A harmonically weighted filter for cyclical long memory processes

Federico Maddanu¹

Received: 12 November 2020 / Accepted: 10 February 2021 / Published online: 8 March 2021
© The Author(s) 2021

Abstract

The estimation of the long memory parameter d is a widely discussed issue in the literature. The harmonically weighted (HW) process was recently introduced for long memory time series with an unbounded spectral density at the origin. In contrast to the most famous fractionally integrated process, the HW approach does not require the estimation of the d parameter, but it may be just as able to capture long memory as the fractionally integrated model, if the sample size is not too large. Our contribution is a generalization of the HW model, denominated the Generalized harmonically weighted (GHW) process, which allows for an unbounded spectral density at $k \geq 1$ frequencies away from the origin. The convergence in probability of the Whittle estimator is provided for the GHW process, along with a discussion on simulation methods. Fit and forecast performances are evaluated via an empirical application on paleoclimatic data. Our main conclusion is that the above generalization is able to model long memory, as well as its classical competitor, the fractionally differenced Gegenbauer process, does. In addition, the GHW process does not require the estimation of the memory parameter, simplifying the issue of how to disentangle long memory from a (moderately persistent) short memory component. This leads to a clear advantage of our formulation over the fractional long memory approach.

Keywords Cyclical long memory · Harmonically weighted filter · Whittle estimation · Prediction · Paleoclimatic temperatures

1 Introduction

Many environmental and financial time series show dependence between distant observations, denoting a phenomenon that is known in the literature as long memory. Adopting a common definition (see Beran (1994) and Hassler (2019)), a

✉ Federico Maddanu
federico.maddanu@gmail.com

¹ Department of Economics and Finance, Università di Roma Tor Vergata, via Columbia 2, 00133 Rome, Italy

stochastic process exhibits long memory if the autocovariance function is not absolutely summable or, equivalently, if the spectral density is unbounded. When the singularity in the spectrum is located at a frequency away from the origin, then the process exhibits cyclical long memory, denoting some periodicity.

In the last decades, cyclical long memory processes have been proposed in the literature to model time series characterized by a strong quasi-periodic behaviour. In particular, the well-known Gegenbauer autoregressive moving average (GARMA) process was introduced in the pioneer work of Hosking (1981) and then formalized in Anel (1986) and Gray et al. (1989). It generalizes the fractionally integrated ARMA (FARIMA) model by Hosking (1981) allowing for an unbounded spectral density also at a frequency away from zero. Woodward et al. (1998) extended this result to the k -factor GARMA process, where the spectral density is unbounded at $k \geq 1$ frequencies. The following equation describes the latter:

$$\prod_{j=1}^k (1 - 2 \cos(\lambda_{gj})L + L^2)^{d_j} (y_t - \mu) = x_t \quad (1)$$

where k is a positive integer and $d_j \in (0, 1)$, while x_t is assumed to be an ARMA(p, q) process, such that $x_t = \frac{\psi(L)}{\phi(L)} \varepsilon_t = c(L) \varepsilon_t$ with $\sum_{j=0}^{\infty} |c_j| < \infty$ and ε_t is a white noise (WN) sequence. The process in (1) displays cyclical long memory, and the contribution of each pole to the periodicity of the process is controlled by the d_j parameters. According to the literature (see Woodward et al. (1998) and also Dissanayake et al. (2018) for a review), the k -factor GARMA process is covariance stationary and exhibits long memory if $d_j \in (0, \frac{1}{2})$ and $|\cos(\lambda_{gj})| < 1$ for $j = 1, \dots, k$ (or if $d_j \in (0, \frac{1}{4})$ whenever $|\cos(\lambda_{gj})| = 1$).

The estimation of the memory parameter d is a widely discussed issue in the literature, and plenty of methods were developed to address it (see Hassler (2019) for a review). For instance, in the case of a singularity in the spectrum at the zero frequency, Geweke and Porter-Hudak (1983) and Robinson (1995b) proposed log-periodogram regression methods. On the other hand, a full Whittle likelihood approach (c.f. Whittle (1953)) can be found in Beran (1994), while a semiparametric local Whittle method has been formalized in Robinson (1995a). Pseudo-Whittle estimates have been proposed in Velasco and Robinson (2000) for the no-stationary case. Regarding cyclical long memory processes, Chung (1996a, b) proposed the conditional sum of squares method for the parameters estimation of the GARMA model, while Arteche and Robinson (2000) generalized the local Whittle and log-periodogram methods allowing for a singularity in the spectrum away from the origin. A full Whittle likelihood approach is discussed in Ferrara and Guegan (1999). See also Palma and Chan (2005) and Reisen et al. (2006) for further study.

In practice, when it comes to estimating d , the issue is to separate long memory from a (moderately persistent) short memory component. This is complicated by the difficulty of estimating d correctly. For instance, it often happened that different researchers estimated different values of the memory parameter analysing the same data set. Recently, Hassler and Hosseinkouchack (2020) introduced the harmonically weighted (HW) process that theoretically displays less persistence and weaker

memory than the fractionally integrated model of order d . Although the two processes have different qualities asymptotically, in terms of persistence and memory, they are hard to distinguish given realistic sample sizes. The HW process is characterized by an unbounded spectral density at the origin and does not require the estimation of the memory parameter, even though the HW approach may be just as able to capture long memory as the fractionally integrated model, if the sample size is not too large. In this case, the long memory could be captured just by filtering the data.

Our purpose is to extend the contribution of Hassler and Hosseinkouchack (2020) considering also cyclical long memory time series characterized by a strong quasi-periodic behavior. The novel approach, denominated the generalized harmonically weighted (GHW) process, generalizes the HW filter allowing for an unbounded spectral density at $k \geq 1$ frequencies away from zero. We provide analytical expressions for the spectrum and the autocovariance function of the GHW process, along with simulation and estimation methodologies and the convergence in probability of the Whittle estimator. Finally, an empirical application on paleoclimatic temperature reconstructions is discussed, comparing the results performed both via the GHW and the k -factor GARMA models. Our main conclusion is that the GHW model is able to capture the long memory, as well as the standard GARMA does. In addition, the GHW process does not require the estimation of the d parameter, simplifying the issue of how to disentangle long memory from a short memory component. This leads to a clear advantage of our formulation over the GARMA approach; for instance, the estimation by exact maximum likelihood can be easily implemented if the frequency parameters are supposed to be known.

We organize the rest of the paper as follows. In sect. 2, we recall the main characteristics of the HW process. In sect. 3, we introduce the GHW process, providing analytical formulas for the autocovariance function and the spectral density. Multiple periodicity and simulation methods are also considered in this section. Section 4 discusses estimation methodologies, providing the convergence in probability of the Whittle estimator. In sect. 5, an empirical application on the EPICA Dome C series of paleoclimatic temperature reconstructions is examined. The last section concludes the paper. Mathematical proofs are relegated to the appendix.

2 Review of the harmonically weighted process

Let us start with a definition of long memory that is very common in the literature (see for instance Beran (1994), Palma (2006) and, more recently, Hassler (2019)).

Definition 1 A stationary process $\{y_t\}$ is said to display long memory if its autocovariance function is not absolutely summable or, equivalently, if its spectral density is unbounded at one or more frequencies. Otherwise, it is said to have short memory.

Definition 1 meets the FARIMA and GARMA processes introduced by Hosking (1981). Both of them depend on a fractional parameter d and can capture the

long memory from the data via the estimation of a fractional filter, depending on d . Hassler and Hosseinkouchack (2020) introduced an alternative way to model long memory in time series, which does not require the estimation of d . They defined the harmonically weighted filter $h(L)$ by the formal expansion of $\ln(1 - L)$:

$$h(L) = -\frac{\ln(1 - L)}{L} = \sum_{j=0}^{\infty} \frac{L^j}{j+1} \quad (2)$$

Basically, the filter inputs result to be harmonically weighted. Let us assume $x_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$ as a general short memory process with $\varepsilon_t \sim WN(0, \sigma^2)$. This assumption meets all stationary and invertible ARMA processes, since c_j is geometrically bounded in the ARMA case.

