New CO_2 Concentration Predictions and Spectral Estimation Applied to the Vostok Ice Core

David Bonacci and Bernard Lacaze

***Abstract*— The Vostok ice core provides measurements of the CO_2 concentration during the last 414×10^3 yr (year). Estimations of power spectra show peaks, with the strongest one corresponding to a time period of around 100×10^3 yr. In this paper, a new reconstruction method from irregular sampling is used, allowing more accurate estimation of spectral peaks. This method intrinsically decomposes the analyzed signal as a sum of sines, providing amplitudes but also phases measurements of periodic tendencies (due to the nature of the studied phenomena). This decomposition can be conducted with noisy and inaccurate measurements of the sampling instants and the concentrations. The widely used Vostok data was chosen as an example but the method could also be applied to data from other places (e.g. dome C, Antarctica) or to study other phenomena as nitrogen dioxide NO_2, methane CH_4, oxygen isotope ^{18}O (closely linked to temperature), deuterium ^2H or dust concentrations.**

***Index Terms*— Ice, Signal sampling, Signal reconstruction, Spectral analysis, Prediction methods.**

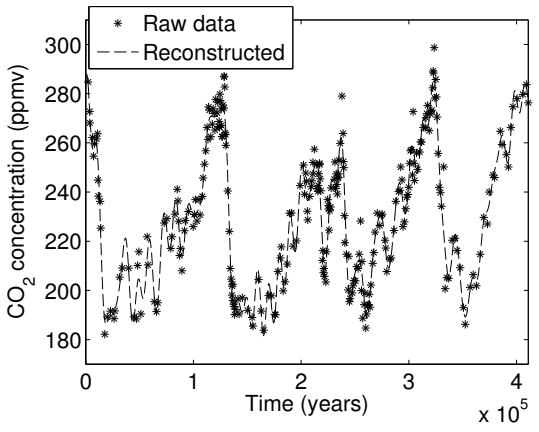

Fig. 1. Reconstructed and raw Vostok ice core data.

I. Introduction

THE Vostok ice core provides information for the estimation of CO_2 time variations during the last 414×10^3 yr (years). The ice core was a thickness larger than three kilometers above a deep lake in Antarctica [1], [2]. The observations are summarized by 283 measurements of the CO_2 concentration at irregular time instants estimated from the depth of the samples (in this paper, only the first 282 values are used due to the necessity of an even number of measurements inherent to the proposed method). Note that the raw data and sampling instants are noisy. Moreover, a periodic component is visible with a period of around 100×10^3 yr. Classical spectrum estimation methods confirm this spectral peak around frequency 10^{-5} yr^{-1} with a secondary spectral peak around 2.5×10^{-5} yr^{-1} and other smaller ones [3]. It is worth noting that the stationarity of the studied phenomena is a necessary condition to make use of these spectral estimation techniques (and that the obtained results encourage the validity of this assumption). The used data set is depicted in Fig. 1 together with a reconstruction using the method of Section II.

The literature on spectral estimation for irregularly sampled data is rich. Recent publications include papers by Selva [4] and Eldar [5] for multiband signals, and by Aldroubi for compressive sampling, data smoothing and interpolation by cubic splines [6], [7]. Spectral estimation (spectrum being the Fourier transform of the autocorrelation function associated with CO_2 concentration time variations) has been studied by authors like Scargle [8]. An interesting classification of spectral analysis methods for nonuniform data based on the sampling pattern, signal model and type of spectrum can be found in [3]. A more recent reference summarizes state-of-the-art methods linked to the study of ice core data [9]. Note that hyperspectral techniques [10] [11] could be interesting for this problem. However, they would require several data sets (one for each considered spectral band) as they make use of hyperspectral sensors. In this work (Vostok ice core) we only have one (irregularly sampled) data set (provided by analysis of air bubbles in ice carrots), making the use of hyperspectral-based techniques less appropriate.

This paper investigates a reconstruction method based on theoretical results coming from previous papers [12], [13]. These previous works were dealing with signal reconstruction with a low number of irregular samples. Here, the increased number of data allows us to get a more accurate spectral estimator and to reconstruct the spectral peaks of the Vostok ice core as separated components with phase information, allowing signal prediction outside the observation interval. This original method allows us to reconstruct not only the irregularly sampled process, but also the process filtered with ideal bandpass filters whose bandwidth is a multiple of $1/2L$ (where $2L$ is the number of samples). This reconstruction is possible provided the average sampling rate satisfies the Nyquist condition. As a consequence, it is possible to track spectral peaks by defining these bandpass filters. Moreover, when L is large enough, the filtering operation provides a sinusoid whose amplitude and phase can be easily estimated. The amplitude provides the peak power whereas the phase defines the natural tendency of the studied phenomenon at the origin (present days) and in the future, under the stationarity

Corresponding author: David Bonacci, email: david.bonacci@tesa.prd.fr
Address: TESA Laboratory, 7 Boulevard de la Gare, 31500 Toulouse, France. Tel: +33 561 247381

hypothesis (not taking into account anthropogenic factors).

The paper is organized as follows. Section II introduces the reconstruction formulas used in this work. The application of these formulas to the analysis of the Vostok ice core is studied in Section III. The way of predicting sinusoidal components from the filtered CO_2 concentration signal versus time is also presented. Conclusions and future works are reported in Section IV.

