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.

Limits of Soil-Moisture-Based Estimation of DCP-Derived Bearing Capacity of Unpaved Surfaces: Temporal and Cross-Site Validation

Petr Kalvoda

petr.kalvoda@unob.cz

University of Defence

Luděk Čičmanec

University of Defence

David Demydenko

University of Defence

Research Article

Keywords:

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11674565/v1

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

Additional Declarations: No competing interests reported.

Limits of Soil-Moisture-Based Estimation of DCP-Derived Bearing Capacity of Unpaved Surfaces: Temporal and Cross-Site Validation

Abstract

Purpose: This study tested whether profile volumetric water content can estimate contemporaneous dynamic-cone-penetrometer-derived California Bearing Ratio (DCP–CBR) of unpaved surfaces across time and sites.
Methods: Site 1 supplied 311 paired observations on 71 dates and Site 2 supplied 167 quality-controlled pairs on 61 dates. A depth median, monotonic log-linear model, ridge regression, and artificial neural network (ANN) were evaluated by date-grouped cross-validation, a final-15-date holdout, and no-recalibration cross-site testing. Raw DCP conversion, replicate-test reliability, PR2 channel stability, rainfall, and a calendar-placebo negative control were audited retrospectively.
Results: Grouped ANN MAE was 2.100 CBR points versus 2.451 for the median ($R^2 = 0.451$). In the forward test, previous CBR performed best (MAE 2.256), while ANN achieved MAE 2.639 and $R^2 = -0.175$. On all Site 2 observations, ANN was worse than the Site 1 median (MAE 4.355 versus 3.941; $R^2 = -1.043$). Common-support restriction reduced ANN MAE to 2.696, but R^2 remained 0.071 and ANN did not outperform ridge directly. Replicate-test SD was 0.73 on the log-CBR scale and ICC was 0.19–0.25. Deep PR2 channels showed large within-day voltage swings. An apparent external weather gain was reproduced by a calendar placebo and was not stable after temporal exclusion.
Conclusion: Water content carried local information but did not provide transferable bearing-capacity estimates. Measurement noise and time–site confounding further limit interpretation; direct penetration testing and prospective, temporally separated validation remain necessary.

Keywords: unpaved surface, California Bearing Ratio, dynamic cone penetrometer, volumetric water content, cross-site validation, artificial neural network

1 Introduction

The bearing response of an unpaved airport surface is relevant to maintenance planning and to the performance of graded areas adjacent to paved movement surfaces. Local loss of strength can promote rutting, uneven deceleration, and damage when a vehicle or aircraft leaves a paved surface. Direct field methods such as the dynamic cone penetrometer (DCP) are therefore used to characterize near-surface resistance,

frequently by converting the penetration index to an estimated California Bearing Ratio (CBR) [1, 2]. The conversion is convenient, but the resulting quantity is an empirical DCP-derived estimate rather than a laboratory CBR.

Water content is a plausible state variable for repeated monitoring. In unsaturated soils, increasing water content is generally associated with reduced matric suction and a lower contribution of suction to effective stress. The strength response nevertheless depends on grading, plasticity, dry density, fabric, stress history, and the soil-water retention relationship. Simultaneous penetration and water-content measurements have consequently shown strong local associations without implying a universal water-content-to-strength law [3]. Recent studies in *Transportation Infrastructure Geotechnology* similarly demonstrate that matric suction, material class, and compaction state materially alter DCP or CBR response [4, 5].

Data-driven CBR models often combine several geotechnical descriptors. Tilahun et al. [6] used moisture, dry density, and elevation in a spatial artificial-intelligence model, while an uncertainty-aware study by Adjei-Yeboah et al. [7] emphasized maximum dry density, optimum water content, and Atterberg limits. Instrument configuration is another source of variability: cone geometry and hammer mass affect penetration resistance and therefore the transferability of empirical CBR relationships [8]. High accuracy within a single dataset cannot, by itself, establish temporal stability or transfer to another site.

An earlier airport study compared linear regression, fuzzy logic, and artificial neural networks (ANNs) using rainfall, temperature, and antecedent bearing-capacity observations [9]. The more flexible models improved local estimates, but profile water content and cross-site validation were not available. The present study addresses a different and deliberately parsimonious question: how much information about contemporaneous DCP–CBR is contained in profile volumetric water content and nominal depth alone, and does that relationship survive a later time block and a spatially separated control site?

The analysis was framed as a retrospective validation study rather than a confirmatory hypothesis test. Four questions were examined:

1. Does water content improve date-grouped estimates over a depth-specific constant?
2. Does any improvement persist in a forward holdout and on all quality-controlled observations at a second site?
3. Does a regularized ANN provide evidence of added value over transparent monotonic and ridge models?
4. Do antecedent rainfall records add information beyond measured water content?

The main methodological contributions are: (i) depth-aligned pairing of profile water content with DCP–CBR; (ii) reconstruction of the DCP conversion from raw blow and penetration increments; (iii) audits of repeatability and PR2 channel stability; (iv) date-grouped, forward, and cross-site validation against temporal and locally calibrated comparators; and (v) a calendar-placebo test that distinguishes plausible meteorological information from shared seasonality. The primary all-observation cross-site test remains separate from post-hoc sensitivities. This design is intended to expose, rather than conceal, limits to transferability.

2 Materials and methods

2.1 Study design and sites

Two grass-covered unpaved sites in the Czech Republic were analysed. Site 1 was an unpaved area at an airport and served exclusively for model development, grouped validation, and the forward holdout. Site 2 was a non-airport control area and served as the spatial transfer test. Data from the two sites were not pooled when fitting the transferable models.

The national Evaluated Soil-Ecological Unit (BPEJ) codes were 4.11.10 at Site 1 and 5.12.00 at Site 2 [10, 11]. These mapping units provide pedological and terrain context but are not engineering classifications of samples from the measurement points. No site-specific particle-size distributions, Atterberg limits, specific gravity, or dry bulk density were available. Consequently, neither USCS nor AASHTO classes were assigned retrospectively; doing so from regional mapping alone would create unsupported precision. Table 1 distinguishes observed or mapped information from parameters that remain unknown.

Table 1 Site context, sampling role, and available soil information.

