Temperature Monitoring in Concrete Dams: A Survey on Data Preprocessing and Anomaly Detection

www.surveyx.cn

Abstract

This survey paper systematically reviews methodologies for preprocessing and anomaly detection in temperature monitoring data from concrete dams, focusing on handling outliers and missing values to enhance structural health assessments. The paper underscores the critical role of temperature monitoring in maintaining dam integrity and identifies key challenges, such as data quality issues stemming from noise, outliers, and missing values. It highlights robust statistical techniques, convex optimization, regression methods, clustering, and filtering as essential preprocessing tools for ensuring data reliability. The paper also explores various anomaly detection methods, including statistical, machine learning, and hybrid approaches, emphasizing their applicability in real-time monitoring scenarios. Case studies illustrate the practical utility of these methods, demonstrating their effectiveness in diverse contexts. Emerging trends suggest advancements in real-time monitoring systems, unified frameworks for time series data, and hybrid approaches combining robustness with computational efficiency. Future research directions include optimizing algorithms for high-dimensional datasets, developing automated outlier detection tools, and exploring data fusion technologies. This survey contributes significantly to dam safety and environmental monitoring, offering insights into improving the reliability and accuracy of structural health assessments through advanced data processing and anomaly detection techniques.

1 Introduction

1.1 Importance of Temperature Monitoring

Temperature monitoring is vital for maintaining the structural integrity and environmental safety of concrete dams, primarily to prevent cracking and other issues arising from thermal stresses that could jeopardize stability and functionality [1]. Timely detection through monitoring enables the mitigation of potential problems, thus preserving both the dam's operational lifespan and the surrounding environment. Integrating temperature data with other sensor inputs enhances the understanding of dam behavior, contributing to the development of comprehensive monitoring solutions. This approach aligns with the broader goal of multimodal data fusion, which aims to create standardized frameworks for integrating diverse data types across various disciplines [2]. Consequently, effective temperature monitoring not only addresses immediate structural concerns but also aids in developing more robust and adaptive monitoring systems.

1.2 Challenges in Sensor Data

Analyzing temperature monitoring data from concrete dams presents challenges, primarily due to outliers and missing values that compromise data quality and reliability. Outliers, or data points significantly deviating from the majority [3], can skew results, particularly in multivariate analyses [4]. This distortion is exacerbated in classical Functional Linear Regression (FLR) models, where

www.surveyx.cn

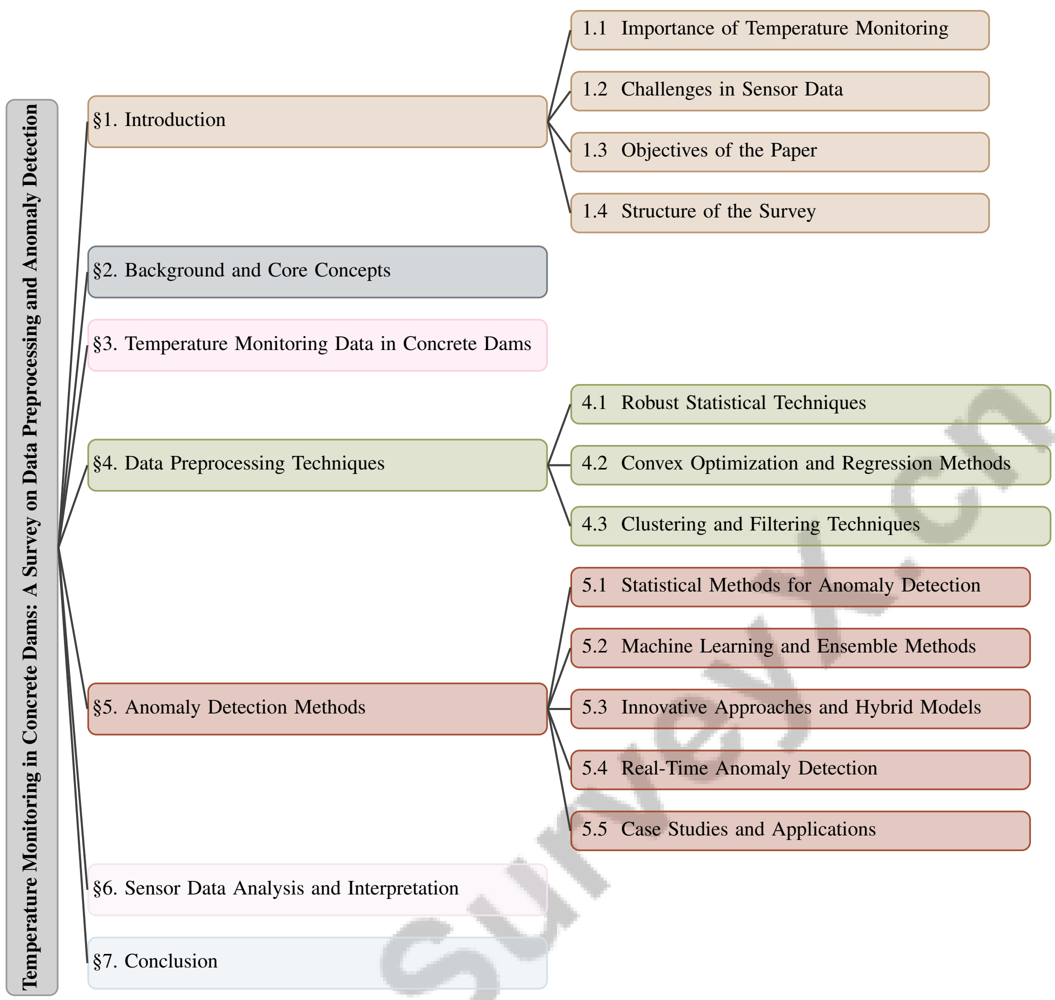


Figure 1: chapter structure

www.SurveyX.cn

outliers can severely impair performance [5]. Furthermore, traditional robust estimators struggle with contaminated cells or cases, complicating outlier detection [6].

The challenge intensifies with the necessity for computational scalability in real-time monitoring, especially in large-scale data streams where unexpected outliers may arise [7]. Existing anomaly detection techniques often lack interpretability, making it difficult for auditors to understand the reasons behind detected anomalies [8]. Additionally, casewise and cellwise outliers in multivariate functional data significantly undermine the effectiveness of statistical process monitoring techniques [9].

Handling missing values further complicates the scenario, particularly in incomplete multi-way data that includes both rowwise and cellwise outliers [10]. This is especially critical in time series data, where identifying significant deviations from overall distributions remains challenging [11]. The lack of a consensus framework for outlier detection complicates identification, as the presence and nature of outliers are often unknown [12]. These challenges highlight the need for robust statistical methods and advanced preprocessing techniques to enhance the reliability and accuracy of temperature monitoring data, thereby improving structural health assessments of concrete dams.

