INTERNATIONAL JOURNAL OF ADVANCED RESEARCH
International Peer-Reviewed Research Journal
ISSN (Online): 3106-7980 | ISSN (Print): 3106-7972
Volume 3 • Issue 7 • November 2026

ESTIMATING VILLAGE-LEVEL DRINKING-WATER ACCESSIBILITY IN SINDH, PAKISTAN, FROM OPEN GEOSPATIAL DATA: WHY ANNUAL RAINFALL INVERTS THE RANKING

Arsalan Kaleem[*1], Sidra Soomro[2], Umar Jan[3], Sneha Haseja[4]

[*1,3,4]*Department of Computer Science, The Shaikh Ayaz University, Shikarpur, Sindh Pakistan.*
[2]*M.E. Department of Software Engineering, Mehran University of Engineering & Technology Jamshoro, Sindh Pakistan*

[*1]arsalanabbasi.here@gmail.com, orcid: 0009-0002-9054-6965

DOI: https://doi.org/10.5281/zenodo.24134151

Keywords
Water accessibility, composite index, remote sensing, CHIRPS, seawater intrusion, open data

Article History
Received: 05 August 2026
Accepted: 19 October 2026
Published: 10 November 2026

Copyright @Author
Corresponding Author: *
Arsalan Kaleem

Abstract

Household surveys are the main way Pakistan tracks progress toward Sustainable Development Goal 6.1, but they are expensive, run infrequently, and cannot resolve drinking-water access below the district level. This study builds the Sindh Water Access Index (SWAI), a reproducible composite index scoring relative drinking-water accessibility for 5,159 villages in Sindh from open geospatial data alone: HydroRIVERS and OpenStreetMap hydrography, WorldPop population, CHIRPS precipitation, Sentinel-2 NDWI, and SRTM elevation. Indicators were log-transformed, winsorized at the 1st and 99th percentiles, min-max normalized, then combined with rainfall and canal proximity treated as substitutes rather than as additive terms. The main finding is methodological. Using annual precipitation, which is standard practice in composite water indices, inverts the district ranking and places Tharparkar - the most water-insecure district in the province - among the best served. Tharparkar is not dry by annual total. Mithi receives about 277 mm a year against 148-169 mm in the canal-command districts, but it has no perennial supply, deeply saline groundwater, and no storage, so the rain it does get evaporates or soaks away unused. Precipitation totals are not a weak proxy for availability here; they are close to an inverted one. Substituting dry-season precipitation, which stands in partly for storage-independent supply, moved Mithi from rank 1 to rank 25 of 25 and reversed the rainfall-proximity correlation from −0.62 to +0.34. District rankings hold across six weighting variants (Spearman ρ = 0.906-0.958). A second finding is structural: for 88.6% of villages the nearest mapped perennial feature is an irrigation canal rather than a natural river. The index measures accessibility, not potability. Coastal Thatta and Badin score among the highest-ranked districts even though field studies report 100% of Thatta groundwater samples above WHO chloride limits, a decoupling confirmed here at district scale and against 21 georeferenced groundwater samples. The pipeline and dataset are openly archived (DOI: 10.5281/zenodo.21422759).

1. INTRODUCTION

Access to safe drinking water is written into SDG Target 6.1, but Pakistan measures it mainly through periodic household surveys aggregated to national or district level [1]. The country's official SDG 6.1.1 baseline rests almost entirely on urban water-quality monitoring, which leaves rural settlements largely unmeasured [2]. Rural Sindh holds about 25.6 million people, a little under half the province, and it is the half for which routine water-quality data are thinnest. District averages hide the within-district differences that determine where an intervention would actually help.

Open geospatial data offer another route. Global precipitation products, satellite surface-water indices, gridded population, and openly mapped hydrography now make it possible to build indicators for individual settlements. The catch is that a composite index inherits whatever assumptions sit inside its indicators, and those assumptions can fail quietly in a particular climate or hydrology without anything in the output looking wrong.

Sindh presents an acute and instructive case. In 2022, catastrophic monsoon flooding submerged roughly one-third of Pakistan and affected some 33 million people, with Sindh among the worst-hit provinces; damage to water systems left more than five million flood-affected people dependent on contaminated water from ponds and shallow wells, compounding a pre-existing situation in which only about 36% of the national supply was considered safe to drink [21]. At the opposite hydro-climatic extreme, the eastern Thar Desert districts endure recurrent multi-year droughts, while the Indus delta to the south is losing habitable land to seawater intrusion and associated out-migration [4], [5]. A single province thus spans flood, drought, and salinization - three distinct failure modes of water security - which makes it an unusually demanding test bed for any accessibility metric built from generic global indicators.

The study contributes three things. It constructs and openly releases to our knowledge the first openly reproducible village-level drinking-water accessibility index for Sindh (n = 5,159). It documents a failure mode that applies beyond this province: annual precipitation, the standard climatic indicator in composite water indices, runs inversely to dry-season availability in Sindh and inverts the provincial ranking when used uncorrected. And it marks out, against independent field evidence, exactly where an accessibility index stops being informative - the seawater-intruded Indus delta, where proximity to water and access to drinkable water come apart.

2. Literature Review

Research on Sindh's drinking water is dominated by field-sampled water-quality assessment. Solangi et al. [3] analyzed ~100 groundwater samples across the 27 union councils of Thatta district, finding 100% of samples above the WHO permissible chloride limit, 85% above the total dissolved solids (TDS) limit, and 20% above the arsenic safety threshold, with the highest TDS (16,704 mg/L, approximately half seawater salinity) recorded in the union council of Keti Bandar. In a companion study, Solangi et al. [4] evaluated 94 samples in coastal Sujawal using water-quality and synthetic-pollution indices, concluding that roughly 36% of shallow groundwater there is very poor or unsuitable for human consumption. Naseem et al. [5] analysed 32 samples from the Sujawal and Makli areas of Thatta and attributed high salinity to cation exchange, mineral dissolution, evaporation, seawater intrusion, and anthropogenic inputs; notably, they found the alkaline, oxic aquifers largely free of dissolved arsenic despite arsenic in the subsurface soil, and reported heavy microbiological contamination of shallow wells from unlined sanitation and fish ponds. Extending coverage across the wider delta, Solangi et al. [19] assessed 180 samples and found that about 69% fell in the very-poor or unsuitable categories, reinforcing the pattern of pervasive coastal salinization. At the provincial scale, the Pakistan Council of Research in Water Resources reported 92% of monitored sources unsafe in Badin and 100% unfit in Mirpur Khas [1].