The harmonically weighted (HW) process is described by:

$$y_t = \mu + h(L)x_t = \mu + \sum_{j=0}^{\infty} \frac{1}{j+1} x_{t-j}$$

where y_t is a second-order stationary harmonically weighted process with mean μ and moving average representation:

$$h(L)c(L)\varepsilon_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} = b(L)\varepsilon_t$$

with

$$b_j = \sum_{k=0}^j \frac{c_k}{j+1-k}$$

The autoregressive representation is given by:

$$c(L)^{-1}g(L)(y_t - \mu) = c(L)^{-1}\left(1 - \sum_{j=1}^{\infty} g_j L^j\right)(y_t - \mu) = a(L)(y_t - \mu)$$

with

$$g_j = \frac{1}{j+1} - \sum_{i=1}^{j-1} \frac{g_i}{j-i+1} \quad ; \quad j \geq 1 \quad ; \quad g_0 = 1$$

where $g(L) = h(L)^{-1}$ is defined as the harmonic inverse transformation (HIT) function. For $\lambda > 0$, the spectral density is given by:

$$f(\lambda) = \left[\ln^2 \left(2 \sin \frac{\lambda}{2} \right) + \left(\frac{\pi - \lambda}{2} \right)^2 \right] f_x(\lambda)$$

where

$$f(\lambda) \sim \ln^2(\lambda)f_x(0) \quad \text{as } \lambda \rightarrow 0$$

As $k \rightarrow \infty$, the autocovariance function behaves like:

$$\gamma(k) \sim 2\pi f_x(0) \frac{\ln k}{k}$$

The above process displays standard long memory with an unbounded spectral density at the zero frequency and not absolutely summable autocovariance function, which decays to zero slowly (see Hassler and Hosseinkouchack (2020) for further details).

3 The generalized harmonically weighted process

In this section, we introduce a novel process that generalizes the HW filter to admit cyclical long memory. In substance, it consists in allowing for an unbounded spectral density also at a frequency away from the origin. Let $\lambda_g \in [0, \pi] \setminus \{\frac{\pi}{2}, \frac{\pi}{3}\}$ be the Gegenbauer frequency (see Gray et al. (1989)), the following filter generalizes the Hassler and Hosseinkouchack (2020) filter allowing for a singularity in the spectrum away from zero:

$$\begin{aligned} h(L) &= - \frac{\ln(1-2\eta L+L^2)}{2L} \\ &= \sum_{j=0}^{\infty} \frac{\cos(\lambda_g(j+1))}{j+1} L^j \end{aligned} \quad (3)$$

where $\eta = \cos(\lambda_g)$. Let us observe that if $\lambda_g = 0$ (which implies $\eta = 1$) the expression in (3) is reduced to the HW filter in (2).

Proposition 1 Let x_t be an ARMA(p, q) process such that $x_t = \frac{\psi(L)}{\phi(L)}\varepsilon_t = c(L)\varepsilon_t$ with $\sum_{j=0}^{\infty} |c_j| < \infty$, $\phi_0 = \psi_0 = 1$ and $\varepsilon_t \sim WN(0, \sigma^2)$. Let $\eta = \cos(\lambda_g)$ with $\lambda_g \in [0, \pi] \setminus \{\frac{\pi}{2}, \frac{\pi}{3}\}$, then the process:

$$\begin{aligned} y_t &= \mu + h(L)x_t = \mu - \frac{\ln(1-2\eta L+L^2)}{2L}x_t \\ &= \mu + \sum_{j=0}^{\infty} \frac{\cos(\lambda_g(j+1))}{j+1} x_{t-j} \end{aligned}$$

is a second-order stationary generalized harmonically weighted (GHW) process with mean μ and moving average representation:

$$h(L)c(L)\varepsilon_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} = b(L)\varepsilon_t$$

with

$$b_j = \sum_{k=0}^j c_{j-k} \frac{\cos(\lambda_g(k+1))}{(k+1)}$$

The autoregressive representation is given by:

$$c(L)^{-1}g(L)(y_t - \mu) = c(L)^{-1} \sum_{j=0}^{\infty} g_j(y_{t-j} - \mu) = a(L)(y_t - \mu)$$

with

$$g_j = - \sum_{k=1}^j \frac{\cos(\lambda_g(k+1))}{\eta(k+1)} g_{j-k} \quad \text{and} \quad g_0 = \eta^{-1}$$

where $g(z) = h(z)^{-1}$ is defined as the generalized harmonic inverse transform (GHIT) function.

The spectral density of $\{y_t\}$ is:

$$f(\lambda) = \frac{1}{4} (\ln^2(2|\cos(\lambda) - \eta|) + (\alpha\pi - \lambda)^2) f_x(\lambda) \quad \text{with} \quad \lambda \in [-\pi, \pi] \setminus \{\lambda_g\}$$

where $f_x(\lambda)$ is the spectral density of the short memory component x_t and

$$\alpha = \begin{cases} 1, & \text{if } \lambda > \lambda_g \\ 0, & \text{if } -\lambda_g < \lambda < \lambda_g \\ -1, & \text{if } \lambda < -\lambda_g \end{cases}$$

As $\lambda \rightarrow \lambda_g$, the spectral density behaves like:

$$f(\lambda) \sim \begin{cases} \frac{1}{4} \ln^2 |\lambda - \lambda_g| f_x(\lambda_g), & \text{if } \lambda_g \notin \{0, \pi\} \\ \ln^2(\lambda) f_x(0), & \text{if } \lambda_g = 0 \\ \ln^2(\lambda - \pi) f_x(\pi), & \text{if } \lambda_g = \pi \end{cases}$$

The autocovariance function admits the Wold representation:

$$\gamma(k) = \sum_{j=0}^{\infty} b_j b_{j+k} \sigma^2 \quad \text{with} \quad k \in \mathbb{Z}$$

Proof The proof of Proposition 1 is given in the Appendix. $\square$

It is immediately clear that the spectrum of the process in Proposition 1 is unbounded at λ_g . According to Definition 1, both the HW and the GHW processes exhibit long memory, but let us say that the latter exhibits cyclical long

memory if $\lambda_g > 0$. This implies a singularity in the spectrum away from the origin, denoting some periodicity in the process.

We also point out that the GHIT function is not defined if $\lambda_g = \frac{\pi}{2}$. Moreover, the infinite sum of the g_j coefficients converges only if $\lambda_g \neq \frac{\pi}{3}$, because of:

$$g(1) = -\frac{2}{\ln(2(1 - \cos(\lambda_g)))}$$

Therefore, if we have a time series that shows a fundamental frequency or a harmonic close to $\frac{\pi}{2}$ or to $\frac{\pi}{3}$, then the $\{g_j\}$ sequence diverges as j increases, prejudicing a good filtering. In this case, we cannot compute the short memory component by filtering the data.

3.1 Multiple periodicity with the GHW process

In this section, a multiplicative model that generalizes the process in Proposition 1 is introduced. It allows for an unbounded spectral density at $k \geq 1$ frequencies, and its main properties are summarized by the following proposition.

Proposition 2 *Let x_t be an ARMA(p, q) process such that $x_t = \frac{\psi(L)}{\phi(L)}\varepsilon_t = c(L)\varepsilon_t$ with $\sum_{j=0}^{\infty} |c_j| < \infty$, $\phi_0 = \psi_0 = 1$ and $\varepsilon_t \sim WN(0, \sigma^2)$. Consider the set of frequencies $\{\lambda_{g,j}\}_{j=1}^k \in [0, \pi] \setminus \{\frac{\pi}{2}, \frac{\pi}{3}\}$, then the process:*

$$y_t = \mu + \left[\prod_{j=1}^k h_j(L) \right] x_t$$

with

$$h_j(L) = \sum_{i=0}^{\infty} \frac{\cos(\lambda_{g,j}(i+1))}{(i+1)} L^i$$

is a second-order stationary k -factor generalized harmonically weighted (k -GHW) process with mean μ and spectral density :

$$f(\lambda) = \prod_{j=1}^k \left(\frac{1}{4} (\ln^2(2|\cos(\lambda) - \cos(\lambda_{g,j})|) + (\alpha_j \pi - \lambda)^2) \right) f_x(\lambda) \quad \lambda \in [-\pi, \pi] \setminus \{\lambda_{g,j}\}_{j=1}^k$$

where $f_x(\lambda)$ is the spectral density of the process x_t and

$$\alpha_j = \begin{cases} 1, & \text{if } \lambda > \lambda_{g,j} \\ 0, & \text{if } -\lambda_{g,j} < \lambda < \lambda_{g,j} \\ -1, & \text{if } \lambda < -\lambda_{g,j} \end{cases}$$

The autocovariance function admits the Wold representation:

$$\gamma(h) = \sum_{j=0}^{\infty} b_j b_{j+h} \sigma^2 \quad \text{with } h \in \mathbb{Z}$$

where the polynomial $b(L)$ is given by the convolution of $c(L)$ and $\prod_{j=1}^k h_j(L)$ such that:

$$c(L) \prod_{j=1}^k h_j(L) = b(L).$$

Proof It follows directly from the proof of Proposition 1. $\square$

Basically, we generalized the GHW process similarly as Woodward et al. (1998) generalized the Gegenbauer process via the k -factor GARMA model. The process in Proposition 2 exhibits cyclical long memory with an unbounded spectral density in the set of frequencies $\{\lambda_{g,j}\}_{j=1}^k$. We collect these latter into a vector defined as the frequency parameters vector: $\lambda_g = [\lambda_{g,1}, \dots, \lambda_{g,k}]^\top$.

