Environmetrics

WILEY

RESEARCH ARTICLE OPEN ACCESS

A Hybrid Nonparametric Framework for Outlier Detection in Functional Time Series

David Solano[1,2] | Ruben Dario Guevara[3] | Sergio A. Calderón V.[3] | Marc Saez[1,2] | Maria A. Barceló[1,2]

[1]Research Group on Statistics, Econometrics and Health (GRECS), University of Girona, Girona, Spain | [2]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain | [3]Departamento de estadística, Universidad Nacional de Colombia, Bogotá, Colombia

Correspondence: Maria A. Barceló (antonia.barcelo@udg.edu)

Received: 17 January 2026 | **Revised:** 1 April 2026 | **Accepted:** 8 April 2026

Keywords: directional outlyingness | functional depth | functional time series | magnitude and shape anomalies | Moving Block Bootstrap | outlier detection | sliding window methods

ABSTRACT

Outlier detection in functional time series is challenging due to temporal dependence and the simultaneous presence of magnitude, shape, and partial anomalies. Existing methods often assume independence or rely on model based approaches, such as the Standard Smoothed Bootstrap on Residuals (SmBoR), which may not work well if the model is misspecified. Model free alternatives, based on the moving block bootstrap, improve robustness but may detect only a limited number of magnitude anomalies. This work proposes a fully model free pipeline with two components. First, the Directional Outlyingness (DirOut) framework is extended by recalibrating its cutoff via an outlier detection procedure based on the moving block bootstratp (MBBo), improving the detection of shape and partial outliers while controlling false positives. Second, a Sliding Window Functional Boxplot (SWOD) is used to focus on local temporal neighborhoods and detect magnitude anomalies that other methods may miss. Simulations show that SWOD has high detection rates for magnitude outliers, while MBBo calibrated DirOut achieves almost perfect detection for shape and partial anomalies, outperforming SmBoR. The method is also tested on a real temperature dataset, showing its practical usefulness.

1 | Introduction

Throughout this paper, an outlier refers to an observation that is unlikely to have been generated by the same underlying stochastic mechanism that produces the majority of the data. Such departures may arise from measurement errors, exogenous shocks, or structural changes that modify the data generating process. In the context of time-dependent functional data, the impact of these atypical observations can be considerable big. They may deform global envelopes, bias parameter estimates, and lead to incorrect interpretations of temporal evolution.

Outlier detection in functional and multivariate functional data has become an increasingly active research area over the past decade, driven by the growing availability of high frequency temporal data in environmental, industrial, and social systems. Functional Data Analysis (FDA) models data as smooth functions, enabling a more comprehensive analysis than traditional pointwise methods. This approach helps to detect unusual behaviors, whether they come from magnitude shifts, shape deviations, or more complex patterns that standard multivariate methods might miss (López-Oriona and Vilar 2021).

Early contributions to functional outlier detection define the theoretical foundations of functional depth and its application. Originally introduced for multivariate data, the concept of statistical depth was later extended to functional data to provide a center outward ordering of curves, enabling statistical analysis.

Depth based approaches play an important role in functional data analysis, as they can be used for tasks such as exploratory analysis, outlier detection, or classification (Zhao et al. 2024; Jiménez-Varón et al. 2024; Harris et al. 2021; Wang et al. 2024).

Several depth notions have been proposed to order functional data. For instance, Cuevas et al. (2006) introduced Modal Depth (MD) and Random Projection Depth (RP), while López-Pintado and Romo (2009) proposed Modified Band Depth (MBD) as an extension of band-based methods. Other important depth measures include integrated depth (Fraiman and Muniz 2001), half-region depth (López-Pintado and Romo 2011), spatial depth (Chakraborty and Chaudhuri 2014), extremal depth (Narisetty and Nair 2016), and total variation depth (Huang and Sun 2016). More recent developments include elastic depth (Harris et al. 2021), model-based statistical depth (Zhao et al. 2024), pointwise data depth (Jiménez-Varón et al. 2024), quantile integrated depth (Luo et al. 2026), and regularized halfspace depth (Yeon et al. 2025), among others. These methods provide different ways to measure how central or extreme a curve is within a sample.

Building on depth concepts, functional boxplots provide intuitive visualization and detection of magnitude and shape outliers by extending classical boxplots to the functional setting (Sun and Genton 2011). However, when curves exhibit temporal or spatial dependence, standard functional boxplots may misidentify outliers due to biased fences. To address this issue, adjusted functional boxplots have been proposed, where simulation based calibration of the fences accounts for correlation structures, improving control of false positives in spatio temporal data (Sun and Genton 2012). These developments highlight the importance of both robust depth measures and dependence aware adjustments in functional outlier analysis.

In the same vein, Raña et al. (2015) proposed a bootstrap based approach to detect outliers in functional time series that explicitly accounts for temporal dependence. By iteratively resampling depth based residuals, their method is capable of detecting both magnitude and shape anomalies while maintaining accurate control of false-positive rates (FPRs). Later, Vilar et al. (2016) extended this idea using robust Functional Principal Component Analysis (FPCA), performing outlier identification through projection scores or residual norms while preserving the temporal structure of functional coefficients.

More recent research has focused on improving robustness and adaptability in high dimensional and complex settings. Liu et al. (2022) proposed the max–min functional outlier detection algorithm (MM-FOD), a two stage procedure combining clean set selection and false discovery rate (FDR) control, enabling adaptive thresholding without strong distributional assumptions. Similarly, Elías et al. (2023) proposed the TDEPTH (or STDEPTH when using scaled depth) framework to identify *evolution outliers*, namely observations whose temporal progression deviates from the population trend, highlighting the importance of dynamic depth evaluation in high dimensional functional time series. In the industrial domain, Mun et al. (2024) proposed the Domain Knowledge Informed Sequential Transformations (DK-ST) method, integrating expert driven transformations such as derivatives and Fourier components to transform shape anomalies into magnitude outliers, achieving high sensitivity to latent defects in manufacturing systems. Most recently, Rigueira et al. (2025) developed a multivariate functional approach (MMSA) that extends directional outlyingness measures to jointly capture magnitude, shape, and amplitude anomalies in environmental sensor networks.

Despite these advances, a persistent challenge remains: most existing functional outlier detection methods assume independence among observations, an assumption that rarely holds in real world functional time series where temporal or spatial dependence is intrinsic. Aside from the bootstrap based approach of Raña et al. (2015), few existing methods explicitly model dependence, which may lead to biased thresholds and poor control of false discoveries. In addition, model based approaches, such as autoregression based SmBoR (Raña et al. 2015), may produce biased results if the underlying model is misspecified, and the extent of this bias and its impact on outlier detection can vary with the type of anomalies present, making it difficult to quantify. Many approaches either rely on univariate functional representations, thus losing potentially valuable inter variable relationships, or treat derivative information as auxiliary rather than integral to the detection process. From a calculus perspective, derivatives encode key aspects of the geometry of a function: its slope and curvature; therefore, they provide natural descriptors of shape variability. Ignoring these features may mask subtle but meaningful deviations, particularly in complex or partially contaminated functional series. Additionally, methods such as DirOut (Dai and Genton 2019), although effective for multivariate functional data, are primarily designed under frameworks that focus on the marginal properties of curves. As a result, their test statistics may be sensitive to temporal or cross curve dependence, which could affect their direct applicability to functional time series.

To address these issues, this work proposes a two component model free pipeline for detecting outliers in functional time series, designed to jointly handle shape, magnitude, and partially contaminated anomalies. The first component extends the Directional Outlyingness (DirOut) framework by recalibrating its detection threshold using the bootstrap procedure of Raña et al. (2015). This combination mitigates DirOut's independence assumption while enhancing its sensitivity to geometric deviations, making it particularly effective for detecting shape related and partial outliers.

The second component introduces a Sliding Window Functional Boxplot approach aimed at identifying magnitude outliers in temporally dependent observations. By constructing symmetric temporal windows around each observation and computing its local functional depth, this method captures localized magnitude deviations that may appear "central" in a global functional boxplot but are temporally inconsistent. For example, in a daily temperature series, a winter curve similar to an early spring pattern might lie within the global envelope, yet still represent anomalous temporal behavior. The sliding window approach effectively captures such context specific magnitude outliers while maintaining extremely low FPRs (see Figure 1). The adoption of sliding windows is motivated by prior studies on scalar time series, where window based approaches have proven effective for capturing local temporal dependencies and detecting context specific

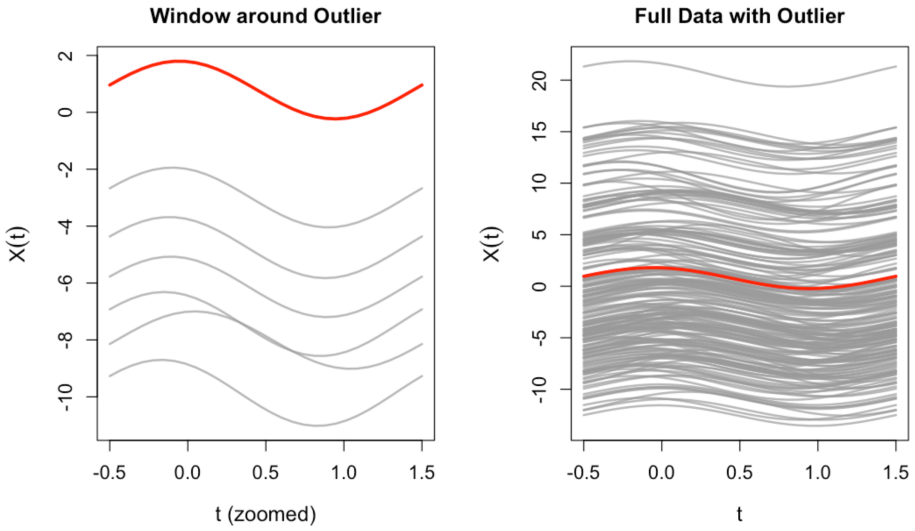

FIGURE 1 | Left panel: Application of a sliding window approach reveals the magnitude outlier by leveraging temporal context. Right panel: The same outlier remains undetected when the temporal sequence is ignored, highlighting the importance of time-dependent analysis.

anomalies (Alimohammadi and Chen 2022; Mejri et al. 2024). Extending this principle to functional data allows the method to preserve the temporal structure while maintaining low FPRs, helping to identify local magnitude anomalies accurately in functional time series.