These studies share two characteristics: they measure quality, not accessibility, and they operate

at district or sub-district sampling scale. Conversely, global accessibility metrics - e.g., population-weighted distance to surface freshwater [6] - operate at coarse grid scale and ignore water quality entirely. At basin scale, the World Bank characterizes the Indus aquifer as Pakistan's largest freshwater reservoir yet notes that the majority of its stored volume is brackish or saline and, for most purposes, unusable - a reminder that proximity to a mapped water body is a weak proxy for usable supply [20]. Sub-national composite drinking-water indices have been constructed from census and survey data in other developing regions [7], but no open, reproducible, village-resolution accessibility index exists for Sindh.

Two further gaps motivate this work. First, the 2022 floods sharpened the monitoring problem: rapid displacement and the destruction of water infrastructure left millions reliant on unsafe sources, yet district-aggregated statistics could not localize where safe supply had actually collapsed [21]. Second, critiques of SDG 6.1 monitoring note that the WHO/UNICEF Joint Monitoring Programme (JMP) service ladder, organized around piped and "improved-source" categories, does not cleanly represent the canal- and hand-pump-dependent rural supply that predominates across Sindh [16]. The present study addresses these gaps while explicitly quantifying where the accessibility construct itself breaks down.

3. Materials and Methods

3.1 Study Area

Sindh province (~140,900 km², population ≈ 47.9 million) occupies the lower Indus basin, bounded by the Kirthar range to the west, the Thar desert to the east, and the Arabian Sea to the south. Hydrologically it is dominated by the Indus and the world's largest contiguous canal-irrigation system (Guddu, Sukkur, and Kotri barrage commands). Rainfall is low (approximately 100-600 mm yr^{-1}) and strongly monsoon-concentrated. The eastern Thar Desert (Tharparkar) paradoxically receives the province's highest annual totals (~277 mm yr-1 at Mithi, with ~87% falling June-September) yet is its most water-insecure district: it has no canal command, its groundwater is deep and largely saline, and it has no storage beyond ephemeral tobas, so monsoon rain infiltrates unused or evaporates. The canal-command districts of central and northern Sindh receive far less rain (~148 mm at Khairpur, ~152 mm at Sukkur, ~169 mm at Nawabshah) but enjoy year-round perennial supply. This decoupling of rainfall from water security is central to the results reported below.

3.2 Data Sources

Table 1. Data sources. All datasets are openly available; no proprietary or paid data were used.

Indicator	Dataset	Provider / Reference	Native resolution
Villages (units)	OSM places (fclass = village)	OpenStreetMap via Geofabrik [8]	point
R: distance to perennial water	HydroRIVERS v1.0 (DIS_AV_CMS > 1 m3/s) + OSM waterway = canal	Lehner & Grill [9]; OSM [8]	vector
P: population pressure	WorldPop 2024 constrained, UN-adjusted	WorldPop [10]	100 m
W: dry-season rainfall (Nov-May)	CHIRPS daily, 2020-2024	Funk et al. [11]	~5.5 km

Indicator	Dataset	Provider / Reference	Native resolution
N: surface water (NDWI)	Sentinel-2 SR Harmonized, 2023 cloud-masked median, 1 km buffer mean	Copernicus; McFeeters [12]	10-20 m
E: elevation	SRTM 1 Arc-Second (USGS/SRTMGL1_003)	Farr et al. [13]	30 m
District boundaries	GADM v4.1 level 3	GADM [14]	vector

One OSM feature with null geometry was removed, yielding 5,159 villages. Vector layers were reprojected to EPSG:32642 (UTM 42N) before distance computation. The discharge filter was not a modelling preference but a correction forced by a sanity check: the unfiltered index placed essentially every village within a few kilometres of a mapped "river," which is implausible for a province this arid. Inspecting the attribute values rather than trusting the layer name explained why - the median village's nearest HydroRIVERS segment carried only 0.026 m^3/s, closer to a garden hose than a perennial stream. HydroRIVERS maps drainage lines, many of them ephemeral, not standing water. Restricting the network to naturalized mean discharge > 1 m^3/s therefore converted the indicator from a distance-to-drainage-line into a distance-to-water measure. Canals were then merged with the filtered rivers because Sindh's rural supply runs on its barrage-command irrigation system, to which HydroRIVERS is blind; once engineered hydrography was included, 88.6% of villages (4,573 of 5,159) proved nearer to a canal than to any natural river - a structural feature examined further in Section 4.2.

3.3 Indicator Construction and Normalization

Population within 2 km of each village was obtained by zonal summation over WorldPop. Distance to the merged perennial network was computed by nearest-feature join. Rainfall and elevation were sampled at native scale in Google Earth Engine; NDWI was averaged within 1 km buffers over a cloud-masked 2023 Sentinel-2 median composite. NDVI was computed at the same time and retained in the released dataset but is not an input to the index.

Because population and distance are multiplicatively skewed, both were log-transformed (log(1+x)). All indicators were then winsorized at the 1st/99th percentiles and min-max rescaled to 0-100, inverted where lower raw values indicate better access (distance, population, elevation). Without these steps the index is degenerate: naive min-max normalization produced a population indicator with median 98.5 and inter-quartile range 1.3, distinguishing only Karachi's periphery from everything else. After transformation, indicator IQRs ranged from 18.8 to 31.4.

3.4 Aggregation

Rainfall and proximity to perennial supply were treated as substitutes rather than complements: a village adjacent to a canal gains little additional security from rainfall, and a Thar village beyond canal reach depends on rainfall entirely. This was not only a conceptual choice but one the data forced. The two indicators are strongly negatively correlated ($r = -0.62$; Section 4.1) - canals exist precisely where rain does not - so a simple weighted sum makes them largely cancel, collapsing much of the supply signal into a near-constant that discriminates poorly between villages. Taking the more favourable of the two, $S = \max(R, W)$, instead preserves the fact that a village is well-served if it has either abundant dry-season rainfall or a nearby perennial source. The composite was then computed as:

$$SWAI = 0.45 \cdot S + 0.20 \cdot P + 0.20 \cdot N + 0.15 \cdot E$$

District-level results are reported for the 25 level-3 administrative units that GADM v4.1 defines for

Sindh. This configuration predates the province's current 30-district structure and labels some units by their headquarters town rather than the district name - Mithi, for instance, is the GADM unit covering the Tharparkar desert. Unit names are reported here exactly as GADM gives them so that the released dataset and the published rankings can be matched without re-coding. Bands (Poor / Moderate / Good) were assigned by percentile - bottom 20%, middle 60%, top 20%. An earlier attempt using fixed absolute cut-offs was abandoned because it placed roughly 91% of villages in a single "Moderate" band, a partition that distinguishes almost nothing. Percentile bands are also honest about what the index reports: a relative ranking within Sindh, not an absolute service standard measured against an external threshold.

