Journal of Environmental & Earth Sciences
https://journals.bilpubgroup.com/index.php/jees

REVIEW

Computing the Planet: Integrating Machine Learning, Remote Sensing, and Sensor Data Fusion for Environmental Insights

Kai Mao

College of Artificial Intelligence, Chongqing Technology and Business University, Chongqing 400067, China

ABSTRACT

Indeed, a range of systems in the environment requires timely, spatially explicit, and credible information to support its environmental decision-making, but no one observing system can give the complete and reliable measures of the Earth system across scales. This review summarizes how the realization of the Compute the Planet is underway in the form of machine learning, remote sensing, and sensor data fusion to generate decision-ready environmental insights. We use the application-first approach, which considers remote sensing, in situ and Internet of Things (IoT) sensing, and physics-based models as complementary streams of evidence with similar strengths and failures. We look critically at how an integrated system can convert heterogeneous observations to action products across three high impact application areas: atmosphere and air quality, water–land–ecosystem dynamics, and hazards. Rapid-response situational awareness, ecosystem condition metrics, drought and flood indicators, exposure maps, and hazard/extreme indicators are key products. The integrated systems to environment interface in three high impact application areas: atmosphere and air quality, water-land-ecosystem dynamics, and hazard Examine Our operational requirements can often determine real-life value such as latency, time stability, smooth degradation in the presence of missing or degraded inputs, and calibrated uncertainty usable in threshold-based decisions. These pitfalls are common across fields: mismatch in the scale between a point sensor and a gridded product, objectives on proxies in remotely sensed measurements, domain shift in the extremes and changing baselines, and evaluation aspects, which overestimate generalization because of spatiotemporal autocorrelation. Based on these lessons, we present cross-domain proposals for strong validation, uncertainty quantification, provenance, and versioning, as well as

*CORRESPONDING AUTHOR:
Kai Mao, College of Artificial Intelligence, Chongqing Technology and Business University, Chongqing 400067, China; Email: maokai@ctbu.edu.cn

ARTICLE INFO
Received: 12 December 2025 | Revised: 1 January 2026 | Accepted: 4 January 2026 | Published Online: 28 January 2026
DOI: https://doi.org/10.30564/jees.v8i1.12946

CITATION
Mao, K., 2026. Computing the Planet: Integrating Machine Learning, Remote Sensing, and Sensor Data Fusion for Environmental Insights. Journal of Environmental & Earth Sciences. 8(1): 277–297. DOI: https://doi.org/10.30564/jees.v8i1.12946

fair performance evaluation. We conclude that the next era of environmental intelligence will see a reduction in average accuracy improvement and an increase in terms of robustness, transparency, and operational responsibility, thus allowing the integrated environmental intelligence system to be deployed, which may be relied on to monitor human health, resource allocation, and survival in a more climate-adapted world.

Keywords: Machine Learning; Remote Sensing; Sensor Data Fusion; Environmental Monitoring; Uncertainty Quantification

1. Introduction

Environmental change has ceased to be a far-off indicator occurring only on long-term averages, and it is becoming more and more a matter of personal experience as compounding shocks, heatwaves which lengthen wildfire seasons, floods which follow drought-hardened soils, and episodes of air pollution that are more and more a product of local emissions along with transboundary transport[1]. The control of these risks requires some environmental information that is comprehensive yet detailed in the locality and timely in making decisions. But the system of the Earth is not easy to observe. Most of the most important climate adaptation and sustainable management variables, such as particulate exposure, root-zone soil water, or feel of ecosystem stress, inundation under cloud cover, or heat risk of the neighborhoods, cannot be measured in all places and at all times directly in just one instrument[2]. Instead, they are created out of a mosaic of incomplete measurements, one that is restricted by physics and sampling, cost, and operational realities.

And this is where Computing the Planet fits: the notion that insights into the environment are becoming more and more a product of the combination of heterogeneous observations and models of machine learning and data fusion, and no longer of any particular sensing technology alone[3]. In the past ten years, there have been three trends that have converged. To begin with, there has been an increase in Earth observation (EO) activities in the public and commercial sectors, which has resulted in growing catalogs of multispectral images, radar, thermal, atmospheric trace gas recoveries, and near-real-time products with decreasing revisit times. Second, in situ sensing has also come to be more diverse in its use, and is not limited to traditional regulatory stations and research networks, encompassing low-cost air quality sensors, connected weather stations, hydrometric gauges, soil probes, buoys, flux towers, and citizen science measurements. Third, machine learning has come out to be more of a toolkit of predictive tools to a more comprehensive representation learning, multimodal integration, spatiotemporal prediction, uncertainty estimation, and scaled deployment. These developments combined have offered a viable opportunity, namely to study the planet as a unified computational problem where remote sensing offers coverage, ground and IoT sensors offer local fidelity, and models offer physical coherence and counterfactual structure, where ML fills in missing gaps, corrects biases, and converts streams of observations into decision-ready products[4]. **Figure 1** summarizes the planetary viewpoint of the computing framework embraced in this review, where the evidence that senses each other are combined to generate decision-ready environmental products, clearly exposing uncertainty and provenance.

Nevertheless, the potential of built-in environmental intelligence has not yet necessarily been converted into a credible vision. The information streams possess unique strong and weak points. Vast areas can be imaged with satellite imagery consistently, but it can be obscured by clouds, limited by how frequently the satellite revisits an area, and be misconstrued by complicated conditions of the surface and retrieval assumptions[5]. Radar is able to penetrate through the clouds, and it may not be easy to interpret through vegetation and different soil conditions. In situ networks quantify quantities frequently inaccessible to satellites, near-surface concentrations, vertical structure, or subsurface environments, but which are selected sparsely, unevenly, and often preferentially in those locations accessible or rich. Cheap sensors are enhanced, but they come at the cost of drift, calibration, and variation of data quality. Physical models and reanalysis products provide fields all over the world with physical consistency, totally over time; but like parameterizations, coarse resolution, and imperfect initiation, uncertainties in physical models have systematic and decision-forcing errors at the extremes. The main difficulty is not just the larger data scale, but the ability to integrate data with alternative error formats, solutions, and latencies, and representativeness in a unified estimate of the environmental state and hazard[6].

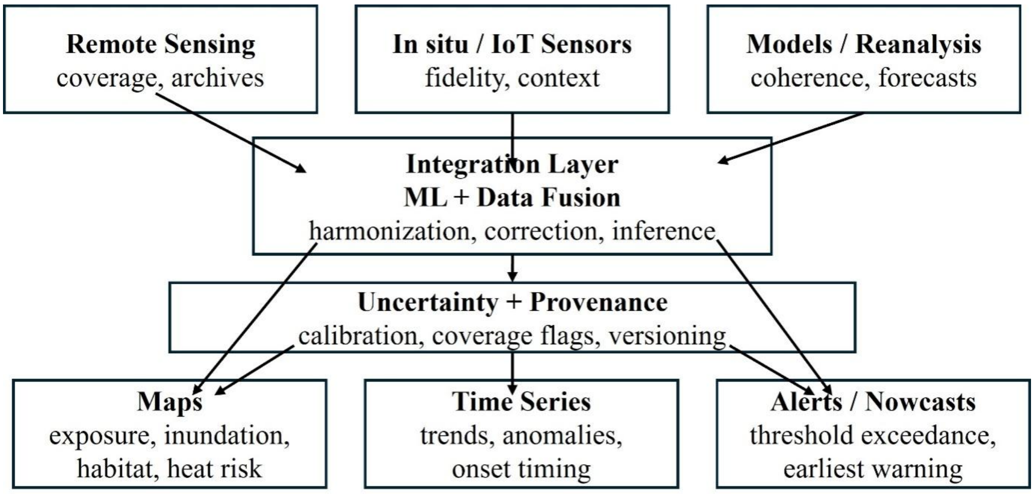

Figure 1. Computing the planet's conceptual framework.

The necessity of integration is the sharpest in the spheres where precision and speed are required in decisions. Neighborhood-level air quality estimates with uncertainty limits are needed by the agencies of health, especially in periods of wildfire smoke, when exposure may be characterized by a sharp distance or time constraint. The water managers should be in a position to get a warning of the beginning of drought, snowpack deviations, and evapotranspiration trends, preferably at a basin-wide scale with locally implementable detail. Under operational constraints (lost data, broken communications, rarely seen events), emergency responders and utilities require quick mapping of flood extent, or burn severity and heat stress conditions that should not be modeled based on a theoretically based average-case. Conservation professionals require sound tools to monitor changes in habitat state and seasonal changes, as well as observe subtle location information and administrative limitations. In all these applications, environmental acumen is now evaluated not ontologically, in terms of scientific accuracy, after all, but in terms of decision-making aptitude: latency, reliability, interpretability, endurance over time, and transparent uncertainty[7].

This review takes an application-first approach. You can start with the list of questions that practitioners can ask and the limits that they have to achieve, rather than compiling an exhaustive list of algorithms. The series of practical steps tends to be the opposite of a methodology-oriented account: variables of interest are given by decisions; variables are given by what can be observed; we see what can be observed but not gaps, thus combining strategies that can be used to fill in the gaps. It is because of this reason that the center of the article is structured around areas in which integrated sensing has taken a center stage: (i) atmosphere and air quality, (ii) water, land, and ecosystems, and (iii) hazards, extremes, and rapid response. In each field, we integrate the way remote sensing, in situ sensing, and model-based products are most commonly integrated, what decision-ready products can be anticipated, and where integration fails most frequently. We stress those patterns that are present across domains, coverage-versus-fidelity trade-offs, scale mismatch, the persistence of domain shift, and the difficulty of validating products in the case of incomplete or biased ground truth.

