A&A 627, A120 (2019)
https://doi.org/10.1051/0004-6361/201935560

Astronomy & Astrophysics

Discrete-time autoregressive model for unequally spaced time-series observations

Felipe Elorrieta[1,3], Susana Eyheramendy[2,3,4,5], and Wilfredo Palma[2,3]

[1] Departmento de Matemáticas, Facultad de Ciencia, Universidad de Santiago de Chile, Av. Libertador Bernardo O'Higgins 3663. Estacion Central, Santiago, Chile
e-mail: susana@mat.puc.cl
[2] Departmento de Estadística, Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 7 820436 Macul, Santiago, Chile
[3] Millennium Institute of Astrophysics, Santiago, Chile
[4] Max-Planck-Institut für Astronomie, Heidelberg, Germany
[5] Faculty of Engineering and Sciences, Universidad Adolfo Ibañez, Diagonal Las Torres 2700, Peñalolén, Santiago, Chile

Received 27 March 2019 / Accepted 14 June 2019

ABSTRACT

Most time-series models assume that the data come from observations that are equally spaced in time. However, this assumption does not hold in many diverse scientific fields, such as astronomy, finance, and climatology, among others. There are some techniques that fit unequally spaced time series, such as the continuous-time autoregressive moving average (CARMA) processes. These models are defined as the solution of a stochastic differential equation. It is not uncommon in astronomical time series, that the time gaps between observations are large. Therefore, an alternative suitable approach to modeling astronomical time series with large gaps between observations should be based on the solution of a difference equation of a discrete process. In this work we propose a novel model to fit irregular time series called the complex irregular autoregressive (CIAR) model that is represented directly as a discrete-time process. We show that the model is weakly stationary and that it can be represented as a state-space system, allowing efficient maximum likelihood estimation based on the Kalman recursions. Furthermore, we show via Monte Carlo simulations that the finite sample performance of the parameter estimation is accurate. The proposed methodology is applied to light curves from periodic variable stars, illustrating how the model can be implemented to detect poor adjustment of the harmonic model. This can occur when the period has not been accurately estimated or when the variable stars are multiperiodic. Last, we show how the CIAR model, through its state space representation, allows unobserved measurements to be forecast.

Key words. methods: statistical – methods: data analysis – stars: general

1. Introduction

Time-series theory is vast and provides a myriad of methods to model the serial correlation of a temporal sequence. Most of these methods assume that the observed values are measured regularly in time. However, there are many cases where this assumption is not valid. When this happens, we say that the time series is sampled irregularly.

An irregular time series is defined as a real valued sequence $\{y_{t_i}\}_{i=1}^{n}$ measured at observational times $t_1, \ldots, t_n$, such that the sequence $t_1, \ldots, t_N$ is strictly increasing and the distance between consecutive times $t_j - t_{j-1}$ is not constant.

Irregular time series can be observed in many disciplines. For example, natural disasters such as earthquakes do not occur regularly in time. Paleoclimate data are often sampled irregularly due to the difficulty in obtaining historical information. In astronomy, objects can be observed irregularly in time because of, for example, the dependency of optical telescopes on clear skies.

Analyzing irregular time series is challenging because there are still few robust statistical tools available. In astronomy, some efforts have been made that attempt to model irregular time series. One approach is to transform the irregular times into regular ones by interpolation (Rehfeld et al. 2011). Other approaches implement Gaussian processes to fit the light curves in the time domain (Foreman-Mackey et al. 2017), but this solution can be computationally expensive. In some cases, regular time-series models have also been used to model astronomical data (for a review of such methods see e.g., Feigelson et al. 2018). Alternatively, the CARMA family of models has been popular for fitting astronomical time series (Kelly et al. 2014), however these models are defined as the solution of a differential stochastic equation, underlying the assumption of small time gaps between observations. Therefore, we consider it important to develop alternative models that can fit irregularly spaced time series under the assumption of discrete time gaps.

In Eyheramendy et al. (2018) we introduced a new model called the Irregular Autoregressive (IAR) model to fit unequally spaced time series. The IAR model is a discrete representation of the continuous autoregressive model of order 1 (CAR(1)), which is strictly stationary and ergodic. The IAR has some extra flexibility when compared with the CAR(1). In the model assumptions, the IAR can assume distributions other than Gaussian[1].

The IAR model has two parameters that need to be estimated from the observed series. One parameter represents the variance of the process and the other one the value of the autocorrelation function (ACF). A limitation of the IAR model (and also of

[1] We use the term Gaussian data to make reference to normally distributed time-series data.

the CAR(1) model) is that the ACF derived from the estimated model parameters – via Eq. (1) – is positive only. This is a limitation that the regular autoregressive model does not have; the latter can detect both positive and negative values of the ACF.

Negative values of the ACF appear often in some disciplines. For instance, they have been detected in some physical phenomena, such as the velocity for hard spheres (e.g., Williams et al. 2006; Alder & Wainwright 1970). In financial time series, it is common to find significant negative autocorrelation for weekly and monthly stock returns (further discussion can be found in Sewell 2011). Negative values of the ACF are also generally observed in stocks with a high trading frequency (e.g., Conrad et al. 1994; Campbell et al. 1997).

Well-known examples of time series with negative values for the ACF are the antipersistent processes, which are characterized as having negative values of the ACF for all positive lags (Bondon & Palma 2007). One of the best known anti-persistent processes is the Kolmogorov energy spectrum of turbulence (Gao et al. 2007). There are several examples of antipersistent processes in meteorology. For instance, Ausloos & Ivanova (2001) detect antipersistence in the fluctuations of the Southern Oscillation Index, such as the sea level pressure. Carvalho et al. (2007) also detected antipersistence in temperature anomalies. Other examples are in financial data; for example, the electricity prices in some Canadian provinces show an antipersistent behavior (e.g., Uritskaya & Uritsky 2015).

The problem of detecting negative values of the ACF in irregularly sampled time series has scarcely been addressed in the literature. Chan & Tong (1987) showed that a discrete-time regular autoregressive model of order 1 (AR(1)) with negative coefficient can always be embedded in a suitably chosen continuous-time ARMA(2,1) process, but this is not necessarily a parsimonious solution. Alternatively, when an irregular time series has an antipersistent behavior it can be fitted by the continuous-time ARFIMA (CARFIMA) process (Tsai 2009) with an intermediate memory,that is, the Hurst parameter H is such that $0 < H < 1/2$. In this work we propose another alternative that corresponds to an extension of the IAR model. We call this model the Complex Irregular Autoregressive (CIAR) model because it needs complex numbers to represent the series.

Complex stochastic processes have been addressed by several authors, such as Miller (1974) and Picinbono & Bondon (1997). These processes are often used in signal-processing analyses since for example in telecommunication systems it is common to find complex signals. In addition, Dubois & Glanz (1986) and Sekita et al. (1991) introduced a complex extension of a regular autoregressive model, where they proposed to use the coefficients of these models for shape recognition. Furthermore, Martin (1999) proposed a method to estimate the parameters of the CAR(1) model in a complex time series. However, in these studies it is assumed that both the real and the imaginary part of the complex process are observed. The model proposed here assumes that only the real part of the process is observed and the imaginary part is a latent process.

In this study, we apply the CIAR model to astronomical data, but the model can be applied in any other field in which irregular time series are available. In astronomy, the analysis of the temporal behavior of variable stars, transients, or supernovae attracts significant interest because several properties of the objects can be obtained from the analysis of their light curves (e.g., Richards et al. 2011; Kelly et al. 2014; Elorrieta et al. 2016).

In particular, in the analysis of variable stars the temporal modeling of the light curves is an important step to classify them. From a harmonic model fitted to the light curves of periodic variable stars, several features are extracted which describe the physical behavior of a specific class of variable stars. We implement the CIAR model on the residuals of the harmonic model fitted on variable stars from OGLE (Udalski et al. 1999) and Hipparcos (Perryman et al. 1997). Under this scenario, our aim is to assess whether the harmonic model is capable of describing the temporal structure of the light curves of variable stars or some autocorrelation remains on the residuals.

The structure of this paper is as follows. In Sect. 2 we present the irregular discrete time series models: the IAR model proposed by Eyheramendy et al. (2018) and the CIAR model. In both cases the maximum likelihood estimation procedure is described. In the case of the CIAR model, we also provide expressions for forecasting unobserved measurements. In Sect. 3 we assess the variance and bias of the estimated parameters of the CIAR model using Monte Carlo simulations. Furthermore, we perform simulations in order to compare the performance in fitting negative values of the ACF of well-known time series models and the CIAR model. In Sect. 4 we apply the CIAR model to light curves of variable stars from the OGLE and Hipparcos surveys. We also implement the CIAR model on an AGN to demonstrate the forecasting procedure. We compare the results obtained with the IAR model and illustrate some applications for these data. The article ends with a discussion and proposals for future work in Sect. 5.

2. Methods

2.1. Irregular Autoregressive (IAR) model

Letting $\{y_{t_j}\}$ be a time series observed at irregular times, where $\{t_j\}$ is an increasing sequence of observational times for $j = 1, \ldots, n$, the irregular autoregressive (IAR) process is defined as

$$y_{t_j} = \phi^{t_j - t_{j-1}}\, y_{t_{j-1}} + \sigma_y \sqrt{1 - \phi^{2(t_j - t_{j-1})}}\, \varepsilon_{t_j}, \tag{1}$$

where ϕ is the parameter of the ACF and ε_{t_j} is a white noise sequence[2] with zero mean and unit variance. We note that

$$\mathbb{E}(y_{t_j}) = 0 \text{ and } \mathbb{V}(y_{t_j}) = \sigma_y^2 \quad \forall t_j.$$

The autocovariance function between two observational times, s and t, with $s < t$, is given by $\gamma(t-s) = E(y_t\, y_s) = \sigma_y^2\, \phi^{t-s}$, and the ACF, $\rho(t-s) = \frac{\gamma(t-s)}{\gamma(0)} = \phi^{t-s}$. Therefore, if $0 < \phi < 1$ the sequence $\{y_{t_j}\}$ corresponds to a second-order or weakly stationary process, that is, the time series has constant mean and finite second moment, and possesses an autocovariance function.

