

CNPq/UFRGS
Apoio

William Hepp (william.hepp@ufrgs.br)
Bolsista

Sílvia R.C. Lopes (silvia.lopes@ufrgs.br)
Orientadora

Introdução

Este trabalho apresenta uma análise univariada de séries temporais de medições de radiação solar fornecidas pelo *National Solar Radiation Database*, nos Estados Unidos. O banco de dados fornece um total de 233 série temporais de medidas da radiação solar, observadas mês a mês. Cada série é referente a uma localidade dos EUA. As observações tiveram início em janeiro de 1960 e término em dezembro de 2010. Para analisar os dados seguimos duas propostas. Primeiramente, dada a série $\{y_t\}_{t=0}^N$, aplicamos uma normalização da série original por meio das médias e variâncias observadas em relação aos meses, isto é, calculamos

$$\mu_i = \frac{1}{n} \sum_{k=0}^{n-1} y_{12k+i} \quad \text{e} \quad \sigma_i^2 = \frac{1}{n-1} \sum_{k=0}^{n-1} (y_{12k+i} - \mu_i)^2, \quad i = 1, \dots, 12, \quad (1)$$

e definimos a série normalizada $\{z_t\}_{t=0}^N$ por

$$z_t = \frac{y_t - \mu_i}{\sigma_i}, \quad t = 1, \dots, N, \quad \text{para} \quad i = t - 12 \left\lfloor \frac{t}{12} \right\rfloor. \quad (2)$$

Devido a existência de uma quebra estrutural nos dados a partir de janeiro de 1991 a normalização que descrevemos foi aplicada separadamente para as observações de janeiro de 1960 a dezembro de 1990 e para aquelas a parti de janeiro de 1991 até dezembro de 2010. Prosseguimos, analisando os resíduos com uso do periodograma cumulativo e no caso de rejeição da hipótese nula de que o resíduo é um ruído branco, um modelo ARMA(p, q) foi ajustado, com $p, q \in \{0, 1, 2\}$. A segunda proposta de análise foi remover a aparente componente determinística oscilatória de cada série. Para tanto, utilizamos um modelo harmônico de períodos fixos $p_1 = 12, p_2 = 8$ e $p_3 = 6$. Nos casos em que os resíduos do modelo harmônico apresentaram correlação significativa, em geral o periodograma da série apresentava decaimento lento, propriedade característica de longa dependência. Com essa motivação, para a modelagem dos resíduos, fizemos uso de modelos ARFIMA(p, d, q), com $p, q \in \{0, 1, 2\}$. Em todo o trabalho, utilizamos o periodograma cumulativo para decidir se os resíduos de um modelo podiam ser considerados um ruído branco. Lembramos que o periodograma da série temporal $\{x_t\}_{t=0}^N$ é a transformada discreta de Fourier da série, a saber,

$$I(\omega) = \frac{1}{2\pi N} \left| \sum_{t=0}^N x_t e^{-i\omega t} \right|^2,$$

e o periodograma cumulativo padronizado é a função dada por

$$C(x) = \begin{cases} 0, & x < 1, \\ \frac{\sum_{k=1}^x I(\omega_k)}{\sum_{k=1}^q I(\omega_k)}, & 1 \leq x < i+1, \quad i = 1, \dots, q := \lfloor \frac{N-1}{2} \rfloor, \\ 1, & x \geq q, \end{cases}$$

onde $\omega_j = 2\pi j/N$ são as frequências de Fourier. O procedimento para se testar se uma dada série é um ruído branco consiste no estudo do gráfico do periodograma cumulativo padronizado e rejeitar a hipótese nula de que a série não é um ruído branco ao nível $\alpha > 0$ se, para algum $x \in [1, q]$, $C(x)$ fica fora dos limites de Kolmogorov-Smirnov

$$y = \frac{x-1}{q-1} \pm \frac{k_\alpha}{\sqrt{q-1}},$$

onde k_α denota o α -ésimo quantil da distribuição da estatística de Kolmogorov-Smirnov.