There are two clarifications to make. To begin with, by integration we mean a wide set of practices, such as bias correction, gap filling, downscaling, calibration transfer, cross-sensor harmonization, and joint inference, but we do not consider any single technical method to be the best. Integration may be as straightforward as making low-cost sensors waveform resonant to some sound regulatory network, or as narrow as integrating radar, optical imagery, and states of hydrologic models into the estimation of inundation under cloud cover[8]. The important thing is that the integrated product is more helpful than its components to a properly defined decision task, and the limitation of the product is clear. Second, the current review is not attempting to give a separate methodology tutorial. Already, the literature covering ML architectures, data assimilation variants, or fusion algorithms is quite extensive and fast developing. Rather, we tend to concentrate on the application logic that is used to decide on the integration choices that are defensible, the success measurement, and what constitutes good practice when products are operated beyond the training region, season, or sensor environment.

One of the common barriers in the literature is that the enhancements of the literature are not necessarily applicable to operational reliability. Environmental data are highly autocorrelated both spatially and over time, which can bloat performance when performance splits into train and test samples that contain correlated data[9]. Labels are either infrequent, noisy, or indirect (e.g., the state of air quality based on monitors that are unevenly distributed). Shaun's events take up risk but are not represented in the training information. Idly motion sensors at the sensor stage. It is in perspective that sensors do not stay the same: new satellites come out, new algorithms are used to retrieve or store things, ground networks get larger or older. This is called a distribution shift, silently decaying the model performance. Moreover, integrated products need to be frequently calibrated (i.e., uncertainty estimates should correspond to observed error rates), robust (i.e., behave sensibly in the presence of missing or degraded inputs), and auditable (i.e., stakeholders should be able to know what a product is, and what it is not). These necessities drive the discipline in the direction of a non-optimism of the highest average result, more towards an operational mental state, that is, a performance in extremes, stability in geographies, and openness in matters of uncertainty and failure states[10].

The question that is not necessarily technical is also brought into play by the phrase Computing the Planet. Scalability Unified data streams put ethics and governance on the frontline. Inequality is reflected in sensor networks; lack of coverage in some areas may prejudice environmental intelligence systems. Certain environmental data are sensitive, such as the whereabouts of endangered species, native territories, or important infrastructure vulnerability, and this type of information must be well-managed and restricted by access control[11]. Overconfident outputs may cause harm to the decision systems, particularly when they are applied to resource allocation, enforcement, or responding to an emergency. With more competent integrated systems, responsible practice should not just be stronger models, nor should responsible practice stay silent: provenance of data, model versioning, description of intended use, and a way to challenge or upgrade products in encountering different conditions.

Although there is a significant literature review of machine learning techniques in remote sensing, sensor fusion models, and digital Earth or digital twin systems, these surveys are subject to organization based on algorithms, model architectures, or computational processes. Consequently, they are more likely to stress methodological innovation and relatively less importance on the operational settings that integrated environmental intelligence systems eventually have to operate[12]. This review uses a complementary approach and uses an application-first and failure-mode-conscious approach to frame machine learning and data fusion. Instead of listing algorithms, the discussion is organized by the decision questions in the key areas of the environment, such as air quality, water and ecosystem monitoring, and hazard response, and the real-world limitations that are associated with integrating heterogeneous observations into operational products. Special consideration is made of questions like spatial and time scale incompatibilities, expression of uncertainty, performance with missing or corrupting observations, and graceful performance during extraordinary events. This review is intended to address the gap between methodological research and the use of operational environmental intelligence by concentrating on the design, validation, and utilization of the integrated systems in real-world monitoring and decision work processes. Therefore, the target audience will also include environmental product developers, practitioners with operational agencies, and managers or funding organizations who need a systematic outlook of how to construct trustworthy, decision-ready systems and not merely more accurate predictor models[13–15].

These motivations make this review add a systematic synthesis with targeted researchers and practitioners who construct or consume integrated environmental products. Section 2 elaborates on the creation of a single perspective of streams of evidence of the environment and the contribution each makes to decision-making. Parts 3–5 discuss 3 high-impact application areas: air quality, water/land/ecosystems, and hazards/extremes—including integration patterns, operation requirements, and failure modes in each. Section 6 closes it by relating by giving cross-domain values and also giving an overarching agenda on reliability, uncertainty, and real-time integrated sensing. The common denominator is straightforward, though: environmental understanding more and more relies on the extent of our combination of perfect observations, rather than on the capacities of a particular sensor[16,17]. It is the next generation of systems—more and more data-rich, coherent, transparent, and decision-ready—that will compute the planet in ways that can be relied upon when it matters the most.

2. A Unified View of Environmental Evidence

Environmental understanding is seldom the result of one instrument or information. It can better be defined as an inferential product built up out of a variety of imperfect perspectives on the Earth system that vary in their measurements, their sampling of space and time, and the way their errors work[18]. Practically, modern environmental intelligence is constructed on three interacting streams of evidence: one remote sensing, one in situ sensing, and one physics-based models (with reanalysis and operational predictions). Both streams offer a specific type of information, including coverage, fidelity, and physical coherence, respectively, and both streams present typical blind spots. A single perspective does not consider these streams as rival sources of the so-called truth, but offers them as complementary restrictions on the latent environmental states and processes. The present framing explains why integration is both a necessity and a basis: the process through which observational incompleteness, scale incompatibility, and uncertainty are converted into estimates useful in connection to a decision.

2.1. Remote Sensing as a Backbone for Coverage and Consistency

Remote sensing provides a wide, recurring, and standardized observation layer for planetary monitoring[19]. The systematic spatial coverage of satellite and airborne sensors has political boundaries and decades-long coverage, allowing comparative analysis, trend detection, and consistent mapping at scales that the in situ networks could not possibly match. In most environmental variables, remote sensing is the only viable way to monitor patterns of large areas, e.g., land cover change, surface temperature gradients, vegetation phenology, snow extent, coastal turbidity, or atmospheric column properties. It is also important with respect to its temporal continuity: lengthy archives will allow building of baselines and anomaly detection, which are necessary to detect the departure of the normal conditions in a changing climate[20].

Nevertheless, the power of remote sensing cannot be separated from its limitations. Numerous satellite products are not direct measurements of the variable of interest but retrievals based on the radiances or backscatter signals. Assumptions can be stored in the retrieval process concerning surface properties, atmospheric conditions, or sensor geometry and can collapse under the very conditions most decision-relevant, e.g., heavy smoke, complicated urban surfaces, high aerosol loads, heterogeneous vegetation canopies, or snow-forest mixtures. Sampling and occlusion are also limitations to remote sensing. The optical sensors are constrained by cloud cover and brightness, the frequency of revisit might not be adequate to rapidly changing hazards, and even when the frequency of the data is high, the spatial resolution might be too coarse to make decisions at the neighborhood scale. The value of remote sensing has not been diminished by these limitations; instead, the limitations form the guidelines within which remote sensing will be used together with other streams of evidence to generate reliable information that can be used locally to take action[21].

2.2. In Situ and IoT Sensing as a Source of Fidelity and Process Context

In situ observations give the most immediate measurements of the conditions of the environment in which they are located; they are often highly time resolved, and their measurement definitions closely conform to regulatory or scientific standards[22]. Local truth anchors can be obtained with ground-based air quality observations, weather stations, hydrometric, ocean buoys, eddy covariance towers, soil probes, and those in the ecological field to validate remotely sensed products, offer additional calibration, and support context. They are also required where only proxies are offered by satellites. Pollutant concentrations in the near-surface, subsurface moisture movements, precipitation response, streamflow, and local microclimate effects are hard to infer; only space and in situ networks provide the required reference.

Recent developments in low-cost and connected sensors have increased the density of the observations and decreased response time, pushing environmental monitoring toward more concentrated and community-scale applications. But there is a price, as this growth brings new complications. Compromised comparability of time and space. Low-cost devices may drift, have cross-sensitivity, or vary between devices, without applying extreme calibration and quality control. Hybrid networks are costly and may have uneven distributions, even those of good quality, especially when available in situ. Consequently, in situ data can prove to be not representative of the larger landscape, and over-sample some land use and socioeconomic environments and under-sample others. It poses a minor but significant limitation: in situ data are invaluable in supporting and correcting larger-scale products, but they should be viewed as samples with their own biases, rather than being a homogenously distributed truth [23].

In situ sensing provides process context in addition to fidelity to measurement. A lot of in situ observations have ancillary metadata, local knowledge, and site characterization, which can describe why a signal changes and whether it is significant or not [24]. This context is of special significance to the process of separating sensor artifacts and vulnerability to environmental processes and the perception of localized drivers like topography, land management, or built environment impacts. In situ networks are thus used as reference points for measurements as well as interpretive scaffolding for translating integrated data streams into plausible environmental discourse.

2.3. Models and Reanalysis as Providers of Physical Coherence and Counterfactual Structure

Physics-based models and reanalysis products take a different place: they provide estimates with global completeness, physically consistent representations of processes that bring together variables using conservation laws of processes. Reanalyses combine observations and numerical models to generate temporally continuous meteorological fields and associated variables, and operational forecasts extend these fields into the near future [25]. Hydrologic, land-surface, ocean, and ecosystem models also give coherent descriptions of states that can be sparsely monitored, e.g., soil moisture distributions, evapotranspiration quantities, snowpack dynamics, or ocean circulation quantities [26].

Coherence is the most important input of models. They implement interactions of variables—between precipitation, temperature, radiation, and winds to evapotranspiration, runoff, transport by the atmosphere, such that approximations are not just interpolations but dynamically good states. Such coherence makes it possible to achieve capabilities inaccessible to pure observation, namely forecasting, exploration of scenarios, and counterfactual reasoning. Practically, much of the environmental decision process uses a model format, either consciously or unconsciously, since the planning process has to demand projections of conditions that do not yet exist or quantities that cannot be directly measured [26].

