

Dados Solar

Author ^{*a*}

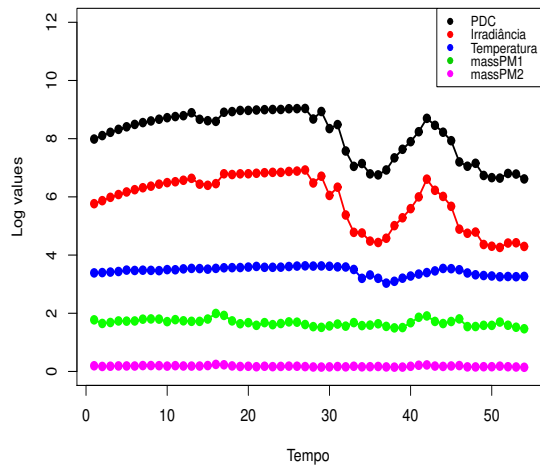
a

Abstract

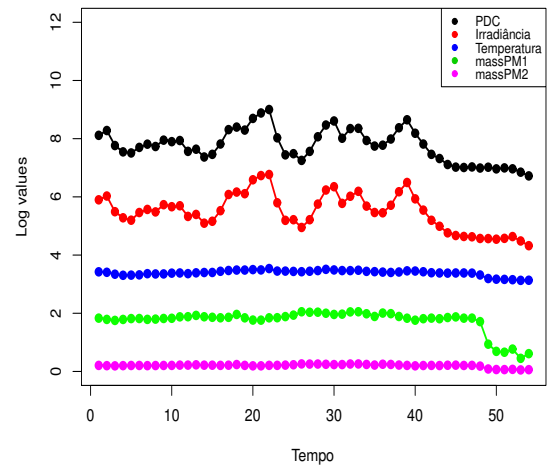
In this paper,

Keywords: text ...

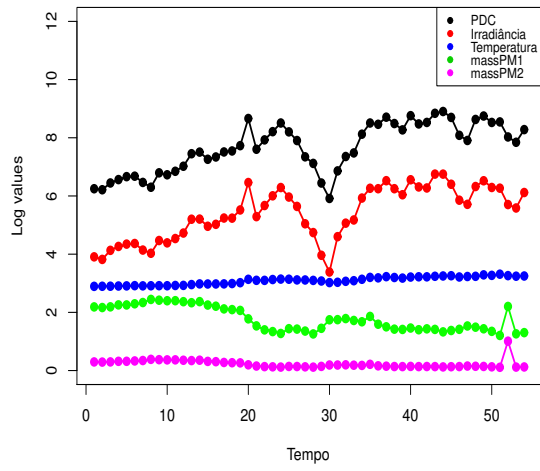
1. Introdução



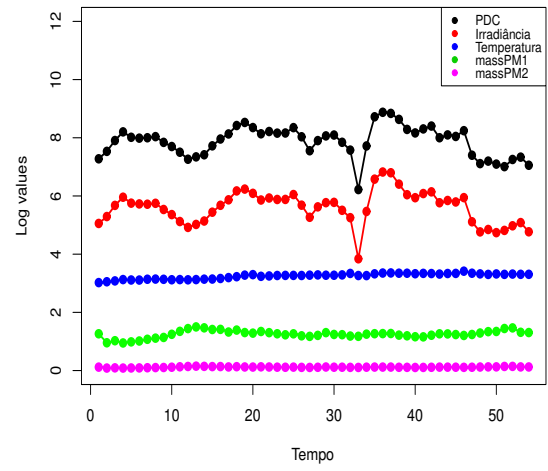
(a) Dia 1.



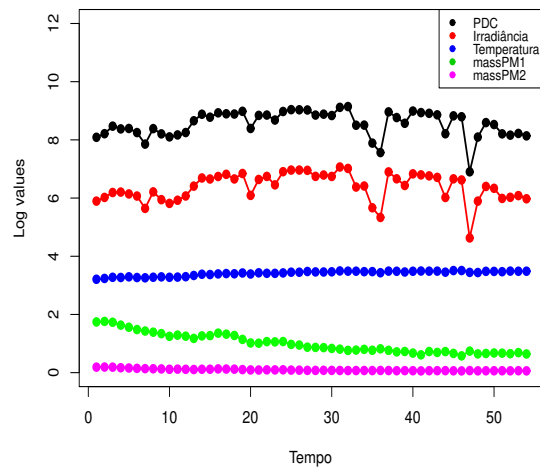
(b) Dia 2.



