Dados Solar

Author [a]

Abstract

In this paper,

Keywords: text ...

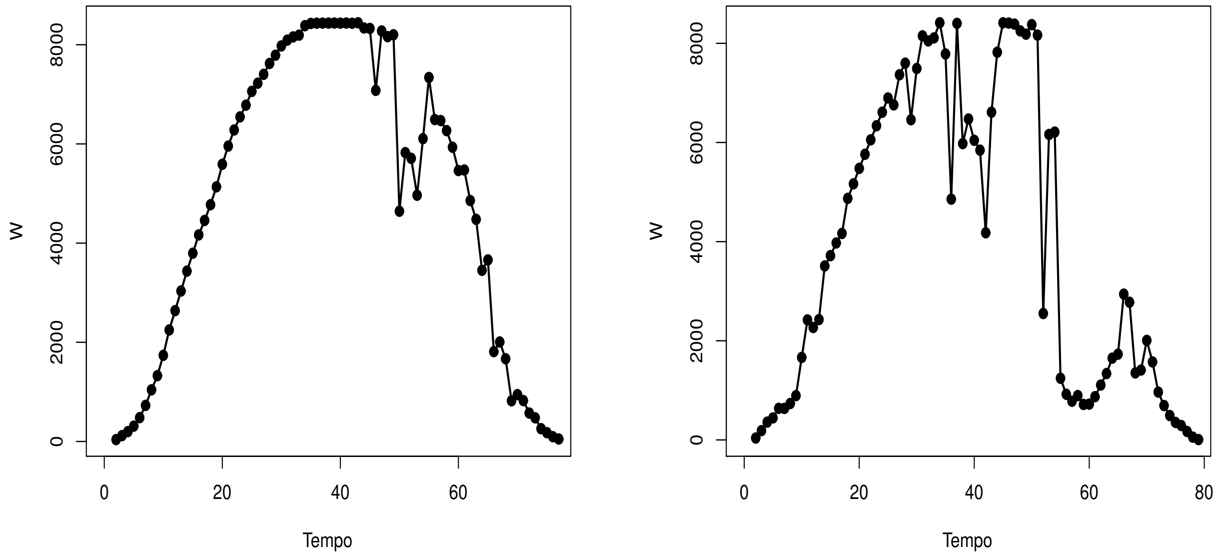
1. Introdução

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 $j = 1, \ldots, n$. Assim, W_{ij} é a potência gerada no i-ésimo instante de tempo do dia j, para $i = 0, \ldots, k_j$ e $j = 1, \ldots, n$. 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 $k_1 = 76$ medições. Para o dia 2, a primeira medição foi feita as 5 horas e 29 minutos e a ultima medição foi feita as 17 horas e 49 minutos, totalizando $k_2 = 78$ medições. No total, o experimento foi realizado durante $n = 21$ dias (três semanas).



(a) Dia 1. (b) Dia 2.

Figure 1: Potências geradas nos dias 1 e 2.

3. Modelos

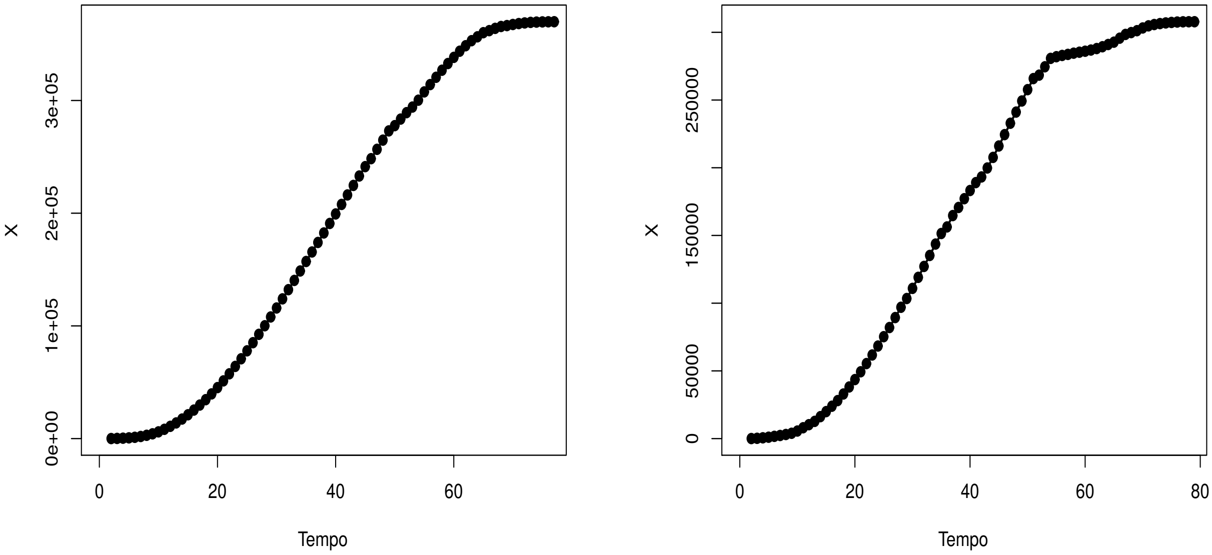
Como podemos observar na Figura 1, as potências geradas apresentam um comportamento um tanto instável ao longo tempo. Isto ocorre devido a fatores ambientais, tais como, temperatura, irradiação, quantidade de sujeiras nas placas, entre outras.