Simultaneously, there are structural uncertainties in models. The parameterizations can be distortions, spatial resolution may not be bridged to local decisions, and their initiation errors can chatter. Creating a model can ideally smooth out extremes, can give an incorrect representation of what happens locally, or may have systematic biases across the regions and seasons [27]. They are not failures but features that need to be considered when evidence streams of models are used. A single point-of-view considers models, however, not as substitutes for observation but as quantified a priori information that can be fixed and regulated by remote-validation and supplementary data, though also with dynamical consistency and predictive ability. **Table 1** summarizes the roles each stream plays in producing decision-ready environmental intelligence.

2.4. Why Does Each Evidence Stream Fail Alone

A single evidence point of view is strongest when one takes into account the failures of each stream singly. Remote sensing may be international but localized, particularly when retrievals are confounded or when the variable of interest cannot be easily viewed in space. Precise but geographically sparse in situ sensing, geographically sparse in situ sensing is the most accurate, but it is subject to the realities of network placement and maintenance [28]. These models are consistent and complete but systematically biased, especially those at the local levels, which are where decisions are made and where the limits and heterogeneity prevail.

These modes of failure interact with the nature of environmental phenomena. Most environmental hazards are sharp and nonlinear. Air quality everywhere within a city block can vary radically; topography to micro-topography to drainage; wildfire behavior anywhere can vary quickly with wind, fuel, water; ecological stress can be a compound measure, and often operators interact in a compounding way. In those environments, one stream of evidence may be correct on one front and wrong on another. A ground monitor can detect a spike of pollution but miss the neighborhoods around it; a model can forecast the general meteorological pattern but not a local circulation or the terrain effect; optical images can show us in the clearest ways where the burns occurred, but overlook the active fire behavior because it is covered in smoke; on the other hand a ground monitor (utilized differently) will show a pollution spike but not all of the neighborhoods surrounding it. For example, it will show us where all the responders are located but fail to capture a local circulation or the terrain influence on fire behavior[29,30].

Table 1. Evidence streams and their complementary roles in environmental intelligence.

Evidence Stream	What It Measures Best	Strengths	Typical Limitations	Common Failure Modes	What Integration Should Use It for
Remote sensing (satellite/airborne/UAV)	Spatial patterns, surface properties, column signals, large-area change	Broad coverage, consistency, long archives, cross-border comparability	Occlusion (cloud/smoke), revisit constraints, retrieval assumptions, indirect proxies	Retrieval confounding (surface/atmosphere), cloud gaps, geometry-driven bias, scale mismatch	Provide coverage, spatial context, trend baselines, event footprints
In situ & IoT sensing (stations, gauges, probes, buoys, low-cost sensors)	Near-surface exposure, point-scale states, high-frequency dynamics	High fidelity at measurement point, continuous time series, process context	Sparse and uneven placement, representativeness bias, maintenance/calibration burden	Sensor drift, siting bias, outages during extremes, inconsistent protocols	Anchor/calibrate large-scale products, validate, capture micro-scale gradients
Models & reanalysis (NWP, hydrology, land-surface, ecosystem)	Physically consistent fields and forecasts	Physical coherence, global completeness, predictive capability, counterfactual structure	Resolution limits, parameterization error, regime-dependent bias	Smoothed extremes, wrong local dynamics, bias under specific regimes (e.g., stable boundary layer, complex terrain)	Provide dynamic context, fill gaps, support forecasting and scenario reasoning

This suggests that environmental intelligence cannot be reduced to a question of the selection of the best dataset[31]. It is an exercise in how to reconcile the incomplete evidence within some constrained view, which consists of various streams written down to code different structures of errors. Error sources in remote sensing can be connected to clouds, surface types, or viewing geometry; error sources in sensor networks can be connected to device type, maintenance practices, or siting; error sources in models can be connected to regimes such as stable boundary layers, convective events, or snow-covered surfaces. When managed explicitly by seeking to exploit complementary strengths, then integration succeeds.

2.5. From Measurements to Insight: What Environmental Products Must Deliver

The concept of environmental insight is an outcome concept: it refers to the information that modifies the decisions made, the decision-making time, or the confidence with which the decisions are made[31]. As a matter of fact, knowledge assumes various forms. They contain spatial products, e.g., exposure, inundation, vegetation condition, and land cover change maps; temporal products, e.g., trends, seasonal measures, and anomaly time series; and event-oriented products, e.g., alerts, nowcasts, and post-event evaluations. They also have explanatory products that relate observed change to drivers, either by attribution analysis or transport diagnostics, or by relating to known physical processes.

An integrated perception of evidence underlines that such products are not entirely descriptive. They have implicit assertions regarding uncertainty, representativeness, as well as validity in both space and time. A map is not only a best estimate, but it is a statement of the locations where that estimate is to be believed and of where it is extrapolated. A trend is not merely a track across the points of data but an assertion that the phenomenon that has been observed is not merely a product of sensor improvements, recovery modifications, the migration of stations, or alterations in the sampling procedural patterns[32]. An alert is a detection, as well as a policy-relevant decision threshold, the false-alarm rate, and the missed-event rate of which are of concern. It is for this reason that the shift from data to insight should be considered relative to constraints on decision-making, i.e., latency, stability in the face of missing inputs, seasonality, and geography, as well as the capacity to represent uncertainty in a manner useful to decision-makers.

The integration in making decision-grade products is also explained by the unified evidence framing. Integration is not a step to accuracy improvement but rather the process whereby coverage, fidelity, and coherence are made to conform to the requirements of the application. It permits spatial completeness in cases where there are dense observations, temporal continuity in cases where sampling is sporadic, and physical plausibility in cases where the only support is the correlation[33]. Above all, it enables environmental products to be designed in line with the facts of decision making: how to have timely updates, explicit caveats, and how to perform under extremes. Having this homogenized vision of evidence streams and their functions, the following sections discuss the practical implementation of integrated environmental intelligence in key areas of application, and what distinguishes a scientifically impressive demonstration from those that are operationally reliable.

3. Application Domain I: Atmosphere and Air Quality

3.1. Decision Questions in Air Quality Monitoring

One of the most developed areas of remote sensing, in situ measurements, and computational modelling integration is air quality monitoring. The questions that influence decision-making in this area are associated with the overall protection of the population, the regulation of the situation, and the response to emergency pollution incidents. The government will need estimates on the exposure to pollutants that are spatially resolved to provide advisory, enforcement of emission standards, and mitigation measures. The formation, transport, and elimination of atmospheric pollutants, however, take place at various spatial and time scales so that exposure may be difficult to measure directly[34].

Conventional regulatory monitoring networks give an extremely accurate measurement of the pollutants of particulate matter ($PM_{2.5}$), ozone (O_3), and nitrogen dioxide (NO_2), but lack spatial coverage and tend to be biased towards cities or areas of regulatory concern. Satellite remote sensing, on the other hand, provides regional or global measurements of the composition of the atmosphere, but generally the measurements are column-integrated values or proxies, not the values near the surface. The necessity to combine these disparate sources thus emerges in coming up with continuous air quality maps that are spatially exhaustive and locally precise[35–38].

The growing need for high-resolution exposure estimates can be attributed to the growing awareness of the fact that there is a high level of variability in air pollution across neighborhoods as a result of local sources of emissions, the form of urban structure, and weather (meteorological conditions). Therefore, there is a growing use of integrated environmental intelligence systems that can merge various observations into coherent exposure estimates with definite uncertainty[39]. We use this structure in the review, linking streams of evidence to decision-ready products (**Table 2**).

3.2. Targets and Observational Constraints

The major targets of the atmosphere about integrated air quality monitoring are the following: particulate matter (especially $PM_{2.5}$), ozone, nitrogen dioxide, and aerosol optical properties that follow smoke or dust events. All these targets have different measurement issues.

Satellite observations often access aerosol optical depth (AOD), which is the integration of the extinction of light by airborne particles available column-wise. Though AOD is a useful source of information about aerosol loading, it has to be converted into surface-based $PM_{2.5}$ concentrations, taking into consideration the effects of vertical aerosol distribution, atmospheric mixing, and humidity. Likewise, trace gas satellite observations of NO_2 biomarkers can obtain column densities as opposed to surface concentrations that are most susceptible to exposure. Ground-based surveillance networks have direct concentrations of the surface but are low-density in most areas. Even in countries where the station distance is carefully controlled, the distance between stations can be tens of kilometers apart, and this is not resolved. Recently, low-cost sensor networks have also been extended to cover a greater range, but sensor networks also add new issues of maintaining calibration stability, environmental sensitivity, and data quality assurance[40].

Weather situations increase the exposure-observation relation further. The concentrations of pollutants are affected by the frontier layer height, wind arrangement, temperature inversion, and atmospheric chemistry, among others. Consequently, there is a need to combine measurements of atmospheric composition and meteorological data to produce credible exposure estimates[41].

Table 2. Domain overview: Decision questions, targets, evidence sources, and typical outputs.