Na segunda metodologia, iniciamos com o ajuste de um modelo harmônico aos dados. Para selecionar o modelo mais adequado, seguimos o método de eliminação passo a passo para trás (*backward stepwise elimination*) baseado nos p -valores dos coeficientes, onde o modelo completo considerado foi um modelo harmônico com três frequências. Para todos os casos, as frequências utilizadas no modelo completo foram 1/12, 1/6 e 1/8, sugeridas a partir da análise do periodograma da série original. Os ajustes e cálculos dos p -valores dos coeficientes foram feitos pela função `lm` do R.

Exemplo de Análise

Iniciamos normalizando a série pelas médias e desvios-padrão relativos aos meses, respectivamente denotadas por μ_i e σ_i , $i = 1, \dots, 12$ e calculadas como em (1).

A partir dos gráficos das médias e desvios-padrão mensais, fica clara a natureza não-estacionária da série.

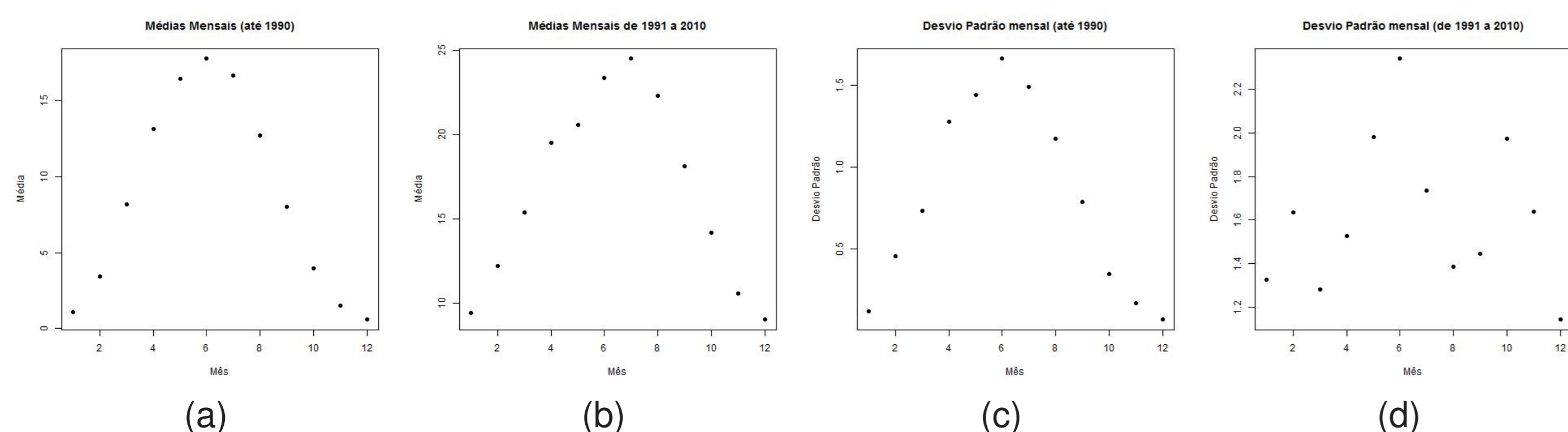


Figura: (a) Médias relativas aos meses até 1990; (b) Médias relativas aos meses de 1991 a 2010; (c) Desvios-padrão relativos aos meses até 1990; (d) Desvios-padrão relativos aos meses de 1991 a 2010.

No próximo passo, definimos a série normalizada $\{z_t\}$ como em (2). A seguir, apresentamos gráficos da série original $\{y_t\}$, onde claramente se observa uma componente sazonal, da série normalizada $\{z_t\}$. Notamos que nenhuma sazonalidade óbvia pode ser percebida na série normalizada.

