Multi-domain Anomalous Relationships in Heterogeneous Temporal Data

Tolulope Ale
University of Maryland, Baltimore County

Vandana Janeja (vjaneja@umbc.edu)
University of Maryland, Baltimore County

Research Article

Keywords: time series, anomaly detection, rule mining, multi-domain relation, delayed correlation

Posted Date: December 6th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3705904/v1

Additional Declarations: No competing interests reported.

Multi-domain Anomalous Relationships in Heterogeneous Temporal data

Tolulope Ale[1*] and Vandana Janeja[1*]

[1]Information System, University of Maryland, Baltimore County, 1000 Hilltop Cir, Baltimore, 21250, Maryland, United States.

*Corresponding author(s). E-mail(s): tale2@umbc.edu; vjaneja@umbc.edu;

Abstract

The Arctic region is crucial to global climate stability. However, recent years have witnessed periods of extreme snow and ice melt, with rising temperatures that double the global average. These are not isolated events. They are the result of intricate interconnections across distinct domains. The challenge, therefore, lies not in understanding these individual domains, such as temperature, and radiation, but in decoding the inter-domain relationships inducing these polar anomalies. To address this, our study presents a novel framework aimed at mining these inter-domain relationships to explain such anomalies and the relationship across time series features comprehensively. These features may be selected from the same or different domains. Such anomalous relationships across features could help detect interesting phenomena such as extreme snow melt, and cloud cover and help identify time periods of interest when such relationships are more prevalent. We extracted the anomalous intervals in each domain using the Poisson Distribution model of rSatScan, then leveraged the concept of Direct Overlap and Proximity of anomalies to identify the direct and time-delayed temporal association (delayed correlation) between anomalies across features. The concept helps us understand how events in one domain may be associated with events in another domain during specific time periods using association rule mining. We evaluated our approach using ERA5 reanalysis data, and validated the identified anomalies against ground truth and evaluated the strength of the generated association rules using metrics like confidence and lift. Notably, several of our identified rules were consistent with findings confirmed by domain experts.

Keywords: time series, anomaly detection, rule mining, multi-domain relation, delayed correlation

1 Introduction

Anomaly detection techniques in a temporal dataset generally identify patterns in single domain. However, the complexity of real-world phenomena cannot be fully explained from the analysis of a single domain; an observed anomaly in a domain can have implications in other domains as well [1]. Therefore, relationships between related domains must also be discovered to discover the complete pattern. A multi-domain analysis has its challenges due to the heterogeneity of data. Hence the need for a technique that can identify and quantify relations between multiple disparate domains.

Our work is motivated by the melting ice caps that contribute to the sea level rise and how this is a result of several sub-components and their relationships to each other.

A comprehensive satellite study confirms that the melting ice caps are raising sea levels at an accelerating rate. Several methods have been used to analyze ice sheet changes. The study of [2], analyzing satellite data on the polar ice and changes in the Earth's crust, revealed that the melting of the polar ice is not only raising sea level, it is also causing an anomalous shifting of the Earth's crust. A model based on Zernike moments and Mann-Kendal test techniques was used to detect abrupt changes over the Antarctic ice sheet [3], and [4] used a combination of clustering algorithm and least squares fitting model to extract elevation changes of the Antarctic ice sheet. In July 2012, low-level clouds over Greenland increased downward infrared radiation and allowed enough solar radiation to penetrate, resulting in surface temperatures above melting point during an extended surface melt event [5]. The meltwater drains into the bottom of the ice, causing the ice to slip more quickly into the sea, contributing to sea level rise. Thus, one must look at several interrelated phenomena to study the phenomena of sea level rise.

To understand the ice sheet melting that contributes to the sea level rise we also need to understand the underlying phenomena which contribute to it and find linkages between the ice sheet melt and other phenomena such as radiation, snow albedo (whiteness of the snow), and other factors. There may be some time periods when these are unusual and correspond to an unusual snow melt.

In this paper, we propose an approach to uncover temporal relationships among multiple associated domains using anomalous time series data within those domains. We present an extended analysis with comprehensive results and ground truth validations. Our objective is to identify and analyze temporal anomalies in multi-domain datasets. We begin by identifying anomalous windows in each domain and then employ our novel algorithm to discover temporal relationships across distinct domains. Overlap and proximity concepts were utilized to detect temporal relationships between domains. Time-delayed relationship is identified by applying delayed correlation and shifting one domain by a time-delay coefficient δ. We investigated the associated relationship of Greenland (Arctic) atmospheric variables. We use anomalous clusters discovered in individual domains to find relationships between snow melt and other variables. If an anomalous cluster of one domain overlaps fully or partially with anomalous clusters in other domains, then we identify direct relations for them. Otherwise, we check for proximity between anomalous cluster sets and identify the relationships between domains by shifting one domain by time-delay width δ. Through this approach, we aim to identify the underlying influence of ice sheet melting.

The rest of the paper is organized as follows: Section 2 reviews the related work regarding temporal data mining. Then Section 3 provides a comprehensive analysis of the concept of overlap and proximity, and the introduction of our algorithm. In Section 4, we present the findings of our study based on the analysis of two real-world datasets. We describe the evaluation techniques employed in the study and provide an in-depth analysis of the results obtained. Finally, Section 5 summarizes the key findings of our study and highlights the potential of our model to be applied in the study of climate change.