Characteristic	Site 1	Site 2
Analytical role	Model development, grouped validation, forward holdout	Primary cross-site transfer test
Surface context	Grass-covered unpaved airport area	Grass-covered non-airport control area
BPEJ code	4.11.10	5.12.00
Mapped soil context	Luvisol-type soil on loess-derived cover	Luvisol-to-Cambisol context on mixed slope deposits
Climatic region in BPEJ	Region 4: moderately warm, dry	Region 5: moderately warm, moderately humid
Mapped texture and profile	Medium texture with heavier subsoil; deep profile	Medium-to-heavy texture; deep profile
Mapped slope	3–7°	0–3°
Mapped skeletal fraction	$\leq 10\%$	$\leq 10\%$
Paired-data period	22 January–15 December 2023	14 September 2021–12 November 2023
USCS/AASHTO class, Atterberg limits, grading, dry density	Not measured at the sampling point	Not measured at the sampling point

BPEJ describes a national mapping unit, not a laboratory-tested specimen. Approximate pedological wording is provided only as site context and was not used as a model input.

2.2 Measurement concept and depth pairing

Volumetric water content was measured with a PR2/6 profile probe at nominal depths of 10, 20, 30, 40, 60, and 100 cm. DCP soundings were made within approximately 1 m of the access tube. The separation reduced the risk of damaging the tube but introduced an unavoidable support-volume mismatch: the PR2 senses a local volume around the tube, whereas DCP soundings sample adjacent penetration paths.

The analytical unit was a site–date–depth pair $(\theta_{\mathrm{v}}, \mathrm{CBR})_d$ at $d = 10, 20, 30, 40$, or 60 cm (Fig. 1). The 100-cm water-content reading had no comparable CBR target and was excluded from modelling. At Site 2, multiple DCP soundings on the same date and depth were summarized by their median, preventing dates with more soundings from receiving disproportionate weight.

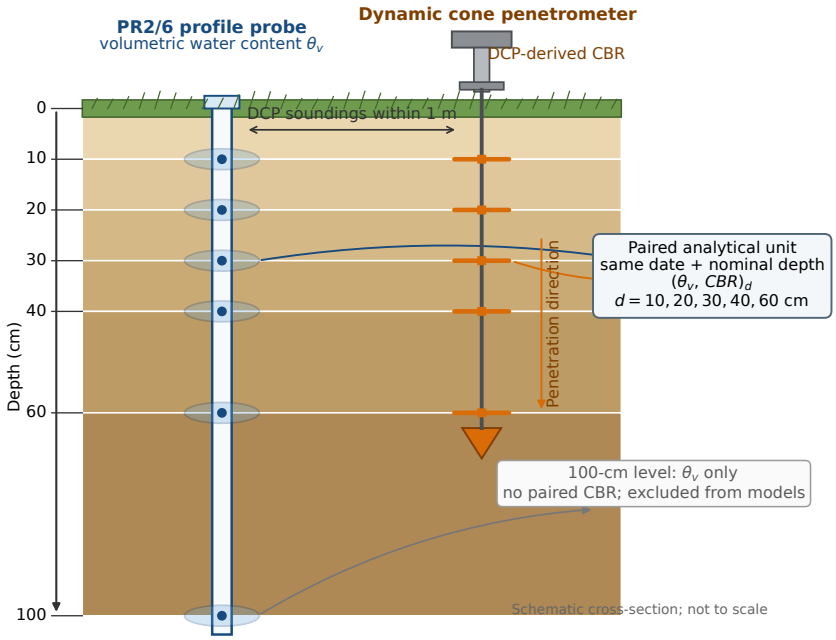

Fig. 1 Schematic measurement concept. PR2/6 water-content readings (blue) were aligned by date and nominal depth with DCP-derived CBR values (orange). DCP soundings were taken within approximately 1 m of the access tube. The 100-cm water-content level had no paired CBR target. The drawing is not to scale.

2.3 Profile water content and quality control

The PR2/6 probe was connected to a DL6 logger. Its electromagnetic response is sensitive to the dielectric properties of the surrounding soil, which are strongly influenced by water content [12, 13]. Water content is reported here as volumetric percent, $100 \times \mathrm{m}^3\,\mathrm{m}^{-3}$.

For raw-voltage quality control, the manufacturer's generic mineral-soil polynomial was used:

$$\theta_{\mathrm{v}} = -0.057 - 0.66V + 8.00V^2 - 27.91V^3 + 49.23V^4 - 42.46V^5 + 14.47V^6, \qquad (1)$$

where V is voltage and θ_v is in $m^3\,m^{-3}$ [12]. The sixth-order coefficient was fixed at 14.47. Readings with $V > 1.088$ V or $\theta_v > 1.00$ $m^3\,m^{-3}$ were flagged as outside the manufacturer's conversion range. They were neither truncated to 100% nor imputed. Twelve of 358 Site 2 voltage readings (3.4%) were flagged; four had a paired CBR and were removed before aggregation.

An archived DL6 export from Site 2 (8–29 May 2023; 2,025 timestamps per channel) enabled a separate stability audit. For each depth, valid connected readings, the share exceeding 1.088 V, median voltage, and the median and maximum within-day voltage range were calculated. This short diagnostic period was not used to delete additional analytical pairs; it was used to assess whether nominally valid values could nevertheless be affected by unstable probe–soil contact. The generic calibration was not checked gravimetrically, and possible access-tube air gaps could not be quantified. These limitations are relevant because site-dependent calibration error can appear as failed spatial transfer.

2.4 DCP measurement and reconstruction of DCP-derived CBR

Bearing response was measured with a Kessler K-100 DCP fitted with a 60° cone, a 4.6-kg hammer, and a nominal drop height of 574 mm. The MR100 logger exported penetration information and a derived CBR. ASTM D6951/D6951M permits the 4.6-kg configuration for softer materials but requires the instrument configuration and conversion to be documented [1, 14].

Raw text files were available for an audit subset. Each penetration increment contained a blow count N_j and cumulative depth z_j in inches. Penetration per blow for the 4.6-kg hammer was reconstructed as

$$\mathrm{PR}_{4.6,j} = 25.4\,\frac{z_j - z_{j-1}}{N_j} \quad [\mathrm{mm\,blow^{-1}}], \tag{2}$$

then converted to an 8-kg-equivalent index using the factor specified for the lighter hammer,

$$\mathrm{DCPI}_{8,\mathrm{eq},j} = 2\,\mathrm{PR}_{4.6,j}, \qquad \mathrm{CBR}_j = \frac{292}{\mathrm{DCPI}_{8,\mathrm{eq},j}^{1.12}}. \tag{3}$$

For a nominal depth d, the assigned value was the increment spanning that level, $z_{j-1} < d \le z_j$, rather than an average over an arbitrary fixed window. This definition explains why a single increment can supply the value at a nominal depth. It also makes the target discrete: integer blow counts and depth increments map to repeated horizontal CBR levels. For example, penetrations of 4, 7, 8, 9, 10, and 11 mm per blow with the 4.6-kg hammer produce recurring CBR values of approximately 28.4, 15.2, 13.1, 11.5, 10.2, and 9.2%.

