

## Introdução

Este trabalho apresenta uma análise univariada de séries temporais de medições de radiação solar fornecidas pelo *National Solar Radiation Database*, nos Estados Unidos. Os dados foram coletados em 249 localidades do país, a 213 diferentes latitudes, entre os anos de 1960 e 1990, sendo cada valor de medição a média mensal da radiação solar medida diariamente às 14h do horário local de cada posto de medição. Para analisar os dados seguimos duas propostas. Primeiramente, dada a série  $\{y_t\}_{t=1}^{360}$ , aplicamos uma normalização da série original por meio das médias e variâncias observadas em relação aos meses, isto é, calculamos

$$\mu_i = \frac{1}{30} \sum_{k=0}^{29} y_{12k+i} \quad \text{e} \quad \sigma_i^2 = \frac{1}{29} \sum_{k=0}^{29} (y_{12k+i} - \mu_i)^2, \quad i = 1, \dots, 12, \quad (1)$$

e definimos a série normalizada  $\{z_t\}_{t=1}^{360}$  por

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

Após a normalização, analisamos 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  $\{x_t\}_{t=1}^N$  é a transformada discreta de Fourier da série, a saber,

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

e o periodograma cumulativo padronizado é a função dada por

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

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  $k_\alpha$  denota o  $\alpha$ -ésimo quantil da distribuição da estatística de Kolmogorov-Smirnov.

Na segunda metodologia, começamos 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 variâncias relativas aos meses, respectivamente denotadas por  $\mu_i$  e  $\sigma_i^2$ ,  $i = 1, \dots, 12$  e calculadas como em (1).