2 Related Work

Anomaly or outlier detection is discovering deviations or unusual behavior from normalcy [6]. An anomaly in temporal data is a novel or unseen behavior or sequence of behaviors that are different from the rest of the time-series data [7]. While many researchers have addressed anomaly detection for time series data for a single domain [8], the methods used in the past have included window-based, proximity-based, prediction-based, Hidden Markov Model-based, and segmentation-based [7, 8]. Recent advancements in temporal anomaly detection include deep learning-based models such as Autoencoders and Graph Neural Networks (GNNs) [9], which have proven effective in uncovering complex patterns across multiple heterogeneous variables. A detailed analysis of these recent developments can be found in [10].

The objective of temporal data mining is to identify the hidden relations between sequences and subsequences of events [1]. These relations are discovered using association rule mining to identify patterns frequently occurring together. Compared to conventional association rule mining, temporal association rule adds time information which might be a time point or time range [11]. Time information is paramount while mining

associations for time-varying data sets [12]. This allows us to handle chunks of timeframes separately, and different rules can be found for different timeframes [1].

Episodal association mining has been implemented to discover the periodic occurrence of exciting events [13], and Genetic Network Programming (GNP) has been used for time series association rule mining on traffic data [14]. These methods employed chi-square and support measures to determine the significance of association rules. However, they did not take into account the adaptation of confidence and lift in temporal mining. Temporal association rule mining focuses on discovering rules within a given timeframe. However, what we are interested in is identifying temporal relationships in which an occurrence of one unusual event is associated with other unusual events that happen simultaneously or after a specific time interval, i.e., a delayed effect. To address this issue, Delayed Correlation Analysis (DCA) was previously employed to analyze software evolution [15], assuming that a change in one variable during a specific time would affect other variables after some time delay. However, none of the previous methods capture the anomalous associations in a multivariate heterogeneous dataset.

The contribution of this work are:

1. Developed a framework that mined the temporal relationship between distinct domains using co-occurrence of anomalies, offering a novel approach to understanding complex environmental phenomena by linking disparate data sources.
2. Validated our model on real-world data and applied it on climate data, demonstrating its effectiveness in analyzing critical environmental issues such as snow melting.
3. Employed advanced analytical techniques, including delayed correlation and shifting domain analysis, to identify time-delayed relationships and interdependencies among various environmental factors.
4. Extended the application of temporal data mining techniques to environmental studies, paving the way for more nuanced and informed responses to climate change challenges.

3 Methodology

Our methodology for multi-domain temporal association is as follows: Firstly, we discretized the data into time bins to focus on different time periods e.g. seasons and to explore periods of extreme events identified by domain scientists. Then we detect anomalies in each bin and individual domain using the Poisson Distribution model in the rSatScan package[16, 17]. The model utilizes the probability mass function (PMF) to calculate the probability of observing a specific event. A predefined threshold is used to identify the period where the observed event significantly deviates from the expected. A threshold is determined based on the Poisson probability through several experiments on historical data. We then use the method of overlap, proximity, and delayed correlation to identify temporal relationships in the anomalous windows of different domains. We first check for the number of overlaps between anomalous windows. We discover associations if more than a certain percentage of anomalous window pairs overlap. Otherwise, we check for delayed correlation across those domains. This helps us understand how events in one domain may be associated with events in another domain during specific time periods. We then quantify the strength of these temporal relationships using association rule mining. The methodology workflow is shown in Figure 1.

3.1 Data Preprocessing

In our analysis, we utilized the ERA5 reanalysis dataset comprising a region in Greenland (Arctic). The region encompassed the Arctic geolocations at 80.8N - 79.8N and 60.8W - 59.8W. This provided us with longitude and latitude data for 25 different points spanning the time period from 2001 to 2022. To focus on the temporal aspect of the analysis, we aggregated these geolocations into a single entity, resulting in 7305 consecutive timesteps or samples. For the purpose of our analysis, we discretized the data into 22 bins, with each bin representing a time span of three summer months (June, July, and August). This particular choice was motivated by the fact that the majority of snow melting occurs during this period. We experimented with three window sizes, seven, 10, and 15, to assess their impact on the results. We

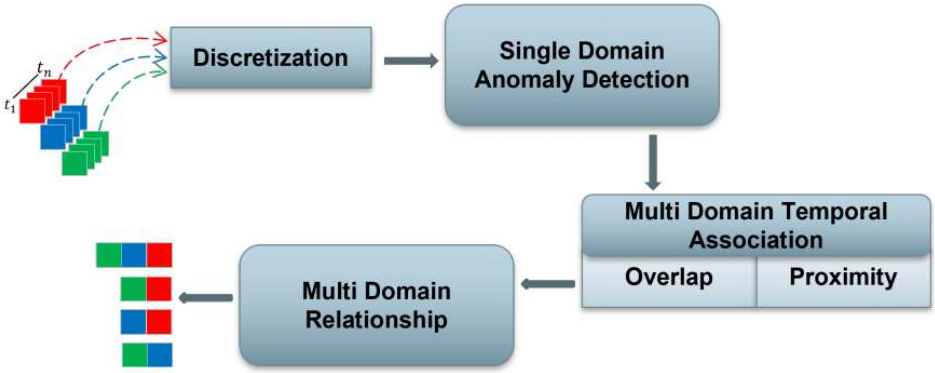

Fig. 1 Multi-domain Temporal Association Process

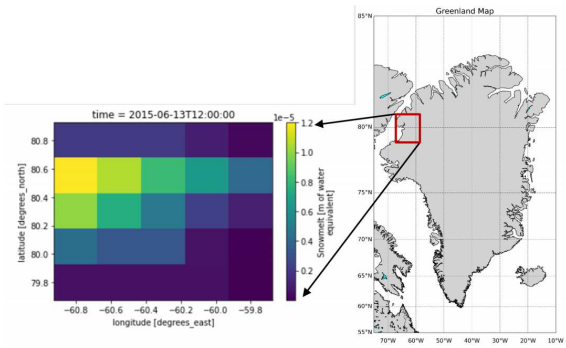

Fig. 2 Location in Northwest Greenland used in this study

determined that the window size of 15 yielded the most optimal outcome, exhibiting statistical significance at a 95% confidence level. To detect anomalous windows, we employed the Poisson Distribution model within each bin. Further details regarding the methodology for identifying these anomalies can be found in Section 3.