Domain (Section)	Primary Decision Questions	Core Targets	Typical Integrated Evidence	Operational Cadence (Typical)	Decision-Ready Outputs
Atmosphere & air quality (Section 3)	Advisories, exposure management, compliance tracking, wildfire smoke response	$PM_{2.5}$, O_3, NO_2, AOD/smoke proxies	Satellites + regulatory monitors + low-cost sensors + meteorology/reanalysis	Hourly–daily (events), seasonal–annual (trends)	Exposure surfaces, plume nowcasts, exceedance probability, trend indicators
Water, land & ecosystems (Section 4)	Drought triggers, flood mitigation, irrigation/reservoir ops, habitat condition	Soil moisture, ET, SWE/snowmelt, inundation, turbidity/chlorophyll, phenology	Optical/SAR/thermal + gauges/wells/towers + hydro/land/ecosystem models	Daily–weekly (monitoring), seasonal (allocation), event-based (floods)	Drought indices, flood extent maps, ET products, ecosystem stress/phenology metrics
Hazards & extremes (Section 5)	Early warning, situational awareness, resource deployment, recovery tracking	Fire activity/severity, extreme heat risk, flood extent, landslide susceptibility	SAR/optical/thermal + gauges/stations + forecasts + incident/human signals	Minutes–hours (response), daily (recovery)	Alerts, threshold exceedance risk, rapid damage mapping, operational dashboards

3.3. Integrated Evidence in Practice

The operational air quality intelligence systems normally incorporate four types of evidence, namely, satellite observations, regulatory monitoring networks, dense low-cost sensor arrays, and meteorological or chemical transportation models. These are then fed into machine learning and data assimilation techniques to provide spatially continuous pollutant estimates. **Figure 2** shows a representative air quality integration process in wildfire smoke events: probabilities of satellites, ground monitoring, low-cost sensors, and meteorological background are combined with near-surface exposure products and exceedance-risk products[42].

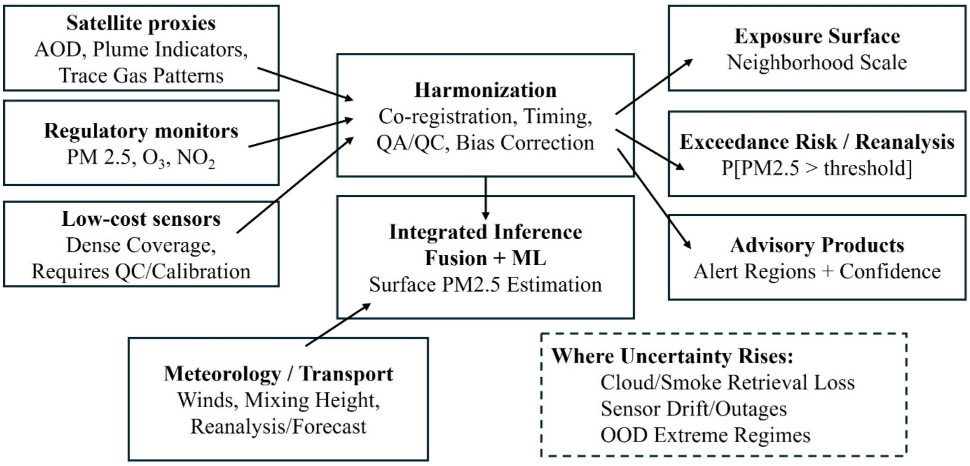

Figure 2. Air quality integration workflow (smoke events example).

Satellite measurements are wide-spatial, and they can observe the patterns of pollution on a regional scale, like wildfire smoke puffs or transport of aerosol over long distances. The ground monitoring stations are used as calibration points, which convert the satellite proxies to surface-level concentration estimates. Low-cost sensor networks provide higher spatial coverage and record variability on a scale (neighborhood), but again, their measurements need to be subsequently refined with statistical calibration techniques. Physically coherent representations of interactions between transport and chemistry are given by atmospheric models and reanalysis products, which are useful in interpolating between observations and in bridging temporal gaps[43].

- **Case Study 1: The NASA/NOAA GEOS-CF Atmospheric Composition System**

One practical illustration of how atmospheric integration is actually realized is the GEOS-Composition Forecast (GEOS-CF) system that has been created by working together with NASA and the U.S. National Oceanic and Atmospheric Administration (NOAA). By combining satellite data of atmospheric content with an international chemical transportation model and meteorological forecasts, GEOS-CF estimates pollutants like $PM_{2.5,}$ ozone, and nitrogen dioxide within a few seconds of the present situation.

The system incorporates satellite retrievals of various instruments and, at the same time, relies on meteorological analysis to model atmospheric transport as well as chemical transformations. The output is a set of globally consistent data, which offers both observational and physical consistency in estimating atmospheric composition on an hourly basis. These systems demonstrate the ability of integrated modeling systems to transform a heterogeneous measurement into a continuous environmental intelligence product (temporally) that can be analyzed scientifically and monitored operationally[44].

- **Case Study 2: Machine Learning-Driven Air Quality Mapping**

The other use case is Google's environmental insights and air quality, which takes in satellite imagery, ground monitoring data, land-use data, and meteorological variables through machine learning models to generate hyperlocal air quality maps in urban areas across the globe. These models involve aerosol indicators based on satellites along with traffic patterns, population density, and weather variables in order to conclude the spatial patterns of pollution on a much smaller scale than the traditional monitoring networks can be effective.

Such systems have been deployed on operational bases to show how integrated data pipelines can create exposure surfaces that are updated daily or hourly and present useful information to the agencies in charge of communicable disease activities and urban planners[45,46].

3.4. Operational Requirements and User Needs

The usefulness of integrated air quality products in real-world contexts is not limited to their predictive capabilities, but also to the extent of their compliance with the operational requirements. During episodes of pollution, the public health agencies need timely updates, in most cases hourly, to give advisories or recommended protective measures. Mixed products should then be able to consume data streams in near real time and generate trusted products despite incomplete or poor observations.

The spatial resolution is also critical. The estimates of exposure should be able to reflect differences among the urban neighborhoods, especially the neighborhoods around a major source of emissions, like highways or an industrial area. Integrated systems should thus be able to reconcile measurements made with very different scales, i.e., satellite footprints which are kilometers apart, with point-based ground sensors.

Transparency on uncertainty is another important need. The decision-makers should know the way confidence changes with regions and conditions, particularly in times when there are high-impact events of pollution. Explicit reporting systems allow the authorities to convey the risks more efficiently and to favor additional measurements in case of high uncertainty[47,48].

3.5. Failure Modes and Integration Challenges

Nonetheless, even after much work, the integrated air quality monitoring systems still have specific challenges. The indirect connection between pollution at the surface and that of the satellite is one of the most widespread problems. In some cases, like when smoke from wildfires happens, the aerosol distribution in the vertical direction can fluctuate over a short period, reducing the relationship between column optical depth and ground-level concentrations.

Ligament reliability is also problematic. The low-cost sensors can drift with time or react nonlinearly to the variations in humidity and temperature. Such sensors may cause systematic biases in integrated products in the absence of nonstop calibration.

Domain shift in extreme events is yet another challenge. The situation can be substantially different during atmospheric conditions during wildfires, dust storms, or temperature inversion compared to periods of normal monitoring. The likelihood of machine learning models operating on historical data to generalize to such unusual regimes might be hindered[49].

3.6. Characteristics of Robust Integrated Air Quality Systems

Effective air quality intelligence systems have a number of distinguishing features. To begin with, they have a high observational anchoring by being in constant comparison with ground monitoring stations. Second, they include meteorological context and atmospheric transport modeling in order to ensure physical validity where there are no observations. Third, they explicitly follow the uncertainty and provenance of data, such that the users can know how confidence changes through space and time[50].

This principle of graceful degradation is also increasingly accepted in operational systems, i.e., when the performance does not crash suddenly when some of the data streams are unavailable, but gradually declines. As an example, when there is a temporary cloud or heavy smoke cover on the satellite retrievals, integrated systems can call upon more ground sensors and model forecasts until new observations are available.

These attributes collectively indicate how built-in sensing systems can convert discontinuous measurements into plausible and spatially complete estimates of air pollution exposure. Such all-encompassing methods are expected to form the new paradigm of atmospheric environmental intelligence, as the power of computation and observational networks keep increasing[51].

4. Application Domain II: Water and Ecosystem Dynamics

4.1. Decision Questions in Water and Ecosystem Monitoring

Tracking of water and ecosystem dynamics is the key to managing agriculture, biodiversity, and water resources in the changing climate conditions. Some of the common questions that decision-makers in this field may have are when a drought is developing, the changes in water availability over a basin, or whether the ecosystems are under stress, which might be threatening biodiversity or food production. The environmental intelligence systems needed to address these questions should have the ability to combine observations at large spatial scales and retain the local hydrological detail.

Water and ecosystem processes exist through interrelated compartments such as soil moisture, vegetation productivity, snow storage, groundwater, and surface water flows. This is hardly what individual sensors can capture. The satellite records have a wide spatial extent of the surface condition of the land, whereas in situ records of hydrological gauges, flux towers, and soil sensors have local accuracy and time continuity. Integrated sensing systems are important in converting piecemeal observations into recognizable representations of environmental conditions[52–54].

4.2. Observational Targets and Constraints

The important variables of integrated monitoring are soil moisture, evapotranspiration, vegetation productivity, snow water equivalent, and surface water extent. All these indicators give partial consideration to how land-water-ecosystem systems operate.

The satellite platforms can provide measurements on these variables using various modalities. Microwave instruments deliver soil moisture records that can cut through the moderate vegetation cover, whereas optical sensors yield vegetation indices with regard to vegetation health and productivity. The thermal observations give details related to the temperature and evapotranspiration of the land surface, and radar observations can monitor the inundation or wetness of the soil beneath the clouds.

Regardless of such possibilities, observational constraints are still important. Spatial resolution or revisit frequency may often be a limiting factor in satellite retrievals, and uncertainties may be introduced by the environment, including dense vegetation or complicated terrain. Ground-based measurements are more accurate, but they are usually very sparse and unbalanced. It is important to combine these heterogeneous observations; hence, models and machine learning models are needed that can balance differences in scale and gaps in observation[54]. **Figure 3** organizes the water-land-ecosystem domain in the form of linked states, fluxes, and impacts, which explains how the various modes of observation limit various segments of the chain and the need to integrate them coherently[55].