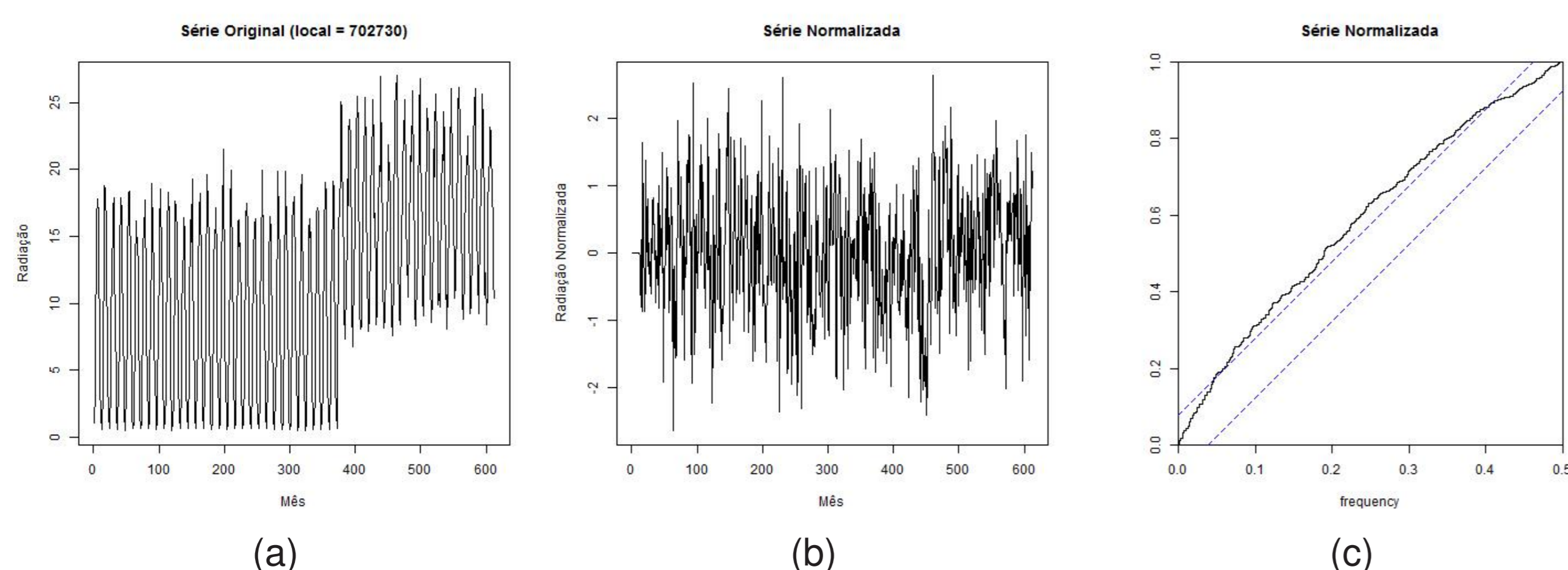


Figura: (a) Série original $\{y_t\}$; (b) Série normalizada $\{z_t\}$; (c) Periodograma cumulativo da série dessazonalizada.

Em seguida, procedemos com uma análise residual da série normalizada utilizando o periodograma cumulativo.

