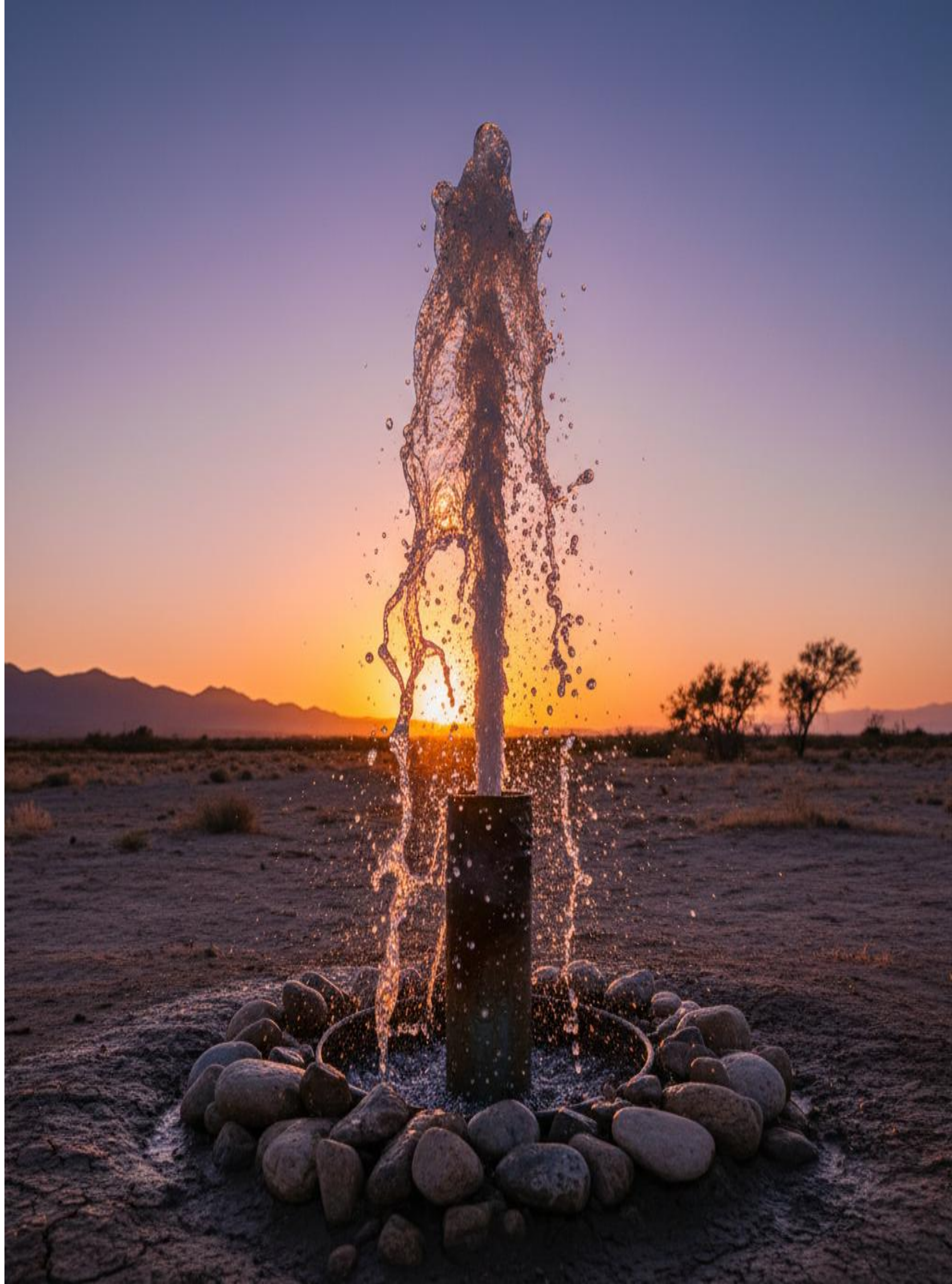
Scientific Policy-making for Groundwater Management

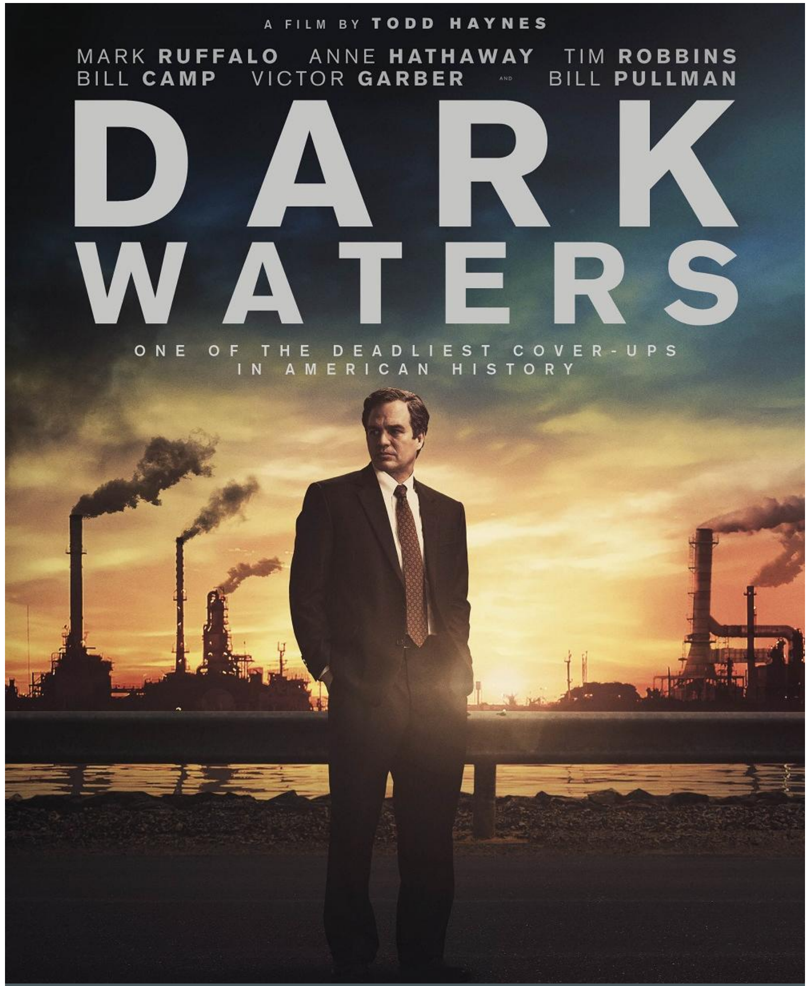
From Data to Insight, Insight to Action

Integrating Public Data, AI, and Geospatial Analysis to transform how we monitor, predict, and protect our most vital hidden resource.

Prof. Tae Kwon Lee

Yonsei University, MIRAE Campus





Safe Enough?

Primary Drinking Water Source

The Core Problem:
A Hidden, Vital, Resource under Treat

Groundwater supplies drinking water to over 2 billion people globally, **yet remains invisible and poorly monitored unlike surface water**, contamination often goes undetected until irreversible damage occurs

Anthropogenic pollutants (nitrates, pharmaceuticals, PFAS) and microbial pathogens are increasingly infiltrating aquifers with residence times of decades to centuries, making remediation extremely challenging and costly

Climate change intensifies the crisis through altered recharge patterns and saltwater intrusion while growing agricultural and industrial demands accelerate depletion rates faster than natural replenishment

50%+ Of rural wells unfit for drinking

45% PFAS detected of US DW

The Importance of Groundwater: Our Invisible Lifeline

Public Health

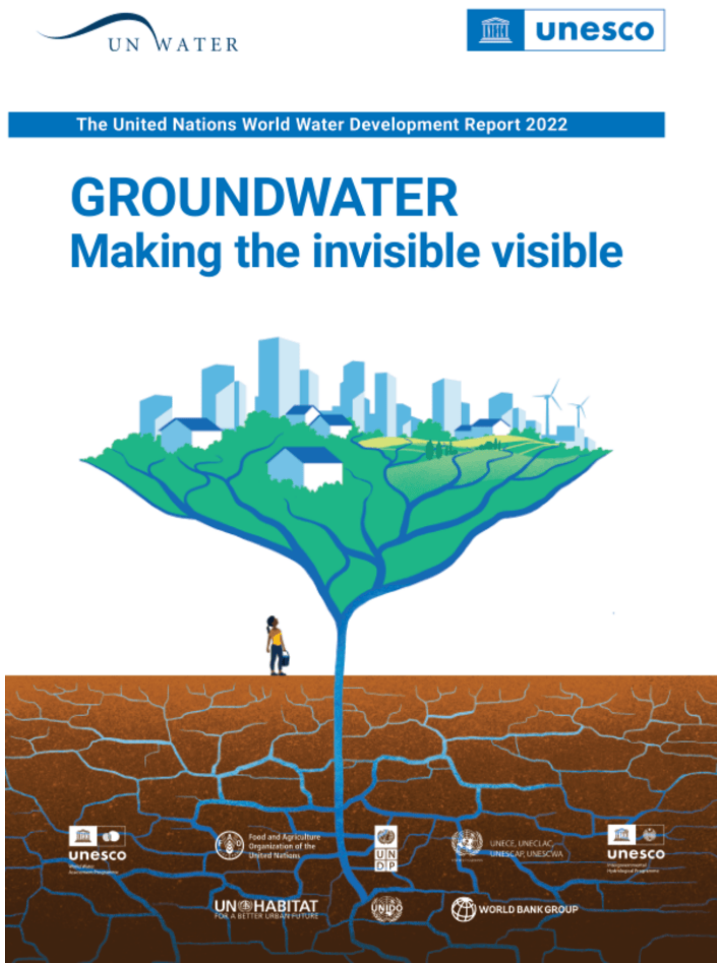
Groundwater contamination directly causes disease outbreaks and chronic poisoning among 2+ billion people who depend on it for drinking water, disproportionately affecting vulnerable populations without treatment alternatives.

Food Security

Aquifer depletion threatens crop production for 800 million food-insecure people in regions where 40% of global irrigated agriculture relies entirely on groundwater extraction.

Economic Stability

Groundwater depletion causes land subsidence, infrastructure damage, and costly industrial relocations that threaten over $3 trillion in annual economic activity dependent on stable aquifer resources.



The Challenge: Complexity and Spatial Variability

Aquifer heterogeneity makes contamination unpredictable, requiring site-specific monitoring that exceeds the capacity of conventional policy frameworks.

$100M+
Per site

Average remediation cost for contaminated groundwater

Three-dimensional flow systems interact

across multiple geological layers with different recharge rates, residence times, land-use, hydrology, and climate complicating source attribution and remediation strategy design.

Heterogeneous aquifer properties

create unpredictable contaminant transport pathways across spatial scales from meters to kilometers, making conventional monitoring networks inadequate for early detection.

Geochemical gradients vary

dramatically within single aquifers due to lithology, redox conditions, and microbial activity, requiring site-specific assessment rather than generalized management approaches.

The Limitations of Traditional Analysis

Traditional groundwater monitoring are insufficient for modern challenges. They are reactive rather than predictive without identifying broader patterns, and cannot effectively identify spatial hotspots or predict risk in unmonitored areas.

99.7% of contaminants escape routine detection

M/Q Sampling miss contamination pulses

< 1% overlooking unculturable pathogens or ARGs

01 Standard Chemical Ignore Emerging Threats

Pharmaceuticals, personal care products, pesticide metabolites, and industrial chemicals remain unmonitored despite widespread occurrence and endocrine-disrupting potential.

02 Snapshot Sampling Fails to Capture Temporal Dynamics

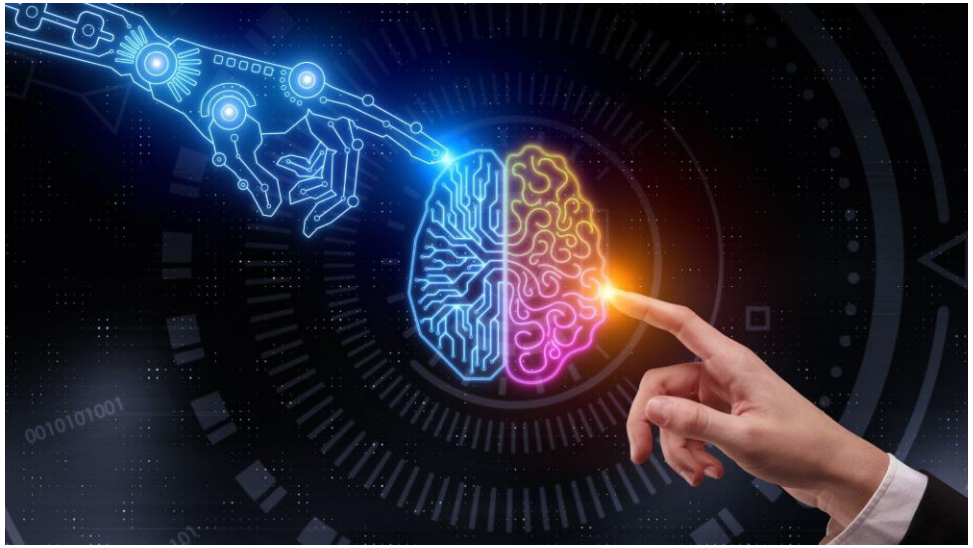
Monthly or quarterly grab samples miss contamination pulses from storm events, seasonal redox fluctuations, and episodic discharges.