Furthermore, under some regularity conditions, the process $\{y_{t_j}\}$ is strictly stationary and ergodic (Eyheramendy et al. 2018).

We note that the IAR process is an extension of the regular autoregressive process. If $t_j - t_{j-1} = 1$ is assumed, the IAR process becomes the autoregressive model of order 1 (AR(1)). Also, the IAR process is equivalent to the continuous autoregressive process of order 1 (CAR(1)) when a Gaussian distribution is assumed on the white noise sequence ε_{t_j}. However, the IAR process is more flexible since it allows also nonGaussian data (Eyheramendy et al. 2018).

The finite past predictor of the process at time t_j is given by,

$$\widehat{y}_{t_j} = \phi^{t_j - t_{j-1}}\, y_{t_{j-1}}, \text{ for } j = 2, \ldots, n, \tag{2}$$

where the initial value is $\widehat{y}_{t_1} = 0$. Consequently, $e_{t_1} = y_{t_1}$ and $v_{t_1} = \mathrm{Var}(e_{t_1}) = \sigma_y^2$. Furthermore, $e_{t_j} = y_{t_j} - \widehat{y}_{t_j}$ is the innovation with variance $v_{t_j} = \mathrm{Var}(e_{t_j}) = \sigma_y^2[1 - \phi^{2(t_j - t_{j-1})}]$.

[2] A white noise is an uncorrelated weakly stationary process.

The estimation of the model parameters $\theta = (\sigma_y^2, \phi)$ can be performed by maximum likelihood. Minus the log-likelihood of the process when a Gaussian distribution is assumed on ε_{t_j} is given by

$$\ell(\theta) = \frac{n}{2}\log(2\pi) + \frac{1}{2}\sum_{j=1}^{n}\log v_{t_j} + \frac{1}{2}\sum_{j=1}^{n}\frac{e_{t_j}^2}{v_{t_j}}. \tag{3}$$

For other distributional assumptions it can be written similarly. We can obtain the maximum likelihood estimator of σ_y^2 by directly maximizing the log-likelihood (3),

$$\hat{\sigma_y}^2 = \frac{1}{n}\sum_{j=1}^{n}\frac{(y_{t_j} - \widehat{y}_{t_j})^2}{\tau_{t_j}}, \text{ where } \tau_{t_j} = v_{t_j}/\sigma_y^2, \tag{4}$$

but it is not possible to find a closed form expression for the maximum likelihood estimator of ϕ. However, iterative methods can be used.

A drawback of this model is that ϕ can only take values in the interval Alder & Wainwright (1970), since a negative ϕ to the power of a real (and not integer) number is a complex number. Therefore, the IAR model only allows us to estimate positive values of the autocorrelation. To detect negative values of the ACF, the IAR model must be extended. In the following section we introduce the CIAR which allows for negative as well as positive autocorrelation to be modeled.

2.2. Complex Irregular Autoregressive (CIAR) model

To derive a complex extension of model (1), we follow the approach of Sekita et al. (1991), which builds a complex autoregressive model for regular times. Supposing that x_{t_j} is a complex valued sequence, such that, $x_{t_j} = y_{t_j} + iz_{t_j}\ \forall j = 1, \ldots, n$, and likewise, $\phi = \phi^R + i\phi^I$ is the complex coefficient of the model and $\varepsilon_{t_j} = \varepsilon_{t_j}^R + i\varepsilon_{t_j}^I$ is a complex white noise, we define the CIAR process as,

$$y_{t_j} + iz_{t_j} = (\phi^R + i\phi^I)^{t_j - t_{j-1}}(y_{t_{j-1}} + iz_{t_{j-1}}) + \sigma_{t_j}(\varepsilon_{t_j}^R + i\varepsilon_{t_j}^I), \tag{5}$$

where $\sigma_{t_j} = \sigma\sqrt{1 - |\phi^{t_j - t_{j-1}}|^2}$ and $|.|$ is the modulus of a complex number. We assume that only the real part y_{t_j} is observed and that the imaginary part z_{t_j} is a latent process. In addition, $\varepsilon_{t_j}^R$ and $\varepsilon_{t_j}^I$, the real and imaginary part of ε_{t_j}, respectively, are assumed to be independent with zero mean and positive variance $\mathbb{V}(\varepsilon_t^R) = 1$ and $\mathbb{V}(\varepsilon_t^I) = c$, respectively, where c is a fixed parameter that takes values in $\mathbb{R}^+$. Generally, we assume $c = 1$, and the initial values are $y_{t_1} = \sigma\varepsilon_{t_1}^R$ and $z_{t_1} = \sigma\varepsilon_{t_1}^I$. In the following lemma we state some of the properties of this process.

Lemma 1. Consider the CIAR process x_{t_j} described by Eq. (5). Define $\gamma_0 = \mathbb{E}(\overline{x}_{t_j}x_{t_j})$, $\gamma_k = \mathbb{E}(\overline{x}_{t_{j+k}}x_{t_j})$, and ρ_k as the variance, autocovariance, and autocorrelation, respectively, of the process x_{t_j}. Subsequently, the expected value, the variance, the autocovariance, and autocorrelation of the process respectively satisfy

(a) $\mathbb{E}(x_{t_j}) = 0$,
(b) $\mathbb{V}(x_{t_j}) = \gamma_0 = \mathbb{E}(\overline{x}_{t_j}x_{t_j}) = \sigma^2(1 + c)$,
(c) $\gamma_k = \mathbb{E}(\overline{x}_{t_{j+k}}x_{t_j}) = \overline{\phi^{\Delta_k}}\sigma^2(1 + c)$,
(d) $\rho_k = \overline{\phi^{\Delta_k}}$,

where $\Delta_k = t_{j+k} - t_j$ denotes the time differences between the observational times t_{j+k} and t_j. In addition, $\overline{x}_{t_j}$ is the complex conjugate of x_{t_j}. See Appendix A for a proof of this lemma.

If $|\phi| = |\phi^R + i\phi^I| < 1$, the results on Lemma 1 above show that the complex sequence x_{t_j} is a weakly stationary process. We note that the ACF ρ_k of the CIAR process decays at a rate $\phi^{t_{j+k} - t_j}$ (the so-called exponential decay). This autocorrelation structure is different to an antipersistent or intermediate memory CARFIMA process, since the ACF of the latter decays more slowly than an exponential decay. Thus, although both models can fit irregular time series with negative values of the ACF, the appropriate use of these models will depend on the correlation structure of the data. To make a decision about the most suitable model for the data, techniques based on the likelihood of the data, such as AIC, BIC, and so on, can be used.

In what follows, we express the CIAR model in terms of a state-space system. This representation allows us: (i) to implement the Kalman filter to obtain maximum likelihood estimators of the parameters of the model; and (ii) to forecast unobserved measurements.

2.3. State-space system

A linear state-space system may be described by the following equations.

$$X_t = F_t X_{t-1} + V_t, \tag{6}$$
$$Y_t = GX_t + W_t, \tag{7}$$

where Eq. (6) is known as the state equation which determines a v-dimensional state variable X_t. Equation (7) is called the observation equation, which expresses the w-dimensional observation Y_t. In addition, F_t is a sequence of $v \times v$ matrices called the transition matrices, and $G \in \mathbb{R}^{w \times v}$ is the observation linear operator of the observation matrix. Finally, $W_t \sim WN(0, R_t)$, $V_t \sim WN(0, Q_t)$ and V_t is uncorrelated with W_t.

In order to represent the CIAR model in a state-space system we need to rewrite Eq. (5). In the following lemma an alternative way of writing the CIAR model is proposed.

Lemma 2. The CIAR process described by Eq. (5) can be expressed by the following equation.

$$y_{t_j} + iz_{t_j} = (\alpha_{t_j}^R + i\alpha_{t_j}^I)(y_{t_{j-1}} + iz_{t_{j-1}}) + \sigma_{t_j}(\varepsilon_{t_j}^R + i\varepsilon_{t_j}^I), \tag{8}$$

where $\alpha_{t_j}^R = |\phi|^{\delta_j}\cos(\delta_j\psi)$, $\alpha_{t_j}^I = |\phi|^{\delta_j}\sin(\delta_j\psi)$, $\delta_j = t_j - t_{j-1}$, $\psi = \arccos\left(\frac{\phi^R}{|\phi|}\right)$ and $\phi = \phi^R + i\phi^I$. See Appendix B for a proof of this lemma.

By following the representation of the CIAR model described in Lemma 2, we can express the observed process as

$$y_{t_j} = \alpha_{t_j}^R y_{t_{j-1}} - \alpha_{t_j}^I z_{t_{j-1}} + \sigma_{t_j}\varepsilon_{t_j}^R, \tag{9}$$

and the latent process as

$$z_{t_j} = \alpha_{t_j}^I y_{t_{j-1}} + \alpha_{t_j}^R z_{t_{j-1}} + \sigma_{t_j}\varepsilon_{t_j}^I. \tag{10}$$

We note that the observed process y_{t_j} is an IAR with parameter ϕ, if we assume $\alpha_{t_j}^I = 0$ and $\alpha_{t_j}^R = \phi^{t_j - t_{j-1}}$. In addition, it is straightforward to show that $\alpha_{t_j}^I = 0$ is equivalent to $\phi_I = 0$.

