CHAPTER 2

SOFTWARE METHODS FOR ENVIRONMENTAL THREAT ASSESSMENT BASED ON REAL-TIME DATA ANALYSIS

Oleshchenko L. M.
DOI https://doi.org/10.30525/978-9934-26-653-9-2

INTRODUCTION

The rapid intensification of anthropogenic activities, urbanization, and climate change has led to a significant increase in the frequency and severity of environmental threats, including air and water pollution, extreme weather events, and ecosystem degradation. These challenges pose serious risks to public health, economic stability, and sustainable development at both regional and global levels. As a result, timely detection, assessment, and mitigation of environmental threats have become critical priorities for modern societies.

In recent years, the widespread deployment of environmental sensors, IoT devices, satellite platforms, and citizen sensing systems has enabled the continuous collection of large volumes of heterogeneous environmental data.

Unlike traditional environmental monitoring, which relies on periodic measurements and offline analysis, contemporary monitoring infrastructures generate high-velocity data streams that reflect rapidly changing environmental conditions in near real time. This shift has fundamentally transformed environmental data from static datasets into dynamic streaming data, requiring new computational paradigms for effective analysis and decision support. The availability of streaming environmental data alone does not guarantee timely or reliable threat assessment. Conventional software systems and analytical methods are often designed for batch processing and are not capable of handling the velocity, volume, and variability inherent in continuous data streams.

Environmental threats are frequently characterized by complex spatiotemporal patterns, nonlinear dynamics, and uncertainty, which further complicate real-time interpretation and forecasting. These limitations highlight the need for advanced software methods that can operate under real-time constraints while ensuring scalability, robustness, and analytical accuracy.

Software methods for environmental threat assessment based on streaming data analysis provide a promising solution to these challenges. By integrating stream processing architectures, intelligent data analytics, and adaptive modeling techniques, such methods enable continuous monitoring, early detection of anomalous environmental events, and predictive assessment of potential risks. The use of machine learning and data-driven models within streaming environments allows systems to learn evolving patterns, adapt to changing conditions, and support proactive rather than reactive decision-making.

The relevance of this research is further reinforced by its practical importance for critical domains such as public health protection, disaster risk reduction, smart city management, and environmental policy development.

Real-time environmental threat assessment systems can support rapid response to pollution incidents, inform early warning mechanisms, and enhance the resilience of urban and natural ecosystems. In this context, the development of scientifically grounded and software-oriented methods for streaming data analysis represents not only a technological challenge but also a societal necessity.

The research of software methods for environmental threat assessment based on streaming data analysis addresses fundamental gaps in existing environmental monitoring systems and contributes to the advancement of intelligent, real-time decision-support technologies capable of responding to the growing complexity and urgency of environmental challenges in the modern world.

The development of software-based environmental monitoring systems is closely linked to advances in sensor networks and data transmission technologies. One of the foundational contributions in this area was provided by Akyildiz I.F., Melodia T. and Chowdhury K.R., who presented a comprehensive survey of wireless multimedia sensor networks. Their work analyzed architectures, communication protocols, and data processing challenges associated with high-volume and heterogeneous sensor data, forming the technological basis for continuous environmental data acquisition and streaming analysis[1].

The transition from data collection to predictive analytics based on streaming data was further demonstrated by Ramakrishnan N., Butler P., Muthiah S., Self N., Khandpur R.P., Saraf P., et al., who introduced the EMBERS system for forecasting civil unrest using open-source indicators. Although focused on social phenomena, this study established important methodological principles for real-time threat assessment, event detection, and large-scale

[1] Akyildiz I.F., Melodia T., Chowdhury K.R. A survey on wireless multimedia sensor networks. Computer Networks, 2007, vol. 51, no. 4, pp. 921–960. https://doi.org/10.1016/j.comnet.2006.10.002.

stream processing that are directly applicable to environmental threat analysis[2].

The integration of sensing technologies into urban-scale systems was examined by Zanella A., Bui N., Castellani A., Vangelista L. and Zorzi M., who explored the role of the IoT in smart cities. Their research highlighted the importance of distributed sensing, data fusion, and real-time processing for monitoring urban environments, including air quality, noise pollution, and energy efficiency. This work emphasized the need for scalable software architectures capable of handling continuous data streams[3].

The societal relevance of environmental monitoring systems is reinforced by studies linking environmental indicators to economic and health outcomes. Mujtaba G. and Shahzad S.J.H. analyzed long-term air pollution data across OECD countries and demonstrated significant relationships between pollutant levels, economic growth, and public health. Their findings underline the necessity of reliable data-driven assessment tools to support sustainable development and policy decision-making[4]. Complementing macro-level analyses, Keswani A., Akselrod H. and Anenberg S. investigated the clinical and public health impacts of air pollution in the context of climate change. Their work provided evidence that real-time environmental monitoring is critical for early detection of health risks and for informing adaptive mitigation strategies, thereby strengthening the case for continuous, software-based environmental threat assessment systems[5].

A broader conceptual framework for integrating environmental monitoring into urban development was provided by Bibri S.E. and Krogstie J., who conducted an extensive interdisciplinary review of smart sustainable cities. Their study emphasized the role of advanced analytics, real-time data streams, and intelligent decision-support systems in addressing environmental challenges within complex urban ecosystems[6]. Recent advances in machine learning and

---

2 Ramakrishnan N., Butler P., Muthiah S., Self N., Khandpur R.P., Saraf P., et al. Beating the News with EMBERS: Forecasting civil unrest using open source indicators. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014, pp. 1799–1808. DOI: 10.1145/2623330.2623373.

3 Zanella A., Bui N., Castellani A., Vangelista L., Zorzi M. Internet of Things for smart cities. IEEE Internet of Things Journal, 2014, vol. 1, no. 1, pp. 22–32. DOI: 10.1109/JIOT.2014.2306328.

4 Mujtaba G., Shahzad S.J.H. Air pollutants, economic growth and public health: implications for sustainable development in OECD countries. Environmental Science and Pollution Research, 2021, vol. 28, pp. 12686–12698. DOI: 10.1007/s11356-020-11212-1.

5 Keswani A., Akselrod H., Anenberg S. Health and clinical impacts of air pollution and linkages with climate change. NEJM Evidence, 2022, vol. 1, no. 7, article EVIDra2200068. DOI: 10.1056/EVIDra2200068.

6 Bibri S.E., Krogstie J. Smart sustainable cities of the future: An extensive interdisciplinary literature review. Sustainable Cities and Society, 2023, vol. 31, pp. 183–212. DOI: 10.1016/j.scs.2017.02.016.

data stream analysis have further expanded the capabilities of threat detection systems. Yang Y., Wu Q., He B., Peng H., Yang R., Hao Z. and Liao Y. proposed a contrastive learning framework for social bot detection based on structural entropy and multi-view data representations. Although applied to social data, the methodological innovations presented in this work demonstrate the potential of advanced representation learning for analyzing complex, evolving data streams[7].

Sun T., Liu C., Chen L., Qian Z., Li P. and Zhu Q. introduced a unified framework for multi-modal detection that leverages dynamic interactions and evolving stances. The proposed approach illustrates how heterogeneous streaming data sources can be integrated and analyzed in real time, offering transferable concepts for multi-source environmental threat assessment[8].