1.3 Objectives of the Paper

This survey paper aims to systematically review and synthesize methodologies for data preprocessing and anomaly detection in temperature monitoring for concrete dams. A significant focus is on developing efficient data preprocessing techniques capable of addressing missing values and outliers

AI-generated, for reference only.

in time series data, which are critical for enhancing the reliability of energy systems [13]. The paper proposes robust statistical methods to improve the reliability of outlier probability estimates, thereby enhancing data quality and structural health assessments [4].

Additionally, the survey addresses the need for effective explanations of anomalies detected by machine learning models, particularly relevant in complex domains such as banking [8]. This involves advancing state estimation methods to manage model mismatches caused by outliers and proposing new approaches to monitor and visualize temporal discrepancies in data entries [11].

The paper also seeks to accurately estimate loadings and scores in the PARAFAC model, especially in the context of incomplete data containing both rowwise and cellwise outliers [10]. By comparing existing outlier detection techniques, the survey aims to guide data scientists in selecting appropriate algorithms for machine learning model development, facilitating advancements in anomaly detection [14].

Furthermore, the paper emphasizes the necessity for standardized methods to integrate and process multimodal data, addressing inefficiencies and inconsistencies in current practices. By establishing clear objectives, this paper contributes significantly to dam safety and environmental monitoring, enhancing the reliability and accuracy of structural health assessments through advanced methodologies. This includes analyzing thermal stress in roller-compacted concrete dams and systematically generating artificial outliers for improved data analysis. Experimental planning and numerical simulations will address critical factors influencing temperature gradients and cracking during dam construction, thereby providing a comprehensive framework for evaluating structural integrity under various environmental conditions [15, 1].

1.4 Structure of the Survey

The survey is systematically organized into several key sections to comprehensively address the multifaceted aspects of temperature monitoring in concrete dams. The paper begins with an **Introduction** that underscores the significance of temperature monitoring for structural integrity and environmental safety while identifying the challenges posed by outliers and missing values in sensor data. This section sets the stage for the primary objectives of the survey, focusing on reviewing data preprocessing and anomaly detection methodologies.

Following the introduction, the **Background and Core Concepts** section provides foundational knowledge on concrete dam structures and the role of temperature monitoring. It includes definitions of essential terms such as outliers, missing values, data preprocessing, and anomaly detection, highlighting their relevance to dam safety and maintenance.

The **Temperature Monitoring Data in Concrete Dams** section delves into the types of sensors used, the nature and patterns of the collected data, and common issues such as noise, outliers, and missing values, emphasizing the importance of accurate data for assessing dam health.

Subsequently, the **Data Preprocessing Techniques** section reviews various methods for handling outliers and missing values, including robust statistical techniques, convex optimization, regression methods, clustering, and filtering techniques. The strengths and limitations of each method are evaluated in the context of dam monitoring.

The **Anomaly Detection Methods** section explores different techniques applicable to temperature monitoring data, such as statistical methods, machine learning approaches, and hybrid models, including case studies and examples of successful applications in dam monitoring.

The penultimate section, **Sensor Data Analysis and Interpretation**, discusses the analysis of sensor data post-preprocessing and anomaly detection, explaining how the data is interpreted to assess the structural health of the dam.

Finally, the **Conclusion** summarizes the key findings and insights, discussing the implications of effective data preprocessing and anomaly detection for dam safety. The findings indicate not only the identification of current methodologies for temperature monitoring and anomaly detection but also highlight emerging trends and potential future research directions, particularly in leveraging global and local information from normal samples to enhance detection accuracy in various applications, including intrusion detection and fault diagnosis in time series data [16, 17].The following sections are organized as shown in Figure 1.

AI-generated, for reference only.

www.SurveyX.cn

2 Background and Core Concepts

2.1 Overview of Concrete Dam Structures

Concrete dams are vital infrastructures for water management, supporting irrigation, flood control, and hydroelectric power generation. Their design considers factors like height, water capacity, and site geology, with temperature monitoring being essential for maintaining structural integrity. For instance, a study on a 45-meter dam in Northern Vietnam underscores the impact of temperature variations on structural stability [1]. Temperature differentials can cause thermal cracking due to concrete expansion and contraction, necessitating effective monitoring to detect anomalies for timely interventions. By integrating temperature data with other sensor inputs, a comprehensive understanding of dam behavior is achieved, aligning with multimodal data fusion goals to combine diverse data types across environmental and energy systems [2]. This approach not only addresses immediate structural issues but also aids in developing adaptive monitoring systems for long-term dam safety and efficiency.

2.2 Key Terms and Definitions

Understanding key terms is crucial for analyzing and interpreting sensor data in temperature monitoring of concrete dams. Outliers are data points that significantly deviate from expected patterns, potentially skewing analyses and leading to erroneous conclusions. They can manifest as rowwise or cellwise outliers in multivariate data, complicating preprocessing and analysis [10]. Detecting these anomalies is critical, as they may indicate structural issues or data collection errors [14]. Missing values, often due to sensor malfunctions or transmission errors, can introduce biases if not addressed, necessitating techniques for maintaining dataset integrity [10]. Data preprocessing involves cleaning and preparing raw data for analysis, addressing outliers and missing values to ensure reliability for subsequent modeling [10]. Anomaly detection focuses on identifying deviations that may signal structural or environmental issues, using various models to enhance detection robustness [14]. Sensor data analysis interprets temperature sensor data to assess dam conditions, predicting potential failures and enabling timely interventions, especially in high-dimensional settings where traditional methods may falter [10]. These concepts are essential for managing thermal stresses in roller-compacted concrete dams during construction, which arise from exothermic heat during cement hydration, potentially leading to significant temperature gradients and cracking. Engineers equipped with this knowledge can better ensure the structural integrity and safety of concrete dams, mitigating risks associated with temperature fluctuations during both construction and operational phases [2, 1].

2.3 Significance of Outliers and Missing Values

Outliers and missing values in temperature monitoring data for concrete dams pose significant challenges to data quality and dam safety. Outliers, which deviate markedly from the dataset, can distort analyses and lead to incorrect conclusions about the dam's structural health. This is particularly problematic in multivariate contexts where outliers can undermine the reliability of outlyingness indices [18]. Such distortions may obscure genuine structural anomalies, complicating timely interventions. Methods relying on single data points for estimating material properties are especially susceptible to noise and outliers [19], highlighting the need for robust preprocessing techniques to mitigate their impact and ensure accurate reflections of the dam's condition. Missing values, often due to sensor malfunctions or transmission errors, result in incomplete datasets that can bias analyses if not properly addressed. The lack of comprehensive benchmarks for outlier detection techniques further hampers the development of unified platforms for testing various algorithms [14]. This gap underscores the necessity for advanced preprocessing methods capable of effectively addressing missing values to preserve dataset integrity. Addressing outliers and missing values is crucial for maintaining the accuracy and reliability of temperature monitoring data, enabling the detection of potential issues in concrete dams. Ensuring structural integrity and safety involves identifying thermal stress and cracking—common problems arising from temperature gradients during construction—before they escalate into major structural failures. This proactive approach is vital for the longevity and reliability of hydraulic engineering structures [20, 3, 1].

