

Modelling Cycles in Climate Series: the Fractional Sinusoidal Waveform Process

Tommaso Proietti **Federico Maddanu**
Università di Roma "Tor Vergata"

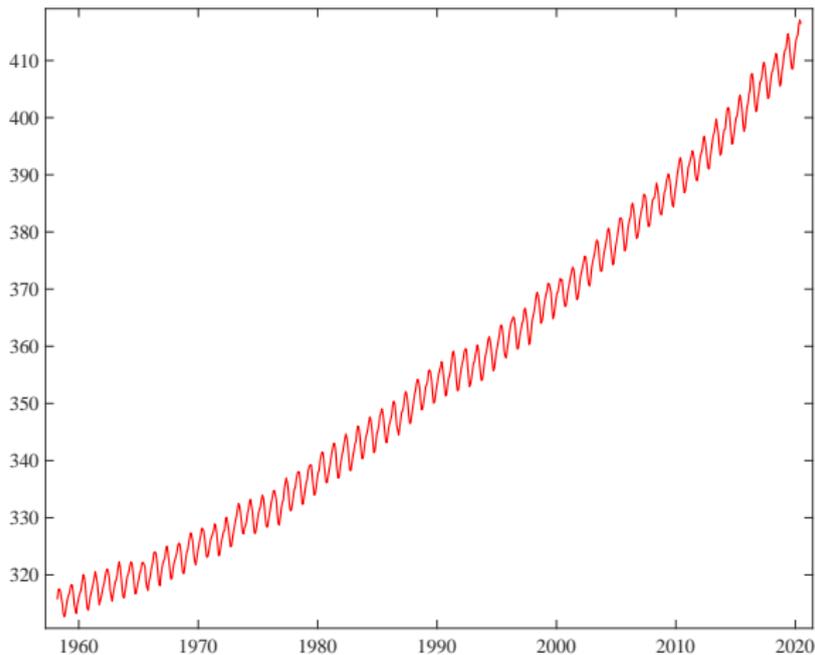
Climate Econometrics Seminar, April 20, 2021

Introduction

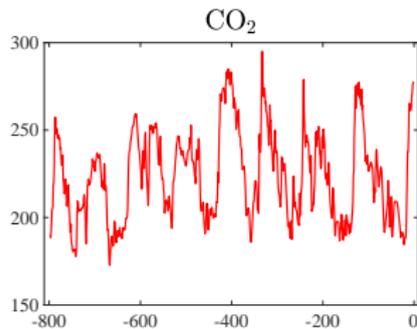
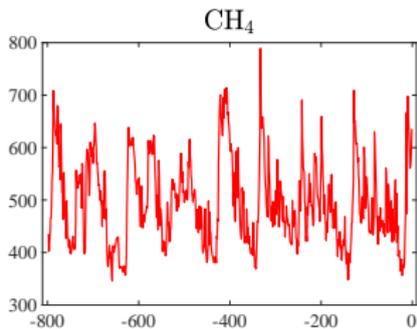
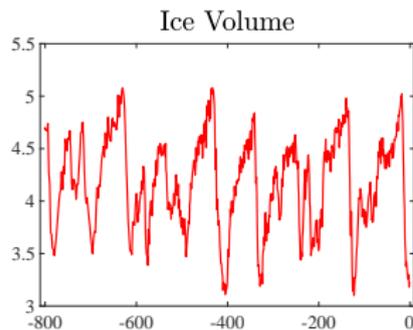
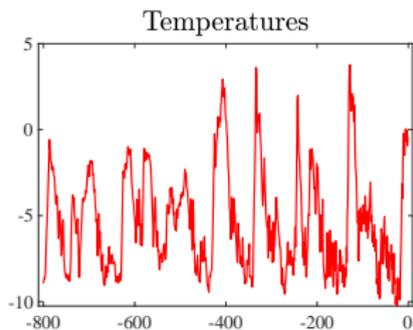
- We propose a novel model for climate time series characterized by persistent stationary cycles, the fractional Sinusoidal Waveform (fSW) process.
- The underlying idea is to allow the parameters that regulate the amplitude and phase to evolve according to fractional noise processes.
- The advantages over the Gegenbauer process are twofold: (i) the autocovariance function is available in closed form; (ii) deterministic cycles arise as a special case.
- When combined additively with other components, such as 'background continuum' red noise, we obtain a model suitable for climate time series with mixed spectra (spectral distribution function with jumps).
- Our illustrations deal with the change in amplitude and phase of the intra-annual component of carbon dioxide concentrations in Mauna Loa, and with the estimation and the quantification of the contribution of orbital cycles to the variability of paleoclimate time series.

Carbon dioxide concentrations (ppm) at Mauna Loa (Hawaii).

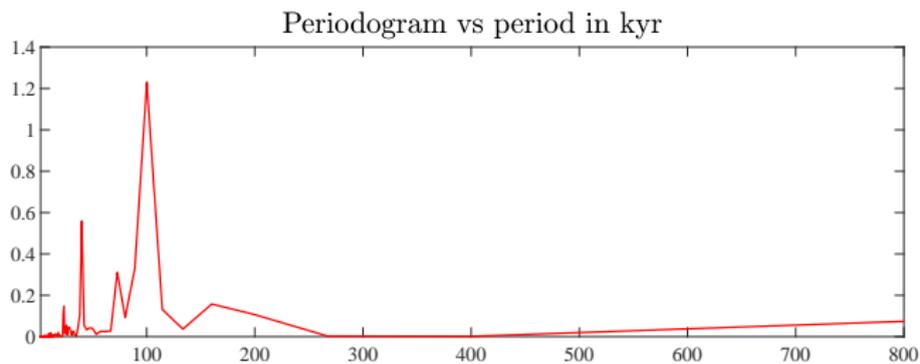
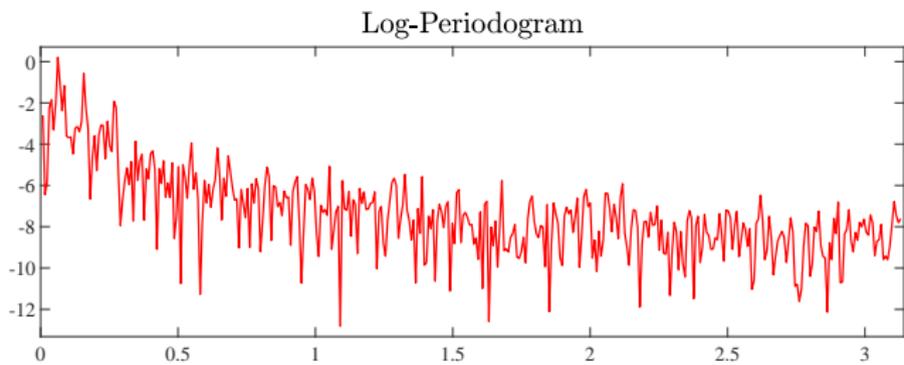
Issues: characterizing the interannual and intra-annual variability of CO₂ (varying amplitude and phase of seasonal cycle).



Four paleoclimate time series dealing with ice cores reconstructions of temperatures, methane (CH_4), carbon dioxide (CO_2) concentrations and ice volume.



