Trainable Filters for the Identification of Anomalies in Cosmogenic Isotope Data

ANDREAS NEOCLEOUS[1], **GEORGE AZZOPARDI**[2], **MARGOT KUITEMS**[1], **ANDREA SCIFO**[1], **AND MICHAEL DEE**[1]

[1]Center for Isotope Research, University of Groningen, 9747 AG Groningen, The Netherlands
[2]Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, 9747 AG Groningen, The Netherlands

Corresponding author: Andreas Neocleous (a.c.neocleous@rug.nl)

ABSTRACT Extreme bursts of radiation from space result in rapid increases in the concentration of radiocarbon in the atmosphere. Such rises, known as Miyake Events, can be detected through the measurement of radiocarbon in dendrochronological archives. The identification of Miyake Events is important because radiation impacts of this magnitude pose an existential threat to satellite communications and aeronautical avionics and may even be detrimental to human health. However, at present, radiocarbon measurements on tree-ring archives are generally only available at decadal resolution, which smooths out the effect of a possible radiation burst. The Miyake Events discovered so far, in tree-rings from the years 3372-3371 BCE, 774-775 CE, and 993-994 CE, have essentially been found by chance, but there may be more. In this paper, we use signal processing techniques, in particular COSFIRE, to train filters with data on annual changes in radiocarbon (Δ^{14}C) around those dates. Then, we evaluate the trained filters and attempt to detect similar Miyake Events in the past. The method that we propose is promising, since it identifies the known Miyake Events at a relatively low false positive rate. Using the findings of this paper, we propose a list of 26 calendar years that our system persistently indicates are Miyake Event-like. We are currently examining a short-list of five of the newly identified dates and intend to perform single-year radiocarbon measurements over them. Signal processing techniques, such as COSFIRE filters, can be used as guidance tools since they are able to identify similar patterns of interest, even if they vary in time or in amplitude.

INDEX TERMS Radiocarbon measurement, digital signal processing, Miyake Events, COSFIRE, pattern matching.

I. INTRODUCTION

The isotope radiocarbon (^{14}C) underpins the eponymous method that enables direct dating of organic remains back to about 50,000 years ago. To apply this method, it is necessary to know how the atmospheric concentration of ^{14}C has varied over time. This is primarily achieved by measuring the ^{14}C concentration of tree-rings of known age, as they retain the signal of the atmospheric CO_2 absorbed each year during photosynthesis. Furthermore, because ^{14}C radioactively decays, in order to reconstruct past concentrations of ^{14}C, it is necessary to correct for the loss due to decay in each of the known-age samples. The estimates of past ^{14}C concentrations that result are denoted Δ^{14}C [1]. It has long been known that Δ^{14}C has fluctuated over time [2]. These fluctuations,

however, were assumed to be minor ($\sim$1–2%) from one year to the next and therefore estimates of Δ^{14}C, have generally been obtained on blocks of 5-10 tree-rings. This assumption was disproven by Miyake *et al.* who made single-year measurements on Japanese tree-rings and found rapid increments of Δ^{14}C ($>$12%) between the years 774 - 775 CE [3] and 993 - 994 CE [4]. These sudden increases were subsequently coined Miyake Events and their amplitude can vary. The first Miyake Event, illustrated in Fig. 1, has since been confirmed by other ^{14}C laboratories on dendrochronogical archives from Germany [5], the USA, Russia [6] and New Zealand [7]; and the second, by teams in Denmark and Poland [8]. Another similar event has been identified by Wang *et al.* [9] in 3372 - 3371 BCE, and an analogous but slightly slower uplift has also been found around 660 BCE by Park *et al.* [10]. The possible reasons for these sudden rises in radiocarbon production have been widely debated in the literature, and

---

The associate editor coordinating the review of this manuscript and approving it for publication was Siddhartha Bhattacharyya.
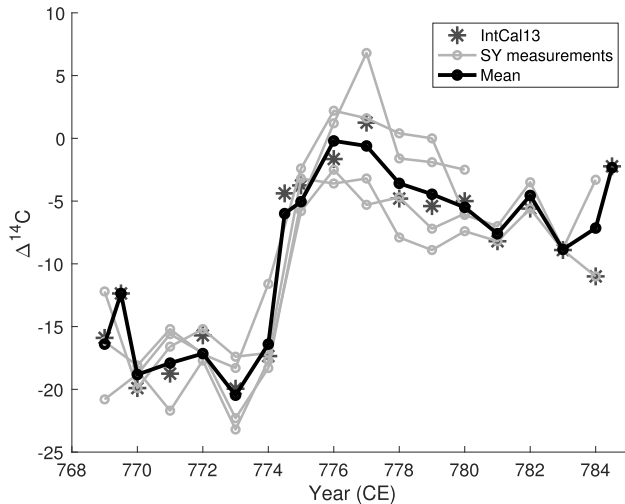
FIGURE 1. Linearly interpolated data from the first Miyake Event around 774 CE. The star symbols show the IntCal13 data and the light gray lines show the single-year data (SY) from different ^{14}C laboratories. The mean value of all the data points per year is shown with black spots.

FIGURE 2. The main steps of the proposed methodology.

the leading hypothesis is that they were caused by extreme solar energetic particle events [5], [7], [11], [12]. Other origins such as γ-ray sources could also have generated similar effects [3], [11], [13], [14].