3.2 Multi-domain Anomaly Detection

Anomalies are sequences or subsequences of events that are not normal with respect to others. This type of anomaly detection offers explainability to complex real-world problems.

Here, in Figure 3, we can see that for the first time-series domain $T_{D1} = t_{D1^1}, t_{D1^2}, \ldots. t_{D1^x}$, where $D1$ and $D2$ represents the first and second domain, and $t_{D1^1} = t_{D2^1} = t_1$. A set of anomalous windows is represented as $A_{D1} = A_{D1^1}, A_{D1^2}, \ldots. A_{D1^i}$, where $A_{D1^i} = t_{D1^{n-p}}, t_{D1^{n-p+1}}, \ldots. t_{D1^{n-q}}$ is the ith anomalous window of the first domain and A_{D1^i} contains a subsequence of time events between t_{D1^1} and t_{D1^x}.

Having identified all the anomalous windows for all the distinct domains using Poisson distribution's probability mass function and an anomaly threshold based on the Poisson probability, the next step is to discover and quantify relations between these domains using the anomalous windows.

Definition 1 (Overlap) *Let T_{D1} and T_{D2} be set of time windows from domain $D1$ and $D2$ respectively. For time windows $T_{D1} = t_{D1^1}, \ldots t_{D1^x}$ and $T_{D2} = t_{D2^1}, \ldots t_{D2^y}$ overlap O_{D1D2} between T_{D1} and T_{D2} exists if both time windows have at least one identical time event, i.e., $t_{D1^x} = t_{D2^y}$ (Fig. 3A).*

Overlaps between anomalous time windows from two distinct domains mean some unusual activities are happening in those domains simultaneously, as shown in Figure 3A. We assume that overlap indicates the direct relationship between these distinct domains. If a certain percentage, measured by a threshold, of the total time events in each anomalous time window are occurring at the same time, then they are said to have an overlap. For example, if more than 50% of the total bins have anomalous time window pairs which overlaps, then a set of domains are said to have significant overlaps.

Definition 2 (Proximity) *For n number of bins, let us take a pair of anomalous time windows with t_{D1} and t_{D2}, where t_{D1} and t_{D2} are anomalous time windows in the nth bin from domain D1 and D2, respectively. Let d_{D1D2} be the distance between t_{D1} and t_{D2}. Proximity P_{D1D2} is defined*

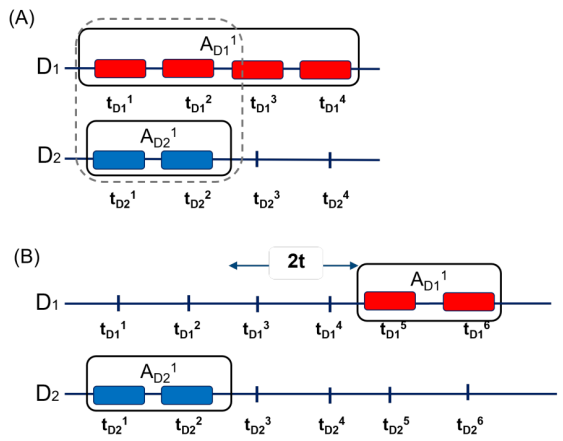

Fig. 3 **A** Overlapped anomalous windows in two distinct time series $D1$ & $D2$. **B** Proximity of two distinct time series windows $D1$ & $D2$

as the threshold used to determine the nearness between two-time windows, t_{D1} and t_{D2}. It is calculated as $P_{D1D2} = T/(n*2)$, where T is the total number of time events in either domain and $T = T_{D1} = T_{D2}$ and n is the number of bins. Time window t_{D2} is said to be in proximity with respect to t_{D1} if $P_{D1D2} > d_{D1D2}$.

If there is no overlap or the overlap is insignificant, then we check if the anomalous window pair is within proximity. If it is within proximity, we check for delayed relations for the domains.

As shown in Figure 3B, anomalous window A_{D1^1} are said to be within proximity with respect to A_{D2^1} if proximity, $P \geq 2t$.

Discovering Multi-domain Relation: For m number of domains $D1, D2, ...Dm$, let us take two domains, $D1$ and $D2$. Each domain is discretized into n number of bins. A set of anomalous time windows in domain $D1$ and $D2$ is represented by A_{D1} and A_{D2} where $A_{D1} = A_{D1^1}, A_{D1^2}, \ldots A_{D1^i}$ and $A_{D2} = A_{D2^1}, A_{D2^2}, \ldots A_{D2^j}$. Each anomalous time window consists of anomalous time events represented as t_{Dm}. For a set of anomalous time windows in nth bin S_{D1D2^n}, we check for overlap O_{D1D2^n} between anomalous time windows of $D1$ and $D2$. If an overlap is observed in more than 50% of bins, then we identify direct multi-domain relations $R_{D1=>D2}$ by using the Apriori algorithm [18]. Otherwise, delayed relations are identified by checking for proximity P_{D1D2^n} and correlation between the anomalous time windows. Then we use the Apriori algorithm to identify and quantify relations $R_{D1=>D2}$ after shifting the delayed domain with the highest correlation time lag.