II. Proposed spectral estimation

In this section, we show how time reconstruction and spectral estimation can be performed using irregular samples of a wide sense stationary random process. These reconstruction and spectral estimation are obtained by defining an ideal filterbank, having as many subbands as samples (see Appendix for mathematical details).

A. Time reconstruction

A sequence of measurements of a (zero-mean) stationary process $\mathbf{Z} = \{Z(t), t \in \mathbb{R}\}$ is considered. More precisely, $Z(t)$ is observed at irregular sampling instants

$$\mathbf{t} = \{t_k, k \in \{-L, ..., -1\} \cup \{1, ..., L\}\}.$$

The power spectral density $s_Z(f)$ of this stationary process $\mathbf{Z}$ (assumed to be band limited with support $[-a, a]$) is defined as [14]

$$\mathrm{E}\left[Z(t) Z^*(t-\tau)\right] = \int_{-a}^{a} e^{2i\pi f\tau} s_Z(f)\, df \tag{1}$$

where E[.] and the superscript * stand for mathematical expectation (or ensemble mean) and complex conjugate.

The bandpass of $\mathbf{Z}$ is divided into $2L$ frequency bands $\Delta_k, 1 \leq |k| \leq L$ defined as

$$\Delta_k = \begin{cases} [(k-1)a/L, ka/L[, k \geq 1 \\ ]ka/L, (k+1)a/L], k \leq -1 \end{cases}. \tag{2}$$

The intervals Δ_k define a partition of $[-a, a]$ in $2L$ intervals of length a/L. We denote as $\mathbf{Z}_k = \{Z_k(t), t \in \mathbb{R}\}$ the signal $Z(t)$ filtered by an ideal bandpass filter with frequential support Δ_k, i.e.,

$$Z_k(t) = Z(t) * h_k(t) \tag{3}$$

with

$$H_k(f) = \mathrm{TF}\left[h_k(t)\right] = I_{\Delta_k}(f) \tag{4}$$

where TF is the Fourier transform and I_{Δ_k} is the indicator function on the interval Δ_k. The signal $Z(t)$ can then be decomposed as

$$Z(t) = \sum_{1 \leq |k| \leq L} Z_k(t). \tag{5}$$

Using this decomposition, Section V (appendix) shows the following result based on a periodic nonuniform sampling (PNS) of order $2L$ [15], [16])

$$\begin{array}{c} \sum\limits_{1 \leq |k| \leq L} Z_k(t)\, e^{-i\pi\alpha_k(t-\theta)} = \\ \sum\limits_{n \in \mathbb{Z}} (-1)^n \,\mathrm{sinc}\left[\frac{\pi a}{L}(t-\theta) - \pi n\right] Z\left(\theta + \frac{nL}{a}\right) \end{array} \tag{6}$$

with $\alpha_k = \frac{a}{L}[2k - \mathrm{sgn}(k)]$, for any real number θ and where $\mathrm{sgn}(k) = k/|k|$, $\mathrm{sinc}(x) = \sin(x)/x$. Note that the measured quantities $Z(t_j)$ appear in the right-hand side of (6) for $n = 0, \theta = t_j$. If L/a is large enough and provided that t is in the neighborhood of the observation interval $[t_{-L}, t_L]$, it is possible to neglect the terms associated with $n \neq 0$ in the right-hand side of (6) (the sinc function converges to 0), leading to

$$\sum_{1 \leq |k| \leq L} Z_k(t) e^{-i2\pi\alpha_k(t-t_j)} = \mathrm{sinc}\left[\frac{\pi a}{L}(t-t_j)\right] Z(t_j) + r_j(t) \tag{7}$$

where the residue $r_j(t)$ is defined as

$$r_j(t) = \sum_{n \in \mathbb{Z}^*} (-1)^n \,\mathrm{sinc}\left[\frac{\pi a}{L}(t-\theta) - \pi n\right] Z\left(\theta + \frac{nL}{a}\right). \tag{8}$$

In order to solve (7), it is easier reparametrize the problem as a function of the variables $Z_k(t) e^{-i2\pi\alpha_k t}, 1 \leq |k| \leq L$. This reparameterization allows (7) to be rewritten in matrix form by varying the parameter j over the interval $-L, ..., -1, 1, ..., L$ (for a given value of t)

$$\boldsymbol{M}\boldsymbol{x}(\boldsymbol{t}) = \boldsymbol{b}(\boldsymbol{t}) + \boldsymbol{r}(\boldsymbol{t}) \tag{9}$$

with

$$\boldsymbol{M} = \begin{bmatrix} e^{i2\pi\alpha_{-L}t_{-L}} & \dots & e^{i2\pi\alpha_L t_{-L}} \\ \vdots & \ddots & \vdots \\ e^{i2\pi\alpha_{-L}t_L} & \dots & e^{i2\pi\alpha_L t_L} \end{bmatrix}$$

$$\boldsymbol{x}(\boldsymbol{t}) = \begin{bmatrix} \widehat{Z}_{-L}(t) e^{-i2\pi\alpha_{-L}t} \\ \vdots \\ \widehat{Z}_L(t) e^{-i2\pi\alpha_L t} \end{bmatrix}$$

$$\boldsymbol{b}(\boldsymbol{t}) = \begin{bmatrix} \mathrm{sinc}\left[\frac{\pi a}{L}(t-t_{-L})\right] Z(t_{-L}) \\ \vdots \\ \mathrm{sinc}\left[\frac{\pi a}{L}(t-t_L)\right] Z(t_L) \end{bmatrix}, \boldsymbol{r}(\boldsymbol{t}) = \begin{bmatrix} r_{-L}(t) \\ \vdots \\ r_L(t) \end{bmatrix}.$$