The determination of the length k of the frequency parameters vector is an issue that requires further discussion. In the case of a k -factor GARMA process, Leschinski and Sibbertsen (2019) proposed a novel procedure based on sequential tests of the maximum of the periodogram and semiparametric estimators of the model parameters. A similar approach could be implemented also in the k -GHW framework. A preliminary analysis on the residuals periodogram is suggested here as a starting point. If k is chosen too low, then the residuals should display poles in the periodogram. In contrast, if k is chosen too high, the residual process should exhibit antipersistence. At the end, we should try to balance between these two possible scenarios. In sect. 3.2, we will see as the frequency parameters can affect persistence in the k -GHW process.

3.2 Comparison with respect to the GARMA process

Before proceeding further, we need to introduce the concept of “persistence” in time series analysis. According to Hassler (2019), it is related but not equivalent to the concept of “memory” given in Definition 1. Indeed, while the definition of “memory” focuses on the absolute summability of the autocovariance function, the concept of “persistence” considers just its sum.

Definition 2 Let $\{y_t\}$ be a second-order stationary process and let us define $\omega^2 = \sum_{k=-\infty}^{\infty} \gamma(k) = 2\pi f(0)$ as the long-run variance, where $f(0)$ is the spectral density of y_t evaluated at the zero frequency. Considering $\omega^2 \geq 0$ (as implied by Theorem 1.5.1 in Brockwell and Davis (1986)), we have that the process y_t is:

- (i) *moderately persistent* if $\omega^2 < \infty$,
- (ii) *anti-persistent* if $\omega^2 = 0$,
- (iii) *strongly persistent* if $\omega^2 = \infty$.

Now, let us compare the behaviour at the origin between the spectra of a GHW noise (defined by assuming $f_x(\lambda) = \sigma^2/2\pi$ in Proposition 1) and of a fractionally differenced Gegenbauer noise (that is the GARMA process in (1) with $k = 1$ and $x_t = \varepsilon_t \sim WN(0, \sigma^2)$). The spectrum of the fractionally differenced Gegenbauer noise is given by the following equation (see Andel (1986)):

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left(4 \sin \left(\frac{\lambda - \lambda_g}{2} \right) \sin \left(\frac{\lambda + \lambda_g}{2} \right) \right)^{-2d} \quad (4)$$

The function in (4) is proportional to $\left(4 \sin^2 \left(\frac{\lambda_g}{2} \right) \right)^{-2d}$ as λ approaches to zero. On the other hand, the spectrum of the GHW noise behaves at the origin as a square logarithmic function with $f_{GHWn}(0) \propto \ln^2(2 \sin(\frac{\lambda_g}{2}))$. In Fig. 1, we compare the two spectra for the true values of $\sigma = 1$, $\lambda_g = 0.6$ and for different values of the d parameter. According to Definition 2, the GHW noise seems to be the least persistent with respect to the others. For any $d \in (0, 1)$ and unit variance parameter for both processes, we can state that the GHW noise is the least persistent if the following inequality holds:

$$\ln^2 \left(2 \sin \left(\frac{\lambda_g}{2} \right) \right) < \left(4 \sin^2 \left(\frac{\lambda_g}{2} \right) \right)^{-2d} \quad (5)$$

Fig. 1 Comparison between the spectra of a fractionally differenced Gegenbauer noise and of a GHW noise with $\lambda_g = 0.6$ and $\sigma = 1$

Obviously, we know that both the left-hand side (LHS) and the right-hand side (RHS) of (5) diverge to $+\infty$ as $\lambda_g \rightarrow 0$, leading to an indeterminate form of the limit ratio. However, applying the Hopital's rule the RHS dominates the other one as λ_g approaches to zero. In Fig. 2, we plot both sides of (5) as a function of λ_g for different values of d in the RHS. Observe that the GHW noise is the least persistent for low values of the frequency parameter λ_g ; however, as λ_g becomes larger, the RHS and the LHS of (5) seem to be closer. In this case, the GHW and Gegenbauer noise display similar persistence (notice that if $d > 0.5$, we refer to the pseudo-spectrum of the fractionally differenced Gegenbauer noise, see Velasco and Robinson (2000) for further details).

In conclusion, similar arguments can be generalized to include also multiple periodicity, pointing out that the persistence of the GHW process is affected by the frequency parameters. This is in contrast to the GARMA model, where the d parameters allow for a better control on the contribution of each pole to the memory and persistence of the process.

Fig. 2 Considering the inequality in (5), we plot the LHS versus the RHS for different values of d

3.3 Simulation of the GHW process

A simple way to simulate T realizations from a Gaussian GHW process is to truncate its MA representation according to:

$$y_t = \sum_{j=0}^t b_j \varepsilon_{t-j} \quad (6)$$

for $t = 1, \dots, T$ and $\varepsilon_t \sim N(0, \sigma^2)$. The b_j coefficients are defined in Proposition 1. Following Hassler and Hosseinkouchack (2020), we can simulate a sample size of length T from a (type I) GHW process generating $5000 + T$ realizations from the (type II) GHW process in (6), and then discarding the first 5000 observations. However, this simulation approach may be computationally expensive if the b_j coefficients are given by the convolution of many polynomials.

An alternative way to simulate T realizations from the Gaussian GHW process is to use approximate methods, like for instance, the spectral method described in Dieker and Mandjes (2003). The latter is particularly useful in the GHW case because we have a closed form of the spectral density but not of the autocovariance function. According to the spectral analysis of a stationary discrete-time Gaussian process, see Dieker and Mandjes (2003), we can represent y_t in terms of its spectral density $f(\lambda)$ as:

$$y_t = \int_0^\pi \sqrt{2f(\lambda)} \cos(t\lambda) dB_1(\lambda) - \int_0^\pi \sqrt{2f(\lambda)} \sin(t\lambda) dB_2(\lambda) \quad (7)$$

where $B_1(\lambda)$ and $B_2(\lambda)$ are mutually independent Brownian motions.

Notice that the two integrand functions $r_1(\lambda) = \sqrt{2f(\lambda)} \cos(t\lambda)$ and $r_2(\lambda) = \sqrt{2f(\lambda)} \sin(t\lambda)$ in (7) are simple functions in the sense that, given a positive integer l and a strictly increasing sequence of real numbers $\{u_j\}_{j=0}^{l-1}$ with $u_0 = 0$ and $u_l = \pi$, as well as given two bounded sequences of real numbers $\{r_{1,j}\}_{j=0}^{l-1}$ and $\{r_{2,j}\}_{j=0}^{l-1}$, we can write $r_1(z) = r_{1,j}$ and $r_2(z) = r_{2,j}$ for $z \in (u_j, u_{j+1}]$. As a consequence, the stochastic integrals in (7) have the following definitions:

$$\begin{aligned} \int_0^\pi r_1(\lambda) dB_1(\lambda) &:= \sum_{j=0}^{l-1} r_{1,j} (B_1(u_{j+1}) - B_1(u_j)) \\ \int_0^\pi r_2(\lambda) dB_2(\lambda) &:= \sum_{j=0}^{l-1} r_{2,j} (B_2(u_{j+1}) - B_2(u_j)) \end{aligned} \quad (8)$$

Using (8) and the properties of the standard Brownian motion, we can approximate (7) as:

$$y_t \simeq \sum_{j=0}^{l-1} \sqrt{2\pi f(u_{j+1})/l} \left(\cos(tu_{j+1}) \epsilon_j - \sin(tu_{j+1}) \epsilon_j^* \right) \quad (9)$$

where $u_j = \pi j/l$ and ϵ_j, ϵ_j^* are *i.i.d.* standard normal random variables.

Obviously, by simulating a GHW process through (9) we have to consider that $f(\lambda)$ is not defined in its poles. To address this issue, we approximate $f(\lambda_g) \simeq \left(\tilde{\gamma}(0) + 2 \sum_{k=1}^{T-1} \tilde{\gamma}(k) \cos(k\lambda_g) \right) f_x(\lambda_g)$ where $\tilde{\gamma}(k)$ is computed by truncating at T the Wold representation of the autocovariance function of the standardized GHW noise, while $f_x(\lambda_g)$ is defined in Proposition 1.

All the previous relations can be generalized to simulate the k -GHW process.

Let us consider again the GHW process in Proposition 1, assume $\mu = 0$ and an ARMA(1, 1) as short memory component. Figure 3 shows a simulation of this process obtained through the spectral method in (9) with $l = \lfloor \frac{T}{2} \rfloor$ and a comparison between the periodogram and the spectral density along with a comparison between the log-periodogram and the log-spectrum.