In the context of assessing the health of concrete dams, the integration of temperature monitoring data plays a crucial role. As illustrated in Figure 2, the hierarchical structure of this data encompasses various sensor types, distinct data characteristics, and the challenges associated with data collec-

AI-generated, for reference only.

www.SurveyX.cn

tion. Furthermore, the figure highlights the significance of accurate temperature data in the overall assessment of dam health, emphasizing its importance in ensuring structural integrity and safety. This visual representation not only aids in understanding the complexities involved in temperature monitoring but also reinforces the critical nature of precise data in the maintenance and evaluation of concrete dams.

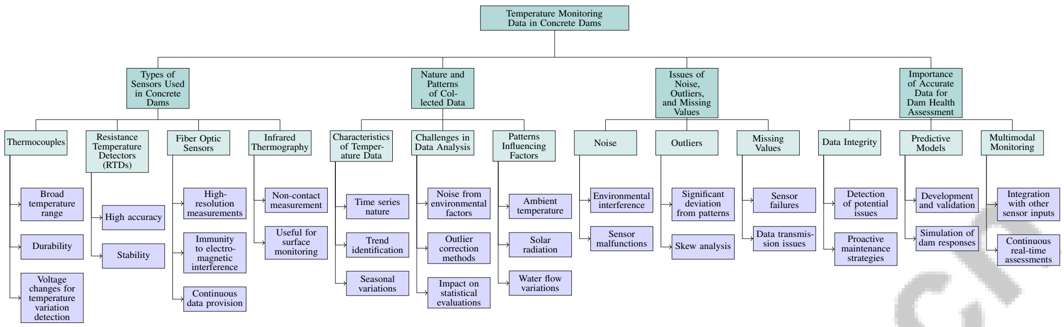


Figure 2: This figure illustrates the hierarchical structure of temperature monitoring data in concrete dams, encompassing sensor types, data characteristics, challenges, and the significance of accurate data for dam health assessment.

3 Temperature Monitoring Data in Concrete Dams

3.1 Types of Sensors Used in Concrete Dams

Temperature monitoring in concrete dams utilizes various sensor technologies crucial for assessing structural health and operational safety. Thermocouples are widely used due to their broad temperature range and durability, generating measurable voltage changes in response to temperature variations, which are beneficial for outlier detection and robust regression modeling [5, 2, 15]. Resistance Temperature Detectors (RTDs) are also common for their high accuracy and stability, based on predictable resistance changes with temperature.

Fiber optic sensors represent advanced technology, providing high-resolution temperature measurements along the fiber's length, enhancing data collection in automated manufacturing and environmental monitoring [21, 2, 22, 17, 23]. Their immunity to electromagnetic interference and ability to provide continuous data make them advantageous for large structures like dams, where localized temperature changes can indicate structural issues.

Infrared thermography offers a non-contact temperature measurement method, useful when direct contact is impractical, such as in medical diagnostics and industrial inspections [4, 24, 3, 17, 25]. While effective for surface monitoring, its limitations in depth data collection necessitate complementary methods for comprehensive structural analysis.

Selecting appropriate sensors for dam monitoring requires consideration of measurement precision, environmental conditions, and spatial distribution needs. With advancements in smart, miniaturized sensors and high-speed data communication, careful evaluation of these factors is vital to optimize data collection and processing, especially when integrating multimodal data for effective monitoring and analysis [3, 2, 26, 7]. Integrating diverse sensor technologies into a cohesive framework fosters a comprehensive understanding of concrete dams' thermal behavior, promoting proactive maintenance and long-term structural integrity.

3.2 Nature and Patterns of Collected Data

Temperature data from sensors in concrete dams exhibit distinct characteristics essential for understanding their thermal behavior. Typically time series in nature, these data capture continuous measurements at regular intervals, facilitating trend identification and seasonal variations critical for establishing baseline operational conditions and detecting anomalies. Recent advancements in outlier detection techniques enhance the reliability of analyses across various research contexts [3, 16].

AI-generated, for reference only.

www.SurveyX.cn

Temperature data are susceptible to noise from environmental factors or sensor malfunctions, complicating accurate analysis. Noise can obscure genuine temperature variations and impact statistical evaluations, especially when outlier correction methods are applied, as these can inflate Type I errors and distort population parameter estimates. Preprocessing steps, such as filtering, are essential for enhancing data quality and ensuring accurate assessments of thermal behavior [21, 3, 15, 24].

Patterns in temperature data often reflect the dam's response to external conditions like ambient temperature, solar radiation, and water flow variations, which can induce thermal stresses leading to expansion and contraction. Understanding these patterns is crucial for forecasting the dam's response to environmental scenarios and for effectively identifying outliers that may indicate underlying issues needing further investigation. This insight enhances the reliability of predictive models and contributes to better management and decision-making regarding dam safety and performance [3, 15, 17].

Spatial variations in temperature data necessitate multiple sensors distributed throughout the dam to capture a comprehensive thermal state. Integrating data from these sensors aids in detecting localized anomalies, such as hotspots, potentially indicating structural issues. However, the presence of outliers and missing values complicates data pattern analysis, necessitating robust statistical methods and advanced preprocessing techniques to ensure analyses accurately reflect the dam's true condition. High-breakdown robust methods can effectively handle outliers in multivariate data, ensuring reliable estimation of parameters such as covariance and regression coefficients. Employing appropriate outlier detection tools, like median absolute deviation for univariate outliers and Mahalanobis-MCD distance for multivariate outliers, is crucial for maintaining data integrity [3, 27, 24]. Understanding the nature and patterns of temperature data enables engineers and researchers to develop more effective monitoring strategies to ensure concrete dams' safety and integrity.

3.3 Issues of Noise, Outliers, and Missing Values

Temperature monitoring data in concrete dams often face challenges related to noise, outliers, and missing values, compromising data integrity. Noise, caused by environmental interference or sensor malfunctions, obscures the true signal, necessitating preprocessing techniques like filtering to avoid erroneous thermal behavior interpretations [10].

Outliers, or data points significantly deviating from expected patterns, can skew analysis and lead to incorrect conclusions about structural health. Their presence is particularly problematic in multivariate analyses, distorting statistical models and indices' reliability [18]. Detecting and managing outliers is critical for ensuring data accurately reflects dam conditions, requiring robust statistical methods to differentiate genuine structural anomalies from data errors [14].