The most recent developments focus explicitly on environmental applications of machine learning (ML). Rajesh M., Babu R.G., Moorthy U. and Easwaramoorthy S.V. presented a ML-driven framework for real-time air quality assessment and predictive environmental health risk mapping. Their work integrates streaming sensor data with predictive models to generate continuous risk assessments, representing a mature example of software-based environmental threat analysis grounded in real-time data processing[9].

Many advanced learning models suffer from limited interpretability, which complicates the explanation and justification of system decisions to stakeholders, policymakers, and domain experts. This issue is particularly critical in environmental applications, where transparency and trust are essential.

The validation of software-based environmental threat assessment systems remains challenging, especially in the context of rare and extreme events. Such events are often underrepresented in historical data, making it difficult to evaluate model robustness and generalization capabilities.

Addressing these limitations requires further research into explainable artificial intelligence, adaptive real-time processing techniques, and resilient software architectures capable of maintaining reliable operation under uncertain and evolving conditions.

---

7 Yang Y., Wu Q., He B., Peng H., Yang R., Hao Z., Liao Y. SeBot: Structural entropy guided multi-view contrastive learning for social bot detection. Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2024, pp. 3841–3852. DOI: 10.1145/3637528.3671871.

8 Sun T., Liu C., Chen L., Qian Z., Li P., Zhu Q. A unified framework for multi-modal rumour detection via multi-level dynamic interaction with evolving stances. Information Processing & Management, 2025, vol. 62, no. 3, article 104066. DOI: 10.1016/j.ipm.2025.104066.

9 Rajesh M., Babu R.G., Moorthy U., Easwaramoorthy S.V. Machine learning-driven framework for realtime air quality assessment and predictive environmental health risk mapping. Sci Rep. 2025, 15(1):28801. DOI: 10.1038/s41598-025-14214-6.

1. Software methods for environmental threat assessment

Environmental threat assessment is a multidisciplinary process aimed at identifying, evaluating, and forecasting risks posed by natural and anthropogenic factors to ecosystems, human health, and critical infrastructure.

From a software engineering perspective, environmental threats are represented as complex events emerging from continuous streams of heterogeneous data, including sensor measurements, satellite observations, meteorological inputs, and socio-environmental indicators.

Traditional assessment approaches relied on retrospective analysis of static datasets, which limited their ability to support timely decision-making. In contrast, modern environmental monitoring systems increasingly operate in real time, requiring software methods capable of continuous data ingestion, on-the-fly analysis, and adaptive response. This shift necessitates the development of scalable, fault-tolerant, and intelligent software solutions that can process high-velocity data streams while maintaining analytical accuracy and robustness.

Software methods for environmental threat assessment can be classified according to several key dimensions: data processing paradigm, analytical approach, temporal characteristics, and level of automation (Tables 1-2).

Table 1

Classification by data processing paradigm

Data processing paradigm	Description	Typical use cases
Batch processing	Analysis of accumulated historical data	Long-term trend analysis, policy evaluation
Micro-batch processing	Near-real-time processing with short delays	Periodic air quality updates
Stream processing	Continuous, event-driven data processing	Real-time pollution detection, early warning systems

Stream processing paradigms are particularly relevant for environmental threat assessment, as they enable immediate reaction to rapidly evolving environmental conditions.

Machine learning–based approaches dominate contemporary systems due to their adaptability and ability to handle non-stationary data streams.

Environmental threat assessment systems are commonly implemented using layered or pipeline-based architectures designed to support continuous data flow.

Typical architectural layers are next.

1. *Data acquisition layer* – sensors, IoT devices, remote sensing platforms.
2. *Stream ingestion layer* – message brokers (e.g., event queues, data streams).

3. *Stream processing layer* – real-time analytics engines.

4. *Analytics and intelligence layer* – ML models, anomaly detection, forecasting.

5. *Visualization and decision support layer* – dashboards, alerts, APIs.

Table 2

Classification by data analytical methods

Method category	Key characteristics	Examples of applications
Statistical methods	Linear/nonlinear modeling, descriptive analytics	Threshold-based pollution alerts
Rule-based systems	Expert-defined rules and heuristics	Regulatory compliance monitoring
Machine learning methods	Data-driven pattern recognition	Air quality prediction, anomaly detection
Deep learning methods	Hierarchical feature learning	Spatiotemporal pollution forecasting
Hybrid methods	Combination of statistical and ML techniques	Robust real-time threat assessment

Such architectures enable modularity and scalability while supporting real-time analytics under high data velocity conditions.

Real-time environmental analytics requires methods that operate under strict latency constraints and can adapt to evolving data distributions (Table 3).

Table 3

Real-time analytical tasks

Task	Objective	Software methods
Event detection	Identify abnormal environmental conditions	Change-point detection, stream clustering
Anomaly detection	Detect deviations from normal behavior	Autoencoders, isolation-based methods
Prediction	Forecast future threat levels	Online regression, recurrent neural networks
Risk assessment	Estimate potential impact	Probabilistic modeling, fuzzy inference

Software-based approaches to environmental threat assessment offer several significant advantages that make them well suited for modern monitoring scenarios. One of the primary benefits is the capability for continuous monitoring and early detection of environmental threats.

By processing streaming data in real-time, software systems can identify emerging anomalies and hazardous conditions at early stages, enabling timely responses and mitigation actions. This continuous operation contrasts with traditional batch-oriented methods, which often fail to capture rapidly evolving environmental phenomena. Another important advantage lies in the scalability of software-based methods. Contemporary stream processing architectures are designed to handle large volumes of heterogeneous data originating from diverse sources, such as sensor networks, remote sensing platforms, and meteorological services. This scalability allows environmental monitoring systems to expand in both spatial and temporal dimensions without fundamental redesign.

Current research in software-based environmental threat assessment is oriented toward enhancing system intelligence, adaptability, and robustness. One prominent trend is the integration of explainable ML techniques into real-time analytics frameworks, aiming to improve model transparency and user trust without sacrificing predictive performance. Another active research direction involves the development of self-adaptive and self-healing environmental monitoring systems that can automatically adjust their behavior in response to changes in data patterns, system performance, or environmental conditions.

In parallel, there is growing interest in the fusion of physical models with data-driven approaches. By combining domain-specific knowledge with machine learning, hybrid models can improve prediction accuracy and ensure consistency with known environmental processes. The emergence of distributed and edge-based analytics platforms is reshaping the architectural landscape of environmental monitoring systems. These platforms enable data processing closer to data sources, reducing latency and enhancing scalability in large-scale deployments.

SaveEcoBot platform represents a significant milestone in the development of environmental information systems in Ukraine.

As the first comprehensive environmental monitoring system of its kind in the country, SaveEcoBot integrates data from multiple sources to provide near real-time information on key environmental indicators.

By aggregating, processing, and visualizing diverse environmental data, the platform supports both public awareness and scientific analysis, thereby addressing a critical gap in national environmental information infrastructure[10].

The importance of SaveEcoBot is grounded in its ability to transform raw environmental data into actionable knowledge. In many regions, particularly those with limited access to centralized monitoring systems, fragmented or delayed environmental data have historically hindered timely decision-making.

