Journal of Natural Resources and Farming
https://doi.org/10.7770/jonraf-v2-art2907

SPECIAL ISSUE

Multiyear Validation and Mass-Balance Auditing of Richards Simulations on Uniform and Graded Finite-Volume Meshes Using USDA SCAN Observations

José D. Zúniga-Varela[a*], Marlon M. López-Flores[b], William Campillay-Llanos[c]

Escuela de Matemática y Ciencias de la Computación, Universidad Nacional Autónoma de Honduras, Tegucigalpa[a], Honduras
Laboratório de Inteligência Artificial, Robótica e Cibernética, Instituto Militar de Engenharia, Rio de Janeiro[b], Brazil.
Departamento de Ciencias Matemáticas y Físicas; Núcleo de Investigación en Producción Alimentaria (NIPA), Universidad Católica de Temuco, Temuco[c], Chile.

Article history
Received: 10/02/2026
Accepted: 08/06/2026
Published: 22/09/2026

Corresponding author
*José D. Zúniga-Varela
jzuniga@unah.edu.hn

ABSTRACT
Purpose: We tested whether surface-focused mesh grading reduced spatial discretization error in a finite-volume Richards model, whether grid rankings persisted across seasons, and whether hourly water-balance residuals decreased under temporal refinement. **Methods:** Hourly observations from the U.S. Department of Agriculture (USDA) Soil Climate Analysis Network (SCAN) at Mahantango Creek, Pennsylvania, were quality controlled and aggregated daily for June–September 2023–2025. Effective parameters were selected from 2023; 2024–2025 were held out. Uniform, logarithmic, and algebraic grids were compared with a 40–640-cell reference sequence. An hourly audit compared 32- and 64-substep predictor–corrector configurations. **Results:** The combined held-out root mean square error (RMSE) was 0.0292 $m^3\ m^{-3}$, and depth-specific Nash–Sutcliffe efficiency (NSE) values were 0.603, 0.777, and 0.875 at 5.1, 10.2, and 20.3 cm. At 40 cells, logarithmic grading reduced mean profile error by 47.9% in calibration and 62.1–71.8% in validation; algebraic grading had the lowest error in all three events. Grid errors were below 0.19% of held-out model–observation RMSE. Doubling boundary substeps reduced the maximum substep residual from 1.41×10^{-9} to 8.48×10^{-10} m but increased the maximum hourly residual from 1.26×10^{-8} to 2.13×10^{-8} m and the cumulative residual from 2.46×10^{-7} to 3.74×10^{-7} m.7. **Conclusions:** Surface-focused grading proved transferable across seasons, but its field-scale effect was secondary to forcing, structural, and observational uncertainty. The accounting was internally consistent, but strict temporal convergence was not demonstrated; the benchmark is mass-balance audited, not asymptotically mass-conservative.

Keywords: Richards equation; temporal validation; soil moisture; finite-volume method; mesh grading; mass-balance audit

1. Introduction

Soil-water dynamics govern infiltration, root-zone storage, drainage, and solute transport. Numerical models for agricultural and environmental applications should therefore preserve water-content bounds, represent hydraulic fluxes, and close the water balance while resolving sharp near-surface gradients. The Richards equation with van Genuchten–Mualem hydraulic functions remains a standard continuum description of variably saturated flow (Richards, 1931; Mualem, 1976; van Genuchten, 1980). Its nonlinear and sometimes degenerate behavior makes the numerical solution sensitive to time integration, constitutive regularization, and spatial resolution (Farthing and Ogden, 2017; Ireson et al., 2023).

Mass-conservative formulations are essential because a numerically stable pressure-head trajectory can still produce unacceptable water-balance errors if storage is updated inconsistently. Classical mixed-form and conservative time-stepping studies established this principle (Milly, 1985; Celia et al., 1990; Kavetski et al., 2001). Finite-volume methods are attractive because shared face fluxes cancel exactly in the global spatial balance and can be used on nonuniform meshes (Eymard et al., 1999; Caviedes-Voullième et al., 2013; Svyatsky and Lipnikov, 2017). A graded mesh can concentrate cells near the soil surface, but a favorable result for one storm does not demonstrate that a specific mapping is robust or practically significant.

A defensible grid study must distinguish among numerical discretization error, process-model discrepancy, and observational uncertainty. In situ networks such as the U.S. Department of Agriculture (USDA) Soil Climate Analysis Network (SCAN) provide soil-moisture and meteorological observations at multiple depths, but sensor calibration, support volume, site heterogeneity, and quality-control status affect interpretation (Bogena et al., 2010; Mittelbach et al., 2012; Dorigo et al., 2011, 2013). Hydraulic inversion is also nonunique; calibration against short records may compensate for forcing or structural errors through correlated effective parameters (Scharnagl et al., 2011; Gao et al., 2019). Temporal holdout data are therefore required before a mesh result is described as transferable.

This study develops a reproducible, field-informed benchmark using three warm-season records from one SCAN station. The objectives were to (i) calibrate effective upper-profile parameters using 2023 data only and evaluate model performance in 2024 and 2025 without refitting; (ii) demonstrate spatial convergence before designating a numerical reference; (iii) compare uniform, logarithmic, and algebraic mappings over multiple cell counts and grading parameters; (iv) compare grid error with the held-out model–observation discrepancy; and (v) audit whether hourly water-balance residuals decrease when the predictor–corrector boundary substeps are doubled. We hypothesized that surface-focused grading would reduce discretization error across independent rainfall windows but that no single mapping would be universally optimal. We do not claim a new finite-volume stencil; rather, the contribution is a transparent benchmark for validation, mesh selection, and water-balance auditing.

2. Materials and Methods

2.1 Site, sensors, and archived data

The case study was conducted at Mahantango Creek SCAN station 2028 in Northumberland County, Pennsylvania (40.67309 N, 76.66702 W; elevation 720 ft). The station has reported since 27 April 1999 and lies within the Mahantango Creek agricultural watershed (Buda et al., 2011; USDA NRCS, n.d.-b).. The sensor-description export identified Stevens HydraProbe sensors for volumetric soil moisture at nominal depths of 2, 4, 8, 20, and 40 in (5.1, 10.2, 20.3, 50.8, and 101.6 cm).

Two hourly USDA Natural Resources Conservation Service (NRCS) Report Generator exports covered 1 January 2023 through 31 December 2025: meteorological forcing and instantaneous soil-moisture observations. The expected complete sequence contained 26,304 hours. The forcing file contained 26,294 timestamps and the soil file 26,291, with no duplicate timestamps. Data were exported under NRCS provisional-data conditions (USDA NRCS, 2024, n.d.-a).

