Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 7, No. 2, August 2019, pp.907-923
Available online at: http://pen.ius.edu.ba

Time and frequency analysis of vostok ice core climate data

Migdat Hodzic[1] and Ivan R. Kennedy[2]

[1] American University in Bosnia and Herzegovina, Sarajevo, Bosnia and Herztegovina
[2] University of Sydney, Sydney, NSW, Australia

Article Info

Received May 5, 2019

Keyword:

Vostok Data
Time Analysis
Frequency Analysis
Sampling Period
Signal Energy
Machine Learning

ABSTRACT

The periodicity of Vostok ice core climate temperature and gas concentration data indicate inherent long term past regularity of Earth's climate, with a period of around 100,000 years, warming around 15,000 and cooling of around 85,000 years. At this point we are at the top of one of the warming periods. Vostok data cover around 430,000 years, ie 4 climate cycles (warming-cooling), of similar but not quite the same duration. In this paper we perform a detailed time and frequency analysis of these data for each of the cycles as well as their various combinations, including a full tested period of 430,000 years. Time correlation analysis allows for more accurate time lag estimate in each cycle already noted between temperature change and carbon dioxide content. We estimate these lags to lie between 1000-2500 years, longer than previously concluded. On the frequency side we perform Fast Fourier Analysis and identify full spectrum of harmonics for various cycles, and then perform energy analysis to identify which of the harmonics contributes the most. The idea is to reduce the computational load for further modeling and analysis using Kalman Filter based prediction method. Once the prediction model is defined (a follow up paper) data will be split into two segments, Learning and Testing, in preparation of a Machine Learning fine tunning methodology. We can use last three or last two or even just last cycle to learn on, and then the current on going cycle to test on. This will result in real time prediction of relevant climate data. Assuming causal time regularity, more of these cycles are employed in training, more the prediction error for the next cycle should be reduced. Hence it is critical to perform very detailed time and frequency analysis of Vostok data as a solid data base for the prediction model to follow.

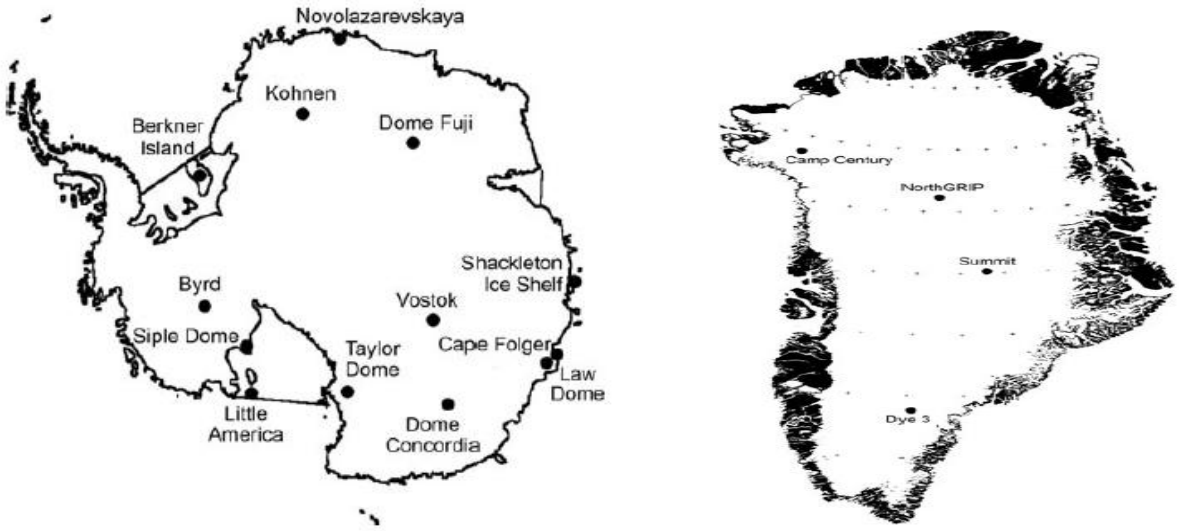
Corresponding Author:

Migdat HODZIC,
AUBIH, Sarajevo, Bosnia and Herzegovina
Email: migdathodzic@gmail.com

1. Introduction

Extensive climatic data on the past four ice ages and beyond is available from various studies, commencing from mid 1950s until now. Figure 1 shows 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 in the two areas. Currently, intensive ice core drilling is being conducted in other areas as well, so an even larger data set is anticipated. Reference [3] and [12] describe the history of ice core drilling in detail. Our purpose in this paper is to employ Vostok Station ice core data for a variety of time and frequency related analyses. The Vostok ice core data set includes derivations of relative temperature, carbon dioxide (CO_2),

methane (CH_4), oxygen and solar variation (insolation), during the last 430,000 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 [15]. Reference [10] presented the first record to span and full glacial-interglacial from an ice core drilled in the Russian Vostok station in Antarctica. A 430,000 year record was constructed from the Petit et al. study on a 3 km deep core of ice (Fig. 2). Another source of similar ice core data is available from European EPIC drilling project (European Project for Ice Coring in Antarctica) which lasted from 1998 until 2005 (Fig. 3). In this paper we focus on specific analysis related to only two Vostok data variables, namely relative temperature and CO_2 content. Data used are as corrected in [4] and [5]. The variation of atmospheric CO_2 and temperature is shown in Fig. 2. Although originally thought that the CO_2 data might be considered as proof of its causal role in global warming, it is now widely considered that CO_2 lags temperature change and its lower rate of solution in and release from sea water is more likely the cause of the relationship.



.

Fig. 1 Ice Core drilling sites in Antarctica and Greenland. Note that two maps are not to scale, for Greenland is less than 1/6th of the size of Antarctica.

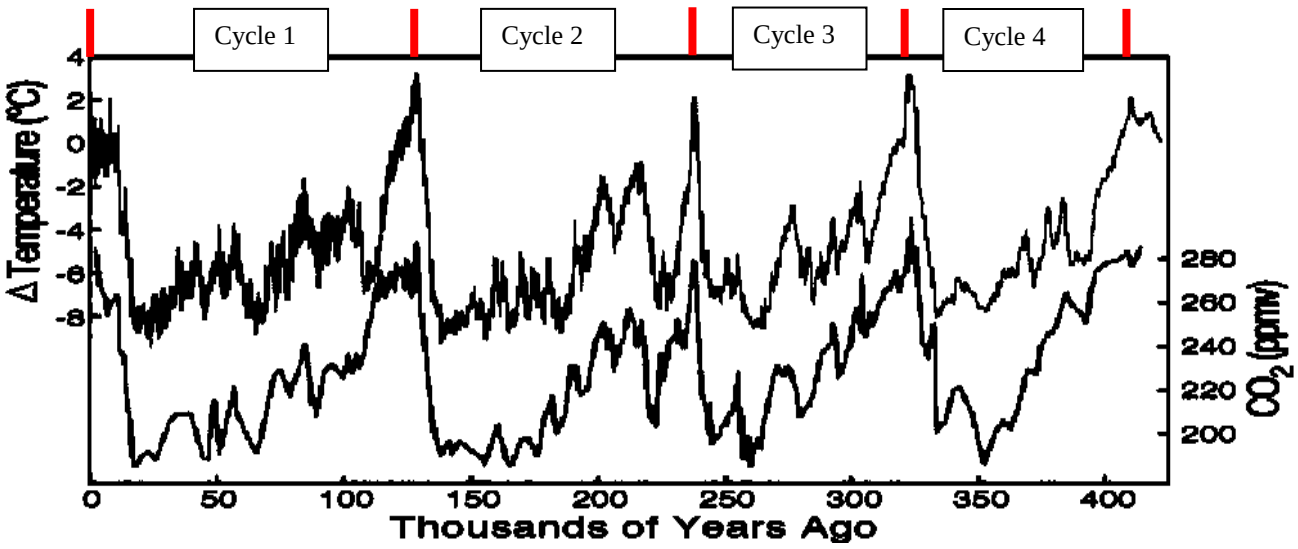


Fig. 2 Vostok data variation in global relative temperature and CO_2 content (Barnola at al. 1999, 2003)

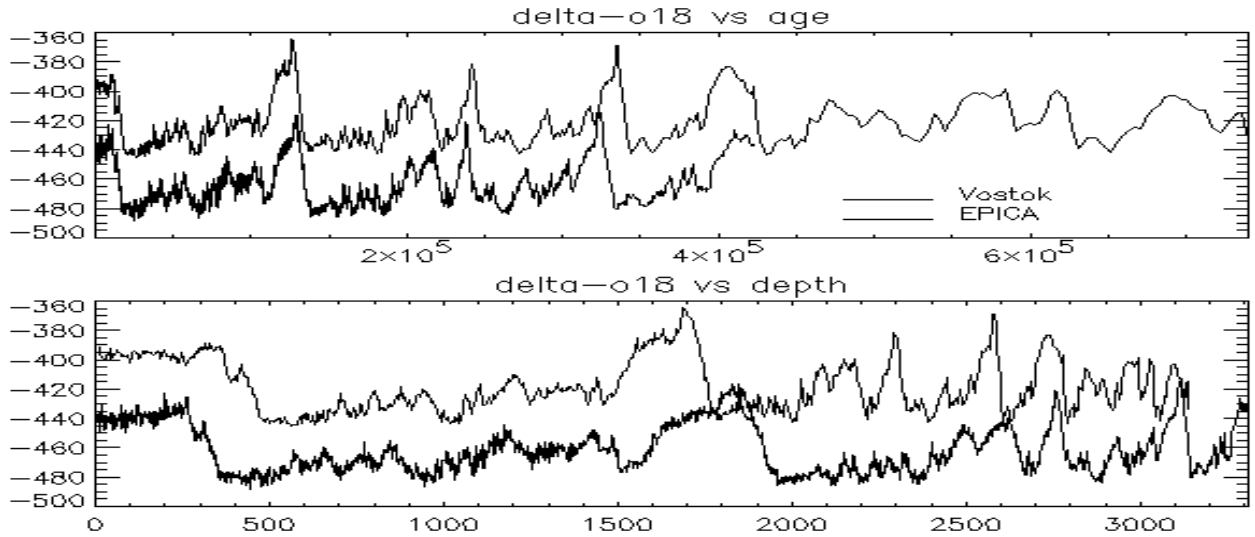


Fig. 3 EPICA ice core temperature and CO_2 data compared to Vostok data; Vostok data from source: http://www.usgcrp.gov/usgcrp/images/Vostok.jpg

