University of Groningen

Identification of possible $\delta^{14}\text{C}$ anomalies since 14 ka BP

Neocleous, Andreas; Azzopardi, George; Dee, Michael W.

Published in:
Science of the Total Environment

DOI:
[10.1016/j.scitotenv.2019.01.251](https://doi.org/10.1016/j.scitotenv.2019.01.251)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Neocleous, A., Azzopardi, G., & Dee, M. W. (2019). Identification of possible $\delta^{14}\text{C}$ anomalies since 14 ka BP: A computational intelligence approach. *Science of the Total Environment*, 663, 162-169.
<https://doi.org/10.1016/j.scitotenv.2019.01.251>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Identification of possible $\Delta^{14}\text{C}$ anomalies since 14 ka BP: A computational intelligence approach

Andreas Neocleous^{a,*}, George Azzopardi^b, Michael Dee^a

^a Center for Isotope Research, University of Groningen, Groningen, the Netherlands

^b Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, the Netherlands

HIGHLIGHTS

- Miyake events are natural phenomena causing failures to telecommunication systems.
- We investigated computational techniques that effectively model such events.
- COSFIRE filters are effective for modeling Miyake events.
- Spectral features are also suitable for identifying Miyake events.
- Our methods identify 75% of the known Miyake events with FPR <10%.

GRAPHICAL ABSTRACT

ARTICLE INFO

Article history:

Received 1 November 2018

Received in revised form 18 January 2019

Accepted 18 January 2019

Available online 23 January 2019

Editor: Mae Sexauer Gustin

Keywords:

COSFIRE

Digital signal processing

Feature extraction

Miyake events

Pattern matching

Radiocarbon dating

ABSTRACT

Rapid increments in the concentration of the radiocarbon in the atmosphere ($\Delta^{14}\text{C}$) have been identified in the years 774–775 CE and 993–994 CE (Miyake events) using annual measurements on known-age tree-rings. The level of cosmic radiation implied by such increases could cause the failure of satellite telecommunication systems, and thus, there is a need to model and predict them. In this work, we investigated several intelligent computational methods to identify similar events in the past. We apply state-of-the-art pattern matching techniques as well as feature representation, a procedure that typically is used in machine learning and classification. To validate our findings, we used as ground truth the two confirmed Miyake events, and several other dates that have been proposed in the literature. We show that some of the methods used in this study successfully identify most of the ground truth events (~1% false positive rate at 75% true positive rate). Our results show that computational methods can be used to identify comparable patterns of interest and hence potentially uncover sudden increments of $\Delta^{14}\text{C}$ in the past.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

When radiocarbon (^{14}C) dating was first developed in the 1940s, it was assumed that the concentration of radiocarbon in the atmosphere ($\Delta^{14}\text{C}$) had been constant over time (Libby et al., 1949). A few years later, it was discovered from analyzing known-age tree-rings that

atmospheric ^{14}C concentrations had in fact been fluctuating (de Vries, 1958). Nevertheless, the changes were still expected to be minor (~1%) from one year to the next, so records of $\Delta^{14}\text{C}$ back through time were obtained on blocks of 5–10 tree rings. However, Miyake et al. (2012) found a rapid increase (~12%) in $\Delta^{14}\text{C}$ in Japanese tree-rings between the years 774 and 775 CE (Common Era). Other groups confirmed this anomaly using dendrochronological archives from Germany (Usoskin et al., 2013), the USA and Russia (Jull et al., 2014), and New Zealand (Güttler et al., 2015). A similar anomaly in the

* Corresponding author at: Nijenborgh 6, Groningen, the Netherlands.
E-mail address: a.c.neocleous@rug.nl (A. Neocleous).

concentration of ^{14}C has since been found by the group of Miyake et al. for the years 993–994 CE (Miyake et al., 2013). These events have also been observed in records of the cosmogenic isotope ^{10}Be in ice cores from Greenland and Antarctica (Sukhodolov et al., 2017). Single-year measurements over the periods 770–780 CE and 990–1000 CE and detailed studies of these events were completed by (Büntgen et al., 2018) and (Uusitalo et al., 2018).

The cause of these increases in cosmogenic isotope production has not been proven categorically, but some plausible explanations have been proposed. The level of ^{14}C in the atmosphere can be elevated by extreme solar energetic particle events (Usoskin et al., 2013; Gütler et al., 2015; Dee et al., 2017; Mekhaldi et al., 2015). This is now the leading hypothesis, although supernovae and other γ -ray sources (Miyake et al., 2012; Hambaryan and Neuhäuser, 2013; Pavlov et al., 2013) could also generate the same effect. Terrestrial events, such as volcanic eruptions, fossil fuel burning and oceanic upwelling can only result in decreases in atmospheric $\Delta^{14}\text{C}$.

Attempts at identifying further sudden increases in the past have been made and reported in the literature. A common method is to analyze the long time-series of $\Delta^{14}\text{C}$ data (IntCal) that are publicly available.¹ The first systematic search of different records, including the $\Delta^{14}\text{C}$ IntCal09 series was performed by (Usoskin and Kovaltsov, 2012) using a visual inspection. The most up-to-date record is the IntCal13 dataset (Reimer et al., 2013), although it is mostly composed of decadal measurements, creating difficulties in identifying yearly changes in the signal.