Missing values, often due to sensor failures or data transmission issues, complicate analyses by introducing gaps in datasets. These gaps can bias results if not properly addressed, potentially masking underlying patterns [10]. Managing missing values is particularly challenging in complex data structures like time series, where preserving temporal dependencies is essential for accurate analysis.

A comprehensive data preprocessing strategy is crucial for addressing these quality challenges. This strategy should incorporate advanced techniques for filtering noise, accurately detecting and managing both univariate and multivariate outliers, and systematically imputing missing values. Utilizing robust detection methods, such as median absolute deviation for univariate outliers and Mahalanobis-MCD distance for multivariate outliers, can significantly enhance statistical analysis reliability and improve research findings' reproducibility. Additionally, pre-registering outlier management plans can help maintain analytical rigor and reduce flexibility in data analysis, leading to more valid inferences [3, 15, 23]. By improving data quality through these methods, researchers and engineers can ensure more reliable assessments of a dam's structural health, contributing to the safety and longevity of this critical infrastructure.

3.4 Importance of Accurate Data for Dam Health Assessment

Accurate temperature monitoring data are essential for effectively evaluating dam health and safety. The integrity of such data influences the ability to detect and address potential structural issues before they escalate. Precise data facilitate the identification of thermal anomalies indicating underlying

AI-generated, for reference only.

www.Survey...

weaknesses or environmental impacts, enabling timely interventions and proactive maintenance strategies [18].

In concrete dams, where temperature-induced stresses can lead to cracking and other structural problems, data reliability is critical. Distorted or incomplete data, resulting from noise, outliers, or missing values, can obscure genuine trends, leading to incorrect assessments of dam conditions [10]. Employing robust preprocessing techniques is vital to mitigate these issues, ensuring data accurately reflects the dam's true thermal behavior.

Moreover, accurate data are crucial for developing and validating predictive models in structural health monitoring systems. These models rely on high-quality data to simulate the dam's response to various environmental conditions and anticipate potential failure scenarios. Inaccurate data can compromise model performance, reducing predictive accuracy and reliability [19].

Integrating precise temperature data with other sensor inputs enhances the understanding of dam behavior, contributing to more comprehensive monitoring frameworks. This multimodal approach supports adaptive and robust monitoring systems capable of providing continuous, real-time assessments of dam health [2].

The importance of accurate data extends beyond immediate structural concerns, playing a critical role in ensuring the long-term safety and efficiency of concrete dams. By facilitating precise evaluations of structural health through advanced multimodal data collection and processing techniques, accurate data enable effective maintenance strategies that protect infrastructure integrity and minimize potential environmental impacts, ultimately enhancing safety and sustainability across sectors such as transportation, energy systems, and environmental management [2, 15, 19].

4 Data Preprocessing Techniques

Category	Feature	Method
Robust Statistical Techniques	Outlier Management	CRCA[28], OMF[3]
	Resilient Data Modeling	RSS[29]
Convex Optimization and Regression Methods	Robust Optimization Techniques	L1-Splines[30], ExLasso[31], WDRO[32]
	Simulation and Modeling	OTM-CD[1], RGP-BM[33]
	Statistical and Probabilistic Methods	LRR[19]
Clustering and Filtering Techniques	Iterative Methods	STAIR[34], MP[10]
	Clustering Techniques	AWT[35], UBF-DDC-GRE-C[36]
	Outlier and Noise Handling	DPMRPF[37], QTF[38], GARD[39]

Table 1: The table provides a comprehensive summary of various data preprocessing methods categorized into robust statistical techniques, convex optimization and regression methods, and clustering and filtering techniques. Each category details specific features and the associated methods, highlighting their relevance in enhancing the quality of temperature monitoring data in concrete dams. The methods are referenced by recent studies, underscoring their academic and practical significance.

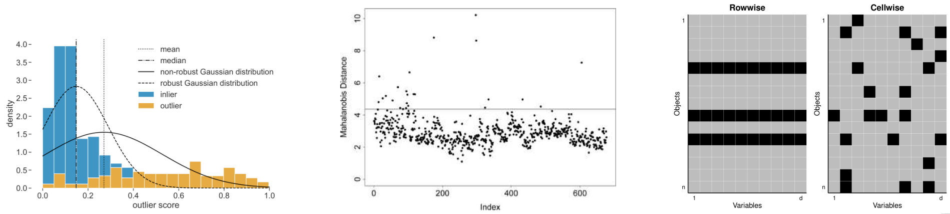
Data preprocessing for temperature monitoring in concrete dams is crucial for ensuring data integrity and reliability, particularly through robust statistical techniques. Table 3 offers a comprehensive classification of data preprocessing techniques essential for maintaining data integrity in temperature monitoring of concrete dams, underscoring the importance of robust statistical methods in managing outliers and enhancing data quality. Outliers can significantly distort analytical outcomes; thus, employing resilient methods is essential. The following subsections explore various robust statistical techniques that enhance the quality of temperature monitoring data, addressing their methodologies and practical implications.

4.1 Robust Statistical Techniques

Robust statistical techniques are essential for enhancing the quality of temperature monitoring data in concrete dams, especially amidst outliers. These methods offer resilience against data contamination, ensuring accurate interpretations and reliable decision-making. The Mahalanobis minimum covariance determinant method identifies multivariate outliers by considering data covariance structures, crucial for maintaining analytical integrity in multivariate analyses [3]. Quantile Tracking Filters (QTFs) provide a novel approach for real-time processing and effective outlier mitigation in dynamic monitoring environments [38]. In regression analysis, the Greedy Algorithm for Robust Denoising (GARD) combines least squares optimization with Orthogonal Matching Pursuit (OMP) for

AI-generated, for reference only.

robust data denoising [39]. The Distributionally Robust Optimization (DRO) framework minimizes worst-case expected loss over distributions close to the empirical distribution, enhancing regression model robustness [32]. These techniques improve data accuracy and reliability, ensuring critical temperature-related anomalies are detected and addressed [29, 1, 7, 23, 27].



(a) Robustness of Gaussian distribution to outliers[29]

(b) Mahalanobis Distance Scatter Plot[27]

(c) Two Graphs Representing Data in Different Ways[28]

Figure 3: Examples of Robust Statistical Techniques

As illustrated in Figure 3, robust statistical techniques are crucial for maintaining data integrity and reliability amid outliers. The first image compares robust and non-robust Gaussian distributions, highlighting the former's superior outlier mitigation. The second showcases Mahalanobis distance's role in identifying multivariate outliers, while the third emphasizes diverse visualization techniques' importance in enhancing data preprocessing.