The least squares (LS) estimator of $\boldsymbol{x}(\boldsymbol{t})$ can then obtained by minimizing $\sum\limits_{1 \leq |j| \leq L} r_j^2(t)$ leading to

$$\widehat{\boldsymbol{x}}(\boldsymbol{t}) = \boldsymbol{M}^{-1}\boldsymbol{b}(\boldsymbol{t}). \tag{10}$$

It is important to note that the matrix $\boldsymbol{M}$ (that needs to be inverted) does not depend on t. As a consequence, this matrix only needs to be inverted once. After (10) has been solved, estimations of $Z_k(t)$ (denoted as $\widehat{Z}_k(t)$) and of $Z(t)$ (denoted as $\widehat{Z}(t)$ can be easily obtain using (5).

Remarks: the matrix $\boldsymbol{M}$ can be ill-conditioned (mainly for large values of L) depending on the irregularity of the time instants. For instance, Fig. 2 shows the eigenvalues of $\boldsymbol{M}$ for the Vostok ice core data. Many eigenvalues are very low so that the condition number (ratio between the larger and smaller eigenvalues) is very large, leading to numerical instabilities. To avoid numerical problems for the inversion of $\boldsymbol{M}$, its low rank approximation is performed before its inversion [17]. The rank p of $\boldsymbol{M}$ is estimated considering only eigenvalues having a value greater than one hundredth of the largest one. Note that small variations around the estimated rank $p = 93$ do not affect significantly the reconstruction of the signal $Z(t)$ nor its spectral estimation. Denote as

$$\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \in \mathbb{R}^{2L \times 2L} \tag{11}$$

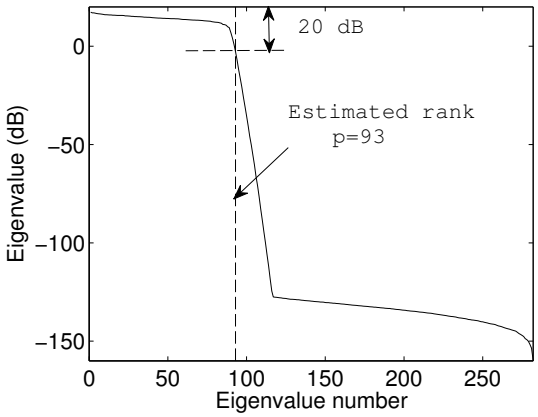

Fig. 2. Sorted eigenvalues of the linear system corresponding to the (non uniform) time samples

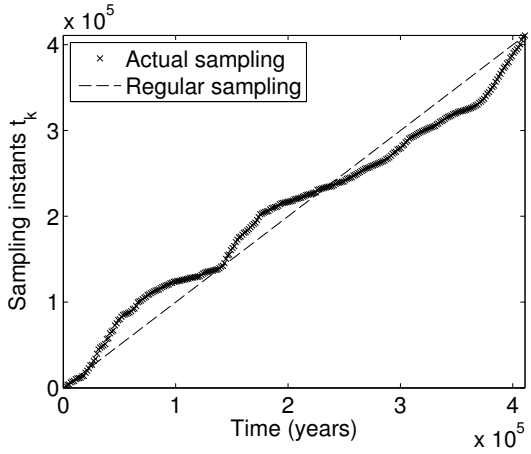

Fig. 3. Sampling instants repartition

the singular value decomposition of $\boldsymbol{M}$ and partition $\boldsymbol{U}, \boldsymbol{\Sigma}$ and $\boldsymbol{V}$ as follows

$$\boldsymbol{U} = [\boldsymbol{U_1 U_2}], \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma_2} \end{bmatrix}, \text{ and } \boldsymbol{V} = [\boldsymbol{V_1 V_2}] \quad (12)$$

where the diagonal matrix $\boldsymbol{\Sigma_1}$ is $p \times p$, $\boldsymbol{U_1}$ is $2L \times p$ and $\boldsymbol{V_1}$ is $2L \times p$. Then the matrix, obtained from the truncated singular value decomposition

$$\widehat{\boldsymbol{M}} = \boldsymbol{U_1 \Sigma_1 V_1}^T \quad (13)$$

is the matrix of rank p minimizing the Frobenius norm $\|\boldsymbol{M} - \widehat{\boldsymbol{M}}\|_F$. A robust estimator of the inverse of $\boldsymbol{M}$ is then defined as

$$\widehat{\boldsymbol{M}}^{-1} = \boldsymbol{V_1 \Sigma_1}^{-1} \boldsymbol{U_1}^T. \quad (14)$$

B. Spectral estimation

The power of $Z(t)$ in the frequency band Δ_k (defined in (2)) is

$$\mathbf{P}_{\Delta_k} = \mathrm{E}\left[|Z_k(t)|^2\right] = \int_{\Delta_k} s_Z(f)\, df \quad (15)$$

which can be estimated by standard estimators such as

$$\frac{1}{N} \sum_{n=1}^{N} \left| \widehat{Z}_k(nT_s) \right|^2$$

where T_s is a suitable sampling period.