In (Miyake et al., 2017a), the authors identified twelve potential events based on the percentage difference between successive samples in the dataset. They proceeded with single-year measurements over the period of highest variance, around 5480 Before Common Era (BCE). We contend that the approach taken in (Miyake et al., 2017a) can be improved in several different respects. First, we identify the difficulty in measuring the difference between samples when using the IntCal13 dataset, due to the fact that in most of the cases we have multiple $\Delta^{14}\text{C}$ values for the same calendar year. This is mainly because the data come from different research groups. If we also take into account the uncertainty that comes from the measurement procedure, the magnitude of the difference between samples varies greatly, making the method even less informative. Moreover, we believe that Miyake events are only valid when there is a tail in the signal for some years after the sudden increment of the $\Delta^{14}\text{C}$ and not just an anomalous jump that lasts one year, as this may simply be an outlier. Lastly, one could ask what the optimal value is for the percentage difference to be considered a potential Miyake event. In Miyake et al. (2017a) they used a threshold of 0.4%, but the derivation of that value is not wholly convincing.

One further example of the weakness of the aforementioned approach is found in (Wang et al., 2017) where the authors had to perform single-year measurements over 30 years to identify an event in 3372 BCE. This is obviously very costly and the risk of not identifying an event is high. Similarly, Miyake et al. (2017b) analyzed four periods (40 dates in total) and no anomalous $\Delta^{14}\text{C}$ were found. This fact shows that the IntCal13 dataset is unreliable when using the percentage difference between consecutive samples as an indicator for possible Miyake events. To enhance this line of reasoning, we observed that in IntCal13 there is an increment of about 7‰ between the years 124 and 129 CE. Then, we performed single-year measurements of about ten years around 128 CE and no Miyake event appeared during this period.

In this work we investigate state-of-the-art signal processing methods and feature representation to the IntCal13 dataset in order to identify patterns similar to Miyake events in the past, and identify the best method for this task.

2. Methods

2.1. Overview

Several peculiarities of the IntCal13 dataset have to be taken into account, such as missing data, multiple values for the same dates, and the uncertainty that comes unavoidably from the $\Delta^{14}\text{C}$ measurement process. To overcome these issues, we apply an averaging, as well as simple linear interpolation between existing $\Delta^{14}\text{C}$ values. Then, we use several signal processing methods by applying a sliding window technique. We used the first Miyake event as a training pattern and the remaining data as a test pattern. Then, we applied distance metrics based on Combination of Shifted Filter Responses (COSFIRE filters), Euclidean distance, cross correlation and other state-of-the-art methods. Furthermore, we extracted a number of spectral features from the entire signal and we visualize their responses. The pipeline of our methodology is shown in Fig. 1.

2.2. Data

From the IntCal13 dataset we used atmospheric data that consists of ^{14}C measurements on tree-rings made by the University of Washington (Stuiver et al., 1998), Queen's University Belfast (McCormac et al., 2008), University of Waikato (Hogg et al., 2009), University of Groningen (van der Plicht et al., 1995), Heidelberger Akademie der Wissenschaften (Hua et al., 2009), CSIR, Pretoria (Vogel and van der Plicht, 1993) and Center for Accelerator Mass Spectrometry and the University of California, Irvine (Taylor and Southon, 2013). We also used single-year data from Miyake et al. (2012), Jull et al. (2014), Usoskin et al. (2013) and Wang et al. (2017).

The IntCal13 data can be viewed either as conventional ^{14}C ages (yr BP) or $\Delta^{14}\text{C}$ (‰) values. The former is predominantly used for ^{14}C dating purposes. The $\Delta^{14}\text{C}$ values are corrected for the radioactive decay of ^{14}C , and can be thought of as the change in atmospheric ^{14}C concentration in each year. We use the dataset of $\Delta^{14}\text{C}$ values for our investigations.

2.3. Data interpolation

In the IntCal13 dataset, single-year measurements are available from 1510 CE onwards while the sample frequency before that date is between 5 and 10 years. Furthermore, data collected from different groups, is not precisely identical, resulting in multiple values for the same years. Taking into account the statistical uncertainty that comes with measurements on tree-rings, we need to represent the data as a 1D signal, which we achieve by linearly interpolating consecutive data points.

Fig. 1. The main steps of the proposed methodology.

¹ <http://intcal.qub.ac.uk/intcal13/>

In Fig. 2, we present a chunk of the raw signal between 390 and 415 CE. The stars and triangles show the $\Delta^{14}\text{C}$ data from two different research groups. In the years 395 and 405 CE we have two values from the University of Washington and one from Queen's University, Belfast. The statistical uncertainty is shown with error bars. The linearly interpolated signal is shown in Fig. 2 with dots connected by straight lines. In cases of multiple data within the same year, we compute the arithmetic mean between the 1σ uncertainties of the lowest and highest data points.

