Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Choosing groundwater quality testing methods in low-income settings: field kits, national and accredited international laboratories compared in Lao PDR

Paul Pavelic
paul.pavelic@flinders.edu.au
Flinders University
Eddie Banks
Flinders University
Okke Batelaan
Flinders University
Somphasith Douangsavanh
International Water Management Institute, Lao People's Democratic Republic
Sithong Kathsana
Department of Water Resources, Lao PDR
Keoduangchai Keokhamphui
National University of Laos
Latsamee Kinglattana
National University of Laos
Saskia Noorduijn
Flinders University
Keomanivone Phuthakhanty
Department of Water Resources, Lao PDR
Lammone Sayavong
Department of Water Resources, Lao PDR
Manivanh Suyavong
Gender Development Association, Lao People's Democratic Republic
Phaphone Sivongxay
Gender Development Association, Lao People's Democratic Republic
Simone Stewart
Flinders University
Mathieu Viossanges
International Water Management Institute, Lao People's Democratic Republic
Sounipha Xaiyakeo
Department of Water Resources, Lao PDR
Ounakone Xayviliya
Department of Water Resources, Lao PDR

Research Article

Keywords: Groundwater quality monitoring, field-based testing, method comparison, resource-constrained settings, Lao PDR

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11804597/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Groundwater monitoring in low-income countries often relies on field-based tools due to high costs and limited laboratory access, yet systematic evidence comparing field methods against professional laboratories remains scarce. This study evaluates five testing approaches: in-situ sensors, test strips, field kits, a national laboratory, and accredited international laboratories, across 20 groundwater sites in Lao PDR, with both laboratory approaches compared across a six-site pilot subset. Analyzing selected physicochemical, major ion, nutrient, trace element, and microbiological parameters, we assessed method reliability under real-world conditions.

Field methods effectively screen parameters when concentrations substantially exceed detection limits and produce distinct color changes against reference scales. Field kits generally outperform test strips, while in-situ sensors reliably measure basic physicochemical parameters in real time. Test strips perform poorly for nutrients and trace elements at low concentrations, causing systematic bias due to limited sensitivity and high inter-observer variability. In the pilot subset, national laboratory results aligned more closely with accredited laboratories than field tools, though high detection limits and quality assurance practices may have introduced variability for certain parameters. Costs per sample ranged from USD 4 for test strips to USD 80 for field kits and USD 155 for the national laboratory, with turnaround times from immediate results to weeks or months.

These findings highlight the need to select testing tools based on fit-for-purpose monitoring objectives rather than analytical precision alone. Combining field-based screening with targeted laboratory confirmation can expand monitoring coverage while preserving data integrity, offering practical guidance for groundwater management in resource-constrained settings.

Introduction

Groundwater provides drinking water and domestic supply for hundreds of millions of people worldwide, particularly in low-income countries where water supplies often rely upon decentralized systems of abstraction (WHO & UNICEF 2025; UNWWDR 2022). In these settings, systematic water quality monitoring is constrained by limited finances, inadequate laboratory capacity, and lack of trained personnel. In response, a wide range of relatively low-cost, field-based water quality testing methods have become available, including multi-parameter test strips, portable test kits, and handheld sensor-based instruments (Pichel et al. 2023; Bain et al. 2021). These methods are promoted for rapid screening and decentralized monitoring, and are increasingly used by government agencies, researchers, non-governmental organizations, and community groups (Wright et al. 2004; Naigaga et al. 2017). Despite their growing uptake, there is limited empirical evidence on how well these field-based methods perform relative to laboratory-based analyses across a broad range of parameters, particularly under field conditions in low-income countries.

Existing groundwater quality studies in low-income countries largely focus on characterizing hydrochemical conditions or identifying contamination risks, with analytical methods often treated as a secondary consideration. While comparative assessments of groundwater quality methods do exist, these have largely been conducted in high-income countries or under highly controlled conditions (Guigues et al. 2024). Studies systematically evaluating the performance, consistency, and limitations of different analytical methods under real-world conditions in low-income countries remain scarce. Whilst direct comparisons between field-based methods and laboratory analyses have been previously reported (e.g., Reddy et al. 2020), comparisons between national laboratories and internationally accredited laboratories are rarely reported, despite the important roles that national laboratories play in regulatory monitoring and in supporting decision-making.

This study contributes towards closing these gaps by systematically comparing multiple field- and laboratory-based groundwater quality testing methods applied in parallel across a wide range of physical, chemical, and microbiological indicators. The comparison was undertaken under operational conditions typical of low-income country contexts, where limited opportunities for instrument calibration, delays in sample transport, and the absence of formal laboratory accreditation are common challenges. Lao PDR, listed by the UN as a Least Developed Country[1] (UNCTAD 2026), and where groundwater quality monitoring remains at an early stage of institutional development, provides a typical case study for examining these challenges.

The emphasis of this paper is on evaluating the trade-offs between alternative analytical methods in terms of performance, cost, and practicality of implementation. By combining statistical comparisons with logistical and institutional constraints, the study supports more informed method selection for groundwater quality monitoring in resource- and capacity-constrained settings. Field

methods are known to be less accurate than laboratory analyses; however, the issue addressed here is where, parameter by parameter, those differences remain small enough for practical use, and at what cost and turnaround time.

Materials and Methods

Study area and sampling design

Groundwater sampling was conducted in Khammouane Province, Lao PDR, in two areas with contrasting hydrogeological and land use conditions (Fig. 1). The larger southern area is characterized by intensive paddy cultivation, higher population density, and some urban and industrial development, while the smaller northern study area comprises lowland plains adjacent to karst-limestone terrain, with lower population density and less intensive agriculture.

The southern area encompasses parts, or all, of Thakhek, Nongbok, and Xebangfai Districts and is underlain by a combination of Mesozoic sandstones and Quaternary alluvial sediments (Viossanges et al. 2018). The northern area lies within Khounkham District and consists of low-lying alluvial plains underlain by Paleozoic sandstones that are likely hydraulically connected to the adjacent karst systems.

A total of 20 groundwater abstraction sites were sampled, comprising 14 sites in the southern area and 6 sites in the northern area. Sampling locations were selected in consultation with relevant provincial and district authorities, prioritizing tube wells with pumped abstraction and accessible wellheads. Where tube wells were unavailable, dug wells were sampled. The groundwater from these sites was primarily used for domestic purposes, either for households (65%) or through small community supply systems (15%), with the remainder used by government offices, guesthouses, and a bottled water facility.

Field sampling procedures

Fieldwork was conducted between 2 and 6 June 2025. Groundwater samples were collected from the pump outlet after purging until electrical conductivity and temperature had stabilized, ensuring that formation water was being sampled. Field measurements of electrical conductivity (EC), pH, temperature, dissolved oxygen (DO), and oxidation-reduction potential (ORP) were recorded in-situ using handheld multi-parameter instruments.

For chemical analyses, groundwater samples were collected in pre-cleaned polyethylene bottles. Samples for major cation and trace metal analysis were filtered in the field through 0.45 µm membrane filters; however, only metal samples were acidified with ultrapure (70%) nitric acid. All samples destined for laboratory analysis were stored on ice during transportation. Samples analyzed at internationally accredited laboratories were subsequently transported by air to Australia.

Microbiological samples were collected separately and analyzed using field-based methods due to the remoteness of the sampling locations and the difficulty of returning samples to laboratories within the recommended time.

Analytical methods

Five analytical approaches were evaluated: in-situ sensor-based measurements (IS), multi-parameter test strips (TS), portable field test kits (TK), a national laboratory (NL), and internationally accredited laboratories (IL). Table 1 lists the parameters determined using each method.

All 20 sites were analyzed using the three field-based methods, whilst only a subset of 6 sites were analyzed using both laboratory-based methods. These six sites form the basis for all direct field-laboratory method comparisons and, given their number, are treated as a pilot comparison rather than a representative sample. The resulting dataset, therefore, comprises a complete field-based dataset across all sites, with partial laboratory coverage. This distinction is accounted for in subsequent comparisons.