A peak in $\mathbf{P}_{\Delta_k}$ at frequency f_0 corresponds to a large amount of signal power concentrated at f_0. In this case

$$s_Z(f_0) = \sigma^2 \delta(f_0)$$

where $\delta(f)$ is the Dirac delta function and σ^2 is the weight of the power peak. Actually, widened peaks (covering a set of several frequency bands containing a larger amount of power) are observed in the Vostok ice core data. This is a relative notion: when L is large, peaks have a power distributed on more intervals Δ_k. Moreover, as the width of the frequency bands decreases, $\widehat{Z}_k(t)$ appears as a monochromatic wave with a weaker amplitude modulation (this can be viewed as a non constant amplitude varying slowly). For a real process $\mathbf{Z}$, $s_Z(f)$ is even allowing the sums

$$Z_k(t) + Z_{-k}(t)$$

to be considered (instead of $Z_k(t)$ only). These sums will be assumed to be real sinusoids. If A is the amplitude of one of these sinusoids at frequency f_0, the power of the two spectral peaks at $\pm f_0$ is $A^2/4$ for each of them. Note that the phase of this sinusoid, defining the position of the sine component versus time, is an information that is not available with all spectral estimation methods [3], [18]. Here this extra information given by the proposed method is exploited to provide a prediction of the atmospheric CO_2 concentration in the future. In the following section, this estimation formula is exploited to provide an analysis of the Vostok ice core data. The phase of the spectral components (provided by the proposed method) is used to perform prediction of the signal behavior, which could for instance give clues in order to try to assess the predominance of the anthropic factor in the climate evolution.

III. Results about Vostok ice core data

A. Data description

The Vostok ice core data for carbon dioxide is composed of a sequence of measurements of atmospheric concentrations of CO_2 (in ppmv: parts per million by volume) acquired in the past. It is observed during the last 414×10^3 yr at irregular sampling instants t_k. The unit for time is the year (yr), where present corresponds to $t_k = 0$, past is for $t_k > 0$ and future for $t_k < 0$. A sequence of 283 measurements $Z(t_k)$ is provided, among which $2L = 282$ are used going from 0 to 411×10^3 yr. Fig. 3 shows the sequence of time instants t_k as a function of the time t to show how the curve is departing from the regular sampling case (that would correspond to a straight line).

In addition to CO_2 concentrations, t_k values can be also estimated thanks to carbon isotopes dating measurements. It is difficult to quantify the estimation error of t_k especially because air bubbles and surrounding ice can have different ages. The average sampling time for these measurements is 1.45×10^3 yr whereas the minimum and maximum time

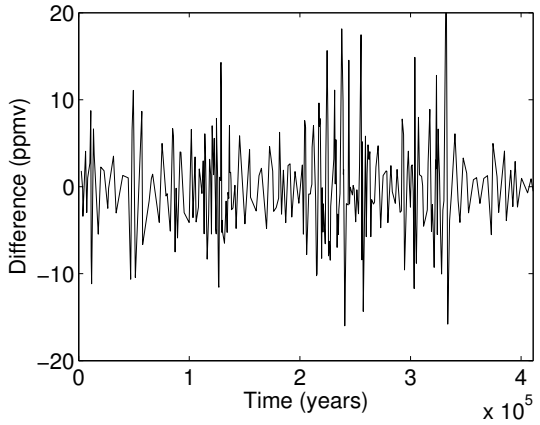

Fig. 4. Difference between reconstructed and raw data at sampling instants for $L = 141$ (noise)

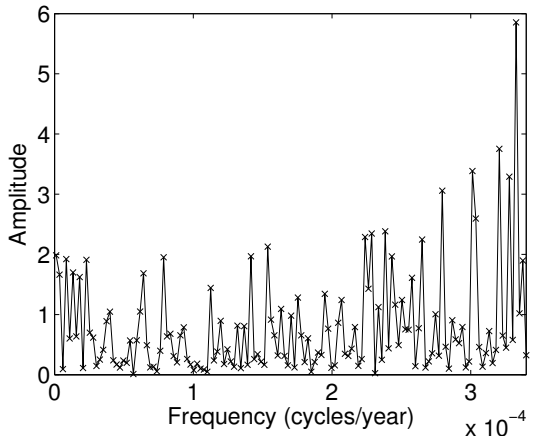

Fig. 5. Noise spectrum

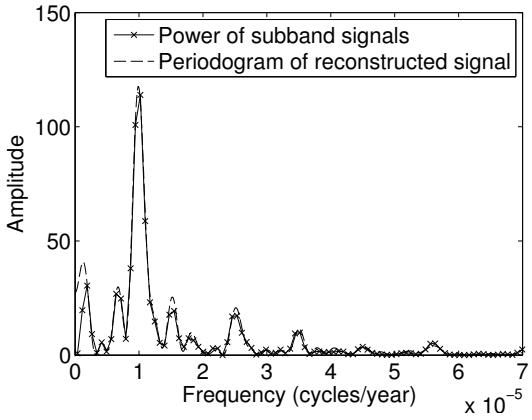

Fig. 6. Spectral power estimator obtained with $L = 141$ individual subband signals and classical periodogram of the reconstructed signal with 10,000 samples

intervals are 44 and 5996 yr (which corresponds to highly irregular sampling). This section considers the formulas of Section II to compute temporal and spectral properties of the data, and to obtain predictions about the behavior of the concentration in the future. Note that this analysis assumes that $Z(t)$ is a stationary random process, which should be validated by specialists.

