

# **GIS-FA na Prática: Predição e Recomendação de Cultivares com Base em Dados Ambientais**

Thiago Brommonschenkel  
Gabriel Mazetti Blasques  
João Marcos Amario



# Conteúdo

01

## **Modelos Fator Analítico**

- Fundamentos teóricos dos modelos FA aplicados à interação G×E
- Vantagens dos modelos FA frente a outras estruturas de covariância
- Implementação prática em R

02

## **Ambientômica**

- Conceitos fundamentais de dados ambientais no melhoramento genético
- Fontes e tipos de dados ambientais
- Aplicação prática: coleta, limpeza e análise exploratória dos dados ambientais

03

## **GIS-FA para Recomendação de Cultivares**

- Conceito e formulação do GIS-FA
- Interpolação
- Similaridade ambiental
- Previsão geoespacial da performance genotípica

# Idealizadores do GIS-FA

Theoretical and Applied Genetics (2024) 137:80  
<https://doi.org/10.1007/s00122-024-04579-z>

ORIGINAL ARTICLE



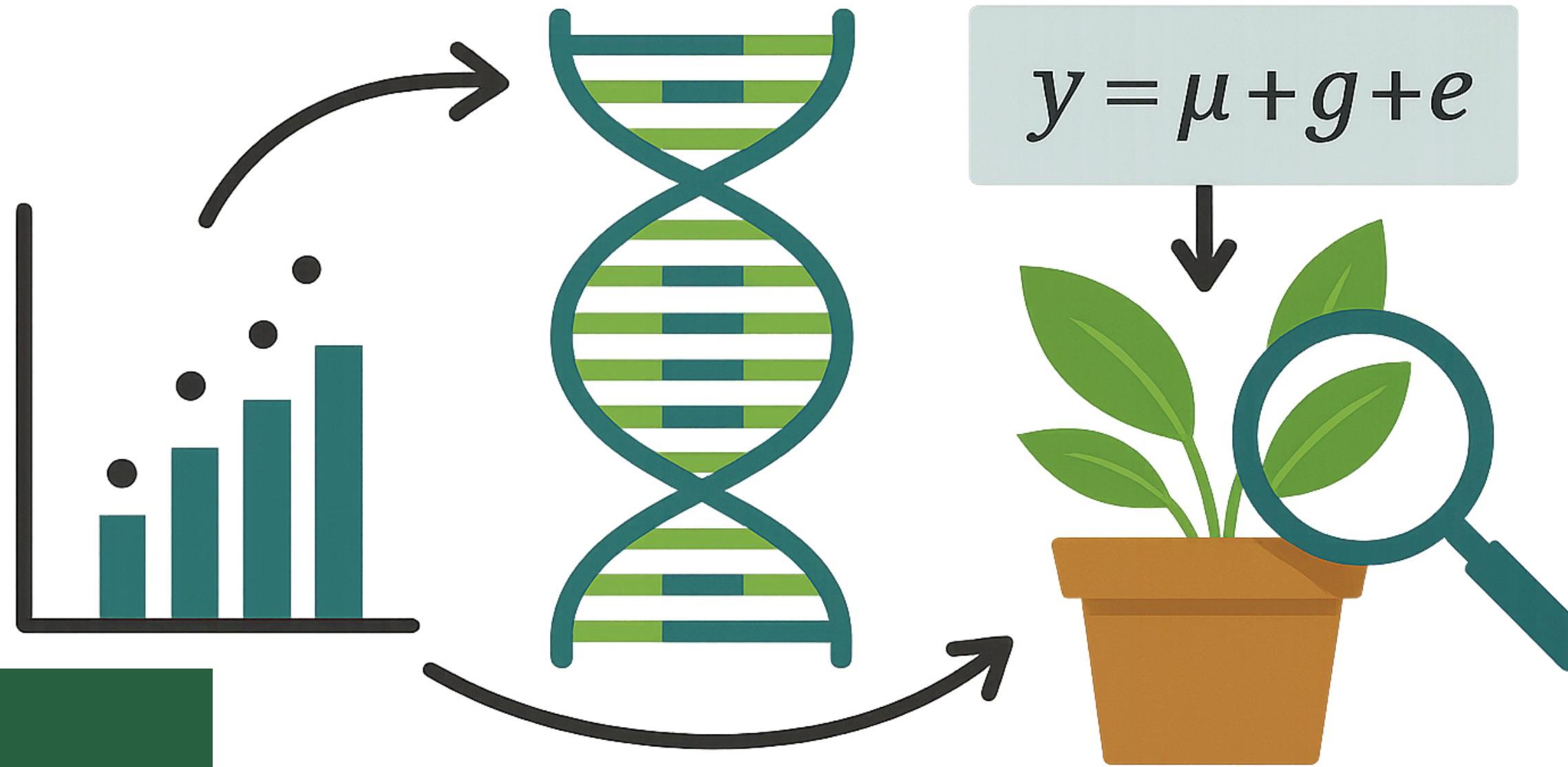
## GIS-FA: an approach to integrating thematic maps, factor-analytic, and envirotyping for cultivar targeting

Maurício S. Araújo<sup>1</sup> · Saulo F. S. Chaves<sup>1</sup> · Luiz A. S. Dias<sup>1</sup> · Filipe M. Ferreira<sup>2</sup> · Guilherme R. Pereira<sup>1</sup> · André R. G. Bezerra<sup>3</sup> · Rodrigo S. Alves<sup>4</sup> · Alexandre B. Heinemann<sup>6</sup> · Flávio Breseghello<sup>6</sup> · Pedro C. S. Carneiro<sup>4</sup> · Matheus D. Krause<sup>7</sup> · Germano Costa-Neto<sup>5</sup> · Kaio O. G. Dias<sup>4</sup>

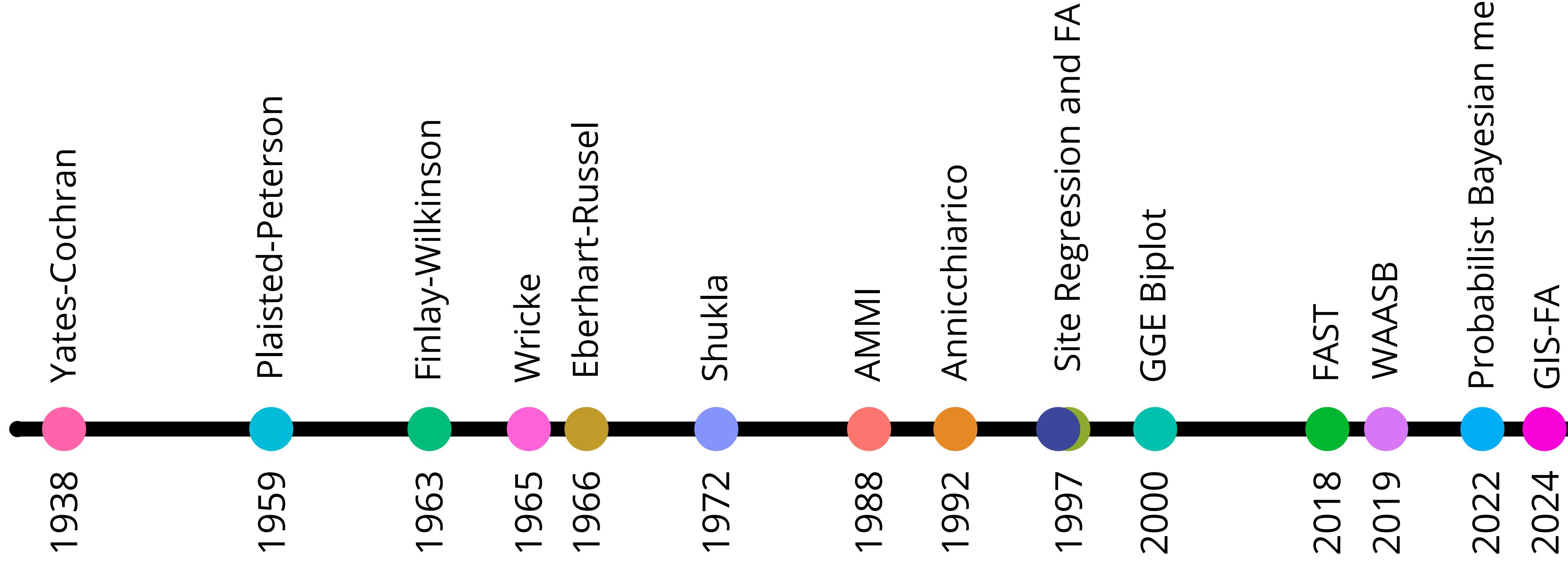
Received: 15 July 2023 / Accepted: 6 February 2024 / Published online: 12 March 2024  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Parte 1:

# Modelos Fator Analítico



# Métodos de Interação GxA

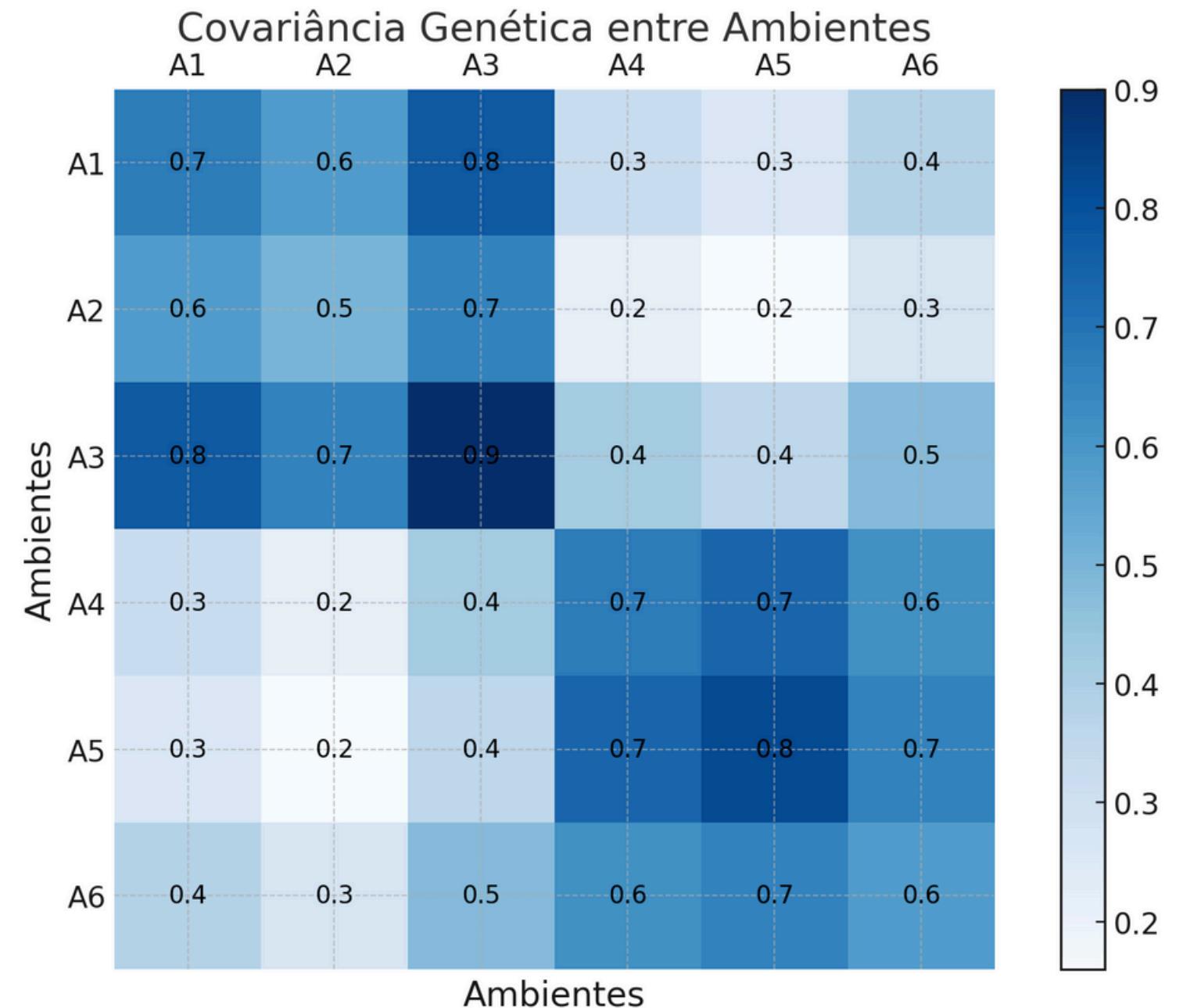


# **Modelos de Fator Análítico**

- Proposto por Piepho (1997, 1998) e Smith et al. (2001)

# Modelos de Fator Análítico

- Proposto por Piepho (1997, 1998) e Smith et al. (2001)
- Vantagens:
  - Captar os efeitos genéticos comuns ao longo dos ambientes



# Modelos de Fator Análítico

- Proposto por Piepho (1997, 1998) e Smith et al. (2001)
- Vantagens:
  - Captar os efeitos genéticos comuns ao longo dos ambientes
  - Aproximar-se do modelo multivariado mantendo a parsimônia

# Modelos de Fator Análítico

- Proposto por Piepho (1997, 1998) e Smith et al. (2001)
- Vantagens:
  - Captar os efeitos genéticos comuns ao longo dos ambientes
  - Aproximar-se do modelo multivariado mantendo a parsimônia
    - Exemplo para 49 ambientes:

# Modelos de Fator Análítico

- Proposto por Piepho (1997, 1998) e Smith et al. (2001)
- Vantagens:
  - Captar os efeitos genéticos comuns ao longo dos ambientes
  - Aproximar-se do modelo multivariado mantendo a parsimônia
    - Exemplo para 49 ambientes:

$$\frac{J(J + 1)}{2} = \frac{49(49 + 1)}{2} = \frac{49 \times 50}{2} = \frac{2450}{2} = 1225 \text{ (Modelo Multivariado)}$$

# Modelos de Fator Análítico

- Proposto por Piepho (1997, 1998) e Smith et al. (2001)
- Vantagens:
  - Captar os efeitos genéticos comuns ao longo dos ambientes
  - Aproximar-se do modelo multivariado mantendo a parsimônia
    - Exemplo para 49 ambientes:

$$\frac{J(J + 1)}{2} = \frac{49(49 + 1)}{2} = \frac{49 \times 50}{2} = \frac{2450}{2} = 1225 \text{ (Modelo Multivariado)}$$

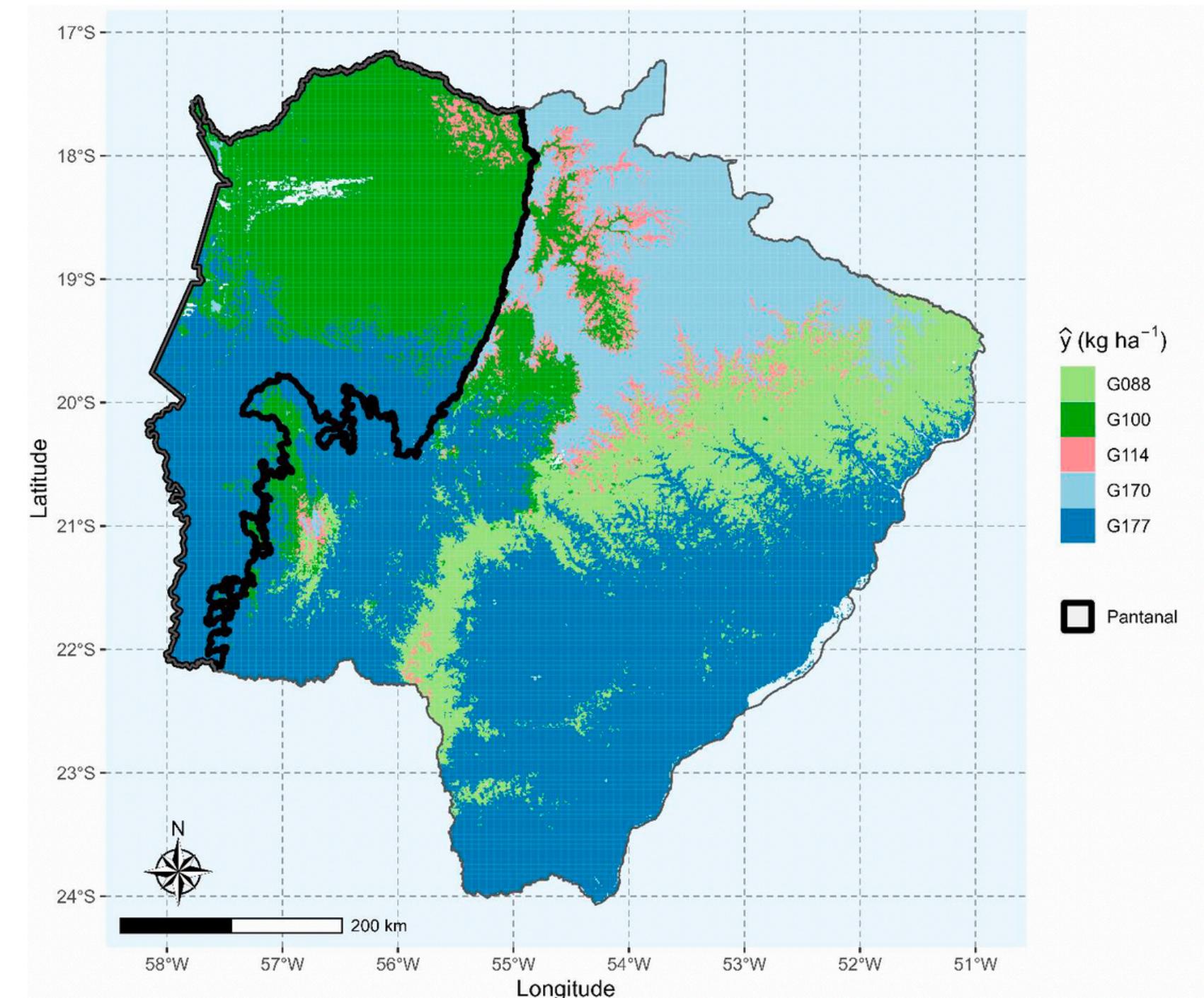
$$(J \times k) + J = (49 \times 4) + 49 = 196 + 49 = 245 \text{ (Modelo FA)}$$