Finally, beyond methodological contributions, this study provides a comprehensive computational framework that ensures full reproducibility and accessibility. All algorithms, including those by Raña et al. (2015) and the proposed extensions, have been completely reimplemented in the R language, offering a unified, modular platform for functional outlier detection. This open source implementation allows users to apply, compare, and expand methods with minimal effort, encouraging transparency and replicability in functional data analysis.

The remainder of the paper is organized as follows. Section 2 presents the theoretical framework, including functional depth notions, functional directional outlyingness (FDO), and the methodological tools used in this study. Section 3 reports on the simulation study, describing the simulated models, hyperparameter selection, and results. Section 4 illustrates the proposed approach through an application to temperature data.

2 | Theoretical Framework

2.1 | Definitions

In this subsection, the basic definitions that will be used throughout the remainder of the paper are collected. These notions provide the terminology and framework needed for subsequent results.

2.1.1 | Functional Depth

Functional depth is a measure designed to provide a systematic way of ordering functional data. It defines a center outward ranking, where the curves with the highest depth lie near the center of the sample, while curves with low depth appear at the extremes. Many depth functions have been proposed in the literature to establish this ordering. In this work, these ordering properties are used to rank curves and identify potential outliers using different strategies (see Section 2.3). The approaches used are presented below.

Modal Depth (MD). The MD, introduced by Cuevas et al. (2006), provides a measure of centrality for functional data. Let $\{\chi_i\}_{i=1}^n$ be a sample of n functional observations defined in L^2 space. The MD assigns to each curve a depth value that reflects its relative centrality with respect to the entire sample.

$$\mathrm{MD}_n(\chi, h) = \frac{1}{nh}\sum_{i=1}^{n} \frac{K\left(\|\chi - \chi_i\|\right)}{h},$$

where $\|\cdot\|$ is a norm, K is a kernel function, and h is a bandwidth. In this work, the L_2 norm and the truncated Gaussian kernel are used.

Modified Band Depth (MBD). MBD (López-Pintado and Romo 2009) quantifies the centrality of a functional observation by evaluating how often its trajectory remains within bands formed by subsets of j curves from a given sample. Let $\{\chi_i\}_{i=1}^m$ be functional observations belonging to L^2 space. For a curve $\chi \in L^2$, the MBD is defined as

$$\mathrm{MBD}_m^J(\chi) = \sum_{j=2}^{J} \mathrm{MBD}_m^{(j)}(\chi),$$

where

$$\mathrm{MBD}_m^{(j)}(\chi) = \binom{m}{j}^{-1} \sum_{1\le i_1<\ldots<i_j\le m} \lambda_r\left\{A\left(\chi_{i_1}, \ldots, \chi_{i_j}\right)\right\},$$

and

$$A\left(\chi_{i_1}, \ldots, \chi_{i_j}\right) = \left\{t \in I : \min_r \chi_{i_r}(t) \le \chi(t) \le \max_r \chi_{i_r}(t)\right\},$$

with λ_r denoting the normalized Lebesgue measure on $I = [a, b]$.

For multivariate functional data $\chi = (\chi_1, \ldots, \chi_s)$, where each component $\chi_k \in L^2$, the depth extends componentwise as

$$\mathrm{MBD}_m^J(\chi) = \sum_{k=1}^{s} p_k \, \mathrm{MBD}_{m,k}^J(\chi_k),$$

where the non-negative weights p_k sum up to one and may incorporate prior knowledge or the correlation structure among components (Ieva and Paganoni 2013).

2.1.2 | Multivariate Functional Depth (MFD)

Multivariate functional depth (MFD) (Claeskens et al. 2014) extends the notion of univariate depth to multivariate functional observations. Let

$$\chi(t) = \left(\chi_1(t), \ldots, \chi_K(t)\right)^\top, \qquad t \in I,$$

denote a multivariate functional observation, where each component $\chi_k \in L^2$ space and $\chi(t) \in \mathbb{R}^K$. The MFD of χ with respect to the distribution F_χ is defined as

$$\mathrm{MFD}\left(\chi, F_\chi\right) = \int_I D\left(\chi(t); F_\chi(t)\right) \, w(t) \, \, dt,$$

where $F_\chi(t)$ is the K-dimensional distribution of the random vector $\chi(t)$, D is a multivariate depth in $\mathbb{R}^K$ (here the Tukey depth), and $w(t)$ is a weight function in I.

2.2 | Functional Directional Outlyingness (FDO)

The ***FDO*** framework (Dai and Genton 2019) extends traditional functional depth measures by incorporating both the magnitude and the direction of deviation from the functional median. For each time point t, the directional outlyingness of a p-multivariate functional data $\chi(t)$ is defined as

$$\mathbf{O}\left(\chi(t), F_{\chi(t)}\right) = \left(\frac{1}{d\left(\chi(t), F_{\chi(t)}\right)} - 1\right)\mathbf{v}(t),$$

where $\boldsymbol{d}(\cdot)$ is a statistical depth and $\boldsymbol{v}(t)$ is the spatial sign from the median function $\mathbf{Z}(t)$ toward $\chi(t)$. Integrating this quantity over the domain yields two descriptive measures:

$$\mathbf{MO} = \int_I \mathbf{O}\left(\chi(t), F_{\chi(t)}\right) w(t)\, dt, \quad VO = \int_I \| \, \mathbf{O}\left(\chi(t), F_{\chi(t)}\right) - \mathbf{MO}\|^2 w(t)\, dt,$$

representing the *magnitude* and *shape* components of outlyingness, respectively. The total outlyingness decomposes as $\mathbf{FO} = \| \mathbf{MO}\|^2 + VO$, providing interpretable information on whether an anomaly arises from level shifts or structural deviations. $w(t)$ is a weight function, and the authors defined it as $w(t) = \{\lambda(\cdot)\}$, where $\lambda(\cdot)$ represents the Lebesgue measure.

In this study, FDO was used to obtain **MO** and VO for univariate and multivariate functional data, using projection depth for robustness. These quantities provide a natural decomposition of outlyingness into magnitude and shape components, which is key to the detection strategy adopted in this work.

Test-Based DirOut Detection. Let $\mathbf{Y}_{k,n} = \left(\mathbf{MO}_{T_k,n}^\intercal, VO_{T_k,n}\right)^\intercal$, where $\mathbf{MO}_{T_k,n}$ and $VO_{T_k,n}$ denote the finite-dimensional representations of the mean directional outlyingness and its variation, respectively, evaluated over the design set $T_k = \{t_1, \ldots, t_k\}$. Assuming that functional observations are generated from a p-dimensional stationary Gaussian process, the joint distribution of $\mathbf{Y}_{k,n}$ can be well approximated by a normal distribution of $(p+1)$-dimensional (Dai and Genton 2019). This approximation enables a test based procedure for functional outlier detection based on the robust Mahalanobis distance and the F-distribution.

Following Rousseeuw (1985), the robust Mahalanobis distance is computed as

$$\mathrm{RMD}^2\left(\mathbf{Y}_{k,n}, \overline{\mathbf{Y}}_{k,n,J}^*\right) = \left(\mathbf{Y}_{k,n} - \overline{\mathbf{Y}}_{k,n,J}^*\right)^\top \left(\mathbf{S}_{k,n,J}^*\right)^{-1} \left(\mathbf{Y}_{k,n} - \overline{\mathbf{Y}}_{k,n,J}^*\right), \tag{1}$$

where J denotes the subset of $h \le n$ observations that minimizes the determinant of the corresponding covariance matrix,

$$\overline{\mathbf{Y}}_{k,n,J}^* = h^{-1} \sum_{i \in J} \mathbf{Y}_{k,n,i}, \quad \text{and} \quad \mathbf{S}_{k,n,J}^* = h^{-1} \sum_{i \in J} \left(\mathbf{Y}_{k,n,i} - \overline{\mathbf{Y}}_{k,n,J}^*\right)\left(\mathbf{Y}_{k,n,i} - \overline{\mathbf{Y}}_{k,n,J}^*\right)^\top.$$

The parameter h controls the robustness of the estimator, with a maximum finite-sample breakdown point of $\lfloor (n-p+1)/2 \rfloor / n$.

According to Hardin and Rocke (2005), the tail of the RMD^2 distribution can be approximated by a F-distribution:

$$\frac{c(m-p)}{m(p+1)} \mathrm{RMD}^2\left(\mathbf{Y}_{k,n}, \overline{\mathbf{Y}}_{k,n,J}^*\right) \sim F_{p+1,m-p}, \tag{2}$$

where c and m are scale and degrees of freedom parameters estimated through Monte Carlo simulation. The cutoff value C is then defined as the α-quantile of the $F_{p+1,m-p}$ distribution, with $\alpha = 0.993$, following the convention used in classical boxplot thresholds under normality.

