

Análises de Agrupamento para Modelos de Produtividade de Soja

Viviana Giampaoli e José Victor Bartol Rodrigues
IME-USP

**1^{er} Workshop do Seguro Agrícola:
Modelagem Estatística e Precificação**

CAPES-PROCAD

Belo Horizonte

25 a 27 de Novembro de 2009

Análises de Agrupamento

Entropia

Se C e C' são diferentes agrupamentos de uma amostra, com K e K' grupos então as seguintes podem ser definidas:

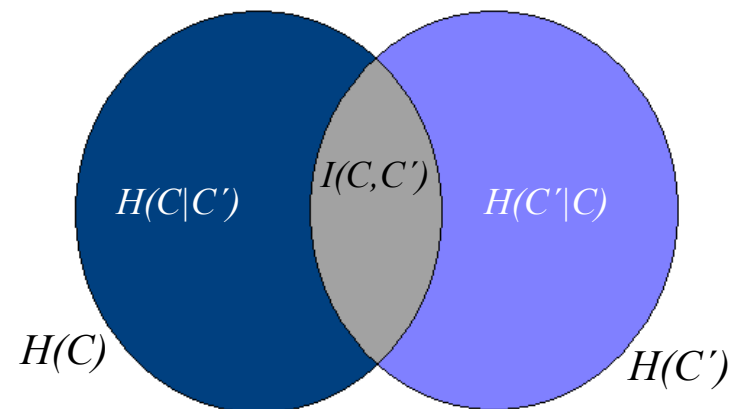
$$H(C) = - \sum_{k=1}^K P(k) \log[P(k)] \quad \text{entropia}$$

$$H(C'|C) = - \sum_{k=1}^K \sum_{k'=1}^{K'} P(k, k') \log [P(k'|k)] \quad \text{entropia condicional}$$

$$I(C, C') = H(C) - H(C|C') = H(C') - H(C'|C)$$

informação mútua

Figura 1 – Diagrama de Venn para as entropias



Varição da Informação

$$\begin{aligned} VI(C, C') &= H(C) + H(C') - 2 \cdot I(C, C') \\ &= H(C | C') + H(C' | C) \end{aligned}$$

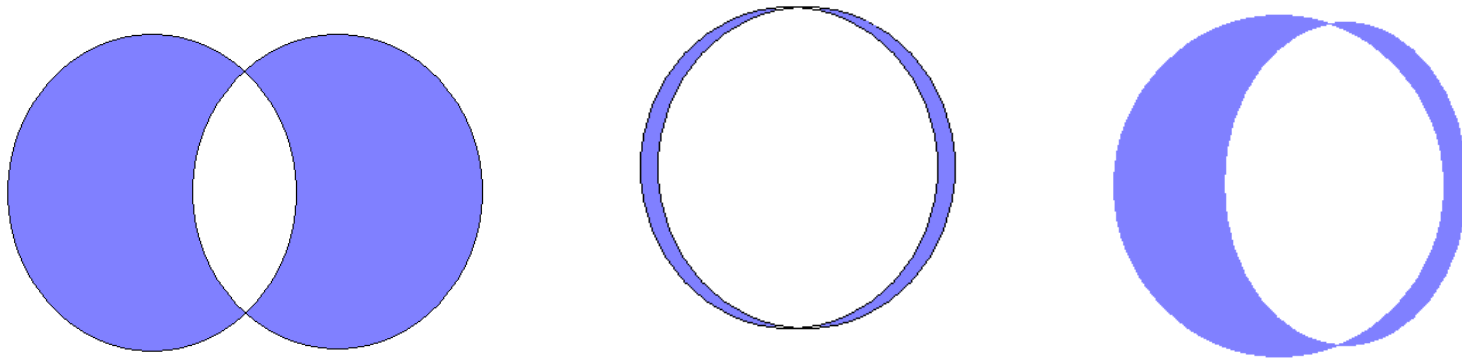


Figura 2 – Representação da Variação da informação

Diferença das entropias condicionais

Suponhamos que se defina que existem K grupos na amostra sob estudo. Com um método de agrupamento temos $1, 2, \dots, K$ clusters, definimos

$$dHC(k) = [H(C_{k+1} | C_k) - H(C_k | C_{k-1})] \times H(C_k),$$
$$k = 2, \dots, K.$$

k^* valor mínimo de $dHC(k)$ é o valor ótimo do número de grupos

Diferença das entropias condicionais

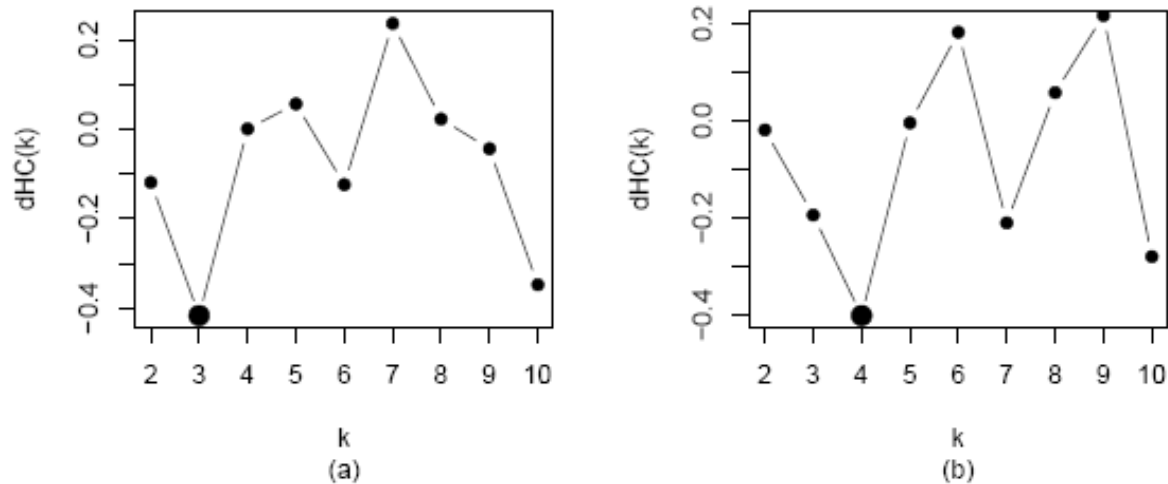


Figura 3 – Exemplos de k vs $dH(C_k)$

A comparação do desempenho desta proposta com outras técnicas pode ser encontrada em Freitas (2007).

Variáveis consideradas para o agrupamento

- ✓ Temperatura mínima (°C);
- ✓ Temperatura máxima (°C);
- ✓ Temperatura média (°C);