2. Spectral Analysis Methodology

The famous statistician Fisher [7] was an early exponent of testing statistical significance in harmonic analysis. Some preliminary spectral analysis has been conducted on Vostok ice core data set in [21] but without major impact. Here, spectral analysis in R as described in [18] and [20] was applied in [4] and [5] for their data set. In such analyses, time series are decomposed into underlying sine and cosine functions into the most important different frequencies. Various texts on fast Fourier transforms were also sources for frequency determination. These approaches allow construction of periodograms quantifying the contributions of of the individual frequencies to the time series regression. The GeneCycle R Package [19] developed for bioinformatics has general application for time series and was employed in this study. Some recent time series analysis research deals with data classification as well as data series use in environment, production and financial research [24],[25],[26]. In our paper we perform our own Vostok data analysis using R and Excel, in both time and frequency domains. Our aim is to use it for Machine Learning training and testing to follow. We analyzed all four Vostok cycles, C1, C2, C3 and C4 as indicated in Fig. 2 above. The cycles are determined by locating maximum values for relative temperature and CO_2. Becuase of a lag determined between relative temperature and CO_2 content the number of data points is slightly different in two data sets. The overall number of data points for both variables is 363. Individual cycles differ in data points slightly. Table 1 summarises the number of data points in each cycle. Note that the number of data poiunts is not the same for each cycle. This will have important effect on our Machine Learning method and it has to be addressed properly. Also note that the cycle C1, the current cycle we live in, is still evolving and new data may be added if required. This data may be skewed by „global warming effects“ which did not exist in previous cycles. Comparison of two data sets may be very useful in estimating global effects of temperature and CO2 content. In addition to the above, the uneven data distribution (non uniform data sampling times within each separate cycle) may pose some numerical issues as well for the analysis in the context of Machine Learning prediction methodology. This will be dealth with in our follow up paper. There are various methods to deal with this as described in [22] [23]. One of the simpler methods, yet effective one, is to „fill in“ missing data by some approximation method. In any case in this initial paper no additional data insertions are done for smplicity. Our aim is to gain more insight into Vostok data in time and frequency domains, and check the corresponding „energy“ content in order to reduce number of significant harmonic components which will reduce complexity of our Machine Learning prediction model.

Table 1 Number of data points for each cycle

No of Data Points	Cycle C1	Cycle C2	Cycle C3	Cycle C4
Temperature	74	136	99	54
CO_2	80	135	99	49

3. Data Analysis of Vostok Ice Core Data

3.1 Data cycles

In this Section we summarise the ice core data analyses. Figs. 4-7 show data for relative temperature and CO_2 for each cycle, analyzed individualy as far as time correlations and also harmonic and energy content.

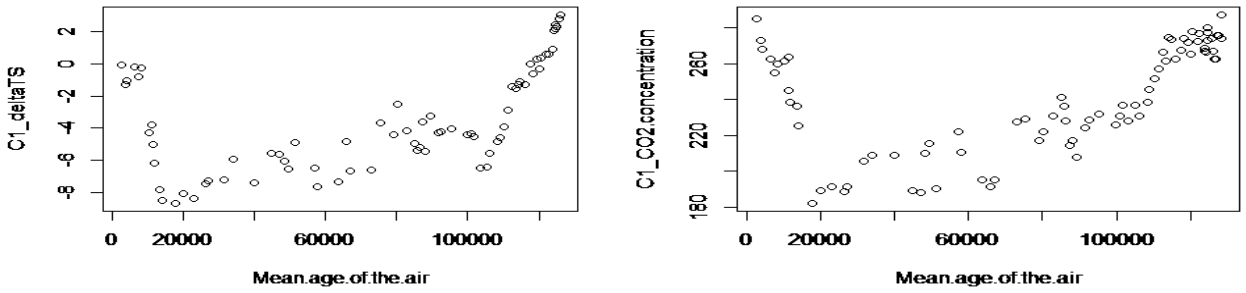


Fig. 4: Cycle C1 relative temperature and CO_2 content

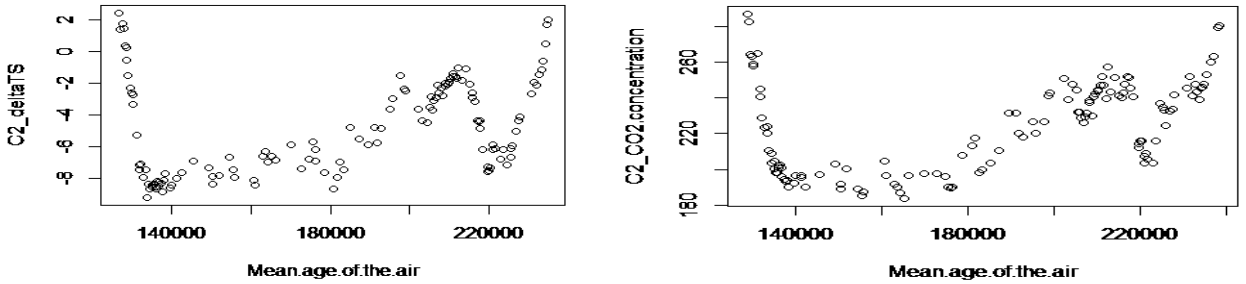


Fig. 5: Cycle C2 relative temperature and CO_2 content

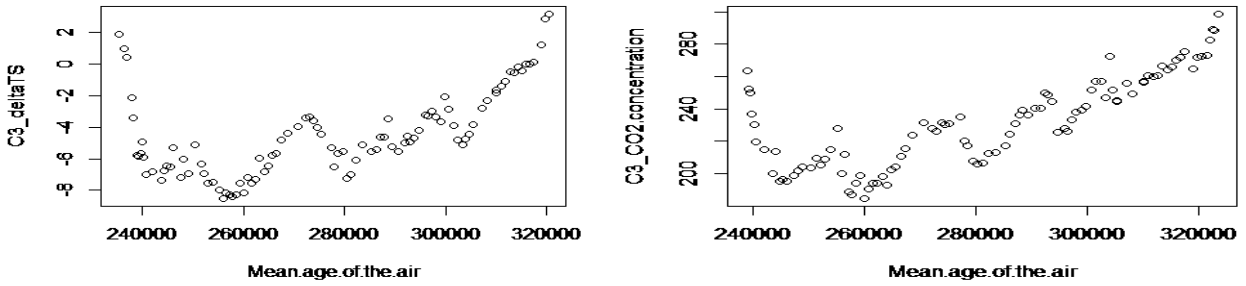


Fig. 6: Cycle C3 relative temperature and CO_2 content

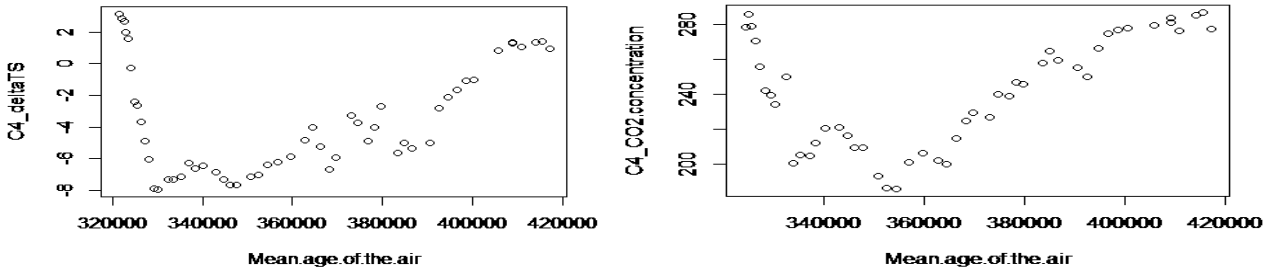


Fig. 7: Cycle C4 relative temperature and CO_2 content

3.2 Duration and sampling times in cycles

As Table 1 indicates the number of data points for each cycle is quite different. This has to be taken into account when analysis is performed. For example some data will not be used in cross correlation calculations between temperature and CO_2. The number of such data is small and will not affect the analysis very much. From Figs. 4-7 we can note general similarity between various cycles data sets. The difference is in the length of each cycle in thousands of years. Table 2 summarizes approximate duration of each cycle based on our analysis (Section 3.4).

Table 2 Cycles approximate duration in years

Duration in Years	Cycle C1	Cycle C2	Cycle C3	Cycle C4
Temperature	127,726	115,156	86,462	96,782
CO_2	128,399	109,800	86,148	95,587

Note that CO_2 lags relative temperature in C4, C3, and C2 (Table 2). In C1 the data may be a little skewed due to number of recent data points where maximum values are not clearly identified in Vostok data (left most CO_2 data). Another feature to keep in mind is very different time differences when the data is obtained from ice core readings. In some cases difference between two data samples are only few years, or 400-500 years, but in some as much as 5,000 years. This adds to the imprecision of any analysis. For each cycle we calculated average time sampling values as summarized in Table 3. In our follow up paper this will be corrected.