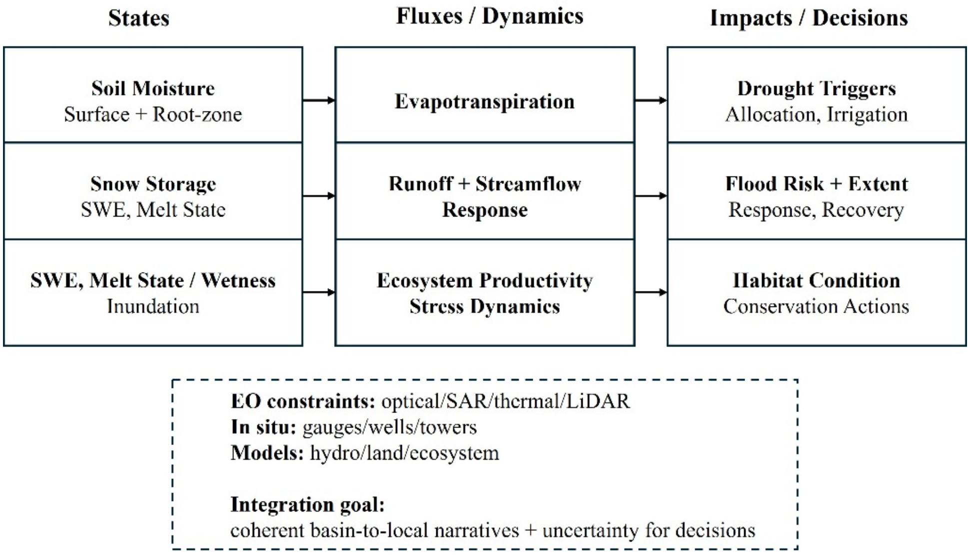

Figure 3. Water-land-ecosystem—state-flux-impact linkage diagram.

4.3. Integrating Observations across Hydrological and Ecological Scales

The use of satellite data, ground-based observations, and hydrological models in the format of operational environmental monitoring systems is becoming more and more popular to produce consistent estimates of the availability and state of the ecosystems. Machine learning models are frequently employed in incorporating varied environmental indicators and making inferences that are hard to quantify.

As an illustration, the application of satellite microwave sensors in measuring soil moisture could be associated with the use of precipitation, vegetation indices, and meteorological records to estimate the severity of drought in the agricultural landscapes. On the same note, it is also possible to generate estimates of evapotranspiration using thermal satellite observations along with meteorological inputs and land surface models.

The reconciliation between observations taken at various spatial and time scales must be done in an integrated system. Regional snapshots can be offered by satellite observations at a frequency of a few days, and ground sensors can be used to record measurements continuously at a specific location. To motivate a combination of these sources, calibration and statistical modeling are necessary to guarantee that integrated products are physically and operationally viable[56,57].

Case Study: Satellite Ground Integration in Global Drought Monitoring

Another example of this form of environmental monitoring is the application of the NASA Soil Moisture Active Passive (SMAP) satellite mission alongside ground-based measurements and land surface models to drought detection. SMAP offers measurements of soil moisture all over the world using microwave radiometry, which can be used to cover large areas consistently. These observations are, however, supplemented by ground meteorological networks and hydrological models that use data on precipitation, temperature, and vegetation.

These various streams of evidence are combined by operational drought monitoring systems like the U.S. Drought Monitor and other regional systems in order to determine the severity of droughts. The anomalies of soil moisture measured by satellites can be used as an early warning of water stress, whereas ground-based measurements and hydrological modeling can support and put these results into perspective. This combination enables the monitoring systems to identify new cases of drought in areas that would be inadequately measured by ground-based measurements[58–60].

4.4. Operational Requirements and Decision Context

The environmental intelligence systems to aid water and ecosystem management should meet several operational demands. To begin with, tracking products should be able to record regional trends and local differences. The availability of water within a watershed may be very different because of the differences in soil type, land cover, and irrigation methods. The integrated systems should thus strike a balance between space coverage and adequate resolution to make a decision about local management.

Second, it is imperative that there is continuity over time. Numerous hydrological processes develop in the course of weeks or months, and, as a consequence, the observation system has to keep the records constant. The failure of a satellite connection or sensors can also interrupt the time series analysis, and having backups among data sources is a valid design requirement.

Lastly, product monitoring should be able to give interpretable indicators as far as policy and management are concerned. The simplified measures that decision-makers normally want are drought severity category, vegetation stress index, or basin-wide water balance estimates. The key issue in operational environmental intelligence has been the transformation of complex environmental observations into actionable indicators [61].

4.5. Failure Modes and Integration Challenges

Although more advanced integrated monitoring has been made, several failure modes still exist. One of these issues is the discrepancies in scale between satellite measurements and local weather. The satellite measurements can smooth the environmental signals over areas with different land uses or hydrological conditions, and therefore, local water stress may be obscured.

The other challenge is data gaps and the uncertainty of retrieval. Satellite measurements of the surface are impossible when the cloud cover is extensive, and the retrieval of the microwave can be influenced by vegetation cover or the stiffness of the surface. If such constraints are present, an integrated system has to be more dependent on other sources of data, like ground measurements or model predictions [62].

Domain shifts that are related to extreme events also prove to be challenging. As an example, anomalous drought patterns can change the behavior of vegetation that is not captured in the historical fields of training data, making machine learning models based on historical data unreliable.

4.6. Characteristics of Robust Monitoring Systems

Strong water and ecosystem surveillance procedures integrate several complementary data streams in such a way that the demise or subordination of one of the observational streams does not significantly restrict the general system function. This concept of graceful degradation is used to ensure that the monitoring products are not wasted when some of the observations cannot be made. Good systems are also capable of integrating observational evidence with process-based models that ensure physical consistency among the hydrological variables. Monitoring systems can have coherent representations of the balance of water and the dynamics of the ecosystem by combining satellite data with meteorological measurements and hydrological simulations [63].

Lastly, clear uncertainty reporting and confirmation by independent observations are also necessary in developing confidence in operational environmental intelligence products. Through such practices, it is guaranteed that integrated monitoring systems are credible tools for managing water resources and protecting ecosystems in a changing environment.

5. Application Domain III: Environmental Hazards and Disaster Response

5.1. Decision Questions in Hazard Monitoring

Floods, fires, and other dangerous weather events are environmental hazards that cause instant threats to human life, infrastructure, and nature. The disaster response monitoring systems should hence be able to deliver timely and reliable information regarding the emergence of hazards, spatial extent, and potential effects. Unlike more mundane types of environmental monitoring, the hazard intelligence systems are subject to very stringent time limitations and conditions in which observations are incomplete or degraded. Satellite imagery, sensor networks, and model frameworks should therefore be incorporated within a very short period of time to create situational awareness in the developing crisis [64].

5.2. Observational Evidence Streams

The hazard monitoring systems are based on several categories of observations. Satellite capabilities offer quick evaluation of hazardous areas using optical information, radar, and thermal images. The use of synthetic aperture radar (SAR) is especially beneficial as the device is capable of locating surface water and structural damage even when clouds are in place.

On the ground measurements, such as river gauges, rainfall sensors, and meteorological stations, give very important local measurements that can be used to confirm early warning systems and model validation. These observations are, in most instances, combined in hydrological or atmospheric models to generate hazard evolution forecasts. By putting these heterogeneous observations together, the monitoring systems can build coherent situational awareness notwithstanding the constraints of any one of the data sources[65].

5.3. Multi-Source Evidence during Major Flood Events

The value of integrated sensing frameworks becomes particularly evident during large-scale flood disasters, where observational constraints frequently challenge monitoring systems.

Case Study: Multi-Sensor Monitoring during the 2022 Pakistan Floods

The examples of the Pakistan floods that have hit millions of people in the Indus River basin in 2022 demonstrate the practicality of multi-source environmental intelligence systems. Throughout the incident, the activity was hampered by constant cloud cover, which hindered the use of the optical satellite imagery to map the extent of the floods, making it difficult to use the traditional remote sensing techniques. To overcome this obstacle, missions like Sentinel-1 with synthetic aperture radar (SAR) are essential to monitor operations, as the radar can observe surface water irrespective of whether there are clouds or not.

These radar images were used with optical images where there were images, river gauges, and satellite-based rainfall product estimates. Introducing hydrological models into the system integrated such inputs and estimated the further development of floods and the areas that were at risk of further inundation. These multi-source datasets were used by international surveillance activities like the Copernicus Emergency Management Service and other humanitarian mapping projects to produce near-real-time flood extent maps and damage estimates[66].

This is where the concept of graceful degradation in integrated hazards monitoring is evident. At the time, the optical imagery was not reliable because of cloud cover, and radar images and ground measurements were other sources of evidence that enabled situational awareness to be preserved. The combination of these sources of data allowed emergency responders and humanitarian organizations to coordinate their actions in response, even when there were not full observational conditions.

5.4. Operational Requirements for Hazard Intelligence

Systems used in the operational hazard monitoring processes will have to meet high-performance standards. Timeliness is important because delays in hazard detection or mapping may have a great impact on the evacuation planning, resource distribution, and emergency response. Integrated systems should be able to process incoming observations quickly, therefore, and come up with new hazard assessments within hours, not days.

It is also important that spatial coverage is good. Hazards often serve large geographical locations, and as such, they may need geographic surveillance systems that are able to integrate both regional satellite images and local measurements to determine the worst hit regions. It is also important to note that in most instances, the decision-makers need to have forecasts of the hazard evolution, and this means that predictive models must be integrated with this.

There should also be effective communication of uncertainty. The forecasts and the hazard maps should communicate the confidence levels in such a way as to enable decision-makers to be certain about the trustworthiness of the information they have in case of the events that have to take place in a short time[67].