Temperature series



Cycle models: a brief review

- The following **harmonic process** is a model for stationary **deterministic** cycles:

$$y_t = \alpha \cos(\lambda t) + \alpha^* \sin(\lambda t), \quad \alpha \sim N(0, \sigma_\alpha^2), \quad \alpha^* \sim N(0, \sigma_\alpha^2),$$

λ is a fixed frequency in $[0, \pi]$ and $E(\alpha\alpha^*) = 0$.

The process is characterized by ACF $\rho(k) = \cos(\lambda k)$, $k \in \mathbb{Z}$, and a purely discrete (or line, or narrow-band) spectrum.

- The following AR(2) process is a model for stationary **short memory stochastic** cycles:

$$y_t = 2\rho \cos(\lambda)y_{t-1} - \rho^2 y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } N(0, \sigma^2), \quad |\rho| < 1,$$

The spectral density is continuous and bounded, with a spectral peak around the frequency λ .

- If $\rho = 1$, the cycle is **nonstationary** and it is said to be **integrated of order 1** at the frequency λ . Introducing the lag operator L , $L^k y_t = y_{t-k}$, $k \in \mathbb{Z}$, the integrated cycle is written $(1 - 2 \cos(\lambda)L + L^2)y_t = \varepsilon_t$.

- The following **Gegenbauer process** is a model for stationary **long memory** cycles:

$$(1 - 2 \cos(\lambda)L + L^2)^d y_t = \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } N(0, \sigma^2),$$

where $d \in \mathbb{R}$ is the memory parameter.

- The process is stationary if $0 < \lambda < \pi$ and $d < 1/2$, or when $\lambda = 0, \pi$ and $d < 1/4$.
- The spectral density

$$f(\omega) = \frac{\sigma^2}{2\pi} \left| 2 \sin \left(\frac{\omega - \lambda}{2} \right) \right|^{-2d} \left| 2 \sin \left(\frac{\omega + \lambda}{2} \right) \right|^{-2d}, \omega \in [-\pi, \pi].$$

has poles at $\pm\lambda$.

- One limitation of the Gegenbauer process and its k -factor generalizations is the lack of a closed form expression for the autocovariance function. This prevents exact maximum likelihood estimation and optimal signal extraction. Several computationally efficient algorithms have been proposed.
- k -factor Gegenbauer ARMA models are multiplicative. The estimation of components of variability is not easy.

The Fractional Sinusoidal Waveform Process

The fractional Sinusoidal Waveform process with memory parameter d , frequency λ and disturbance variance σ_η^2 , is defined as

$$\begin{aligned}y_t &= \alpha_t \cos(\lambda t) + \alpha_t^* \sin(\lambda t), & t \in \mathbb{N}, \\ \alpha_t &= (1 - L)^{-d} \eta_t, & \eta_t \sim \text{i.i.d. } N(0, \sigma_\eta^2), \\ \alpha_t^* &= (1 - L)^{-d} \eta_t^*, & \eta_t^* \sim \text{i.i.d. } N(0, \sigma_\eta^2),\end{aligned} \tag{1}$$

where η_t and η_t^* are mutually independent.

- The process is stationary when $d < 1/2$, in which case $E(y_t) = 0$, $\text{Var}(y_t) \equiv \sigma_\alpha^2$, with $\sigma_\alpha^2 = \sigma_\eta^2 \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}$.
- Generalization of the harmonic process, similar in spirit to Hannan's integrated model of trigonometric seasonality (Hannan, 1964).

Autocovariance and autocorrelation function

- Let $\gamma(k) = E(y_t y_{t-k})$ and $\gamma_\alpha(k) = E(\alpha_t \alpha_{t-k})$, $k \in \mathbb{Z}$. Then,

$$\gamma(k) = \gamma_\alpha(k) \cos(\lambda k), \quad (2)$$

where

$$\gamma_\alpha(k) = \sigma_\eta^2 \frac{\Gamma(1-2d)\Gamma(d+k)}{\Gamma(1+k-d)\Gamma(d)\Gamma(1-d)}.$$

- Let $\rho(k) = \gamma(k)/\gamma(0)$

$$\rho(k) = \frac{\Gamma(1-d)\Gamma(d+k)}{\Gamma(1-d+k)\Gamma(d)} \cos(\lambda k), \quad k \in \mathbb{Z}.$$

- As $k \rightarrow \infty$, $\rho(k) \sim ck^{2d-1} \cos(k\lambda)$, $c > 0$,
- For a proof of these formulae, see section 2.1 of the paper.

Spectral density

- The spectral density function is

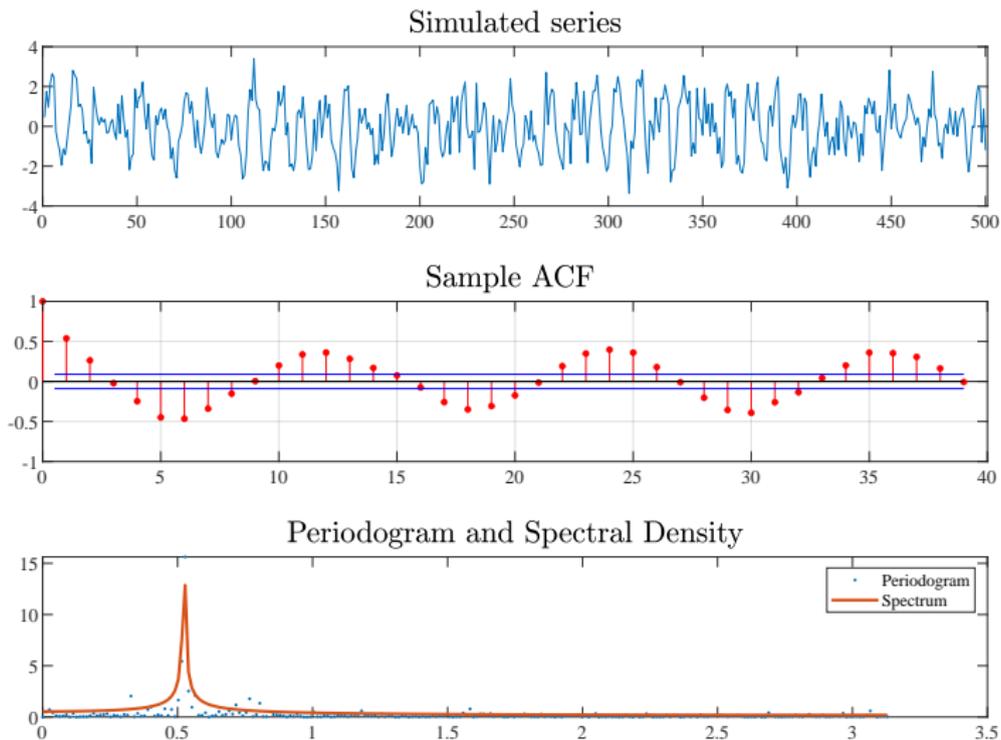
$$f(\omega) = \frac{\sigma_\eta^2}{4\pi} \left\{ \left| 2 \sin \left(\frac{\omega - \lambda}{2} \right) \right|^{-2d} + \left| 2 \sin \left(\frac{\omega + \lambda}{2} \right) \right|^{-2d} \right\}, \omega \in [-\pi, \pi]. \quad (3)$$

- $f(\omega) \sim \frac{\sigma_\eta^2}{4\pi} |\omega - \lambda|^{-2d}$ as $\omega \rightarrow \lambda$.
- For a proof see section 2.2 of the paper. The derivation is based on the equivalent representation of the fSW process:

$$y_t = \frac{1}{\sqrt{2}}(w_t + \bar{w}_t),$$

where w_t is the complex fractional noise process $(1 - e^{-i\lambda}L)^d w_t = \zeta_t$, $\zeta_t = 2^{-1/2}(\kappa_t + i\kappa_t^*)$, with κ_t and κ_t^* i.i.d. Gaussian variables, and $(1 - e^{i\lambda}L)^d \bar{w}_t = \bar{\zeta}_t$, $\bar{\zeta}_t = 2^{-1/2}(\kappa_t - i\kappa_t^*)$.