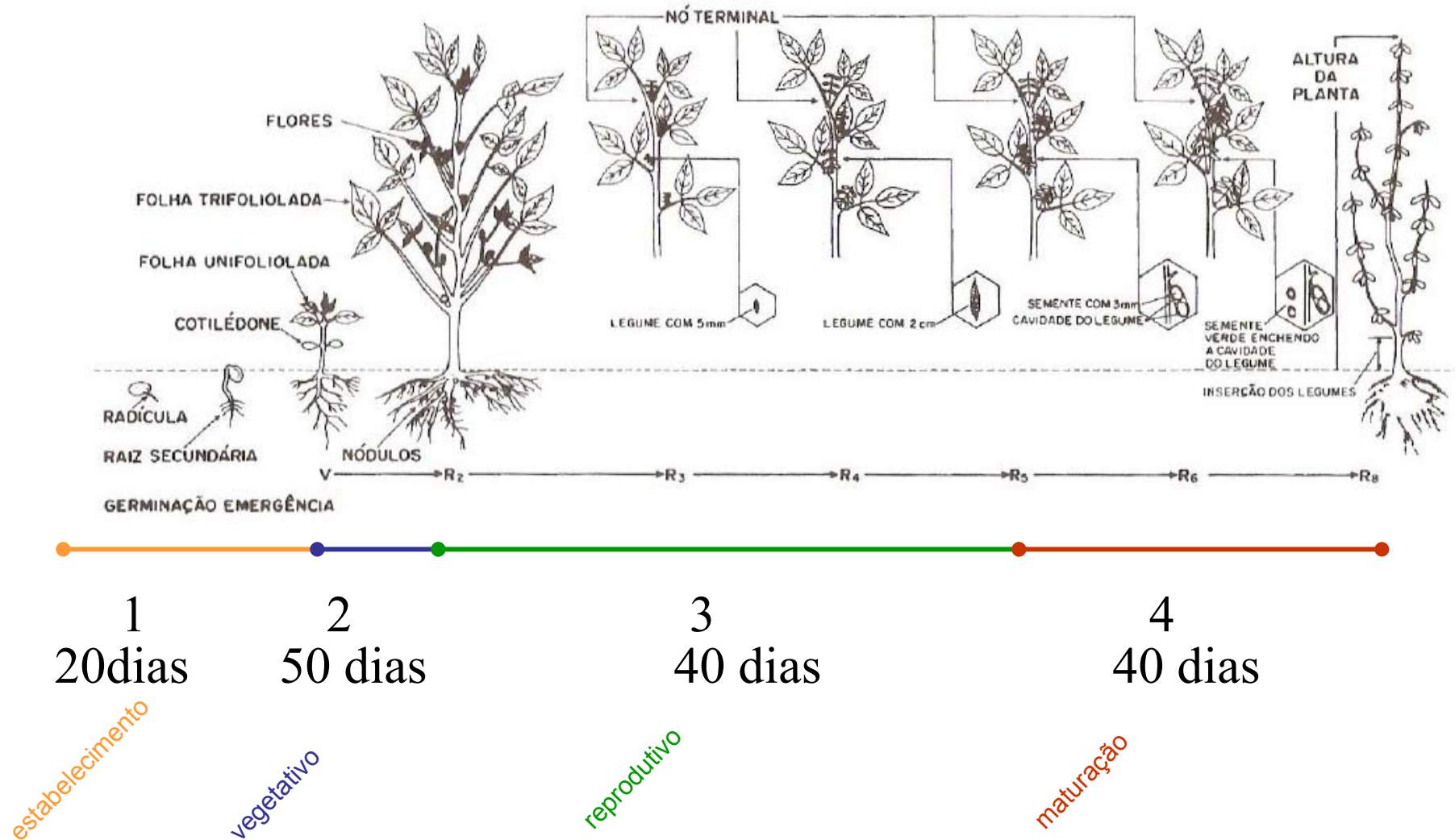
- ✓ Precipitação acumulada (mm);
- ✓ Precipitação máxima (mm);

- ✓ Todas as variáveis foram calculadas utilizando intervalos de cinco dias.

Cultivo de Soja

Estádios fenológicos de um ciclo de cultura de soja

Figura 4: Adaptada de Costa (1996)



Estádios fenológicos de um ciclo de cultura de soja

Tabela 1. Estádios vegetativos e reprodutivos da soja¹.

Estádios vegetativos	Estádios reprodutivos
VE - Emergência	R ₁ - Início do florescimento
VC - Cotilédone	R ₂ - Pleno florescimento
V ₁ - Primeiro nó	R ₃ - Início da formação das vagens
V ₂ - Segundo nó	R ₄ - Plena formação das vagens
V ₃ - Terceiro nó	R ₅ - Início do enchimento das sementes
*	R ₆ - Pleno enchimento das vagens
*	R ₇ - Início da maturação
V(n) - enésimo nó	R ₈ - Maturação plena

¹ Este sistema identifica exatamente os estádios da planta de soja. Porém, nem todas as plantas em um dado campo estarão no mesmo estágio ao mesmo tempo. Quando se divide em estádios um campo de soja, cada estágio específico V ou R é definido somente quando 50% ou mais das plantas no campo estão nele ou entre aquele estágio.

Estádios fenológicos de um ciclo de cultura de soja

- 1) Da sementeira até 20 dias após a emergência, coincidindo com a primeira folha trifoliada desenvolvida, denominado período de estabelecimento da cultura (S-V₂);
- 2) De 20 dias após a emergência até o início do florescimento, denominado período vegetativo da cultura (V₂-R₁);
- 3) Do início do florescimento ao início do enchimento de grãos, denominado período reprodutivo da cultura (R₁-R₅);
- 4) Do enchimento de grãos até a maturação fisiológica, denominado período de maturação de grãos da cultura de soja (R₅-R₇);

Variação da Informação segundo os estádios fenológicos e os anos

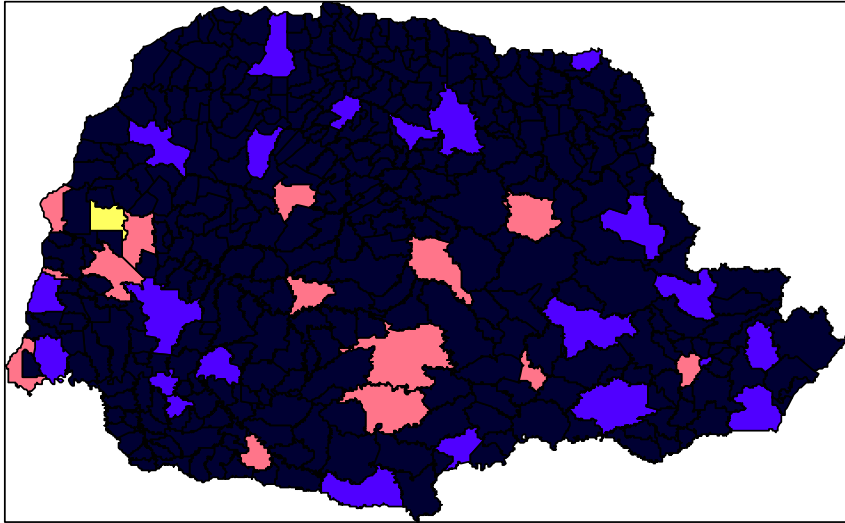
VI fase 1	2004	2005	2006	2007	2008	VI fase 3	2004	2005	2006	2007	2008
2003	0,451	0,398	0,261	0,199	0,315	2003	0,304	0,479	0,172	0,343	0,179
2004		0,360	0,530	0,467	0,515	2004		0,367	0,315	0,324	0,273
2005			0,436	0,419	0,526	2005			0,520	0,364	0,481
2006				0,231	0,408	2006				0,357	0,179
2007					0,352	2007					0,378
2008						2008					
VI fase 2	2004	2005	2006	2007	2008	VI fase 4	2004	2005	2006	2007	2008
2003	0,448	0,258	0,273	0,317	0,459	2003	0,391	0,281	0,301	0,280	0,247
2004		0,275	0,420	0,494	0,575	2004		0,378	0,380	0,401	0,385
2005			0,251	0,376	0,466	2005			0,205	0,199	0,205
2006				0,275	0,471	2006				0,209	0,145
2007					0,579	2007					0,180
2008						2008					

Variação da Informação segundo os estádios fenológicos e os anos

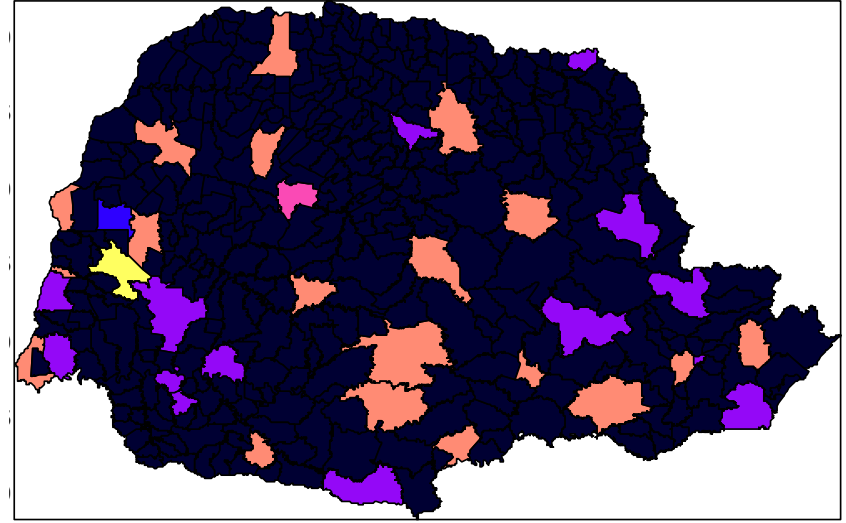
ANO 2003	Fase 2	Fase 3	Fase 4	ANO 2006	Fase 2	Fase 3	Fase 4
Fase 1	0,277	0,207	0,249	Fase 1	0,229	0,292	0,226
Fase 2		0,233	0,313	Fase 2		0,231	0,206
Fase 3			0,247	Fase 3			0,229
Fase 4				Fase 4			
ANO 2004	Fase 2	Fase 3	Fase 4	ANO 2007	Fase 2	Fase 3	Fase 4
Fase 1	0,485	0,356	0,436	Fase 1	0,260	0,329	0,124
Fase 2		0,470	0,473	Fase 2		0,356	0,303
Fase 3			0,375	Fase 3			0,357
Fase 4				Fase 4			
ANO 2005	Fase 2	Fase 3	Fase 4	ANO 2008	Fase 2	Fase 3	Fase 4
Fase 1	0,421	0,369	0,379	Fase 1	0,493	0,232	0,378
Fase 2		0,471	0,179	Fase 2		0,462	0,460
Fase 3			0,447	Fase 3			0,179
Fase 4				Fase 4			