Hourly values were reindexed to the complete sequence of timestamps. Broad physical range checks excluded soil-moisture values outside 0–60%, air temperatures outside −130F, hourly precipitation outside 0–5 in, relative humidity outside 0–100%, and negative solar radiation. Daily precipitation was summed; daily mean temperature and soil moisture were calculated, and daily maximum and minimum temperatures were retained. A daily row was considered complete when each essential variable had at least 20 hourly values. Only the data from 19 June 2024 failed this condition. Its 17 available hourly precipitation values were all zero, so precipitation was set to zero and temperature was linearly interpolated; soil-moisture observations from that day were excluded from the evaluation. No soil-moisture values were interpolated for scoring.

The analysis used June–September in 2023, 2024, and 2025 to avoid frozen-soil processes omitted by the model. Each season contained 122 days; the first seven days were used to initialize and spin up the seasonal column and were excluded from performance metrics. This yielded 115 evaluated days in 2023, 114 in 2024, and 115 in 2025. The 2023 season was the calibration set. The complete 2024 and 2025 seasons were held out from parameter and mesh selection.

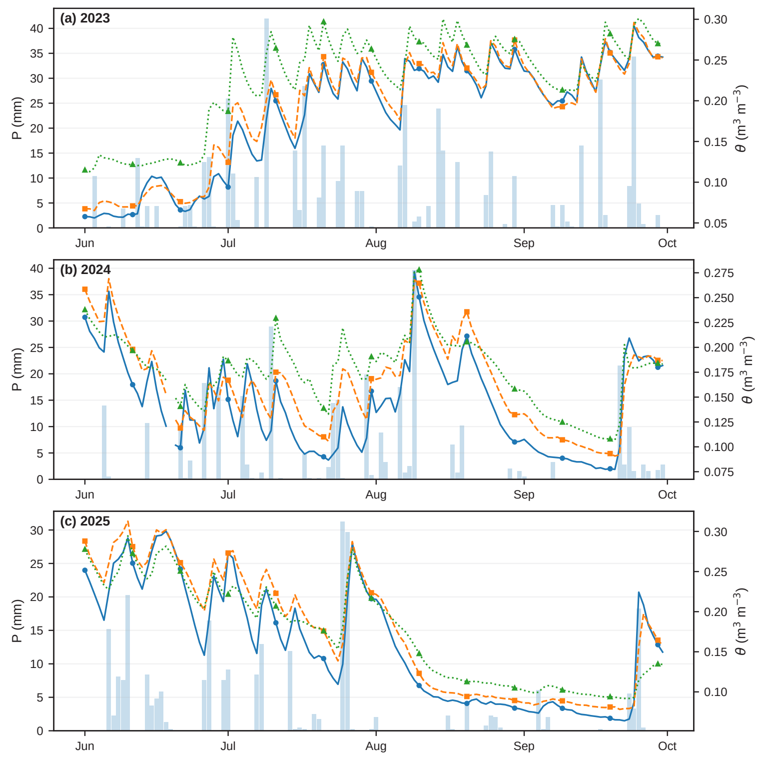
Table I

Data coverage and quality-control summary

Item	Value
Expected hourly timestamps, 2023–2025	26,304
Meteorological timestamps present	26,294
Soil-moisture timestamps present	26,291
Duplicate timestamps	0
Daily rows, June–September	366
Incomplete daily rows	1 (19 June 2024)
Calibration observations	115 days × 3 depths
Held-out validation observations	229 days × 3 depths

FIgure 1.

Daily precipitation and observed volumetric water content during 2023–2025. Panels (a)–(c) show 2023, 2024, and 2025. Bars denote daily precipitation (P), whereas solid circles, dashed squares, and dotted triangles denote volumetric water content (θ) measured by the 2-, 4-, and 8-in sensors, respectively



2.2 Richards model and hydraulic functions

Depth z is positive downward and the modeled domain is $0 \le z \le L = 0.35$ m. Conservation is

$$\frac{\partial\theta}{\partial t} + \frac{\partial q}{\partial z} = -S(z,t), \qquad q = K(h) \quad 1 - \frac{\partial h}{\partial z} \quad , \tag{1}$$

where θ is volumetric water content, h is matric pressure head, q is downward flux, and S is root uptake. The van Genuchten–Mualem functions are

$$S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r} = \left[1 + (\alpha|h|)^n\right]^{-m}, \quad m = 1 - \frac{1}{n}, \tag{2}$$

$$K(S_e) = K_s S_e^{\ell}\left[1 - \quad 1 - S_e^{1/m} \quad ^m\right]^2. \tag{3}$$

The numerical state is the logit $Y = \log[S_e/(1{-}S_e)]$. For stable evaluation, S_e was restricted to $[\varepsilon, 1-\varepsilon]$, with $\varepsilon = 10^{-7}$, and a chain-rule denominator floor of 10^{-12} was applied.

Reference evapotranspiration (ET_0) was calculated using the temperature-based Hargreaves relation and extraterrestrial radiation computed according to the Food and Agriculture Organization's Irrigation and Drainage Paper 56 (FAO-56) (Hargreaves and Samani, 1985; Allen et al., 1998). With evapotranspiration multiplier c_E and transpiration fraction f_T, potential evaporation and transpiration are $E_p = (1 - f_T)c_E ET_0$ and $T_p = f_T c_E ET_0$.

2.3 Finite-volume discretization and operational boundaries

For cell i with width Δz_i,

$$\frac{d\theta_i}{dt} = -\frac{q_{i+1/2} - q_{i-1/2}}{\Delta z_i} - S_i. \tag{4}$$

Interior face conductivity is the harmonic mean,

$$K_{i+1/2} = \frac{2K_iK_{i+1}}{K_i + K_{i+1} + 10^{-30}}, \quad q_{i+1/2} = K_{i+1/2} \quad 1 - \frac{h_{i+1} - h_i}{z_{i+1} - z_i} \quad . \tag{5}$$

The lower boundary is free drainage, $q_{N+1/2} = K_N$.