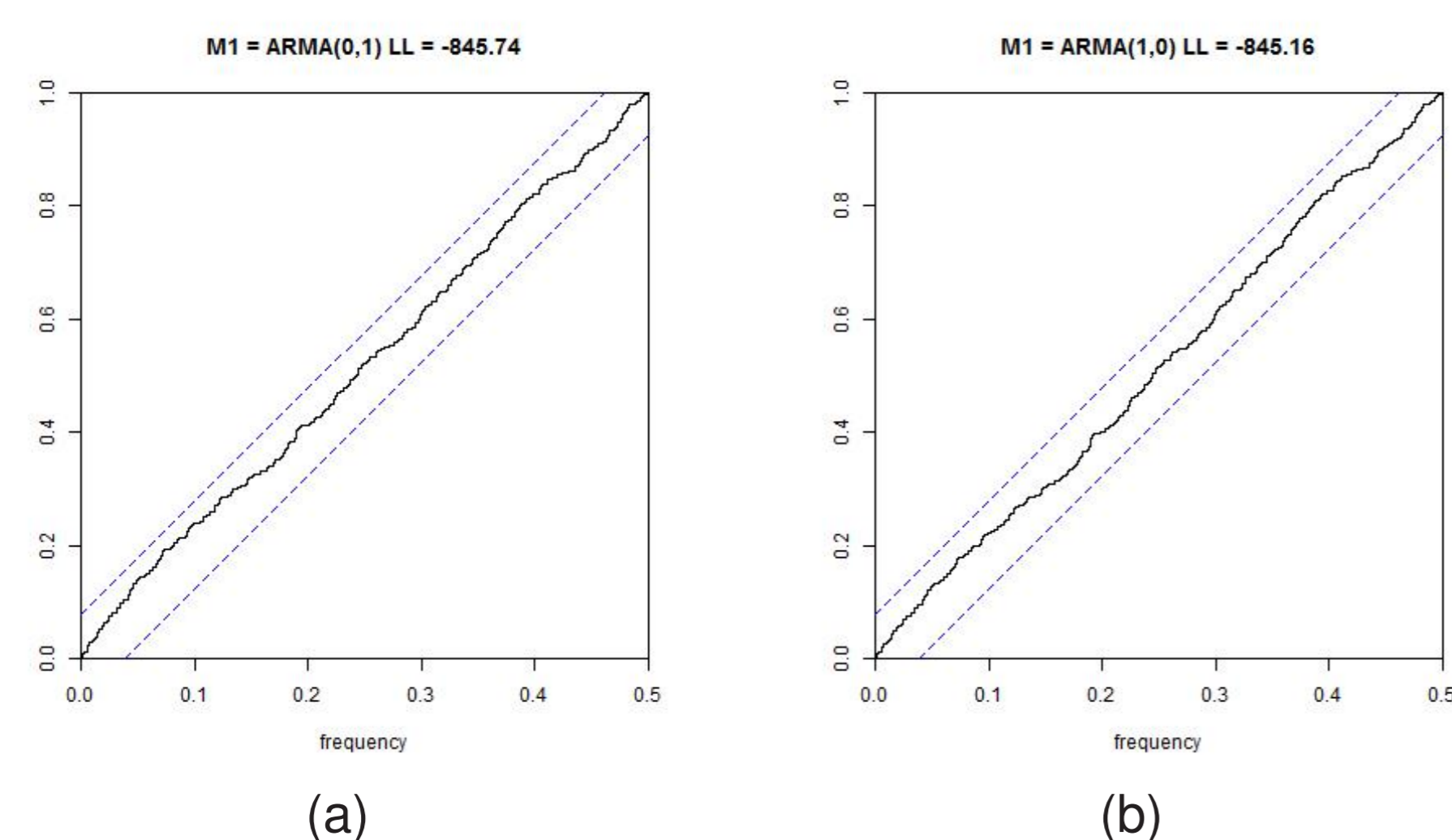


Figura: Periodogramas cumulativos dos modelos (a) ARMA(0, 1); (b) ARMA(1, 0) ajustados.