# Modelos de Fator Análítico

- Proposto por Piepho (1997, 1998) e Smith et al. (2001)
- Vantagens:
  - Captar os efeitos genéticos comuns ao longo dos ambientes
  - Aproximar-se do modelo multivariado mantendo a parsimônia
  - Heterogeneidade de covariâncias genéticas

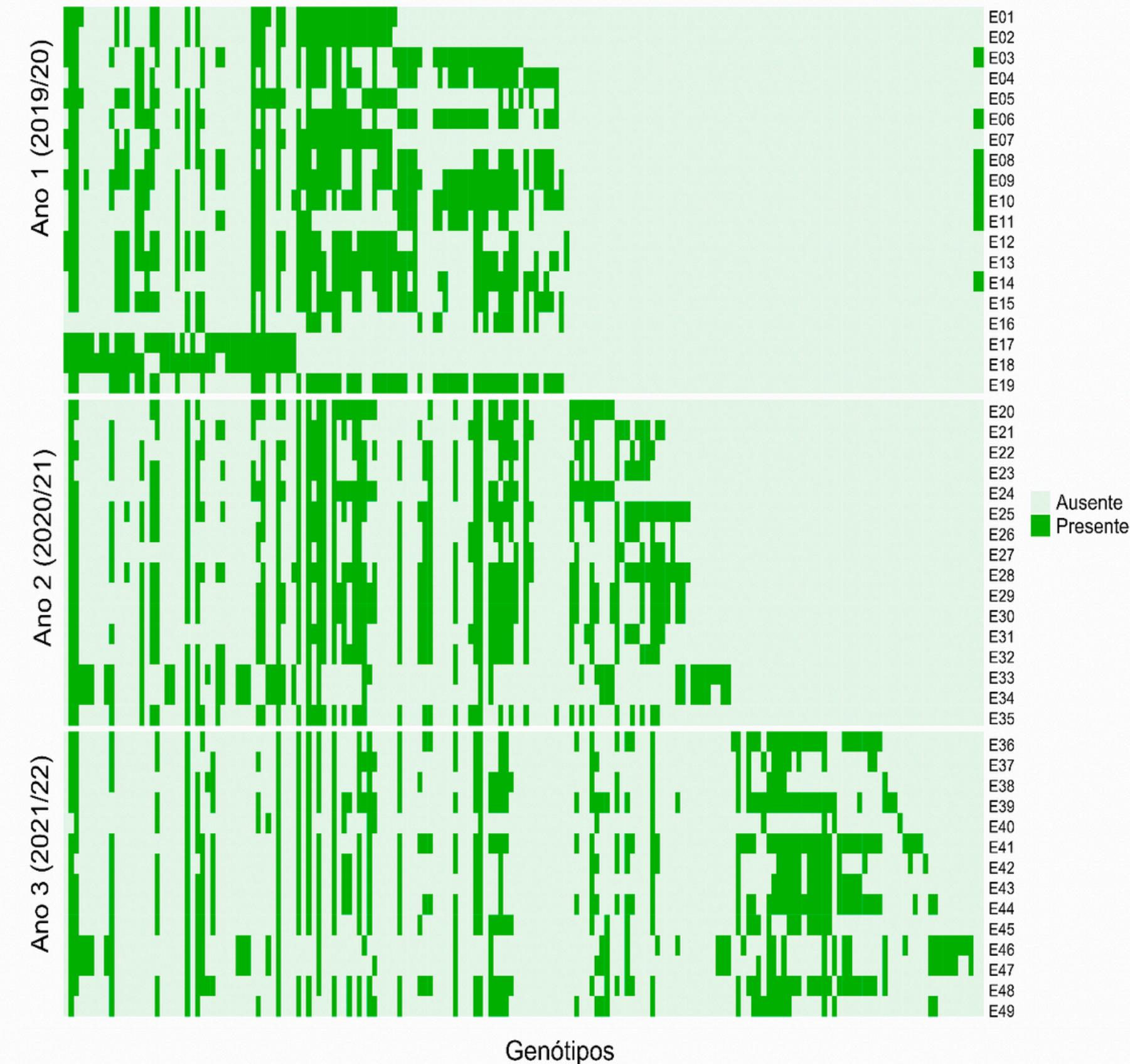
# Modelos de Fator Análítico

- Proposto por Piepho (1997, 1998) e Smith et al. (2001)
- Vantagens:
  - Captar os efeitos genéticos comuns ao longo dos ambientes
  - Aproximar-se do modelo multivariado mantendo a parsimônia
  - Heterogeneidade de covariâncias genéticas
  - Efeito genético (BLUP)



# Modelos de Fator Análítico

- Proposto por Piepho (1997, 1998) e Smith et al. (2001)
  - Vantagens:
    - Captar os efeitos genéticos comuns ao longo dos ambientes
    - Aproximar-se do modelo multivariado mantendo a parsimônia
    - Heterogeneidade de covariâncias genéticas
    - Efeito genético (BLUP)
    - Desbalanceamento e dependência residual



# Modelos de Fator Análítico

- Qual é a ideia central dos modelos FA?

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}_1\mathbf{a} + \mathbf{X}_2\mathbf{r} + \mathbf{Z}_1\mathbf{g} + \boldsymbol{\varepsilon}$$

# Modelos de Fator Análítico

- Qual é a ideia central dos modelos FA?

$$\begin{aligned} \mathbf{y} &= \mathbf{1}\mu + \mathbf{X}_1\mathbf{a} + \mathbf{X}_2\mathbf{r} + \mathbf{Z}_1\mathbf{g} + \boldsymbol{\varepsilon} \\ &\quad \downarrow \\ \mathbf{g} &= (\boldsymbol{\Lambda} \otimes \mathbf{I}_v)\mathbf{f} + \boldsymbol{\delta} \end{aligned}$$

# Modelos de Fator Análítico

- Qual é a ideia central dos modelos FA?

$$\begin{aligned} \mathbf{y} &= \mathbf{1}\mu + \mathbf{X}_1\mathbf{a} + \mathbf{X}_2\mathbf{r} + \mathbf{Z}_1\mathbf{g} + \boldsymbol{\varepsilon} \\ &\quad \downarrow \\ \mathbf{g} &= (\boldsymbol{\Lambda} \otimes \mathbf{I}_v)\mathbf{f} + \boldsymbol{\delta} \end{aligned}$$

- O que seria cada componente presente na decomposição do efeito genético?

# Modelos de Fator Análítico

$$\mathbf{g} = (\Lambda \otimes \mathbf{I}_v) \mathbf{f} + \boldsymbol{\delta}$$

$\Lambda$ : matriz de cargas fatorais

# Modelos de Fator Análítico

$$\mathbf{g} = (\Lambda \otimes \mathbf{I}_v)\mathbf{f} + \boldsymbol{\delta}$$

$\Lambda$ : matriz de cargas fatorais (o quanto cada ambiente se “associa” a um fator latente)

	FA1	FA2	FA3	FA4
E01	418.66	0.00	0.18	0.10
E02	142.05	266.11	0.18	0.10
E03	294.38	-88.59	-23.77	0.10
E04	104.27	-2.16	150.01	-39.46
E05	296.17	-79.58	-158.29	41.37
E45	138.29	48.72	141.10	32.56
E46	-14.30	-138.26	22.86	-504.65
E47	-40.05	-313.84	177.47	-335.38
E48	237.16	28.69	-81.99	176.76
E49	157.94	60.75	26.68	79.17

# Modelos de Fator Análítico

$$\mathbf{g} = (\Lambda \otimes \mathbf{I}_v)\mathbf{f} + \boldsymbol{\delta}$$

**f** : score dos genótipos (como cada genótipo responde aos fatores)

	FA1	FA2	FA3	FA4
<b>C006</b>	0.32	-0.47	0.88	-1.40
<b>C049</b>	0.94	0.14	-0.17	-0.83
<b>C054</b>	1.29	0.79	0.69	0.00
<b>C084</b>	0.65	-0.29	-0.18	1.10
<b>G001</b>	-0.58	0.78	0.53	-0.06
<b>G191</b>	0.02	0.23	-0.05	0.83
<b>G192</b>	0.00	0.05	-0.01	0.17
<b>G193</b>	-0.10	-0.76	0.49	-0.48
<b>G194</b>	0.66	-1.61	-1.03	-0.55
<b>G195</b>	0.52	-0.83	-0.55	0.10

# Modelos de Fator Análítico

$$\mathbf{g} = (\boldsymbol{\Lambda} \otimes \mathbf{I}_v) \mathbf{f} + \boldsymbol{\delta}$$

$\boldsymbol{\Lambda}$ : matriz de cargas fatorais (o quanto cada ambiente se “associa” a um fator latente)

$\mathbf{f}$  : matriz dos escores fatoriais (como cada genótipo responde aos fatores)

$\boldsymbol{\delta}$  : efeito específico de cada ambiente

# Modelos de Fator Análítico

- Matriz de co(variância) genética

$$\mathbf{G} = (\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}) \otimes \mathbf{P}$$

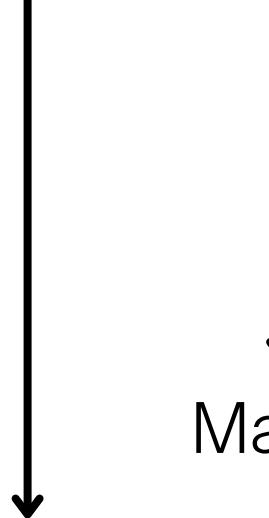


Matriz de covariância explicada pelos fatores comuns entre os ambientes.

# Modelos de Fator Análítico

- Matriz de co(variância) genética

$$\mathbf{G} = (\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}) \otimes \mathbf{P}$$



Matriz de variância específica dos ambientes

Matriz de covariância explicada pelos fatores comuns entre os ambientes.

# Modelos de Fator Análítico

- Matriz de co(variância) genética

$$\mathbf{G} = (\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}) \otimes \mathbf{P}$$

Matriz de Similaridade genética entre os genótipos

Matriz de variância específica dos ambientes

Matriz de covariância explicada pelos fatores comuns entre os ambientes.

# Modelos de Fator Análítico

- Matriz de co(variância) genética

$$\mathbf{G} = (\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}) \otimes \mathbf{P}$$

- Qual o resultado?

$$\mathbf{G} = \begin{bmatrix} \sum_{k=1}^K \lambda_{1k}^2 + \psi_1 & \sum_{k=1}^K \lambda_{1k}\lambda_{2k} & \cdots & \sum_{k=1}^K \lambda_{1k}\lambda_{mk} \\ \sum_{k=1}^K \lambda_{2k}\lambda_{1k} & \sum_{k=1}^K \lambda_{2k}^2 + \psi_2 & \cdots & \sum_{k=1}^K \lambda_{2k}\lambda_{mk} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^K \lambda_{mk}\lambda_{1k} & \sum_{k=1}^K \lambda_{mk}\lambda_{2k} & \cdots & \sum_{k=1}^K \lambda_{mk}^2 + \psi_m \end{bmatrix} \otimes \mathbf{P}$$

# Rotação das cargas e dos escores

- Fazer com que os fatores sejam ortogonais
- Método utilizado: SVD proposto por Cullis et al., 2010

$$\Lambda = \mathbf{UDV}^\top$$

$\mathbf{U}$ : matriz de autovetores de  $\Lambda\Lambda^\top$

$\mathbf{D}$ : matriz diagonal com a raiz quadrada dos autovalores de  $\Lambda\Lambda^\top$  (ordem decrescente)

$\mathbf{V}$ : matriz de autovetores de  $\Lambda^\top\Lambda$

$$\Lambda^* = \mathbf{c}\Lambda\mathbf{V}$$

$$\mathbf{f}^* = (\mathbf{c}\mathbf{V}^\top \otimes \mathbf{I}_v)\mathbf{f}$$

# Rotação das cargas e dos escores

$\Lambda^*$  =

	FA1	FA2	FA3	FA4
E01	0.19	0.08	-0.20	0.07
E02	0.07	-0.20	-0.20	0.00
E03	0.14	0.13	-0.09	0.02
E04	0.02	0.04	-0.13	0.22
E05	0.16	0.10	-0.02	-0.15
E45	0.05	-0.02	-0.10	0.24
E46	-0.12	0.29	-0.39	-0.20
E47	-0.12	0.39	-0.18	0.12
E48	0.16	-0.04	0.06	0.00
E49	0.09	-0.05	-0.04	0.10

$f^*$  =

	FA1	FA2	FA3	FA4
C006	-962.44	1,111.87	-1,038.42	287.96
C049	1,018.84	390.53	-884.67	-212.10
C054	1,894.12	-444.46	-867.87	477.10
C084	2,024.63	-25.41	631.54	169.37
G001	-1,239.74	-823.04	-226.07	234.63
G191	716.10	-544.03	463.76	130.52
G192	145.93	-110.86	94.51	26.60
G193	-797.71	922.15	-107.90	222.13
G194	1,080.95	1,858.49	198.17	-625.84
G195	1,169.59	845.95	296.53	-241.80

# Escolha do Modelo FA

- Como escolher o número de fatores “ideal” para o nosso Modelo?

# Escolha do Modelo FA

- Como escolher o número de fatores “ideal” para o nosso Modelo?
- Critérios de Informação

$$AIC = -2 \log L + 2p$$

# Escolha do Modelo FA

- Como escolher o número de fatores “ideal” para o nosso Modelo?
- Critérios de Informação

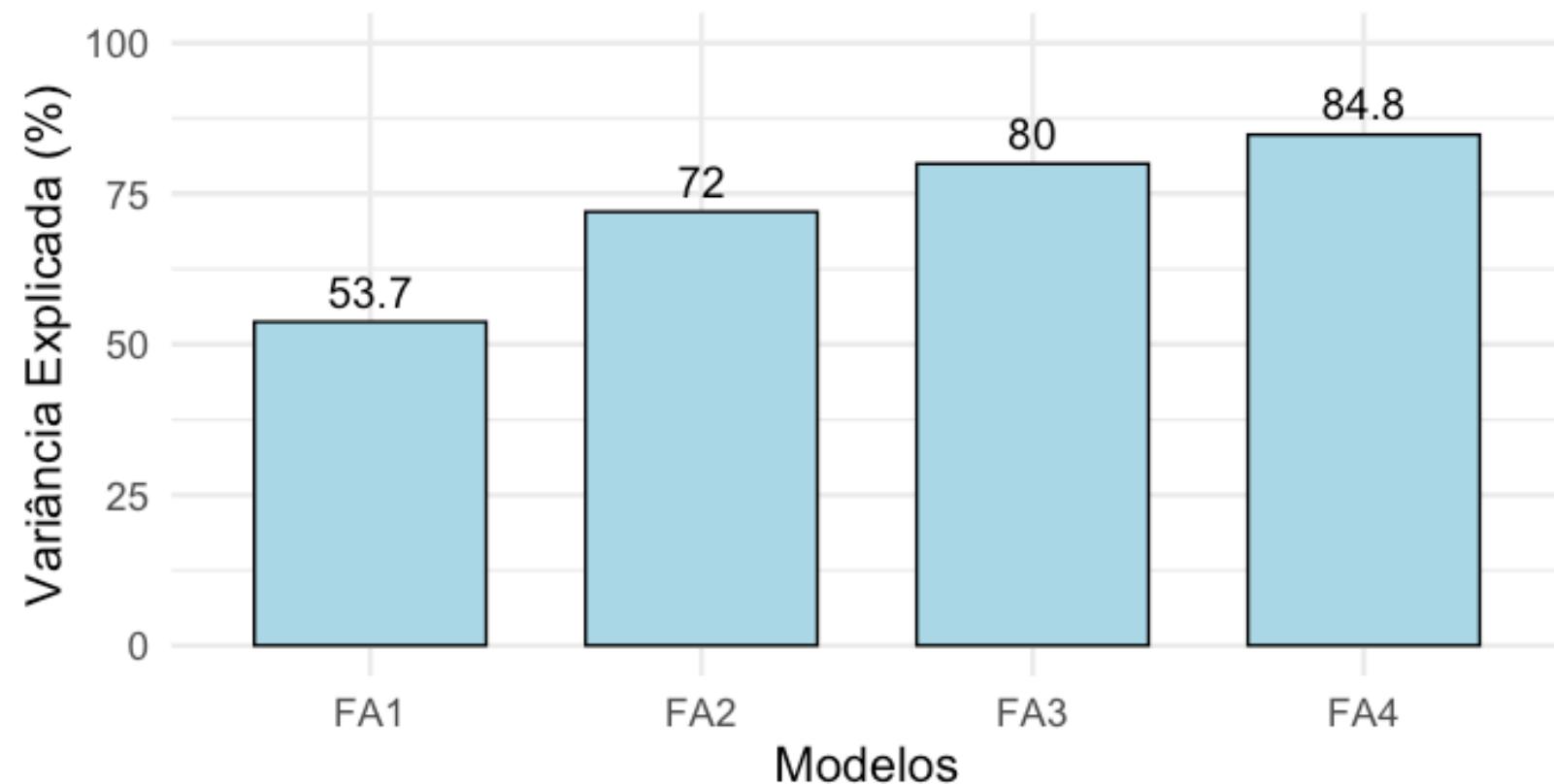
$$AIC = -2 \log L + 2p$$

- \* Podem não selecionar o melhor modelo em termos de interpretação biológica ou desempenho preditivo