Field-based methods

Prior to field testing, seven potential suppliers of field-based testing kits from Australia, India, and the United States were reviewed. Final selection was based on availability, cost, and ease of use under field conditions. Costs per sample (Table 5) cover consumables and shipping for the field methods and the commercial analytical charge for NL; staff time is not included. Commercial brands are not specifically identified, as the analysis focuses on the performance of different methods rather than an evaluation of specific products.

In-situ sensors (IS): Field parameters (EC, pH, temperature, DO, and ORP) were measured using a handheld water quality analyzer (IS-1). The DO sensor was calibrated prior to use; however, calibration of EC and pH was constrained by the unavailability of calibration reagents during fieldwork. To assess potential calibration bias, parallel measurements of EC and pH were also taken using a secondary handheld EC–pH analyzer (IS-2) that had been calibrated prior to fieldwork conducted in the preceding week.

Test strips (TS): Multi-parameter colorimetric test strips were used for rapid, on-site screening of selected physicochemical parameters, major ions, nutrients, and selected trace elements. Test strips were immersed directly in groundwater samples and left for the manufacturer-specified reaction time (< 1 minute). Each strip was photographed and independently interpreted by three observers following a standardized protocol to evaluate consistency amongst readers.

Test kits (TK): Portable test kits were used for field analysis of selected physicochemical parameters, major ions, nutrients, trace elements, and microbiological indicators. Following collection, samples were analyzed at the field base using manufacturer-supplied reagents, glassware and color comparison charts, following standard operating procedures. Microbiological quality was assessed using hydrogen sulphide presence/absence tests for fecal indicator bacteria, with results interpreted after a 24-hour incubation period.

Laboratory-based methods

National laboratory (NL): A subset of six samples was analyzed at a university-affiliated water quality laboratory in Laos. The laboratory primarily supports research activities and provides limited commercial services. While it had not applied for formal accreditation, it follows internal analytical protocols and supports national research and monitoring programs. Analyses were restricted to parameters that could be reliably assessed using available instrumentation, including major ions and selected trace elements.

Internationally accredited laboratories (IL): The same subset of six samples was analyzed at two NATA-accredited laboratories in Australia operating under ISO/IEC 17025-compliant quality management systems. Major cations and trace metals were analyzed using inductively coupled plasma optical emission spectrometry (ICP-OES) and inductively coupled plasma mass spectrometry (ICP-MS), while major anions were analyzed by ion chromatography.

Quality assurance, detection limits, and data limitations

Formal quality assurance procedures differed among the methods. Field-based methods relied on manufacturer protocols, parallel measurements where feasible, and inter-observer comparison (for test strips) to assess consistency. National and internationally accredited laboratories applied their routine internal quality control procedures; however, laboratory-specific quality control data (e.g., blanks or duplicates) were not used for independent evaluation in this study.

For some trace elements (e.g., lead), detection limits at the national laboratory exceeded guideline values, limiting the interpretability of non-detect results. These limitations were taken into account during method interpretation. In addition, iron and other trace element results are not equally comparable across methods: national and accredited laboratory samples were filtered (0.45 µm) prior to analysis, whereas test strip and test kit measurements were performed on unfiltered waters, as is usual for field kits; their reagents respond mainly to dissolved iron unless a digestion step is included, so neither approach measures total iron. This difference in sample treatment should be noted when comparing results across methods.

Statistical analysis

Analytical methods were compared using both visual plots and statistical tests. Because all methods were tested on the same six sites, paired tests were used to account for natural differences in water chemistry between locations. Prior to analysis,

concentrations were base-10 logarithmically transformed to stabilize variance across wide concentration ranges. Where three or more methods were compared, we applied a two-way ANOVA (accounting for both analytical method and sampling site), followed by Tukey's honestly significant difference (HSD) post-hoc test, and confirmed the results using the non-parametric Friedman test to ensure findings were not skewed by outliers or non-normal data. Where only two methods were compared, we used a paired *t*-test on the log-transformed values alongside the exact Wilcoxon signed-rank test as a non-parametric check. A difference between methods was treated as statistically supported only when both the parametric and non-parametric tests yielded $p < 0.05$. To quantify the magnitude of differences, we report geometric mean ratios relative to the international laboratory (IL), for which total hardness was calculated from calcium and magnesium concentrations (as $CaCO_3$) to match the national laboratory (NL) procedure.

Values below the limit of detection were set to half this value for laboratory methods, and to half the lowest readable color-scale value observed across all 20 sites for field methods. If half or more of a method's results for a given parameter fell below detection, that method was excluded from statistical testing so that comparisons reflected actual measurements rather than substituted values; parameters left with fewer than two testable methods were described qualitatively. Finally, because the small sample size ($n = 6$) allows only large differences to be detected, and because the exact Wilcoxon test cannot return a *p*-value below 0.031, a non-significant result does not demonstrate that two methods are equivalent. Accordingly, these laboratory comparisons are reported as a pilot assessment and interpreted with appropriate caution.

To assess the reliability of test strips as screening tools, pairwise comparisons were conducted between TS results and other methods using error metrics (mean bias error, mean absolute error, and root mean square error), correlation coefficients, regression slopes and measures of relative variability and rank preservation.

Table 1
Parameters determined by the five analytical methods: IS, TS, TK, NL and IL. Light shading indicates parameters suitable for qualitative comparison across different methods.

Parameter	IS (*n* = 20)	TS (*n* = 20)	TK (*n* = 20)	NL (*n* = 6)	IL (*n* = 6)
EC [1]					
Temperature [1]					
DO					
ORP					
pH[1]					
Turbidity					
Total Alkalinity					
Total Hardness					
Bicarbonate					
Calcium					
Chloride					
Sodium					
Sulphate					
Sulphite					
Magnesium					
Potassium					
Fluoride					
Bromide		Bromine			
Iodine					
Ammonia [2]				Ammonium	
Nitrate					
Nitrite					
Phosphate [3]					Total Phosphorous
Aluminium					
Arsenic					
Cadmium					
Chromium					
Cobalt					
Copper					
Iron					
Lead					

[1] two in-situ instruments used; [2] TK “Ammonia” and NL “Ammonium” report the same species on different bases; [3] TK “Phosphate” and IL “Total Phosphorus” are not necessarily equivalent, as total phosphorus includes organic and particulate bound fractions in addition to orthophosphate; [4] semi-quantitative determination by presence or absence only

Parameter	IS (*n* = 20)	TS (*n* = 20)	TK (*n* = 20)	NL (*n* = 6)	IL (*n* = 6)
Manganese					
Mercury					
Zinc					
Total Chlorine					
E. Coli [4]			Fecal Bacteria [2]		
Silica					
Antimony					
Boron					
Molybdenum					
Nickel					
Selenium					
Strontium					
Uranium					
Oxygen-18					
Deuterium					

[1] two in-situ instruments used; [2] TK “Ammonia” and NL “Ammonium” report the same species on different bases; [3] TK “Phosphate” and IL “Total Phosphorus” are not necessarily equivalent, as total phosphorus includes organic and particulate bound fractions in addition to orthophosphate; [4] semi-quantitative determination by presence or absence only

Results

Overview of groundwater quality across sampled sites

Groundwater quality across the sampled sites spans a broad, but generally low-to-moderate concentration range for most physico-chemical parameters (Table 2). EC values range from 27 to 1,496 µS/cm, with a mean of 423 µS/cm, indicating predominantly fresh groundwater. pH values measured with IS-1 range from 3.5 to 7.8, with 85% of sites below the drinking water range of 6.5–8.5. Because IS-1 could not be calibrated and read on average 0.51 units lower than the calibrated IS-2, this proportion is likely overstated; on IS-2 readings it is 60%. Either way, most of the groundwater sampled is mildly acidic. DO concentrations are generally low (mean 2.4 mg/L), and ORP values indicate mildly oxidizing conditions at most sites. Major ion concentrations are low to moderate, with calcium, magnesium, sodium, and chloride values well below drinking water guideline values.