In the following, we compare the two simulation methods involving (6) and (9) for the above process by computing the square of the Euclidean distance between the theoretical log spectral density and the log expected value of the periodogram, such that:

$$\mathcal{D}_n = 2 \sum_{\substack{j=1 \\ \lambda_j \neq \lambda_g}}^{\lfloor \frac{T}{2} \rfloor} \left(\ln f(\lambda_j) - \ln E[I(\lambda_j)] \right)^2 \quad (10)$$

where $\left\{ \lambda_j \right\}_{j=1}^{\frac{T}{2}} = \left\{ \frac{2\pi j}{T} \right\}_{j=1}^{\frac{T}{2}}$ is the grid of Fourier frequencies and $I(\lambda_j)$ is the periodogram defined as:

$$I(\lambda_j) = \frac{1}{2\pi T} \left| \sum_{t=0}^{T-1} y_{t+1} e^{it\lambda_j} \right|^2$$

For obvious reason, we do not consider the Gegenbauer frequency when we are computing (10). To get an estimate of $E[I(\lambda_j)]$, we simulate the process with each method for $n = \{250, 500, 1000\}$ reps, then we take the expectation of the periodogram. As is known, see Percival and Walden (1993), the expected value of the periodogram is an asymptotically unbiased estimator of the spectral density. Table 1 shows the ratio between the $\mathcal{D}_n$ measure computed through the spectral method in (9) over the same measure computed via the method involving (6). Let us observe that we do not see a huge difference between the two methods, therefore the spectral approximation approach can be used as a valid alternative.

4 Estimation of the GHW parameters

According to Quinn and Hannan (2001), the periodogram is a visual powerful tool for locating a frequency pole in the theoretical spectral density. However, spurious peaks could mistakenly indicate the presence of a periodicity in the process. This section discusses the estimation in the frequency domain of the

Fig. 3 Simulation of 10^3 realizations from a GHW Gaussian process with an ARMA(1, 1) as short memory component. The true values of the parameters are $\lambda_g = 0.6$, $\phi = 0.8$, $\psi = -0.4$ and $\sigma = 1$ (ϕ and ψ are the AR and MA coefficients, respectively). On the bottom, we show a comparison between the theoretical spectral density and the periodogram of the simulated series

Table 1 Ratio between the $\mathcal{D}_n$ measure defined in (10) computed through the spectral simulation method in (9) over the same measure computed via the truncation method involving (6)

	T=500	T = 1000	T = 1500	T = 2000	T = 2500
$n = 250$					
$\lambda_g = \pi/132$	0.8120	1.0397	1.0398	0.9826	1.0693
$\lambda_g = \pi/8$	1.1518	0.9924	0.9372	0.9285	1.0486
$\lambda_g = \pi/4$	1.0551	1.0099	0.9769	1.0158	0.9998
$\lambda_g = \pi/2$	0.8964	0.9826	1.0572	1.1065	1.0798
$\lambda_g = \pi$	0.9324	0.8646	1.0278	0.9313	0.9007
$n = 500$					
$\lambda_g = \pi/132$	1.0742	1.0967	1.0986	0.9683	0.9444
$\lambda_g = \pi/8$	0.8659	0.9387	0.8794	0.9470	1.0447
$\lambda_g = \pi/4$	0.9221	0.9992	1.0026	0.9634	0.8954
$\lambda_g = \pi/2$	0.8944	0.9453	0.8218	0.9638	1.0205
$\lambda_g = \pi$	1.0461	1.0116	1.0319	1.0447	1.0104
$n = 1000$					
$\lambda_g = \pi/132$	1.1486	1.0860	1.0787	0.9741	0.9400
$\lambda_g = \pi/8$	0.8636	0.9384	1.0367	0.9049	0.9836
$\lambda_g = \pi/4$	0.9938	0.7718	1.0891	0.9537	0.9286
$\lambda_g = \pi/2$	0.9145	1.0828	1.1043	0.9864	1.0484
$\lambda_g = \pi$	0.9931	1.1355	1.0908	0.9712	1.0041

Values lower than one are in bold text

T is the sample size and n is the number of repetitions

GHW model via the maximization of the standard Whittle likelihood (see Whittle (1953)), where we set the local maximizers of the periodogram as initial guess of the frequency parameters vector in the optimization algorithm.

Let $y = [y_1, \dots, y_T]^\top$ be T realizations of a random variable $Y \sim N(0, \Sigma_\theta)$ where Σ_θ is the covariance matrix. According to Dzhaparidze and Yaglom (1983), the exact log-likelihood function can be written as:

$$LL(\theta) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} y^\top \Sigma_\theta^{-1} y - \frac{1}{2} \ln |\Sigma_\theta| \quad (11)$$

where θ is the parameters vector.

The Whittle likelihood is an approximation of (11) where the covariance terms in the time domain are substituted by spectral terms in the frequency domain. It is given from the following function:

$$L_W(\theta) = -\frac{T}{4\pi} \int_{-\pi}^{\pi} \ln f(\lambda; \theta) d\lambda - \frac{T}{4\pi} \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda; \theta)} d\lambda \quad (12)$$

where $f(\lambda; \theta)$ is the spectral density evaluated at the θ parameters and T is the sample size. The integral terms in (12) can be approximated via Riemann sums (c.f. Beran (1994)), such that:

$$\tilde{L}_W(\theta) = - \sum_{t=1}^{\lfloor \frac{T}{2} \rfloor} \ln f(\lambda_t; \theta) - \sum_{t=1}^{\lfloor \frac{T}{2} \rfloor} \frac{I(\lambda_t)}{f(\lambda_t; \theta)}$$

Relatively to maximum exact likelihood estimation, the Whittle approach simplifies considerably computations (from $O(T^2)$ to $O(T \ln(T))$). Furthermore, difficulties encountered with a non-zero mean, which is hard to estimate under long memory, can be avoided by performing Whittle estimation, since the periodogram is invariant with respect to an additive constant like the mean (see Hassler (2019) for further details).

Consider now the k -GHW process in Proposition 2 and let $f_x(\lambda_t) = \frac{\sigma^2}{2\pi}$ be the spectral density of the short memory component. We can estimate the model parameters $\theta = [\lambda_{g,1}, \dots, \lambda_{g,k}, \sigma^2]^\top$ via the maximization of (12), with $\theta \in \Theta \subset S$ where Θ is assumed to be a compact set and S follows from Proposition 2. Defining the k -GHW spectral density as $f(\lambda_t) = \sigma^2 h(\lambda_t)$, it is easy to verify that $\frac{\partial \tilde{L}_W(\theta)}{\partial \sigma^2} = 0$ implies $\hat{\sigma}^2 = \frac{2}{T} \sum_{j=0}^{\lfloor \frac{T}{2} \rfloor} \frac{I(\lambda_j)}{h(\lambda_j)}$. Therefore, the parameter σ^2 can be concentrated outside the likelihood.

In the following equations, we provide the gradient and the Hessian of the Whittle likelihood function for a k -GHW noise. The gradient is given by:

$$\frac{\partial L_W(\theta)}{\partial \theta_i} = \frac{T}{4\pi} \int_{-\pi}^{\pi} \left(\frac{I(\lambda)}{f(\lambda)} - 1 \right) \frac{\partial \ln f(\lambda)}{\partial \theta_i} d\lambda \quad (13)$$

The ij -element of the $(k+1) \times (k+1)$ Hessian matrix can be computed as:

$$H_{ij}(\theta) = -\frac{T}{4\pi} \int_{-\pi}^{\pi} \frac{\partial \ln f(\lambda)}{\partial \theta_i} \frac{\partial \ln f(\lambda)}{\partial \theta_j} \left(\frac{I(\lambda)}{f(\lambda)} \right) d\lambda \quad (14)$$

for $i, j \in \{1, 2, \dots, k+1\}$. The terms in (13) and (14) can be obtained as follows:

$$\frac{\partial \ln f(\lambda)}{\partial \lambda_{g,i}} = \frac{2 \ln(2|\cos(\lambda) - \cos(\lambda_{g,i})|) \sin(\lambda_{g,i})}{\cos(\lambda) - \cos(\lambda_{g,i})} (\ln^2(2|\cos(\lambda) - \cos(\lambda_{g,i})|) + (\alpha_i \pi - \lambda)^2)^{-1};$$

for $i, j = 1, 2, \dots, k$ and $\frac{\partial \ln f(\lambda)}{\partial \sigma^2} = \frac{1}{\sigma^2}$.