5.5. Failure Modes in Hazard Monitoring Systems

Environmental hazard intelligence systems are also prone to several types of failure modes. One of them is the loss or degradation of observation, e.g., cloud cover that hides optical imagery or the failure of ground sensors during severe weather. Redundancy of sources of observation must thus be incorporated in the integrated monitoring systems to ensure that there is situational awareness. A second challenge is a result of data latency. Satellite images and surface observations can arrive at various times, and as such, this may delay the integration process. The real-time hazard intelligence systems should, therefore, strike the right balance between the necessity of prompt updates and the provision of good observations[68].

Lastly, machine learning models with historical data could not perform well in cases of extreme events, which are not similar to the historical events. It is also a challenge to operational environmental intelligence to ensure that models are not compromised by this situation.

5.6. Governance, Equity, and Trust in Environmental Hazard Intelligence

Since integrated environmental monitoring systems are progressively important in disaster response and risk management, the issue of governance and equity emerges as a core part of the design and assessment of integrated environmental monitoring systems. The products of environmental intelligence affect the process of the identification, interpretation, and prioritization of hazards; i.e., the technical performance cannot be relied on to provide credible decision support. An important mode of failure is due to imbalanced coverage of observations. Although satellite monitoring has a wide geographic coverage, several important elements of monitoring systems, including gauge networks, communication infrastructure, and ground sensors, are not evenly distributed. Areas with poor density of observation networks can thus be subjected to more uncertainty or slower hazard detection. The consequences of such inequalities directly relate to environmental justice because the community with low monitoring infrastructure might also experience a greater exposure to environmental hazards[69–72].

One more problem is related to the overconfidence in automated hazard products. Systems that incorporate machine learning, remote sensing, and sensor networks are usually used to provide visual insights in the form of maps and predictions. Unless uncertainty and data drawbacks are clearly communicated, however, such products can be construed as certain characterizations of risk. This overconfidence is likely to influence operational decision-making, especially in a fast-changing disaster.

To overcome these challenges, there is a need to have governance practices that come with technical development. Specifically, equity-conscious validation must be used to supplement conventional performance measures to look at the variability in system reliability in different regions that have different densities of observation or socio-economic backgrounds. They also require reporting of uncertainty, data provenance, as well as observational gaps in a transparent way so as to allow informed interpretation by operational users.

The incorporation of these principles of governance in the framework of environmental monitoring systems strengthens the overall goal of decision-ready environmental intelligence, namely the generation of information that is technically sound as well as credible, fair, and responsive to the societal settings in which environmental risks are perceived and addressed. Since an integrated product is only as helpful as it is reliable in the actual working conditions, in **Table 3** we summarize suggested validation designs, metrics, and uncertainty reporting practices. In all domains, the failures are limited by a common set of failure modes, causing a lack of transferability and operational confidence; **Table 4** highlights these issues and the corresponding mitigation methods[73–75].

Table 3. Evaluation and uncertainty reporting template for integrated environmental products.

Product Type	**Validation Split That Matches Use**	**Performance Metrics That Matter**	**Uncertainty Outputs to Report**	**Stress Tests to Include**	**Minimum Reporting for Reproducibility**
Gridded maps (e.g., $PM_{2.5}$, ET, inundation)	Spatial holdout (leave-location/region out) + seasonal holdout	RMSE/MAE plus spatial skill, bias by land cover/urbanicity, extremes error	Calibrated predictive intervals or exceedance probabilities	Cloud/smoke gaps, sensor dropouts, new geography	Data versions, QA/QC rules, tiling/resampling, train/test geography
Time series indicators (e.g., drought onset, phenology)	Temporal holdout across years + regime holdout	Timing error (onset/recovery), persistence skill, trend stability	Uncertainty on change points/trends, reliability over time	Shifting baseline periods, sensor/retrieval updates	Baseline definition, stationarity assumptions, versioning policy

Table 3. *Cont.*

Product Type	Validation Split That Matches Use	Performance Metrics That Matter	Uncertainty Outputs to Report	Stress Tests to Include	Minimum Reporting for Reproducibility
Alerts/nowcasts (e.g., smoke, flood warning)	Event-based holdout (unseen events) + near-miss negatives	Precision/recall, false alarm rate, detection latency	Probability of threshold exceedance + confidence degradation flags	Rare extremes, compound events, degraded comms/coverage	Latency targets, failure behavior, threshold definitions and rationale

Table 4. Cross-domain failure modes and practical mitigations.

Recurrent Challenge	How It Shows Up (Symptom)	Why It Happens (Root Cause)	Diagnostics That Reveal It	Mitigation Strategies (Integration-Oriented)
Scale mismatch	Good performance at stations, poor elsewhere	Point measurements vs. pixel/basin averages	Error vs. land cover/topography; holdout-by-region	Representativeness-aware training; hierarchical/scale-aware products; multi-resolution fusion
Proxy ambiguity	Same signal implies different states	EO signals conflate drivers (e.g., greenness vs. harvest)	Counterexample analysis; subgroup errors	Add contextual constraints (meteorology/hydrology/land use); multi-source corroboration
Domain shift in extremes	Biggest errors during disasters	Rare regimes underrepresented	Event-only benchmarks; tail metrics	Extreme-focused training sets; uncertainty up-weighting; drift monitoring
Sensor drift and heterogeneity	Gradual bias growth, inconsistent hotspots	Device aging, calibration changes	Co-location checks; temporal bias trends	Continuous calibration; QC gating; robust aggregation and sensor credibility weighting
Missing/degraded inputs	Output instability or silent failure	Clouds, outages, pipeline breaks	Coverage flags; performance vs. missingness	Graceful degradation design; redundancy across modalities; explicit confidence reduction
Unequal coverage and equity risks	Worse accuracy in under-monitored areas	Network placement bias	Error maps vs. socioeconomic/land use strata	Coverage-aware evaluation; uncertainty surfacing; targeted sensing strategies and transparent caveats

6. Conclusion and Outlook

The computation of the planet is becoming less and less about accumulating more and more environmental information and increasingly more about transforming heterogeneous, imperfect observations into coherent decision-ready information. With atmosphere and air quality, water-land-ecosystem interactions, hazards, and extremes, a general trend is evident: there is no uniform data stream adequate. Remote sensing is accessible and offers longevity, but it is based on retrieval questions, obscuration, and size. Sensing in situ and IoT is not dense and covers fidelity and local context, but is sparse, uneven, and variable in quality. Models and reanalysis offer regime-specific predictive structure and offer physical coherence at computational costs in the form of regime-specific biases and cutoffs. Integrated systems - bringing together such streams with the aid of machine learning and fusion are not the best extension; instead, they will become the foundational pillars of environmental understanding at the types, latencies, and levels of reliability required by contemporary administration, population health, and more robust planning.

Concurrently, the review points out the fact that integration does not mean trust. The most significant failures happen to systems that are good in average modes but silently fail in domain shift, seldom in extremes, or under poor observational coverage. In the air quality, this is most commonly manifested in the form of an inappropriate calibration in column observations and near-surface exposure to wildfire smoke or stagnant inversion. It manifests itself in water and ecosystems in the form of proxy ambiguity, scale mismatch, and shifting baselines that weaken the association of indicators of remote sensing with underlying hydrological or ecological conditions. In hazards, it manifests itself in the form of rare-event bias and pipelines that are brittle and burst at the very time when time pressure and impact are the greatest. The implication for the field is obvious that the subsequent stage of development is going to be characterized not by improvement in terms of incremental gains on typical standards but by strength, measurement of uncertainty, and accountability in work.

This synthesis generates a number of cross-domain recommendations. First, the assessment should be in line with the intended use. Spatially and temporally structured data demand validation designs, which test performance based on generalization to new geographic positions, seasons, and event regimes, as opposed to the performance-inflationary random splits based on randomized mixes. Second, the idea of uncertainty is to be discussed as one of the main product requirements. Uncertainty is the interface between prediction and decision required to do integrated environmental intelligence: it needs to be empirically tuned, receptive to incomplete or bad data, and formulated into forms accessible as threshold conditioned actions and risk messages. Third, versioning and provenance need to be the norm. The integrated products require the development of sensors, retrieval algorithms, and model components; without the provision of transparent documentation and revalidation, even scientifically valid systems may lose credibility due to the untraceable changes in behavior.

Three directions are expected to define the notion of computing on the planet in the future, both as a research agenda and as an operational infrastructure. The former is the emergence of multimodal foundation models and large-scale representation learning specific to Earth data, which can both decrease the reliance on labeled data and also allow a wider range of sensors and geographies to be used more consistently. The second one is a more intricate connection between learning and physics, including hybrid methods that maintain physical constraints but make use of data-driven flexibility, enhancing extrapolation with changing climates and management regimes. The third one is based on maturing real-time environmental intelligence systems- streaming ingestion, edge-to-cloud sensor systems, and continuously monitored models that may deliver early warning, situational awareness with explicit confidence, and graceful degradation.

Lastly, the calculation of the planet cannot be done without the government. The data coverage is as real as social and economic facts, and integrated products may unintentionally recreate inequities when there are systematic differences in performance and uncertainty across communities. Environmentally sensitive data should be stewarded, and the decision support system should be structured in such a way that it does not give overconfident results that can lead to harm. With the increasing capability of environmental intelligence, the discipline should expand its definition of success by encompassing not just the accuracy, but also transparency, equity, reproducibility, and the ability to make decisions in stressful environments.

Altogether, the combination of machine learning, remote sensing, and sensor data fusion provides a window to the environment that can provide a detailed view of the Earth system and the sense of urgency of contemporary environmental issues. It is not an ideal map of the planet at all times, impossible in the face of limitations of observation and uncertainties but a good evidence infrastructure: an infrastructure that is clear on what it knows, candid on what it does not know, and strong in the face of decisions that must be made in the face of change and when the stakes are high.