Nutrient concentrations are generally low, with nitrate values ranging from non-detectable to 20.8 mg/L and ammonium below detection limits at all six laboratory sites. Sulphate and phosphate concentrations are also low. Trace elements, including arsenic, cadmium, copper, and lead, are predominantly below detection limits or present at very low concentrations. Iron and manganese show greater variability, with iron concentrations exceeding the national guideline of 1 mg/L at one site. Microbiological testing detected no fecal contamination at all sites. As this result is based on a qualitative presence/absence method, it indicates that fecal contamination was not detected within the test's limits of sensitivity, rather than definitively confirming absolute microbiological safety.

Spatial patterns in groundwater quality

The spatial distribution of groundwater quality, as assessed by the test kit (TK) method, reveals a clear hydrogeological contrast between the northern and southern study areas (Fig. 2). In the northern area, the groundwater is characterized by low EC values (< 500 µS/cm), moderately hard water, and predominantly low pH. These trends are consistent with the region's karstic terrain, where higher transmissivity aquifers and rapid recharge processes limit the residence time for extensive mineral dissolution. In contrast,

the southern area exhibits a more complex hydrochemical profile, with a wider range of EC values (30 – 1,500 µS/cm) and greater variability in hardness. This is consistent with lower recharge rates and longer residence times in the floodplain aquifers. Elevated EC levels (> 1000 µS/cm) in the south likely reflect dissolution of evaporites (Thangon Formation) underlying the southern area (Sha et al. 2026). Increased hardness near the Mekong River suggests the dissolution of calcium- and magnesium- bearing minerals within the alluvial sediments. Vertical stratification is evident in the south; the presence of saline groundwater at depth, overlain by shallow freshwater lenses, aligns with recent geophysical evidence of salinity layering (Xayavong et al. 2023).

Nitrate concentrations are generally low across the province, with higher values observed at a small number of southern sites. Trace element concentrations show limited spatial variability, with fluoride below the TK detection limit at all but one site. TK arsenic readings do not exceed 0.01 mg/L, although the laboratory comparison indicates that TK overestimates arsenic (Table 3). Detectable iron concentrations are observed at several southern sites.

Visual comparison of analytical methods

Comparison of results from the six sites where all methods were applied reveals both consistencies and systematic discrepancies in the dataset (Fig. 3). For EC and pH, the two in-situ instruments (IS-1 and IS-2) yield broadly comparable values, though IS-2 consistently records slightly higher readings. Test strip (TS) and test kit (TK) results for pH generally fall within the range of in-situ measurements, but with reduced variability for the TK method.

For temperature, IS-2 measurements are systematically higher than IS-1, reflecting differences in measurement protocols rather than instrument performance. TS alkalinity results are mostly zero, and TS hardness is lower than laboratory values at the hardest sites, whereas TK hardness exceeds laboratory values in soft waters. TK alkalinity broadly follows alkalinity inferred from laboratory bicarbonate but reads high at low alkalinity sites.

Nutrient concentrations differed substantially depending on the analytical method used. Nitrate values obtained using TK, NL, and IL are broadly consistent, while TS results are predominantly zero, indicating limited sensitivity. For phosphate, NL results are below detection at all six sites, so any apparent similarity with TK reflects the substituted value; IL results are lower than TK. This comparison should be interpreted with caution, as “phosphate” (TK, NL) and “Total Phosphorus” (IL) represent distinct phosphorus pools and thus are not necessarily equivalent analytes (Table 1).

This comparison should be interpreted with caution, as “phosphate” (TK, NL) and “Total Phosphorus” (IL) represent distinct phosphorus pools and thus are not necessarily equivalent analytes (Table 1).

Trace metal concentrations show the greatest variation by method. Iron concentrations for NL and IL are substantially higher than those of TS and TK, while lead and arsenic concentrations are generally below detection limits across all methods. This contradicts expectations based on filtration alone, as removing particulate matter should decrease rather than increase laboratory values relative to unfiltered field measurements. Mean iron concentrations for NL and IL are substantially higher than those of TS and TK, but the difference arises from a single site, where NL and IL measured 7.7 and 4.5 mg/L and both field methods read 0.3 mg/L. At the other five sites all methods reported 0.3 mg/L or less. Filtration cannot explain the discrepancy, as removing particulate matter would lower, rather than raise, laboratory values relative to unfiltered field measurements. Instead, differences in iron speciation and sample preservation provide a more likely explanation. Laboratory samples were filtered and acidified at collection to keep dissolved iron in solution, whereas field tests were carried out on unpreserved water (and for TK, only after transport to the field base). In this low-oxygen groundwater (DO = 0.23 mg/L at that site), dissolved ferrous iron exposed to air can oxidize and precipitate as ferric oxyhydroxide, which colorimetric reagents detect poorly without an acid dissolution step. Compression of the field color scales at higher concentrations, also evident in the low TS regression slopes for iron (Table 4), would further compound this underestimate. However, this mechanism could not be directly verified with the data collected. Lead is below detection in all field and NL results, and arsenic is below the NL detection limit at all six sites.

Inspection of Fig. 3 shows that agreement between analytical methods is highly parameter-dependent. Overall, these discrepancies likely reflect a combination of method-specific detection limits, differences in analytical resolution, and subjective color-scale interpretation rather than a uniform bias across all parameters.

Statistical comparison of analytical methods

Paired comparisons were possible for eight parameters (Table 3). Differences between methods were supported by both parametric and non-parametric tests for three: total hardness, magnesium, and fluoride. For hardness, TK results exceeded IL by a factor of about 3.4 on geometric means (Tukey $p = 0.047$), with the largest overestimates in soft waters (315 mg/L by TK compared with 22 mg/L by IL at one site), whereas TS compressed the range, reading low at the hardest sites. For magnesium, TS results exceeded both laboratories by more than an order of magnitude (geometric mean ratio 17 relative to IL). For fluoride, NL results were lower than IL (geometric mean ratio = 0.38), though both remained well below the 1.5 mg/L guideline, making the difference of no practical importance.

TK arsenic results had a geometric mean roughly 19 times higher than IL (paired *t*-test $p = 0.001$), but the Wilcoxon test did not reach significance ($p = 0.063$), so the difference was not considered statistically significant. Nevertheless, this matters in practice: TK reported the WHO guideline value of 0.01 mg/L at two sites where IL measured below 0.0004 mg/L. No differences were supported for calcium, chloride, sulphate, or nitrate. Nitrite, phosphate, copper, iron, and lead could not be tested because fewer than two methods had sufficient detections. With six samples, the absence of a supported difference is not evidence of equivalence; for chloride, for example, the TK geometric mean was 3.3 times the IL value without reaching significance.

Inter-observer variability in test strip interpretation

To assess the consistency of test strip (TS) interpretation, digital images of TS results were independently interpreted by three observers. Results were classified as full agreement (all three observers reported the same value), partial agreement (two observers agreed), or no agreement (all three differed).

Levels of agreement vary markedly across parameters (Fig. 4). High levels of full agreement (≥ 75%) are observed for parameters with distinct color changes, such as nitrate and total alkalinity. In contrast, very low levels of full agreement (≤ 10%) are observed for calcium and magnesium (not shown in Fig. 4), with frequent partial or no agreement. Parameters such as pH, total hardness, sulphite, and copper show moderate to low agreement. These results demonstrate that inter-observer variability can be a significant source of uncertainty for this method, and is consistent with previous findings (Ramesh et al. 2024).

Reliability of test strips as screening indicators

The reliability of test strips as screening tools was evaluated through pairwise comparisons of TS with test kits and laboratory methods for nitrate, fluoride, and iron. These comparisons utilized datasets of up to 20 sites for the TS versus TK analysis, and up to six sites for the laboratory comparisons, focusing on parameters common to all methods (Tables 1 and 4).

