Small-Mutation Screening of WEPP Hillslope Peak Flow

Remediation Strategies and Validation Findings

Roger Lew*

Version 0.2 — October 6, 2026

Abstract

The October update evaluates two responses to the surface-return timing defect: increasing initial random roughness and correcting return-water peak assignment. Roughness inflation can avoid the problematic positive-excess-duration coupling for some events, but changes physical model inputs and is not a general repair. An implemented, parameter-free candidate instead uses hourly return information when available and a 24-hour return rate otherwise. The repeated Topanga census reduces the fivefold peak-departure rate from 0.303% to 0.108%, without eliminating twofold sensitivity. A matched Cedar Creek watershed comparison preserves long-term water yield closely but reveals an unresolved routed hydrograph-volume discrepancy: the 10 cm candidate's whole-record shortfall increases from 0.463% to 1.321%, and a selected five-day deficit increases from 0.83% to 18.30%. The candidate is implemented and research-tested, not unconditionally validated for hydrograph use or approved for release.

The original August census and its methods are retained as historical evidence. The repeated study applies the same one-at-a-time ±1% conductivity and ±0.01 cover mutations across burned and unburned Topanga scenarios. These enriched-discovery experiments diagnose numerical sensitivity, not observed-flow accuracy or cross-site prevalence.

1 Purpose and Status

WEPP reduces each modelled runoff event to quantities used by erosion and routing calculations, including runoff volume and a peak runoff rate. A peak is a statement about subdaily timing: the same event volume can produce a low peak when spread over time or a high peak when compressed. Previous Topanga work found controlled cases in which event runoff changed little while the reported peak changed dramatically. Source instrumentation traced one such case to the treatment of daily soil-water surface return and a second discontinuity to peak-solver selection.

The present study asks a broader but deliberately simpler question: how often does the local hillslope peak estimate respond abnormally to a very small one-at-a-time parameter perturbation? It is a screening census, not a calibration exercise and not a parameter uncertainty analysis. The perturbations are numerical probes chosen to interrogate continuity near each frozen baseline.

The original census is an interim site study. Topanga is the first, enriched-discovery site in a planned multi-site audit. Additional rain- and snow-dominated sites remain future work. Their addition does not require expanding the mutation magnitude: the Topanga results show that the original small probes are already sensitive enough to expose the response of interest.

The October assessment adds a repeated Topanga mutation census and a separate full-watershed Cedar Creek comparison. It does not convert the original enriched census into a cross-site prevalence

*University of Idaho; report prepared with agentic coauthor support from OpenAI Codex.

estimate. In the historical sections, "present WEPP" refers to the unpatched implementation under study, not to the October candidate. The current remediation disposition is in Section 3; the August results remain explicitly historical.

2 Hydrologic and Implementation Background

2.1 Expected first-order response

For infiltration-excess surface runoff, increasing K_{sat} ordinarily increases infiltration capacity and reduces rainfall excess. Increasing ground cover ordinarily reduces and delays surface runoff through interception, surface protection, and hydraulic resistance. The first-order expectation for both mutations is therefore inverse:

$$\Delta K_{\text{sat}} > 0 \quad \Rightarrow \quad \Delta\texttt{PeakRO} < 0, \qquad \Delta C > 0 \quad \Rightarrow \quad \Delta\texttt{PeakRO} < 0. \tag{1}$$

Antecedent saturation can complicate the size of these responses. Water that enters an already wet profile can be returned to the surface. It does not, however, follow that a daily returned-water volume has a uniquely determined subdaily timing or peak rate.

2.2 What the official documentation describes

WEPP was designed primarily to predict erosion for conservation and land management analysis. Chapter 4 says that surface hydrology supplies the erosion component with runoff volume, peak discharge, rainfall-excess intensity, and rainfall-excess duration [4]. In continuous simulation, most peaks are estimated with an approximation to kinematic-wave routing rather than read from a retained, fully routed hydrograph. The erosion calculation then assumes steady discharge at the peak and defines an effective duration

$$D_e = \frac{Q_v}{q_p}. \tag{2}$$

This rectangle has the calculated peak as its height and the event runoff as its area. Effective duration is therefore calculated after the peak. It is not an independently routed hydrograph duration or time to peak. A very short reported duration is a consequence of a large calculated peak, not separate evidence that the physical runoff pulse was short.