Notice that if $\lambda_{g,i}$ approaches to zero, then the Hessian in (14) becomes singular because of the sine function. According to Dzharidze and Yaglom (1983), defining $\Gamma^{-1}(\theta)$ as the asymptotic covariance matrix for the Whittle estimator $\hat{\theta}_T = \operatorname{argmax}_{\theta \in \Theta} L_W(\theta)$, we can obtain the ij -element of $\Gamma(\theta)$ in the usual way as:

$$\lim_{T \rightarrow \infty} -\frac{1}{T} E(H_{ij}(\theta)) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial \ln f(\lambda)}{\partial \theta_i} \frac{\partial \ln f(\lambda)}{\partial \theta_j} d\lambda = \Gamma_{ij}(\theta)$$

where the first equality is implied by Theorem 3.4 in Zygmund and Fefferman (2003).

4.1 Consistency of the parameters

In this subsection, we will see that, under some regularity conditions, the Whittle estimator is a consistent estimator of the k -GHW process with an ARMA(p, q) as short memory component. Following the asymptotic theory in Hannan (1973), Giraitis and Leipus (1995) and Giraitis et al. (2001) provided the consistency of the Whittle estimator for a k -factor GARMA process with unknown poles in the spectrum.

Consider the k -GHW process in Proposition 2 and let $f_x(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{\psi(e^{-i\lambda})}{\phi(e^{-i\lambda})} \right|^2$ be the spectral density of the short memory ARMA(p, q) component. Let $\theta = [\sigma^2, \lambda_{g,1}, \dots, \lambda_{g,k}, \phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q]^\top$ be the parameters vector such that $\theta \in \Theta \subset S$ where Θ is assumed to be a compact set and S follows from Proposition 2. The next theorem establishes the consistency of the Whittle estimator $\hat{\theta}_T$.

Theorem 1 *Let y_t be a k -factor GHW process such that:*

$$y_t = \sum_{i=0}^{\infty} b_i \varepsilon_t = \left(\prod_{j=1}^k h_j(L) \right) \frac{\psi(L)}{\phi(L)} \varepsilon_t$$

where $h_j(L) = \sum_{i=0}^{\infty} \frac{\cos(\lambda_{g,j}(i+1))}{i+1} L^i$ and $\varepsilon_t \sim WN(0, \sigma^2)$.

Let $\theta = [\sigma^2, \lambda_{g,1}, \dots, \lambda_{g,k}, \phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q]^\top$ be the vector of the parameters such that $\theta \in \Theta \subset S$ where Θ is assumed to be a compact set and $S = (0, \infty) \times \Lambda^k \times \mathbb{R}^{p+q}$ with $\Lambda = [0, \pi]$. Under the following conditions

- (i) $\sum_{i=0}^{\infty} b_i^2 < \infty$,
- (ii) $\phi(0) = \psi(0) = 1$,
- (iii) The polynomials $\phi(\cdot)$ and $\psi(\cdot)$ have no common zeros such that $\phi(z) \neq 0$ and $\phi(0) \neq 0$ for $|z| < 1$,
- (iv) The spectral density $f(\lambda; \theta)$ of the process y_t is uniquely determined by the parameters in θ ,

the Whittle estimator $\hat{\theta}_T$ converges in probability to the true value of the parameters:

$$\hat{\theta}_T \xrightarrow{p} \theta_0 \quad \text{as } T \rightarrow \infty$$

Proof The proof of Theorem 1 is given in the Appendix. □

4.2 A simulation experiment

In the following, we implement a Monte Carlo experiment simulating 10^3 times a sample size of $T = \{100, 200, 600, 1000\}$ realizations generated by the GHW process in Proposition 1, with $\mu = 0$ be known and an ARMA(1, 1) as short memory component. At the end, the model is estimated via the standard concentrated Whittle likelihood.

Process simulations have been carried out via the spectral method in (9), considering different values of λ_g and setting the short memory parameters as $\sigma = 1$, $\phi = 0.8$ and $\psi = -0.4$, where ϕ and ψ are the AR and MA coefficients, respectively. Initial guesses in the optimization algorithm are set equal to the maximizer of the periodogram for λ_g , while they are obtained through a grid search procedure for the ϕ and ψ parameters under the restrictions in Proposition 1. Table 2 shows the parameters bias and mean square error (MSE). Let us observe that they becomes smaller for each parameter as the sample size increases, matching the statement in Theorem 1.

4.3 How the GHW model fits the Gegenbauer process

In this subsection, we investigate the issue of how the GHW model fits data generated from a Gegenbauer process. Let us start by considering a fractionally differenced Gegenbauer noise as the data generating process (DGP), whose theoretical spectral density is defined in (4). Data are simulated considering a type I and type II processes as in Sect. 3.3 for different values of the d parameter, while the Gegenbauer frequency is set to $\pi/4$ and $\sigma^2 = 1$. Then, the GHW filter in (3) is applied to the simulated series assuming $\mu = 0$ and $\lambda_g = \pi/4$ be known.

In the following, a Monte Carlo experiment has been carried out considering 10^3 reps, where the Gegenbauer fractional noise is simulated at each repetition and then filtered via (3). Finally, an ARMA(p, q) component is estimated from the filter output via the exact likelihood method.

Table 3 shows the squared Euclidean distance between the log theoretical spectral density of the Gegenbauer DGP and the log estimated spectral density of the GHW(p, q) model, where p and q refer to the order of the ARMA short memory component. The squared Euclidean distance is computed via the numerical approximation of the following integral:

$$\mathcal{D}_T = 2 \int_0^\pi (\ln f_{\text{GARMA}}(\lambda) - \ln \hat{f}_{\text{GHW}}(\lambda))^2 d\lambda$$

where $f_{\text{GARMA}}(\lambda)$ and $f_{\text{GHW}}(\lambda)$ are the spectral densities of the Gegenbauer and GHW processes, respectively. Let us observe that $\mathcal{D}_T$ is lower when the d parameter is larger and an AR(1) component is assumed (see columns GHW(1,0) and GHW(1,1)). Therefore, it should be feasible to state that the weaker persistence characterizing the GHW process with respect to the Gegenbauer may be compensated by assuming an autoregressive component. This is also confirmed by the visual examination of Fig. 4, where it is clear that the GHW log estimated spectrum

Table 2 Bias and MSEs of the Whittle estimator for 10^3 simulated series generated by a GHW process with an ARMA(1, 1) as short memory component

		$Bias(\hat{\lambda}_g)$	$Bias(\hat{\sigma})$	$Bias(\hat{\phi})$	$Bias(\hat{\psi})$
$T = 100$	$\lambda_g = \pi/132$	0.5018	0.0388	-0.0637	0.1780
	$\lambda_g = \pi/8$	0.2292	-0.0203	-0.2103	0.2517
	$\lambda_g = \pi/4$	0.0382	-0.0354	-0.1734	0.1359
	$\lambda_g = \pi/2$	-0.0308	-0.0583	-0.1459	0.1208
$T = 200$	$\lambda_g = \pi/132$	0.2047	0.0252	-0.0097	0.0382
	$\lambda_g = \pi/8$	0.1034	-0.0182	-0.1203	0.1272
	$\lambda_g = \pi/4$	0.0132	-0.0141	-0.0446	0.0279
	$\lambda_g = \pi/2$	0.0079	-0.0625	-0.0992	0.0921
$T = 600$	$\lambda_g = \pi/132$	0.0260	0.0079	-0.0080	0.0016
	$\lambda_g = \pi/8$	-0.0198	-0.0053	-0.0313	0.0173
	$\lambda_g = \pi/4$	-0.0018	-0.0022	-0.0033	-0.0048
	$\lambda_g = \pi/2$	0.0111	-0.0161	-0.0249	0.0196
$T = 1000$	$\lambda_g = \pi/132$	0.0117	0.0035	-0.0063	0.0024
	$\lambda_g = \pi/8$	-0.0150	-0.0029	-0.0214	0.0147
	$\lambda_g = \pi/4$	0.0009	0.0007	-0.0022	0.0024
	$\lambda_g = \pi/2$	-0.0133	-0.0064	-0.0113	-0.0006
		$MSE(\hat{\lambda}_g)$	$MSE(\hat{\sigma})$	$MSE(\hat{\phi})$	$MSE(\hat{\psi})$
$T = 100$	$\lambda_g = \pi/132$	1.2161	0.0126	0.0866	0.2076
	$\lambda_g = \pi/8$	0.7296	0.0085	0.2080	0.2549
	$\lambda_g = \pi/4$	0.2125	0.0191	0.2120	0.2060
	$\lambda_g = \pi/2$	0.2691	0.0329	0.1245	0.1622
$T = 200$	$\lambda_g = \pi/132$	0.3975	0.0050	0.0207	0.0594
	$\lambda_g = \pi/8$	0.3875	0.0044	0.0854	0.1132
	$\lambda_g = \pi/4$	0.0598	0.0074	0.0537	0.0605
	$\lambda_g = \pi/2$	0.2910	0.0289	0.0697	0.1342
$T = 600$	$\lambda_g = \pi/132$	0.0107	0.0011	0.0033	0.0060
	$\lambda_g = \pi/8$	0.0066	0.0013	0.0115	0.0096
	$\lambda_g = \pi/4$	0.0006	0.0015	0.0015	0.0048
	$\lambda_g = \pi/2$	0.0871	0.0084	0.0182	0.0443
$T = 1000$	$\lambda_g = \pi/132$	0.0014	0.0006	0.0016	0.0028
	$\lambda_g = \pi/8$	0.0051	0.0008	0.0072	0.0060
	$\lambda_g = \pi/4$	0.0002	0.0007	0.0008	0.0024
	$\lambda_g = \pi/2$	0.0340	0.0033	0.0077	0.0116