2.4. Ground truth

In order to test our methods we used as ground truth (GT) the Miyake events for the periods 774–775 CE and 993–994 CE (Miyake et al., 2012), (Miyake et al., 2013) and the period 3372–3371 BCE (Wang et al., 2017) where a similar, as yet unconfirmed, sudden increase in $\Delta^{14}\text{C}$ has been found. We also took into consideration the 660 BCE found by Park et al. (2017), even though it seems that the increment in $\Delta^{14}\text{C}$ spans three years and thus one may argue that it is not a Miyake event.

2.5. Pattern matching

2.5.1. Sliding window technique

For the pattern matching techniques applied in this work, we used a sliding window method, as shown in Fig. 3. We first broke the original signal $f(t) = \Delta^{14}\text{C}$ into windows of the same length w and overlap of size h . Then, every window was compared with a prototype signal using a pattern matching technique such as Dynamic Time Warping (DTW) (Berndt and Clifford, 1994), or a distance based measurement such as Euclidean.

The window has typically the same size as the prototype signal, unless the method also allows different sizes between window and prototype signals. Here, we used a window size of 15 years that is based on the size of the single-year data that are available from several research groups. In the supplementary material (Fig. S1) we showed that changing the size of this window will only change the amplitude of the response vectors and not their positions in time. We used an overlap h of one year for making all the possible comparisons.

Fig. 2. Data interpolation procedure. IntCal13 data points are marked with stars and triangles. The straight lines connecting the dots indicate linear interpolation between consecutive data points. In years where multiple data exist, we consider the arithmetic mean value between the highest and lowest uncertainty values of the extreme data points.

Fig. 3. Sliding window technique. The original signal is separated into windows of size w and overlap of size h . Then, each window is compared with a prototype signal of size p . The result of every comparison is stored in a vector representing the response signal.

2.5.2. Window rescaling

The values in the prototype and the window signals were rescaled to the range $[0, 1]$. This is necessary because absolute values for $\Delta^{14}\text{C}$ vary through time. One example of such rescaling is shown in the supplementary material in Fig. S2.

Here, we need to mention that rescaling the $\Delta^{14}\text{C}$ values over the time windows may cause complications, because small fluctuations in quiet periods will be amplified and it is possible that similar patterns to Miyake events could be created. In such cases the methods will return high similarity to Miyake events where in truth there is no similarity. We assume that such phenomena are rare.

2.6. Signal processing

We apply the following five methods to the IntCal13 dataset, using a sliding window technique: Euclidean distance, correlation coefficients, cross correlation, dynamic time warping and COSFIRE. Since only COSFIRE filters yielded satisfactory results we provided an extended explanation of that method in Section 2.6.5. The plots in Fig. S3 in the supplementary material indicate that none of the other methods work well for the task at hand. In the following definitions we denote by W the signal in a given window and by P the prototype signal.

2.6.1. Euclidean distance

The Euclidean distance is one of the simplest measurements of similarity between two vectors or signals. It is very successful for many applications and it is a common practice to use it as a first attempt in many tasks. It is defined as the square root of the sum of the squares of the differences between the corresponding points of two signals:

$$d(W, P) = \sqrt{\sum_{i=1}^n (W_i - P_i)^2} \quad (1)$$

2.6.2. Correlation coefficients

For every comparison between a window and a prototype signal, we calculated the so-called Pearson correlation coefficient $\rho(W, P)$ that is a measure of their linear dependence:

$$\rho(W, P) = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{W_i - \mu_W}{\sigma_W} \right) \left(\frac{P_i - \mu_P}{\sigma_P} \right) \quad (2)$$

where μ_W , μ_P , σ_W and σ_P are the means and standard deviations of the signals, respectively.

2.6.3. Cross correlation

2.6.3.1. Raw correlations. The cross correlation function (Vorburger et al., 2011) is used in signal processing for a variety of applications including weather forecasting (Leese et al., 1971), water traffic tracking (Hunaidi, 2000), bio-signals (Musha, 2002), telecommunications (Traill et al., 2000) and others. It measures the degree to which two time signals are correlated, and it is defined as their dot product:

$$(W * P)(\tau) \triangleq \sum_{t=-\infty}^{\infty} W^*(t)P(t + \tau) \quad (3)$$

where W^* indicates the complex conjugate² of W , t indicates the time and τ the time lag of one signal relative to the other.

2.6.3.2. Normalized correlations. In our experiments, we made a grid search between the following normalization options:

$$\text{Biased} = \frac{1}{N} (W * P)(\tau) \quad (4)$$

$$\text{Unbiased} = \frac{1}{N - |\tau|} (W * P)(\tau) \quad (5)$$

$$\text{Coeff} = \frac{1}{(W * W)(0)(P * P)(0)} (W * P)(\tau) \quad (6)$$

where N is the size of vector W . The zero lag in Eq. (6) means that the two signals are aligned.

2.6.4. Dynamic time warping (DTW)

The DTW method has been widely used in speech recognition (Amin and Mahmood, 2008), but it can be applied to any time signals (Berndt and Clifford, 1994). The main advantage of this method is that it allows for temporal tolerance and for temporal variances along the given signals. In other words, similar patterns that may vary in duration or in their accelerations or decelerations are still considered similar to some extent.