Finally, a curve is declared an outlier when

$$\frac{c(m-p)}{m(p+1)} \mathrm{RMD}^2\left(\mathbf{Y}_{k,n}, \overline{\mathbf{Y}}_{k,n,J}^*\right) > C. \tag{3}$$

This F-distribution based criterion defines the original DirOut test. Although effective under independence, its distributional assumption may not hold for temporally dependent functional data, motivating the bootstrap calibration introduced in Section 2.4.1. In particular, when strong serial dependence is present, the subset J of size h that minimizes the covariance determinant may not adequately preserve the temporal structure of the data. This misalignment can reduce the sensitivity of the test and lead to inconsistent threshold estimation. The proposed bootstrap based calibration allows DirOut to adapt to dependence while maintaining robustness and directional interpretability.

2.3 | Methods

2.3.1 | The Functional Boxplot

Let $\chi_1, \ldots, \chi_n$ be functional observations defined in L^2 space. To establish a center-outward ordering of the sample, a functional

depth $D(\chi_i)$ is assigned to each curve. Denote by $\chi_{[1]}, \ldots, \chi_{[n]}$ the curves arranged from the highest to the lowest depth.

In the functional boxplot of Sun and Genton (2011), the 50% central region aggregates the deepest half of the sample and is defined as

$$C_{0.5} = \left\{ (t, y(t)) : t \in I, \ \min_{r=1,\ldots,\lfloor n/2 \rfloor} \chi_{[r]}(t) \le y(t) \le \max_{r=1,\ldots,\lfloor n/2 \rfloor} \chi_{[r]}(t) \right\}. \tag{4}$$

Rather than relying on quartiles or interquartile ranges as in the classical univariate boxplot, the functional version uses the envelope of $C_{0.5}$ as its analogue of the "central box". An inflation factor F is then applied to this envelope to produce an expanded band that represents the non-outlying region. A curve is an outlier if there exists $t \in I$ such that it falls outside the inflated band, with $I \subset \mathbb{R}$ a compact interval.

The approach is depth agnostic: different notions of depth induce different notions of centrality and may yield distinct sensitivity to atypical curves, while the overall mechanism of the functional boxplot remains unchanged. In this document, the functional boxplot will serve as a straightforward, baseline method for outlier detection.

2.3.2 | The Multivariate Functional Boxplot

Let $\chi(t)$ represent a multivariate functional object composed of p component curves. For each index $j = 1, \ldots, p$, the function χ^j lies in the space $L^2(I)$, so all components are square integrable on the interval I. Following the framework introduced in Ieva and Paganoni (2013), the construction of a multivariate functional boxplot is based on marginal depth evaluations and their corresponding central envelopes. The procedure may be summarized as follows:

- For each coordinate process $\chi^{(j)}$, compute a univariate functional depth and obtain the associated center outward ordering.
- Using these marginal depth orderings, construct the univariate functional boxplot for every component $\chi^{(j)}$, obtaining its inflated non-outlying region.
- A multivariate curve χ is classified as an outlier if there are a time point $t_0 \in I$ and a component j such that $\chi^{(j)}(t_0)$ lies outside the non-outlying region defined by the jth marginal functional boxplot.

2.3.3 | Bootstrap for Dependent Data

Bootstrap methods provide a flexible approach to detect outliers in functional data by resampling observed curves and estimating a depth based cutoff C that separates typical curves from potential outliers. In the general framework (Febrero et al. 2008, 2007), low depth curves are iteratively identified and removed until no further outliers are detected.

In functional time series, dependence between observations must be taken into account when estimating the cutoff C. The Moving Block Bootstrap for Outlier detection (MBBo) proposed by Raña et al. (2015) extends this framework by generating bootstrap samples that preserve the temporal dependence structure of the data. The procedure consists of three components: (i) remove potential outliers from the original sample, (ii) generate bootstrap functional time series samples using the moving block bootstrap (MBB), and (iii) derive the cutoff $\widehat{C}$ from the empirical distribution of depth values across bootstrap samples.

Hyperparameters. The MBBo method depends on the following hyperparameters:

- Functional depth measure $D_s(\cdot)$, which determines the centrality of the curves.
- Bootstrap block length l, determining the dependence structure preserved by the MBB.
- Number of bootstrap replications B.
- Depth cutoff quantile α, used to determine the bootstrap based threshold.
- Initial outlier removal method used to obtain the cleaned sample S_{clean}.

These hyperparameters control the sensitivity of the method to dependence, sampling variability, and the definition of centrality.

Procedure

1. **Step 1**: Let S be the set of curves and O the empty set.
2. **Step 2**: Compute the depth of each curve in S using a functional depth $D_s(\cdot)$. Let A be the set of curves with depth below the cutoff C. Update $O = O \cup A$ and $S = S \backslash A$.
3. **Step 3**: Repeat Step 2 until A is empty.
4. **Step 4**: The curves in O are declared outliers.

MBB . Following Künsch (1989), let l denote the block length and n the sample size. The original series $\{X_1, \ldots, X_n\}$ is divided into all possible overlapping contiguous blocks of length l:

$$B_{i,l} = \{X_i, X_{i+1}, \ldots, X_{i+l-1}\}, \qquad i = 1, \ldots, n - l + 1.$$

To generate a bootstrap replicate, draw i.i.d. indices $I_1, \ldots, I_k \sim \text{Unif}\{1, \ldots, n - l + 1\}$, where k is the smallest integer such that $kl \ge n$. The bootstrap series is then obtained by pasting together the blocks

$$B_{I_1,l}, B_{I_2,l}, \ldots, B_{I_k,l}$$

and truncating to the first n observations. This resampling scheme reproduces the dependence structure at the block level while preserving the marginal distribution of the curves.

Estimating the cutoff C using MBBo. Let S^{clean} denote the sample after removing visually obvious outliers (e.g., using the functional BoxPlot; Sun and Genton 2011). The MBBo cutoff estimative proceeds as follows:

1. Generate B bootstrap samples $\{S^{*(b)}\}_{b=1}^{B}$ from S^{clean} using the MBB with block length $l = 4$, as recommended by Raña et al. (2015).
2. For each bootstrap sample $S^{*(b)}$, compute the depth $D_s(\cdot)$ of all curves and determine its bootstrap-based cutoff $C^{(b)}$. This cutoff is typically chosen as the empirical α-quantile of the depth distribution, with $\alpha = 0.01$ following Febrero et al. (2008).
3. Aggregate the bootstrap estimates by taking the median:

$$\widehat{C} = \text{median}\{C^{(1)}, \ldots, C^{(B)}\}.$$

This mechanism produces a cutoff that adapts to the dependence structure of the data and reduces the likelihood of depth underestimation due to serial correlation. Only the MBBo version is considered here, as Raña et al. (2015) report minimal empirical differences between the moving block and stationary bootstrap-based outlier detection procedures (MBBo and StBo).

Multivariate Bootstrap for Dependent Data (multiMBBo). The multiMBBo method extends the MBBo framework to multivariate functional data, where each observation (or individual) is represented by a p-dimensional vector of functional trajectories. In this setting, the MBB is applied to the sequence of individuals, and for every selected index within a bootstrap block, the *entire* set of its p marginal curves is resampled together. This ensures that the multivariate structure of each individual is preserved and that the temporal dependence across the series of multivariate observations is maintained. After generating B multivariate bootstrap samples, MFDs (e.g., multivariate MBD or MFD) are computed to calculate centrality across dimensions jointly, and depth based cutoffs are obtained following the same aggregation procedure used in the univariate MBBo.

2.4 | Proposed Method

Motivated by the limitations of existing functional outlier detection methods, particularly their independence assumptions and limited handling of temporal dependence, this paper proposes a two component, model free pipeline for functional time series. The framework is designed to detect shape, magnitude, and partially contaminated anomalies while preserving temporal and multivariate structure. The first component extends the Directional Outlyingness (DirOut) framework to dependent data by recalibrating detection thresholds via a moving block bootstrap (MBBo), thereby retaining sensitivity to geometric deviations while taking into account serial dependence. The second component, called Sliding Window Outlier Detection (SWOD), complements this global approach with a localized, magnitude based strategy. SWOD identifies context specific magnitude anomalies that may appear central under global depth measures. Together, the two components provide a unified framework that combines global shape detection with local temporal adaptivity.

2.4.1 | DirOut Based Detection Under Temporal Dependence

MBBo Calibration of DirOut Threshold. Once the directional outlyingness measures ($\mathbf{MO}, VO$) are computed for each observation where the p-variate case is considered here, although the methodology applies analogously to the univariate setting. The MBBo procedure is adapted to determine outliers in the multivariate functional framework. For each bootstrap sample:

1. Compute the directional outlyingness vector $\mathbf{MO} = (MO_1, \ldots, MO_p)$ and the scalar VO for each curve.
2. For each element MO_j ($j = 1, \ldots, p$) and for VO, compute the α-quantile (with $\alpha = 0.99$, consistent with the original proposal of Febrero et al. 2008) across the bootstrap sample, giving a cutoff vector $C^b = \left(C_1^b, \ldots, C_p^b, C_{VO}^b\right)$.
3. Repeat the multivariate MBB B times and take the median of the cutoffs across all samples to define the final cutoff vector $\widehat{\mathbf{C}} = \left(\widehat{C}_1, \ldots, \widehat{C}_p, \widehat{C}_{VO}\right)$.