The technique of discovering multi-domain relations is illustrated in Figure 4, where A_{Dm^x} represents anomalous windows, and dashed lines represent overlaps. For the temperature domain, we have anomalous windows, $A_{D1^1} = t_{D1^1}$, t_{D1^2}, t_{D1^3}, $t_{D1^4}$and $A_{D1^2} = t_{D1^7}, t_{D1^8}, t_{D1^9}$, where t_{D1^x} is an unusual time event recorded at time x in $D1$. For snow melt domain, we have anomalous windows, $A_{D2^1} = t_{D2^1}$, t_{D2^2}, t_{D2^3}, $A_{D2^2} = t_{D2^6}$, t_{D2^7}, t_{D2^8} and $A_{D2^3} = t_{D2^{10}}$. The anomalous windows for these domains are overlapped at t_1, t_2, t_3, t_7, and t_8. Then, we generate a transaction where anomalous temporal events are treated as transactions, and domains with an anomaly in those temporal events are treated as items in a typical transaction. The Apriori algorithm computed the association, support, confidence, and lift.

Algorithm 1 outlines the Multi-domain Temporal Anomaly Association approach. We use a list of anomalous temporal windows for each domain as input for our algorithm. In line 1 to line 10, we calculate overlaps in anomalous windows pair and compute the number of pairs which indicate direct relation. In line 11, if the counter value exceeds the given percentage overlap threshold (θ), indicating that the overlap is significant, we created a transaction in lines 13 - 16, indicating the relationship between domains $D1$ and $D2$ for overlapped and non-overlapped time event. In line 19, we use the association rule mining and calculate confidence, support, and lift. As an additional measure of validation, we also calculate the correlation for the anomalous time window pairs in line 20.

3.3 Time-delayed Relation

Unusual activity in a domain can cause unusual time-delayed activity in another domain. So, in this section, for each bin, we check for delayed correlation between anomalous time windows if they are within proximity. If a correlation is found, we follow the same steps as Algorithm 1 for discovering relations. The illustration in Figure 5 shows that no anomaly in domain $D1$ has overlapping time events with anomalies in domain $D2$, which implies that there is no direct relationship between these two domains. However, there still can be a

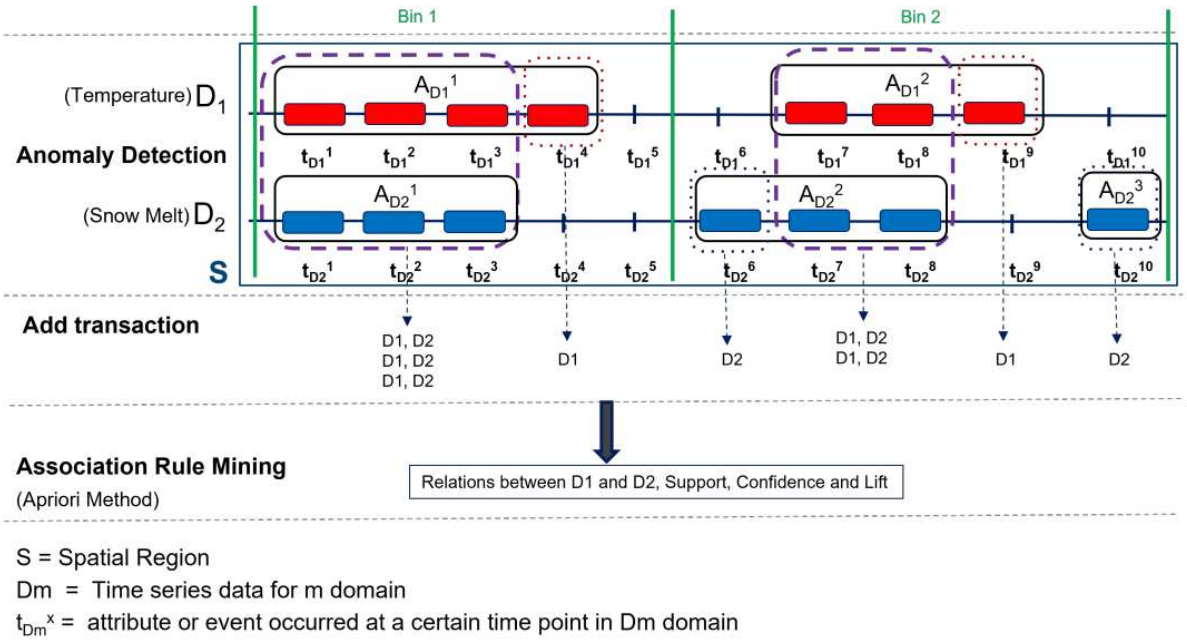

Fig. 4 Framework for detecting overlapping multi-domain anomalous relationships. $D1$ & $D2$ are two distinct domains, the dashed lines represent overlaps between anomalous window A_{D1^1} & A_{D2^1} and A_{D1^2} & A_{D2^2}

time-delayed relationship. So, in the second step of Figure 5, we check for proximity. The anomalous window pairs (A_{D1^1}, A_{D2^1}) and (A_{D1^2}, A_{D2^2}) are within proximity of $P = 2t$. Then we check if anomalous windows in a bin are correlated by using cross-correlation with a lag of δ; identify the time lag with maximum correlation δ_{max} and shift a domain with the δ_{max} value. We then create a transaction and use the Apriori approach on that transaction as we did in Algorithm 1.