The archive contained 141 raw files representing 103 unique files by SHA-256 checksum; every unique file reported a 4.6-kg hammer and 574-mm drop. Device metadata labelled 90 unique files as CL and 13 as other, but laboratory classification was unavailable and the metadata label was not used as a predictor. Forty dates

could be matched unambiguously to complete five-depth exported profiles. The reconstructed and exported values differed by a median profile MAE of 0.15 CBR points (90th percentile 0.57); 34 of 40 profiles had MAE ≤ 0.50. This audit supports the generic conversion for the matched subset but does not establish the firmware branch for every analytical date or validate DCP–CBR against laboratory CBR. Throughout the paper, the outcome is therefore termed *DCP-derived CBR*.

2.5 Replicate-test reliability audit

Several Site 2 dates included independent DCP soundings associated with one water-content reading. Repeatability was therefore quantified before the date–depth median was formed. For each date–depth cell g with at least two tests, $x_{ig} = \log(\mathrm{CBR}_{ig})$ and the pooled within-cell standard deviation was

$$s_{\mathrm{within}} = \sqrt{\frac{\sum_g \sum_i (x_{ig} - \bar{x}_g)^2}{\sum_g (n_g - 1)}}. \tag{4}$$

A one-way random-effects ICC(1) was estimated within nominal depth. The primary audit used the CBR value from the increment spanning the nominal depth, consistent with Eq. 3; a consecutive-blow-window aggregation was retained as a sensitivity analysis. The audit comprised 43 repeated date–depth cells and 93 residual degrees of freedom. Site 1 had no corresponding replicate series, so reliability could not be assumed to be identical there. Under the classical measurement-error interpretation, ICC describes attenuation in a single sounding and an approximate ceiling for cell-level explanatory power; it is not a universal deterministic upper bound for every model or aggregation.

2.6 Analytical dataset

Site 1 contributed 311 pairs from 71 dates: 71 observations at 10 cm and 60 observations at each of 20, 30, 40, and 60 cm. Site 2 contributed 236 valid paired soundings after PR2 quality control. Median aggregation by date and depth yielded 167 analytical observations from 61 dates.

The full set of 167 Site 2 observations was specified as the primary spatial transfer test because it reflects actual deployment without silent deletion of difficult cases. A depth-specific common-support flag was then defined from the minimum and maximum Site 1 water content. Of the Site 2 observations, 138 were inside and 29 outside those ranges. The restricted analysis was treated as post-hoc sensitivity, not as the primary external result.

2.7 Models and comparator hierarchy

The main task was contemporaneous estimation,

$$\widehat{\mathrm{CBR}}_i = f(\theta_{\mathrm{v},i}, d_i), \tag{5}$$

rather than forecasting future CBR. Four models were evaluated:

1. **Depth-specific median baseline:** the median CBR at each depth in the relevant training data, without water content.
2. **Monotonic log-linear model:** $\log(\mathrm{CBR}) = \alpha_d + \beta_d \theta_v$, fitted separately by depth with $\beta_d \leq 0$. This transparent benchmark encoded only the expected direction of the water-content relationship.
3. **Ridge regression:** depth-specific linear or quadratic water-content curves. Polynomial degree (1 or 2) and L_2 penalty $\alpha \in \{0.01, 0.1, 1, 10, 100\}$ were selected within grouped validation using MAE.
4. **ANN:** a feed-forward multilayer perceptron with standardized water content, one-hot encoded depth, ReLU activation, L-BFGS optimization, and standardized target. Candidate hidden layers were (4), (8), and (8,4), with the same penalty grid. The final cross-site model used one eight-neuron layer and $\alpha = 0.01$; predictions were averaged across seven prespecified random initializations.

Two context-specific comparators were added. In the forward holdout, the previous observed CBR at the same depth represented an exploratory persistence-type benchmark. It was not used in grouped or cross-site validation because observation intervals were irregular and the study task was contemporaneous estimation. At Site 2, a depth-specific median estimated by grouped cross-validation within Site 2 represented the performance achievable after minimal local calibration. All models were implemented in Python using scikit-learn [15].

2.8 Validation, metrics, and uncertainty

Site 1 internal performance was estimated by nested grouped cross-validation. The outer loop had five folds and produced one out-of-sample estimate for each of the 311 observations. Hyperparameters were selected in an inner four-fold loop. All depths from the same date remained in the same fold, limiting leakage between measurements from one profile [16].

For the forward test, the earliest 56 dates (240 pairs; 22 January–6 November 2023) were used for fitting and tuning, while the final 15 dates (71 pairs; 7 November–15 December 2023) remained untouched until evaluation. Models refitted to all Site 1 data were then applied without recalibration to all Site 2 observations.

Primary metrics were MAE, RMSE, and R^2; Spearman's ρ was included to assess ranking. Model differences were evaluated directly, not inferred from whether separate intervals did or did not cross zero. Paired MAE differences were resampled by complete measurement date using 10,000 cluster-bootstrap replicates. The intervals are exploratory because the analysis was retrospective, several models and protocols were compared, and no multiplicity adjustment was applied.

Calibration was described using mean bias, $\overline{\widehat{y} - y}$, and

$$y_i = a + b\widehat{y}_i + \varepsilon_i. \tag{6}$$

Ideal calibration gives bias 0, $a = 0$, and $b = 1$. With this regression direction, $b < 1$ indicates that fitted changes are too large relative to their calibration

information; it is not, by itself, evidence that the marginal prediction distribution is compressed. Prediction-to-observation standard-deviation ratios were therefore reported separately.

2.9 Exploratory meteorological extension

Daily rainfall recorded locally at Site 1 was compared with quality-controlled daily precipitation from a professional station 4.69 km away, obtained from the Czech Hydrometeorological Institute open-data archive [17]. The 2023 totals were 484.7 and 529.7 mm, respectively. Same-date daily correlation was 0.275, whereas shifting the local record forward by one day increased correlation to 0.817, suggesting a difference in date convention. Official station dates were retained for the primary sensitivity analysis; the one-day shift was labelled post hoc.

At Site 2, blank rainfall cells were confirmed to represent 0 mm. Natural rainfall totalled 483.6 mm over 2023, with 74 wet and 291 zero-rainfall days. Two experimental watering events of 20 mm each were flagged separately and excluded from natural rainfall. METAR/TAF records were not used because they do not provide a consistent local daily accumulation series for this purpose.