03 Single-parameter Approaches Overlook Synergistic Effects

Analyzing contaminants individually ignores co-occurring mixtures where combined toxicity exceeds additive predictions and geochemical interactions amplify risks.

A New Mindset

From "Let's Use AI" to "How Can AI Help?"



01 **Define the Policy Question**

"Where should we focus our limited well inspection budget for maximum public health impact?"

02 **Identify the Analytical Need**

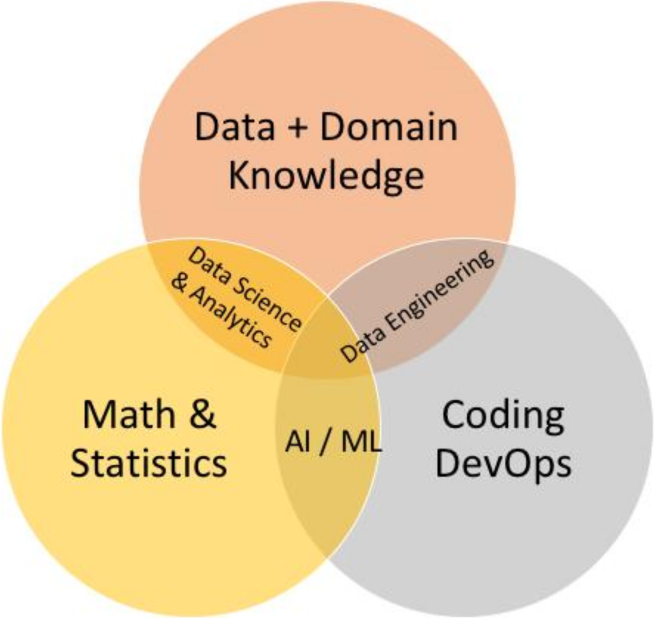
"We need a model that predicts which wells are at the highest risk of failure."

03 **Apply the Right Tool**

"AI can build a predictive risk model using land use, well age, and historical data to answer our specific question."

Problem-First Approach Succeeds: Start with policy needs, not technology possibilities. The most effective applications of AI begin with clear questions, not algorithms.

Why Field Experts Are More Important Than Ever?



AI is a Tool, Not an Oracle

Machine learning models predict contamination risk based on training data, but cannot account for unprecedented geological conditions, novel pollutant interactions, or microbial community shifts.

Policy-science gap widens without translators

Field experts bridge laboratory findings and regulatory implementation, interpreting complex data (e.g., emerging contaminants) for policymakers while ensuring scientifically-sound monitoring networks.

Risk Analysis of Co-occurring Inorganic Chemicals

Project Overview

NC State & Arizona State University collaboration spanning 140+ years (from 1875 to 2021) of historical monitoring data (20M+ data points, 50+ parameters) across North Carolina & Arizona

Key Innovation

Machine learning model predicts co-occurring inorganic pollutants (arsenic, lead, phosphorus) using limited water quality parameters enabling strategic resource allocation for monitoring.

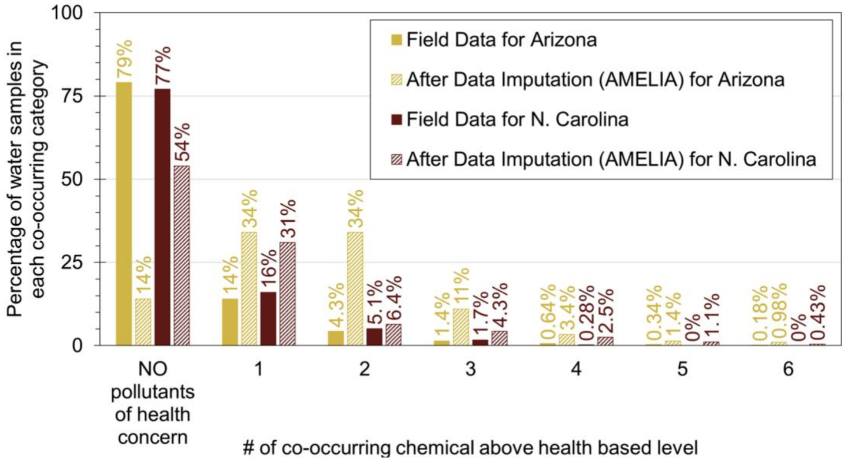
Limited Parameter Data → ML Analysis → Risk Prediction → Targeted Monitoring

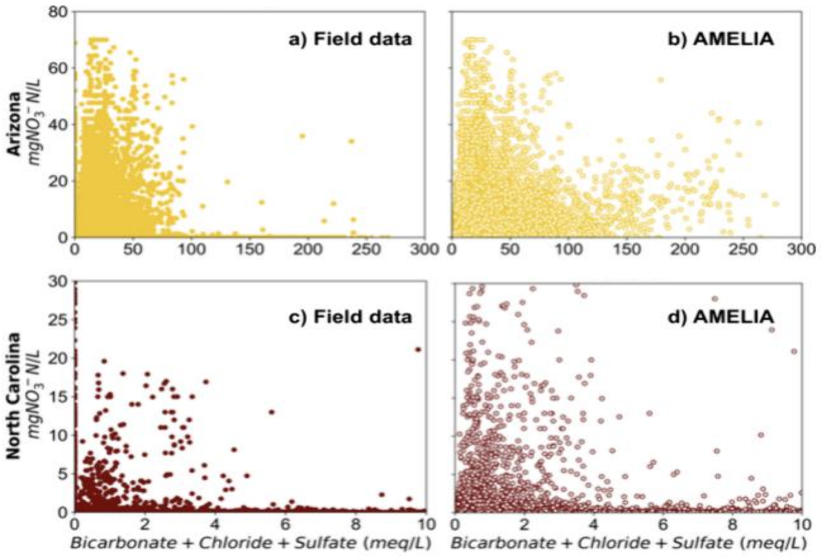
Policy Impact

Hidden Risk Revealed: Traditional monitoring showed 75-80% sites safe, ML predicts only 15-55% truly risk-free

Enhanced Detection: 2-5x increase in identified high-risk sites requiring testing

Optimized Strategy: State agencies now prioritize testing based on AI-identified hotspots





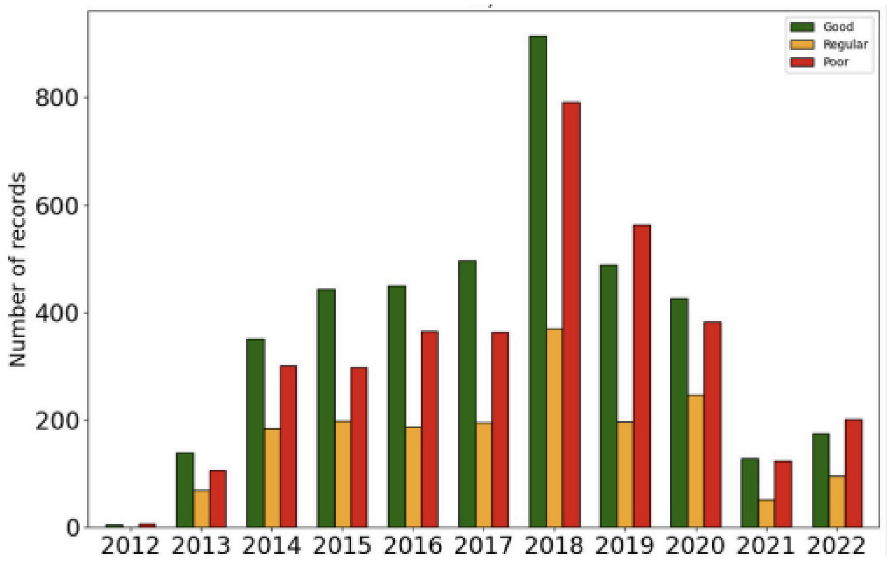
Mahmood et al., 2024, Environ Sci Technol

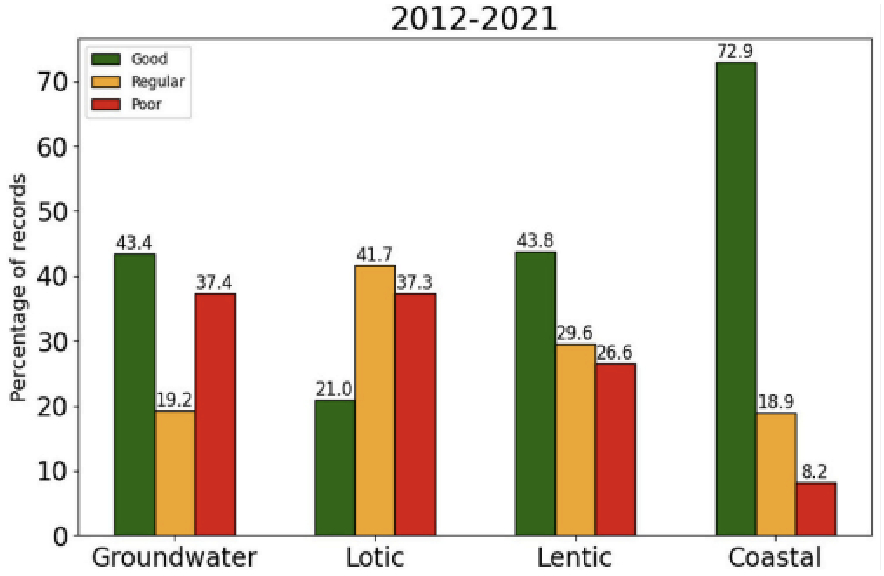
National-Scale Platform for Water Management

CONAGUA's AQuA-P platform represents the first machine learning-based national water quality assessment tool compliant with Mexican Official Norms. This comprehensive system integrates data from 2012-2021 across 13 hydrologic regions and 4 water body types including groundwater, lotic, lentic, and coastal waters.

AQuA-P integrates Decision Tree models achieving remarkable 99.3-100% for real-time water quality classification, replacing subjective Water Quality Index calculations with probability-based assessments.

100% Water Classification Accuracy





Díaz-González et al., 2025, J Cont Hydro

Our Goal:
A New Paradigm

A complete data-driven workflow that transforms groundwater management from reactive to proactive.

Can we predict the overall safety of any well in the nation?

Ahn et al., 2023, Ecotoxicology and Environmental Safety

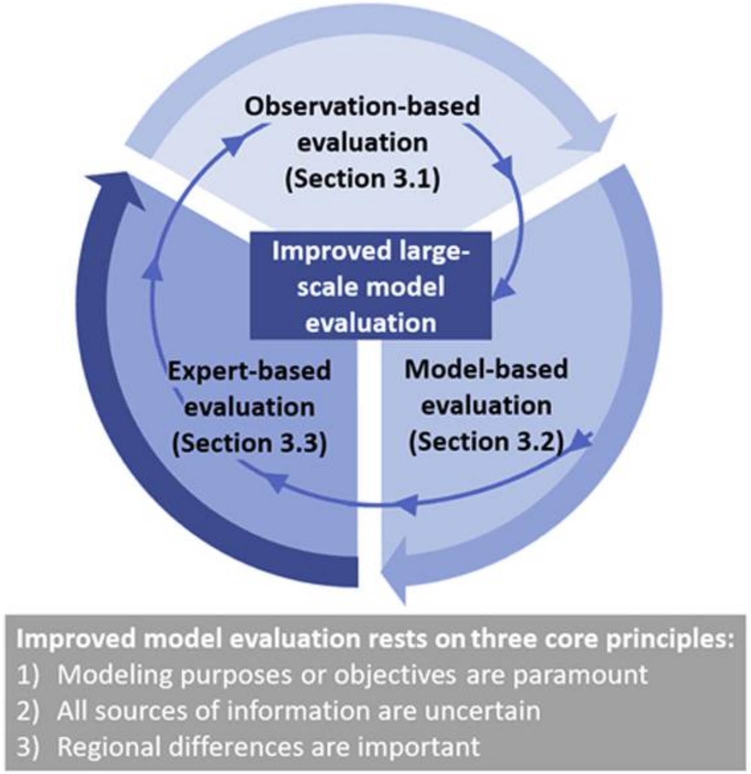
Is pollution a local issue or a regional one? Can we find the hotspots?

Jahan et al., 2025, Water Research X

Can we build reliable, and automated early-warning systems?

Youn et al., 2025, Journal of Environmental Management

CASE STUDY 1: Groundwater Quality Index



Gleeson et al., 2021, Geo Model Devel

The Challenge of Holistic Assessment

Simple pass/fail tests on individual parameters like nitrate are insufficient for true water safety. A well might pass one standard but fail on bacteria, heavy metals, or turbidity creating an incomplete picture of overall quality.