The DTW arranges two sequences in the form of a grid where the tested sequence is placed on the top left and the test sequence on the bottom of the grid. Then, in each cell of the grid a distance measure is placed which compare the corresponding elements of the two sequences. These comparisons create an optimal path between the two sequences that minimizes all the distance measures, and finally the total distance between the two sequences is calculated using the so-called “wrapping function”. For more detailed explanation of the method we refer to (Sakoe and Chiba, 1978).

2.6.5. Combination of shifted filter responses (COSFIRE filters)

2.6.5.1. Overview. COSFIRE filters have been shown to be effective in image recognition (Azzopardi and Petkov, 2013), and in digital signal processing for 1D musicological signals (Neocleous et al., 2015).

A COSFIRE filter is configured by using some properties of keypoints within a given prototypical signal. The midpoint of this set of keypoints lies at the center of the prototype. To illustrate the effectiveness of a COSFIRE filter in terms of amplitude and noise tolerance, we present an example of a synthetic signal in Fig. 4a, consisting of four different patterns. The first pattern shows a step function signal that we used as a prototype to configure a COSFIRE filter. In the next pattern, we used a random signal to show that the filter is not sufficiently responsive.

Fig. 4. (a) Input 1D signal. The first step function is used as a prototype to configure a COSFIRE filter. (b) The responses of the filter to the input signal. The highest responses are achieved for the prototype signal, but strong responses are also recorded at the center of similar but noisy patterns. Negligible responses are achieved for the random pattern.

Finally, the two last patterns in Fig. 4a are similar to the prototype with 50% added noise. The last one is also amplified by 150%. In Fig. 4b we present the response signal of the COSFIRE filter that was configured using the first pattern (prototype). It is shown that it returns maximum responses to the center of the prototype and high peaks in the third and fourth patterns. Lower responses are returned obtained along the random pattern.

2.6.5.2. Configuration of a COSFIRE filter. Each point i of a COSFIRE filter is described by a pair (C_i, t_i) , where C_i is the value of the signal at time point t_i , around the center of the filter support. We denote by A_c a COSFIRE filter that is defined as a set of such pairs:

$$A_c = \{(C_i, \rho_i) | i = 1 \dots n\} \quad (7)$$

where $\rho_i = \delta(i - (n + 1)/2)$, n is the total number of considered time points and δ is the length of the interval between the time points.

Similarity function. We used a Gaussian kernel function to compute a similarity value for each pair in the set A_c that defines a COSFIRE filter at a given point in time t of a test signal T :

$$D_i(t) = \exp \left(-\frac{(C_i - T_{t+\rho_i})^2}{2\sigma_i^2} \right) \quad (8)$$

where C_i is the preferred value of the i -th pair in the set A_c , and $T_{t+\rho_i}$ is the corresponding value in the concerned neighborhood of a signal T at time t .

The standard deviation $\hat{\sigma}_i$ of the Gaussian kernel function is defined as the multiplication factor of the standard deviation between the samples of the first Miyake event that are available from different groups. Therefore, for every point away from the support of the filter, the standard deviation $\hat{\sigma}_i$ may have a different value. The multiplication factor is optimized using a grid search.

We denote by $R(t)$ the response of a COSFIRE filter at time t that is computed as the geometric mean of all involved Gaussian kernel responses:

$$R(t) = \left(\prod_{i=1}^n D_i(t) \right)^{\frac{1}{n}} \quad (9)$$

² The complex conjugate of a complex number has equal real part and an imaginary part with opposite sign.

Fig. 5. The raw $\Delta^{14}\text{C}$ signal is shown with a black line and the detrended signal with a grey line. Better feature representation is achieved using the detrended signal. The ground truth data (Miyake events) are shown with vertical dashed lines.

For a detailed explanation of the 1D COSFIRE filters we refer to Neocleous PhD³ (Neocleous, 2016).

2.7. Feature extraction

We use a number of spectral features to explore the potential value for the identification of similar Miyake events in the $\Delta^{14}\text{C}$ signal. While we examined seventeen spectral features only eight yielded compelling results: 1) Spectral centroid, 2) Spectral spread, 3) Spectral skewness, 4) Spectral kurtosis, 5) Spectral roll-off 95, 6) Spectral entropy, 7) Spectral flatness, and 8) Spectral roughness.

We achieved better results when we applied a detrending function to the signal. In Fig. 5 we show the raw $\Delta^{14}\text{C}$ in black signal with black line and the detrended signal in grey. This detrending function essentially subtracts the mean or a best-fit line from the respective dataset and the detrended signal is used for the feature extraction.

2.7.1. Spectral centroid

The spectral centroid feature describes the central tendency of the spectrum and it is defined as:

$$\text{Centroid} = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)} \quad (10)$$

where $x(n)$ represents the amplitude of the frequency of bin number n , $f(n)$ is the center frequency of that bin, and N is the size of the discrete Fourier transform (DFT).

