

Chapter

Kalman Filter Harmonic Bank for Vostok Ice Core Data Analysis and Climate Predictions

Migdat Hodzic and Ivan Kennedy

Abstract

The Vostok ice core data cover 420,000 years indicating the natural regularity of Earth's surface temperature and climate. Here, we consider four major cycles of similar duration, ranging from 86,000 to 128,000 years, comprising 15% of periods for the warming interglacials compared to some 85% of cooling periods. Globally, we are near the peak of a rapid warming period. We perform a detailed frequency analysis of temperature and CO₂ cycles, as a primary stage in building a logical Climate Prediction Engine (CPE), illustrated with specific harmonics. This analysis can be repeated for all harmonics and various cycle combinations. Our time correlation estimates the CO₂ time lag for temperature at 400–2300 years, depending on the cycle, longer on average than previously concluded. We also perform Fast-Fourier transform analysis, identifying a full harmonic spectrum for each cycle, plus an energy analysis to identify each harmonic amplitude – to achieve further prediction analysis using a Kalman filter harmonic bank. For Vostok data we can use combinations of different cycles compared to the most recent for learning and then the current ongoing cycle for testing. Assuming causal time regularity, more cycles can be employed in training, hence reducing the prediction error for the next cycle. This results in prediction of climate data with both naturally occurring as well as human forced CO₂ values. We perform this detailed time and frequency analysis as a basis for improving the quality of our climate prediction methodologies, with particular attention to testing alternative hypotheses of the possible causes of climate change. These include the effect on albedo of suspended dust and increasing water vapor with temperature in initiating interglacial warming, the effect of temperature and pH values of surface water on ambient level of CO₂ in the atmosphere and finding a larger latent heat capacity in the atmosphere required to sustain its circulatory motions, leading to friction and turbulent release of heat in boundary layer. All these potentials can be examined in an effective CPE.

Keywords: Vostok data, time and frequency analysis, Kalman filter harmonic bank, climate prediction engine, machine learning

1. Introduction

Extensive climatic data for the past four ice ages and earlier, the period in which *Homo sapiens* evolved, is available in various scientific reports analyzing ice cores, commencing from the mid-1950s. There are various sites on Antarctica and

Greenland where intensive ice core drilling has occurred since 1956, with several countries supporting more than a score of different drilling projects. Currently, intensive ice core drilling is being conducted in other areas as well, so an even larger data set is anticipated. In previous papers [1–4] the history and limitations of ice core drilling are described in detail. Our purpose in this paper is to employ Vostok ice core data for time and frequency related analyses to set a basis for improving prediction of climate variations. The data sets include derivations of relative temperature, carbon dioxide (CO_2), methane (CH_4), oxygen, dust and solar variation (insolation), during the past 420,000 years. Because isotopic fractionation of oxygen-18 and deuterium in snowfall is temperature dependent and a strong spatial correlation exists between mean annual temperature and mean isotopic ratios it is possible to derive ice-core climate records. References [5, 6] presented the first record for full glacial–interglacial period from an ice core drilled in the Russian Vostok station in Antarctica. However, it is important to establish criteria for controlling the quality of this new science that we claim could eventually lead to a Climate Prediction Engine (CPE), based on verified causes.

A 420,000 year record was constructed from the [5] study on a 3 km deep core of ice (**Figure 1**). Another source of similar ice core data is available from European EPICA drilling project [7] which lasted from 1998 until 2005. EPICA data are comparable with Vostok data. In this paper we focus on specific analysis related to only two of the Vostok data variables, namely relative temperature and CO_2 content. We also touch upon possible effect of the dust on various points along Vostok timeline. We will include other data at a later date. Data sets used are from [8] and, with corrections, [9]. The variation of atmospheric CO_2 , temperature and dust are shown in **Figure 1** together with our definition of four cycles (C1, C2, C3, C4) formed from the maxima of the variables. These cycles could also be defined starting from the variable minima. Before presenting very detailed time and frequency analysis of specific subset of Vostok data, we make some general observations related to climate fluctuations of CO_2 , temperature and dust.

Variations in global climate measured by temperature change automatically involve the thermal energy content and heat capacity of the atmosphere. These highly variable systems can be contrasted with the energy content of conservative physical systems like planetary orbits, where the principle of least action defines the trajectories favored; conservative systems can be reliably modeled as an interaction

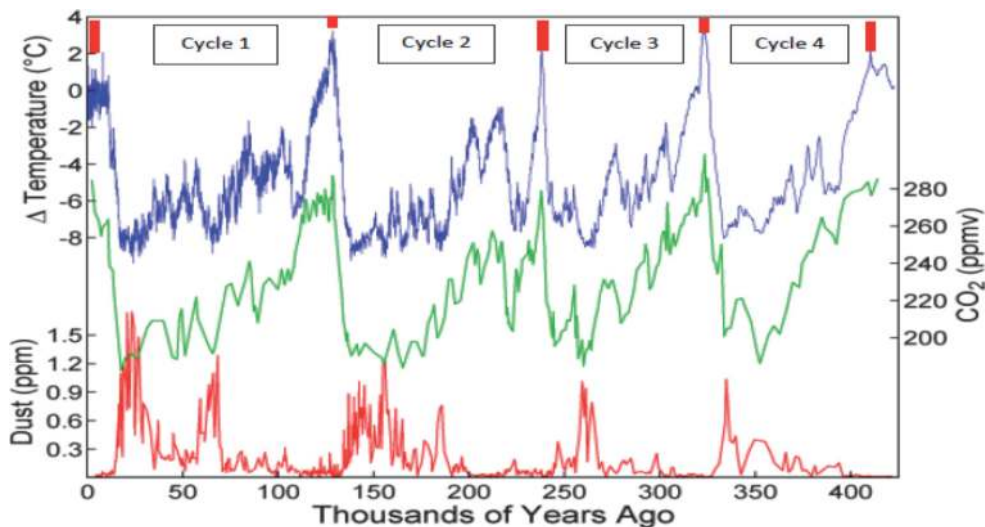


Figure 1.
Vostok data variation in global relative temperature and CO_2 content [8, 9].

between gravity and inertial forces, with the angular momentum and action held constant as a function of mass and the gravitational constant [10]. By contrast, the molecular states in the global ecosystem are dominated by transient energy flows, oscillating in a complex set of time scales as short as hourly to as long as geological epochs. Such variations in energy flows invite frequency analysis, characterized with Fourier transforms to provide underlying information.

A “snowball” Earth [11], more or less covered by ice and snow as the ice cores were fashioned clearly has diminished total energy in terms of molecular quantum states affecting molecular vibration, rotation, translation – as well as sustaining the more coherent circulatory motions in the atmosphere and the ocean, transferring thermal energy towards the poles. A warm Earth as at present has all these flows in higher quantum states, raising their entropy as we have quantified for atmospheric molecules elsewhere [12]. Variation in physical parameters such as temperature is obvious but complex in causes, with any longer-term trends overlying daily or seasonal trends as a function of latitude.

One state property that does not vary over time on Earth except geologically is atmospheric pressure, given that it is effectively based on the weight of the atmosphere, apart from variation for water, some 4% during the El Nino cycle [13]. The current IPCC consensus concludes that the most significant warming is from greenhouse gas content, such as that of CO₂, methane and nitrous oxide, but this is not necessarily the major controlling cause of increasing temperature, given how negotiable thermal energy is. Globally, we are conducting a large uncontrolled experiment to see if CO₂ is so important. Although originally thought that the CO₂ data from ice cores might be considered as proof of its causal role in global warming, but given that changes in CO₂ lag temperature changes its resultant release or absorption from solution in sea water is more likely the cause of correlations in the relationship.

An additional causative factor controlling the atmospheric pressure of CO₂ that seems to have been overlooked is varying acidity of waters, which can also respond to temperature. In warm periods with high oxygen levels, acidification by oxidation of reduced sulfur, nitrogen and carbon compounds is favored [14]. Alkaline conditions are favored anaerobically. The reduction of dimethyl sulfate, a major oxidant in sea water, can also lead to the evolution of reduced sulfur compounds, which can be converted to sulfuric acid in the atmosphere by ultraviolet oxidation. The alkaline pH of the ocean about 8.2 is effectively the same as that obtained in aerated water equilibrated with limestone (CaCO₃). Sources of alkaline carbonated salts in the ocean is basaltic volcanic rocks and soils, in contrast to more acidic granitic deposits that tend to acidify water.

We have shown using equilibrium theory that a lowering in surface ocean pH of 0.01 units can lead to an increased CO₂ pressure in the atmosphere of 8–10 ppm by volume. So, a change of 100 ppm could occur locally if the ocean surface or water on land was acidified by 0.10 unit, say from pH 8.20 to 8.10 as observed recently. Much larger local changes could occur, from a catastrophic event such as major volcanic eruption. The annual oscillation in CO₂ pressure at 3500 m altitude on Mona Loa, Hawaii could be a result of the change in the high surface temperature of the nearby ocean between winter and summer, rather than imbalance between photosynthetic assimilation and respiratory evolution of CO₂, often claimed in general climate models. The CO₂ pressure in Hawaii peaks after winter in May when the sea temperature is about 23°C and reaches its minimum in late October when the sea is about 27°C. A peculiarity of calcite is that its solubility declines with higher temperature [15], so its precipitation removing CO₂ from the atmosphere could partly or fully explain the oscillation. By contrast, at the sampling site at Cape Grim in Tasmania, where the annual sea temperature has a mean temperature

of 15°C with variation between 11 and 19, a range which does not lead necessarily to precipitation of calcite, there is only slight evidence of an oscillation. While the burning of fossil fuels would be the main current cause of pH variation in the ocean, any other acidifying processes in the atmosphere or on land such as from agriculture, could also be controlling influences.