4.2 Convex Optimization and Regression Methods

Method Name	Optimization Techniques	Noise Mitigation	Application Contexts
L1-Splines[30]	Convex Optimization Techniques	Laplacian Distribution Noise	Temperature Monitoring
OTM-CD[1]	Finite Element Analysis	Temperature Control Optimization	Concrete Dam Construction
LRR[19]	Robust Regression	Noise And Outliers	Material Property Estimation
RGP-BM[33]	Maximum Likelihood Estimation	Robust Estimation	Environmental Data
ExLasso[31]	Robust Regression	Thresholding Techniques	Temperature Monitoring
EMORF[40]	Gaussian Filtering Results	Outlier-resilient Filtering	Target Tracking Application
WDRO[32]	Convex Optimization	Outlier Detection	Temperature Monitoring

Table 2: Overview of various convex optimization and regression methods, detailing their optimization techniques, noise mitigation strategies, and application contexts in temperature monitoring and construction. The table highlights the diversity of approaches, including L1-Splines and Extreme Lasso, and their specific uses in enhancing data preprocessing and structural health assessments.

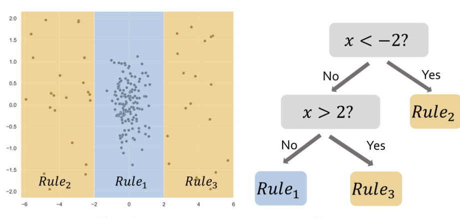
Convex optimization and regression methods are integral to preprocessing temperature monitoring data in concrete dams, particularly for mitigating outliers and noise. These techniques optimize fitting processes, ensuring robust and reliable analyses. Table 2 presents a comprehensive comparison of convex optimization and regression methods utilized in preprocessing temperature monitoring data, illustrating their respective optimization techniques, noise mitigation strategies, and application contexts. One approach applies convex optimization to fit splines robust to outliers, assuming a Laplacian noise distribution [30]. Integrating experimental planning with finite element analysis optimizes temperature control during dam construction, preserving structural integrity [1]. Robust regression methods, like k-nearest neighbors in Local Robust Regression (LRR), enhance data preprocessing by considering multiple data points [19]. Transforming robust Gaussian Process (GP) regression problems into standard GP challenges through bias modeling facilitates solutions for predictive mean and variance [33]. The Extreme Lasso approach, minimizing an p norm loss function with a Lasso penalty, provides a robust framework for modeling extreme values [31]. The Expectation-Maximization Outlier Robust Filtering (EMORF) method addresses correlated measurement noise and outliers using the Expectation-Maximization framework [40]. The Wasserstein Distributionally Robust Optimization (WDRO) approach formulates regression problems as minimizing worst-case expected absolute loss over a Wasserstein ball of distributions [32]. Collectively, these methods ensure accurate reflections of structural health in concrete dams and support reliable safety assessments.

www.SurveyX.cn

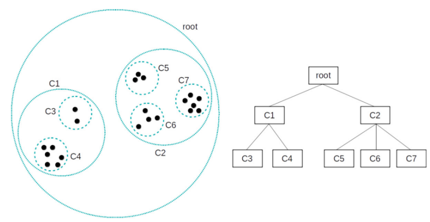
AI-generated, for reference only.

4.3 Clustering and Filtering Techniques

Clustering and filtering techniques are vital for preprocessing temperature monitoring data in concrete dams, enabling effective outlier and noise management. Advanced methods, such as Dirichlet process mixture models, automatically determine components for accurate clustering and outlier detection, enhancing robustness against non-Gaussian data challenges [6, 35, 41, 23]. Clustering-based subsampling organizes large datasets into manageable clusters, improving processing efficiency [36]. Innovative clustering methods, including autoencoders, utilize network structures to compute outlier scores based on reconstruction errors, effectively distinguishing anomalies [42]. Filtering techniques like Quantile Tracking Filters (QTFs) offer dynamic estimation of signal quantiles, providing robust protection against outliers [38]. The DPMRPF method adapts to outlier structures, improving state estimation robustness [37]. MacroPARAFAC iteratively estimates loadings and scores while managing missing values and outliers, maintaining dataset integrity [10]. The Greedy Algorithm for Robust Denoising (GARD) minimizes the 0 norm of outlier noise, enhancing data preprocessing robustness [39]. These techniques enable systematic cleaning and organization of temperature monitoring data, supporting proactive maintenance strategies and addressing significant temperature-induced stresses [1, 16, 7, 35, 41].



(a) Decision Tree for Classification[34]



(b) Clustering and Hierarchical Structure[35]

Figure 4: Examples of Clustering and Filtering Techniques

As depicted in Figure 4, clustering and filtering techniques play a crucial role in organizing and simplifying complex datasets for effective analysis. The first image illustrates a decision tree for classification, demonstrating data categorization into distinct clusters governed by specific rules. The second image showcases a clustering and hierarchical structure, grouping data into clusters and organizing them into a tree-like hierarchy. These examples highlight the effectiveness of clustering and filtering techniques in transforming raw data into structured insights, essential for informed decision-making and advanced data analysis [34, 35].

Feature	Robust Statistical Techniques	Convex Optimization and Regression Methods	Clustering and Filtering Techniques
Outlier Management	Multivariate Detection	Noise Mitigation	Dynamic Estimation
Optimization Method	Covariance-based	Least Squares	Autoencoders
Application Context	Temperature Monitoring	Structural Health	Data Organization

Table 3: This table provides a comparative analysis of various data preprocessing techniques used in temperature monitoring of concrete dams. It highlights the features of robust statistical techniques, convex optimization and regression methods, and clustering and filtering techniques, focusing on their approaches to outlier management, optimization methods, and application contexts. These techniques are crucial for enhancing data integrity and reliability in critical monitoring applications.

5 Anomaly Detection Methods

Anomaly detection methods are essential for effectively monitoring temperature data in concrete dams, addressing deviations that may indicate structural issues. This section examines key methodologies, beginning with statistical techniques that identify deviations from expected patterns, facilitating timely interventions and ensuring structural integrity.

AI-generated, for reference only.

5.1 Statistical Methods for Anomaly Detection

Statistical methods are pivotal in detecting anomalies within temperature monitoring data from concrete dams, highlighting deviations that suggest structural concerns or environmental impacts. The Mahalanobis distance is crucial for identifying anomalies in complex datasets, offering a scale-invariant measure of distance between a data point and the mean of a multivariate distribution, particularly effective in high-dimensional data [4]. Nonparametric strategies, such as those by [43], provide robust anomaly detection without requiring prior data distribution knowledge, enhancing adaptability in dynamic settings. The HS-filter by [6] excels in distinguishing true anomalies from noise in high-dimensional data.

Outliagnostics minimizes masking and swamping effects by utilizing data point relationships and their contributions to the outlying score [11], enhancing anomaly detection robustness. Furthermore, [44] proposes model adaptability during outlier detection and model fitting, addressing traditional inference challenges for reliable outcomes.