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

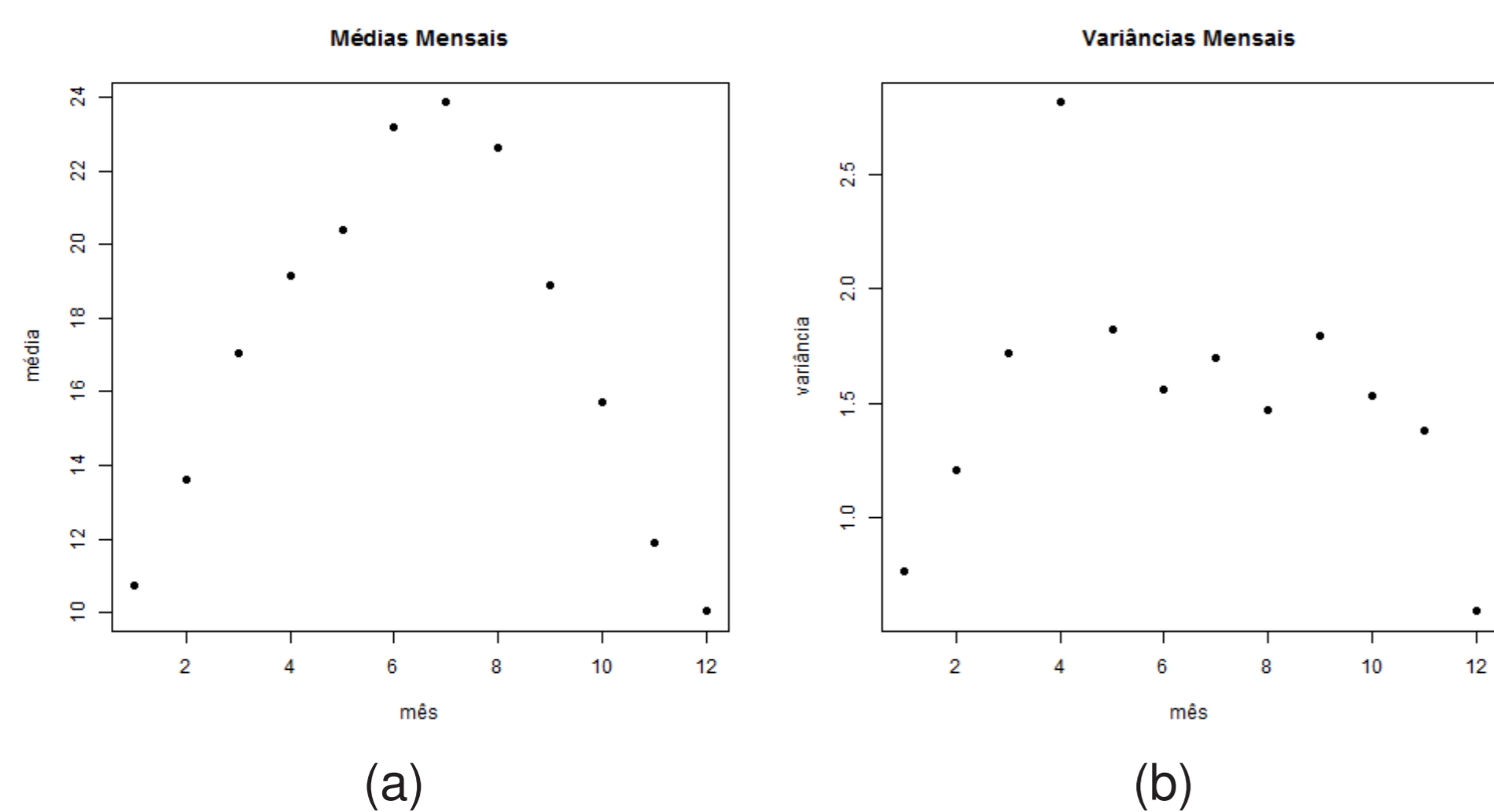


Figura: (a) Médias relativas aos meses; (b) Variâncias relativas aos meses.

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.

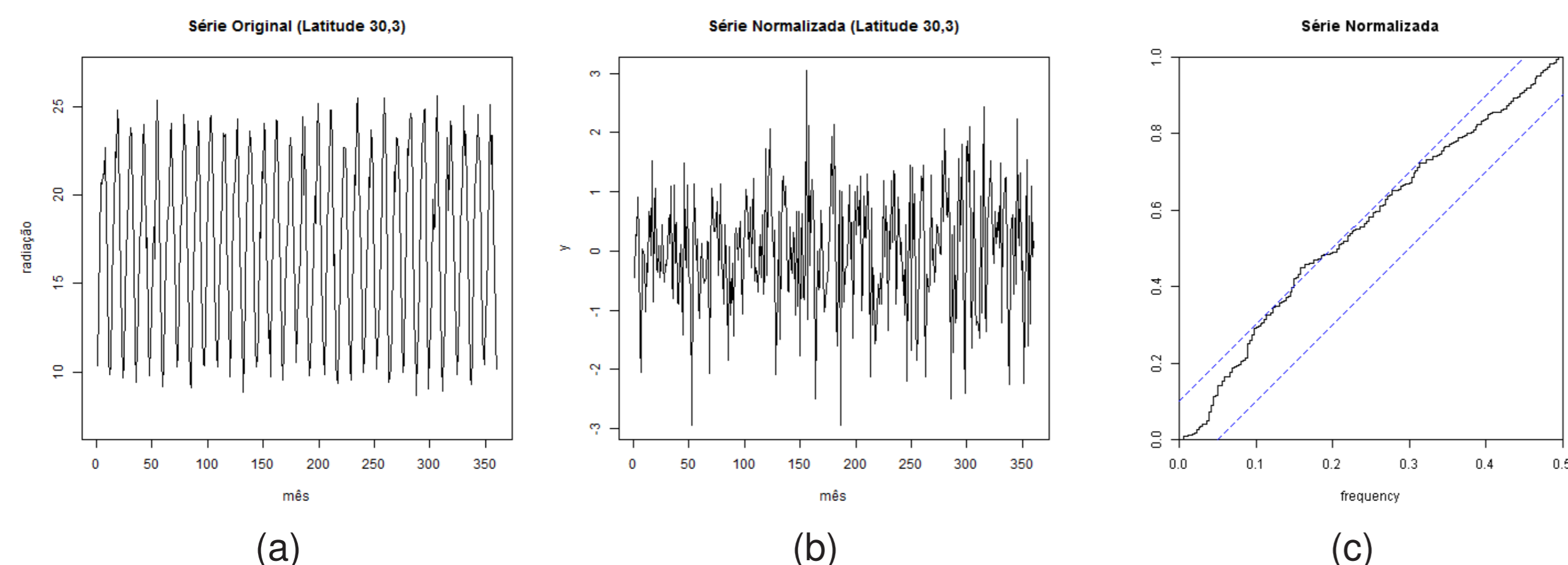


Figura: (a) Série original  $\{y_t\}$ ; (b) Série normalizada  $\{z_t\}$ ; (c) Periodograma cumulativo da série dessazonalizada.