(c) Dia 3.



(d) Dia 4.



(e) Dia 5.

Figure 1: Valores observados de PDC e variáveis explicativas.

2. Modelo Estatístico

Considere que a potência gerada por placas fotovoltaicas limpas sejam monitoradas por n dias, $n \geq 1$. Suponha que para cada um dos n dias, sejam feitas $(k + 1)$ medições das potência geradas. Sem perda de generalidade, assuma que as medições são feitas nos instantes de tempos (t_0, t_1, \dots, t_k) definidos de maneiras regularmente espaçada em um intervalo de tempo (H_1, H_2) previamente fixado, *i.e.*, t_0 é o instante inicial de medição e h é uma constante que define o intervalo de medição. Por exemplo, em nosso experimento, fixamos $H_1 = 6$ and $H_2 = 17$, ou seja, as medições foram feitas das 6 horas até as 17 horas, em intervalos de $h = 15$ minutos. Assim, para cada dia temos $k = 40$ instantes de medição.

Seja Y_{ij} a potência gerada no j -ésimo instante do dia i , para $j \in \{t_1, \dots, t_k\}$ e $i = 1, \dots, n$. Considere as seguintes variáveis explicativas

$$\begin{aligned} I &= \text{Irradiancia;} \\ X &= \text{Temperatura;} \\ T &= \text{Tempo;} \\ P &= \text{massa PM1} \\ M &= \text{massa PM2.} \end{aligned}$$

Como primeiro modelo, denominado de modelo M_1 , considere o seguinte modelo linear com interações

$$Y_{ij} = \beta_0 + \beta_1 I_{ij} + \beta_2 X_{ij} + \beta_3 T_{ij} + \beta_4 P_{ij} + \beta_5 M_{ij} + \Psi_2^{M_1} + \Psi_3^{M_1} + \Psi_4^{M_1} + \Psi_5^{M_1} + \varepsilon_{ij} \quad (1)$$

onde $\Psi_2^{M_1} = \Psi_{2a}^{M_1} + \Psi_{2b}^{M_1}$, $\Psi_3^{M_1} = \Psi_{3a}^{M_1} + \Psi_{3b}^{M_1}$, $\Psi_4^{M_1}$ e $\Psi_5^{M_1}$ representam os termos com as interações, *i.e.*,

$$\begin{aligned} \Psi_{2a}^{M_1} &= \beta_6 I_{ij} X_{ij} + \beta_7 I_{ij} T_{ij} + \beta_8 I_{ij} P_{ij} + \beta_9 I_{ij} M_{ij} + \beta_{10} X_{ij} T_{ij}; \\ \Psi_{2b}^{M_1} &= \beta_{11} X_{ij} P_{ij} + \beta_{12} X_{ij} M_{ij} + \beta_{13} T_{ij} P_{ij} + \beta_{14} T_{ij} M_{ij} + \beta_{15} P_{ij} M_{ij}; \\ \Psi_{3a}^{M_1} &= \beta_{16} I_{ij} X_{ij} T_{ij} + \beta_{17} I_{ij} X_{ij} P_{ij} + \beta_{18} I_{ij} X_{ij} M_{ij} + \beta_{19} I_{ij} T_{ij} P_{ij} + \beta_{20} X_{ij} T_{ij} P_{ij}; \\ \Psi_{3b}^{M_1} &= \beta_{21} I_{ij} T_{ij} M_{ij} + \beta_{22} X_{ij} T_{ij} M_{ij} + \beta_{23} I_{ij} P_{ij} M_{ij} + \beta_{24} X_{ij} P_{ij} M_{ij} + \beta_{25} T_{ij} P_{ij} M_{ij}; \\ \Psi_4^{M_1} &= \beta_{26} I_{ij} X_{ij} T_{ij} P_{ij} + \beta_{27} I_{ij} X_{ij} T_{ij} M_{ij} + \beta_{28} I_{ij} X_{ij} P_{ij} M_{ij} + \beta_{29} I_{ij} T_{ij} P_{ij} M_{ij} + \beta_{30} X_{ij} T_{ij} P_{ij} M_{ij}; \\ \Psi_5^{M_1} &= \beta_{31} I_{ij} X_{ij} T_{ij} P_{ij} M_{ij}. \end{aligned}$$