Another important consideration is that the observed process y_{t_j} does not depend directly on $\varepsilon_{t_j}^I$. Consequently, the variance of the imaginary part c is a nuisance parameter, in the sense that it takes any value in $\mathbb{R}^+$ and does not cause significant changes in the model.

Equation (8) can be represented by the state-space system of Eqs. (6)–(7) assuming $t = t_j$ and $X_{t_j} = \begin{pmatrix} y_{t_j} \\ z_{t_j} \end{pmatrix}$. Given that only y_{t_j} are actually observed, we obtain $Y_t = y_{t_j}$. Therefore, $G = (1\,0)$ is the observation matrix under this representation.

To complete the specification we define the transition matrix as $F_{t_j} = \begin{pmatrix} \alpha^R_{t_j} & -\alpha^I_{t_j} \\ \alpha^I_{t_j} & \alpha^R_{t_j} \end{pmatrix}$ and the noise of Eqs. (6) and (7) as $V_{t_j} = \sigma_{t_j} \begin{pmatrix} \varepsilon^R_{t_j} \\ \varepsilon^I_{t_j} \end{pmatrix}$ and $W_{t_j} = 0$, respectively. Finally, the observation and state equations of the state-space representation of the CIAR model are,

$$\begin{pmatrix} y_{t_j} \\ z_{t_j} \end{pmatrix} = \begin{pmatrix} \alpha^R_{t_j} & -\alpha^I_{t_j} \\ \alpha^I_{t_j} & \alpha^R_{t_j} \end{pmatrix} \begin{pmatrix} y_{t_{j-1}} \\ z_{t_{j-1}} \end{pmatrix} + \sigma_{t_j} \begin{pmatrix} \varepsilon^R_{t_j} \\ \varepsilon^I_{t_j} \end{pmatrix}, \tag{11}$$

$$y_{t_j} = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t_j} \\ z_{t_j} \end{pmatrix}. \tag{12}$$

We note that in this representation, the transition matrix and the variance of noise term of the state equation $Q_{t_j} = |\sigma_{t_j}|^2 \mathbb{V}\left(\varepsilon_{t_j}\right)$ depend on time.

Lemma 3. Now we let $\alpha_{t_j} = \alpha^R_{t_j} + i\alpha^I_{t_j}$. If $\sup|\alpha_{t_j}| < 1$, the process in Eqs. (11)–(12) is stable and has a unique stationary solution given by

$$X_{t_j} = V_{t_j} + \sum_{k=1}^{\infty} V_{t_{j-k}} \prod_{i=0}^{k-1} F_{t_{j-i}} \tag{13}$$

where $V_{t_{j-k}} = \sigma_{t_{j-k}} \begin{pmatrix} \varepsilon^R_{t_{j-k}} \\ \varepsilon^I_{t_{j-k}} \end{pmatrix}$. Appendix C presents a proof of this lemma.

2.4. Estimation

For the state-space model of Eqs. (6)–(7), the one-step predictors $\hat{X}_{t_j} = P_{t_{j-1}}(X_{t_j})$ and their error covariance matrices $\Omega_{t_j} = \mathbb{E}[(X_{t_j} - \hat{X}_{t_j})(X_{t_j} - \hat{X}_{t_j})']$ are unique and determined by the initial values: $\hat{X}_{t_1} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\Omega_{t_1} = \mathbb{E}[(X_{t_1} - \hat{X}_{t_1})(X_{t_1} - \hat{X}_{t_1})']$. Using the properties of model (5) we can rewrite Ω_{t_1} as,

$$\Omega_{t_1} = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix}.$$

The Kalman recursions, for $j = 1, \ldots, n-1$ are defined as follows.

$$\begin{aligned} \Lambda_{t_j} &= G_{t_j} \Omega_{t_j} G'_{t_j}, \\ \Theta_{t_j} &= F_{t_j} \Omega_{t_j} G'_{t_j}, \\ \Omega_{t_{j+1}} &= F_{t_j} \Omega_{t_j} F'_{t_j} + Q_{t_j} - \Theta_{t_j} \Lambda^{-1}_{t_j} \Theta'_{t_j}, \\ \nu_{t_j} &= y_{t_j} - G_{t_j} \hat{X}_{t_j}, \\ \hat{X}_{t_{j+1}} &= F_{t_j} \hat{X}_{t_j} + \Theta_{t_j} \Lambda^{-1}_{t_j} \nu_{t_j}, \end{aligned} \tag{14}$$

where $\{\nu_{t_j}\}$ is called the innovation sequence.
The maximum likelihood estimators of the CIAR model parameters ϕ^R and ϕ^I can be obtained by minimizing the reduced likelihood defined as,

$$\ell(\phi) \propto \frac{1}{n} \sum_{j=1}^{n} \left(\log(\Lambda_{t_j}) + \frac{\nu^2_{t_j}}{\Lambda_{t_j}} \right),$$

where Λ_{t_j} and ν_{t_j} come from the Kalman recursion. We developed scripts in the statistical language/software R and in Python to perform the estimation of the parameters of the model using Kalman recursions.

Another feature of this model is that it allows forecasting, the process of which is explained below.

2.5. Forecasting

Using the space-state Eqs. (11)–(12) it is straightforward to define a forecasting expression for the CIAR model. We note that from the Kalman filter recursions (14) we obtain the estimates of the state $\hat{X}_{t_j}$, Λ_{t_j}, Θ_{t_j} and ν_{t_j} at the last time point t_n. A one-step forecasting for the CIAR model can be obtained from the following equations.

$$\begin{aligned} \hat{X}_{t_{n+1}} &= F_{t_{n+1}} \hat{X}_{t_n} + \Theta_{t_n} \Lambda^{-1}_{t_n} \nu_{t_n} \\ \hat{Y}_{t_{n+1}} &= G_{t_{n+1}} \hat{X}_{t_{n+1}}, \end{aligned} \tag{15}$$

where $F_{t_{n+1}} = |\phi|^{\delta} \begin{pmatrix} \cos(\delta\psi) & -\sin(\delta\psi) \\ \sin(\delta\psi) & \cos(\delta\psi) \end{pmatrix}$ and $\delta = t_{n+1} - t_n$.

In addition, a confidence interval for the forecast $\hat{Y}_{t_{n+1}}$ can be defined by,

$$\hat{Y}_{t_{n+1}} \pm z_{1-\alpha/2} \sqrt{\mathbb{V}(e_{t_{n+1}})}, \tag{16}$$

where $e_{t_{n+1}} = Y_{t_{n+1}} - \hat{Y}_{t_{n+1}}$ is the forecast error with variance $\mathbb{V}(e_{t_{n+1}}) = \sigma^2(1 - \left|\phi^{t_j - t_{j-1}}\right|^2)$ and $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution.

3. Simulation results

3.1. Assessing the estimation performance of the CIAR model

In this section, we assess the performance of the estimation procedure proposed for the CIAR model. We perform Monte Carlo experiments based on 1000 repetitions of each simulation. In each repetition we generate a CIAR sequence from the model (5) using coefficients with different positive and negative values for the real part ϕ^R. In addition, both the imaginary part of the coefficient and the imaginary variance are set to $\phi^I = 0$ and $c = 1$, and the real and imaginary errors are assumed to follow a Gaussian distribution with a mean equal to zero and a variance of one. The irregular times are generated using the following mixture of two exponential distributions,

$$f(t|\lambda_1, \lambda_2, \omega_1, \omega_2) = \omega_1 g(t|\lambda_1) + \omega_2 g(t|\lambda_2). \tag{17}$$

We choose $\lambda_1 = 15$ and $\lambda_2 = 2$ as the means of each exponential distribution, respectively, and $\omega_1 = 0.15$ and $\omega_1 = 0.85$ as their respective weights. Under these parameters, the mean of the time differences in the simulated data is ≈3.95. The rationale for choosing this distribution came from mimicking time gaps observed in surveys such as the VVV[3] where some observations are very close to each other followed by observations that are more separated from each other.

Table 1 shows the results of the Monte Carlo simulations, which suggest that the finite-sample performance of the proposed methodology is accurate, both for positive and negative values of the parameter ϕ^R. In addition, we also estimate the parameter ϕ of the IAR model in each simulation. A precise IAR

[3] https://vvvsurvey.org/

Table 1. Maximum likelihood estimation of complex ϕ computed by the CIAR model in the real part of the simulated CIAR data.

Case	N	ϕ^R	$\widehat{\phi^R}$	SD $(\widehat{\phi^R})$	ϕ^I	$\widehat{\phi^I}$	SD $(\widehat{\phi^I})$	$\widehat{\phi}_{\text{IAR}}$	SD $(\widehat{\phi}_{\text{IAR}})$	DCF	SD (DCF)
1	300	0.999	0.9949	0.0036	0	0.0009	0.0030	0.9949	0.0036	0.9743	0.1273
2	300	0.9	0.8960	0.0187	0	0.0116	0.0413	0.8950	0.0188	0.8809	0.1315
3	300	0.7	0.6967	0.0412	0	0.0557	0.0819	0.6948	0.0406	0.6909	0.1279
4	300	0.5	0.4942	0.0596	0	0.0849	0.1111	0.4965	0.0569	0.5004	0.1240
5	300	−0.999	−0.9984	0.0012	0	0.0001	0.0009	0.0626	0.0265	−0.6260	0.1038
6	300	−0.9	−0.8991	0.0154	0	0.0014	0.0134	0.0643	0.0299	−0.5644	0.1166
7	300	−0.7	−0.6991	0.0414	0	0.0061	0.0354	0.0628	0.0289	−0.4382	0.1114
8	300	−0.5	−0.4971	0.0717	0	0.0091	0.0607	0.0589	0.0283	−0.3152	0.1089

Notes. The observational times are generated using a mixture of exponential distribution with $\lambda_1 = 15$ and $\lambda_2 = 2$, $\omega_1 = 0.15$ and $\omega_2 = 0.85$.

coefficient estimate is obtained when ϕ^R is positive, but when the CIAR process is generated with negative ϕ^R the estimation of the IAR coefficient is close to zero. This result is important since it shows that the IAR model cannot detect negative values of the ACF, unlike the CIAR model. Finally, we note that the accuracy of the estimated values does not depend on the magnitude of the coefficient.