The continuous-simulation approximation was evaluated over the domain shown in Table 1. Its three equation branches had reported mean errors of 1%, 10%, and 5% relative to kinematic-wave routing, or 6.6% combined. These were comparisons with another model calculation, not with observed event peaks. Chapter 4 did not prescribe warnings, clamps, or another method outside the tested domain, and it did not report tests of continuity under small changes in K_{sat}, vegetation, or cover.

Table 1: Domain used to develop the Chapter 4 continuous-simulation peak approximation.

Dimensionless quantity	Minimum	Maximum
Discharge time, t^*	0.09	10.0
Mean-to-maximum excess ratio, v^*	0.08	1.0
Peak ratio, q^*	0.07	8.0

The documentation also gave substantial warnings. Chapter 4 says WEPP does not explicitly represent variable contributing areas or return flow, calculates rainfall excess before routing, does not fully couple routing and recession infiltration in all cases, and expects more peak error for multipeaked hydrographs. It nevertheless states that surface hydrology should perform best for large events on one OFE, assuming correct parameters. Chapter 6 reports that runoff volume compared more favourably with observations than peak rate in one drainage evaluation; peak validity remained undetermined because hydraulic roughness had not been measured [5]. These statements support caution, but they do not describe the later surface-return timing problem.

2.3 How the present implementation constructs a peak

The source and Topanga instrumentation show the following sequence:

1. Rainfall and infiltration produce a subdaily rainfall-excess series.
2. `REID` sums only intervals with positive excess to obtain the selected duration, excluding zero-excess gaps between bursts.
3. The daily water balance calculates `surdra`, water that cannot remain in the full surface soil layer.
4. The peak path calculates

$$\text{return rate} = \frac{\text{daily surface-return depth}}{\text{positive-excess duration}} \tag{3}$$

   and adds that constant rate only to intervals already having positive infiltration excess.
5. An exactly zero stored ponding time selects the approximate `APPMTH` solver; a positive value selects `HDRIVE` characteristic routing. No check requires the two methods to agree at the boundary.
6. After the peak is returned, WEPP calculates $D_e = Q_v/q_p$, capped at one day.

The operation conserves the daily water amount, but it does not preserve when that water reached the surface. For example, if 20 mm is returned on a day with two hours of positive infiltration excess, the code adds 10 mm/h to those two hours. If a small parameter change leaves the same depth but shortens the selected duration to one hour, the imposed rate doubles. The water volume is unchanged, but the constructed peak forcing is not.

Enabling the hourly water balance does not preserve the missing timing in this path. The hourly routine calculates surface return in `ui_scrunf`, but then sums those hourly values into daily `surdra`. The peak routine consumes that daily sum rather than the hourly return series. It assigns new timing from the infiltration-excess intervals even though hourly return values had already been calculated.

When surface return exists without any positive-excess duration, the code must choose another time window. It tries a storm-derived duration, then an upslope flow duration, and finally a 24-hour fallback. These choices preserve volume, but none records when the soil actually returned the water.

2.4 Controlled experiments that identified the defect

The census followed several controlled Hill 106 experiments. They address different questions and should not be treated as demonstrations of one common cause.

Table 2: Controlled Topanga evidence preceding the census.

Experiment	Controlled change	Peak response	What it establishes
1980 burned K_{sat}	20 to 35 mm/h	47.71 to 92.71 mm/h	Instrumented surface-return compression; solver switch secondary
1986 canopy screen	LAI 5 to 6; canopy 0.70 to 0.90	3.56 to 294.42 mm/h	Severe discontinuity; immediate branch unresolved
1986 cover bracket	Paired cover 0.80 to 0.90	312.29 to 3.56 mm/h	Compact boundary for replay
1986 cover matrix	36 canopy/cover combinations	3.516 to 323.029 mm/h	Two regimes while runoff remained 42.834–44.115 mm
2005 cover step	Paired cover 0.80 to 0.90	97.74 to 157.62 mm/h	Recurrence on another date with nearly fixed runoff

The 1980 event was traced with diagnostic logging placed between the water balance and the peak solvers. The logging recorded values ordinary WEPP output does not expose and did not change the hydrology or peak equations.

