

Dados Solar

Author ^a

Abstract

In this paper,

Keywords: text ...

1. Introdução

text

2. Dados

Considere que a potência gerada por placas fotovoltaicas limpas, denotada por W , sejam monitoradas por n dias, $n \geq 1$. Suponha que para o j -ésimo dia sejam feitas $(k_j + 1)$ medições das potências geradas, para $j = 1, \dots, n$. Assim, W_{ij} é a potência gerada no i -ésimo instante de tempo do dia j , para $i = 0, \dots, k_j$ e $j = 1, \dots, n$. Assuma como instante inicial de medição do dia j , $i = 0$, o horário em que a primeira medida de potência é gerada, $W_{0j} > 0$. A partir do instante da primeira medição, registramos as potências geradas em intervalos de $h = 10$ minutos, até o instante $i = k_j$, onde k_j é o último horário de registro antes da placa deixar de gerar potência, $W_{(k_j+1),j} = 0$.

Como ilustração dos dados gerados, a Figura 1 mostra as potências geradas nos dias 1 e 2 de realização do experimento. No dia 1, a primeira medição foi feita as 5 horas e 29 minutos e a última medição foi feita as 17 horas e 39 minutos, totalizando $k_1 = 74$ medições. Para o dia 2, a primeira medição foi feita as 5 horas e 29 minutos e a última medição foi feita as 17 horas e 49 minutos, totalizando $k_2 = 75$ medições. No total, o experimento foi realizado durante $n = 21$ dias (três semanas).

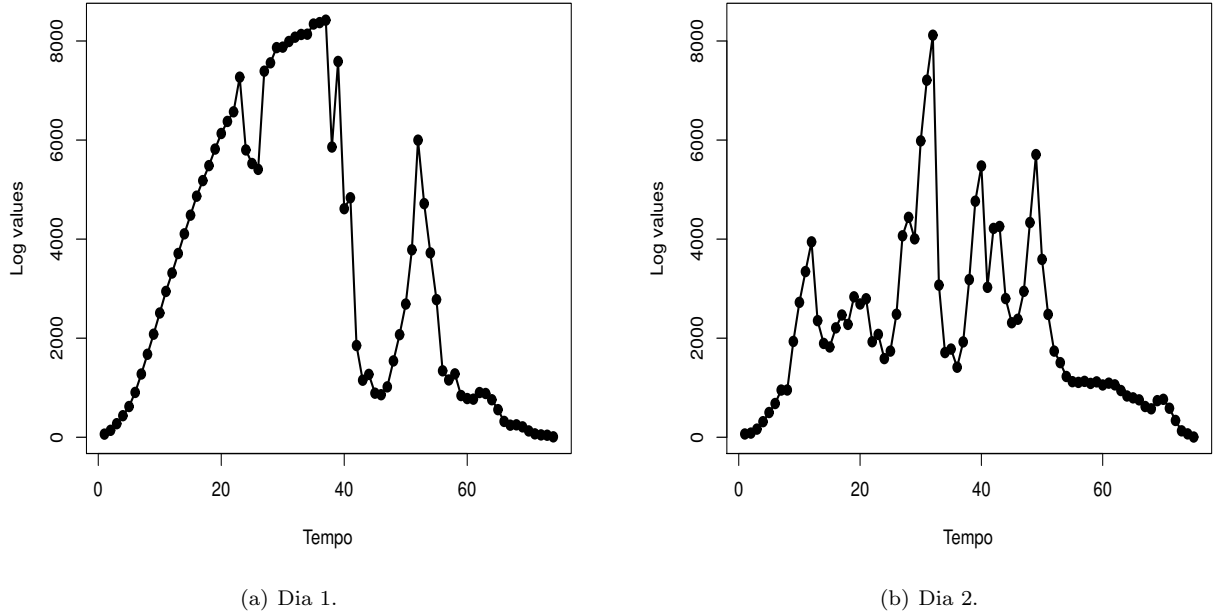


Figure 1: Potências geradas nos dias 1 e 2.

