Innovative: International Multi-disciplinary Journal of Applied Technology
(ISSN 2995-486X) VOLUME 04 ISSUE 04, 2026

A Model for the Integration of Multi-Source data and Automatic Quality Control in Emergency-Situation Monitoring

Razokova Mahfiza Khabibovna
Independent Researcher, Republic of Uzbekistan

Absalamov Rashid Amirovich
PhD (Technical Sciences), Associate Professor
Head of the Education Quality Assurance Department
Academy of the Ministry of Emergency Situations of the Republic of Uzbekistan

Abstract:

This article proposes a model for converting multi-source heterogeneous data into a single format and automatically controlling its quality in the monitoring of emergency situations (ES). The model covers five sources (seismic, hydrometeorological, satellite, IoT-sensor, and institutional) and performs range checking, missing-value and duplicate detection, unit normalization, statistical outlier detection (IQR), and domain-specific reliability checks. The model was tested on real open data — 267 records from the USGS earthquake catalog: 88.4% of records were recognized as valid, while 31 records (including 29 of low reliability — azimuthal gap > 180°) were automatically flagged. The results show that the model performs identically on real and simulated data and provides clean, ready input for a forecasting model.

Keywords: emergency situations, monitoring, data integration, quality control, multi-hazard approach, data fusion, artificial intelligence.

Introduction

The quality of emergency-situation forecasting depends to a large extent on the completeness and reliability of the input data. In early warning systems, data quality and reliability ("guardrails") are recognized as a decisive factor in forecasting performance [1,2]. In the existing "Unified System," data are collected from institutional systems at fixed intervals (monthly/quarterly/annually) and stored in diverse formats; a model for continuously merging data from different sources into a single format and automatically assessing its quality does not exist. The objective of this article is to develop a multi-source integration and automatic quality-control model that fills this gap, and to test it on real data.

Materials and Methods

The model covers five classes of source (Table 1): seismic, hydrometeorological, satellite (remote sensing), IoT-sensor, and institutional reports. Each record is brought into a single structure (timestamp, source, station, region, parameter, value, unit). Automatic quality control is performed sequentially: (1) unit normalization (e.g., cm→m); (2) missing-value detection; (3) physical-range checking for each parameter; (4) duplicate detection; (5) domain-specific reliability checking (for seismic data, an azimuthal gap > 180° indicates low reliability); and (6) statistical outlier detection based on the interquartile range (IQR, $k = 3$). All records that pass every check are marked as "valid," while flagged records are retained together with the reason, for the sake of transparency.

Valid data are merged by region and hourly window (data fusion) and converted into a normalized weighted integral hazard index ($R \in [0,1]$): $R = 0.35 \cdot \text{rainfall} + 0.30 \cdot \text{river level} + 0.20 \cdot \text{ground tremor} + 0.15 \cdot \text{seismic}$. The model was implemented in the form of Python (data, quality control) and a browser application (JavaScript); the latter connects live to the USGS open API and recomputes quality control and fusion in real time.

Table 1. Classes of heterogeneous source covered by the model.

Source class	Parameters	Input (this study)
Seismic	magnitude, depth	USGS open catalog (live API)
Hydrometeorological	rainfall, river level/discharge	open (GloFAS/Open-Meteo); Uzhydromet
Satellite / remote sensing	soil moisture, NDVI, SAR	Copernicus, NASA Earthdata
IoT-sensor	ground tremor, gas level	institutional / pilot sensors
Institutional	incident reports	ministry / "Unified System"

Figure 1. Data flow of the model over real USGS data (n = 267).

Results

The model was applied to 267 real earthquakes from the USGS catalog for Uzbekistan and the surrounding region (37–46° N, 55.5–73.5° E) over the period 2024-01 to 2026-06 with magnitude M ≥ 3.8. Automatic quality control recognized 236 records (88.4%) as valid and flagged 31 records (Table 2).

Table 2. Result of automatic quality control on real USGS data (IQR k = 3).