Table 3 Average data sampling time in years for each cycle

Average Sampling Time in Years	Cycle C1	Cycle C2	Cycle C3	Cycle C4
Temperature	1,703	800	865	1,792
CO_2	1,605	813	862	1,911

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 Machine Learning approach all other possible data samples have been also created, i.e. any combination of available cycles not necesarily in time order as they transpired. The idea is to add to the richness of all available data for training purposes of a Machine Learning approach. For example Table 4 indicates respectively average sampling times for Cycle C1 and C2 combined (noted as C12), Cycle 2 and C3 combined (C23), as well as Cycle 2, Cycle 3 and Cycle 4 combined (C234), plus total C1234. The total number of data points, duration and can be easily obtaind from Tables 1 and 2. Note from Table 4 that more cycles we add the more uniform data sampling times become.

Table 4 C12, C23, C234 and total C1234 average data sampling times in years

Average Sampling Time in Years	Cycle C12	Cycle C23	Cycle C234	Cycle C1234
Temperature	1,087	827	1,008	1,149
CO_2	1,108	834	1,020	1,149

3.3 Time correlation analysis

Time correlation analysis produces a variety of useful information about periodicity and correlation strenght among data samples of a given quantity. In particular autocorrelations produce the measure of self correlation of a data series. Standard definition of the autocorrelation for a data time series x(t), t = 1,2,...,N, such as relative temperature or CO_2 content, is given as:

$$R_{xx}(m) = \sum_t x(t)x(t-m),\ t = 1,2,...,N \quad (1)$$

where m stands for the lag (delay), and m=1,2,...,N-1. Standard definition of crosscorrelation for two discrete time data series x(t) and y(t), such as relative temperature and CO_2 concentration in our case, is given as:

$$R_{xy}(m) = \sum_t x(t)y(t+m),\ m= -N+1,...,-2,-1,0,1,2,...,N-1 \quad (2)$$

Here m stands fot the lag (delay) and t =1,2,...,N. The above formulas are implemented in various ways depending on the software tool used, such as R, Matlab or Python. In our work reported in this paper we used mostly R for coding and analysis. One property of crosscorrelations 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 i.e.:

$$\tau_{delay} = \arg\max_m R_{xy}(m) \quad (3)$$

Fig. 8 illustrates short term calculations for autocorrelations for relative temperature and CO_2 (first two diagrams) as well as the corresponding cross correlation between the two variables (third diagram) for the entire Vostok data data set. We can read the value of τ_{delay} between temperature and CO_2 as approximately equal to 2 lag units. Since the calculation is done for the entire Vostok data set, from Table 4 we can read an average sampling time for C1234 as 1,149 years, hence we can make an upper limit approximation of the size of the delay to be:

$$\tau_{delay} < 2 \text{ times } 1{,}149 \text{ years} = 2{,}298 \text{ years} \quad (4)$$

One needs to note here that the calculation is not very precize primarily due to very non uniform dsitribution of Vostok data. We will address how to harmonize these data in our subsequent work. The point is that the delay is of order of 2,000 years and not of 100 or 200 years. That is an important finding which can influence our thinking about global climate change caused by natural processes that happened prior to CO_2 increase caused by human actions. Instead of calculating cross correlation point by point, one can also calculate correlation coefficient, which are single numbers and they can be used as a simple measures of cross correlation intensity between two variables. There are several coefficients named after their inventors such as Spearman, Pearson and Kendal [6] and they all indicate certain statistical properties which tie two data series together. Table 5 summarizes standard cross coefficient (x) between relative temperature and CO_2 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 0.82 for the entire set.

Table 5 Cross correlation coefficient for individual and entire cycle

Cross Correlation Coefficient	Cycle C1	Cycle C2	Cycle C3	Cycle C4	Entire Cycle C1234
Temperature, CO_2	0.83885075	0.861018891	0.876113	0.744963	0.82097999

Continuing with our analysis, Fig. 9 indicates entire (long) term auto and cross correlation. It is clearly visible that we are dealing with a periodic process with non uniform periods. This is also confirmed in harmonic analysis in Section 5. In order to determine average time delays between relative temperature and CO_2 for individual we look at Figs. 10-13. The first two diagrams (in Fig. 10 part a) are autocorrelations and they also indicate certain periodicity within the each cycle but obviously not as much on the entire data set. Again this will be confirmed with our harmonic analysis. The third (fourth in Fig. 10) diagram shows short term and long term cross correlation, and they indicates periodicity between two varaibles. The time delay can be read from short term cross correlation. For Cycle 1 it is less than one lag but more than zero lag, we can estimate it at a bit less than half of one lag. From Table 3 for both relative temperature and CO_2 the average data sampling times are 1,703 and 1,605 years which puts the delay at around 800 – 850 years. Similar calculations can be done for other cycles as well. To get a more precise approximation we would need more data and finer resolution around the zero lag where the cross correlation is maximum. It is also worth noting that that the maximum values of autocorrelation and cross correlation indicate average energy in the data series itself or between the two series. Next to each harmonic there is an indication (in parenthesis) of harmonic. For example 1 is the basic one, 2 indicates half of it (in years), and so on. For example if the 1st harmonic is 75,000 years, the second will be half of that, i.e. 37,500 years, 10th harmonic will be equal to 7,500 years, and so on. Considering frequency values, 75,000 years corresponds to 1/75,000 = 0.000013333 = 1.3333×10^{-5} Hz. For the entire Vostok data set some of the harmonics, per our anlysis in Section 5, are of order of 375,000 years which corresponds to 1/375,000 = 0.0000026666 = 2.6666×10^{-6} Hz. Typically, first several harmonics carry most of the energy, as given in Section 4 on energy considerations.