Para a série considerada, o periodograma cumulativo não está confinado dentro dos limites de confiança. Isso sugere a necessidade de modelagem adicional da série. Prosseguimos então com o ajuste dos modelos ARMA(p, q), para todas as combinações de $p, q \in \{0, 1, 2\}$. Após o ajuste, excluímos os modelos que possuem algum coeficiente cujo p -valor for maior que 0,05. Neste caso, os únicos modelos que satisfazem esta condição são o ARMA(0, 1) e o ARMA(1, 0). Consideramos adequados os modelos cujos gráficos do periodograma cumulativo dos resíduos estejam situados dentro dos limites de confiança. Na Figura ?? apresentamos os periodogramas cumulativos dos resíduos referentes aos modelos ARMA(0, 1) e ARMA(1, 0). Em ambos os casos, os resíduos podem ser considerados um ruído branco Gaussiano. Além disso, os valores da log-verossimilhança de ambos são praticamente iguais: -845,74 e -845,16, respectivamente. Desta forma, o modelo final escolhido é o mais simples/parcimonioso. Portanto, neste exemplo, escolhemos como modelo adequado o ARMA(0, 1) = AR(1).

Apresentamos abaixo o sumário da função `lm` para o modelo harmônico selecionado.

```
----- Harmonic Model 1 summary ----- :
Call:
lm(formula = serie1 ~ G - 1)
Residuals:
    Min       1Q   Median       3Q      Max
-3.1182 -0.4468  0.0076  0.3497  3.5436
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
GA1  -8.78916    0.07281  -120.706 < 2e-16 ***
GA2   0.55994    0.07282    7.690 1.35e-13 ***
GB2  -0.17105    0.07282   -2.349  0.0193 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.9931 on 369 degrees of freedom
Multiple R-squared:  0.9754, Adjusted R-squared:  0.9752
F-statistic: 4878 on 3 and 369 DF, p-value: < 2.2e-16
----- Harmonic Model 2 summary ----- :
Call:
lm(formula = serie2 ~ G - 1)
Residuals:
    Min       1Q   Median       3Q      Max
-5.6984 -1.0453 -0.0097  1.3030  4.9014
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
GA1  -7.2715    0.1534  -47.407 < 2e-16 ***
GB1  -1.6618    0.1534  -10.832 < 2e-16 ***
GA2  -0.4488    0.1534   -2.926 0.003773 **
GB2   0.5240    0.1534    3.416 0.000748 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 1.68 on 236 degrees of freedom
Multiple R-squared:  0.91, Adjusted R-squared:  0.9084
F-statistic: 596.3 on 4 and 236 DF, p-value: < 2.2e-16
```

Neste caso, portanto, o modelo harmônico ajustado não inclui a frequência 1/8. Na Figura ?? apresentamos os gráficos da série original, do modelo harmônico ajustado, da série dessazonalizada (resíduos após o ajuste do modelo harmônico) e do periodograma cumulativo da série dessazonalizada. Note que nenhuma sazonalidade óbvia pode ser percebida na série dessazonalizada e que o periodograma cumulativo da série dessazonalizada aponta a necessidade de uma modelagem adicional.