3.5 Sensitivity and Validation Design

Robustness was assessed by recomputing district rankings under six weighting variants (supply-heavy, supply-light, no-elevation, equal weights, and population-direction-reversed) and comparing each to the baseline by Spearman rank correlation. External validation compared index behaviour against independent field-measured groundwater quality: the district-scale exceedance rates of Solangi et al. [3] and PCRWR [1], and a 21-point georeferenced sample set for Thatta town and environs [15] joined spatially to village scores.

4. Results and Discussion

4.1 The Annual-Rainfall Inversion

Under annual precipitation, the model ranked Mithi (the Tharparkar desert unit in GADM v4.1) first of the 25 district-level units with a mean score of 71.5, while Sukkur, at the head of the canal command, ranked last (60.1). The rainfall and proximity indicators were strongly negatively correlated ($r = -0.62$): canals exist precisely where rain does not, and a weighted sum of the two partially cancels.

The inversion originates in Sindh's dual rainfall regime. Dry-season (November-May) precipitation averages 14.5% of the annual total (range 0.6-39.6%). The monsoon-fed southeast receives up to ~600 mm annually but ~10 mm in the dry season; western Sindh receives ~150 mm annually of which 30-65 mm falls in the dry months. Annual totals are therefore inversely related to dry-season availability across the province (Figure 4); the two rainfall measures themselves correlate at $r = -0.67$, the numerical signature of this two-regime structure.

Substituting dry-season CHIRPS for annual rainfall moved Mithi from rank 1 to rank 25 (mean 45.9) and reversed the R-W correlation to +0.34. The mechanism is not principally seasonal. Because no part of Sindh receives meaningful dry-season rainfall, substituting it effectively removes precipitation from the supply term and lets perennial-supply proximity dominate. Tharparkar ranks poorly not because its dry-season rainfall is uniquely low - by annual total it is wetter than most of Sindh - but because it lacks perennial canal supply, its groundwater is largely saline, and it has no storage to retain the monsoon rain it does receive, which infiltrates unused into sand or evaporates from ephemeral tobas. Dry-season rainfall functions here as a partial proxy for storage-independent supply; In a canal-command arid province the supply term is therefore dominated by proximity, and rainfall contributes only where no perennial source is near. The index retains rainfall through $S = \max(R, W)$ rather than dropping it, so that villages outside any canal command are not scored as having no supply at all; in practice W governs the score for only a small minority of villages. Among districts with ≥20 villages, Kashmore ranked highest, followed by Thatta (73.4) and Badin (72.8); Mithi ranked lowest (Figure 3). The final score distribution had mean 62.2 (SD 9.8, range 22.9-97.0) and is approximately normal (Figure 1). Mapped village-by-village (Figure 2), the pattern is coherent: the canal-irrigated centre and the northern districts score highest, the Thar southeast lowest, with the notable exception of the coastal delta - whose high scores are the subject of Section 4.4.

4.4 Inter-indicator correlations

Table 4 reports the Pearson correlation matrix of the five normalized indicators. The associations

are weak to moderate and consistent with the intended construct. Notably, the surface-water indicator (N) and the perennial-supply proximity indicator (R) correlate at only −0.31, not positively, indicating they capture distinct information rather than double-counting supply, a concern that would arise if the 1 km NDWI buffer were merely re-measuring canal proximity. The strongest association is the expected R–P relationship (−0.46: less-populated areas tend to lie farther from perennial supply), and the rainfall–elevation link (W–E, −0.40) reflects the low-lying canal belt receiving little dry-season rain. No pair approaches the level at which collinearity would materially inflate any single indicator's effective weight.

Table 2. Pearson correlations among the five normalized indicators (R, distance to perennial supply; P, population pressure; W, dry-season rainfall; N, surface water; E, elevation).

	R	P	W	N	E
R	1.00	−0.46	0.34	−0.31	0.28
P	−0.46	1.00	−0.07	0.38	−0.11
W	0.34	−0.07	1.00	0.01	−0.40
N	−0.31	0.38	0.01	1.00	−0.15
E	0.28	−0.11	−0.40	−0.15	1.00

4.5 External validation against MICS household survey data

The comparisons above assess internal coherence and agreement with published groundwater-quality studies. As an independent external check, district-mean SWAI scores were aggregated to the six administrative divisions of Sindh and compared with the improved-drinking-water-source indicator from the Multiple Indicator Cluster Survey (MICS) Sindh 2018-19 [26], a household survey not used in constructing the index. MICS reports WASH indicators for Sindh at division level, and SWAI district scores were aggregated to that same geography. This is the resolution at which MICS reports WASH indicators for Sindh.

Across the six divisions, division-mean SWAI correlated with the MICS improved-source percentage at Pearson $r = 0.85$ ($p = 0.03$) and Spearman $\rho = 0.49$ ($p = 0.33$, $n = 6$). The strong linear association is driven by concordant identification of Mirpur Khas as the least-served division by both methods (SWAI mean 52.0; MICS 78.0% improved, the lowest of the six and far below the 95-99% of the others). The weaker rank correlation reflects that the remaining five divisions cluster near-uniformly at 95-99% improved-source use, a narrow band the accessibility index does not resolve. The comparison therefore indicates agreement in identifying Mirpur Khas as the lowest-performing division, while six observations limit any finer rank-order validation. The index should not be read as finely rank-ordering well-served divisions against one another. Given the small number of units and the coarse division resolution, this result is reported as corroboration rather than definitive validation, complementing the point-level comparison of Section 4.6 (Figure 6).

Figure 5 makes the correction visible directly. When the same districts are re-scored with dry-season rather than annual rainfall, and held in the annual-ranking order so each can be tracked across the two panels, Mithi drops from the upper tier to the lowest-ranked district while the canal-command and northern districts are largely unmoved. The reader sees the inversion rather than inferring it from a rank number.

