellite imagery: A systematic review. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 15, p. 3292-3312, 2022.

KHAN, M.H.U.; WANG, S.; WANG, J.; AHMAR, S.; SAEED, S.; KHAN, S. U.; XU, X.; CHEN, H.; BHAT, J. A.; FENG, X. Applications of artificial intelligence in climate-resilient smart-crop breeding. **International Journal of Molecular Sciences**, v. 23, n. 19, p. 11156, 2022.

KHANAL, S.; KC, K.; FULTON, J. P.; SHEARER, S.; OZKAN, E. Remote sensing in agriculture—accomplishments, limitations, and opportunities. **Remote sensing**, v. 12, n. 22, p. 3783, 2020.

KHOSLA, R. **Big data in agriculture**: agronomic considerations. Soil and Crop Sciences Section Seminar Series (Cornell University), 2023.

KIM, H.J.; DINOY, P.R.T.; CHOI, H.S.; LEE, K.B.; MISSION, J.L.C. Spatial interpolation of SPT data and prediction of consolidation of clay by ANN method. **Coupled Syst. Mech.** v. 8, p. 523–535, 2019.

KOKALY, R.F.; DESPAIN, D.G.; CLARK, R.N.; LIVO, K. Mapping vegetation in Yellowstone National Park using spectral feature analysis of AVIRIS data. **Remote sensing of environment**, v. 84, n. 3, p. 437-456, 2003.

KOUZIS-LOUKAS, D. **Learning Scrapy**. Livery Place (Packt Publishing), 2016.

KRAUSE, M. D.; DIAS, K. O. G.; SINGH, A. K.; BEAVIS, W. D. Using soybean historical field trial data to study genotype by environment variation and identify mega-environments with the integration of genetic and non-genetic factors. **Agronomy Journal**, v. 117, n. 1, e70023, 2025.

KRAUSE, M. D.; DIAS, K. O. D. G.; SANTOS, J. P. R.; OLIVEIRA, A. A., GUIMARÃES, L. J. M., PASTINA, M. M., MARGARIDO, G. R. A.; GARCIA, A. A. F. Boosting predictive ability of tropical maize hybrids via genotype-by-environment interaction under multivariate GBLUP models. **Crop Science**, v. 60, n. 6, p. 3049-3065, 2020.

KRINITSKIY, M.; SPRYGIN, A.; ELIZAROV, S.; NARIZHNAYA, A.; SHIKHOV, A.; CHERNOKULSKY, A. Towards the accurate automatic detection of mesoscale convective systems in remote sensing data: from data mining to deep learning models and their applications. **Remote Sensing**, v. 15, n. 14, p. 3493, 2023.

LECHNER, A. M.; FOODY, G. M.; BOYD, D. S. Applications in remote sensing to forest ecology and management. **One Earth**, v. 2, n. 5, p. 405-412, 2020.

LEVCHENKO, I.; BAZAKA, K.; DING, Y.; RAITSES, Y.; MAZOUFFRE, S.; HENNING, T.; KLAR, P.J.; SHINOHARA, S.; SCHEIN, J.; GARRIGUES, L.; et al. Space micropropulsion systems for Cubesats and small satellites: From proximate targets to furthestmost frontiers. **Applied Physics Reviews**, v. 5, n. 1, 2018.

LI, X.; GUO, T.; BAI, G.; ZHANG, Z.; SEE, D.; MARSHALL, J.; GARLAND-CAMPBELL, K.A.; YU, J. Genetics-inspired data-driven approaches explain and predict crop performance fluctuations attributed to changing climatic conditions. **Molecular Plant**, v. 15, n. 2, p. 203-206, 2022.

Li, X.; Guo, T.; Wang, J.; Bekele, W. A.; Sukumaran, S.; Vanous, A. E.; McNellie, J. P.; Tibbs-Cortes, L. E.; Lopes, M. S.; Lamkey, K. R.; et al. An integrated framework

reinstating the environmental dimension for GWAS and genomic selection in crops. **Molecular Plant**, v. 14, n. 6, p. 874-887, 2021.

LUND, H. Most Common Types of Data Integration Methods, 2020.  
<https://www.rapidonline.com/blog/most-common-types-of-data-integration-methods>.

MALOSETTI, M.; RIBAUT, J. M.; VAN EEUWIJK, F. A. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. **Frontiers in physiology**, v. 4, p. 44, 2013.

MARCATTI, G. E.; RESENDE, R. T.; RESENDE, M. D. V.; RIBEIRO, C. A. A.; SANTOS, A. R.; CRUZ, J. P.; LEITE, H. G. GIS-based approach applied to optimizing recommendations of Eucalyptus genotypes. **Forest ecology and management**, v. 392, p. 144-153, 2017.

MARSH, J. I.; HU, H.; GILL, M.; BATLEY, J.; EDWARDS, D. Crop breeding for a changing climate: Integrating phenomics and genomics with bioinformatics. **Theoretical and Applied Genetics**, v. 134, n. 6, p. 1677-1690, 2021.

MESSINA, C.; COOPER, M. Plant Breeding in the face of climate change. Preprint at bioRxiv, 2022. <https://doi.org/10.1101/2022.10.07.511293>.

MILLET, E. J.; KRUIJER, W.; COUPEL-LEDRU, A.; ALVAREZ PRADO, S.; CABRERA-BOSQUET, L.; LACUBE, S.; CHARCOSSET, A.; WELCKER, C.; VAN EEUWIJK, F.; TARDIEU, F. Genomic prediction of maize yield across European environmental conditions. **Nature genetics**, v. 51, n. 6, p. 952-956, 2019.

MISHRA, V. N.; PRASAD, R.; KUMAR, P.; SRIVASTAVA, P. K.; RAI, P. K. Knowledge-based decision tree approach for mapping spatial distribution of rice crop using C-band synthetic aperture radar-derived information. **Journal of Applied Remote Sensing**, v. 11, n. 4, p. 046003-046003, 2017.

MONTESINOS-LÓPEZ, O. A.; MONTESINOS-LÓPEZ, A.; CROSSA, J. **Multivariate statistical machine learning methods for genomic prediction**. Springer Nature: Cham, 2022.

NAGHIBI, S. A.; HASHEMI, H.; BERNDTSSON, R.; LEE, S. Application of extreme gradient boosting and parallel random forest algorithms for assessing groundwater spring potential using DEM-derived factors. **Journal of Hydrology**, v. 589, p. 125197, 2020.

NATTINO, F.; KU, O.; GROOTES, M. W.; IZQUIERDO-VERDUGUIER, E.; GIRGIN, S.; ZURITA-MILLA, R. CGC: a scalable Python package for co-and tri-clustering of geodata cubes. **Journal of open source software**, v. 7, n. 72, p. 4032, 2022.

NEGUS, K. L.; LI, X.; WELCH, S. M.; YU, J. The role of artificial intelligence in crop improvement. **Advances in agronomy**, v. 184, p. 1-66, 2024.

- NEISSE, A. C.; KIRCH, J. L.; HONGYU, K. AMMI and GGE Biplot for genotype× environment interaction: a medoid–based hierarchical cluster analysis approach for high–dimensional data. **Biometrical Letters**, v. 55, n. 2, p. 97-121, 2018.
- NOACK, S.; KNOBLOCH, A.; ETZOLD, S. H.; BARTH, A.; KALLMEIER, E. Spatial predictive mapping using artificial neural networks. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 40, p. 79-86, 2014.
- OLIVOTO, T.; LÚCIO, A. D.; SILVA, J. A.; MARCHIORO, V. S.; SOUZA, V. Q.; JOST, E. Mean performance and stability in multi-environment trials I: combining features of AMMI and BLUP techniques. **Agronomy journal**, v. 111, n. 6, p. 2949-2960, 2019.
- OUYANG, Z.; MOWERS, R. P.; JENSEN, A.; WANG, S.; ZENG, S. Custer analysis for genotype × environment interaction with unbalanced data. **Crop science**, v. 35, n. 5, p. 1300-1305, 1995.
- PATTEN, B. C. Holoecology: the genotype–phenotype–envirotpe complex in genetic and extragenetic inheritance. **Holoecology: The Unification of Nature by Network Indirect Effects**, Unpublished manuscript, Complexity in Ecological Science Series, Columbia University Press: New York, 1991.
- PAZÚR, R.; PRICE, B.; ATKINSON, P. M. Fine temporal resolution satellite sensors with global coverage: an opportunity for landscape ecologists. **Landscape Ecology**, v. 36, n. 8, p. 2199-2213, 2021.
- PEBESMA, E; BIVAND, R. **Spatial data science: With applications in R**. Chapman and Hall/CRC, 2023.
- PENG, B.; GUAN, K.; TANG, J.; AINSWORTH, E. A.; ASSENG, S.; BERNACCHI, C. J.; COOPER, M.; DELUCIA, E. H.; ELLIOTT, J. W.; EWERT, F.; et al. Towards a multiscale crop modelling framework for climate change adaptation assessment. **Nature Plants**, v. 6, n. 4, p. 338-348. 2020.
- PICARD, M.; SCOTT-BOYER, M. P.; BODEIN, A.; PÉRIN, O.; DROIT, A. Integration strategies of multi-omics data for machine learning analysis. **Computational and Structural Biotechnology Journal**, v. 19, p. 3735-3746, 2021.
- PIEPHO, H. P. Prediction of and for new environments: What's your model? **Molecular Plant**, v. 15, n. 4, p. 581-582, 2022.
- PIEPHO, H.P.; BLANCON, J. Extending Finlay–Wilkinson regression with environmental covariates. **Plant Breeding**, v. 142, n. 5, p. 621-631, 2023.
- PINTO, F.; ZAMAN-ALLAH, M.; REYNOLDS, M.; SCHULTHESS, U. Satellite imagery for high-throughput phenotyping in breeding plots. **Frontiers in Plant Science**, v. 14, p. 1114670, 2023.

POGGIO, L.; SOUSA, L.M.; BATJES, N.H.; HEUVELINK, G.B.M.; KEMPEN, B.; RIBEIRO, E.; ROSSITER, D. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. **Soil**, v. 7, n. 1, p. 217-240, 2021.