Mapas resultantes do agrupamento segundo o estágio fenológico

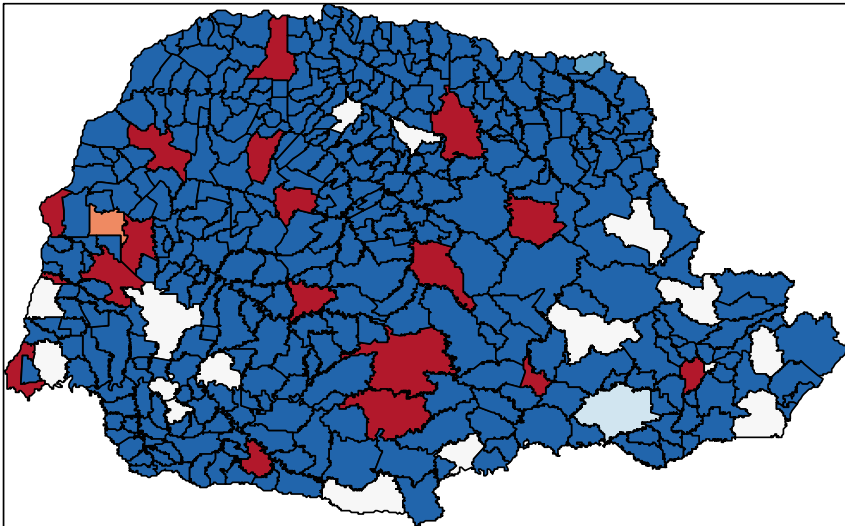
Agrupamento para Fase 1 de 2003



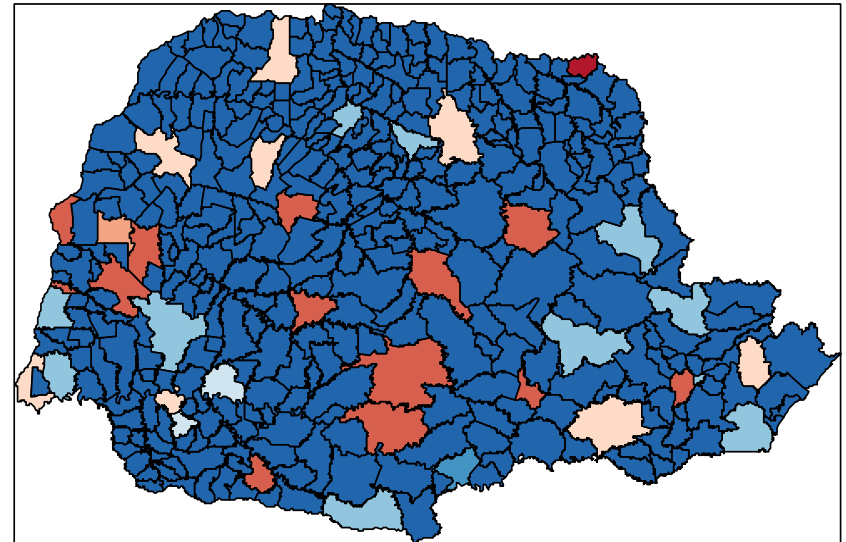
Agrupamento para Fase 2 de 2003



Agrupamento para Fase 3 de 2003



Agrupamento para Fase 4 de 2003



Evapotranspiração

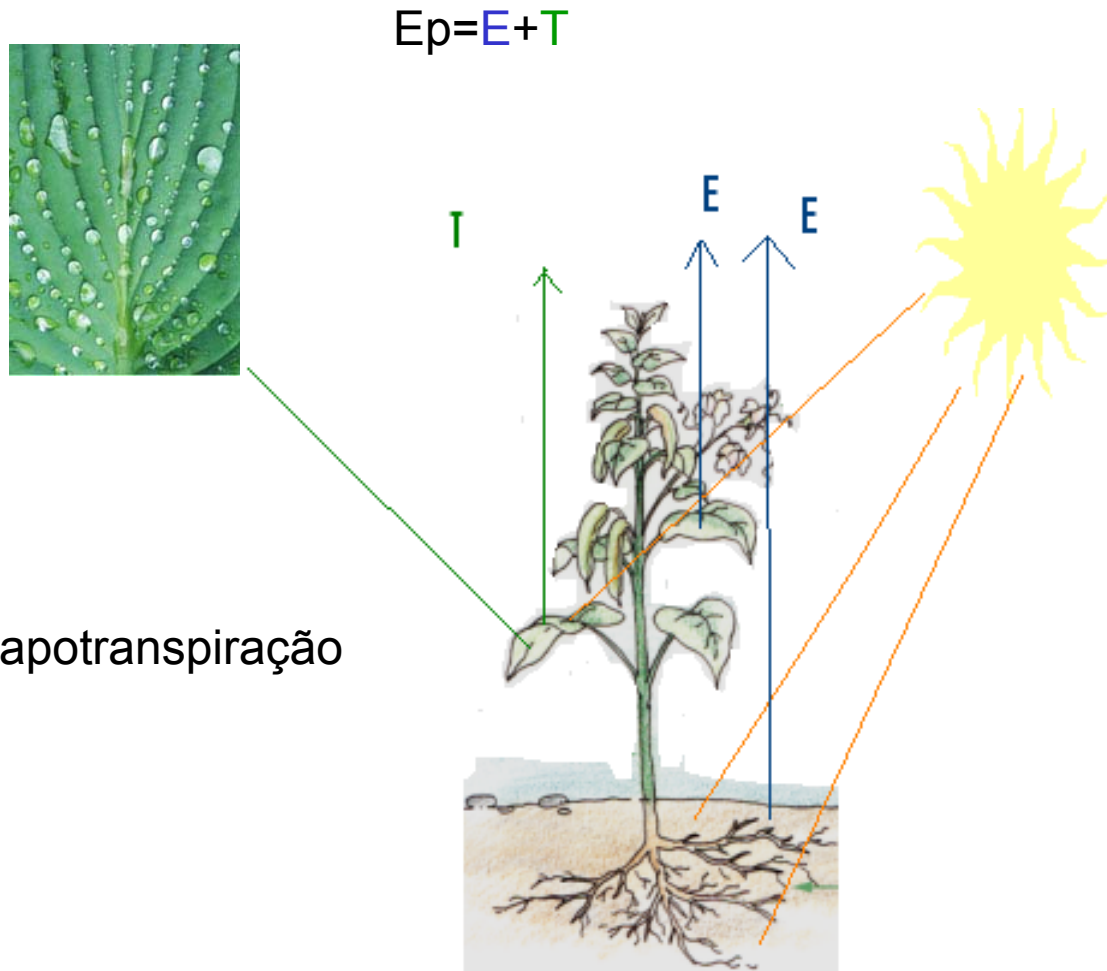