These statistical methods provide a comprehensive toolkit for effective anomaly detection in temperature monitoring, addressing challenges like outlier identification amidst normal patterns, employing unsupervised techniques in high-dimensional datasets, and optimizing scoring functions to quantify abnormality. Aligning techniques with data characteristics enhances structural health assessment reliability, contributing to the safety and longevity of concrete dams [45, 16, 17, 20].

5.2 Machine Learning and Ensemble Methods

Machine learning and ensemble methods have advanced anomaly detection in temperature monitoring data from concrete dams, surpassing traditional techniques by uncovering intricate patterns and relationships, thus improving detection accuracy and resilience in noisy environments. GALDetector leverages global similarities and local sparsity scores from normal samples to identify anomalies without anomalous data, with data-driven scoring functions aiding classification based on abnormality, crucial for intrusion detection and predictive maintenance [17, 20].

Robust models like Robust Function-on-Function Linear Regression (RFLR) maintain predictive model integrity by capturing regression behaviors while identifying outliers without their influence [39]. Ensemble methods, such as the ALTBI framework, enhance stability and efficiency in outlier detection, consistently outperforming existing techniques [4].

Unsupervised techniques, incorporating random subspace and subsampling ensembles of Dirichlet process Gaussian mixtures, bolster robustness and computational efficiency in outlier detection [10]. The SPINEX algorithm enhances detection accuracy in high-dimensional data through similarity principles and higher-order interactions [44]. Robust Partial Least Squares (RPLS) mitigates outlier influence, yielding reliable estimates crucial for structural health assessments [32].

Collectively, these machine learning and ensemble methods offer a robust framework for anomaly detection in temperature monitoring data, significantly enhancing structural health assessment accuracy and reliability for concrete dams. This improvement is vital for identifying potential thermal stresses and cracks during construction and operation, ensuring infrastructure safety and longevity while accounting for noise and outliers in material properties [19, 1].

5.3 Innovative Approaches and Hybrid Models

Innovative approaches and hybrid models have emerged as effective strategies for enhancing anomaly detection in temperature monitoring data of concrete dams. These methods integrate advanced techniques, such as robust statistical measures and machine learning algorithms, to improve detection accuracy and interpretability of outliers, addressing limitations of traditional methods often hindered by contaminating data or the presence of normal samples [24, 3, 20, 34, 17].

The STAIR framework exemplifies innovation by focusing on outlier summarization and interpretation-aware optimization, contrasting with conventional decision tree methods prioritizing classification accuracy over interpretability [34]. Robust statistical methods transform outlier scores, enhancing reliability in datasets with high variability [29].

GALDetector integrates global and local information for anomaly detection without anomalous data [17], while repeated inversion of covariance matrices reduces computational complexity, facilitating

AI-generated, for reference only.

efficient data fitting [46]. FastPCS enhances the reliability of the outlyingness index in multivariate datasets [18]. The Outlier-Insensitive Kalman Filter (OIKF) maintains optimal performance in dynamic environments, effectively managing sporadic data anomalies [47].

DTOR provides understandable rules for explaining anomalies, improving interpretability in decision-making processes [8]. SPINEX incorporates higher-order feature interactions and a robust similarity framework, significantly improving detection accuracy [48]. The ECOD method captures rare events in distribution tails, simplifying outlier detection while maintaining accuracy [45]. Depth-filters utilizing statistical data depth functions enhance outlier detection in multivariate data [6].

Integrating these innovative approaches and hybrid models significantly enhances anomaly detection capabilities in temperature monitoring data, improving structural health assessment accuracy and reliability for concrete dams. Techniques like GALDetector utilize global and local information from normal samples to identify potential anomalies, even without genuine outlier data. The generation of artificial outliers further aids in benchmarking detection algorithms and refining identification processes, contributing to more effective monitoring and risk management of critical infrastructure [15, 17].

5.4 Real-Time Anomaly Detection

Real-time anomaly detection in temperature monitoring for concrete dams is crucial for timely interventions and maintaining structural integrity. The dynamic nature of temperature data necessitates robust methods capable of continuous operation. The L-CUSUM method, recognized for its robustness to infrequent outliers, allows recursive implementation suitable for real-time monitoring [7].

Cumulative sum techniques continuously monitor statistical evidence over time, declaring anomalies only when substantial evidence accumulates, thereby minimizing false positives [43]. The GALDetector method enhances real-time detection through a three-stage process: calculating global and local scores, selecting potential anomalies, and constructing a weighted detection model [17]. By integrating these techniques, real-time anomaly detection systems for temperature monitoring in concrete dams achieve higher accuracy and reliability, facilitating proactive maintenance and safeguarding structural health.

5.5 Case Studies and Applications

The application of advanced anomaly detection methods in dam monitoring is illustrated through various case studies, demonstrating the efficacy of these techniques in real-world scenarios. The implementation of SPINEX has shown superior performance compared to existing algorithms, achieving outstanding results on synthetic datasets and competitive outcomes on real datasets, emphasizing its robustness and adaptability in diverse monitoring environments [48].

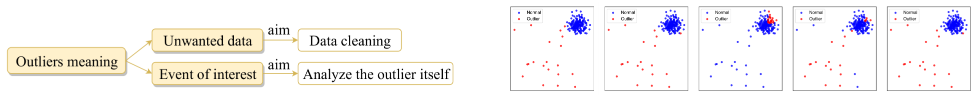
The ECOD method, evaluated across 30 benchmark datasets, has proven effective against 11 state-of-the-art techniques, confirming its proficiency in identifying anomalies [45]. The Outliagnostics approach, assessed with real-world datasets, has demonstrated its capability in visualizing outlying temporal profiles, vital for detecting potential structural issues [11].

Additionally, applying a robust multivariate functional control method in monitoring resistance spot welding in the automotive industry illustrates the practical utility of advanced anomaly detection methods [9]. The real-time nonparametric anomaly detection method tested on cyber-attack datasets further demonstrates efficacy in detecting anomalies without prior data distribution knowledge [43].

These case studies underscore the importance of selecting and applying appropriate anomaly detection techniques tailored to specific dam monitoring challenges. By employing advanced methodologies proven effective across various applications, engineers and researchers can significantly improve the reliability and accuracy of structural health assessments for concrete dams. This is crucial for mitigating temperature-induced stresses and potential cracking during construction and accurately analyzing material properties amidst noise and outliers, ensuring the safety and longevity of concrete dams [24, 1, 15, 19, 44].

As shown in Figure 5, anomaly detection is a critical research area involving identifying data points that deviate significantly from expected patterns. The examples presented illustrate two

AI-generated, for reference only.



(a) Data Cleaning and Analyzing Outliers[16] (b) Outlier Detection in a Multivariate Data Set[45]

Figure 5: Examples of Case Studies and Applications

distinct approaches: the first example, "Data Cleaning and Analyzing Outliers," outlines a systematic flowchart for managing unwanted data and analyzing outliers. The second example, "Outlier Detection in a Multivariate Data Set," visually represents outlier detection within complex data structures, emphasizing the distinction between normal data points and anomalies. Together, these examples provide valuable insights into the methodologies and applications of anomaly detection, showcasing its relevance and utility across various scenarios [16, 45].