For nitrate, TS results show a moderately strong correlation with TK ($r = 0.81$), but with a regression slope of 0.40, indicating systematic underestimation, particularly at higher concentrations. Comparisons with laboratory methods show consistent negative bias but limited variability, reflecting TS readings of zero at all six laboratory sites. For fluoride, TS results show poor agreement with all reference methods, with negative correlations and a positive bias of about 2 mg/L: TS read 4 mg/L at 11 of 20 sites, whereas TK detected fluoride at only one site and laboratory values did not exceed 0.14 mg/L. These results indicate that TS methods do not reliably quantify fluoride within the observed ranges. For iron, the high correlations between TS and laboratory methods rest on a single high concentration site and tied substituted values (Table 4), so they do not indicate agreement; the low regression slopes show that TS greatly underestimates iron at that site. Comparisons between TS and TK show weaker correlations and underrepresentation of variability at higher concentrations.

Overall, the TS method preserves broad spatial patterns determined from other methods for nitrate but performs poorly for fluoride and iron under the concentration ranges observed. These findings support the use of TS for rapid screening of selected parameters, while highlighting clear limitations for others.

Table 2
Summary of groundwater quality relative to national and WHO drinking water guidelines where given. EC, pH, temperature, DO and ORP are from IS-1, and turbidity, total alkalinity and bacteria from TK, at all 20 sites; all other parameters are from NL at the 6 sites. Exceedance values are underlined.

Parameters	National Guidelines	WHO Guidelines	Mean	Minimum	Maximum
EC (µS/cm)	1000		422.6	26.9	1496
pH	6.5–8.5		5.7	3.5	7.8
Temperature (°C)			27.4	24.4	29.2
DO (mg/L)			2.4	0.2	7.2
ORP (mV)			173.7	-68.6	1159
Turbidity (NTU)	10		5.3	2.5	> 25
Total Alkalinity (mg/L)			232.5	0.0	650
Bicarbonate (mg/L)			123.9	8.2	360.8
Calcium (mg/L)			31.7	0.8	109.6
Chloride (mg/L)	250		40.6	< 0.05	143.6
Fluoride (mg/L)	1.5	1.5	0.04	< 0.03	0.07
Total Hardness (mg/L)	300		107.5	2.9	277.3
Magnesium (mg/L)			6.9	< 0.1	29.3
Potassium (mg/L)			6.4	0.1	25.9
Sodium (mg/L)			20.5	0.2	63.9
Sulphate (mg/L)			11.2	0.3	33.4
Ammonium (mg/L)	1.5	1.5	< 0.08	< 0.08	< 0.08
Nitrate (mg/L)	50	50	5.4	< 0.05	20.8
Nitrite (mg/L)	3	3	0.2	< 0.09	0.9
Phosphate (mg/L)			< 0.21	< 0.21	< 0.21
Arsenic (mg/L)	0.05	0.01	< 0.007	< 0.007	< 0.007
Cadmium (mg/L) [1]		0.003	< 0.008	< 0.008	< 0.008
Copper (mg/L)		2	< 0.02	< 0.02	< 0.02
Iron (mg/L)	1		1.4	< 0.27	7.7
Lead (mg/L) [2]	0.01	0.01	< 0.15	< 0.15	< 0.15
Manganese (mg/L)		0.08	0.07	< 0.02	0.23
Bacteria	0 [3]		ND [4]	ND	ND

[1] detection limits for NL exceeded guideline, however IL results indicate values of < 0.004 mg/L; [2] detection limits for NL exceeded guideline, however IL results indicate values of < 0.0005 mg/L; [3] guideline refers to fecal coliforms (MPN/100mL); [4] ND = not detected. Means include values below detection set at half the detection limit.

Table 3
Paired comparisons of analytical methods at the six sites analyzed by all methods.

Parameter [1]	**Methods Compared (*n*)**	**Methods Excluded (below detection ratio)**	**Test Statistic**	***p***	**Rank Test *p***	**Supported Pairwise Differences (*p*)**	**Geometric Mean Ratio to IL**
Total Hardness	TS, TK, NL, IL [2] (6)	-	$F(3,15) = 4.89$	**0.014**	**0.008**	TK > TS (0.016); TK > IL (0.047)	TS 0.79; TK 3.39; NL 1.04
Calcium	TS, NL, IL (6)	-	$F(2,10) = 1.05$	0.385	0.042	-	TS 2.21; NL 1.17
Chloride	TK, NL, IL (6)	-	$F(2,10) = 2.47$	0.135	0.513	-	TK 3.33; NL 0.60
Magnesium	TS, NL, IL (5) [3]	-	$F(2,8) = 13.84$	**0.003**	**0.022**	TS > NL (0.003); TS > IL (0.010)	TS 17.0; NL 0.50
Sulphate	NL, IL (6)	TS (6/6)	$t(5) = -1.10$	0.323	0.844	-	NL 0.71
Fluoride	NL, IL (6)	TS (3/6); TK (5/6)	$t(5) = -5.90$	**0.002**	**0.031**	NL < IL	NL 0.38
Nitrate	NL, IL (6)	TS (6/6); TK (3/6)	$t(5) = -0.79$	0.467	0.156	-	NL 0.68
Arsenic	TK, IL (6)	NL (6/6)	$t(5) = 6.48$	0.001	0.063	_ [4]	TK 19.1

F: randomized block ANOVA on log10 values with site as block, rank test Friedman, pairwise Tukey HSD on the block model. *t*: paired *t* test on log10 values, rank test exact Wilcoxon signed rank (minimum attainable $p = 0.031$ for $n = 6$). Bold values indicate parameters where both the parametric and non-parametric tests reached $p < 0.05$. Geometric mean ratio: method relative to IL; values above 1 indicate higher readings than IL. Methods with half or more results below detection were excluded. [1] Nitrite, phosphate, copper, iron and lead were not tested as fewer than two methods had most results above detection; [2] IL hardness calculated from calcium and magnesium as $CaCO_3$; [3] TS missing at one site; [4] paired *t* test significant but rank test not.

Table 4
Results of the comparative analysis of the absolute accuracy and relative sensitivity of the TS method for water quality monitoring

Method comparison	*n*	MBE (mg/L)	MAE (mg/L)	RMSE (mg/L)	*r*	Slope	R^2	SD ratio	Rank r
Nitrate									
TS vs TK	20	-8.50	8.50	11.73	0.813	0.402	0.662	0.494	0.34
TS vs NL	6	-5.40	5.40	9.08	-	-	-	0.773	-
TS vs IL	6	-4.80	4.80	8.27	-	-	-	0.837	-
Fluoride									
TS vs TK	20	1.98	2.03	2.83	-0.229	-4.44	0.062	18.4	-0.23
TS vs NL	6	1.96	2.02	2.81	-0.772	-73.85	0.649	95.2	-0.80
TS vs IL	6	1.90	2.02	2.77	-0.745	-66.67	0.694	83.8	-0.78
Iron									
TS vs TK	19	0.024	0.055	0.124	0.395	0.507	0.156	1.28	0.35
TS vs NL[(1)]	5	-1.585	1.585	3.31	1.00	0.040	1.00	0.04	1.00
TS vs IL	5	-0.921	0.921	1.90	0.998	0.067	0.997	0.06	0.71

[(1)] *r* = 1.00 is an artifact of tied substitute values and a single data pair, not true method agreement

Discussion

Performance and constraints of field-based methods

Field-based testing agreed reasonably with laboratory analyses for some parameters, such as nitrate by TK, where concentrations exceeded detection limits and color matching to reference scales was unambiguous, and the two in situ sensors gave consistent EC readings. This supports their use for rapid screening and preliminary site assessments.