Algorithm 2 outlines the Multi-domain Temporal Time-delayed Anomaly Association approach. In line 1, we compute the proximity and check if time windows in anomalous window pairs are within the defined proximity. In line 2, for each pair, we compute the average time difference between anomalous time windows in a pair for each pair. If the average time difference is lower than proximity, then we move forward to check for correlation. In line 6, we compute the cross-correlation between domains. If a correlation is found, then that indicates some delayed relationship between these domains. So, we identify the time lag (δ) with the highest correlation δ_{max}; then in line 7, we shift one domain by δ_{max}, and for each pair of anomalous windows, we compute the correlation. In line 18, we apply association rule mining to discover and quantify the relations.

4 Results and Discussion

This section describes the results obtained from the implementation of the method described in the previous sections for multi-domain anomaly detection. The metrics used to evaluate our algorithm's performance are support, confidence, and lift. To measure the accuracy of the anomaly detection technique by the Poisson Distribution model in rSatscan, we used precision, recall, and f1 score on labeled anomalies. We present the results we got after applying our novel algorithm to the anomalies to discover relationships between domains in each dataset. Then using the Apriori algorithm, we computed the relation between those domains. We use correlation to see if the overlapped pairs of anomalous windows are related. As an added validation we compared the overall correlation to see if the correlation is stronger within the bins depicting the anomalous overlaps.

We used two multi-domain real-world datasets, NJDOT (New Jersey Department of Transportation) [19], and ERA5 reanalysis atmospheric data

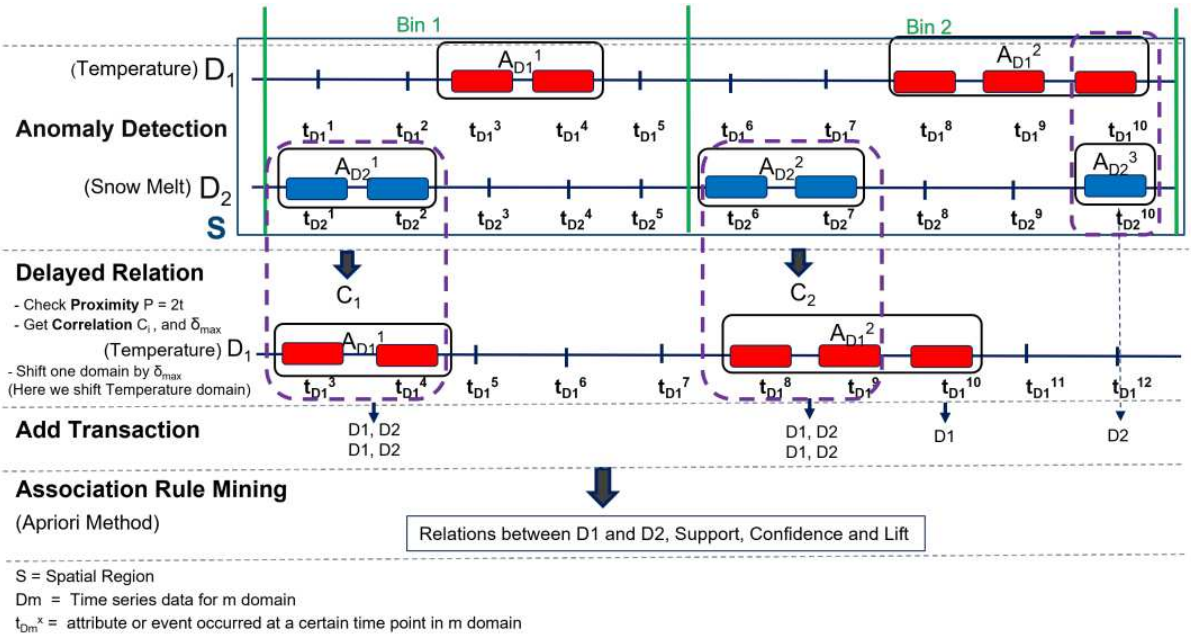

Fig. 5 Framework for identifying time-delayed relationships in domains $D1$ & $D2$. Domain $D1$ was shifted by the time lag with maximum correlation δ_{max} to observe an overlap

Table 1 Dataset description

	Domain	Temporal Resolution
NJDOT	Total Injured Light Condition Surface Condition	Jun-2014
ERA5	2-meter Temperature $t2m$ Surface Thermal Radiation Downwards $strd$ Surface Solar Radiation Downwards $ssrd$ Total Cloud Cover tcc Snowmelt $smlt$	Jun, July, August (2001 - 2022)

[20]. We also used a variation of the NJDOT data with synthetic anomalies to validate our method.

4.1 NJDOT Data

New Jersey Department of Transportation maintains the data on crashes, drivers, vehicles, occupants, and pedestrians from 2001 to date. For our analysis, we used the Total Injured, Light Condition, and Surface Condition domains of Bergen County in June 2014 for the multi-domain anomalies association approach. Each domain has 2502 rows. We discretized the data in the domains into three bins and identified the anomalies using the described techniques. The identified anomalies used in our analysis have statistical significance at the 95% level of confidence ($p < 0.05$). We converted the identified anomalous windows into a transaction file where the anomalous temporal events are treated as transactions, and domains with an anomaly in those temporal events are treated as items in a typical transaction and used the Apriori method. Relationships found in the domains are shown in Table 2. We also observe significant overlaps between the domains (Fig. 6),

Algorithm 1 Multi-domain Temporal Anomaly Association Algorithm.