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 $i = 0, \ldots, k_j$ e $j = 1, \ldots, n$. Isto é,

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

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

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 Logístico e Gompertz.



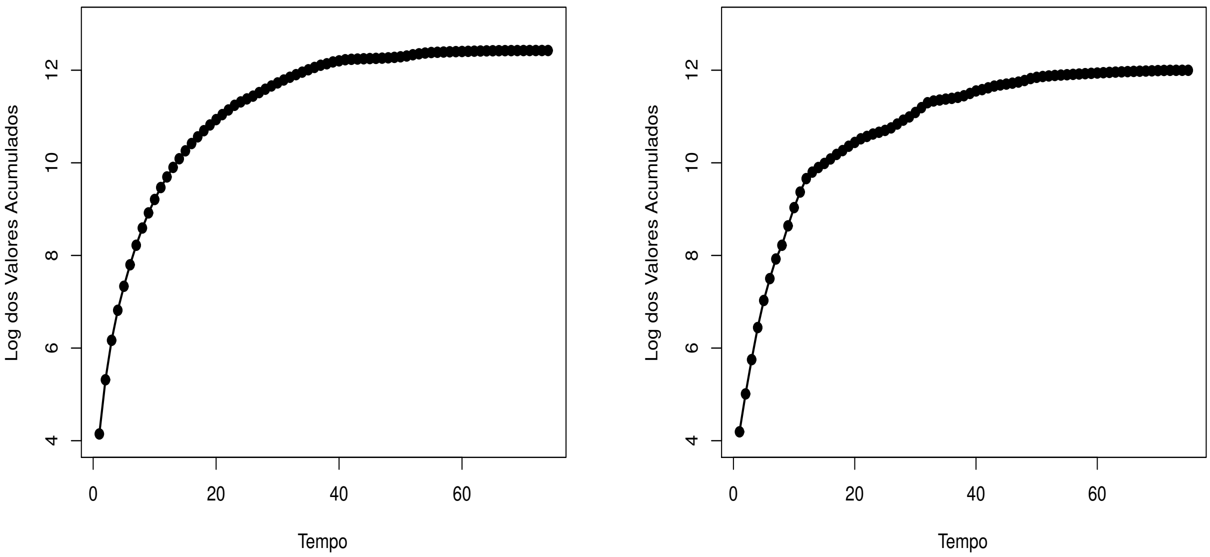
(a) Dia 1.

(b) Dia 2.

Figure 2: Potências acumuladas, dias 1 e 2.

Contudo, antes de descrevermos os modelos, considere a transformação logarítmica para os valores das potências acumuladas. Este 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 e o procedimento de estimação dos parâmetros.

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



(a) Dia 1.

(b) Dia 2.