To avoid leakage from unknown measurement time, rainfall on the field-measurement date was excluded. Candidate meteorological features were 1-, 3-, 7-, and 14-day antecedent totals; the professional-station analysis additionally included a 30-day antecedent precipitation index. These features were added to the ridge model because its behaviour is transparent and it was comparable to ANN in the forward test. Date-grouped and forward validation were repeated. Site 2 rainfall was evaluated only for the complete 2023 subset (92 pairs, 47 dates).

For the 2023 cross-site subset, a negative-control calendar model added only $\sin(2\pi\,\mathrm{DOY}/365.25)$ and $\cos(2\pi\,\mathrm{DOY}/365.25)$ to water content and depth. These terms contain no rainfall or soil-state measurement. Of 311 Site 1 records, 199 (64%) fell on the same 47 calendar dates as Site 2 tests. A stricter sensitivity analysis therefore removed every Site 1 record within ± 7 days of a Site 2 test date, leaving 95 training records after 216 exclusions. A weather improvement comparable to the calendar placebo, or one that disappeared under this temporal exclusion, was interpreted as time–site confounding rather than evidence of a transportable physical weather effect.

3 Results

3.1 Data distribution and DCP target audit

Table 2 shows substantial differences between sites and depths. Site 2 included very high water-content readings that remained within the instrument conversion limit but were far outside the Site 1 training range. Without measured porosity, values below the 100% instrument threshold could not be rejected solely on physical grounds. They were retained in the primary external analysis and flagged as outside model support.

Exported CBR ranged from 1.0 to 28.4% at Site 1 and from 0.5 to 28.4% at Site 2. Repeated horizontal bands were consistent with discrete penetration increments and

Table 2 Descriptive statistics of paired water content and DCP-derived CBR.

Site	Depth (cm)	n	Water content, mean ± SD (%)	CBR, median [IQR] (%)
1	10	71	28.1 ± 5.6	5.30 [4.20–8.30]
1	20	60	24.1 ± 5.8	5.00 [4.00–6.78]
1	30	60	19.7 ± 5.0	3.15 [2.48–4.78]
1	40	60	14.9 ± 3.6	2.85 [2.38–4.20]
1	60	60	22.8 ± 4.8	4.85 [3.78–7.78]
2	10	60	37.1 ± 19.3	6.50 [4.38–10.53]
2	20	25	32.3 ± 9.3	9.20 [6.20–13.10]
2	30	27	23.2 ± 9.9	6.00 [3.70–9.20]
2	40	28	16.8 ± 17.0	4.70 [3.13–6.96]
2	60	27	27.7 ± 17.3	7.00 [5.00–9.20]

IQR = interquartile range. Site 2 rows are medians of repeat soundings within a date and depth.

the conversion in Eq. 3. The raw-file reconstruction removed the earlier black-box ambiguity for the matched subset: the median discrepancy of 0.15 CBR points was small relative to model MAE. The audit does not remove target quantization or establish equivalence to laboratory CBR.

3.2 PR2 stability and replicate-test reliability

The short DL6 audit showed a stable 10-cm channel but marked instability at several deeper levels (Table 3). The 10-cm median within-day swing was 20 mV and no valid reading exceeded 1.088 V. At 20–100 cm, median within-day swings ranged from 137 to 380 mV; 74% of valid readings exceeded 1.088 V at both 20 and 100 cm. Changes of several hundred millivolts within a day are unlikely to represent real profile-water-content dynamics and are compatible with intermittent contact, air gaps, or a connection problem. The audit does not identify a unique failure mechanism, but it weakens confidence in deep-channel transfer.

Table 3 Archived PR2/DL6 channel-stability audit at Site 2, 8–29 May 2023.

Depth (cm)	Valid readings	> 1.088 V (%)	Median (mV)	Median daily range (mV)	Maximum daily range (mV)
10	745	0	828	20	194
20	173	74	1127	177	238
30	332	30	1005	244	636
40	609	16	953	380	686
60	277	34	1040	206	259
100	174	74	1126	137	242

Each channel contained 2,025 timestamps before connection flags were applied. Daily range is maximum minus minimum among valid connected readings on a calendar day.

For the point definition, replicate-test s_{within} on the log-CBR scale was 0.73, equivalent to a multiplicative factor $\exp(0.73) = 2.08$. The alternative blow-window definition gave 0.68, and a robust estimate was 0.55. ICC(1) was 0.19 for point values and 0.25 for the window sensitivity. Repeatability varied by depth (within-cell SD 0.57 at 10 cm and 1.13 at 60 cm) and by period (0.41 in 2021 and 0.74 in 2023), with the latter influenced by ten soundings on 12 November 2023. These results show substantial single-sounding noise but should not be generalized to Site 1 because its replicate reliability was not measured.

3.3 Water-content–CBR relationships

At Site 1, Pearson correlations between water content and CBR were negative at all depths: −0.596, −0.246, −0.235, −0.637, and −0.565 at 10, 20, 30, 40, and 60 cm, respectively. Within common support at Site 2, the corresponding correlations were −0.530, −0.288, −0.676, −0.593, and −0.437. Despite consistent direction, the scatter was wide and the distributions differed between sites (Fig. 2).