onde, β_r são os parâmetros do modelo e ε_{ij} é o erro aleatório proveniente de uma distribuição normal com média 0 e variância σ^2 , $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, para $r = 0, \dots, 31$, $j = \{t_0, \dots, t_k\}$ e $i = 1, \dots, n$.

Para obter as estimativas para os parâmetros do modelo M_0 utilizando o software R e o comando `lm`. Para a realização dos testes de hipótese, consideramos um nível de significância $\alpha = 5\%$. A Tabela 1 mostra as estimativas dos parâmetros. Note que, ao nível de 5% de significância, nenhuma das variáveis é significativa para o modelo ($p\text{-valor} < \alpha = 0.05$).

Table 1: Estimates for parameters of the modelo M_1 .

Parameter	Estimate	Std. Error	t value	p-value	Parameter	Estimate	Std. Error	t value	p-value
β_0	-39.2685	25.0893	-1.565	0.1189	β_{16}	0.0578	0.0542	1.066	0.2876
β_1	7.8116	4.2875	1.822	0.0697	β_{17}	0.1109	3.0441	0.036	0.9710
β_2	13.4767	8.1909	1.645	0.1012	β_{18}	-0.0746	0.2615	-0.285	0.7755
β_3	1.1371	1.0024	1.134	0.2578	β_{19}	-0.1162	0.4715	-0.246	0.8055
β_4	-5.7486	57.3781	-0.100	0.9203	β_{20}	28.8381	62.8280	0.459	0.6467
β_5	647.6059	1165.4227	0.556	0.5789	β_{21}	3.7890	6.2064	0.610	0.5421
β_6	-2.2201	1.3779	-1.611	0.1084	β_{22}	6.5276	11.3478	0.575	0.5657
β_7	-0.1832	0.1738	-1.054	0.2928	β_{23}	33.1919	55.1170	0.602	0.5476
β_8	-0.3560	0.3157	-1.127	0.2607	β_{24}	64.9768	100.2741	0.648	0.5176
β_9	-0.0426	10.0597	-0.004	0.9966	β_{25}	7.0641	10.47400	0.674	0.5007
β_{10}	1.1018	17.6224	0.063	0.9502	β_{26}	0.0181	0.0802	0.225	0.8221
β_{11}	0.4737	1.5202	0.312	0.7556	β_{27}	-1.1001	1.9266	-0.571	0.5685
β_{12}	-94.7567	203.7012	-0.465	0.6422	β_{28}	-10.2384	17.1650	-0.596	0.5514
β_{13}	-195.3480	365.5237	-0.534	0.5935	β_{29}	-1.2215	1.7977	-0.679	0.4975
β_{14}	-22.6207	36.2073	-0.625	0.5327	β_{30}	-2.0587	3.2927	-0.625	0.5324
β_{15}	-212.7318	316.6061	-0.672	0.5023	β_{31}	0.3593	0.5600	0.642	0.5218

Como a variável temperatura possui um comportamento praticamente constante durante os seis dias considerados e p -valor maior do que o nível de significância α , considere o modelo M_1 sem a variável temperatura. Denominamos este modelo de M_2 ; sua expressão matemática é dada por

$$Y_{ij} = \beta_0 + \beta_1 I_{ij} + \beta_2 T_{ij} + \beta_3 P_{ij} + \beta_4 M_{ij} + \Psi_2^{M_2} + \Psi_3^{M_2} + \Psi_4^{M_2} + \varepsilon_{ij}, \quad (2)$$

onde $\Psi_2^{M_2}$, $\Psi_3^{M_2}$ e $\Psi_4^{M_2}$ são os termos com interações, dados por