Simulated time series of length $n = 500$, generated by an fSW process with $d = 0.40$, $\lambda = \pi/6$, $\sigma_{\eta}^2 = 1$.



Deterministic Cycles

- Recall $\sigma_\alpha^2 = \sigma_\eta^2 \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}$. Assume that for $c > 0$ we can write $\sigma_\eta^2 = 4c(1-2d)/\pi$. Then,

$$\lim_{d \rightarrow \frac{1}{2}^-} \sigma_\alpha^2 = c > 0$$

(as $\lim_{x \rightarrow 0^+} x\Gamma(x) = 1, x = 1 - 2d$).

- The autocovariance function

$$\gamma(k) = \sigma_\alpha^2 \frac{\Gamma(1-d)\Gamma(d+k)}{\Gamma(1+k-d)\Gamma(d)} \cos(\lambda k), \quad \sigma_\alpha^2 > 0,$$

for $d \rightarrow (1/2)^-$, tends to $\gamma(k) = \sigma_\alpha^2 \cos(\lambda k)$.

- Also, writing the spectral density function (3) in terms of σ_α^2 ,

$$f(\omega) = \frac{\sigma_\alpha^2 \Gamma^2(1-d)}{4\pi \Gamma(1-2d)} \left\{ \left| 2 \sin \left(\frac{\omega - \lambda}{2} \right) \right|^{-2d} + \left| 2 \sin \left(\frac{\omega + \lambda}{2} \right) \right|^{-2d} \right\},$$

$\lim_{d \rightarrow \frac{1}{2}^-} f(\omega) = \frac{\sigma_\alpha^2}{2} \delta(\omega - \lambda)$, where $\delta(\cdot)$ is Dirac's Delta function.

Contemporaneous aggregation and cyclical long memory

We iterate the argument by Granger (1980) to illustrate a possible generating mechanism for cyclical long memory.

- Let us consider a panel of N short memory cyclical processes

$$y_{it} = \alpha_{it} \cos(\lambda t) + \alpha_{it}^* \sin(\lambda t), \quad i = 1, 2, \dots, N,$$

where $\lambda \in [0, \pi]$ is fixed and $\alpha_{it} = \phi_i \alpha_{i,t-1} + \eta_{it}$, $\eta_{it} \sim \text{i.i.d. } N(0, \sigma_\eta^2)$,
 $\alpha_{it}^* = \phi_i \alpha_{i,t-1}^* + \eta_{it}^*$, $\eta_{it}^* \sim \text{i.i.d. } N(0, \sigma_\eta^2)$.

- The individual coefficients ϕ_i^2 are an i.i.d. sample from a Beta distribution, $\phi_i^2 \sim \text{Beta}(p, q)$
- We are interested in characterizing the behaviour of cross-sectional aggregate

$$y_{Nt} = \frac{1}{\sqrt{N}} \sum_{i=1}^N y_{it}.$$

By the law of large numbers

$$\lim_{N \rightarrow \infty} \gamma_N(k) \approx \sigma_\eta^2 \frac{\Gamma(q-1)}{B(p, q)} k^{1-q} \cos(\lambda k), \quad k \rightarrow \infty.$$

- Hence, y_{Nt} displays long memory with parameter $d = 1 - q/2$ in the covariance sense at the frequency λ .

Multiple periodicities, red noise and exogenous effects

A more general specification for periodic time series, including the possibility of a *red noise* component, multiple periodicities and strongly exogenous regression effects, is the following:

$$\begin{aligned}y_t &= \beta' \mathbf{x}_t + u_t, \quad t = 1, 2, \dots, n, \\u_t &= \alpha_{0t} + \sum_{j=1}^M (\alpha_{jt} \cos(\lambda_j t) + \alpha_{jt}^* \sin(\lambda_j t)) + \varepsilon_t,\end{aligned}\tag{4}$$

- The process α_{0t} is the *red noise* component of the series, capturing low frequency movements, formulated as

$$\alpha_{0t} = \phi \alpha_{0,t-1} + \eta_{0t}, \quad \eta_{0t} \sim \text{i.i.d. } N(0, \sigma_{\eta 0}^2).\tag{5}$$

- The j -th component cycle, $\alpha_{jt} \cos(\lambda_j t) + \alpha_{jt}^* \sin(\lambda_j t)$, is a Gaussian fSW process with memory parameter d_j , frequency parameter λ_j , disturbance error variance $\sigma_{\eta j}^2$, and unconditional variance $\sigma_{\alpha j}^2$.
- Finally ε_t is a Gaussian white noise process, $\varepsilon_t \sim \text{i.i.d. } N(0, \sigma_{\varepsilon}^2)$.
- The components are assumed to be mutually independent.

Maximum likelihood estimation

Let $\mathbf{y} = [y_1, y_2, \dots, y_n]'$, $\mathbf{X} = [\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_n]'$, $\mathbf{u} = [u_1, u_2, \dots, u_n]'$.

The linear model (4) implies the representation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, with $\mathbf{u} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Gamma}_n)$, where $\boldsymbol{\Gamma}_n$ is the Toeplitz matrix

$$\boldsymbol{\Gamma}_n = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \ddots & \ddots & \gamma(n-2) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \gamma(n-2) & \ddots & \ddots & \ddots & \gamma(1) \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(1) & \gamma(0) \end{bmatrix},$$

with elements

$$\gamma(k) = \phi^k \frac{\sigma_{\eta 0}^2}{1 - \phi^2} + \sum_{j=1}^M \sigma_{\alpha_j}^2 \frac{\Gamma(1 - d_j) \Gamma(d_j + k)}{\Gamma(1 - d_j + k) \Gamma(d_j)} \cos(\lambda_j k) + I(k=0) \sigma_{\varepsilon}^2,$$

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}'_0, \boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_m, \sigma_\varepsilon^2)'$, $\boldsymbol{\theta}_0 = (\phi, \sigma_{\eta_0}^2)'$, $\boldsymbol{\theta}_j = (d_j, \lambda_j, \sigma_{\eta_j}^2)'$, $j = 1, \dots, M$.
 The Gaussian log-likelihood is

$$\ell(\boldsymbol{\theta}, \boldsymbol{\beta}; \mathbf{y}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Gamma}_n| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Gamma}_n^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

- Maximising with respect to $\boldsymbol{\beta}$ yields the generalized least squares estimator

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}' \boldsymbol{\Gamma}_n^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Gamma}_n^{-1} \mathbf{y}.$$

- Replacing into the previous expression gives the profile likelihood

$$\ell_\beta(\boldsymbol{\theta}; \mathbf{y}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Gamma}_n| - \frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{y},$$

with $\mathbf{P} = \boldsymbol{\Gamma}_n^{-1} - \boldsymbol{\Gamma}_n^{-1} \mathbf{X} (\mathbf{X}' \boldsymbol{\Gamma}_n^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Gamma}_n^{-1}$.