The ability to identify and predict Miyake Events is important because it could help mitigate potentially dangerous cosmic radiation impacts, especially for aeronautical avionics and global telecommunication systems. In the literature, there have been some attempts at identifying similar events using the IntCal13 dataset [15], which is the most comprehensive dataset of decadal Δ^{14}C measurements. The most common method, applied by Wang *et al.* [9], Miyake *et al.* [16] and others, is to compute the percentage change between successive samples in the IntCal13 data. That approach is not always sufficiently reliable, however, for several reasons. Firstly, because the sample rate is five to ten years and therefore taking the percentage between successive samples can yield many false outcomes. Indeed, the difference between successive samples can sometimes extend to decades. Additionally, in many cases data from different ^{14}C laboratories vary substantially.

In this work, we use a signal processing technique to identify and predict similar patterns to the Miyake events. In particular, we use the Combination of Shifted Filter Responses (COSFIRE) filters, as our previous work showed they outperformed other important signal processing techniques [29]. The strengths of COSFIRE filters lie in their trainable character and their tolerance to some temporal and magnitudinal deviations. In this work, we cross-validate our results across a number of *established* and *speculative* events and finally we suggest a list of new *speculative* Miyake Events that have not been considered previously in the literature.
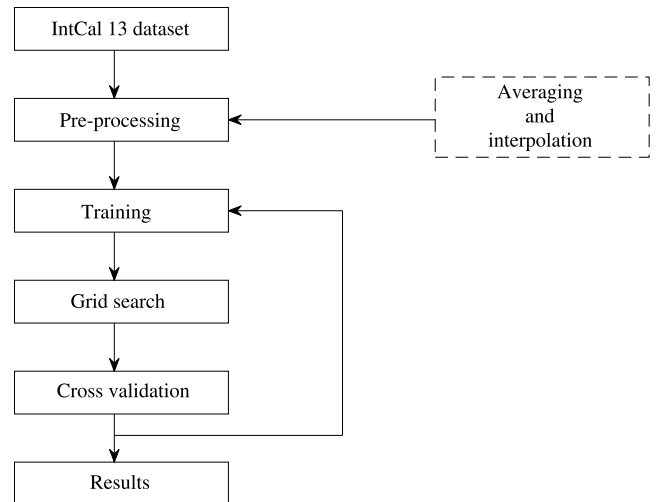
II. METHODS
A. OVERVIEW
The IntCal13 dataset consists of radiocarbon measurements on tree-rings provided by several ^{14}C laboratories. As these ^{14}C laboratories used various tree species that grew in different parts of the world, and because of the natural statistical variability in the measurement of radiocarbon, the raw IntCal13 values scatter to some extent. This means that for the same years there are sometimes multiple values. In order to mitigate this issue, we simply compute their average, such that for each year we deal with a single value. Another major issue is the fact that the sample rate is between 5 and 10 years. We address this matter by interpolating between the averaged values and use the resulting signal in our experiments. Then, we use the COSFIRE method to train a detector that is selective for the three established Miyake Events (774 - 775 CE, 993 - 994 CE and 3372 - 3371 BCE). In our previous work [29] we have shown that COSFIRE filters perform very well on the task of the anomaly detection in cosmogenic data. We fine tune the COSFIRE parameters by a grid search[1] and use the Miyake Events as the validation set. The pipeline of our method is shown in Fig. 2.

B. GROUND-TRUTH DATA
We created two groups of ground-truth (GT), namely *established* and *speculative* Miyake Events. In the *established* group, we include those events that are reported in the literature: a) 774 - 775 CE in [3], b) 994 - 995 CE in [4] and c) 3372 - 3371 BCE in [9]. Single-year measurements across those years are available from the ^{14}C laboratories who carried out the studies and are also used in the training procedure.

---

[1]We define a range of possible values for each one of the COSFIRE parameters and we compute the results on the training set. For the validation, we use the filters that were configured with the parameter values that returned the best training results.

In the *speculative* group, we include the hypothesized events: a) 10750 BCE, b) 10720 BCE, c) 5480 BCE, d) 3077 BCE, e) 1835 BCE, f) 1677 BCE, g) 1588 BCE, h) 660 BCE, i) 400 BCE, j) 544 CE, k) 1220 CE and i) 1859 CE (Carrington flare. A major solar event but a Δ^{14}C spike is visually absent).

The dates 10750 BCE, 10720 BCE and 1220 CE were discussed by Wacker [17]. The events of 3077 BCE, 1677 BCE and 544 CE are speculated by Dee and Pope [18], and the ones in 1835 BCE and 400 BCE are hypothesized by Sturt Manning [personal communications]. The anomalies at 660 BCE and 5480 BCE are reported in Park *et al.* [10] and Miyake *et al.* [16] but the rise in Δ^{14}C appears gradually within 3 to 10 years. Therefore, we cannot consider them as Miyake Events.

C. DATA

We use the atmospheric data from the IntCal13 dataset, which is available online and consists of ^{14}C measurements on tree-rings made by the University of Washington [19], Queen's University Belfast [20], University of Waikato [21], University of Groningen [22], Heidelberger Akademie der Wissenschaften [23], CSIR, Pretoria [24], Center for Accelerator Mass Spectrometry and University of California, Irvine [25]. Single-year data are also available from Miyake *et al.* [3], Usoskin *et al.* [5], and Jull *et al.* [6].

It is common practice to view the IntCal13 data as Δ^{14}C (%) values, instead of the conventional ^{14}C ages (yr BP) used for dating purposes. The Δ^{14}C (%) values are corrected for the radioactive decay of ^{14}C, and can be thought of as the change in atmospheric ^{14}C concentration from one year to the next. The Δ^{14}C record is the dataset used in our study.