A curve is identified as an outlier if any of its $\mathbf{M}O$ components or its VO is greater than the corresponding cutoff in $\widehat{\mathbf{C}}$.

On the Removal of the Initial Cleaning Step. The original MBBo formulation includes an initial removal of visually detected outliers using tools such as the functional boxplot. While this may be acceptable in the i.i.d. setting, deleting curves in functional time series breaks the temporal dependence structure and affects the core assumption that the MBB must preserve dependence. As a result, the preliminary cleaning step can compromise the validity of the procedure.

The extension based on directional outlyingness addresses this limitation by removing the preliminary filtering step. Instead, bootstrap based cutoff values are computed using the entire sample, which preserves the temporal structure of the data and results in a resampling procedure that is consistent with the data behavior.

2.4.2 | Sliding Window Outlier Detection (SWOD)

The *SWOD* method is proposed to identify outliers in functional data by utilizing local information through sliding windows and functional depth. This approach is designed to capture deviations that may be localized within the domain of the functions, making it particularly suitable for functional time series or ordered functional data. The methodological foundation for using sliding windows in time series outlier detection is well established in the literature, particularly in streaming and time ordered contexts, where sliding window frameworks enable localized modeling and robust detection of anomalous behavior (Yu et al. 2014).

Hyperparameters. The SWOD method depends on the following hyperparameters:

- Sliding window size w, which determines the number of neighboring curves included around each focal observation.
- Outlier persistence threshold $\tau \in [0, 1]$, specifying the minimum proportion of windows in which a curve is classified as a global outlier.
- A functional depth measure $D(\cdot; W_i)$, which defines how centrality is measured in each local window.
- Local boxplot rule, that is, the threshold used by the functional boxplot within each window (this implicitly defines how local outliers are detected).

These hyperparameters control the locality of detection, the required consistency of anomalous behavior, and the sensitivity of the window level outlier identification.

Procedure. Let $\{\chi_i\}_{i=1}^n$ be a sample of n temporally dependent functional observations defined on $L^2(I)$ space. The SWOD procedure consists of the following steps:

1. For each observation χ_i, a symmetric sliding window of size w is defined around the index i. The window contains the set of curves

   $$W_i = \{\chi_j : j = \max(1, i - \lfloor w/2 \rfloor), \ldots, \min(n, i + \lfloor w/2 \rfloor)\}.$$

   The hyperparameter w determines the level of locality in the analysis. Smaller values of w emphasize local, fine scale variations. However, larger values produce smoother and more stable depth estimates.

2. Within each window W_i, a functional depth measure $D(\chi_j; W_i)$ is calculated for each curve $j \in W_i$, providing a local center outward ranking of the curves.

3. Outliers within the window are identified using a traditional functional boxplot (Sun and Genton 2011), identifying curves with depth values below the boxplot threshold.

4. The global indices of the curves identified as outliers are recorded, and a counter tracks the number of windows in which each curve is flagged.

5. Steps 1–4 are repeated for all observations $i = 1, \ldots, n$.

An observation is considered a global outlier if it is identified in at least a fraction τ of the windows it appears in. The parameter $\tau \in [0, 1]$ sets how often a curve must show unusual behavior to be labeled as a global outlier: small values catch curves that are extreme only sometimes, while large values require the curve to be consistently unusual across many windows.

2.4.3 | Extension to Multivariate Functional Data

SWOD extends naturally to multivariate functional data, $\chi_i = \left(\chi_i^{(1)}, \ldots, \chi_i^{(p)}\right)$, where p denotes the number of functional components. In this context, the sliding windows are defined as before, but outlier detection considers all components jointly:

1. For each window W_i, the indices of curves identified as outliers are determined using MFD or multivariate MBD, although any suitable multivariate functional depth measure can be used.

2. In window iteration, the indices of selected individuals are recorded. All corresponding marginal functional data for those individuals are sampled together, preserving the multivariate structure.

3. The global outlier status is assigned based on the fraction of windows in which the individual is identified as an outlier, analogous to the univariate case.

This extension ensures that the full multivariate information of each individual is preserved while detecting outliers, enabling the identification of anomalies that may happen only when considering multiple functional components jointly.

2.4.4 | Algorithm Overview

The proposed framework combines global and local strategies to detect outliers in functional time series. The overall procedure can be summarized as follows:

1. Compute directional outlyingness measures (**MO**, VO) for all observations.
2. Apply the moving block bootstrap (MBBo) to estimate cutoff thresholds.
3. Identify global outliers using the calibrated DirOut criteria.
4. For each observation, construct sliding windows and compute the functional depth within each window.
5. Detect local outliers in each window using the functional boxplot rule or an alternative depth based outlier detection criterion.
6. Aggregate window-level results and compute the proportion of times each curve is detected as an outlier.
7. Classify observations as outliers if they exceed the threshold τ.
8. Combine the outputs of DirOut and SWOD to obtain the final set of detected outliers.

3 | Simulation Study

3.1 | Simulated Models

Functional time series were generated following the setup in Raña et al. (2015). First, an uncontaminated series was defined as

$$\zeta_i(t) = \begin{cases} \cos(\pi t), & \text{if } i = -n + 1, \\ \cos(\pi t)(1 - \rho) + \rho\zeta_{i-1}(t) + a_i(t) + b_i, & \text{if } -n + 1 < i \le n, \end{cases} \quad (5)$$

where $t \in [-0.5, 1.5]$, $a_i(t) = X_i \sin(\pi t)$ with X_i i.i.d. Gaussian with mean 0 and standard deviation 0.3, and $\{b_i\}$ is a scalar Gaussian AR(1) process with correlation ρ and standard deviation $(1 - \rho^2)^{-1/2}$. Three outliers were then introduced at random indices $\{\zeta_i\}_{i=1}^n$ to construct the contaminated time series. The contaminated models were defined as follows:

- **Magnitude Outliers:**

  $$\chi_i(t) = \zeta_i(t) + kI\{i \in \{I_1, I_2, I_3\}\}, \quad 1 \le i \le n, \quad (6)$$

  where k is a contamination size constant, and I_j ($j = 1,2,3$) are i.i.d. uniform random indices from $\{1, \ldots, n\}$.

- **Shape Outliers:**

  $$\chi_i(t) = \zeta_i(t) + k\cos(3\pi t)I\{i \in \{I_1, I_2, I_3\}\}, \quad 1 \le i \le n. \quad (7)$$

TABLE 1 | Average true-positive rate (TPR) and false-positive rate (FPR) under magnitude contamination for different contamination sizes k.

		$k = 10$		$k = 15$		$k = 20$		$k = 25$	
Method	**Depth**	**TPR (SD)**	**FPR**	**TPR (SD)**	**FPR**	**TPR (SD)**	**FPR**	**TPR (SD)**	**FPR**
MultiMBBo	MultiMBD	22.00 (0.24)	13.89	28.67 (0.28)	12.96	25.67 (0.29)	12.48	26.33 (0.26)	13.28
	DirOut	33.67 (0.26)	3.41	58.33 (0.26)	3.12	74.00 (0.19)	3.46	80.67 (0.17)	2.11
MBBo	MBD	46.00 (0.26)	4.67	69.67 (0.31)	4.70	90.67 (0.20)	5.26	97.33 (0.10)	5.10
	MD	36.33 (0.29)	1.84	67.00 (0.29)	2.06	87.33 (0.22)	1.78	96.67 (0.11)	2.52
	DirOut	34.33 (0.25)	1.34	57.67 (0.24)	2.05	70.67 (0.20)	0.82	78.67 (0.17)	0.70
BoxPlot	MBD	18.00 (0.22)	1.18	48.67 (0.34)	1.33	78.00 (0.26)	1.82	90.33 (0.19)	1.62
	MultiMBD	0.00 (0.00)	0.10	0.00 (0.00)	0.05	0.00 (0.00)	0.04	0.00 (0.00)	0.06
	MD	17.66 (0.24)	1.55	45.33 (0.38)	2.31	75.67 (0.31)	2.25	86.33 (0.26)	2.35
SWOD	MBD	**84.67 (0.21)**	0.50	**93.67 (0.16)**	0.44	**99.33 (0.05)**	0.51	**100 (0.00)**	0.55
	MultiMBD	36.67 (0.31)	1.13	46.00 (0.31)	1.10	61.00 (0.33)	1.07	63.33 (0.27)	1.10
	MD	83.67 (0.22)	4.13	94.33 (0.14)	4.18	98.00 (0.09)	4.37	100.00 (0.00)	4.43
Test Based	DirOut	28.00 (0.29)	14.04	35.00 (0.32)	14.15	46.00 (0.35)	13.03	64.00 (0.35)	14.10

Note: The values in parentheses correspond to the standard deviation of the TPR. Bold value indicates the best performing method in terms of detection accuracy.

- **Partial Outliers:**

$$\chi_i(t) = \zeta_i(t) + kI\{i \in \{I_1, I_2, I_3\} \cap \{t \geq T_i\}\}, \quad 1 \leq i \leq n, \tag{8}$$

where T_i is a random variable uniformly distributed on $[-0.5,1.5]$.

The curves were discretized on a grid of 30 equispaced points in $[-0.5,1.5]$. The sample size was $n = 200$ and the AR(1) correlation coefficient was set to $\rho = 0.8$. The contamination sizes were chosen following Raña et al. (2015) to allow a direct comparison with their results. In particular, $k = 10,15,20,25$ is considered for magnitude and partial outliers, and $k = 4,5,6,7$ for shape outliers. The use of different ranges is motivated by the nature of the contamination: magnitude and partial outliers require larger values of k to produce noticeable amplitude deviations, whereas shape outliers can be identified even for smaller values due to changes in the curve geometry. For each configuration, 100 Monte Carlo replications were performed, and the average true-positive rate (TPR) and FPR were calculated.