RANGHETTI, L.; BOSCHETTI, M.; NUTINI, F.; Busetto, L. “sen2r”: An R toolbox for automatically downloading and preprocessing Sentinel-2 satellite data. **Computers & Geosciences**, v. 139, p. 104473, 2020.

REICHLER, R.H.; DRAPER, C. S.; LIU, Q.; GIROTTO, M.; MAHANAMA, S. P. P.; KOSTER, R.D.; DE LANNOY, G. J. M. Assessment of MERRA-2 land surface hydrology estimates. **Journal of Climate**, v. 30, n. 8, p. 2937-2960, 2017.

REITZ, K. Requests Documentation Release 2.31.0. 2024.

<https://requests.readthedocs.io/en/latest/>.

RESENDE, R. T.; CHENU, K.; RASMUSSEN, S. K.; HEINEMANN, A. B.; FRITSCHENETO, R. Editorial: Enviromics in Plant Breeding. **Frontiers in Plant Science**, v. 13, n. 935380, p. 1-3, 2022.

RESENDE, R.T.; PIEPHO, H. P.; ROSA, G. J. M.; SILVA-JUNIOR, O. B.; SILVA, F. F.; RESENDE, M. D. V.; GRATTAPAGLIA, D. Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. **Theoretical and Applied Genetics**, v. 134, n. 1, p. 95-112, 2021.

RESENDE, R. T. Balancing genomic selection efforts for allogamous plant breeding programs. **Journal of Crop Science and Biotechnology**, v. 27, n. 3, p. 349-358, 2024.

REZAEI, E. E.; WEBBER, H.; ASSENG, S.; BOOTE, K.; DURAND, J. L.; EWERT, F.; MARTRE, P.; MACCARTHY, D. S. Climate change impacts on crop yields. **nature reviews earth & environment**, v. 4, n. 12, p. 831-846, 2023.

RINCENT, R.; MALOSETTI, M.; ABABAEI, B.; TOUZY, G.; MINI, A.; BOGARD, M.; MARTRE, P.; LE GOUIS, J.; VAN EEUWIJK, F. Using crop growth model stress covariates and AMMI decomposition to better predict genotype-by-environment interactions. **Theoretical and Applied Genetics**, v. 132, n. 12, p. 3399-3411, 2019.

RIZZO, R.; WADOUX, A. M. C.; DEMATTÊ, J. A.; MINASNY, B.; BARRÓN, V.; BEN-DOR, E.; FRANCO, N.; SAVIN, I.; POPPIEL, R.; SILVERO, N. E.; et al. Remote sensing of the Earth's soil color in space and time. **Remote Sensing of Environment**, v. 299, p. 113845, 2023.

RUSTOWICZ, R. M. **Crop classification with multi-temporal satellite imagery**. Final Reports. Univ. Stanford, 2017.

SALTZ, J. B.; BELL, A. M.; FLINT, J.; GOMULKIEWICZ, R.; HUGHES, K. A.; KEAGY, J. Why does the magnitude of genotype-by-environment interaction vary? **Ecology and Evolution**, v. 8, n. 12, p. 6342-6353, 2018.

SAMUI, P. Application of artificial intelligence in geo-engineering. *In: Information Technology in Geo-Engineering: Proceedings of the 3rd International Conference (ICITG)*, Guimarães: Springer International Publishing, 2020.

SCHMIDT, P.; HARTUNG, J.; RATH, J.; PIEPHO, H.P. Estimating broad-sense heritability with unbalanced data from agricultural cultivar trials. **Crop Science**, v. 59, n. 2, p. 525-536, 2019.

SELBY, P.; ABBELOOS, R.; BACKLUND, J. E.; BASTERRECHEA SALIDO, M.; BAUCHET, G.; BENITES-ALFARO, O. E.; BIRKETT, C.; CALAMINOS, V. C.; CARCELLER, P.; CORNUT, G.; et al. BrAPI — an application programming interface for plant breeding applications. **Bioinformatics**, v. 35, n. 20, p. 4147-4155, 2019.

SHAIKH, T. A.; RASOOL, T.; LONE, F. R. Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. **Computers and Electronics in Agriculture**, v. 198, p. 107119, 2022.

SHMUELI, G. To explain or to predict? **Statistical science**, p. 289-310, 2010.

SILVA, V. S.; SALAMI, G.; SILVA, M. I. O.; SILVA, E. A.; MONTEIRO JUNIOR, J. J.; ALBA, E. Methodological evaluation of vegetation indexes in land use and land cover (LULC) classification. **Geology, Ecology, and Landscapes**, v. 4, n. 2, p. 159-169, 2020.

SKOVSGAARD, J. P.; VANCLAY, J. K. Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands. **Forestry**, v. 81, n. 1, p. 13-31, 2008.

SOM-ARD, J.; IMMITZER, M.; VUOLO, F.; ATZBERGER, C. Sugarcane yield estimation in Thailand at multiple scales using the integration of UAV and Sentinel-2 imagery. **Precision Agriculture**, v. 25, n. 3, p. 1581-1608, 2024.

SMITH, A.; NORMAN, A.; KUCHEL, H.; CULLIS, B. Plant variety selection using interaction classes derived from factor analytic linear mixed models: models with independent variety effects. **Frontiers in Plant Science**, v. 12, p. 737462, 2021.

SONG, Y.; KALACSKA, M.; GAŠPAROVIĆ, M.; YAO, J.; NAJIBI, N. Advances in geocomputation and geospatial artificial intelligence (GeoAI) for mapping. **International Journal of Applied Earth Observation and Geoinformation**, v. 120, p. 103300, 2023.

SOUSA, M.; GALLI, G.; LYRA, D.; GRANATO, Í.; MATIAS, F.; ALVES, F.; FRITSCHÉ-NETO, R. Increasing accuracy and reducing costs of genomic prediction by marker selection. **Euphytica**, v. 215, n. 2, p. 18, 2019.

SPARKS, A. H. nasapower: a NASA POWER global meteorology, surface solar energy and climatology data client for R. **Journal of Open Source Software**, v. 3, n. 30, p. 1035, 2018.

TERENTEV, A.; DOLZHENKO, V.; FEDOTOV, A.; EREMENKO, D. Current state of hyperspectral remote sensing for early plant disease detection: A review. **Sensors**, v. 22, n. 3, p. 757, 2022.

TOMAR, V.; MANDAL, V. P.; SRIVASTAVA, P.; PATAIRIYA, S.; SINGH, K.; RAVISANKAR, N.; SUBASH, N.; KUMAR, P. Rice equivalent crop yield assessment using MODIS sensors' based MOD13A1-NDVI data. **IEEE sensors journal**, v. 14, n. 10, p. 3599-3605, 2014.

UTHES, S.; KELLY, E.; KÖNIG, H. J. Farm-level indicators for crop and landscape diversity derived from agricultural beneficiaries data. **Ecological Indicators**, v. 108, p. 105725, 2020.

VAN EEUWIJK, F. A.; BUSTOS-KORTS, D. V.; MALOSETTI, M. What should students in plant breeding know about the statistical aspects of genotype  $\times$  environment interactions? **Crop Science**, v. 56, n. 5, p. 2119-2140, 2016.

VELASTEGUI-MONTOYA, A.; MONTALVÁN-BURBANO, N.; CARRIÓN-MERO, P.; RIVERA-TORRES, H.; SADECK, L.; ADAMI, M. Google Earth Engine: a global analysis and future trends. **Remote Sensing**, v. 15, n. 14, p. 3675, 2023.

VENANCIO, L. P.; MANTOVANI, E. C.; AMARAL, C. H.; USHER NEALE, C. M.; GONÇALVES, I. Z.; FILGUEIRAS, R.; CAMPOS, I. Forecasting corn yield at the farm level in Brazil based on the FAO-66 approach and soil-adjusted vegetation index (SAVI). **Agricultural Water Management**, v. 225, p. 105779, 2019.

VANCE, T. C.; HUANG, T.; BUTLER, K. A. Big data in Earth science: Emerging practice and promise. **Science**, v. 383, n. 6688, p. eadh9607, 2024.

VOITIK, A.; KRAVCHENKO, V.; PUSHKA, O.; KUTKOVETSKA, T.; SHCHUR, T.; KOCIRA, S. Comparison of NDVI, NDRE, MSAVI and NDSI indices for early diagnosis of crop problems. **Agricultural Engineering**, v. 27, 2023.

WANG, A.; WO, J.; LUO, X.; WANG, Y.; CONG, W.; DU, P.; ZHANG, J.; ZHAO, B.; ZHANG, J.; ZHU, Y.; ET AL. Ka-band microwave photonic ultra-wideband imaging radar for capturing quantitative target information. **Optics express**, v. 26, n. 16, p. 20708-20717, 2018.

WEISS, M.; JACOB, F.; DUSSEILLER, G. Remote sensing for agricultural applications: A meta-review. **Remote sensing of environment**, v. 236, p. 111402, 2020.

WESTRA, E. **Python geospatial development**. Birmingham, UK: Packt Publishing, 2010.

XU, Y. **Molecular plant breeding**. Cabi, 2010.

XU, Y. Envirotyping for deciphering environmental impacts on crop plants. **Theoretical and Applied Genetics**, v. 129, n. 4, p. 653-673, 2016.

XU, Y.; ZHANG, X.; LI, H.; ZHENG, H.; ZHANG, J.; OLSEN, M. S.; VARSHNEY, R. K.; PRASANNA, B. M.; QIAN, Q. Smart breeding driven by big data, artificial

intelligence, and integrated genomic-enviromic prediction. **Molecular Plant**, v. 15, n. 11, p. 1664-1695, 2022.

YANG, B.; ZHU, W.; REZAEI, E. E.; LI, J.; SUN, Z.; ZHANG, J. The optimal phenological phase of maize for yield prediction with high-frequency UAV remote sensing. **Remote Sensing**, v. 14, n. 7, p. 1559, 2022.

YUE, J.; TIAN, J.; TIAN, Q.; XU, K.; XU, N. Development of soil moisture indices from differences in water absorption between shortwave-infrared bands. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 154, p. 216-230, 2019.

ZHANG, W.; WU, C.; ZHONG, H.; LI, Y.; WANG, L. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. **Geoscience Frontiers**, v. 12, n. 1, p. 469-477, 2021.