Figs. 14 and 15 of [3] provide the power spectral density obtained using results from a review of state-of-the-art spectral estimators adapted to nonuniformly sampled data such as the Vostok data. Those figures show that the power spectrum density of the process $\mathbf{Z}$ (i.e., $s_Z(f)$) is very weak (negligible) for frequency values $f > 7\times10^{-5}$ yr^{-1} (this value is common to all spectral estimators). The frequency a defining the power spectral density support can then be chosen greater or equal to this value. Our results were obtained using $a = 10^{-4}$ yr^{-1} to make sure that there is no power in the spectrum after 7×10^{-5} yr^{-1}, in order to avoid any aliasing.

B. Signal reconstruction

The purpose of the signal reconstruction is to estimate $Z(t)$ (CO_2 concentration) at some given time t in the observation interval $[t_{-L}, t_L]$. Fig. 1 shows the reconstruction of $Z(t)$ in the interval of measurements $\left[0, 411\times10^3\right]$ yr whereas the difference between the actual data and its reconstruction is displayed in Fig. 4. The spectrum of this error, depicted in Fig. 5, shows that this difference may be considered as measurement noise (noisy flat spectrum). These results were obtained by solving (10) with a low rank approximation.

C. Spectral estimation

The estimated power spectrum of $Z(t)$ is displayed in Fig. 6. It is a function of the frequency f expressed in yr^{-1} (or cyles/yr). Each cross corresponds to the power in the considered frequency band versus its central frequency f (total of 282 crosses). The width of each frequency subband is $a/L = \frac{10^{-4}}{141} \approx 7.5\times10^{-7}$ yr^{-1}.

For comparison purposes, the results of [3] (obtained with the same Vostok data) are reproduced in this paper. These results were obtained for different signal models (nonparametric or parametric), different sampling pattern for the signal of interest (discrete or continuous). These methods were classified into 4 broad categories: methods based on least squares (LS), interpolation techniques or slotted resampling and methods based on continuous time models.

The Vostok ice core CO_2 concentration data studied in this paper is characterized by a nonparametric signal model, a sampling pattern involving arbitrary irregular sampling (not only missing data) and a discrete spectrum. As a consequence, only the first 3 categories of methods studied in [3] can be used. Moreover, due to large gaps between sampling instants, methods based on slotted resampling cannot be applied or simply fail to locate any peak in the spectrum [3]. For these reasons, a technical comparison is provided in this section only with methods based on LS and on interpolation techniques. Results with methods based on LS are presented in Fig. 7. These methods consist in estimating the spectral parameters at each considered frequency through an LS minimization involving known irregular samples (note that details on the Schuster periodogram can be found in [3], on the Lomb-Scargle in [8] and and on the iterative adaptive approach (IAA) in [19]).

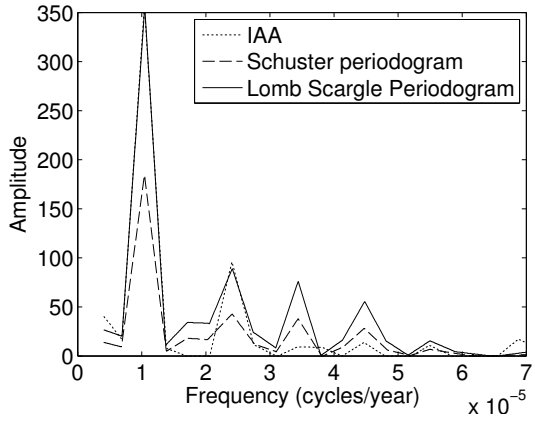

Fig. 7. Spectrum based on LS methods [3]

TABLE I
FUNDAMENTAL AND HARMONIC PEAKS

Frequency	Subbands	Power
$f_1 = 10.10^{-6}$ yr^{-1}	12-15	$\mathbf{P}_{f_1}^{\text{tot}} = 0.410$
$f_2 = 15.10^{-6}$ yr^{-1}	20-22	$\mathbf{P}_{f_2}^{\text{tot}} = 0.059$
$f_3 = 25.10^{-6}$ yr^{-1}	32-36	$\mathbf{P}_{f_3}^{\text{tot}} = 0.073$
$f_4 = 35.10^{-6}$ yr^{-1}	45-48	$\mathbf{P}_{f_4}^{\text{tot}} = 0.034$
$f_5 = 45.10^{-6}$ yr^{-1}	59-61	$\mathbf{P}_{f_5}^{\text{tot}} = 0.011$
$f_6 = 55.10^{-6}$ yr^{-1}	73-76	$\mathbf{P}_{f_6}^{\text{tot}} = 0.020$