6 Sensor Data Analysis and Interpretation

6.1 Data Interpretation Techniques

Interpreting sensor data is crucial for assessing the health of concrete dams, necessitating robust methodologies that accurately reflect structural conditions and risks. Weighted Likelihood Estimation (WLE) enhances functional time series forecasting by integrating weighted likelihood principles, thereby increasing robustness against outliers and ensuring accurate interpretations amidst anomalies [49]. The Directional Outlyingness (DO) method computes outlyingness for each point in a functional dataset, capturing both local and global deviations, which is essential for identifying structural issues requiring further investigation [25].

Robust Function-on-Function Linear Regression (RFLR) minimizes outlier impact through robust estimation, capturing underlying functional relationships to predict dam behavior under varying conditions [5]. The Adaptive Wavelet Transform (AWT) algorithm identifies cluster sizes based on data structure, incorporating implicit outlier detection and determining optimal cluster numbers with user-defined thresholds. This is particularly effective in analyzing crowd-sourced temperature data and revealing correlations with urban land-use characteristics [35, 45, 7]. Such adaptability aids in understanding complex interactions affecting temperature and stress in dams, facilitating informed decision-making to mitigate risks like cracking.

Robust Gaussian Process Regression (GPR) enhances data interpretation by estimating bias terms within a Gaussian process framework, ensuring computational efficiency and reliability [33]. Future research should explore various statistical depth functions for developing new filters, as suggested by [6], to further improve robustness and accuracy in sensor data interpretation, thereby enhancing structural health assessments of concrete dams.

6.2 Sensor Data Analysis in Dam Monitoring

Analyzing sensor data is critical for understanding and predicting dam behavior, facilitating timely interventions and maintenance. Advanced statistical and computational techniques extract meaningful insights from collected data, addressing challenges such as outlier presence that can distort results and lead to inaccurate interpretations of structural health. Quantile regression's robustness against outliers provides a foundation for addressing issues posed by censored data [50].

Significant advancements in outlier detection techniques for time series data, common in temperature monitoring, enhance prediction accuracy and reliability [16]. These techniques enable engineers to better interpret sensor data, identify structural issues, and implement proactive measures for dam safety and longevity. Incorporating advanced anomaly detection techniques into monitoring systems supports sophisticated predictive models that simulate dam behavior under varying conditions. Techniques leveraging global and local information from normal samples allow anomaly detection even without genuine outlier data. For instance, GALDetector's three-stage process assesses normal behavior and identifies anomalies, significantly improving predictive capabilities and response strategies [15, 17]. This capability is essential for anticipating potential failure scenarios and implementing

www.SurveyX.cn

AI-generated, for reference only.

effective maintenance strategies, contributing to the structural integrity and operational efficiency of concrete dams.

7 Conclusion

7.1 Emerging Trends and Future Directions

The field of temperature monitoring and anomaly detection for concrete dams is witnessing significant advancements, particularly through the integration of sophisticated real-time monitoring systems. These systems facilitate dynamic adjustments in temperature management during both construction and operational phases, thereby bolstering structural integrity and efficiency. The pursuit of unified frameworks for handling diverse time series data is becoming increasingly prominent, enhancing the adaptability and robustness of detection methodologies across various applications. Future research should focus on refining data preprocessing techniques, such as the development of constrained splines and the exploration of L1-Optimal Splines' sparsity properties.

Efforts are also directed toward creating efficient algorithms for robust methods, with hybrid approaches that combine robustness and computational efficiency gaining attention, especially for high-dimensional data. Investigating adaptive methods for parameter optimization, including the extension of these techniques to address complex material behaviors, presents a promising research avenue. Furthermore, enhancing methods for detecting and managing cellwise outliers in high-dimensional datasets through automated strategies could significantly improve anomaly detection accuracy and efficiency.

Optimizing algorithms like the fqn for outlier detection in complex data streams is essential, with potential applications spanning diverse monitoring environments. Developing novel methods to distinguish between relevant statistical outliers and contaminating outliers, alongside evaluating non-parametric tests' efficacy, remains a critical area for exploration. Additionally, extending analyses to multidimensional subspaces and employing weighted PCA approaches to mitigate high noise samples' influence could further enhance detection capabilities.

Research should also focus on optimizing design parameters to boost detection accuracy and resilience against environmental variations, with applications in various IoT domains presenting promising opportunities. Integrating the strengths of existing outlier detection techniques while addressing their limitations is crucial for advancing the field. Exploring the theoretical properties of PCS and its application in dam monitoring could provide valuable insights.

The development of standardized frameworks, fostering interdisciplinary collaboration, and exploring emerging trends in data fusion technologies are vital for the continued evolution of temperature monitoring and anomaly detection. Extending methodologies to accommodate non-normal distributions and complex high-dimensional scenarios will further improve these systems' effectiveness. Exploring alternative regularization techniques, refining hyperparameter selection, and applying the Extreme Lasso across different fields are also potential research avenues. Integrating the Outlier-Insensitive Kalman Filter with data-driven variants and examining the impact of various priors could enhance real-time detection capabilities. Further refinements in model selection procedures and robust techniques are likely to yield improvements in anomaly detection methodologies, suggesting promising directions for future research.

AI-generated, for reference only.

www.surveyx.cn

References

[1] Nikolay Aniskin and Chuc Nguyen Trong. The thermal stress of roller-compacted concrete dams during construction. In *MATEC Web of Conferences*, volume 196, page 04059. EDP Sciences, 2018.

[2] Chun-An Chou, Xiaoning Jin, Amy Mueller, and Sarah Ostadabbas. Mmdf2018 workshop report, 2018.

[3] Christophe Leys, Marie Delacre, Youri L Mora, Daniël Lakens, and Christophe Ley. How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration. *International Review of Social Psychology*, 32(1), 2019.

[4] Hamid Ghorbani. Mahalanobis distance and its application for detecting multivariate outliers. *Facta Universitatis, Series: Mathematics and Informatics*, pages 583–595, 2019.

[5] Harjit Hullait, David S. Leslie, Nicos G. Pavlidis, and Steve King. Robust function-on-function regression, 2019.

[6] Giovanni Saraceno and Claudio Agostinelli. Robust multivariate estimation based on statistical depth filters, 2021.

[7] Ruizhi Zhang, Yajun Mei, and Jianjun Shi. Robust change detection for large-scale data streams, 2021.

[8] Riccardo Crupi, Daniele Regoli, Alessandro Damiano Sabatino, Immacolata Marano, Massimiliano Brinis, Luca Albertazzi, Andrea Cirillo, and Andrea Claudio Cosentini. Dtor: Decision tree outlier regressor to explain anomalies, 2024.