Another method commonly used for estimating autocorrelation in irregular time series is the discrete correlation function (DCF; Edelson & Krolik 1988). This method is based on computing a Pearson correlation coefficient between data pairs. Each data pair is composed by observations with a time difference in the interval $(\tau - \delta/2, \tau + \delta/2)$ where τ is the order of the autocorrelation and δ is the bin window size. In order to assess whether this method can detect the autocorrelation of the CIAR process, we implement the DCF using the package *sour* of R (Edelson et al. 2017). The last two columns in Table 1 correspond to the mean of the DCF estimates of order $\tau = 1$ and its standard deviation, respectively. We note that the DCF estimates are close to the ϕ^R parameter when it is positive, but with a large standard deviation. On the other hand, when ϕ^R is negative, the DCF estimates are not accurate.

3.2. Comparison of the CIAR with other time series models

We perform a simulation experiment to compare the performance of well-known time series models, that is, IAR, AR(1), ARFIMA and CAR(1), in fitting a CIAR process. This experiment consists in generating 1000 sequences of the CIAR process $\{y_1, \ldots, y_n\}$ with length $n = 300$. In order to generate positive and negative values of the ACF, $\phi^R = 0.99$ and $\phi^R = -0.99$ are used. The remaining parameters are set as $\phi^I = 0$, $c = 1$ and the irregular times are generated with a mixture of two exponential distributions. We then fit each sequence using the different time-series models, and the CIAR model. The AR(1), ARFIMA, and CAR(1) models were estimated by maximum likelihood using the functions `arima`, `arfima`, and `cts`, respectively, from the statistical software R. To compare its performance we compute the root mean squared error (RMSE). Figure 1a shows that the RMSE estimated by the irregular time-series models is smaller in comparison with that of either the AR or ARFIMA model which assume regular sampling. However, as can be seen in Fig. 1b the CIAR model shows significantly better performance in fitting the negatively correlated processes than the other implemented models. Therefore, we verify that a CIAR process with a large and negative value of ϕ^R cannot be correctly modeled with the conventional time-series models, including those that assume irregular sampling.

3.3. Computation of the autocorrelation in a harmonic model

An important advantage of the CIAR process over other time-series models for unequally spaced observation is its ability to model weakly stationary time series with negative as well as positive values of the ACF. A well-known example of a time series that can be negatively correlated is the following harmonic process,

$$y_{t_i} = A \sin(f t_i + \psi) + \epsilon_{t_i}, \tag{18}$$

where f is the frequency of the process and ϵ_{t_i} is a Gaussian sequence with a mean of zero and a variance of σ^2. In addition, the amplitude A is a fixed parameter and the phase ψ is a random variable with uniform distribution between $-\pi$ and π. We note that the weakly stationarity of the process y_{t_i} is guaranteed when this distribution for ψ is assumed (for more details, see e.g., Lindgren et al. 2013). Assuming regular observational times, the one-step autocorrelation is given by $\rho_1 = \cos(f)$ (Broersen 2006). This result is also satisfied under irregular times. We note that this autocorrelation is negative for $f \in (\pi/2, \pi)$. In addition, we note that for higher-frequency values the harmonic process (18) becomes more anti-persistent (Alperovich et al. 2017).

We performed a simulation study in order to assess whether the CIAR model can detect the correlation structure of an irregular harmonic model. We generated the irregular observational times t_i with $i = 1, \ldots, n$ using the mixture of exponentials distributions (Eq. (17)). The process y_{t_i} is simulated with length $n = 300$, amplitude $A = 20$, and a unit variance for ϵ_{t_i}. Simulations using (18) are run with $k = 200$ different frequencies taken equally spaced from the interval $(0, \pi)$. We fit a CIAR model to each simulated sequence. In Fig. 2 the parameters estimated (y-axis) from the CIAR model for each of the 200 frequencies (x-axis) are shown as the black line, and the mapping of the frequency values to the theoretical autocorrelation $\cos(f)$ is depicted as a red line. We note that the ϕ^R parameter estimated by the CIAR model fits the theoretical values almost perfectly.

4. Application of the CIAR model to astronomical data

In this section we show some results of the implementation of the CIAR model in astronomical time series. Astronomical data are naturally irregularly sampled since it is not always possible to get observational data from optical telescopes due to a dependency on clear skies. Many astronomical objects, such as variable stars, transients, and supernovae, can be characterized by their brightness. Generally, this brightness is represented by a

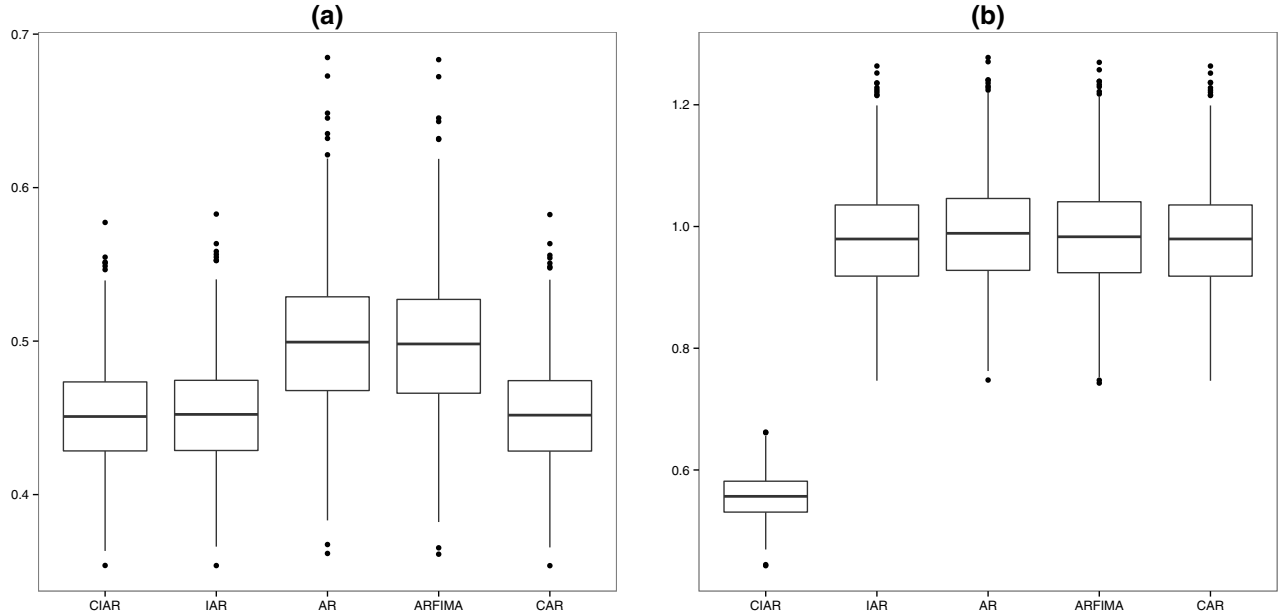

Fig. 1. Boxplot of the root mean squared error computed for the fitted models on the 1000 sequences simulated of the real part of the CIAR process. In *panel a*, each CIAR process was generated using $\phi^R = 0.99$. In *panel b*, each CIAR process was generated using $\phi^R = -0.99$. The other parameters of the models are defined as $\phi^I = 0$, $c = 0$ and length $n = 300$. The observational times are generated using a mixture of Exponential distribution with $\lambda_1 = 15$ and $\lambda_2 = 2$, $\omega_1 = 0.15$ and $\omega_2 = 0.85$.

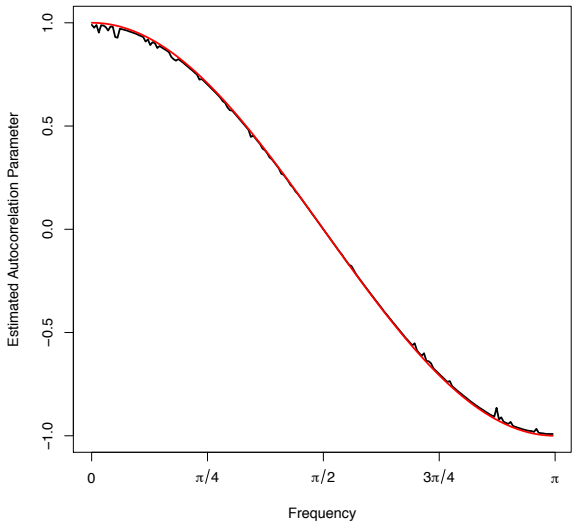

Fig. 2. Estimated coefficients ϕ^R (y-axis) by the CIAR model in $k = 200$ harmonic processes generated using frequencies (x-axis) in the interval $(0, \pi)$. The black line corresponds to the coefficients estimated by the CIAR model. The red line is the theoretical autocorrelation of the process y_{t_i}

time series (light curve) of this object. We apply the CIAR model to the residuals of a harmonic model fitted to the light curves of variable stars.