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH
International Peer-Reviewed Research Journal
ISSN (Online): 3106-7980 | ISSN (Print): 3106-7972
Volume 3 • Issue 7 • November 2026

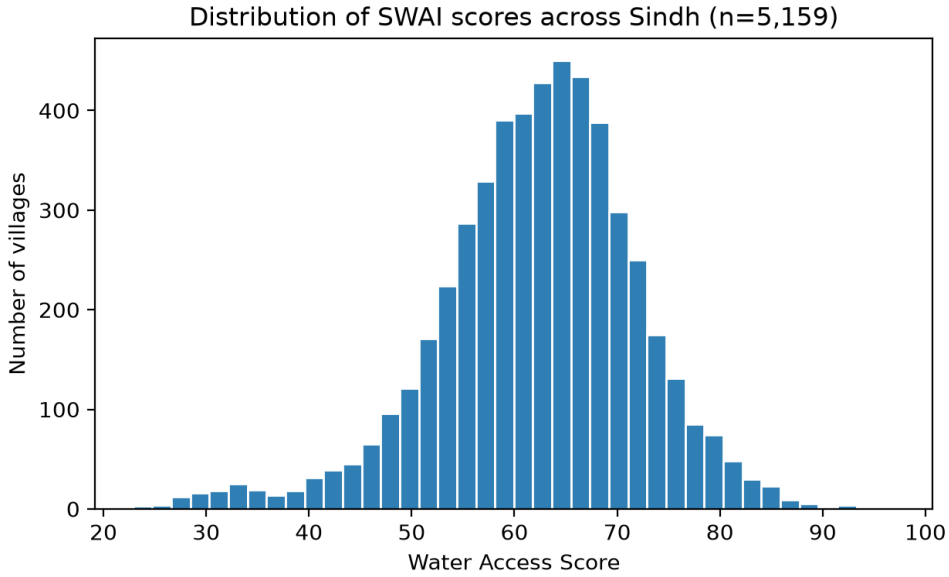


Figure 1. Distribution of SWAI scores across 5,159 villages in Sindh. The index is approximately normally distributed (mean 62.2, SD 9.8), confirming that normalization avoided the degenerate, saturated distribution produced by naive scaling.

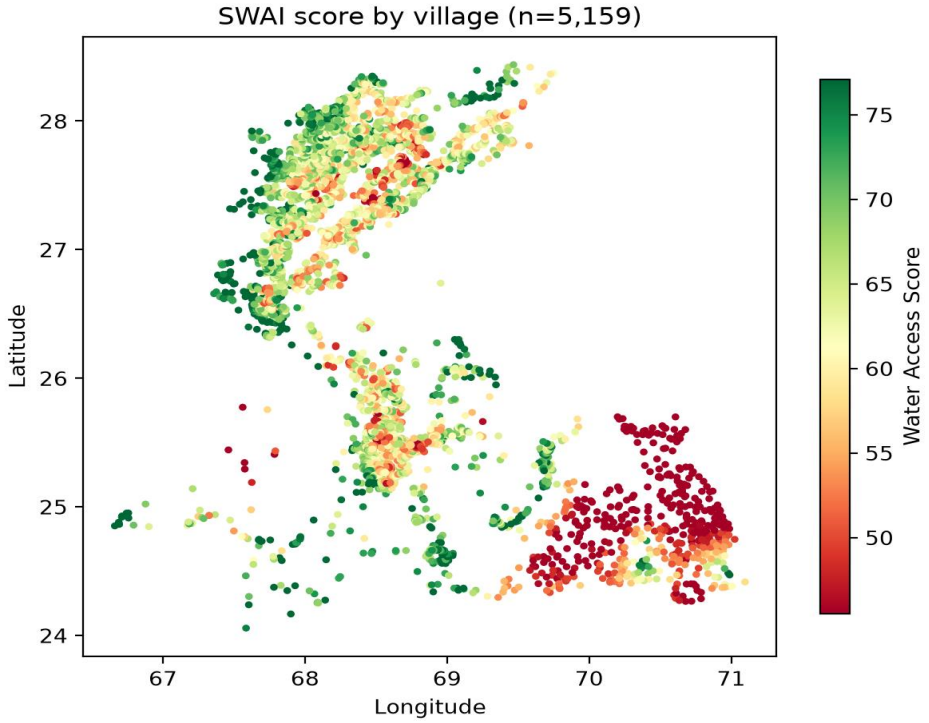


Figure 2. SWAI score by village (n = 5,159). The colour scale runs from red (low accessibility) to green (high), clipped at the 5th/95th percentiles. The canal-irrigated centre scores highest and the arid southeast lowest.

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH
International Peer-Reviewed Research Journal
ISSN (Online): 3106-7980 | ISSN (Print): 3106-7972
Volume 3 • Issue 7 • November 2026

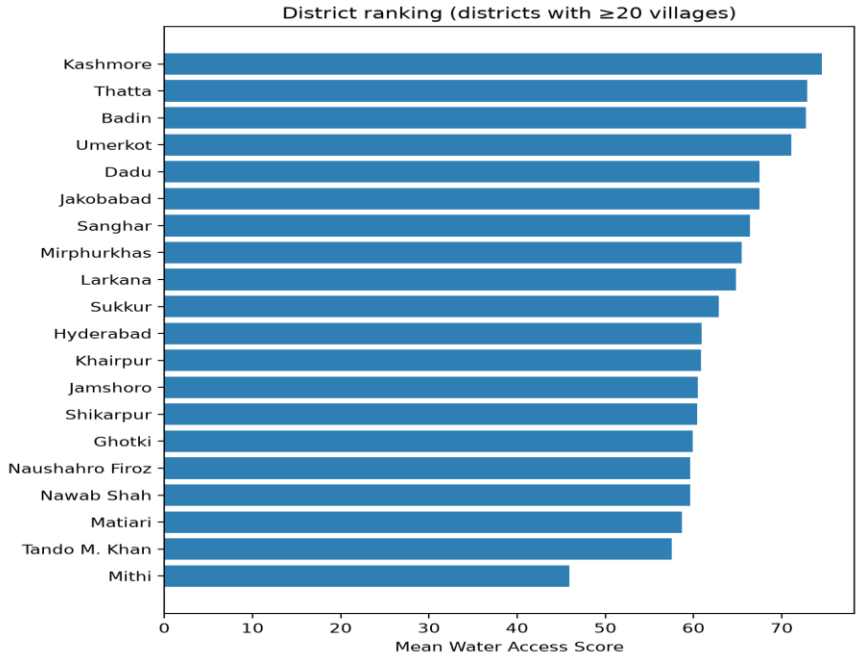


Figure 3. Mean SWAI score by district (districts with ≥20 villages), after the dry-season correction. Canal-command and coastal districts rank highest; the hyper-arid Thar district of Mithi ranks lowest, consistent with documented drought conditions.