# Escolha do Modelo FA

- Como escolher o número de fatores “ideal” para o nosso Modelo?
  - Variância explicada + Parsimônia

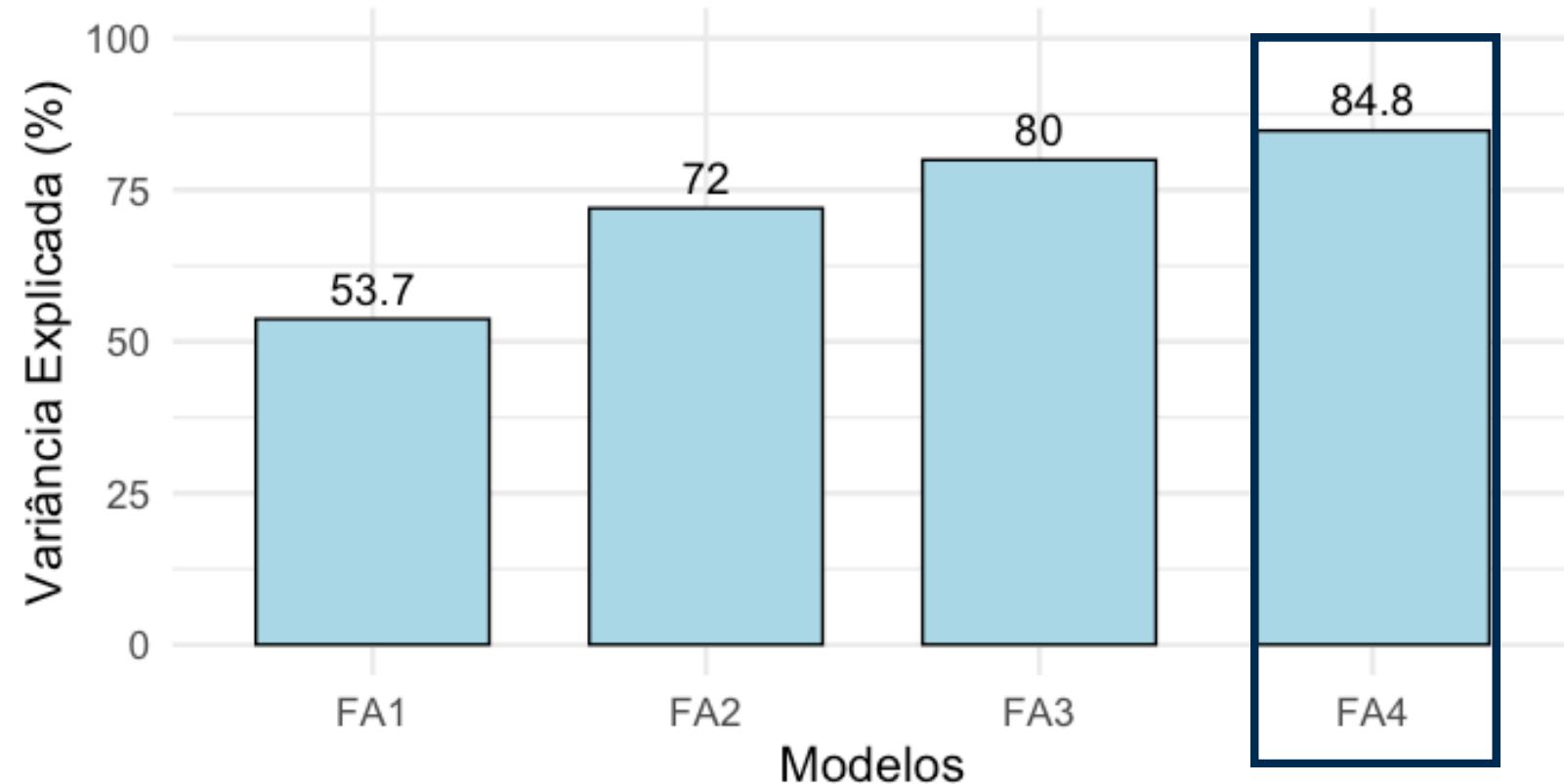
$$\bar{v} = \frac{\text{tr}(\Lambda^* \Lambda^{*\top})}{\text{tr}(\mathbf{G})}$$



# Escolha do Modelo FA

- Como escolher o número de fatores “ideal” para o nosso Modelo?
  - Variância explicada + Parsimônia

$$\bar{v} = \frac{\text{tr}(\Lambda^* \Lambda^{*\top})}{\text{tr}(\mathbf{G})}$$



# Variância explicada por cada Fator

$$v_{Kj} = \frac{\sum_{k=1}^K \lambda_{kj}^2}{\sum_{k=1}^K \lambda_{kj}^2 + \psi_j}$$

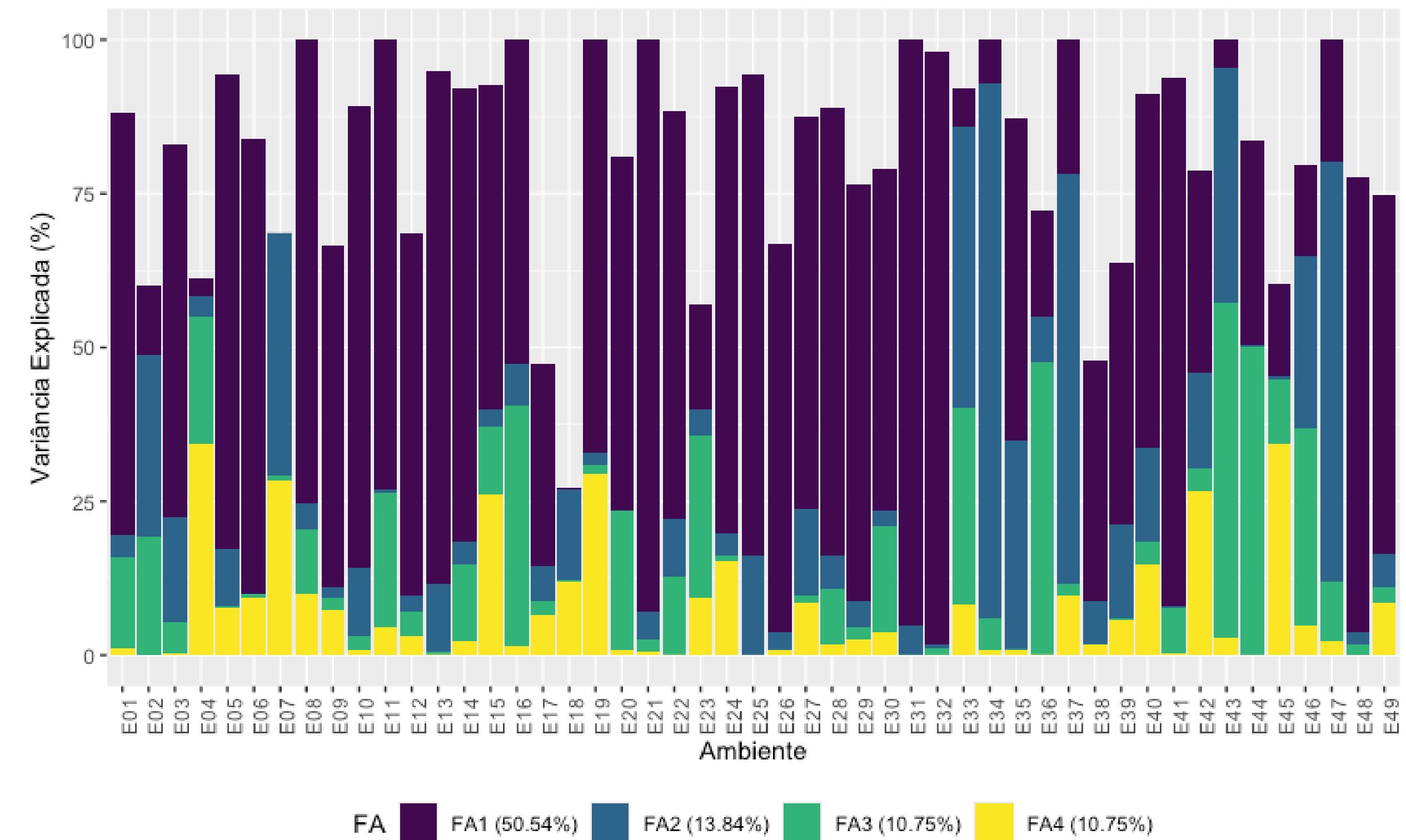
# Variância explicada por cada Fator

$$v_{Kj} = \frac{\sum_{k=1}^K \lambda_{kj}^2}{\sum_{k=1}^K \lambda_{kj}^2 + \psi_j}$$

	FA1	FA2	FA3	FA4
E01	68.71	3.69	14.76	1.05
E02	11.35	29.53	19.20	0.00
E03	60.70	16.96	5.21	0.20
E04	2.87	3.42	20.57	34.39
E05	76.99	9.56	0.24	7.58
E06	74.06	0.01	0.46	9.37
E07	0.04	39.31	0.73	28.41
E08	75.31	4.18	10.67	9.84
E09	55.67	1.66	1.95	7.37
E10	75.15	10.96	2.42	0.79

# Variância explicada por cada Fator

	FA1	FA2	FA3	FA4
E01	68.71	3.69	14.76	1.05
E02	11.35	29.53	19.20	0.00
E03	60.70	16.96	5.21	0.20
E04	2.87	3.42	20.57	34.39
E05	76.99	9.56	0.24	7.58
E06	74.06	0.01	0.46	9.37
E07	0.04	39.31	0.73	28.41
E08	75.31	4.18	10.67	9.84
E09	55.67	1.66	1.95	7.37
E10	75.15	10.96	2.42	0.79



# Correlação entre os Ambientes

$$\rho_{g_{jj'}} = \frac{\sum_{k=1}^K \lambda_{kj}^* \lambda_{kj'}^*}{\sqrt{\sigma_{g_j}^2 \sigma_{g_{j'}}^2}} = DGD$$

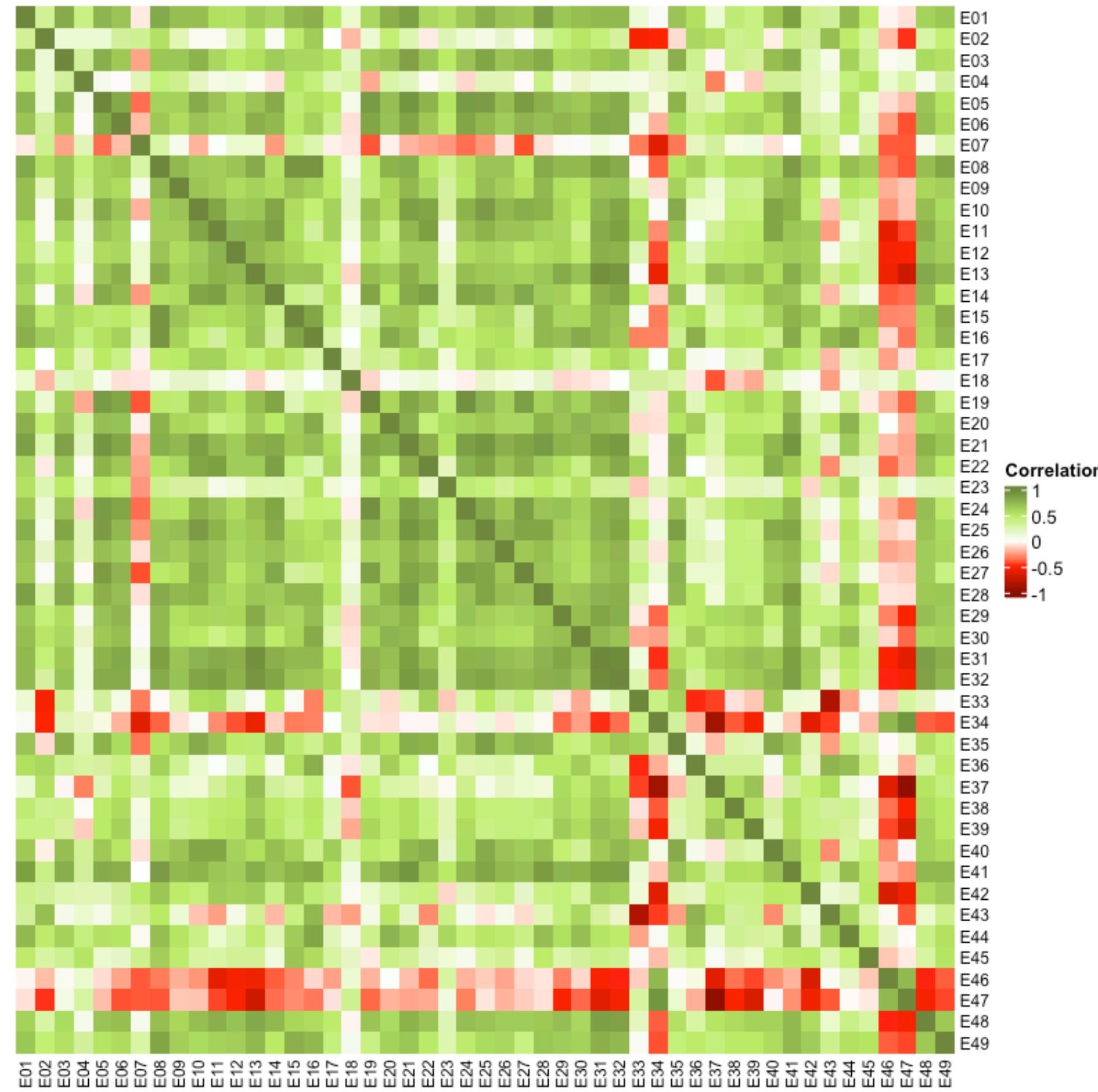
D : matriz diagonal cujos elementos são  $\frac{1}{\sqrt{\sigma_{g_j}^2}}$

G : matriz de (co)variâncias genéticas entre os ambientes

# Correlação entre os Ambientes

	E01	E02	E03	E04	E05	E06	E07	E08	E09	E10
E01	1.00	0.34	0.82	0.41	0.78	0.71	-0.05	0.84	0.72	0.73
E02	0.34	1.00	0.14	0.15	0.15	0.33	0.37	0.55	0.24	0.04
E03	0.82	0.14	1.00	0.34	0.81	0.67	-0.23	0.68	0.68	0.78
E04	0.41	0.15	0.34	1.00	0.07	0.00	0.23	0.44	0.37	0.19
E05	0.78	0.15	0.81	0.07	1.00	0.84	-0.35	0.63	0.63	0.83
E06	0.71	0.33	0.67	0.00	0.84	1.00	-0.17	0.67	0.57	0.71
E07	-0.05	0.37	-0.23	0.23	-0.35	-0.17	1.00	0.31	0.06	-0.19
E08	0.84	0.55	0.68	0.44	0.63	0.67	0.31	1.00	0.75	0.66
E09	0.72	0.24	0.68	0.37	0.63	0.57	0.06	0.75	1.00	0.69
E10	0.73	0.04	0.78	0.19	0.83	0.71	-0.19	0.66	0.69	1.00

# Correlação entre os Ambientes



# Ferramentas de Seleção (FAST)

- Desempenho geral:
  - Medida de performance média dos híbridos

$$OP_i = \frac{1}{m} \sum_{j=1}^m \lambda_{1j}^* f_{1i}^*$$

# Ferramentas de Seleção (FAST)

- Desempenho geral:

- Medida de performance média dos híbridos

$$OP_i = \frac{1}{m} \sum_{j=1}^m \lambda_{1j}^* f_{1i}^*$$

- Desvio quadrático médio:

- Medida de estabilidade agronômica dos híbridos

$$RMSD_i = \sqrt{\frac{1}{m} \sum_{j=1}^m \varepsilon_{ij}^2}$$

# Ferramentas de Seleção (FAST)

- Desempenho geral:

- Medida de performance média dos híbridos

$$OP_i = \frac{1}{m} \sum_{j=1}^m \lambda_{1j}^* f_{1i}^*$$

- Desvio quadrático médio:

- Medida de estabilidade agronômica dos híbridos

$$RMSD_i = \sqrt{\frac{1}{m} \sum_{j=1}^m \varepsilon_{ij}^2}$$

- Confiabilidade:

$$r_i = 1 - \frac{PEV_i}{\sigma_g^2}$$

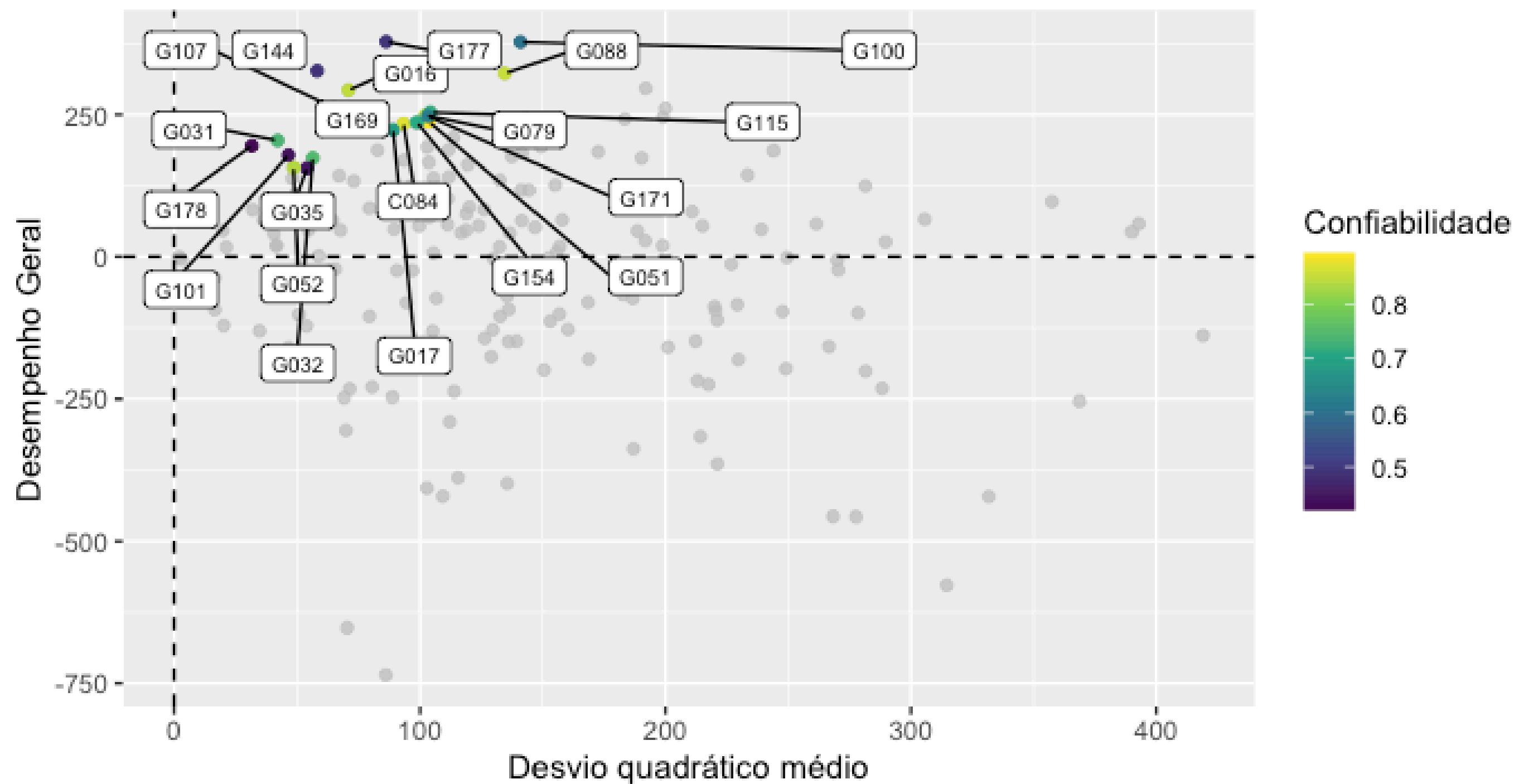
# Ferramentas de Seleção (FAST)

- Como utilizar as ferramentas de seleção?

$$IS = 2 \times \left( \frac{OP - \overline{OP}}{\sqrt{\text{Var}(OP)}} \right) - \left( \frac{RMSD - \overline{RMSD}}{\sqrt{\text{Var}(RMSD)}} \right)$$

# Ferramentas de Seleção (FAST)

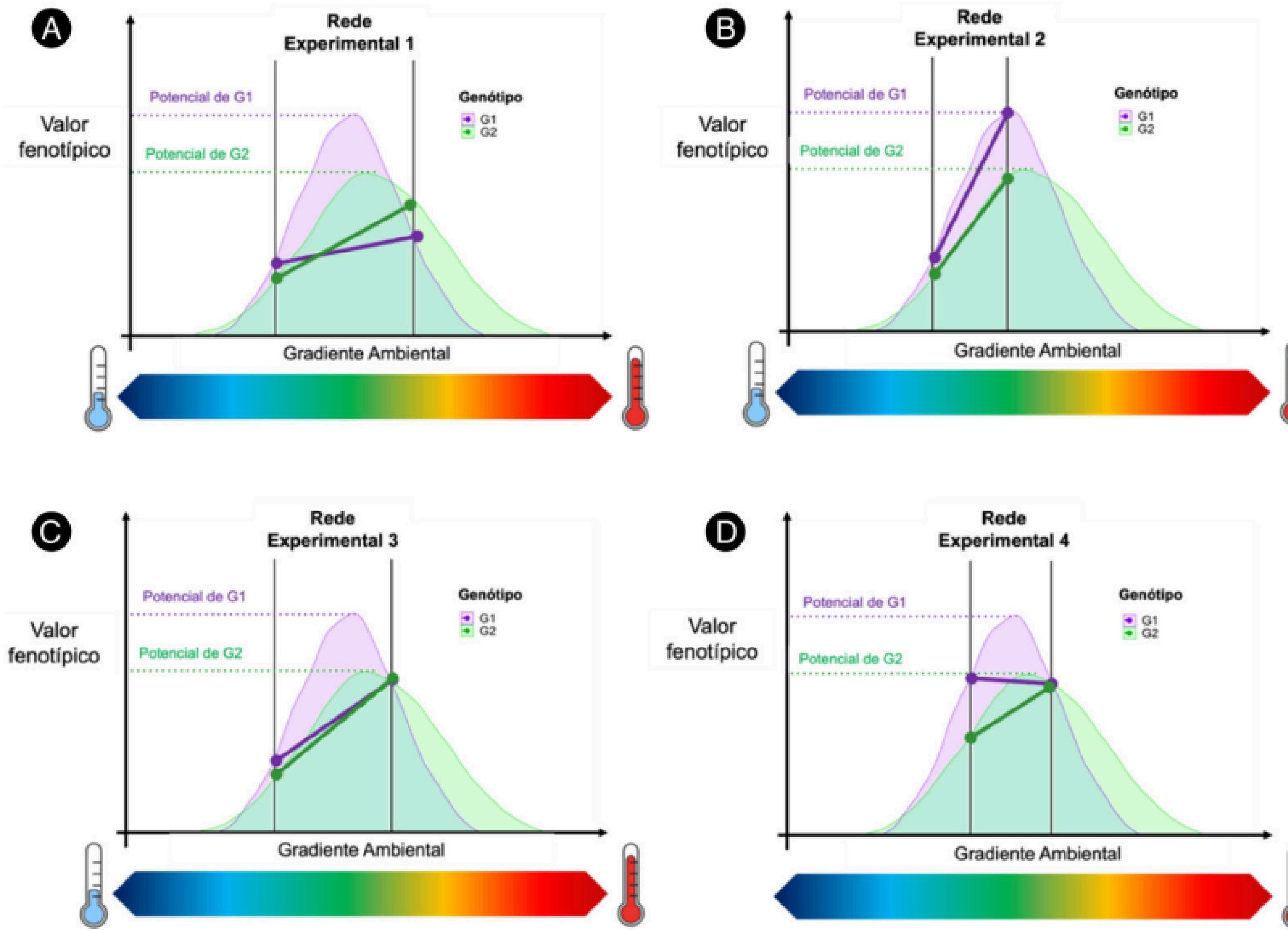
- Como utilizar as ferramentas de seleção?



The background of the slide is a vibrant aerial photograph of a rural landscape. It features numerous rectangular agricultural fields of various sizes and colors, ranging from deep green to golden yellow, indicating different crops or stages of growth. The fields are separated by a network of dark brown dirt roads and small clusters of trees. In the lower-left foreground, there is a solid dark green rectangular area containing the title text.

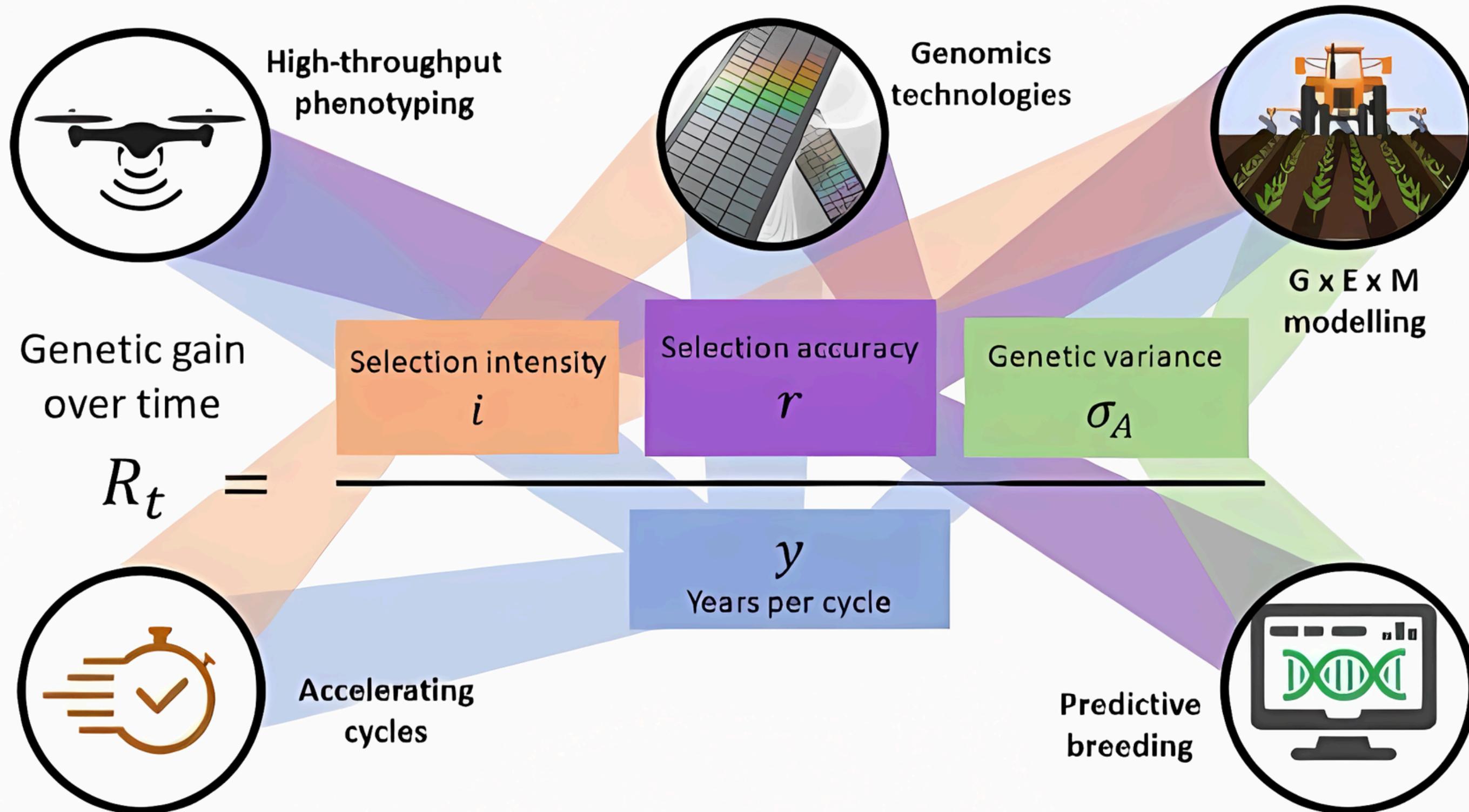
Parte 2:

# Ambientômica



**Figura 2.** A inter-relação entre plasticidade fenotípica (curvas) e norma de reação (retas entre pontos) em função da diversidade ambiental de cada rede experimental e seu restrito gradiente ambiental (linhas verticais pretas).

Fonte: Adaptado de Costa-Neto et al. (2021b).





Where ?

# Easy access

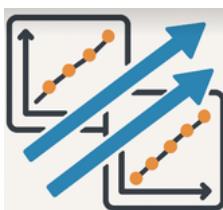
- APIs públicas e portais web intuitivos (NASA POWER, Earthdata, SoilGrids, WorldClim);
- Pacotes R (por ex. EnvRtype, raster, terra);
- Downloads em formatos padrão (CSV, GeoTIFF, NetCDF) e resolução customizável;
- Documentação e tutoriais disponíveis para cada fonte, com exemplos de código.