Figura 5: Evapotranspiração

Evapotranspiración

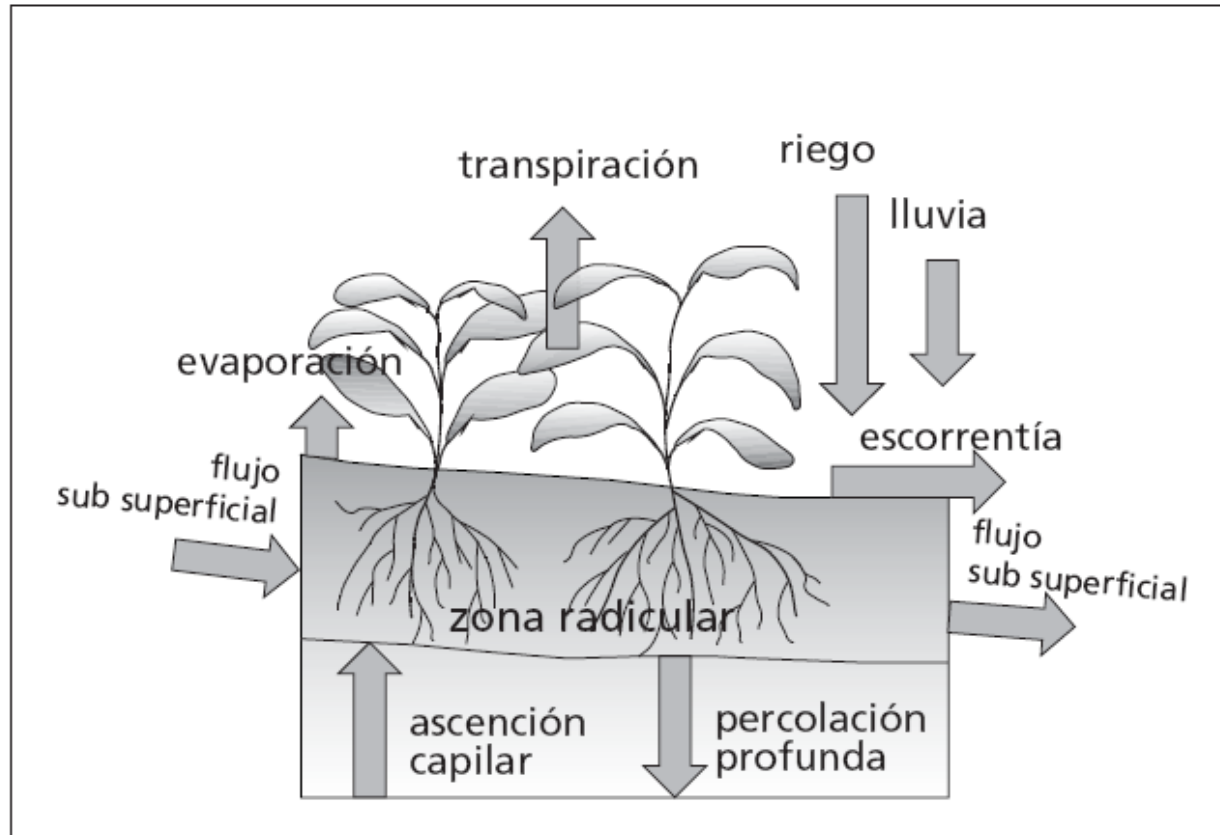


Figura 6: Balanço da água do solo da zona radicular adaptada FAO 2006

Evapotranspiração

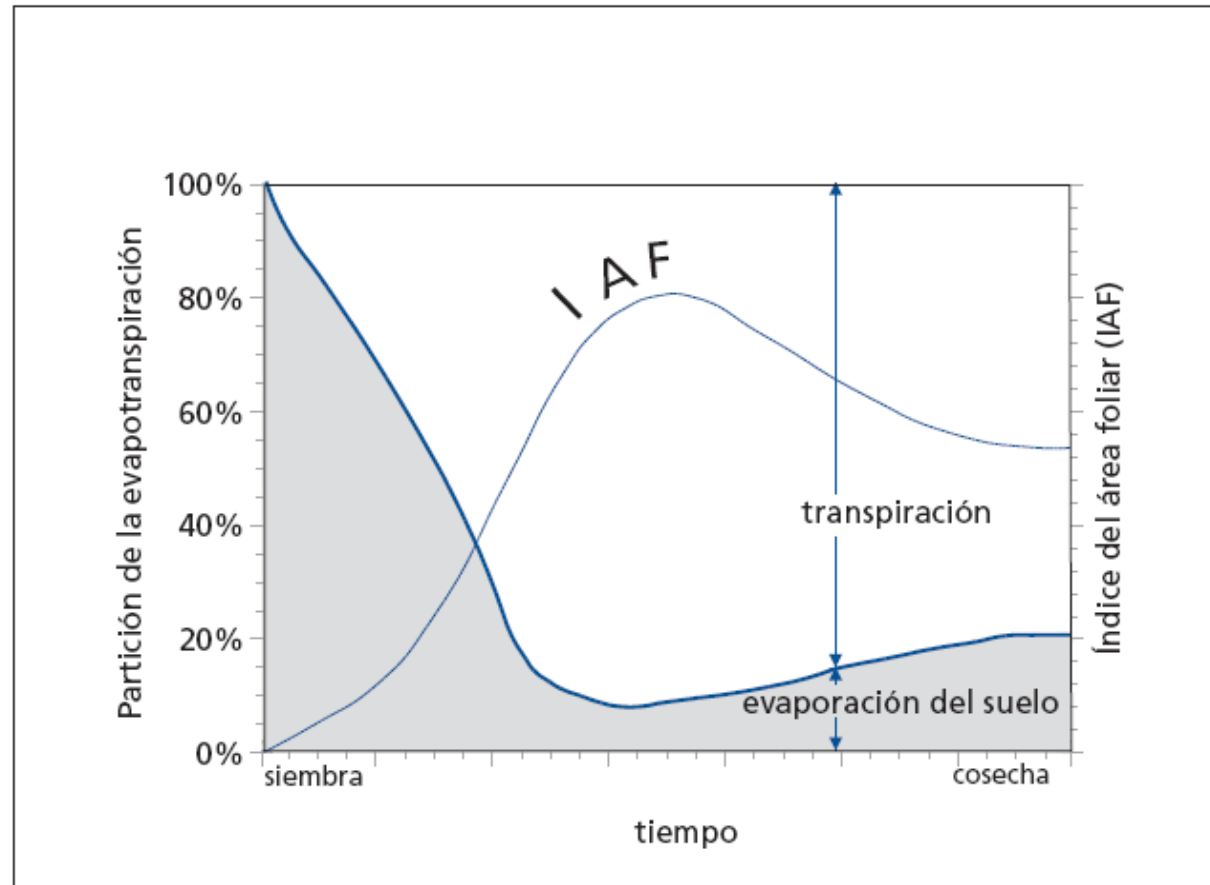


Figura 7: Divisão da evapotranspiração em transpiração e evaporação durante o crescimento de um cultivo anual- adaptada FAO 2006