Input: *Set of anomalous windows* $A_{D1} = A_{D1^1}$, A_{D1^2}, ... A_{D1^n}, $A_{D2} = A_{D2^1}$, A_{D2^2}, ... A_{D2^n} *for two domains; Percentage overlap threshold* θ

Output: *relations* $R_{D1=>D2}$ *between domain* $D1$ *and* $D2$*, support* S*, confidence* C*, and lift* L*.*

Pseudo code:

for each pair of anomalous windows in a bin, A_{D1^n} and A_{D2^n}, where $A_{D1^n} = t_x, t_{x+1}, \ldots t_k$ and $A_{D2^n} = t_y, t_{y+1}, \ldots t_l$ **do**
    Counter = 0
    Check for overlaps of time events in the anomalous windows pair:
    for $t_x \in A_{D1^n}$ **do**
        for $t_y \in A_{D2^n}$ **do**
            if **then** $> 50\%$ of $t_x = t_y$, then overlap is found
                Counter++
            end if
        end for
    end for
    if Counter $> \theta$ **then**
        for each time event in overlapped time subsequence **do**
            Add a transaction, $tr_i = D1, D2$
        end for
        for each non-overlapped time event in domain p and q **do**
            Add a transaction, $tr_i = D1$ for domain $D1$, $tr_i = D2$ for domain $D2$
        end for
    end if
    Apply Apriori on transaction tr_i
    Calculate correlation.
end for

specifically in bin 2, we observed a strong relationship across all three domain corresponding to the confidence and lift score in Table 2.

To evaluate the performance of our method, we used a version of the NJDOT data with synthetic anomalies. These anomalies were labeled in the data as unusual with respect to the rest of the data. Figure 7 shows the precision, recall, and f1 score indicating that our approach can identify most anomalies in the data and each domain.

Algorithm 2 Multi-domain Temporal Time-delayed Anomaly Association Algorithm.

Input: *set of anomalous windows* $A_{D1} = A_{D1^1}$, A_{D1^2}, ... A_{D1^n}, $A_{D2} = A_{D2^1}$, A_{D2^2}, ... A_{D2^n}, *time delay coefficient* δ*.*

Output: *relation* $R_{D1=>D2}$ *between domains* $D1$ *and* $D2$*, correlation, support* S*, confidence* C*, lift* L*.*

Pseudo code:

Compute proximity threshold: Proximity $P_{D1D2} = T/(n * 2)$, n = no. of bins in the domain
for each set of anomalous windows pair **do**
    Calculate the average time gap t_{avg}, between those anomaly pairs
end for
if (time gap $t_{avg} <$ proximity) **then**
    Use cross-correlation to find time lag (δ) with the highest correlation δ_{max}
    Shift one domain by the width of δ
    for each pair of anomalous windows **do**
        Compute the correlation between the pair
        for each overlapped time event **do**
            Add a transaction, $tr_i = D1, D2$
        end for
        for each non-overlapped time event of domain p and q **do**
            Add a transaction, $tr_i = D1$ for domain 1, $tr_i = D2$ for domain 2
        end for
    end for
end if
Apply the Apriori method to tr_i.

4.2 ERA5 Reanalysis Data

ERA5 Reanalysis dataset is a multi-contextual dataset with many atmospheric, land-surface, and ocean-wave variables. The dataset combines model data and observations across the world using the laws of physics. Therefore we use it as a multi-domain dataset for our purposes. We analyzed the temporal relationships for domains *smlt*, *t2m*, *strd*, *ssrd*, and *tcc* described in Table 1 for a location in Northwest Greenland, (Fig. 2). In this case, we analyzed the anomalous windows in the domains for the summer months of June, July, and August from 2001 to 2022. We used a sliding window sizes of 15 for the analysis.

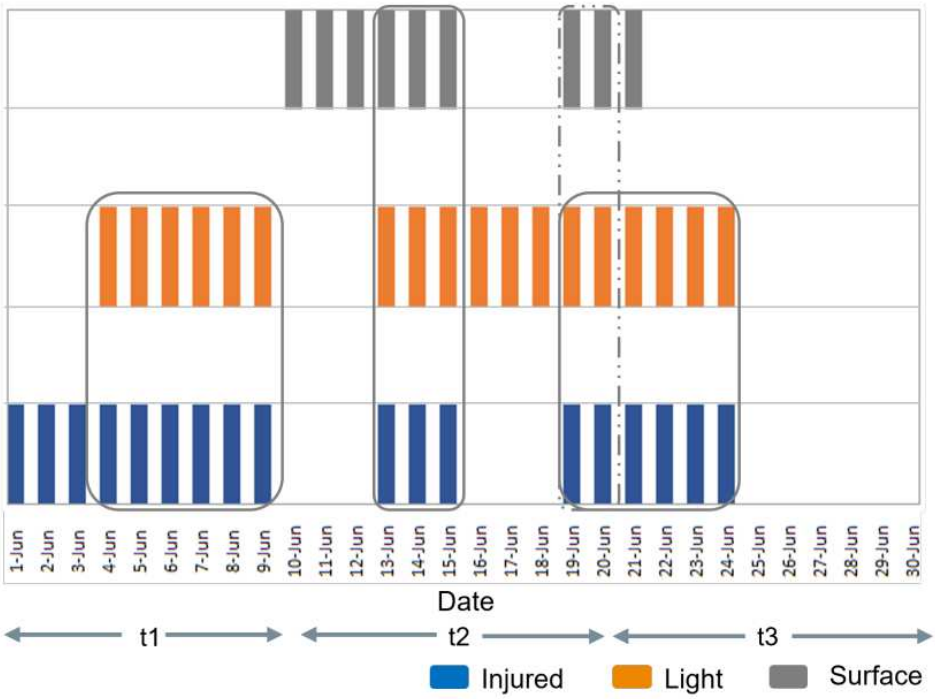

Fig. 6 Multi-domain relationships in NJDOT data for total injured (injured), light condition (light), and surface condition (surface) domains. The three bins are represented as t1, t2, and t3

Table 2 Support, Confidence, and Lift for multi-domain relationships in NJDOT data.