$$\begin{aligned} \Psi_2^{M_2} &= \beta_5 I_{ij} T_{ij} + \beta_6 I_{ij} P_{ij} + \beta_7 T_{ij} P_{ij} + \beta_8 I_{ij} M_{ij} + \beta_9 T_{ij} M_{ij} + \beta_{10} P_{ij} M_{ij}; \\ \Psi_3^{M_2} &= \beta_{11} I_{ij} T_{ij} P_{ij} + \beta_{12} I_{ij} T_{ij} M_{ij} + \beta_{13} I_{ij} P_{ij} M_{ij} + \beta_{14} T_{ij} P_{ij} M_{ij}; \\ \Psi_4^{M_2} &= \beta_{15} I_{ij} T_{ij} P_{ij} M_{ij}. \end{aligned}$$

em que, β_r são os parâmetros do modelo e ε_{ij} é o erro aleatório proveniente de uma distribuição normal com média 0 e variância σ^2 , $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, para $r = 0, \dots, 15$, $j = \{t_0, \dots, t_k\}$ e $i = 1, \dots, n$.

A Tabela 2 mostra as estimativas dos parâmetros do modelo M_2 . Note que, ao nível de 5% de significância, somente a variável irradiância é significativa para o modelo (p -valor $< \alpha = 0.05$).

Table 2: Estimates for parameters of the modelo M_2 .

Parameter	Estimate	Std. Error	t value	p-value					
β_0	2.0914	1.0718	1.951	0.0521	β_8	-13.3988	9.4276	-1.421	0.1565
β_1	1.0237	0.1906	5.371	1.77e-07	β_9	-2.7027	1.6378	-1.650	0.1001
β_2	0.0543	0.0399	1.363	0.1739	β_{10}	-20.3404	14.7226	-1.382	0.1683
β_3	-4.2208	2.7093	-1.558	0.1205	β_{11}	-0.0201	0.0121	-1.654	0.0995
β_4	77.1893	55.1195	1.400	0.1626	β_{12}	0.4706	0.2832	1.662	0.0978
β_5	-0.0095	0.0070	-1.359	0.1753	β_{13}	3.4972	2.5364	1.379	0.1692
β_6	0.7413	0.4561	1.625	0.1053	β_{14}	0.7894	0.4619	1.709	0.0886
β_7	0.1146	0.0708	1.620	0.1066	β_{15}	-0.1367	0.0800	-1.707	0.0890

Com o parâmetro da variável tempo possui o maior valor de p -valor (sem os termos com interação), retiramos esta variável do modelo. Assim, obtemos o modelo M_3 ; sua expressão matemática é dada por

$$Y_{ij} = \beta_0 + \beta_1 I_{ij} + \beta_2 P_{ij} + \beta_3 M_{ij} + \Psi_2^{M_3} + \Psi_3^{M_4} + \varepsilon_{ij}, \quad (3)$$

onde $\Psi_2^{M_4} = \beta_4 I_{ij} P_{ij} + \beta_5 I_{ij} M_{ij} + \beta_6 P_{ij} M_{ij}$ e $\Psi_3^{M_3} = \beta_{15} I_{ij} P_{ij} M_{ij}$ são os termos com interação, β_r são os parâmetros do modelo e ε_{ij} é o erro aleatório proveniente de uma distribuição normal com média 0 e variância σ^2 , $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, para $r = 0, \dots, 7$, $j = \{t_0, \dots, t_k\}$ e $i = 1, \dots, n$.

A Tabela 3 mostra as estimativas dos parâmetros do modelo M_3 . Ao nível de 5% de significância, o intercepto e a variável irradiância são significativas para o modelo (p -valor $< \alpha = 0.05$).

Table 3: Estimates for parameters of the modelo M_1 .