The compared methods include the original MBBo framework using MD and MBD, and the MBBo to estimate the DirOut thresholds, the proposed multivariate extensions (multiMBBo), the new SWOD approach, and the functional BoxPlot baseline (univariate and multivariate, see Sun and Genton (2011) and Ieva and Paganoni (2013), respectively).

In the multivariate case, all methods labeled as *multi* incorporate the first and second derivatives of the functional trajectories to enrich the functional information available for depth computation. From a calculus perspective, derivatives capture the local geometry of functional trajectories slope through the first derivative and curvature through the second. These geometric features are particularly informative for identifying shape outliers that may not be evident from the raw curves alone. All methods were evaluated using the same contamination sizes k as in the reference study. Tables 1–3 report the mean TPR, FPR, and standard deviations for all configurations.

3.2 | Hyperparameter Selection

3.2.1 | MBBo

The implementation of the MBBo method followed the configuration proposed by Raña et al. (2015), using a block length $l = 4$, $B = 200$ bootstrap samples, and a cutoff quantile of $\alpha = 0.01$, the latter consistent with the recommendation in Febrero et al. (2008). The functional boxplot (Sun and Genton 2011) was applied as an initial cleaning step to obtain S_{clean}. In the univariate setting, the MBD and MD depths were used, as these are the depth measures originally considered in the MBBo applications to real data.

For the multivariate extension (multiMBBo), the parameters l, B, and α were kept unchanged. Because MD based approaches are known to be effective for detecting shape and partial outliers, and given that multiMBBo incorporates derivative information to capture geometric features, only the multiMBD depth was used. Finally, when MBBo was integrated into the DirOut framework, the cutoff parameter α was set to 0.99, following standard practice in functional outlier detection, since larger values in the DirOut method are more likely to be classified as outliers.

3.2.2 | SWOD

To select the hyperparameters of the SWOD method, a systematic simulation study was realized using the same data generation mechanism described in Section 3.1. Specifically, for each outlier model (magnitude, shape, and partial), the simulation design was replicated by discretizing the curves in 30 equispaced points in $[-0.5,1.5]$, with sample size $n = 200$, AR(1) correlation coefficient $\rho = 0.8$, and contamination levels k as defined for each type of outlier. For every configuration, 100 Monte Carlo replications were generated and the average TPR and FPR were computed.

To determine the optimal pair of hyperparameters, window size ω and outlier threshold τ, all combinations in the grid

TABLE 2 | Average true-positive rate (TPR) and false-positive rate (FPR) under shape contamination for different contamination sizes k.

Method	Depth	$k = 4$		$k = 5$		$k = 6$		$k = 7$	
		TPR (SD)	FPR	TPR (SD)	FPR	TPR (SD)	FPR	TPR (SD)	FPR
MultiMBBo	MultiMBD	67.33 (0.41)	13.37	59.33 (0.45)	13.26	62.33 (0.43)	12.10	63.33 (0.45)	12.66
	DirOut	**100 (0.00)**	1.55	**100 (0.00)**	1.60	**100 (0.00)**	1.60	**100 (0.00)**	1.67
MBBo	MBD	1.00 (0.06)	5.21	0.33 (0.33)	5.24	0.33 (0.33)	5.21	1.00 (0.00)	5.00
	MD	32.00 (0.32)	1.58	71.67 (0.38)	0.48	93.00 (0.21)	0.39	99.00 (0.07)	0.30
	DirOut	98.00 (0.14)	2.12	**100 (0.00)**	1.78	**100 (0.00)**	2.02	**100 (0.00)**	1.83
BoxPlot	MBD	5.33 (0.14)	1.15	5.67 (0.14)	1.28	14.33 (0.28)	1.65	15.33 (0.25)	1.38
	MultiMBD	**100 (0.00)**	**0.10**	**100 (0.00)**	**0.05**	**100 (0.00)**	**0.05**	**100 (0.00)**	**0.05**
	MD	7.33 (0.18)	1.43	9.33 (0.21)	2.15	13.00 (0.22)	2.19	18.00 (0.26)	2.30
SWOD	MBD	53.33 (0.28)	0.46	59.33 (0.31)	0.46	72.33 (0.30)	0.53	81.33 (0.22)	0.54
	MultiMBD	**100 (0.00)**	1.06	**100 (0.00)**	1.01	**100 (0.00)**	1.00	**100 (0.00)**	1.01
	MD	65.00 (0.29)	4.13	76.00 (0.26)	4.17	79.33 (0.25)	4.38	89.00 (0.18)	4.44
Test based	DirOut	100 (0.00)	13.18	100 (0.00)	11.61	100 (0.00)	12.56	100 (0.00)	13.18

Note: The values in parentheses correspond to the standard deviation of the TPR. Bold values highlight methods achieving optimal performance in shape contamination scenario (TPR = 100%), facilitating visual comparison among approaches.

TABLE 3 | Average true-positive rate (TPR) and false-positive rate (FPR) under partial contamination for different contamination sizes k.

Method	Depth	$k = 10$		$k = 15$		$k = 20$		$k = 25$	
		TPR (SD)	FPR	TPR (SD)	FPR	TPR (SD)	FPR	TPR (SD)	FPR
MultiMBBo	MultiMBD	59.00 (0.36)	11.70	65.33 (0.37)	10.32	71.67 (0.30)	9.60	76.67 (0.31)	11.51
	DirOut	**100 (0.00)**	1.49	99.67 (0.03)	1.45	**100 (0.00)**	1.37	99.67 (0.03)	1.24
MBBo	MBD	1.67 (7.30)	5.13	2.67 (9.10)	4.20	2.33 (8.55)	3.56	3.33 (11.11)	3.23
	MD	81.67 (0.23)	0.77	94.00 (0.15)	0.85	99.67 (0.10)	1.61	99.67 (0.03)	1.83
	DirOut	84.33 (0.17)	1.34	85.33 (0.17)	2.05	85.33 (0.17)	0.82	86.67 (0.17)	0.69
BoxPlot	MBD	17.00 (0.24)	1.24	48.00 (0.31)	1.99	70.00 (0.29)	1.52	82.00 (0.27)	1.61
	MultiMBD	**100 (0.00)**	**0.06**	**100 (0.00)**	**0.03**	**100 (0.00)**	**0.02**	**100 (0.00)**	**0.04**
	MD	12.33 (0.21)	1.98	47.33 (0.35)	2.76	73.67 (0.32)	3.08	83.33 (0.27)	2.39
SWOD	MBD	64.00 (0.28)	0.39	75.67 (0.25)	0.49	78.67 (0.26)	0.69	77.33 (0.24)	0.54
	MultiMBD	**100 (0.00)**	1.22	**100 (0.00)**	1.23	**100 (0.00)**	1.26	**100 (0.00)**	1.23
	MD	87.67 (0.22)	4.51	97.00 (0.10)	4.25	99.33 (0.05)	4.45	99.00 (0.06)	4.27
Test based	DirOut	100 (0.00)	13.78	100 (0.00)	12.75	100 (0.00)	12.91	100 (0.00)	13.70

Note: Values in parentheses indicate the standard deviation of the TPR. Bold values highlight methods achieving optimal performance in partial contamination scenario (TPR = 100%), facilitating visual comparison among approaches.

$$\omega \in \{4,6,8,10,12,16\}, \qquad \tau \in \{0.3,0.4,0.5,0.6,0.7\}$$

were evaluated. For each model (magnitude, shape, partial), the following procedure was applied:

1. Functional samples were generated according to the simulation scheme illustrated previously.
2. For every pair (ω, τ) in the grid, the sliding window detection algorithm was run across the 100 Monte Carlo replications.
3. The average TPR and FPR were computed for each hyperparameter combination.
4. The optimal combination was selected as the one that simultaneously maximizes TPR and minimizes FPR (see Tables A3–A5).

The results reveal a consistent pattern across all outlier models. The window size shows a clear influence on performance, with optimal values typically ranging from $\omega = 10$ to 16, corresponding to approximately 5%–8% of the sample size. In contrast, the threshold parameter τ has a minor effect on detection accuracy; therefore, a common value of $\tau = 0.5$ is adopted for all models. These recommendations hold for both univariate and multivariate functional data.

3.3 | Results

The following subsection presents the results of the simulation study implemented to evaluate the performance of the proposed detection methods. The analysis summarizes the detection accuracy across all outlier scenarios and hyperparameter configurations considered.

For magnitude outliers, the SWOD method shows strong performance across all contamination levels, with high detection rates ranging from about 84% under low contamination to 100% under high contamination, and a consistently low and stable FPR (around 0.5% when using the MBD depth function). The variability of the TPR decreases as contamination increases, reaching standard deviations below 5% for medium to high levels. The proposed adaptation of DirOut using an MBBo based bootstrap threshold improves the original F-test formulation by reducing the FPR and slightly increasing the TPR, indicating that bootstrap calibration provides a more reliable adjustment under temporal dependence.

Under shape contamination, DirOut + MBBo approaches achieve nearly perfect detection ($TPR \approx 100\%$) across all scenarios while maintaining a low and stable FPR ($\approx 1.6\%$ for the multivariate case and $\approx$2.0% in the univariate case), confirming the high sensitivity of DirOut based methods to geometric deviations in functional trajectories. The bootstrap calibration once again ensures distribution free reliability, effectively compensating for the independence assumption of the original test. Both SWOD and BoxPlot in their multivariate versions perform strongly, achieving perfect detection rates. While SWOD shows slightly higher FPRs, BoxPlot attains near zero false positives.