Policymakers and citizens need a comprehensive, intuitive metric that summarizes multiple contaminant risks similar to an "Air Quality Index" but for groundwater. This unified indicator would enable more effective communication, prioritization, and resource allocation.

Why Use Machine Learning for a Groundwater Quality index?

The Data Challenge

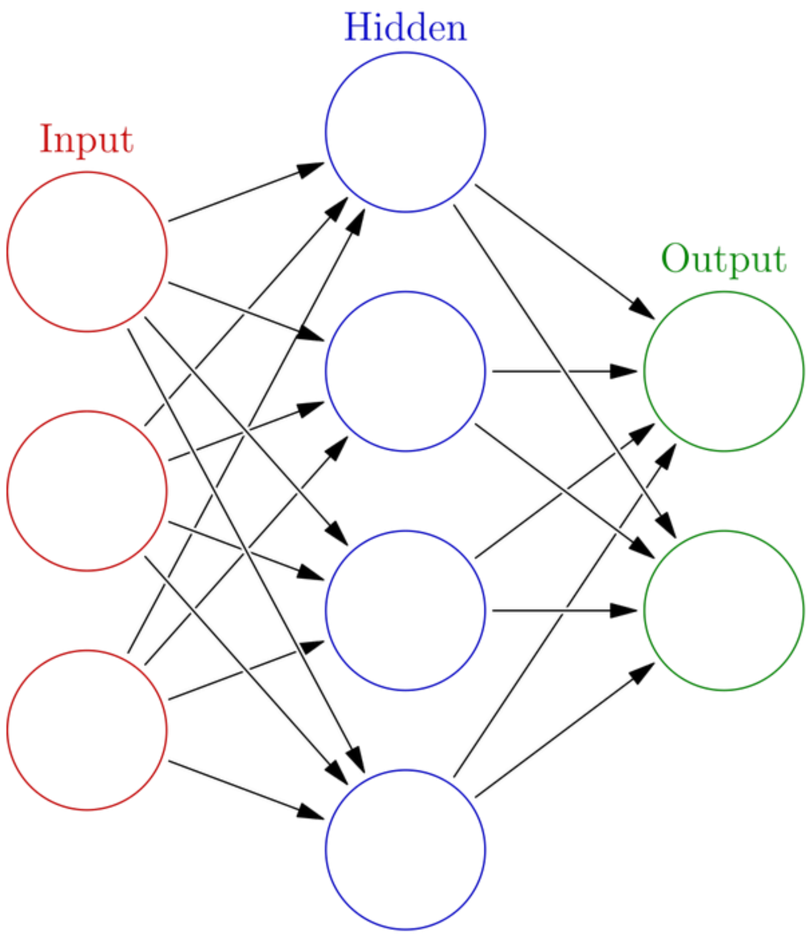
Calculating GQI for every well is extremely difficult due to missing or incomplete data. Of 8,326 wells sampled, only 3,552 had complete parameter sets for traditional GQI calculation (Ahn et al., 2023).

ML Solution

Machine learning can identify patterns between available parameters and overall quality, accurately predicting GQI grades even with partial data. This allows us to assess wells that would otherwise be impossible to classify.

Rapid Screening

By predicting risk levels from limited parameters, ML enables rapid, nationwide screening of wells, immediately flagging those at highest risk for priority sampling and intervention.



Is This GW Really Suitable for Drinking Water?

Sample	WQ indicator 1	WQ indicator 2	WQ indicator 3	Decision
A	0.2	0.1	0.1	Potable
B	0.2	1.1	0.3	Unsuitable
C	0.95	0.95	0.95	Potable

Can the water quality results analyzed from a single sampling be trusted?

Need to provide reliable information to citizens actually using groundwater as a major source of drinking water.

Dataset: The "Safe Groundwater Project"

National Institute of Environmental Research (2017-2020)

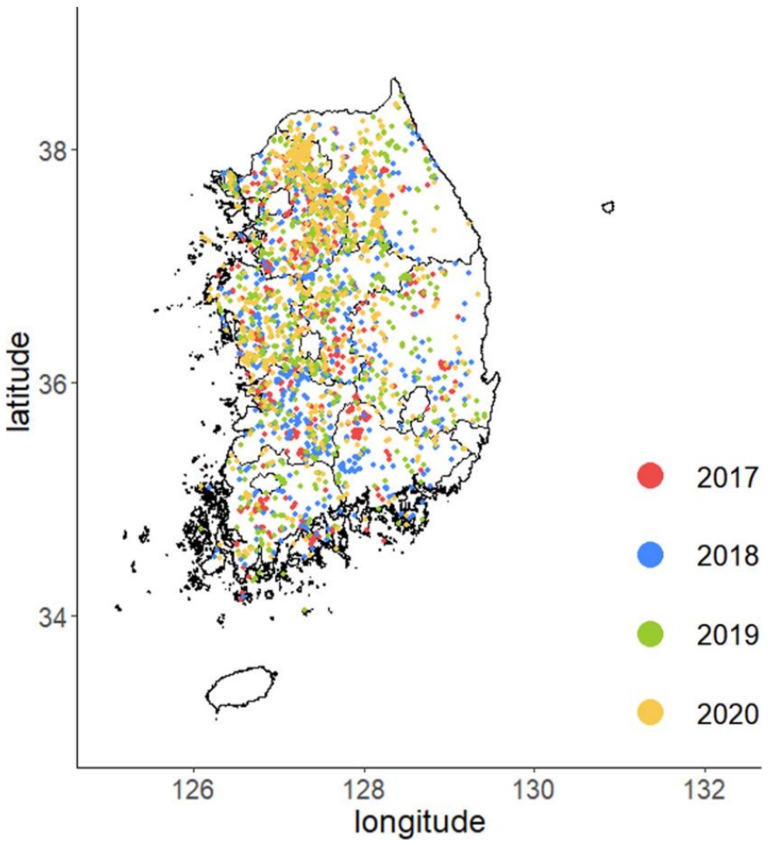
Comprehensive nationwide sampling of groundwater quality in unsupplied areas across South Korea. A massive dataset providing unprecedented insight into drinking water quality for rural communities.

Scale & Parameters

8,326 wells analyzed across four years (2017: 2,061; 2018: 2,142; 2019: 2,019; 2020: 2,104). Initially measured 47 drinking water parameters with 28 key parameters used for final modeling after data preprocessing.

Critical Finding

Over 57% of wells (4,774 out of 8,326) were found inappropriate for drinking purposes, with at least one parameter exceeding safety standards. Our GQI analysis focused on the remaining 3,552 potable wells.



Groundwater Quality Index (GQI): A Quantitative and Qualitative Assessment Framework

3 Groups

Worrisome / Good / Very Good

Convert raw parameter values into a GQI

The lower the deviation from standard values, the higher the GQI (better quality)

GQI scores differ significantly among the three quality groups

Contamination loaded by GQI groups

Worrisome wells often contaminated by ≥3 parameters, But very good wells mostly show 0–1 contamination.

GQI grades not only capture numerical scores but reflect real-world contamination burden

Data preparation

Parameter Normalization

$$P_n = \frac{P_e - P_{e\,min}}{P_s - P_{e\,min}}$$

P_e : Value for each parameter
P_s : Standard value for each parameter
P_n : Normalization value for paramter

↓

Calculate the Distance value

$$P_d = 1 - P_n$$

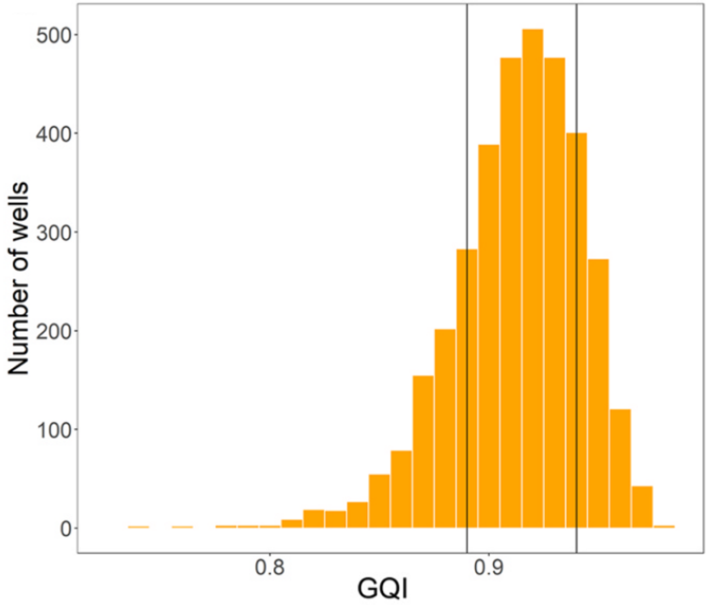
P_d : Deviation for parameter
P_n : Normalization value for paramter
1 : Standard value
∵ Normalization max value is Standard value for parameter

$$W_d = \frac{\sum_{i=1}^{x} (P_d)^2}{x}$$

W_d : Distance value for well
P_d : Deviation for paramter
x : Number of Parameter

↓

Set Grades:
Bottom 20 %: Worrisome (W_d < 0.89)
Top 20 %: Very good (W_d > 0.94)
Between : good (0.89 < W_d < 0.94)



Key Finding

The GQI effectively integrates multi-parameter deviations into a unified metric, enabling robust classification of wells into three quality tiers, and providing a quantitative yet interpretable framework for nationwide groundwater safety assessment.

Groundwater Quality Index (GQI): A Quantitative and Qualitative Assessment Framework

3 Groups

Worrisome / Good / Very Good

Convert raw parameter values into a GQI

The lower the deviation from standard values, the higher the GQI (better quality)

GQI scores differ significantly among the three quality groups

Contamination loaded by GQI groups

Worrisome wells often contaminated by ≥3 parameters, But very good wells mostly show 0–1 contamination.

GQI grades not only capture numerical scores but reflect real-world contamination burden

Data preparation

Parameter Normalization

$$P_n = \frac{P_e - P_{e\,min}}{P_s - P_{e\,min}}$$

P_e : Value for each parameter
P_s : Standard value for each parameter
P_n : Normalization value for paramter

↓

Calculate the Distance value

$$P_d = 1 - P_n$$

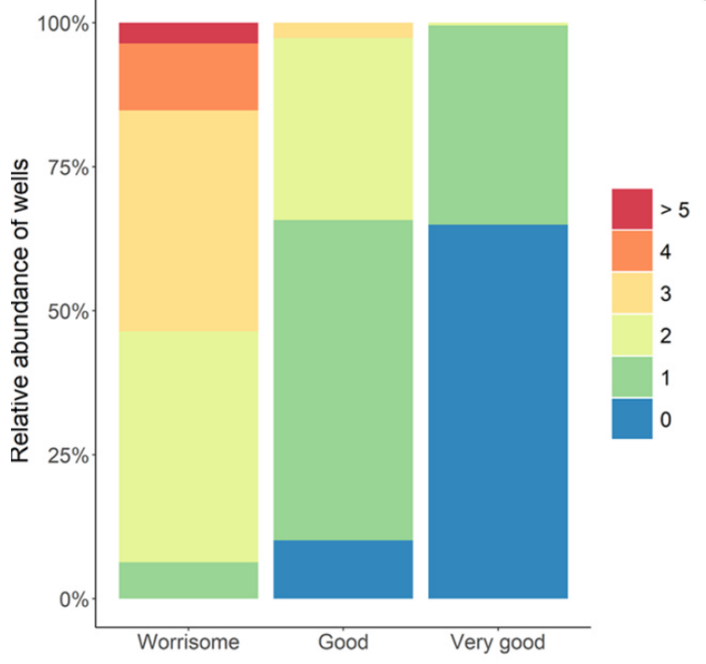
P_d : Deviation for parameter
P_n : Normalization value for paramter
1 : Standard value
∵ Normalization max value is Standard value for parameter

$$W_d = \frac{\sum_{i=1}^{x} (P_d)^2}{x}$$

W_d : Distance value for well
P_d : Deviation for paramter
x : Number of Parameter

↓

Set Grades:
Bottom 20 %: Worrisome (W_d < 0.89)
Top 20 %: Very good (W_d > 0.94)
Between : good (0.89 < W_d < 0.94)



Key Finding

The GQI effectively integrates multi-parameter deviations into a unified metric, enabling robust classification of wells into three quality tiers, and providing a quantitative yet interpretable framework for nationwide groundwater safety assessment.