Evapotranspiração

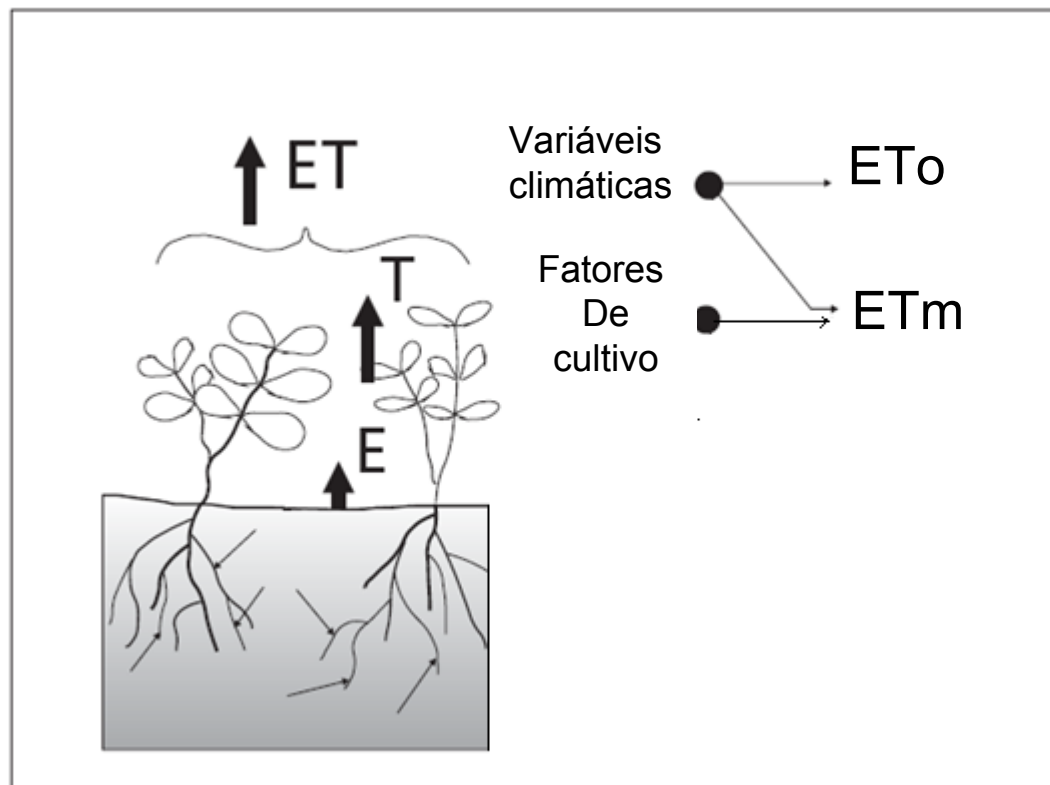


Figura 8: Fatores que afetam a evapotranspiração - adaptada FAO 2006

Evapotranspiração

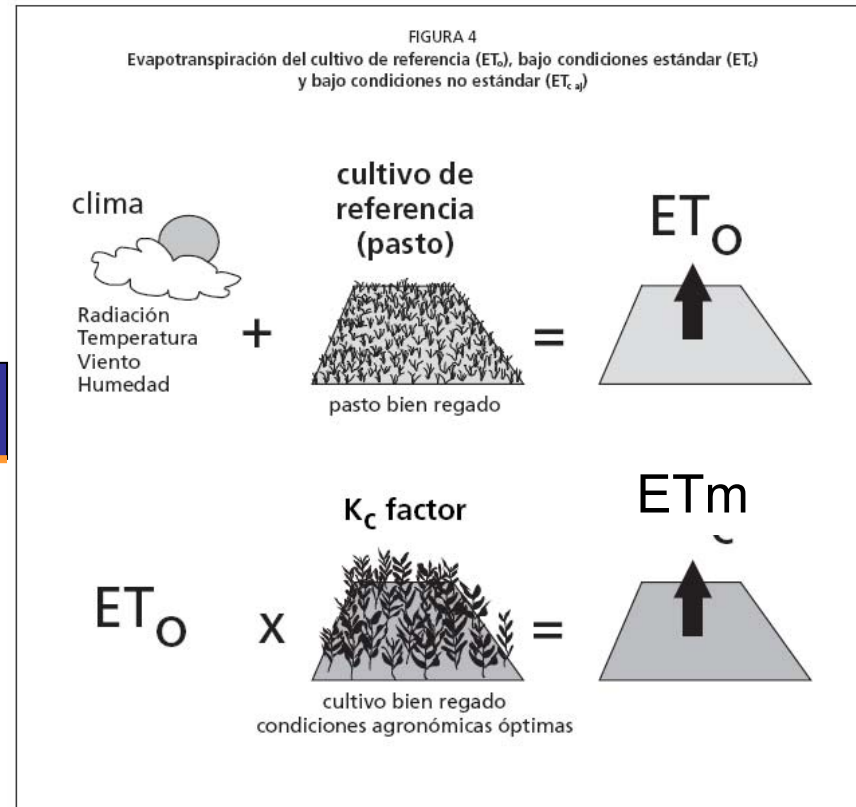


Figura 9: Evapotranspiração do cultivo de referência (ET_0) e do cultivo sob condições normais (ET_m)- adaptada FAO 2006

Evapotranspiração

- Evapotranspiração real (**ET_r**): É aquela que ocorre numa superfície vegetada, independente de sua área, de seu porte e das condições de umidade do solo. Calculados pelo método de THORNTHWAITE & MATHER (1955).
- Evapotranspiração de referência (**ET_o**): É a taxa de evapotranspiração de uma superfície de referência, que ocorre sem restrições de água, a superfície de referência corresponde a um cultivo hipotético de pasto com características específicas. Observados e utilizando Penman.
- Evapotranspiração máxima (**ET_m**): É aquela que ocorre em qualquer fase do ciclo de desenvolvimento de uma cultura, sem deficiência de água no solo. Sendo que ET_m pode ser relacionada com a ET_o, através de um coeficiente denominado coeficiente de cultura (**K_c**), definido como:

$$K_c = ET_m / ET_o$$

Evapotranspiração

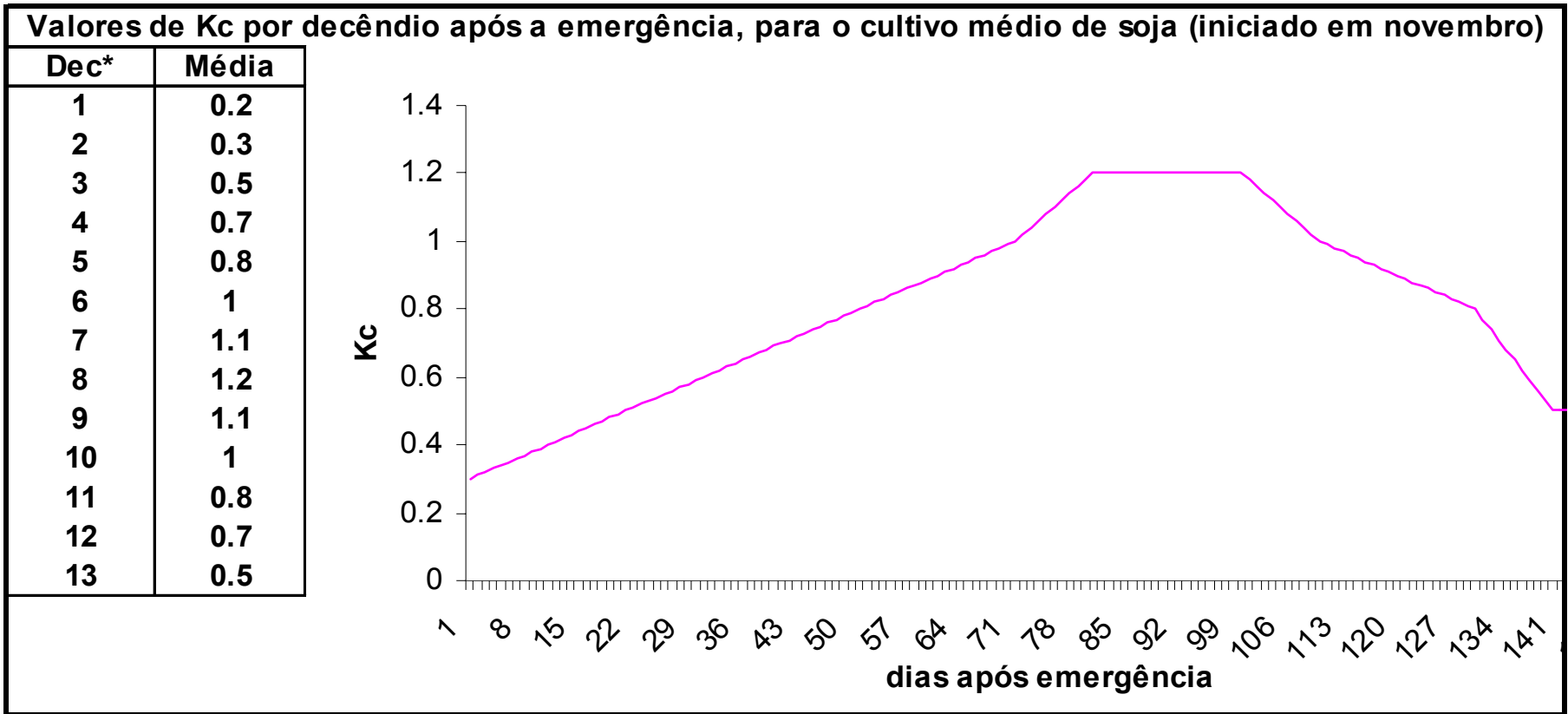


Figura 10: Coeficiente de cultura em relação ao tempo

Modelo de Jensen

Modelo de Jensen

O modelo de predição de rendimento segundo os quatro períodos fenológicos em que foi dividido de desenvolvimento da soja é representado pela seguinte equação

$$\frac{Y}{Y_m} = \left(\frac{ETr}{ETm} \right)_1^{\lambda_1} \times \left(\frac{ETr}{ETm} \right)_2^{\lambda_2} \times \left(\frac{ETr}{ETm} \right)_3^{\lambda_3} \times \left(\frac{ETr}{ETm} \right)_4^{\lambda_4}$$

Y: Rendimento esperado de grãos de soja em um determinado ano agrícola (kg ha⁻¹);

Y_m: Rendimento máximo de grãos de soja obtido na região (kg ha⁻¹) do ano anterior;

ETr: Evapotranspiração real (mm) no período considerado;

ETm: Evapotranspiração da soja (mm) no período considerado;

λ_i : os coeficientes que representam a sensibilidade relativa da planta ao déficit hídrico segundo o período (i) considerado.