ZHANG, X.; ZHAO, J.; YANG, G.; LIU, J.; CAO, J.; LI, C.; ZHAO, X.; GAI, J. Establishment of plot-yield prediction models in soybean breeding programs using UAV-based hyperspectral remote sensing. **Remote Sensing**, v. 11, n. 23, p. 2752, 2019.

ZHAO, Q.; YU, L.; DU, Z.; PENG, D.; HAO, P.; ZHANG, Y.; GONG, P. An overview of the applications of earth observation satellite data: impacts and future trends. **Remote Sensing**, v. 14, n. 8, p. 1863, 2022.

### **3 DESENVOLVIMENTO DE PIPELINE INTEGRADO DE AMBIENTIPAGEM 3,5K PARA MODELAGEM G×A**

#### **RESUMO**

A análise da interação genótipos × ambientes (G×A) no melhoramento genético vegetal pode ser muito beneficiada a partir de uma caracterização ambiental detalhada e multiescalar. No entanto, a integração sistemática de dados de diferentes fontes ambientais ainda representa um desafio técnico e operacional, dificultando a vida do melhorista. Este trabalho desenvolveu um *pipeline* metodológico que automatiza a aquisição e o processamento de grandes volumes de dados ambientais, harmonizando dados provenientes das plataformas SoilGrids, WorldClim, NASA POWER. São gerados 3.507 (3,5K) covariáveis ambientais por pixel, abrangendo atributos de solo, clima, relevo e sensoriamento remoto, prontos para análises estatísticas e modelagem G×A. Análises de Componentes Principais (PCA) revelaram padrões de correlação entre agrupamentos temáticos de parâmetros coerentes com a variabilidade ecológica das diferentes plataformas. E ampla explicação por parte das 3,5K covariáveis em todo espectro dos componentes principais PC1 e PC2. A plataforma também permitirá atualizações contínuas do banco de dados e futuramente acesso remoto por meio de API, conferindo flexibilidade para diferentes aplicações no melhoramento. O sistema oferece uma base para compreender e modelar a variabilidade ambiental que afeta o desempenho das culturas e materiais genéticos melhorados.

Palavras-chave: Tipagem Ambiental; Genótipos x Ambientes (G×E); Soilgrids; NASAPOWER; WORLDCLIM

#### **ABSTRACT**

#### **DEVELOPMENT OF AN INTEGRATED 3,5K ENVIROTYPING PIPELINE FOR A G×E MODELING**

The analysis of genotype × environment (G×E) interaction in plant breeding can greatly benefit from detailed and multiscale environmental characterization. However, the systematic integration of data from different environmental sources still poses a technical and operational challenge, complicating the breeder's work. This study developed a methodological pipeline that automates the acquisition and processing of large volumes of environmental data, harmonizing information from the SoilGrids, WorldClim, and NASA POWER platforms. A total of 3,507 (3.5K) environmental covariates per pixel are generated, covering soil, climate, topography, and remote sensing attributes, ready for statistical analyses and G×E modeling. Principal Component Analysis (PCA) revealed correlation patterns among thematic parameter groupings consistent with the ecological variability captured by the different platforms, with a broad explanatory power of the 3.5K covariates across the first two principal components (PC1 and PC2). The platform will also enable continuous updates to the database and, in the future, remote access via API, providing flexibility for various breeding applications. The system offers a foundation for understanding and modeling the environmental variability that affects crop and improved genetic material performance.

Keywords: Envirotyping; Genotype × Environment (G×E); SoilGrids; NASA POWER; WorldClim

### 3.1 INTRODUÇÃO

O ambiente como um todo desempenha o papel ecofisiológico na expressão de características fenotípicas desejáveis. A plasticidade fenotípica, que é a capacidade das plantas de alterar suas características em resposta às variações ambientais, destaca-se para a adaptação e resiliência das culturas. Vários estudos enfatizam a importância da interação genótipos-ambientes (G×E), sendo fundamental para otimizar a seleção de genótipos em programas de melhoramento (van Eeuwijk et al., 2016; Bančić et al., 2024; Resende et al., 2024a). Apenas modelos que incorporam covariáveis ambientais, tais como os modelos ambientômicos, são capazes de prever o comportamento diferencial dos materiais genéticos em diferentes cenários climáticos não experimentados, aumentando a precisão das previsões genotípicas e melhorando o processo de seleção genética (Resende et al., 2021).

Com o avanço das geotecnologias de sensoriamento remoto e a crescente disponibilidade de dados ambientais, a ambientipagem desponta como uma estratégia importante para a compreensão detalhada dos fatores ambientais que afetam a produção agrícola (Cooper et al., 2014; Xu, 2016). Essa abordagem, que integra dados ambientais em larga escala com informações genéticas, permite uma análise preditiva do desempenho das culturas em distintos ambientes e em frente às mudanças climáticas (Messina & Cooper, 2022; Rezaei, 2023). No contexto dos Sistemas de Informações Geográficas (SIG), o uso de dados no formato raster, composto por uma matriz de células chamadas pixels, é fundamental para representar e analisar variáveis ambientais espacialmente distribuídas (Resende et al., 2024b). Cada pixel carrega informações ambientais específicas, como temperatura ou umidade do solo, em uma localização geográfica precisa, permitindo desvendar padrões na interação G×E e facilitando a seleção de genótipos adaptados a diferentes condições.

A implementação de *pipelines* computacionais utilizando linguagens de programação como Python e bibliotecas específicas de SIG mostram ganho em eficiência no processamento e análise de dados ambientais. Bibliotecas como GeoPandas ([pypi.org/project/geopandas](https://pypi.org/project/geopandas/)), RioxArray ([pypi.org/project/rioxarray](https://pypi.org/project/rioxarray/)) e Xarray ([pypi.org/project/xarray](https://pypi.org/project/xarray/)) permitem a manipulação de grandes volumes de dados geoespaciais de forma ágil e precisa, o que é necessário para o desenvolvimento de modelos ambientais aplicados ao melhoramento genético de plantas. Além disso, o uso de plataformas

de dados ambientais como NASA POWER (<https://power.larc.nasa.gov/>), WorldClim (<https://www.worldclim.org/>) e SoilGrids (<https://soilgrids.org/>), SRTM (<https://www.earthdata.nasa.gov/sensors/srtm>) facilita o acesso a dados ambientais em larga escala, fundamentais para a ambientipagem.

Cada uma dessas plataformas poderá contribuir para a análise ambientômica, permitindo que diferentes aspectos do ambiente sejam integrados e analisados em conjunto (Xu et al., 2022). Por exemplo, SoilGrids fornece dados sobre propriedades do solo em diferentes profundidades, enquanto o NASA POWER disponibiliza dados climáticos históricos e atuais. Além disso, o WorldClim oferece dados climáticos de alta resolução, o SRTM (*Shuttle Radar Topography Mission*) fornece modelos digitais de elevação, e o MODIS (*Moderate Resolution Imaging Spectroradiometer*) disponibiliza informações sobre a cobertura da terra e a vegetação (Resende et al., 2024b). A combinação de dados de ambientipagem de diferentes plataformas permite uma análise integrada do ambiente, essencial para alimentar os modelos ambientômicos, otimizando a seleção de genótipos com alto potencial adaptativo (Crossa et al., 2021).

A integração entre a biologia, especialmente no campo do melhoramento de plantas e animais, e as ferramentas computacionais apresenta desafios, principalmente para profissionais que não possuem habilidades em programação (Bernardo, 2020). No entanto, o uso de dados ambientais é essencial para o desenvolvimento de estratégias de ambientipagem que otimizem a seleção de genótipos. Uma alternativa prática para esses usuários é o acesso a banco de dados que disponibilizam informações geoprocessadas de forma simplificada, eliminando a necessidade de programação direta em SIG para aquisição de dados descentralizados, facilitando a integração de múltiplas fontes de dados, e.g., como dados de biodiversidade informatizados (Norton, 2021); e genômicos (Swaminathan et al., 2016). Assim, melhoristas poderiam utilizar ferramentas para realizar análises ambientômicas e tomar decisões embasadas (Resende et al., 2021). A adoção de soluções que combinem Banco de Dados ambientais com estruturas amigáveis pode reduzir a lacuna entre o conhecimento biológico e as ferramentas computacionais, promovendo maior adesão às práticas de ambientipagem e aprimorando os resultados em programas de melhoramento genético.

A importância deste trabalho se manifesta em duas frentes diferentes. A primeira é a concatenação de diferentes conjuntos de geodados ambientais, de forma a representar o ambiente de maneira exaustiva. A segunda é a otimização do processo de aquisição de dados