3. Modelos

Como podemos observar na Figura 1, as potências geradas apresentam um comportamento um tanto instável ao longo tempo. Isto ocorre devido a fatores ambientais, tais como, temperatura, irradiação, quantidade de sujeiras nas placas, entre outras.

Com o objetivo de obter uma descrição mais estável das potências geradas, considere X_{ij} a potência registrada de maneira acumulada até o i -ésimo instante de tempo do dia j , para $i = 0, \dots, k_j$ e $j = 1, \dots, n$. Isto é,

$$X_{0j} = W_{0j} \quad \text{e} \quad X_{ij} = \sum_{i'=0}^i W_{i'j},$$

para $i = 0, 1, \dots, k_k$ e $j = 1, \dots, n$

A Figura 2 mostra os gráficos das potências registradas nos dias 1 e 2, de maneira acumulada. Note que, utilizando os valores acumulados, obtemos um comportamento mais estável dos dados. Além disso, a maneira

como os dados se comportam ao longo do tempo sugere algum tipo de modelo matemático, em que, o gráfico é uma **curva de crescimento do tipo sigmoidal (em forma de S)**. Dois modelos matemáticos que apresentam esta característica são os modelos do **tipo Logístico e do tipo Gompertz**.

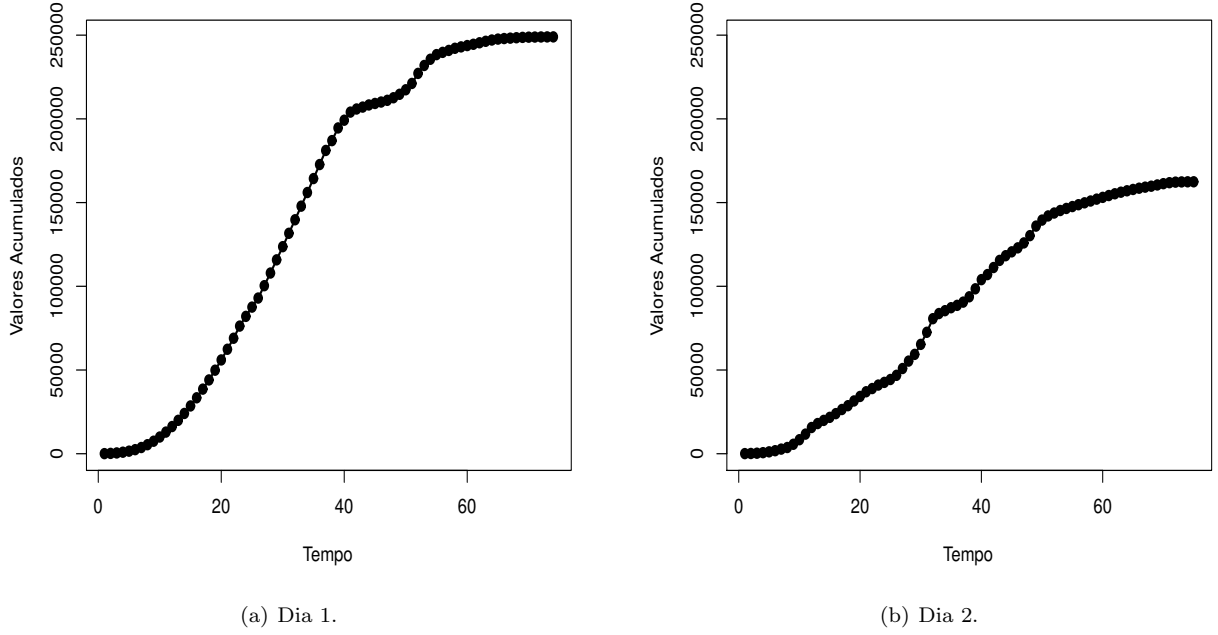


Figure 2: Potências geradas nos dias 1 e 2.