Parameter	Estimate	Std. Error	t value	p-value
β_0	2.8944	0.4909	5.896	1.15e-08
β_1	0.8753	0.0874	10.018	< 2e-16
β_2	-0.4217	0.8457	-0.499	0.618
β_3	5.4750	19.8351	0.276	0.783
β_4	0.0759	0.1481	0.513	0.609
β_5	-0.7807	3.5008	-0.223	0.824
β_6	-1.4847	5.5053	-0.270	0.788
β_7	0.1878	0.9765	0.192	0.848

Como quarto modelo, M_4 , considere o modelo M_3 sem a variável M , *i.e.*,

$$Y_{ij} = \beta_0 + \beta_1 I_{ij} + \beta_2 P_{ij} + \beta_3 I_{ij} P_{ij} + \varepsilon_{ij}, \quad (4)$$

onde β_r são os parâmetros do modelo e ε_{ij} é o erro aleatório proveniente de uma distribuição normal com média 0 e variância σ^2 , $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, para $r = 0, 1, 2, 3$, $j = \{t_0, \dots, t_k\}$ e $i = 1, \dots, n$.

A Tabela 4 mostra as estimativas dos parâmetros do modelo M_4 . Ao nível de 5% de significância, o intercepto, a variável irradiância, a variável massa PM1 e a interação IP são significativas para o modelo (p-valor $< \alpha = 0.05$).

Table 4: Estimates for parameters of the modelo M_1 .

Parameter	Estimate	Std. Error	t value	p-value
β_0	3.1005	0.0647	47.919	$< 2e - 16$
β_1	0.8498	0.0111	76.863	$< 2e - 16$
β_2	-0.2422	0.0396	-6.121	$3.32e - 09$
β_3	0.0439	0.0069	6.314	$1.14e - 09$

Como modelo M_5 , considere o modelo M_3 , sem a variável P , *i.e.*,

$$Y_{ij} = \beta_0 + \beta_1 I_{ij} + \beta_2 P_{ij} + \beta_3 I_{ij} M_{ij} + \varepsilon_{ij}, \quad (5)$$

onde β_r são os parâmetros do modelo e ε_{ij} é o erro aleatório proveniente de uma distribuição normal com média 0 e variância σ^2 , $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, para $r = 0, 1, 2, 3$, $j = \{t_0, \dots, t_k\}$ e $i = 1, \dots, n$.

A Tabela 5 mostra as estimativas dos parâmetros do modelo M_5 . Ao nível de 5% de significância, o intercepto, a variável irradiância, a variável massa PM2 e a interação IM são significativas para o modelo (p-valor $< \alpha = 0.05$).

Table 5: Estimates for parameters of the modelo M_1 .

Parameter	Estimate	Std. Error	t value	p-value
β_0	2.9890	0.0457	65.394	$< 2e - 16$
β_1	0.8696	0.0079	109.731	$< 2e - 16$
β_2	-1.5104	0.2377	-6.354	$9.04e - 10$
β_3	0.2765	0.0429	6.446	$5.39e - 10$

Os resultados acima nos mostram que os modelos M_4 e M_5 possuem variáveis significativas para explicação da variável resposta Y . Para escolher um destes dois modelos, consideramos os critérios de seleção de modelos AIC e BIC. A Tabela 6 mostra os valores AIC e BIC de cada um dos modelos. A terceira coluna desta Tabela apresenta o valor do erro quadrático médio (REQM) dos modelos. De acordo com estas três medidas, o modelo M_5 apresenta uma leve vantagem em relação ao modelo M_4 . Assim, a partir deste ponto, consideramos o modelo M_5 como o melhor modelo dentre os 5 considerados até este ponto do trabalho.

Table 6: Valores AIC, BIC e REQM para os modelo M_4 e M_5 .

Model	AIC	BIC	REQM
M_4	-918.5599	-900.5678	0.043348
M_5	-919.7066	-901.7145	0.043256

Uma das presunções básicas de um modelo de regressão, como o modelo M_5 , é a de que os erros são não correlacionados, *i.e.*, $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{in}) \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I})$, onde $\mathcal{N}_n(\cdot)$ representam a distribuição normal n -variada com vetor de média $\mathbf{0}$ de dimensão $n \times 1$ e \mathbf{I}_n é a matriz identidade de ordem n . Contudo, uma vez que as observações são feitas longitudinalmente em cada dia, a suposição de erros não correlacionados pode não ser adequada. Devido a isto, considere o modelo M_6 , que é dado pelo modelo M_5 porém considerando uma correlação do tipo autoregressiva de ordem 1 para os erros, *i.e.*, $\varepsilon_i \sim \mathcal{N}_n(\mathbf{0}, \Sigma)$, onde