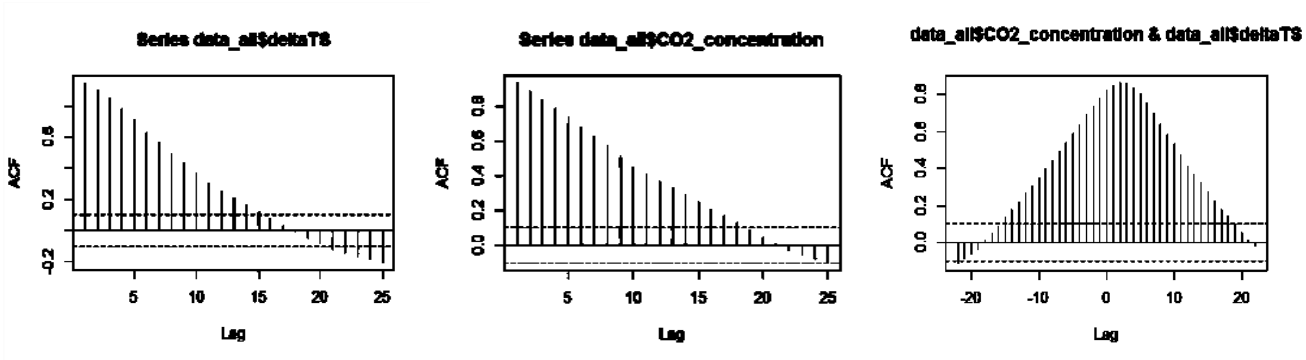


Fig. 8 Short term temperature, CO_2 autocorrelations and cross correlation

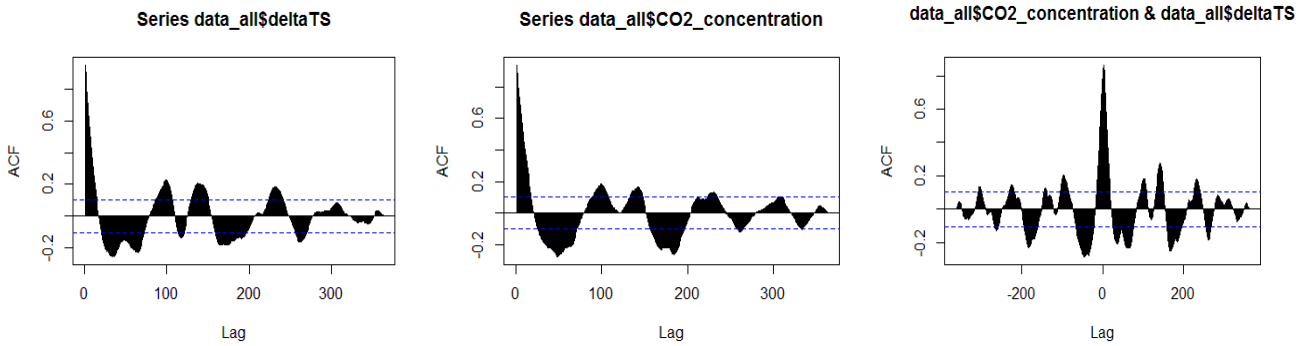


Fig. 9 Total long term temperature, CO_2 autocorrelations and cross correlation

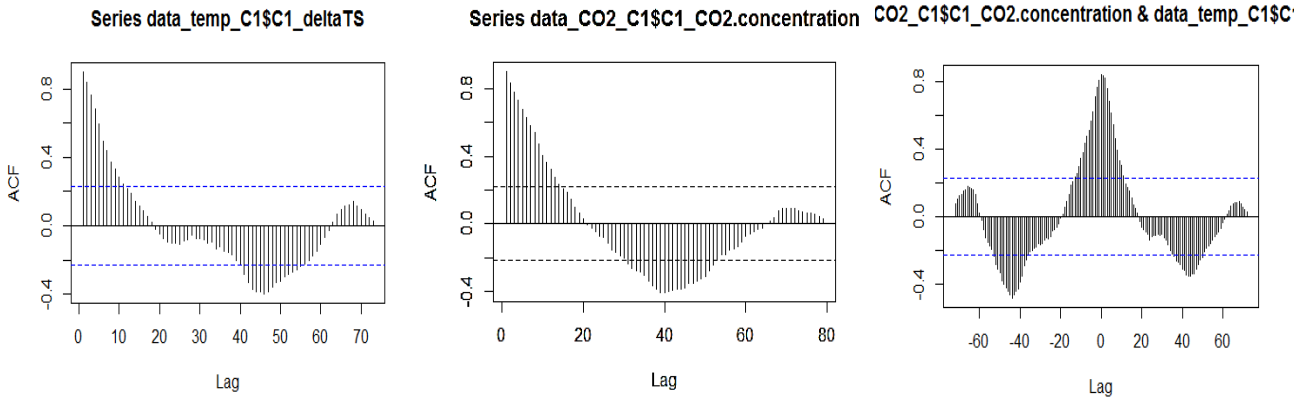


Fig. 10 Cycle C1 temperature, CO_2 autocorrelations and cross correlation

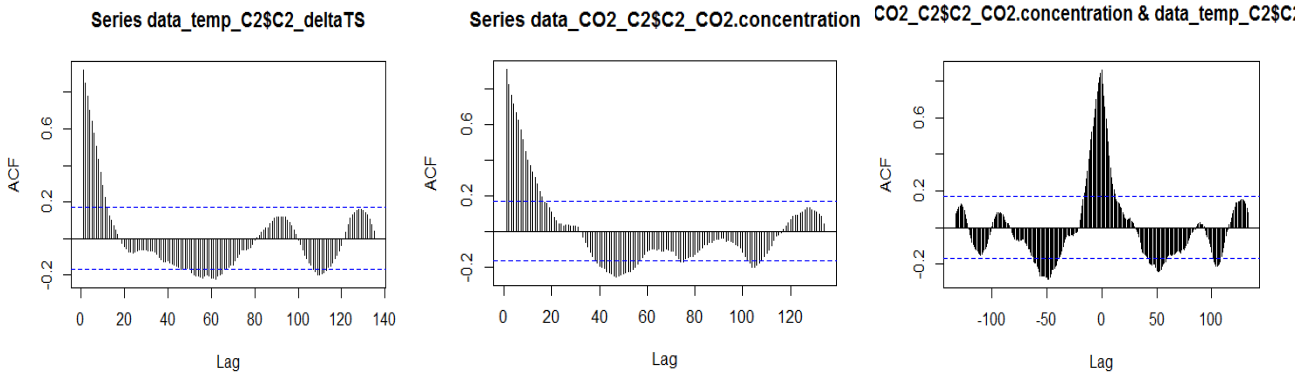


Fig. 11 Cycle C2 temperature, CO_2 autocorrelations and cross correlation

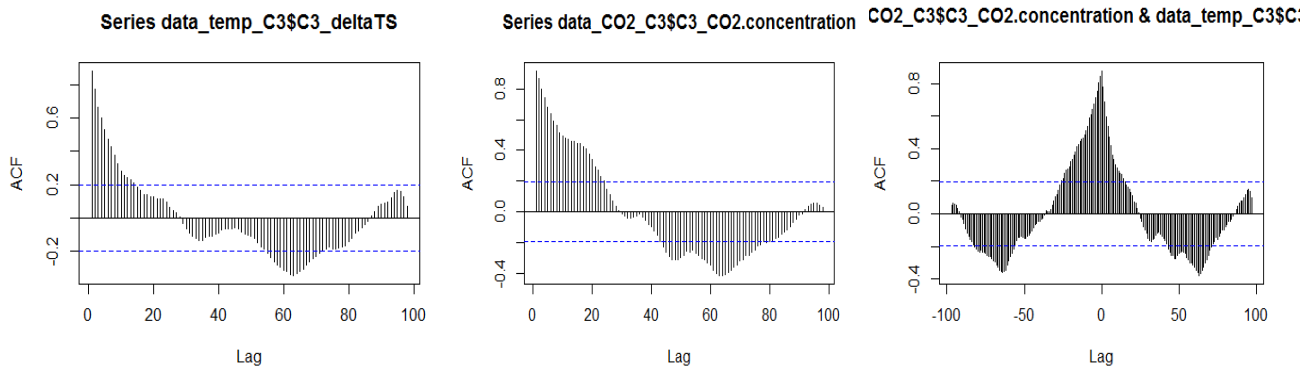


Fig. 12 Cycle C3 temperature, CO_2 autocorrelations and cross correlation

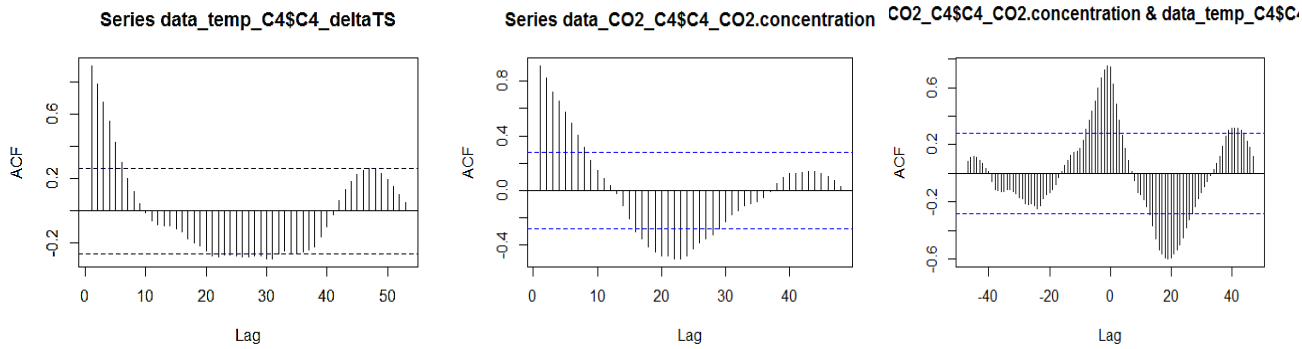


Fig. 13: Cycle C4 temperature, CO_2 autocorrelations and cross correlation

3.4 Harmonic Analysis

The key item in our Machine Learning Vostok ice core data preprocessing summarized in this paper lies in both time correlation analysis (Section 3.3) as well harmonic analysis in this Section. The methodology used is based on calculating FFT (Fast Fourier Transform) using an R language tool and then recalculating to real frequencies. The „real“ frequencies (harmonics) are approximations due to a non uniform data points distributions. There are methods described in literature to deal with certain non uniform data sets, but we did not pursue that in this paper. Our approach was to use an average data sampling due to a wide sampling time variability in Vostok data set. Our follow up work addresses nonuniformity of data with filling for missing data. Figs. 14-18 indicate FFT diagrams for both relative temperature as well as CO_2. The frequencies on horizhontal axes are normalized FFT frequencies which need to recalculated to correspond to real frequencies assuming correctness of average sampling times. We also included data sets for individual cycles to have a feel for their variability within each cycle. Note that vertical lines indicate Fourier coefficients which indicate energy level of a specific harmonics. It is obvious from all FFT diagrams that the energy is concentrated in very small frequency range which has several harmonics carrying majority of the energy content of a specific sdata set. Energy analysis is presented in Section 4 with an indication for each cycle or combination of cycles as to the energy distribution across various harmonics. This distribution is a base for making a reasonable approximation in our Machine Learning approach and it will be used to train and learn from past data to predict the future ones. In this case „future“ refers to both actual past data which will be used to test how the „training“ is sucessfull as well as real future of new data sets to be predicted by our methodology. In order to calculate real frequencies from the FFT ones indicated in Figs. 14 – 18 (obtained from data in Figures 4-7) we perform the following simple calculation:

$$f_R = f_{FFT} \text{ times } T_s \qquad (6)$$

where f_R is a real frequency in Hertz, f_{FFT} is normalized FFT frequency and T_s is an average sampling time, different for each different cycle scenario. Normalized FFT frequencies go from zero up to 0.5, and FFT algorithm all produce number of normalized frequencies which correspond to half of the number of data

points available for analysis. The time periods which correspond to each individual harmonics are calculated by simple inverse operation, i.e.

$$T_R = 1/f_R \quad (7)$$

FFT assumes that number of data points is always a power of 2, such as $2^6 = 64$ data points, or $2^8 = 256$. If data set does not have precise power of 2 data points, then certain number of zeroes are appended to the end of the series to make up to power of 2. In our case, since the total number of data points is 363, and the next higher power of 2 above 363 is $2^9 = 512$, a total of 149 zeros are appended to the end of both relative temperature and CO_2. This is all taken care of by R statistical software tool. As an example Table 6 shows the first five most energy containing harmonics for temperature and CO_2 for the entire Vostok data set as well as for four individual cycles. Due to very small frequency values, Table 6 shows the corresponding time duration of one period of each harmonics per formula 7 (harmonic wave length). This is also more natural to visualize and associate to the number of (thousands of years, as well as to the each cycle duration. The numbers are rounded for simplicity. Several observations can be made from Table 6. First, it is not always the case the first harmonic is the most „energetic". For example for the entire Vostok data set cycle (428,000 years approximately) for relative temperature (Table 6), the most energy is carried in the 5th harmonic followed by the 3rd, whereas for CO_2 it is vice versa, and so on. More is described in Section 4 on energy considerations. Second oof all, in most cases both temperature and CO_2 carry very similar, few times equivalent harmonics, but not necessarily with the same enrgy. We also note from the Figs. 14 – 18 that longer the cycle or cycle combinations are, more harmonics are present which carry considerable energy in the total signal. Another observation is that, for example, for Cycle 4 temperature shows only two very significant harmonics and CO_2 only one. That could due to lack of data points in Vostok set (Table 1), as there are only around 50 data points collected, whereas for other cycles it is 100 or more, for Cycle 2 and Cycle 3, and around 75 for Cycle 1. It is worth nothing that Cycle 1 is still ongoing and more data in the future will be collected. Plus it is not clear if we are on top of Cycle 1, it will take many years to test this. One of important items for future research is to try to identify known climate related events, including Milankovitch Cycles of certain duration which may coincide with the harmonics in our analysis.