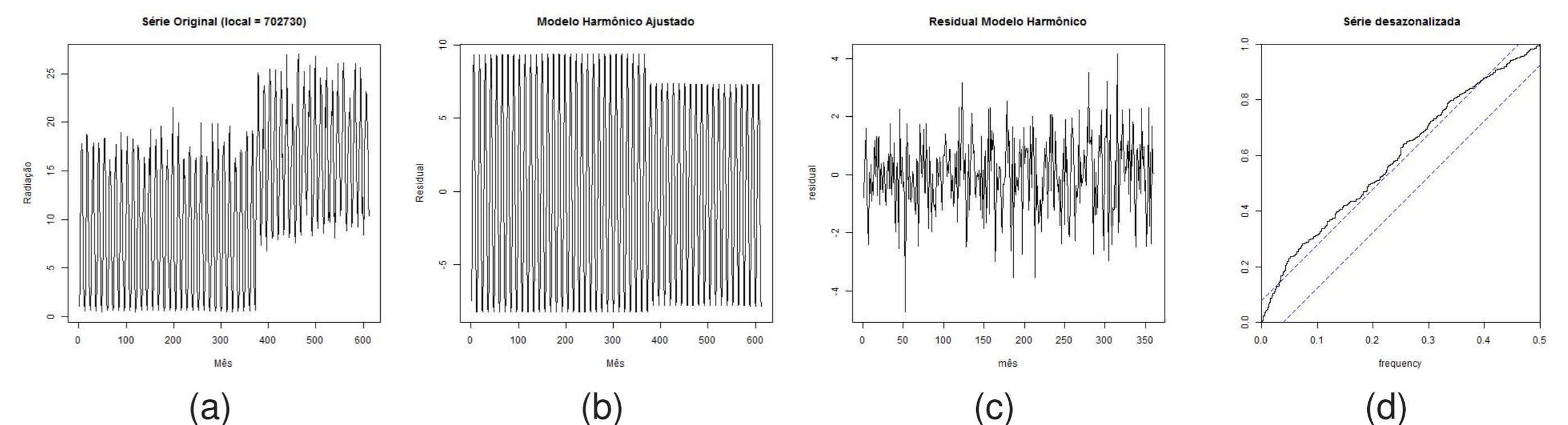


Figura: (a) Série original; (b) Modelo harmônico ajustado; (c) Série dessazonalizada; (d) Periodograma cumulativo da série dessazonalizada.

Seguimos com o ajuste dos modelos ARFIMA(p, d, q), para todas as combinações de $p, q \in \{0, 1, 2\}$, fazendo uso da função `fracdiff` do pacote homônimo do R. Neste caso, apenas os modelos ARFIMA(0, d , 0) e ARFIMA(1, d , 1) apresentaram todos os seus coeficientes significativos. Após a exclusão dos modelos com algum parâmetro não-significativo, procedemos a análise de resíduos através do periodograma cumulativo de cada modelo restante. Como regra geral, daqueles que bem ajustaram os resíduos do modelo harmônico, foi escolhido o modelo com maior log-verossimilhança ou, no caso em que estes são muito próximos entre si, o modelo mais parcimonioso. Na Figura ?? mostramos os gráficos do periodograma cumulativo dos resíduos dos modelos restantes. Observamos que o modelo ARFIMA(0, d , 0) e ARFIMA(1, d , 1) se ajustaram bem à série. Além da diferença dos valores da log-verossimilhança entre os modelos ser muito pequena, o modelo ARFIMA(0, d , 0) leva grande vantagem em ser mais parcimonioso e é o modelo escolhido como mais adequado aos dados.