D. DATA PRE-PROCESSING
1) AVERAGING AND INTERPOLATION

In time series, low resolution is a common problem and it needs to be addressed before proceeding with any analysis. A typical approach for addressing this issue is to interpolate consecutive values to increase the resolution.

In the IntCal13 dataset, the sample frequency is between 5 and 10 years (low resolution), but in our training patterns we have single-year data (higher resolution). Therefore, the training pattern has many more values than any part of the test signal, as explained above. In order to mitigate this problem, we performed a linear interpolation between values, with frequency every six months. Before doing this, however, in some cases we needed to average values where multiple data exist in the same years.

In Fig. 3, we illustrate one example of this approach. We show a part of the IntCal13 data that is converted into Δ^{14}C, between the years 550 and 580CE. In this example, for the years 555, 565 and 575 CE there are multiple data points coming from the ^{14}C laboratories, for which we simply take the mean value. The linearly interpolated signal is shown with dots that are connected with straight lines and it is the one that is used in our experiments.
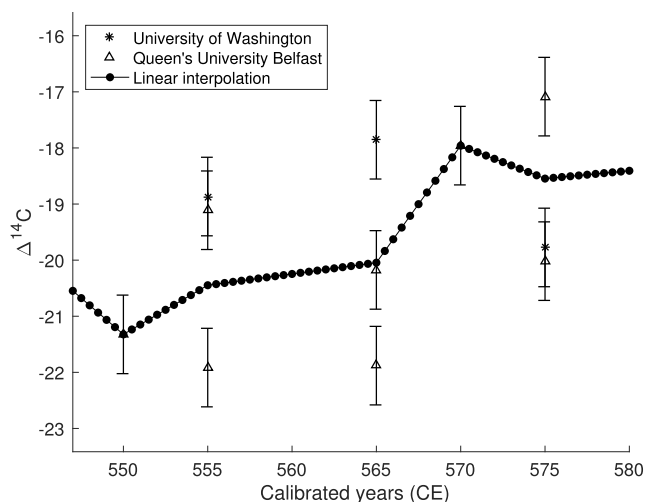

FIGURE 3. Pre-processing procedure. In this example, we present data from the University of Washington (stars) and the Queen's University of Belfast (triangles). The black dots that are connected with straight lines represent the mean and linearly interpolated signal that is used in our experiments.

2) SLIDING WINDOW AND RESCALING

Even though the major feature of the Miyake Events is the sudden increment in Δ^{14}C between consecutive years, a wider pattern which consists of few years before and few years after the event is commonly considered. The Δ^{14}C after the event decreases roughly linearly until it reaches the values that it had beforehand. We need to know that one sudden increase in Δ^{14}C is not just an outlier, which "jumps" back to normal values immediately after, which sometimes can happen because of the natural materials used for radiocarbon analysis.

Most of the single-year measurements around the known Miyake Events that are provided in the literature span a range of about 10 years around the event. For the pattern matching method, we use a training pattern with a window size equal to the data points that are provided. Then, the validation is done by taking the same amount of data from the interpolated IntCal13 dataset (test window), starting from the beginning to the end of the signal and shifting the test window one point at a time. This procedure is repeated until all test windows are validated and the responses of the COSFIRE filters are stored for further analysis.

The training and test windows are rescaled in the range between 0 and 1 as required by the COSFIRE filtering approach.

E. COSFIRE FILTERS
1) OVERVIEW

The COSFIRE filtering approach was initially introduced for the detection of patterns in images [26], and later in digital signal processing for 1D musicological signals [27]. We have shown their effectiveness in 1D cosmogenic data in [29] where COSFIRE filters outperformed other state-of-the-art signal processing techniques. In [26], it is shown that

they are very effective for tasks such as detection of vascular bifurcations and the detection and recognition of traffic signs, for instance. They are trainable and they allow for temporal and amplitudinal tolerance that can be defined with a set of parameters. In this work, we use the 1D COSFIRE filtering approach as introduced in [27].

2) CONFIGURATION OF A COSFIRE FILTER

A COSFIRE filter is configured by determining a set of parameter values from a given prototype signal. These parameters are in the form of pairs (C_i, ρ_i). The parameter C_i contains the value of the prototype (preferred) signal at time point ρ_i, around the center of the filter support which lies at the center of the prototype. We denote by A_c a COSFIRE filter that is defined as a set of such pairs:

$$A_c = \{(C_i, \rho_i)|i = 1 \ldots n\} \qquad (1)$$

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

3) APPLYING COSFIRE FILTERS

A COSFIRE filter is applied to a signal by computing a similarity function between each pair of the filter and the values of the signal. We choose our similarity function to be a Gaussian kernel function because it allows for some amplitudinal tolerance. Then, the response of the COSFIRE filter is computed as the geometric mean of all the similarity values.

4) SIMILARITY FUNCTION

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

$$D_i(t) = \exp^{-\frac{(C_i - T_{t+\rho_i})^2}{2\sigma^2}}, \quad \sigma = \sigma_0 + \alpha(|\rho_i|) \qquad (2)$$

where C_i is the preferred value of the i-th pair in 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 (σ) of the Gaussian kernel function increases linearly with increasing distance from the center of the filter. In this way, we allow more tolerance to the values of time points that are on the periphery of the support of the filter than those that are closer to its center. The constant parameters σ_0 and α are determined empirically.