- The marginal likelihood of the $n - p$ linear transformation of the data, $\mathbf{y}^* = \mathbf{A}' \mathbf{y}$, which is invariant to $\boldsymbol{\beta}$, where \mathbf{A} is an $(n - k) \times n$ matrix spanning the null space of \mathbf{X} , i.e., it is chosen so that $\mathbf{A}(\mathbf{A}' \mathbf{A})^{-1} \mathbf{A}' = \mathbf{I} - \mathbf{X}(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'$,

$$\ell(\boldsymbol{\theta}; \mathbf{y}^*) = -\frac{n-p}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Gamma}_n| - \frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{y} - \frac{1}{2} |\mathbf{X}' \boldsymbol{\Gamma}_n^{-1} \mathbf{X}|,$$

- The value $\tilde{\boldsymbol{\theta}}$ maximizing is known in the literature as a REML estimator. See Verbyla (1990). See also Doornik and Ooms (2003) for the discussion of the merits of various likelihoods for the estimation of ARFIMA models.

- The evaluation of the log-likelihood entails the inversion and the determinant of a possibly large dimensional covariance matrix.
- The Durbin–Levinson algorithm performs the factorization:

$$\mathbf{\Gamma}_n^{-1} = \mathbf{\Phi}'_n \mathbf{D}_n \mathbf{\Phi}_n,$$

where $\mathbf{D}_n = \text{diag}(v_0^{-1}, v_1^{-1}, \dots, v_{n-1}^{-1})$, $v_k = \text{Var}(u_t | u_{t-1}, \dots, u_{t-k})$, i.e., the conditional variance of $u_t = y_t - \beta' \mathbf{x}_t$, given k past values, and

$$\mathbf{\Phi}_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_{11} & 1 & 0 & \cdots & 0 \\ -\phi_{22} & -\phi_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ -\phi_{n-1,n-1} & -\phi_{n-1,n-2} & -\phi_{n-1,n-3} & \cdots & 1 \end{bmatrix}. \quad (6)$$

- In an appendix we deal with the maximum likelihood estimation in the case when the fSW process collapses to a deterministic cycle with discrete spectrum.
- In the paper we also deal with Whittle likelihood estimation and compare the two estimators by a Monte Carlo experiment.

Signal Extraction and Prediction

Consider the problem of estimating the signal $\mathbf{s} = (s_1, \dots, s_t, \dots, s_{n+h})'$, $h \geq 0$, where, e.g., $s_t = \alpha_{kt} \cos(\lambda_k t) + \alpha_{kt}^* \sin(\lambda_k t)$, for a given k , conditional on $\tilde{\boldsymbol{\theta}}$ and (\mathbf{y}, \mathbf{X}) .

The optimal estimator is

$$\hat{\mathbf{s}} = \boldsymbol{\Gamma}_{s,y} \tilde{\boldsymbol{\Phi}}_n' \tilde{\mathbf{D}}_n \tilde{\boldsymbol{\Phi}}_n (\mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\beta}})$$

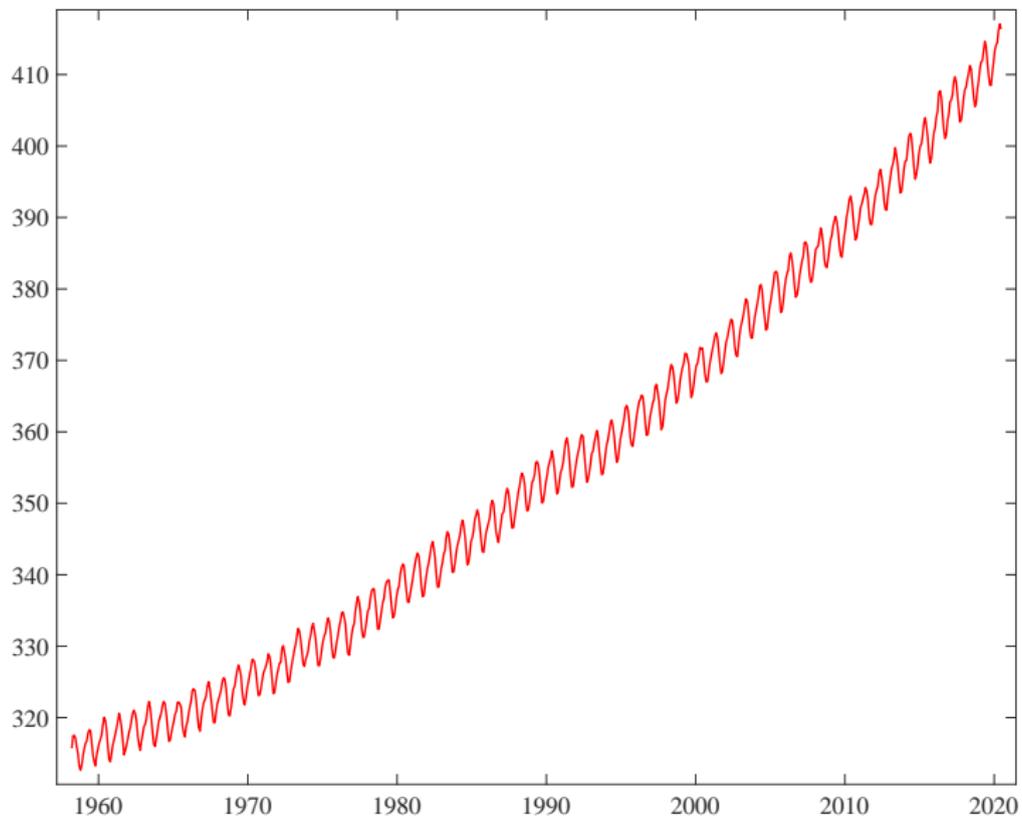
where $\boldsymbol{\Gamma}_{s,y} = \text{Cov}(\mathbf{s}, \mathbf{y})$ has (i, j) element

$$\tilde{\sigma}_{\eta k}^2 \frac{\Gamma(1 - 2\tilde{d}_k) \Gamma(\tilde{d}_k + |i - j|)}{\Gamma(1 - \tilde{d}_k + |i - j|) \Gamma(\tilde{d}_k) \Gamma(1 - \tilde{d}_k)} \cos(\tilde{\lambda}_k (|i - j|)).$$

The minimum mean square estimator of the other components and the prediction of y_t follows straightforwardly.

Mauna Loa Atmospheric CO₂ Data

- The series consists of monthly atmospheric carbon dioxide measurements collected at the summit of Mauna Loa mountain (Hawaii), dealing with concentrations in parts per million (ppm), over the period January 1958 - June 2020 .
- It is very relevant for climate change discussion, being the longest instrumental record available of atmospheric CO₂; it is also a testbed for the class of *k*-factor generalized Gegenbauer processes, which have been fitted to the second differences of the series. See Wood et al (1998) and McElroy and Holan (2012, 2016).
- The series displays important inter-annual and intra-annual movements.
- Seasonality is indeed prominent, and the changes in the amplitude and phase of the annual cycle have been the subject of a rich debate.
- The seasonal cycle, which peaks in May and has a trough in October, is driven by the metabolic activity of terrestrial plants and soils: the process of carbon uptake and release of the land biosphere is such that CO₂ concentrations increase in winter, when plant respiration dominates, and decreases in summer, when the photosynthesis uptake dominates.