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

3.1. Modelos Log-Logístico e Log-Gompertz

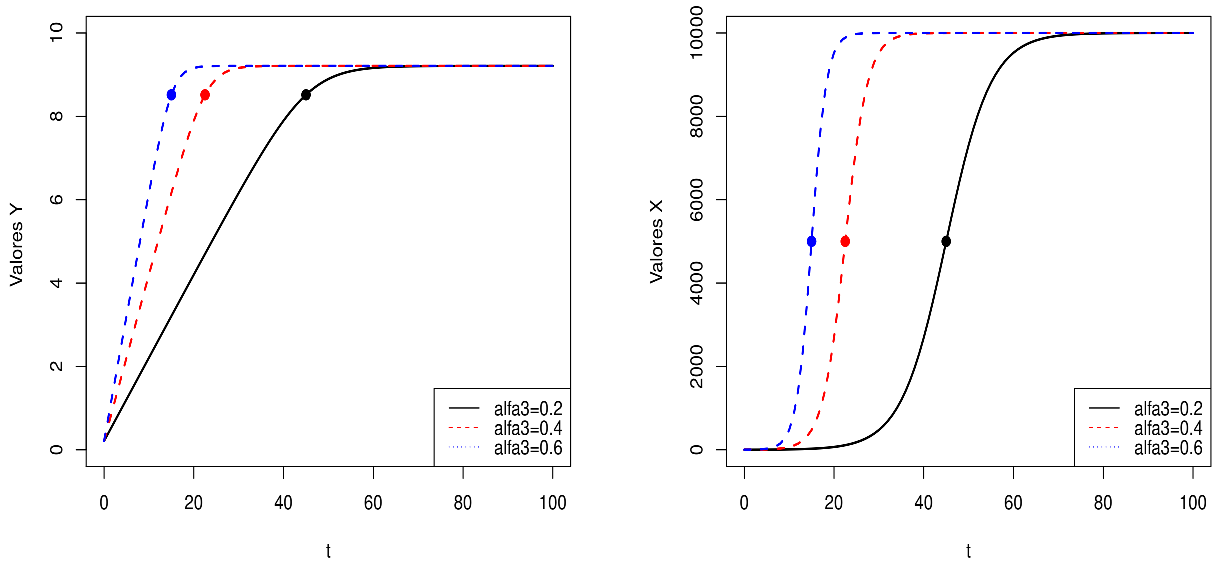
Considere que os valores Y_{ij} sejam gerados de acordo com o modelo Log-Logístico de parâmetro $\boldsymbol{\theta} = (\alpha, \beta_0, \beta_1)$, *i.e.*,

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

onde ε_{ij} é um erro aleatório com média 0 e variância σ^2, para $i = 0, 1, \ldots, k_j$ e $j = 1, \ldots, n$. 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 $y_T = \frac{\alpha_1}{2}$, 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 $i = 0, 1, \ldots, k_j$ e $j = 1, \ldots, n$. A Figura 4(b) mostra o mesmo gráfico na escala original, *i.e.*, $X_{i,j} = \exp\{Y_{ij}\}$, para $i = 0, 1, \ldots, k_j$ e $j = 1, \ldots, n$. 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.

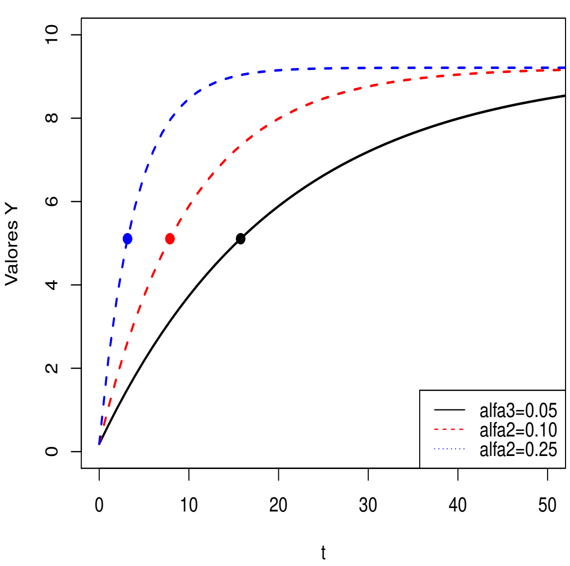
Como segundo modelo, considere que os valores Y_{ij} sejam gerados de acordo com o modelo Log-Gompertz de parâmetro $\boldsymbol{\theta} = (\alpha, \beta_0, \beta_1)$, *i.e.*,