Spectral estimators based on interpolation techniques are presented in Fig. 8, also reproduced from the results of [3]. These estimators were obtained by applying spectral estimation methods adapted to uniform sampling, i.e., Periodogram and Capon estimators applied to covariance (or autocorrelation) interpolation results with sinc kernels. Compared to previous spectral estimations (LS methods, methods based on interpolation techniques), the proposed method (power of filtered signal in each subband) produces a cleaner estimator. It highlights a widened main spectral peak around $f_1 = 10^{-5}$ yr^{-1}, in subbands $k = 12, 13, 14, 15$. By normalizing the total power to 1, and denoting as $\mathbf{P}_k$ the power in subband $|k|$ (positive and negative subbands), the following results can be obtained

$$\mathbf{P}_{12} = 0.050, \mathbf{P}_{13} = 0.13, \mathbf{P}_{14} = 0.15, \mathbf{P}_{15} = 0.077.$$

The total power of these peaks is around $\mathbf{P}_{f_1}^{\text{tot}} = 0.41$. Other widened spectral peaks linked to the previous one appear at frequencies f_2 to f_6 in sets of subbands and with total powers $\mathbf{P}_{f_j}^{\text{tot}}$ given in Table I. These 6 spectral peaks contain 60.7% of the total power. They are harmonics of a single fundamental signal component at frequency $f_0 = 5\times10^{-6}$ yr^{-1} (harmonics 2, 3, 5, 7, 9 and 11) with very low power ($\mathbf{P}_{f_0}^{\text{tot}} = 0.0073$). Two others spectral peaks appear at frequencies 2.5×10^{-6} yr^{-1} and 7.5×10^{-6} yr^{-1} with respective powers 0.0780 and 0.0680. Note that these spectral components were not detected in [3]. They correspond to a very slow trend of CO_2 concentration and have no impact on prediction in the next 50×10^3 yr.

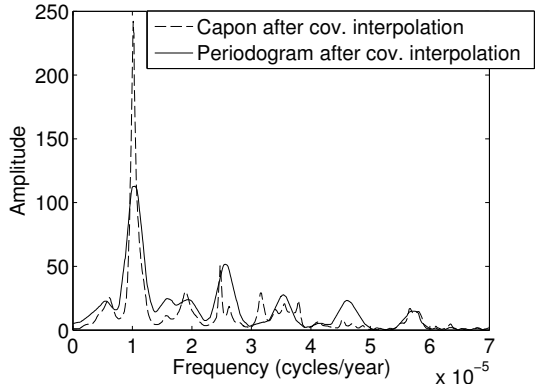

Fig. 8. Spectrum obtained by covariance interpolation and applying non-parametric uniform sampling spectral estimation methods (Periodogram and Capon) [3]

D. Signal prediction (extrapolation)

The principle of signal prediction is explained in Fig. 9. This figure shows the subband signals components $Z_k(t)$ in subbands 3, 9, 10, 12, 13 and 14 (added to subbands -3, -9, -10, -12, -13 and -14 to obtain real signals). They are quasi periodic functions (quasi pure sinusoids) because subbands have narrow bandwidth. Thanks to these components corresponding to the first 3 spectral peaks of Fig. 6 (low pass components), extrapolations are computed thanks to a Fourier Transform to predict future CO_2 concentrations (dashed left parts of curves for $t < 0$). In the same way as for reconstruction (consisting in summing all subband signals (5)), signal prediction can be obtained by summing these sine extrapolations. When the low frequency subbands are used for the prediction, a filtering operation is also performed. The proposed prediction scheme was validated to predict the CO_2 concentration during the last 50×10^3 yr using only data for $t > 50 \times 10^3$ yr. Results are plotted Fig. 10 where we can see that the predictions are very close to the true data.

In a second experiment, considering again all available samples, thanks to extrapolations of the 6 components of Fig. 9 (subbands 3, 9, 10, 12, 13 and 14) corresponding to low-frequency spectral peaks, a first filtered prediction of CO_2 concentration in the atmosphere can be done for the next 50×10^3 yr. As can be seen in Fig. 11, the CO_2 concentration should decrease during 50×10^3 yr (without taking into account human activities). The same prediction can also be done taking into account more components than just the main peak. Fig. 12 shows the same extrapolation using components of the 8 main visible spectral peaks or the 6 peaks of Table I, leading to the same interpretation (decrease of CO_2 concentration during the next 50×10^3 yr).

IV. CONCLUSION

The CO_2 concentration curve of the past 411×10^3 yr from the Vostok ice core record exhibits periodic components in the presence of noise. Though the spectral noise content seems richer in recent past, the stationarity hypothesis remains necessary as it is required for extracting Fourier components. The proposed estimation method can be used to predict

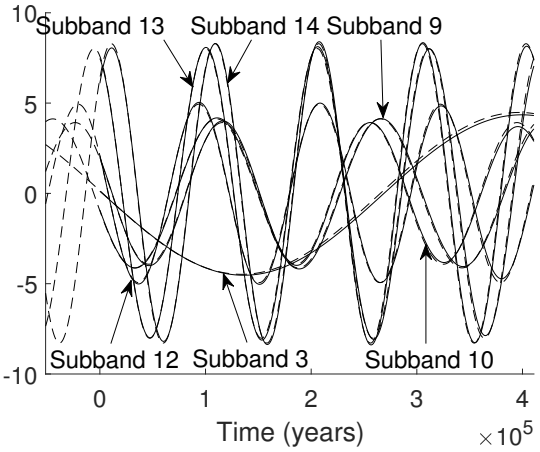

Fig. 9. Main peak subbands components (solid lines) and sinusoidal extrapolations (dashed)