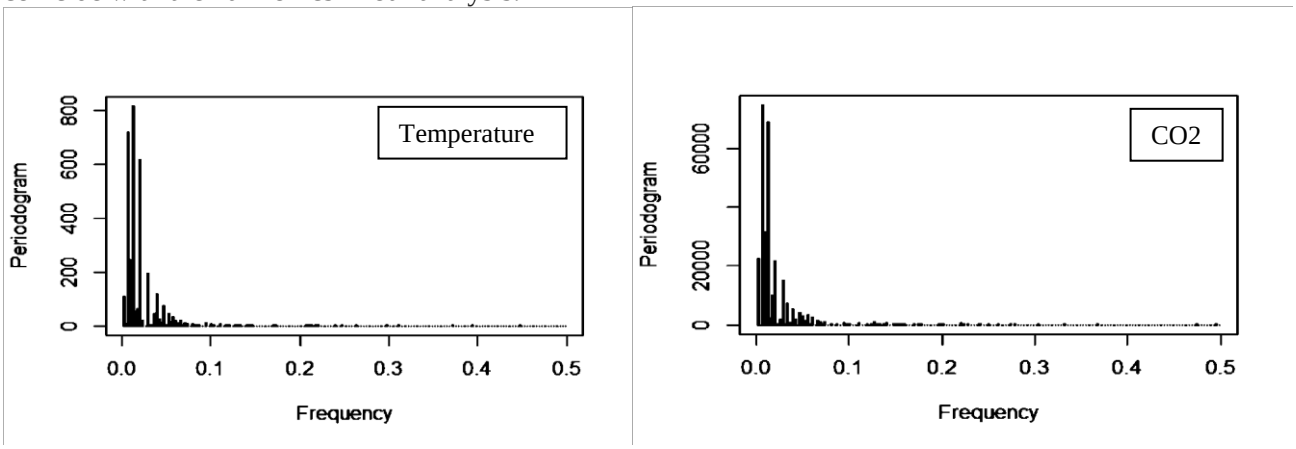


Fig. 14 Full Vostok ice core temperature and CO_2 harmonic content, FFT normalized

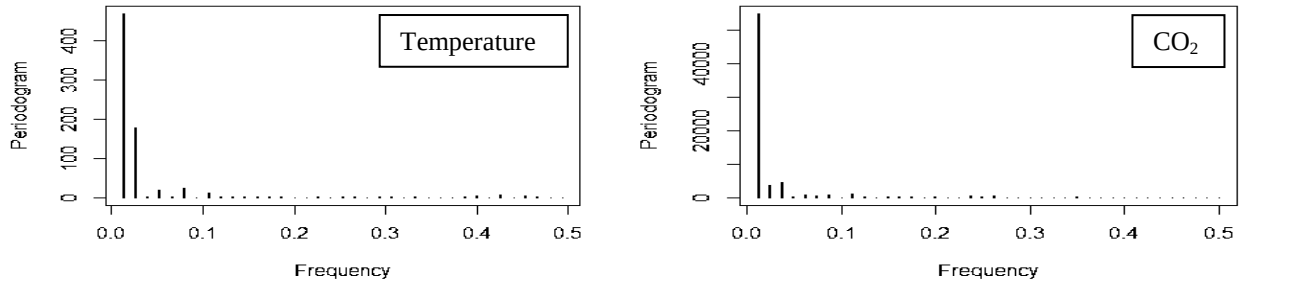


Fig. 15 Cycle 1 temperature and CO_2 harmonic components, FFT normalized

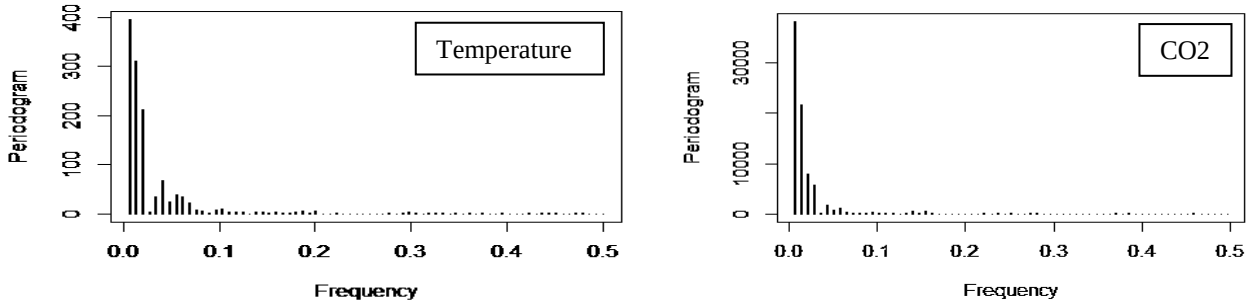


Fig. 16 Cycle 2 temperature and CO_2 Harmonic components, FFT normalized

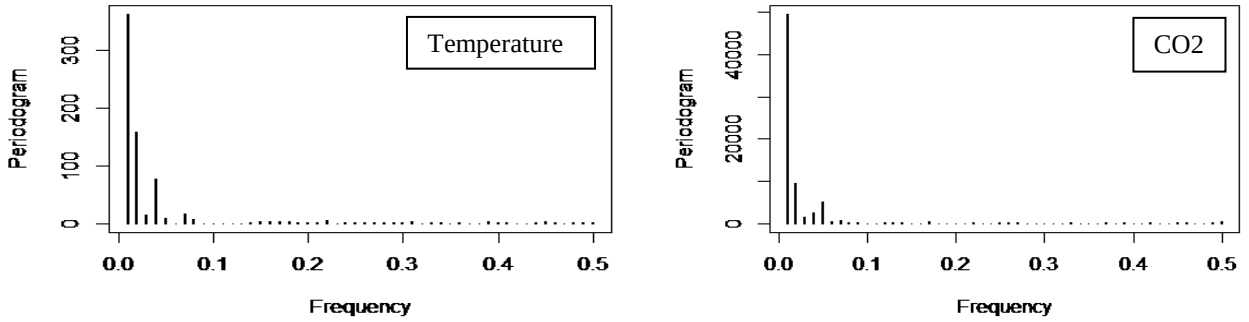


Fig. 17 Cycle 3 temperature and CO_2 harmonic components, FFT normalized

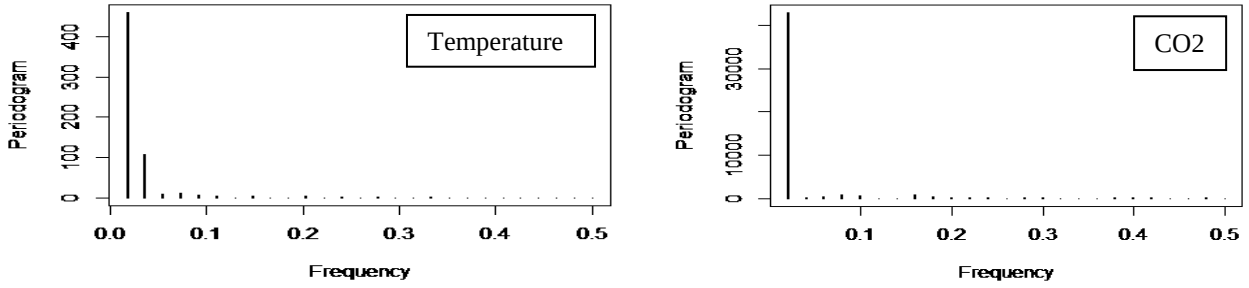


Fig. 18 Cycle 4 temperature and CO_2 harmonic components, FFT normalized

Table 6 First five harmonics for individual and entire Vostok ccyles

Harmonics (in years)	Cycle C1	Cycle C2	Cycle C3	Cycle C4	Entire C1234
Temperature	127,726 (1)	115,156 (1)	86,462 (1)	96,782 (1)	86,175 (5)
	63,863 (2)	57,578 (2)	43,231 (2)	48,391 (2)	143,625 (3)
	21,288 (6)	38,385 (3)	21,615 (4)	24,196 (4)	53,859 (8)
	31,932 (4)	19,193 (6)	12,352 (7)	32,261 (3)	107,718 (4)
	15,966 (8)	14,394 (8)	28,821 (3)	19,356 (5)	39,170 (11)
					28,725 (18), 430875 (1)
CO_2	128,399 (1)	109,800 (1)	86,147 (1)	95,587 (1)	143,625 (3)
	42,800 (3)	54,900 (2)	43,074 (2)	11,948 (8)	86,175 (5)
	64,200 (2)	36,600 (3)	17,229 (5)	23,897 (4)	107,719 (4)
	14,267 (9)	27,450 (4)	21,537 (4)	19,117 (5)	430,875 (1)
	18,343 (7)	18,300 (6)	28,716 (3)	31,862 (3)	53,859 (8)

Figures 19 – 23 have the results of harmonic analysis of selected cycle combinations.