In addition to providing an overview of progress in the field of machine learning and environmental sensing, this review has attempted to highlight a significant change in the conceptualization of integrated Earth observation systems design and assessment. Although most of the current surveys are relatively algorithm-centric, architecture-centric, or data-modality-centric, the approach in this paper is operational decision requirements as the unifying factor in the integration of machine learning, remote sensing, and in situ sensing networks. Through air quality monitoring, water and ecosystem dynamics, and environmental hazards, the review demonstrates how a successful environmental intelligence system needs to balance the heterogeneous observations and maintain physical and statistical coherence across scales, and its ability to stand up under uncertainty, sparse observations, and extreme events. This application-oriented and failure-mode-sensitive framing does not just present a roadmap on how methodological research can be advanced, but also how operational systems can be developed to be trusted enough. In this respect, the target audience is not only the community of academic scholars working in the field of data fusion and Earth observation but also environmental product developers, agencies engaged in operational monitoring, and funding organizations that will design next-generation environmental intelligence infrastructures. The development of the future will be based not only on enhanced predictive accuracy but also on creating integrated systems that will not only be transparent and reliable but also able to support planetary-scale real-world environmental decision-making.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

No new data were created or generated in this study. As this is a review, it is based on data and information from previously published sources, which are cited in the reference list.

Conflicts of Interest

The author declares no conflict of interest.

References

[1] Gössling, S., Scott, D., 2025. Tourist Demand and Destination Development under Climate Change: Complexities and Perspectives. Journal of Sustainable Tourism. 1–31. DOI: https://doi.org/10.1080/09669582.2025.2543953

[2] Aguado, E., Burt, J.E., 2015. Understanding Weather and Climate. Pearson: Boston, MA, USA.

[3] Gettelman, A., Geer, A.J., Forbes, R.M., et al., 2022. The Future of Earth System Prediction: Advances in Model–Data Fusion. Science Advances. 8(14), eabn3488.

[4] Han, H., Liu, Z., Li, J., et al., 2024. Challenges in Remote Sensing Based Climate and Crop Monitoring: Navigating the Complexities Using AI. Journal of Cloud Computing. 13(1), 34.

[5] Chuvieco, E., 2020. Fundamentals of Satellite Remote Sensing: An Environmental Approach, 3rd ed. CRC Press: Boca Raton, FL, USA.

[6] Fascista, A., 2022. Toward Integrated Large-Scale Environmental Monitoring Using WSN/UAV/Crowdsensing: A Review of Applications, Signal Processing, and Future Perspectives. Sensors. 22(5), 1824.

[7] Kreinovich, V., Ouncharoen, R., 2015. Fuzzy (and Interval) Techniques in the Age of Big Data: An Overview with Applications to Environmental Science, Geosciences, Engineering, and Medicine. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 23(Suppl. 1), 75–89.

[8] Tao, Y., Tian, B., Adhikari, B.R., et al., 2024. A Review of Cutting-Edge Sensor Technologies for Improved Flood Monitoring and Damage Assessment. Sensors. 24(21), 7090.

[9] Salazar, J.J., Garland, L., Ochoa, J., et al., 2022. Fair Train–Test Split in Machine Learning: Mitigating Spatial Autocorrelation for Improved Prediction Accuracy. Journal of Petroleum Science and Engineering. 209, 109885.

[10] Welsh, M., 2014. Resilience and Responsibility: Governing Uncertainty in a Complex World. The Geographical Journal. 180(1), 15–26.

[11] Layode, O., Naiho, H.N.N., Adeleke, G.S., et al., 2024. Data Privacy and Security Challenges in Environmental Research: Approaches to Safeguarding Sensitive Information. International Journal of Applied Research in Social Sciences. 6(6), 1193–1214.

[12] Li, X., Feng, M., Ran, Y., et al., 2023. Big Data in Earth System Science and Progress Towards a Digital Twin. Nature Reviews Earth & Environment. 4(5), 319–332.

[13] Wu, X., Lu, G., Wu, Z., 2023. Remote Sensing Technology in the Construction of Digital Twin Basins: Applications and Prospects. Water. 15(11), 2040.

[14] Nica, E., Popescu, G.H., Poliak, M., et al., 2023. Digital Twin Simulation Tools, Spatial Cognition Algorithms, and Multi-Sensor Fusion Technology in Sustainable Urban Governance Networks. Mathematics. 11(9), 1981.

[15] Poliak, M., Poliakova, A., Crîșmariu, O.-D., et al., 2023. Visual Analytics and Digital Twin Modeling Tools, Spatio-Temporal Fusion and Predictive Modeling Algorithms, and Deep Learning-Based Sensing and Image Recognition Technologies in Data-Driven Smart Sustainable Cities and Immersive Multisensory Virtual Spaces. Geopolitics, History, and International Relations. 15(1), 91–105.

[16] Rundel, P.W., Graham, E.A., Allen, M.F., et al., 2009. Environmental Sensor Networks in Ecological Research. New Phytologist. 182(3), 589–607.

[17] Lein, J.K., 2011. Environmental Sensing. In Environmental Sensing: Analytical Techniques for Earth Observation. Springer: New York, NY, USA. pp. 23–49.

[18] Dixon, B., McHan, C.R., 2025. Geographic Problem Solving: A Summary of Key Environmental Problems that Need Integrative Approaches and the Notion of Uncertainty Associated with 'Ways of Knowing'. In Interdisciplinary Environmental Solutions: Using Geospatial Technologies for Bridging Disciplines, Scale and Data. Palgrave Macmillan: Cham, Switzerland. pp. 381–444.

[19] Dash, J., Ogutu, B.O., 2016. Recent Advances in Space-Borne Optical Remote Sensing Systems for Monitoring Global Terrestrial Ecosystems. Progress in Physical Ge-

ography. 40(2), 322–351.

[20] Thomas, Z.A., 2016. Using Natural Archives to Detect Climate and Environmental Tipping Points in the Earth System. Quaternary Science Reviews. 152, 60–71.

[21] De Leeuw, J., Georgiadou, Y., Kerle, N., et al., 2010. The Function of Remote Sensing in Support of Environmental Policy. Remote Sensing. 2(7), 1731–1750.

[22] Aplin, P., 2006. On Scales and Dynamics in Observing the Environment. International Journal of Remote Sensing. 27(11), 2123–2140.

[23] Zhang, J., Atkinson, P., Goodchild, M.F., 2014. Scale in Spatial Information and Analysis. CRC Press: Boca Raton, FL, USA.

[24] Balsamo, G., Agusti-Panareda, A., Albergel, C., et al., 2018. Satellite and In Situ Observations for Advancing Global Earth Surface Modelling: A Review. Remote Sensing. 10(12), 2038.

[25] Crespi, M., Ferrari, G., 2020. Reanalysis, or Retrospective Analysis, in Meteorology. Open Meteorology Notebook. 2. Available from: https://www.hypermeteo.com/wp-content/uploads/2021/12/2-Reanalysis-or-retrospective-analysis-in-meteorology.pdf

[26] Gibbs, D.A., Flotemersch, J., 2019. How Environmental Futures Can Inform Decision Making: A Review. Futures. 108, 37–52.

[27] Sillmann, J., Thorarinsdottir, T., Keenlyside, N., et al., 2017. Understanding, Modeling and Predicting Weather and Climate Extremes: Challenges and Opportunities. Weather and Climate Extremes. 18, 65–74.

[28] Zhang, P., Martonosi, M., 2008. Locale: Collaborative Localization Estimation for Sparse Mobile Sensor Networks. In Proceedings of the 2008 International Conference on Information Processing in Sensor Networks (IPSN 2008), St. Louis, MO, USA, 22–24 April 2008; pp. 195–206.

[29] Scheffran, J., 2015. Complexity and Stability in Human–Environment Interaction: The Transformation from Climate Risk Cascades to Viable Adaptive Networks. In: Kavalski, E. (Ed.). World Politics at the Edge of Chaos: Reflections on Complexity and Global Life. State University of New York Press: Albany, NY, USA. pp. 229–252.

[30] Wu, Y., 2003. Mechanism Analysis of Hazards Caused by the Interaction between Groundwater and Geo-Environment. Environmental Geology. 44(7), 811–819.

[31] Malakar, N., 2022. Environmental Intelligence Using Machine Learning Applications. Journal of Nepal Physical Society. 8(2), 42–47.

[32] Chandler, R., Scott, M., 2011. Statistical Methods for Trend Detection and Analysis in the Environmental Sciences. John Wiley & Sons: Hoboken, NJ, USA.

[33] Wang, J.-F., Stein, A., Gao, B., et al., 2012. A Review of Spatial Sampling. Spatial Statistics. 2, 1–14.

[34] Kumar, P., Debele, S.E., Sahani, J., et al., 2021. An Overview of Monitoring Methods for Assessing the Performance of Nature-Based Solutions against Natural Hazards. Earth-Science Reviews. 217, 103603.

[35] Bagkis, E., Debele, S.E., Sahani, J., et al., 2025. Evolving Trends in Application of Low-Cost Air Quality Sensor Networks: Challenges and Future Directions. npj Climate and Atmospheric Science. 8(1), 335.

[36] Javan, F.D., Samadzadegan, F., Toosi, A., 2025. Air Pollution Observation—Bridging Spaceborne to Unmanned Airborne Remote Sensing: A Systematic Review and Meta-Analysis. Air Quality, Atmosphere & Health. 18(8), 2481–2549.

[37] Mashala, M.J., Dube, T., Mudereri, B.T., et al., 2023. A Systematic Review on Advancements in Remote Sensing for Assessing and Monitoring Land Use and Land Cover Changes Impacts on Surface Water Resources in Semi-Arid Tropical Environments. Remote Sensing. 15(16), 3926.