4.1. Modeling light curves

One of the most important challenges in the analysis of variable stars is to classify them based on their temporal behavior. The pulsating variables represent a particular class of variable stars that are characterized by having a periodical behavior. Therefore, the light curves of pulsating variable stars are generally fitted by a harmonic model (see e.g. Debosscher et al. 2007; Richards et al. 2011; Elorrieta et al. 2016). The p-harmonic model is defined as

$$y(t) = \beta_0 + \sum_{j=1}^{p} (\alpha_{1j} \sin(2\pi f_1 jt) + \beta_{1j} \cos(2\pi f_1 jt)) + \epsilon(t), \qquad (19)$$

where f_1 corresponds to the dominant frequency estimated by the generalized Lomb-Scargle periodogram (GLS; Zechmeister & Kürster 2009). Following Debosscher et al. (2007), for example, $p = 4$ is assumed to fit the light curves. The main goal of implementing the irregular time series models in light curves from periodic variable stars is to detect whether the harmonic model is sufficient to capture all the temporal dependency in the light curve. Otherwise, the residuals of the harmonic model remain with autocorrelation.

We apply the CIAR model to the residuals of the harmonic model fitted on light curves of variable stars from the optical surveys OGLE and Hipparcos. We use these data since many classes of variable stars are available in the catalogue of these surveys.

In order to fit both models on these data, each light curve of the OGLE and Hipparcos surveys was fitted by a four-harmonic model (Eq. (19)). The residuals of each harmonic model are then fitted using both the IAR and CIAR models. In Fig. 3, we note that there is a high correlation between the estimated coefficients $\hat{\phi}_{\text{IAR}}$ and $\hat{\phi}^R$ of the IAR and CIAR models, respectively, on the light curves when both values are positive, which is consistent with the results of the Monte Carlo simulations.

However, we observe that several cases identified as uncorrelated by the IAR model, that is, $\phi_{\text{IAR}} \approx 0$, have a high but negative autocorrelation estimated by the CIAR model. In other words, these light curves remain with negative dependency structure on the residuals after fitting the harmonic model. Double-mode cepheids (DMCEP) and eclipsing binaries (EB) make up the majority of these variable stars.

The time structure detected can be due to the fact that a light curve was incorrectly fitted by the harmonic model, for example

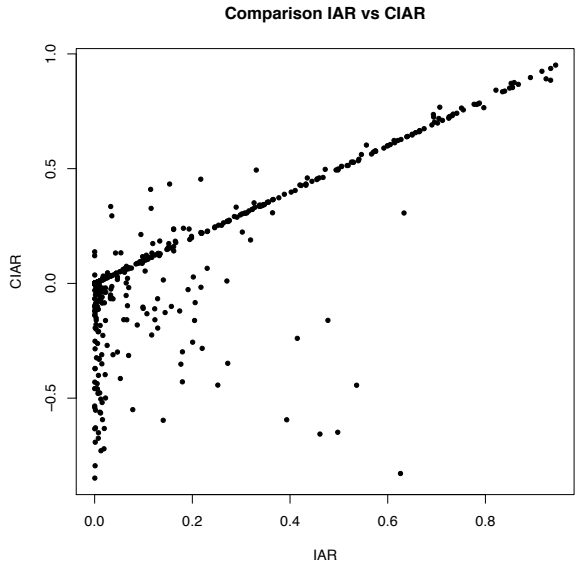

Fig. 3. Plot of ϕ^R vs. ϕ_{IAR}, the parameter coefficients estimated by the CIAR and IAR models, respectively, from light curves in OGLE and Hipparcos.

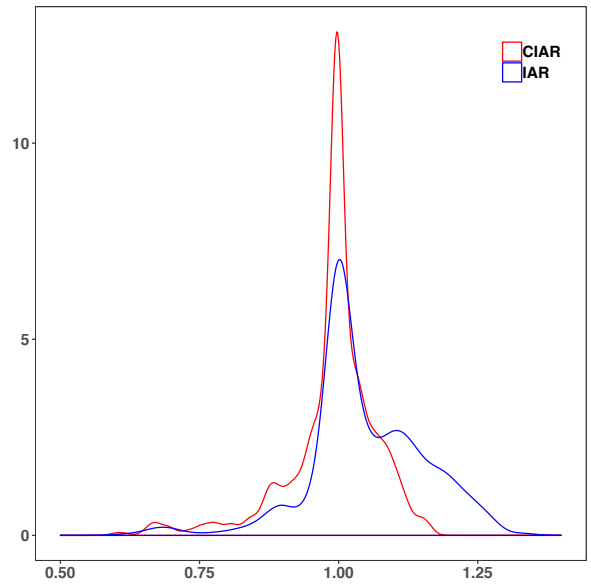

Fig. 4. Kernel density of the RMSE computed for the residuals of harmonic fit in the light curves when the CIAR coefficient is negative. The red density corresponds to the RMSE computed using the CIAR model, and the blue density corresponds to the RMSE computed using the IAR model.

when using an incorrect period. Also the CIAR model can detect a correlation structure in the residuals of a harmonic model fitted to the light curve of a multiperiodic variable star. Intuitively, if the light curve has two or more periods, a harmonic model fitted using only the dominant frequency is not sufficient to account for all its temporal dependency structure.

It is also important to assess the goodness of fit performance of the IAR and CIAR models on these data. We compute the RMSE after fitting each irregular model on the residuals of the harmonic model. These results do not vary significantly when ϕ^R is positive. However, if this coefficient is negative, the RMSE estimated by the CIAR model is smaller than the one obtained when we fit these data using the IAR model, as can be seen in Fig. 4. This result is consistent with those obtained in Sect. 3.2.

4.2. Statistical test for the autocorrelation parameter

The magnitude of the coefficients estimated by the irregular time series models depends on the period of the light curve. Light curves with short periods have, in general, smaller estimated coefficients. In other words, a small value of the estimated coefficient does not always imply uncorrelated residuals. If the light curve has a large period, a small coefficient could mean uncorrelated residuals. However, if the variable star has a shorter period, it could mean the opposite. Therefore, we cannot make decisions based only on the value of the parameter estimated by the model.

To overcome this problem we designed a statistical test that can assess when a value of ϕ is significantly different from (larger than) the values that can arise by chance. To do that, we estimated the distribution of the parameter $\widehat{\phi^R}$ for time series with autocorrelation. We then assessd how likely it is to observe a particular value of $\widehat{\phi^R}$ under this distribution. Following the methodology proposed by Eyheramendy et al. (2018) we selected 38 different frequencies to fit each light curve incorrectly, where each of them is a variation of the correct period in the interval $(f_1 - 0.5f_1, f_1 + 0.5f_1)$. The factor f_1 corresponds to the correct frequency of the light curve. To develop the test we assumed that the $\log(|\widehat{\phi^R}|)$ follows a Gaussian distribution, estimated using the 38 $\widehat{\phi^R_i}, i = 1, \ldots, 38$ obtained from the CIAR fit of the time series that each assume a different incorrect frequency from the interval. Consequently, the null hypothesis here is that the $\log(|\widehat{\phi^R}|)$ belongs to this Gaussian distribution. We note that the test is formulated so that each estimated coefficient is compared only with the estimated ones in the same light curve fitted incorrectly.

Using this test, there are some examples in which the IAR model cannot distinguish if the model is correctly fitted or not, but the CIAR model can always do this. Figure 5 shows an example of a RRc star observed by the Hipparcos survey. Figure 5a shows the perfect fit of the harmonic model to the light curve. Figures 5b and c show that the CIAR and IAR models, respectively, have an estimate of the parameter close to zero on the residuals of the 39 fitted models. However, the CIAR model gives a value of $|\widehat{\phi^R}|$ significantly smaller than the remaining ones when the light curve is correctly fitted, giving a p-value close to zero (Fig. 5b).

4.3. Classification features estimated from the irregular time-series models

One of the most important aims in the light curve analysis from variable stars is to find features that can discriminate one class of variable star from another. Finding a good feature is the key to building a classifier with good performance to detect stars of a given class. Generally, these features are extracted from the temporal behavior of the brightness of each variable star.

In this study, we propose to use the parameter estimated by the CIAR model as a feature. For example, for a multiperiodic variable star, it can then be expected that a harmonic model fitted using only the dominant frequency will not be enough to describe its temporal dependency structure. In other words, if we fit the CIAR and IAR models to the residuals of the harmonic model in Eq. (19) we should obtain values for the parameter of the ACF that are significantly different from zero. Therefore, it is natural to think that these coefficients are capable of distinguishing multiperiodic variable stars from other classes. In the

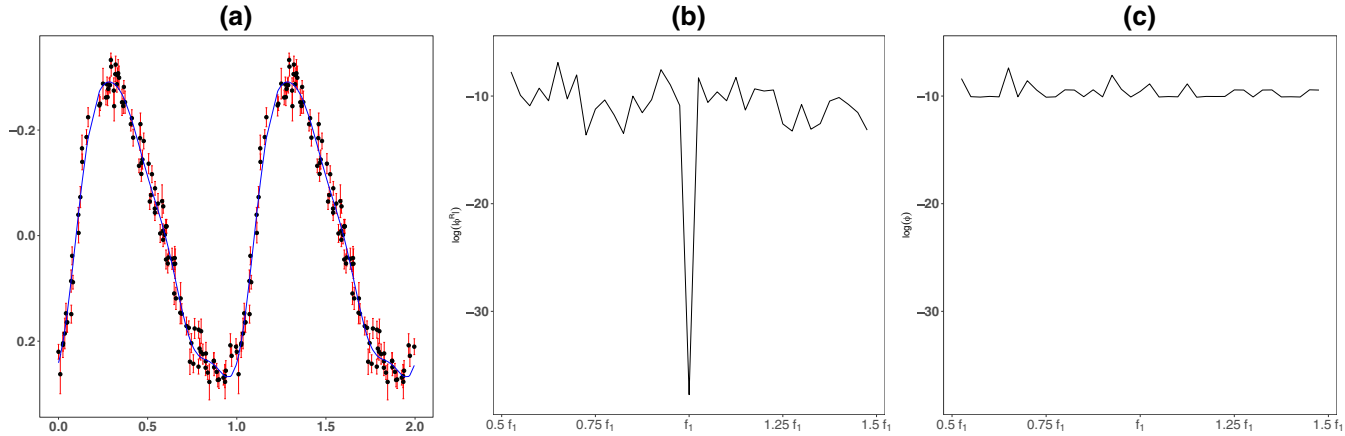

Fig. 5. *Panel a*: light curve of a RR*c* star observed by the Hipparcos survey. The continuous blue line is the harmonic best fit. *Panel b*: natural logarithm of the absolute value of the estimated parameter $\hat{\phi}^R$ by the CIAR model on the residuals of the harmonic model fitted with different frequencies; *x*-axis shows percentage variation from the correct frequency, *y*-axis shows the natural logarithm of $\hat{\phi}^R$. *Panel c*: natural logarithm of the estimated parameter $\hat{\phi}_{\text{IAR}}$ by the IAR model on the residuals of the harmonic model fitted with different frequencies; *x*-axis shows the percentual variations from the correct frequency, *y*-axis shows the natural logarithm of $\hat{\phi}_{\text{IAR}}$.