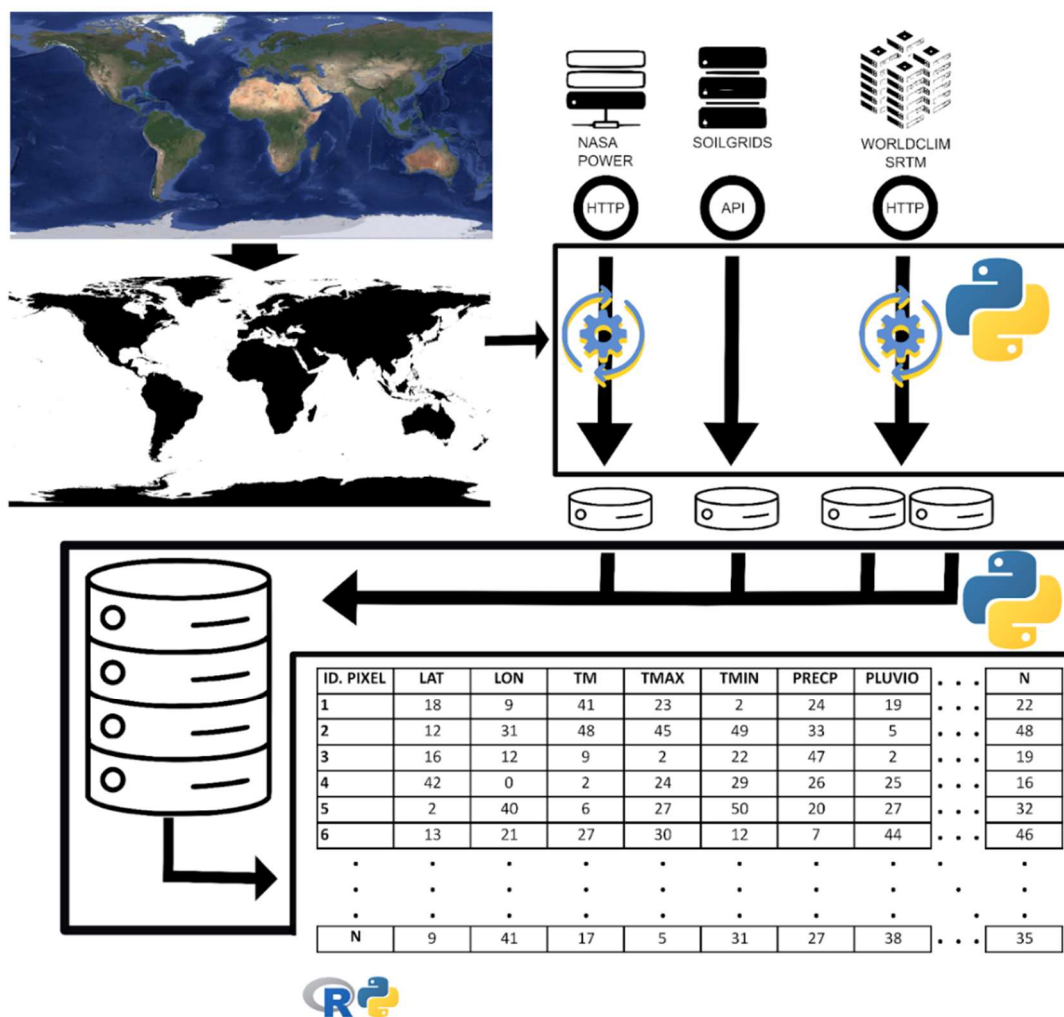
para ambientipagem, uma vez que esse processo costuma ser complexo, trabalhoso e moroso. Além disso, o trabalho facilita a disponibilização desses dados em uma estrutura matricial já familiar aos melhoristas, permitindo sua utilização sem a necessidade de conhecimentos aprofundados em linguagens de programação aplicadas aos Sistemas de Informação Geográfica (SIG).

## 3.2 MATERIAIS E MÉTODOS

Para a execução desse *pipeline* foram utilizadas bibliotecas que abrangem tanto módulos padrão do Python quanto bibliotecas externas especializadas. Entre as bibliotecas padrão, os ([pypi.org/project/os/](https://pypi.org/project/os/)) e `pathlib` ([pypi.org/project/pathlib/](https://pypi.org/project/pathlib/)) permitem manipular caminhos e diretórios, enquanto `json` ([pypi.org/project/json5/](https://pypi.org/project/json5/)) e `datetime` ([pypi.org/project/DateTime/](https://pypi.org/project/DateTime/)) oferecem ferramentas para trabalhar com dados em formato JSON (JavaScript Object Notation – que são uma representação leve e estruturada de dados em formato texto, organizados em pares de chave-valor, amplamente utilizada para a troca de informações entre sistemas e APIs), datas e fluxos de entrada e saída. A biblioteca `requests` ([pypi.org/project/requests/](https://pypi.org/project/requests/)) foi utilizada para realizar requisições HTTP, e `zipfile` ([pypi.org/project/zipfile2/](https://pypi.org/project/zipfile2/)) permite a manipulação de arquivos ZIP. Esses módulos fornecem a base para a manipulação e processamento de dados requisitados nas diferentes plataformas de dados ambientais.

Dentre as bibliotecas externas, `pandas` ([pypi.org/project/pandas/](https://pypi.org/project/pandas/)) e `numpy` ([pypi.org/project/numpy/](https://pypi.org/project/numpy/)) foram empregadas para a manipulação e análise de dados tabulares e computação científica, respectivamente. `rioxarray` e `rasterio` ([pypi.org/project/rasterio/](https://pypi.org/project/rasterio/)) foram usadas para o trabalho com dados geoespaciais, devido a sua tratabilidade com arquivos raster. `dotenv` ([pypi.org/project/dotenv/](https://pypi.org/project/dotenv/)) auxiliou a gerenciar variáveis de ambiente, enquanto `xarray` forneceu as funções para processamento de grandes conjuntos de dados multidimensionais. Complementando essas ferramentas, os módulos customizados como `'logger'`, `'time_it'`, foram criados com as funcionalidades específicas de controle de eficiência no processo de aquisição de dados mediante requisição. Esses componentes formam a base de bibliotecas necessárias para a aquisição via *download* dos dados brutos, seu processamento e posterior disponibilização dos dados de Ambientipagem em formato para utilizar diretamente em modelos estatísticos de ambientômica para análises de interação  $G \times A$  (Resende et al., 2024b).

Para criar a base geoespacial rasterizada para concatenação dos diferentes bancos de dados, foi confeccionado um arquivo de grade mundial por meio de um dos script que converte dados vetoriais de toda a superfície do globo obtido por meio da plataforma do Worldclim em uma grade raster com resolução ajustável em graus (e.g., 0,05 grau  $\approx$  5,55 km; 0,1 grau  $\approx$  11,10 km; 0,5 grau  $\approx$  55,5 km). As configurações essenciais, como o caminho dos arquivos de entrada (`world_countries.geojson`) e o destino dos arquivos de saída (`world_raster.tif`), foram carregadas a partir de um arquivo `.env`, sendo esse padrão aplicado para todas as etapas. A classe `WorldBase` foi implementada para realizar esse processo, utilizando a biblioteca `geocube` para rasterizar os polígonos com uma projeção EPSG:4326. O raster resultante, com valores constantes atribuídos a cada célula, foi gerado e salvo em formato TIFF, preparado para padronização de dados geoespaciais nas próximas etapas. Cada célula do raster representa uma unidade espacial uniforme. Essa etapa inicial de estruturação espacial do pipeline está representada na Figura 1, que detalha o fluxograma do processo de ambientipagem.



**Figura 1.** Fluxograma do *pipeline* de ambientipagem, demonstrando a coleta automatizada de dados ambientais multiescalares das plataformas WorldClim, NASA POWER e SoilGrids. As variáveis são adquiridas com base em uma grade espacial global, armazenadas como arquivos .tiff para cada covariável. Após a aquisição dos arquivos foram padronizados e concatenados, sendo armazenados em um banco de dados robusto. A padronização foi realizada a extração dos valores ambientais por pixel, com base na interseção espacial das camadas. Utilizando scripts Python, foi extraído valores centralizados por coordenadas, correspondentes às células da grade global, resultando em uma matriz tabular estruturada por pixel e covariável.

As plataformas SoilGrids, NASA POWER e WorldClim, disponibilizam fontes de dados ambientais processados e confiáveis, que oferecem uma gama diversa de informações necessárias para a análise de diferentes aspectos do ambiente (Figura 1-A). Cada plataforma contribui com banco de dados específicos desde propriedades do solo, clima e topografia, até a cobertura da terra e reanálises atmosféricas, permitindo uma compreensão exaustiva e integrada do ambiente. Essas plataformas fornecem dados em larga escala, que podem ser utilizados em estudos de ambientipagem. A aquisição desses dados pode ser feita como o uso de suas interfaces web ou APIs, com opções para personalizar os

parâmetros de consulta, possibilitando a implementação de seus downloads em scripts de rotina para processamento e análises.

### 3.2.1 DETALHAMENTO TÉCNICO POR PLATAFORMA

Os dados de solo foram adquiridos por meio do pacote Python oficial da plataforma SoilGrids, mantida pelo ISRIC (*International Soil Reference and Information Centre*). As requisições abrangeram dez variáveis edáficas fundamentais para a caracterização ambiental, como densidade aparente (bdod), capacidade de troca catiônica (cec), teor de argila (clay), pH em H<sub>2</sub>O (phh2o), teor de matéria orgânica (ocd, soc), teor de nitrogênio (nitrogen), entre outras. Cada variável foi solicitada em múltiplas profundidades do perfil do solo (0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm, 100–200 cm) e em diferentes estatísticas (média, quantis 5% e 95%). O script utilizado para essa tarefa permitiu o download direto dos arquivos em formato GeoTIFF, facilitando a integração com rotinas de processamento posteriores. Os arquivos foram organizados por tipo de variável, profundidade e estatística. Um mecanismo de controle baseado em JSON registrou os downloads já realizados, evitando redundância de execução e otimizando a requisição do pipeline.

O conjunto de dados climáticos da plataforma WorldClim, versão 2.1, foi utilizado para integrar variáveis atmosféricas e topográficas ao banco de dados. Os arquivos foram obtidos no formato compactado (.zip), abrangendo dados de elevação (altitude), temperatura média (tavg), máxima (tmax) e mínima (tmin), precipitação acumulada (prec), radiação solar (srad), pressão de vapor (vapr) e velocidade do vento (wind). Todos os dados foram obtidos na resolução de 2,5 minutos de grau (~4,5 km), padrão compatível com a escala do raster de referência definido no *pipeline*. Após o download, os arquivos foram extraídos e convertidos para GeoTIFF, e em seguida reamostrados com interpolação cúbica para alinhamento espacial ao pixel definido em grade. Posteriormente, os rasters foram recortados para os limites definidos pelo raster base global, assegurando compatibilidade espacial com os demais conjuntos de dados.

A plataforma NASA POWER (*Prediction Of Worldwide Energy Resource*) foi utilizada como fonte principal de dados atmosféricos modernos derivados de reanálises. O acesso se deu por meio de chamadas à API da NASA em blocos espaciais (tiles) e temporais (anuais), com parâmetros definidos para precipitação corrigida (PRECTOTCORR),

temperatura média (T2M), máxima (T2M\_MAX) e mínima (T2M\_MIN), umidade relativa (RH2M), temperatura de orvalho (T2MDEW), radiação de onda curta e longa (ALLSKY\_SFC\_SW\_DWN e LW\_DWN), e velocidade do vento (WS2M). Os arquivos foram inicialmente baixados no formato .JSON e agregados por parâmetro e ano em arquivos únicos. A conversão para GeoTIFF foi realizada com interpolação espacial do tipo nearest, respeitando a grade de referência definida no raster base, etapa necessária para integração com os dados obtidos das outras plataformas.