The atmospheric boundary condition uses a fixed 1-cm surface control layer, independent of grid geometry. Let w_i^s be the overlap of cell i with $0 \le z \le d_s$, where $d_s = 0.01$ m. Define storage capacity, extractable storage, and root uptake within this layer as

$$C_s = \sum_i [\theta_s(1-\varepsilon) - \theta_i]w_i^s, \quad A_s = \sum_i [\theta_i - \theta_r(1+\varepsilon)]w_i^s, \quad B_s = \sum_i S_i w_i^s. \tag{6}$$

Here, ε is the dimensionless numerical buffer defined above; it keeps the effective saturation and surface-layer storage bounds away from their exact limiting values. The flux interpolated at d_s is q_δ. The admissible daily surface bounds are $q^+ = q_\delta + C_s + B_s$ and $q^- = q_\delta - A_s + B_s$. Surface stress is $\omega_i = \mathrm{clip}[(S_{e,i} - 0.05)/0.35, 0, 1]$, and actual atmospheric demand is $a = P - E_p\omega_1^2$, where P is precipitation. The Darcy capacity diagnostic is

$$I_c = \max \quad K_s, K_1 \quad 1 + \frac{\max(-h_1, 0)}{\max(z_1, 10^{-4})} \quad . \tag{7}$$

For $a \ge 0$, $q_{1/2} = \min[a, \max(0, \min(I_c, q^+))]$ and runoff is $\max(a - q_{1/2}, 0)$. For $a \ge 0$, $q_{1/2} = \max(a, q^-)$ and evaporation is $P - q_{1/2}$. This operational boundary prevents any grid from receiving additional surface storage merely because its first cell is larger.

Root density is exponential and cell-normalized,

$$\beta_i = \frac{\exp(-z_i/0.15)}{\sum_j \exp(-z_j/0.15)\Delta z_j}, \qquad S_i = T_p\beta_i\omega_i. \tag{8}$$

This parsimonious sink does not represent compensated physiological root uptake (Šimůnek and Hopmans, 2009). The ordinary differential equation (ODE) system and cumulative top flux, drainage, uptake, evaporation, and runoff were integrated using SciPy's backward differentiation formula (BDF) solver (Virtanen et al., 2020), using a relative tolerance (rtol) of 10^{-6}, an absolute tolerance (atol) of 10^{-8}, and a maximum step of 0.05 d. A Jacobian sparsity pattern was supplied. The daily residual was

$$R_M = [M(t_{k+1}) - M(t_k)] - [Q_{\mathrm{top}} - Q_{\mathrm{bottom}} - Q_{\mathrm{sink}}], \quad M = \sum_i \theta_i \Delta z_i. \tag{9}$$

Here, Q_{top}, Q_{bottom}, and Q_{sink} are the cumulative top-boundary flux, bottom-boundary flux, and root uptake over the daily interval, respectively.

2.4 Effective parameter selection and validation design

The initial profile in each season was a shape-preserving interpolation through the first 2-, 4-, and 8-in observations, extended with the 2-in value to the surface and the 8-in value to 0.35 m. Seasonal reinitialization was necessary because the omitted October–May processes were not simulated; after 1 June, no observation reset was applied.

Eight effective parameters define the model (Table II). The final vector was the candidate with the smallest standardized sum of squared errors in a documented 14-member local design evaluated exclusively against 2023 data after the seven-day spin-up. The candidate design perturbed θ_s, α, n, K_s, c_E, and f_T around the base vector, while θ_r and ℓ were fixed. The selected vector had an objective value of 94.221, compared with 94.719 for the next-best candidate. This narrow difference underscores that these are effective rather than uniquely identified parameters. No observations from 2024 or 2025 were used to select parameters, the grid mapping, or solver tolerances.

Table II

Effective parameters used in the multiyear simulations

Parameter	Value	Status	Unit
θ_r	0.0500	fixed	$m^3\,m^{-3}$
θ_s	0.3621	selected	$m^3\,m^{-3}$
α	0.1700	selected	m^{-1}
n	1.6024	selected	–
K_s	0.0700	selected	$m\,d^{-1}$
ℓ	0.5000	fixed	–
ET_0 multiplier c_E	0.6800	selected	–
Transpiration fraction f_T	0.1500	selected	–

Performance was summarized by root mean square error (RMSE), mean absolute error (MAE), bias, Pearson's correlation coefficient (r), and Nash–Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970). For the combined 2024–2025 metrics, 95% confidence intervals (CIs) were estimated with 2,000 seven-day moving-block bootstrap replicates (seed 20260711), preserving short-range serial dependence. A secondary event analysis used predeclared rainfall triggers of at least 5 mm d^{-1} and extended each event through three days after the latest trigger; overlapping windows were merged.

2.5 Grid families, reference convergence, and transfer tests

Uniform, logarithmic, and algebraic edge locations were

$$z_j^U = Lj/N, \tag{10}$$

$$z_j^L = z_*[(1 + L/z_*)^{j/N} - 1], \tag{11}$$

$$z_j^A = L(j/N)^p. \tag{12}$$

The primary mappings used $z_\star = 0.04$ m and $p = 1.5$. The largest three-day rainfall totals were 52.1 mm on 7–9 July 2023, 43.4 mm on 7–9 August 2024, and 61.2 mm on 25–27 July 2025. The corresponding 20-day numerical test windows were 1–20 July 2023 (144.0 mm), 1–20 August 2024 (91.2 mm), and 19 July–7 August 2025 (67.8 mm), where the values in parentheses are total precipitation over each 20-day window. All grids used identical forcing and parameters.

Uniform-grid solutions with $N \in \{40, 80, 160, 320, 640\}$ were compared using a relative tolerance of 3×10^{-7}, an absolute tolerance of 3×10^{-9}, and a maximum step of 0.03 d. After convergence was quantified, the $N = 640$ solution was adopted as the reference. Profile error was

$$E_{L^2}(t) = \left[\frac{1}{L}\int_0^L (\theta_N - \theta_{\mathrm{ref}})^2 dz\right]^{1/2}. \tag{13}$$

Primary mappings were evaluated at $N = 40, 80, 160$ in 2023. A predeclared $N = 40$ sweep used $z_\star \in \{0.02, 0.04, 0.08, 0.16\}$ m and $p \in \{1.2, 1.3, 1.5, 1.8\}$. Transferability was then tested at $N = 40$ on the 2024 and 2025 event windows without changing the mappings.

2.6 Hourly mass-balance refinement audit

To test whether water-balance closure was robust to temporal refinement, an additional June–September 2023 audit was performed on a uniform $N = 320$ grid at hourly resolution (2,928 intervals). Both runs used the BDF solver with Radau fallback, rtol = 3×10^{-8}, atol = 3×10^{-11}, a maximum step fraction of 0.03125, a predictor–corrector atmospheric-boundary treatment, saturation-excess correction, a mass-correction tolerance of 10^{-10} m, 48 correction iterations, and a zero-flux-floor acceptance tolerance of 2×10^{-9} m. Hereafter, PC32 and PC64 denote predictor–corrector configurations using 32 and 64 atmospheric-boundary substeps per hourly interval, respectively. Apart from the number of boundary substeps, the two configurations were identical. We denote the hourly residual by r_h and the physical corrector substep residual by r_j. Strict audit criteria were maxh $|r_h| \leq 10^{-8}$ m and $|\Sigma_h r_h| \leq 10^{-8}$ m.