SaveEcoBot addresses this challenge by offering an accessible, web-based

10 SaveEcoBot. The first environmental system in Ukraine. URL: https://www.saveecobot.com/en.

interface through which users, researchers, and policymakers can observe current environmental conditions, track trends, and identify emerging threats.

The map in Fig. 1 indicates a spatial cluster of elevated air quality index AQI values concentrated around the industrial zone, suggesting a localized pollution hotspot likely associated with industrial activity and limited dispersion conditions.

Fig. 1 reveals a potential environmental hazard manifested as a localized concentration of elevated air pollution indicators in the vicinity of the industrial area, where multiple monitoring stations report consistently higher values compared to surrounding regions. This spatial pattern suggests an increased risk of adverse air quality impacts, likely linked to industrial emissions and insufficient pollutant dispersion, which may pose a threat to public health and requires continuous monitoring and further investigation.

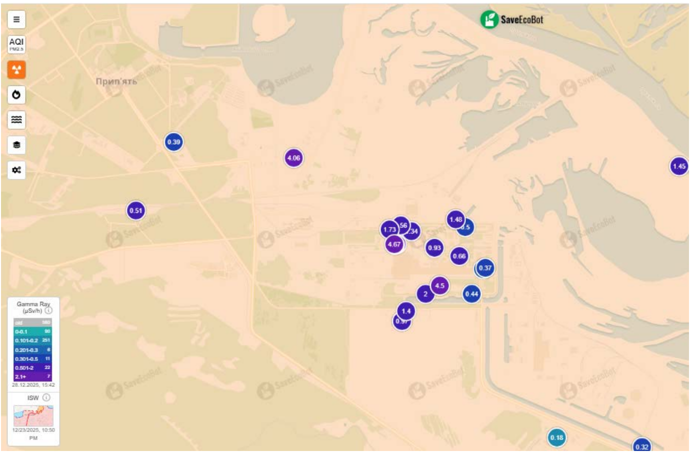

Fig. 1. Spatial distribution of air quality monitoring stations and AQI levels in the industrial area (SaveEcoBot platform, 28.12.2025)

The platform's emphasis on transparency and broad accessibility enhances environmental literacy and fosters engagement with environmental issues across society. From a technological perspective, SaveEcoBot exemplifies the potential of software platforms to support streaming data analysis and real-time environmental assessment. Although not originally designed as a research tool, the system demonstrates how harmonized data flows from sensor networks, governmental databases, and citizen reporting can be synthesized into a coherent analytical environment. This approach aligns with contemporary paradigms of environmental threat assessment, in which data velocity and heterogeneity demand flexible and scalable software solutions. By enabling continuous data ingestion and visualization, SaveEcoBot foreshadows more advanced analytical frameworks that incorporate real-time anomaly detection, predictive modeling, and automated alerts.

The platform's architecture highlights the potential for future enhancements through the integration of ML and real-time processing techniques. For example, predictive modules could be developed to forecast critical environmental events, while anomaly detection algorithms could provide early warnings of pollution spikes or hazardous exposures. Such capabilities would extend SaveEcoBot's utility from a primarily informational system to an intelligent decision-support infrastructure capable of proactive environmental threat assessment. In this regard, SaveEcoBot offers an invaluable case study for the application of advanced software methods in real-world environmental monitoring.

SaveEcoBot is not only a pioneering environmental system within the Ukrainian context but also a practical demonstration of how software platforms can support continuous environmental assessment. Its role in aggregating heterogeneous data, promoting transparency, and enabling broad access to environmental information underscores the importance of integrated digital solutions in addressing complex socio-environmental challenges.

As environmental monitoring evolves toward more intelligent, adaptive models, platforms such as SaveEcoBot provide both inspiration and foundational infrastructure for next-generation analytical systems.

The presented in Fig. 2 fragment of SaveEcoBot data illustrates a snapshot of multi-parameter environmental monitoring collected from a single device at a specific time point. The measurements indicate moderate particulate matter concentrations, with PM_{10} = 52.7 μg/m³ and $PM_{2.5}$ = 28.1 μg/m³, which may exceed recommended health guidelines and suggest potential short-term air quality concerns. In contrast, gaseous pollutants such as formaldehyde and total volatile organic compounds (TVOC) are reported at negligible levels, while basic meteorological parameters (temperature and pressure) remain within normal ranges.

This data analysis fragment demonstrates the platform's capability to integrate heterogeneous sensor readings for comprehensive, time-synchronized environmental assessment and highlights the importance of particulate matter as a key indicator of localized pollution risk.

SaveEcoBot platform provides capabilities for real-time environmental data analysis by continuously collecting, synchronizing, and visualizing heterogeneous sensor streams, including air quality indicators and meteorological parameters. Its analytical functionality enables users to observe short-term dynamics, detect abnormal patterns, and assess correlations between pollutants and environmental conditions as they evolve over time.

	device_id	phenomenon	value	logged_at
0	24764	pm1	2.618	2025-07-31 10:17:00
1	24764	pm25	3.805	2025-07-31 10:17:00
2	24764	pm10	6.515	2025-07-31 10:17:00
3	24764	temperature	17.831	2025-07-31 10:17:00
4	24764	humidity	82.502	2025-07-31 10:17:00
5	24764	pressure_pa	984315.110	2025-07-31 10:17:00
6	24764	pm1	1.998	2025-07-31 10:18:00
7	24764	pm25	3.076	2025-07-31 10:18:00
8	24764	pm10	5.510	2025-07-31 10:18:00
9	24764	temperature	17.910	2025-07-31 10:18:00

```
def read_csv_safely(path: str) -> pd.DataFrame:
    df = pd.read_csv(path)
    df.columns = [c.strip() for c in df.columns]
    return df

dfs_raw = {p: read_csv_safely(p) for p in paths}

for p, d in dfs_raw.items():
    print("\nFILE:", p)
    display(d.head(10))
    print("shape:", d.shape)
    print("columns:", list(d.columns))
```

Fig. 2. Fragment of SaveEcoBot data analysis from IoT device 24764

By integrating time-series visualization, threshold comparison with regulatory limits, and interactive exploration, SaveEcoBot supports rapid situational awareness and evidence-based decision-making for environmental risk assessment (Fig. 3).

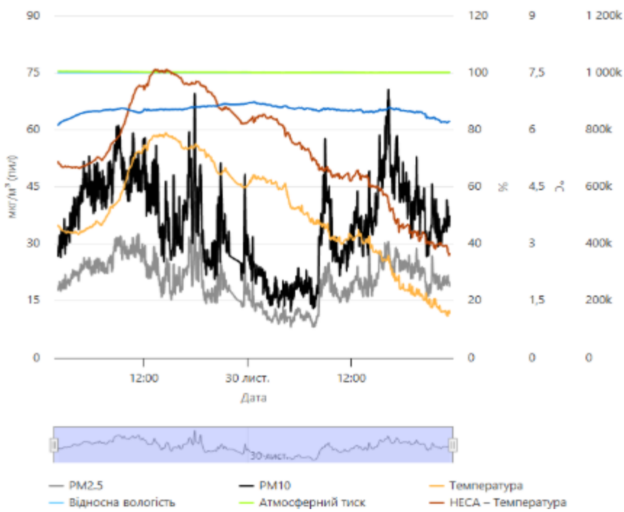

Fig. 3. Air quality data at the address Yuriy Popravky Street, 4 in Kyiv (for the period from 28.11.2025 to 30.11.2025)

The presented figure exemplifies these capabilities by showing the temporal evolution of $PM_{2.5}$ and PM_{10} concentrations alongside temperature, atmospheric pressure, and relative humidity. The chart reveals pronounced short-term fluctuations in particulate matter, with several peaks approaching or exceeding recommended safety thresholds, while meteorological variables change more smoothly. Notably, increases in particulate concentrations coincide with periods of relatively stable pressure and decreasing temperature, suggesting limited dispersion conditions that may exacerbate pollution accumulation.