Todos os arquivos raster resultantes do processamento das três plataformas foram então organizados e consolidados em um banco de dados multidimensional no formato .ZARR. Esse formato foi escolhido pela sua eficiência no armazenamento e leitura de grandes volumes de dados ambientais, permitindo consultas rápidas e estruturação escalável. A organização por bandas respeitou a origem e o nome das variáveis, facilitando o rastreamento de cada camada ambiental. Complementarmente, com o objetivo de monitorar e otimizar o desempenho do *pipeline*, foi implementado um decorador personalizado denominado *time\_it*. Esse componente registra automaticamente o tempo de execução das funções críticas envolvidas nas etapas de download, amostragem e estruturação. Os tempos computacionais foram armazenados em logs organizados, possibilitando o diagnóstico de gargalos e a avaliação da eficiência das rotinas implementadas.

### 3.2.2 ANÁLISE ESTATÍSTICA E EXPLORATÓRIA DAS COVARIÁVEIS AMBIENTAIS

Antes das análises multivariadas, todas as variáveis ambientais foram submetidas à padronização por meio do cálculo de escores  $Z$  ( $Z$ -score), etapa necessária para assegurar a comparabilidade entre variáveis com escalas e unidades distintas. A transformação foi realizada utilizando a fórmula  $Z = (x - \mu)/\sigma$ , onde  $\mu$  representa a média e  $\sigma$  o desvio padrão de cada variável. A padronização elimina o efeito de magnitude e evita que variáveis com maior amplitude numérica dominem a variância total em técnicas como a Análise de Componentes Principais (PCA). Esse procedimento foi implementado com a biblioteca *pandas*, e aplicado diretamente sobre a matriz final de pixels  $\times$  variáveis, resultando em um novo conjunto de dados normalizado. A padronização garante equidade na contribuição de cada covariável para os eixos de variação, sendo necessário para diagnósticos posteriores, como detecção de *outliers* através do desvio padrão, testes de distribuição e verificação da homogeneidade entre grupos de covariáveis ambientais.

A matriz padronizada foi submetida a análises exploratória das covariáveis ambientais. Inicialmente, foram calculadas estatísticas descritivas clássicas: média, desvio padrão, valor mínimo, valor máximo, mediana e intervalo interquartil (IQR). As estatísticas foram produzidas tanto para o conjunto completo de variáveis quanto agrupadas em domínios temáticos previamente definidos com base em padrões de nomenclatura e conhecimento funcional: temperatura do ar, precipitação e umidade, radiação e energia, solo (textura, estrutura e química), topografia, entre outros. Essa agregação temática permitiu observar padrões de variabilidade interna dentro de grupos de variáveis relacionados, facilitando a interpretação ecológica dos dados. As análises foram realizadas com pandas e os resultados organizados em tabelas exportáveis (.csv).

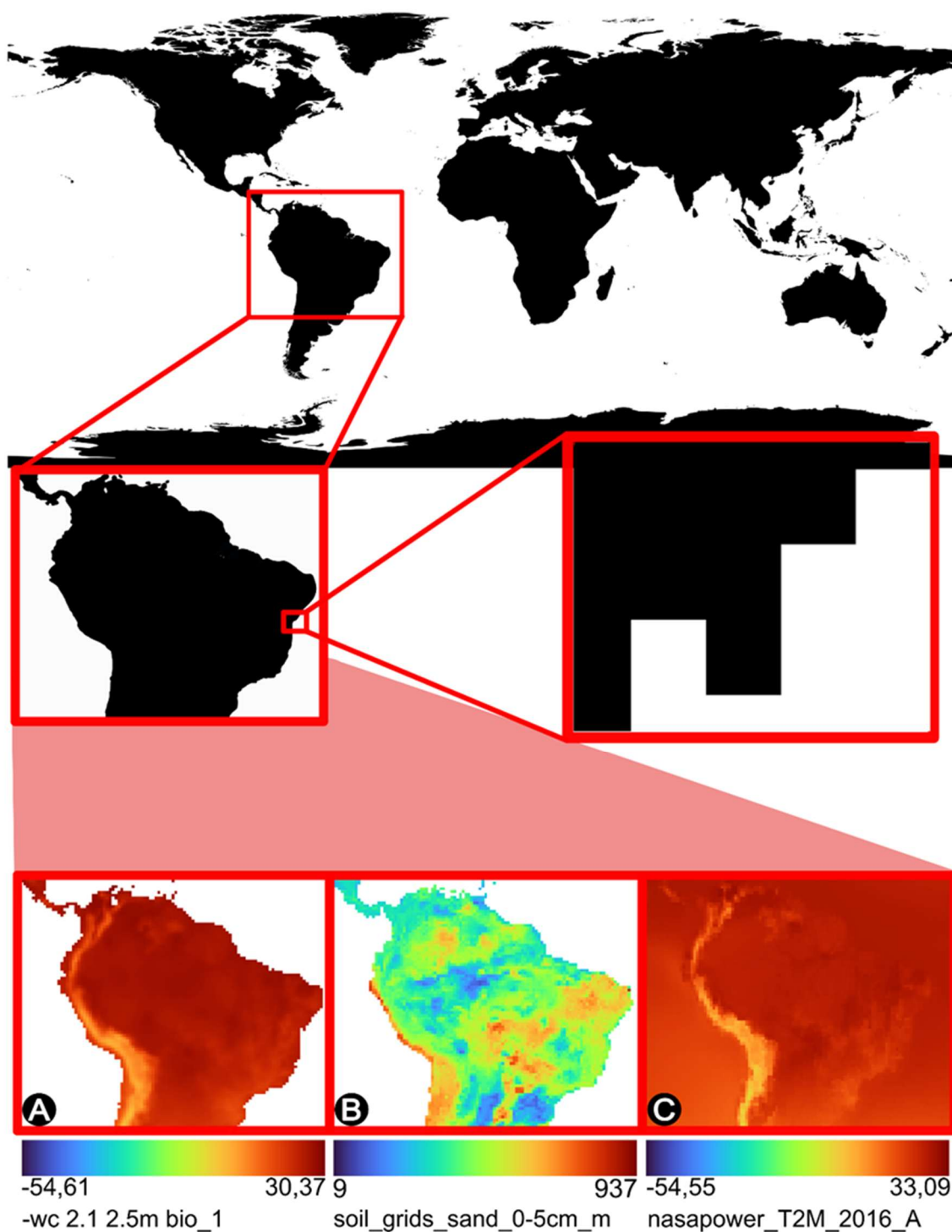
Com a matriz padronizada e as estatísticas descritivas, foi realizada a Análise de Componentes Principais (PCA), com o objetivo de reduzir a dimensionalidade dos dados e identificar os principais gradientes ambientais presentes no conjunto multivariado. A PCA foi implementada com a biblioteca scikit-learn, utilizando os dois primeiros componentes principais (PC1 e PC2), que juntos explicaram 68,2% da variância total. O procedimento incluiu a remoção de outliers extremos com base em desvios-padrão das projeções, assegurando maior estabilidade visual dos vetores no biplot. O gráfico resultante representa os pixels amostrados no plano PC1 × PC2, juntamente com vetores das variáveis agrupadas por plataforma (SoilGrids, WorldClim, NASA POWER e SRTM). Essa visualização também evidenciou redundâncias e colinearidades. Os eixos PC1 e PC2 foram exportados com coordenadas geográficas dos pixels e convertidos em arquivos raster (GeoTIFF), permitindo análises espaciais e sobreposição com dados genéticos e ecológicos.

### **3.3 RESULTADOS**

Foi possível automatizar a aquisição e o processamento dos dados de elevação (SRTM), da plataforma WorldClim, da plataforma SoilGrids e da plataforma NASA POWER. A aquisição desses dados foi integrada ao *pipeline* metodológico, permitindo padronização e consistência na obtenção e processamento das informações. No total, o banco de dados ambiental multiescalar atualmente compreende 3.507 camadas ambientais provenientes de diferentes fontes: foram adquiridas 244 covariáveis do SoilGrids, 104 covariáveis do WorldClim, incluindo o perfil de elevação SRTM. Além disso, incorporaram-se 3.159 camadas climáticas históricas do NASA POWER. Essas informações foram obtidas

utilizando o protocolo de requisição HTTP para aquisição de dados do WorldClim e APIs para aquisição de dados SoilGrids e NASA POWER, com extração em formato JSON, posterior agregação, interpolação espacial e conversão para GeoTIFF. O sistema automatizado garante reprodutibilidade, rastreabilidade e controle de qualidade no armazenamento das variáveis ambientais que compõem a base.

O processo de automação dos dados iniciou-se com a criação de uma linha de grade global em formato GeoTIFF, derivada do processamento dos limites continentais disponibilizados pelo WorldClim. A resolução dessa grade pode ser ajustada de acordo com as necessidades do projeto, tendo como limite mínimo um pixel de 5 km, devido à complexidade de geração e processamento de grades com resoluções menores. Essa grade garante compatibilidade entre os diferentes conjuntos de dados adquiridos, permitindo a padronização das covariáveis em uma mesma matriz de referência. A flexibilidade para ajustar a resolução permite que o *pipeline* seja adaptado a diferentes contextos de pesquisa, dependendo da escala espacial necessária. A Figura 2 ilustra visualmente essa malha global com recortes progressivos para América do Sul e um pixel da grade, exemplificando a padronização espacial com três camadas ambientais extraídas: uma variável bioclimática (WorldClim), uma edáfica (SoilGrids) e uma climática histórica (NASA POWER).



**Figura 2.** Visualização da malha global construída para o *pipeline* de ambientipagem, com destaque progressivo para a América do Sul e subsequente ampliação de um pixel da grade, ilustrando a resolução espacial de aproximadamente 0,5° (cerca de 50 km). Na parte inferior, são apresentados exemplos de covariáveis extraídas das três plataformas ambientais: (A) **WorldClim** – temperatura média anual (BIO1), com valores variando de  $-54,61$  °C a  $30,37$  °C; (B) **SoilGrids** – teor de areia na camada de 0–5 cm (sand\_0-5cm), com valores entre  $0$  e  $937$  g/kg; e (C) **NASA POWER** – temperatura média anual de 2016 (T2M), com variação de  $-54,55$  °C a  $33,09$  °C. As imagens ilustram a padronização espacial e a diversidade de atributos ambientais considerados no pipeline. Os códigos nos gradientes das 3 imagens são as codificações lógicas utilizadas nas 3,5k variáveis.