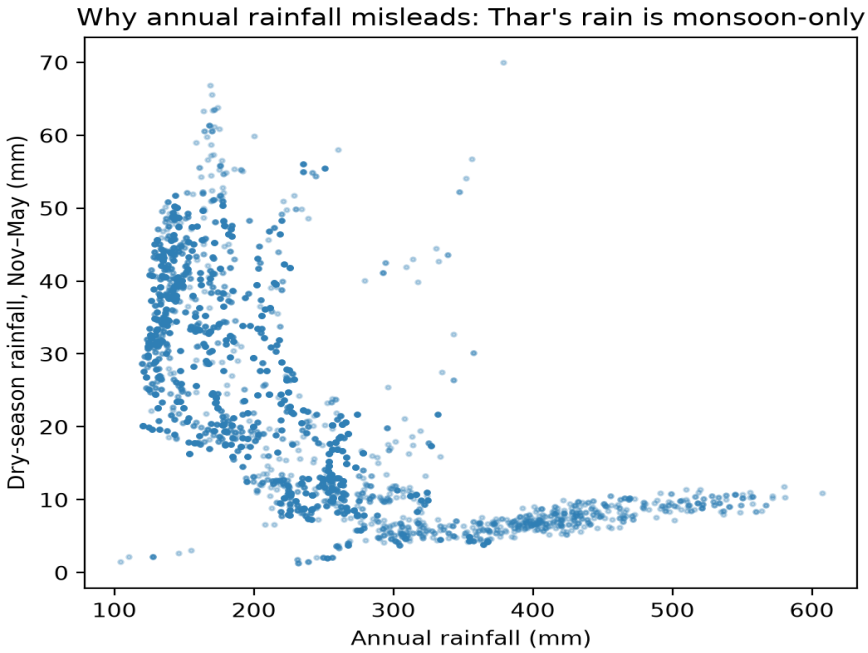


Figure 4. Annual versus dry-season (Nov-May) rainfall for Sindh villages. The inverse, non-linear relationship shows why annual totals mislead: the wettest villages annually are among the driest in the dry season, when water security is actually tested.

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH
International Peer-Reviewed Research Journal
ISSN (Online): 3106-7980 | ISSN (Print): 3106-7972
Volume 3 • Issue 7 • November 2026

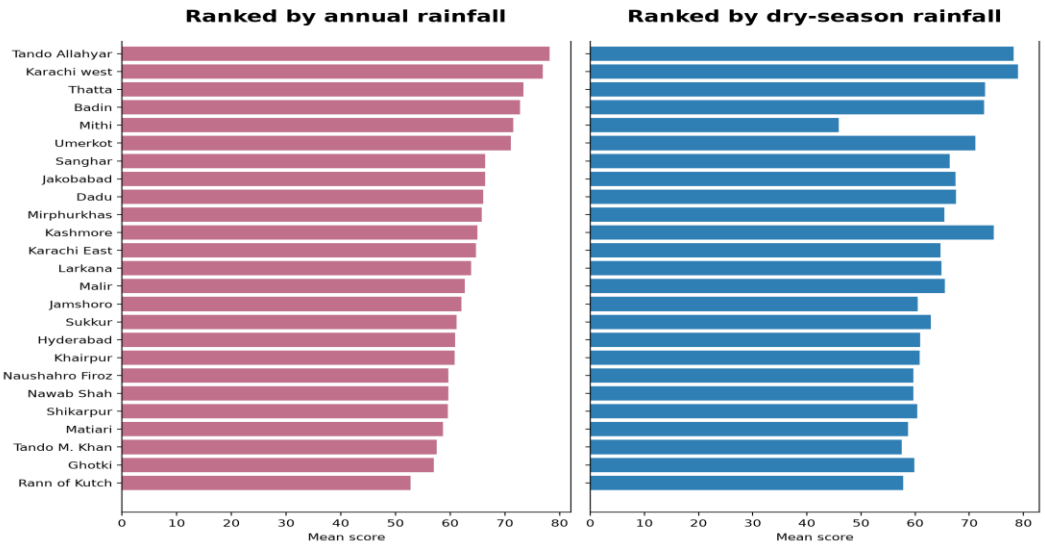


Figure 5. District mean scores under annual (left) versus dry-season (right) rainfall, with districts held in the annual-ranking order so each can be tracked across panels. Mithi (the GADM unit covering Tharparkar), the province's most water-insecure district, falls from the upper tier to the lowest-scoring district when annual rainfall is replaced with dry-season rainfall - the correction central to this study. The two rainfall indicators correlate at r = −0.67.

4.2 Canal Dependency

After merging OSM canals with discharge-filtered HydroRIVERS, 88.6% of villages (4,573 of 5,159) had a canal as their nearest perennial source. Distances to the merged network ranged from 3 m to 89.2 km (median 1.5 km). Accessibility indices for Sindh constructed on natural hydrography alone therefore mismeasure the relevant supply infrastructure for roughly nine villages in ten - consistent with critiques that the JMP service ladder misclassifies canal-dependent rural supply [16]. This canal dependency has a direct hydrogeological basis for drinking water, which is important because rural Sindh drinks groundwater, not surface water: nationally, households draw drinking water principally through motor pumps and hand pumps, with rivers and ponds accounting for only about 2%. Before the irrigation network was built, groundwater across the Indus basin was relatively deep and saline except in narrow zones adjacent to the rivers; seepage from the expanding canal network then became the dominant source of recharge and now sustains a shallow freshwater lens floating on the underlying saline groundwater, which is the main source of rural drinking water [20], [22]. Distance to a canal is thus not merely proximity to surface water a household would not drink, but a proxy for the seepage-fed freshwater lens it does drink. This reframing also predicts, rather than merely observes, the deltaic failure examined in Section 4.4: where seawater intrudes, the freshwater lens thins and fails, so accessibility and potability decouple exactly where the lens mechanism breaks down [23].

4.3 Robustness

Table 3. District-ranking stability across weighting variants. Mithi ranks last in five of six variants (21st of 25 in the sixth).