Contudo, antes de descrevermos os modelos considere a transformação logarítmica para os valores das potências acumuladas. Esta transformação é feita com o objetivo de obter valores em uma escala numérica que seja matematicamente e computacionalmente mais adequada para se desenvolver a modelagem.

Assim, considere **Seja $Y_{ij} = \log(X_{ij})$, i.e.**, logaritmo da potência acumulada até o i -ésimo instante do dia j , para $i = 0, 1, \dots, k_j$ e $j = 1, \dots, n$. A Figura 3 mostra os gráficos da Figura 2 nas escala logarítmica.

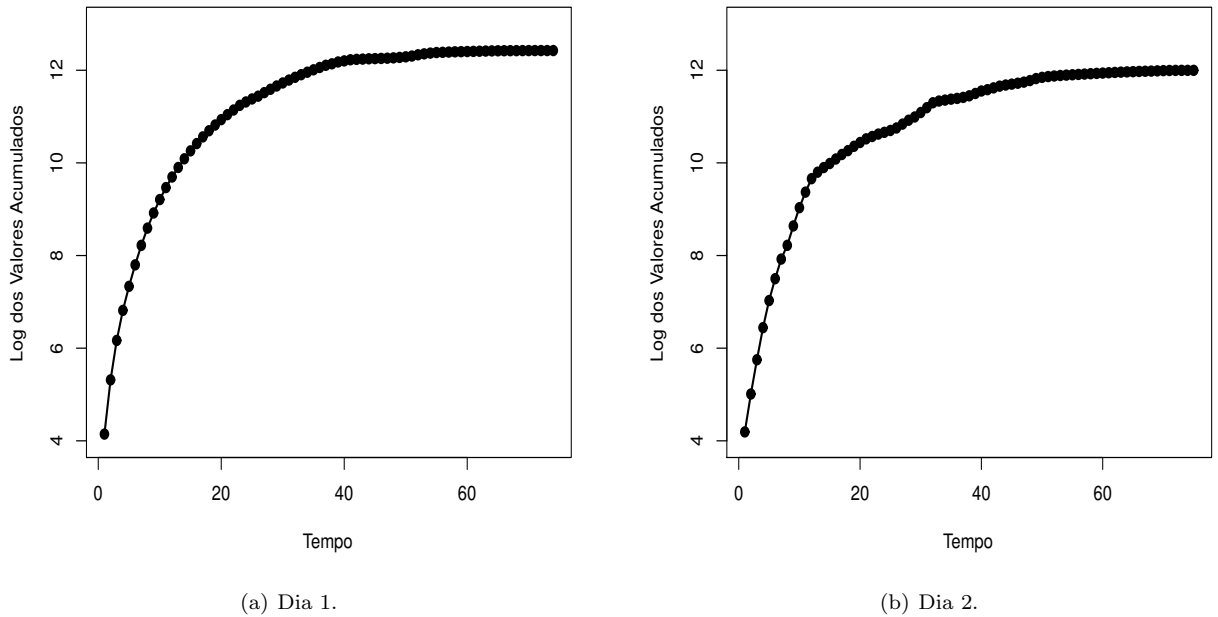


Figure 3: Logaritmo das potências acumuladas, dias 1 e 2.