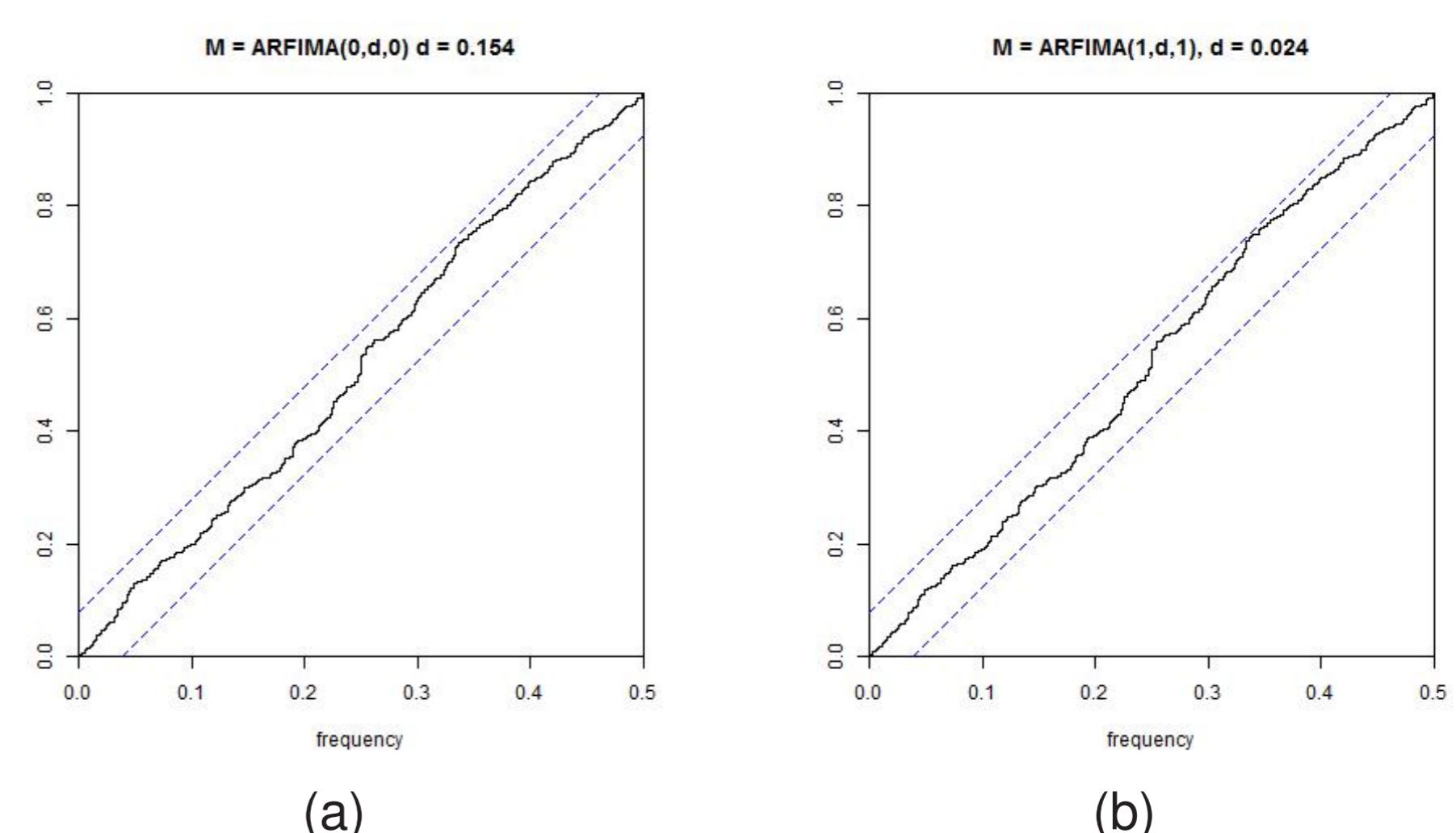


Figura: Análise residual dos modelos ARFIMA restantes: (a) ARFIMA(0, d , 0); (b) ARFIMA(1, d , 1).

Resultados e Conclusão

Os resultados da análise feita foram bastante positivos. Ambos os modelos possibilitaram o ajuste de todas as série do banco de dados e, em maioria, os modelos selecionados foram os mais parcimoniosos dentre os considerados. Com o primeiro modelo em 68 % dos casos a série normalizada foi modelada pelo modelo de menor complexidade, AR(1), e com o segundo em 28% dos casos, após dessazonalização da série pelo modelo harmônico, um simples ARFIMA(0, d , 0) foram suficientes.

Referências

- Beran, J. (1994). *Statistics for long-memory processes*. New York: Chapman & Hall.
- Bloomfield, P. (2000). *Fourier Analysis of Time Series: an Introduction*. 2nd edn. New Jersey: Wiley.
- Brockwell, P.J. e Davis, R.A. (1991). *Time Series: theory and methods*. 2nd edn. New York: Springer-Verlag.
- Doukhan, P., Oppenheim, G. e Taqqu, M.S. (2003). *Theory and Applications of Long-range Dependence*. Boston: Birkhäuser.
- Graybill, F. (2000). *Theory and Applications of the Linear Model*. Boston: Duxbury.
- Lopes, S.R.C. (2008). *Long-Range Dependence in Mean and Volatility: Models, Estimation and Forecasting*. Progress in Probability, v. 60, p. 497-525.