Generally, the lower the values of these two parameters the more similar a signal has to be to the prototype signal in order for the filter to achieve a high response. With low values of σ_0 and α, a small deviation in shape between a test signal and the prototype signal affect a substantial drop in the COSFIRE response.

5) RESPONSE

We denote by $R(t)$ the response of a COSFIRE filter at time t, which we define as the geometric mean of all

Gaussian kernel responses:

$$R(t) = \left(\prod_{i=1}^{n} D_i(t)\right)^{\frac{1}{n}} \qquad (3)$$

For further technical details on the 1D COSFIRE filters we refer to Neocleous PhD[2] [28].

III. RESULTS

A. OVERVIEW

We use the three *established* Miyake Events as training patterns to configure the COSFIRE filter parameters, as explained in Section III-C. Then, we evaluate the COSFIRE filters on the *speculative* ground-truth group that is presently being suggested by several researchers. We consider these events as "test data" since their existence has not been proven, by way of laboratory single-year radiocarbon measurements.

B. EVALUATION PROTOCOL

To quantify the results, we start by defining the terms that we use, namely true positive (TP), true negative (TN), false positive (FP) and false negative (FN) classifications. We use a threshold value that we apply to the COSFIRE responses in order to obtain positive and negative classifications. A COSFIRE response is considered a TP if it is above the threshold and is within at most five years of a GT year. A FN classification denotes when the response value at a GT position is lower than the threshold. A FP and a TN classification arises when the response values occur at a distance of more than five years from the nearest GT year, and they have values above and below the threshold, respectively. Then, we compute the true positive rate (TPR) and the false positive rate (FPR):

$$TPR = \frac{TP}{TP + FN} \qquad (4)$$

$$FPR = \frac{FP}{FP + TN} \qquad (5)$$

From the TPR and the FPR we generate a receiver operating characteristic curve (ROC), which is obtained by computing the TPR and the FPR for a set of threshold values in a specific range. Typically, a range of different thresholds between the minimum and the maximum value of a response signal is used to measure the values of the TPR and FPR. Both the TPR and the FPR decrease with increasing threshold value. The best results, however, are when TPR is at a maximum and FPR is at a minimum. The ROC curve is the plot of the FPR against TPR and the area under the curve (AUC) is the integral of that function, which can be computed by trapezoidal approximations of that curve.

C. GRID SEARCH AND CROSS-VALIDATION

To cross-validate our system, we configure six COSFIRE filters: three from the IntCal13 dataset, and three from the

---

[2]http://www.rug.nl/research/portal/en/publications/computing-experts-intelligence(9775d485-9396-42cb-aa1d-34737e33da2f).html

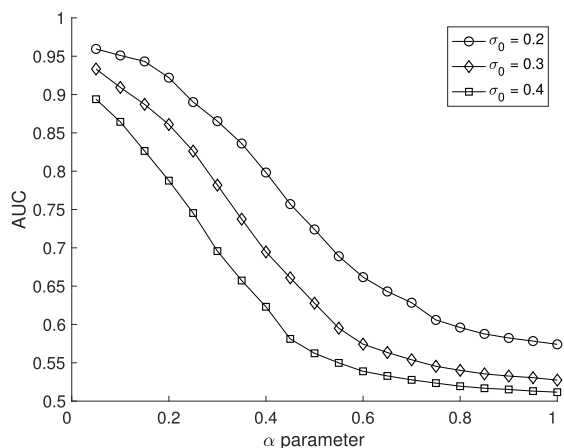**FIGURE 4.** The values of the area under the ROC curve (AUC) for different values of the parameters σ_0 and α of the COSFIRE filters. We performed a grid search in the range between 0 and 1 for both σ_0 and α. Best results are achieved with low values of α but are slightly affected by the σ_0 parameter.

TABLE 1. Grid search results for the six COSFIRE filters and their application in the IntCal13 (IC) dataset and in the single-year data (SY). Here, we present the values of σ_0 and α parameters and the results in terms of the maximum AUC.

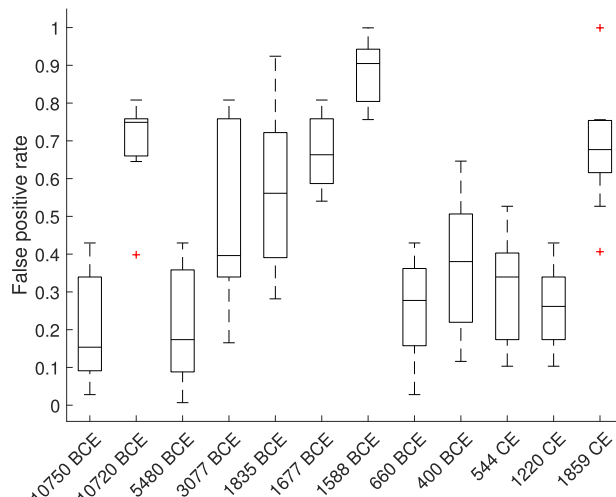| Training event	Applied in IC			Applied in SY		
	σ_0	α	AUC	σ_0	α	AUC
Wang IC	0.24	0.01	0.87	0.3	0.01	0.79
Wang SY	0.27	0.01	0.77	0.14	0.01	0.91
Miyake 1 IC	0.09	0.01	0.97	0.04	0.11	0.78
Miyake 1 SY	0.25	0.01	0.89	0.27	0.02	0.86
Miyake 2 IC	0.28	0.03	0.83	0.23	0.02	0.78
Miyake 2 SY	0.26	0.06	0.82	0.10	0.05	0.99

FIGURE 5. Distributions of the FPR of the *speculative* Miyake Events (test set) for the 12 COSFIRE filter responses. The bottom and top edges of the boxes indicate the 25th and 75th percentiles and the error bars the distribution at 95% probability that a random variable will fall in. The red crosses represent outliers.

single-year dataset, using the *established* Miyake Events from the GT. Then, we apply every COSFIRE filter to both datasets. This produces a total of 12 COSFIRE filter response signals. In the evaluation procedure, the response to the training event is not included in the quantification of TPs, FPs and FNs.