3.1. Modelos Log-Logístico e Log-Gompertz

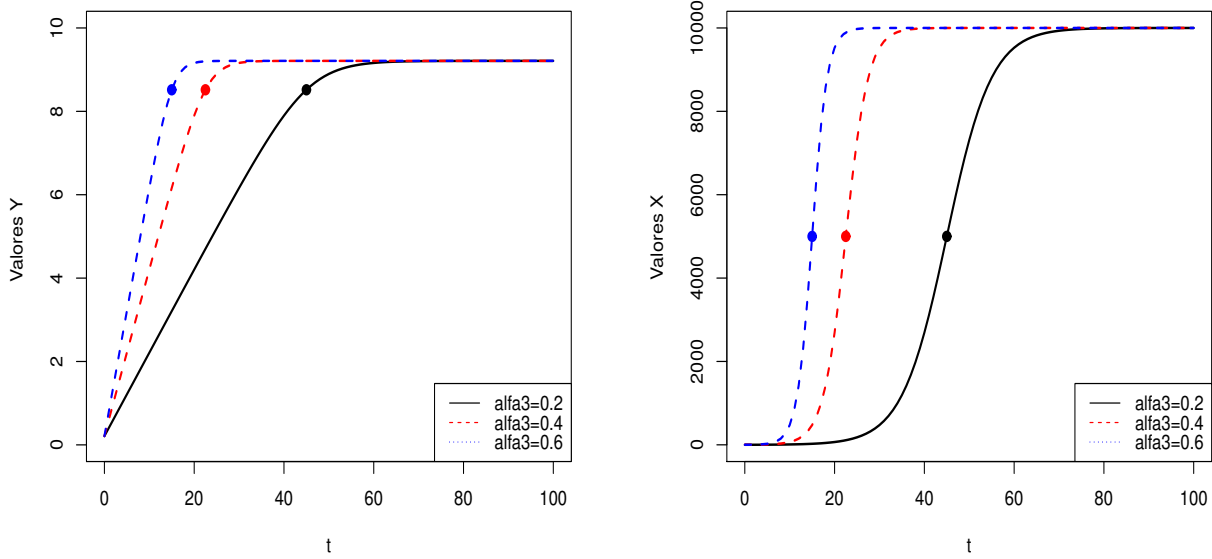
Considere que os valores Y_{ij} sejam gerados de acordo com o modelo Log-Logístico de parâmetro $\theta = (\alpha, \beta_0, \beta_1)$, i.e.,

$$Y_{ij} = \log(\alpha) - \log(1 + \exp\{\beta_0 - \beta_1 t_{ij}\}) + \varepsilon_{ij}, \quad (1)$$

onde ε_{ij} é um erro aleatório com média 0 e variância σ^2 , para $i = 0, 1, \dots, k_j$ e $j = 1, \dots, n$. Neste modelo, o valor de α representa assintota superior do modelo na escala original. Ou seja, este valor representa a estimativa para o total de potência gerada em um dia. O parâmetro β_1 representa a taxa média de crescimento (afeta a inclinação da curva). O parâmetro β_0 está associado as coordenadas do ponto de inflexão da curva através das expressões $T = \frac{\beta_0}{\beta_1}$ e $y_T = \frac{\alpha}{2}$, onde T é o tempo que irá ocorrer a mudança de comportamento da curva. Denominamos o modelo da Equação 1 de Modelo ML_1 .

A Figura 4(a) mostra o gráfico do modelo Log-logístico para $\alpha_1 = 10.000$, $\alpha_2 = 9$ e $\alpha_3 = \{0.20, 0.40, 0.60\}$ sem o erro aleatório ε_{ij} , para $i = 0, 1, \dots, k_j$ e $j = 1, \dots, n$. A Figura 4(b) mostra o mesmo gráfico na escala original, i.e., $X_{i,j} = \exp\{Y_{ij}\}$, para $i = 0, 1, \dots, k_j$ e $j = 1, \dots, n$. Os símbolos \bullet nos gráficos são os pontos de inflexão.

Como podemos notar Figura 4(b) o gráfico é simétricos em relação ao ponto de inflexão das curvas. Aumentando o valor de β_1 (taxa de crescimento) e mantendo α e β_0 fixos, mais rápido é o crescimento da curva. Além disso, note que, os gráficos da Figura 4 apresentam o formato desejado para modelagem dos dados apresentados nos gráficos das Figuras 3 e 2.



(a) Modelo Log-Logístico.

(b) Modelo Logístico.

Figure 4: Gráficos dos modelos Log-logístico e Logístico.