For all of the above reasons regarding the complexity of causes of climate change, we consider it would be beneficial if the ice core data could be subjected to careful Fourier frequency analysis, yielding detailed data regarding long term mechanisms of variation in climate, also pointing to an appropriate dynamic modeling of the underlying processes. In this chapter we focus specifically on CO₂ Vostok data for C1 as defined in **Figure 1**, as an example of our CPE approach, extending this to CO₂ for C2, C3 and C4, or their combinations. Other data such as temperature, methane, dust, and any other available ice core data such as EPICA, can be analyzed similarly.

2. Methodology

The eminent statistician Fisher [16] was an early exponent of testing statistical significance by harmonic analysis. Some preliminary spectral analysis has been conducted on Vostok ice core data set as reported in [17–19]. In this paper, a detailed spectral analysis in R and Excel was applied to Barnola et al. data set [8, 9]. In such analysis, time series are decomposed into underlying sine and cosine functions to establish the most important frequencies. Various texts on Fast Fourier Transforms (FFT) were also sources for frequency determination. These approaches allow construction of periodograms quantifying the contributions of the individual frequencies to the time series regression. The methodology developed for bioinformatics [20] has general application for time series and was employed in this study. In this paper we perform our own time and frequency analysis using R and Excel. Trudinger et al. [21, 22] have also applied Kalman filter analysis to ice core data. This allowed a more rigorous analysis of CO₂ variability for the Law Dome ice cores of Antarctica over the recent 1000 years than the usual deconvolution method. These authors pointed out that the Kalman filter allows better calculation of uncertainties in the deduced sources. The uncertainties correspond to the selected range of frequencies. They claimed [22] that it allows investigation of statistical properties that are directly related to physical properties.

Our aim is to apply Kalman filter based methodology to define the logic for a Climate Prediction Engine (**Figure 2**) based on 4 data cycles, **Figure 1**. Cycle C1 is the most recent period, the current interglacial warming period nearing its maximum but still incomplete. The overall number of published data points for both variables (relative temperature and CO₂) is 363. The individual cycles are determined by locating maximum absolute values for relative temperature and CO₂. Individual cycles differ in the number of data points slightly. Also, because of a lag observed between the maxima for temperature and CO₂ data, the number of data points differs slightly in two data sets in each cycle. This is also confirmed in our time correlation analysis below. **Table 1** summarizes the number of data points in each cycle. Note that each cycle is defined as top-to-top data values for both CO₂ and temperature, resulting in a different number of data points for CO₂ and temperature. However, the beginning of each interglacial where temperatures commences climbing might prove superior for some analyses, given that the controlling causes for reversal may be more consistent for these periods. Obviously, the current cycle is still evolving and new data may be added as required for further analysis. We can achieve that by appending the original Vostok data with additional current points, that might skew the previous natural data progression. Comparison of two

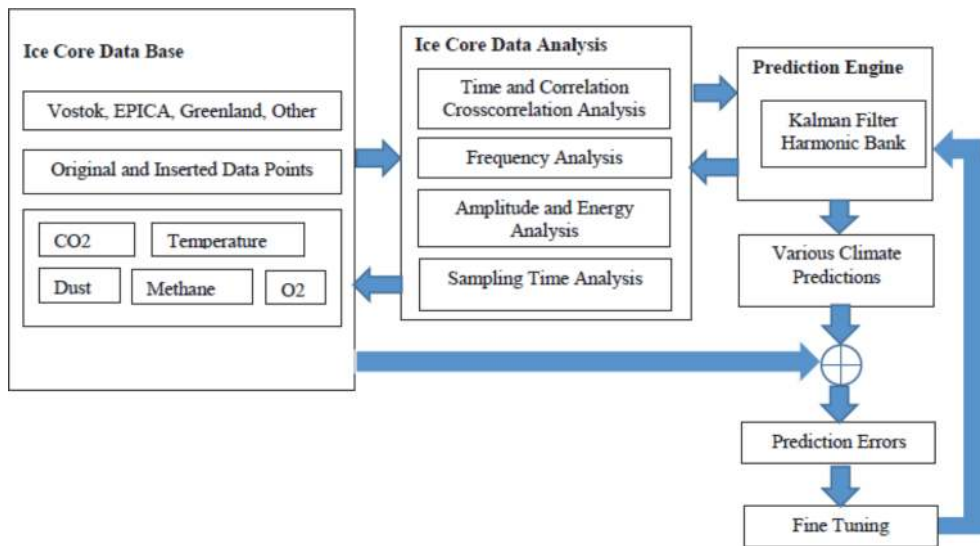


Figure 2.
 Climate prediction engine (CPE) logic.

No of Data Points	Cycle C1	Cycle C2	Cycle C3	Cycle C4	Total
Temperature	74	136	99	54	363
CO₂	80	135	99	49	363

Boldfaced entries point to CO₂ and Cycle 1 related values.

Table 1.
 Number of original Vostok data points for each cycle.

data sets including current data may be very useful in estimating global near and far future effects of factors like temperature and CO₂ content – an ultimate goal in this research.

Note that one such additional current data point may not make a large difference on the current C1, as far as harmonic analysis, which has long term time behavior built in, but the short term time effects may be more prominent. Our approach covers both short and long term effects and it can be used to perform sensitivity analysis using current climate data together with original Vostok data. The uneven data sampling times within each cycle may pose some numerical issues as well for the analysis in the context of Kalman filter (KF) prediction methodology. There are various methods missing data problem and one of the simpler methods, yet effective one, is to enter missing data by some (linear or nonlinear) approximation method. We inject additional data using linear interpolation, also achieving Nyquist sampling requirements and in the process minimize the number of required harmonics used by the Kalman filter harmonic bank (KFHB). For C1 which originally had 80 data points for CO₂ we added additional data points for the total of 128 suitable for FFT but also corresponding to the length of 128,000 years so far in the cycle.

An aim is to gain more insight into Vostok data in time and frequency domains, and check the corresponding amplitude and energy content in order to reduce the number of significant harmonic components (**Figure 2**). Other authors used similar energy consideration but in our case, we propose the simple and effective mathematical model of a stochastic harmonic oscillator, based on which a KFHB can be built and used to further analyze Vostok data as well as predict near and far future data behavior. Note that besides C1 other cycles can be used to improve the precision by combining two or three cycles based on past data for C1, C2, C3. Amplitude

and energy analysis performed here reduces a number of required individual Kalman Filter Harmonic Oscillators (KFHO) as individual blocks for the KFHB. It is important to note that using this approach rather than just combining amplitudes and harmonic components directly as described for example in [23], is to accommodate stochasticity of the data and the overall probabilistic nature of the prediction problem using KFs. This approach adds to the robustness of the method and provides better probabilistic accuracy both for short and long term periods, as well as allowing for prediction sensitivity analysis using various values, such as CO₂ levels as measured now and estimated for some future periods. Our methodology can be applied for both research as well as for policy making tools for the future climate related societal and technological decisions.

In this paper we perform time analysis of all cycles and C1 CO₂ frequency analysis (boldfaced in **Table 1**) to illustrate the KFHB approach (Section 7).

3. A climate prediction engine

The block diagram in **Figure 2** summarizes our CPE. It seeks the ability to predict various quantities using both past data (from one or more cycles) as well as the current data available. This also allows to run various sensitivity analysis by using a variety of future scenarios as far as CO₂ and temperature, for example. The CPE can be applied to any ice core data, Vostok, EPICA or any other and for any variable, CO₂, temperature, methane, O₂, or dust.

The CPE consists of three major blocks. First one is a data base of various ice core data, both original, conditioned and inserted as required to make the data more uniformly distributed across the time span of each cycle. The second block is data analysis, in time and frequency domains, including various correlation measures. The third block is the prediction engine which consists of a set of oscillators (KFHO) that produce a KFHB (Section 8). Various predictions as well as sensitivity analysis are performed in this block. Finally prediction parameters are fine-tuned by original ice core data vs. the prediction errors. This is primarily done to fine tune KFHO gains. Section 8 contains all the mathematical details of KFHB.

4. Average sampling time analysis

First, we considered the entire set of Vostok data regarding time sampling. This will yield some initial indication of important issues in dealing with the data and how to perform further Vostok data filling to minimize the effects of non-uniform data distribution. As **Table 1** indicates the number of data points and the corresponding time periods for each cycle is quite different. This has to be taken into account when analysis is performed, as far as machine learning use of individual cycles for the benefit of estimating on going C1 data values. **Table 2** summarizes approximate duration of each cycle based on our definition of four cycles.

Duration in years	Cycle C1	Cycle C2	Cycle C3	Cycle C4
Temperature	127,726	115,156	86,462	96,782
CO₂	128,399	109,800	86,148	95,587

Boldfaced entries point to CO₂ and Cycle 1 related values.

Table 2.
Climatic cycles approximate duration in years.