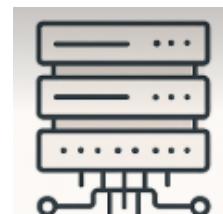
# Desafios



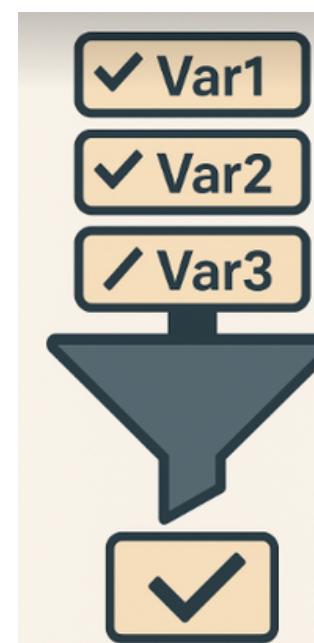
Preenchimento de Dados Faltantes



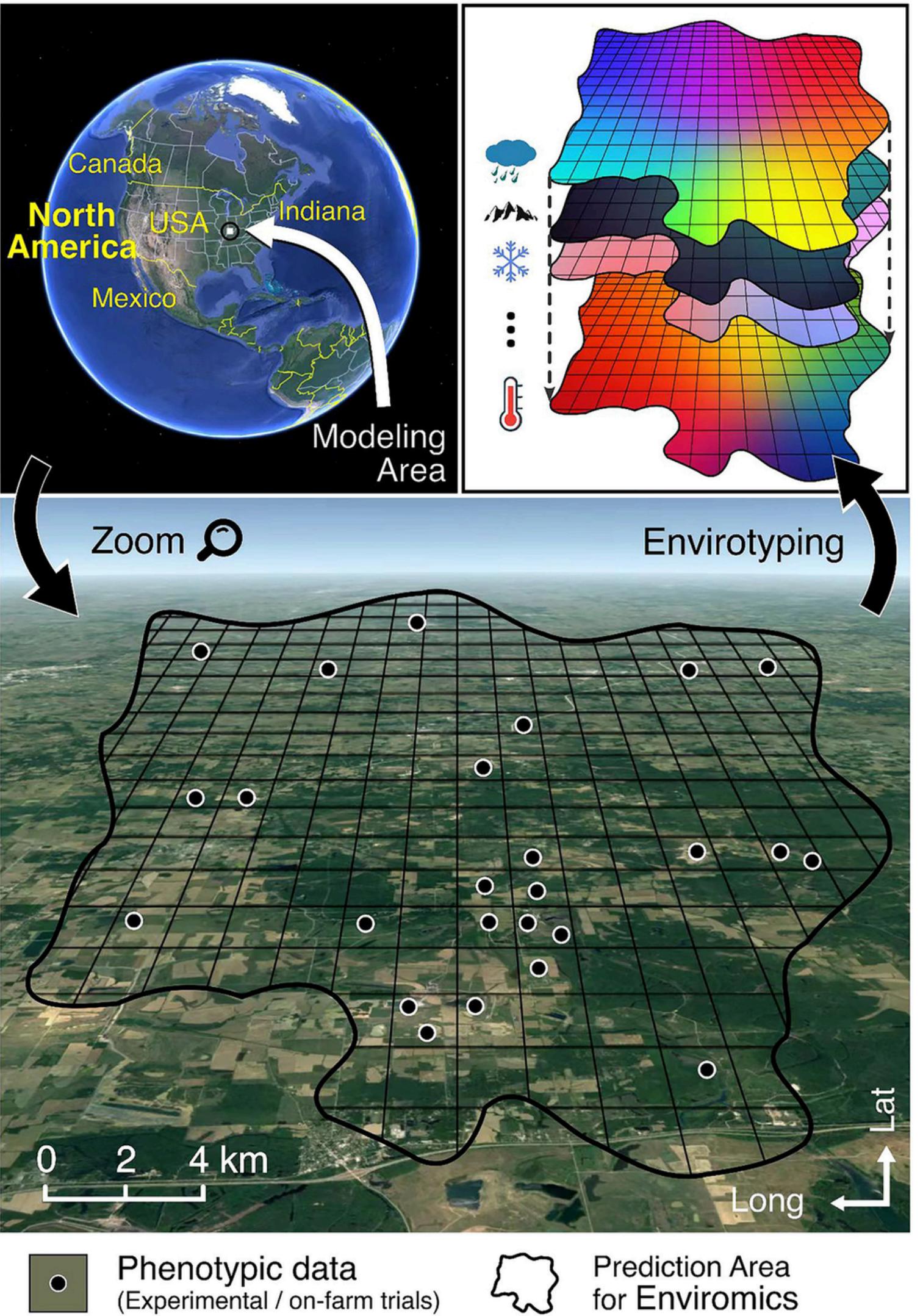
Multicolinearidade

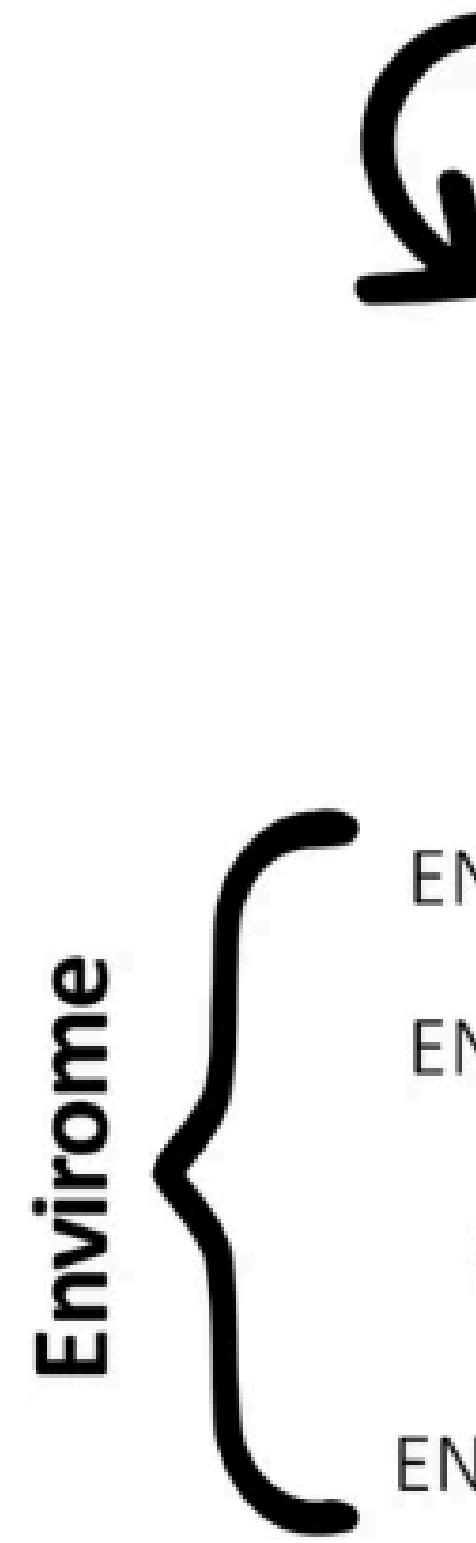


Complexidade de processamento de big data



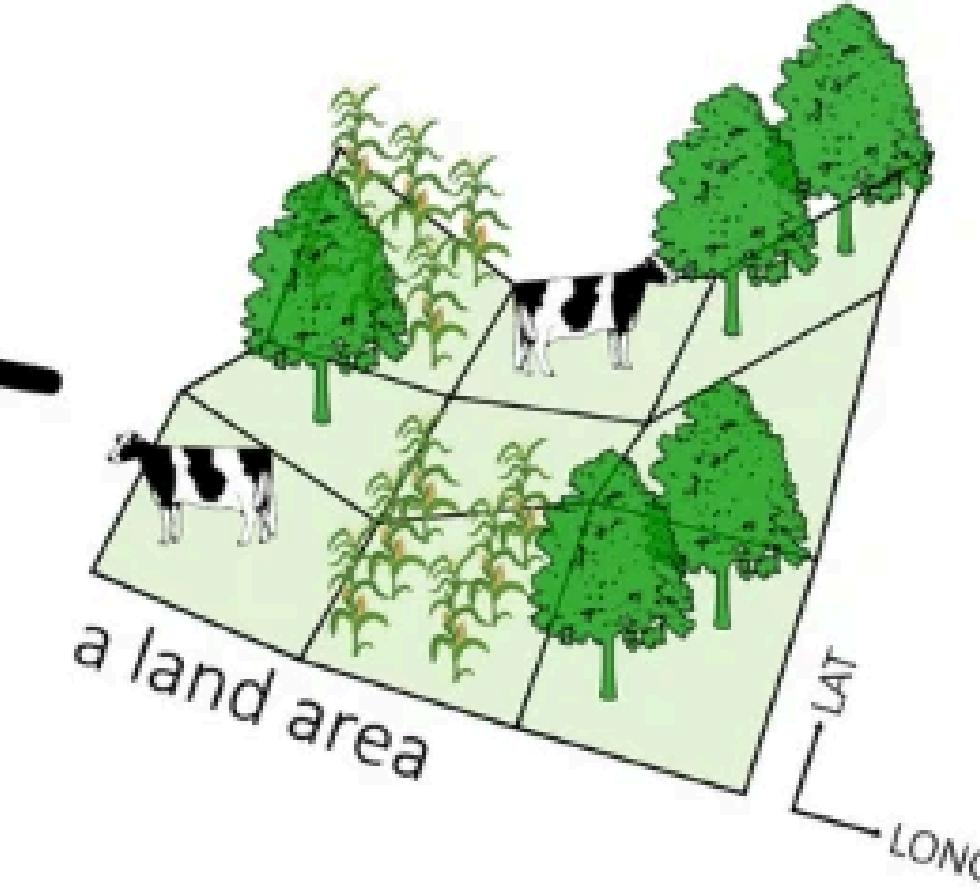
Seleção de Variáveis



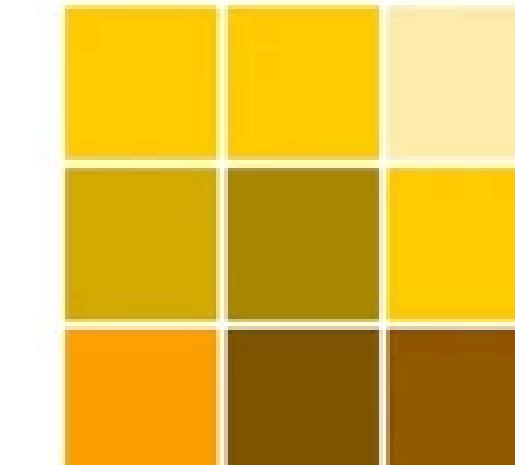


**Envirotyping  
techniques**

overlapping  
**Enviromic markers**



**Envirotypic  
polymorphism  
assisted breeding**

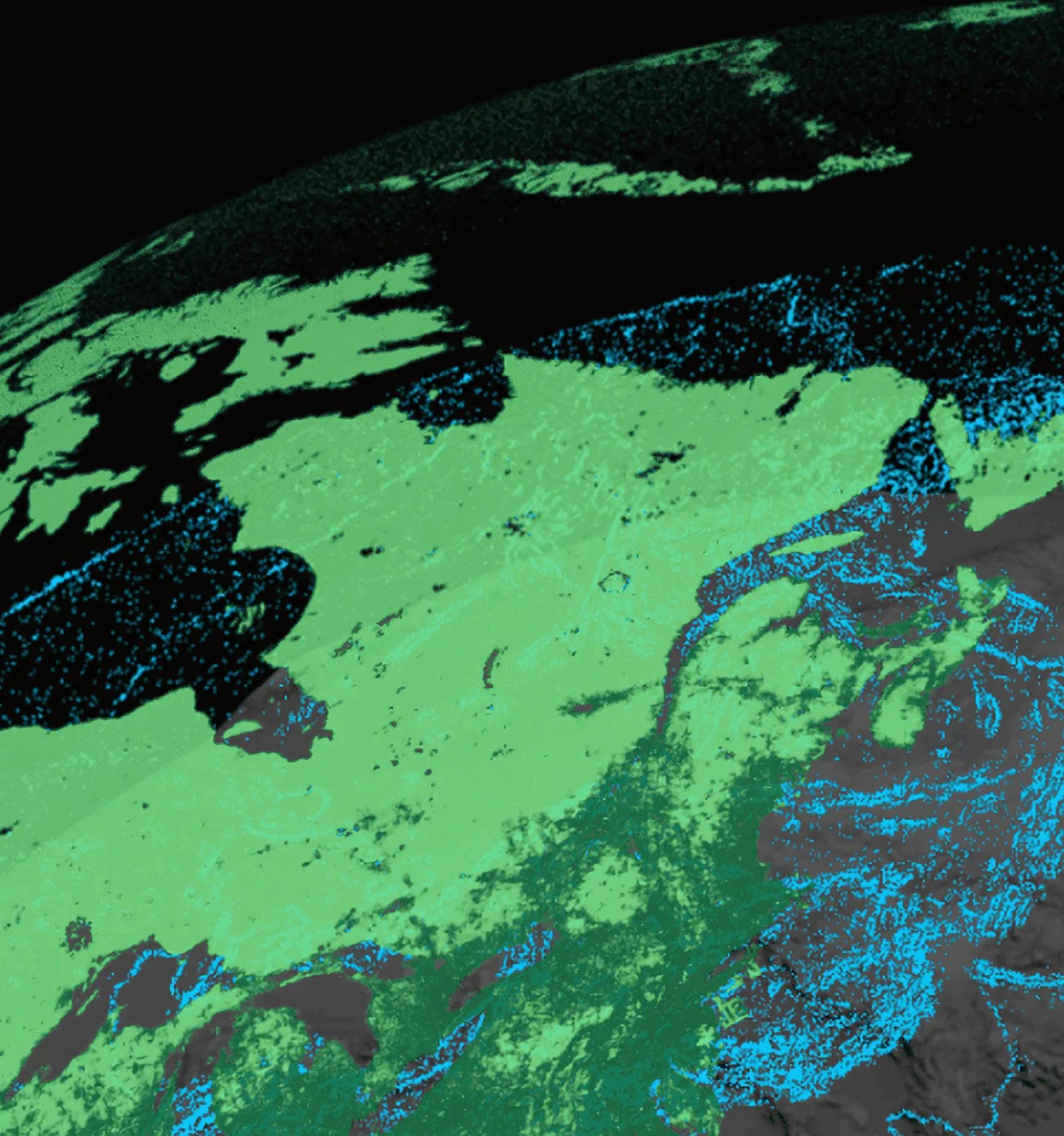




Parte 3:  
**GIS-FA para**  
**Recomendação**  
**de Cultivares**

# Sistema de informação geográfica (GIS)

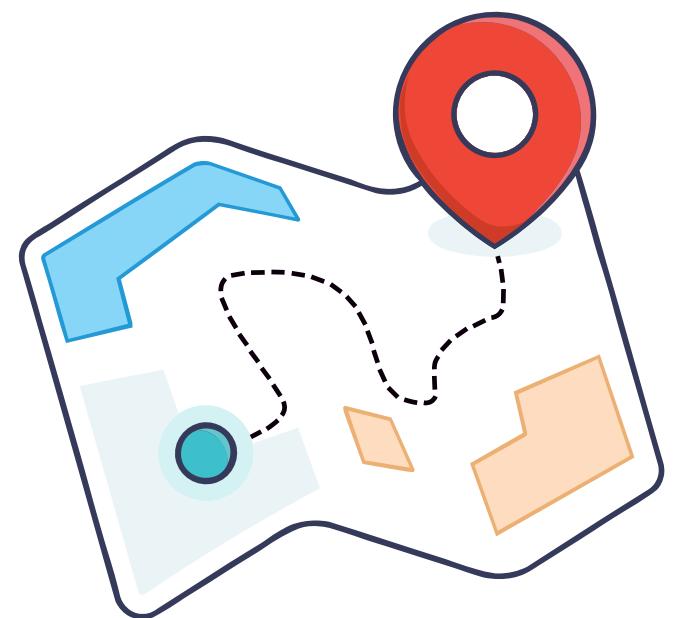
- Captura, armazenamento e manipulação dados georreferenciados
- Integra camadas de informação espacial
- Dados Vetoriais: Pontos, linhas, polígonos.
- Rasters: Imagens de satélite, modelos de elevação, NDVI etc.



# Conceito do GIS-FA

GIS

---



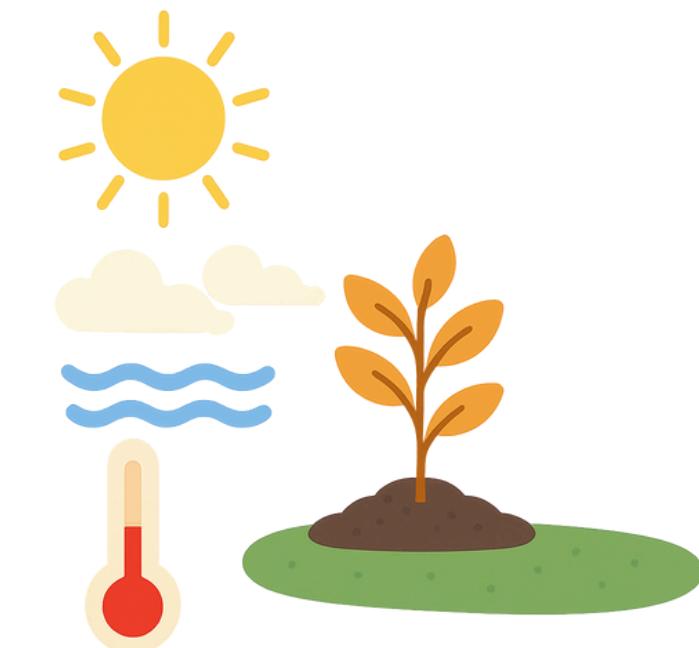
FA

---

$$g_u = (\Lambda_u^* \otimes I_t) f^*$$

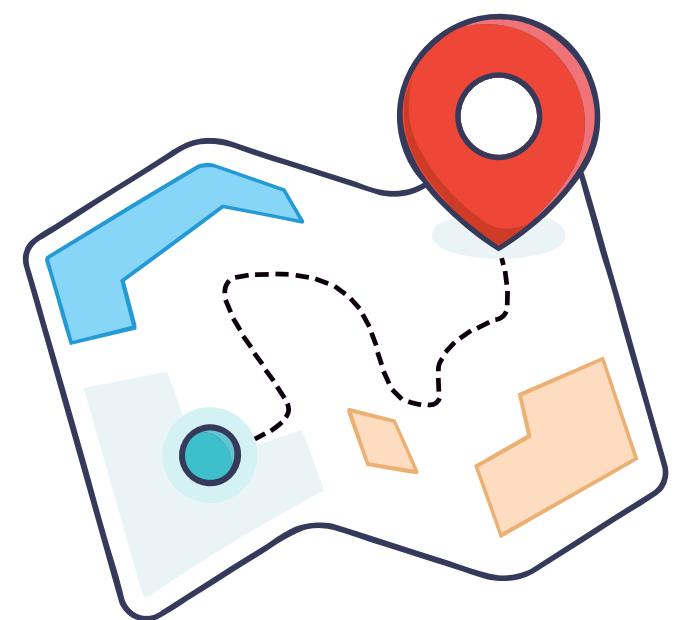
Ambientômica

---



# Conceito do GIS-FA

GIS

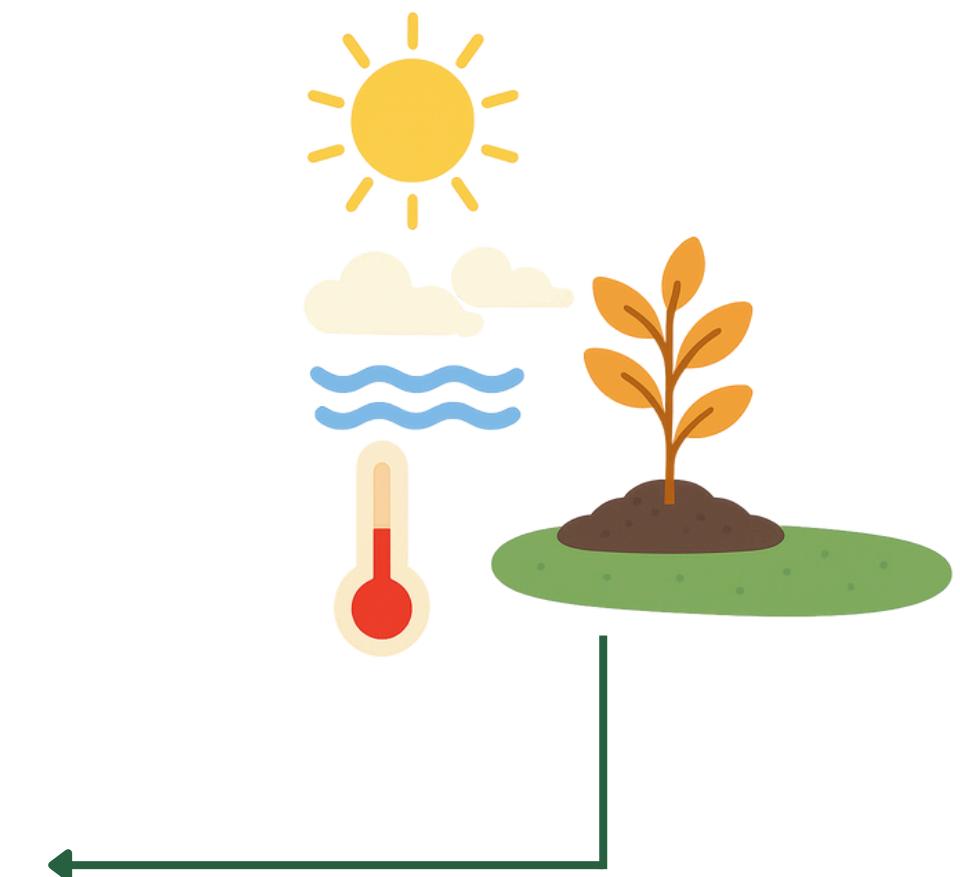
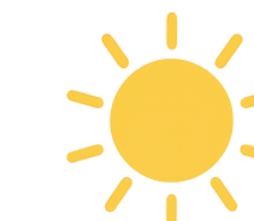