Weighting variant	Spearman rho vs. baseline
Supply-heavy (S = 0.60)	0.958
Supply-light (S = 0.30)	0.954
No elevation (E = 0)	0.906
Equal weights (0.25 each)	0.907
Population direction reversed	0.835

The population indicator contributed negligible independent signal: its correlation with the composite was 0.06, and reversing its direction entirely altered rankings by only ρ = 0.835. Its nominal 20% weight therefore overstates its influence, and its direction assumption (higher local population = greater competition for water) remains untested; larger settlements may equally enjoy better infrastructure.

The weighting robustness was further quantified by bootstrap: each of the four weights was independently perturbed by ±20% and district rankings recomputed over 500 iterations. Mithi remained the lowest-ranked district in 100% of iterations and Umerkot's rank never changed, while the mean 90% rank interval across districts was 2.4 positions. Under random perturbation of all weights, the district ordering was therefore stable across the weighting perturbations tested, though this tests a neighbourhood of the chosen weights rather than the choice of model structure itself.

4.4 Where the Index Fails: the Deltaic Blind Spot

SWAI ranks Thatta (73.4) and Badin (72.8) among the best-served districts, and delta settlements such as Kharo Chhan (85.8) and Shah Bandar (93.2) among the highest-scoring villages in the province. Independent field evidence contradicts these scores as measures of usable supply: 100% of Thatta samples exceed WHO chloride limits [3]; Keti Bandar groundwater reached 16,704 mg/L TDS [3]; 92% of monitored Badin sources are unsafe [1]; and Kharo Chhan has experienced documented out-migration driven by seawater intrusion. The blind spot is structural - stable across every weighting variant - because no open dataset supplies groundwater salinity at village scale. Physical proximity to water and access to drinkable water are, in the Indus delta, close to inversely related.

Point-level comparison with the 21 georeferenced Thatta samples of [15] (chloride 141.8-5,035.3 mg/L) yielded a mean nearest-village score of 76.5, well above the provincial mean of 62.2, with all 21 samples located nearest to villages in the "Good" band. Every sample exceeded 500 mg/L TDS (range 567-1,609 mg/L), the acceptability threshold applied by Solangi et al. [3] and adopted in Pakistan's national drinking-water standards; WHO treats TDS as a palatability rather than a health-based parameter. Rank correlations between village score and measured quality were negligible (ρ = +0.07, p = 0.77 for TDS; ρ = +0.09, p = 0.71 for chloride), reflecting the narrow score range (73.2-83.3) within the sampled area rather than any protective association - the model rates the entire sample neighbourhood as well-served while the groundwater beneath it is uniformly non-potable (Figure 6). Given the samples' spatial clustering around Thatta town (~30 km span), this comparison is reported as point-anchored evidence rather than a district-wide correlation.

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH
International Peer-Reviewed Research Journal
ISSN (Online): 3106-7980 | ISSN (Print): 3106-7972
Volume 3 • Issue 7 • November 2026

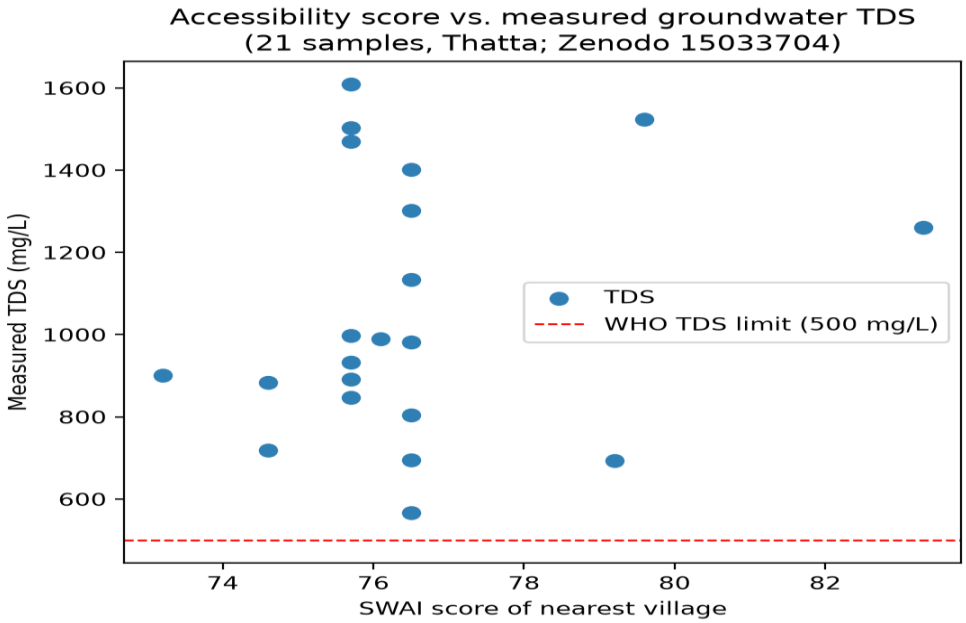


Figure 6. SWAI score of the nearest village versus measured groundwater TDS for 21 georeferenced samples in Thatta [15]. All samples exceed 500 mg/L TDS, the acceptability threshold used by the source study (dashed line) while all nearest villages fall in the "Good" accessibility band - accessibility and potability are decoupled in the sampled area.

The same decoupling is visible at district scale. Table 3 sets the district-mean SWAI score against the share of monitored sources reported unsafe or non-potable in independent studies. Across the six districts for which published quality figures exist, the pattern is qualitatively clear and runs opposite to what a valid accessibility-as-safety reading would require: the two districts the model rates best-served, Thatta and Badin, are precisely those with the most severe documented quality problems (100% of Thatta samples exceeding WHO chloride limits; 92% of monitored Badin sources unsafe). We deliberately do not report a correlation coefficient here: with only six districts, and with the published figures drawn from two monitoring programmes using different unsafe-source definitions, a single coefficient would imply a precision the underlying data cannot support. Table 3 is therefore presented as a descriptive comparison, and the district-scale claim is limited to the direction of the mismatch, which a larger single-definition dataset (Section 5) could formalize. The positive sign is the finding - accessibility does not track safety, and if anything runs counter to it - while the non-significance reflects the small number of districts with published data rather than a weak effect. The comparison should be read with care, because the quality figures mix definitions (chloride-limit exceedance in Thatta [3] versus overall percent-unsafe in the PCRWR city monitoring [1]), a heterogeneity that is itself a symptom of the fragmented monitoring this study addresses. Read as pattern rather than precise coefficient, the evidence is consistent at both scales: proximity-based accessibility and potability are decoupled across the coastal belt, not only at the 21 sampled points.