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{k-1} \\ \rho & 1 & \rho & \dots & \rho^{k-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{k-3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho^{k-1} & \rho^{k-2} & \rho^{k-3} & \dots & 1 \end{bmatrix}$$

Para obter as estimativas dos parâmetros do modelo M_6 utilizamos o comando *gls* do software R. A Tabela 7 mostra as estimativas dos parâmetros do modelo M_6 . Como para o modelo M_5 , o intercepto, a variável irradiância, a variável massa PM2 e a interação IM são significativas para o modelo (p-valor $< \alpha = 0.05$). A estimativa para o parâmetros ρ é $\hat{\rho} = 0.3271$. Como os modelos M_5 e M_6 são modelos “encaixados”, comparamos estes dois modelos através do teste da razão de máxima verossimilhança. Para realizações deste teste utilizamos o comando *anova* do software R.

A Tabela 8, mostra os resultados do teste. Como p-valor $< \alpha$, rejeitamos a hipótese nula de que $\rho = 0$. Ou seja, o modelo M_6 é preferível em relação ao modelo M_5 . A Tabela 8, também mostra os valores AIC e BIC dos dois modelos. Como podemos observar, os valores AIC e BIC do modelo M_6 são menores do que os valores do modelo M_5 ; indicando o modelo M_6 como sendo o melhor modelo.

Table 7: Estimates for parameters of the modelo M_1 .

Parameter	Estimate	Std. Error	t value	p-value
β_0	2.9404	0.0566	51.9354	0
β_1	0.8785	0.0098	89.7209	0
β_2	-1.2731	0.2983	-4.2670	0
β_3	0.2315	0.0533	4.3368	0

Table 8: Comparação dos modelos M_5 e M_6 .

Model	df	AIC	BIC	logLik	L.Ratio	p-value
M_6	6	-916.9866	-895.4857	464.4933	28.37866	< 0.0001
M_5	5	-890.6080	-872.6905	450.3040	-	-

2.1. Modelo linear misto

Considere agora o modelo M_6 acrescido de um componente aleatório no intercepto. Denominamos este modelo de M_7 ; sua expressão matemática é dada por

$$Y_{ij} = (\beta_0 + U_{0j}) + \beta_1 I_{ij} + \beta_2 M_{ij} + \beta_3 I_{ij} M_{ij} + \varepsilon_{ij}, \quad (6)$$

onde $U_{0j} \sim \mathcal{N}(0, \Sigma_0)$, para $j = \{t_0, \dots, t_k\}$ e $i = 1, \dots, n$.

Para obter as estimativas dos parâmetros do modelo M_7 , utilizamos o pacote *nlme* do software R e o comando *lme*. A Tabela 9 mostra as estimativas para os efeitos fixos. Realizamos o teste da razão de verossimilhanças para verificar se a introdução do componente aleatório é essencial. Ou seja, comparamos os modelo M_6 e M_7 . A Tabela 10 mostra os resultados do teste. Como p-valor $> \alpha$, rejeitamos a presença do componente aleatório no intercepto. O modelo M_6 , permanece como melhor modelo, dentre os testados.

Table 9: Estimates for parameters of the modelo M_1 .

Parameter	Estimate	Std. Error	t value	p-value
β_0	2.9108	0.0609	47.8125	0
β_1	0.8838	0.0106	83.7382	0
β_2	-1.1718	0.3200	-3.6615	3e-04
β_3	0.2120	0.0573	3.7011	3e-04

Table 10: Comparação dos modelos M_6 e M_7 .

Model	df	AIC	BIC	logLik	L.Ratio	p-value
M_7	7	-913.3974	-888.3130	463.6987	1.589219	0.2074
M_6	6	-916.9866	-895.4857	464.4933	-	-

Repetimos o procedimento do modelo M_7 , considerando a presença do termo aleatório nos parâmetros β_1 , β_2 (um parâmetro por vez). Porém, a presença dos termos aleatórios foram rejeitadas os dois casos, como mostrado nas Tabelas 11 e 12.