Ref: Jensen (1968), Nishijima, T. (2004)

Modelo de Jensen

$$\text{Log}(Y / Y_m) = \lambda_1 \log(ETr / ETm)_1 + \lambda_2 \log(ETr / ETm)_2 + \lambda_3 \log(ETr / ETm)_3 + \lambda_4 \log(ETr / ETm)_4$$

Trabalhando com o logaritmo do modelo proposto por Jensen podemos estimar os λ_i 's do modelo.

	Estimativas para alguns anos e respectivos (erro padrão)		
	2003	2004	2005
Intercepto	0,113 (0,18)	-0,229 (0,25)	0,159 (0,21)
lambda1	-0,394 (0,12)	0,011 (0,25)	-0,099 (0,19)
lambda2	0,376 (0,18)	0,372 (0,58)	-0,110 (0,23)
lambda3	-0,230 (0,33)	-0,396 (0,24)	-0,095 (0,14)
lambda4	0,178 (0,24)	0,296 (0,14)	0,281 (0,10)

Modelo de Jensen Modificado

- Incorporando os agrupamentos de cada período fenológico adotado, passamos a trabalhar com o seguinte modelo:

$$\log\left(\frac{Y}{Y_m}\right) = \beta + \lambda_1 \log\left(\frac{ETr}{ETm}\right)_1 + \lambda_2 \log\left(\frac{ETr}{ETm}\right)_2 + \lambda_3 \log\left(\frac{ETr}{ETm}\right)_3 + \lambda_4 \log\left(\frac{ETr}{ETm}\right)_4 + A_1 + A_2 + A_3 + A_4 + \varepsilon$$

Em que A_i representa o agrupamento do i -ésimo período com $i = 1, 2, 3$ e 4 .

Modelo de Jensen Modificado

Ajuste Ano 2004

Agrupamento
Período 2

Agrupamento
Período 4

```
Call:
glm(formula = PREVRen2004 ~ X12004 + X22004 + X32004 + X42004 +
     A22004 + A42004)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.787e-01 -2.295e-03 -5.551e-17  8.818e-03  1.270e-01

Coefficients: (2 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.41868    0.19047  -2.198  0.07028 .
X12004       0.11083    0.14845   0.747  0.48353
X22004      -1.97742    0.54501  -3.628  0.01099 *
X32004       0.32755    0.22040   1.486  0.18779
X42004       0.33383    0.13412   2.489  0.04722 *
A220042     -0.48962    0.19395  -2.524  0.04502 *
A220043     -0.46450    0.16596  -2.799  0.03121 *
A220044     -0.17805    0.17147  -1.038  0.33910
A220045      0.41912    0.19502   2.149  0.07520 .
A220046     -0.49692    0.21570  -2.304  0.06079 .
A220047      0.04390    0.18350   0.239  0.81889
A220048      0.42458    0.17804   2.385  0.05441 .
A220049     -0.09323    0.36914  -0.253  0.80904
A2200410     0.69269    0.15637   4.430  0.00442 **
A420042      1.22872    0.21944   5.599  0.00138 **
A420043      0.35065    0.15988   2.193  0.07076 .
A420044      0.81481    0.18999   4.289  0.00516 **
A420045      0.34329    0.14578   2.355  0.05668 .
A420046     -0.30907    0.14919  -2.072  0.08369 .
A420047           NA           NA       NA       NA
A420048           NA           NA       NA       NA
A420049      0.35137    0.17797   1.974  0.09578 .
A4200410     0.83722    0.23543   3.556  0.01198 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for gaussian family taken to be 0.01517628)
Null deviance: 2.477994 on 26 degrees of freedom
Residual deviance: 0.091058 on 6 degrees of freedom
AIC: -33.064
```

> A42004

```
[1] 8 1 3 1 1 1 1 1 1 9 10 5 5 3 10 7 4 10 2 3 1 10 1 1 3  
[26] 5 9 5 5 3 1 6 3 5 1 3 5
```

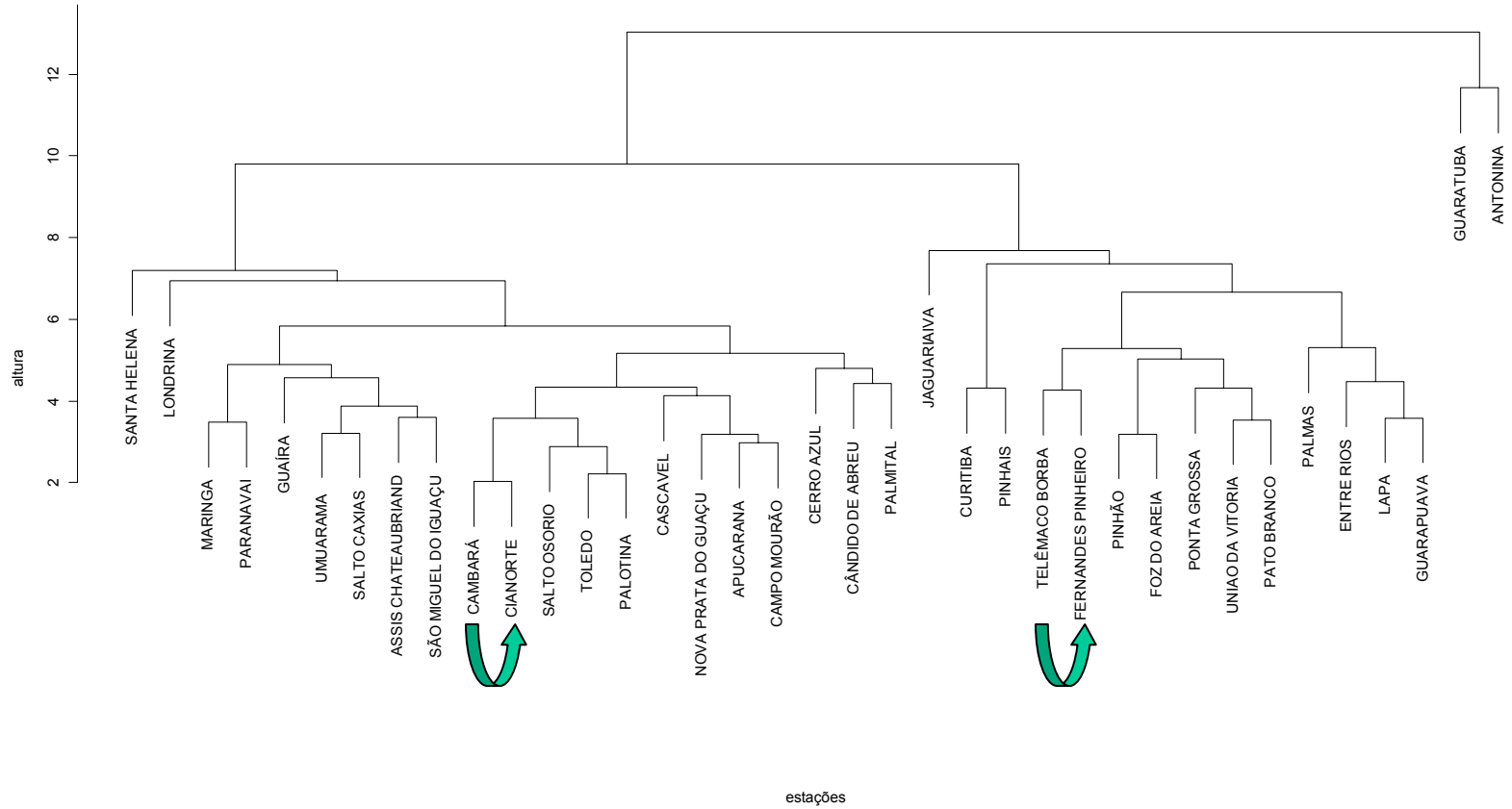


Figura 11: Dendrograma do Agrupamento relacionado ao ano 2004

Modelo de Jensen Modificado

```
glm(formula = PREVRen2004 ~ X12004 + X22004 + X32004 + X42004 +  
A22004 + A42M2004)
```

