Recurrence Analysis of Climate Variation in the Northern Hemisphere

Project Report – GEW-DAP02: Nonlinear Data Analysis Concepts

Toni Schmidt, Nikolaos (Nikos) Kordalis, Kathia Ivett Cuevas Martínez

01 May 2020

1—INTRODUCTION

The states of a dynamic system such as the climate system, recur, which provides information about the systems behavior. For this, phase space trajectories are investigated as part of nonlinear data analysis. There are a series of steps to obtain a graphical representation of a dynamical system and observe their underlying nature. Such steps include (1) collecting time series data, (2) reconstructing the one-dimensional time series data into a multidimensional phase space, (3) transforming the embedded time series into a recurrence plot (RP), and (4) quantify the patterns within the RP using specific metrics or measuring the attributes (Knight et al., 2016).

Phase space trajectories represent the state of a system of a given time point in a multidimensional space. A system can be projected to a phase space by using a time-delayed embedding in which the original one dimension is unfolded into multiple time-lagged dimensions. When working with just one of the many dimensions that characterize a system, there are unobserved components. This results in an approximation of the system behavior, although aiming at reconstructing the most accurate. The decisions made regarding the embedding parameters are very important for further investigations and should be suited for the specific type of data.

RPs are two-dimensional representations of the embedded time series and illustrates the revisite of a state over time. A line of identity (LOI), i.e. the main diagonal, is part of every RP and separates the past from the future recurrences for a given time point. A RP is a binary matrix, whose elements can cluster to lines or various kinds of geometrical patterns. Recurrences can be constituted in different matters, e.g. level of proximity or distance in relation to the original state. For this, a norm between states and a threshold ε is crucial for constructing a RP to keep relevant patterns. Attributes that can be quantified in the a RP are, e.g. recurrence rate (how often a system revisits the same states), determinism (predictability of the system dynamics), maximum diagonal length (predictability and constancy of the system), or entropy (complexity of the system structure).

2—DATA

Deep ice core samples pose the problem that they can be folded near the bedrock. Andersen et al. (2004) published an undisturbed climate record as part of the North Greenland Ice Core Project (NGRIP). Fig. 1 depicts the location of the drilling (75.1 °N,

42.3 °W, 2917 m a.s.l.), which started in 1996 and ended in 2003. The core had a length of 3085 m, which is equivalent to a time span of 123 kyrs, i.e. it extends back to the last interglacial period. The data covers the Holocene and the entire last glacial period. For this examination work, we used the NGRIP stable oxygen isotopic ($\delta^{18}O$) time series (Fig. 2), which is a frequently used climate proxy for the northern hemisphere.

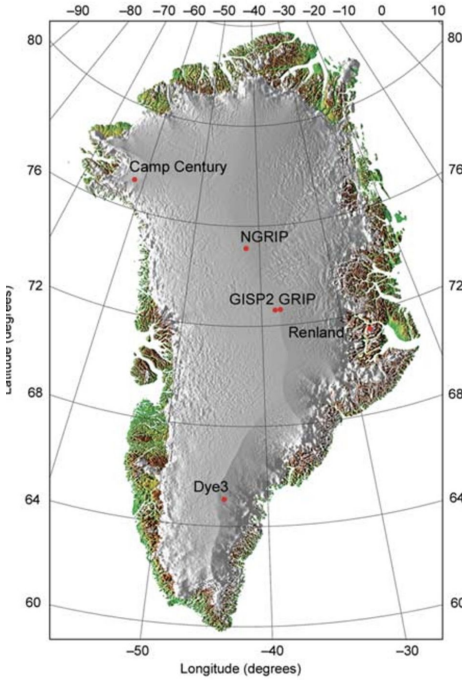

Figure 1. Core sampling location for the NGRIP data set in northern Greenland at 75.1 °N, 42.3 °W (Fig. 1, Andersen et al., 2004).

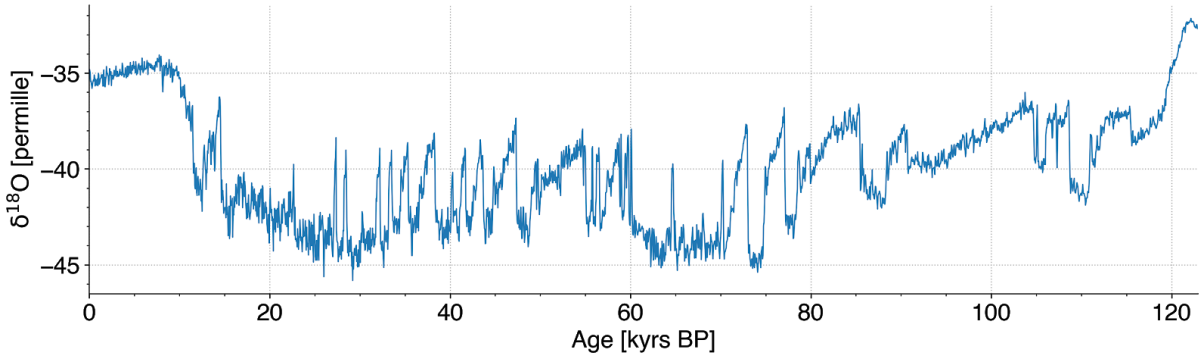

Figure 2. The NGRIP stable oxygen isotopic time series. Isotopic values ($\delta^{18}O$) are expressed in permille with respect to Vienna Standard Mean Ocean Water (V-SMOW). An accuracy of ±0.1 permille is stated by the authors (Andersen et al., 2004).