For every COSFIRE filter, we performed a grid search by changing the values of the COSFIRE parameters σ_0 and α between 0 and 1 at intervals of 0.04. We keep the parameter δ constant to ($\delta = 1$), because we want to include all the available single-year data for training. The other parameter to test is the size of the COSFIRE filter. We performed experiments to examine the effect of different filter sizes and we found out that the responses of the COSFIRE filters differ in amplitude and not in temporal positions. Therefore, the filter size is insensitive to the results.

We observed that in most of the cases the AUC has the highest value for $\alpha < 0.1$. The results do not change significantly with the σ_0 parameter. One example of the AUC values for different parameter values is shown in Fig. 4. The figure shows three plots, in one we fix σ_0 to 0.2, another one with $\sigma_0 = 0.3$ and the last one when σ_0 is 0.4. The x-axis shows the values for the parameter α and the y-axis shows a standard performance measurement known as the area under the ROC curve (AUC), which takes into account the number of true positives, false positives and false negatives. These three plots demonstrate that the best results are obtained with small values of α.

In Table 1 we present the values of the σ_0 and α parameters that contribute to the maximum AUC. We then apply the COSFIRE filters with the determined parameters σ_0 and α to the test data.

D. TEST DATA

We use the second group of GT for testing data which consists of *speculative* Miyake Events. For every COSFIRE filter,

we compute the similarity response, which essentially indicates how similar a given pattern is to the Miyake Events used for training. The higher the response, the more likely there is an event of interest. From those response signals, we compute the FPR at the detection of each individual *speculative* Miyake Event.

In Fig. 5, we present the distribution of the FPR that is computed from the 12 COSFIRE responses, for every *speculative* event. It is shown that the events in the years 10750 BCE, 5480 BCE, 660 BCE, 400 BCE, 544 CE and 1220 CE return low FPR. On the contrary, the dates 10720 BCE, 1677 BCE, 1588 BCE and 1859CE return high FPRs and the dates 3077 BCE and 1835 BCE have wide distributions. Indeed, if we compute the results with only the above mentioned *speculative* events (*speculative* test set 1) that return the lowest FPRs, the AUC increases. If we then also remove the date 400 BCE which is less likely than all the others, the AUC increases further still (speculative test set 2).

In Fig. 6, we show the ROC curves for the speculative events. The dashed line with circle markers shows the FPR and the TPR across a range of different thresholds of the entire test set. The events in the speculative test set 2 can be identified with an FPR of 18%.
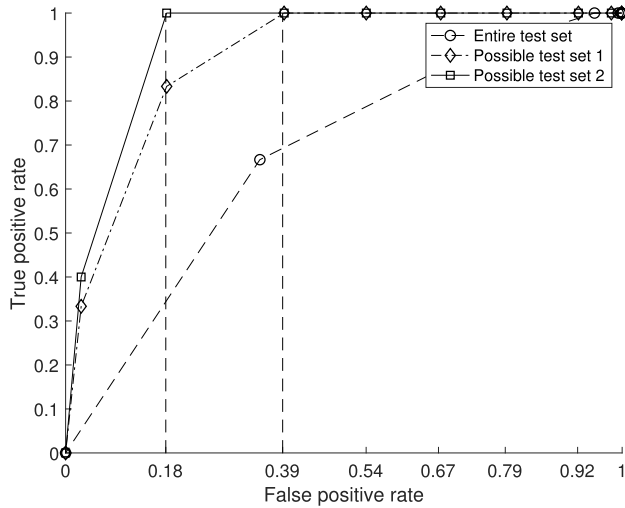
FIGURE 6. ROC curves of the entire test set (dashed line with circle data points), the *speculative* test set 1 (dotdashed line with diamond data points) and the *speculative* test set 2 (solid line with square data points). The vertical dashed lines show the FPR at 100% TPR for the *speculative* test sets 1 and 2.
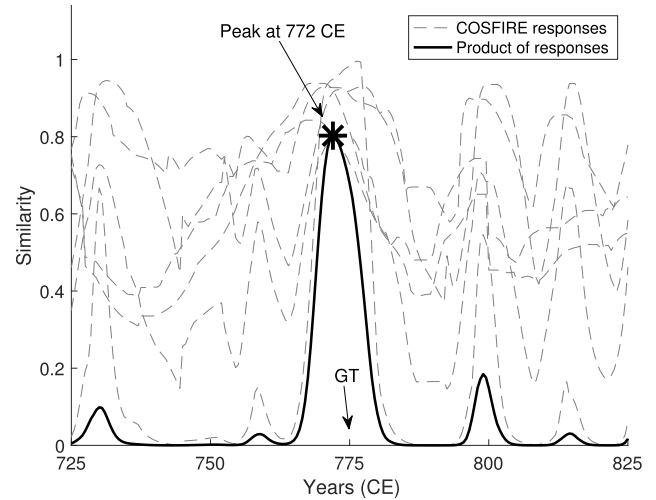