FA

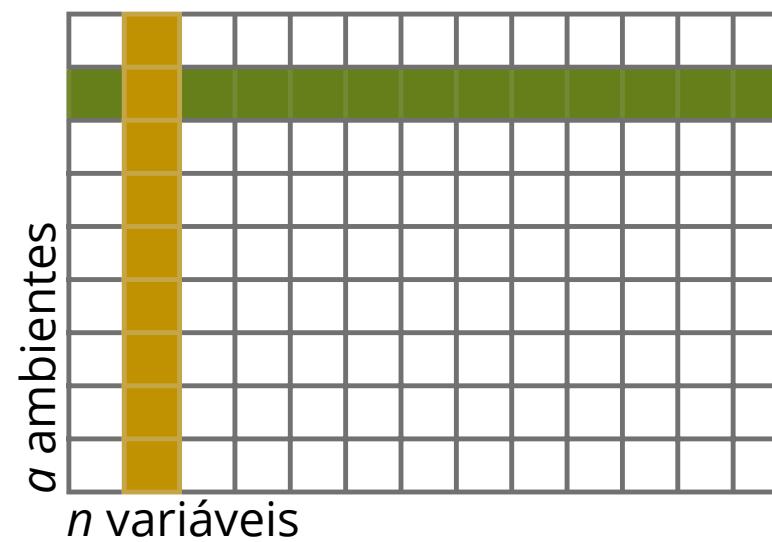
$$g_u = (\Lambda_u^* \otimes I_t) f^*$$

PLS

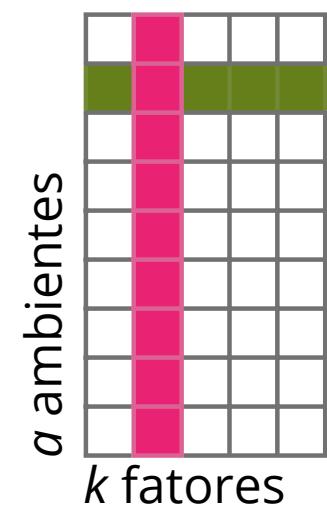
Ambientômica



## Variáveis



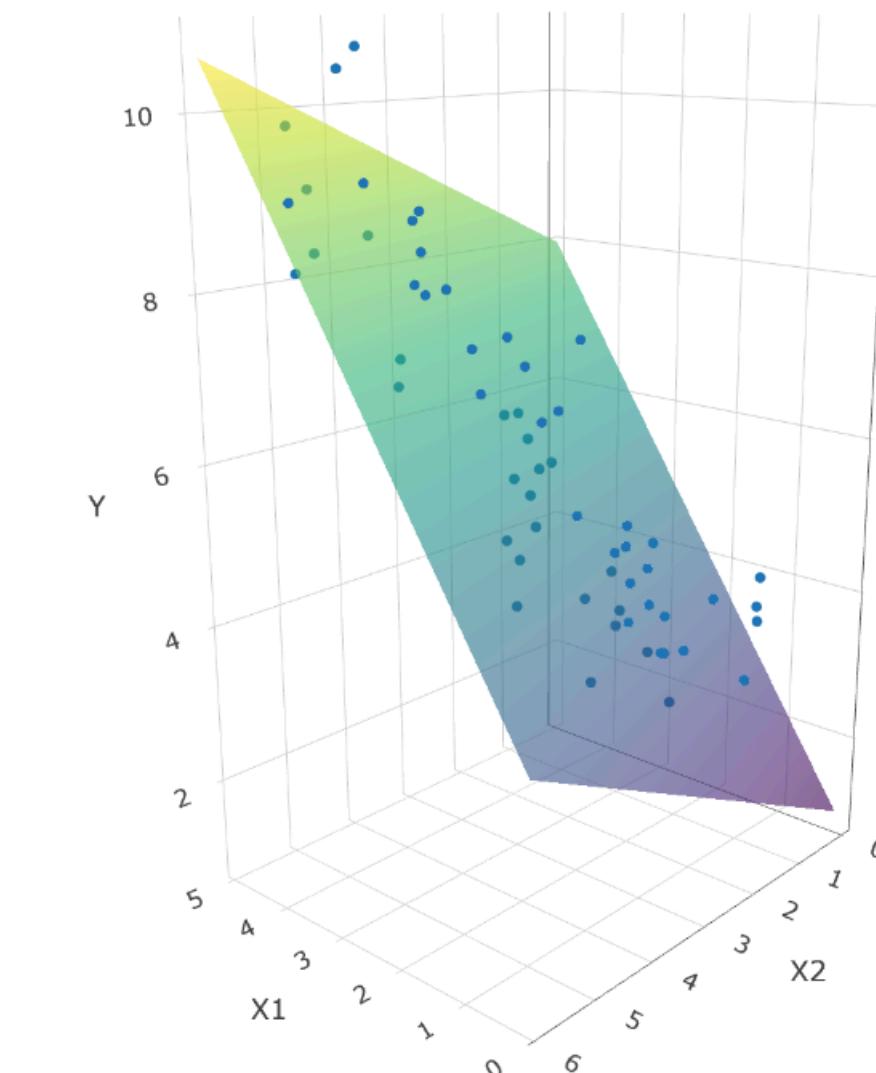
## Cargas



$$y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_n \cdot X_n + \epsilon$$

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

$$\text{Var}(\hat{\beta}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$$

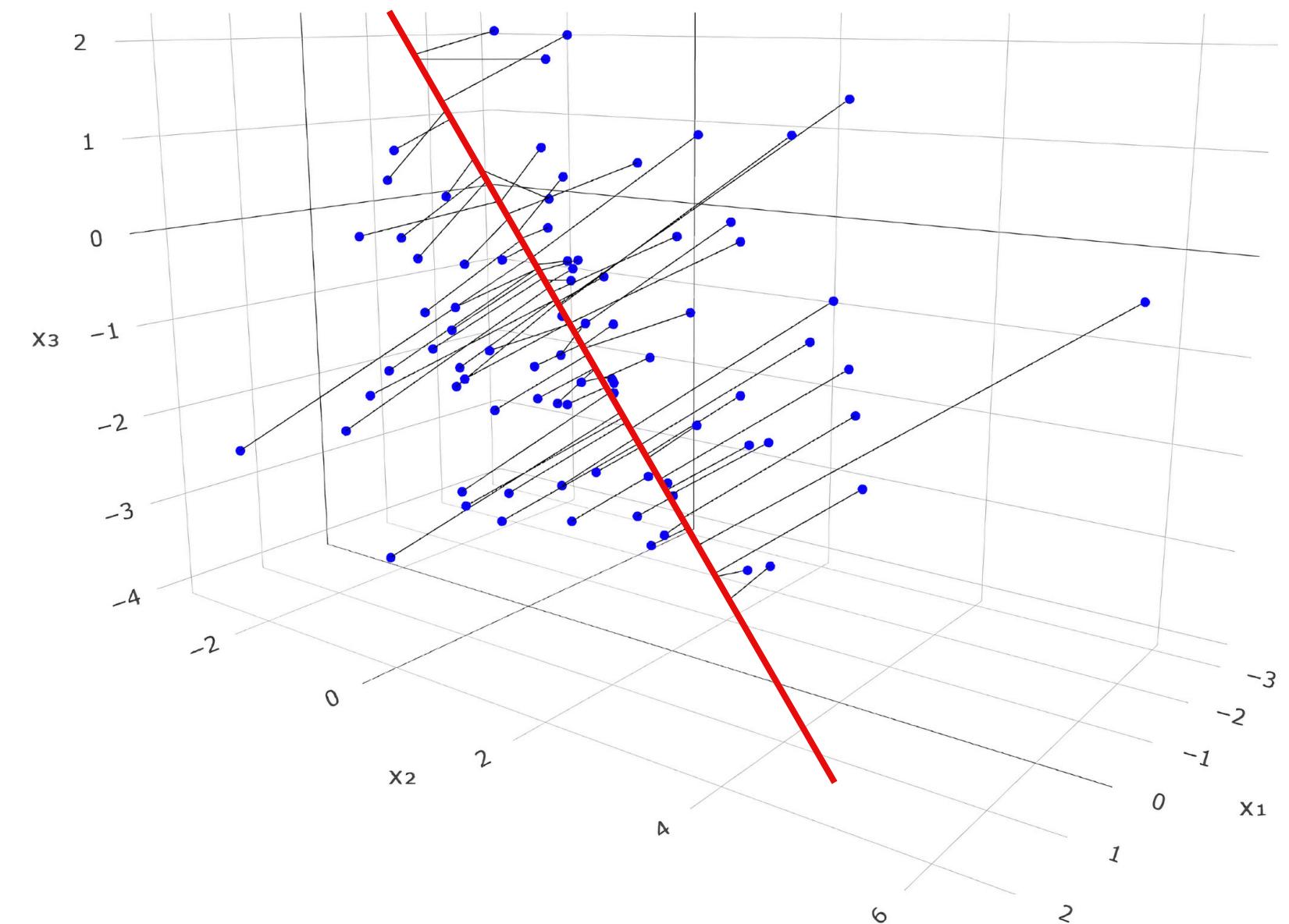


## Predição

- Predizer as cargas dos fatores através de variáveis ambientais
- Por que PLS?

# Regressão PLS

- 1 Encontrar fator que maximiza covariância entre X e Y
- 2 Usar componentes como variáveis explicativas em um modelo clássico de MLR para prever cada coluna de Y
- 3 Combinar os coeficientes da MLR para obter o vetor final de coeficientes reais para cada variável



# Regressão PLS

Variáveis



Cargas FA



$$C = X^T Y \quad \xrightarrow{\text{red arrow}} \quad w_1 = W[:, 1]$$

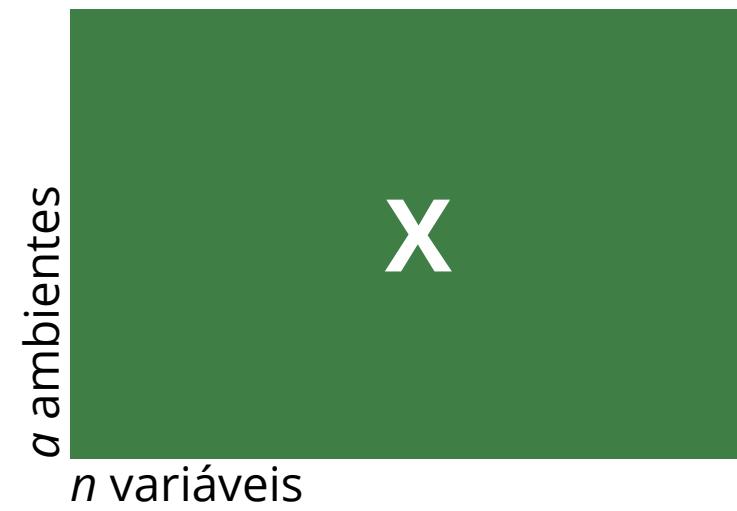
$$X^T Y = W D Z^T \quad t_1 = X w_1$$

$$\mathbf{p}_1 = \frac{\mathbf{X}^T \mathbf{t}_1}{\mathbf{t}_1^T \mathbf{t}_1} \quad \mathbf{q}_1 = \frac{\mathbf{Y}^T \mathbf{t}_1}{\mathbf{t}_1^T \mathbf{t}_1}$$

$$\mathbf{X}^{(1)} = \mathbf{X} - \mathbf{t}_1 \mathbf{p}_1^T \quad \mathbf{Y}^{(1)} = \mathbf{Y} - \mathbf{t}_1 \mathbf{q}_1^T$$