Como segundo modelo, considere que os valores Y_{ij} sejam gerados de acordo com o modelo Log-Gompertz de parâmetro $\theta = (\alpha, \beta_0, \beta_1)$, i.e.,

$$Y_{ij} = \log(\alpha) - \exp\{\beta_0 - \beta_1 t_{ij}\} + \varepsilon_{ij}, \quad (2)$$

onde α_1 , α_2 e α_3 são os parâmetros do modelo. De maneira similar ao modelo Logístico, no modelo Gompertz o valor do parâmetro α_1 é a assintota superior. O parâmetro β é a taxa média de crescimento e o parâmetro β_0 determina as coordenadas do ponto de inflexão através das expressões $T = \frac{\beta_0}{\beta_1}$ e $y_T = \frac{\alpha}{e}$. Denominamos este modelo de MG_1 .

A Figura 5(a) mostra o gráfico do modelo Log-logístico para $\alpha_1 = 10.000$, $\alpha_2 = 2.2$ e $\alpha_3 = \{0.05, 0.10, 0.25\}$ sem o erro aleatório ε_{ij} , para $i = 0, 1, \dots, k_j$ e $j = 1, \dots, n$. A Figura 5(b) mostra o mesmo gráfico na escala original, i.e., $X_{i,j} = \exp\{Y_{ij}\}$, para $i = 0, 1, \dots, k_j$ e $j = 1, \dots, n$. Os símbolos \bullet nos gráficos são os pontos de inflexão. Aumentando o valor de β_1 e mantendo α e β_0 fixos, mais rápido é o crescimento da curva.

A principal diferença entre o modelo Logístico e o modelo Gompertz é que o modelo Gompertz não é simétrico em relação ao ponto de inflexão, ver Figura 5(b). O valor da ordena do ponto de inflexão do modelo Logístico é sempre maior do que o do modelo Gompertz. No modelo Logístico o valor é igual a metade do valor α , $\frac{\alpha_1}{2}$, e no modelo Gompertz é 36,79% do valor α .

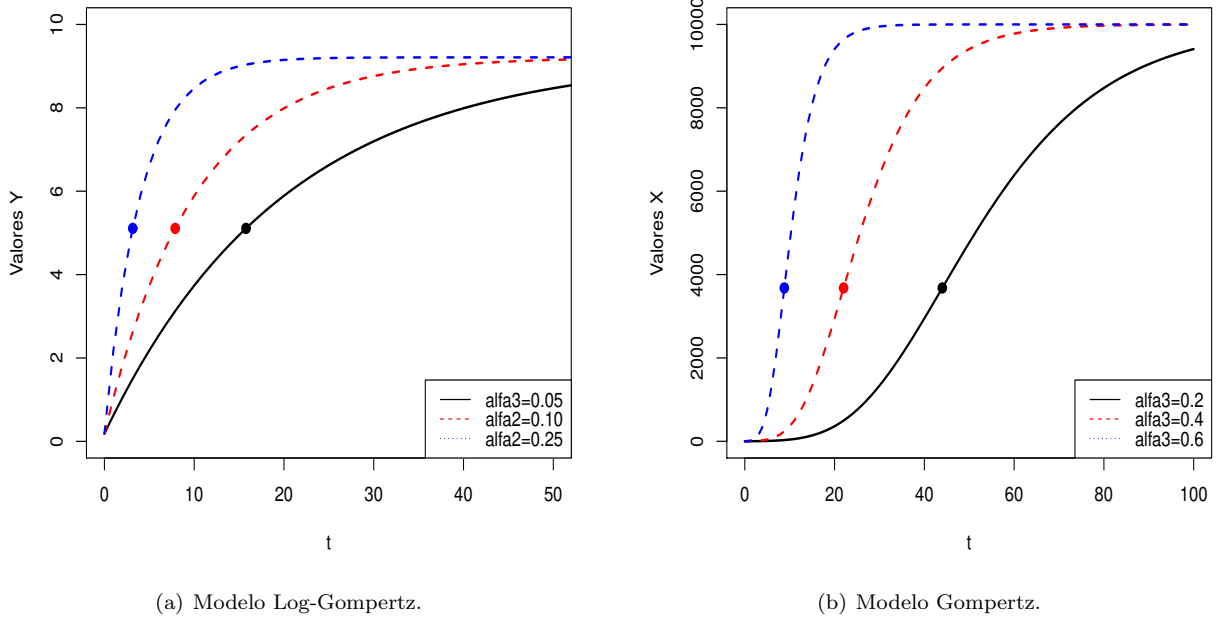


Figure 5: Gráficos dos Modelos Log-Gompertz e Gompertz.