Table 3: Internal operands for the instrumented February 14, 1980 event.

Recorded value	K_{sat} 20	K_{sat} 35
Final runoff	60.122 mm	58.630 mm
Daily surface return	36.999 mm	51.637 mm
Positive-excess duration	4,536.60 s	2,478.44 s
Imposed return rate	29.36 mm/h	75.00 mm/h
Maximum excess before insertion	34.69 mm/h	19.35 mm/h
Maximum forcing after insertion	64.05 mm/h	94.35 mm/h
Stored ponding time	0 s	5,100 s
Selected method	`APPMTH`	`HDRIVE`
Published `PeakRO`	47.710 mm/h	92.714 mm/h

The higher-K_{sat} run behaved as expected before the surface-return insertion: it had less runoff and a lower maximum infiltration-excess rate. The reversal appeared when a larger daily return was divided by a duration that was 45% shorter. Separate diagnostic builds then forced both already-constructed event inputs through the same solver.

Table 4: Forced-solver counterfactual for the 1980 K_{sat} event.

Solver assignment	K_{sat} 20 peak	K_{sat} 35 peak
Normal selection	47.710 mm/h	92.714 mm/h
Both `HDRIVE`	61.751 mm/h	92.714 mm/h
Both `APPMTH`	47.710 mm/h	85.162 mm/h

The solver switch enlarged the difference but did not create it. Under either common solver, the higher-K_{sat} event retained the larger peak because its surface return had already been assigned a much higher rate. This controlled case motivated a census across the complete Topanga hillslope population.

2.5 Shared implementation and source scope

The public Daily Erosion Project repository contains a verbatim import of the September 2024 Windows WEPP source. Comparison with the reviewed WEPP-Forest source found byte-identical `REID`, `APPMTH`, `HDRIVE`, `GRNA`, and shared hydrology-state definitions. Both lineages contain the same daily-surplus insertion, exact-zero solver selection, and effective duration calculation [6]. The Topanga runtime instrumentation used WEPP-Forest. The Windows comparison is static source evidence; it does not claim that every Windows executable was run on the Topanga fixtures.

2.6 The missing coupling and why a direct repair was abandoned

The fundamental shortcoming is not simply that WEPP calculates saturation excess. It is that the peak-flow path does not carry the calculated hourly saturation-excess series forward as a time-varying source to the kinematic-wave calculation. The hourly water balance knows when water is returned to the surface, but the legacy path reduces those hourly values to the daily scalar `surdra`. After the original infiltration-excess event has been constructed, the peak routine attempts to reconcile that missing runoff by dividing the daily `surdra` depth by a selected duration and adding the resulting constant rate to the existing rainfall-excess array. The kinematic-wave solver, when selected, therefore receives a reconstructed forcing series rather than the actual hourly combination of infiltration excess and saturation excess. This distinction is important: the addback occurs before the final `HDRIVE` call, but after the model has discarded the timing needed to define the source correctly. Volume can still balance for the day while the imposed rate, event shape, and resulting `PeakRO` depend on an artificial duration.

The production Papanicolaou (Lane D) water-balance implementation in openWEPP takes a different approach. It retains a closing 24-bin surface-runoff ledger and preserves saturation-return water in the hour in which the soil-water calculation produced it. That same closed hourly ledger supplies the active Lane D routing source. The published hillslope peak is defined more conservatively as the maximum hourly mean flow from the ledger, rather than as an instantaneous subhourly kinematic-wave peak. The complete frozen Topanga small-mutation design was then repeated in openWEPP: all 1,088 trials completed, with 1,913,158 finite paired-positive peak rows, and no row combined runoff volume within 5% of baseline with a peak change of twofold or greater. Thus, the openWEPP study did not reproduce the unexplained, volume-stable `PeakRO` anomalies found in legacy WEPP. This is evidence for the benefit of preserving hourly source timing, not a claim of observed-flow accuracy, subhourly peak resolution, or watershed-outlet validation [10].