# Regressão PLS

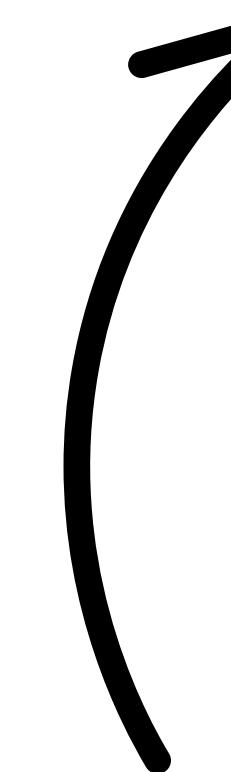
Variáveis



Cargas FA



SVD



$$C = X^T Y$$

$$X^T Y = W D Z^T$$

$$\mathbf{p}_1 = \frac{\mathbf{X}^T \mathbf{t}_1}{\mathbf{t}_1^T \mathbf{t}_1}$$

$$\mathbf{q}_1 = \frac{\mathbf{Y}^T \mathbf{t}_1}{\mathbf{t}_1^T \mathbf{t}_1}$$

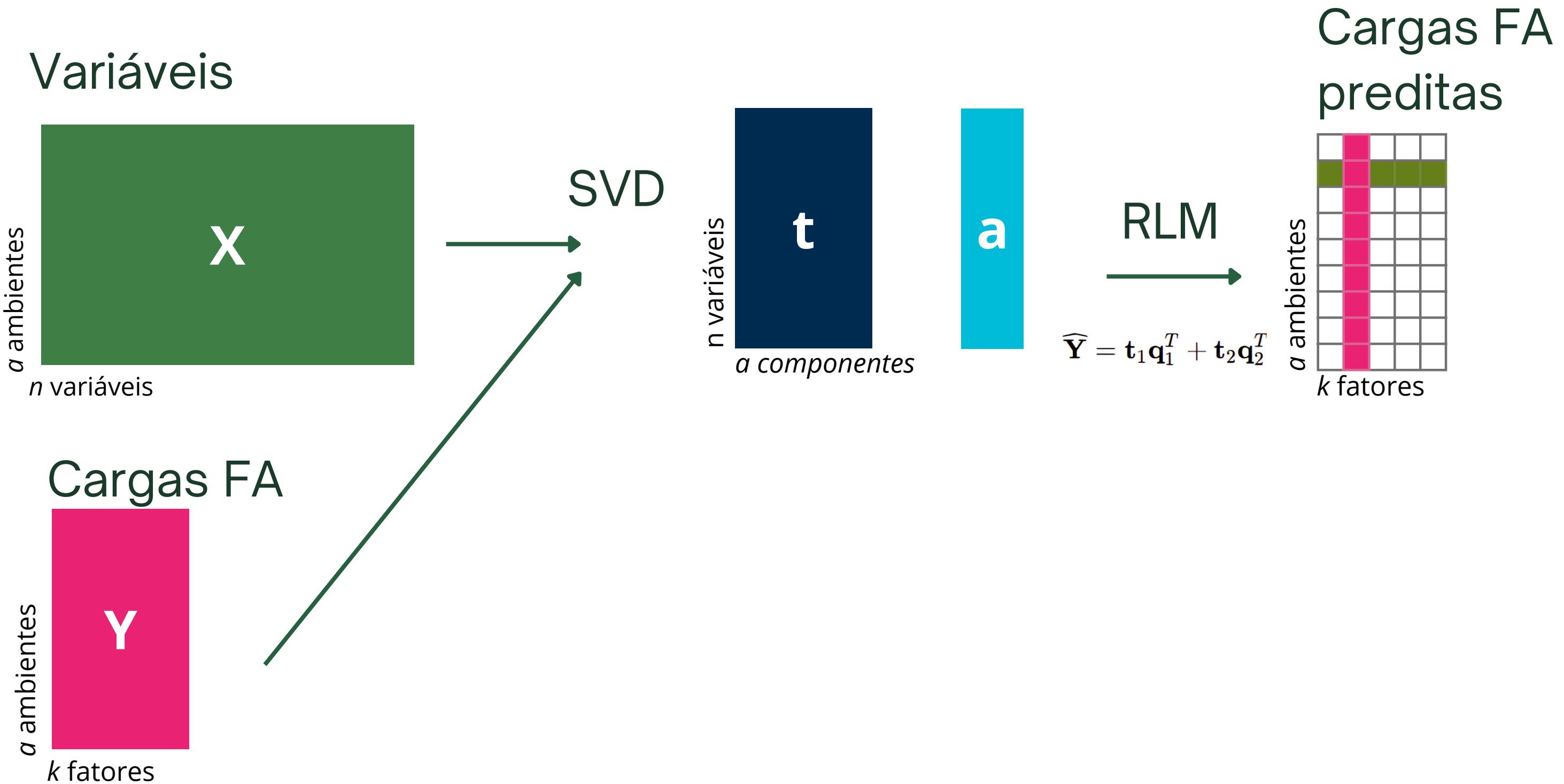
$$w_1 = W[:, 1]$$

$$t_1 = X w_1$$

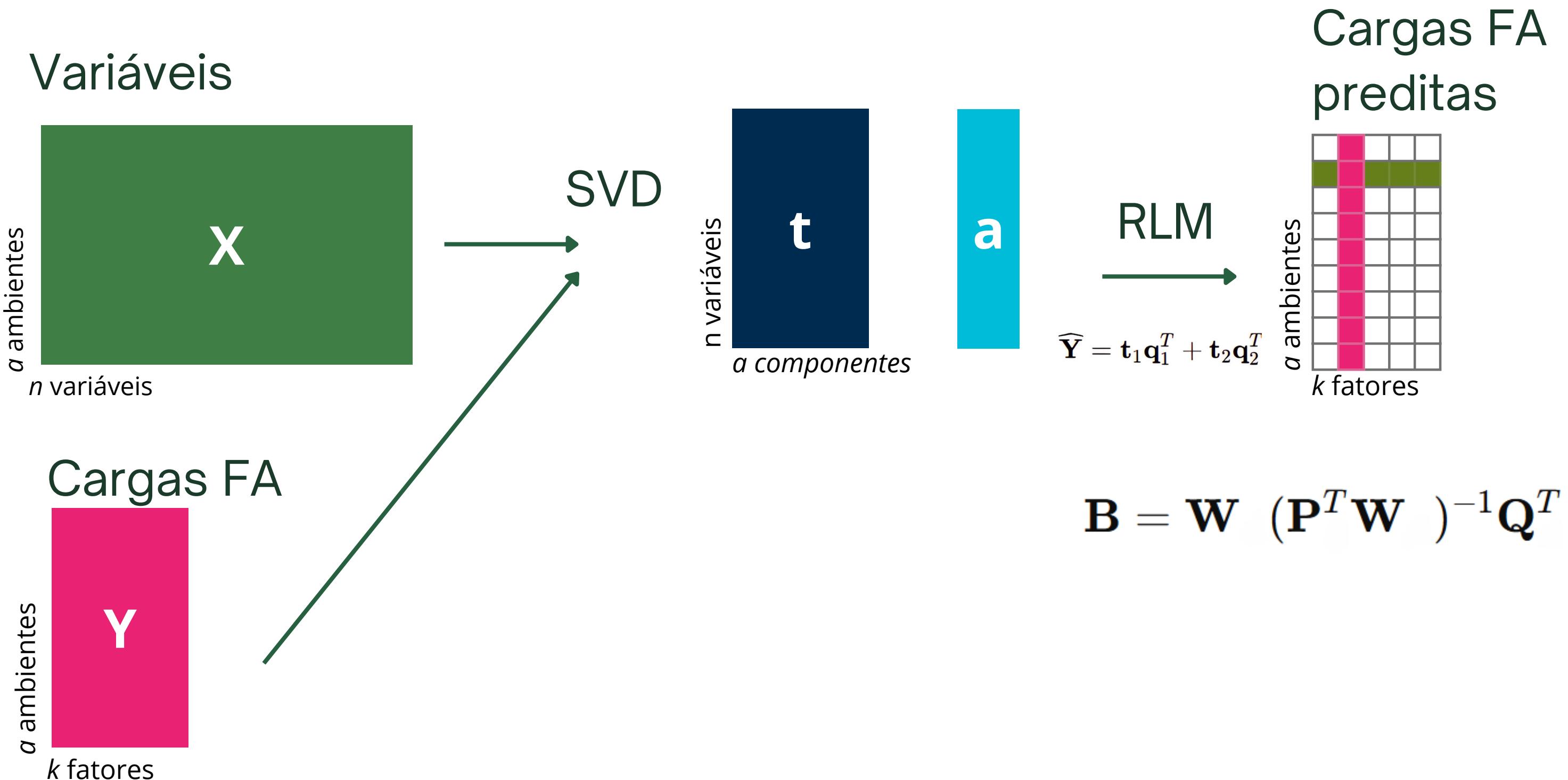
$$\mathbf{X}^{(1)} = \mathbf{X} - \mathbf{t}_1 \mathbf{p}_1^T$$

$$\mathbf{Y}^{(1)} = \mathbf{Y} - \mathbf{t}_1 \mathbf{q}_1^T$$

# Regressão PLS

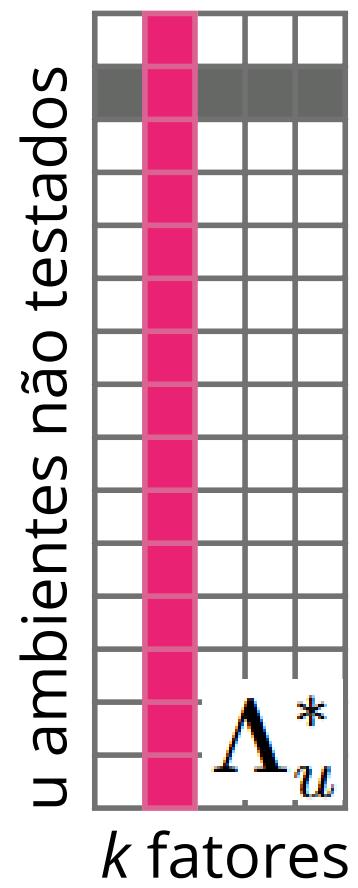


# Regressão PLS

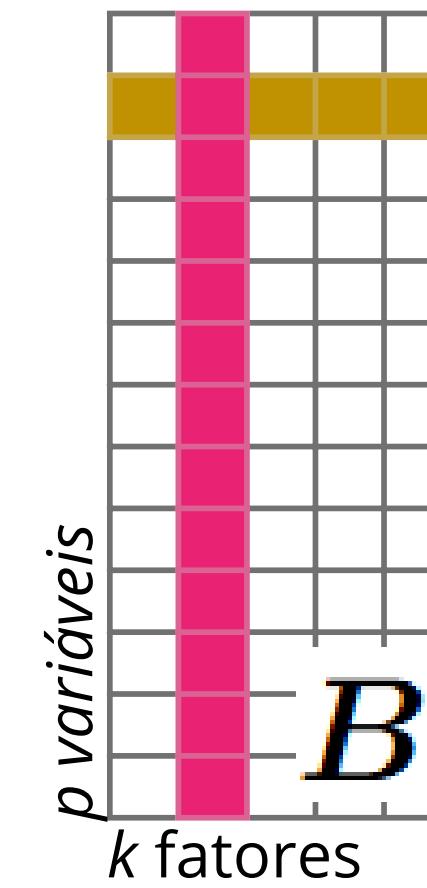
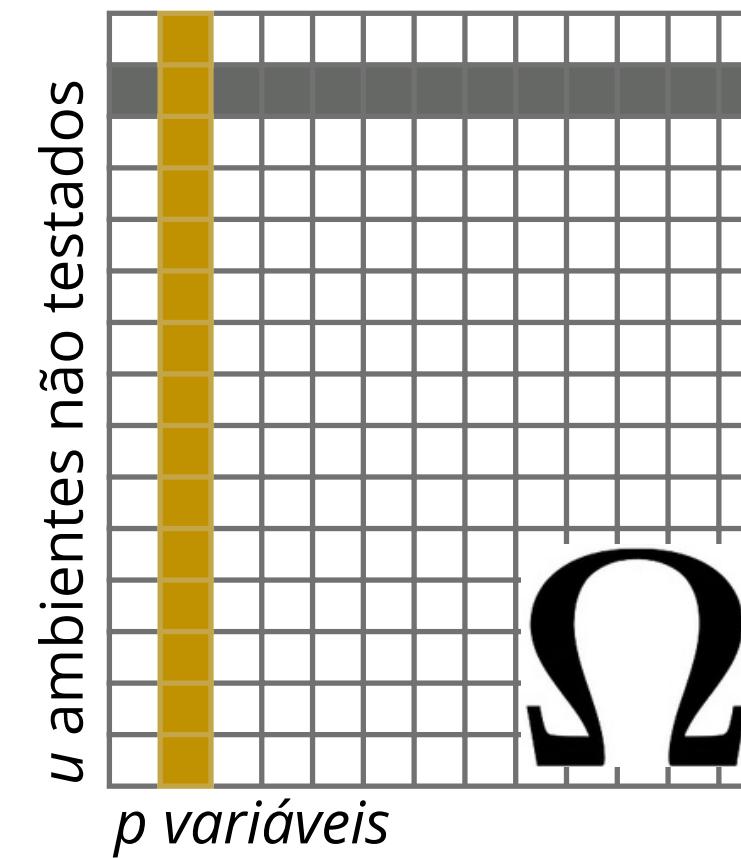


# Regressão PLS

$$\hat{\Lambda}_U^* = \Omega B^*$$



||



# Regressão PLS

## CvLOO



# Regressão PLS

## CvLOO



# Regressão PLS

## CvLOO

$$\text{X teste} \quad X \quad = \quad Y$$

- Salvamos os Y preditos e calculamos acurárias médias

# Porque a PLS

## Vantagem 1

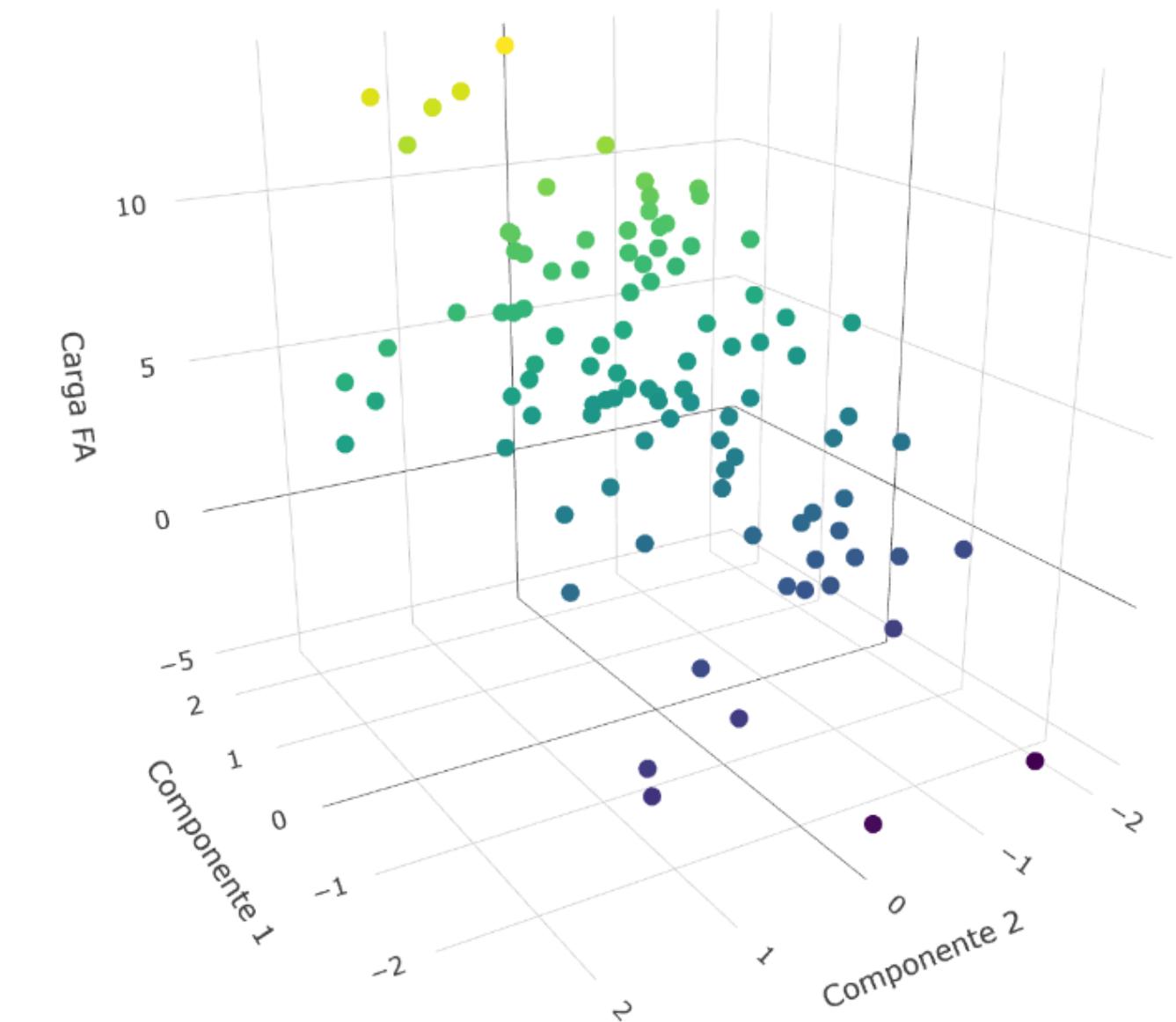
Redução da dimensionalidade da variáveis ambientais em componentes independentes

## Vantagem 2

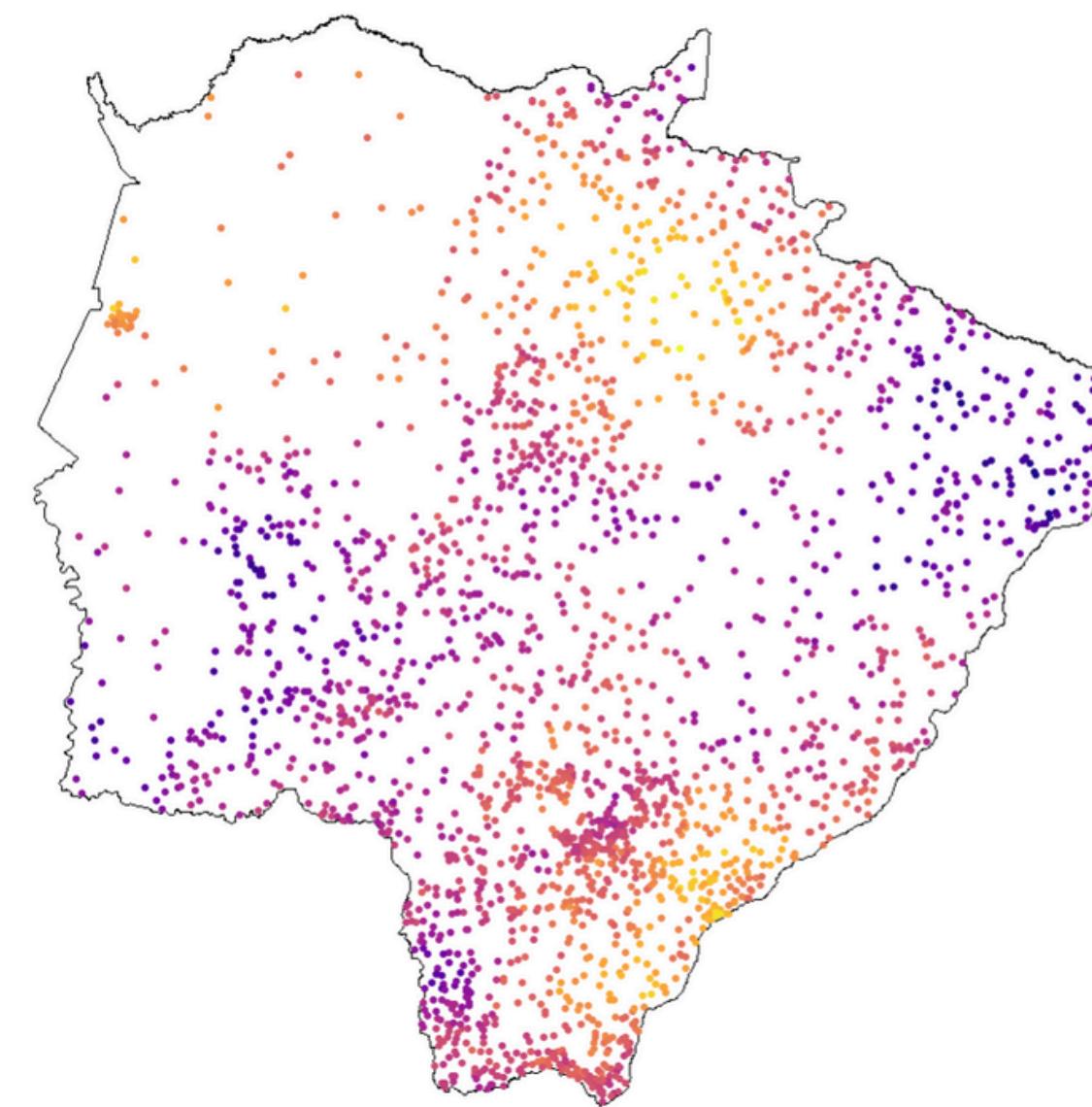
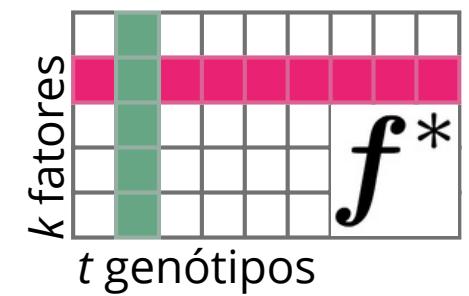
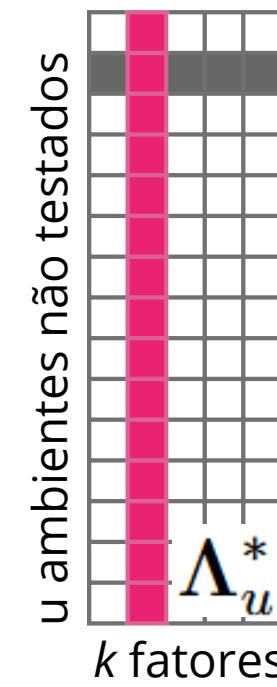
Permite interpretar as cargas para descobrir quais fatores ambientais mais influenciam variável resposta

## Vantagem 3

Evita overfitting, ideal para predição



# Predição dos valores genotípicos



$$\mathbf{g}_U = (\hat{\Lambda}_U^\star \otimes \mathbf{I}_V) \tilde{\mathbf{f}}^\star$$