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

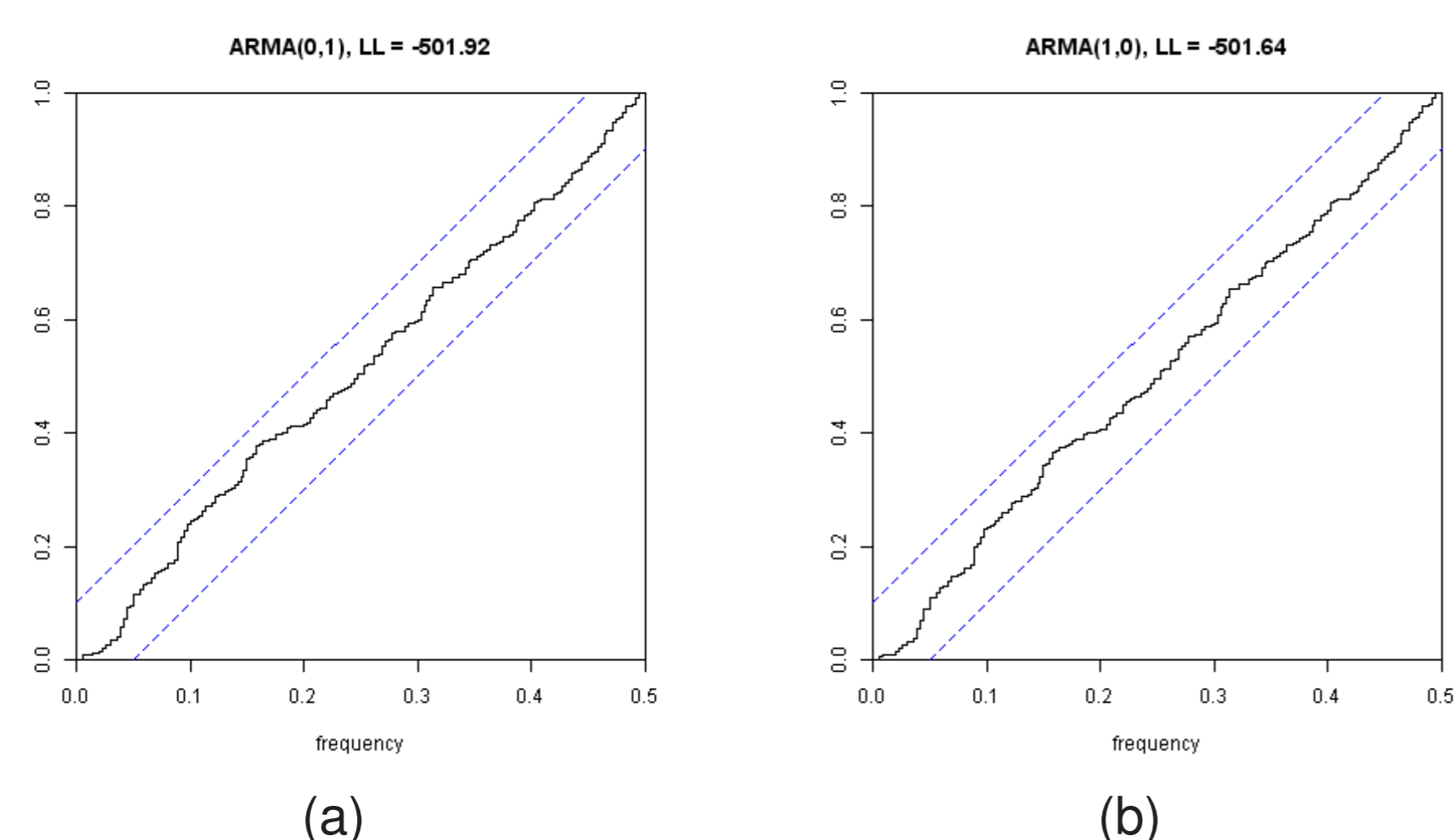


Figura: Periodogramas cumulativos dos modelos (a) ARMA(0, 1); (b) ARMA(1, 0) ajustados.

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 3 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: -501,92 e -501,64, 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.

```
Residuals:
  Min       1Q   Median       3Q      Max
-4.7284 -0.7796  0.0445  0.8565  4.1484

Coefficients:
  Estimate Std. Error t value Pr(>|t|)
A1 -6.34770  0.09562  -66.385  < 2e-16 ***
B1 -1.48146  0.09562  -15.494  < 2e-16 ***
A2 -0.78429  0.09562   -8.2020  4.36e-15 ***
B2  0.56166  0.09561   5.8740  9.74e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.283 on 356 degrees of freedom
Multiple R-squared: 0.9303, Adjusted R-squared: 0.9295
F-statistic: 1187 on 4 and 356 DF, p-value: < 2.2e-16
```

Neste caso, portanto, o modelo harmônico ajustado não inclui a frequência 1/8. Na Figura 4 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.