3—METHODS

First, we sliced the data set using every second row, as both variables within the set are repeated for two consecutive rows, but with a shift of one row with respect to each other (e.g. the first five rows are 0, –35.19; 50000, –35.19; 50000, –34.83; 100000, –34.83; 100000, –35.05). Then we assigned variables time and stable oxygen isotopes (δ^{18}O). The time data originally is in myrs (10^{-3} yrs)[1], so it was converted to kyrs (10^{3} yrs). The sampling period then is 0.05 kyrs, the sampling frequency thus 20 kyrs^{-1}. Then we normalized the time series and filtered it, as explained below.

Then we estimated appropriate embedding parameters, reconstructed the phase space trajectories, and constructed RPs of the embedded (filtered) time series. For this we conducted trial and error approaches to determine the best embedding parameters, recurrence criteria, and thresholds ε by choosing different values. We selected the ones that lead to results with the best information contents. Using the RPs we computed the τ-recurrence rates (RR_τ) of the embedded (filtered) time series and made Fourier transforms (FTs) of those to derive the recurrence based power spectra. The impact of each filter over the recurrence metrics has been assessed. Finally, we made FTs of the (filtered) time series to compare the power spectra with the recurrence based power spectra.

3.1—FILTERING

Before filtering the NGRIP time series, we normalized it to better compare the unfiltered with the low-, high-, and band-pass filtered time series (Fig. 3). The low-pass filter (Fig. 3B) smoothes the original data, but loses high-frequent details. High-pass filtering (Fig. 3C) results in a high-frequent signal, i.e. a detrended time series that lacks the ability to show long-term patterns in the data. Whereas band-pass filtering (Fig. 3D) outputs a smooth signal that retains the strongest features within the time series. Cut-off frequencies for each filter option are 0.125 kyrs^{-1} for the low-pass filter, 0.5 kyrs^{-1} for the high-pass filter, and the same values as minimum and maximum respectively for the band-pass filter.

[1] The time variable could also be published as unsigned integers with a precision of three decimal places if normalized by 1000. No further explanations could be found concerning this.

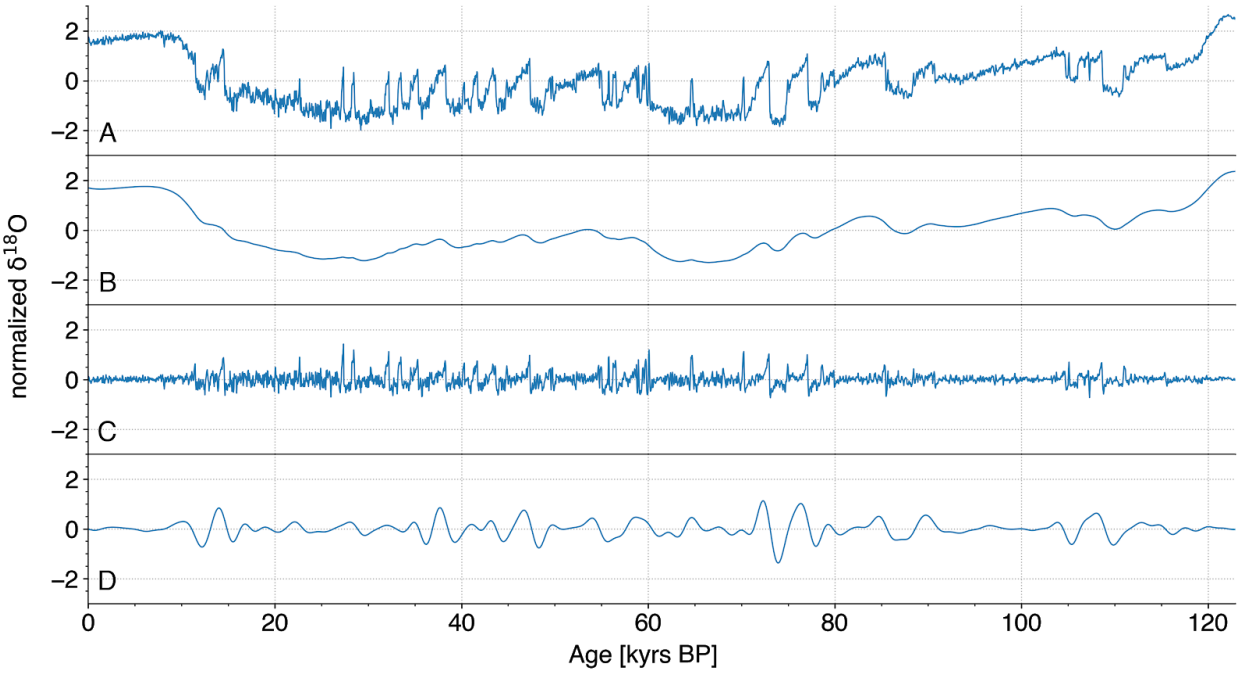

Figure 3. Filtering outputs of normalized NGRIP time series. **A** shows the unfiltered, **B** the low-pass filtered (cut-off frequency of 0.125 $kyrs^{-1}$), **C** the high-pass filtered (0.5 $kyrs^{-1}$), and **D** the band-pass filtered (0.125 $kyrs^{-1}$ and 0.5 $kyrs^{-1}$) time series.

3.2—ESTIMATION OF EMBEDDING PARAMETERS

For the reconstruction of the phase spaces we used the time delay method. This requires two embedding parameters: the dimension m and the delay τ. We applied two different functions to estimate τ: the mutual information function and the autocorrelation coefficient function. For estimating m, we used the false nearest neighbors algorithm.

The mutual information function is a nonlinear, unitless, and non-normalized measure of the dependency between two variables, where both variables are the same, i.e. the NGRIP time series with a τ-lagged version of itself. A low mutual information corresponds to a high independence. It is a function of τ and decreases as τ increases. The first local minimum of the curve matches with an appropriate τ due to the relatively low correlation between the time series. Fig. 4 shows that τ of the unfiltered time series is supposed to be 2 (which corresponds to 0.10 kyrs, taking the sampling period of 0.05 kyrs).