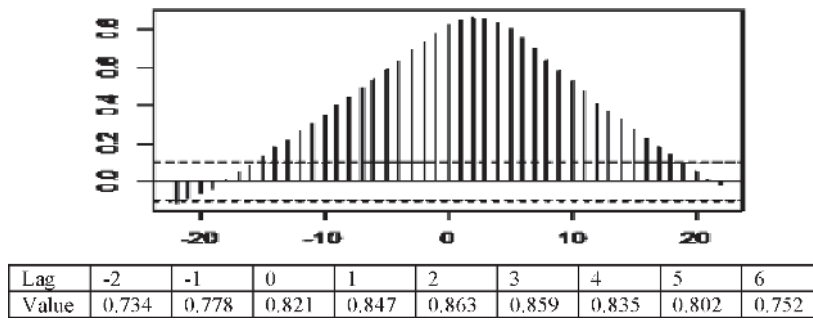


Figure 3.
 Short term temperature and CO₂ cross correlation.

Detailed visual inspection of maximum values in **Figures 2** and **3** indicates that CO₂ lags relative temperature in C4, C3, and C2. In C1, the data may be a little skewed due to the number of recent data points where maximum values are still not clearly identified in Vostok data (left-most CO₂ data, **Figure 1**), as C1 is still evolving. Better lag estimates can be obtained by cross correlation analysis as done in Section 4 below. Another feature is the varied time differences when the data is obtained from ice core readings. In some cases difference between two data samples are only 400–500 years, but in some as much as 5000 years. To start the analysis for each cycle we calculated average time sampling values as summarized in **Table 3**. In Section 5 this will be corrected to some extent by data filling and determining an ideal Nyquist and data sampling rate.

The differences in average sampling times obviously come from the number of data points collected and the duration of each individual cycle. To facilitate our general approach all other combination of available cycles data are also considered. The idea is enrich with the varied of all available data for prediction purposes and also training purposes of machine learning approach. For example **Table 4** indicates respectively average sampling times for C1 and C2 combined (C12), C2 and C3 combined (C23), as well as C2, C3 and C4 combined (C234), plus the total data set C1234. Note from **Table 4** that more cycles we add the more uniform average data sampling times become, until it becomes equivalent for both CO₂ and temperature. Further refinement of average sampling time is given in Section 5.2 below.

Average sampling time in years	Cycle C1	Cycle C2	Cycle C3	Cycle C4
Temperature	1703	800	865	1792
CO₂	1605	813	862	1911

Boldfaced entries point to CO₂ and Cycle 1 related values.

Table 3.
 Average data sampling time in years for each cycle.

Average sampl. time, years	Cycles C12	Cycles C23	Cycles C234	Cycles C1234
Temperature	1087	827	1008	1149
CO₂	1108	834	1020	1149

Boldfaced entries point to CO₂ and combined Cycle C12.

Table 4.
 Combined cycles C12, C23, C234 and C1234 mean data sampling times in years.

5. Time correlation analysis

As a starting point in time cross correlation analysis we examine correlation coefficients – single numbers that can be used as simple measure of cross correlation intensity between two variables. There are several coefficients named after their inventors such as Spearman, Pearson and Kendal [24] and they all indicate certain statistical properties that can relate two data series. **Table 5** summarizes standard cross coefficient between relative temperature and CO₂ for individual cycles as well as for the entire Vostok data set. The intensity of the cross correlation is quite high, on average more than 0.8 for the entire set. If we split the cycles into up and down sub cycles we obtain **Table 6** which indicates cross correlation coefficients for up and down cycle parts. Overall these coefficients indicate bigger spread between up and down sub cycles, and are very sensitive to where the break between up and down parts is chosen.

In general, one of the coefficients (up or down) is considerably larger than the overall single cycle coefficient. This might indicate that the usefulness of the individual up and down cross correlation analysis may be limited of the current C1 cycle. In the context of machine learning methodology this points to putting less emphasis on the cycles from longer in the past compared to the ongoing C1 cycle. To complete this analysis, the down-up period (boldfaced) ratios in **Table 6** indicate that the descending period is on average 6 to 8.6 times longer than the ascending period during interglacials, given that the cooling period is that much longer than the warming part. See also **Figure 9** for C1 CO₂.

Cross correlation coefficient	Cycle C1	Cycle C2	Cycle C3	Cycle C4	Entire cycle C1234
Temperature vs. CO ₂	0.8389	0.8094	0.8069	0.8191	0.821

Table 5.
Cross correlation coefficient for individual and entire cycle.

Cycle 1	Down	Correlation	0.869	Down	Years	108,328
					Ratio	7.05582
	Up	Correlation	0.8262	Up	Years	15,353
	Total	Correlation	0.8389	Total	Years	123,681
Cycle 2	Down	Correlation	0.8008	Down	Years	97,087
					Ratio	8.653802
	Up	Correlation	0.9014	Up	Years	11,219
	Total	Correlation	0.8094	Total	Years	108,306
Cycle 3	Down	Correlation	0.8558	Down	Years	72,931
					Ratio	5.961337
	Up	Correlation	0.6503	Up	Years	12,234
	Total	Correlation	0.8069	Total	Years	85,165
Cycle 4	Down	Correlation	0.8519	Down	Years	83,533
					Ratio	6.824034
	Up	Correlation	0.827	Up	Years	12,241
	Total	Correlation	0.8191	Total	Years	95,774

Table 6.
Cross correlation coefficients for individual sub cycles.

Time correlation analysis produces a variety of useful information about periodicity and correlation strength among data samples of a given quantity. In particular autocorrelations produce the measure of self correlation of a data series, and the cross correlations indicate how two different data sets are correlated. We used standard programs such as R and Excel to generate those. One property of crosscorrelations R_{xy} is very useful in analysis of relative temperature vs. CO_2 content and that is the estimate of the time delay between the two. In general one needs to locate the maximum value point for R_{xy} and locate the corresponding argument, time lag in our case, between relative temperature and CO_2 content. **Figure 3** (with numerical values around zero lag) illustrates short term calculations for cross correlation between the two variables for the total C1234 cycle. We can read the value of the delay τ_{delay} between temperature and CO_2 as approximately equal to two lag units. Since the calculation is done for the entire data set, from **Table 4** we can read an average sampling time for C1234 as 1149 years, hence we can make a rough upper limit approximation of the size of the delay for the entire data set to be:

$$\tau_{\text{delay}} < 2 \text{ times } 1,149 \text{ years} \sim 2,300 \text{ years}$$

The calculation is approximate, primarily because the non-uniform distribution of the Vostok data. Some data is separated by only hundred years rather than thousands. This points to a need to make the data more uniform by inserting additional data. We address how to harmonize these data in Section 5. The total delay appears to be of order of 2000 or more years and not 100–200 years. That may be an important finding which can influence our thinking about the role of CO_2 increase caused by human actions. **Figure 4** indicates entire data set auto and cross correlation. It is clear that the data exhibits some periodicity. To determine average time delays between relative temperature and CO_2 for individual cycles we examine **Figures 5–8** which also have numerical values around zero lag. The first two diagrams in **Figure 5** are autocorrelations and they also indicate certain periodicity within the each cycle but obviously not as well as the entire data set. The third diagram shows cross correlation between two variables. The time delay can be read from cross correlation and for C1 it is less than one lag period but maybe more than zero lag, due to the non uniformity of data. We can estimate it as less than half of one period lag. From **Table 3** for both

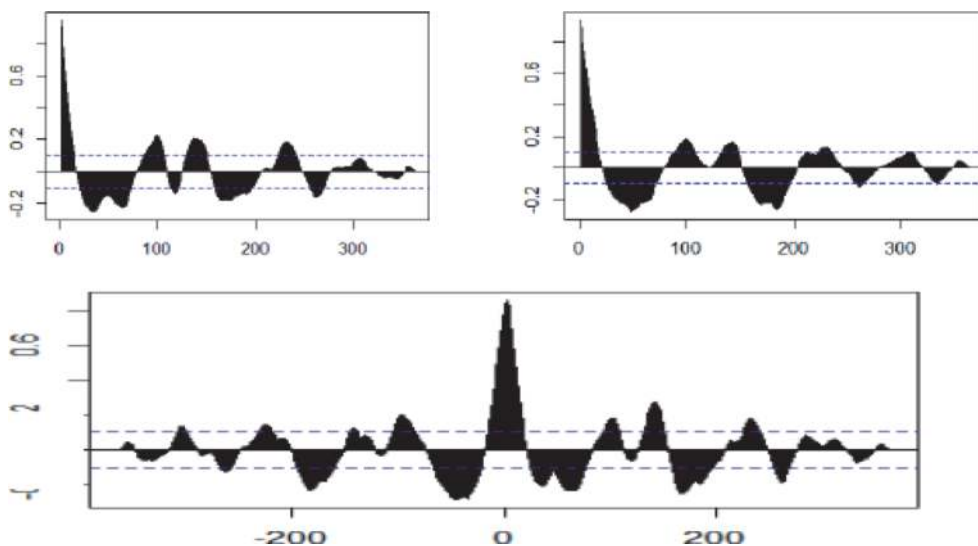


Figure 4. Total long term temperature and CO_2 autocorrelations (left to right, above) and cross correlation (bellow) indicating inherent data periodicity.

relative temperature and CO₂ the average data sampling times are 1703 and 1605 years putting the absolute delay at around 800–850 years or less. Similar approximate estimates can be done for other cycles. Reading from C2 and C3 cross correlations, **Figures 7 and 8 and Table 3**, by the same consideration the delay is less than half the cycle, which translates into 400–435 years on average, or less. For C4 (**Figure 8 and Table 3**) the delay appears to be of order of one cycle sampling time, a delay around 1800 years. To get a more precise approximations we would need more uniform data and finer resolution around the zero lag where the cross correlation is at its maximum. Note that on the individual cycle up and down parts this delay may differ from the average cycle level, also indicated by larger correlation coefficients spread in **Table 6**. We can identify no specific pattern in these coefficients regarding up and down sub cycles having larger or smaller coefficient. Overall, there is a significant CO₂ delay across total Vostok data C1234 compared to the relative temperature. For individual cycles a more detailed cross correlation analysis should be done, especially following data insertion. We made very rough estimates above.