OGLE and Hipparcos catalogs there are two classes of multiperiodic variable stars: the double-mode RR Lyraes (RRDs) and the double-mode Cepheids (DMCEPs).

We implement the IAR and the CIAR models on the residuals after fitting a harmonic model with only one period to the light curves of these classes. We found several examples of negative autocorrelation in the residuals of the harmonic model, which cannot be detected using the IAR model. One of these examples is a bi-periodic double-mode Cepheid (OGLE ID:175210) in which the IAR and the CIAR coefficients estimated on these residuals are $\hat{\phi}^R = -0.561$ and $\hat{\phi}_{\text{IAR}} = 0.011$. Consequently, we focus on the CIAR model estimates.

The features that can be extracted from the CIAR model are the parameter ϕ^R and the p-value associated to the test performed on ϕ^R. It is interesting to assess whether these features can separate the multiperiodic classes from the other RR-Lyraes and Cepheids, respectively. Figures 6a and b show the distribution of the p-values computed by the CIAR model. As can be seen from these figures there are significant differences between the classes of RR-Lyraes and Cepheids in the distributions of the computed p-values. We note that the RRD and DMCEP classes take larger values in comparison to the other classes.

The use of the p-value as a feature reflects the multiperiodic behavior of the RRD and DMCEP classes.

4.4. Forecasting astronomical data

In this section, we illustrate the forecasting procedure implemented for the CIAR model. For this we use the light curve of the AGN MCG-6-30-15 (Lira et al. 2015) measured in the *K*-band. The time series of this object has $n = 237$ observations taken over a period of approximately 4.5 years.

Initially, we normalize the time series and use the first 90% of the data to estimate the parameters of the CIAR model. We forecast the next observation at the time t_{j+1} given the observational times $t_1, \ldots, t_j$ corresponding to the first 90% of the data. Later, we include the observation at the moment t_{j+1} and re-estimate the model to forecast the observation at the following time t_{j+2}. This procedure is repeated iteratively until the remaining 10% of the data is forecasted, obtaining the vector of one-step forecasted values $(\hat{y}_{j+1}, \ldots, \hat{y}_n)$. Figure 7a shows the normalized MCG-6-30-15 light curve and the forecasted values represented with red dots.

The parameters estimated for this light curve were $\widehat{\phi^R} = 0.9859$ and $\widehat{\phi^I} \approx 0$ for the first 90% of the data and $\widehat{\phi^R} = 0.9863$ and $\widehat{\phi^I} \approx 0$ for all the data. In addition, we also compute the confidence interval at a 90% level for each forecasted value, which is shown in Fig. 7b. We note that the interval size is larger for larger time gaps.

5. Discussion

In this work we present an extension of the irregular autoregressive (IAR) model (Eyheramendy et al. 2018), that we call the Complex Irregular Autoregressive (CIAR) model. As opposed to the IAR model that can only estimate positive values of the ACF, the CIAR model can estimate both positive and negative values of the ACF. We have shown that this model is weakly stationary and its state-space representation is stable under regular assumptions. We propose a maximum likelihood estimation procedure for the parameters of this model, where the solution is reached using Kalman recursions. We developed a code in R and Python to perform an estimation of the model.

There is a strong connection between the three models: CAR(1), IAR, and CIAR. Assuming a null imaginary part and positive real part of the complex autocorrelation parameter, the CIAR model becomes the IAR model. The connection between both models is verified by Monte Carlo simulations in Sect. 3.1. Also, when Gaussian data is fitted in the models, the IAR and the CAR(1) are equivalent. Both the IAR and the CIAR models fit the irregularity of the time gaps between observations as discrete times as opposed to the CAR(1) model which considers time as a continuous variable. At this stage, only the IAR model can fit data that is not Gaussian; the CIAR and the CAR models can only fit Gaussian data.

We illustrate the contribution of the CIAR model in two applications. First, the CIAR process can fit irregular time series with negative values of the ACF, the shape of which is an exponential decay. Second, the CIAR process can identify series with negative values of the ACF such as a high frequency harmonic model or an antipersistent process. In both applications, we show

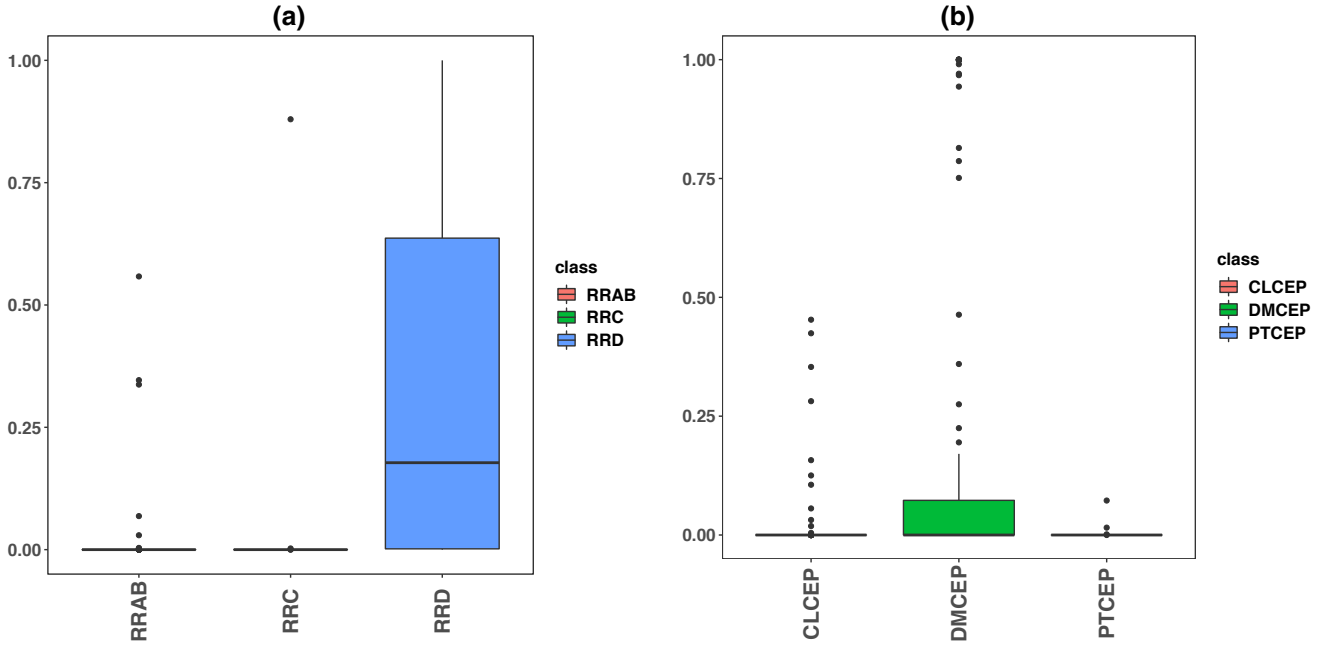

Fig. 6. *Panel a*: boxplot of the p-value estimated from the CIAR model in the RR-Lyraes variable stars separated by subclass. *Panel b*: boxplot of the p-value estimated from the CIAR model in the Cepheids variable stars separated by subclass.

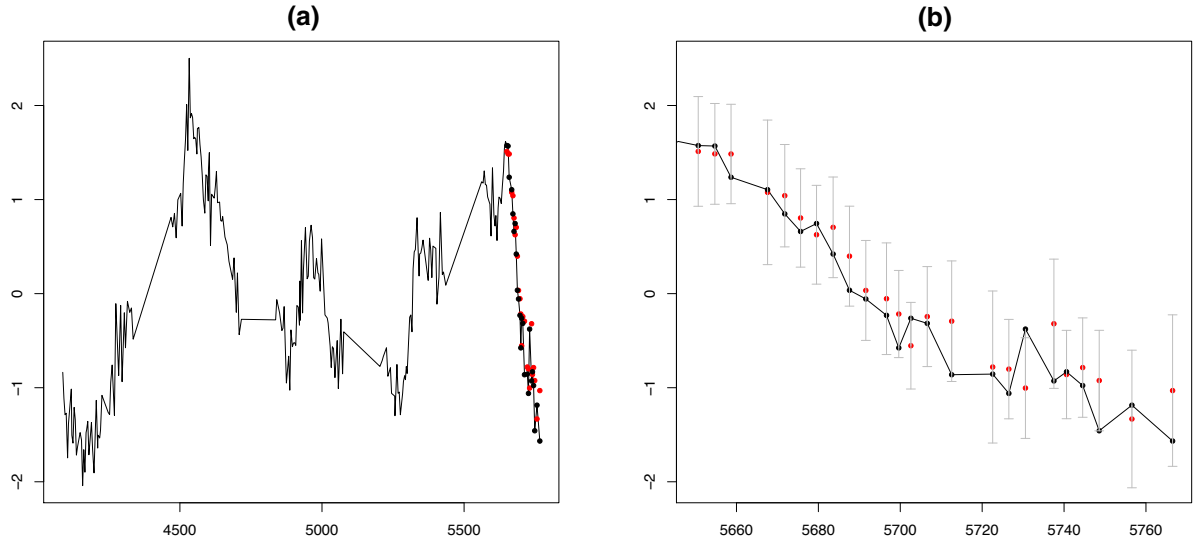

Fig. 7. *Panel a*: normalized K-band MCG-6-30-15 light curve. The red dots are the forecasted values. *Panel b*: zoom of the last 10% of the MCG-6-30-15 light curve. The red dots are the forecasted values using the CIAR model and the gray bars are the confidence intervals at the 90% level.

with simulated data that the CIAR model performs better than other popular time-series models.