However, performance is highly parameter-dependent and subject to major constraints. Systematic bias was observed for several parameters, notably fluoride and magnesium by TS and hardness by TK, and TS greatly underestimated iron at the one site with high concentrations. This was evidenced by compressed response ranges in test strips and low regression slopes (Fig. 3; Table 4). These differences are consistent with well-known issues associated with colorimetric resolution, field conditions, and operator interpretation (Swinney et al. 2024). While field methods are valuable for initial scoping, they are not a substitute for laboratory analysis where regulatory compliance, public health risk assessment, or long-term data trends are required. Test kits generally outperformed test strips due to better analytical resolution and procedural controls. These results emphasize that field-based methods should be evaluated based on their individual suitability for specific objectives, rather than being treated as a uniform class of analytical tools.

The field methods evaluated exhibited distinct performance limitations. Because test strips and kits rely on colorimetric reactions, their accuracy is governed by chemical stability and color resolution rather than hardware performance. Sensor accuracy, by contrast, depends heavily on routine calibration. To directly evaluate the sensitivity of our data to instrument calibration status, electrical conductivity (EC) and pH were measured in parallel using a second, independently calibrated in situ instrument (IS-2) alongside the primary, uncalibrated field instrument (IS-1). This comparison revealed a systematic bias, with IS-1 deviating from IS-2 by an average of 16 ± 20 µS/cm for EC and 0.51 ± 0.26 for pH.

Trade-offs between performance and operational feasibility

Internationally accredited laboratories provided the reference dataset, but results took weeks to months to be received, and their use involved substantial logistical challenges related to sample preservation, transport and regulatory border controls. These

constraints can limit the exclusive reliance on accredited laboratories for routine groundwater monitoring in many low-income country settings.

National laboratories play an important role. In this pilot, their results generally aligned more closely with those from the accredited laboratories than did field-based methods, but with greater variability for certain parameters. At USD 155 per sample and a turnaround of weeks, the national laboratory was also the most expensive option costed here, at around twice the cost of test kits. The variability in NL results is likely attributable to differences in analytical protocols, calibration frequency, detection limits, and quality assurance procedures, rather than to fundamental technical constraints. Given the diverse applications of national laboratories in academic research, regulatory monitoring and decision-making, targeted investments in quality assurance systems, staff training and inter-laboratory comparison efforts have the potential to lead to substantial improvements in data reliability at relatively low cost.

Field-based methods, including TS and TK, offer clear advantages in terms of affordability, speed and spatial coverage, despite lower analytical accuracy. When deployed intentionally for purposes such as initial screening, reconnaissance surveys or community-based monitoring, they can substantially expand the opportunities for groundwater quality assessments.

These results indicate that inter-observer variability can be a significant source of uncertainty for the TS method. While not used in this study, smartphone applications can reduce observer bias by replacing visual color-matching with automated, algorithm-based determination. This standardizes the process of reading strips and minimizes subjective human judgment (Pattinson et al. 2023; Topping and Kolok 2021). Whilst machine-based methods increase consistency relative to direct visual reading, improvements in absolute accuracy over trained visual observation are not always apparent and depend on a host of factors such as strip type, parameters considered, and light conditions.

Implications for groundwater quality monitoring

These results have potential implications for the design of groundwater monitoring programs in Lao PDR and similar resource-constrained settings. Primarily, they demonstrate that analytical method selection should go beyond a narrow focus on accuracy to a more balanced consideration between technical monitoring outcomes and operational factors such as method availability, local familiarity, and cost.

Given that field and laboratory results agreed for some parameters, preliminary assessments that are aimed at identifying broad areas of concern can apply field-based methods. However, it is also important that the limitations of field methods are adequately understood by practitioners. When the objective shifts towards research or regulatory decision-making, the data from this study indicate that laboratory-based methods remain necessary to mitigate the risks of systematic bias and inaccuracies at low concentrations.

The study supports mixed methods of monitoring. Field-based methods serve as a high-efficiency prioritization tool to identify specific sampling locations, parameters or time intervals that warrant the higher cost and logistical requirements of laboratory analysis. In the context of Lao PDR, where the policy and institutional drivers for groundwater monitoring are in their early stages, such periodic surveys using field-based tools can offer a valuable baseline for sustainable resource management (Brindha et al. 2019). This approach suits data-scarce settings with limited resources (Hinton et al. 2024). The matrix presented in Table 5 categorizes the various methods according to their strengths and limitations, and, with cost per sample and time to result, provides a basis for method selection.

Transferability and limitations

The challenges and trade-offs identified in this study are not limited to the case study areas. Decentralized groundwater use, limited monitoring infrastructure and resource constraints are common across Lao PDR and other low- or middle-income countries (Karandish et al. 2025). The findings may therefore apply in comparable contexts, although this remains to be tested.

Several important limitations should be noted. Due to budget constraints, the laboratory comparison was restricted to a pilot of six sites and the evaluation was limited to a single brand of test strip, one field test kit and one national laboratory. Performance may differ across alternative products, manufacturers or laboratory facilities, as well as under different hydrogeochemical conditions. The results should therefore be interpreted as indicative rather than broadly representative.

This study also could not account for broader operational and temporal factors, such as inter-operator variability, instrument drift, or seasonal shifts in water chemistry. These gaps highlight the need for more comprehensive research that tracks a wider range of field and laboratory methods across diverse environmental conditions and extended timescales.

Table 5
Decision matrix for selecting analytical methods by monitoring purpose, with cost per sample and time to result as incurred in this study

Analytical method	Main purpose	Cost per sample (USD)	Time to result	Suitable parameters (indicative)	Main strengths	Main limitations	Recommended use
Test strips (TS)	Initial screening	4	Immediate	EC, pH, major ions	• Very low cost • Rapid results • Minimal training required	• Low precision • Color interpretation subjective • Poor performance at low concentrations	• Rapid reconnaissance • Community screening • Prioritizing sites
Test kits (TK)	Screening / semi-quantitative	80	Same day	Physicochemical parameters; selected nutrients; bacterial indicators	• Better precision than TS • Deployable in field • Reasonably rapid results	• Reagent stability • Operator dependence • Limited trace element capability	• Field surveys • Contingency monitoring where lab access is constrained
In-situ sensors (IS)	Real-time field measurement	0 (1)	Immediate	EC, pH, temperature, DO, ORP	• Rapid results • High repeatability	• Calibration • Drift • Limited parameter range	• Field verification • Supporting laboratory sampling
National laboratories (NL)	Routine analysis	155	Weeks	Broad range	• Lower cost than IL • Institutional integration	• Variable QA/QC • Inter-lab variability	• Regulatory monitoring • Confirmation of field results
Accredited laboratories (IL)	Reference analysis	-	Weeks to months	Full suite	• Higher accuracy & consistency • Accredited QA/QC	• High cost • Long turnaround • Logistical challenges	• Baseline studies • Compliance • Validation

(1) IS cost excludes instrument purchase

Conclusions

This study has evaluated the performance and operational trade-offs of field- and laboratory-based groundwater quality testing methods using a case study from Khammouane Province, Lao PDR. By comparing results from test strips, field kits, in-situ sensors, a national laboratory, and internationally accredited laboratories, the study provides an assessment of how different types of methods perform under the range of constraints typical of low-income countries.

The results demonstrate that field-based methods can provide useful screening-level information for selected physicochemical parameters where concentrations are well above detection limits and the resulting color changes are distinct and easily matched to reference scales. However, their performance is highly parameter-dependent, with substantial limitations observed for nutrients and trace elements at low concentrations. Field-based methods should therefore not be used as substitutes for laboratory analyses in situations where rigorous evidence is necessary.

Among the field-based approaches evaluated, test kits generally performed more reliably than test strips, highlighting the importance of distinguishing between different classes of field methods. In the pilot of six sites, the national laboratory produced results broadly comparable to accredited laboratories but with greater variability for some parameters. This highlights the value of targeted investments in quality assurance and analytical capacity.