3. Results

3.1 Seasonal calibration and held-out validation

The model reproduced the broad wetting and drying trajectories but retained depth- and year-specific biases (Fig. 2; Table III). Calibration RMSE was 0.0318 m^3 m^{-3} overall. Held-out overall RMSE values were 0.0315 and 0.0268 m^3 m^{-3} in 2024 and 2025, respectively. The 2024 2-in series had a negative NSE (−0.074) and a positive bias of 0.0373 m^3 m^{-3}, whereas NSE was positive at all three depths in 2025. Thus, temporal validation revealed a shallow-layer weakness that was not apparent from 2023 calibration alone.

Table III

Depth-specific calibration and validation metrics. RMSE, root mean square error; MAE, mean absolute error; r, Pearson's correlation coefficient; NSE, Nash–Sutcliffe efficiency. RMSE, MAE, and bias are in m^3 m^{-3}

Period	Depth	RMSE	MAE	Bias	r	NSE
Calibration 2023	2 in	0.0375	0.0301	0.0145	0.884	0.667
	4 in	0.0262	0.0207	0.0106	0.933	0.826
	8 in	0.0307	0.0243	-0.0211	0.941	0.680
Validation 2024	2 in	0.0446	0.0396	0.0373	0.826	-0.074
	4 in	0.0273	0.0246	0.0195	0.885	0.557
	8 in	0.0154	0.0116	0.0041	0.902	0.796
Validation 2025	2 in	0.0298	0.0249	0.0144	0.962	0.830
	4 in	0.0298	0.0270	0.0000	0.956	0.840
	8 in	0.0194	0.0169	0.0032	0.961	0.899

The combined held-out RMSE decreased with depth (Table IV). The seven-day moving-block bootstrap 95% CIs for bias were entirely above zero at the two shallow depths, whereas the 8-in bias CI included zero. The combined NSE values were 0.603, 0.777, and 0.875 at 2, 4, and 8 in. The event analysis identified 11 validation windows in 2024 and eight in 2025. Median event-wide RMSE values were 0.0247 and 0.0294 m^3 m^{-3}, respectively. Event NSE was often unstable because short windows had low observed variance; event RMSE and bias were therefore treated as the more interpretable diagnostics.

3.2 Water balance

Maximum absolute daily mass residuals were 4.35×10^{-8}, 3.95×10^{-9}, and 1.43×10^{-8} m for 2023, 2024, and 2025, respectively (Table V). These values are negligible relative to the seasonal water-balance components and indicate internally consistent budget accounting for the reported configuration. They establish neither strict temporal convergence nor a unique field-scale partition of evaporation, drainage, root uptake, and runoff.

3.3 Hourly temporal-refinement audit

Both audited configurations completed the 2,928-hour season, but neither satisfied the strict 10^{-8} m criteria (Table VI). PC64 reduced the maximum local substep residual by 40.1%,

Figure 2.

Observed and simulated daily volumetric water content. Columns show the 2023 calibration, 2024 validation, and 2025 validation periods; rows show depths of 2, 4, and 8 in. Solid lines with circles denote observations, dashed lines denote simulations, and hatched intervals denote the seven-day spin-up excluded from the metrics. RMSE, root mean square error; NSE, Nash–Sutcliffe efficiency

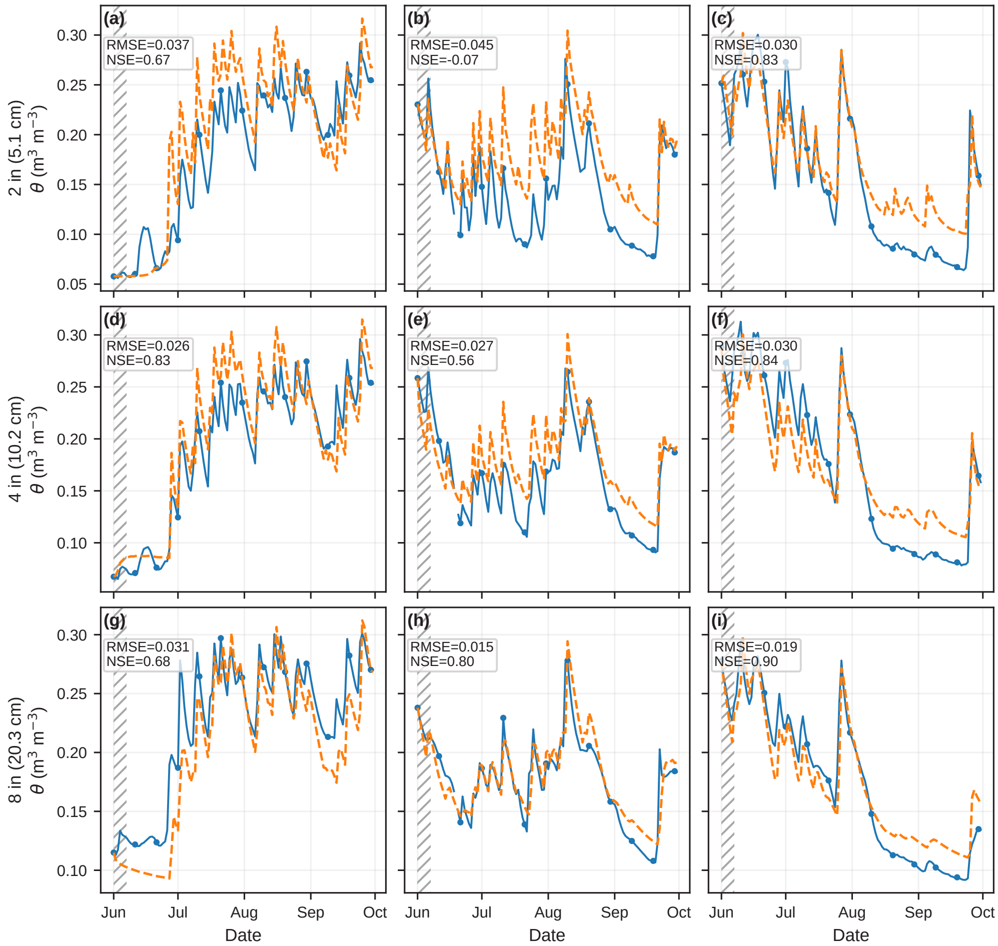


Table IV

Combined 2024–2025 validation metrics and moving-block bootstrap 95% confidence intervals. RMSE, root mean square error; CI, confidence interval; *n*, number of observations; NSE, Nash–Sutcliffe efficiency. RMSE and bias are in $m^3\ m^{-3}$