Real-time analytics in SaveEcoBot facilitate the identification of hazardous episodes, support the interpretation of pollution–weather interactions, and enable early warning of potentially adverse air quality events.

2. Real-time radiation monitoring integrated with data processing and analysis of solar activity and solar flares

The research of radiation emissions in real time, conducted in parallel with the processing and analysis of solar activity and solar flare data, is of critical scientific and practical importance. Solar activity is a primary driver of variations in cosmic radiation, including high-energy particles and electromagnetic emissions that directly affect the near-Earth space environment. Sudden increases in radiation levels caused by solar flares and coronal mass ejections can pose significant risks to space-borne and ground-based technological systems, as well as to human health in aviation and space missions. Therefore, integrating real-time radiation monitoring with continuous solar data analysis enables timely detection of hazardous events and supports rapid decision-making processes.

From a methodological perspective, the parallel analysis of radiation measurements and solar activity parameters allows for the identification of causal relationships and temporal correlations between solar phenomena and radiation dynamics. Real-time data streams from radiation sensors, when synchronized with observations of solar flares, sunspot activity, and solar flux indices, provide a comprehensive framework for modeling radiation variability.

Such an approach improves the accuracy of predictive models and enhances the understanding of energy transfer mechanisms from the Sun to the Earth's magnetosphere and atmosphere.

The real-time integration of radiation monitoring and solar data analytics is essential for the development of advanced early-warning systems. These systems can support space weather forecasting by detecting precursors of extreme solar events and estimating their potential radiological impact. The ability to process and analyze large volumes of heterogeneous data in parallel is particularly important in the context of modern observational infrastructures, which rely on distributed sensors, satellite platforms, and high-performance computing environments.

The results of such integrated studies have long-term significance for the design of resilient technological systems and radiation protection strategies. By improving situational awareness of radiation conditions influenced by solar activity, this research contributes to safer satellite operations, more reliable communication and navigation systems, and enhanced protection of critical infrastructure. Consequently, real-time radiation research combined with intelligent analysis of solar activity represents a key interdisciplinary direction at the intersection of space physics, data science, and applied engineering.

Research of streaming data on solar radiation and analysis of indicators from open platforms using software methods was conducted. The basis of the air quality monitoring system is a block of three metal-oxide semiconductor sensors of the MQ series, located on the right side of the board. This block includes sensors MQ-3, MQ-7 and MQ-9, each of which has a similar structure: a ceramic tube with a built-in heating element, covered with a sensitive layer of tin dioxide. The principle of their operation is based on a change in the electrical conductivity of this layer.

These sensors are mounted on expansion boards and enclosed in cylindrical stainless-steel mesh housings, which serve both as particulate filters and as explosion-protection elements due to the presence of internal heating components.

Each sensor contains a micro-ceramic tube coated with a tin dioxide (SnO_2) sensitive layer and an embedded nickel–chromium heating coil.

During operation, the heating element raises the temperature of the SnO_2 layer to activate surface chemical reactions. In clean air, oxygen molecules adsorb onto the sensor surface and capture free electrons, resulting in high electrical resistance. When target gases are present, such as ethanol vapors (MQ-3), carbon monoxide (MQ-7), or carbon monoxide and combustible gases like methane and LPG (MQ-9), they react with the adsorbed oxygen. This reaction releases trapped electrons back into the conduction band, causing a decrease in sensor resistance. The sensor is configured as part of a voltage divider, and the resulting voltage change is sampled by the ADC. An increase in gas concentration corresponds to an increase in the measured output voltage and, consequently, higher ADC readings. The MQ-7 sensor operates with a cyclic heating profile, alternating between high-temperature cleaning and low-temperature sensing phases to improve selectivity for carbon monoxide.

Radiation measurement using a Geiger–Müller tube

Ionizing radiation detection is performed using a Geiger–Müller (GM) tube, specifically the J-305 glass tube, located in the lower section of the device. The tube consists of a sealed glass envelope filled with a low-pressure inert gas mixture and equipped with a central anode wire and a conductive cathode layer. A high voltage, typically in the range of 380–400 V, is applied between the electrodes, establishing a strong electric field without sustaining a continuous current.

When ionizing radiation (primarily beta or gamma particles) enters the tube, it ionizes gas atoms, producing free electrons. These electrons are accelerated toward the anode, initiating a Townsend avalanche that results in a short current pulse. Each pulse corresponds to a single detected radiation event.

Ambient temperature and relative humidity are measured using a digital capacitive sensor module mounted on the SD card adapter. The humidity-sensing element is based on a polymer dielectric whose capacitance varies with absorbed water vapor. As relative humidity increases, changes in dielectric permittivity are converted into precise digital humidity readings by the sensor's internal circuitry.

Temperature measurement is typically realized using an integrated silicon-based temperature sensor or thermistor. Unlike the analog MQ sensors, this module communicates with the microcontroller via a digital interface, ensuring stable, noise-resistant data transmission and eliminating the need for ADC conversion.

In clean air, the resistance of the sensor is high, but when heated and in contact with the target gas, a chemical oxidation reaction occurs on the surface. This leads to a decrease in electrical resistance in proportion to the gas concentration. Each sensor is doped for specific impurities: MQ-3 reacts mainly to alcohol (ethanol) vapors, MQ-7 is specialized for detecting carbon monoxide (CO), and MQ-9 is a combined sensor for carbon monoxide and combustible hydrocarbons such as methane and propane-butane. The signal from the gas sensors is analog and is fed to the analog-to-digital converter (ADC) of the microcontroller.

A 10-bit bit depth with a threshold voltage of 3.3 V is used, which allows converting the sensor output voltage into a digital range of values from 0 to 1023.

The higher the numerical value obtained, the higher the concentration of the detected gas in the air (Fig. 4).

A digital temperature and humidity sensor mounted on an adapter is used to measure the microclimate. Humidity measurement is carried out by the capacitive method: the sensor contains a capacitor with a polymer dielectric capable of absorbing water vapor from the environment.

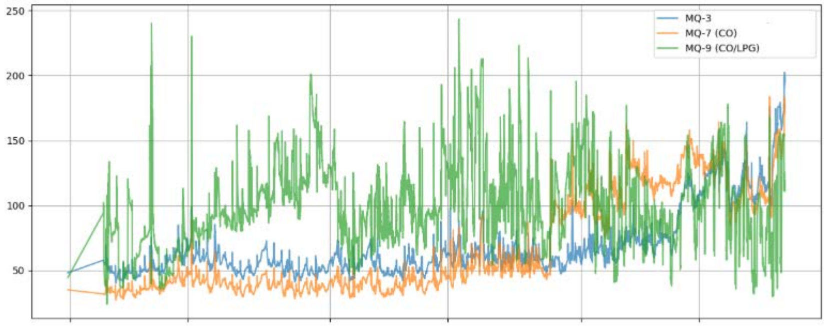

Fig. 4. Gas concentration dynamics

The change in humidity affects the dielectric constant of the polymer and changes the capacitor capacitance, which is then converted into percentage values. The temperature is recorded by a built-in semiconductor element, and all data is transmitted to the controller via a noise-resistant digital interface.