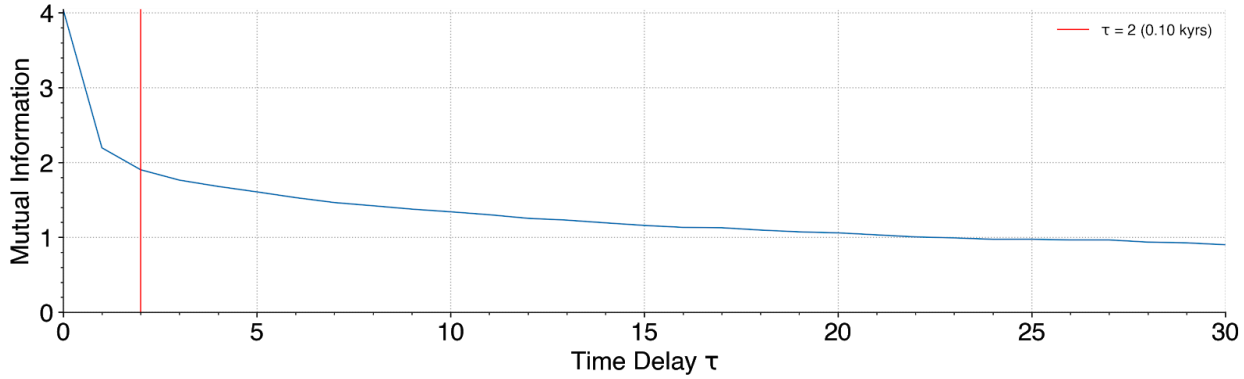

Figure 4. Estimation of time delay τ for the unfiltered time series using the mutual information function. τ = 2 was chosen, as it corresponds to the first local minimum of the curve.

The autocorrelation coefficient function is based on the same principle as the mutual information function, as both measure the dependency between two variables. But the former uses the variance-normalized covariance of the time series with a τ-lagged version of itself. It ranges from –1 to 1. This facilitates relative comparisons among, e.g., the filtered time series. An appropriate τ is where the autocorrelation coefficient falls below 1/e. Fig. 5 shows that τ of the unfiltered time series using this approach is supposed to be 171 (which corresponds to 8.5 kyrs). For the unfiltered and filtered time series we observed relatively high τ compared to outputs of the mutual information function, too. After embedding the time series and discussing further processing outputs, we decided to keep the τ as derived using the mutual information function.

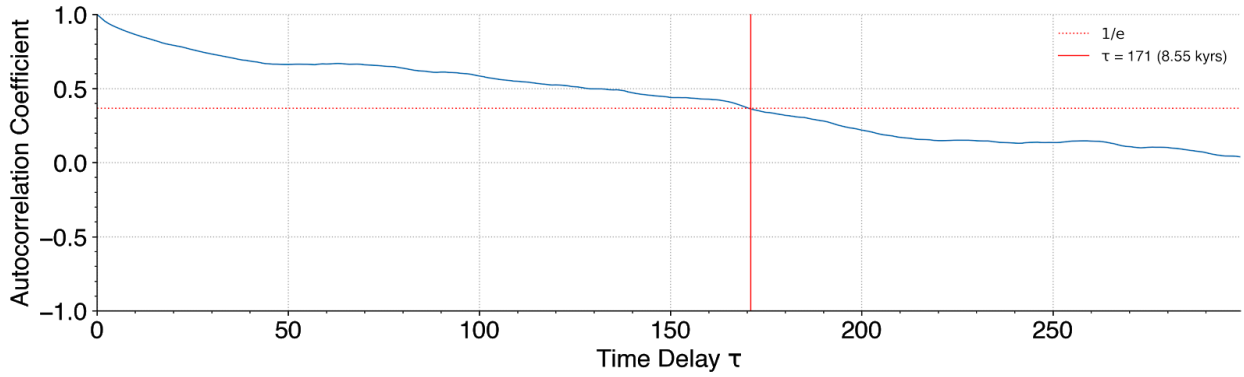

Figure 5. Estimation of time delay τ for the unfiltered time series using the autocorrelation coefficient function. An optimal τ is where the autocorrelation coefficient falls below 1/e, i.e. τ = 171. This τ is not used for embedding the time series, but the one derived using the mutual information function (Fig. 4).

The false nearest neighbors algorithm is used to calculate the appropriate embedding dimension m. For this, the integer dimensions are increasing from 1 while the number of neighbors within a fixed radius is calculated. It is normalized by the one-dimensional

case, such that the maximum relative number of false nearest neighbors is 100%. For a dimension where only true neighbors remain, i.e. for a low percentage of false nearest neighbors, the appropriate m is found. Fig. 6 shows that we estimated m = 11 for the original time series.