Ajuste Ano 2004

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.787e-01	-2.295e-03	-5.551e-17	8.818e-03	1.270e-01

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-0.41868	0.19047	-2.198	0.07028	.
X12004	0.11083	0.14845	0.747	0.48353	
X22004	-1.97742	0.54501	-3.628	0.01099	*
X32004	0.32755	0.22040	1.486	0.18779	
X42004	0.33383	0.13412	2.489	0.04722	*
A220042	-0.48962	0.19395	-2.524	0.04502	*
A220043	-0.46450	0.16596	-2.799	0.03121	*
A220044	-0.17805	0.17147	-1.038	0.33910	
A220045	0.41912	0.19502	2.149	0.07520	.
A220046	-0.49692	0.21570	-2.304	0.06079	.
A220047	0.04390	0.18350	0.239	0.81889	
A220048	-0.41264	0.28862	-1.430	0.20274	
A220049	-0.09323	0.36914	-0.253	0.80904	
A2200410	0.69269	0.15637	4.430	0.00442	**
A42M20042	1.22872	0.21944	5.599	0.00138	**
A42M20043	0.35065	0.15988	2.193	0.07076	.
A42M20044	0.81481	0.18999	4.289	0.00516	**
A42M20045	0.34329	0.14578	2.355	0.05668	.
A42M20046	-0.30907	0.14919	-2.072	0.08369	.
A42M20049	0.35137	0.17797	1.974	0.09578	.
A42M200410	0.83722	0.23543	3.556	0.01198	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

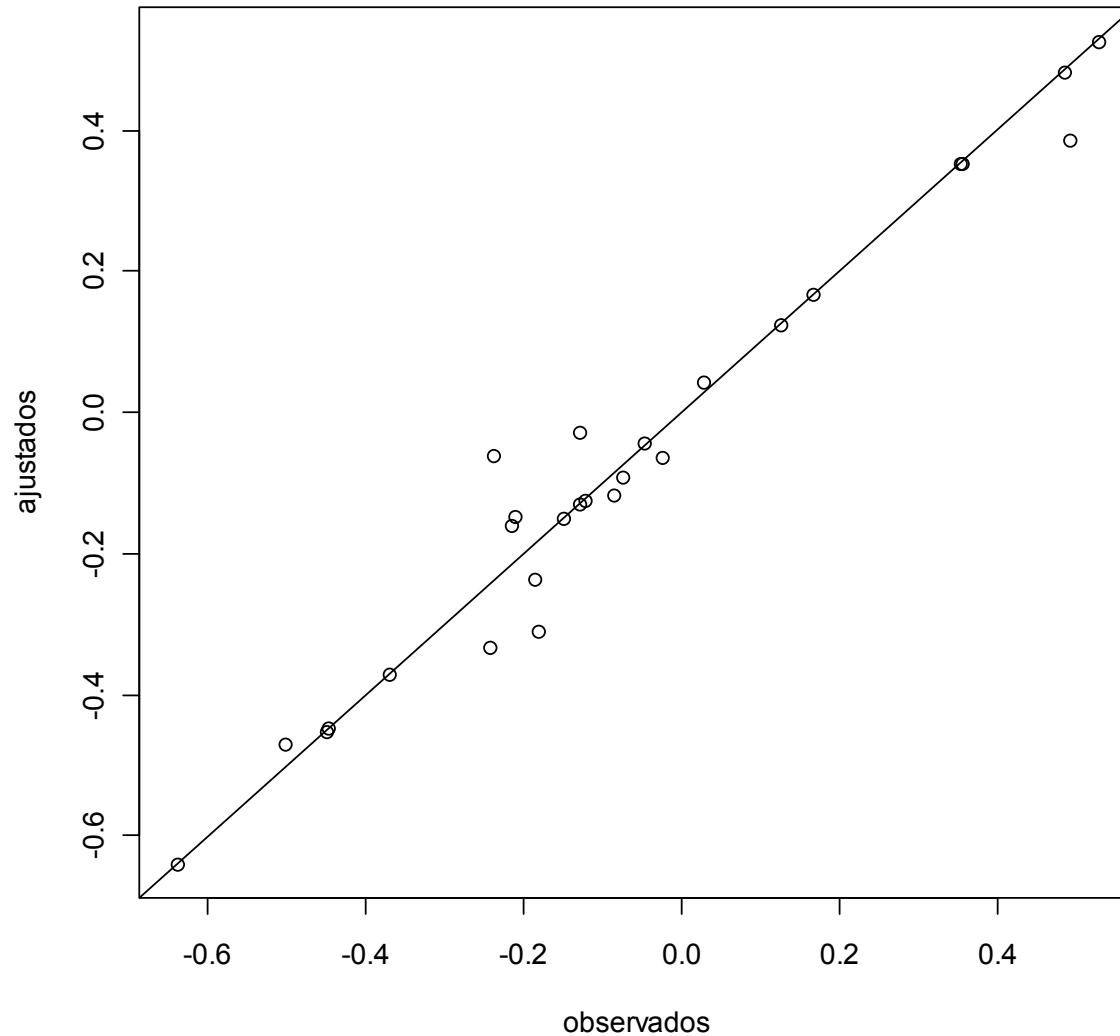
(Dispersion parameter for gaussian family taken to be 0.01517628)

Null deviance: 2.477994 on 26 degrees of freedom
Residual deviance: 0.091058 on 6 degrees of freedom
AIC: -33.064

Modelo de Jensen Modificado

Ajuste Ano 2004

Figura 12: Observados Vs Ajustados



Modelo de Jensen Modificado

Ajuste Ano 2004

- ✓ Coeficientes de Sensibilidade Hídrica
- ✓ Coeficientes do Agrupamento

$$\left(\frac{\hat{Y}}{Ym}\right)_{2005} = \exp \hat{\beta} \times \exp\left(\frac{ETr}{ETm}\right)_{1,2005}^{\hat{\lambda}_1} \times \exp\left(\frac{ETr}{ETm}\right)_{2,2005}^{\hat{\lambda}_2} \times \exp\left(\frac{ETr}{ETm}\right)_{3,2005}^{\hat{\lambda}_3} \\ \times \exp\left(\frac{ETr}{ETm}\right)_{4,2005}^{\hat{\lambda}_4} \times \exp(A_{2,2004} + A_{4,2004})$$