AI Model Predicts Well Safety with ~99% Accuracy

98.6%

Classification Accuracy

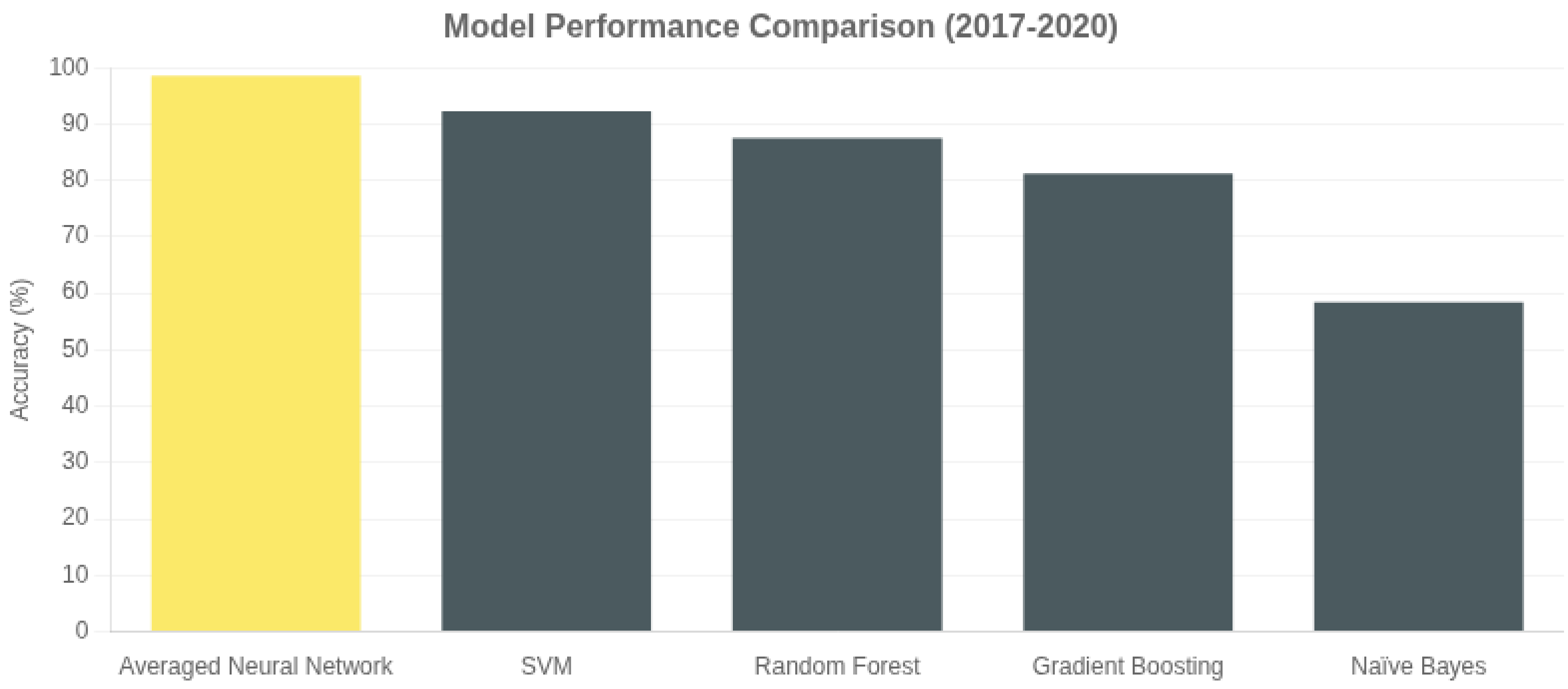
Averaged Neural Network Model

Outperformed other classification models including SVM (90%), Random Forest, and Decision Trees (<60%).

Robust Performance

Confusion matrices show minimal misclassifications across all four years (2017-2020).

In 2020: correctly identified all 163 "Worrisome" wells and 619 of 620 "Good" wells.



Key Finding

The Averaged Neural Network demonstrates exceptional stability in classifying groundwater quality across multiple years, providing a reliable tool for nationwide assessment of well safety.

A Blueprint for Cost-effective Monitoring

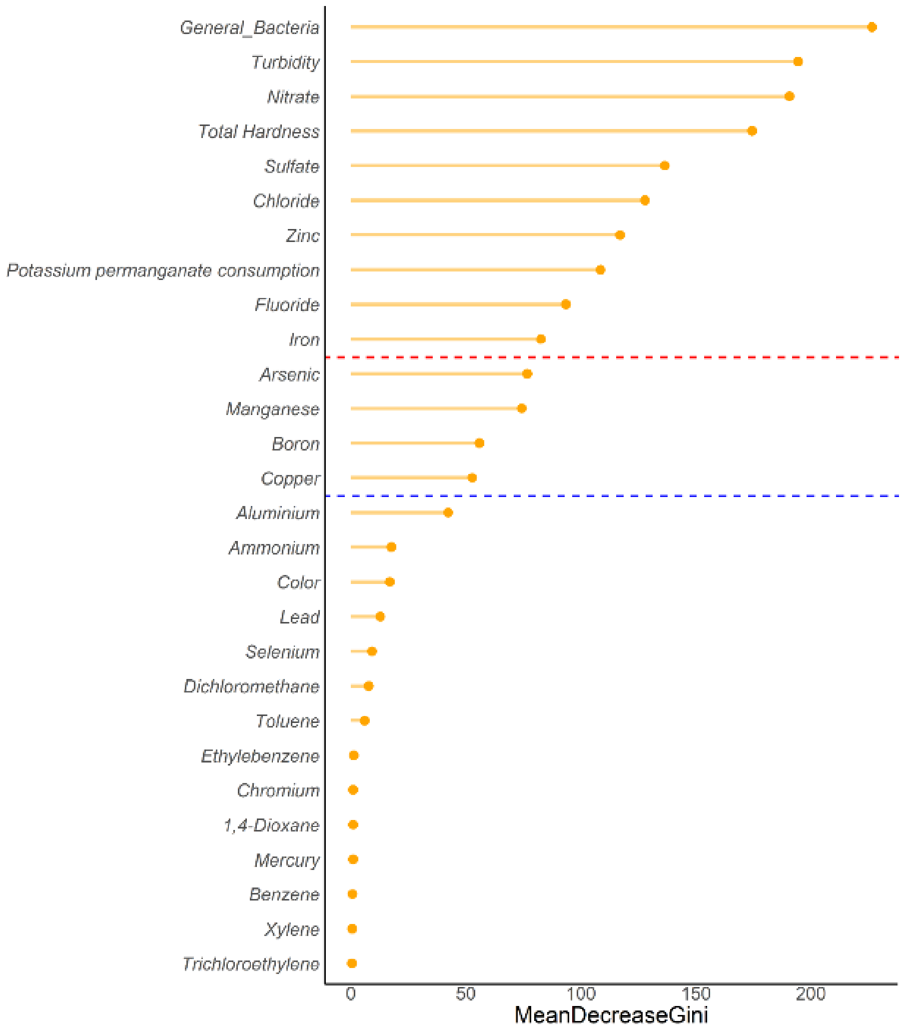
Using a Random Forest model for feature selection, we identified the most influential parameters for predicting groundwater quality. This allows for targeted, efficient monitoring while maintaining high accuracy.

90%

Accuracy with just 10 parameters

95% Accuracy with 14 parameters

Using fewer parameters drastically reduces monitoring costs while maintaining reliable quality assessment.



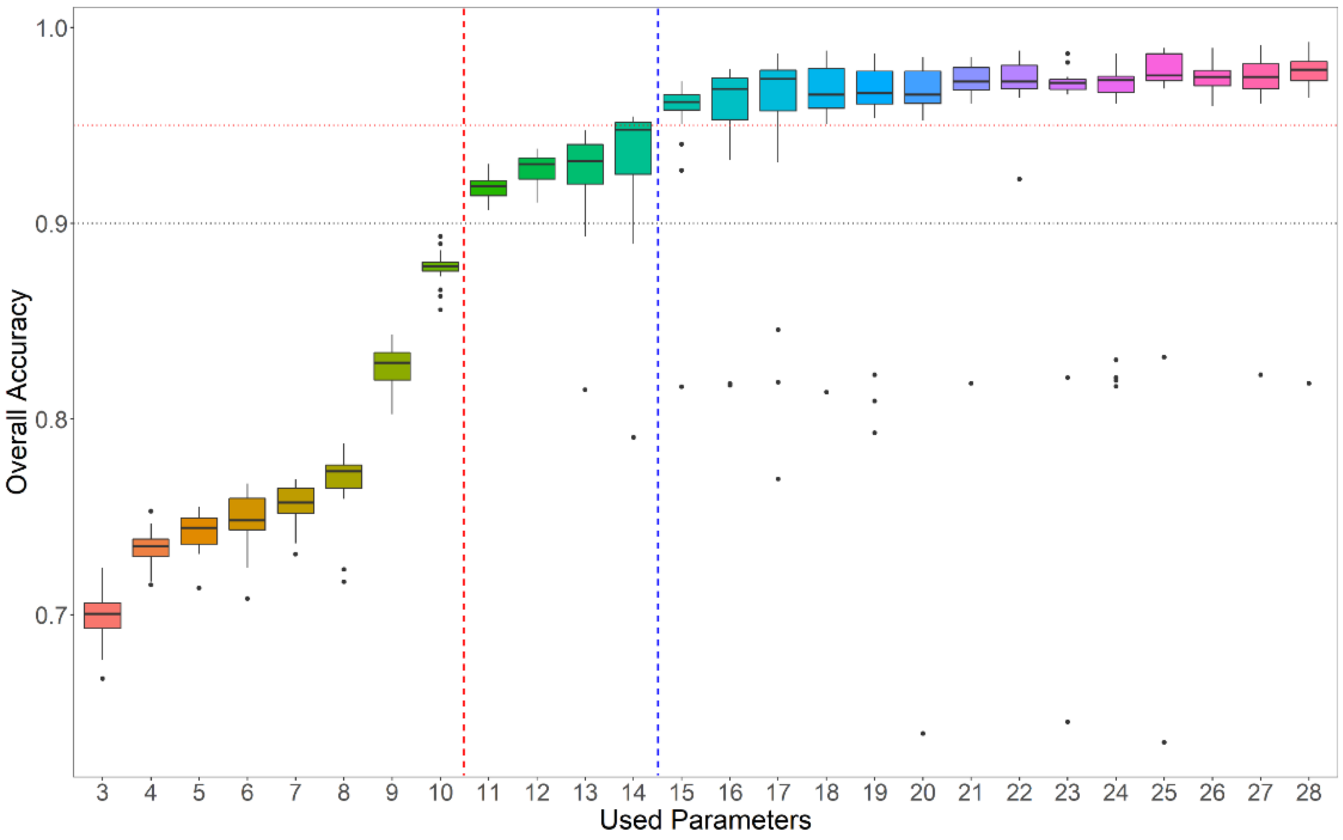
A Blueprint for Cost-effective Monitoring

Using a Random Forest model for feature selection, we identified the most influential parameters for predicting groundwater quality. This allows for targeted, efficient monitoring while maintaining high accuracy.

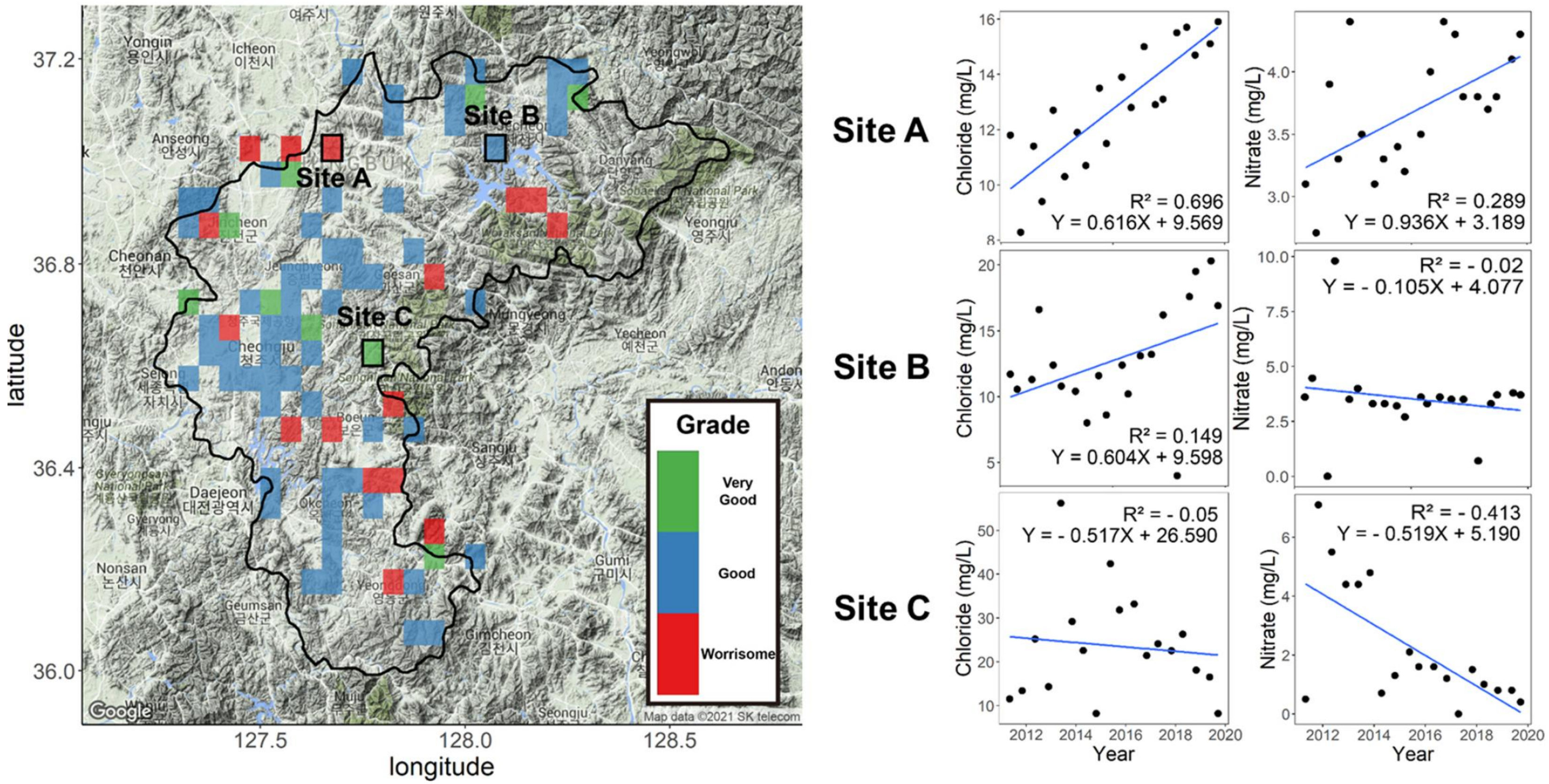
90%
Accuracy with just 10 parameters

95% Accuracy with 14 parameters

Using fewer parameters drastically reduces monitoring costs while maintaining reliable quality assessment.