A plataforma SoilGrids fornece informações sobre as características físicas e químicas do solo, pertinentes para compreender a resposta das plantas ao ambiente edáfico. Foram extraídas 244 covariáveis, incluindo textura do solo (proporção de areia, silte e argila), teor de matéria orgânica, densidade do solo, carbono orgânico, pH e capacidade de retenção de água. Esses dados são fornecidos para seis camadas de profundidade padrão (0–5, 5–15, 15–30, 30–60, 60–100 e 100–200 cm), cada uma com quatro estatísticas preditivas (quantis Q0.05, Q0.95, mediana e média). Essa estrutura representa a incerteza das estimativas e fornece um intervalo de confiança para os parâmetros do solo. A composição dessas combinações resulta em 244 arquivos GeoTIFF. Além disso, variáveis como teor de nitrogênio total e capacidade de troca catiônica (CTC) enriquecem o conjunto de dados, proporcionando uma visão abrangente do perfil edáfico em escala global, com potencial para análise da variabilidade do solo em diferentes contextos.

Os dados climáticos do WorldClim complementam a base com informações representativas da climatologia de longo prazo. Essa plataforma fornece dados médios mensais de sete variáveis meteorológicas: temperatura média (tavg), temperatura máxima (tmax), temperatura mínima (tmin), precipitação (prec), pressão de vapor (vap), radiação solar (srad) e velocidade do vento (wind). Cada uma dessas variáveis é representada por 12 camadas mensais, totalizando 84 arquivos. A partir dessas camadas foram derivados 19 índices bioclimáticos (BIO1 a BIO19), que resumem padrões anuais, sazonais e extremos do clima. Entre essas variáveis destacam-se a temperatura média anual (BIO1), a amplitude térmica diária média (BIO2), a isotermalidade (BIO3), a precipitação anual (BIO12), a precipitação do mês mais seco (BIO14) e a do trimestre mais quente (BIO18). As variáveis bioclimáticas foram utilizadas na resolução de 2,5 minutos de grau ( $\approx 5$  km), mantendo compatibilidade com as demais fontes de dados do pipeline.

Além disso, para incorporar a dimensão altitudinal foi integrado os dados topográficos de elevação (SRTM), derivados da mesma base do WorldClim. Esses dados são disponibilizados em resoluções de até 1 km<sup>2</sup> e permitem agregar a dimensão altitudinal às análises ambientais. A presença da variável de relevo contribui para representar a heterogeneidade topográfica, com influência direta sobre padrões de temperatura, umidade e escoamento. Além disso, as variáveis do WorldClim, especialmente os índices bioclimáticos, permitem análises aprofundadas de tendências anuais, sazonalidades e fatores ambientais extremos – como a temperatura do mês mais frio (BIO6) ou a precipitação do

trimestre mais seco (BIO17) – que influenciam diretamente o desempenho de espécies vegetais e sistemas agrícolas.

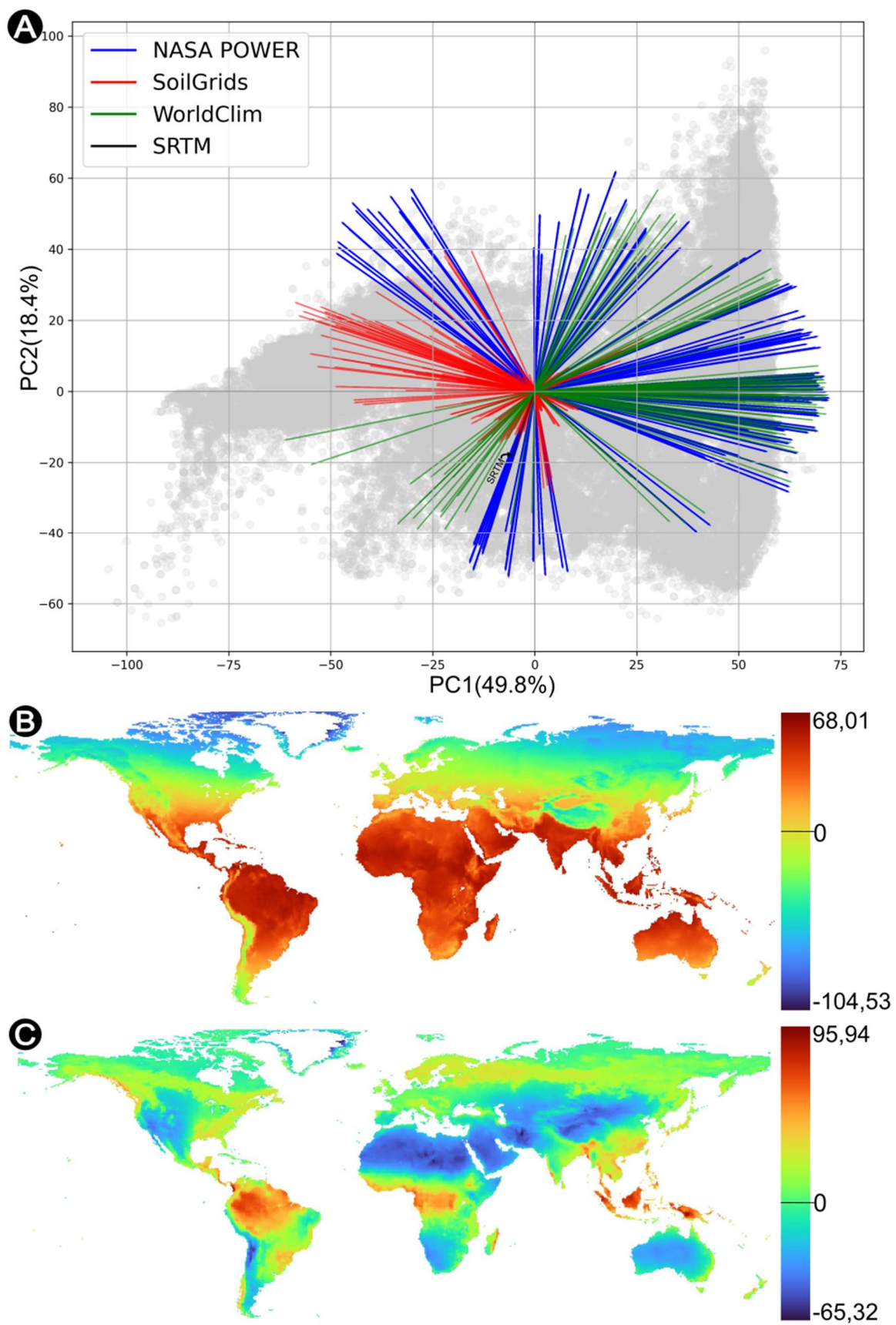
A plataforma NASA POWER foi incorporada com o objetivo de expandir a dimensão temporal da base climática. Por meio de sua API, foram obtidas variáveis meteorológicas mensais e anuais de 1990 a 2016, incluindo: temperatura média do ar a 2 m (T2M), temperatura máxima (T2M\_MAX), mínima (T2M\_MIN), precipitação total corrigida (PRECTOTCORR), velocidade do vento a 2 m (WS2M), umidade relativa (RH2M), temperatura de orvalho (T2MDEW) e irradiâncias de onda curta e longa em céu aberto (ALLSKY\_SFC\_SW\_DWN e ALLSKY\_SFC\_LW\_DWN). Cada variável foi segmentada em 13 intervalos temporais por ano (12 meses + média anual), totalizando 351 arquivos por variável ao longo de 27 anos. Com nove variáveis extraídas, o volume final foi de 3.159 arquivos raster em formato GeoTIFF. Essa base climática histórica, distribuída globalmente, viabiliza análises de tendências, variabilidade interanual e simulações ambientais retrospectivas com diferentes abordagens estatísticas e modelagens ambientais baseadas em séries temporais.

A síntese dessas informações esta apresentada na Tabela 1, que sistematiza os principais parâmetros ambientais extraídos de cada plataforma, categorizando-os por tipo, e contabilizando número de covariáveis por grupo temático, resolução espacial, abrangência temporal e frequência de coleta. Permite uma visão sintética e comparativa da composição da base de dados multiescalar construída. Observa-se, por exemplo, que a plataforma SoilGrids contribuiu com dados edáficos em alta resolução (250 m), organizados por atributos físicos e químicos do solo, enquanto o WorldClim forneceu variáveis bioclimáticas e climáticas médias de longo prazo com resolução de 1 km. Por sua vez, a NASA POWER, apesar de apresentar menor resolução espacial ( $\approx 55$  km), abrange uma extensa série temporal (1990–2016) com dados diários, o que confere alta densidade temporal à base. A organização temática dos parâmetros permite maior clareza para análises posteriores, como seleção de covariáveis, redução de dimensionalidade e aplicação em modelos preditivos baseados em interações genótipo-ambiente.