Depth	n	RMSE	RMSE 95% CI	Bias	Bias 95% CI	NSE	NSE 95% CI
2 in	229	0.0379	0.0338–0.0439	0.0258	0.0206–0.0344	0.603	0.302–0.709
4 in	229	0.0286	0.0259–0.0312	0.0097	0.0031–0.0190	0.777	0.672–0.822
8 in	229	0.0175	0.0150–0.0195	0.0036	−0.0009–0.0083	0.875	0.827–0.907

Table V

Seasonal water-balance components for the 40-cell production simulations. Flux components are in mm; R_M is the daily mass residual

Season	Rain	Net top flux	Drainage	Root uptake	Evaporation	Runoff	Max. $\lvert R_M \rvert$ (m)
2023 calibration	516.89	200.85	95.01	44.44	262.23	55.69	4.35×10^{-8}
2024 validation	322.07	53.86	20.04	50.80	240.86	29.18	3.95×10^{-9}
2025 validation	269.24	40.84	35.72	47.26	218.18	11.15	1.43×10^{-8}

yet increased the maximum hourly residual by 68.4%, the 99th percentile of hourly absolute residuals by 58.2%, and the cumulative residual by 52.1%. The estimated refinement exponents based on the hourly maximum and cumulative residual were therefore negative and were not interpreted as convergence orders. The zero-flux-floor branch was accepted once in PC32, with a residual of -1.2054×10^{-10} m, less than 0.1% of the cumulative seasonal residual and opposite in sign to the final positive bias.

Table VI

Hourly mass-balance audit for predictor–corrector configurations with 32 (PC32) and 64 (PC64) atmospheric-boundary substeps. The hourly and substep residuals are denoted by r_h and r_j, respectively

Metric	PC32	PC64	Ratio (PC64/PC32)
Maximum hourly $\lvert r_h \rvert$ (m)	1.2645×10^{-8}	2.1295×10^{-8}	1.684
Cumulative $\Sigma_h r_h$ (m)	2.4576×10^{-7}	3.7391×10^{-7}	1.521
99th percentile hourly $\lvert r_h \rvert$ (m)	2.1833×10^{-9}	3.4547×10^{-9}	1.582
Maximum substep $\lvert r_j \rvert$ (m)	1.4144×10^{-9}	8.4776×10^{-10}	0.599
Positive hourly residuals	1,895	1,900	–
Negative hourly residuals	1,033	1,028	–

Trace accounting was internally consistent. The sum of physical corrector residuals matched each hourly budget within 1.3×10^{-17} m, and the four-term reconstruction error was below 9.84×10^{-17} m. At 9 July 14:00, 31 of 32 PC32 residuals and 63 of 64 PC64 residuals were positive; the corresponding ratio of mean signed substep residuals was 0.832, insufficient to offset the doubling of substeps. At 21 July 00:00, the ratio was 0.955. The PC64–PC32 residual differences were primarily explained by the storage–bottom-flux interaction: on 9 July, a $+6.0755 \times 10^{-7}$ m bottom-flux change was nearly cancelled by a -5.9915×10^{-7} m storage change; on 21 July, bottom flux and storage contributed $+7.7433 \times 10^{-9}$ and $+2.4008 \times 10^{-9}$ m, respectively. Top flux and uptake changes were negligible. Paired-interval endpoint gaps of 1.93×10^{-5} and 9.74×10^{-7} m showed that PC32 and PC64 followed different trajectories, so the audit associated the nonconvergent residual behavior with storage–drainage coupling but did not isolate a defective lower-boundary quadrature. Detailed trace results are supplied in Supplementary Tables S1–S3.

3.4 Reference convergence

The mean L^2 error relative to the 640-cell solution decreased from 5.48×10^{-5} m^3 m^{-3} at $N = 40$ to 3.46×10^{-6} m^3 m^{-3} at $N = 320$(Table VII; Fig. 3). The observed convergence rates were not monotonic, likely reflecting nonlinear boundary switching and comparison with a finite reference. Nevertheless, the sequence demonstrates that the earlier choice of a 240-cell solution as a numerical reference, without an accompanying convergence study, would have been insufficiently justified.

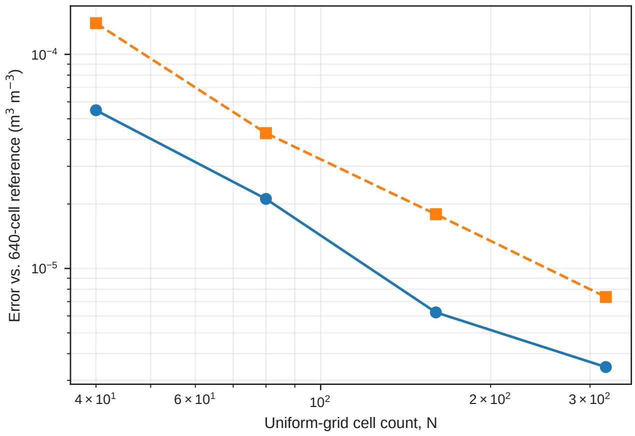
Table VII

Uniform-grid convergence for the 2023 calibration event relative to $N = 640$. The L^2 and L^∞ error metrics are in m^3 m^{-3}

N	Mean L^2	Max L^2	Mean L^∞	Observed order
40	5.482×10^{-5}	1.399×10^{-4}	3.485×10^{-4}	–
80	2.113×10^{-5}	4.286×10^{-5}	1.731×10^{-4}	1.38
160	6.233×10^{-6}	1.792×10^{-5}	7.224×10^{-5}	1.76
320	3.463×10^{-6}	7.357×10^{-6}	2.262×10^{-5}	0.85
640	reference	reference	reference	–

Figure 3.

Uniform-grid convergence during the 2023 calibration event. Solid circles show mean L^2 error and dashed squares show maximum L^2 error



3.5 Grid comparison and transfer to validation events

At $N = 40$ in the 2023 event, logarithmic grading reduced mean L^2 error by 47.9% relative to the uniform grid, whereas algebraic grading reduced it by 72.4%. Error decreased with cell count for all primary mappings (Fig. 4). The parameter sweep showed that $z_\star = 0.08$ m yielded a lower mean error than the primary logarithmic mapping ($z_\star = 0.04$ m), and $p = 1.8$ produced the smallest algebraic mean error but a larger maximum error than $p = 1.5$ (Fig. 6). This sensitivity shows that the mapping parameter is part of the numerical design and should not be tuned on validation data.