2.7.2. Spectral spread

The spectral spread is a measure of the dispersion of the distribution of the DFT. It is described as the standard deviation of the spectrum:

$$\text{Spread} = \sqrt{\frac{1}{N} \sum_{n=1}^N (x_i - \bar{x})^2} \quad (11)$$

2.7.3. Spectral roll-off 95

The roll-off 95 feature is a measure of the high frequency content in the spectrum. It defines the frequency in which the energy of the spectrum below that frequency is 95% of the total energy.

2.7.4. Entropy

The entropy of a spectrum was introduced by Shannon (1948) and it has been used abundantly in information theory. It can, however, be used for any type of data. The Shannon entropy is defined as:

$$\text{Entropy} = - \sum_{i=1}^n P(x_i) \log_b P(x_i) \quad (12)$$

where $p(x_i)$ is the probability for the signal to take values x_i , and b is the base of the logarithm. Here we use $b = 10$.

2.7.5. Flatness

This is the ratio between the geometric mean and the arithmetic mean and it is a measure of the “smoothness” of the spectrum:

$$\text{Flatness} = \frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\frac{\sum_{n=0}^{N-1} x(n)}{N}} \quad (13)$$

2.7.6. Spectral skewness

The coefficient of the skewness of the data is a measure of the symmetry of the spectrum. It is defined as the ratio of the skewness to the standard deviation σ , raised to the third power:

$$\text{Skewness} = \frac{\sum_{i=1}^N (x_i - \bar{x})^3 / N}{\sigma^3} \quad (14)$$

2.7.7. Spectral kurtosis

Kurtosis or “excess kurtosis” is a measure of whether the data have outliers (heavy-tailed) or not (light-tailed), relative to a normal distribution:

$$\text{Kurtosis} = \frac{\sum_{i=1}^N (x_i - \bar{x})^4 / N}{\sigma^4} - 3 \quad (15)$$

2.7.8. Spectral roughness

This feature is mainly used for music information retrieval (MIR) since it aims to explain a psychoacoustical effect which occurs when two tones appear close to each other and the human brain perceives them as one tone. The sound in this case is characterized as rough and this feature aims to give a rank to this roughness. It was proposed by Sethares (1998) and it is essentially the average of the dissonance between all possible pairs of spectral peaks. Although this feature has been introduced for MIR, it may also be effective in other applications, as turned out to be the case in our experiments.

3. Results

3.1. Overview

In Figs. 6–8 and in Fig. S4 in the supplementary material we present the results of the methods used in this study. In Fig. 6 we present the response signal from (a) a COSFIRE filter configured with the $\Delta^{14}\text{C}$ data in the period 774–775 CE (first Miyake event), (b) the values of the “Centroid” feature that is representative of the features “Spread”, “Roll-off 95%”, “Entropy” and “Flatness”, (c) the values of the “Skewness” feature that is representative to the “Kurtosis” (see Fig. S4 in the supplementary material) and (d) the “Roughness” feature. The COSFIRE filter response

³ https://www.rug.nl/research/portal/files/37549216/Complete_thesis.pdf

Fig. 6. (a) COSFIRE response signal. (b) Spectral centroid. This feature is similar to the spectral spread, roll-off 95%, entropy and flatness. (c) Spectral skewness. This feature is similar to spectral kurtosis. (d) Spectral roughness. Spectral features are rescaled by dividing with the maximum value. The arrows indicate the four ground truth dates.

signal, the “Centroid” and the “Roughness” features have local maxima around the two Miyake events. This is shown in more detail in Fig. 7. The COSFIRE filter has maximum response of 1 around the first Miyake event (774–775 CE), because this one is the period we use for training.

The only spectral feature that returned a different response signal (less noisy) is the spectral roughness. It is interesting that even though only very few peaks appear across the feature vector, we observed some low responses precisely at the ages of both Miyake events (Fig. 7i). We need to emphasize that the feature representation is just a transformation of the original signal and that no training procedure is involved. Therefore, the peaks in the preferred dates as shown in Fig. 7i are among the most interesting results of this work. However, most of the

Fig. 7. COSFIRE response signal and feature values for the two Miyake events. COSFIRE achieves maximum response for the first Miyake event because it was used for training. Another smaller peak is returned in the second Miyake event. It is shown that spectral features are also able to identify the two Miyake events (maximum values in (b, c, f, g, h and i) and minimum values in (d and e)). The red vertical lines indicate the two Miyake events.

Fig. 8. False positive rate of the spectral features and the signal processing techniques used in this work, at the true positive rate of 75%. For the signal processing techniques one of the ground truth is used for training and therefore it is always identified by the methods. Therefore the FPR reported here is actually at a TPR of 50%.

spectral features identified both Miyake events. The features “Skewness” and “Kurtosis” returned a minimum instead of maximum at those dates and they are generally noisy. From Figs. 6, 7 and S4 (in the supplementary material), we observe that the values of the spectral features, with the exception of the “roughness feature”, contain local maxima at the points of interest, but they also contain local maxima in many other irrelevant dates, resulting in many false positives.