# Interpolação

- IDW
- Pondera pela distância a pontos conhecidos

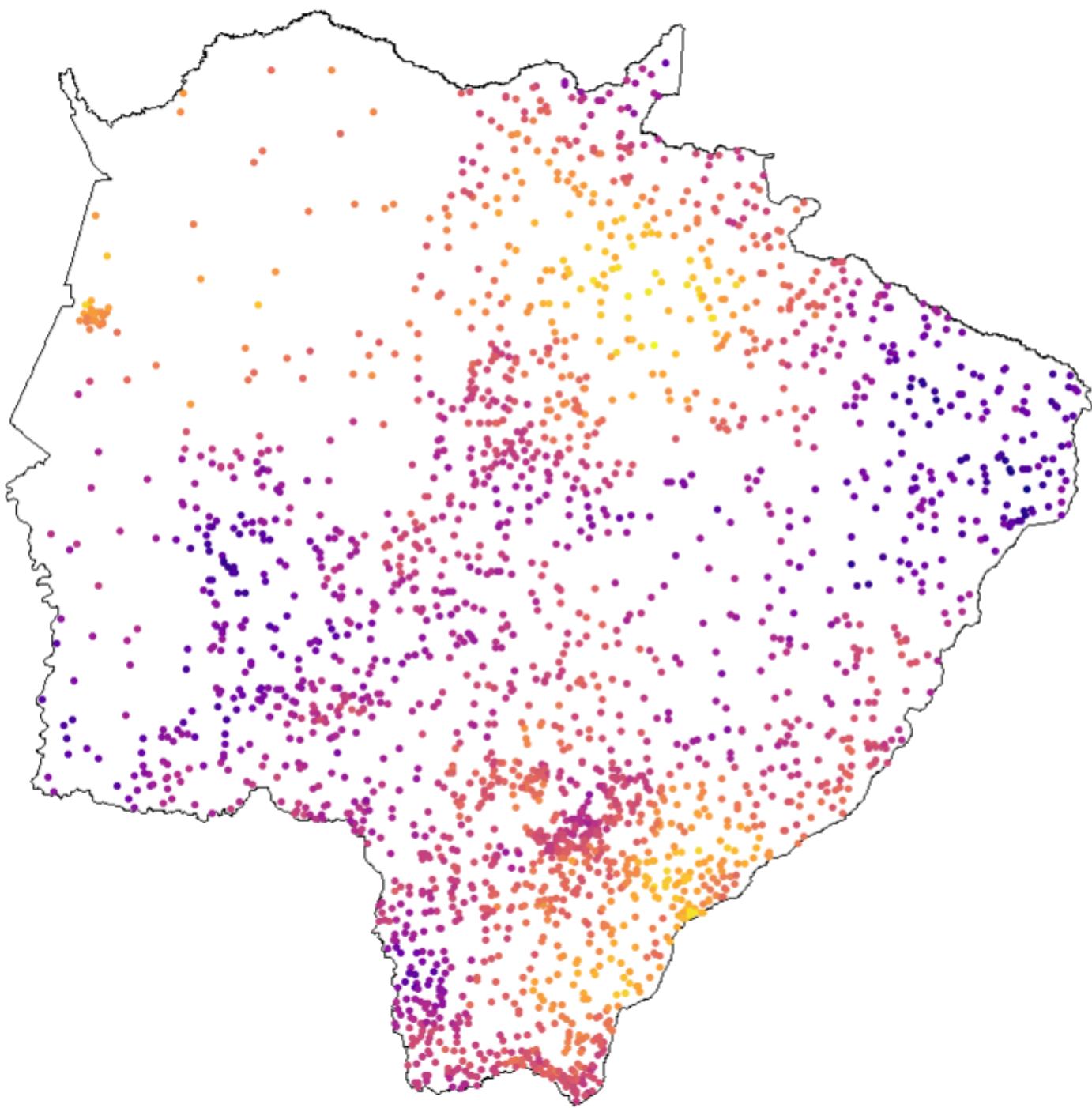
$$\hat{Z}(s_0) = \frac{\sum_{i=1}^n w_i(s_0) Z(s_i)}{\sum_{i=1}^n w_i(s_0)}$$

$Z(s_i)$  = valor observado no ponto  $s_i$ ,

$w_i(s_0)$  = peso dado ao ponto  $s_i$  com relação ao ponto  $s_0$ ,

$n$  = número de pontos considerados.

$$w_i(s_0) = \frac{1}{d(s_0, s_i)^p}$$



# Interpolação

- IDW
- Pondera pela distância a pontos conhecidos

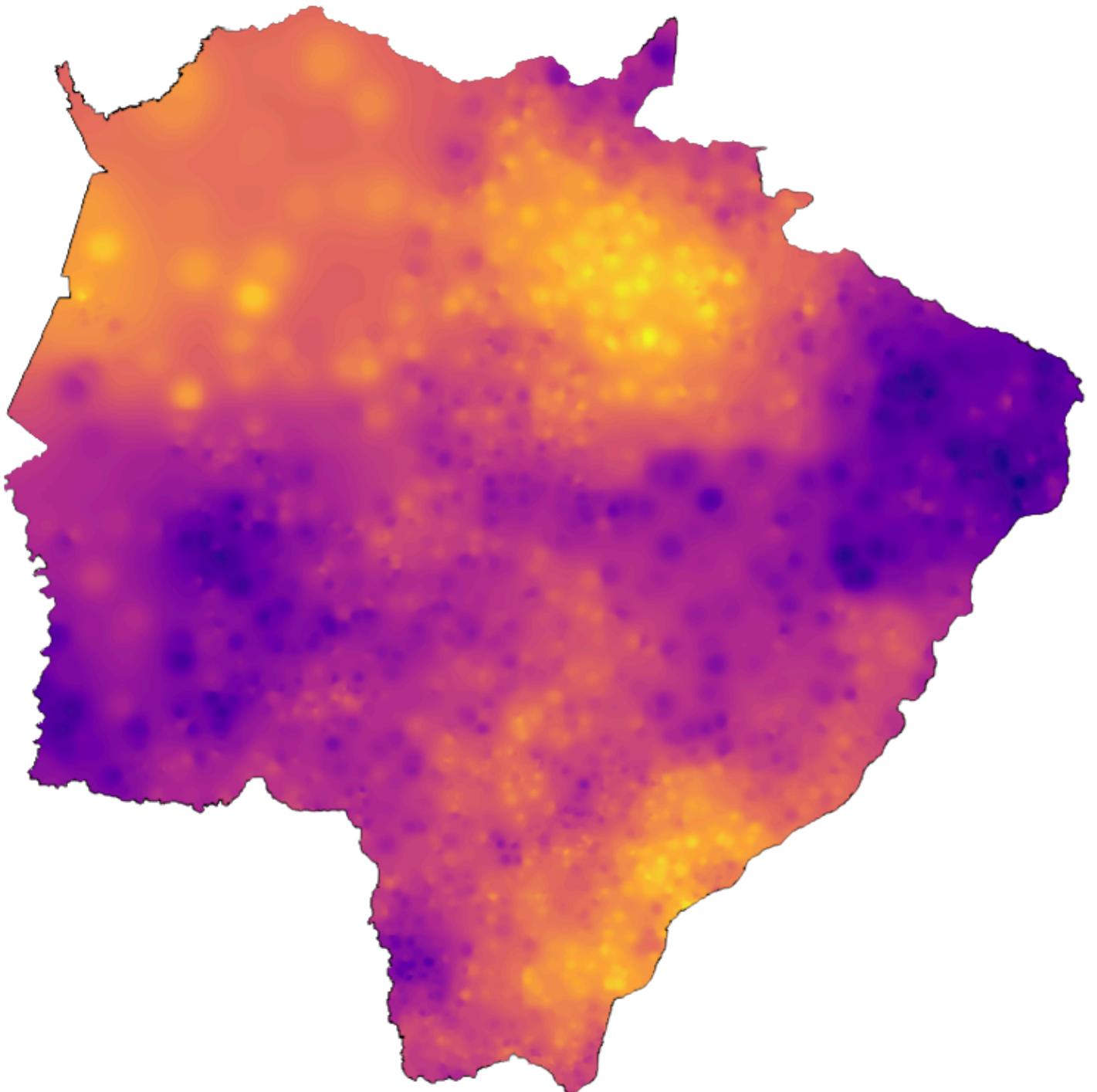
$$\hat{Z}(s_0) = \frac{\sum_{i=1}^n w_i(s_0) Z(s_i)}{\sum_{i=1}^n w_i(s_0)}$$

$Z(s_i)$  = valor observado no ponto  $s_i$ ,

$w_i(s_0)$  = peso dado ao ponto  $s_i$  com relação ao ponto  $s_0$ ,

$n$  = número de pontos considerados.

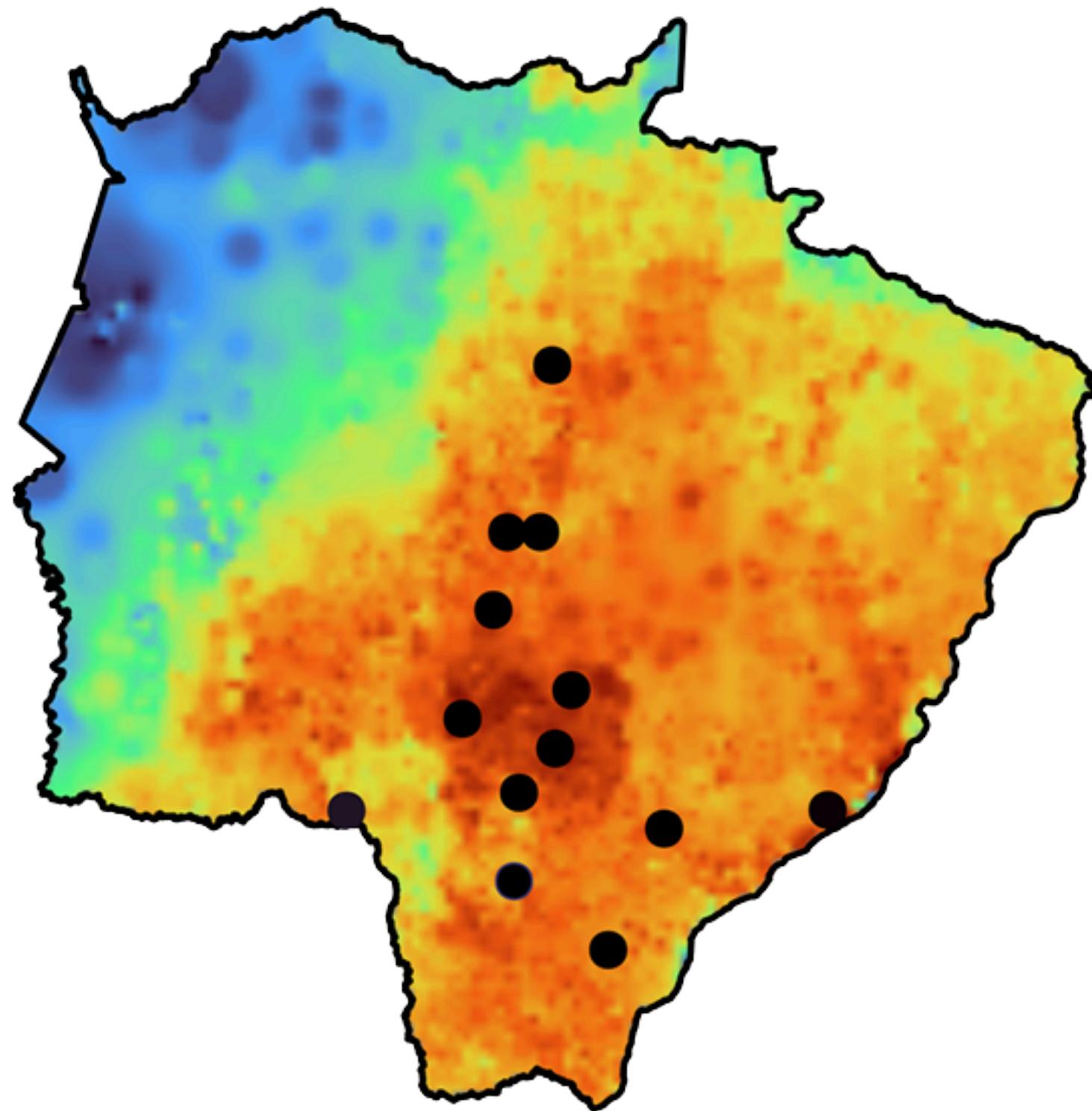
$$w_i(s_0) = \frac{1}{d(s_0, s_i)^p}$$



# Similaridade Ambiental

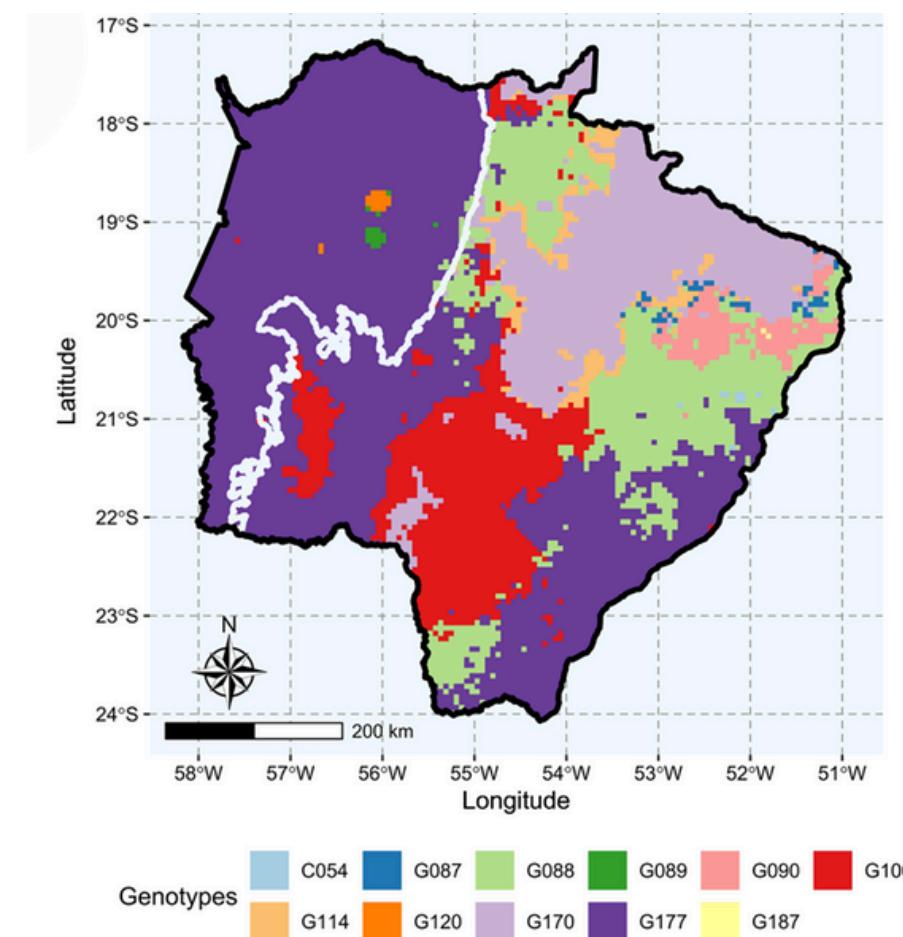
$$D_{ju} = \sqrt{\sum_{p=1}^P (w_{jp} - \omega_{up})^2}$$

$$S_u = \min(D_{uj}) \quad S_{u^*} = \min(D_{u^*j})$$

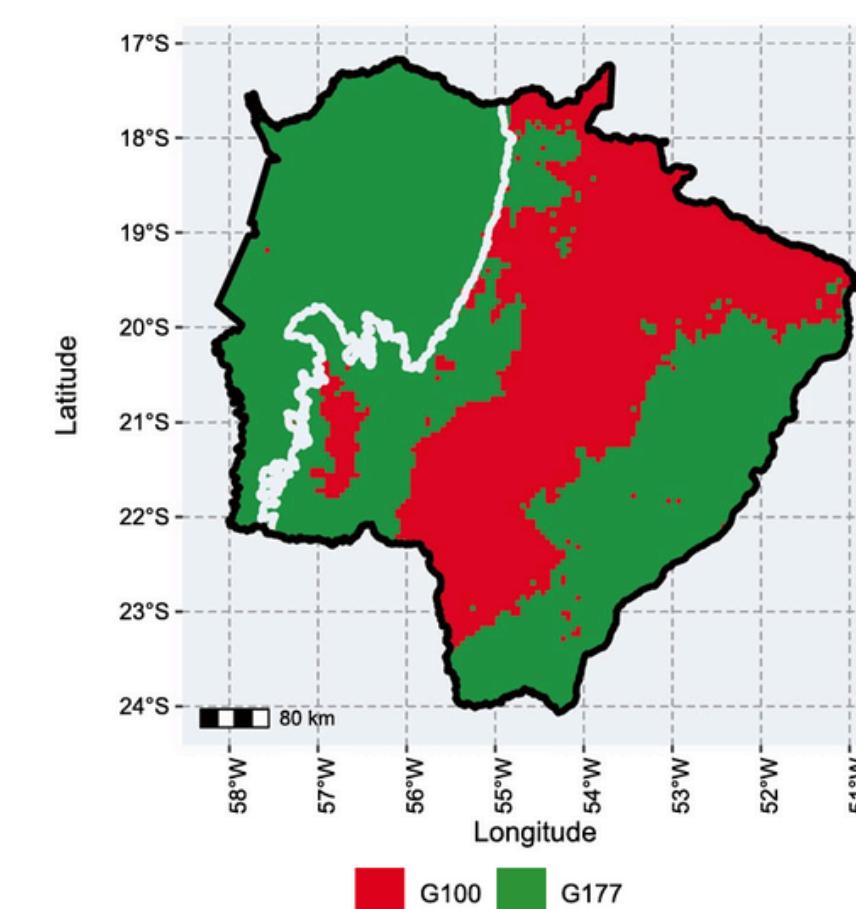


# Mapas Temáticos

## Quem ganha onde



## Comparação par a par



## Zonas de adaptação