3.2. Modelo Hierárquicos

Considere agora o interesse em relacionar Y_{ij} com a variável tempo (T) e as seguintes variáveis ambientais:

- I = Irradiancia acumulada;
- M = Massa acumulada de particulados.

Para mantermos um escala adequada para todas as variáveis, também consideramos a transformação logarítmica para os valores observados das variáveis I e M .

Utilizando o modelo Logístico dado na Equação (1), assuma a seguinte relação hierárquica entre a variável resposta Y e as variáveis explicativas

$$Y_{ij} = \log(\alpha_1) - \log(1 + \exp\{\beta_{ij}^* - \beta_4 T_{ij}\}) + \varepsilon_{ij} \quad (3)$$

$$\beta_{ij}^* = \beta_0 + \beta_1 I_{ij} + \beta_2 M_{ij} + u_{ij}, \quad (4)$$

onde, ε_{ij} e u_{ij} são erros aleatórios com a suposição de média igual a 0 e variâncias σ^2 e σ_u^2 , respectivamente. Além disso, também assumimos que ε_{ij} e u_{ij} são independentes, para $i = 0, 1, \dots, k_j$ e $j = 1, \dots, n$.

Das expressões em (3) e (4), temos que

$$Y_{ij} = \log(\alpha_1) - \log(1 + \exp\{\beta_0 + \beta_1 I_{ij} + \beta_2 M_{ij} - \beta_3 T_{ij} + u_{ij}\}) + \varepsilon_{ij} \quad (5)$$

Denominamos este modelo de ML_2 .

Como as medidas de potência gerada dentro de cada dia são feitas longitudinalmente, é plausível assumir que os erros aleatórios ε_{ij} são correlacionados. Assim, considere que ε_{ij} e $\varepsilon_{(i+s)j}$ são dois erros do j -ésimo dia separado por s unidades de tempo. **Assumindo um modelo auto-regressivo de primeira ordem (AR_1), a correlação entre os dois erros é dada por $\text{corr}(\varepsilon_{ij}, \varepsilon_{(i+s)j}) = \rho^{|s|}$, para $i = 0, 1, \dots, k_j$ e $j = 1, \dots, n$. Denominamos o modelo da Equação (5) com estrutura de correlação AR_1 de modelo ML_3 .**

Considere agora o modelo Log-Gompertz dado na Equação (2) com a seguinte relação hierárquica entre a variável resposta Y e as variáveis explicativas

$$Y_{ij} = \log(\alpha_1) - \exp\{\beta_{ij} - \beta_4 T_{ij}\} + \varepsilon_{ij} \quad (6)$$

$$\beta_{ij}^* = \beta_0 + \beta_1 I_{ij} + \beta_2 M_{ij} + u_{ij}, \quad (7)$$

onde, ε_{ij} e u_{ij} são erros aleatórios com a suposição de média igual a 0 e variâncias σ^2 e σ_u^2 , respectivamente. Além disso, também assumimos que ε_{ij} e u_{ij} são independentes, para $i = 0, 1, \dots, k_j$ e $j = 1, \dots, n$.