Solar radiation from the height of the sun is shown in the Fig. 5.

Python methods were used to find the regression dependence:

```
df_resampled = df.resample('10T').mean().dropna()
def calculate_sun_altitude(row):
    date_with_tz = row.name.replace(tzinfo=datetime.timezone.utc)
    return get_altitude(row['lat'], row['lon'], date_with_tz)
df_resampled['sun_altitude'] = df_resampled.apply(calculate_sun_altitude, axis=1)
df_daylight = df_resampled[df_resampled['sun_altitude'] > 0].copy()
X = df_daylight['sun_altitude'].values.reshape(-1, 1)
y = df_daylight['radiation'].values
reg = LinearRegression().fit(X, y)
y_pred = reg.predict(X)
r2_score = reg.score(X, y)
plt.figure(figsize=(10, 6))
plt.scatter(df_daylight['sun_altitude'], df_daylight['radiation'], alpha=0.3, s=10)
plt.plot(df_daylight['sun_altitude'], y_pred, color='red', linewidth=2,
start_date = df_resampled.index.min()
end_date = start_date + datetime.timedelta(days=7)
week_df = df_resampled[(df_resampled.index >= start_date) & (df_resampled.index < end_date)].copy()
print(f"\n Analysis of the week{start_date.date()} по {end_date.date()}")
90
```

Sample time interval analysis was from 2023-04-30 to 2023-05-07 (Fig. 6).

Python methods were used to find the regression relationship:

```
week_df['hour_of_day'] = week_df.index.hour + week_df.index.minute / 60
```

```
X_time = week_df['hour_of_day'].values.reshape(-1, 1)
y_rad = week_df['radiation'].values
poly = PolynomialFeatures(degree=2)
X_poly = poly.fit_transform(X_time)
poly_reg = LinearRegression().fit(X_poly, y_rad)
X_range = np.linspace(0, 24, 100).reshape(-1, 1)
y_poly_pred = poly_reg.predict(poly.fit_transform(X_range))
```

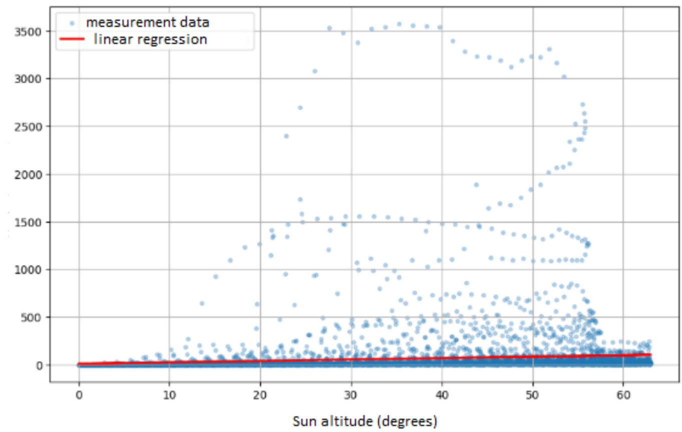

Fig. 5. Dependence of radiation (flares) on the height of the sun

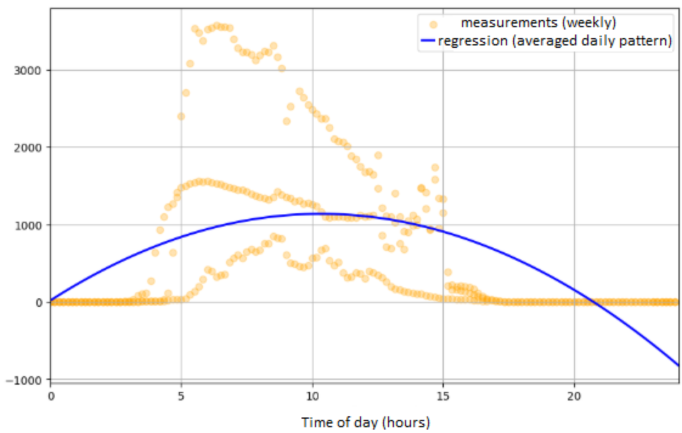

Fig. 6. Radiation dependence on time of day (weekly sample)

The dependence of gas concentrations and radiation on the time of day was studied. Average values by hour:

hour	radiation	MQ3	MQ7	MQ9
0	1.628346	68.391634	67.022900	99.138550
1	1.646679	68.290962	66.838103	98.082266
2	2.050192	68.690332	67.229613	96.802576
3	5.066239	69.274220	67.648767	96.113950
4	27.436708	69.658421	67.827370	95.393173
5	68.353680	70.530418	68.492415	95.075105
6	99.605249	70.509736	68.121539	94.486783
7	106.204988	69.695175	67.984081	94.128687
8	109.893291	68.687844	68.203765	94.863063
9	91.192113	68.025165	67.803630	96.321176
10	94.261812	67.231018	66.583220	97.258566
11	87.988099	67.591331	67.197946	103.845882
12	92.266487	67.168665	66.660335	106.132874
13	76.979410	66.839487	66.084558	109.827603
14	61.979529	66.795793	65.927512	108.319749
15	23.289983	66.616441	65.780667	106.127355
16	7.119139	67.138495	66.505668	105.753083
17	2.238965	67.339127	66.653917	103.355099
18	1.639621	67.516913	66.607686	98.559791
19	1.629242	67.574867	66.895444	96.233144
20	1.619183	67.803843	67.263903	95.050285
21	1.628695	68.690631	68.443665	96.962783
22	1.630391	68.674213	67.914109	97.275175
23	1.623799	68.602496	67.297990	100.783763

The observed diurnal pattern of average radiation is primarily driven by the daily cycle of solar elevation and the associated modulation of solar and secondary cosmic radiation reaching the Earth's surface. Radiation levels increase rapidly after sunrise as the Sun rises higher above the horizon, reducing atmospheric path length and attenuation, which leads to enhanced penetration of high-energy particles and electromagnetic radiation. Peak values are typically observed around local midday, when solar elevation and solar-driven ionization processes are strongest. As the Sun sets, atmospheric absorption increases and the contribution of solar radiation diminishes, resulting in a sharp decline in measured radiation during the evening and nighttime hours, when only background terrestrial and cosmic radiation remains.

According to NOAA, moderate solar activity was observed on May 6, 2023 with several active regions on the Sun (regions 3293, 3294, 3296, 3297, 3299). The largest region, 3297, had an area of 0.460 millionths of a sphere and a complex magnetic configuration (Beta). This activity could have caused an increase in radiation levels, which is confirmed by our data – a record value of 2627 (Fig. 7).