We frame method selection as a balance between analytical accuracy, cost, and operational factors. Cost per sample spanned USD 4 to USD 155 and turnaround time ranged from immediate to weeks or months, meaning method selection directly determines how many sites a budget can cover and how quickly decisions can be made. Despite the limitations of the study, the findings broadly support the use of mixed approaches whereby field-based methods are used for initial screening and prioritization, complemented by laboratory-based analyses for confirmation and decision-making. Such approaches offer a pragmatic way to expand groundwater quality monitoring coverage while keeping laboratory confirmation where decisions depend on it.

Declarations

Funding

This research was funded by the Australian Water Partnership through the project “Support to DWR – Sustainable Groundwater Management”.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author contributions

Paul Pavelic - Methodology, Investigation, Project Administration, Visualization, Writing–Original Draft; Eddie Banks - Methodology, Data Curation, Resources; Okke Batelaan - Methodology, Funding Acquisition, Supervision, Writing–Original Draft; Somphasith Douangsavanh - Investigation, Visualization; Sithong Kathsana - Investigation, Resources, Validation; Keoduangchai Keokhamphui - Methodology, Resources; Latsamee Kinglattana - Data Curation, Investigation, Resources, Validation; Saskia Noorduijn - Data Curation, Investigation, Resources, Writing–Original Draft; Keomanivone Phuthakhanty - Investigation, Resources, Validation; Lammone Sayavong - Investigation, Resources, Validation; Manivanh Suyavong – Investigation; Phaphone Sivongxay – Investigation; Simone Stewart - Writing–Original Draft; Mathieu Viossanges – Visualization; Sounipha Xaiyakeo - Investigation, Resources, Validation; Ounakone Xayviliya - Investigation, Resources, Validation.

Data Availability

The data supporting the findings of this study are available from the corresponding author on reasonable request.

Acknowledgements

The authors gratefully acknowledge the support provided by the relevant provincial and district authorities in Khammouane for their assistance in identifying sampling locations and generously sharing their local knowledge. We also thank the generosity of community members within the study areas for giving their consent for wells to be sampled for this study.

References

1. Bain R, Johnston R, Khan S, Hancioglu A, Slaymaker T (2021) Monitoring drinking water quality in nationally representative household surveys in low- and middle-income countries: Cross-sectional analysis of 27 multiple indicator cluster surveys 2014-2020. Environ Health Perspect 129(9), 97010

2. Brindha K, Pavelic P, Sotoukee T (2019) Environmental assessment of water and soil quality in the Vientiane Plain, Lao PDR. Groundw Sustain Dev 8, 24-30
3. Guigues N, Uysal E, Raveau S, Magar J, Billon G, Kahiluoto J, Schneider RJ, Liv L, Tosun SG, Canteau G, Lalere B (2024) Insights on the validation of alternative tools for water quality monitoring: the case of on-site test kits, portable devices and continuous measuring devices. Accredit Qual Assur 29,163-173; https://doi.org/10.1007/s00769-023-01570-x
4. Hinton RGK, Kalin RM, Banda LC, Kanjaye MB, Macleod CJA, Troldborg M, Phiri P, Kamtukule S (2024) Mixed method analysis of anthropogenic groundwater contamination of drinking water sources in Malawi. Sci Total Environ 957, 177418; https://doi.org/10.1016/j.scitotenv.2024.177418
5. Karandish F, Liu S, de Graaf I (2025) Global groundwater sustainability: A critical review of strategies and future pathways. J Hydrol 657, 133060; https://doi.org/10.1016/j.jhydrol.2025.133060
6. Naigaga S, Boyd CE, Gaillard P, Abdelrahman HA, Molnar JJ (2017) Assessing the reliability of water-test kits for use in pond aquaculture. J World Aquac Soc 48(4), 555-562. https://doi.org/10.1111/jwas.12377
7. Pattinson NB, Taylor J, Dickens CWS, Graham PM (2023) Digital innovation in citizen science to enhance water quality monitoring in developing countries. Colombo, Sri Lanka: International Water Management Institute (IWMI). 37p. (IWMI Working Paper 210); https://doi.org/10.5337/2024.201
8. Pichel N, Hymnô de Souza F, Sabogal-Paz LP, Shah PK, Adhikari N, Pandey S, Shrestha BM, Gaihre S, Pineda-Marulanda DA, Hincapie M, Luwe K, Kumwenda S, Aguilar-Conde JC, Cortes MALRM, Hamilton JWJ, Byrne JA, Fernandez-Ibañez P (2023) Field-testing solutions for drinking water quality monitoring in low- and middle-income regions and case studies from Latin American, African and Asian countries. J Environ Chem Eng 11(6), 111180; https://doi.org/10.1016/j.jece.2023.111180
9. Ramesh R, Frank E, Padmavilochanan A, Barda Y, Eldar I, Wolf H, Pras A, Pousty D, Anita P, Shekar L, von Lieres JS, Rao RB, Mamane H, Fishman R (2024) Reliable water quality monitoring by women in low-resource communities. ACS ES&T Water 4(9), 3832–3841
10. Reddy RR, Rodriguez GD, Webster TM, Abedin MJ, Karim MR, Raskin L, Hayes KF (2020) Evaluation of arsenic field test kits for drinking water: Recommendations for improvement and implications for arsenic affected regions such as Bangladesh. Water Res 170, 115325; https://doi.org/10.1016/j.watres.2019.115325
11. Sha J, Yin H, Zeng X, Zhang H (2026) Trace elements and REEs of the Late Cretaceous Halite from Thakhek Basin, Laos and its paleoenvironmental implication. Minerals 16(4), 346; https://doi.org/10.3390/min16040346
12. Swinney GG, Mullins SA, Martin J, Kinsey JB, King DA (2024) Assessment of the precision and accuracy of at-home multi-test water quality test strips. J Undergrad Chem Res 23(4), 54-58
13. Topping M, Kolok A (2021) Assessing the accuracy of nitrate concentration data for water quality monitoring using visual and cell phone quantification methods. Citiz Sci Theor Prac 6(1), 5, 1-9. https://doi.org/10.5334/cstp.346
14. UNCTAD (2026) The Least Developed Countries Report 2025: Are services the new path to structural transformation? United Nations Conference on Trade and Development, UNCTAD/LDC/2025
15. UNWWDR (2022) Groundwater: Making the invisible visible. United Nations, The United Nations World Water Development Report 2022, UNESCO, Paris
16. Viossanges M, Pavelic P, Rebelo L-M, Lacombe G, Sotoukee T (2018) Regional mapping of groundwater resources in data-scarce regions: The case of Laos. Hydrology 5(1), 2; https://doi.org/10.3390/hydrology5010002
17. WHO & UNICEF (2025) Progress on household drinking water, sanitation and hygiene 2000–2024: Special focus on inequalities. Geneva: World Health Organization (WHO) and the United Nations Children's Fund (UNICEF). Licence: CC BY-NC-SA 3.0 IGO
18. Wright J, Gundry S, Conroy R (2004) Household drinking water in developing countries: a systematic review of microbiological contamination between source and point-of-use. Trop Med Int Health 9(1), 106–117; https://doi.org/10.1046/j.1365-3156.2003.01160.x
19. Xayavong V, Minh VD, Singsoupho S, Duong NA, Prasad KND, Tuan VM, Chung DA (2023) Combination of 2D-Electrical Resistivity Imaging and Seismic Refraction Tomography methods for groundwater potential assessments: A case study of Khammouane province, Laos. Vietnam J Earth Sci 45(2), 1-13. https://doi.org/10.15625/2615-9783/18348

Footnotes

1. Lao PDR is scheduled to graduate from Least Developed Country status in November 2026 (UN General Assembly Resolution 76/8)