FIGURE 8. Second approach for suggesting other *speculative* Miyake events. We use the product of the responses of the 12 COSFIRE filter responses and we identify the highest values. The responses are shown with dashed gray lines and their product with solid black line. The star indicates the global peak above a certain threshold.
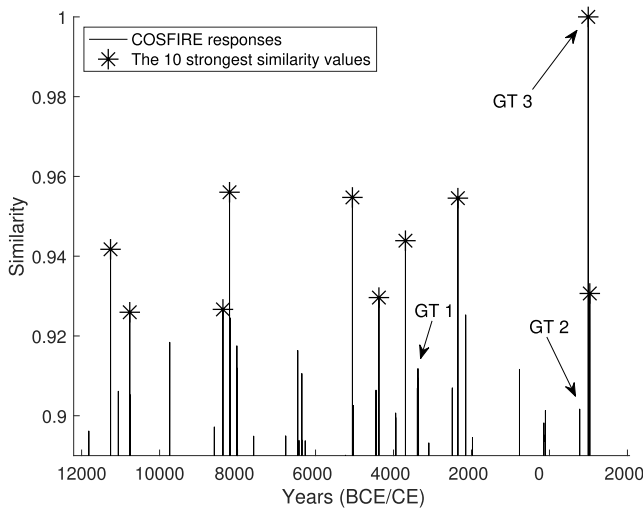

FIGURE 7. First approach for suggesting other *speculative* Miyake events. Here we simply collect the 10 highest COSFIRE responses that was trained with the second Miyake event (= GT 3), as suggested by the grid search and the cross validation. The years of the ground-truth data (GT) are shown with arrows. We mention that the response in the second Miyake event returns maximum value because is the one that has been trained with.

E. SUGGESTIONS FOR NEW MIYAKE EVENTS

We use two different approaches for compiling a list of years that COSFIRE filters suggest exhibit similar patterns in $\Delta^{14}C$ to the ones around the Miyake Events.

For the first approach, we choose the years that correspond to the ten highest responses of the COSFIRE filter that was trained with the second Miyake Event (774 - 775 CE). For reasons of clarity, in Fig. 7 we plot the COSFIRE responses of 90% and over. The ten dates of greatest similarity are shown with stars at their peak values. The positions of the GT are indicated by text arrows, with GT1 being the 3372 BCE event, GT2 the 775 CE event, and GT3 the 994 CE event. The GT3 event in this example has a value of 1 (highest)

because it is the date that was used for training the COSFIRE filter. The GT1 and GT2 events are among the 30 highest responses.

For the second approach, we use the product of all the 12 COSFIRE responses to obtain a new result that has peaks of at the points where all the 12 responses have high values. Then, we rescale this final response in the range [0,1] by dividing the whole response signal by its maximum value. The possible dates chosen here are the ones that return a higher value than 0.8. One example of this approach is shown in Fig. 8. Here, we illustrate with dashed gray lines the COSFIRE responses for the 100 years around the first Miyake Event, and with a solid bold line the product of those responses. For clarity, we only illustrate 6 response signals. The star represents the global peak. It is shown that after the multiplication of all the response signals together, the resulting responses have clearer peaks. In this example, COSFIRE returns a high peak around the first Miyake event and the responses elsewhere are reduced significantly.

Therefore, the second approach involves taking into account more than one response, in contrast with the first approach, so its results could be considered more robust or trustworthy.

In Table 2 we present 26 dates, ten of which are suggested with the first approach and the remaining 16 with the second. In the fourth column, we show whether the suggested dates are also in the GT.

IV. DISCUSSION

In this study, we demonstrate the effectiveness of the COSFIRE filters for the identification of Miyake Events in dendrochronological data. We used data from three *established* events that are used as GT to configure COSFIRE filters that respond to similar patterns. We evaluate 12 *speculative* events and we compute the FPR for

TABLE 2. List of other *speculative* Miyake events that our system suggests. We use two approaches to make these suggestions. In the last column we indicate whether any of the suggested dates are also in the GT.

Suggested dates	Approach 1	Approach 2	ground-truth
11258 BCE	v		
11030 BCE		v	
10765 BCE	v		
8542 BCE		v	
8492 BCE		v	
8375 BCE	v		
8206 BCE	v		
6123 BCE		v	
5481 BCE		v	v
5054 BCE	v		
4375 BCE	v		
3960 BCE		v	
3694 BCE	v		
3653 BCE	v		
3375 BCE		v	v
2467 BCE		v	
2349 BCE	v		
381 CE		v	
656 CE		v	
772 CE		v	v
995 CE	v		v
1023 CE		v	
1035 CE	v		
1151 CE		v	
1262 CE		v	
1390 CE		v	

every individual event, and we suggest a subset of 5 of those 12 *speculative* events that are identified with considerably low FPR.

Additionally, based on the COSFIRE filter responses to the IntCal13 dataset and single-year data, we suggest a number of new hypothetical events that our system finds most like to be Miyake Events. We present two different ways of doing this. One is taking the ten strongest responses of the COSFIRE filter that was trained and optimized the best, and the second method takes into account the responses of 6 COSFIRE filters that were configured using the three *established* Miyake Events. Based on the results, it seems the second approach is more accurate since it suggests three dates that are also in the GT.

Additional reassurance was gained from new information relating to the year 1218 CE. This date, which was included in the GT, returned a very high probability of being a Miyake Event on the basis of our system (Shown as low FPR in Fig. 5). During the course of our work, single-year tree-rings were measured over this year and a new Miyake Event was indeed discovered.