For partial contamination, which simultaneously affects amplitude and shape, the MBBo + DirOut approaches maintain near perfect detection (TPR $\approx$ 100%) with FPR mostly below 1.5%, demonstrating its robustness to complex deviations. The SWOD method also performs competitively, particularly in its multivariate configuration, and the BoxPlot with derivative information continues to show very good detection. Together, DirOut and SWOD based methods provide a complementary and robust pipeline, effectively covering both shape related and magnitude driven anomalies in functional time series.

Within the SWOD framework, both MBD and MD depth functions yield similar TPRs across all scenarios; however, the MD depth consistently produces higher FPRs. This suggests that the MBD based configuration achieves a better trade-off between sensitivity and specificity.

Given that DirOut + MBBo and its multivariate extension (DirOut + multiMBBo) exhibit similar performance across all simulated scenarios, and considering the superior balance

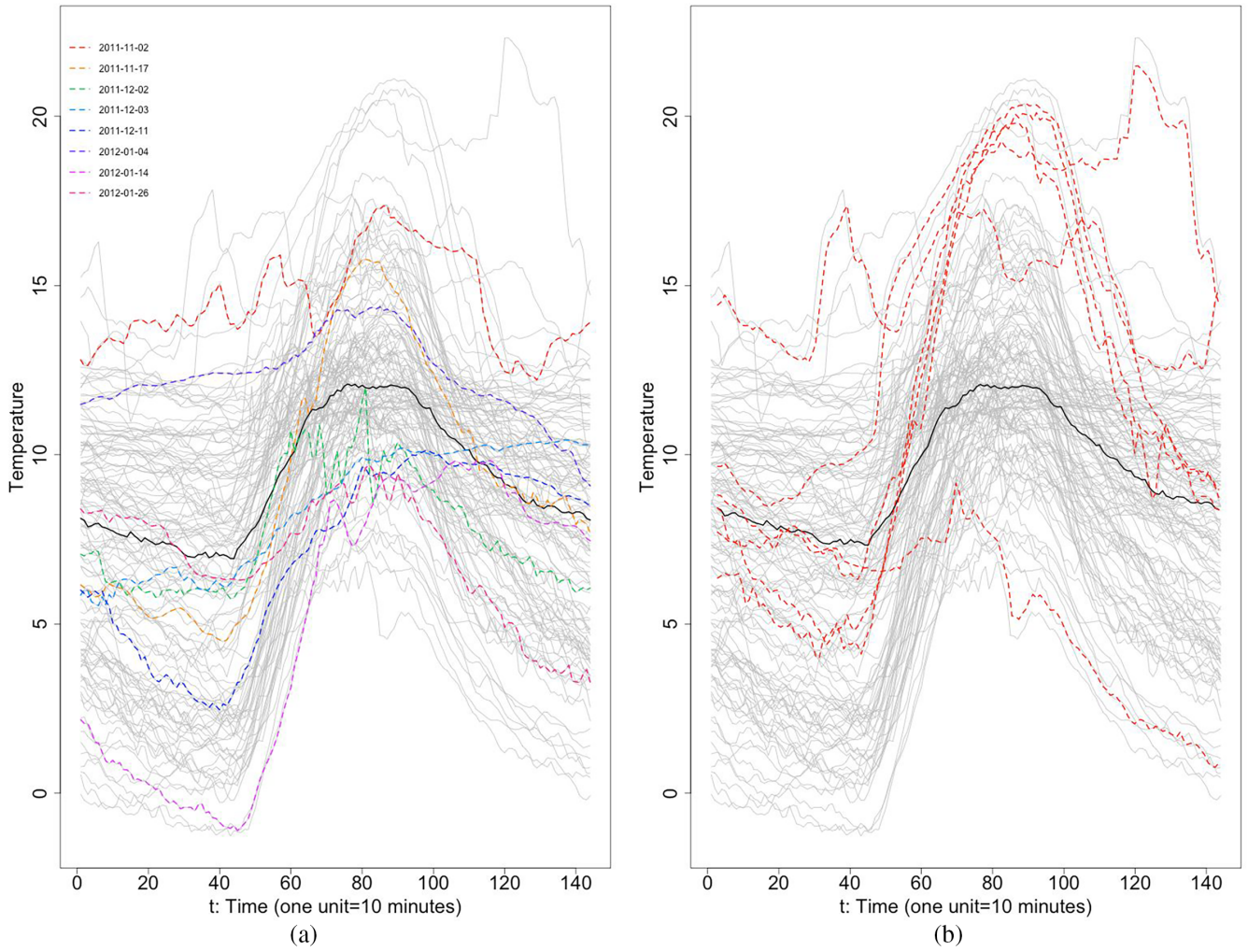

FIGURE 2 | Daily temperature curves showing detected outliers. Right panel (a) shows jointly detected outliers, and left panel (b) shows additional outliers detected only by the proposed methods. The median curve is shown in black for reference.

achieved by SWOD + MBD, only the univariate DirOut + MBBo and SWOD + MBD configurations are retained for the real data analysis.

To further assess the effect of the contamination size, an additional experiment was conducted using smaller values of k. In particular, we considered $k = 1,5,7$ for magnitude and partial outliers, and $k = 1,2,3$ for shape outliers.

The results reveal a clear dependence on both the contamination type and the data representation. For magnitude outliers, all methods exhibit very low TPRs when $k = 1$, reflecting the difficulty of detecting deviations that are nearly indistinguishable from the underlying process. As k increases to moderate values (e.g., $k = 5$), detection performance improves substantially across methods.

In contrast, for shape and partial outliers, methods that incorporate multivariate information through the inclusion of derivatives perform very well even for small values of k. This is because derivatives capture geometric features such as slope and curvature, making shape deviations easier to detect even under low contamination levels. These findings reinforce that the difficulty of the detection problem depends not only on the contamination size but also on the type of anomaly and the representation of the functional data.

4 | Temperature Data

To allow comparison, the analysis used the same dataset of daily temperature curves used by Raña et al. (2015). The data consist of 10 min temperature observations recorded at the Santiago de Compostela station (A Coruña, Spain) from November 1, 2011 to February 29, 2012. The dataset can be accessed from MeteoGalicia (http://www.meteogalicia.es).

A graphical overview of the data, highlighting detected outliers, is presented in Figure 2. Figure 2a shows the curves corresponding to outliers jointly identified by Raña et al. (2015) and our proposed methods, while Figure 2b highlights additional outliers detected exclusively by the proposed approach.

From the results summarized in Table 4 (with temperature summaries for outlier days provided in Table A1, and monthly temperature summary statistics in Table A2), and illustrated in Figure 2, it can be observed that the proposed approach is capable of correctly detecting all outliers identified by the original proposal, with the single exception of November 13. In addition, the proposed method identifies several other outliers not detected by the original approach. As shown graphically in Figure 2b, these additional curves exhibit clearly atypical behavior compared to the majority of daily temperature patterns (and with the median curve), confirming the effectiveness of the proposed detection procedure in capturing previously unrecognized anomalies.

The parameter selection for the SWOD component was guided by the sensitivity study (Tables A3–A5), which indicates that an optimal window corresponds to 5%–8% of the dataset size. For this analysis, a window length of 9 was chosen, representing approximately 7.4% of the 121 daily temperature curves. The detection threshold (th) was set equal to 0.5 in accordance with the sensitivity study.

TABLE 4 | Outliers detected by the original proposal (SmBoR with MD or MBD combined with time series outlier detection methods) compared with other approaches from this study.

Method	Detected outliers (dates)
Original proposal	November: 12, 13
	January: 15
	February: 4, 23, 24, 27
SWOD + MBD	November: 2, 12, 13, 17
	December: 2, 3, 11
	January: 4, 14, 15, 26
MBBo + DirOut	November: 12
	February: 23, 24, 27
BoxPlot + multiMBD	November: 12, 13, 14
	December: 13

In addition to the methods discussed previously, results from a baseline BoxPlot approach applied with multiMBD are also presented. While BoxPlot does not inherently account for temporal dependencies, enriching the analysis with derivative information allows it to effectively capture shape deviations in the curves. Indeed, this approach identified two additional shape outliers (see Figure 3).

5 | Conclusions

The results of this study demonstrate that the proposed SWOD approach is a robust and effective alternative to detect outliers in functional time series, particularly for magnitude anomalies. In comparison with existing methods, SWOD achieves higher TPRs and lower FPRs across a range of contamination scenarios, confirming its stability and accuracy. In real world applications, SWOD was also able to identify previously unseen outliers that were not detected by the original proposal, highlighting its practical utility.

For the DirOut approach, integrating MBBo for the estimation of the cutoff reduces FPRs while slightly increasing TPRs, resulting in a more favorable balance between sensitivity and specificity. Notably, computing the directional outlyingness on the full dataset without removing visually identified outliers beforehand does not appear to deteriorate the performance, indicating that the method remains robust without a preliminary cleaning step.

Although the SmBoR method proposed in Raña et al. (2015) was not replicated in this study, Table A6 presents the results reported by the original authors, allowing the readers to compare them with the approaches proposed here (SWOD and MBBo calibration of DirOut). Consistent with the comparisons involving MBBo and MBD or MD, the proposed methods show superior performance: SWOD achieves higher TPRs and lower FPRs than SmBoR, particularly for low contamination scenarios involving magnitude and shape outliers, while MBBo DirOut outperforms SmBoR for shape and partial outliers.