3.2. Evaluation protocol

As explained in Section 2.4, we used a number of dates that we call ground truth data for validating the methods used in this work. We compute our results in terms of True Positive Rate (TPR) and False Positive Rate (FPR) (Eqs. 16 and 17). To compute the evaluation measures, we needed to identify the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). We call a TP the outcome where there is a local maximum in the output signal within a distance of three years from a GT data point. A FP is defined to occur when a local maximum of the output signal appears at a date that does not coincide within three years of a GT data point. A TN is defined when no GT data are defined and no output signal is present at a given date. A FN classification is defined to occur when a GT data point is not matched by any local maximum point of the output signal.

A positive output signal is considered of interest if that value is above a certain threshold. The higher the value of the threshold, the lower FPR and TPR.

$$\text{TPR} = \text{TP} / (\text{TP} + \text{FN}) \quad (16)$$

$$\text{FPR} = \text{FP} / (\text{FP} + \text{TN}) \quad (17)$$

3.3. Fine-tuning of parameters

In order to fine-tune the parameters of the methods used in this study, we performed a grid search with different values and we measured the results in terms of false positive rate. We use as training pattern the first Miyake event and as test the second Miyake event. Then, we chose the parameters that identify the second Miyake event at the lowest false positive rate.

For the COSFIRE method, we applied a grid search of the $\hat{\sigma}$ parameter by multiplying it with factors between 0.5 and 3.5. Here, the best results are achieved with a multiplication factor of 2.5.

For the Cross Correlation method, we used three different normalization options and one without normalization: a) “none” which correlates the two sequences without normalization, b) “biased”, c) “unbiased” for biased or unbiased estimates of the cross-correlation function (Bendat and Piersol, 2011) and d) a normalization option so that the autocorrelations at zero lag equal 1, abbreviated as “coeff”. The latter returned the best results.

For the DTW method, we used four different distance metrics: a) Euclidean b) Manhattan, c) square of Euclidean and d) symmetric Kullback-Leibler. The best results were achieved by using the square of Euclidean.

For the spectral features we make a grid search by changing the window size of the Fast Fourier Transform using the values 5, 10, 15, ..., 65 data points.

In general, the bigger the window for the FFT, the higher frequency resolution is achieved, with the cost of having lower temporal resolution. For this task, we prefer to have high temporal resolution since the Miyake events appear within one year, and this is what we aim to achieve. In Figs. S5–S7 in the supplementary material we show some examples of how the spectrum, the spectral centroid and the roll-off95 are influenced by changing the window of the FFT. In these examples we present the outputs of the above-mentioned features for windows of 25 and 250 data points.

The best window size for “Centroid”, “Spread”, “Roll-off95”, “Entropy” and “Flatness” is 5, for “Skewness” and “Kurtosis” is 10 and for “Roughness” is 25. The other two methods, namely correlation coefficients and Euclidean distance, did not have any parameters to tune.

3.4. Comparison with the ground truth data

In Fig. 8 we present the false positive rates of all methods used in this study (feature representation and signal processing techniques) at the 75% TPR. It is shown that the best results are achieved by the COSFIRE filters, but comparable results were also yielded by the spectral features entropy, flatness and roughness. Further, we need to emphasize that for the signal processing techniques one of the ground truth data is used for training (first Miyake event) and therefore the FPR here is essentially at 50% TPR.

4. Discussion

In this work we apply several state-of-the-art signal processing techniques, as well as spectral feature representations to the atmospheric $\Delta^{14}\text{C}$ data in order to propose dates for anomalies similar to the two published Miyake events. Our study spanned the years ~12,000 BCE to 1950 CE. From the signal processing point of view, most of the methods do not perform well. As can be seen in Fig. S3 in the supplementary material, four popular methods return a noisy output that cannot have a meaningful interpretation.

COSFIRE filters appear to perform well for this particular task. We configure a COSFIRE filter that is trained with the first Miyake event (774–775 CE) and we applied that filter to the remaining signal. A local maximum appeared in the years 993–994 CE (second Miyake event), where no such peak is easily observable in the IntCal13 data. This result confirms the effectiveness and potential of such filters. Furthermore, in order to strengthen this conclusion we applied cross validation by training a COSFIRE filter with the second Miyake event and, indeed, we got a local maximum response at the first event.

From the feature representation techniques, next to the second Miyake event we observed another local peak, which is visible in Fig. 7. One possible reason that we get peaks in the spectral features is because when there is a discontinuity in the signal usually there is

energy in the whole spectrum due to the noise that is created from the fast Fourier transform. This fact also explains why most of the spectral features yield a similar shape over time, but at different magnitudes, as shown in Fig. S4 in the supplementary material.

Regarding the pre-processing done in the data, we needed to apply some interpolation the data was not captured with a consistent frequency. In this paper we applied linear interpolation, however, other and more sophisticated non-linear interpolation methods could be investigated.

5. Future work

We find it interesting that the response signals of the methods used in this study achieve peaks at dates that are not part of the ground truth. These peaks warrant closer investigation to see whether they might actually be Miyake events that have not been discovered. One direction could be to compare the response signals of all methods and analyze the relevance of common peaks for the same dates. Also, we are interested in investigating if there is a possibility of a repetitive pattern of such Miyake events, over time. This can be achieved by studying the most common peaks from the response signals.