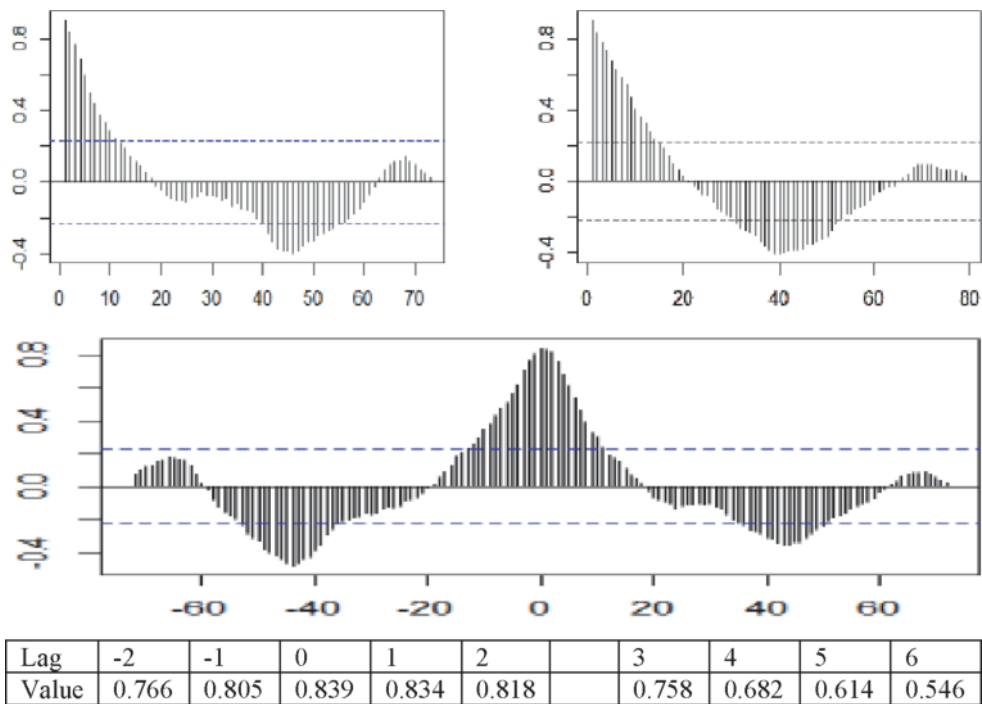


Figure 5. Cycle C1 temperature and CO₂ autocorrelations (left to right, above) and cross correlation (bellow) indicating some periodicity within the cycle C1 itself.

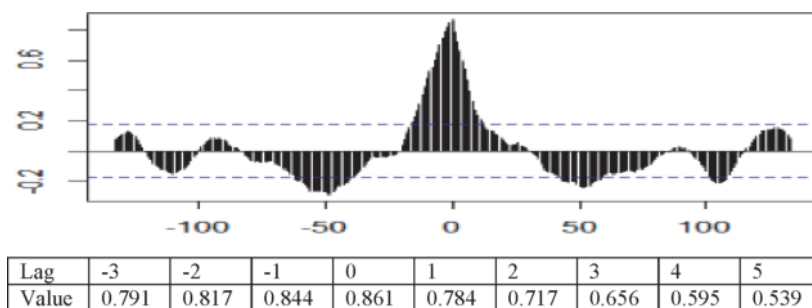


Figure 6. Cycle C2 temperature and CO₂ cross correlation indicating some periodicity within the cycle C2 itself.

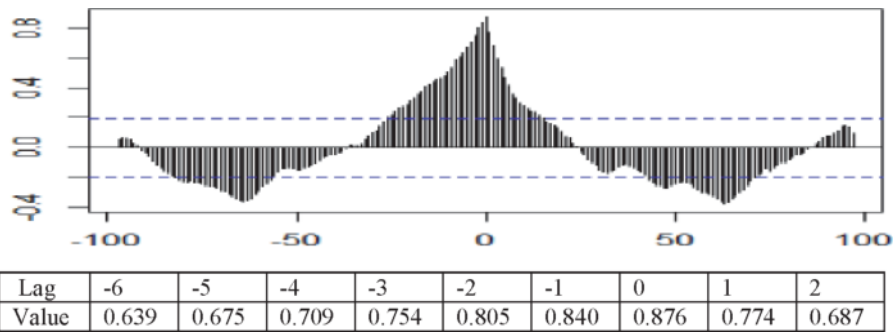


Figure 7.
 Cycle C3 temperature and CO₂ cross correlation indicating some periodicity within the cycle C3 itself.

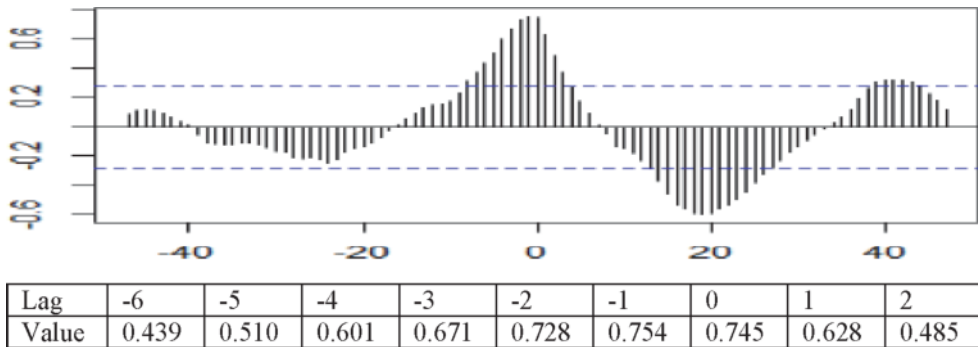


Figure 8.
 Cycle C4 temperature and CO₂ cross correlation indicating some periodicity within the cycle C4 itself.

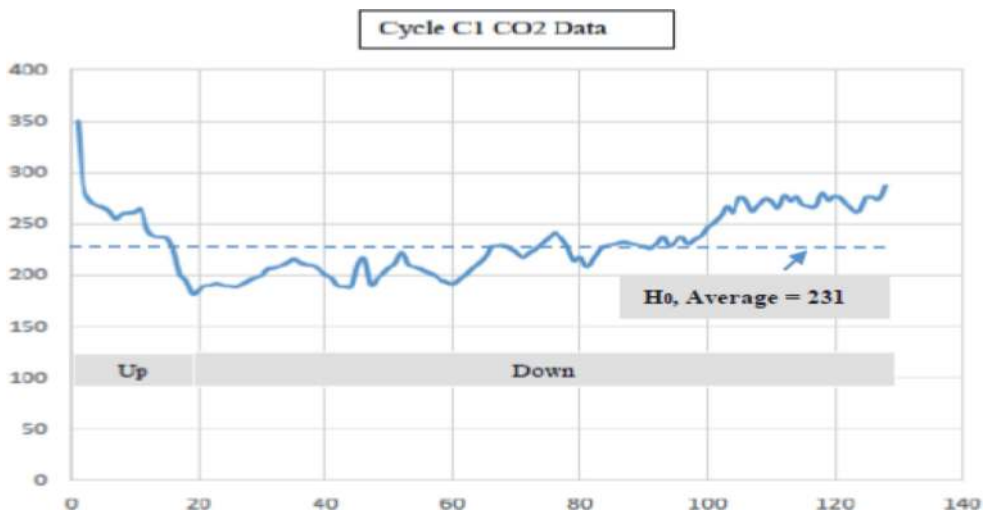


Figure 9.
 Cycle C1 CO₂ content.

For practical reasons, the most important figure to keep in mind is the current cycle C1 delay which points to 800–850 years or less.

Note at the end of this Section that the way various cycles are defined (**Figure 1**) also affects the analysis. In a follow up work we aim at repeating this analysis by defining cycles from minimum to minimum CO₂ values. Choice of maximum or minimum values may assist in determining what triggers cycle reversals. We have early indications that the dust might have a significant role in this reversal.

6. Cycle C1 CO₂ frequency analysis

As an illustration of frequency analysis we examine Cycle C1 CO₂ data in details. Similar approach can be used for other data, such as temperature, methane, oxygen and dust. The first concern is to make the C1 CO₂ data more homogeneous hence the sampling time can be more precise and useful.