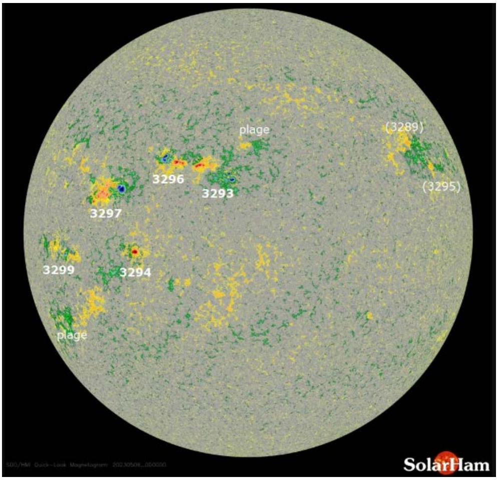

Fig. 7. SDO/HMI Quick-Look Magnetogram (May 6, 2023)[11]

The magnetogram of the Sun is image of the distribution and polarity of the magnetic field in the photosphere. Such images show areas with positive and negative magnetic polarity (active regions, sunspots), but not the flares themselves.

Solar flares are usually observed: in the ultraviolet or X-ray range (SDO/AIA), as bright local flashes of radiation associated with a sharp restructuring of the magnetic fields. HMI magnetograms are used to analyze the prerequisites of flares, since a complex and tense configuration of the magnetic field can indicate an increased probability of their occurrence, but the flares themselves are not directly visible in them[12]:

```
:Issued: 2023 May 06 0030 UTC
# Prepared jointly by the U.S. Dept. of Commerce, NOAA,
# Space Weather Prediction Center and the U.S. Air Force.
#
Joint USAF/NOAA Solar Region Summary
SRS Number 126 Issued at 0030Z on 06 May 2023
Report compiled from data received at SWO on 05 May
I.  Regions with Sunspots.  Locations Valid at 05/2400Z
Nmbr Location  Lo  Area  Z   LL   NN Mag Type
3293 N13E07   150  0110 Dai  10   06 Beta
```

[11] SDO/HMI Quick-Look Magnetogram: 20230506_000000. URL: https://solarham.com/pictures/2023/may6_2023_mag.jpg

[12] National Centers for Environmental Information. URL: https://www.ngdc.noaa.gov/stp/space-weather/swpc-products/daily_reports/solar_region_summaries/2023/05/20230506SRS.txt.

3294 S08E30 127 0120 Hsx 02 01 Alpha
3296 N16E20 137 0280 Dki 05 11 Beta
3297 N08E36 121 0460 Ekc 11 18 Beta
3299 S06E55 102 0050 Dao 05 04 Beta
IA. H-alpha Plages without Spots. Locations Valid at 05/2400Z May
Nmbr Location Lo
3289 N20W55 212
3292 N14W54 211
3295 N16W61 218
3298 S16E18 139
II. Regions Due to Return 06 May to 08 May
Nmbr Lat Lo
3279 S19 048
3282 N12 043

Solar Elevation Function (Required for Day/Night Filtering):

```
def calculate_solar_elevation(row):
    """Calculates solar elevation in degrees."""
    lat = row['Lat']
    lon = row['Lon']
    time = row['Timestamp']
    doy = time.timetuple().tm_yday
    gamma = (2 * np.pi / 365) * (doy - 1 + (time.hour - 12) / 24)
    eqtime = 229.18 * (0.000075 + 0.001868 * np.cos(gamma) - 0.032077 * np.sin(gamma)
             - 0.014615 * np.cos(2 * gamma) - 0.040849 * np.sin(2 * gamma))
    decl = 0.006918 - 0.399912 * np.cos(gamma) + 0.070257 * np.sin(gamma) \
           - 0.006758 * np.cos(2 * gamma) + 0.000907 * np.sin(2 * gamma) \
           - 0.002697 * np.cos(3 * gamma) + 0.00148 * np.sin(3 * gamma)
    time_offset = eqtime + 4 * lon
    tst = time.hour * 60 + time.minute + time.second / 60 + time_offset
    ha_rad = np.radians((tst / 4) - 180)
    lat_rad = np.radians(lat)
    cos_zenith = np.sin(lat_rad) * np.sin(decl) + np.cos(lat_rad) * np.cos(decl) * np.cos(ha_rad)
    return np.degrees(np.arcsin(np.clip(cos_zenith, -1, 1)))
# --- Data Loading and Cleaning ---
column_names = [
    'CallSign', 'Time', 'Date', 'Lat', 'Lon', 'Alt',
    'Radiation_Flashes', 'Temp', 'Humidity',
    'MQ3_ADC', 'MQ7_ADC', 'MQ9_ADC', 'Unused', 'Empty'
]
df = pd.read_csv('LOG.TXT', sep=';', names=column_names, header=None)
df.drop(columns=['CallSign', 'Unused', 'Empty'], inplace=True)
numeric_cols = ['Lat', 'Lon', 'Alt', 'Radiation_Flashes', 'Temp', 'Humidity', 'MQ3_ADC', 'MQ7_ADC', 'MQ9_ADC']
for col in numeric_cols:
    df[col] = pd.to_numeric(df[col], errors='coerce')
df['Timestamp'] = pd.to_datetime(df['Date'] + ' ' + df['Time'], errors='coerce')
df = df[(df['Lat'] != 0) & (df['Lon'] != 0)].dropna(subset=['Timestamp', 'Lat', 'Lon']).copy()
df['Sun_Elevation'] = df.apply(calculate_solar_elevation, axis=1)
```

```
# --- 2. SOLAR STORM DETECTION LOGIC ---
# 1. NIGHTTIME FILTER: Sun must be significantly below the horizon
# (e.g., lower than -10 degrees, which is past astronomical twilight).
NIGHT_ELEVATION_THRESHOLD = -10

# 2. RADIATION THRESHOLD: Must be significantly above the normal nighttime background
# (Normal background max was 3 flashes. A storm should be > 5,
# or maybe 3x the average background). We set a conservative threshold of 10 flashes.
STORM_RADIATION_THRESHOLD = 10
# Apply filters
night_data = df[df['Sun_Elevation'] < NIGHT_ELEVATION_THRESHOLD]
solar_storm_events = night_data[night_data['Radiation_Flashes'] >= STORM_RADIATION_THRESHOLD]
```

The solar storm detection technique successfully filtered out the vast majority of noise caused by daylight leakage, leaving only events that recorded high levels of radiation when the sensor was actually in complete darkness.

Event frequency: the events were detected over a two-month period, indicating several separate high-energy particle events occurring near Kyiv.

Typical peak value: most events show peaks in the range of 10 to 22 flashes (in 5 seconds). This is approximately 3-7 times the normal nighttime maximum, indicating a significant increase in the flux of ionizing particles.

Peak event (exception): the event that occurred on 03.07.2023 at 19:45:24 is distinguished by 54 flashes. This value is almost 5 times higher than the next largest event, and approximately 18 times higher than the normal background.