A bounded WEPP-Forest repair was also attempted. Its intended change was to retain hourly saturation excess separately, combine it with infiltration excess, and let the existing characteristic solver route that time-varying lateral inflow. The feasibility work was stopped before production wiring because the existing solver could not accept the required series without changing deeper contracts. In the mixed Hill 106 test, its normal 95%-volume stopping rule terminated at 75,600 seconds even though forcing continued through 86,400 seconds; only 93.58% of the combined input had arrived by that time. In a saturation-only test, exact zero-flow intervals caused a floating-point trap in the characteristic integration, where cumulative terms are divided by the current step rate. Retaining the legacy substitute of a tiny positive rate would avoid the trap only by erasing the

dry gaps that the repair was meant to preserve. The attempted change also crossed established ownership boundaries for infiltration, depression storage, recession infiltration, runoff volume, event duration, and erosion inputs. It was therefore not a safe, local patch: it would require a zero-aware characteristic formulation or another explicitly time-series solver, followed by renewed water-balance, erosion, and routing validation. The experimental code was removed and WEPP-Forest was restored byte-for-byte to the diagnostic release; the failed feasibility evidence was retained in the closed work package [11].

3 Remediation Strategies and October Validation

The practical choice is between changing a physical input to avoid a defective pathway and correcting the rate assigned to surface-return water. The completed experiments support the targeted correction as the more appropriate strategy, but not an unconditional release claim. The candidate's routed hydrograph has an unresolved volume inconsistency despite nearly unchanged water-yield ledgers.

Table 5: Remediation options and current evidence boundary.

Option	Mechanism	Disposition
Raise initial roughness	Change depression storage and rainfall-excess timing; some events leave the compressed-return path	Diagnostic workaround only; changes runoff and resistance, with no universal safe threshold
Return-peak estimator	Use hourly return ledger, or daily return divided by 86,400 s; retain existing water-transfer path	Implemented, no new flags; mutation improvement, unresolved watershed hydrograph closure
Fully coupled routing	Route combined rainfall excess and timed return with consistent storage and recession bookkeeping	Earlier feasibility attempt stopped at solver-contract failures; not the implemented candidate

3.1 What initial random roughness changes

`rrinit` initialises random surface roughness, in metres. It is not itself a reservoir depth or a return-flow duration. Current roughness `rrc` contributes to depression storage in `IRS` and to surface friction in `FRCFAC`; roughness and erosion feedback remain active. The implemented depression-depth relation is

$$d_{\mathrm{dep}} = \max\{0,\ 0.112r + 3.1r^2 - 1.20rS\}, \tag{4}$$

where r is current random roughness in metres and S is slope as a fraction. For a uniform 30% slope, evaluating the formula at the initial roughness gives:

Table 6: Illustrative depression depth at 30% slope, before equivalent-plane averaging or subsequent roughness evolution.

Initial roughness	Depression depth	Increase over 10 cm
10 cm	6.20 mm	—
17 cm	47.43 mm	41.23 mm
60 cm	967.20 mm	961.00 mm

These depths are not storage through which all watershed streamflow must pass. `GRNA` applies depression storage to rainfall excess; sufficiently large storage can remove that excess event. The

subsequent soil-water balance can still return water from a full profile through `surdra`, without applying the same depression-storage subtraction again. Lateral flow and groundwater baseflow are also not routed through this rainfall-excess storage calculation. Consequently, a very large nominal storage increase need not remove a comparable depth of total streamflow.

In the original return branch, positive rainfall-excess duration makes the imposed return rate equal to daily return divided by that duration. Eliminating positive excess can move the calculation to a different return-duration path: storm duration, upslope duration, or ultimately 24 hours. This can prevent an anomalous peak without eliminating surface return or repairing its timing. It can also change the selected peak solver. Raising roughness does not guarantee a monotonic reduction of peaks.

Anurag's email identifies 17 cm as a threshold in his undisturbed-forest comparison, after initially testing 60 cm against a 6 cm baseline. That is a case-specific transition, not a WEPP constant or a physical discontinuity at 17 cm. Its exact event mechanism requires tracing the operands. His burned–unburned comparison also changes other scenario properties; it is not the matched roughness-only Cedar Creek experiment below. The embedded screenshot was not available in the retained text attachment, so no quantitative claim about its axes, aggregation, or executable identity is made here.

The disturbed templates examined during this investigation are not described by a single universal 6 cm default. OldForest is 10 cm unburned, 4 cm at low severity, and 6 cm at moderate and high severity; shrub is 6 cm throughout, and tall grass is 2 cm throughout. The unburned `hand-to-mouth-drought` landscape is predominantly shrub (92.67%), explaining its 6 cm value. The 10 cm records in `warming-championship` are consistent with the existing forest template; they are not evidence that Marta changed the default.