- A significant amplitude increase was documented already by Bacastow (1985), while Keeling et al (1996) also detected a change in the phase, implying an advance of the seasonal cycle of about 7 days.
- These references attributed the changes to global warming and longer growing seasons.
- Singular spectrum analysis of the Mauna Loa time series (Dettinger, 1998) provided further support for the changes in the seasonal cycle.
- Kaufmann (2007) identifies statistically significant variation in the anomalies pertaining to the monthly concentrations of April and October.
- Amplitude trends and phase changes reflect changes in the global carbon cycle and its response to climate change; thus, their attribution is an important matter of investigation, see Forkel et al (2016), Bastos et al (2019), Wang et al (2020), for some recent contribution and discussion.

- The interannual variation of CO₂ concentrations has been related to the El Niño Southern Oscillation (ENSO) phenomenon already by Bacastow (1976), who used the Southern Oscillation Index (SOI) as a measure of ENSO.
- ENSO originates in the tropical Pacific Ocean, but is one of the main drivers of interannual global climate variability.
- The correlation has been confirmed by alternative methods (Dettinger and Ghil, 2000) and measurements (Chatterjee et al, 2017).
- As highlighted by Zeng et al (2005), the strong correlation between the interannual variation and the SOI index is quite remarkable, considering the chain of causal links that relates the two phenomena.
- Volcanic eruptions also contribute to interannual variation of CO₂.
- Hendry and Pretis (2013) conclude, however, that natural factors are not sufficient to explain the changes in CO₂ concentrations. They identify significant anthropogenic contributions using an autoregressive distributed lag model selected by a general to specific modelling approach.

Denoting y_t the Mauna Loa CO₂ monthly time series, we fit the model:

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \alpha_{0t} + \sum_{j=1}^5 (\alpha_{jt} \cos(\lambda_j t) + \alpha_{jt}^* \sin(\lambda_j t)) + \alpha_{6t} \cos(\pi t).$$

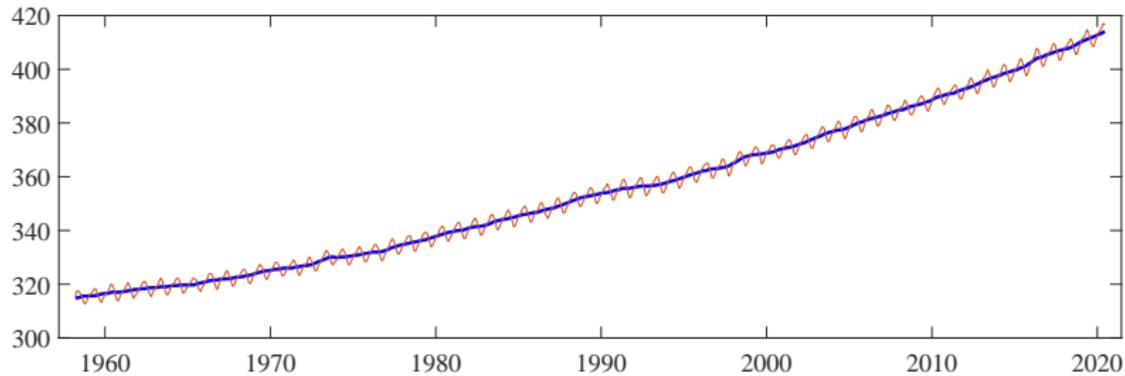
- α_{0t} is the AR(1) red noise process (5), modelling the low-frequency variability.
- The MLE of ϕ is 0.9998, so that the red noise process actually turns *brown*, and $\tilde{\sigma}_{\eta 0}^2 = 0.0212$.
- The seasonal component is modelled by the six fSW cycles defined at the seasonal frequencies $\lambda_j = \frac{\pi}{6}j, j = 1, \dots, 6$. The MLEs of the parameters are

j	λ_j	\tilde{d}_j	$\tilde{\sigma}_{\eta j}^2$	$\tilde{\sigma}_{\alpha j}^2$
1	$\pi/6$	0.4995	0.0127	4.0165
2	$\pi/3$	0.4980	0.0041	0.3297
3	$\pi/2$	0.5000	0.0000	0.0032
4	$2\pi/3$	0.5000	0.0000	0.0027
5	$5\pi/6$	0.0569	0.0241	0.0243
6	π	0.5000	0.0000	0.0000

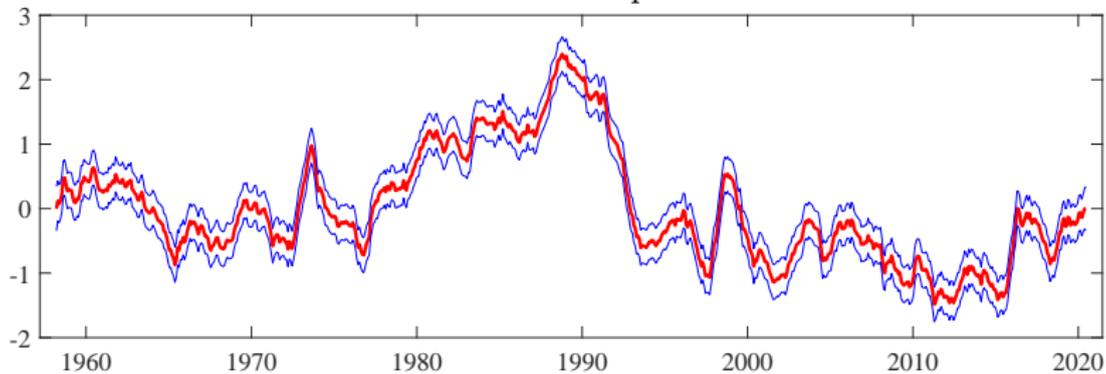
- The broad band components are identified at the fundamental frequency (1 cycle per year), and at $\pi/3$, the semiannual frequency. The seasonal cycle at $5\pi/6$ (2.4 cycles per year) is also estimated as a broadband component, although it shows little persistence and variability.

- The model provides a good fit: the estimated prediction error variance is $\tilde{v}_{n-1} = 0.0913$, which represent 6.04% of the variance of Δy_t . The standardized residuals show no significant autocorrelation, the sample autocorrelations at lags 1 and 12 resulting 0.0052 and 0.0199, respectively, and those of their squares being equal to 0.0120 and 0.0775.
- The estimated components provide useful insight.
- As a measure of interannual variability we consider $\Delta\alpha_{0t}$. To investigate its relation with the ENSO phenomenon, we compare its estimates with the SOI index.
- It could be argued that the cross-correlation is inflated by the fact that the estimates of $\Delta\alpha_{0t}$ are conditional on the full available sample. To address this point we also present the cross-correlation function of the SOI and the real time estimates of the interannual variation in CO_2 .