Potential Solar Storm Events Detected (Radiation > 10 flashes at night):

	Timestamp	Sun_Elevation	Radiation_Flashes
67336	2023-05-08 19:17:34	-14.3566	11
136836	2023-05-12 19:59:23	-17.0644	22
258585	2023-05-19 21:21:49	-19.5333	10
535912	2023-06-04 23:12:14	-15.234	11
570763	2023-06-06 23:41:12	-13.3423	10
897493	2023-06-25 22:12:51	-16.0957	10
1033739	2023-07-03 19:45:24	-10.6102	54

In this research, near–real-time event stream processing is implemented using the Kappa architecture, which relies on a single, immutable event log for both micro-batch and replay-based processing. Unlike the Lambda architecture, which separates batch and streaming layers and requires duplicated business logic, the Kappa approach unifies processing semantics through a common event log. This design reduces development and maintenance costs, simplifies testing procedures, and eliminates inconsistencies between real-time and batch computations.

The immutable nature of the log facilitates the implementation of idempotent processors, enabling safe reprocessing of identical data without side effects – an essential property for fault recovery. The physical decoupling of write and read paths allows independent scaling of ingestion and analytics workloads, while micro-batch support provides elasticity under bursty event arrivals without the operational complexity of heavyweight message brokers. Spatial unification is achieved through the H3 hierarchical hexagonal grid system, which offers near-equidistant neighborhood relationships and stable aggregation operations. Hexagonal geometry more naturally approximates circular influence zones and reduces distortion effects commonly associated with square grids. The hierarchical structure enables seamless transitions across spatial resolutions, as higher-level aggregates are computed by consolidating finer-grained cells, thereby simplifying the construction of multiscale analytical views. Efficient neighborhood operations, such as *k*-rings and distance disks, are particularly important for calculating spatial risk aggregates in quasi–real-time scenarios (Fig. 8).

The use of a fixed grid also simplifies storage indexing: daily and weekly materializations rely on identical keys, avoiding expensive spatial joins – an advantage in applications where proximity and coverage are prioritized over exact geometric precision.

The presented architecture illustrates an integrated data pipeline for real-time ingestion, processing, and analytical exploitation of heterogeneous geospatial and radiation-related data streams. At the ingestion layer, live and near–real-time events are collected from multiple sources, including social and news channels, satellite-based fire detection systems (FIRMS), and radiation monitoring sensors. All incoming data are normalized and persisted in a centralized event store, which functions as an immutable events log and provides a reliable foundation for downstream processing, replay and auditability.

The processing layer is organized as a micro-batch workflow that performs a sequence of domain-specific transformations and analytics.

Key stages include deduplication of fire radiative power (FRP) events within defined temporal windows, baseline estimation and anomaly detection for gamma radiation measurements, and morphologically aware geocoding to ensure robustness across linguistic and semantic variations in location references.

In parallel, cross-signal corroboration is applied to correlate independent data streams, enabling the identification of consistent patterns and the reduction of false positives by validating events across multiple sensing modalities.

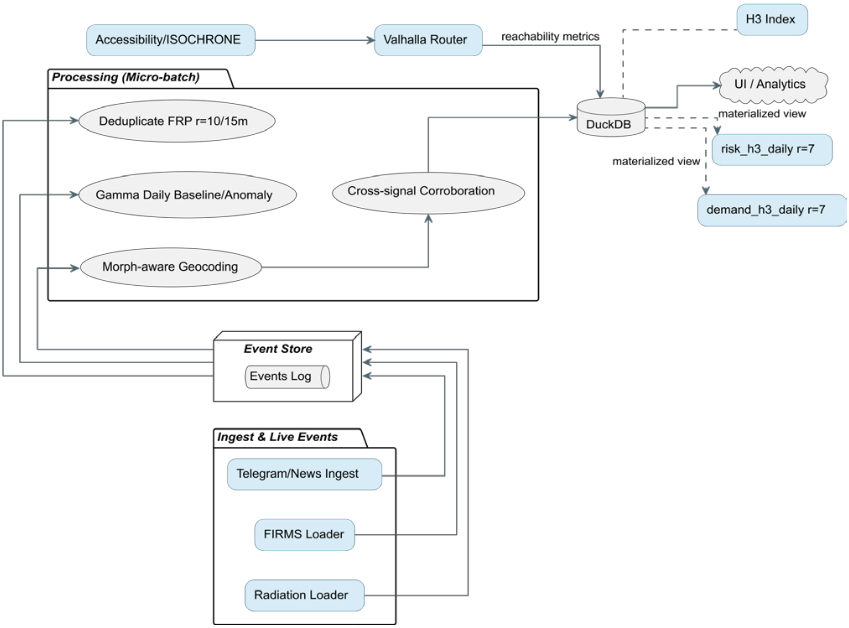

Fig. 8. Diagram of the connections between the elements of the proposed software

Geospatial accessibility and reachability analysis is incorporated through isochrone-based modeling and routing via the Valhalla router, producing reachability metrics that characterize spatial influence and response potential. The processed and enriched data are then materialized in an analytical storage layer based on DuckDB, where spatial indexing H3 supports efficient aggregation and querying.

Materialized views, such as daily risk and demand indicators at H3 resolution, provide a structured interface for user-facing analytics and visualization.

Overall, the architecture emphasizes scalability, temporal coherence, enabling near–real-time situational awareness and decision support in environments influenced by radiation dynamics and geospatial factors.

As a lightweight analytical core, DuckDB is employed as an embedded, column-oriented database optimized for high-performance local OLAP queries. Its vectorized execution model, column compression, and support for modern storage formats enable high query throughput on a single node without deploying distributed clusters. Embedding the database engine directly within the application process simplifies transaction management of local artifacts and enhances reproducibility by co-locating data and code versions. This configuration is particularly well suited for autonomous and local analytics use cases, where inter-node coordination is unnecessary and data volumes do not justify the operational overhead of distributed processing frameworks. For network-based computations of distance, reachability time, and isochrone construction, the research utilizes the Valhalla routing engine, which operates on OpenStreetMap graph data and supports fully offline execution. Isochrone calculations are performed on preprocessed map tiles, producing temporal accessibility metrics that are directly integrated into coverage models and visualized on the same H3 tiles. Although preprocessing introduces an upfront cost, it is offset by the robustness and speed of accessibility estimation in critical situations, which is especially important for autonomous deployments. The preservation of all intermediate artifacts supports auditing and result reproducibility, in line with contemporary principles of open and reproducible data analytics. The integration of these components results in a coherent architectural framework: the immutable event log ensures reliable data lineage; CQRS decouples analytical queries from write operations and accelerates projection generation; H3 indexing resolves geometric heterogeneity; DuckDB provides fast local aggregations without additional servers; and Valhalla enables offline accessibility assessment.

This hybrid approach, combining micro-batch stream processing, H3-based spatial indexing, embedded OLAP analytics, and offline routing – achieves a balanced trade-off between functionality and architectural complexity for environmental threat monitoring systems. Ultimately, the overall architectural risk profile is reduced: failure of any individual module does not compromise historical data, as all events are preserved in the immutable log and analytical views can be reconstructed through replay. Infrastructure requirements are minimized, node autonomy is increased, and dependency on external services is reduced, which is particularly critical under crisis and emergency conditions.

CONCLUSIONS

The research directions indicate a clear transition from traditional reactive monitoring systems toward intelligent, autonomous software platforms capable of continuous and proactive environmental threat assessment. Such systems represent a critical step toward more resilient and sustainable environmental management in the face of increasing environmental complexity and uncertainty.