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

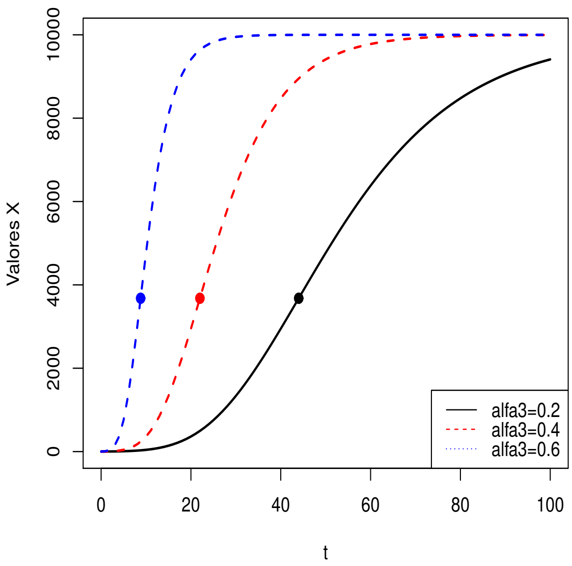
onde α_1, α_2 e α_3 são os parâmetros do modelo. De maneira similar ao modelo Logístico, no modelo Gompetz o valor do parâmetro α é a assintota superior. O parâmetro β_1 é a taxa média de crescimento e o parâmetro β_0 determina o valor da abcissa do do ponto de inflexão através da expressão $T = \frac{\beta_0}{\beta_1}$. 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 $i = 0, 1, \ldots, k_j$ e $j = 1, \ldots, n$. A Figura 5(b) mostra o mesmo gráfico na escala original, *i.e.*, $X_{i,j} = \exp\{Y_{ij}\}$, para $i = 0, 1, \ldots, k_j$ e $j = 1, \ldots, n$. Os símbolos • 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, $\frac{\alpha}{2}$, e no modelo Gompertz é $36,79\%$ do valor α.



(a) Modelo Log-Gompertz.



(b) Modelo Gompertz.

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:

$$\begin{aligned} I &= \text{Irradiancia acumulada;} \\ M &= \text{Massa acumulada de particulados.} \end{aligned}$$

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

$$\begin{aligned} Y_{ij} &= log(\alpha) - log\left(1 + \exp\left\{\beta^*_{ij} - \beta_3 T_{ij}\right\}\right) + \varepsilon_{ij} && (3) \\ \beta^*_{ij} &= \beta_0 + \beta_1 I_{ij} + \beta_2 M_{ij} + u_{ij}, && (4) \end{aligned}$$

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

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

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

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 $corr(\varepsilon_{ij}, \varepsilon_{(i+s)j}) = \rho^{|s|}$, para $i = 0, 1, \ldots, k_j$ e $j = 1, \ldots, n$. 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

$$
\begin{aligned}
Y_{ij} &= log(\alpha) - \exp\{\beta_{ij} - \beta_3 T_{ij}\} + \varepsilon_{ij} && (6)\\
\beta^*_{ij} &= \beta_0 + \beta_1 I_{ij} + \beta_2 M_{ij} + u_{ij}, && (7)
\end{aligned}
$$

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

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

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

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* (Pinheiro *et al.*, 2020). Utilizando os modelo ML_1 e MG_1 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 modelo ML_3 e MG_3 utilizamos os modelos ML_2 e MG_2 estimados e o comando *update* com a 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 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	2.096,099	-6.393,345	**-7.839,235**	-614,723	-7.004,732	-7.543,194
BIC	2.150,044	-6.355,583	**-7.796,079**	-560,778	-6.966,971	-7.500,038

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-valor$< \nu$, com exceção para β_2. Isto indica que a variável M (massa cumulada) 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.