We chose to work with COSFIRE filters because we had already completed a study on the identification of the best signal processing methods for Miyake Event detection [29]. In that work, we showed that COSFIRE filters outperformed other possible approaches, namely Euclidean distance, cross correlation and dynamic time warping. The COSFIRE filters are trainable, in that they allow us to configure a detector that is selective to any pattern of interest. The generalization of COSFIRE filters can be controlled by temporal and

amplitude - related parameters, which can be determined empirically.

From a signal processing point of view, the IntCal13 dataset is hard to handle. In many cases, there are multiple values for the same years with varying spreads. Also, every data point has a probability distribution around the mean value. Typically, one or three standard deviations of this distribution are considered. In this work, we simply take the mean values between multiple data points. Moreover, the sampling frequency is between 5 and 10 years, where the training examples have a frequency of 6 months. In signal processing this is typically referred as a "low resolution" or "missing values" problem. Since the majority of the signal processing techniques require data without missing values, in such cases, several techniques can be applied to fill, or subtract data. We use linear interpolation between missing values. In future, we aim to investigate non-linear interpolation techniques too.

For physical validation of our results, we will now obtain single-year measurements of $\Delta^{14}C$ in tree-rings over a selection of the *speculative* Miyake Events that our proposed method identifies, as shown in Table 2.

V. CONCLUSION

The use of signal processing techniques in datasets is important when patterns of interest need to be identified. Here, we demonstrate that computational methods and COSFIRE filters are suitable for the identification of the Miyake Events. This proposed system can be used as a tool for discovering and predicting such events. Its trainable character also allows us to adapt the same approach for the identification of other patterns of interest.

REFERENCES

[1] M. Stuiver and H. A. Polach, "Discussion reporting of ^{14}C data," *Radiocarbon*, vol. 19, no. 3, pp. 355–363, 1977.

[2] H. de Vries, "Variation in concentration of radiocarbon with time and location on earth," in *Proc. Koninkl. Nederl. Akad. Wetenschappen*, vol. 61. 1958, pp. 1–9.

[3] F. Miyake, K. Nagaya, K. Masuda, and T. Nakamura, "A signature of cosmic-ray increase in ad 774–775 from tree rings in Japan," *Nature*, vol. 486, no. 7402, pp. 240–242, 2012.

[4] F. Miyake, K. Masuda, and T. Nakamura, "Another rapid event in the carbon-14 content of tree rings," *Nature Commun.*, vol. 4, Apr. 2013, Art. no. 1748.

[5] I. G. Usoskin *et al.*, "The AD775 cosmic event revisited: The sun is to blame," *Astron. Astrophys.*, vol. 552, p. L3, Feb. 2013.

[6] A. J. T. Jull *et al.*, "Excursions in the ^{14}C record at A.D. 774–775 in tree rings from Russia and America," *Geophys. Res. Lett.*, vol. 41, no. 8, pp. 3004–3010, 2014.

[7] D. Güttler *et al.*, "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.*, vol. 411, pp. 290–297, Feb. 2015.

[8] A. Fogtmann-Schulz, S. M. Østbø, S. G. B. Nielsen, J. Olsen, C. Karoff, and M. F. Knudsen, "Cosmic ray event in 994 C.E. recorded in radiocarbon from Danish oak," *Geophys. Res. Lett.*, vol. 44, no. 16, pp. 8621–8628, 2017.

[9] F. Wang, H. Yu, Y. Zou, Z. G. Dai, and K. S. Cheng, "A rapid cosmic-ray increase in BC 3372–3371 from ancient buried tree rings in China," *Nature Commun.*, vol. 8, no. 1, 2017, Art. no. 1487.

[10] J. Park, J. Southon, S. Fahrni, P. P. Creasman, and R. Mewaldt, "Relationship between solar activity and Δ^{14}C peaks in AD 775, AD 994, and 660 BC," *Radiocarbon*, vol. 59, no. 4, pp. 1147–1156, 2017.

[11] M. Dee, B. Pope, D. Miles, S. Manning, and F. Miyake, "Supernovae and single-year anomalies in the atmospheric radiocarbon record," *Radiocarbon*, vol. 59, no. 2, pp. 293–302, 2017.

[12] F. Mekhaldi *et al.*, "Multiradionuclide evidence for the solar origin of the cosmic-ray events of AD 774/5 and 993/4," *Nature Commun.*, vol. 6, Oct. 2015, Art. no. 8611.

[13] V. V. Hambaryan and R. Neuhäuser, "A galactic short gamma-ray burst as cause for the ^{14}C peak in AD 774/5," *Monthly Notices Roy. Astronomical Soc.*, vol. 430, no. 1, pp. 32–36, 2013.

[14] A. K. Pavlov *et al.*, "Gamma-ray bursts and the production of cosmogenic radionuclides in the Earth's atmosphere," *Astron. Lett.*, vol. 39, no. 9, pp. 571–577, 2013.

[15] P. J. Reimer *et al.*, "IntCal13 and marine13 radiocarbon age calibration curves 0-50,000 years cal BP," *Radiocarbon*, vol. 55, no. 4, pp. 1869–1887, 2013.

[16] F. Miyake *et al.*, "Large ^{14}C excursion in 5480 BC indicates an abnormal sun in the mid-Holocene," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 5, pp. 881–884, 2017.

[17] L. Wacker, "Towards a new radiocarbon calibration curve based on annually resolved data," in *Proc. 14th Int. Conf. Accel. Mass Spectrometry (AMS)*, Ottawa, ON, Canada, 2017.