From Data to Decision:
Translating GQI into Actionable Policy Insights

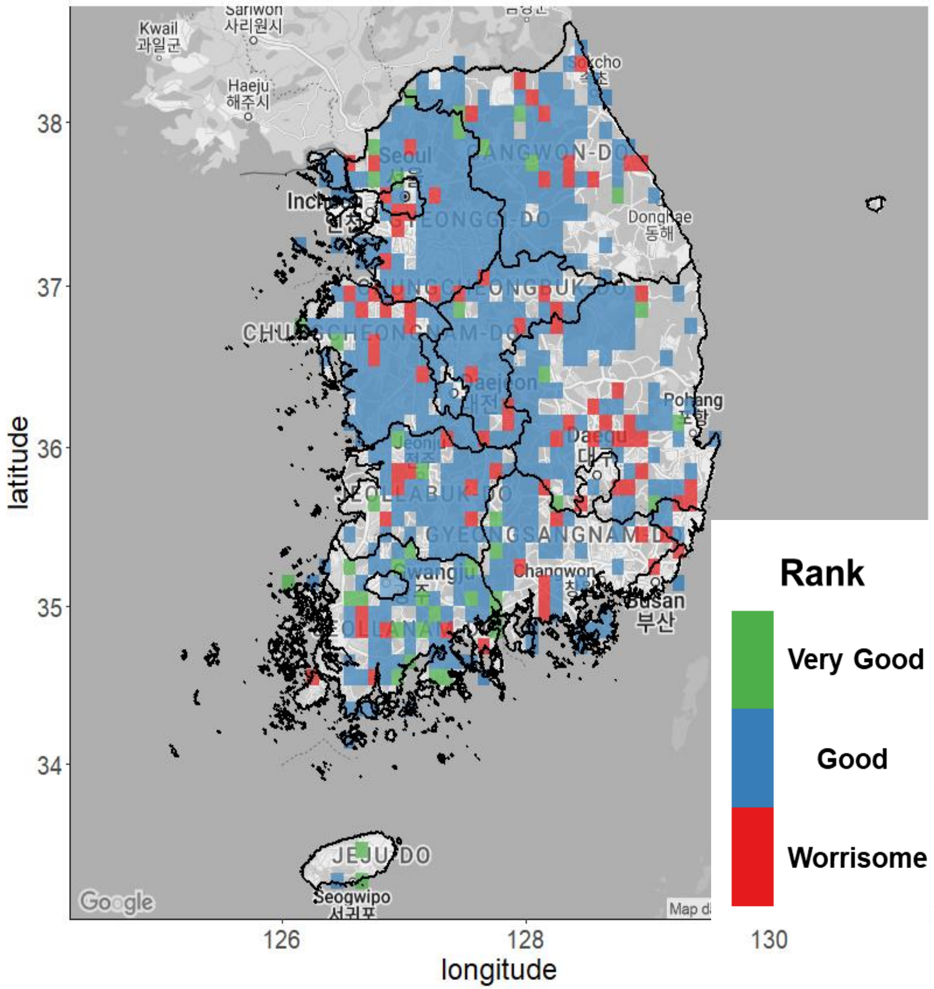


Worrisome

Good

Very Good

- By applying our GQI predictive model to thousands of wells and grouping results into administrative-level grids (a process called spatial binning), we transformed complex point data into intuitive, actionable visualization maps. Long-term monitoring data of nitrate and chloride concentrations over the past decade at representative sites (A–C) showed strong agreement with GQI-based classifications.
- The GQI-based mapping framework effectively bridges the gap between scientific modeling and environmental policy implementation by transforming high-dimensional hydrochemical data into accessible, decision-ready formats.



CASE STUDY 1: Conclusion

We have demonstrated a complete workflow to reliably and efficiently assess the overall quality of thousands of wells and visualize this risk spatially. This moves us from single-parameter checks to a holistic, predictive risk assessment.

National Risk Mapping

GQI + AI + spatial analysis provides the scientific basis for a national groundwater risk map.

Optimized Monitoring

Just 10-14 parameters can predict quality with 90-95% accuracy, reducing monitoring costs.

Enhanced Public Communication

Intuitive color-coded maps make risk levels easily understandable to citizens and officials.

CASE STUDY 2: Local VS. Regional Debate

Traditional Approach

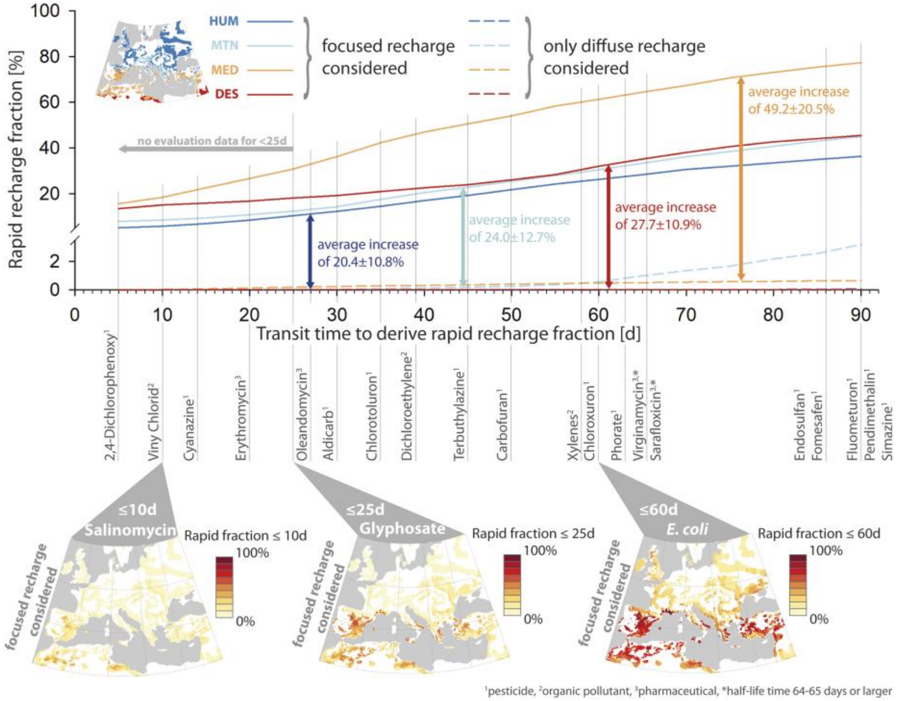
Current policy treats groundwater contamination as a localized issue that can be addressed at the individual well or farm level (focusing on point sources).

The Critical Question

Is this approach sufficient for highly mobile, agricultural contaminants like nitrate? Does contamination truly stay confined to the property where it originates, or does it form regional patterns that require broader management?

Our Research Hypothesis

We hypothesized that for diffuse pollutants like nitrate from agriculture, contamination could be transported over significant distances within the aquifer, creating large-scale, interconnected regional patterns requiring watershed-level management.



Hartmann et al., 2021, PNAS

Why Use Geospatial Analysis?

Beyond Simple Mapping

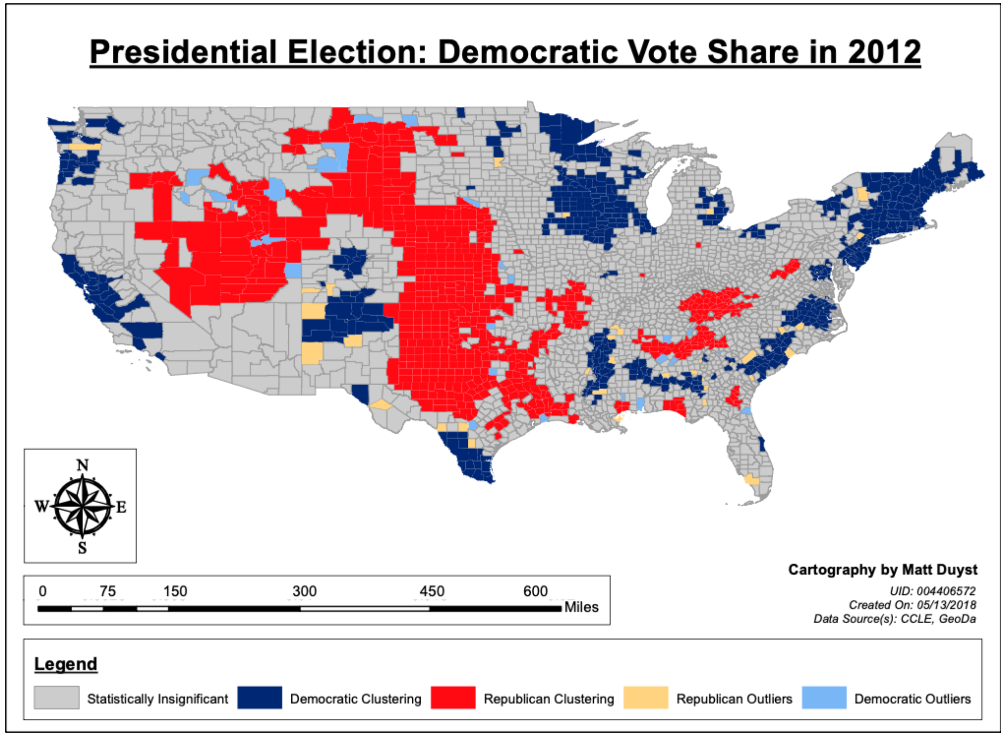
Simply mapping high nitrate values doesn't prove a connection. Clusters could be coincidental. Spatial statistics can prove whether the pattern is statistically significant or random chance.

Statistical Certainty

Tools like Moran's I and LISA (Local Indicators of Spatial Association) can mathematically verify that high-value wells are located near other high-value wells more often than random chance would allow.

Administrative Targeting

Geospatial analysis can identify precisely which administrative areas are true "hotspots" of contamination creating legally and scientifically defensible priority zones for intervention.



Spatial autocorrelation as a solution for identifying groundwater vulnerable Areas

DATASET: High-density Sampling in Chungcheongnam-do

2,349 Wells Sampled

Study Area & Scale

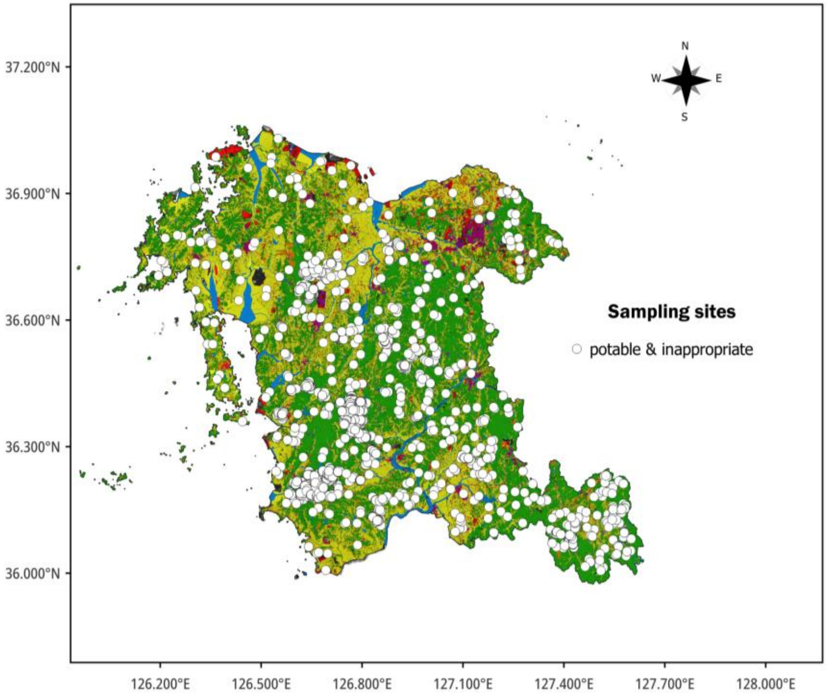
Chungcheongnam-do Province, a region with intensive agricultural activity in central South Korea. Data collection period: 2017–2022 from the "Safe Groundwater Project"

Sampling Density

2,349 groundwater wells sampled across the province, providing an unprecedented high-density dataset ideal for spatial analysis. Each well was analyzed for multiple contaminants, with Nitrate (NO3-N) as primary focus.