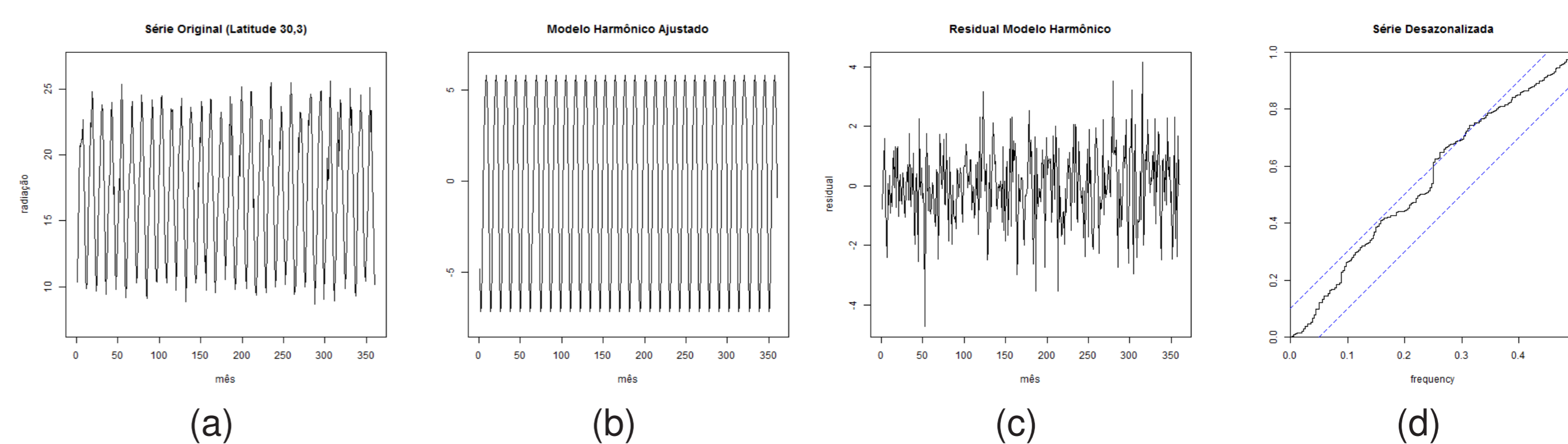


Figura: (a) Série original; (b) Modelo harmônico ajustado; (c) Série dessazonalizada; (d) Periodograma cumulativo da série dessazonalizada.

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 5 mostramos os gráficos do periodograma cumulativo dos resíduos dos modelos restantes. Observamos que o modelo ARFIMA(0,  $d$ , 0) se ajustou bem à série, ainda que o periodograma cumulativo de seus resíduos esteja apenas marginalmente dentro dos limites. O modelo ARFIMA(1,  $d$ , 1) apresenta um ajuste bom, porém o coeficiente da parte MA fica muito próximo do círculo unitário, o que indica que houve uma super-diferenciação no parâmetro de longa dependência ajustado. O modelo ARFIMA(0,  $d$ , 0) é o mais parcimonioso, portanto foi o escolhido como mais adequado aos dados.