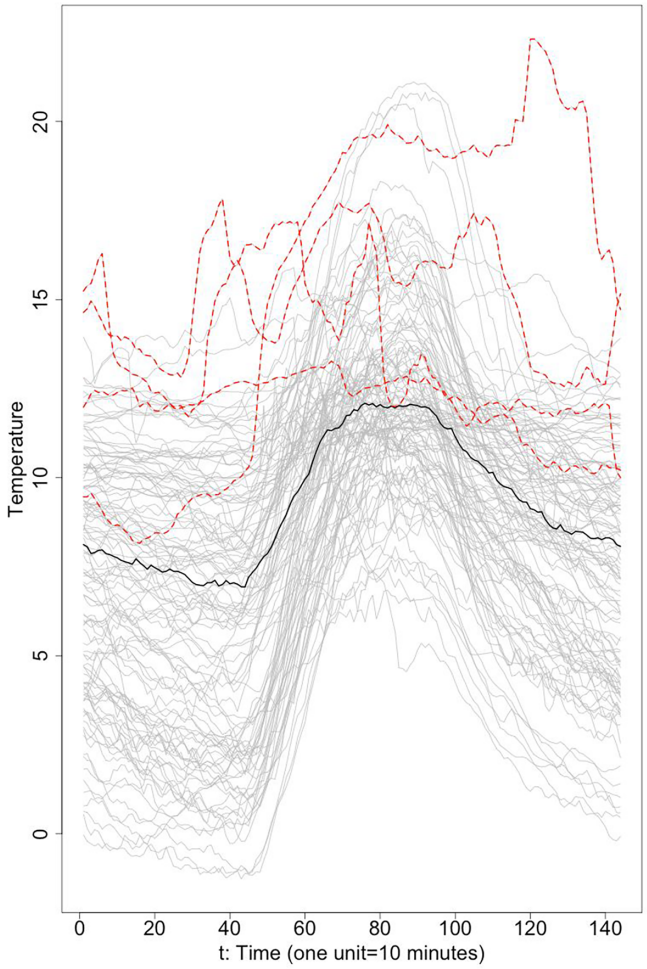

FIGURE 3 | Daily temperature curves. Red dotted curves denote the outliers identified using BoxPlot and multiMBD depth for the original process plus its two first derivatives. The median curve is shown in black for reference.

While SWOD shows strong performance, this does not imply that it universally outperforms all other methods. Rather, the combination of SWOD for magnitude outliers with shape outlier detection mechanisms, including approaches that do not explicitly account for temporal dependence (such as BoxPlot enriched with derivative information), provides a highly robust framework. Incorporating derivative based features allows these simpler methods to identify atypical curve behaviors effectively, complementing time-dependent detection techniques.

From an applied perspective, the ability to identify atypical temperature curves is relevant for several downstream tasks, including automated quality control of meteorological data, detection of unusual thermal events, environmental monitoring, and the refinement of heat-health warning systems, where unusual intra-day temperature patterns may have important operational implications beyond what is captured by daily summary indicators.

Overall, these findings suggest that a hybrid strategy combining SWOD, MBBo calibrated DirOut, or derivative enriched shape detection methods constitutes a comprehensive and reliable mechanism for outlier identification in functional time series. Finally, to facilitate reproducibility and adoption, an open R implementation of all methods and simulation code used in this study is provided here.

Beyond methodological performance, the results highlight the practical value of functional outlier detection for identifying unusual thermal patterns that may be relevant for climate studies, data quality control, and heat–health warning systems.

6 | Discussion

When applied to the derivatives of the original process in real data, the multivariate version of SWOD detected an unusually high proportion of outliers. This behavior may be attributed to the complexity and high variability of the real world data. For this reason, the results from this approach were not included in the final analyses; however, the multivariate SWOD could still be a promising option for less complex phenomena or for processes that are naturally multivariate.

For the univariate SWOD, outlier detection is currently based on a functional boxplot. While this approach performs very well for magnitude outliers, there is room for improvement when detecting shape outliers. As future work, it would be interesting to explore integrating a method more specifically designed for shape outliers such as DirOut based approaches directly into the SWOD framework rather than as an external addition. This could complement the logic of SWOD from its construction and potentially improve overall detection performance, providing a coherent framework analogous to combining SWOD with DirOut and bootstrap methods for time series.

From an applied climatological perspective, the detected outliers correspond to days with clearly atypical temperature dynamics compared to the seasonal pattern. Several of the detected days coincide with unusually high temperatures for the time of year. When extreme temperatures are defined as those exceeding the 95th percentile of the distribution of daily maximum temperatures (Ministerio de Sanidad, Gobierno de España 2023; Barceló and Saez 2025), six extreme days are identified in the dataset (November 12, 2011; November 13, 2011; November 24, 2011; February 23, 2012; February 24, 2012; and February 27, 2012). Notably, several of these days are also detected as outliers by the proposed methods, indicating that the procedure can capture thermally anomalous days that may correspond to unseasonably warm events.

Importantly, analyzing the full daily temperature curve provides more information than relying solely on daily summary statistics such as maximum temperature. Two days may present similar maximum temperatures but very different intra-day temperature trajectories; for example, a day with sustained high night-time temperatures may have very different environmental and health implications than a day with a short afternoon peak. Functional outlier detection allows these differences to be identified.

From an applied point of view, the identification of these atypical curves is relevant for several downstream tasks, including automated quality control of meteorological data, detection of

unusual thermal events, environmental monitoring, and the refinement of heat–health warning systems, where unusual intra-day temperature patterns may have important operational implications.

Finally, simulation studies revealed that SWOD's FPR is sensitive to the choice of functional depth, as seen in the differences between MBD and MD. Although this paper focuses on these two commonly used depth functions, future research should further investigate how alternative depth measures may impact SWOD's robustness and detection performance.

Acknowledgments

This study was carried out within the "Atlas of Social and Environmental Determinants of Health" subprograms of CIBER of Epidemiology and Public Health (CIBERESP). The authors declare that GenAI was used only for summarization, text editing, correction of typographical errors, and improving clarity of writing. It was not used for statistical analysis, simulations, or reference searching. We are grateful to the editor and to two anonymous reviewers for their constructive comments on an earlier version of this work, which have undoubtedly helped us to improve it. The usual disclaimer applies. Open Access funding provided thanks to the CRUE-CSIC agreement with Wiley.

Funding

This work was partially financed by AGAUR, the Department of Climate Action, Food and Rural Agenda, and by the Department of Research and Universities, both part of the Government of Catalonia (Generalitat de Catalunya) (Grant Number 2023 CLIMA 00037). The funding bodies did not participate in the design or conduct of the study; the collection, management, analysis, or interpretation of the data; or the preparation, review, and approval of the manuscript.

Ethics Statement

The authors have nothing to report.

Consent

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest. The manuscript is an original contribution that has not been published before, whole or in part, in any format. All authors will disclose any actual or potential conflicts of interest including any financial, personal, or other relationships with other people or organizations that could inappropriately influence or be perceived to influence their work.

Data Availability Statement

Temperature data are available from MeteoGalicia, the official meteorological service of Galicia (Xunta de Galicia). MeteoGalicia provides weather forecasts, monitors climate and air quality, and develops numerical models to better understand the atmosphere and ocean. Historical observations can be accessed through its data portal: MeteoGalicia Data Portal (last accessed: December 10, 2025). The code will also be available in the research group webpage.

References

Alimohammadi, H., and S. N. Chen. 2022. "Performance Evaluation of Outlier Detection Techniques in Production Timeseries: A Systematic Review and Meta-Analysis." *Expert Systems with Applications* 191: 116371.

Barceló, M., and M. Saez. 2025. "Assessing Excess Mortality and Heat-Attributable Risk During the Summer of 2022 in Catalonia, Spain: A Bayesian Spatiotemporal Analysis." *Journal of Geographical Systems*. https://doi.org/10.1007/s10109-025-00475-2.

Chakraborty, A., and P. Chaudhuri. 2014. "The Spatial Distribution in Infinite Dimensional Spaces and Related Quantiles and Depths." *Annals of Statistics* 42, no. 3: 1203–1231.

Claeskens, G., M. Hubert, L. Slaets, and K. Vakili. 2014. "Multivariate Functional Halfspace Depth." *Journal of the American Statistical Association* 109, no. 505: 411–423.

Cuevas, A., M. Febrero, and R. Fraiman. 2006. "On the Use of the Bootstrap for Estimating Functions With Functional Data." *Computational Statistics & Data Analysis* 51, no. 2: 1063–1074.

Dai, W., and M. G. Genton. 2019. "Directional Outlyingness for Multivariate Functional Data." *Computational Statistics & Data Analysis* 131: 50–65.

Elías, A., J. Morales, and S. Pineda. 2023. "A High Dimensional Functional Time Series Approach to Evolution Outlier Detection for Grouped Smart Meters." *Quality Engineering* 35, no. 3: 371–387.

Febrero, M., P. Galeano, and W. González-Manteiga. 2007. "A Functional Analysis of Nox Levels: Location and Scale Estimation and Outlier Detection." *Computational Statistics* 22, no. 3: 411–427.

Febrero, M., P. Galeano, and W. González-Manteiga. 2008. "Outlier Detection in Functional Data by Depth Measures, With Application to Identify Abnormal Nox Levels." *Environmetrics: The Official Journal of the International Environmetrics Society* 19, no. 4: 331–345.

Fraiman, R., and G. Muniz. 2001. "Trimmed Means for Functional Data." *Test* 10, no. 2: 419–440.