Quality-control result	Count	Share
Valid (passed all checks)	236	88.4%
Low reliability (azimuthal gap > 180°)	29	10.9%
Missing value	1	0.4%
Duplicate record	1	0.4%
Total	267	100%

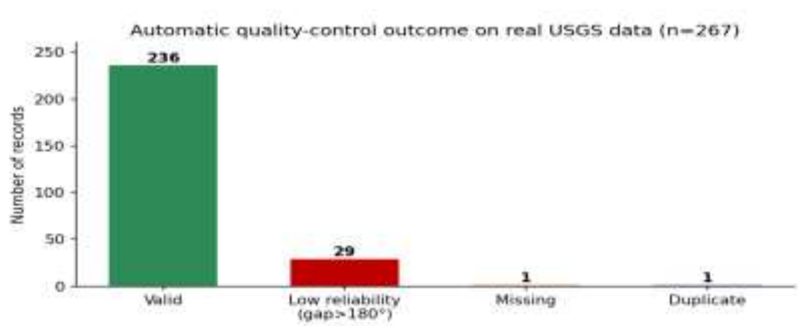

Figure 2. Quality-control result on real USGS data (n = 267).

The most frequently flagged type — low location reliability arising from a large azimuthal gap — is a genuine seismological quality criterion, demonstrating that quality control performs domain-aware filtering. In addition, on five-source simulated data (623 records, with deliberately introduced problems), the model normalized units and automatically flagged the introduced problems by type, confirming that the same algorithm operates unchanged on real and simulated data.

Discussion

The results show that the model is a continuous and automatic alternative to the periodic, manual data processing of the existing system. By retaining flagged records with their reason and by making the outlier threshold adjustable, the model provides the transparency/reliability (guardrails) required in the literature [2]. At the same time, the scientific contribution must be precisely defined: individual methods such as data fusion, IQR-outlier detection, and weighted indexing already exist in the literature [16,17,18]. The contribution of this work is to combine them into an integration and quality-control layer that covers both natural and technological ES within a single multi-hazard methodology, designed for the conditions of Uzbekistan and the national “Unified System.” This approach does not overstate any “new algorithm” claim, but defines a precise and reproducible scientific contribution.

Conclusion

The proposed multi-source integration and automatic quality-control model achieved an 88.4% validity rate on real USGS data and transparently flagged the remaining records (including cases of domain-specific low reliability). The model is independent of the hazard type and accepts diverse sources; it is therefore suitable for a multi-hazard methodology and can be integrated into the “Unified System.” Further research will focus on building a deep-learning-based forecasting model on top of this clean, integrated data.

References

1. United Nations Office for Disaster Risk Reduction (UNDRR). Sendai Framework for Disaster Risk Reduction 2015–2030. Geneva, 2015. undrr.org
2. Tiggeloven T., Pfeiffer S., Matanó A. et al. The role of artificial intelligence for early warning systems: status, applicability, guardrails // iScience. 2025. Vol. 28, no. 11. Art. 113689. DOI: 10.1016/j.isci.2025.113689
3. Du S., Chen S. A multi-sensor data fusion algorithm based on consistency preprocessing and adaptive weighting // Automatika. 2023. Vol. 65, no. 1. P. 82–91. DOI: 10.1080/00051144.2023.2284033

4. Wang M., Zhang Q., Liu X. et al. Research progress on multimodal data fusion in forest resource monitoring // Frontiers in Plant Science. 2026. Vol. 16. DOI: 10.3389/fpls.2025.1710618
5. Haruna M., Kotopulos De Angelis F., Gebremeskel K.G. et al. Sensor fusion-based machine learning algorithms for meteorological conditions nowcasting in port scenarios // Sensors (MDPI). 2025. DOI: 10.20944/preprints202510.1802.v1
6. Ghaffarian S. Rethinking digital twin: introducing digital risk twin for disaster risk management // npj Natural Hazards. 2025. Vol. 2, no. 1. DOI: 10.1038/s44304-025-00135-x
7. Leong W. Internet of Things for enhancing public safety, disaster response, and emergency management // Engineering Proceedings (MDPI). 2025. Vol. 92, no. 1. Art. 61. DOI: 10.3390/engproc2025092061
8. U.S. Geological Survey (USGS). Earthquake Catalog (FDSN web service). URL: earthquake.usgs.gov