Relationship Found	Supp	Conf	Lift
$Injured \Rightarrow Light$	0.601	0.833	1.157
$Light \Rightarrow Injured$	0.601	0.833	1.157
$Injured, Surface \Rightarrow Light$	0.242	1.000	1.389
$Light, Surface \Rightarrow Injured$	0.242	1.000	1.389
$Injured, Light \Rightarrow Surface$	0.242	0.400	1.111

As shown in Table 3, we observe a strong association between the domains, $t2m$, tcc, and $strd$ such that whenever there is an anomaly in $t2m$ and tcc, there is 100% Confidence of an anomaly being observed in $strd$, with a lift of 3.23. We compared the domain pair correlation of the overlapping anomalous windows and the rest of the data and found a stronger correlation among the overlapping anomalous windows (Table 5).

4.3 Evaluation

To validate the ERA5 data, we conducted a correlation analysis of pairs of overlapping anomalous windows and used this as a validation layer, as described in Table 4. We found that more than

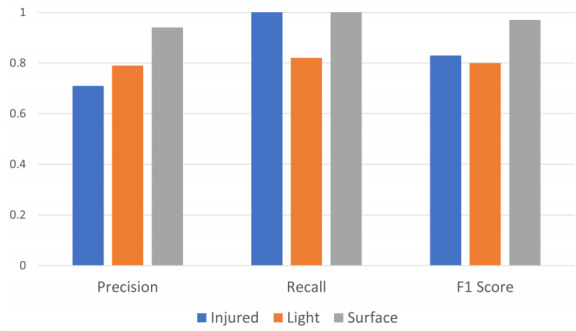

Fig. 7 Precision, Recall, and Accuracy score for total injured, light condition, and surface condition domains of the labeled anomalies in NJDOT

50% of anomalous window pairs had overlaps, indicating a strong direct relationship between the domains in this dataset. We computed the correlation in the bins with overlapping anomalies, between each domain pair to measure performance using Pearson correlation. We observed a substantial improvement in the correlation between the variables in these bins with overlapping windows compared to the correlation of the variables in general. This indicates that the relationships are

Table 3 Support, Confidence, and Lift for the multi-domain relationships.

Relationship Found	Support	Confidence	Lift
$t2m, tcc \Rightarrow strd$	0.15	1.00	3.23
$tcc \Rightarrow strd$	0.20	0.64	2.05
$strd \Rightarrow tcc$	0.20	0.64	2.05
$ssrd \Rightarrow smlt$	0.13	0.41	1.32
$smlt \Rightarrow ssrd$	0.13	0.41	1.32
$t2m \Rightarrow ssrd$	0.10	0.32	1.03
$ssrd \Rightarrow t2m$	0.10	0.32	1.03
$t2m \Rightarrow strd$	0.10	0.32	1.03
$strd \Rightarrow t2m$	0.10	0.32	1.03

Table 4 Absolute correlation coefficient at bins with overlapping anomalous windows.

Bin No.	strd and tcc	smlt and ssrd	ssrd and t2m	strd and t2m
Bin 1	0.780	0.560	0.213	
Bin 2	0.727			0.749
Bin 3	0.761		0.677	
Bin 4	0.773	0.194		
Bin 5		0.250		0.322
Bin 6	0.701	0.287		0.725
Bin 7		0.146		0.391
Bin 8	0.724		0.685	
Bin 9		0.430		0.428
Bin 10	0.805		0.314	
Bin 11	0.768	0.350		
Bin 12				0.555
Bin 13			0.479	
Bin 14	0.864		0.310	
Bin 15		0.138		
Bin 16	0.517			
Bin 17	0.366	0.510		
Bin 18	0.746			
Bin 19				0.401
Bin 20			0.266	
Bin 21	0.425			
Bin 22	0.550	0.348		

Table 5 Domain pair correlation across all bins and bins with overlapped anomalous windows.

Domain pair	Overlapped windows	Base value
smlt and *ssrd*	0.321	0.129
strd and *tcc*	0.679	0.526
strd and *t2m*	0.510	0.407
ssrd and *t2m*	0.421	0.240

indeed highlighting a stronger correlation during these time periods as compared to the entire time series.

We used the correlation of each domain pair as the base value for evaluating the correlation score in the overlapping anomalous windows. Specifically, we compared the Pearson correlation coefficient of the anomalous windows with the base value. We present the results of our analysis in Table 5, which show that the bins with overlapped anomalous windows are more strongly correlated. Overall, our correlation analysis provides valuable insights into the relationships between different domains in the ERA5 dataset. Our findings detailed in Table 6 are confirmed by several work.

5 Conclusion

We have proposed an algorithm to discover temporal associations across multiple distinct domains using the time windows with unusual events. We employ the concept of overlap of anomalies and proximity of anomalies to discover the direct and time-delayed relations. We evaluated the accuracy of our multi-domain anomaly detection model on ERA5, and NJDOT data. In ERA5 data, there was a significant overlapping and delayed relationship between the domains, a strong mutual association between snow melt, 2-meter air temperature, surface thermal radiation downward, and total cloud cover was observed across multiple locations. Our findings are corroborated by previous findings (6). In our future work, we plan to analyze spatiotemporal anomalous windows across multiple locations to identify anomalous associations. Hence, we propose to create a spatial-neighborhood segmentation [23] of Greenland and deploy our model to explore the spatiotemporal association between the anomalous windows of these domains in the spatial neighborhood. The disparity between West Antarctica losing ice at an accelerating rate and East Antarctica's growing ice sheet is an exciting research area where our model can provide explainability.