A negative autocorrelated time series is characterized by more fluctuations (Box et al. 2015). Take for example $\Delta_t = y_t - y_{t-1}$ and the variance $\mathrm{Var}(\Delta_t) = 2\sigma_y^2(1-\phi)$. This variance is larger for $\phi < 0$ than for positive values of ϕ. Figure 8 illustrates the behavior mentioned above from a CIAR process simulated with $\phi^R = 0.99$ and $\phi^R = -0.99$. We note that there are more oscillations in the CIAR process generated with negative autocorrelation.

The application of the proposed model is performed on astronomical data. Particularly, we use the light curves of variable stars observed by the OGLE and Hipparcos surveys. As many variable stars show periodical behavior in terms of their brightness, it is common to fit the light curves by a harmonic model. We have discussed that the irregular time series models are useful to determine whether or not the residuals of the harmonic fit are still autocorrelated. When a time dependency structure remains on the residuals, we can conclude either that the light curve was not correctly fitted or that it corresponds to a multiperiodic variable star. Furthermore, we noticed that the estimated coefficients are highly related to the frequency of the light curves. In order to interpret the coefficient correctly, we have developed a statistical test.

After fitting both models to the light curves, we can see a strong relationship between the IAR and CIAR models when the coefficient ϕ^R of the CIAR model is positive. However, we have also found several cases of negative values of the ACF of light

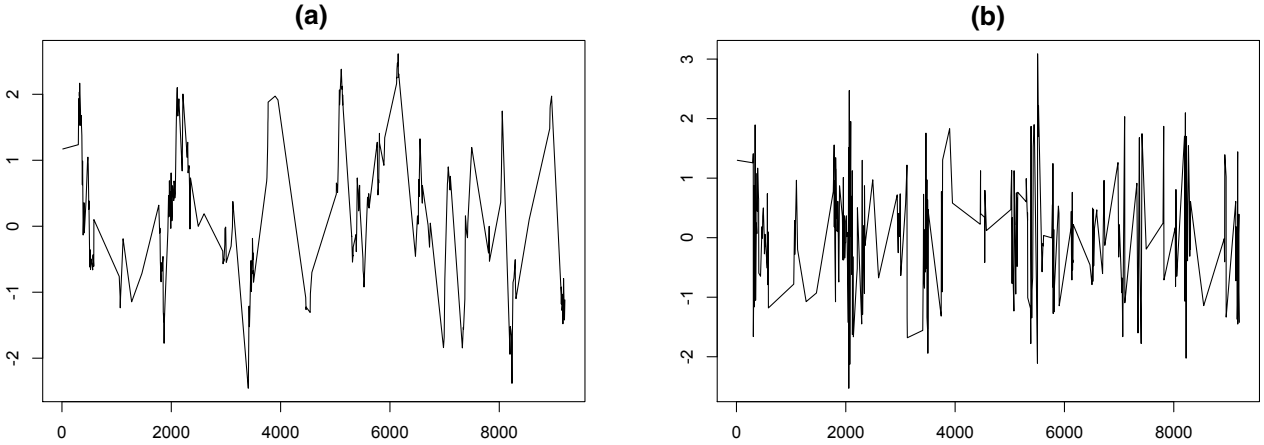

Fig. 8. *Panel a*: CIAR process with positive autocorrelation generated using parameters $\phi^R = 0.99$, $\phi^I = 0$, $c = 0$ and length $n = 300$. *Panel b*: CIAR process with negative autocorrelation generated using parameters $\phi^R = -0.99$, $\phi^I = 0$, $c = 0$ and length $n = 300$.

curves that the IAR model is not capable of identifying. In addition, we show examples where the inability to estimate negative values of ϕ from the IAR model prevents us from finding multiperiodic variable stars or correctly detecting whether the harmonic model is incorrectly specified.

It is expected that negative values of the ACF will be less frequent than positive values. This hypothesis is reflected in the astronomical datasets analyzed. However, there are several examples of negative values obtained from time series of light curves from the OGLE and Hipparcos surveys. This result shows how important it is to have a model that can identify the negative as well as the positive time dependencies in an irregular time series. Among the time-series models that assume irregular sampling, only the CARFIMA process can fit a negative autocorrelated time series if this has an antipersistent behavior. The contribution of the model proposed here is that it can model series with negative exponential decay in the ACF. In other words, both processes can be negatively autocorrelated, but still have different correlation structures.

In addition, we have shown that the p-values obtained from the test proposed in this work for the CIAR model are useful for characterizing the multiperiodic classes of RR-Lyraes (RRDs) and Cepheids (DMCEPs). This result indicates that the p-value can be an important feature for a machine-learned classifier implemented on these classes. In a future study, a classifier for multiperiodic variable stars will be implemented using this feature.

The main aim of this work is to continue developing models for irregularly observed time series where time is considered discrete as opposed to continuous. This will allow us to extend the popular ARMA models for regular observations to the irregular case. We have shown that the CIAR model extends the IAR model by allowing negative as well as positive autocorrelations to be captured. We will continue working on expanding the scope of irregularly observed time series.

Acknowledgements. Support for this research was provided by grant IC120009, awarded to The Millennium Institute of Astrophysics, MAS, and from Fondecyt grant 1160861. F.E. acknowledges support from CONICYT-PCHA (Doctorado Nacional 2014- 21140566).

References

Alder, B., & Wainwright, T. 1970, Phys. Rev. A, 1, 18
Alperovich, Y., Alperovich, M., & Spiro, A. 2017, Tenth International Conference Management of Large-Scale System Development (MLSD), 1
Ausloos, M., & Ivanova, K. 2001, Phys. Rev. E, 63, 047201
Bondon, P., & Palma, W. 2007, J. Time Ser. Anal., 28, 261
Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. 2015, Time Series Analysis: Forecasting and Control, 5th edn. (John Wiley & Sons, Inc.)
Brockwell, P., & Davis, R. 2002, Introduction to Time Series and Forecasting (New York: Springer-Verlag)
Broersen, P. M. T. 2006, Automatic Autocorrelation and Spectral Analysis (Secaucus, NJ, USA: Springer-Verlag, New York, Inc.)
Campbell, J. Y., Lo, A. W. C., & MacKinlay, A. C. 1997, The Econometrics of Financial Markets (Princeton University Press), 632
Carvalho, L. M. V. D., Tsonis, A. A., Jones, C., Rocha, H. R. D., & Polito, P. S. 2007, Nonlinear Processes Geophys., 14, 723
Chan, K. S., & Tong, H. 1987, J. Time Ser. Anal., 8, 277
Conrad, J. S., Hameed, A., & Niden, C. 1994, J. Finance, 49, 1305
Debosscher, J., Sarro, L. M., Aerts, C., et al. 2007, A&A, 475, 1159
Dubois, S. R., & Glanz, F. H. 1986, IEEE Trans. Pattern Anal. Mach. Intell., 8, 55
Edelson, R. A., & Krolik, J. H. 1988, ApJ, 333, 646
Edelson, R., Gelbord, J., Cackett, E., et al. 2017, ApJ, 840, 41
Elorrieta, F., Eyheramendy, S., Jordán, A., et al. 2016, A&A, 595, A82
Eyheramendy, S., Elorrieta, F., & Palma, W. 2018, MNRAS, 481, 4311
Feigelson, E. D., Babu, G. J., & Caceres, G. A. 2018, Front. Phys., 6, 80
Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, AJ, 154, 220
Gao, J., Cao, Y., Tung, W. W., & Hu, J. 2007, Multiscale Analysis of Complex Time Series: Integration of Chaos and Random Fractal Theory, and Beyond (Wiley-Interscience)
Kelly, B. C., Becker, A. C., Sobolewska, M., Siemiginowska, A., & Uttley, P. 2014, ApJ, 788, 33
Lindgren, G., Rootzén, H., & Sandsten, M. 2013, Stationary Stochastic Processes for Scientists and Engineers (Chapman and Hall)
Lira, P., Arévalo, P., Uttley, P., McHardy, I. M. M., & Videla, L. 2015, MNRAS, 454, 368
Martin, R. 1999, Sign. Proces., 77, 139
Miller, K. 1974, in Complex Stochastic Processes: an Introduction to Theory and Application (Addison-Wesley Publishing Company, Advanced Book Program), Adv. Book Program
Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, A&A, 323, L49
Picinbono, B., & Bondon, P. 1997, IEEE Trans. Sign. Proces., 45, 411
Rehfeld, K., Marwan, N., Heitzig, J., & Kurths, J. 2011, Nonlinear Processes Geophys., 18, 389
Richards, J. W., Starr, D. L., Butler, N. R., et al. 2011, ApJ, 733, 10
Sekita, I., Kurita, T., & Otsu, N. 1991, Complex Autoregressive Model and its Properties (Electrotechnical Laboratory)
Sewell, M. 2011, Characterization of Financial Time Series
Tsai, H. 2009, Bernoulli, 15, 178
Udalski, A., Soszynski, I., Szymanski, M., et al. 1999, Acta Astron., 49, 223
Uritskaya, O. Y., & Uritsky, V. M. 2015, Energy Econ., 49, 72
Williams, S. R., Bryant, G., Snook, I. K., & van Megen, W. 2006, Phys. Rev. Lett., 96, 087801
Zechmeister, M., & Kürster, M. 2009, A&A, 496, 577