Table 11: Comparação dos modelos M_6 e M_7 ($\beta_1 + U_{1j}$).

Model	df	AIC	BIC	logLik	L.Ratio	p-value
M_7	7	-913.6095	-888.5250	463.8048	1.3771	0.2406
M_6	6	-916.9866	-895.4857	464.4933	-	-

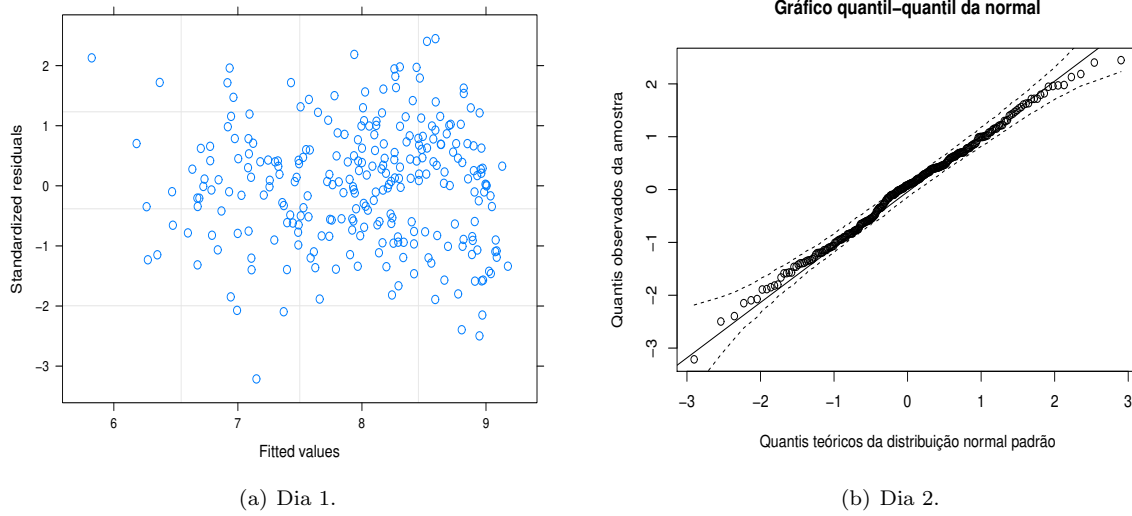
Table 12: Comparação dos modelos M_6 e M_7 ($\beta_2 + U_{2j}$).

Model	df	AIC	BIC	logLik	L.Ratio	p-value
M_7	7	-912.5222	-887.4377	463.2611	2.4644	0.1164
M_6	6	-916.9866	-895.4857	464.4933	-	-

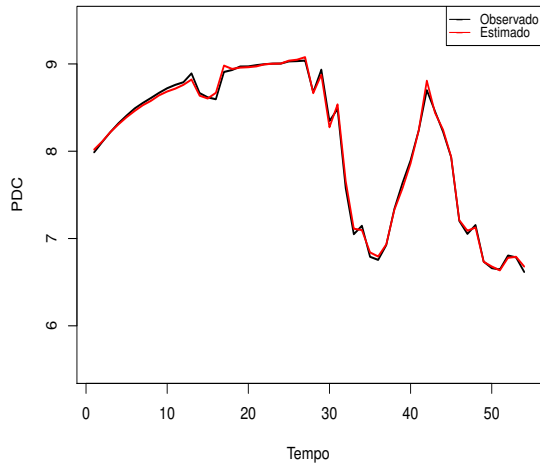
Table 13: Comparação dos modelos M_6 e M_7 ($\beta_2 + U_{2j}$).