6.1 Data filling

The original Cycle 1 Vostok data contains a total of 80 CO₂ data points are very unevenly distributed in time. This poses issues with FFT or DFT harmonic analysis. Hence we corrected the situation somewhat by inserting certain number of additional “data points” using linear approximation between the original data points with the biggest time difference between the neighboring points, hence generating a total of 128 original and inserted data points. In particular, the very first point we added to the original Vostok data is the current value of CO₂ which we stated at 350 [25] for the year 1990, as an example. Current levels are estimated at over 400 ppm. Our CPE model can be used for a variety of sensitivity analysis by playing “what if” and processing the data through CPE.

Next data is the first original Vostok data, 2362 years in the past. This will correspond to 64 FFT harmonics described next. Note that the time period of C1 CO₂ data is 128,399 years (**Table 2**), time difference between the oldest C1 data and the current time, **Figure 9** above.

6.2 Sampling time

For sampling time, we proceeded as follows. As stated in Section 2 the cycle duration is measured between the maximum (CO₂ and temperature) points in each cycle. Hence, we have an average of a bit over 1000 years time sampling for Cycle 1, i.e. $T = \frac{128,399}{128-1} = 1011$ years, which is used as the CO₂ data sampling time. The FFT analysis on 128 total points produces 64 harmonics with the highest frequency of $f_{max} = \frac{f}{2} = \frac{1}{2T} = \frac{1}{2022} 10^{-3} = 0.000495 \times 10^{-3}$ Hz. For the analysis we use $T = 1.011$ with the understanding it is in 1000s of years. **Table 7** summarizes Cycle 1 harmonic content ordered by the amplitude, from the largest to the smallest to identify each harmonic energy contribution. First harmonic H_1 is boldfaced. As shown in Section 7.1, for the discretization purposes, it is important to satisfy an additional sampling time requirement given in inequality (7) bellow which produces equivalent resonant frequency of continuous and discrete models, and in turn makes for more precise discrete KFHO and KFHB models as well. This increases required sampling frequency and reduces time sampling period bellow 1011 years, and we deal with it using energy analysis as described in **Table 8**. For other approaches to non-uniform data analysis see [26, 27] as well as general literature on non uniform and optimal Fourier analysis.

6.3 Amplitude and energy analysis

Table 7 shows that the average signal value (DC or H_0) is 231 and the rest of the energy content (reflected in a corresponding harmonic amplitude) of the harmonics follows the highest-to-lowest amplitude pattern $H_1, H_2, H_4, H_3, H_7, H_{10}, H_5$, and so on.

The cumulative column in **Table 7** indicates amplitude sum percentage of subsequent harmonics compared to the sum of all harmonic amplitudes. For example to reach 90% + of the total amplitude we need a total of 25 harmonics, including H_0 .

H No	Frequency f	Period 1/f	Phase	Amplitude	Cumulative %
0	0	N/A	0	231.0068	58.83917
1	0.007727498	129.408	0.754987	36.27814	68.07948
2	0.015454995	64.704	0.555383	15.87177	72.12214
4	0.03090999	32.352	-1.11877	9.409128	74.51872
3	0.023182493	43.136	0.478204	6.815518	76.25468
7	0.054092483	18.48686	1.435168	5.576594	77.67508
10	0.077274975	12.9408	0.316939	5.098453	78.97369
5	0.038637488	25.8816	-1.14182	4.847574	80.20841
12	0.09272997	10.784	-0.20146	4.780349	81.426
6	0.046364985	21.568	-2.02131	3.49376	82.31588
17	0.131367458	7.612235	0.471006	3.462987	83.19793
14	0.108184965	9.243429	-0.58073	3.138693	83.99738
15	0.115912463	8.6272	0.489932	2.959198	84.75111
26	0.200914936	4.977231	-0.08324	2.30261	85.3376
25	0.193187438	5.17632	0.527865	2.031371	85.85501
38	0.293644906	3.405474	-0.48454	2.018039	86.36902
32	0.247279921	4.044	-0.171	1.841643	86.8381
11	0.085002473	11.76436	0.041037	1.827369	87.30354
34	0.262734916	3.806118	0.371439	1.794662	87.76066
37	0.285917409	3.497514	0.402143	1.71814	88.19828
60	0.463649852	2.1568	0.050809	1.694368	88.62985
21	0.16227448	6.162286	-1.80194	1.617602	89.04186
23	0.177732443	5.626435	0.231643	1.60276	89.4501
28	0.216369931	4.621714	0.236801	1.570401	89.85009
27	0.208642433	4.792889	-0.21626	1.560695	90.24761

Boldfaced entries correspond to the first harmonic H_1 .

Table 7.
 Summary of cycle 1 CO_2 harmonic analysis.

Highest frequency	Maximum frequency	Cumulative amplitude %	Total H's	Sampling Ts <	Chosen Ts
15	0.293645	86.369	15	1.083996	1.011
15	0.293645	88.1983	19	1.083996	1.011
20	0.46365	90.2476	24	0.686531	0.650
63	0.486832	96.6477	44	0.653839	0.650

Boldfaced entries point to the lowest limit on Sampling Time Ts.

Table 8.
 Cycle 1 CO_2 amplitude and sampling time.

The last (boldfaced) harmonic in this list is H_{27} . For a bit smaller cumulative amplitude, say 88%+, we only need 19 harmonics, with H_{37} as the last one. **Table 8** further summarizes energy analysis. For example, a chosen sampling time of 1011 satisfies the requirement in inequality (7) below all the way to cumulative energy

MAPE	Harmonics						
	1	1 + 2	1 + 2 + 4	1 + 2 + 4 + 3	1 + 2 + 4 + 3 + 7	1 + 2 + 4 + 3 + 7 + 10	1 + 2 + 4 + 3 + 7 + 10 + 5
y	0.058	0.043	0.038	0.036	0.034	0.033	0.031
x Corrected	0.058	0.043	0.037	0.036	0.034	0.033	0.030
x Predicted	0.056	0.044	0.039	0.037	0.035	0.034	0.033

Table 9.
Kalman filter harmonic bank estimation errors.

of 88.1983%. To reach the energy level of 90.2476%, much lower sampling time would be required, i.e. 650 years instead of 1011. In order to satisfy this requirement we would need to almost double the total number of data points and use more harmonics. Fortunately, not all the amplitudes are at their maximums all the time, hence the above percentages are only very conservative lower bounds. In reality we can calculate the errors by MAPE (Maximum Absolute Percentage Error) and these will be much more realistic, actually of order of 3–5% off of 100%, per **Table 9**. Also, energy of these first seven harmonics is of order of 99% + due to its calculation based on amplitude squares. This analysis gives us an idea of various issues at hand related to non-uniformity of Vostok data as well as number of harmonics to be used based on energy analysis. Our aim is to show that properly designed KFHB based CPE can deal with these issues in a very effective way. In Section 7 we illustrate various points raised here using first 7 harmonics, $H_1, H_2, H_4, H_3, H_7, H_{10}, H_5$.

7. Kalman filter harmonic bank

In an illustrative example in this paper we focus on C1 CO₂ harmonic analysis and the corresponding KFHO for the strongest, amplitude wise, first harmonic H_1 in **Table 7**. The same analysis can be repeated for any harmonic as we elaborate in Sections below. We proceed with the development of KFHB which models a series of harmonic components obtained by the energy analysis, which approximates to a reasonable degree (say 90% + or higher cumulative amplitude percentage) the original Vostok data. First step is to define a general Markov model [28] for a generic harmonic oscillator in discrete time described in the following Section. Other authors also used the Kalman filter approach to analyze ice core data, but with a different emphasis, [21, 22] as mentioned in Introduction.

7.1 Discrete time harmonic oscillator

To start we introduce a continuous model of a Harmonic Oscillator described in [29, 30]:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega_0^2 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad (1)$$

where ω_0 is radial frequency, $\omega_0 = 2\pi f_0$, with f_0 frequency in Hz. The solution to the above state equation is sine or cosine function depending on the initial conditions. The state variables x_1 and x_2 are “position” and “velocity” of whatever variable we are dealing with, such as CO₂ content and its rate of change. The first

level of continuous-to-discrete approximation is based on one point derivative approximation and sampling time T , whereas we obtain:

$$\begin{pmatrix} x_1(t+1) \\ x_2(t+2) \end{pmatrix} = \begin{pmatrix} 1 & T \\ -\omega_0^2 T & 1 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \quad (2)$$

with “ $t+1$ ” standing for $(t+1)T$, and similarly $(t+2)T$ for “ $(t+2)$ ”, where T is dropped for simplicity. To the above discretized time model we can also add model uncertainty via additional stochastic zero mean Gaussian white inputs r_1 and r_2 , with certain variance values, V_1 and V_2 which can be fine-tuned. Hence we have:

$$\begin{pmatrix} x_1(t+1) \\ x_2(t+1) \end{pmatrix} = \begin{pmatrix} 1 & T \\ -\omega_0^2 T & 1 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + \begin{pmatrix} r_1(t) \\ r_2(t) \end{pmatrix} \quad (3)$$