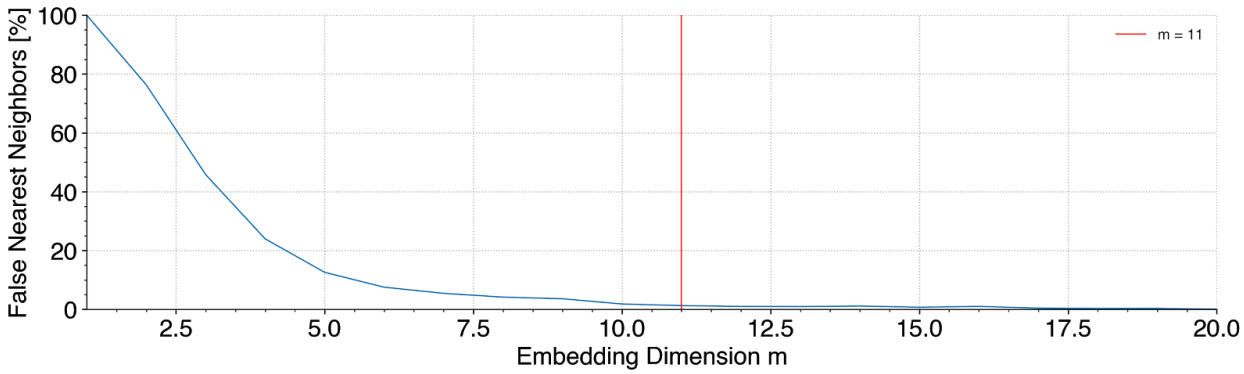

Figure 6. Estimation of embedding dimension m for the unfiltered time series using the false nearest neighbors algorithm using a radius of 1. m = 11 was chosen, as it corresponds to the first local minimum of the curve.

Embedding parameters for the filtered time series were derived using the same functions and algorithm. Though τ-outputs of the autocorrelation function were not used for embedding the time series, as explained above. To conclude, we obtained the following embedding parameters:

- Unfiltered time series: m = 11, τ = 2
- Low-pass filtered time series: m_{lp} = 2, τ_{lp} = 30
- High-pass filtered time series: m_{hp} = 16, τ_{hp} = 1
- Band-pass filtered time series: m_{bp} = 9, τ_{bp} = 16

4—RESULTS

By using the estimated embedding parameters we embedded the (filtered) time series to reconstruct the phase spaces using the time delay method and to construct the RPs. From those we computed the τ-recurrence rates (RR_τ) and made FTs to derive the recurrence based power spectra, which in the end were compared to the power spectra of the original (filtered) time series.

4.1—PHASE SPACE TRAJECTORIES

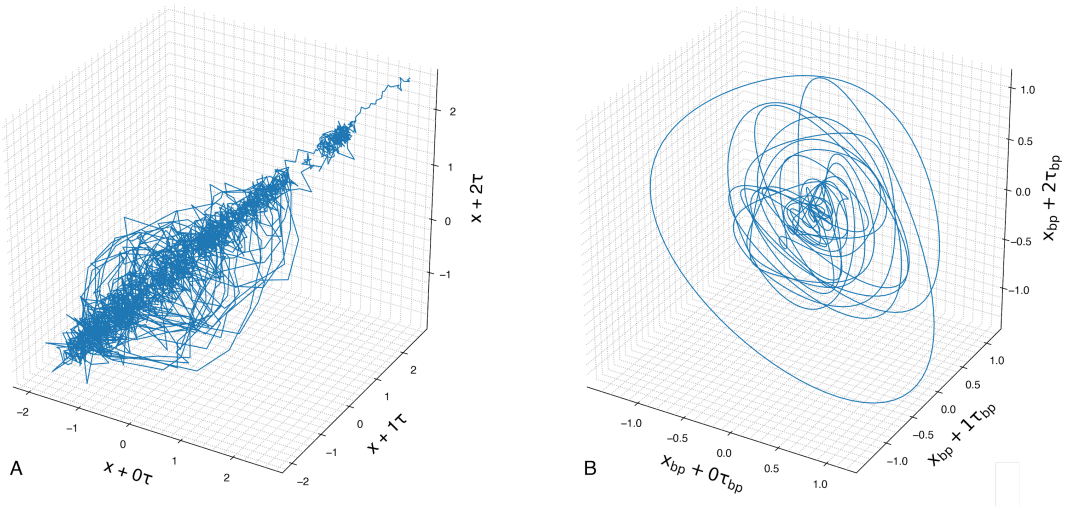

Figure 7. Phase space trajectories of **A** the unfiltered time series (m = 11, τ = 2) and **B** the band-pass filtered time series (cut-off frequencies 0.125 kyrs^{-1} and 0.5 kyrs^{-1}, m = 9, τ = 16).

We only show the phase space trajectories of the unfiltered and band-pass filtered time series (Fig. 7), as the latter is more convincing than those of the low- or high-pass filtered time series. The time delay of the low-pass filtered time series was derived 2, i.e. the phase space trajectory is 2-dimensional. Due to the low amplitude of the high-pass filtered time series, the phase space trajectory of it is very dense and thus shows information that is difficult to grasp. The phase space trajectory of the embedded band-pass filtered time series (Fig. 7B) shows a clear trajectory, whereas that of the unfiltered time series (Fig. 7A) is influenced by signals with high-frequency and is stretched, which leads to a disrupted RP (Fig. 8A).

4.2—RECURRENCE PLOTS (RPs)

After embedding the (filtered) time series, we constructed the RPs for each of them. Choosing an appropriate ε, i.e. it must not be too small, where information is lost from the system, or too large, where many artefacts disturb the investigation of the system, is crucial for the construction of RPs. We decided on setting a fixed ε = 0.1 using the fixed amount of nearest neighbours (FAN) recurrence criterion for the construction of all RPs, after using other different ε and criteria. This resulted in extensive information within the RPs by the presence of distinguishable typologies and textures.