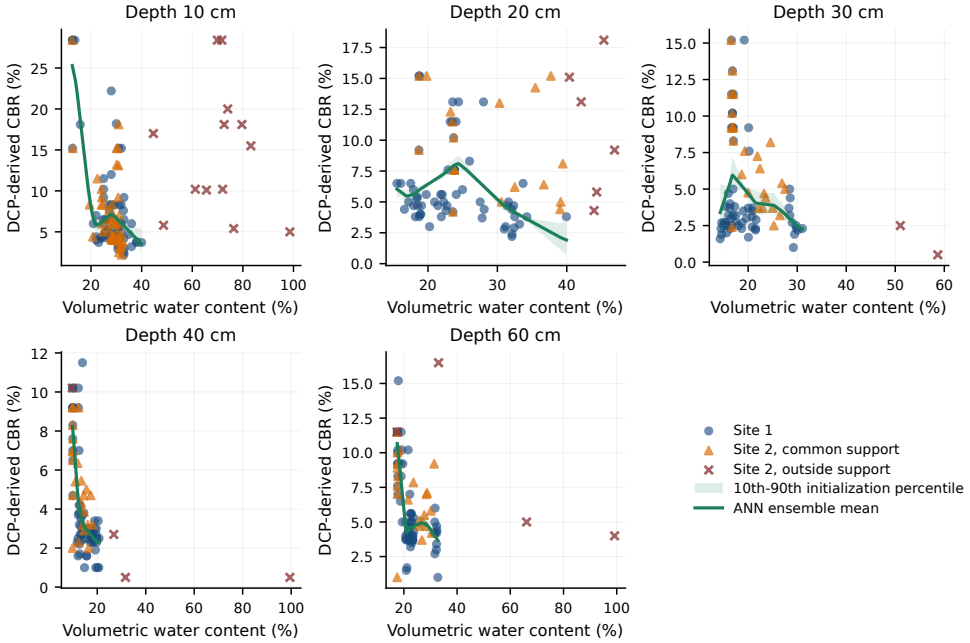

Fig. 2 Volumetric water content and DCP-derived CBR by depth. The green curve is the mean of seven ANN initializations fitted only at Site 1; shading is the 10th–90th percentile across initializations, not a predictive interval. Crosses identify Site 2 observations outside the Site 1 water-content range. Local angular changes at 20 and 30 cm arise from the piecewise-linear ReLU network, sparse support, and the absence of a monotonicity constraint; they should not be interpreted as physical thresholds.

3.4 Grouped, forward, and cross-site performance

Table 4 presents the benchmark hierarchy. In grouped validation, ANN achieved the lowest MAE (2.100) and RMSE (3.042), with $R^2 = 0.451$. Its MAE difference from ridge was −0.151 CBR points (95% cluster-bootstrap interval −0.268 to −0.038). The monotonic model was competitive in MAE (2.225), indicating that much of the local signal was consistent with a simple decreasing relationship.

In the forward holdout, ANN and ridge were indistinguishable in MAE: the direct ANN-minus-ridge difference was −0.003 (−0.305 to 0.321). ANN R^2 fell to −0.175, and the previous-CBR benchmark had better point performance (MAE 2.256, $R^2 = 0.205$, $\rho = 0.647$). The ANN-minus-previous-CBR MAE difference was 0.383 (−0.469 to 1.308), so the point advantage of persistence was not conclusive but contradicted any claim that ANN was clearly the strongest operational benchmark.

Table 4 Performance under the three validation protocols.

Validation	Model	n	MAE	RMSE	R^2	Spearman ρ
Grouped CV, Site 1	Depth median	311	2.451	4.204	−0.048	0.360
	Monotonic log-linear	311	2.225	3.414	0.309	0.580
	Ridge	311	2.251	3.161	0.408	0.566
	ANN	311	**2.100**	**3.042**	**0.451**	**0.642**
Forward holdout, Site 1	Depth median	71	3.535	4.713	−0.893	−0.155
	Monotonic log-linear	71	3.035	4.152	−0.470	0.191
	Ridge	71	2.642	3.533	−0.064	0.284
	ANN	71	2.639	3.712	−0.175	0.416
	Previous CBR	71	**2.256**	**3.054**	**0.205**	**0.647**
Cross-site, all Site 2	Site 1 depth median	167	3.941	5.818	−0.379	0.231
	Monotonic log-linear	167	3.923	6.155	−0.543	0.166
	Ridge	167	14.121	55.769	−125.688	0.293
	ANN	167	4.355	7.081	−1.043	0.123
	Site 2 local median CV	167	**3.548**	**5.091**	**−0.056**	0.220

The cross-site block is the primary spatial test and includes every quality-controlled Site 2 observation. Site 2 local median CV is a grouped local-calibration benchmark, not a transferable Site 1 model.

The primary cross-site test failed. ANN MAE (4.355) exceeded the Site 1 median baseline (3.941), and $R^2 = -1.043$. The locally calibrated Site 2 median was also better than ANN; the ANN-minus-local-median MAE difference was 0.807 (0.186 to 1.440). Quadratic ridge extrapolated catastrophically at high Site 2 water contents, confirming that unconstrained polynomial transfer outside support was unsafe.

Restricting Site 2 post hoc to the 138 observations within depth-specific Site 1 ranges produced lower errors: baseline MAE 3.351, monotonic model 2.839, ridge 2.788, and ANN 2.696. ANN R^2 was nevertheless only 0.071. ANN did not directly outperform ridge (difference −0.091, −0.300 to 0.115) or the local Site 2 median (difference −0.208, −0.611 to 0.175). Among the 29 excluded observations, ANN MAE rose to 12.246 and the monotonic model to 9.084. Common-support results therefore describe conditional behaviour, not general cross-site success.

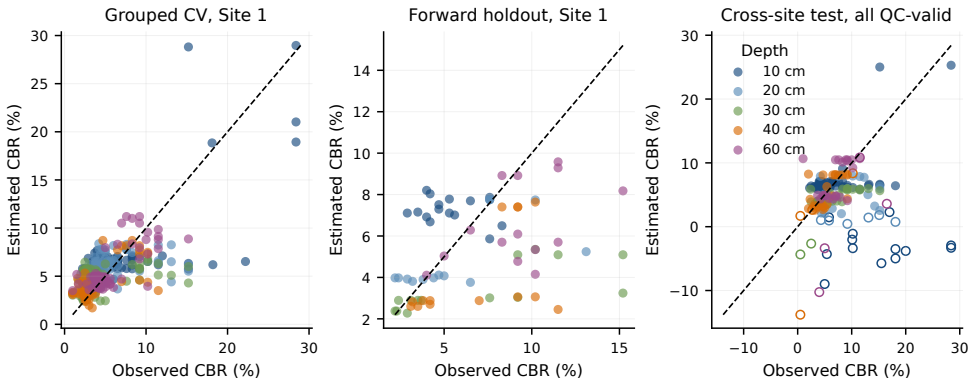

Fig. 3 Observed versus ANN-estimated DCP-derived CBR. The cross-site panel includes all quality-controlled Site 2 observations; open symbols indicate values outside the depth-specific Site 1 water-content range. The dashed line is identity.

3.5 Calibration and residual structure

Internal ANN bias was 0.050 CBR points and its calibration slope was 0.951. In the forward holdout, bias was −1.486 and slope fell to 0.517; in the common-support cross-site subset, bias was −0.903 and slope was 0.603. Because Eq. 6 regresses observations on estimates, the sub-unity slopes should not be described as simple prediction compression. The separate standard-deviation ratios were 0.708 internally, 0.623 forward, and 0.778 cross-site, while correlations weakened. Together with Fig. 3, these diagnostics indicate systematic underestimation and failure to recover the upper tail, rather than a single calibration defect.