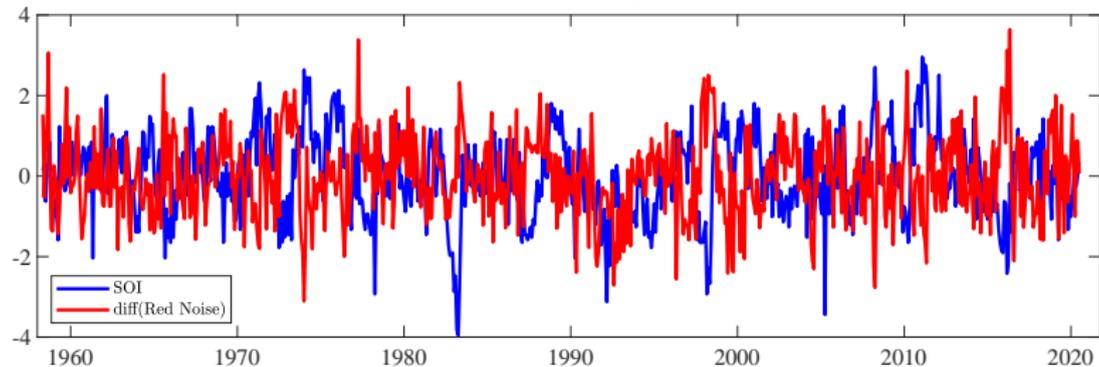
Series and Trend



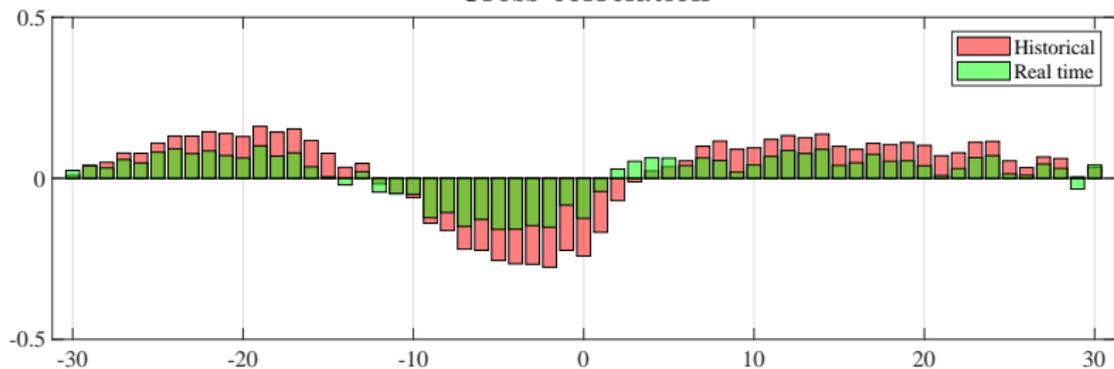
Red Noise Component



Soi and 1st diff. Red noise



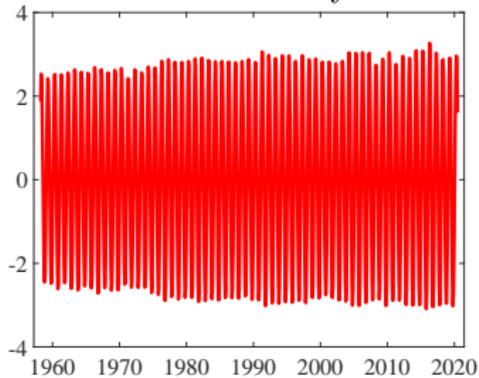
Cross-correlation



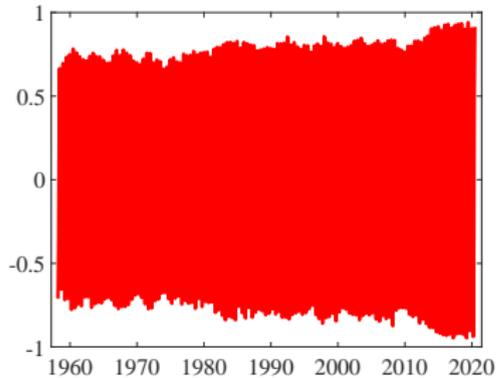
Seasonal changes in Mauna Loa CO₂ concentrations

- The overall change in amplitude during the period considered amounts to 0.5 ppm for the fundamental cycle and to 0.2 ppm for the semiannual one.
- The changes imply that CO₂ release in winter months has relatively increased and uptake in summer months has declined.
- The trend in the amplitude shows a deceleration after the 1980s, consistent with Wang et al (2020), but it is subject to a more rapid increase in the recent years.
- Our estimates imply a phase advance of 20 days for the fundamental cycle. The semiannual cycle is also subject to a phase advance, but of only 6 days.
- The combined effects imply that the May annual peak becomes more prominent with time, while September emerges as the seasonal trough.

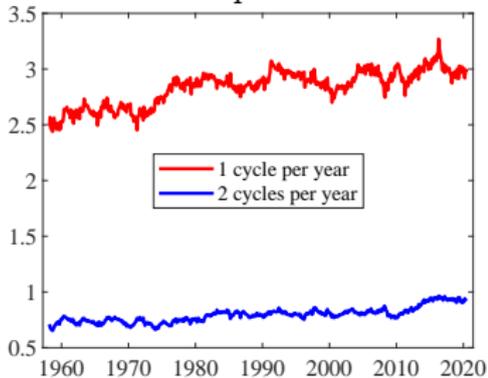
Fundamental cycle



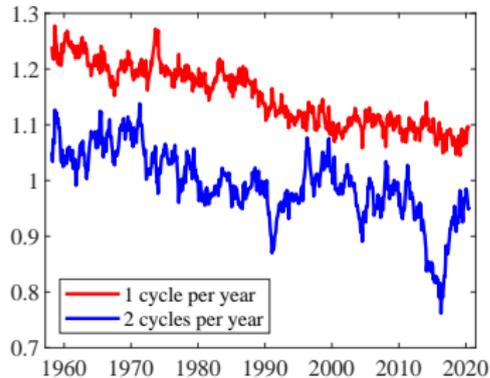
First harmonic



Amplitude



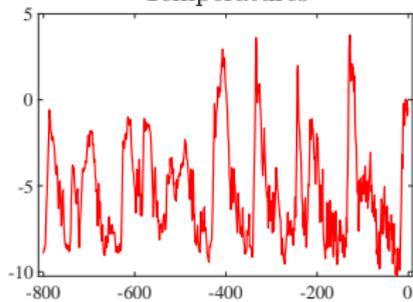
Phase



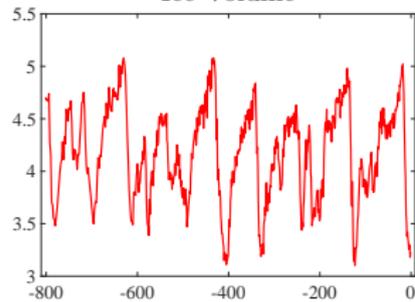
Cycles and Variability in Paleoclimate Data

- We consider four paleoclimate time series dealing with ice cores reconstructions of temperatures, methane (CH_4), carbon dioxide (CO_2) concentrations and ice volume. The series were obtained by European Project for Ice Core in Antarctica (EPICA), and Lisiecki and Raimo (2005).
- The series display substantial recurrent co-movements referred to as *glacial cycles*.
- According to the paleoclimatic literature and the Milankovitch theory (Hays et al., 1976), glacial cycles are attributed to changes in Earth's orbital geometry over time, which affects incoming solar radiation. The three main sources of variation are:
 - **Eccentricity** of the Earth orbit round the Sun, due to gravitational effects of other planets in the solar system, which varies deterministically with a periodicity of about 100 kyr.
 - **Obliquity** or tilt of the Earth's axis of rotation, which varies with a period of 41 kyr.
 - **Precession** of the equinoxes. This component has periodicities of about 23 and 19 kyr.