Furthermore, cumulative analysis of the temporal anomalous windows could provide an understanding and explainability of the relationship between melting polar ice and shifting Earth's crust. We could also deploy our model to study oceanography to create new knowledge about the rising sea level.

Table 6 Ground truth validation: Comparison of our findings with previous findings

Research Item	Our Findings	Previous Findings
Domain Relationship	Table 3 shows that there is a strong association between $t2m$, tcc, and $strd$ with a confidence of 1.00 and lift of 3.23. We also observed a mutual association between tcc and $strd$ with confidence of 0.64 and lift of 2.05. Analyzing the correlation between domain pairs with overlapping anomalous relationship Table 5 indicates that the pair of $strd$ and tcc has the highest correlation score, the correlation score at the overlapping anomalous windows is higher than correlation of all the data points.	[5, 21, 22] analyzed and discussed the relationship between cloud cover and thermal radiation. Cloud cover and thermal radiation are some of the significant features that influenced two of the biggest extreme snowmelt events in the Arctic.

Declarations

- **Funding** This work is funded in part by NSF award #2118285

- **Conflict of interest** On behalf of all authors, the corresponding author state that there is no conflict of interest.
- **Authors' contributions** All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by [Tolulope Ale]. The first draft of the manuscript was written by [Tolulope Ale] and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

[1] Blázquez-García, A., Conde, A., Mori, U., Lozano, J.A.: A review on outlier/anomaly detection in time series data (2020)

[2] Coulson, S., Lubeck, M., Mitrovica, J.X., Powell, E., Davis, J.L., Hoggard, M.J.: The global fingerprint of modern ice-mass loss on 3-d crustal motion. Geophysical Research Letters **48**(16), 2021–095477 (2021)

[3] Liang, D., Guo, H., Zhang, L., Wang, M., Wang, L., Liang, L., Shirazi, Z.: Analyzing antarctic ice sheet snowmelt with dynamic big earth data. International Journal of Digital Earth **14**(1), 88–105 (2021)

[4] Li, S., Liao, J., Zhang, L.: Extraction and analysis of elevation changes in antarctic ice sheet from cryosat-2 and sentinel-3 radar altimeters. Journal of Applied Remote Sensing **16**(3), 034514 (2022)

[5] Bennartz, R., Shupe, M.D., Turner, D.D., Walden, V., Steffen, K., Cox, C.J., Kulie, M.S., Miller, N.B., Pettersen, C.: July 2012 greenland melt extent enhanced by low-level liquid clouds. Nature **496**(7443), 83–86 (2013)

[6] Li, J., Izakian, H., Pedrycz, W., Jamal, I.: Clustering-based anomaly detection in multivariate time series data. Appl. Soft Comput. **100**, 106919 (2021)

[7] Teng, M.: Anomaly detection on time series. In: 2010 IEEE International Conference on Progress in Informatics and Computing, vol. 1, pp. 603–608 (2010). IEEE

[8] Golmohammadi, K., Zaiane, O.R.: Time series contextual anomaly detection for detecting market manipulation in stock market. In: 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 1–10 (2015). IEEE

[9] Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. Advances in neural information processing systems **29** (2016)

[10] Li, G., Jung, J.J.: Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges. Information Fusion **91**, 93–102 (2023) https://doi.org/10.1016/j.inffus.2022.10.008

[11] Segura-Delgado, A., Gacto, M.J., Alcalá, R., Alcalá-Fdez, J.: Temporal association rule mining: An overview considering the time variable as an integral or implied component. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **10**(4), 1367 (2020)

[12] Nair, B.B., Mohandas, V., Nayanar, N., Teja, E., Vigneshwari, S., Teja, K.: A stock trading recommender system based on temporal association rule mining. SAGE Open **5**(2), 2158244015579941 (2015)

[13] Dai, H., Wang, Z.: Data mining of serial-episode association rules using gap-constrained minimal occurrences. International Journal of Business Intelligence and Data Mining **8**(3), 288–305 (2013)

[14] Zhou, H., Wei, W., Shimada, K., Mabu, S., Hirasawa, K.: Time related association rules mining with attributes accumulation mechanism and its application to traffic prediction. In: 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence), pp. 305–311 (2008). IEEE

[15] Yamtani, Y., Ohira, M.: An exploratory analysis for studying software evolution: Time-delayed correlation analysis. In: 2014 6th International Workshop on Empirical Software Engineering in Practice, pp. 13–18 (2014). IEEE

[16] SaTScan, K.M.: v9. 6: software for the spatial, temporal, and space-time scan statistics (2018)

[17] Kleinman, K.: rsatscan: tools, classes, and methods for interfacing with satscan standalone software. See https://CRAN. R-project. org/package= rsatscan (2015)

[18] Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207–216 (1993)

[19] Commission, N.J.M.V., et al.: Njtr-1 form field manual [accessed 2015 may 1]. http://www. state. nj. us/transportation/ref-data/accident/pdf. NJTR-1Field_Manual. pdf (2011)

[20] Hersbach, H.: The era5 atmospheric reanalysis. In: AGU Fall Meeting Abstracts, vol. 2016, pp. 33–01 (2016)

[21] Djoumna, G., Mernild, S., Holland, D.: Meteorological conditions and cloud effects on surface radiation balance near helheim glacier and jakobshavn isbræ (greenland) using ground-based observations. Frontiers in Earth Science **8**, 616105 (2021)

[22] Large ice loss on the greenland ice sheet in 2019. National Snow and Ice Data Center-Greenland Ice Sheet Today (2019)

[23] Janeja, V.P., Palanisamy, R.: Multi-domain anomaly detection in spatial datasets. Knowledge and information systems **36**, 749–788 (2013)