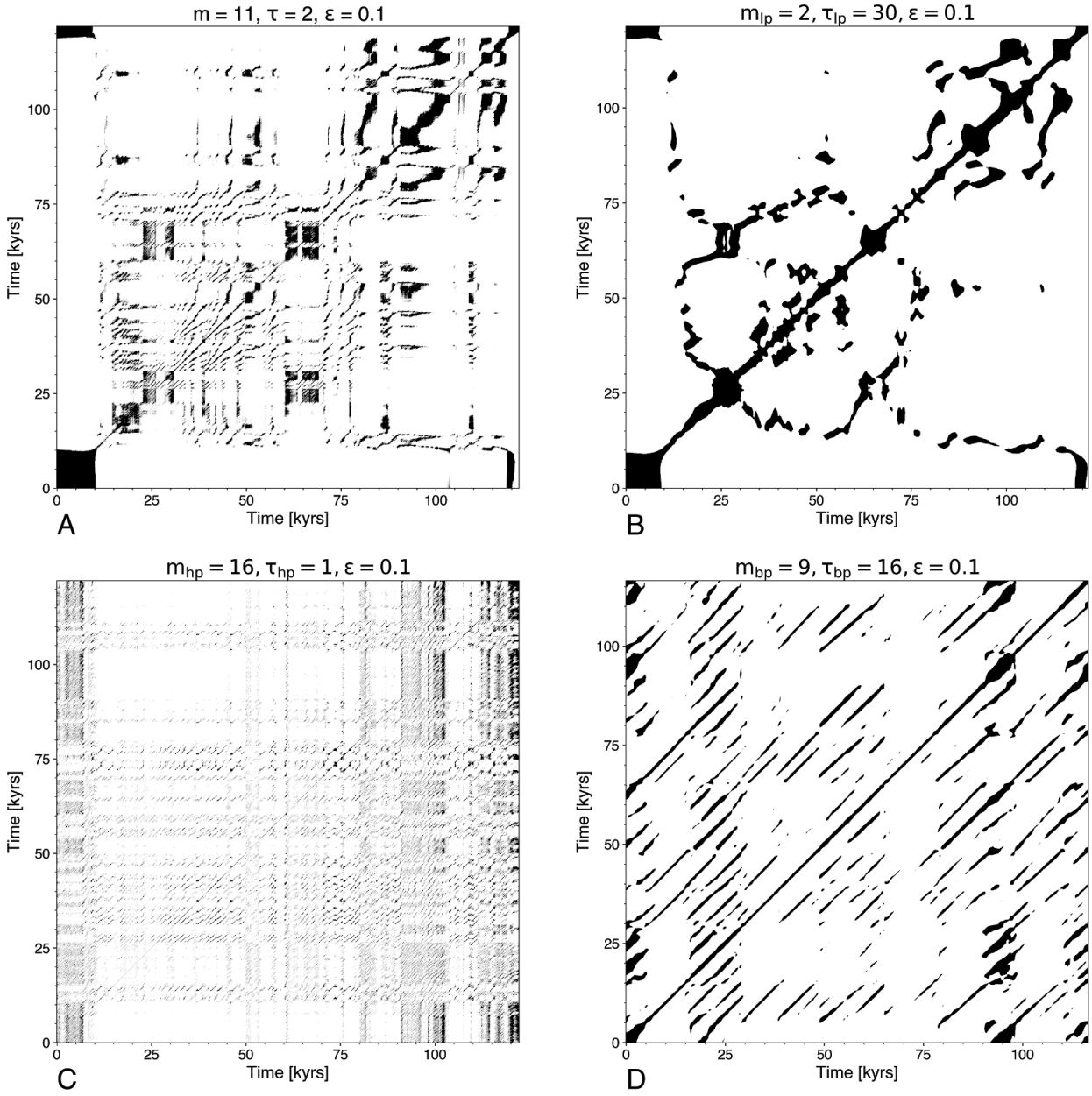

Figure 8. Recurrence plots of the time series. **A** is based on the unfiltered, **B** on the low-pass filtered, **C** on the high-pass filtered, and **D** on the band-pass filtered time series. Cut-off frequencies are found in Fig. 3.

The typologies, i.e. the large scale patterns within the RPs, differ between the unfiltered and filtered embedded time series. The RP of the unfiltered and low-pass filtered time series are disrupted (Fig. 8A and C), whereas the RPs of the low- and band-pass filtered time series are periodic (Fig. 8B and D). Textures, i.e. small scale structures within the RPs, are manifold. The RP of the unfiltered time series (Fig. 8A) has regions with clusters. These clusters repeat in similar way, specially at 25–30 kyrs and 60–70 kyrs. Although not

very defined, there are other clusters around 0–10 kyrs and 15–23 kyrs. In addition, there are diagonal lines of different lengths. The geometric definition of the dataset starts to be lost after 75 kyrs. Though it recurs still at ~75–80 kyrs, ~90–100 kyrs, and ~110–120 kyrs. In the area between 80 and 120 kyrs a periodic behaviour is observed. The RP of the embedded low-pass filtered time series (Fig. 8B) is very periodic and harmonic. In contrast, the RP of the embedded high-pass filtered time series (Fig. 8C) shows individual points which are sharply defined. A challenge is to detect the recurrent behavior within the phase space using this visual approach. The LOI is very thin. The RP of the embedded band-pass filtered time series (Fig. 8D) allow us to distinguish different recurrence-patterns within the data. The deterministic typology has a relatively constant periodicity in relation to the revisit of the same state over time. The apparent entropy level can be interpreted as relatively low, given the fact that the length of the observed diagonals does not change much especially before 75 kyrs.