The ranking persisted in the held-out events (Table VIII; Fig. 5). Relative to the uniform 40-cell grid, logarithmic grading reduced mean error by 62.1% in 2024 and 71.8% in 2025. Algebraic grading remained the lowest-error primary mapping in both years. Thus, the result supports the transferability of surface-focused refinement but does not establish that logarithmic grading is uniquely superior.

Figure 4.

Mean profile error versus cell count for the primary mappings in the 2023 calibration event. Solid circles denote uniform, dashed squares logarithmic ($z_* = 0.04$ m), and dotted triangles algebraic ($p = 1.5$)) grids

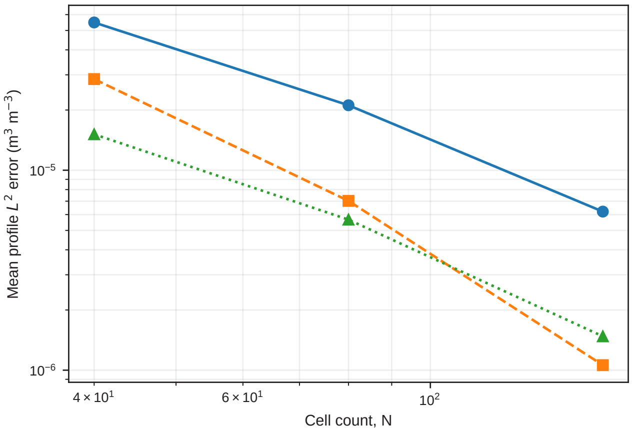


Figure 5.

Transfer of the 40-cell grid ranking from the 2023 calibration event to the 2024 and 2025 validation events. Within each grid group, bars from left to right represent 2023, 2024, and 2025; hatching provides an additional distinction beyond color

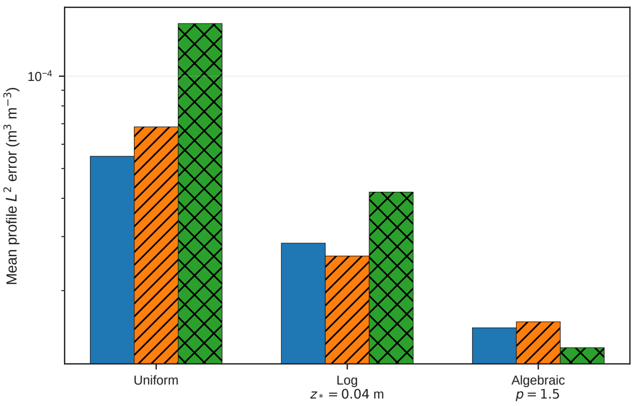


Figure 6.

Predeclared 40-cell grading-parameter sweep in the 2023 calibration event. Validation years were not used to select z_* or p

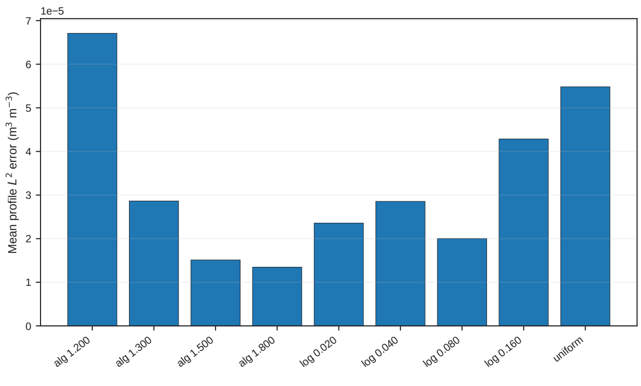


Table VIII

Mean profile L^2 error for 40-cell grids in independent seasonal event windows (m^3 m^{-3})

Grid	2023 calibration	2024 validation	2025 validation
Uniform	5.482×10^{-5}	6.836×10^{-5}	1.485×10^{-4}
Logarithmic, $z_* = 0.04$ m	2.856×10^{-5}	2.593×10^{-5}	4.190×10^{-5}
Algebraic, $p = 1.5$	1.513×10^{-5}	$1.582 \times 10^{-5}5$	1.303×10^{-5}

3.6 Practical magnitude of the numerical gain

The depth-aggregated held-out RMSE was 0.02922 m^3 m^{-3}. The mean $N = 40$ grid errors in the 2023 event were only 0.188%, 0.0977%, and 0.0518% of this value for uniform, logarithmic, and algebraic grids, respectively. Event-integrated top fluxes differed by less than 0.03 mm among the three mappings, and bottom fluxes by about 0.01 mm. Therefore, the effect of grid grading is numerically measurable and reproducible, but it does not materially change the model–observation discrepancy at the resolution tested.

4. Discussion

4.1 Implications of independent seasons

The multiyear design addresses a key limitation of evaluating fitted parameters only against the data used for calibration. The 2024 and 2025 seasons were never used for parameter selection, yet the model retained positive combined NSE at all depths. Nevertheless, the negative 2024 NSE at 2 in shows that temporal transfer was not uniformly successful. A homogeneous 0.35-m column with daily forcing did not reproduce all near-surface states across seasons, and the positive shallow bias suggests compensating effects among effective hydraulic properties, evaporation, and boundary storage.

Performance at 8 in was the most consistent across seasons, whereas performance at shallower depths varied. This depth pattern is compatible with stronger near-surface sensitivity to rainfall timing, evaporation, sensor support, and fine-scale layering (Bogena et al., 2010; Mittelbach et al., 2012; Dorigo et al., 2013). It does not prove a specific mechanism. The deeper 20- and 40-in sensors were deliberately excluded because they lie outside the modeled domain and may reflect layered or preferential pathways not represented by single-domain Richards flow (Šimůnek et al., 2003, 2016, 2024).

4.2 Interpretation of grid transfer

Surface-focused grids reduced profile discretization error in all three event windows. This transferability provides stronger evidence of robustness than a one-event comparison because the mapping parameters were fixed before the validation events. Nevertheless, the algebraic mapping outperformed the selected logarithmic mapping in every event, and the parameter sweep changed the ranking within each family. The defensible conclusion therefore concerns the location of refinement, not the universal superiority of a specific logarithmic mapping.

The convergence sequence also informs the interpretation. A 240-cell solution may be adequate for many purposes, but it should not be declared a reference without quantifying its distance from finer solutions. The nonmonotonic observed order near the finest grids likely reflects nonlinear boundary switching and comparisons with a finite reference; it argues against reporting a single asymptotic order as if it were universal.