Das expressões em (6) e (7), temos que

$$Y_{ij} = \log(\alpha_1) - \exp\{\beta_0 + \beta_1 I_{ij} + \beta_2 M_{ij} - \beta_3 T_{ij} + u_{ij}\} + \varepsilon_{ij} \quad (8)$$

Denominamos este modelo de MG_2 . **Além disso, considere como modelo MG_3 o modelo da expressão (8) com estrutura de correlação AR_1 .**

4. Resultados

Para estimação dos parâmetros dos modelos ML_1 e MG_1 utilizamos o software R e o comando *nlme* do pacote *nlme*. Utilizando os modelos ML_1 e ML_2 estimados, obtemos as estimativas para os parâmetros dos modelos ML_2 e MG_2 utilizando o comando *update* com a opção *fixed=c(alpha ~ 1, beta0 ~ I + M, beta1 ~ 1), random=(beta0 ~ 1)*. Para estimação dos parâmetros dos modelos ML_3 e MG_3 utilizamos os modelos ML_2 e MG_2 estimados e o comando *update* com a opção *autocorrelation = corCAR1*.

Comparamos os modelos ajustados, utilizando os critérios de seleção de modelos AIC e BIC. A Tabela 1 mostra os valores dos critérios de seleção de modelos AIC e BIC para os seis modelos. O modelo com menor valor AIC e BIC é o modelo que melhor explica os dados observados. Os valores destacados em negrito são os menores valores AIC e BIC. Ou seja, de acordo com estes três critérios, o modelo ML_3 é o melhor modelo dentre os seis modelos considerados.

Table 1: Estimativas para os efeitos fixos.

Modelo	ML_1	ML_2	ML_3	MG_1	MG_2	MG_3
AIC	1.967,92	-7.013,15	-10.209,86	-644,18	-6.063,26	-7.308,71
BIC	1.021,62	-6.975,66	-10.166,89	-590,47	-6.025,66	-7.265,75

A Tabela 2 mostra as estimativas para os parâmetros dos efeitos fixos do modelo ML_3 . Para um nível de significância $\nu = \{1\%, 5\%, 10\%\}$, temos $p\text{-valor} < \nu$, com exceção para β_2 . **Isto indica que a variável M (massa acumulada) não é significativa.** Contudo, isto era esperado pois o modelo foi ajustado utilizando medições em placas fotovoltaicas limpas. Devido a isto, optamos por manter esta variável no modelo. As estimativas para os desvios-padrão dos erros aleatórios são $\hat{\sigma} = 0,0453$ e $\hat{\sigma}_u = 9,4556 \cdot 10^{-6}$. A estimativa do coeficiente de correlação é $\hat{\rho} = 0.9778$, indicando uma correlação forte.

Table 2: Estimativas para os efeitos fixos.

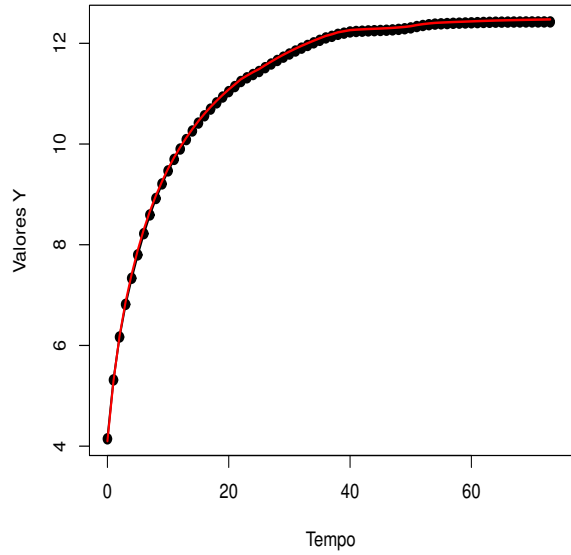
Parâmetros	Estimativa	Erro padrão	G.L.	Estatística t	p-valor
α	14.0456	0.0755	1563	185.9782	0.0000
β_0	12.0198	0.0760	1563	158.2226	0.0000
β_1	-1.0219	0.0013	1563	-814.5961	0.0000
β_2	0.0043	0.0047	1563	0.9211	0.3571
β_3	0.0026	0.0004	1563	6.2523	0.0000