The initial conditions are given as a transposed vector $(x_1(0)|x_2(0))$ with:

$$x_1(0) = A_0 \cos(\theta_0), x_2(0) = -A_0 \sin(\theta_0) \quad (4)$$

where A_0 and θ_0 are the amplitude and the phase of the harmonic ω . Better discrete approximation can be obtained by 2 point derivative approximation whereas we obtain:

$$x_1(t+1) = -a_0 x_1(t) - x_1(t-1) \quad (5)$$

where parameter a_0 is calculated to match discrete and continuous resonant frequencies:

$$a_0 = -2 \cos(\omega_0 T) \quad (6)$$

It is important to note that in this case one needs to choose sampling time $T = 1/f$ to satisfy:

$$f > \pi f_0, \omega_0 = 2\pi f_0 \quad (7)$$

which is higher than the standard Nyquist frequency, $f > 2f_0$. We show in the next Section an example of this. Eq. (5) produces cosine function with the proper initial conditions $x_1(0)$ in (4). The corresponding equation for $x_2(t+1)$ is equivalent to (5) with the initial condition $x_2(0)$ in (4) to produce sin function. Instead of proceeding with two state model (3) we can produce another two state model using (5):

$$\begin{pmatrix} x_1(t+1) \\ x_1(t) \end{pmatrix} = \begin{pmatrix} -a_0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_1(t-1) \end{pmatrix} + \begin{pmatrix} w_1(t) \\ w_2(t) \end{pmatrix} \quad (8)$$

$$y_0(t) = x_1(t) + v_0(t) \quad (9)$$

where $w_1(t)$ and $w_2(t)$ are zero mean Gaussian white inputs with joint covariance symmetric matrix $Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{12} & Q_{22} \end{pmatrix}$, [2]. The measurement $y_0(t)$ of $x_1(t)$ with the measurement error $v_0(t)$ is simply defined in (9) and it corresponds to the harmonic ω_0 . Measurement error $v_0(t)$ is zero mean stochastic process with variance R_0 assumed constant across all time samples, which is a reasonable assumption, unless

there is a compelling reason to make it time varying. The model (8) and (9) above remains the same. Obviously equivalent model holds for $x_2(t+1)$ with the proper initial conditions. The model as given by (8) and (9) is our starting point for KFHO described next. The consideration holds for any harmonic ω .

7.2 Kalman filter harmonic oscillator

To facilitate the next step, we rewrite (8) as:

$$\begin{pmatrix} x_1(t+1) \\ x_2(t+1) \end{pmatrix} = \begin{pmatrix} -a_0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + \begin{pmatrix} w_1(t) \\ w_2(t) \end{pmatrix} \quad (10)$$

where $x_2(t+1)$ is just an auxiliary notation for $x_1(t)$ and it is not $x_2(t)$ in (3). Then the standard KF equations in the above case produce [2]:

Prediction Step:

$$\hat{x}_1(t+1/t) = -a_0\hat{x}_1(t/t) - \hat{x}_2(t/t) \quad (11)$$

$$\hat{x}_2(t+1/t) = \hat{x}_1(t/t) \quad (12)$$

Correction Step:

$$\hat{x}_1(t+1/t+1) = \hat{x}_1(t+1/t) + K_{11}(t+1)\tilde{y}_0(t+1) \quad (13)$$

$$\hat{x}_2(t+1/t+1) = \hat{x}_2(t+1/t) + K_{21}(t+1)\tilde{y}_0(t+1) \quad (14)$$

In (13), (14) above, $\tilde{y}_0(t) = y_0(t) - \hat{x}_1(t/t-1)$ is Innovation Sequence and filter gains are:

$$K_{11}(t) = P_{11}(t/t-1)/[P_{11}(t/t-1) + R] \quad (15)$$

$$K_{21}(t) = P_{12}(t/t-1)/[P_{11}(t/t-1) + R] \quad (16)$$

The corresponding Prediction Step and Correction Step variances and covariances of the estimation error $\tilde{x}_1(t/t-1) = x_1(t) - \hat{x}_1(t/t-1)$, and $\tilde{x}_1(t/t) = x_1(t) - \hat{x}_1(t/t)$, and similarly for the state $x_2(t)$, are:

Prediction Step:

$$p_{11}(t+1/t) = a_0^2 p_{11}(t/t) + 2a_0 p_{12}(t/t) + p_{22}(t/t) + Q_{11}. \quad (17)$$

$$p_{12}(t+1/t) = -a_0 p_{11}(t/t) - p_{12}(t/t) + Q_{12} \quad (18)$$

$$p_{22}(t+1/t) = p_{11}(t/t) + Q_{22} \quad (19)$$

Correction Step:

$$p_{11}(t+1/t+1) = [1 - K_{11}(t)] p_{11}(t+1/t) \quad (20)$$

$$p_{12}(t+1/t+1) = [1 - K_{11}(t)] p_{12}(t+1/t) \quad (21)$$

$$p_{22}(t+1/t+1) = -K_{21}(t) p_{12}(t+1/t) + p_{22}(t+1/t) \quad (22)$$

The initial conditions for the above equations are:

$$p_{11}(1/0), p_{12}(1/0), \text{ and } p_{22}(1/0) \quad (23)$$

and they are determined by the initial Kalman Filter design. One way to determine them is to use matrix Q values:

$$p_{11}(1/0) = Q_{11}, p_{12}(1/0) = Q_{12}, \text{ and } p_{22}(1/0) = Q_{22} \quad (24)$$

Here the values of Q_{11} and Q_{22} are assumed to be of the same order because they represent uncertainty in modeling $x_1(t)$ and $x_2(t)$, and they are just one step apart values of the same state. Simple correlation analysis of $x_1(t)$ and $x_2(t)$ indicates that Q_{12} is of order of $-a_0 Q_{11}/4$. **Figure 10** below shows a block diagram of a single ω_0 KFHO. Here we have the total state vector corresponding to specific harmonic ω_0 as:

$$\hat{x}_0(t/t-1) = [\hat{x}_1(t/t-1), \hat{x}_2(t/t-1)]^T \quad (25)$$

7.3 Kalman filter harmonic bank

Once we define single harmonic KF as in **Figure 10** we can proceed and construct a KFHB as an assemblage of a number of individual harmonic filters in parallel with the combine outputs to form the original signal (data). We assume a set of harmonics $\omega = \{\omega_1, \omega_2, \omega_3, \dots, \omega_N\}$ and for each of $\omega_i, i = 1, 2, \dots, N$ we define a separate KFHO as described above. Note that harmonics are related to each other via:

$$\omega_i = i \omega_1, i = 1, 2, \dots, N \quad (26)$$

and the total signal output (such as Votok data) is the sum of individual harmonic $\omega_i = 2\pi f_i$ outputs:

$$y(t) = \sum_1^N y_i(t) \quad (27)$$

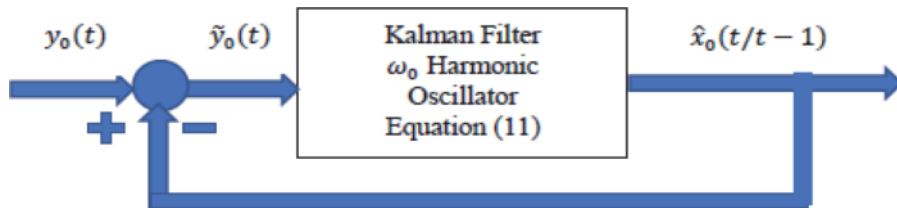


Figure 10.
Kalman filter harmonic oscillator.

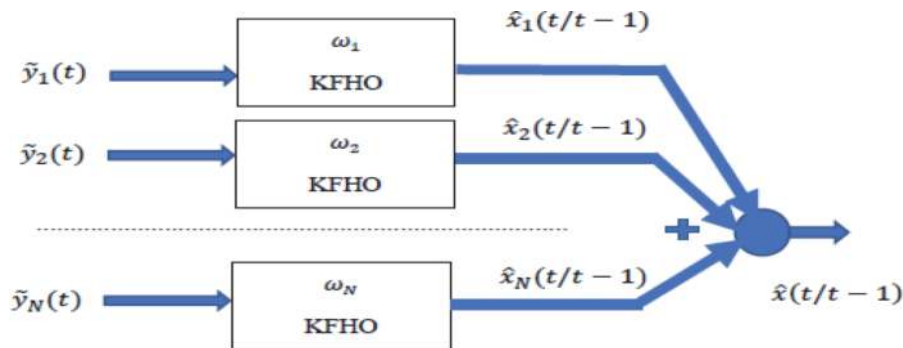


Figure 11.
Kalman filter harmonic bank.

Figure 11 indicates this arrangement with KFHO where we have:

$$\hat{x}(t/t - 1) = \sum_1^N \hat{x}_i(t/t - 1) \tag{28}$$

representing the total KFHB predicted state estimate. These can be used for short and long term prediction purposes for CO2, temperature or other variables of interest. We can similarly define a set of corrected state estimates which are more precise than predicted with the real data available.

$$\hat{x}(t/t) = \sum_1^N \hat{x}_i(t/t) \tag{29}$$

General idea here as compared to a simple sum of harmonic cosine signals (following inverse Fourier Transform) as in (27) with:

$$y(t) = \sum_1^N y_i(t) = \sum_1^N A_i \cos(\omega_i t + \theta_i) \tag{30}$$

is to accommodate stochasticity of the underlying Vostok measurement data as well as a simple linear structure of Kalman Filters Harmonic Oscillator, and its ability to make predictions for the signal future values in the probabilistic (more realistic) environment.