[9] Christian Capezza, Fabio Centofanti, Antonio Lepore, and Biagio Palumbo. Robust multivariate functional control chart, 2024.

[10] Mia Hubert and Mehdi Hirari. Macroparafac for handling rowwise and cellwise outliers in incomplete multi-way data, 2024.

[11] Vung Pham and Tommy Dang. Outliagnostics: Visualizing temporal discrepancy in outlying signatures of data entries, 2019.

[12] Seoyoung Cho, Jaesung Hwang, Kwan-Young Bak, and Dongha Kim. Altbi: Constructing improved outlier detection models via optimization of inlier-memorization effect, 2025.

[13] Michał Narajewski, Jens Kley-Holsteg, and Florian Ziel. tsrobprep - an r package for robust preprocessing of time series data, 2021.

[14] Yue Zhao, Zain Nasrullah, and Zheng Li. Pyod: A python toolbox for scalable outlier detection. *Journal of machine learning research*, 20(96):1–7, 2019.

[15] Georg Steinbuss and Klemens Böhm. Generating artificial outliers in the absence of genuine ones – a survey, 2020.

[16] Ane Blázquez-García, Angel Conde, Usue Mori, and Jose A. Lozano. A review on outlier/anomaly detection in time series data, 2020.

[17] Fan Xu, Nan Wang, and Xibin Zhao. Exploring global and local information for anomaly detection with normal samples, 2023.

[18] Kaveh Vakili and Eric Schmitt. Finding multivariate outliers with fastpcs, 2013.

[19] Yoshihiro Kanno. Simple heuristic for data-driven computational elasticity with material data involving noise and outliers: A local robust regression approach, 2018.

[20] Myrto Limnios, Nathan Noiry, and Stéphan Clémençon. Learning to rank anomalies: Scalar performance criteria and maximization of two-sample rank statistics, 2021.

[21] David Hong, Laura Balzano, and Jeffrey A. Fessler. Towards a theoretical analysis of pca for heteroscedastic data, 2016.

www.SurveyX.cn

AI-generated, for reference only.

[22] Mohsen Ebadi, Shojaeddin Chenouri, and Stefan H. Steiner. Phase i analysis of high-dimensional processes in the presence of outliers, 2022.

[23] Amulya Agarwal and Nitin Gupta. Comparison of outlier detection techniques for structured data, 2021.

[24] Marjorie Fonnesu and Nicola Kuczewski. Doubts on the efficacy of outliers correction methods, 2019.

[25] Peter J. Rousseeuw, Jakob Raymaekers, and Mia Hubert. A measure of directional outlyingness with applications to image data and video, 2017.

[26] Carina Gerstenberger. Robust discrimination between long-range dependence and a change in mean, 2020.

[27] Mia Hubert, Peter J. Rousseeuw, and Stefan Van Aelst. High-breakdown robust multivariate methods, 2008.

[28] Jakob Raymaekers and Peter J. Rousseeuw. Challenges of cellwise outliers, 2023.

[29] Philipp Röchner, Henrique O. Marques, Ricardo J. G. B. Campello, Arthur Zimek, and Franz Rothlauf. Robust statistical scaling of outlier scores: Improving the quality of outlier probabilities for outliers (extended version), 2024.

[30] Masaaki Nagahara and Clyde F. Martin. L1-optimal splines for outlier rejection, 2013.

[31] Andersen Chang, Minjie Wang, and Genevera Allen. Sparse regression for extreme values, 2021.

[32] Ruidi Chen and Ioannis Ch. Paschalidis. A robust learning algorithm for regression models using distributionally robust optimization under the wasserstein metric, 2018.

[33] Chiwoo Park, David J. Borth, Nicholas S. Wilson, Chad N. Hunter, and Fritz J. Friedersdorf. Robust gaussian process regression with a bias model, 2020.

[34] Yu Wang, Lei Cao, Yizhou Yan, and Samuel Madden. Interpretable outlier summarization, 2023.

[35] Awt – clustering meteorological time series using an aggregated wavelet tree.

[36] Andy Leung, Victor J. Yohai, and Ruben H. Zamar. Multivariate location and scatter matrix estimation under cellwise and casewise contamination, 2016.

[37] Bin Liu. Robust particle filtering via bayesian nonparametric outlier modeling, 2019.

[38] Alexei V. Nikitin and Ruslan L. Davidchack. Quantile tracking filters for robust fencing in intermittently nonlinear filtering, 2019.

[39] George Papageorgiou, Pantelis Bouboulis, Sergios Theodoridis, and Kostantinos Themelis. Robust linear regression analysis - a greedy approach, 2015.

[40] Aamir Hussain Chughtai, Muhammad Tahir, and Momin Uppal. Emorf/s: Em-based outlier-robust filtering and smoothing with correlated measurement noise, 2024.

[41] Dongwook Kim, Juyeon Park, Hee Cheol Chung, and Seonghyun Jeong. Unsupervised outlier detection using random subspace and subsampling ensembles of dirichlet process mixtures, 2024.

[42] Tung Kieu, Bin Yang, Chenjuan Guo, and Christian S Jensen. Outlier detection for time series with recurrent autoencoder ensembles. In *Ijcai*, pages 2725–2732, 2019.

[43] Mehmet Necip Kurt, Yasin Yilmaz, and Xiaodong Wang. Real-time nonparametric anomaly detection in high-dimensional settings, 2020.

AI-generated, for reference only.

www.SurveyX.cn

[44] Toshiaki Tsukurimichi, Yu Inatsu, Vo Nguyen Le Duy, and Ichiro Takeuchi. Conditional selective inference for robust regression and outlier detection using piecewise-linear homotopy continuation, 2022.

[45] Zheng Li, Yue Zhao, Xiyang Hu, Nicola Botta, Cezar Ionescu, and George H. Chen. Ecod: Unsupervised outlier detection using empirical cumulative distribution functions, 2022.

[46] M. de Jong. On the repeated inversion of a covariance matrix, 2017.

[47] Shunit Truzman, Guy Revach, Nir Shlezinger, and Itzik Klein. Outlier-insensitive kalman filtering using nuv priors, 2022.

[48] MZ Naser and Ahmed Z Naser. Spinex: Similarity-based predictions with explainable neighbors exploration for anomaly and outlier detection, 2024.

[49] Ufuk Beyaztas and Han Lin Shang. Forecasting functional time series using weighted likelihood methodology, 2019.

[50] Soo-Heang Eo, Seung-Mo Hong, and HyungJun Cho. Identification of outlying observations with quantile regression for censored data, 2014.

www.SurveyX.cn

AI-generated, for reference only.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

www.SurveyX.cn

AI-generated, for reference only.