4.3— τ -RECURRENCE RATE (RR_τ)

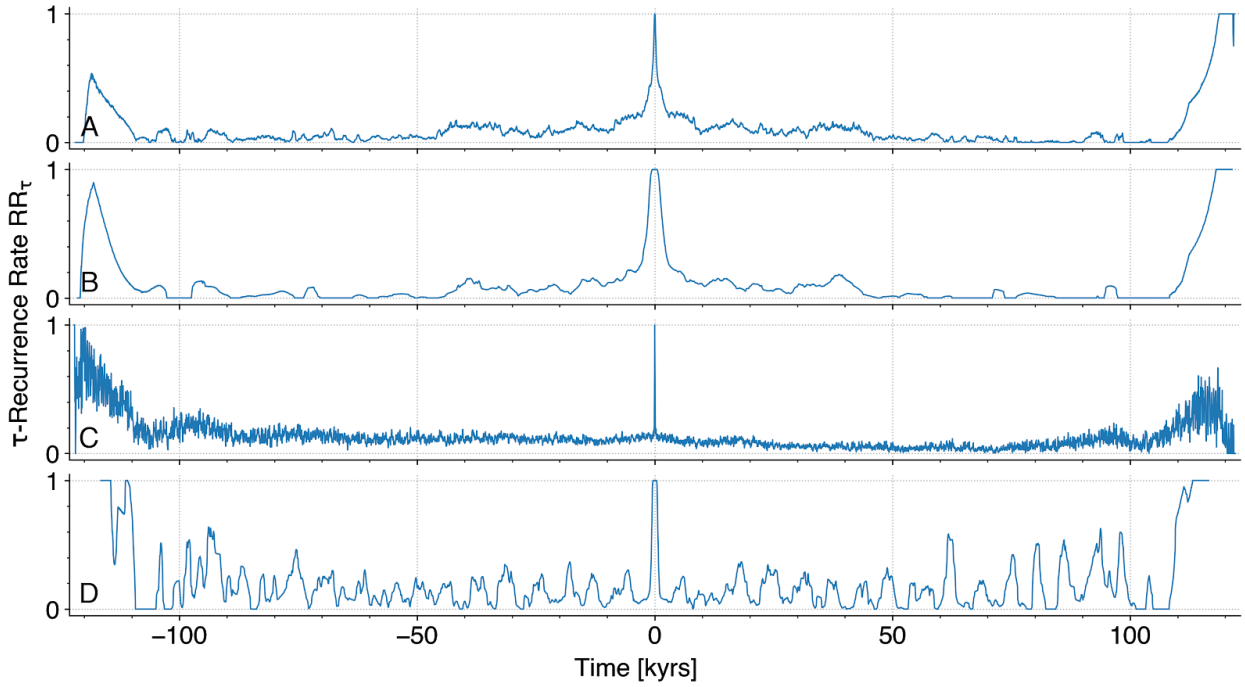

Figure 9. τ -recurrence rates RR_τ of the recurrence plots of the time series. **A** is based on the unfiltered, **B** on the low-pass filtered, **C** on the high-pass filtered, and **D** on the band-pass filtered time series.

The RR_τ refers to the probability of a system to recur in a given number of time steps τ within its ε -neighbourhood (Marwan et al., 2007). Fig. 9 shows the RR_τ of the embedded (filtered) time series, which are not symmetric, as the chosen recurrence criterion, i.e.

the fixed amount of nearest neighbours (FAN), is not a metric criterion. Simple euclidean distance as a metric criterion on the other hand would provide symmetric RPs, thus symmetric RR_τ. Predominant peaks are mainly apparent when using the embedded band-pass filtered time series. There, a periodic behavior of the probability of the climate system to recur in relatively constant time steps is evident. The RR_τ curve of the embedded unfiltered time series on the other hand is not showing this periodicity.

4.4—RECURRENCE BASED POWER SPECTRUM

The time series analysis in the time domain can be transformed to the frequency domain which is a complementary analysis but can explore features that are not detected by the former one. These features mainly are related to frequencies that coexist in the time series signal. FT is a method to switch these domains and FT algorithms are the appropriate tool for spectral analysis. The power spectral density (PSD) of a signal is defined as the FT of the autocorrelation function. A classic, non-parametric algorithm estimation of the PSD is the periodogram, which is used to identify the dominant power densities per frequencies (or periods) of a time series. The maximum power density of a periodogram is an indicator of periodicity.

As shown in Fig. 10 and 11, no power densities below 0.1 kyrs, i.e. above frequencies of 10 $kyrs^{-1}$, are calculated due to the Nyquist criterion, which states that the sampling frequency (here: 20 $kyrs^{-1}$) must be at least twice the maximum component in the signal (Cerna and Harvey, 2000).

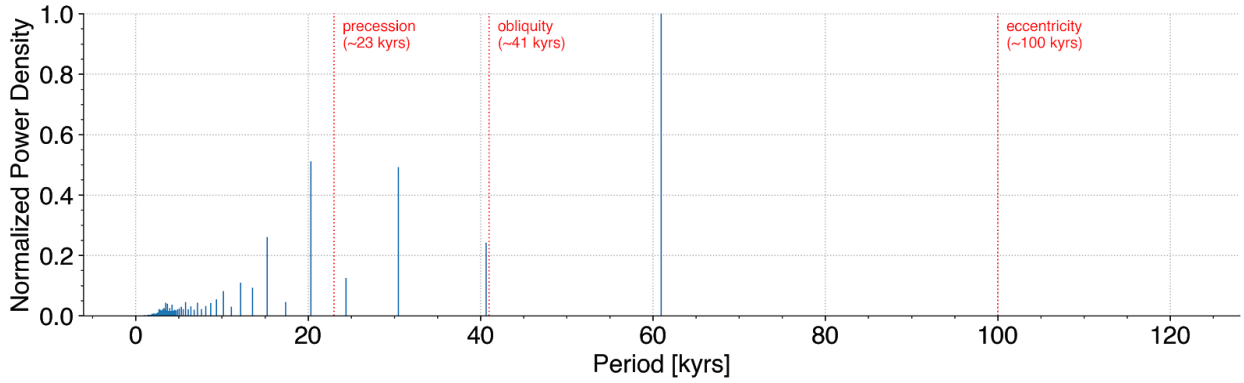

Figure 10. Recurrence based power spectrum of the unfiltered time series showing the normalized power density (i.e. normalized by the maximum power density) over the period [kyrs]. Three orbital cycles as posited by Milankovitch are highlighted by the dashed red lines.