[38] World Health Organization, 2021. WHO Global Air Quality Guidelines: Particulate Matter ($PM_{2.5}$ and PM_{10}), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide. World Health Organization: Geneva, Switzerland.

[39] Stratoulias, D., Nuthammachot, N., Dejchanchaiwong, R., et al., 2024. Recent Developments in Satellite Remote Sensing for Air Pollution Surveillance in Support of Sustainable Development Goals. Remote Sensing. 16(16), 2932.

[40] Fawole, O.G., Ogunjo, S.T., Olabode, A., et al., 2026. Estimation of Surface PM2.5 Concentration from Satellite Aerosol Optical Depth Using a Constrained Observation-Based Model. Climate. 14(1), 1.

[41] Lolli, S., 2025. Urban $PM_{2.5}$ Concentration Monitoring: A Review of Recent Advances in Ground-Based, Satellite, Model, and Machine Learning Integration. Urban Climate. 63, 102566.

[42] Krecl, P., Akutagawa, G., Targino, A.C., et al., 2026. Integrating Low-Cost Sensors, Fixed and Mobile Monitoring with Satellite Data to Assess the Impact of Biomass Burning on Fine, Ultrafine and Black Carbon Particle Concentrations. Atmospheric Environment. 372, 121838.

[43] Duncan, B.N., Prados, A.I., Lamsal, L.N., et al., 2014. Satellite Data of Atmospheric Pollution for U.S. Air Quality Applications: Examples of Applications, Summary of Data End-User Resources, Answers to FAQs, and Common Mistakes to Avoid. Atmospheric Environment. 94, 647–662.

[44] Gonzalez Abad, G., Souri, A.H., Bak, J., et al., 2019. Five Decades Observing Earth's Atmospheric Trace Gases Using Ultraviolet and Visible Backscatter Solar Radiation from Space. Journal of Quantitative Spectroscopy and Radiative Transfer. 238, 106478.

[45] Alvarez, C.I., Ulloa Vaca, C.A., Echeverria Llumipanta, N.A., 2025. Machine Learning for Urban Air Quality Prediction Using Google AlphaEarth Foundations

Satellite Embeddings: A Case Study of Quito, Ecuador. Remote Sensing. 17(20), 3472.

[46] Karmoude, M., Munhungewarwa, B., Chiraira, I., et al., 2025. Machine Learning for Air Quality Prediction and Data Analysis: Review on Recent Advancements, Challenges, and Outlooks. Science of the Total Environment. 1002, 180593.

[47] Mujtaba, M.A., Munir, M.A., Ali, S., et al., 2025. Using Machine Learning for Air Quality Prediction and Sustainable Urban Planning. Sustainable Futures. 10, 100981.

[48] Bernacki, J., Scherer, R., 2025. A Comprehensive Review of Data-Driven Techniques for Air Pollution Concentration Forecasting. Sensors. 25(19), 6044.

[49] Anggraini, T.S., Irie, H., Sakti, A.D., et al., 2025. Global Air Quality Index Prediction Using Integrated Spatial Observation Data and Geographic Machine Learning. Science of Remote Sensing. 11, 100197.

[50] Imam, M., Adam, S., Dev, S., et al., 2024. Air Quality Monitoring Using Statistical Learning Models for Sustainable Environment. Intelligent Systems with Applications. 22, 200333.

[51] Alsamrai, O., Redel-Macias, M.D., Dorado, M.P., 2025. Real-Time Intelligent Monitoring of Outdoor Air Quality in an Urban Environment Using IoT and Machine Learning Algorithms. Applied Sciences. 15(16), 9088.

[52] Chang, H., Ye, T., Xie, Z., et al., 2025. Application of Environmental DNA in Aquatic Ecosystem Monitoring: Opportunities, Challenges and Prospects. Water. 17(5), 661.

[53] Torres López, S., Barrionuevo, M.d.l.A., Rodríguez-Labajos, B., 2023. A New Operational Approach for Understanding Water-Related Interactions to Achieve Water Sustainability in Growing Cities. Environment, Development and Sustainability. 25(1), 122–137.

[54] Bianconi, A., Furlan, E., Vascon, S., et al., 2026. Harnessing AI for Smarter Water Management under a Changing Climate: A Review of Machine Learning and Deep Learning Applications within EU Water Framework Directive and Marine Strategy Framework Directive. Ocean & Coastal Management. 274, 108093.

[55] Pacetti, T., 2018. Investigating Water–Land–Ecosystems Nexus for Watershed Integrated Management [PhD Thesis]. Technische Universität Braunschweig: Braunschweig, Germany.

[56] Arrighi, C., De Simone, M., Checcucci, G., et al., 2025. Development of an Eco-Hydrological Distance Index and Improved Environmental Flow Assessment by Integrating Ecological Monitoring and Hydrological Modeling. Science of the Total Environment. 969, 178961.

[57] Albuhaisi, Y.A.Y., van der Velde, Y., Pandey, S., et al., 2024. Integrating Satellite Observations and Hydrological Models to Unravel Large TROPOMI Methane Emissions in South Sudan Wetlands. Remote Sensing. 16(24), 4744.

[58] Martínez-Fernández, J., González-Zamora, A., Sánchez, N., et al., 2016. Satellite Soil Moisture for Agricultural Drought Monitoring: Assessment of the SMOS Derived Soil Water Deficit Index. Remote Sensing of Environment. 177, 277–286.

[59] Farahani, A., Moradikhaneghahi, M., Ghayoomi, M., et al., 2022. Application of Soil Moisture Active Passive (SMAP) Satellite Data in Seismic Response Assessment. Remote Sensing. 14(17), 4375.

[60] Xu, L., Abbaszadeh, P., Moradkhani, H., et al., 2020. Continental Drought Monitoring Using Satellite Soil Moisture, Data Assimilation and an Integrated Drought Index. Remote Sensing of Environment. 250, 112028.

[61] Olawade, D.B., Wada, O.Z., Ige, A.O., et al., 2024. Artificial Intelligence in Environmental Monitoring: Advancements, Challenges, and Future Directions. Hygiene and Environmental Health Advances. 12, 100114.

[62] Thies, B., Bendix, J., 2011. Satellite-Based Remote Sensing of Weather and Climate: Recent Achievements and Future Perspectives. Meteorological Applications. 18(3), 262–295.

[63] Hipsey, M.R., Hamilton, D.P., Hanson, P.C., et al., 2015. Predicting the Resilience and Recovery of Aquatic Systems: A Framework for Model Evolution within Environmental Observatories. Water Resources Research. 51(9), 7023–7043.

[64] Zografos, K.G., Vasilakis, G.M., Giannouli, I.M., 2000. Methodological Framework for Developing Decision Support Systems (DSS) for Hazardous Materials Emergency Response Operations. Journal of Hazardous Materials. 71(1–3), 503–521.

[65] Joyce, K.E., Belliss, S.E., Samsonov, S.V., et al., 2009. A Review of the Status of Satellite Remote Sensing and Image Processing Techniques for Mapping Natural Hazards and Disasters. Progress in Physical Geography. 33(2), 183–207.

[66] Mubeen, A., Saeed, S., 2023. GeoAI-Driven Flood Susceptibility Mapping and Exposure Assessment in Pakistan Using Multi-Source Geospatial Big Data and Deep Learning Models. Frontiers in Computational Spatial Intelligence. 1(2), 87–97.

[67] Harder, C., Sutler, J., Sawaryn, S., et al., 2015. Real Time Data Monitoring Experience Results in Enhanced Safety and Efficiency. In Proceedings of the SPE/IADC Drilling Conference and Exhibition, London, UK, 17–19 March 2015. DOI: https://doi.org/10.2118/173090-MS

[68] Tsatsaris, A., Kalogeropoulos, K., Stathopoulos, N., et al., 2021. Geoinformation Technologies in Support of Environmental Hazards Monitoring under Climate Change: An Extensive Review. ISPRS International Journal of Geo-Information. 10(2), 94.

[69] Sarker, S., 2025. Role of Management Information Systems in Environmental Risk Assessment: A Systematic Review of Geographic and Ecological Applications.

American Journal of Interdisciplinary Studies. 6(1), 95–126.

[70] Silva Villanueva, P., 2011. Learning to ADAPT: Monitoring and Evaluation Approaches in Climate Change Adaptation and Disaster Risk Reduction—Challenges, Gaps and Ways Forward. The Institute of Development Studies and Partner Organisations, University of Sussex: Brighton, UK.

[71] Aslam, M.A., Abbas, M.S., Mustaqeem, M., et al., 2024. Comprehensive Assessment of Heavy Metal Contamination in Soil–Plant Systems and Health Risks from Wastewater-Irrigated Vegetables. Colloids and Surfaces C: Environmental Aspects. 2, 100044.

[72] Khan, M.N., Aslam, M.A., Muhsinah, A.B., et al., 2023. Heavy Metals in Vegetables: Screening Health Risks of Irrigation with Wastewater in Peri-Urban Areas of Bhakkar, Pakistan. Toxics. 11(5), 460.

[73] Yazdi, M., 2024. Reliability-Centered Design and System Resilience. In Advances in Computational Mathematics for Industrial System Reliability and Maintainability. Springer: Cham, Switzerland. pp. 79–103.

[74] Zheng, H., Wang, R., Yang, Y., et al., 2019. Cross-Domain Fault Diagnosis Using Knowledge Transfer Strategy: A Review. IEEE Access. 7, 129260–129290.

[75] Baustert, P., Othoniel, B., Rugani, B., et al., 2018. Uncertainty Analysis in Integrated Environmental Models for Ecosystem Service Assessments: Frameworks, Challenges and Gaps. Ecosystem Services. 33, 110–123.