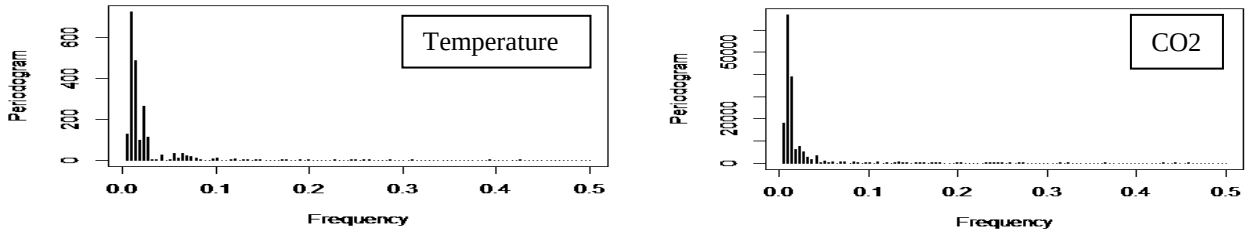


Fig. 19 Cycles 1 and 2, temperature and CO_2 harmonic components, FFT normalized

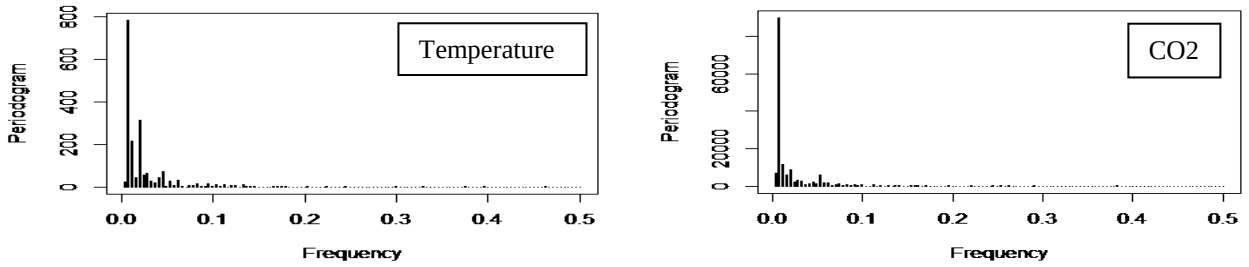


Fig. 20 Cycles 2 and 3, temperature and CO_2 harmonic components, FFT normalized

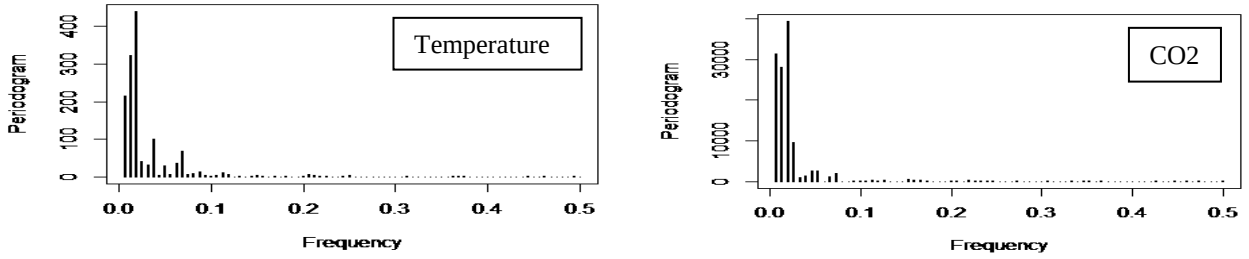


Fig. 21 Cycles 3 and 4, temperature and CO_2 harmonic components, FFT normalized

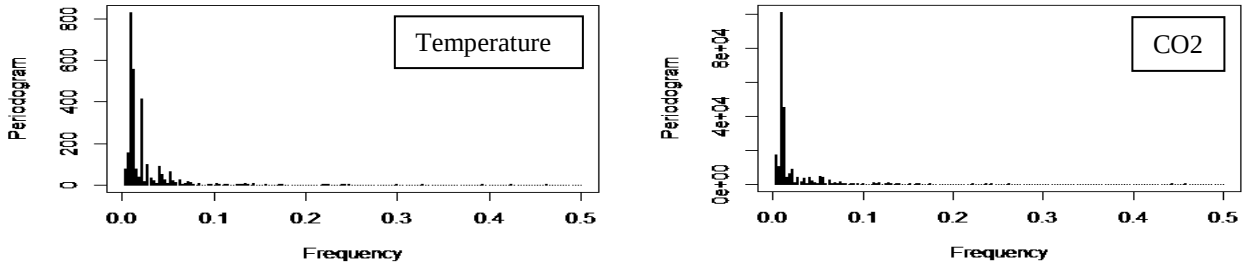


Fig. 22 Cycles 1, 2 and 3, temperature and CO_2 harmonic components, FFT normalized

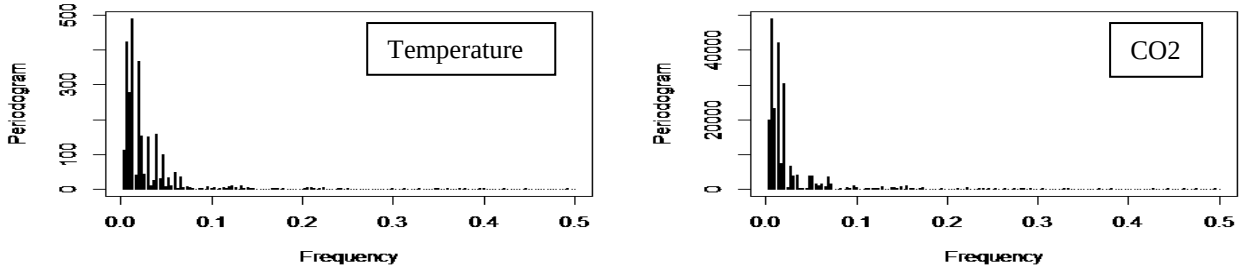


Fig. 23 Cycles 2, 3 and 4, temperature and CO_2 harmonic components, FFT normalized

3.5 Preliminary Time_Frequency Analysis

To conclude this Section we also show very preliminary results on simultaneous time-frequency analysis of Vostok ice core data using Short Term Fourier Transform (STFT). The idea is to se the results for short and long term predictions for both relative temperature and CO_2 via so called Uncertainty Principle which relates resolution in time (short term) vs frequency (long term). One version of the Principle [9] reads as:

$$\Delta t \, \Delta f \geq 1/4\pi \tag{8}$$

where loosely speaking Δt and Δf represent essential time and frequency range around a certain moment in the data series time. There are other versions of the Principle as well. This principle resembles classis Heisenberg Uncertainty Principle in physics and comes from very similar mathematical analysis [9]. Figs. 24 – 27 (normalized frequency vs time) show results for STFT for Cycles 1 through 4, for temperature (first figure), CO_2 (second figure) and cross spectrum between the two (third figure). Darker areas indicate higher harmonic energy. Note dark areas in cross correlation are concentrated around certain time and indicate CO2 delay with respect to temperature (for example longer for C1 compared to C4). Note also more energy (dark area) on the right of second diagram in Figure 24 (CO2) comapred to the first diagram Figure 24 (temperature). Compare this with Figure 4 earlier indicating lots of CO2 data „activity" vs much less for temperature towards the end of the C1 record. These kinds of considerations can be used to further fine tune our algorithm for short and long term training and prediction. As stated in Abstract we are working on Kalman Filter Predictor as a base of Machine Learning prediction methodology which we will report in a follow up paper.