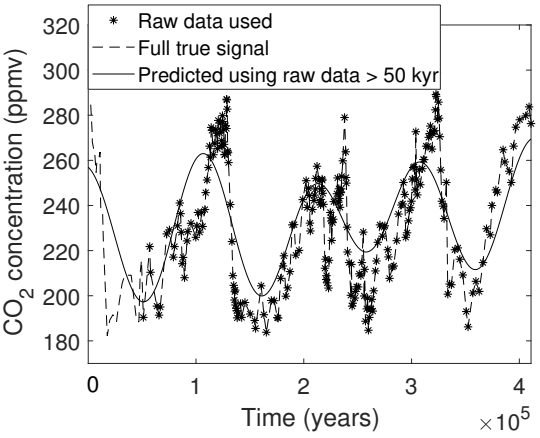

Fig. 10. True data and de-noised/predicted concentration during last 50×10^3 yr using subbands corresponding to the first 3 low-frequency peaks and only samples older than 50×10^3 yr

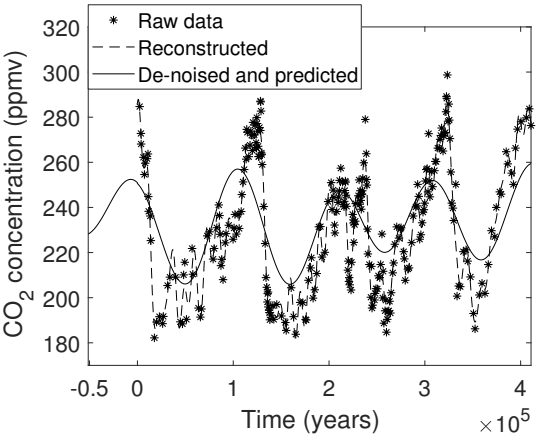

Fig. 11. Reconstructed CO_2 concentration using all subbands and de-noised/predicted using subbands corresponding to the first 3 low-frequency peaks

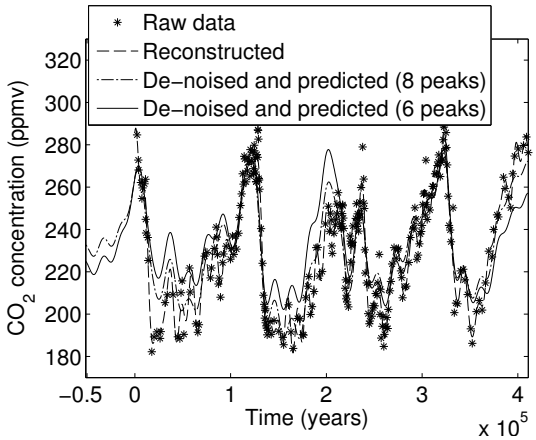

Fig. 12. Reconstructed CO_2 concentration using all subbands, de-noised/predicted using subbands corresponding to the 8 main spectral peaks visible in Fig. 6 and de-noised/predicted using subbands corresponding to the 6 peaks of Table I

hypothetical tendencies of the CO_2 concentration in the future (and then temperature) if human activities do not destroy this stationarity hypothesis.

Many procedures are available to extract the amplitudes of the spectral peaks, mainly based on estimations of correlation functions and power spectra [3], [18]. In this paper, a new reconstruction method giving access to individual filtered components was applied, giving the amplitude and phase of each spectral component (and of partial sums). The width of the spectral peaks was found to be not negligible. Among possible reasons, we can mention a deviation from the stationarity assumption, the finite length of measurements, some errors in the determination of the sampling instants and of the CO_2 concentrations.

V. Appendix

Let us assume that spectral density $s_U(f)$ of $U(t)$ has a bounded support

$$s_U(f) = 0, f \notin [\alpha, \beta].$$

The FSE (Fourier Series Expansion) of $e^{2i\pi ft}$ on $[\alpha,\beta]$ is

$$e^{2i\pi ft} = \sum_{n\in\mathbb{Z}} c_n(t)\, e^{2i\pi \frac{nf}{\beta-\alpha}} \tag{16}$$

with

$$c_n(t) = \frac{1}{\beta-\alpha}\int_\alpha^\beta e^{2i\pi ft} e^{-2i\pi\frac{nf}{\beta-\alpha}} df = e^{i\pi\left[(\beta+\alpha)\left(t-\frac{n}{\beta-\alpha}\right)\right]}\mathrm{sinc}\pi\left[(\beta-\alpha)t-n\right]. \tag{17}$$

The FSE in (16) also leads to

$$e^{2i\pi f(t+\theta)} = \sum_{n\in\mathbb{Z}} c_n(t)\, e^{2i\pi f\left[\theta+\frac{n}{(\beta-\alpha)}\right]}$$
$$e^{2i\pi ft} = \sum_{n\in\mathbb{Z}} c_n(t-\theta)\, e^{2i\pi f\left[\theta+\frac{n}{(\beta-\alpha)}\right]}$$

implying the sampling formula

$$U(t) = \sum_{n\in\mathbb{Z}} c_n(t-\theta)\, U\left(\theta+\frac{n}{\beta-\alpha}\right). \tag{18}$$