**Tabela 1.** Descrição das Fontes de Dados Ambientais e Seus Parâmetros Temáticos

| Fonte             | Categoria                  | Parâmetros | Covariáveis | Resolução Original | Janela Temporal | Frequência |
|-------------------|----------------------------|------------|-------------|--------------------|-----------------|------------|
| NASA<br>POW<br>ER | Temperatura                | 4          | 1352        | 0,5°<br>(~55 km)   | 1990–2024       | Diário     |
|                   | Umidade                    | 1          | 338         |                    |                 |            |
|                   | Precipitação               | 1          | 338         |                    |                 |            |
|                   | Radiação                   | 2          | 676         |                    |                 |            |
|                   | Vento                      | 1          | 338         |                    |                 |            |
| Soil<br>Grids     | Composição Física do Solo  | 4          | 96          | 250 m              | Estático        | Estático   |
|                   | Composição Química do Solo | 6          | 144         |                    |                 |            |
| World<br>Clim     | Temperatura                | 3          | 36          | 1 km               | 1970–2000       | Estático   |
|                   | Umidade                    | 1          | 12          |                    |                 |            |
|                   | Precipitação               | 1          | 12          |                    |                 |            |
|                   | Radiação                   | 1          | 12          |                    |                 |            |
|                   | Vento                      | 1          | 12          |                    |                 |            |
|                   | Elevação                   | 1          | 1           |                    |                 |            |
|                   | Índices Bioclimáticos      | 19         | 19          |                    |                 |            |

Com o intuito de reduzir a dimensionalidade das 3.507 camadas ambientais extraídas e identificar os principais gradientes de variação ambiental, foi realizada uma PCA. Essa abordagem multivariada permitiu sintetizar os dados e evidenciar padrões latentes. A Figura 3 apresenta os resultados obtidos. O painel A mostra o biplot dos dois primeiros componentes principais (PC1 e PC2), responsáveis por 49,8% e 18,4% da variância total, respectivamente. Os vetores estão agrupados por plataforma: NASA POWER (azul), SoilGrids (vermelho), WorldClim (verde) e SRTM (preto), evidenciando a contribuição relativa de cada conjunto de variáveis. A distribuição dos vetores indica a complementaridade entre as fontes de dados ambientais, com PC1 fortemente associado a variáveis climáticas e PC2 relacionado a características edáficas e topográficas. Os painéis B e C da Figura 3 ilustram a espacialização global dos scores de PC1 e PC2, respectivamente, destacando a distribuição geográfica dos gradientes ambientais. O primeiro componente apresenta um padrão latitudinal consistente com a variação térmica e energética, enquanto o segundo revela uma variabilidade regional associada a solos e relevo.



**Figura 3.** Análise de Componentes Principais (PCA) aplicada às covariáveis ambientais extraídas do pipeline de ambientipagem. Em A, o biplot exibe a distribuição dos pixels amostrados no espaço

multivariado dos dois primeiros componentes (PC1: 49,8%; PC2: 18,4%) e os vetores das variáveis agrupadas por plataforma: **NASA POWER (azul)**, **SoilGrids (vermelho)**, **WorldClim (verde)** e **SRTM (preto)**. A dispersão dos vetores evidencia a contribuição relativa de cada plataforma à variância ambiental. Em **B**, o mapa mostra a especialização global do score **PC1** e em **C** o mapa mostra a especialização global do score **PC2**.

A Figura 3A apresenta o biplot dos dois componentes principais (PC1 e PC2), que juntos explicam aproximadamente 68,2% da variância total dos dados ambientais extraídos do banco de dados final. Observa-se que os vetores estão bem distribuídos no espaço multivariado, com agrupamentos nítidos por plataforma: as variáveis do NASA POWER (em azul) concentram-se na porção direita do gráfico, indicando forte correlação com o PC1; já as variáveis do SoilGrids (vermelho) mostram maior dispersão vertical, contribuindo mais intensamente para o PC2. As variáveis do WorldClim (verde) se distribuem em direção ao quadrante superior direito, também com forte correlação positiva com PC1 e moderada com PC2. O vetor do SRTM (em preto) é isolado, com peso principalmente negativo sobre PC2. Essa configuração revela a complementaridade das fontes de dados, com os vetores demonstrando direções distintas e sugerindo que cada plataforma captura aspectos próprios do ambiente — como clima, solo e relevo — que, ao serem integrados, compõem um espaço ambiental robusto e multiescalar.

A Figura 3B mostra a espacialização global dos scores do primeiro componente principal (PC1), responsável por 49,8% da variância nos dados ambientais. Nota-se um padrão latitudinal bem definido: áreas tropicais e subtropicais apresentam valores mais elevados de PC1 (em tons quentes), enquanto regiões de altas latitudes — como o Norte da América, Europa e Ásia — concentram valores mais baixos (em tons frios). A Figura 3C representa os scores do segundo componente principal (PC2), responsável por 18,4% da variância ambiental, e evidencia um padrão espacial mais complexo e regionalizado. Observam-se valores extremos tanto positivos quanto negativos em diferentes regiões do globo, sem um gradiente latitudinal claro. Regiões como o Oeste da América do Norte, África Setentrional, Oriente Médio, Austrália e Ásia Central apresentam valores baixos (em azul), enquanto áreas próximas a linha do equador, como a região Amazônica, floresta do Congo e Indonésia exibem valores positivos (em tons quentes).

### 3.4 DISCUSSÃO

A Figura 2 detalha a resolução espacial da malha gerada, com pixels de aproximadamente 5 km<sup>2</sup>, o que possibilita uma caracterização refinada de ambientes agrícolas. A escolha dessa escala envolve um compromisso entre granularidade ambiental e viabilidade computacional, já que pixels em escalas menores aumentam exponencialmente o tamanho do banco de dados trabalhado, trazendo gargalos de análise devido à capacidade de processar grandes volumes de dados. Por outro lado, pixels maiores podem mascarar variações locais críticas, diminuindo a resolução analítica do modelo. Esse desafio de integração entre diferentes resoluções foi enfrentado por meio da reamostragem e interpolações, conforme também sugerido por Bahia et al. (2025). Essa normalização permitiu que covariáveis com escalas e unidades distintas fossem comparadas de forma equitativa. Jarquin et al. (2021) reforçam que a resolução espacial influencia diretamente na detecção de zonas de adaptação específica, sendo decisiva para separar regiões homogêneas de resposta genotípica.

Na sequência, a Tabela 1 sintetiza as plataformas e suas respectivas covariáveis, evidenciando a integração de variáveis climáticas, edáficas e topográficas em um mesmo arcabouço analítico. Essa abordagem multiescalar é necessária para representar a complexidade dos ambientes agrícolas e entender como os genótipos respondem à variabilidade do meio. O uso de covariáveis derivadas por pixel individual, como mostrado na Figura 2, permite classificar ambientes por similaridade ambiental e não apenas por proximidade geográfica, o que facilita a formação de TPEs mais representativos e estatisticamente robustos (Resende et al., 2021). Essa estratificação ecogeográfica também se alinha aos estudos de Crespo-Herrera et al. (2021), que propõem clusters ambientais para seleção genotípica eficiente. A organização dessas covariáveis em domínios temáticos (clima, solo, relevo) potencializa o uso de PCA para integrar diferentes camadas em eixos latentes coerentes com os padrões ambientais observados em campo.

Com base nessa matriz multiescalar, aplicou-se uma Análise de Componentes Principais (PCA) para reduzir a dimensionalidade das variáveis ambientais, evidenciando padrões espaciais relevantes para o melhoramento genético. O primeiro componente principal (PC1), que explicou 49,8% da variância, demonstrou-se associado a um gradiente climático latitudinal, especialmente temperatura e radiação solar, refletindo o papel predominante das condições atmosféricas na variabilidade macroambiental. A Figura 3B mostra essa transição de ambientes tropicais para temperados, e seus resultados corroboram

os achados de Bahia et al. (2025) e Jarquin et al. (2014), que evidenciam a importância de ambientes contrastantes para detecção de interações G×E e formação de TPEs (Target Population of Environments). Esses autores destacam que a estratificação adequada desses ambientes permite selecionar genótipos mais resilientes a condições extremas. Além disso, Crespo-Herrera et al. (2021) reforçam que a modelagem de TPEs com base em eixos climáticos melhora a capacidade de recomendação de cultivares.

O segundo componente principal (PC2), explicando 18,4% da variância, foi associado a variáveis edáficas como capacidade de troca catiônica (CEC), pH e teor de argila, além da altitude, revelando padrões mais complexos e regionalizados. A Figura 3C evidencia sua distribuição espacial, com valores baixos em partes do Oeste da América do Norte, África setentrional e Ásia Central, e valores elevados em regiões tropicais como Amazônia e Indonésia. Essa variabilidade mais heterogênea reflete a influência de microambientes e de fatores edafoclimáticos mais localizados, cuja caracterização é pertinente para recomendações genotípicas com maior especificidade ambiental. Essa tendência está em consonância com as discussões de Jarquin et al. (2021) e Resende et al. (2025), que enfatizam a relevância de componentes ambientais regionais na predição do desempenho genotípico. Ademais, a inclusão de covariáveis edáficas no delineamento ambiental contribui para o entendimento de limitações específicas do solo que influenciam a eficiência fisiológica das cultivares.

A segmentação ambiental com base nos scores dos componentes principais abre caminho para o delineamento de breeding zones mais precisas e adaptadas. Conforme mostrado por Jarquín et al. (2021), a estratificação via PCA ambiental pode superar abordagens baseadas apenas em localização geográfica, uma vez que reflete a funcionalidade ambiental real de cada região. A Figura 3 ilustra essas zonas potenciais, permitindo uma seleção orientada por função ambiental, o que torna o processo seletivo mais alinhado às exigências do ambiente real de cultivo. Esse enfoque foi recentemente defendido por Resende et al. (2025), que demonstraram ganhos em acurácia preditiva ao integrar dados ambientais em modelos GIS-based de interação G×E. Essa estratificação orientada a função permite, ainda, a otimização da alocação de recursos em ensaios, reduzindo redundâncias experimentais e ampliando a eficiência da seleção.

A predição ambiental orientada por dados multiescalares vem se mostrando como uma estratégia no desenvolvimento de cultivares adaptadas a diferentes zonas agroecológicas. Por meio da caracterização ambiental detalhada, como a promovida pelas

covariáveis derivadas de plataformas como SoilGrids, WorldClim e NASA POWER, torna-se possível antecipar o desempenho de genótipos em locais não testados, reduzindo custos com ensaios multilocais e otimizando o delineamento experimental. Modelos que incorporam informações ambientais em suas estruturas preditivas, como os baseados em regressão por componentes principais ou em kernels ambientais, têm demonstrado ganhos em acurácia preditiva (Resende et al., 2021). Além disso, essa abordagem contribui para a escolha eficiente de ambientes de teste e acelera o ciclo de seleção, uma vez que permite identificar as zonas com maior retorno seletivo para cada perfil genético. A predição ambiental, portanto, não só aprimora a alocação de recursos em programas de melhoramento como também antecipa cenários de adaptação futura frente às mudanças climáticas e à variabilidade edafoclimática interanual.