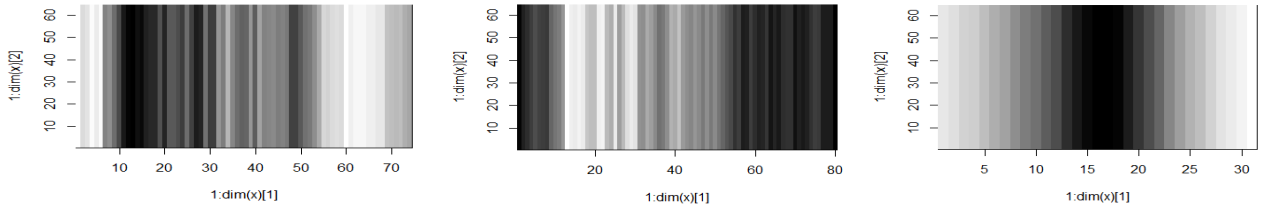


Fig. 24 Cycle 1 Temperature, CO2 and cross harmonics, using STFT

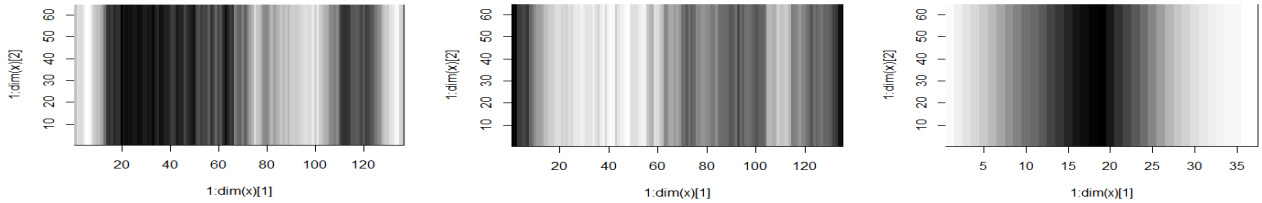


Fig. 25 Cycle 2 Temperature, CO2 and cross harmonics, using STFT

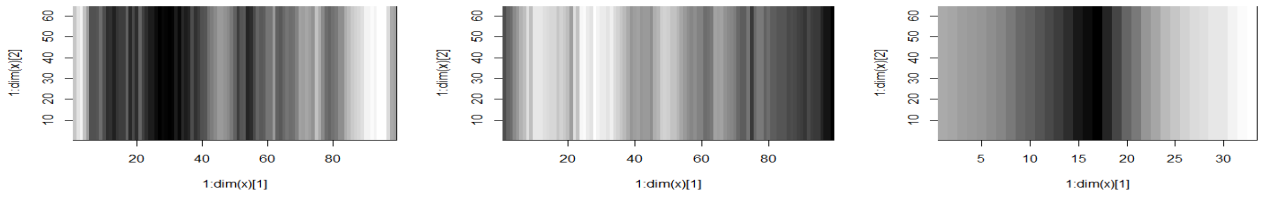


Fig. 26 Cycle 3 Temperature, CO2 and cross harmonics, using STFT

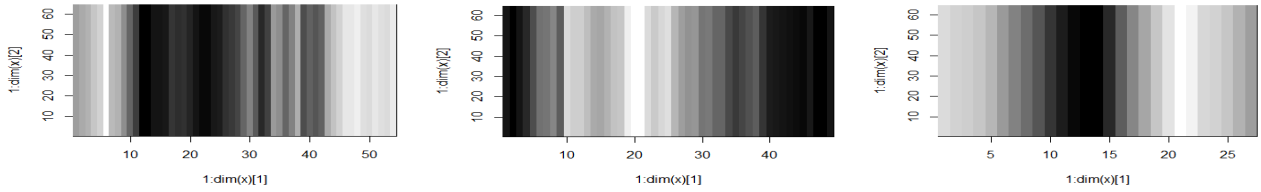


Fig. 27 Cycle 4 Temperature, CO2 and cross harmonics, using STFT

4. Energy Consideration

This section summarizes Vostok ice core data energy analysis. The idea is based on choosing only the highest energy harmonics to simplify future Machine Learning algorithm development. For this we set, as an example, a limit that any number of harmonics need to add up to more than 90% of the total data set energy for each cycle or any of their combinations. We expect that this will result in a considerable reduction of required harmonic components for further inclusion in any training and testing approaches. The FFT algorithm in R produces a set of Fourier transform coeffficients which indicate energy level of each harmonic. They correspond to Table 6 harmonics for each cycle as well as the entire Vostok data set. The values are higher for CO_2 compared to relative temperature due to larger values of original data points. As mentioned earlier, these values are Fourier Transform coefficients. The total energy in the data set can be me3asured using harmonic components or using time data samples. Per Plancherel Theorem (or Parsseval) both approaches produce the same energy content, and for FFT and DFT (Discrete Fourier Transform) this is stated as in:

$$\text{Energy} = \sum_n |x(t)|^2 = (1/N) \sum_k |X(k)|^2), \text{ n and k} = 0,1,...,N-1 \quad (9)$$

where X(k) is the DFT of x(t), both data sets of of length N. The formula applies to both relative temperature as well as CO_2 as they are examples of data sequencies. The Theorem above simply states that there is a one-to-one correspondence between original data set x(t) ND ITS Fourier Transform version in frequency domain. Hence total signal energy in time is equivalent to total ener gy in frequency domain. Because of that we can freely look for energy distribution in frequency domain and use it to choose the most „energetic“ harmonics in temperature or CO_2 data set. As the number of these components will be just few of all present but carrying most of the energy (typically 5 components carry more than 90% of the total energy) per Table 7 bellow, this will simplify choice in Machine Learning approach of an appropriate method to utilize these harmonics. We are working to develop Kalman Filter like methodology to employ these results in the energy domain.

Table 7 First 5 „energy“ harmonics for individual and entire Vostok cycles

Harmonics Energy	Cycle C1	Cycle C2	Cycle C3	Cycle C4	Entire C1234
Temperature	470	395	362	4605	820
	178	311	159	108	719
	23	212	76	12	620
	20	68	18	95	246
	12	39	15	7	197
					121 and 111
CO_2	54,884	38,194	49,542	42926	74,646
	4,577	21,600	9,497	945	68,694
	3,652	7,960	5,056	867	31,585
	1,245	5,743	2,628	579	22,495
	853	1,759	1,635	431	21,783

5. Results summary

Various results of the analysis in this paper are summarized in Table 8 for ease of getting a general and complete view of the results. As stated earlier, all possible combinations of four cycle data were analyzed with the idea to enrich the data set for Machine Learning work in progress. Table 8 shows results for four cycles plus few more cycle combinations, as examples. Laft half of the table is devoted to relative temperature and the nrigh halh to CO_2. C1, C2 etc indicate which cycle the data refers to. On the upper left there is an indication of the length of the cycle period in thousand of years, and also the number of data points for each case, together with the information as to how many „energy“ harmonics were chosen. For example, for C1 temperature portion this is 5 harmonics out of 37, which is 13.5%. This corresponds to 93% of the total energy. There is also an indication of average sample time, for C1 temperature it is 1,703 years.

Table 8. Summary of results

Period 127726 Years		Harm.cs	Temp	in 1000 Y	in 1000 Y	CO2	Harm.cs	Period 128399 Years	
Points 74	C1	1	75	127726	128399	80	1	C1	Points 80
H 5 out of 37		1/2	37,5	63863	42799,7	26,667	1/3		H 6 out of 40
Max E %		1/6	12,5	21287,7	64199,5	40	1/2		Max E %
13,5135135	Energy %	1/4	18,75	31931,5	14266,6	8,8889	1/9	Energy %	15
Average Ts	93,01822	1/8	9,375	15965,8	18342,7	11,429	1/7	93,2569	Average Ts
1703,01351					25679,8	16	1/5		1604,99
Period 115156 Years		Harm.cs	Temp	in 1000 Y	in 1000 Y	CO2	Harm.cs	Period 109800 Years	
	C2	1	144	115156	109800	135	1	C2	
		1/2	72	57577,8	54900	67,5	1/2		
Points 136		1/3	48	38385,2	36600	45	1/3		Points 135
H 10 out of 72		1/6	24	19192,6	27450	33,75	1/4		H 6 out of 67
Max E %	Energy %	1/8	18	14394,4	18300	22,5	1/6	Energy %	Max E %
14	91,10529	1/9	16	12795,1	13725	16,875	1/8	92,217	8,95522388
Average Ts		1/5	28,8	23031,1					Average Ts
799,691176		1/7	20,5714	16450,8					813,33
		1 10th	14,4	11515,6					
		1 15th	9,6	7677,04					
Period 86462 Years		Harm.cs	Temp	in 1000 Y	in 1000 Y	CO2	Harm.cs	Period 86148 Years	
	C3	1	100	86461,6	86147,5	100	1	C3	
Points 99		1/2	50	43230,8	43073,7	50	1/2		Points 99
H 7 out of 50		1/4	25	21615,4	17229,5	20	1/5		H 5 out of 50
14		1/7	14,2857	12351,7	21536,9	25	1/4		10,00
864,616162	Energy %	1/3	33,3333	28820,5	28715,8	33,333	1/3	Energy %	861,47
	90,82682	1/5	20	17292,3				91,4872	
		1/8	12,5	10807,7					
Period 96782 Years		Harm.cs	Temp	in 1000 Y	in 1000 Y	CO2	Harm.cs	Period 95587 Years	
	C4	1	54	96782	95586,7	50	1	C4	
Points 54		1/2	27	48391	11948,3	6,25	1/8		Points 49
H 5 out of 27		1/4	13,5	24195,5	23896,7	12,5	1/4		H 4 out of 25
18,5185185	Energy %	1/3	18	32260,7	19117,3	10	1/5	Energy %	16,00
1792,25926	96,14033	1/5	10,8	19356,4				93,029	1911,73
Period 234781 Years		Harm.cs	Temp	in 1000 Y	in 1000 Y	CO2	Harm.cs	Period 239307 Years	
	C12	1/2	108	117391	119653	108	1/2	C12	
		1/3	72	78260,3	79769	72	1/3		
Points		1/5	43,2	46956,2	239307	216	1		Points
210		1	216	234781	47861,4	43,2	1/5		215,00
H 11 out of 10	Energy %	1/6	36	39130,2	59826,7	54	1/4	Energy %	H 12 out of 10
10,1851852	92,73771	1/4	54	58695,2	39884,5	36	1/6	91,6463	11,1111111
Average Ts		1 12th	18	19565,1	26589,7	24	1/9		Average Ts
1086,94907		1 14th	15,4286	16770,1	34186,7	30,857	1/7		1107,90
		1/9	24	26086,8	29913,4	27	1/8		
		1 15th	14,4	15652,1	21755,2	19,636	1 11th		
		1 16th	13,5	14673,8	18408,2	16,615	1 13th		
					15953,8	14,4	1 15th		