4.3 Numerical accuracy versus agroenvironmental relevance

The grid errors were hundreds to thousands of times smaller than the model–observation discrepancy. Consequently, a 47–72% relative reduction in grid error does not imply a corresponding improvement in field prediction. For the present model and daily data, all three 40-cell grids are numerically adequate relative to structural and observational uncertainty. Grading would become practically relevant if the target were local front position, surface flux under subdaily rainfall, coupled solute transport, or an inverse calculation requiring much tighter numerical accuracy. This distinction prevents a numerical percentage from being presented as an agronomic performance gain.

4.4 Interpretation of the mass-balance audit

The hourly audit limits the strength of the conservation claim. Spatial finite-volume flux cancellation and the diagnostic identity were internally consistent, and trace sums reproduced the hourly budgets to numerical precision. However, doubling the boundary substeps did not reduce the hourly or seasonal residuals. Instead, numerous small corrector residuals had the same sign and accumulated coherently. The benchmark is therefore appropriately described as mass-balance audited, not as demonstrating asymptotic temporal mass conservation under the tested refinement.

The storage–bottom-flux pair accounted for the material PC64–PC32 difference in the two traced hours. The 9 July case was dominated by cancellation between large trajectory-dependent changes, whereas the 21 July case was more directly associated with the bottom-flux integral. Because the endpoint states differed, these results do not justify changing the sign or empirically correcting Q_{bottom}. A definitive follow-up analysis would compare one coarse interval with two half intervals, all restarted from an identical state, and would integrate drainage as an auxiliary cumulative state. This limitation does not alter the observed ranking of spatial grids, but it limits claims about strict closure and about the unique physical partitioning of drainage and storage.

4.5 Uncertainty, identifiability, and limitations

The 14-member candidate design documents parameter selection but does not establish a posterior distribution or global optimum. The small objective difference between the two best candidates and known correlations among soil hydraulic parameters indicate weak identifiability (Scharnagl et al., 2011; Gao et al., 2019). Distributions derived from ROSETTA or the Soil Survey Geographic Database (SSURGO) could provide documented priors, but would not replace site-specific retention and conductivity measurements (Schaap et al., 2001; Zhang and Schaap, 2017).

Other limitations remain. This study considered only one station, the warm seasons were initialized independently, and the forcing was aggregated daily even though hourly data were available. The surface control layer, runoff diagnostic, and exponential uptake are operational benchmark components rather than observed flux models. Interception, crop phenology, hysteresis, snow, layered hydraulic properties, and macropore flow were omitted. The NRCS observations are provisional and the exported tables did not include row-level quality flags in the numerical columns; this study therefore applied transparent completeness and range screens but could not reproduce all internal NRCS validation logic. Bootstrap intervals quantify uncertainty in validation metrics under block resampling, not parameter or structural uncertainty. The hourly PC32–PC64 audit was performed for one season and one spatial grid; it established internal accounting consistency but not a universal temporal convergence order, and it could not separate lower-boundary quadrature error from trajectory sensitivity without identical-state restart experiments.

5. Conclusions

A finite-volume Richards benchmark was calibrated using the 2023 warm season and tested without refitting using 2024 and 2025 data at Mahantango Creek SCAN 2028. Combined held-out NSE values were 0.603, 0.777, and 0.875 at 2, 4, and 8 in, respectively, although NSE was negative for the shallow 2024 series. The validation therefore supports conditional temporal transfer while revealing unresolved near-surface structural limitations and forcing errors.

A 40–640-cell convergence sequence established a defensible finite reference. Surface-focused grids consistently reduced discretization error in the calibration event and in both held-out event windows. The algebraic mapping was more accurate than the selected logarithmic mapping, so logarithmic grading cannot be described as universally optimal. All 40-cell grid errors were below 0.19% of the held-out model–observation RMSE; the mesh result is numerically reproducible but practically secondary to uncertainty in process representation, forcing, and sensor observations in this daily upper-soil application.

The hourly audit confirmed trace reconstruction and trace-to-budget consistency, but it did not demonstrate strict temporal convergence. PC64 reduced the maximum local substep residual while increasing the maximum hourly and cumulative residuals. The increase was associated with coherent same-sign residuals and the storage–bottom-flux balance, whereas the contribution of the single zero-flux-floor acceptance was negligible. Accordingly, the present contribution supports multiyear model validation, spatial mesh selection, and transparent mass-balance auditing, but does not support a claim of asymptotic mass conservation. Identical-state coarse-versus-half-step tests are required before attributing the remaining bias to lower-boundary quadrature.

Declarations

Funding: M.M.L.-F. received partial funding from project EVAAT-GCN under FINEP agreement 01.24.0661.00, ref. 3311/24.

Competing interests: The authors declare no competing interests.

Ethics approval and consent to participate: Not applicable; the study uses public environmental monitoring data and numerical simulations.

Consent for publication: Not applicable.

Data availability: The immutable NRCS exports, station metadata, sensor descriptions, processed daily dataset, quality-control record, and SHA-256 checksums are included in the supplementary reproducibility package to be submitted with the manuscript. SCAN observations are provisional and subject to revision.

Code availability: Versioned Python scripts, environment requirements, cached numerical outputs, commands, and audit tables needed to regenerate the reported results will be included in the supplementary reproducibility package.

Author contributions: Conceptualization, J.D.Z.-V., M.M.L.-F., and W.C.-L.; methodology, all authors; software, validation, formal analysis, and visualization, J.D.Z.-V. and M.M.L.-F.; data curation, J.D.Z.-V.; writing—original draft, J.D.Z.-V.; writing—review and editing, all authors; supervision, W.C.-L. All authors approved the submitted version.

Acknowledgments: The authors thank their institutions for academic support during the preparation of this manuscript.

Use of artificial intelligence: A language model was used to assist with editorial drafting, LaTeX and Word document preparation, and code development. It was not listed as an author. The human authors independently verified the data processing, equations, implementation, numerical outputs, references, interpretations, and conclusions.