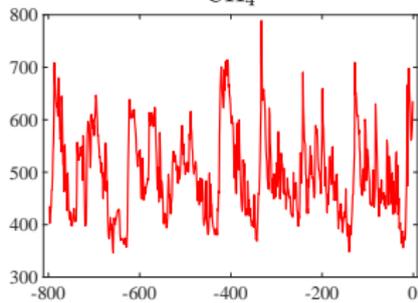
Temperatures



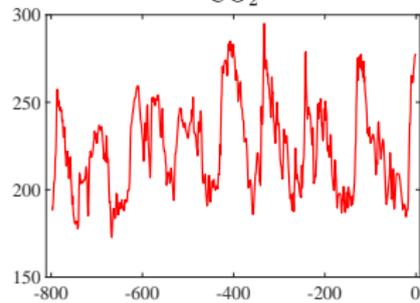
Ice Volume



CH₄



CO₂



- The four series have been recently investigated by Davidson et al (2016) and Castle and Hendry (2020).
- Larger cointegrated VAR systems are adopted by Kaufmann and Juselius (2013, 2016) and Kaufmann and Pretis (2020, 2021).
- A key issue deals with the role of the above three orbital components in explaining the variability in the climate and trace gases.
- It is evident from the plots that glacial cycles in the late Pleistocene occur at intervals of 100 kyr; however, eccentricity accounts only for minor fluctuations in the amplitude of the insolation signal. This is known as the '100 kyr' problem.
- Hence, the detailed mechanism by which small changes in insolation become amplified to drive major climatic changes remains unclear.
- A nonlinear response of the climate system to relatively weak eccentricity variations has been considered. See Paillard (2001).

- An important strand of the literature has aimed at quantifying the contribution of the orbital components via a decomposition of the total variability of paleoclimate series in the frequency domain.
- This entails estimation of a possibly mixed spectrum or pseudo-spectrum, for which purpose the use of Thomson (1982) multitaper spectral method is prominent.
- Given the deterministic nature of orbital forcing, the ability to distinguish narrow-band (discrete spectra) components from background broad-band components is quintessential to the identification of the components of paleoclimate variability. See Mitchell (1976), Mann and Lees (1996), Wunsch (2003), Meyers et al (2008), Ditlevsen et al (2020).
- Our contribution to this literature is to provide an alternative parametric approach, based on the additive model (4), to the quantification of the components of paleoclimate variability via the spectrum.

- We adopt the logarithmic transformation of the variables used by Davidson et al (2016), namely $y_t = \log(x_t + 16)$ for temperatures, $y_t = -\log(8 - x_t)$ for ice, and $y_t = \log(x_t/100)$ for both CH₄ and CO₂, where x_t denotes the original measurement.
- The model features a constant, a red noise component, and four fSW cycles associated to the orbital frequencies $\lambda_1 = \pi/50$ (eccentricity), $\lambda_2 = 2\pi/41$ (obliquity), $\lambda_3 = 2\pi/23$ and $\lambda_4 = 2\pi/19$ (precession):

$$y_t = \beta_0 + \alpha_{0t} + \sum_{j=1}^4 (\alpha_{jt} \cos(\lambda_j t) + \alpha_{jt}^* \cos(\lambda_j t)). \quad (7)$$

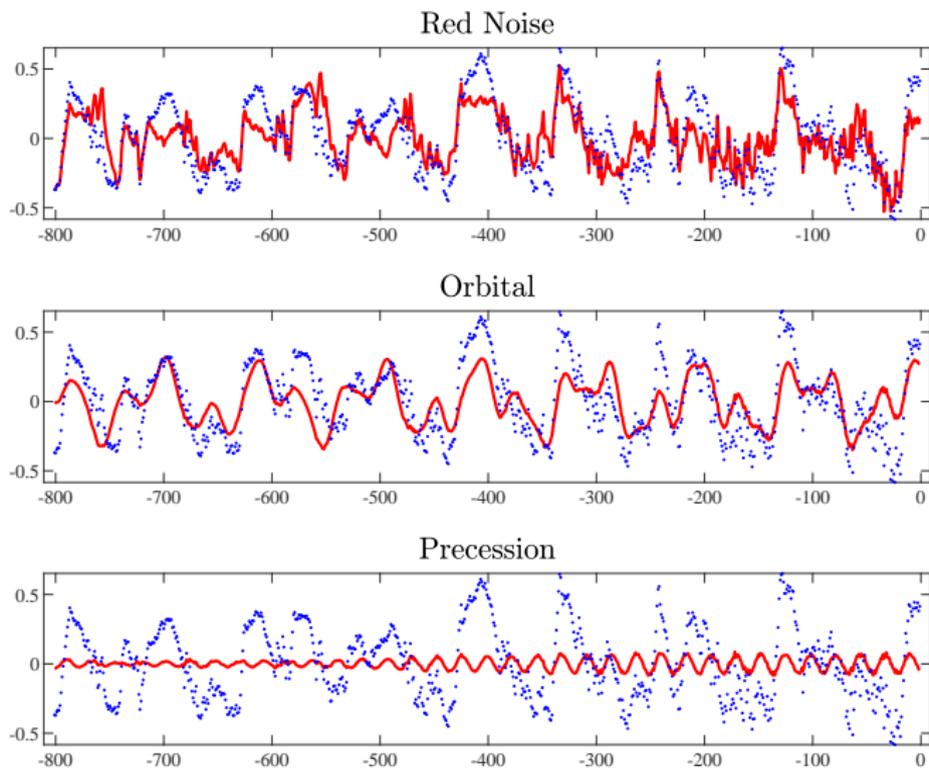
Parameter estimates

	Temp.	Ice v.	CH ₄	CO ₂
$\tilde{\beta}_0$	2.3317	-1.3331	1.5886	0.7967
$\tilde{\phi}$	0.9240	0.9638	0.8705	0.9669
$\tilde{\sigma}_{\eta 0}^2$	0.0057	0.0005	0.0039	0.0006
\tilde{d}_1	0.5000	0.5000	0.5000	0.5000
$\tilde{\sigma}_{\eta 1}^2$	0.0000	0.0000	0.0000	0.0000
\tilde{d}_2	0.5000	0.5000	0.5000	0.5000
$\tilde{\sigma}_{\eta 2}^2$	0.0000	0.0000	0.0000	0.0000
\tilde{d}_3	0.4695	0.4832	0.4467	0.5000
$\tilde{\sigma}_{\eta 3}^2$	0.0006	0.0001	0.0002	0.0000
\tilde{d}_4	0.3753	0.5000	0.5000	0.5000
$\tilde{\sigma}_{\eta 4}^2$	0.0000	0.0000	0.0000	0.0000