Appendix A: Proof of Lemma 1

Consider the CIAR process x_{t_j} described by Eq. (5). We note that $\mathbb{E}(y_{t_j}) = 0$ and $\mathbb{E}(z_{t_j}) = 0$. Therefore, $x_{t_j} = y_{t_j} + iz_{t_j}$ is such that $\mathbb{E}(x_{t_j}) = 0$. According to the definition of the model, we have

$$\begin{aligned}\overline{x}_{t_j} x_{t_j} &= \left(\overline{\phi^{\delta_j}}\,\overline{x}_{t_{j-1}} + \overline{\sigma}_{t_j}\,\overline{\varepsilon}_{t_j}\right)\left(\phi^{\delta_j}\,x_{t_{j-1}} + \sigma_{t_j}\,\varepsilon_{t_j}\right)\\ &= \left|\phi^{\delta_j}\right|^2 \left|x_{t_{j-1}}\right|^2 + \cdots + \left|\sigma_{t_j}\right|^2 \left|\varepsilon_{t_j}\right|^2.\end{aligned}$$

Applying expectation $\mathbb{E}$ and using the properties of the model (5)

$$\gamma_0 = \left|\phi^{\delta_j}\right|^2 \gamma_0 + \left|\sigma_{t_j}\right|^2 (1+c),$$

from which we obtain

$$\begin{aligned}\gamma_0 &= \frac{(1+c)\left|\sigma_{t_j}\right|^2}{1-\left|\phi^{\delta_j}\right|^2}\\ &= \frac{(1+c)\sigma^2\left(1-\left|\phi^{\delta_j}\right|^2\right)}{1-\left|\phi^{\delta_j}\right|^2} = (1+c)\sigma^2.\end{aligned}$$

The autocovariance of the process is defined by $\gamma_k = \mathbb{E}(\overline{x}_{t_{j+k}} x_{t_j})$, such that

$$\begin{aligned}\mathbb{E}(\overline{x}_{t_{j+k}} x_{t_j}) &= \mathbb{E}\left(\left(\overline{\phi^{\delta_{j+k}}}\,\overline{x_{t_{j+k-1}}} + \overline{\sigma_{t_{j+k}}}\,\overline{\varepsilon_{t_{j+k}}}\right) x_{t_j}\right)\\ &= \overline{\phi^{\delta_{j+k}}}\,\mathbb{E}\left(\overline{x_{t_{j+k-1}}}\,x_{t_j}\right)\\ &= \overline{\phi^{\delta_{j+k}}}\,\mathbb{E}\left(\left(\overline{\phi^{\delta_{j+k-1}}}\,\overline{x_{t_{j+k-2}}} + \overline{\sigma_{t_{j+k-1}}}\,\overline{\varepsilon_{t_{j+k-1}}}\right) x_{t_j}\right)\\ &= \overline{\phi^{t_{j+k}-t_{j+k-2}}}\,\mathbb{E}\left(\overline{x_{t_{j+k-2}}}\,x_{t_j}\right).\end{aligned}$$

It can be shown that by recursion the autocovariance function is given by

$$\gamma_k = \overline{\phi^{t_{j+k}-t_j}}\sigma^2(1+c) = \overline{\phi^{\Delta_k}}\sigma^2(1+c),$$

and therefore, the autocorrelation can be expressed as $\rho_k = \overline{\phi^{\Delta_k}}$ □.

Appendix B: Proof of Lemma 2

The CIAR model is defined as $y_{t_j} + iz_{t_j} = (\phi^R + i\phi^I)^{t_j - t_{j-1}}(y_{t_{j-1}} + iz_{t_{j-1}}) + \sigma_{t_j}(\varepsilon^R_{t_j} + i\varepsilon^I_{t_j})$. Let us focus on the term $(\phi^R + i\phi^I)^{t_j - t_{j-1}}$,

$$(\phi^R + i\phi^I)^{t_j - t_{j-1}} = \quad (\phi^R + i\phi^I)^{\delta_j} = |\phi|^{\delta_j}\left(\frac{\phi^R + i\phi^I}{|\phi|}\right)^{\delta_j}.$$

Using the polar representation for complex numbers we obtain

$$\begin{aligned}\left(\frac{\phi^R + i\phi^I}{|\phi|}\right)^{\delta_j} &= (\cos(\psi) + i\sin(\psi))^{\delta_j}\\ \left(\phi^R + i\phi^I\right)^{\delta_j} &= |\phi|^{\delta_j}(\cos(\psi) + i\sin(\psi))^{\delta_j}.\end{aligned}$$

Using the De Moivre formula, we have

$$\begin{aligned}\left(\phi^R + i\phi^I\right)^{\delta_j} &= |\phi|^{\delta_j}(\cos(\delta_j\psi) + i\sin(\delta_j\psi))\\ &= |\phi|^{\delta_j}\cos(\delta_j\psi) + i|\phi|^{\delta_j}\sin(\delta_j\psi)\\ &= \alpha^R_{t_j} + i\alpha^I_{t_j}.\end{aligned}$$

Finally, the Complex IAR model can be represented by the expression

$$y_{t_j} + iz_{t_j} = (\alpha^R_{t_j} + i\alpha^I_{t_j})(y_{t_{j-1}} + iz_{t_{j-1}}) + \sigma_{t_j}(\varepsilon^R_{t_j} + i\varepsilon^I_{t_j})\square.$$

Appendix C: Proof of Lemma 3

At a fixed time t_j, the eigenvalues of the transition matrix of the CIAR process $F_{t_j} = \begin{pmatrix} \alpha^R_{t_j} & -\alpha^I_{t_j} \\ \alpha^I_{t_j} & \alpha^R_{t_j} \end{pmatrix}$ satisfy the following equation.

$$\begin{aligned}|(F_{t_j} - \lambda I)| &= \left|\begin{pmatrix} \alpha^R_{t_j} & -\alpha^I_{t_j} \\ \alpha^I_{t_j} & \alpha^R_{t_j} \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right|\\ &= \left|\begin{pmatrix} \alpha^R_{t_j} - \lambda & -\alpha^I_{t_j} \\ \alpha^I_{t_j} & \alpha^R_{t_j} - \lambda \end{pmatrix}\right|\\ &= (\alpha^R_{t_j} - \lambda)^2 + \alpha^{I2}_{t_j}\\ &= \lambda^2 - 2\alpha^R_{t_j}\lambda + (\alpha^{R2}_{t_j} + \alpha^{I2}_{t_j})\\ &= 0,\\ \Rightarrow \lambda &= \frac{2\alpha^R_{t_j} \pm \sqrt{4\alpha^{R2}_{t_j} - 4(\alpha^{R2}_{t_j} + \alpha^{I2}_{t_j})}}{2} = \alpha^R_{t_j} \pm i\alpha^I_{t_j}.\end{aligned}$$

Since $|\alpha^R_{t_j} + i\alpha^I_{t_j}| = |\alpha^R_{t_j} - i\alpha^I_{t_j}| = |\alpha_{t_j}|$, then the process is stable if $\sup|\alpha_{t_j}| < 1$. Therefore, under this assumption the CIAR process has the unique stationary solution (Brockwell & Davis 2002) given by

$$\begin{aligned}X_{t_j} &= F_{t_j}X_{t_{j-1}} + V_{t_j}\\ &= F_{t_j}(F_{t_{j-1}}X_{t_{j-2}} + V_{t_{j-1}}) + V_{t_j}\\ &= F_{t_j}F_{t_{j-1}}X_{t_{j-2}} + F_{t_j}V_{t_{j-1}} + V_{t_j}\\ &= F_{t_j}F_{t_{j-1}}(F_{t_{j-2}}X_{t_{j-3}} + V_{t_{j-2}}) + F_{t_j}V_{t_{j-1}} + V_{t_j}\\ &= F_{t_j}F_{t_{j-1}}F_{t_{j-2}}X_{t_{j-3}} + F_{t_j}F_{t_{j-1}}V_{t_{j-2}}\\ &\quad + F_{t_j}V_{t_{j-1}} + V_{t_j}.\end{aligned}$$

Therefore, the general form can be written as,

$$X_{t_j} = X_{t_{j-n}}\prod_{k=0}^{n-1}F_{t_{j-k}} + V_{t_j} + \sum_{k=1}^{n-1}V_{t_{j-k}}\prod_{i=0}^{k-1}F_{t_{j-i}}.$$

Since $\left|\prod_{k=0}^{n-1}F_{t_{j-k}}\right| = \prod_{k=0}^{n-1}\left|F_{t_{j-k}}\right|$ and $|F_{t_{j-k}}| < 1$ due to the stability of the process, then $\lim_{n\to\infty}\prod_{k=0}^{n-1}F_{t_{j-k}} = 0$. Finally, if $n \to \infty$ then the unique stationary solution is given by,

$$X_{t_j} = V_{t_j} + \sum_{k=1}^{\infty}V_{t_{j-k}}\prod_{i=0}^{k-1}F_{t_{j-i}}\square.$$