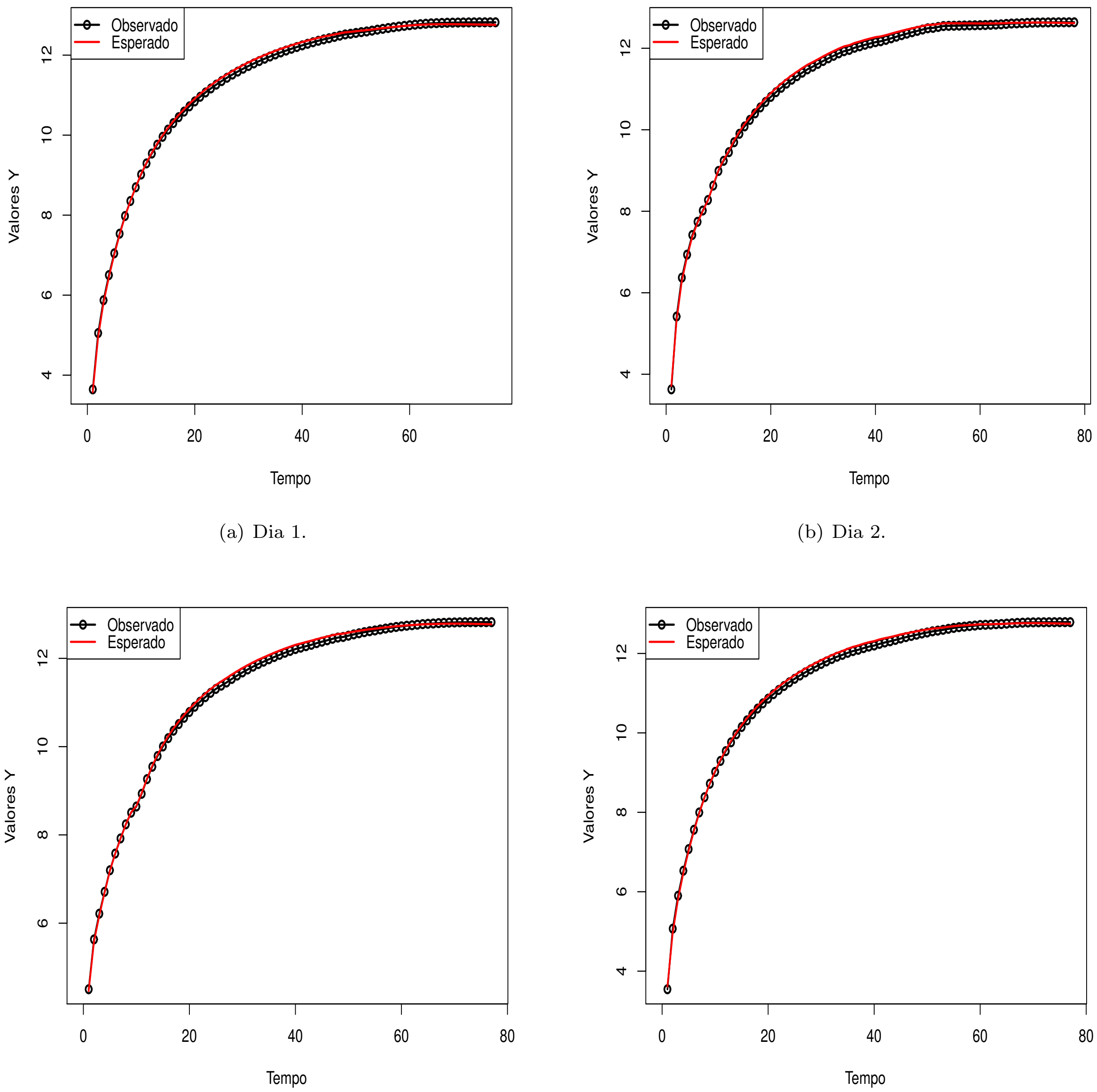
Parâmtros	Estimativa	Erro padrão	G.L.	Estatística t	p-valor
α	14,4664	0,2662	1.602	54,3451	$< 0,0001$
β_0	12,7662	0,2682	1.602	47,6033	$< 0,0001$
β_1	-1,0791	0,0028	1.602	-372,1501	$< 0,0001$
β_2	-0,0064	0,0006	1.602	-1,0171	$< 0,3093$
β_3	-0,0037	0,0009	1.602	-4,1692	$< 0,0001$

O modelo ML_3 estimado é dado por

$$\hat{Y}_{ij} = 14,4664 - \exp\{12,7662 - 1,0791 I_{ij} - 0,0064 M_{ij} + 0,0037 t_{ij}\},$$

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

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 $0,0036$, $0,0061$, $0,0040$ e $0,0046$, respectivamente. Para os dados dos 21 dias, o EQM é de $0,0063$.



(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, dias 1 à 4.

Seja $d_{ij} = \left(\frac{y_{ij} - \hat{y}_{ij}}{y_{ij}}\right) \cdot 100$ a diferença percentual entre o valor observado $y_i j$ e a projeção do modelo $\hat{y}_{ij}$, no i-ésimo instante de medição do dia j; e $\overline{d}_i = \frac{1}{k_k} \sum_{i=1}^{k_j} d_{ij}$ e a diferença percentual média para dia j, para $i = 1, \ldots, k_k$ e $j = 1, \ldots, n$. Para $\overline{d}_j < 0$, temos uma perda percentual em relação ao esperado; e para $\overline{d}_j > 0$, temos um ganho percentual.

A Figura 7, mostra o gráfico em barras das diferenças percentuais médias para os 21 dias considerados para o ajuste do modelo. Note que, somente nos dia 14 e 19 ocorreram ganhos em relação aos esperado pelo modelo. Porém, o ganho foi no máximo de $0,11\%$. A perda em relação ao esperado pelo modelo foi inferior a $1,40\%$. Estes resultados mostram um satisfatório ajuste do modelo.