The data are simulated via the spectral method

The true values of the parameters are $\sigma = 1$, $\phi = 0.8$ and $\psi = -0.4$. We consider different sample sizes of $T = 100, 200, 600$, and 1000

fits the theoretical Gegenbauer log spectral density almost ideally if an AR component is considered along with an high value of the memory parameter. On the other hand, it is also evident that the GHW model fails in fitting the generated data when d

Table 3 Squared Euclidean distance between the log theoretical spectral density of the Gegenbauer DGP and the log estimated spectral density of the GHW(p, q) model

		GHW(0,0)	GHW(1,0)	GHW(0,1)	GHW(1,1)
$T = 250$	$d = 0.15$	4.3978	2.8105	3.6830	0.8071
	$d = 0.25$	5.1919	1.5405	3.2429	0.2980
	$d = 0.35$	7.0572	0.6768	3.2721	0.1574
	$d = 0.45$	10.2464	0.3614	3.8638	0.3028
$T = 500$	$d = 0.15$	4.4455	2.8448	3.7231	0.8106
	$d = 0.25$	5.2112	1.5499	3.2622	0.2973
	$d = 0.35$	6.9841	0.6785	3.2491	0.1560
	$d = 0.45$	10.0779	0.3523	3.8029	0.2865
$T = 1000$	$d = 0.15$	4.4728	2.8596	3.7488	0.8123
	$d = 0.25$	5.2023	1.5522	3.2625	0.2977
	$d = 0.35$	6.9680	0.6791	3.2450	0.1556
	$d = 0.45$	9.9850	0.3490	3.7696	0.2814

assumes small values, even if an AR coefficient is estimated. In this case, we observe that the outcomes may be improved by estimating an ARMA(1,1) component.

5 Empirical application on EPICA dome C temperature reconstructions

The European Project for Ice Coring in Antarctica (EPICA) is a multinational European activity for deep ice core drilling. Its main objective is to report full documentation of the atmospheric record archived in Antarctic ice. Evaluation of these results provides information about the natural climate variability and mechanisms of climatic changes.

Figure 5 shows the transformed series of the paleoclimatic temperature reconstructions since $\sim 800,000$ years ago. The series is obtained by interpolating the data set provided by Jouzel and Masson-Delmotte (2007), who elaborated the EPICA Dome C ice cores records, considering a fix time interval of 1,000 years. Then, data have been transformed according to Davidson et al. (2016) such that $y_t = \ln(z_t + 16)$, where z_t is the original series of length $T = 801$ observations obtained from the interpolation. Previous analysis on this data set can be found in Davidson et al. (2016) and Castle and Hendry (2020).

By visual inspection of the periodogram on the bottom of Figure 5, we easily identify two main periodicities of 100 and 41 thousands of years (kyrs). This result is coherent with respect to the paleoclimatic literature (see Petit et al. (1999)). Although others shorter cycles could be considered for this series, a preliminary analysis on residuals suggests the choice of $k = 2$ and $\lambda_g = [0.06 \ 0.16]^T$ as optimal. In particular, the possibility to include a further cycle (for instance the one related to the tiny peak on the periodogram at frequency $\lambda = 0.27$) is excluded after the residuals analysis of the 3-GHW(p, q) estimation (here, the notation (p, q) refers

GARMA vs GHW(0,0)

GARMA vs GHW(1,0)

GARMA vs GHW(1,1)

Fig. 4 Comparison between the log theoretical spectral density of the Gegenbauer DGP and the log estimated spectral density of the GHW(p, q) model, considering $T = 250$ (left column) and $T = 1000$ (right column) as sample sizes

to the order of the ARMA short memory component). In this case, the square sum of the sample autocorrelations $\sum_{k=1}^{T-1} \hat{\rho}^2(k)$ results to be higher for the 3-GHW(1,0), 3-GHW(0,1) and 3-GHW(1,1) models with respect to the 2-GHW counterparts. At the end, the choice $k = 2$ is hence preferred. Moreover, we assume the frequency

Fig. 5 On the top: transformed series of the EPICA Dome C temperature reconstructions. It consists on 801 observations with time interval length of 1000 years. The zero on the x-axis corresponds to the present age (1950). On the bottom: periodogram of the above series

parameters vector to be known such that $\lambda_g = [0.06 \ 0.16]^\top$ and it will not be considered again in the estimation.

Let us proceed by estimating both a 2-factor GHW and GARMA models for these data. The 2-GARMA model is estimated through the Whittle approach in a one-step procedure. On the other hand, the 2-GHW model is estimated in a 2-step procedure. First, we filter the series via the GHIT function in order to obtain the short memory component x_t . Then, we fit an $\text{ARMA}(p, q)$ process for x_t through the maximization of the exact likelihood. We do not consider p and q orders larger than one since the choice of a higher-order ARMA component does not change significantly our final remarks.

Table 4 shows the estimation results. Let us observe that the memory parameters estimates belong to the stationary region in the GARMA case. The short memory component is obtained via:

$$x_t = F(L)y_t - F(1)\hat{\mu}, \quad t = 1, \dots, T \quad (15)$$

where $F(L) = \prod_{j=1}^2 g_j(L)$ if the 2-factor GHIT filter is considered, while $F(L) = \prod_{j=1}^2 (1 - 2\cos(\lambda_{g,j})L + L^2)^{d_j}$ in the 2-GARMA case. $\hat{\mu}$ is the sample mean estimator computed as $\hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t$. Figure 6 shows the sample autocorrelation function (ACF), the sample partial ACF and the periodogram of the short memory components for the 2-GHW and 2-GARMA models obtained through (15), given the parameters estimates of the first and fifth rows in Table 4. It should be clear

Table 4 Estimation results of the 2-factor GHW and GARMA models with an ARMA(p, q) as short memory components for the EPICA Dome C paleoclimatic temperature reconstructions

	AICs	$\hat{d}_1$	$\hat{d}_2$	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{\psi}$	$\Sigma \hat{\rho}^2(k)$
2-GHW(0,0)	-1963.83			0.0050			1.2347
2-GHW(1,0)	-2042.28			0.0046	0.3192		0.6550
2-GHW(0,1)	-2031.80			0.0046		0.2851	0.7478
2-GHW(1,1)	-2054.68			0.0045	0.7230	-0.5061	0.6277
2-GARMA(0,0)	-19555.06	0.2228	0.3158	0.0051			0.6380
2-GARMA(1,0)	-19698.01	0.1974	0.1440	0.0049	0.4753		0.6140
2-GARMA(0,1)	-19709.82	0.2291	0.1941	0.0048		0.3356	0.5899
2-GARMA(1,1)	-19714.07	0.2171	0.1642	0.0048	0.2191	0.2090	0.6023

AIC is the Akaike information criterion, σ^2 is the variance parameter while ϕ and ψ are the AR and MA coefficients, respectively. The others parameters follows from (15)

Fig. 6 Sample ACF, sample partial ACF and periodogram of the short memory components obtained via (15) considering the 2-GHW(0,0) and 2-GARMA(0,0) models in Table 4

that both approaches are able to reduce the main peaks in the EPICA Dome C periodogram, although the GARMA filter performs slightly better in removing the 100 kyrs periodicity. On the other hand, the two models perform almost equal if we compare the correlogram and the periodogram in Fig. 7 obtained from the 2-GHW(1,0) and 2-GARMA(1,0) estimates in Table 4, where an AR(1) component is considered. In the last column of Table 4, we also show the sum of square of autocorrelations.

Fig. 7 Sample ACF, sample partial ACF and periodogram of the residuals obtained by estimating an AR(1) as short memory component for the 2-GHW(1,0) and 2-GARMA(1,0) models in Table 4

Notice that the two models are very close in controlling for the long-memory features if an AR coefficient is estimated.