Figures

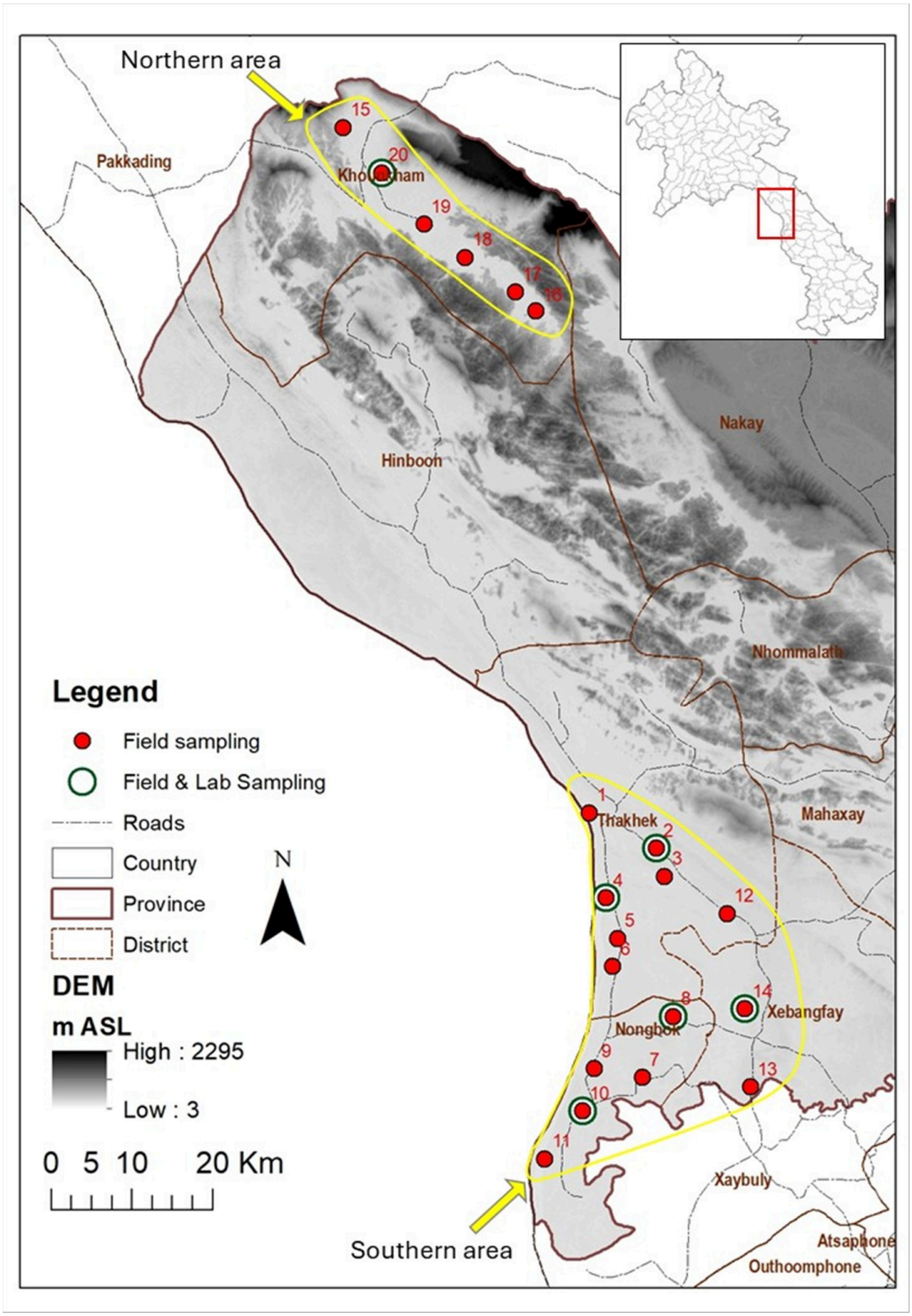


Figure 1

Location of the sampled sites in the southern and northern areas

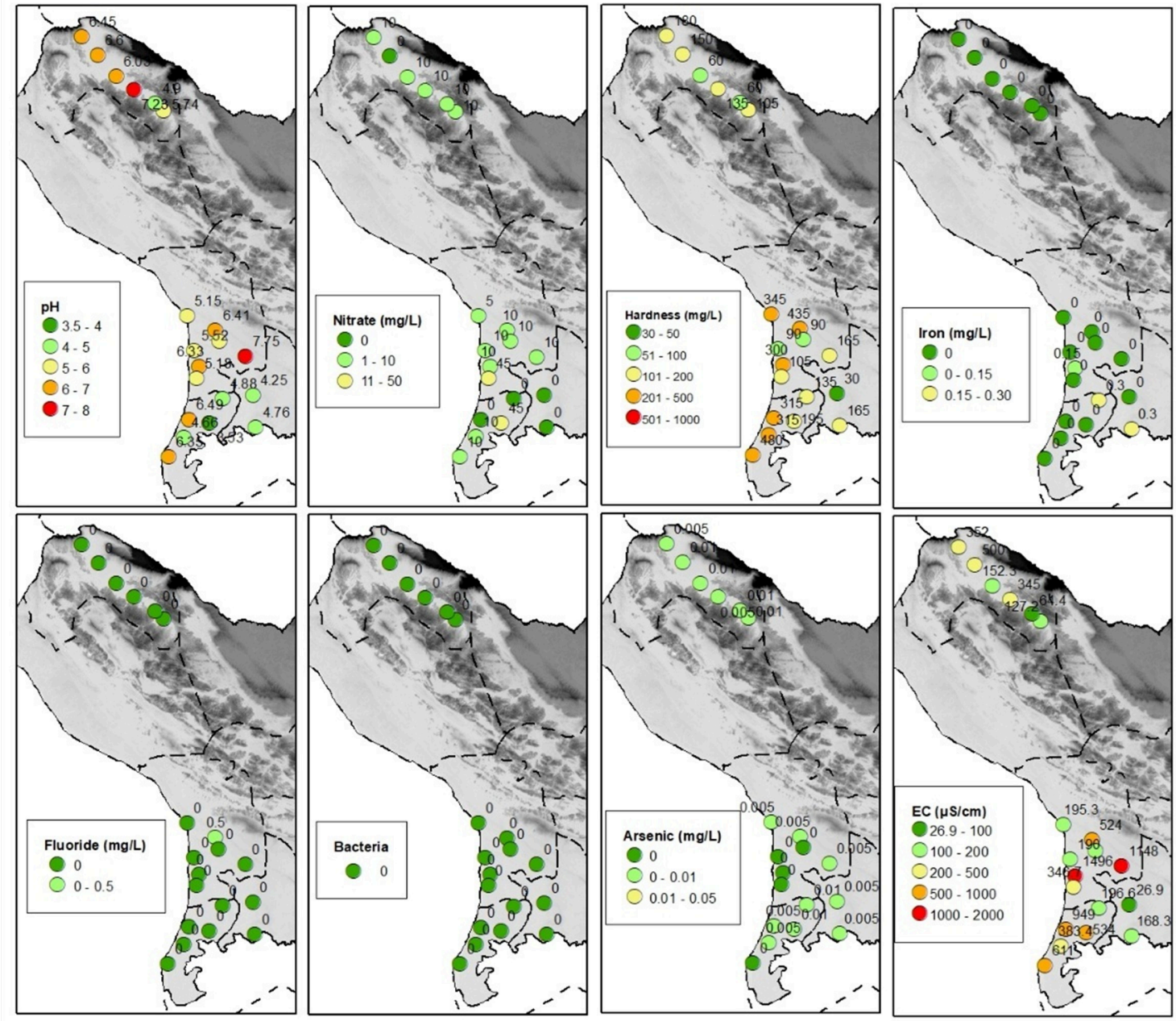


Figure 2

Spatial distribution of eight water quality parameters across northern and southern areas (only TK method data shown)