Milankovitch (orbital) cycles have a major role in determining Earth's climate. Fig. 10

and 11 depict the three cycles, i.e. precession (direction of Earth's axis, periodicity of ~23 kyrs), obliquity (tilt of Earth's axis, ~41 kyrs), and eccentricity (shape of Earth's orbit, ~100 kyrs). At ~40.6 kyrs the fifth-highest power density is very close to the obliquity. The sixth-highest power density at ~24.4 kyrs is very close to the precession. But the maximum power density at ~61.0 kyrs cannot be explained by the Milankovitch cycles. Also the eccentricity, which has a periodicity of ~100 kyrs does not appear in the recurrence based periodogram.

4.5—POWER SPECTRUM

In Fig. 11, at ~41 kyrs the maximum power density matches with the obliquity which has approximately the same periodicity. The third-highest power density at ~24.6 kyrs is very close to the precession which has a periodicity of ~23 kyrs. Here, the second-highest power density at ~61.5 kyrs is close to the period of the maximum power density of the recurrence based periodogram (Fig. 10) and as well cannot be explained by the Milankovitch cycles. Also the eccentricity, which has a periodicity of ~100 kyrs does not appear in the periodogram.

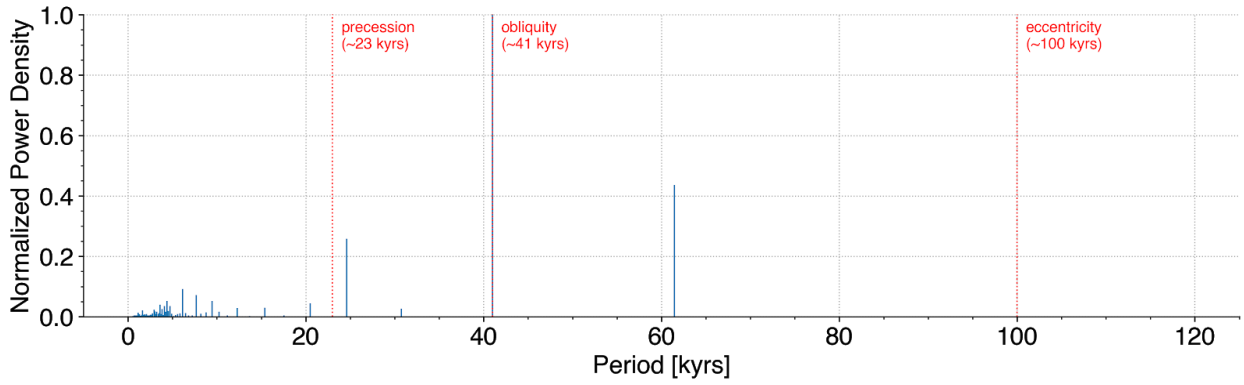

Figure 11. Power spectrum of the unfiltered time series showing the normalized power density (i.e. normalized by the maximum power density) over the period [kyrs]. Three orbital cycles as posited by Milankovitch are highlighted by the dashed red lines.

Comparing Fig. 10 and 11, i.e. comparing the recurrence based power spectrum with the power spectrum of the unfiltered (original) time series, greater detail especially for lower periods can be determined from the former one. As explained above, the periodicity at ~61 kyrs in both periodograms are predominant and must be explained by another phenomenon than the Milankovitch cycles.

The power densities for low periods in Fig. 11 are hard to distinguish, as it is the power

spectrum of the unfiltered time series. Due to this, the periodogram of the high-pass filtered time series is shown in Fig. 12, where more detail for low periods are given. Highest power densities are at ~1.2 and ~1.6 kyrs with relatively equal power densities, where that of the former period is 1.

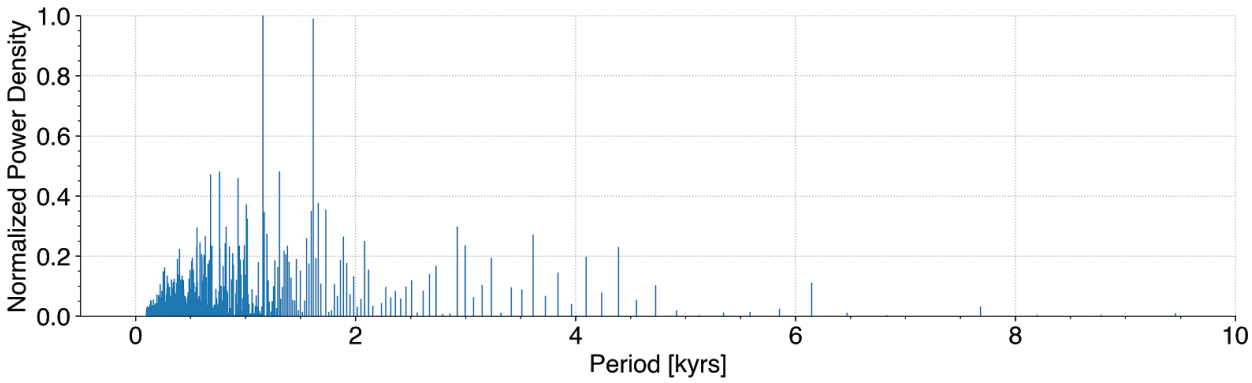

Figure 12. Power spectrum of the high-pass filtered time series (cut-off frequency of 0.5 $kyrs^{-1}$) showing the normalized power density (i.e. normalized by the maximum power density) over the period [kyrs].