6. Conclusion

The identification of similar patterns in a time signal has been a difficult task in the area of signal processing and machine learning. In this work, we showed that some state-of-the-art methods can effectively identify patterns that may prove to be of interest, such as past Miyake events, in a sophisticated manner. COSFIRE filters seemed to be the most appropriate within the signal processing techniques that we study in this paper. Feature extraction and representation looks promising for the task at hand and their use seems to be valuable. Spectral features and COSFIRE filters were able to identify 75% of the anomalies sought with a false positive rate of <10%.

Computational methods are proving more and more promising for such tasks, and they may help provide important insight into these and other palaeoenvironmental data.

Acknowledgements

This work is funded by an ERC research project (ECHOES) with grant number 714679.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2019.01.251>.

References

- Amin, T.B., Mahmood, I., 2008. Speech recognition using dynamic time warping. *Advances in Space Technologies. ICAST 2008, 2nd International Conference on*. IEEE, pp. 74–79.
- Azzopardi, G., Petkov, N., 2013. Trainable COSFIRE filters for keypoint detection and pattern recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 490–503.
- Bendat, J.S., Piersol, A.G., 2011. *Random Data: Analysis and Measurement Procedures*. vol. 729. John Wiley & Sons.
- Berndt, D.J., Clifford, J., 1994. Using dynamic time warping to find patterns in time series. *KDD Workshop*. vol. 10, pp. 359–370 Seattle, WA.
- Büntgen, U., Wacker, L., Galván, J.D., Arnold, S., Arseneault, D., Baillie, M., Beer, J., Bernabei, M., Bleicher, N., Boswijk, G., et al., 2018. Tree rings reveal globally coherent signature of cosmogenic radiocarbon events in 774 and 993 CE. *Nat. Commun.* 9.
- de Vries, H., 1958. Variation in concentration of radiocarbon with time and location on earth. *Proc. Koninkl. Nederl. Akad. Wetenschappen*, B 61, 1–9.
- Dee, M., Pope, B., Miles, D., Manning, S., Miyake, F., 2017. Supernovae and single-year anomalies in the atmospheric radiocarbon record. *Radiocarbon* 59, 293–302.
- Güttler, D., Adolphi, F., Beer, J., Bleicher, N., Boswijk, G., Christl, M., Hogg, A., Palmer, J., Vockenhuber, C., Wacker, L., et al., 2015. Rapid increase in cosmogenic ^{14}C in AD 775 measured in New Zealand kauri trees indicates short-lived increase in ^{14}C production spanning both hemispheres. *Earth Planet. Sci. Lett.* 411, 290–297.
- Hambaryan, V., Neuhäuser, R., 2013. A galactic short gamma-ray burst as cause for the ^{14}C peak in AD 774/5. *Mon. Not. R. Astron. Soc.* 430, 32–36.