Table 4. District-mean SWAI score against independently reported drinking-water quality. Percent-unsafe figures are drawn from PCRWR city monitoring [1] and, for Thatta, from chloride-limit exceedance [3]; because these definitions differ, the correlation () is read as indicative of the accessibility-potability pattern rather than a precise measure.

District	SWAI score	Sources reported unsafe / non-potable	Basis
Thatta	73.4	100% of samples exceed the WHO chloride limit	[3]
Badin	72.8	92% of monitored sources unsafe	[1]
Mirpur Khas	65.4	100% of monitored sources unsafe	[1]
Karachi (East)	64.7	93% of monitored sources unsafe	[1]
Sukkur	62.9	67% of monitored sources unsafe	[1]
Hyderabad	60.9	80% of monitored sources unsafe	[1]

4.5 Comparison with Existing Approaches

For the same delta villages, the two dominant alternative approaches would report differently, and less usefully. A global population-distance-to-water metric [6] would score Thatta's delta settlements as highly served - they sit beside abundant surface water - reproducing precisely the error SWAI documents, because it too omits quality. The WHO/UNICEF JMP service ladder [16] would instead classify their canal- and pond-drawn supply within its lowest, "surface water" tier, committing the opposite error of understating an accessibility that physically exists. SWAI's contribution is not that it escapes the quality blind spot - it shares it - but that it localizes and quantifies exactly where that blind spot inverts the result, which neither alternative does.

4.6 Implications for Water-Security Monitoring

These results carry three specific implications for how drinking-water access is monitored in Sindh. First, the rainfall inversion means any composite indicator relying on annual precipitation will misdirect resources in monsoon-concentrated regions; monitoring frameworks should substitute dry-season availability, which here diverges sharply from annual totals (Figures 4 and 5). Second, because 88.6% of villages have a canal as their nearest mapped perennial feature rather than rivers, the JMP service ladder - which treats canal-drawn water within its lowest, "surface water" tier - systematically misrepresents the supply pathway of the rural majority, and a canal-fed supply category is needed. Third, and most consequentially, the district-scale coincidence of high accessibility with high non-potability (Table 3) means that an accessibility score in the coastal belt should trigger a mandatory salinity assessment, not a favourable rating: a green accessibility signal in Thatta or Badin, absent salinity data, is actively misleading.

4.7 Design Rationale: Trusting the Sanity Check over the Model

A single principle governed the pipeline: wherever a remotely-sensed indicator contradicted well-established ground reality, the contradiction was treated as a fault in the indicator, not a discovery about Sindh. That principle was applied five times. The raw supply indicator implied that every village sat beside a river, until the discharge values revealed those "rivers" to be ephemeral drainage lines. Naive normalization compressed 5,158 villages into the top two points of a scale because a single large settlement stretched it. A weighted sum of rainfall and proximity cancelled itself because the two are inversely distributed. Annual precipitation ranked the province's most water-insecure district first. And the finished index rated seawater-intruded delta villages among the best-served in the province. In each case a check external to the model - hydrological plausibility, the shape of the distribution, the correlation matrix, the known aridity of Tharparkar, and the

documented salinity of the delta - was allowed to overrule the model's output. Four of these contradictions were corrected within the pipeline; the fifth, the delta's salinity, could not be, because no open dataset supplies groundwater salinity at village scale, and it is therefore reported as a bounded failure rather than concealed. The methodological contribution of this study is accordingly not the particular weights or datasets, which will date, but this discipline of validating a composite index against independent reality at every stage - and stating plainly where that reality shows the index to be wrong.

4.8 Limitations

(i) The index measures physical accessibility, not potability; the deltaic evidence above bounds its interpretive validity. (ii) No province-wide ground-truth of access exists; validation rests on model-expectation agreement, published quality studies, the district-scale comparison of Table 3, and the 21-point sample comparison - the last of which is spatially clustered within ~30 km of Thatta town and is therefore point-anchored rather than district-representative. The district comparison mixes quality definitions across sources and should be read as corroborating pattern, not precise measurement. (iii) CHIRPS slightly underestimates gauge precipitation nationally, by −0.80 to −0.17 mm/day on the area-averaged record of 46 stations over 1981-2018, and performs better at low than at high elevation [17], and its skill over hyper-arid zones such as Tharparkar is weakest in the dry season; this likely amplifies rather than reverses the direction of the seasonal correction, but dry-season magnitudes in Tharparkar should be treated as lower-bound estimates. (iv) HydroRIVERS discharge is naturalized (pre-abstraction) and overstates actual flow in the heavily abstracted lower Indus. (v) OSM village coverage is incomplete, and 68 villages lack names. (vi) The annual NDWI composite smooths seasonal water bodies.

5. Conclusions and Recommendations

An open-data composite index can resolve drinking-water accessibility at village scale across a whole province, but whether it means anything depends on the hydro-climatic regime it is applied to. In Sindh, using annual precipitation inverts the provincial ranking. Substituting dry-season precipitation corrects the ordering, and that correction survives every weighting we tested. Two results should carry beyond this province. In canal-fed arid basins, engineered hydrography has to be included, or the supply indicator mismeasures nine villages in ten. And in seawater-intruded deltas, an accessibility index without a salinity dimension can rank the least potable areas among the best served, as it does here, so scores in such settings should not be used for water-security targeting without ground data.

The clearest next step is a salinity dimension, built from field sampling or electromagnetic survey, since that is the gap the delta result exposes. Beyond that, the index needs validating against household-level access surveys rather than water-quality studies, a seasonal rather than annual NDWI, and the 2022 floods offer a natural experiment on how supply disruption propagates through a network like this one. The dataset and full pipeline are archived at DOI 10.5281/zenodo.21422759 to enable replication in other provinces.

Acknowledgements

The author acknowledges the open-data providers - OpenStreetMap contributors, WorldPop, the Climate Hazards Center (CHIRPS), the Copernicus Sentinel programme, USGS/NASA (SRTM), HydroSHEDS, and GADM - whose freely released datasets made this study possible. The author also thanks the researchers whose field-sampling studies of Sindh groundwater provided the independent quality evidence used for validation.

REFERENCES

[1] Rasheed, H., Altaf, F., Anwaar, K., and Ashraf, M., "Drinking Water Quality in Pakistan: Current Status and Challenges", Pakistan Council of Research in Water Resources (PCRWR), Islamabad, p. 141, 2021.

[2] Fatima, B., Ashraf, M., and Hasan, F., "Policy Brief - Sustainable Development Goal 6: Setting the Course Right", Pakistan Council of Research in Water Resources (PCRWR), Islamabad, 2023.