Hardin, J., and D. M. Rocke. 2005. "The Distribution of Robust Distances." *Journal of Computational and Graphical Statistics* 14, no. 4: 928–946.

Harris, T., J. D. Tucker, B. Li, and L. Shand. 2021. "Elastic Depths for Detecting Shape Anomalies in Functional Data." *Technometrics* 63, no. 4: 466–476.

Huang, H., and Y. Sun. 2016. "Total Variation Depth for Functional Data." *arXiv Preprint arXiv:1611.04913*.

Ieva, F., and A. M. Paganoni. 2013. "Depth Measures for Multivariate Functional Data." *Communications in Statistics - Theory and Methods* 42, no. 7: 1265–1276.

Jiménez-Varón, C. F., F. Harrou, and Y. Sun. 2024. "Pointwise Data Depth for Univariate and Multivariate Functional Outlier Detection." *Environmetrics* 35, no. 5: e2851.

Künsch, H. R. 1989. "The Jackknife and the Bootstrap for General Stationary Observations." *Annals of Statistics* 17: 1217–1241.

Liu, C., X. Gao, and X. Wang. 2022. "Data Adaptive Functional Outlier Detection: Analysis of the Paris Bike Sharing System Data." *Information Sciences* 602: 13–42.

López-Oriona, Á., and J. A. Vilar. 2021. "Outlier Detection for Multivariate Time Series: A Functional Data Approach." *Knowledge-Based Systems* 233: 107527.

López-Pintado, S., and J. Romo. 2009. "On the Concept of Depth for Functional Data." *Journal of the American Statistical Association* 104, no. 486: 718–734.

López-Pintado, S., and J. Romo. 2011. "A Half-Region Depth for Functional Data." *Computational Statistics & Data Analysis* 55, no. 4: 1679–1695.

Luo, M., S. Nagy, T. Ogden, and S. López-Pintado. 2026. "The Quantile Integrated Depth With Applications to Noisy Functional Data." *Journal of Computational and Graphical Statistics*: 1–13. https://doi.org/10.1080/10618600.2025.2569651.

Mejri, N., L. Lopez-Fuentes, K. Roy, P. Chernakov, E. Ghorbel, and D. Aouada. 2024. "Unsupervised Anomaly Detection in Time-Series: An Extensive Evaluation and Analysis of State-of-the-Art Methods." *Expert Systems with Applications* 256: 124922.

Ministerio de Sanidad, Gobierno de España. 2023. *Plan Nacional de Actuaciones Preventivas de los Efectos del Exceso de Temperaturas Sobre la Salud 2023*. Technical report, Ministerio de Sanidad. https://www.sanidad.gob.es/ciudadanos/saludAmbLaboral/planAltasTemp/2023/docs/Plan_Excesos_Temperatura_2023.pdf.

Mun, J. H., J. Yoo, H. Kim, N. Ryu, and S. Kim. 2024. "Domain-Knowledge-Informed Functional Outlier Detection for Line Quality Control Systems." *Computers & Industrial Engineering* 189: 109936.

Narisetty, N. N., and V. N. Nair. 2016. "Extremal Depth for Functional Data and Applications." *Journal of the American Statistical Association* 111, no. 516: 1705–1714.

Raña, P., G. Aneiros, and J. Vilar. 2015. "Detection of Outliers in Functional Time Series." *Environmetrics* 26, no. 3: 178–191.

Rigueira, X., D. Olivieri, M. Araujo, A. Saavedra, and M. Pazo. 2025. "Multivariate Functional Data Analysis and Machine Learning Methods for Anomaly Detection in Water Quality Sensor Data." *Environmental Modelling & Software* 190: 106443.

Rousseeuw, P. J. 1985. "Multivariate Estimation With High Breakdown Point." *Mathematical Statistics and Applications* 8: 283–297.

Sun, Y., and M. G. Genton. 2011. "Functional Boxplots." *Journal of Computational and Graphical Statistics* 20, no. 2: 316–334.

Sun, Y., and M. G. Genton. 2012. "Adjusted Functional Boxplots for Spatio-Temporal Data Visualization and Outlier Detection." *Environmetrics* 23, no. 1: 54–64.

Vilar, J. M., P. Rana, and G. Aneiros. 2016. "Using Robust FPCA to Identify Outliers in Functional Time Series, With Applications to the Electricity Market." *SORT: Statistics and Operations Research Transactions* 40, no. 2: 321–348.

Wang, S., Y. Huang, and G. Cao. 2024. "Review on Functional Data Classification." *Wiley Interdisciplinary Reviews: Computational Statistics* 16, no. 1: e1638.

Yeon, H., X. Dai, and S. Lopez-Pintado. 2025. "Regularized Halfspace Depth for Functional Data." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 87, no. 5: 1553–1575.

Yu, Y., Y. Zhu, S. Li, and D. Wan. 2014. "Time Series Outlier Detection Based on Sliding Window Prediction." *Mathematical Problems in Engineering* 2014, no. 1: 879736.

Zhao, W., Z. Xu, Y. Mu, Y. Yang, and W. Wu. 2024. "Model-Based Statistical Depth With Applications to Functional Data." *Journal of Nonparametric Statistics* 36, no. 2: 313–356.

Appendix

TABLE A1 | Temperature summary statistics and percentiles for outlier days.

Date	Mean (SD)	Median (IQR)	Percentile (%)
November 2, 2011	14.51 (1.39)	14.09 (13.48–15.82)	98.35
November 12, 2011	15.67 (4.80)	17.99 (9.57–19.44)	100.00
November 13, 2011	15.10 (1.67)	15.27 (13.61–16.70)	99.17
November 14, 2011	13.25 (2.20)	12.77 (11.71–15.25)	96.69
November 17, 2011	9.15 (3.55)	8.68 (5.84–11.70)	44.63
December 2, 2011	7.71 (1.71)	7.08 (6.07–9.24)	33.88
December 3, 2011	8.46 (1.79)	9.00 (6.42–10.15)	41.32
December 11, 2011	7.07 (2.63)	7.89 (4.57–9.43)	29.75
December 13, 2011	12.30 (0.50)	12.32 (11.98–12.66)	90.91
January 4, 2012	12.32 (1.19)	12.28 (11.71–12.95)	91.74
January 14, 2012	5.04 (4.15)	7.50 (0.55–8.91)	8.26
January 15, 2012	5.06 (2.55)	6.22 (2.92–7.02)	9.92
January26, 2012	6.99 (1.74)	7.31 (6.28–8.28)	28.10
February 4, 2012	2.68 (3.09)	1.86 (−0.24–5.69)	0.83
February 23, 2012	11.10 (5.94)	9.75 (5.46–16.89)	76.86
February 24, 2012	11.33 (5.62)	10.48 (5.85–16.55)	80.99
February 27, 2012	12.38 (5.16)	10.59 (7.59–17.12)	92.56

TABLE A2 | Monthly temperature summary statistics.

Month	Mean (SD)	Median (IQR)
November	11.44 (1.77)	11.22 (10.38–12.48)
December	9.00 (2.23)	9.28 (6.80–11.14)
January	8.05 (2.45)	7.79 (5.93–10.29)
February	7.55 (2.71)	7.70 (5.05–10.02)

TABLE A3 | Optimal hyperparameters for the magnitude model.

Type	*K*	Window size	Threshold	FPR	TPR
Multivariate	10	8	0.3	4.17	64.70
Univariate	10	4	0.4	1.85	95.33
Multivariate	15	12	0.3	2.15	78.67
Univariate	15	6	0.3	1.29	98.00
Multivariate	20	12	0.3	2.32	85.33
Univariate	20	8	0.3	0.55	100.00
Multivariate	25	12	0.3	2.12	88.00
Univariate	25	12	0.4	0.28	100.00
Univariate	25	16	0.7	0.28	100.00

TABLE A4 | Optimal hyperparameters for the shape model.

Type	*K*	Window size	Threshold	FPR	TPR
Multivariate	10	8	0.7	0.23	100.00
Univariate	10	6	0.3	1.14	100.00
Multivariate	15	16	0.6	0.20	100.00
Univariate	15	10	0.3	0.26	100.00
Univariate	15	12	0.3	0.26	100.00
Multivariate	20	12	0.5	0.40	100.00
Multivariate	20	16	0.5	0.40	100.00
Univariate	20	16	0.6	0.11	100.00
Multivariate	25	16	0.5	0.21	100.00
Univariate	25	16	0.4	0.17	100.00
Univariate	25	12	0.7	0.17	100.00

TABLE A5 | Optimal hyperparameters for the Partial model.

Type	*K*	Window size	Threshold	FPR	TPR
Multivariate	10	16	0.5	0.26	100.00
Univariate	10	6	0.4	1.55	72.67
Multivariate	15	16	0.6	0.10	100.00
Univariate	15	10	0.3	0.51	86.00
Multivariate	20	10	0.7	0.27	100.00
Univariate	20	16	0.5	0.18	84.00
Multivariate	25	16	0.5	0.33	100.00
Univariate	25	12	0.4	0.33	90.67

TABLE A6 | SmBoR results Raña et al. (2015) with MD depth.

		Low		Medium to low		Medium to high		High	
Model	Depth	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
SmBoR	Magnitude	33.33	1.11	60.00	1.11	83.67	1.55	95.00	2.24
	Shape	51.67	2.36	82.67	1.60	93.33	0.97	97.33	0.51
	Partial	81.00	0.88	94.00	1.31	96.33	1.89	99.00	2.43

Note: Magnitude: 10–20; shape: 4–7; partial: 10–20 (low to high).