In the following, we perform a rolling forecast experiment with a moving window of length $W = 300$ observations such that we compute predictions of the EPICA Dome C temperature reconstructions according to a 2-GHW and 2-GARMA models with an ARMA(1, 1) as short memory component. The h -step-ahead predictor $\hat{y}_{t+h|t}$ is obtained in the standard way with the support of the Durbin-Levinson (DL) recursion (see Brockwell and Davis (1986)). We set the truncation parameter $\tilde{n}$ in the DL recursion by finding the lower Mean Square Forecast Error (MSFE) for each model. At the end, we set $\tilde{n} = 290$.

Figures 8 and 9 show the correlogram and the periodogram relatively to the residuals obtained from the one-step-ahead linear predictor. Notice that both models are able to capture the cyclical long memory removing the main peaks in the EPICA Dome C periodogram. On the other hand, the correlograms show very weak auto-correlation. On the top of Fig. 10, an illustration of the five-step-ahead forecasts for the two models is provided for example purposes. On the bottom of Fig. 10, the two models are compared in terms of predictability, referring to the following measure:

$$\mathcal{P}(h) = (T - W) \left(\sum_{t=W+h}^T (y_t - \hat{y}_{t|t-h})^2 \right)^{-1}$$

Fig. 8 Residuals obtained from the one-step-ahead forecasts computed via the DL recursion considering the 2-GHW(1,1) model

Fig. 9 Residuals obtained from the one-step-ahead forecasts computed via the DL recursion considering the 2-GARMA(1,1) model

Fig. 10 Five-step-ahead forecasts and predictability for the EPICA Dome C temperature reconstructions

which is basically the reciprocal of the MSFE as a function of the prediction horizon h . Notice that it is almost impossible to distinguish between the two approaches in terms of forecasting performance.

All these results lead to the conclusion that the GHW model behaves more or less similar to its classical competitor, the GARMA process, in controlling for long memory in quasi-periodic time series.

6 Conclusions

The estimation of the d parameter is a widely discussed issue in the long memory literature (see for instance Geweke and Porter-Hudak (1983), Beran (1994), Robinson (1995a, 1995b), Ferrara and Guegan (1999), Arteche and Robinson (2000) and Hassler (2019) for a review). Recently, Hassler and Hosseinkouchack (2020) introduced a new way to model long memory without estimate d . In this paper, a generalization of their contribution is proposed through the introduction of a novel process for cyclical long memory time series. Our main conclusion is that the GHW process is able to model long memory, as well as the GARMA does. In addition, the GHW process does not require the estimation of the memory parameters, simplifying the issue of how to disentangle long memory from a (moderately persistent) short memory component. This leads to a clear advantage of our formulation over its classical competitor, the Gegenbauer process. However, the k -GHW model does not allow for a good control on the single contribution of each pole in the spectrum

to the periodicity of the process. This is in contrast to the GARMA model, where the contribution of each singularity is regulated by the d parameters. On the other hand, although the 1-factor GHW model displays less persistence with respect to the fractionally integrated Gegenbauer process, this weaker persistence may be compensated by estimating a further ARMA(p, q) short memory component. In particular, we observed that the AR coefficient plays a significant role. Furthermore, the GHW model is more parsimonious and allows to consider more frequency parameters in the GHIT filter without estimate further d parameters. Moreover, if we consider the frequency parameters vector to be known (for instance, by estimating it through the periodogram in a preliminary estimation step), we can obtain the short memory component just by filtering the data. Even so, we must be careful when we add frequency parameters to the GHIT filter, in order to prevent overfitting and antipersistence. The selection of the length k of the frequency parameters vector is a crucial step. In fact, if k is too large, it would lead to an antipersistent output and to overfitting. On the other hand, if k is too small, we will not be able to “remove” completely the periodicity from the data. Therefore, a preliminary analysis on filter residuals and on the periodogram of the short memory component is recommended.

Finally, we provide analytical expressions for the spectral density and the autocovariance function of the GHW process along with the consistency of the Whittle estimator. These mathematical tools could be useful in further researches to get worthy results in empirical applications.

Appendix

Proof of Proposition 1

Using $\ln(1 - z) = -\sum_{j=1}^{\infty} \frac{z^j}{j}$ we can write (with $\iota = \sqrt{-1}$):

$$\begin{aligned} \ln(1 - 2\cos(\lambda_g)z + z^2) &= \ln(1 - ze^{\iota\lambda_g}) + \ln(1 - ze^{-\iota\lambda_g}) \\ &= -\sum_{j=1}^{\infty} \left(\frac{z^j e^{-\iota j\lambda_g}}{j} + \frac{z^j e^{\iota j\lambda_g}}{j} \right) \\ &= -2\sum_{j=1}^{\infty} \frac{z^j \cos(\lambda_g j)}{j} \end{aligned}$$

Therefore, the following equations hold:

$$h(L) = -\frac{\ln(1 - 2\eta L + L^2)}{2L} = \sum_{j=0}^{\infty} \frac{\cos(\lambda_g(j+1))}{(j+1)} L^j$$

We get the coefficients of $g(z)$ and $b(z)$ by equalling the following polynomials:

$$1 = h(L)g(L) = \left(\tilde{b}_0 + \frac{\tilde{b}_1}{2}z + \frac{\tilde{b}_2}{3}z^2 + \cdots \right) (g_0 + g_1z + g_2z^2 + \cdots)$$

and

$$\begin{aligned} h(L)c(L) &= \left(\tilde{b}_0 + \frac{\tilde{b}_1}{2}z + \frac{\tilde{b}_2}{3}z^2 + \dots \right) (c_0 + c_1z + c_2z^2 + \dots) \\ &= (b_0 + b_1z + b_2z^2 + \dots) = b(L) \end{aligned}$$

Finally, we prove the equation for the spectral density by considering:

$$f(\lambda) = \left| \frac{\ln(1 - 2\eta e^{-i\lambda} + e^{-i2\lambda})}{2e^{-i\lambda}} \right|^2 f_x(\lambda) = \frac{|\ln(r(\lambda)e^{i\varphi(\lambda)})|^2}{4} f_x(\lambda)$$

We proceed by computing the modulus and the argument of the complex number inside the logarithm. The first one is given by:

$$\begin{aligned} r(\lambda) &= |1 - 2\cos(\lambda_g)\cos(\lambda) + \cos(2\lambda) + i(2\cos(\lambda_g)\sin(\lambda) \\ &\quad - \sin(2\lambda))|^2 = 2|\cos(\lambda) - \cos(\lambda_g)| \end{aligned}$$

On the other hand, we get the argument $\varphi(\lambda)$ considering the ratio:

$$\begin{aligned} &\frac{2\cos(\lambda_g)\sin(\lambda) - \sin(2\lambda)}{1 - 2\cos(\lambda_g)\cos(\lambda) + \cos(2\lambda)} \\ &= \frac{\sin(\lambda)(\cos(\lambda_g) - \cos(\lambda))}{-\cos(\lambda)(\cos(\lambda_g) - \cos(\lambda))} = \frac{v(\lambda)}{u(\lambda)} \end{aligned}$$

where $u(\lambda)$ and $v(\lambda)$ are, respectively, defined as the real and imaginary parts of the complex number inside the logarithm, such that:

$$\varphi(\lambda) = \begin{cases} \arctan\left(\frac{v(\lambda)}{u(\lambda)}\right), & \text{if } u(\lambda) > 0 \\ \arctan\left(\frac{v(\lambda)}{u(\lambda)}\right) + \pi, & \text{if } u(\lambda) < 0 \text{ and } v(\lambda) \geq 0 \\ \arctan\left(\frac{v(\lambda)}{u(\lambda)}\right) - \pi, & \text{if } u(\lambda) < 0 \text{ and } v(\lambda) < 0 \end{cases}$$

with $\varphi(\lambda) \in (-\pi, \pi]$.