Contextual Data

High-resolution land-use data integrated for each well's surroundings: 46.7% forest, 30.5% cropland, and other land uses including livestock areas (1.5%), residential areas (3.8%), and water bodies (7.6%).



Capturing Diverse Land-Use Impacts

16.1% Wells exceeded nitrate standard

- Turbidity: 7.2%
- As: 4.2%
- Fe: 2.3%
- F, Al, Mn: 1.4%
- pH: 1.2%

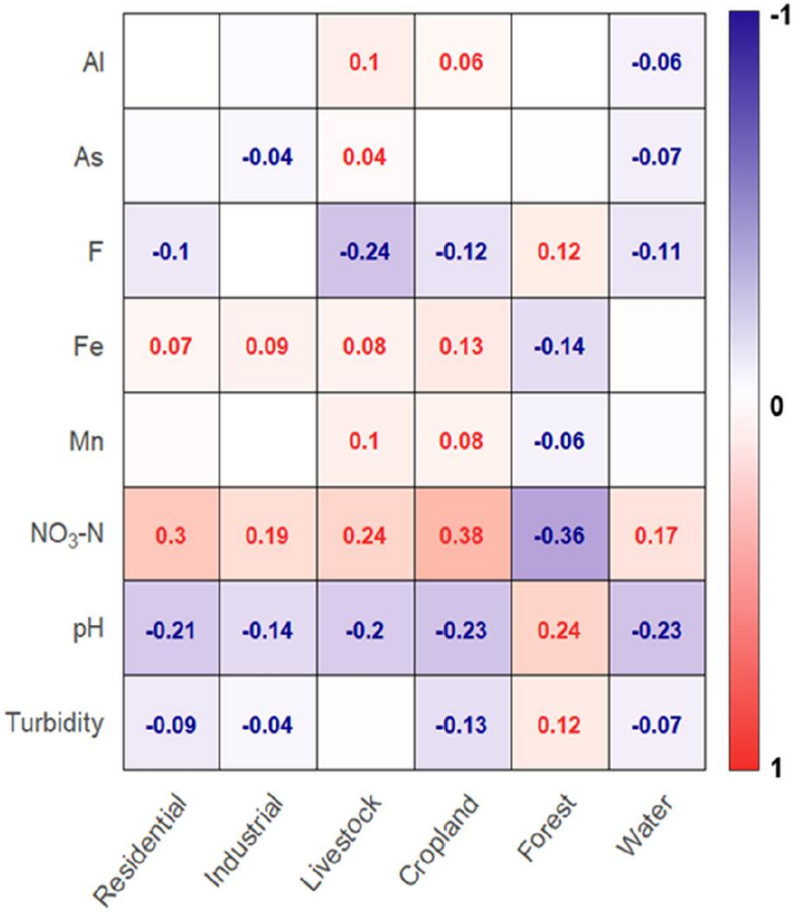
Land-use (%) in Chungcheongnam-do

46.7% Forest

30.5% Cropland

1.2% Livestock

3.8% Residential



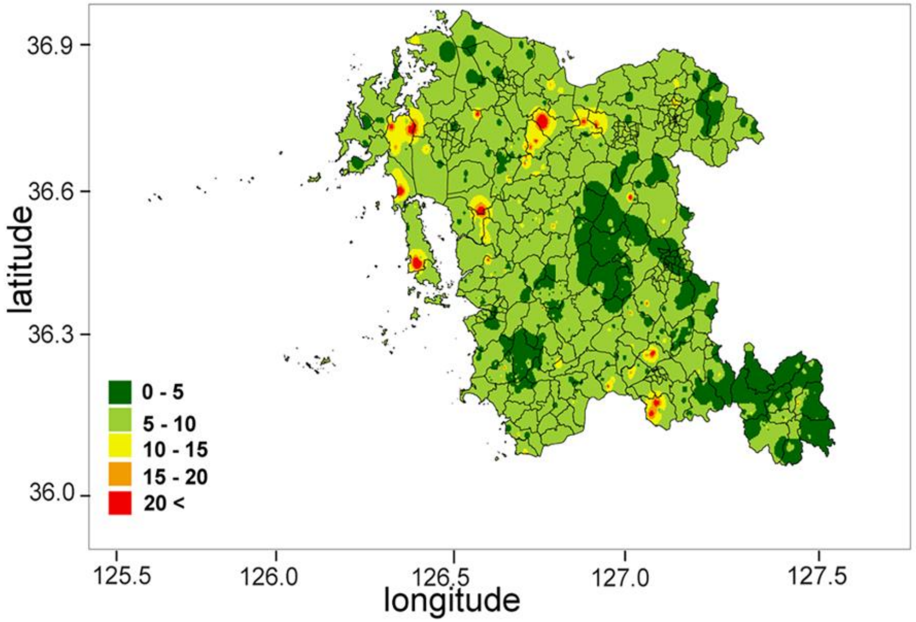
Pinpointing Hotspots with Nitrate Contamination

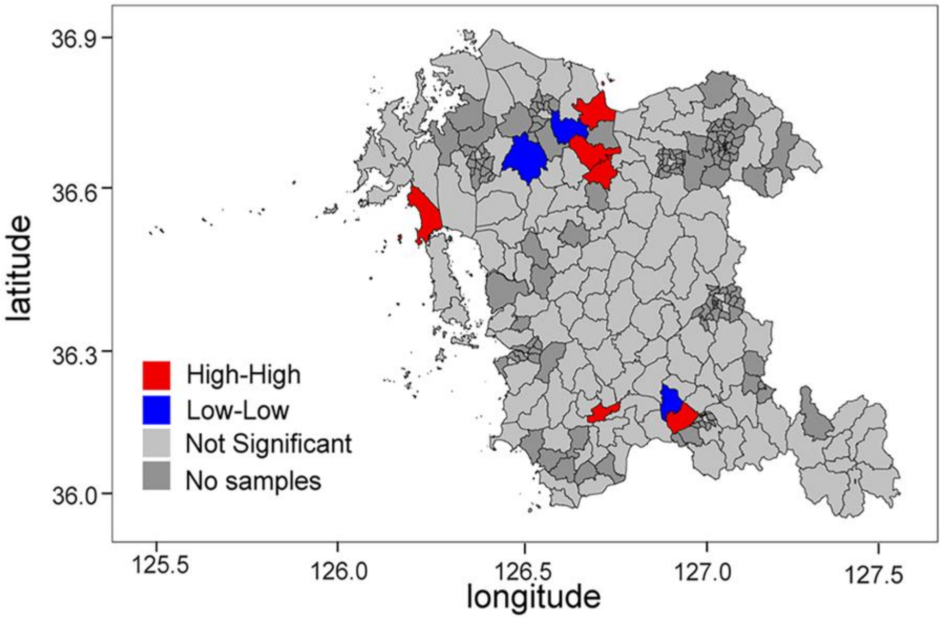
6 hotspots Safe Limit Exceeded

LISA (Local Indicators of Spatial Association) analysis identified 6 specific agricultural townships as statistically significant "High-High" nitrate hotspots—proven epicenters of contamination.

In Hapdok-eup, mean nitrate concentrations reached 64.7 mg/L—over six times the safe drinking water limit. NO_3^--N levels ranged from 59.5 to 70 mg/L, reflecting intensive fertilizer and livestock inputs.

These hotspots exhibited 68% cropland coverage compared to just 33% in non-hotspot areas. This statistical certainty provides clear, defensible targeting for regulatory action and monitoring resources.





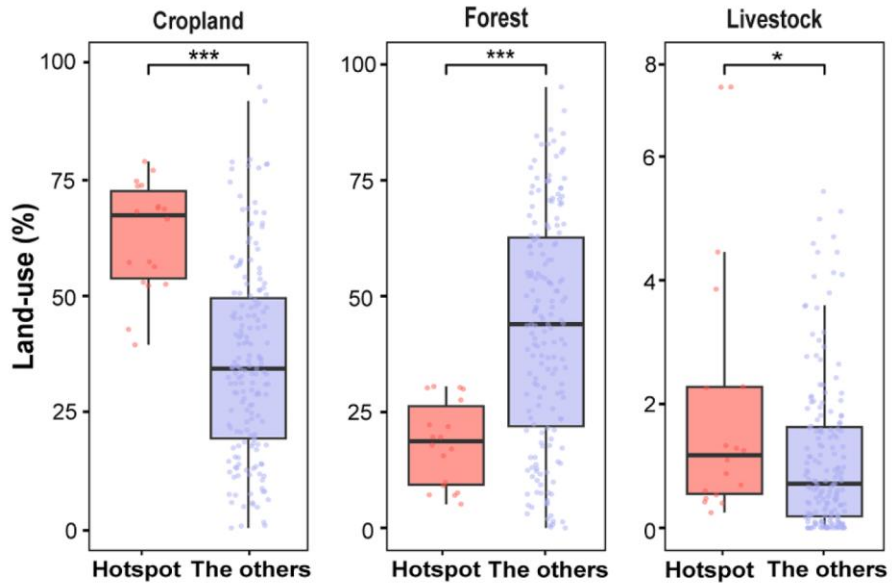
Pinpointing Hotspots with Nitrate Contamination

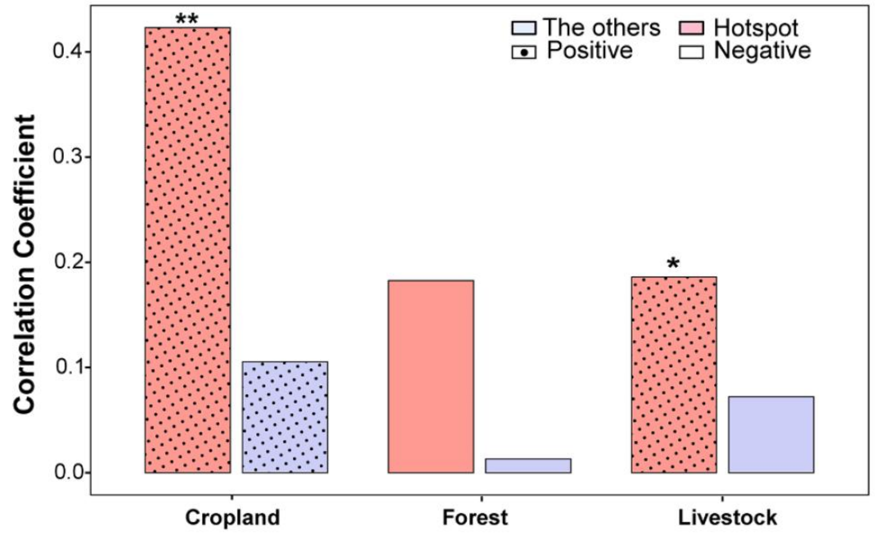
6 hotspots Safe Limit Exceeded

LISA (Local Indicators of Spatial Association) analysis identified 6 specific agricultural townships as statistically significant "High-High" nitrate hotspots—proven epicenters of contamination.

In Hapdok-eup, mean nitrate concentrations reached 64.7 mg/L—over six times the safe drinking water limit. NO_3^--N levels ranged from 59.5 to 70 mg/L, reflecting intensive fertilizer and livestock inputs.

These hotspots exhibited 68% cropland coverage compared to just 33% in non-hotspot areas. This statistical certainty provides clear, defensible targeting for regulatory action and monitoring resources.





Nitrate Contamination is Regional, not Local

Spatial autocorrelation analysis reveals that nitrate contamination forms extensive, interconnected networks spanning far beyond individual farms or wells.