5—DISCUSSIONS

The RPs of the embedded low- and high-pass filtered time series depict the existing relation between the unfolding of the data in the selected number of dimensions. Regarding the embedding parameters, the low-pass filtered time series experiences $m_{lp} = 2$. This results in a two-dimensional phase space trajectory. It is evident from the RP of the embedded low-pass filtered time series (Fig. 8B), that the combination of unfolding in a two dimensions by using a relatively high τ of 30 produces an output with a loss of clarity of the individual periodicity of which the time series is composed. In contrast, when looking at the RP of the embedded high-pass filtered time series (Fig. 8C), the combination of an unfolding in a higher dimension (m = 16) with a low time delay (τ = 1) results in clarity of the individual recurrence-points while losing track of the patterns that would allow to investigate recurrent behavior, if present in the time series. Moreover, the phase space trajectory of the embedded high-pass filtered time series presented a high density that led to a lack of crucial information, due to the low amplitude of this time series. Similarly, the recurrence patterns within the RP of the embedded low-pass filtered time series (Fig. 8B) presents lines with a non-constant slope (bowed lines) and does not contain repeated clusters of data from which we could have had based on our analysis.

One of the major objectives that arises when analyzing the NGRIP time series is whether

it exhibits periodicity, i.e. recurrence. The use of periodograms was introduced for investigating these periods within the time series, which are derived from FT. They provide a break down of the time series into discrete periods, which allows insight into recurrences of the time series. According to the (recurrence based) periodograms (Fig. 10 and 11), the period at ~61 kyrs has a predominant power density. For the recurrence based power spectrum it has the maximum power density, whereas for the power spectrum of the original time series it is the second most significant period. This period does not correspond to Milankovitch cycles, from which two were observed in both periodograms, i.e. precession (~23 kyrs) and obliquity (~41 kyrs). The ~61 kyrs-peak can be related to glacial-interglacial periodicity, i.e. driven by the Milankovitch cycles. The glacials lasted between ~70 and ~90 kyrs, whereas the interglacials lasted ~10 kyrs over the last 450 kyrs (Philander 2008). Thus, the ~61 kyrs-peak can be assigned glacial periodicity. On the other hand, non-linear responses to either internal or external forcings in the system can be observed by periods with a high power density, such as the mid-Pleistocene transition happening at ~0.7–1.25 myrs (Clark et al., 2006), where the periodicity of the glacial cycles changed from ~41 kyrs to ~100 kyrs (Pena and Goldstein, 2014). Muller and MacDonald (1997) stated that the inclination as another orbital cycle has a period of ~70 kyrs, which is close to the ~61 kyrs-peak found in both periodograms.

6—CONCLUSIONS

For this examination project we investigated the NGRIP stable oxygen isotopic time series using nonlinear data analysis techniques. We focussed on the recurrence analysis of the embedded unfiltered, low-, high-, and band-pass filtered time series using RPs, RR_τ, and periodograms. The phase spaces were reconstructed using the time delay method, where m was estimated using the false nearest neighbors algorithm and τ using the mutual information function. Different recurrence criteria and ε were tried, whereas FAN with ε = 0.1 was chosen. The embedded band-pass filtered time series provided more information for the RR_τ.

After decomposing the time series using the FT, we computed the periodograms of the time series, which is an estimator of the power spectrum. The dominant peaks of the periodograms correspond to both precession (~23 kyrs) and obliquity (~41 kyrs) as part of the Milankovitch cycles and to the glacial periodicity (~70 kyrs), whereas predominant peaks were derived at ~61 kyrs.

REFERENCES

Andersen, K., Azuma, N., Barnola, J. et al. (2004) High-resolution record of Northern Hemisphere climate extending into the last interglacial period. Nature, 431, 147–151.

Baraniuk, R. (2015). Signals and Systems. University Press of Florida.

Cerna, M., & Harvey, A. F. (2000). The fundamentals of FFT-based signal analysis and measurement. Application Note 041, National Instruments.

Clark, P. U., Archer, D., Pollard, D., Blum, J. D., Rial, J. A., Brovkin, V., ... & Roy, M. (2006). The middle Pleistocene transition: characteristics, mechanisms, and implications for long-term changes in atmospheric pCO2. Quaternary Science Reviews, 25(23–24), 3150–3184.

Knight, A. P., Kennedy, D. M., & McComb, S. A. (2016). Using recurrence analysis to examine group dynamics. Group Dynamics: Theory, Research, and Practice, 20, 223–241.

Kugiumtzis, D. (2014). Time series analysis. Department of Electrical and Computer Engineering, Faculty of Engineering, Aristotle University of Thessaloniki.

Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. Physics reports, 438(5–6), 237–329.

Muller, R. A., & MacDonald, G. J. (1997). Origin of the 100-kyr Glacial Cycle: eccentricity or orbital inclination. Proc Natl Acad Sci USA, 94, 8329–8334.

Papandreou, N. (2007). Stochastic Signals & Applications. Department of Computer Engineering and Informatics, University of Patras, Greece.

Pena, L.; Goldstein, S. L. (2014). Ice Age Reboot: Thermohaline Circulation Crisis during the Mid-Pleistocene Transition. American Geophysical Union, Fall Meeting 2014, abstract id. PP14B-05.

Philander, S. G. (Ed.) (2008). Encyclopedia of global warming and climate change, vol. 2. Los Angeles: SAGE, 519–522.

Spiegel, D. S., Raymond, S. N., Dressing, C. D., Scharf, C. A., & Mitchell, J. L. (2010).

Generalized Milankovitch cycles and long-term climatic habitability. The Astrophysical Journal, 721(2), 1308.

GITHUB REPOSITORY

A GitHub repository was created for this examination project and is accessible as follow: https://github.com/toschmi/ndaexam/. The data, a Jupyter Notebook with all functions and computations, all figures, the metadata, and a draft version of the report is found there.