Model	df	AIC	BIC	logLik	L.Ratio	p-value
M_7	7	-912.5222	-887.4377	463.2611	2.4644	0.1164
M_6	6	-916.9866	-895.4857	464.4933	-	-

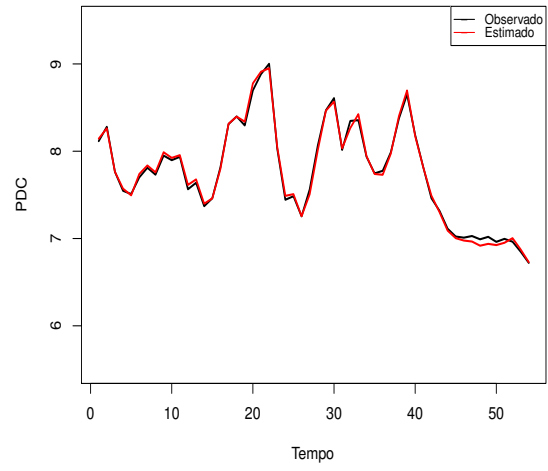
Ou seja, dentre todos os modelos testados, o modelo M_6 é o melhor modelo. A Figura 3, mostra o gráfico dos resíduos e o gráfico quantil-quantil. Note que o gráfico da Figura 3(a) não apresenta nenhuma estrutura, indicando homogeneidade da variância. O gráfico da Figura 3(b), mostra a não rejeição da normalidade dos resíduos. Também realizamos um teste de normalidade dos resíduos, utilizando o comando *shapiro.teste* do software R. O p-valor obtido foi de 0.6633, indicando a não rejeição da normalidade.

Figure 2: Grupos identificados com $k = 3$, $k = 4$ e $k = 5$.

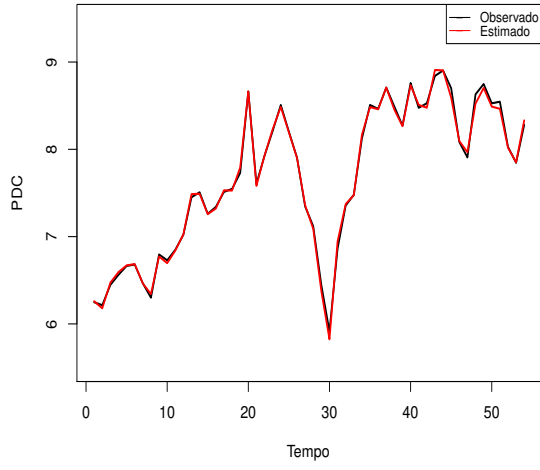
A Figura ?? mostra o gráfico dos valores observados e preditos de PDC pelo modelo M_6 .



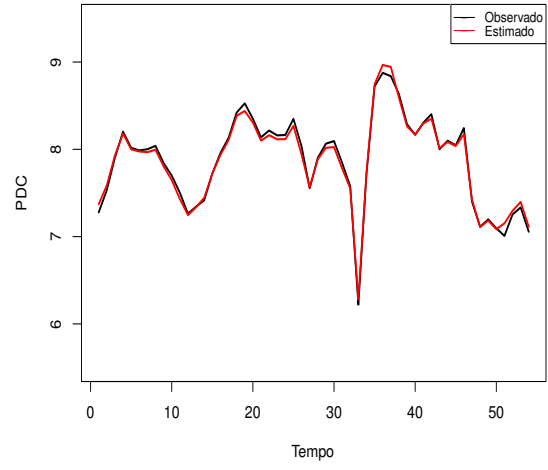
(a) Dia 1.



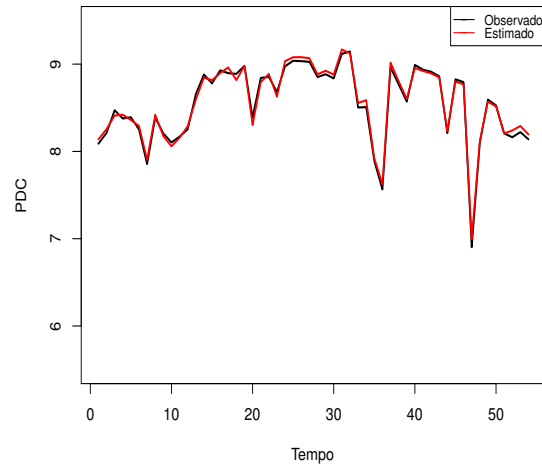
(b) Dia 2.



(c) Dia 3.



(d) Dia 4.



(e) Dia 5.

Figure 3: Valores observados e preditos pelo modelo M_6 .

References

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