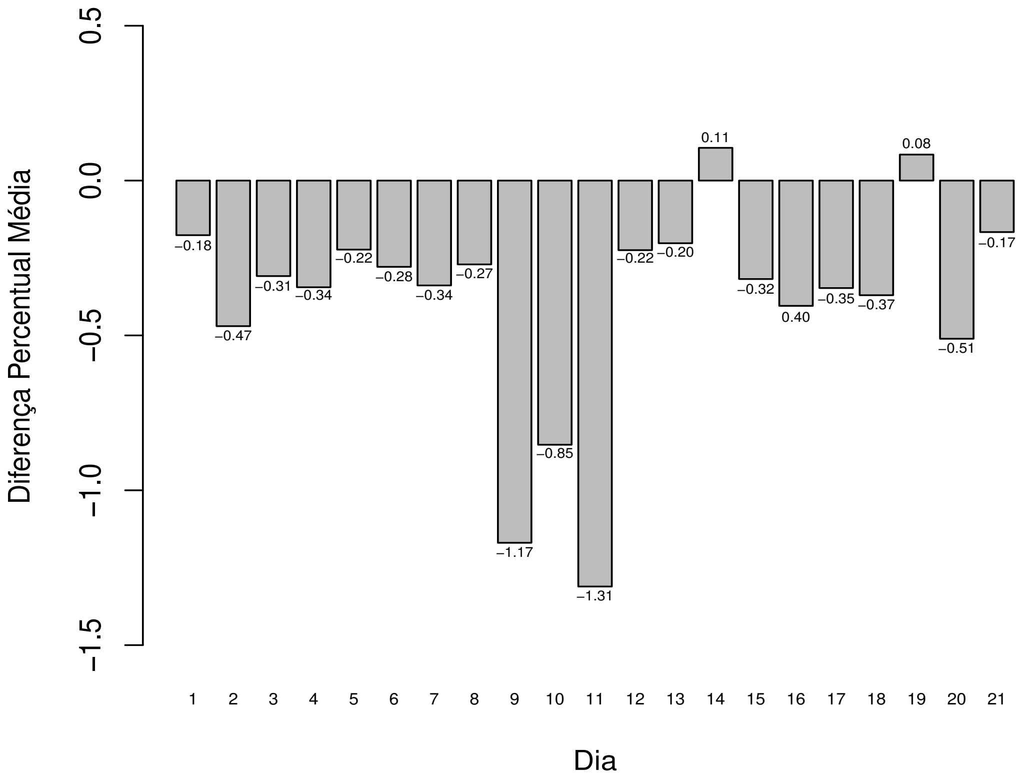


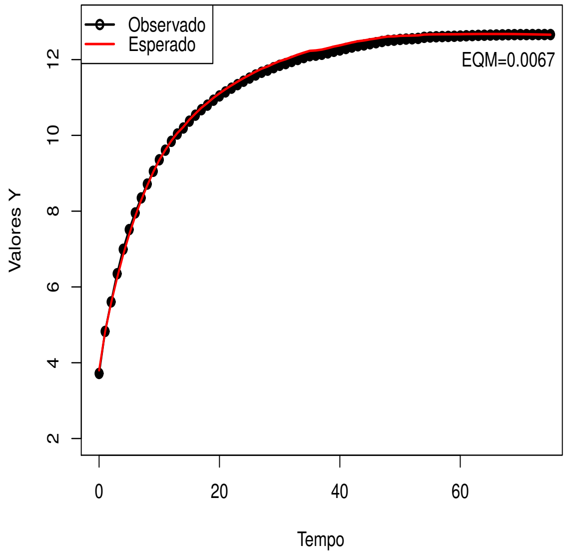
Figure 7: Diferença percentual média por dia.