Residuals at Site 2 increased sharply beyond training support (Fig. 4). The forward block included three dates with recorded air temperature at or below 0°C. Soil temperature was unavailable, so a freezing effect on dielectric response or penetration resistance could not be tested and remains a possible contributor to temporal shift.

3.6 Meteorological sensitivity

Antecedent professional-station rainfall modestly improved the ridge model in grouped validation, but the 95% interval for the MAE change included zero (Table 6). The largest prespecified point improvement occurred in the forward test, where MAE decreased from 2.634 to 2.340 and R^2 changed from −0.062 to 0.261; the MAE-change interval still included zero (−0.608 to 0.031).

The post-hoc one-day alignment produced MAE 2.294 and an interval below zero, but this result was selected after observing the date mismatch and cannot be treated as confirmatory. Site 2 rainfall produced little change. Adding local air temperature with local rainfall did not improve the forward test (MAE 2.957).

The calendar negative control changed the interpretation of the apparent 2023 cross-site weather gain (Table 5). With all Site 1 records available, adding rainfall and temperature reduced ANN MAE by 1.32 points, but the calendar placebo alone reduced it by 1.02 points. After removal of Site 1 observations within ±7 days of Site 2

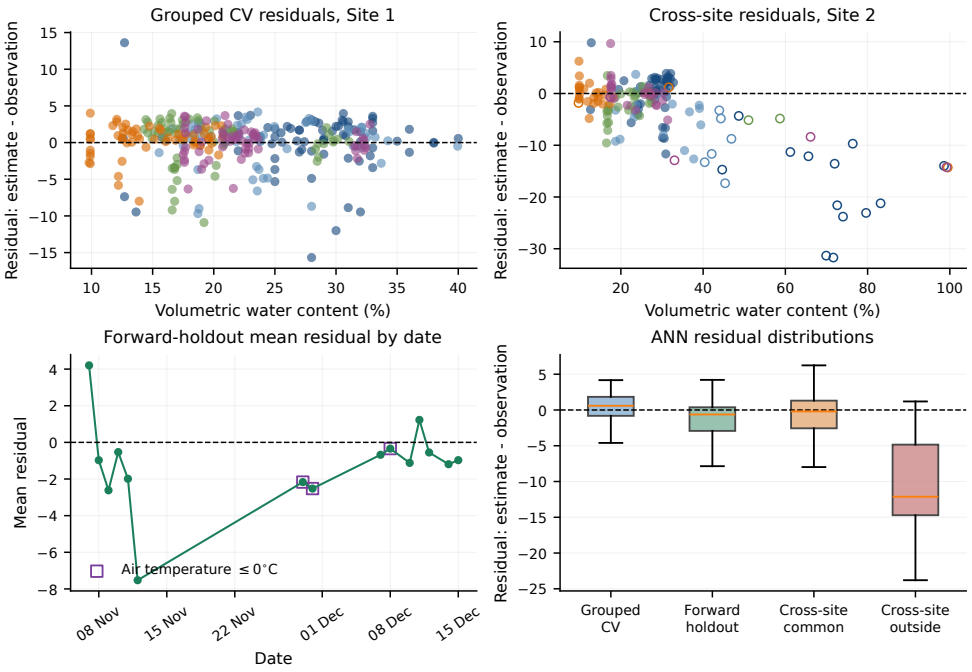

Fig. 4 ANN residual diagnostics. Open cross-site symbols denote observations outside Site 1 water-content support. Squares in the temporal panel mark dates with air temperature at or below 0°C; soil temperature was not measured.

tests, the calendar ANN became numerically unstable and the remaining weather results were inconsistent across ridge and ANN. The external weather improvement therefore cannot be attributed specifically to physical meteorological information. Rainfall remains a plausible state-history covariate for prospective testing, not an established transferable gain.

Table 5 Calendar-placebo diagnostic for the 2023 cross-site subset (92 observations, 47 dates).

Training design	Added information	Model	MAE	ΔMAE (95% interval)
All Site 1 ($n = 311$)	None	Ridge	3.02	Reference
		ANN	3.01	Reference
	Calendar sine/cosine	Ridge	2.82	−0.20 (−0.53; 0.13)
		ANN	1.99	−1.02 (−1.51; −0.50)
	Rainfall + temperature	Ridge	2.78	−0.24 (−0.51; 0.04)
		ANN	1.69	−1.32 (−1.75; −0.87)
Exclude ±7 days ($n = 95$)	None	Ridge	4.07	Reference
		ANN	5.01	Reference
	Calendar sine/cosine	Ridge	4.99	0.92 (−0.23; 2.09)
		ANN	> 1000	Unstable
	Rainfall + temperature	Ridge	4.55	0.48 (−0.20; 1.27)
		ANN	3.55	−1.46 (−2.66; −0.30)

Negative ΔMAE favours the augmented model. This retrospective negative-control analysis is diagnostic; the unstable network and inconsistent strict-design results preclude a causal interpretation.

Table 6 Exploratory effect of antecedent rainfall on ridge estimates.

Validation	Added information	MAE	RMSE	R^2	ΔMAE (95% interval)
Site 1 grouped CV	None	2.250	3.161	0.408	Reference
	Professional-station rainfall, official date	2.156	2.992	0.469	−0.093 (−0.232; 0.051)
Site 1 forward	None	2.634	3.529	−0.062	Reference
	Professional-station rainfall, official date	2.340	2.944	0.261	−0.294 (−0.608; 0.031)
	Professional-station rainfall, +1-day shift	2.294	3.002	0.232	−0.340 (−0.591; −0.063)
Site 2, 2023	None	2.914	4.026	0.092	Reference
	Locally recorded rainfall	2.868	4.012	0.098	−0.046 (−0.203; 0.119)

Negative ΔMAE favours the rainfall-augmented model. The +1-day result is explicitly post hoc.

4 Discussion

4.1 Local information did not imply transferable estimation

The main result is not that ANN was the best model; it is that modest local skill did not survive realistic transfer. Grouped validation at Site 1 supported an association between profile water content and DCP–CBR, and ANN improved MAE by 0.351 CBR points relative to the depth median. Yet the forward R^2 was negative, the previous-CBR benchmark had better point performance, and the primary cross-site ANN result was worse than a constant Site 1 baseline. A minimal Site 2 calibration also outperformed ANN. These comparisons place the apparent internal gain in its proper operational context.