Estimativa para o Ano 2005

versus

Ajuste Ano 2005

Modelo de Jensen Modificado

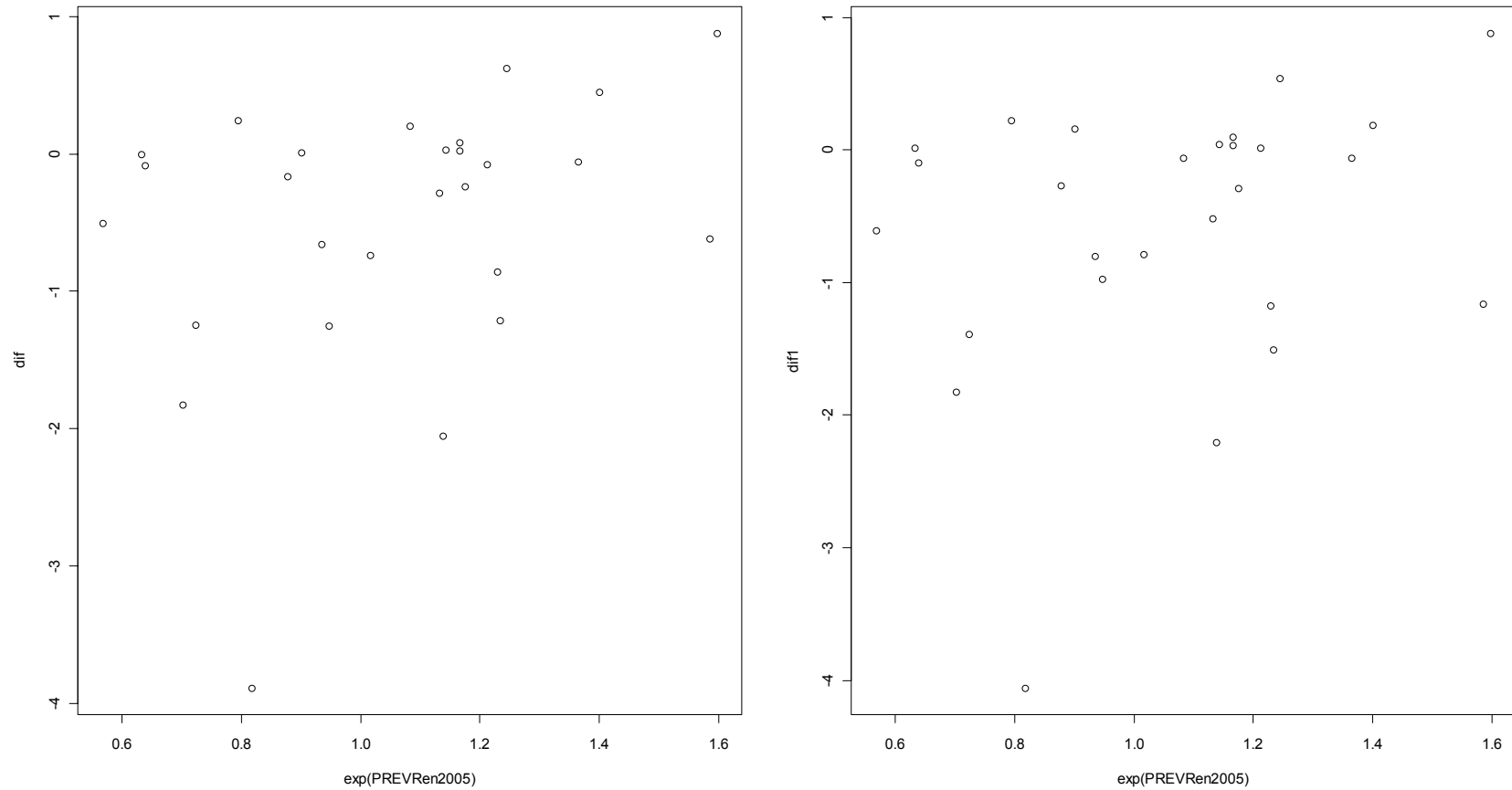
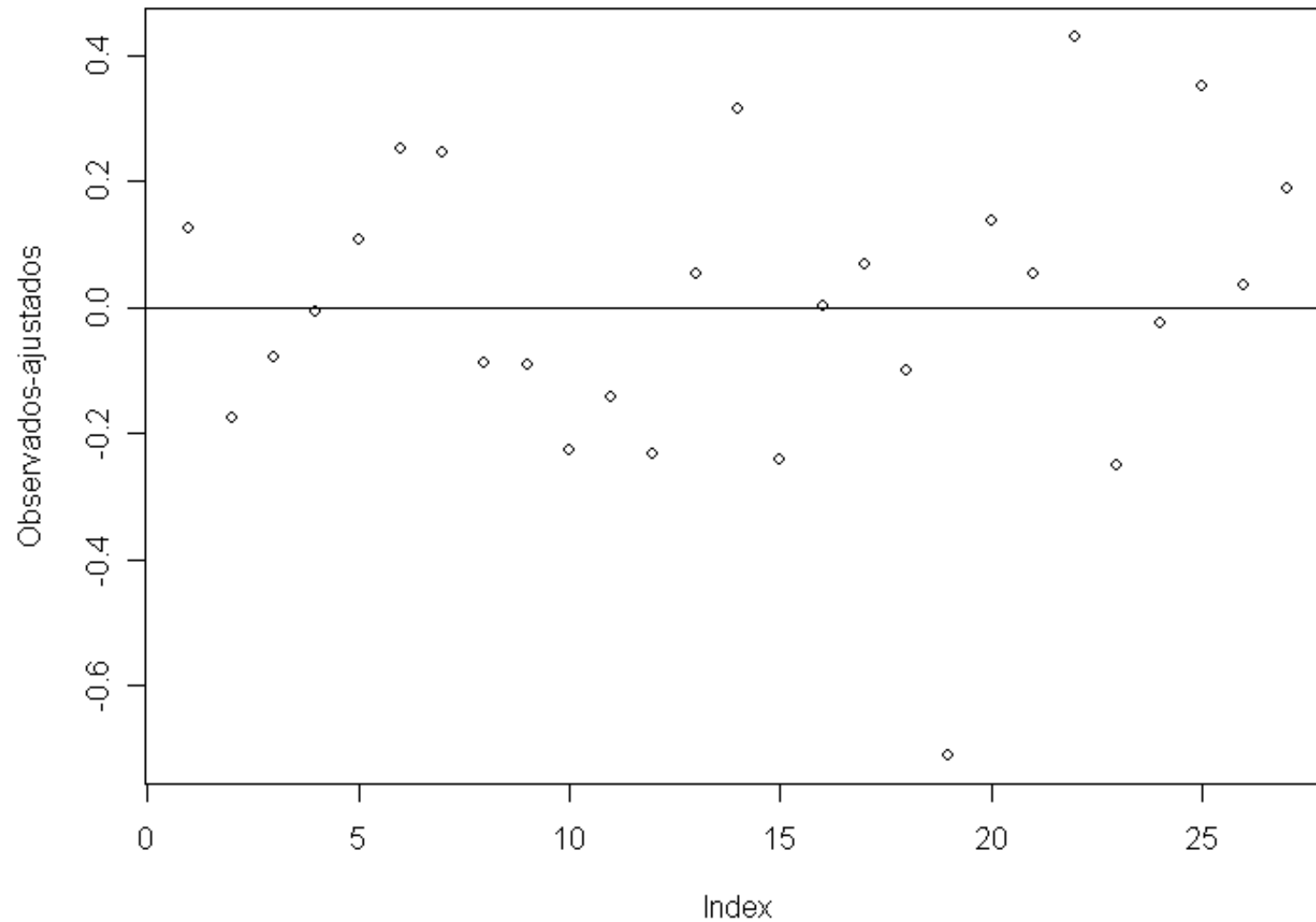


Figura 13: Valores observados de rendimento versus as diferenças estimadas

Modelo de Jensen para o Ano 2005

Figura 14: Resíduos usuais



Conclusões

- Os coeficientes de sensibilidade hídrica do modelo de Jensen variam de ano para ano.
- A variação da informação indica que os agrupamentos segundo as variáveis meteorológicas também mudam de ano para ano.
- O modelo de Jensen modificado estabelece a significância da incorporação de variáveis indicadoras dos grupos.
- Se apresentaram problemas de singularidade e foi necessário o estudo do agrupamento e o estabelecimento de um novo agrupamento.

Próximos Passos

- Estudar a relação entre os agrupamentos de anos consecutivos para que a incorporação de variáveis indicadoras dos grupos seja feita de maneira correta no modelo de Jensen.
- Trabalhar com o rendimento máximo observado, referente ao agrupamento.
- Definir critérios de qualidade da previsão.
- ???

Referências

- [1] Freitas, E. (2007). Variação, comparação e escolha de agrupamentos: uma proposta utilizando a entropia. Dissertação de Mestrado IME-USP. Orientação Viviana Giampaoli.
- [2] COSTA, J. A. **Cultura da soja**. Porto Alegre: I. Manica, J. A. Costa, 1996.
- [3] THORNTHWAITE, C. W.; MATHER, J. R. **The water balance**. Publications in Climatology, New Jersey, v. 1, n. 8, p. 1-104, 1955.
- [4] JENSEN, M. E. Water consumption by agricultural plants. In: KOZLOWSKY, T. T. (ed.). **Water deficits and plant growth**, New York: Academic Press, p. 1-22, 1968.
- [5] Serie de Riego y Drenaje de la FAO No56.
- [6] Nishijima, T. (2004). Modelagem markoviana da precipitação pluvial diária e simulação do rendimento esperado de soja no município de cruz alta (RS). Tese de doutorado UFRS. Orientação André L. L. da Silveria.

Obrigado!