8. Vostok cycle C1 CO₂ Kalman filter harmonic oscillators

8.1 First harmonic H₁

The harmonic analysis of Cycle C1 Vostok data is presented in Section 5. In this Section we proceed and build a specific first harmonic H₁ KFHO, as an example of CPE set up, with the initial conditions:

$$\hat{x}_1(0/0) = A_1 \cos(\theta_1), \hat{x}_2(0/0) = -2\pi f_1 A_1 \sin(\theta_1) \tag{31}$$

Optimal	Constant Gain
k ₁ (+)	k ₂ (+)
N/A	N/A
0.96149	0.48075
0.93616	0.29119
0.93633	0.29926
0.93578	0.30137
0.9357	0.30151
0.9357	0.30151
0.9357	0.30151

Boldfaced entries indicate constant Kalman Filter gains.

Table 10.
Filter gains.

Q_{11}	100
Q_{22}	100
Q_{12}	49.759
R	25
a	-1.99759

Table 11.
Parameters.

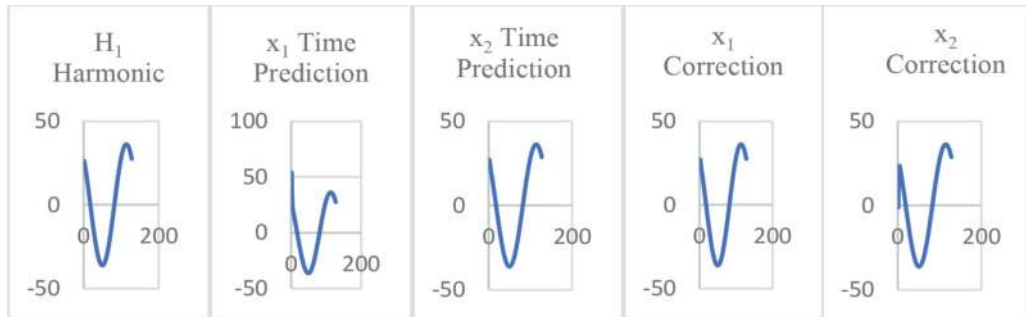


Figure 12.
Complete data, H_1 harmonic, predicted and corrected states.

$x_1(0)$	$x_2(0)$
Initial	Initial
231.0068	0
26.42065	-1.20707
13.48622	-0.81265
4.1098	1.643843
6.050971	-0.45685
0.754028	-1.87793
4.844518	-0.7715
2.016297	1.070199

Table 12.
Initial values.

were A_1 and θ_1 are first harmonic amplitude and phase. In **Table 10** we show H_1 KFHO gains from Eqs. (15), (16) for the first seven time samples. The boldfaced values indicate the moment (after 5 time samples) when the gains become constant. These constant gains can be used in the KFHO design for its simplicity. **Table 11** summarizes various filter parameters. The results for predicted and corrected state estimates \hat{x}_1 and \hat{x}_2 for 128 time samples are shown in **Figure 12**. The reference harmonic H_1 data is calculated using standard cosine function and it is the “measurement” as in (9) (**Table 12**).

8.2 Other harmonics (H_2 , H_4 , H_3 , H_7 , H_{10} , H_5)

The other harmonic calculations are done likewise, with specific amplitude and phase used and we will not give the obvious details here. The initial conditions are

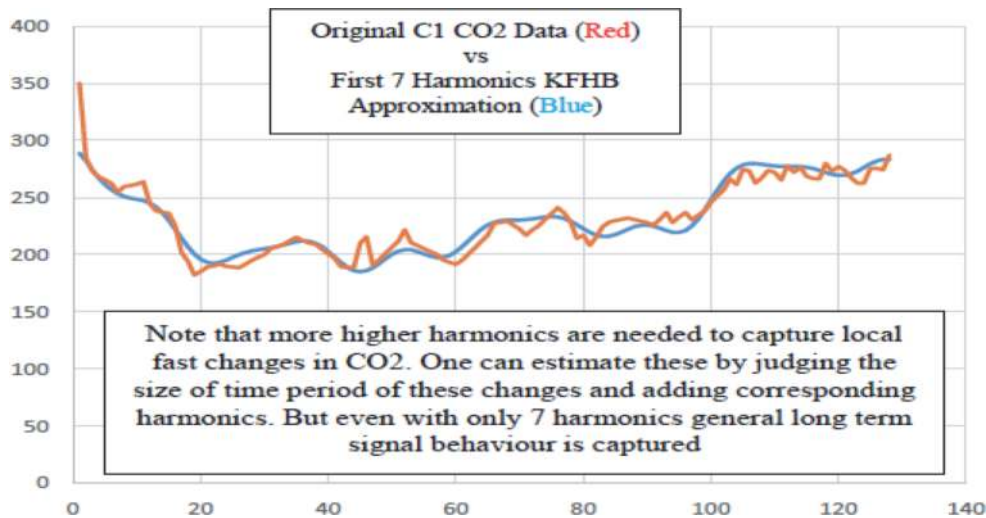


Figure 13.
Comparison of Vostok data vs. approximation with 7 harmonics.

calculated in the equivalent way as given in (31) using different values for amplitude, frequency and the phase angle, using **Table 7** harmonic values. Once calculated, initial conditions drive the KFHO for each harmonic, with the various harmonics KFHO parameters similar to the **Tables 11–12** above for H_1 and the other harmonic signals like in **Figure 12** but with different frequencies and phases. Next all of the predicted and corrected x_1 and x_2 estimates are combined per KFHB output in Eqs. (28) and (29). Resulting values are compared to the known original and inserted Vostok data and appropriate MAPE are generated to check the performance of the method. **Table 9** summarizes resulting MAPE for all 7 harmonics. Note that the results are based on constant filter gains in KFHB similar to the values in **Table 10** which are equivalent to the optimal time varying gains to at least 2nd decimal point. Parameters in **Table 11** were not optimized in any way and they can be tuned further based on errors obtained. This may be done so Q values would be adjusted per specific harmonic amplitude. We made a simple assumption that the standard model deviations and variances are of order of 100 for both states and all harmonics, and cross variance of order of $-a_0 Q_{11}/2$, as noted earlier. Table also indicates what we expected, namely that the corrected errors are smaller than the predicted ones.

Note that the values for y are just specific harmonic amplitudes calculated based on maximum amplitude and phase angle, per Eq. (20). The prediction errors correspond to one sampling interval, in our case 1011 years. This is large enough for long terms calculations and prediction. If we chose to have 2 or more sampling intervals the prediction errors will obviously increase.

As far as corrected errors, they correspond to the situation where we have a specific CO_2 value to use for KFHB correction. For example if we chose to consider a current CO_2 value we can append it to the beginning of the Vostok data, and we actually did that by adding 350 PPM value for CO_2 ‘now’ in front of the ‘newest’ Vostok data, more than 2000 years old. The CPE as we envisioned it allows for all sorts of scenarios, sensitivity analysis, ‘what if’ scenarios for all the variables, CO_2 , temperature, methane, and so on, EPICA and other ice core measurements, both for short term (100+ years) as well as long term (1000 years+). To give a visual impression of Vostok data approximated by first seven harmonics, see **Figure 13** above. Approximation by 7 strongest amplitude harmonics produces very smoothed Vostok data (in blue) without capturing very abrupt changes on a smaller time

scale, as in original Vostok data (in red). This can be improved by addition of more harmonics. The advantage which our CPE offers is its flexibility to treat both natural induced values of CO₂ and other variables as well as modern times human produced effects attributed to global warming.

9. Further considerations

Other cycle contributions can be examined as a base for KFHB based machine learning training and testing. This is important due to different duration of various cycles and the uncertainty about the duration of the ongoing Cycle 1. Then older cycles may carry less useful data for the predictions about the current cycle, especially in a view of current global warming effects due to human activities not present in the climate past. On the other hand the old data carries useful short and long term information which can be used judiciously in the KFHB fine tuning, in particular due to its inherent periodicity. Hence multicycle analysis can be very useful for machine learning implementation of our CPE based on KFHB idea. The minimal choice of harmonics will allow us to devise a reasonably simple machine learning algorithm for training, testing and prediction purposes based on KFHB. For example the longer data sets in C234 can be used as a training data set, as can shorter C23, in order to predict the completion of C1, calculating prediction error E_{1,234} (Error in predicting C1 given C2, C3 and C4 cycles) or E_{1,23} (predicting C1 given only C2 and C3 data). This applies to both temperature and CO₂. This can be repeated for other components in Vostok data set, such as methane, oxygen and insolation. Similarly for the European EPICA data set as well as set of cycles indicated in Milankovich theory [31]. Hence, once we predict C1 using C123 or C23, we obtain errors E_{1,234} and E_{1,23}. Intuitively we can expect that $E_{1,23} > E_{1,234}$, i.e. training based on larger data set ideally would produce smaller test and prediction errors. This has to be confirmed by the further analysis, in particular due to different lengths of the various cycles.