4.1. Predições

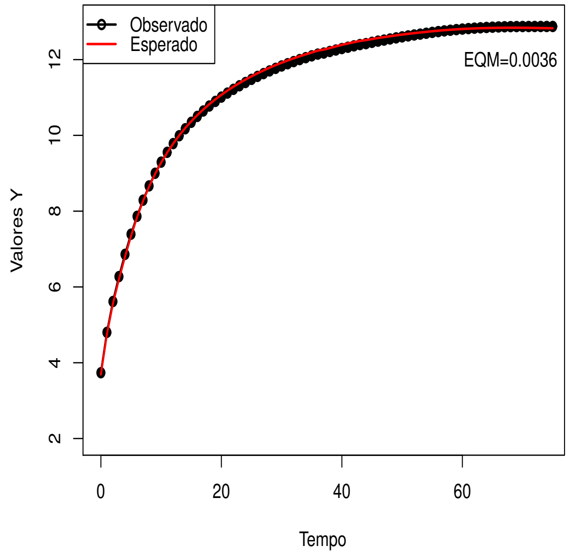
Utilizando o modelo ML_3 ajustado, predizemos os valores $Y's$ para os próximos 21 dias, *i.e.*, dia 22 (DATA) à 42 (DATA). Como ilustração, a Figura 8 mostra os valores observados nos dias 22, 23, 41 e 42 e os valores preditos pelo modelo. Os valores EQM destacados nos gráficos correspondem ao valor do erro quadrático médio.

A Figura 9 mostra o gráfico em pontos dos valores EQM dos 21 dias. A inha tracejada representa a média do EQM para os 21 dias. O EQM para os 21 dias é $0,0083$.

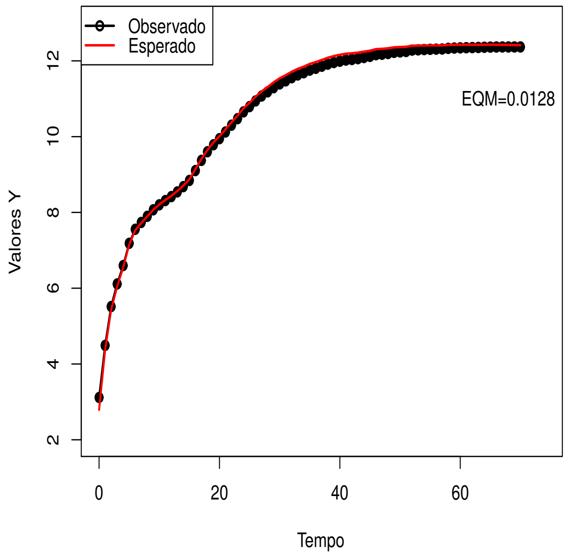
A Figura 10 mostra o gráfico em barras para as diferenças percentuais médias por dia. Para os dias 1 à 20 houve perda percentual em relação ao valor esperado. Contudo a perda foi de no máximo $1,31\%$ e ocorreu no dia 15. A menor perda ocorreu no dia 2 (0.20%). No dia 21 houve um pequeno ganho (0.02%) em relação ao valor esperado



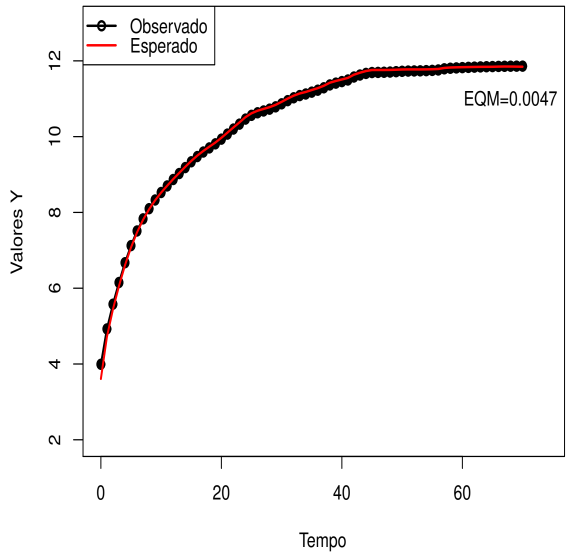
(a) Dia 22.



(b) Dia 23.



(c) Dia 41.



(d) Dia 42.

Figure 8: Gráficos dos valores observados e o valores preditos pelo modelo Log-Logístico.