O modelo ML_3 estimado é dado por

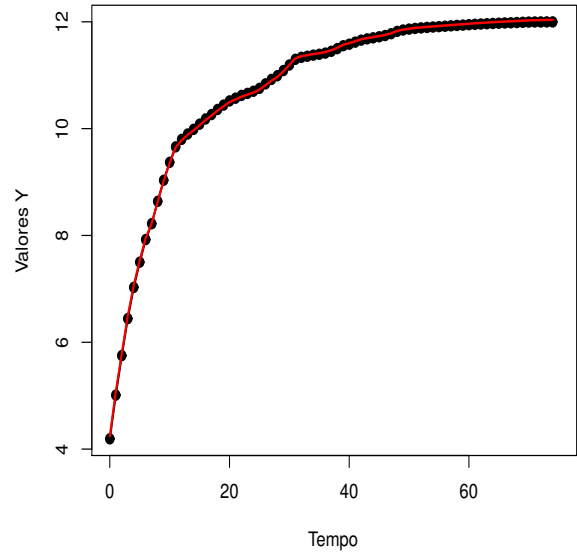
$$Y_{ij} = 14,0455 - \exp\{12,0198 - 1,0219I_{ij} + 0,0043M_{ij} - 0,0026t_{ij}\},$$

para $i = 0, 1, \dots, k_j$ e $j = 1, \dots, n$.

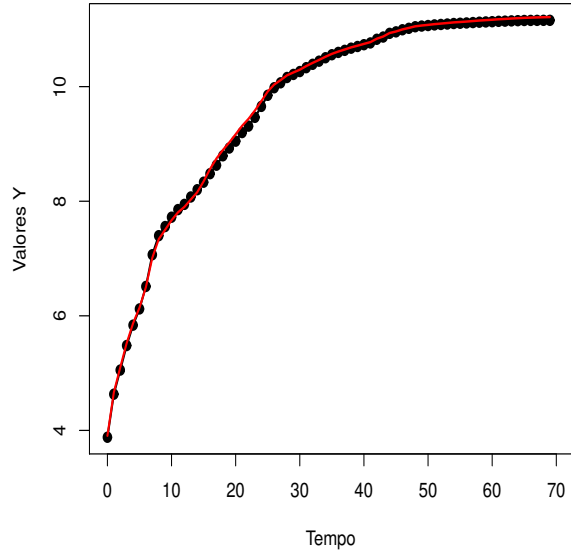
A Figura 6, mostra os valores Y 's observados nos dias 1, 2, 3 e 4 (símbolos \bullet) e o modelo ML_3 ajustado (linha contínua na cor vermelha). O erro quadrático médio (EQM) para os valores destes quatro dias são 0,0019, 0,0006, 0,0025 e 0,0057, respectivamente. Para os dados dos 21 dias, o EQM é de 0,0016.



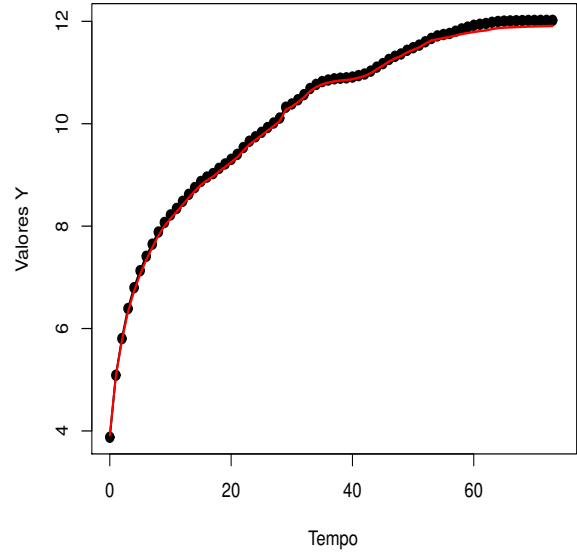
(a) Dia 1.



(b) Dia 2.



(c) Dia 3.



(d) Dia 4.

Figure 6: Gráficos dos valores observados e o valores estimados pelo modelo Log-Logístico.

References

Antoniak, C. E. (1974). Mixture of processes dirichlet with applications to bayesian nonparametric problems.
The Annals of Statistics, **2**, 1142-1174.