Let us consider $\lambda \in (-\pi, \pi)$ and recall that the tangent function is not defined in $\frac{\pi}{2}$ while the identity $\arctan(\tan(x)) = x$ holds only if $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$. Therefore, considering $\tan(\lambda) = \tan(\lambda - \pi)$ if $\lambda > \frac{\pi}{2}$ and $\tan(\lambda) = \tan(\lambda + \pi)$ if $\lambda < -\frac{\pi}{2}$, we get:

- if $\cos(\lambda_g) - \cos(\lambda) > 0$ (which implies $\lambda \notin (-\lambda_g, \lambda_g)$):

$$\varphi(\lambda) = \begin{cases} \arctan(-\tan(\lambda + \pi)) = -\lambda - \pi, & \text{if } -\pi < \lambda < -\frac{\pi}{2} \\ \arctan(-\tan(\lambda)) - \pi = -\lambda - \pi, & \text{if } -\frac{\pi}{2} < \lambda < 0 \\ \arctan(-\tan(\lambda)) + \pi = -\lambda + \pi, & \text{if } 0 < \lambda < \frac{\pi}{2} \\ \arctan(-\tan(\lambda - \pi)) = -\lambda + \pi, & \text{if } \frac{\pi}{2} < \lambda < \pi \end{cases}$$

- if $\cos(\lambda_g) - \cos(\lambda) < 0$ (which implies $\lambda \in (-\lambda_g, \lambda_g)$):

$$\varphi(\lambda) = \begin{cases} \arctan(-\tan(\lambda + \pi)) + \pi = -\lambda, & \text{if } -\pi < \lambda < -\frac{\pi}{2} \\ \arctan(-\tan(\lambda)) = -\lambda, & \text{if } -\frac{\pi}{2} < \lambda < 0 \\ \arctan(-\tan(\lambda)) = -\lambda, & \text{if } 0 < \lambda < \frac{\pi}{2} \\ \arctan(-\tan(\lambda - \pi)) - \pi = -\lambda, & \text{if } \frac{\pi}{2} < \lambda < \pi \end{cases}$$

Finally, we end the proof by solving:

$$f(\lambda) = \frac{\ln^2(r(\lambda)) + \varphi^2}{4} f_x(\lambda)$$

Notice that if $\lambda_g = 0$ (which implies $\eta = 1$), the spectral density in Proposition 1 coincides with the one of the HW process.

We get the asymptotic behaviour of the spectrum as $\lambda \rightarrow \lambda_g$ by taking the Taylor expansion of $\cos(\lambda)$ as $\lambda \rightarrow \lambda_g$. If $\lambda_g \notin \{0, \pi\}$, we consider the first order expansion $\cos(\lambda) \sim \eta - \sin(\lambda_g)(\lambda - \lambda_g)$. If $\lambda_g \in \{0, \pi\}$, we need the second-order expansion $\cos(\lambda) \sim 1 - \frac{(\lambda - \lambda_g)^2}{2}$. $\square$

Proof of Theorem 1

Let us define the spectral density of the process y_t as $f(\lambda, \theta) = \frac{\sigma^2}{2\pi} h(\lambda, \theta)$. Notice that the stationarity condition (i) in Theorem 1 always holds for the k -GHW process in Proposition 2 because $\sum_{i=1}^{\infty} \frac{\cos(i\lambda_{g,j})^2}{i^2} < \infty$ for each $\{\lambda_{g,j}\}_{j=1}^k \in \Lambda$ and $x_t = \frac{\psi(L)}{\phi(L)} \varepsilon_t$ is defined as an ARMA(p, q) process. Moreover, the square summability in i) implies that $\int_{-\pi}^{\pi} \ln h(\lambda, \theta) d\lambda = 0$ holds (see Theorem 5.8.1 in Brockwell and Davis (1986)). Conditions (ii) and (iii) follow directly from the definition of x_t as an ARMA(p, q) process, while condition (iv) holds since the spectral density of the k -GHW process is determined uniquely by θ .

Finally, if conditions (i), (ii), (iii) and (iv) are satisfied, then Theorem 3 in Giraitis and Leipus (1995) holds and the Whittle estimator converges in probability to the true value of the parameters.

Funding Open access funding provided by Università degli Studi di Roma Tor Vergata within the CRUI-CARE Agreement. Not applicable.

Data availability statement The data sets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Code availability All the applications and simulations described in this paper were executed using the 64 bit version of MATLAB-R 2019b. The code is available from the corresponding author on reasonable request.

Declarations

Conflict of interest The author declare that he have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Andel, J.: Long memory time series models. *Kybernetika* **22**(2), 105–123 (1986)
- Arteche, J., Robinson, P.M.: Semiparametric inference in seasonal and cyclical long memory processes. *J. Time Ser. Anal.* **21**(1), 1–25 (2000)
- Beran, J.: *Statistics for Long Memory Processes* Monographs on Statistics and Applied Probability. Chapman & Hall, New York (1994)
- Brockwell, P.J., Davis, R.A.: *Time Series: Theory and Methods*. Springer, Berlin (1986)
- Castle, J.L., Hendry, D.F.: Climate econometrics: an overview. *Found. Trends Econ.* **10**(3–4), 145–322 (2020)
- Chung, C.F.: Estimating a generalized long memory process. *J. Econ.* **73**(1), 237–259 (1996a)
- Chung, C.F.: A generalized fractionally integrated autoregressive moving-average process. *J. Time Ser. Anal.* **17**(2), 111–140 (1996b)
- Davidson, J.E.H., Stephenson, D.B., Turasie, A.A.: Time series modeling of paleoclimate data. *Environmetrics* **27**, 55–65 (2016)
- Dieker, A., Mandjes, M.: On spectral simulation of fractional Brownian motion. *Probab. Eng. Inf. Sci.* **17**(3), 417–434 (2003)
- Dissanayake, G., Peiris, M., Proietti, T., et al.: Fractionally differenced Gegenbauer processes with long memory: a review. *Stat. Sci.* **33**(3), 413–426 (2018)
- Dzhaparidze, K.O., Yaglom, A.M.: Spectrum parameter estimation in time series analysis. In: Krishnaiah, P.R. (ed.) *Developments in Statistics*. Academic Press, New York (1983)
- Ferrara, L., Guegan, D.: Estimation and applications of Gegenbauer processes. Working Papers 99-27, Center for Research in Economics and Statistics (1999)
- Geweke, J., Porter-Hudak, S.: The estimation and application of long memory time series models. *J. Time Ser. Anal.* **4**(4), 221–238 (1983)
- Giraitis, L., Leipus, R.: A generalized fractionally differencing approach in long-memory modelling. *Lith. Math. J.* (1995)
- Giraitis, L., Hidalgo, J., Robinson, P.M.: Gaussian estimation of parametric spectral density with unknown pole. *Ann. Stat.* **29**(4), 987–1023 (2001)
- Gray, H.L., Zhang, N.F., Woodward, W.A.: On generalized fractional processes. *J. Time Ser. Anal.* **10**(3), 233–257 (1989)
- Hannan, E.J.: The asymptotic theory of linear time-series models. *J. Appl. Probab.* **10**(1), 130–145 (1973)
- Hassler, U.: *Time Series Analysis with Long Memory in View*. Wiley, New York (2019)
- Hassler, U., Hosseinkouchack, M.: Harmonically weighted processes. *J. Time Ser. Anal.* **41**(1), 41–66 (2020)
- Hosking, J.R.M.: Fractional differencing. *Biometrika* **68**(1), 165–176 (1981)
- Jouzel, J., Masson-Delmotte, V.: EPICA Dome C Ice Core 800KYr deuterium data and temperature estimates. *PANGAEA* (2007)
- Leschinski, C., Sibbertsen, P.: Model order selection in periodic long memory models. *Econ. Stat.* **9**, 78–94 (2019)
- Palma, W.: *Long-Memory Time Series: Theory and Methods*. Wiley Series in Probability and Statistics, Wiley, New York (2006)
- Palma, W., Chan, N.H.: Efficient estimation of seasonal long-range-dependent processes. *J. Time Ser. Anal.* **26**(6), 863–892 (2005)

- Percival, D.B., Walden, A.T.: Spectral Analysis for Physical Applications. Cambridge University Press, Cambridge (1993)
- Petit, J., Jouzel, J., Raynaud, D., et al.: Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999)
- Quinn, B.G., Hannan, E.J.: The Estimation and Tracking of Frequency. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge (2001)
- Reisen, V.A., Rodrigues, A.L., Palma, W.: Estimation of seasonal fractionally integrated processes. *Comput. Stat. Data Anal.* **50**(2), 568–582 (2006)
- Robinson, P.M.: Gaussian semiparametric estimation of long range dependence. *Ann. Stat.* **23**(5), 1630–1661 (1995a)
- Robinson, P.M.: Log-periodogram regression of time series with long range dependence. *Ann. Stat.* **23**(3), 1048–1072 (1995b)
- Velasco, C., Robinson, P.M.: Whittle pseudo-maximum likelihood estimation for nonstationary time series. *J. Am. Stat. Assoc.* **95**(452), 1229–1243 (2000)
- Whittle, P.: Estimation and information in stationary time series. *Ark. Math.* **2**(5), 423–434 (1953)
- Woodward, W.A., Cheng, Q.C., Gray, H.L.: A k-factor Gamma long-memory model. *J. Time Ser. Anal.* **19**(4), 485–504 (1998)
- Zygmund, A., Fefferman, R.: Trigonometric Series, 3rd edn. Cambridge University Press, Cambridge (2003)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.