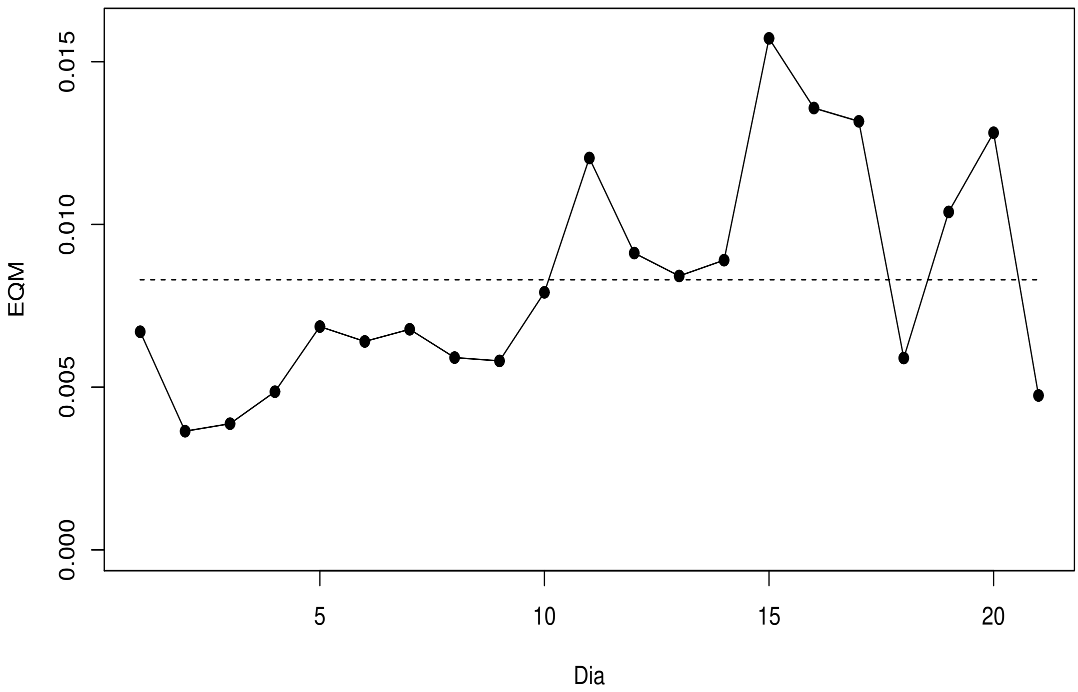


Figure 9: EQM para os 21 dias

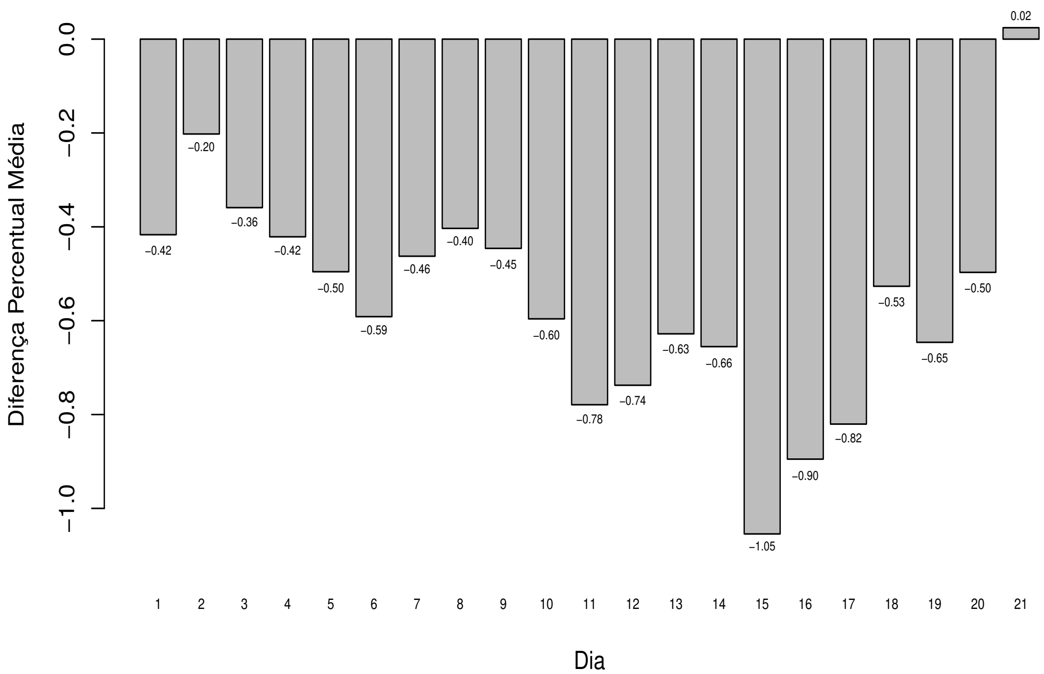


Figure 10: Diferença percentual média.

References

Pinheiro, J., Bates, D., DebRoy, S., Sakar, D. and R Core Team (20202). *Linear and Nonlinear Mixed*. R package version 3.1-147. URL https://CRAN.R-project.org/package-nlme.

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