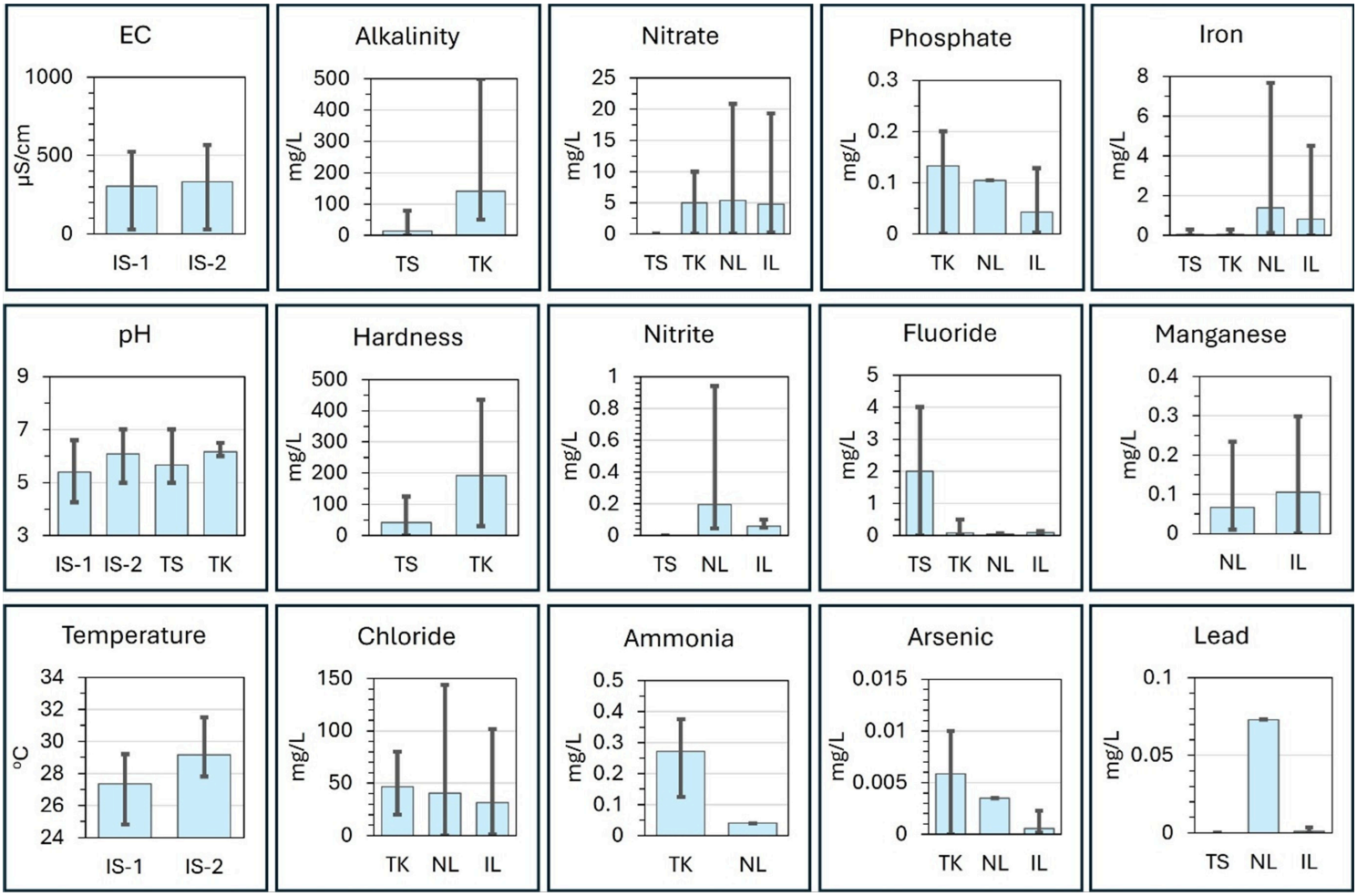


Figure 3

Comparison of the results of the different analytical methods for 15 parameters. Mean, maximum and minimum values on 6 samples tested for all methods are shown. Values that exceed the vertical axis scale are given in brackets.

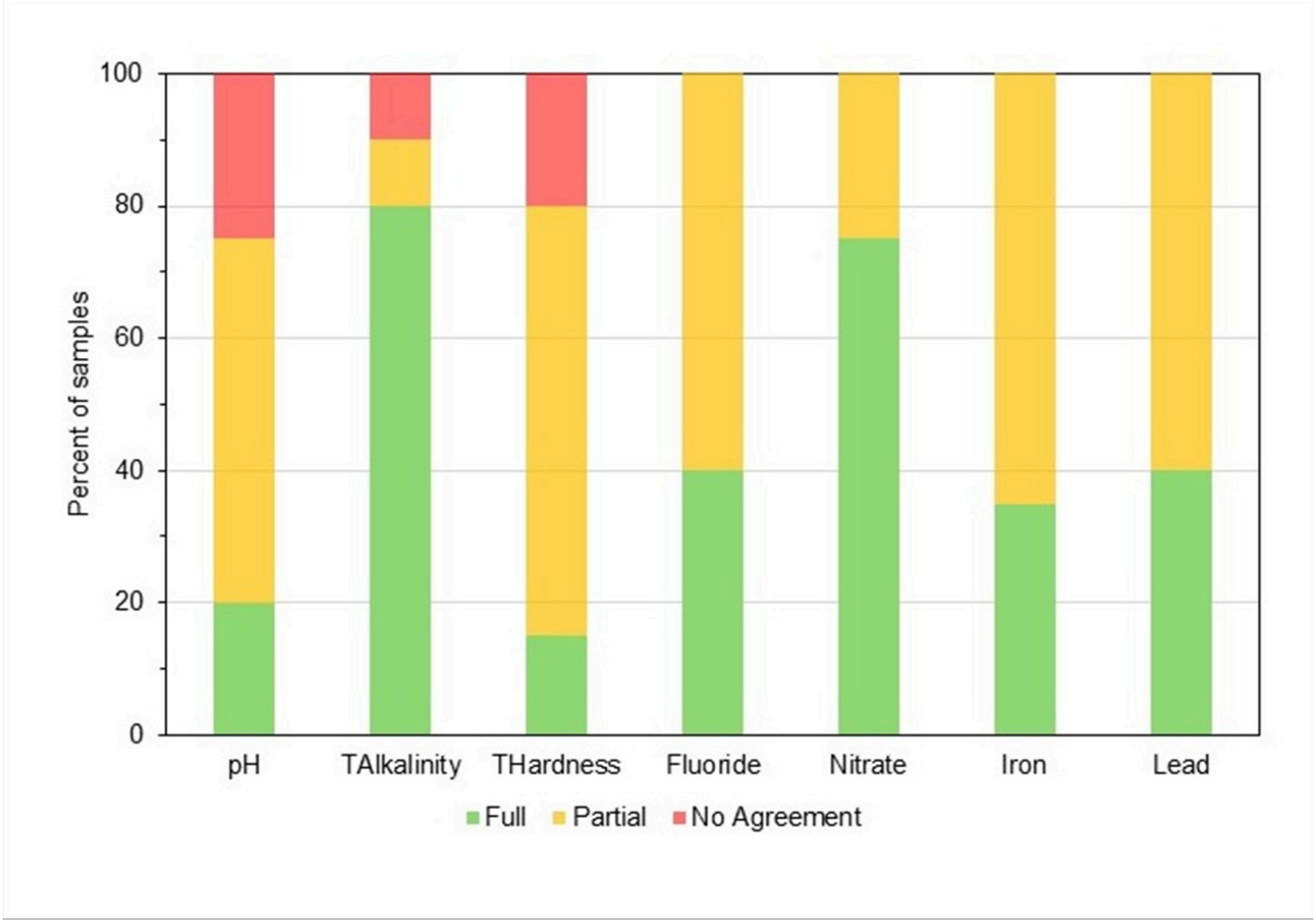


Figure 4

Summary of agreement levels from TS method for seven parameters