[18] M. W. Dee and B. J. S. Pope, "Anchoring historical sequences using a new source of astro-chronological tie-points," *Proc. Roy. Soc. A, Math. Phys. Eng. Sci.*, vol. 472, no. 2192, 2016, Art. no. 20160263.

[19] M. Stuiver, P. J. Reimer, and T. F. Braziunas, "High-precision radiocarbon age calibration for terrestrial and marine samples," *Radiocarbon*, vol. 40, no. 3, pp. 1127–1151, 1998.

[20] F. G. McCormac, A. Bayliss, D. M. Brown, P. J. Reimer, and M. M. Thompson, "Extended radiocarbon calibration in the anglo-saxon period, AD 395-485 and AD 735-805, *Radiocarbon*, vol. 50, no. 1, pp. 11–17, 2008.

[21] A. Hogg, J. Palmer, G. Boswijk, P. Reimer, and D. Brown, "Investigating the interhemispheric ^{14}C offset in the 1st millennium AD and assessment of laboratory bias and calibration errors," *Radiocarbon*, vol. 51, no. 4, pp. 1177–1186, 2009.

[22] J. van der Plicht, E. Jansma, and H. Kars, "The 'Amsterdam Castle': A case study of wiggle matching and the proper calibration curve," *Radiocarbon*, vol. 37, no. 3, pp. 965–968, 1995.

[23] Q. Hua *et al.*, "Atmospheric ^{14}C variations derived from tree rings during the early younger dryas," *Quaternary Sci. Rev.*, vol. 28, nos. 25–26, pp. 2982–2990, 2009.

[24] J. C. Vogel, A. Fuls, E. Visser, and B. Becker, "Pretoria calibration curve for short-lived samples, 1930–3350 BC," *Radiocarbon*, vol. 35, no. 1, pp. 73–85, 1993.

[25] R. E. Taylor and J. Southon, "Reviewing the mid-first millennium BC 14C 'warp' using 14C/bristlecone pine data," *Nucl. Instrum. Methods Phys. Res. B., Beam Interact. Mater. At.*, vol. 294, pp. 440–443, Jan. 2013.

[26] G. Azzopardi and N. Azzopardi, "Trainable COSFIRE filters for keypoint detection and pattern recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 490–503, Feb. 2013.

[27] A. Neocleous, G. Azzopardi, C. N. Schizas, and N. Petkov, "Computer analysis of images and patterns," in *Proceedings 16th International Conference, CAIP* (Image Processing, Computer Vision, Pattern Recognition, and Graphics), vol. 9256, G. Azzopardi and N. Petkov, Ed. Cham, Switzerland: Springer, 2015, pp. 558–569.

[28] A. Neocleous, "Computing expert's intelligence: A case in bio-medicine and a case in musicology," Ph.D. dissertation, Dept. Comput. Sci., Univ. Groningen, Groningen, The Netherlands, 2016.

[29] A. Neocleous, G. Azzopardi, and M. Dee, "Identification of possible Δ^{14}C anomalies since 14 ka BP: A computational intelligence approach," *Sci. Total Environ.*, vol. 663, pp. 162–169, May 2019.

ANDREAS NEOCLEOUS was born in Larnaca, Cyprus. He studied audio signal processing at the Technical University of Crete, Greece, and received the graduate studies at the University of Pompeu Fabra, Spain, and the Ph.D. degree from the University of Groningen, The Netherlands, in 2016, where he is currently a Postdoctoral Researcher with the Center for Isotope Research. He has been collaborating with the University of Cyprus (UCY) as a Research Scientist, since 2011, on research programs funded by the EU, the UCY, and the Cyprus Research Promotion Foundation. He has published articles and has presented his work at international conferences and at high-impact academic journals. His research interests include digital signal processing, machine learning, and computational intelligence.

GEORGE AZZOPARDI received the B.Sc. degree (Hons.) in computer science from Goldsmiths, the M.Sc. degree in computer science from the Queen Mary University of London, and the Ph.D. degree *(cum laude)* in computer science from the University of Groningen, The Netherlands, in 2013, where he is currently an Assistant Professor of computer science. His research interests include pattern recognition, machine learning, signal processing, medical image analysis, and information systems. He is an Associate Editor of the Q1 journal.

MARGOT KUITEMS is currently a Postdoctoral Researcher with the Center for Isotope Research, University of Groningen, The Netherlands. She is an Archaeologist and a Quaternary Scientist on the ECHOES project. She has considerable experience in the analysis of palaeoenvironmental archives and the dynamics of ancient societies. She is examining isotope time-series over the known Miyake Events and comparing them with analogous data from other palaeoastronomical events. She is also taking a lead on using Miyake Events for the exact dating of early societies.

ANDREA SCIFO was born in Catania, Italy. He received the bachelor's degree in physics from the University of Catania, in 2014, and the M.Sc. degree, in 2017, after conducting his thesis project at the Centre for Isotope Research (CIO), University of Groningen, The Netherlands, where he is currently pursuing the Ph.D. degree. In 2017, he started his Ph.D. project at CIO, where his main research interests include cosmogenic isotopes and carbon cycle.

MICHAEL DEE received his first degree in chemistry in his native New Zealand, and the Ph.D. degree in the application of Bayesian statistics to radiocarbon data from the University of Oxford, in 2009. He is currently an Assistant Professor of isotope chronology with the University of Groningen. He was a recipient of the ERC Starter Grant, in 2016, to investigate the origins and applications of short-term anomalies in the atmospheric radiocarbon record.

• • •