Period 198490 Years		Harm.cs	Temp	in 1000 Y	in 1000 Y	CO2	Harm.cs	Period 200088 Years	
	C23	1/2	120	99245,1	100044	120	1/2	C23	
		1/5	48	39698	66696,1	80	1/3		
		1/3	80	66163,4	40017,6	48	1/5		
Points		1 11th	21,8182	18044,6	200088	240	1		Points
235		1/7	34,2857	28355,7	15391,4	18,462	1 13th		234,00
H 17 out of 12	Energy %	1/6	40	33081,7	50022,1	60	1/4	Energy %	H 17 out of 12
14,1666667	91,89298	1/4	60	49622,6	28584	34,286	1/7	91,059	14,1666667
Average Ts		1 10th	24	19849	25011	30	1/8		Average Ts
827,042553		1 15th	16	13232,7	33348	40	1/6		833,70
		1 13th	18,4615	15268,5	18189,8	21,818	1 11th		
		1/8	30	24811,3	14292	17,143	1 14th		
		1	240	198490	13339,2	16	1 15th		
		1/9	26,6667	22054,5	11116	13,333	1 18th		
		1 23rd	10,4348	8630,01	16674	20	1 12th		
		1 20th	12	9924,51	20008,8	24	1 10th		
		1 25th	9,6	7939,61	10004,4	12	1 20th		
		1 27th	8,88889	7351,49	9094,92	10,909	1 22th		

Period 302218 Years		Harm.cs	Temp	in 1000 Y	in 1000 Y	CO2	Harm.cs	Period 293863 Years	
	C234	1/4	75	75554,6	146931	144	1/2	C234	
Points		1/2	150	151109	73465,7	72	1/4		Points
289		1/6	50	50369,7	48977,1	48	1/6		283
H 16 out of 150		1/3	100	100739	97954,3	96	1/3		H 16 out of 14
10,6666667		1 12th	25	25184,9	293863	288	1		11,1111111
Average Ts		1/7	42,8571	43174	58772,6	57,6	1/5		Average Ts
1007,39446	Energy %	1/9	33,3333	33579,8	36732,8	36	1/8	Energy %	1020,35689
	92,66972	1	300	302218	29386,3	28,8	1 10th	92,3698	
		1 14th	21,4286	21587	32651,4	32	1/9		
		1 18th	16,6667	16789,9	19590,9	19,2	1 15th		
		1/8	37,5	37777,3	20990,2	20,571	1 14th		
		1/5	60	60443,7	14693,1	14,4	1 20th		
		1 20th	15	15110,9	13993,5	13,714	1 21th		
		1 16th	18,75	18888,6	16325,7	16	1 18th		
		1 13th	23,0769	23247,6	18366,4	18	1 16th		
		1 11th	27,2727	27474,4	10495,1	10,286	1 28th		

Next three sub columns to the right indicate presence of specific harmonic numbers (1, ½ or 1/3, etc.), plus normalized FFT harmonics in time, and finally real harmonics estimate in 1000s of years. This is then repeated for all other cases of temperature (left side of the Table) as well as for CO_2 on the right side. This way one we can quickly compare temperature data with that of CO_2 as far as how many harmonics each requires, what is the enrgy percentage of the total energy, as well as average time sampling in years, as well as the cycle duration. The table can be expanded to incorporate all other possible cycle combinations for the sake of enriching Machine Learning data sset for training, testing and predicting purposes. We expect more data to be beneficial in order to reduce prediction errors. As the errors get calculated, they can be added to this table for completeness and quick analysis.

6. Conclusion, Future work and Machine Learning

In this paper we analyzed Vostok ice core data using (i) time correlations, (ii) harmonic analysis, as well as (iii) energy consideration. The general approach split the Vostok data set into 4 smaller sets, as per climate periodiciy indicated in the set. The general outcome is a choice of set of high energy harmonics for all cycles and any of their combinations for Kalman Filter based predictor for Machine Learning prediction purposes. The minimal choice of harmonics will allow us to devise a reasonably simple Machine Learning algorithm for training, testing and prediction purposes. For example long data set in C234 can be used as a Training Data Set, so can C23 (shorter though), in order to "predict" C1, calculate prediction error E1,234 (Error in predicting C1 given C2, C3 and C4 cycles) or E1,23 (preedicting C1 given onbly C2 and C3 data). This applies to both temperature and CO_2. We will repeat the above analysis for other components in Vostok data set, such as methane, oxygen and insolation. Similarly for European EPICA data set as well as set of cycles indicated in Milankovich theory. Hence, once we predict C1 using C123 or C23, we obtain errors E1,234 and E1,23. Intuitively we can expect that E1,23 > E1,234, i.e. training based on larger data set ideally would produce smaller test and prediction errors. With this in mind we can use C1234 to predict time now, or in the near or far away climate future of the data set. Using other C combinations not necesarilly next to each otrher in time (such as C13, C14, C24, C123, C134), it is conceivable we will be able to reduce the erros even further, all for the benefit of predicting NOW and in the FUTURE. If a precise numerical correlations are found between Vostok and Milankovitch data sets [14], prediction using our Machine Learning approach can be combined with Milankovitch future cycles as they can be correctly predicted. When chosing harmonics set, various energy targets, such as at 90%, 93%, 95% or 99%, at the expense of incorporating more harmonic components. Another by-product of the temperature data analysis is to estimate the likely water vapour content of atmospheres of different eras. Given that we have proposed positive forcing from irrigation water [11] in addition to other primary sources of warming such as the Milankovitch astronomical cycles. This may prove a more reliable means of correlation using the link already established between water vapour responsible for more than 80% of the heating of air [11].

Acknowledgements The author M. Hodzic would like to thank his PhD student Sabina Brkić at American University Bosnia-Herzigovina in Sarajevo for assisting in testing R code for statistical analysys of various data.

References

[1] Allen MR, Smith LA (1994) Investigating the origins and significance of low frequency modes of climate variability Geophys Res Lett 21:883–886.
[2] Alley RB (2010) Relaibility of ice-core science: historical insights. J Glaciology 56:1095-1103
[3] Bentley CR, Koci R (2010) Drilling to the beds of the Greenland and Antarctic ice sheets: A review. Annals Glaciology 47.
[4] Barnola J-M (1999) Status of the atmospheric CO_2 reconstruction from ice cores analyses: Keynote perspective. Tellus 51:151-155.
[5] Barnola J-M, Raynaud D, Lorius C, Barkov NI (2003) Historical CO_2 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.
[6] Boddy R, Smith G Statistical Methods in Practice: for Scientists and Technologists. John Wiley & Sons 2009.
[7] Fisher RA (1929) Tests of significance of harmonic analysis. Proc. Roy. Soc. A, 125:54-59
[8] Fourteau K, Faïn X, Martinerie P, Landais A,, Ekaykin AA, Lipenkov VY, Chappellaz (2017) Analytical constraints on layered gas trapping and smoothing of atmospheric variability in ice under low-accumulation conditions Clim. Past 13:1815–1830.
[9] Grochenig K (1996) An unceratinty principle related to the Poisson summation formula. Stud Math 121:69.
[10] Jouzel J, Lorius C, Petit JR, Genthon C, Barkov NI, Kotlyakov VM, Petrov VM (1987) Vostok ice core: a continuous isotope temperature record over the last climatic cycle (160,000 years). Nature 329:403-408

[11] Kennedy IR, Hodzic, M. (2018) Testing of the hypothesis that variations in atmospheric water vapour are the main cause of fluctuations in global temperature. This issue.
[12] Langway CC Jr. (2008) The history of early polar ice cores. in J Glaciol 55:385-6
[13] Matlab and R documentation (2018) Fourier Transforms, D Fourier Transforms, Fast Fourier Transforms
[14] MeyersSR, Sageman BB, Pagani M (2008) Resolving Milankovitch: Consideration of signal & noise. Amer J Sci 308:770-786.
[15] Petit JR, Jouzel J, Raynaud D, Barkov NI, Barnola J-M, Basile I, Bender M, Chappellaz J, Davis G, Delayque G, Delmotte M, Kotlyakov VM, Legrand M, Lipenkov VY, Lorius C, Pepin K, Ritz C, Saltzman E, Stievenard M (1999) Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. Nature 399:429-436
[16] Rudinger CM, Enting IG, Rayner PJ (2002a) Kalman filter analysis of ice core data. Method development and testing the statistics. J Geophys Res 107:4422 doi:1029/2001JD001111
[17] Trudinger CM, Enting IG Rayner PJ (2002b) Kalman filter analysis of ice core data 1. Double deconvolution of CO_2 and delta^{13}C measurements. J Geophys Res 107:4423, doi:1029/2001JD001112
[18] Wearing HJ (2010) Spectral Analysis in R
[19] Wichert S, Fokianis K, Strimmer K (2004) Identifying periodically expressed transcripts in microarray time series data. Bioinformatics 20:5-20
[20] Yiou R, Baert E, Loutre MF (1996) Spectral analysis of climate data. Surveys in Geophysics 17: 619–663
[21] Yiou R, FuhrerK, Meeker LD, Jouzely J, Johnson S, Mayerewski PA (1997) Paleoclimatic variability inferred from the spectral analysis of Greenland and Antarctic ice-core data. J Geophys Res 102:26441-26454
[22] Leslie Greengard, June-Yub Lee (2004) Accelerating the Nonuniform Fast Fourier Transform, SIAM REVIEW, Society for Industrial and Applied Mathematic,Vol. 46, No. 3, pp. 443–454.
[23] A. Dutt and V. Rokhlin (1993) Fast Fourier Transforms for Nonequispaced Data, Research Report YALEU/DCS/RR-980
[24] Derbentsev V, Kibalnyk L, Radzihovska Y, " Modelling multifractal properties of cryptocurrency market", PEN, 7:2, 690-701, 2019.
[25] Karthikeyan T, Sitamahalakshmi T, " Optimal and Efficient Time Series Classification with Burrows-Wheeler Transform and Spectral Window Based Transformation", PEN, 6:1, 332-347, 2018.
[26] Sengupta PP, Sinha M, Dutta UP, "Economic and environmental performances in manufacturing industries: A comparative study", PEN, 7:1, 99-108, 2019.