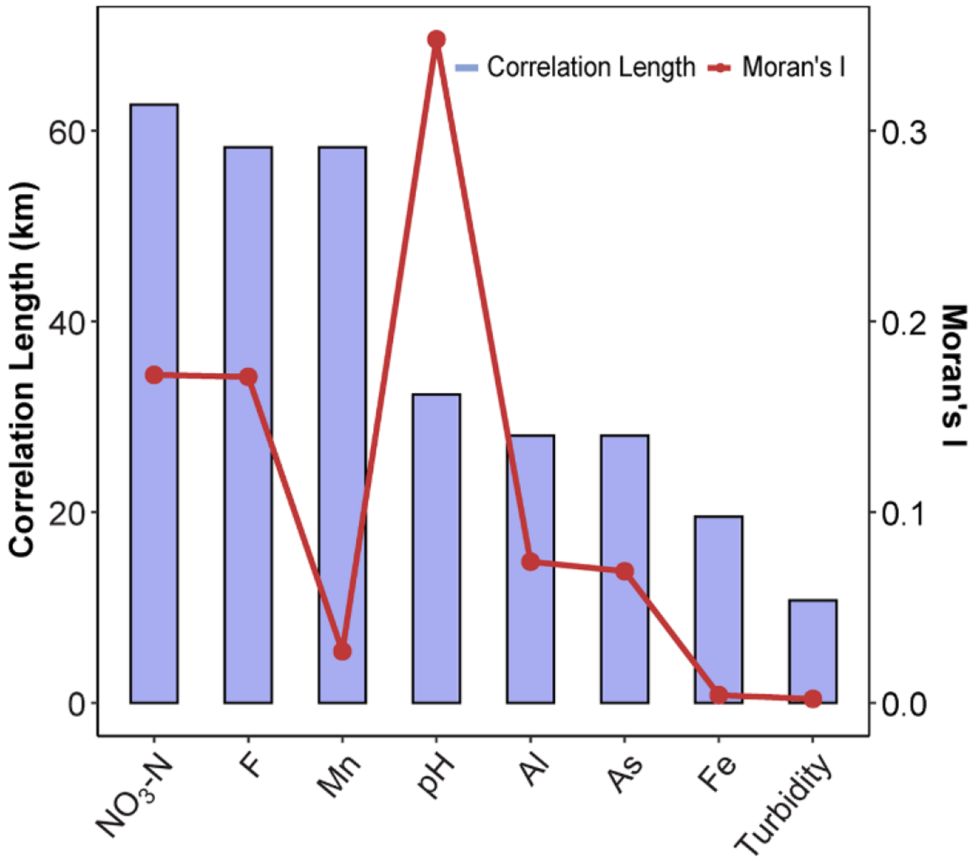
62 km Spatial coherence range for nitrate contamination

Local Contaminants

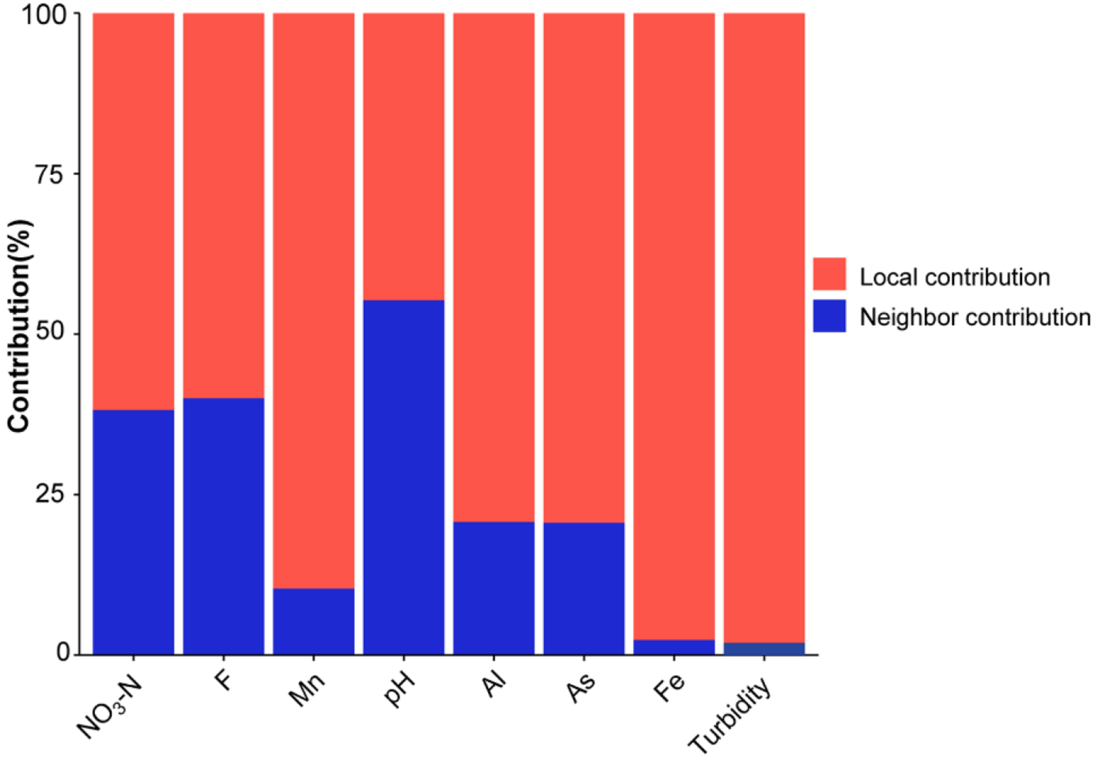
Iron, Arsenic, Aluminum
Spatial range: <20 km
Point-source control effective

Regional Contaminants

Nitrate, Fluoride, Manganese
Spatial range: up to 62 km
Requires watershed-scale management



Quantifying Hidden GW Interconnection



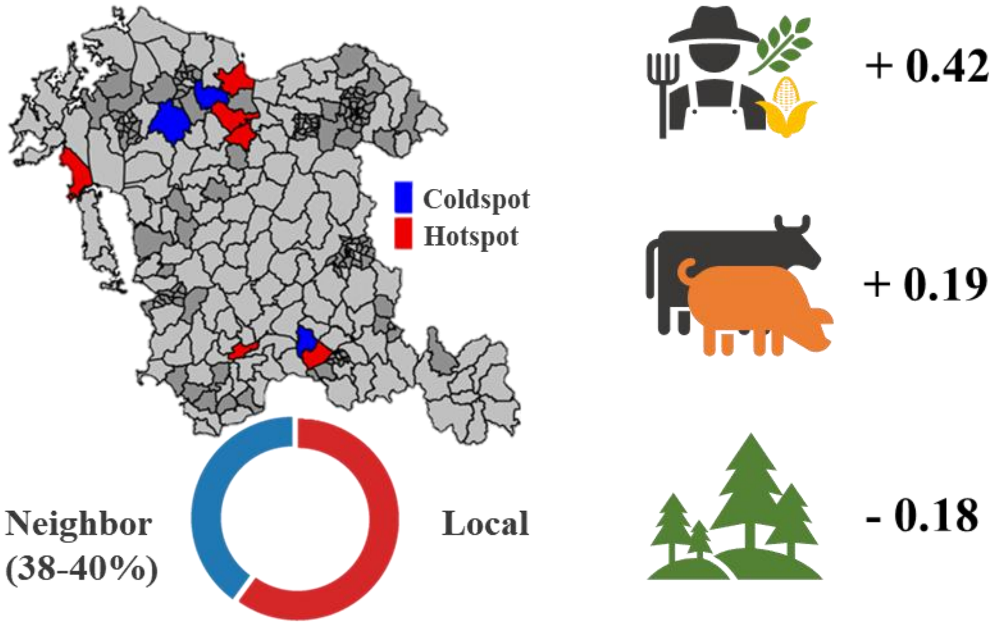
30-40% nitrate Neighbor Influence

Understanding Well Connectivity

Our novel variance partitioning analysis quantified the regional influence on nitrate concentrations. For any given well, 38-40% of its water quality variation is explained by neighboring wells, mathematical proof of extensive hydraulic interconnection across the aquifer.

This stands in stark contrast to trace elements like iron, manganese, and aluminum, where over 80% of the variation is attributed to localized factors (geology, point sources).

The significant neighbor contribution for nitrate mirrors that of pH (55%) and fluoride (38-40%), all indicating widespread agricultural impacts that cannot be addressed on a well-by-well basis.



62km Spatial Coherence

CASE STUDY 2
Conclusion

Our research has transformed understanding of groundwater contamination from local to regional phenomena, necessitating a fundamental shift in management approach.

Nitrate pollution cannot be contained locally

Regional-scale contamination requires watershed-level coordination

Need watershed-scale management

Well-by-well remediation is insufficient when 38-40% of variation comes from neighboring wells

Targeted regulation in hotspot townships

Focus on the six identified agricultural hotspots for maximum policy impact

CASE STUDY 3: The Need for Early Warning System

Time-Critical Monitoring

In high-risk areas, we identified or near potential point sources like carcass burial sites, contamination events can happen rapidly. Traditional monthly or quarterly laboratory sampling is far too slow to catch contamination events before public exposure occurs.

Sensor Reliability Challenge

In-situ sensors are the logical solution for continuous monitoring, but they are notoriously unreliable. Raw sensor data suffers from noise, drift, and calibration issues that make it unsuitable for regulatory decisions or public health alerts without sophisticated processing.

AI as the Solution

Can artificial intelligence transform unreliable sensor data into trustworthy, actionable information for early warning systems?

High-frequency Sensor Data From Anseong

Environmental Test Site

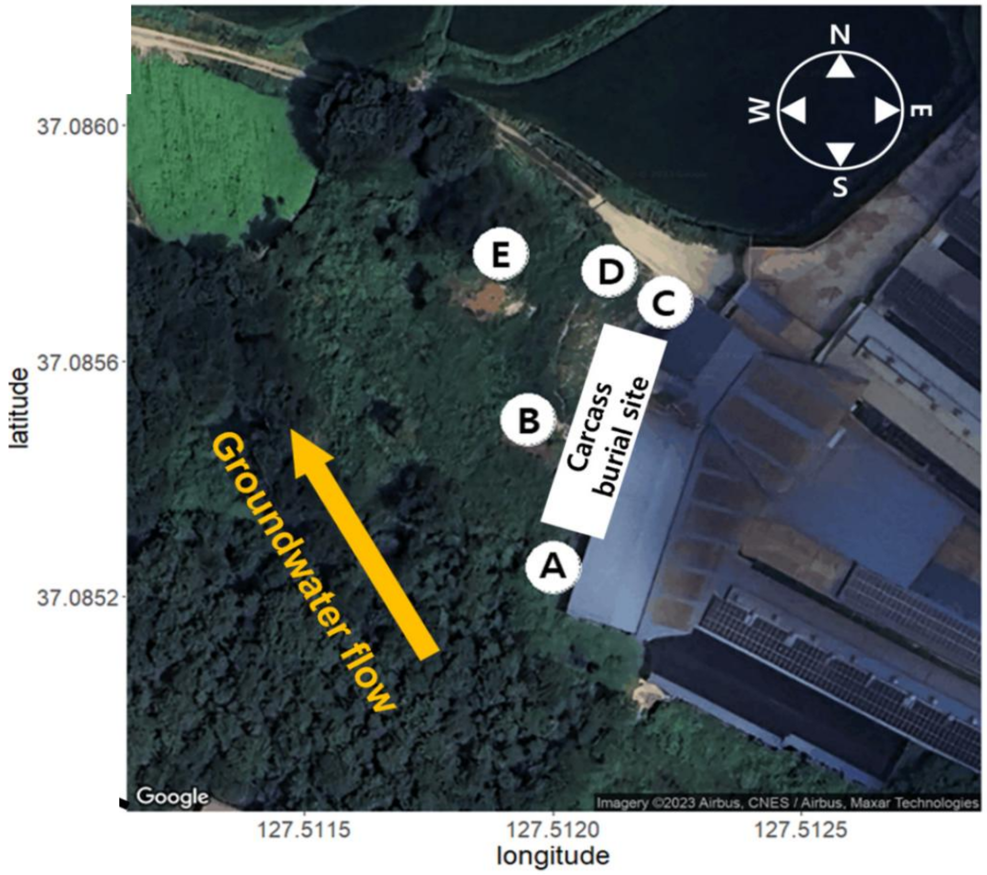
Data collected from a carcass burial site in Anseong, South Korea, a location where rapid detection of potential groundwater contamination is critical for public health protection and environmental monitoring.