- Hogg, A., Palmer, J., Boswijk, G., Reimer, P., Brown, D., 2009. Investigating the interhemispheric ^{14}C offset in the 1st millennium AD and assessment of laboratory bias and calibration errors. *Radiocarbon* 51, 1177–1186.
- Hua, Q., Barbetti, M., Fink, D., Kaiser, K.F., Friedrich, M., Kromer, B., Levchenko, V.A., Zoppi, U., Smith, A.M., Bertuch, F., 2009. Atmospheric ^{14}C variations derived from tree rings during the early younger Dryas. *Quat. Sci. Rev.* 28, 2982–2990.
- Hunaidi, O., 2000. Detecting Leaks in Water Distribution Pipes. Institute for Research in Construction, National Research Council of Canada.
- Jull, A., Panyushkina, I.P., Lange, T.E., Kukarskih, V.V., Myglan, V.S., Clark, K.J., Salzer, M.W., Burr, G.S., Leavitt, S.W., 2014. Excursions in the ^{14}C record at AD 774–775 in tree rings from Russia and America. *Geophys. Res. Lett.* 41, 3004–3010.
- Leese, J.A., Novak, C.S., Clark, B.B., 1971. An automated technique for obtaining cloud motion from geosynchronous satellite data using cross correlation. *J. Appl. Meteorol.* 10, 118–132.
- Libby, W.F., Anderson, E.C., Arnold, J.R., 1949. Age determination by radiocarbon content: world-wide assay of natural radiocarbon. *Science* 109, 227–228.
- McCormac, F., Bayliss, A., Brown, D., Reimer, P., Thompson, M., 2008. Extended radiocarbon calibration in the Anglo-Saxon period, AD 395–485 and AD 735–805. *Radiocarbon* 50, 11–17.
- Mekhaldi, F., Muscheler, R., Adolphi, F., Aldahan, A., Beer, J., McConnell, J.R., Possnert, G., Sigl, M., Svensson, A., Synal, H.-A., et al., 2015. Multiradionuclide evidence for the solar origin of the cosmic-ray events of AD 774/5 and 993/4. *Nat. Commun.* 6, 8611.
- Miyake, F., Nagaya, K., Masuda, K., Nakamura, T., 2012. A signature of cosmic-ray increase in AD 774–775 from tree rings in Japan. *Nature* 486, 240–242.
- Miyake, F., Masuda, K., Nakamura, T., 2013. Another rapid event in the carbon-14 content of tree rings. *Nat. Commun.* 4, 1748.
- Miyake, F., Jull, A.T., Panyushkina, I.P., Wacker, L., Salzer, M., Baisan, C.H., Lange, T., Cruz, R., Masuda, K., Nakamura, T., 2017a. Large ^{14}C excursion in 5480 BC indicates an abnormal sun in the mid-Holocene. *Proc. Natl. Acad. Sci.* 201613144.
- Miyake, F., Masuda, K., Nakamura, T., Kimura, K., Hakozaiki, M., Jull, A.T., Lange, T.E., Cruz, R., Panyushkina, I.P., Baisan, C., et al., 2017b. Search for annual ^{14}C excursions in the past. *Radiocarbon* 59, 315–320.
- Musha, T. (2002). Method and apparatus for will determination and bio-signal control. US Patent 6,349,231.
- Neocleous, A., 2016. Computing Expert's Intelligence: a Case in Bio-medicine and a Case in Musicology, PhD Thesis. University of Groningen.
- Neocleous, A., Azzopardi, G., Schizas, C.N., Petkov, N., 2015. Filter-based approach for ornamentation detection and recognition in singing folk music. *International Conference on Computer Analysis of Images and Patterns*. Springer, pp. 558–569.
- Park, J., Southon, J., Fahrni, S., Creasman, P.P., Mewaldt, R., 2017. Relationship between solar activity and $\Delta^{14}\text{C}$ peaks in AD 775, AD 994, and 660 BC. *Radiocarbon* 59, 1147–1156.
- Pavlov, A., Blinov, A., Vasilyev, G., Vdovina, M., Volkov, P., Konstantinov, A., Ostryakov, V., 2013. Gamma-ray bursts and the production of cosmogenic radionuclides in the Earth's atmosphere. *Astron. Lett.* 39, 571–577.
- Reimer, P.J., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Ramsey, C.B., Buck, C.E., Cheng, H., Edwards, R.L., Friedrich, M., et al., 2013. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* 55, 1869–1887.
- Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Process.* 26, 43–49.
- Sethares, W.A., 1998. Consonance-based spectral mappings. *Comput. Music. J.* 22, 56–72.
- Shannon, C.E., 1948. A note on the concept of entropy. *Bell Syst. Tech. J.* 27, 379–423.
- Stuiver, M., Reimer, P.J., Braziunas, T.F., 1998. High-precision radiocarbon age calibration for terrestrial and marine samples. *Radiocarbon* 40, 1127–1151.
- Sukhodolov, T., Usoskin, I., Rozanov, E., Asvestari, E., Ball, W.T., Curran, M.A., Fischer, H., Kovaltsov, G., Miyake, F., Peter, T., et al., 2017. Atmospheric impacts of the strongest known solar particle storm of 775 AD. *Sci. Rep.* 7, 45257.
- Taylor, R., Southon, J., 2013. Reviewing the mid-first millennium BC ^{14}C “warp” using 14 C/bristlecone pine data. *Nucl. Instrum. Methods Phys. Res., Sect. B* 294, 440–443.
- Traill, D. M., Sheppard, P. J., Bai, J., & Milner, S. (2000). Echo cancellation using cross-correlation of buffered receive and transmit sample segments to determine cancelling filter coefficients. US Patent 6,078,567.
- Usoskin, I.G., Kovaltsov, G.A., 2012. Occurrence of extreme solar particle events: assessment from historical proxy data. *Astrophys. J.* 757, 92.
- Usoskin, I., Kromer, B., Ludlow, F., Beer, J., Friedrich, M., Kovaltsov, G., Solanki, S., Wacker, L., 2013. The AD775 cosmic event revisited: the Sun is to blame. *Astron. Astrophys.* 552, L3.
- Uusitalo, J., Arppe, L., Hackman, T., Helama, S., Kovaltsov, G., Mielikäinen, K., Mäkinen, H., Nöjd, P., Palonen, V., Usoskin, I., et al., 2018. Solar superstorm of ad 774 recorded subannually by arctic tree rings. *Nat. Commun.* 9, 3495.
- van der Plicht, J., Jansma, E., Kars, H., 1995. The “Amsterdam Castle”: a case study of wiggle matching and the proper calibration curve. *Radiocarbon* 37, 965–968.
- Vogel, J., van der Plicht, J., 1993. Calibration curve for short-lived samples, 1900–3900 BC. *Radiocarbon* 35, 87–91.
- Vorburger, T.V., Song, J.-F., Chu, W., Ma, L., Bui, S., Zheng, A., Renegar, T., 2011. Applications of cross-correlation functions. *Wear* 271, 529–533.
- Wang, F., Yu, H., Zou, Y., Dai, Z., Cheng, K., 2017. A rapid cosmic-ray increase in BC 3372–3371 from ancient buried tree rings in China. *Nat. Commun.* 8, 1487.