The post-hoc common-support result is informative but cannot replace the primary analysis. Removing 29 high-water-content observations excluded the cases in which extrapolation failed most severely. Even after restriction, ANN explained only 7.1% of the variance and did not outperform ridge or a local median directly. The result supports a narrow statement: within overlapping water-content ranges, Site 1 water-content relationships contained some transferable average information. It does not support deployment without local calibration.

The benchmark hierarchy also discourages algorithm-centric interpretation. ANN beat ridge directly only in grouped internal validation. It did not do so in the forward or common-support spatial comparisons. The monotonic model had slightly higher internal MAE but avoided the extreme polynomial extrapolation of ridge and offered a physically interpretable direction. Future work should prioritize constrained, calibrated models and uncertainty estimates over additional model complexity.

4.2 Physical interpretation and missing geotechnical variables

Negative water-content–CBR associations are consistent with reduced suction-related strength as wetting progresses, but water content alone does not define suction or effective stress. The TIG studies by Garcia and Tabaroei [4] and Aneke et al. [5] explicitly included matric suction or controlled compaction states and found stronger mechanistic relationships. The present field records lack soil-water retention curves, dry density, grading, and Atterberg limits. Differences in these properties, together with surface management and stress history, are plausible causes of cross-site shift.

BPEJ information improves pedological description but cannot replace local engineering testing. Retrospective USCS or AASHTO labels based only on a mapped soil unit would obscure rather than resolve uncertainty. A future campaign should obtain disturbed and undisturbed samples at each depth, determine particle-size distribution, liquid and plastic limits, specific gravity, dry density, porosity, and compaction characteristics, and calibrate PR2 readings gravimetrically. These measurements would distinguish true material differences from sensor-calibration differences.

4.3 DCP-derived CBR is a quantized surrogate

The raw-file audit materially strengthens reproducibility. It establishes the observed hammer configuration, defines the nominal-depth assignment, and shows that the generic DCP conversion reproduces matched exported profiles closely. It also clarifies why the target contains horizontal bands and a repeated upper value of 28.4%. Standard regression metrics treat the target as continuous, although it is a transformed, discretized penetration response.

The audit does not make DCP–CBR identical to laboratory CBR. Device metadata were not supported by measured plasticity, and specific CL or CH conversion branches could not be independently justified. Adegoke and Okonta [8] show why geometry and hammer configuration must remain explicit. Future modelling should retain raw blow counts and depth increments, analyse DCPI directly or model log(CBR), and validate against an independent bearing test. Censored or interval-aware methods may be preferable if an instrument ceiling is confirmed.

4.4 Measurement noise and PR2 stability

The replicate audit provides a physical explanation for part of the low external R^2. A within-cell SD of 0.73 in log CBR means that two soundings made at the same nominal depth and date can readily differ by approximately a factor of two. ICC values of 0.19–0.25 imply low reliability for a single sounding under the stated variance model. This attenuates relationships with cell-level variables such as water content and weather; it does not prove that every attainable R^2 is bounded by exactly the same number. For field decisions, several soundings and a robust aggregate are preferable to a single penetration, and uncertainty bands should reflect repeatability.

The PR2 audit identifies a second, independent source of noise. The deeper channels combined large daily voltage ranges with frequent excursions beyond the manufacturer's conversion range during the only available logger interval. Because the audit covered one short period, it cannot establish that every deep reading was invalid. It does show that generic range filtering alone is insufficient to guarantee stable tube contact. Future campaigns should document installation quality, perform gravimetric calibration and periodic air/water checks, and retain time-resolved diagnostic voltage alongside converted water content.

4.5 Rainfall and temporal–spatial confounding

Measured water content is a current state indicator, whereas antecedent rainfall provides information about wetting history and redistribution. The forward improvement from professional-station rainfall is therefore plausible. Its interval crossed zero under official dating, however, and the apparently stronger one-day result was post hoc. The high correlation after shifting local rainfall by one day indicates a record-convention issue that should be resolved prospectively with time-stamped gauges.

The calendar placebo exposes an additional design limitation. Nearly two thirds of Site 1 records occurred on the same calendar dates as the 2023 Site 2 tests. Once weather or day-of-year terms enter the model, they can identify shared dates and season rather than a transportable moisture–strength mechanism. The comparable improvements from real weather and calendar sine/cosine terms, followed by unstable or inconsistent results after $\pm$7-day exclusion, are the signature of this confounding. The placebo does not show that weather is irrelevant; it shows that the present cross-site design cannot isolate its physical contribution.

The professional station was preferable to METAR/TAF for a quantitative daily accumulation series, but a station 4.69 km away cannot resolve convective or local surface inputs. Site 2 rainfall added almost no skill, and two experimental watering events had to be separated from natural precipitation. Future work should deploy time-synchronized tipping-bucket gauges, record irrigation separately, include soil temperature and evapotranspiration proxies, and reserve temporally non-overlapping site blocks for evaluation. Meteorological variables and exclusion windows should be specified before evaluating the test period.

4.6 Relation to previous studies

The lower transfer performance relative to recent CBR studies is expected from the restricted predictor set and more demanding validation. Tilahun et al. [6] incorporated dry density and elevation in addition to moisture. Adjei-Yeboah et al. [7] used ten geotechnical variables and quantified predictive uncertainty. Controlled suction studies achieved strong relationships under known material and compaction conditions [4, 5]. The earlier airport model benefited from antecedent bearing capacity, temperature, and rainfall [9]; in the present forward test, previous CBR was likewise a strong benchmark.

The contribution of the present work is therefore primarily diagnostic. By retaining a difficult external set, comparing against local and temporal references, and reporting failed extrapolation, the study identifies what a moisture-only estimator cannot support. Negative transfer results are useful for preventing an internally accurate model from being mistaken for a validated operational tool.

4.7 Practical implications and limitations

No operational CBR threshold was locally validated, so sensitivity, specificity, and decision-curve performance could not be calculated. The modest internal ranking ($\rho = 0.642$) weakened in the forward (0.416) and primary cross-site (0.123) tests. The dataset therefore does not validate moisture-based screening, even though it motivates a future screening study. Current decisions should continue to rely on direct DCP or another accepted bearing-capacity assessment.

The study is limited by one primary measurement area, one external site, approximately one year of Site 1 observations, and irregular bearing tests. Site 2 data span earlier years, so spatial and temporal shifts cannot be fully separated. Date-grouped cross-validation controls leakage among depths on the same date but does not remove all temporal autocorrelation; rolling-origin validation should be used in a longer prospective series. The forward test covered only 15 late-autumn dates, three with air temperature at or below freezing, and soil temperature was not measured.