High-Resolution Monitoring

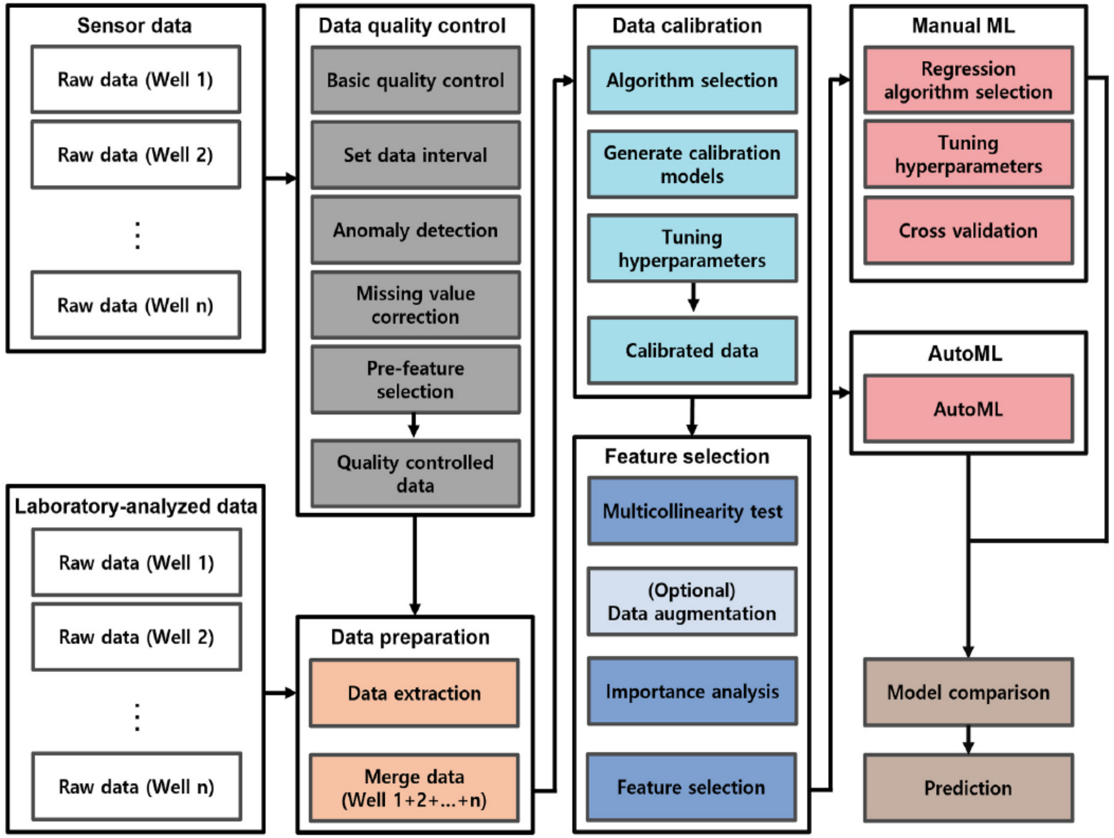
Five strategically placed groundwater monitoring wells equipped with Aqua TROLL 500 multiparameter sondes, collecting data at ~10-minute intervals from December 2018 to February 2023 (over 4 years of continuous monitoring).

Comprehensive Parameter Set

12 key parameters monitored: NH_3-N, NO_3-N, EC, Temperature, pH, DO, Turbidity, Chloride, Water Level, ORP, TDS, and Salinity. Bi-weekly laborator· analyses provided reference values for sensor calibration.



Why Use AutoML?



01 Complex Sensor Challenges

Raw sensor data suffers from noise, drift, and inaccuracies. Manual calibration for each sensor and parameter is complex and not scalable.

02 Automated Solution

ML calibration models correct sensor data using lab samples. AutoML tests hundreds of model types and parameter combinations to find optimal solutions.

03 End-to-End Automation

The framework handles data cleaning, feature selection, model optimization, and deployment, creating a hands-off, reliable system adaptable to any site.

Case Study 3 demonstrates a complete pipeline from noisy sensor input to reliable real-time predictions with minimal human intervention !

Taming Noisy and Unreliable Signals

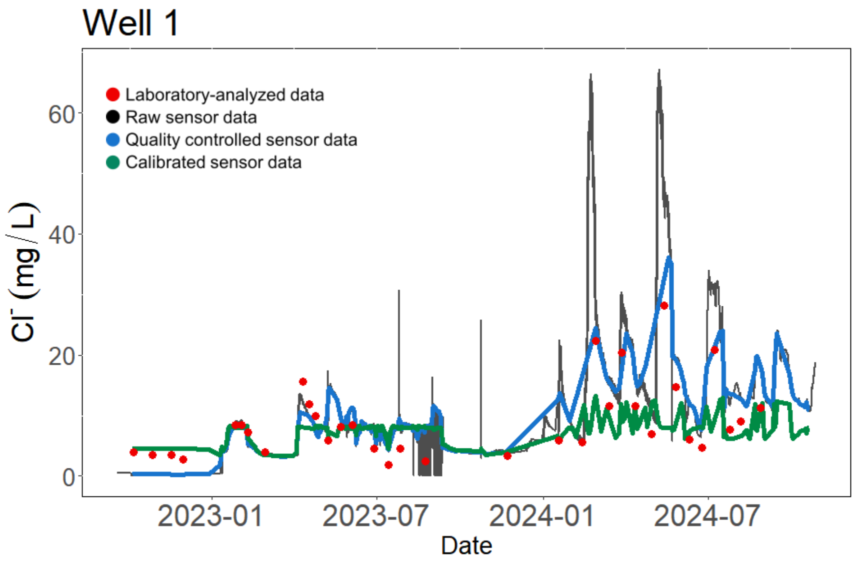
The Challenge of Raw Sensor Data

Before calibration, groundwater sensor data exhibited significant issues that rendered it unsuitable for direct analysis:

- Substantial noise and erratic fluctuations
- Sudden spikes and outliers throughout time series
- Long-term drift causing baseline shifts (e.g., NH3-N RMSE: 487.3 mg/L)
- Frequent data gaps requiring intelligent interpolation

Results After Processing

- NH3-N RMSE improved from 487.3 to just 0.9 mg/L
- Data closely tracks verified laboratory measurements
- Eliminated sensor drift and systematic errors
- Created reliable foundation for predictive modeling



Parameter	Unit	RMSE		Improvement
		Quality controlled data	Calibrated data	
Cl^-	mg/L	105.0	13.3	91.7
Depth	m	2.7	0.7	2.0
EC	μS/cm	307.3	116.2	191.1
NH_3-N	mg/L	487.3	0.9	486.4
NO_3-N	mg/L	16.7	3.6	13.1
ORP	mV	215.2	43.5	171.7
pH	–	0.8	0.3	0.5
Salinity	–	0.10	0.06	0.04
TDS	ppt	393.8	73.1	320.7
Turbidity	NTU	89.6	45.6	44.1

AI Framework Dramatically Improves Accuracy

Our end-to-end framework integrates data preprocessing, rigorous sensor calibration, and AutoML optimization to significantly enhance NH_3-N prediction accuracy in real-time. The comprehensive approach delivers substantially better performance metrics compared to raw sensor readings alone.

Target	Region	Model		Features	Number of features	RMSE	MAE	R^2
		Data type	Algorithm					
NH_3-N	Anseong	Laboratory-analyzed	ML (SVM Radial)	EC, NO_3-N, TDS, depth	4	1.23	0.77	0.58
		Laboratory-analyzed	AutoML (Randomforest)	EC, NO_3-N, TDS, depth	4	0.77	0.52	0.72
		Calibrated	ML (Randomforest)	NH_3-N, EC, temperature, Cl^-	4	0.84	0.57	0.76
		Calibrated	AutoML (ExtraTrees)	NH_3-N, EC, temperature, Cl^-	4	0.65	0.4	0.77
		Calibrated + Augmented	ML (Randomforest)	NH_3-N, EC, temperature, Cl^-	4	0.55	0.32	0.87
		Calibrated + Augmented	AutoML (Randomforest)	NH_3-N, EC, temperature, Cl^-	4	0.38	0.23	0.9
		Calibrated	ML (Randomforest)	EC, temperature, TDS, depth, pH, turbidity, Cl^-	7	1.11	0.83	0.76
		Calibrated	AutoML (ExtraTrees)	EC, temperature, TDS, depth, pH, turbidity, Cl^-	7	0.75	0.53	0.72
		Calibrated + Augmented	ML (Randomforest)	EC, temperature, TDS, depth, pH, turbidity, Cl^-	7	0.6	0.37	0.84
		Calibrated + Augmented	AutoML (Randomforest)	EC, temperature, TDS, depth, pH, turbidity, Cl^-	7	0.4	0.25	0.89
	Hoengseong	Calibrated + Augmented	AutoML (Randomforest)	EC, NH_3-N, temperature, TDS	4	0.008	0.004	0.89
	Uijeongbu	Calibrated + Augmented	AutoML (Randomforest)	NO_3-N, DO, pH, Cl^-, ORP	5	0.02	0.007	0.98

CASE STUDY 3: Conclusion

From Data Noise to Reliable Alerts: Robust & Transferable framework

Consistent Performance

Our automated framework successfully transforms unreliable sensor data into validated, high-confidence risk detection. This enables true real-time monitoring that can detect contamination events before public exposure occurs.

Minimal Recalibration

Once deployed, the system requires minimal site-specific adjustments, making it economically feasible for wide-scale implementation across multiple regions and monitoring networks.

Diverse Applications

Beyond NH_3-N, the framework demonstrates adaptability to other contaminants (NO_3-N, TPH), enabling comprehensive monitoring of various groundwater threats from different sources.

AI-Driven Groundwater Quality Integrating Spatial Data for Groundwater Management

Predictive Gap-Filling

Co-occurring Contaminant Prediction

Challenge: Sensors cannot measure all parameters simultaneously

Approach: ML models predict unmeasured inorganic (heavy metals) and organic (TPH) chemicals

Impact: Reduces monitoring costs while expanding contaminant surveillance coverage

Early Warning Systems

Disaster Risk Forecasting

Challenge: Groundwater contamination events often detected too late

Approach: Spatiotemporal ML models identify pre-disaster warning signals

Impact: Enables preemptive response before public health crises emerge

Probabilistic Risk Mapping

Contamination Probability Assessment

Challenge: Limited resources require strategic monitoring allocation

Approach: Bayesian spatial models estimate site-specific contamination probabilities

Impact: Guides optimal placement of monitoring wells and sampling frequency

Long-term Vulnerability Projections

Regional Hotspot Identification

Challenge: Unknown which regions face highest future contamination risk

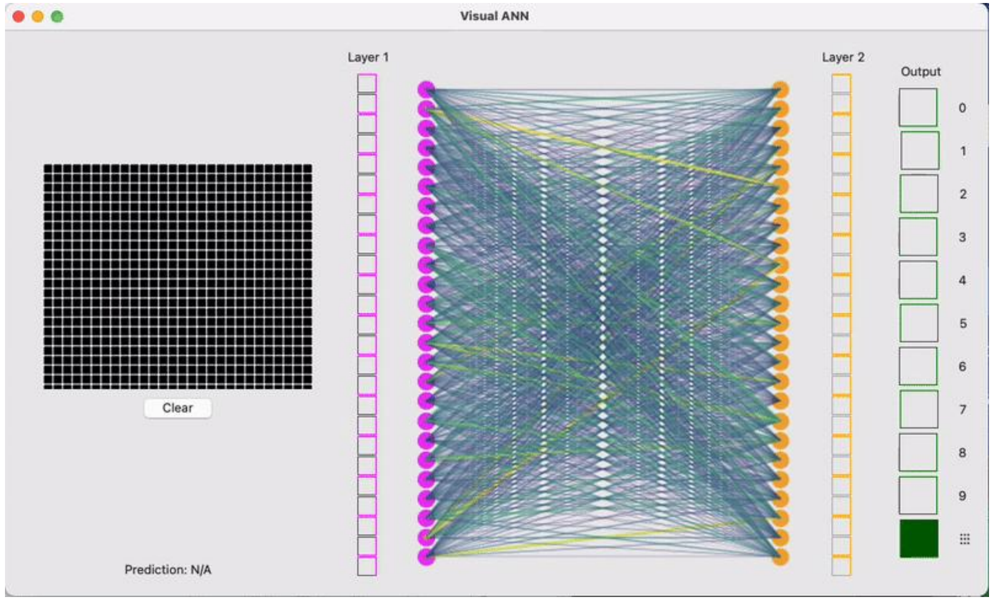
Approach: Ensemble ML + GIS predict region-specific vulnerability over 10-30 years

Impact: Informs preventive zoning policies and aquifer protection strategies

Key Distinguishing Features:

- Multi-scale analysis (well → aquifer → regional → national)
- Multi-contaminant framework (inorganic + organic + emerging)
- Multi-temporal prediction (real-time + seasonal + decadal)

A New Paradigm For Groundwater Policy



From Reactive to Proactive

Old Paradigm:

Static, calendar-based sampling with generalized regulations and response after exposure has occurred. One-size-fits-all approach regardless of risk profiles.

New, Data-Driven Paradigm:

Dynamic & Targeted Monitoring: AI-predicted risk areas receive priority attention.

Evidence-Based Planning: Hotspot maps enable spatially-explicit regulations.

Preventative Interventions: Real-time alerts prevent exposure before it happens.

Key Takeaways For A Data-driven Future

Untapped Potential

South Korea's public groundwater data is a vast, underutilized national asset. Our research shows it holds the key to smarter, more efficient management when properly analyzed and integrated across agencies and disciplines.

Transformative Technology

The combination of AI and Geospatial Analysis is not a future concept; it is a proven, practical toolkit that unlocks the potential of this data, turning it into predictive, actionable intelligence for policymakers.

Complete, Actionable Workflow

We have presented an end-to-end workflow: from national-scale risk assessment, to pinpointing regional hotspots, to monitoring them in real-time. This provides a clear path for implementation that can be adopted today.

Thank You

Questions and Discussions Welcome

tklee@yonsei.ac.kr

Department of Environmental and Energy Engineering

Yonsei University, MIRAE Campus, Wonju, South Korea