Besides our aim at producing an effective CPE methodology, the harmonic analysis spurred a variety of related thinking and ideas which we also summarize in this paper. Some further ones follow. Climate change on the time scales of the ice cores has been considered as consistent with the IPCC's hypotheses [11], focusing on permanent greenhouse gases, particularly CO₂, methane and nitrous oxide. This has included the role of increasing water vapor but viewed only as a secondary amplifying factor. The effect of temperature on water vapor pressure is shown by the Clausius-Clapeyron equation, dictating an exponential increase in vapor pressure as temperature rises. But the role of water vapor in modern global warming is only considered in GCMs as a derivative of primary warming by permanent greenhouse gases. This may be in error, as modern irrigation is now adding an extra 4% of water to land surfaces from 1960. It is possible to estimate water vapor content of atmospheres of different eras from temperature data. We have also hypothesized a positive forcing from irrigation water [32, 33] in addition to other primary sources of warming such as the Milankovich astronomical cycles. This may prove a more reliable means of correlation using the link already established between water vapor responsible for more than 80% of the air heating.

One feature of some of the ice core analyses is the irregular but rising increasing levels of dust as the planet became colder (**Figure 1**), possibly absorbing rather than scattering solar radiation. From frequency analysis, a marked impulse effect can be recognized, given that the peak of the dust samples in ice cores clearly coincides with the commencement of the interglacials, suggesting a role in initiating the warming process. Dust-climate feedbacks have recently been highlighted as having

a role in the final stages of the glacial cooling process [34]; but we have surmised, the opposite that dust on the surface of snow and ice has decreased the albedo, capturing more radiant heat from the Sun. If true, a process in which a decreasing albedo effect from widespread dust in very cold, dry conditions (similar to the Antarctic now) initiates warming that is accentuated by increasing water vapor can be proposed. The rapid upward change of temperature in the interglacial would be directly consistent with the exponential increase in water vapor with temperature. In that case, the release of CO₂ could properly be seen as an effect of Henry's Law [35], rather than cause.

Consequently, the climate sensitivity for CO₂ may be overestimated. It would be of interest to know the scale of wind velocities in the glacials, since the inertial force of wind motion is opposed to gravity in the case of dust suspension and its carriage to high albedo surfaces. When dust particles are dry and disaggregated, current dust storms carrying particles above are most prominent. Furthermore, we have proposed [33, 36] that the atmosphere can store an order of magnitude more thermal energy because of the additional degree of freedom of motion in vertical motion. That is the motion involved in circulating air in anticyclones-cyclones. In a tropical cyclone, condensing water vapor is considered as providing energy to drive the cyclonic motion, using heat derived from the surface of the ocean. This follows from our finding [12] that configurationally entropy (the inverse of free energy) is a logarithmic function of the physical action, a scalar property related to angular momentum but including the dimensionless angular motion [10]. If confirmed this could mean that weather extremes such as very hot days with greater fire risk could be caused by collisions between anticyclones with extra thermal energy released as heat as the laminar flow of air becomes turbulent.

10. Conclusion

In this paper we have analyzed Vostok ice core data using (i) time correlations, (ii) harmonic analysis, as well as (iii) amplitude and energy consideration, and proposed a (iv) general prediction approach using KFHB methodology. In particular we focused on Cycle 1 of CO₂ data in frequency domain as a representative example. The general approach is to split Vostok data set into 4 smaller sets, as per climate periodicity indicated in the set. The outcome is a choice of set of high energy harmonics for all cycles and any of their combinations for designing CPE based on KFHB which is a linear combination of several individual KFHOs, for effective data prediction purposes. This can be incorporated into a practical machine learning methodology for training and testing, as well as data prediction using collected Vostok or other available climate data sets. We believe that our CPE based approach offers advantages in its simplicity and for short as well as long term prediction abilities via KFHB approach which can produce optimal results in the context of stochastic data set. Several issues remain to be solved, in particular (i) uneven cycle lengths as well as better approach to (ii) non uniformity of ice core data. We are addressing both in our ongoing work. Our analysis also seeks to find evidence regarding causes and results in climate science, an essential requirement for more certainty in weather and climate predictions. We consider that current GCMs have considerable uncertainties in such predictions.

Author details

Migdat Hodzic^{1*} and Ivan Kennedy²

1 American University in Bosnia and Herzegovina, Sarajevo,
Bosnia and Herzegovina

2 Institute of Agriculture, University of Sydney NSW, Australia

*Address all correspondence to: migdathodzic@gmail.com

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Bentley CR, Koci R. Drilling to the beds of the Greenland and Antarctic ice sheets: A review, *Annals Glaciology*, 47, 2010.
- [2] Langway CC Jr. The history of early polar ice cores. *Journal Glacial Science*, 55, 385–396, 2008.
- [3] Fourteau K et al, Analytical constraints on layered gas trapping and smoothing of atmospheric variability in ice under low-accumulation conditions, *Climates Past*, 13,1815–1830, 2017.
- [4] Alley RB, Reliability of ice-core science: historical insights. *Journal Glaciology*, 56,1095–1103, 2010.
- [5] Petit JR et al, Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature*, 399, 429–436, 1999.
- [6] Jouzel J et al, Vostok ice core: a continuous isotope temperature record over the last climatic cycle (160,000 years), *Nature*, 329,403–408, 1987
- [7] https://en.wikipedia.org/wiki/European_Project_for_Ice_Coring_in_Antarctica, Accessed December 26, 2019.
- [8] Barnola JM. Status of the atmospheric CO₂ reconstruction from ice cores analyses: Keynote perspective, *Tellus*, 51, 151–155, 1999.
- [9] Barnola JM et al, Historical CO₂ record from the Vostok ice core. In *Trends: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis, Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.* 2003.
- [10] Kennedy IR, *Action in Ecosystems: Biothermodynamics for Sustainability*, Research Studies Press, Wiley, UK 2001.
- [11] Pierrehumbert RT. *Principles of Planetary Climate*. Cambridge Press, UK 2010.
- [12] Kennedy IR, Geering H, Rose M, Crossan. A simple method to estimate entropy and free energy of atmospheric gases from their action. *Entropy*, 21,451–464, 2019.
- [13] Trenberth KE, Smith L. The mass of the atmosphere: A constraint on global analyses. *Journal Climate*, 18, 864–875, 2005.
- [14] Kennedy IR. *Acid Soil and Acid Rain*. Research Studies Press, Wiley, UK 1992.
- [15] Plath DC et al, The solubility of calcite – probably containing magnesium – in seawater, *Marine Chemistry*, 10, 9–29, 1980.
- [16] Fisher RA. Tests of significance of harmonic analysis. *Proceedings Royal Society A*. 125, 54–59, 1929.
- [17] Hodzic M and Kennedy IR, Time and Frequency Analysis of Vostok Ice Core Climate Data, *PEN* 7, 2, 907–923, 2019.
- [18] Robert A. Michael et al, Temperature Impacts on Dengue Emergence in the United States: Investigating the Role of Seasonality and Climate Change, *Epidemics* 28, 2019.
- [19] Yiou R, Baert E, Loutre MF. Spectral analysis of climate data. *Surveys in Geophysics*. 17, 619–663, 1996.
- [20] Wichert S et al, Identifying periodically expressed transcripts in microarray time series data, *Bioinformatics*, 20:5–20, 2004.
- [21] Trudinger CM et al, Kalman filter analysis of ice core data. *Method*

development and testing the statistics, *Journal Geophysical Research*, 107:4422, 2002a.

[22] Trudinger CM et al, Kalman filter analysis of ice core data 1. Double deconvolution of CO₂ and delta¹³C measurements. *Journal Geophysical Research*. 107:4423, 2002b.

[23] Fumi A et al, Fourier analysis for demand forecasting in a fashion company, *INTECH International Journal Engineering Business Management*, 2013.

[24] Boddy R and Smith G. *Statistical Methods in Practice: For Scientists & Technologists*, Wiley, US 2009.

[25] <https://www.co2.earth/>, Accessed December 26, 2019.

[26] Greengard L, Lee JY. Accelerating the Nonuniform Fast Fourier Transform, *SIAM REVIEW*, Society for Industrial and Applied Mathematics, 46, 443–454, 2004.

[27] Dutt A and Rokhlin V. Fast Fourier Transforms for Nonequispaced Data, *Research Report YALEU/DCS/RR-980*, 1993

[28] Anderson and Moore, *Optimal Filtering*, Prentice Hall, US 1986.

[29] Maarten van Walstijn, *Discretization of the Harmonic Oscillator*, PBASS, 2019.

[30] Cieslinski JL. On the exact discretization of the classical harmonic oscillator equation, *Journal of Difference Equations and Applications*, 17, 11, 2009.

[31] Meyers R, Sageman BB, Pagani M. (2008) Resolving Milankovitch: Consideration of signal & noise, *American Journal Science*, 308, 770–786, 2008.

[32] Kennedy IR, Hodzic M. Vortical action and entropy: A new proposal for the turbulent release of heat at the boundary layer. *AMOS*, Fremantle, 2020.

[33] Kennedy IR, Hodzic M. Designing a regional climate model to test the hypothesis that increasing anthropogenic use of water is a contributor to global warming. *International Association Mathematical Geosciences Proceedings*, Penn State, US, 2019.

[34] Shaffer G, Lambert F, *Proceedings National Academy Science*, 115, 2026–2031, 2018.

[35] Bailey N et al, Henry's Law constant for CO₂ in aqueous sodium chloride solutions at 1 atm and sub-zero (Celsius) temperatures, *Marine Chemistry*, 207, 26–32, 2018.

[36] Kennedy IR, Computation of planetary atmospheres by action mechanics using temperature gradients consistent with the virial theorem. *International Journal of Energy Environment*, 9, 2308–1007, 2015.