Por fim, a combinação de big data ambiental e modelos genômicos representa uma vertente para a seleção de cultivares, especialmente em cenários de alta variabilidade climática. Como propõe Jarquín et al. (2014), a modelagem com normas de reação permite entender a resposta genotípica a gradientes ambientais contínuos, utilizando covariáveis ambientais como preditores. A integração dessas abordagens no *pipeline* desenvolvido neste trabalho amplia as perspectivas para a adoção de estratégias de melhoramento mais resilientes, eficientes e adaptadas à heterogeneidade ambiental. Esse tipo de abordagem torna-se cada vez mais relevante frente às mudanças climáticas e à necessidade de recomendação de cultivares robustas em diferentes condições edafoclimáticas. Assim, a convergência entre geotecnologias, modelagem estatística e genética quantitativa representa uma fronteira promissora para a agricultura de precisão e para programas de melhoramento genético adaptativo.

### **3.5 CONCLUSÃO**

Neste trabalho, foram disponibilizadas 3.507 covariáveis ambientílicas, organizadas em uma matriz de pixel por covariável e prontamente aplicáveis a estudos de interação genótipo  $\times$  ambiente (G $\times$ E). As covariáveis foram derivadas de plataformas públicas como NASAPOWER, SoilGrids e WorldClim, abrangendo uma série temporal de 1990 a 2016 e cobrindo variáveis climáticas, edáficas e topográficas. A malha espacial gerada apresenta pixels de aproximadamente 5 km<sup>2</sup>, o que se mostra adequado para representar os talhões e unidades experimentais comumente utilizados em programas de melhoramento vegetal. A

partir dessa base, foi aplicada uma Análise de Componentes Principais (PCA), a qual evidenciou uma ampla representatividade ambiental ao longo de diferentes gradientes ecológicos. Os vetores principais resultantes demonstraram boa dispersão espacial e direcionalidade, capturando variações em múltiplas escalas e refletindo a complexidade dos sistemas agrícolas. Assim, a metodologia proposta consolida-se como uma ferramenta acessível para estruturar análises ambientômicas, contribuindo para a seleção mais eficiente e adaptada de genótipos em cenários edafoclimáticos diversos.

### 3.6 REFERÊNCIAS

BAHIA, M. A. M.; MARCATTI, G. E.; BRESEGHELLO, F.; MELO, P. G. S.; DIAS, K. O. G.; XU, Y.; RESENDE, R. T. Predictive Ability of Enviromic Modeling in G×E Interactions for Upland Rice Site Recommendations. **bioRxiv**, 2025.04.05.647387, p. 1-23, 2025.

BANČIČ, J.; GORJANC, G.; TOLHURST, D. J. A framework for simulating genotype-by-environment interaction using multiplicative models. **Theoretical and Applied Genetics**, v. 137, n. 8, 2024.

BERNARDO, R. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. **Heredity**, v. 125, n. 6, p. 375-385, 2020.

COOPER, M.; MESSINA, C. D.; PODLICH, D.; TOTIR, L. R.; BAUMGARTEN, A.; HAUSMANN, N. J.; WRIGHT, D.; GRAHAM, G. Predicting the future of plant breeding: Complementing empirical evaluation with genetic prediction. **Crop and Pasture Science**, v. 65, n. 4, p. 311-336, 2014.

CRESPO-HERRERA, L. A.; CROSSA, J.; HUERTA-ESPINO, J.; MONDAL, S.; VELU, G.; JULIANA, P.; VARGAS, M.; PÉREZ-RODRÍGUEZ, P.; JOSHI, A. K.; BRAUN, H. J.; SINGH, R. P. Target Population of Environments for Wheat Breeding in India: Definition, Prediction and Genetic Gains. **Frontiers in Plant Science**, v. 12, p. 1-15, 2021.

CROSSA, J.; FRITSCHÉ-NETO, R.; MONTESINOS-LOPEZ, O. A.; COSTA-NETO, G.; DREISIGACKER, S.; MONTESINOS-LOPEZ, A.; BENTLEY, A. R. The Modern Plant Breeding Triangle: Optimizing the Use of Genomics, Phenomics, and Enviromics Data. **Frontiers in Plant Science**, v. 12, n. April, p. 1-6, 2021.

JARQUÍN, D.; CROSSA, J.; LACAZE, X.; DU CHEYRON, P.; DAUCOURT, J.; LORGEOU, J.; PIRAUX, F.; GUERREIRO, L.; PÉREZ, P.; CALUS, M.; BURGUEÑO, J.; DE LOS CAMPOS, G. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. **Theoretical and Applied Genetics**, v. 127, n. 3, p. 595-607, 2014.

JARQUIN, D.; DE LEON, N.; ROMAY, C.; BOHN, M.; BUCKLER, E. S.; CIAMPITTI, I.; EDWARDS, J.; ERTL, D.; FLINT-GARCIA, S.; GORE, M. A.; GRAHAM, C.; HIRSCH, C. N.; HOLLAND, J. B.; HOOKER, D.; KAEPLER, S. M.; KNOLL, J.; LEE, E. C.; LAWRENCE-DILL, C. J.; LYNCH, J. P.; MOOSE, S. P.; MURRAY, S. C.; NELSON, R.; ROCHEFORD, T.; SCHNABLE, J. C.; SCHNABLE, P. S.; SMITH, M.; SPRINGER, N.; THOMISON, P.; TUINSTRA, M.; WISSER, R. J.; XU, W.; YU, J.; LORENZ, A. Utility of Climatic Information via Combining Ability Models to Improve Genomic Prediction for Yield Within the Genomes to Fields Maize Project. **Frontiers in Genetics**, v. 11, n. 592769, p. 1-11, 2021.

MESSINA, C. D.; COOPER, M. Plant Breeding in the face of climate change. **bioRxiv**, 2022-10, 2022.

NORTON, B. APIs: a common interface for the global biodiversity informatics community. **Biodiversity Information Science and Standards**, v. 5, e. 75267, 2021.

RESENDE, R. T.; HICKEY, L.; AMARAL, C. H.; PEIXOTO, L. L.; MARCATTI, G. E.; XU, Y. Satellite-enabled enviromics to enhance crop improvement. **Molecular Plant**, v. 17, n. 6, p. 848-866, 2024.

RESENDE, R. T.; PIEPHO, H. P.; ROSA, G. J. M.; SILVA-JUNIOR, O. B.; E SILVA, F. F.; DE RESENDE, M. D. V.; GRATTAPAGLIA, D. Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. **Theoretical and Applied Genetics**, v. 134, n. 1, p. 95-112, 2021.

RESENDE, R. T.; XAVIER, A.; SILVA, P. I. T.; RESENDE, M. P. M.; JARQUIN, D.; MARCATTI, G. E. GIS-based  $G \times E$  modeling of maize hybrids through enviromic markers engineering. **New Phytologist**, v. 245, n. 1, p. 102-116, 2024.

REZAEI, E. E.; WEBBER, H.; ASSENG, S.; BOOTE, K.; DURAND, J. L.; EWERT, F.; MARTRE, P.; MacCARTHY, D. S. Climate change impacts on crop yields. **Nature Reviews Earth & Environment**, v. 4, n. 12, 831-846, 2023.

SWAMINATHAN, R.; HUANG, Y.; MOOSAVINASAB, S.; BUCKLEY, R.; BARTLETT, C. W.; LIN, S. M. A Review on Genomics APIs. **Computational and Structural Biotechnology Journal**, v. 14, p. 8-15, 2016.

VAN EEUWIJK, F. A.; BUSTOS-KORTS, D. V.; MALOSETTI, M. What should students in plant breeding know about the statistical aspects of genotype  $\times$  Environment interactions? **Crop Science**, v. 56, n. 5, p. 2119–2140, 2016.

XU, Y. Envirotyping for deciphering environmental impacts on crop plants. **Theoretical and Applied Genetics**, v. 129, n. 4, p. 653-673, 2016.

XU, Y.; ZHANG, X.; LI, H.; ZHENG, H.; ZHANG, J.; OLSEN, M. S.; VARSHNEY, R. K.; PRASANNA, B. M.; QIAN, Q. Smart breeding driven by big data, artificial intelligence, and integrated genomic-enviromic prediction. **Molecular Plant**, v. 15, n. 11, p. 1664-1695, 2022.

## 4 CONSIDERAÇÕES FINAIS

A discussão conceitual desenvolvida evidenciou o papel estratégico da ambientômica como campo emergente na interseção entre sensoriamento remoto, estatística ambiental e melhoramento genético. Ao explorar as potencialidades das plataformas abertas de dados e o uso de covariáveis ambientais multiescalares, foi possível consolidar uma base teórica que fundamenta a construção de modelos preditivos mais robustos, sensíveis à variabilidade ecológica. A revisão realizada contribuiu para organizar conceitos dispersos na literatura, esclarecer limites operacionais de diferentes fontes de dados e apontar caminhos para a integração eficiente entre componentes ambientais, genéticos e de manejo. Também se destacou a importância de iniciativas que estimulem a interoperabilidade de dados e o uso de metodologias acessíveis, capazes de promover a familiaridade das ferramentas de ambientipagem nos programas de melhoramento vegetal. Assim, reforça-se a relevância da ambientômica não apenas como um suporte técnico, mas como um eixo estruturante na construção de estratégias resilientes frente às mudanças climáticas e à intensificação da heterogeneidade nos sistemas agrícolas.

Com base nessa fundamentação, foi possível desenvolver um *pipeline* automatizado para aquisição, padronização e organização de covariáveis ambientais aplicáveis à modelagem de G×E. O sistema integrou dados de diferentes plataformas públicas, gerando uma base com 3.507 camadas ambientais em resolução compatível com o uso em experimentos agrônômicos. As análises de componentes principais revelaram gradientes ambientais consistentes, capazes de orientar a segmentação de zonas de recomendação genotípica e a delimitação de populações-alvo de ambientes (TPEs). O *pipeline* possui robustez, escalabilidade e potencial de aplicação prática, superando barreiras operacionais normalmente enfrentadas por melhoristas ao trabalhar com dados ambientais brutos. Ao oferecer uma estrutura reprodutível e adaptável a diferentes contextos, o trabalho contribui para a consolidação da ambientipagem como ferramenta para aumentar a acurácia preditiva, otimizar recursos experimentais e acelerar ciclos seletivos em programas de melhoramento genético de plantas voltados à adaptação em larga escala.