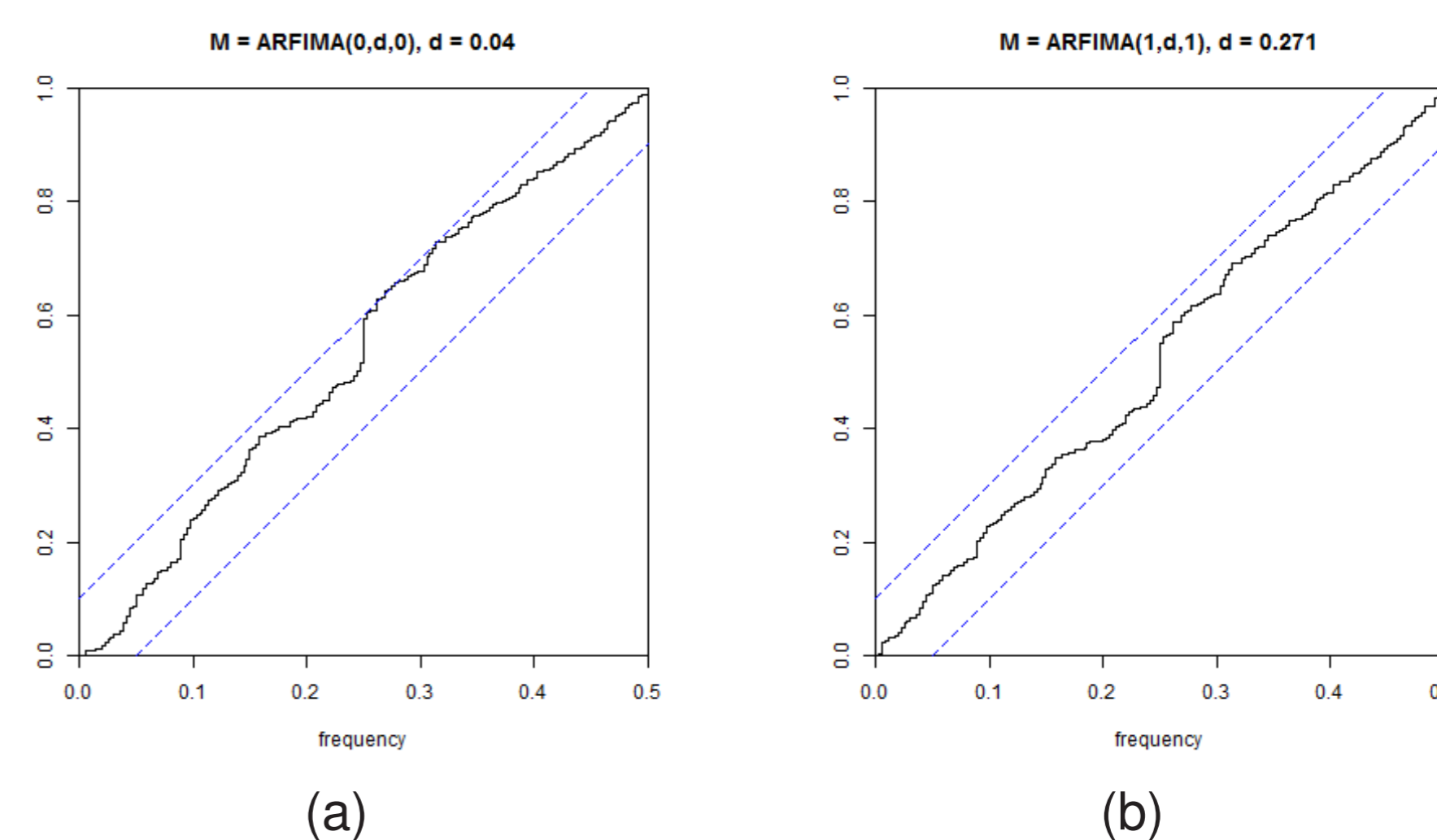


Figura: Análise residual dos modelos ARFIMA restantes: (a) ARFIMA(0,  $d$ , 0); (b) ARFIMA(1,  $d$ , 1).

## Resultados e Conclusão

Até o momento, os resultados mostraram-se bastante positivos. Em geral, modelos muito simples foram suficientes para a análise, uma vez que apenas uma normalização da série ou um simples modelo harmônico já ajustava os dados, na maioria dos casos. Quando necessário ajuste adicional, quase sempre modelos tão simples quanto um AR(1) ou um ARFIMA(0,  $d$ , 0) foram suficientes. Na primeira metodologia, todas as séries foram ajustadas. Em aproximadamente 24% dos casos, a série normalizada já podia ser considerada um ruído branco e em 66% dos casos, um modelo AR(1) ajustava os resíduos. Apenas 1 das 249 séries modeladas necessitou o uso do modelo ARMA(2, 2), o de maior complexidade. Na segunda metodologia, em mais de 55% dos casos, um modelo harmônico foi suficiente para o ajuste e, em mais de 39%, um ARFIMA(0,  $d$ , 0) ajustou os resíduos. Houve 3 latitudes cujas as séries não puderam ser ajustadas pelo segundo procedimento. O prosseguimento deste trabalho conta com a já corrente atualização do banco de dados, que passará a incluir as séries respectivas aos anos de 1990 à 2010. Pretendemos então estender o estudo já feito para os novos dados e avaliar o poder de previsão dos modelos aqui ajustados. Posteriormente, também queremos analisar esses dados como observações de uma série temporal multivariada, em busca de avanços em termos de previsões em relação a modelos univariados.

## Referências

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