References

Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration: guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56. Food and Agriculture Organization of the United Nations, Rome. https://www.fao.org/4/x0490e/x0490e00.htm

Bogena HR, Herbst M, Huisman JA, Rosenbaum U, Weuthen A, Vereecken H (2010) Potential of wireless sensor networks for measuring soil water content variability. Vadose Zone Journal 9:1002–1013. https://doi.org/10.2136/vzj2009.0173

Buda AR, Veith TL, Folmar GJ, Feyereisen GW, Bryant RB, Church CD, Schmidt JP, Dell CJ, Kleinman PJA (2011) U.S. Department of Agriculture Agricultural Research Service Mahantango Creek Watershed, Pennsylvania, United States: long-term precipitation database. Water Resources Research 47:W08702. https://doi.org/10.1029/2010WR010058

Caviedes-Voullième D, García-Navarro P, Murillo J (2013) Verification, conservation, stability and efficiency of a finite volume method for the 1D Richards equation. Journal of Hydrology 480:69–84. https://doi.org/10.1016/j.jhydrol.2012.12.008

Celia MA, Bouloutas ET, Zarba RL (1990) A general mass-conservative numerical solution for the unsaturated flow equation. Water Resources Research 26:1483–1496. https://doi.org/10.1029/WR026i007p01483

Dorigo WA, Wagner W, Hohensinn R, Hahn S, Paulik C, Xaver A, Gruber A, Drusch M, Mecklenburg S, van Oevelen P, Robock A, Jackson T (2011) The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements. Hydrology and Earth System Sciences 15:1675–1698. https://doi.org/10.5194/hess-15-1675-2011

Dorigo WA, Xaver A, Vreugdenhil M, Gruber A, Hegyiová A, Sanchis-Dufau AD, Zamojski D, Cordes C, Wagner W, Drusch M (2013) Global automated quality control of in situ soil moisture data from the International Soil Moisture Network. Vadose Zone Journal 12(3):1–21. https://doi.org/10.2136/vzj2012.0097

Eymard R, Gutnic M, Hilhorst D (1999) The finite volume method for Richards equation. Computational Geosciences 3:259–294. https://doi.org/10.1023/A:1011547513583

Farthing MW, Ogden FL (2017) Numerical solution of Richards' equation: a review of advances and challenges. Soil Science Society of America Journal 81:1257–1269. https://doi.org/10.2136/sssaj2017.02.0058

Gao H, Zhang J, Liu C, Man J, Chen C, Wu L, Zeng L (2019) Efficient Bayesian inverse modeling of water infiltration in layered soils. Vadose Zone Journal 18:190029, 1–13. https://doi.org/10.2136/vzj2019.03.0029

Hargreaves GH, Samani ZA (1985) Reference crop evapotranspiration from temperature. Applied Engineering in Agriculture 1:96–99. https://doi.org/10.13031/2013.26773

Ireson AM, Spiteri RJ, Clark MP, Mathias SA (2023) A simple, efficient, mass-conservative approach to solving Richards' equation (openRE, v1.0). Geoscientific Model Development 16:659–677. https://doi.org/10.5194/gmd-16-659-2023

Kavetski D, Binning P, Sloan SW (2001) Adaptive time stepping and error control in a mass conservative numerical solution of the mixed form of Richards equation. Advances in Water Resources 24:595–605. https://doi.org/10.1016/S0309-1708(00)00076-2

McKay MD, Beckman RJ, Conover WJ (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21:239–245. https://doi.org/10.1080/00401706.1979.10489755

Milly PCD (1985) A mass-conservative procedure for time-stepping in models of unsaturated flow. Advances in Water Resources 8:32–36. https://doi.org/10.1016/0309-1708(85)90078-8

Mittelbach H, Lehner I, Seneviratne SI (2012) Comparison of four soil moisture sensor types under field conditions in Switzerland. Journal of Hydrology 430–431:39–49. https://doi.org/10.1016/j.jhydrol.2012.01.041

Mualem Y (1976) A new model for predicting the hydraulic conductivity of unsaturated porous media. Water Resources Research 12:513–522. https://doi.org/10.1029/WR012i003p00513

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models. Part I—A discussion of principles. Journal of Hydrology 10:282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Richards LA (1931) Capillary conduction of liquids through porous mediums. Physics 1:318–333. https://doi.org/10.1063/1.1745010

Schaap MG, Leij FJ, van Genuchten MT (2001) ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. Journal of Hydrology 251:163–176. https://doi.org/10.1016/S0022-1694(01)00466-8

Scharnagl B, Vrugt JA, Vereecken H, Herbst M (2011) Inverse modelling of in situ soil water dynamics: investigating the effect of different prior distributions of the soil hydraulic parameters. Hydrology and Earth System Sciences 15:3043–3059. https://doi.org/10.5194/hess-15-3043-2011

Šimůnek J, Jarvis NJ, van Genuchten MT, Gårdenäs A (2003) Review and comparison of models for describing non-equilibrium and preferential flow and transport in the vadose zone. Journal of Hydrology 272:14–35. https://doi.org/10.1016/S0022-1694(02)00252-4

Šimůnek J, Hopmans JW (2009) Modeling compensated root water and nutrient uptake. Ecological Modelling 220:505–521. https://doi.org/10.1016/j.ecolmodel.2008.11.004

Šimůnek J, van Genuchten MT, Šejna M (2016) Recent developments and applications of the HYDRUS computer software packages. Vadose Zone Journal 15(7):1–25. https://doi.org/10.2136/vzj2016.04.0033

Šimůnek J, Brunetti G, Jacques D, van Genuchten MT, Šejna M (2024) Developments and applications of the HYDRUS computer software packages since 2016. Vadose Zone Journal 23(4):e20310. https://doi.org/10.1002/vzj2.20310

Svyatsky D, Lipnikov K (2017) A second-order accurate finite volume scheme with the discrete maximum principle for solving Richards' equation on unstructured meshes. Advances in Water Resources 104:114–126. https://doi.org/10.1016/j.advwatres.2017.03.015

USDA Natural Resources Conservation Service (2024) Report Generator Help Guide. National Water and Climate Center. https://www.nrcs.usda.gov/resources/data-and-reports/report-generator. Accessed 11 July 2026.

USDA Natural Resources Conservation Service (n.d.-a) Soil Climate Analysis Network. https://www.nrcs.usda.gov/resources/data-and-reports/soil-climate-analysis-network. Accessed 11 July 2026.

USDA Natural Resources Conservation Service (n.d.-b) Mahantango Ck (2028): site information and reports. https://wcc.sc.egov.usda.gov/nwcc/site?sitenum=2028. Accessed 11 July 2026.

van Genuchten MT (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Science Society of America Journal 44:892–898. https://doi.org/10.2136/sssaj1980.03615995004400050002x

Virtanen P, Gommers R, Oliphant TE, et al. (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods 17:261–272. https://doi.org/10.1038/s41592-019-0686-2

Zhang Y, Schaap MG (2017) Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). Journal of Hydrology 547:39–53. https://doi.org/10.1016/j.jhydrol.2017.01.004