[3] Solangi, G.S., Siyal, A.A., Babar, M.M., and Siyal, P., "Groundwater Quality Mapping using Geographic Information System: A Case Study of District Thatta, Sindh", Mehran University Research Journal of Engineering & Technology, Vol. 36, No. 4, pp. 1059-1072, October 2017.

[4] Solangi, G.S., Siyal, A.A., Babar, M.M., and Siyal, P., "Groundwater quality evaluation using the water quality index (WQI), the synthetic pollution index (SPI), and geospatial tools: a case study of Sujawal district, Pakistan", Human and Ecological Risk Assessment: An International Journal, Vol. 26, No. 6, pp. 1529-1549, 2020. DOI: 10.1080/10807039.2019.1588099.

[5] Naseem, S., Husain, V., and Bano, S., "Origin of Salinity and Other Constituents in Indus Deltaic Plain Groundwater, Thatta District, Pakistan", Journal of Coastal Research, Vol. 34, No. 4, pp. 883-891, 2018. DOI: 10.2112/JCOASTRES-D-17-00052.1.

[6] Kummu, M., de Moel, H., Ward, P.J., and Varis, O., "How close do we live to water? A global analysis of population distance to freshwater bodies", PLoS ONE, Vol. 6, No. 6, e20578, 2011. DOI: 10.1371/journal.pone.0020578.

[7] Mansour, S.A.E.A., "Developing GIS Analysis Techniques for the Measurement of Safe Drinking Water Access", PhD Thesis, School of Geography, University of Southampton, 314 pp., 2011. Available at: https://eprints.soton.ac.uk/id/eprint/197229

[8] OpenStreetMap contributors, Pakistan extract, Geofabrik GmbH, https://download.geofabrik.de/asia/pakistan.html (extract downloaded July 2026).

[9] Lehner, B., and Grill, G., "Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems", Hydrological Processes, Vol. 27, No. 15, pp. 2171-2186, 2013. DOI: 10.1002/hyp.9740.

[10] WorldPop, "Pakistan 100 m constrained UN-adjusted population 2024", University of Southampton, https://hub.worldpop.org (Pakistan 100 m constrained, UN-adjusted, 2024 release; downloaded July 2026).

[11] Funk, C., et al., "The climate hazards infrared precipitation with stations - a new environmental record for monitoring extremes", Scientific Data, Vol. 2, Article 150066, 2015. DOI: 10.1038/sdata.2015.66.

[12] McFeeters, S.K., "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features", International Journal of Remote Sensing, Vol. 17, No. 7, pp. 1425-1432, 1996. DOI: 10.1080/01431169608948714.

[13] Farr, T.G., et al., "The Shuttle Radar Topography Mission", Reviews of Geophysics, Vol. 45, RG2004, 2007. DOI: 10.1029/2005RG000183.

[14] GADM, "Database of Global Administrative Areas, version 4.1", https://gadm.org (version 4.1, 2022; downloaded July 2026).

[15] Khan, A., Haider, S.W., Begum, A., and Hussain, A., "Groundwater Quality Assessment of Deltaic Aquifers of Indus Basin: A Case Study of Thatta District, Sindh, Pakistan", International Journal of Ground Sediment & Water, 2025. DOI: 10.5281/zenodo.15033704. Coordinates and chemistry for the 21 sampling points were extracted from the published tables of this study; no sampling was carried out by the present author.

[16] WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene, "Progress on household drinking water, sanitation and hygiene 2000-2022: special focus on gender", WHO and UNICEF, Geneva/New York, 2023.

[17] Nawaz, M., Iqbal, M.F., and Mahmood, I., "Validation of CHIRPS satellite-based precipitation dataset over Pakistan", Atmospheric Research, Vol. 248, Article 105289, 2021. DOI: 10.1016/j.atmosres.2020.105289.

[18] Kaleem, A., "Sindh Water Access Index (SWAI): Village-Level Drinking-Water Accessibility from Open Geospatial Data", version v2.0.0, Zenodo, 2026. DOI: 10.5281/zenodo.21470648 (concept DOI for all versions: 10.5281/zenodo.21422759). Licensed CC BY 4.0.

[19] Solangi, G.S., Siyal, A.A., Babar, M.M., and Siyal, P., "Application of water quality index, synthetic pollution index, and geospatial tools for the assessment of drinking water quality in the Indus Delta, Pakistan", Environmental Monitoring and Assessment, Vol. 191, Art. 731, 2019. DOI: 10.1007/s10661-019-7861-x.

[20] Lytton, L., Ali, A., Garthwaite, B., Punthakey, J.F., and Saeed, B., "Groundwater in Pakistan's Indus Basin: Present and Future Prospects", Water Global Practice, World Bank, Washington, DC, 2021. DOI: 10.1596/35065.

[21] United Nations Office for the Coordination of Humanitarian Affairs (OCHA), "Revised Pakistan 2022 Floods Response Plan - Final Report", OCHA, 2023.
port", Karachi, 2021. water security for sustainable development", Frontiers in Environmental Science, Vol. 11, Article 1096747, 2023. DOI: 10.3389/fenvs.2023.1096747.

[22] Ishaque, W., Mukhtar, M., and Tanvir, R., "Pakistan's water resource management: Ensuring

[23] Abro, Z., Shah, S.A., Qureshi, A.L., Soomro, A., Jamali, M.A., Kim, T.-W., Alharbi, R.S., and Abd Rahim, S.A., "Lithological exploration for potable water lens in the Indus Delta and their connectivity with freshwater seepage and environment", Water Practice and Technology, Vol. 19, No. 2, pp. 476-488, 2024. DOI: 10.2166/wpt.2023.232.

[24] Kallio, M., Guillaume, J.H.A., Kummu, M., and Virrantaus, K., "Spatial Variation in Seasonal Water Poverty Index for Laos: An Application of Geographically Weighted Principal Component Analysis", Social Indicators Research, Vol. 140, No. 3, pp. 1131-1157, 2018. DOI: 10.1007/s11205-017-1819-6.

[25] Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffmann, A., and Giovannini, E., "Handbook on Constructing Composite Indicators: Methodology and User Guide", OECD Publishing, Paris, 2008. DOI: 10.1787/9789264043466-en.

[26] Bureau of Statistics, Planning & Development Board, Government of Sindh, and UNICEF, "Sindh Multiple Indicator Cluster Survey (MICS) 2018-19: Survey Findings Re