When $U(t)$ is replaced by $Z_k(t)$ with

$$\alpha = (k-1)\,a/L, \beta = ka/L, k > 0$$
$$\alpha = ka/L, \beta = (k+1)\,a/L, k < 0$$

then $\beta - \alpha = a/L$ and the following set of formulas is obtained from (17) and (18)

$$Z_k(t) = \sum_{n\in\mathbb{Z}} c_{nk}\,(t-\theta)\,Z_k\left(\theta + \tfrac{nL}{a}\right)$$
$$e^{-i\pi a(2k-1)\frac{t-\theta}{L}} Z_k(t) = \sum_{n\in\mathbb{Z}} (-1)^n \,\mathrm{sinc}\pi\left[\tfrac{a}{L}(t-\theta) - n\right]$$
$$Z_k\left(\theta + \tfrac{nL}{a}\right), k \geq 1$$
$$e^{-i\pi a(2k+1)\frac{t-\theta}{L}} Z_k(t) = \sum_{n\in\mathbb{Z}} (-1)^n \,\mathrm{sinc}\pi\left[\tfrac{a}{L}(t-\theta) - n\right]$$
$$Z_k\left(\theta + \tfrac{nL}{a}\right), k \leq -1.$$

Using (5), summation over k of the previous equations leads to

$$\sum_{1\leq|k|\leq L} e^{-i\pi a(2k-\mathrm{sgn}k)\frac{t-\theta}{L}} Z_k(t) = \sum_{n\in\mathbb{Z}} (-1)^n \,\mathrm{sinc}\pi\left[\tfrac{a}{L}(t-\theta) - n\right] Z\left(\theta + \tfrac{nL}{a}\right) \tag{19}$$

where $\mathrm{sgn}(k) = 1$ when $k \geq 1$ and -1 for $k \leq -1$, resulting in the expression in (6).

REFERENCES

[1] J. M. Barnola, D. Raynaud, Y. S. Korotkevich, and C. Lorius, "Vostok ice core provides 160,000-years record of atmospheric CO2," *Nature*, vol. 329, pp. 408–414, 1987.

[2] J. R. Petit, J. Jouzel, D. Raynaud, N. I. Barkov, J. M. Barnola, I. Basile, M. Bender, J. Chappellaz, M. Davis, G. Delaygue, M. Delmotte, V. M. Kotlyakov, M. Legrand, V. Y. Lipenkov, C. Lorius, L. Pépin, C. Ritz, E. Saltzman, and M. Stievenard, "Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antartica," *Nature*, vol. 399, pp. 429–436, 1999.

[3] P. Babu and P. Stoica, "Spectral analysis of nonuniformly sampled data - a review," *Digital Signal Processing, Elsevier*, vol. 20, pp. 359–378, 2010.

[4] J. Selva, "Regularized Sampling of Multiband Signals," *Signal Processing, IEEE Transactions on*, vol. 58, no. 11, pp. 5624–5638, Nov 2010.

[5] M. Mishali and Y. Eldar, "From Theory to Practice: Sub-Nyquist Sampling of Sparse Wideband Analog Signals," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 375–391, April 2010.

[6] A. Aldroubi and H. Wang, "An Adaptive and Information Theoretic Method for Compressed Sampling," *CAMSAP, 2009 3rd IEEE International Workshop on*, pp. 193–196, Dec 2009.

[7] A. Aldroubi and H. Feichtinger, "Complete iterative reconstruction algorithms for irregularly sampled data in spline-like spaces," *ICASSP., IEEE International Conference on*, vol. 3, pp. 1857–1860, Apr 1997.

[8] J. Scargle, "Statistical Aspects of Spectral Analysis on Unevenly Spaced Data," *The Astronomical Journal*, vol. 263, pp. 835–853, 1982.

[9] J. Jouzel, "A brief history of ice core science over the last 50 yr," *Climate of the Past*, vol. 9, pp. 2525–2547, 2013.

[10] A. Marinoni, A. Plaza, and P. Gamba, "Harmonic Mixture Modeling for Efficient Nonlinear Hyperspectral Unmixing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, pp. 4247–4256, 2016.

[11] R. Heylen and P. Scheunders, "A Multilinear Mixing Model for Nonlinear Spectral Unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, pp. 240–251, 2016.

[12] D. Bonacci and B. Lacaze, "Lowpass/bandpass signal reconstruction and digital filtering from nonuniform samples," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, Brisbane, Australia, April 2015, pp. 3626–3630.

[13] ——, "A new approach to spectral estimation from irregular sampling," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, Lisbon, Portugal, September 2014, pp. 476–480.

[14] A. Papoulis and S. U. Pillai, *Nonuniform sampling, theory and practice*. McGraw Hill, 2002.

[15] J. Yen, "On Nonuniform Sampling of Bandwidth-Limited Signals," *IRE Transactions on Circuit Theory*, vol. 3, pp. 251–257, 1956.

[16] Y.-P. Lin and P. P. Vaidyanathan, "Periodically nonuniform sampling of bandpass signals," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 45, pp. 340–351, 1998.

[17] I. Markovsky, "Structured low-rank approximation and its applications," *Automatica, Elsevier*, vol. 44, p. 891909, 2008.

[18] P. Babu, "Spectral analysis of nonuniformly sampled data and applications," Ph.D. dissertation, Uppsala University, Sweden, 2012.

[19] P. Stoica, J. Li, and H. He, "Spectral analysis of nonuniformly sampled data: A new approach versus the periodogram," *IEEE Transactions on Signal Processing*, vol. 57, p. 843858, 2011.