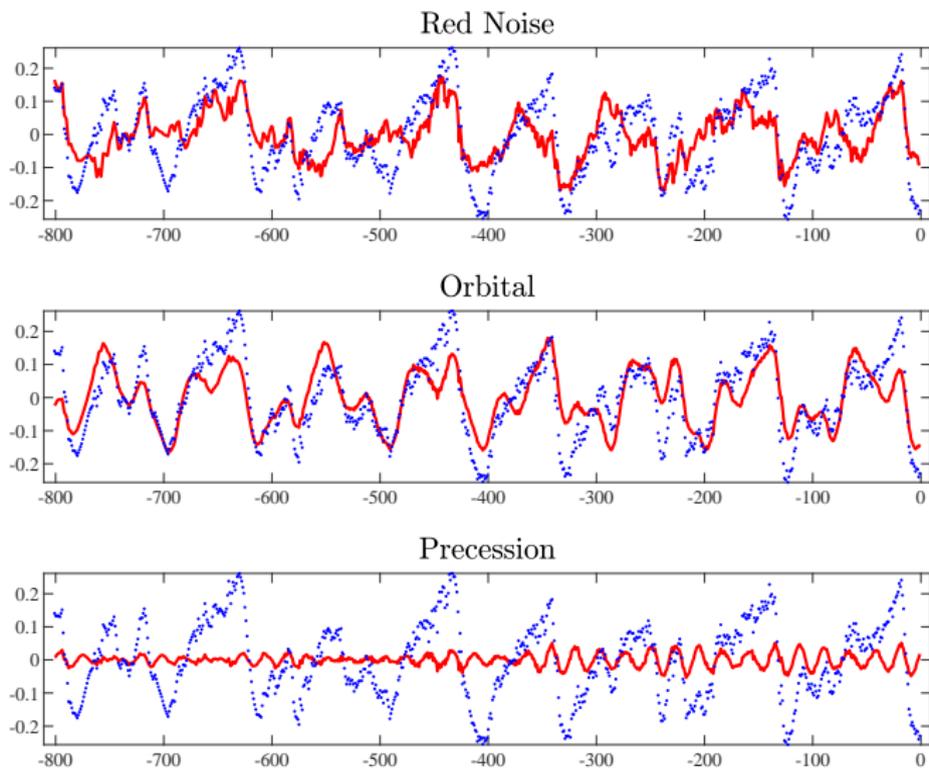
Diagnostics

	Temp.	Ice v.	CH ₄	CO ₂
p.e.v.	0.0065	0.0006	0.0042	0.0006
R^2	0.9075	0.9583	0.8388	0.9559
$r(1)$	0.0382	0.0352	0.0323	0.3493
$r(2 1)$	-0.0537	0.0849	-0.0081	-0.1128
$r_{sq}(1)$	0.1591	0.1811	0.1640	0.2918
Skewness	-0.2548	-0.3451	0.8904	0.5768
Kurtosis	4.3064	4.3159	6.6686	4.8094

Temperatures: MMSEs of the components



Ice volume MMSEs of the components



- Our measurement model allows the quantification of the contribution of the components to the overall variability.
- The following variance decomposition holds: $\text{Var}(y_t) = \sum_{j=0}^4 \sigma_{\alpha j}^2$.
- When the model is estimated by maximum likelihood, the sample counterpart is the additive decomposition $\tilde{v}_0 = \sum_{j=0}^4 \tilde{\sigma}_{\alpha j}$.
- We present the contribution of the red noise component ($\tilde{\sigma}_{\alpha 0}^2$), that of eccentricity ($\tilde{\sigma}_{\alpha 1}^2$), obliquity ($\tilde{\sigma}_{\alpha 2}^2$) and precession ($\tilde{\sigma}_{\alpha 3}^2 + \tilde{\sigma}_{\alpha 4}^2$) to the total variability for the four series (\tilde{v}_0).
- The lower and upper confidence limits are obtained by generating 500 bootstrap samples according to the parametric bootstrap method presented in the Appendix.

Decomposition of total variance: contribution of the components of variability.

		Temperatures	
	Point est.	Lower Conf. Limit	Upper Conf. Limit
Red Noise	0.0392	0.0275	0.0536
Eccentricity	0.0170	0.0001	0.0529
Obliquity	0.0111	0.0000	0.0370
Precession	0.0032	0.0001	0.0080
Total Variance	0.0705	0.0407	0.1084

		Ice Volume	
	Point est.	Lower Conf. Limit	Upper Conf. Limit
Red Noise	0.0064	0.0040	0.0096
Eccentricity	0.0047	0.0001	0.0162
Obliquity	0.0024	0.0001	0.0070
Precession	0.0011	0.0002	0.0030
Total Variance	0.0147	0.0073	0.0271

		CH ₄	
	Point est.	Lower Conf. Limit	Upper Conf. Limit
Red Noise	0.0160	0.0119	0.0207
Eccentricity	0.0064	0.0001	0.0196
Obliquity	0.0027	0.0000	0.0089
Precession	0.0010	0.0000	0.0029
Total Variance	0.0261	0.0159	0.0416

		CO ₂	
	Point est.	Lower Conf. Limit	Upper Conf. Limit
Red Noise	0.0093	0.0056	0.0142
Eccentricity	0.0038	0.0001	0.0116
Obliquity	0.0006	0.0000	0.0019
Precession	0.0001	0.0000	0.0004
Total Variance	0.0138	0.0072	0.0219

- The red noise component accounts for 55.60%, 43.54%, 61.30% and 67.39% of the variance, respectively for temperatures, ice, CH₄ and CO₂.
- These estimates are somewhat above the corresponding estimates for temperature proxy records from the Vostok ice cores obtained by Meyers et al (2008). To explain the difference we notice that we estimate the spectrum of eccentricity and obliquity as a line spectrum, whereas they integrate the multitaper spectral density estimate across a neighbourhood of frequencies around the 100 and 41 kyr frequencies.
- Our estimates are more in line with Wunsch (2003), who questions the notion that paleoclimate variability is predominantly associated with the frequency bands attributed to solar insolation.
- However, this is not quite the end of the story, as the red noise could may also result from the climate response to orbital forcing: the path from the orbital signal to climate and trace gases proxy records goes through several steps, outlined in Meyers et al (2008), and possibly nonlinear and persistent transfer functions can be responsible for the dominance of the stochastic red noise component. A model of low frequency variation featuring hysteresis, such as $\alpha_{0t} = \phi\alpha_{0,t-1} + \alpha_{1t} \cos(\lambda_1 t) + \alpha_{1t}^* \sin(\lambda_1 t) + \eta_{0t}$, cannot be ruled out.

Conclusions

- The paper has proposed a novel time series model for persistent cycles, the fractional sinusoidal waveform (fSW) process.
- The model features stationary cyclical long memory and collapses to a line spectrum component when the parameters are on the boundary of their admissible range.
- Hence, it is suitable to analyze time series with mixed spectra.
- Likelihood inference and optimal signal extraction were discussed with reference to an additive model combining a broadband continuum component of variability with a number of fSW process.
- The application to carbon dioxide concentration and paleoclimate time series have illustrated that the model can address some important questions raised with respect to the quantification of feature changes and the contribution to the total variability of deterministic forcing due to solar radiation.