Additional uncertainty arises from spatial separation between the PR2 tube and DCP soundings, repeated penetration disturbance, turf and root effects in the 10-cm layer, generic PR2 calibration, documented short-period channel instability, possible air gaps, and the quantized DCP target. The repeatability audit was available only at Site 2, and the DL6 stability export covered only 22 days. The study was retrospective and comparisons were not preregistered or multiplicity-adjusted. Results should thus be interpreted as validation evidence for this dataset, not as universal material laws.

A stronger next study would use multiple spatial points at each site and multiple sites, preregister the validation protocol and minimum operational improvement, collect raw DCPI and laboratory CBR, measure the missing geotechnical parameters and soil temperature, and reserve entire sites and seasons for external testing. A small local recalibration set should be compared explicitly with no-recalibration transfer. Only then should threshold classification or uncertainty-aware screening be evaluated.

5 Conclusions

Profile volumetric water content contained useful local information about DCP-derived CBR, but that information was not sufficiently stable for uncalibrated transfer. ANN improved grouped Site 1 MAE relative to a depth median, yet it offered no direct advantage over ridge in the forward or common-support spatial tests. Previous CBR was the strongest forward point benchmark, and a locally estimated Site 2 median outperformed the transferred ANN.

The primary test on all 167 Site 2 observations was negative: ANN MAE was 4.355 CBR points and $R^2 = -1.043$, compared with baseline MAE 3.941. Restriction to common support improved MAE but explained little spatial variance and was post hoc. Replicate ICC of 0.19–0.25 and a log-scale within-cell SD of 0.73 show that single-sounding target noise contributed materially to the limited explainable variation. The short PR2 audit also identified instability in deeper channels.

Antecedent professional-station rainfall produced a promising forward point improvement, but the prespecified interval included no effect; an apparent cross-site weather benefit was reproduced by a calendar placebo and was not stable under temporal exclusion. The raw DCP audit improved traceability while confirming quantization and the need to distinguish DCP-derived from laboratory CBR. Until site-specific soil characterization, sensor calibration, replicated penetration, temporally separated multi-site validation, and operational thresholds are available, profile water content should not replace direct bearing-capacity testing.

References

[1] ASTM International: ASTM D6951/D6951M-18(2023): Standard Test Method for Use of the Dynamic Cone Penetrometer in Shallow Pavement Applications. ASTM International, West Conshohocken, PA (2023). https://doi.org/10.1520/D6951_D6951M-18R23 . ASTM International

[2] Ranieri, V., Berloco, N., D'Auria, D., Disalvo, V., Fedele, V., Intini, P., Colonna, P.: Determination of bearing capacity of cleared and graded areas at airports. Journal of Transportation Engineering, Part B: Pavements **147**(1), 04020086 (2021) https://doi.org/10.1061/JPEODX.0000247

[3] Vaz, C.M.P., Hopmans, J.W.: Simultaneous measurement of soil penetration resistance and water content with a combined penetrometer-tdr moisture probe. Soil Science Society of America Journal **65**(1), 4–12 (2001) https://doi.org/10.2136/sssaj2001.6514

[4] Garcia, K.E., Tabaroei, A.: Matric suction-based mathematical model in predicting california bearing ratio using dynamic cone penetrometer test. Transportation Infrastructure Geotechnology **12**, 49 (2025) https://doi.org/10.1007/s40515-024-00518-z

[5] Aneke, F., Kalumba, D., Hanandeh, S., Shiferaw, H.M.: Quantifying the influence

of matric suction on california bearing ratio of compacted unsaturated subgrades for flexible pavement design. Transportation Infrastructure Geotechnology **13**, 135 (2026) https://doi.org/10.1007/s40515-026-00883-x

[6] Tilahun, Y., Xiao, Q., Ashango, A.A., Han, X., Negewo, M.: Prediction of spatial soil-California Bearing Ratio of subgrade soil using particle swarm optimization-artificial intelligence method. Transportation Infrastructure Geotechnology **12**, 80 (2025) https://doi.org/10.1007/s40515-025-00535-6

[7] Adjei-Yeboah, S., Ankah, M.L.Y., Asare, E.N., Billy, J.E.: Uncertainty-aware estimation of california bearing ratio for tropical subgrade soils using conformal prediction. Transportation Infrastructure Geotechnology **13**, 240 (2026) https://doi.org/10.1007/s40515-026-01008-0

[8] Adegoke, A.H., Okonta, F.: Effect of cone geometry and hammer weight on the penetration resistance of fibre-reinforced unpaved road layer. Transportation Infrastructure Geotechnology **13**, 308 (2026) https://doi.org/10.1007/s40515-026-01070-8

[9] Anonymous: Reference details withheld for double-anonymous review. Published article (2021)

[10] Výzkumný ústav meliorací a ochrany půdy: eKatalog BPEJ: 4.11.10 (2026). https://bpej.vumop.cz/41110 Accessed 2026-08-26

[11] Výzkumný ústav meliorací a ochrany půdy: eKatalog BPEJ: 5.12.00 (2026). https://bpej.vumop.cz/51200 Accessed 2026-08-26

[12] Delta-T Devices Ltd: User Manual for the Profile Probe Type PR2, Version 5.0 edn. Burwell, Cambridge, UK (2016). https://delta-t.co.uk/wp-content/uploads/2017/02/PR2_user_manual_version_5.0.pdf Accessed 2026-08-26

[13] Topp, G.C., Davis, J.L., Annan, A.P.: Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. Water Resources Research **16**(3), 574–582 (1980) https://doi.org/10.1029/WR016i003p00574

[14] Kessler Soils Engineering Products, Inc.: K-100 Dynamic Cone Penetrometer User's Manual. Kessler Soils Engineering Products, Inc., Springfield, VA (2016). Kessler Soils Engineering Products, Inc. https://www.kesslerdcp.com/wp-content/uploads/2018/05/DCP_K100_Manual_SHUFFLED_1JULY2016.pdf Accessed 2026-10-29

[15] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., VanderPlas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É.: Scikit-learn: Machine learning in python. Journal of Machine Learning Research **12**, 2825–2830 (2011)

[16] Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F., Dormann, C.F.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography **40**(8), 913–929 (2017) https://doi.org/10.1111/ecog.02881

[17] Czech Hydrometeorological Institute: Open Data: Historical Daily Precipitation Records (2026). https://opendata.chmi.cz/meteorology/climate/historical_csv/data/daily/precipitation/ Accessed 2026-11-05

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SampleResource1AnalysisData.xlsx