The software methods identified periods of increased radiation that coincided with solar activity data from official NOAA sources.

Proposed software monitoring system demonstrated its effectiveness in detecting solar storms and changes in the radiation background, and also showed the possibility of using inexpensive sensors for environmental monitoring.

This research presents a coherent software-based solution for environmental threat assessment that integrates near–real-time stream processing, spatial unification, and efficient local analytics within a compact and resilient architectural framework. By adopting the Kappa architecture with an immutable event log, the proposed approach ensures consistent processing semantics for both real-time and historical data, eliminating the duplication and synchronization issues inherent to Lambda-based solutions. This design significantly improves system reliability, simplifies maintenance, and enhances fault tolerance through deterministic replay and idempotent processing. Compared to existing environmental monitoring platforms that rely on heavyweight distributed infrastructures or tightly coupled processing pipelines, the proposed solution demonstrates clear advantages in terms of architectural simplicity and operational efficiency.

The use of hierarchical H3 spatial indexing provides stable, multi-resolution spatial aggregation and efficient neighborhood analysis, overcoming the limitations of irregular geometries and expensive spatial joins commonly found in traditional GIS-centered systems. This enables scalable, real-time spatial risk assessment while maintaining computational predictability.

The integration of DuckDB as an embedded analytical engine further distinguishes the proposed approach from conventional server-based or cluster-dependent analytics solutions. High-performance local OLAP queries can be executed without the overhead of distributed coordination, improving reproducibility and enabling autonomous operation.

In parallel, the incorporation of the Valhalla routing engine allows for offline computation of accessibility and isochrone-based metrics, a capability rarely supported by existing systems that depend on continuous network connectivity.

The proposed software solution offers a balanced combination of scalability, resilience, and analytical expressiveness, while minimizing infrastructure requirements. Its ability to operate autonomously, recover from failures without data loss, and support quasi–real-time spatial analytics makes it particularly well suited for environmental threat monitoring under resource constraints and crisis conditions. These advantages position the proposed approach as a robust and adaptable alternative to existing environmental monitoring and decision-support systems.

SUMMARY

Software methods for environmental threat assessment represent a critical foundation for modern environmental monitoring systems.

The transition to streaming data analysis has fundamentally changed the requirements for software architectures and analytical methods, emphasizing real-time processing, adaptability, and scalability. Systematic classification and comparative evaluation of existing methods demonstrate that hybrid, ML–driven approaches integrated into stream processing architectures offer the most promising solutions for real-time environmental threat assessment.

Software-driven approaches can dynamically adapt to changing environmental conditions by updating analytical models and decision rules as new data become available, thereby maintaining relevance in non-stationary environments. Software-based environmental threat assessment systems support proactive and predictive decision-making. By incorporating ML and predictive analytics, these systems can forecast potential threat scenarios and estimate associated risks before critical thresholds are exceeded. Such predictive capabilities are particularly valuable for risk management, early warning systems, and strategic planning in domains related to public health, environmental protection, and critical infrastructure resilience. Despite these advantages, software-based approaches also exhibit several limitations that must be carefully considered.

A fundamental challenge is their strong dependence on data quality and sensor reliability. Inaccurate measurements, missing data, or sensor failures can significantly degrade the performance of analytical models and lead to false alarms or missed detections. This dependency necessitates robust data validation, preprocessing, and fault-tolerance mechanisms within the software architecture. Another limitation is the computational complexity associated with real-time machine learning models, especially deep learning techniques. Processing high-velocity data streams under strict latency constraints requires substantial computational resources and efficient model optimization strategies.

The proposed software solution provides an efficient and resilient framework for environmental threat assessment by combining unified stream processing, hierarchical spatial indexing, and lightweight local analytics. Unlike existing software approaches that rely on complex distributed architectures and duplicated processing logic, the solution ensures consistency, fault tolerance, and reproducibility through an immutable event log and idempotent processing. Its ability to perform near–real-time spatial analytics, operate autonomously with minimal infrastructure, and recover analytical views through replay offers clear advantages in scalability, reliability, and practical applicability, particularly in resource-constrained and crisis scenarios.

Bibliography

1. Akyildiz I.F., Melodia T., Chowdhury K.R. A survey on wireless multimedia sensor networks. *Computer Networks*, 2007, vol. 51, no. 4, pp. 921–960. https://doi.org/10.1016/j.comnet.2006.10.002.

2. Ramakrishnan N., Butler P., Muthiah S., Self N., Khandpur R.P., Saraf P., et al. Beating the News with EMBERS: Forecasting civil unrest using open source indicators. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014, pp. 1799–1808. DOI: 10.1145/2623330.2623373.

3. Zanella A., Bui N., Castellani A., Vangelista L., Zorzi M. Internet of Things for smart cities. *IEEE Internet of Things Journal*, 2014, vol. 1, no. 1, pp. 22–32. DOI: 10.1109/JIOT.2014.2306328.

4. Mujtaba G., Shahzad S.J.H. Air pollutants, economic growth and public health: implications for sustainable development in OECD countries. *Environmental Science and Pollution Research*, 2021, vol. 28, pp. 12686–12698. DOI: 10.1007/s11356-020-11212-1.

5. Keswani A., Akselrod H., Anenberg S. Health and clinical impacts of air pollution and linkages with climate change. *NEJM Evidence*, 2022, vol. 1, no. 7, article EVIDra2200068. DOI: 10.1056/EVIDra2200068.

6. Bibri S.E., Krogstie J. Smart sustainable cities of the future: An extensive interdisciplinary literature review. *Sustainable Cities and Society*, 2023, vol. 31, pp. 183–212. DOI: 10.1016/j.scs.2017.02.016.

7. Yang Y., Wu Q., He B., Peng H., Yang R., Hao Z., Liao Y. SeBot: Structural entropy guided multi-view contrastive learning for social bot detection. *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 3841–3852. DOI: 10.1145/3637528.3671871.

8. Sun T., Liu C., Chen L., Qian Z., Li P., Zhu Q. A unified framework for multi-modal rumour detection via multi-level dynamic interaction with evolving stances. *Information Processing & Management*, 2025, vol. 62, no. 3, article 104066. DOI: 10.1016/j.ipm.2025.104066.

9. Rajesh M., Babu R.G., Moorthy U., Easwaramoorthy S.V. Machine learning-driven framework for realtime air quality assessment and predictive environmental health risk mapping. *Sci Rep*. 2025, 15(1):28801. DOI: 10.1038/s41598-025-14214-6.

10. SaveEcoBot. The first environmental system in Ukraine. URL: https://www.saveecobot.com/en.

11. SDO/HMI Quick-Look Magnetogram: 20230506_000000. URL: https://solarham.com/pictures/2023/may6_2023_mag.jpg.

12. National Centers for Environmental Information. URL: https://www.ngdc.noaa.gov/stp/space-weather/swpc-products/daily_reports/solar_region_summaries/2023/05/20230506SRS.txt.

Information about the author:
Oleshchenko Liubov Mykhailivna,
Candidate of Technical Sciences,
Associate Professor at the Computer Systems Software Department,
National Technical University of Ukraine
“Igor Sikorsky Kyiv Polytechnic Institute”,
37, Beresteiskyi ave., Kyiv, 03100, Ukraine