External consequences of roughness inflation. Changing roughness changes rainfall-excess onset, duration and volume, hydraulic resistance, and potentially erosion and downstream synchronisation. It therefore changes the scientific scenario in addition to changing exposure to the defect. Burned–unburned peak ordering is not sufficient justification for selecting a roughness value, and numerical stabilisation is not evidence that a larger roughness is physically appropriate. Fire-related loss of litter or microtopographic storage must be evaluated as a separate parameterisation question; no universal roughness response by burn severity is inferred here. The 17 and 60 cm values are diagnostic interventions, not recommended defaults.

3.2 The implemented parameter-free return-peak correction

The candidate adds `SURPEAK` and its integration in `IRS`. It changes peak assignment, not the existing daily surface-return transfer or the depression-storage parameterisation. It introduces no user parameter or flag. For non-hourly water balance, the estimator is

$$q_{\text{candidate}} = q_{\text{rain}} + R/86{,}400, \tag{5}$$

where R is daily returned depth and rates use compatible depth/time units. Spreading return over 24 hours is a simple explicit assumption when subdaily timing is unavailable. It can understate a brief return pulse; it is not a validated physical return duration or a claim of instantaneous peak accuracy.

With hourly water balance, the estimator uses the existing hourly return ledger and a volume-closing hourly rainfall-excess distribution. It adds the increase in maximum hourly forcing caused by return to the rain-only routed peak, bounded below by the maximum hourly return rate. Its

result lies between $\max(q_{\text{rain}}, q_{\text{return,max}})$ and $q_{\text{rain}} + q_{\text{return,max}}$. Return is not separately attenuated through a fully coupled kinematic-wave calculation. This bounded estimator is deliberately less ambitious than the abandoned direct-coupling patch described above.

The existing OFE water-transfer and contour paths remain in place. Altered peaks may still change erosion, rill-width feedback, and channel routing; unchanged intended water accounting does not imply byte-identical output. The release-source audit found 485 source/include/build files matching the retained unpatched tree after normalising the build system's capacity-include swap. Candidate differences are confined to `irs.for`, `surpeak.for`, and `makefile`. Complete executable hashes and the compiler-provenance limitation remain in the evidence packages [1, 3].

3.3 Repeated Topanga mutation study

The `hand-to-mouth-drought` repeat completed all 280 baselines and 1,088 eligible mutations over 45 years, with the candidate on both sides of each pair. It retained the original $\pm 1\%$ K_{sat} and ± 0.01 paired-cover probes and did not change roughness or depression storage. Shared inputs match the original census byte-for-byte. The historical comparison is nevertheless descriptive, not a same-lineage patch-only causal estimate: source lineage and observation grain differ [1].

Table 7: Candidate-build mutation response. Denominators apply to each response's eligible positive pairs, not independent storms or hillslopes.

Response	Positive pairs	Outside 0.5–2	Outside 0.2–5
Runoff depth	218,974	453 (0.207%)	34 (0.016%)
Sediment delivery	128,469	372 (0.290%)	90 (0.070%)
Peak rate	200,932	1,021 (0.508%)	218 (0.108%)

The fivefold peak-departure rate falls from 604/199,086 (0.303%) in August to 218/200,932 (0.108%). The high-rate scatter is visibly tighter, and the remaining fivefold tail reaches only 17.39 mm/h on the baseline side and 21.29 mm/h on the mutant side. In the central 0.5–2 band at baseline peaks of 30–100 mm/h, K_{sat} incongruence falls from 24.1% to 4.68% unburned and from 26.0% to 2.24% burned, using the original convention including ties. These findings support the observed reduction and downward shift of the most conspicuous departures.

They do not support saying that all sensitivity decreased. Twofold peak departures remain at 0.508%, versus 0.497% originally; 594 of the 1,021 current twofold departures retain runoff within 5% of baseline. Sediment's twofold-tail rate rises from 0.195% to 0.290%. Of the 218 remaining fivefold peak departures, 184 have no return in either run, the same no-return count reported originally. Residual rows are screening candidates, not individually adjudicated defects.

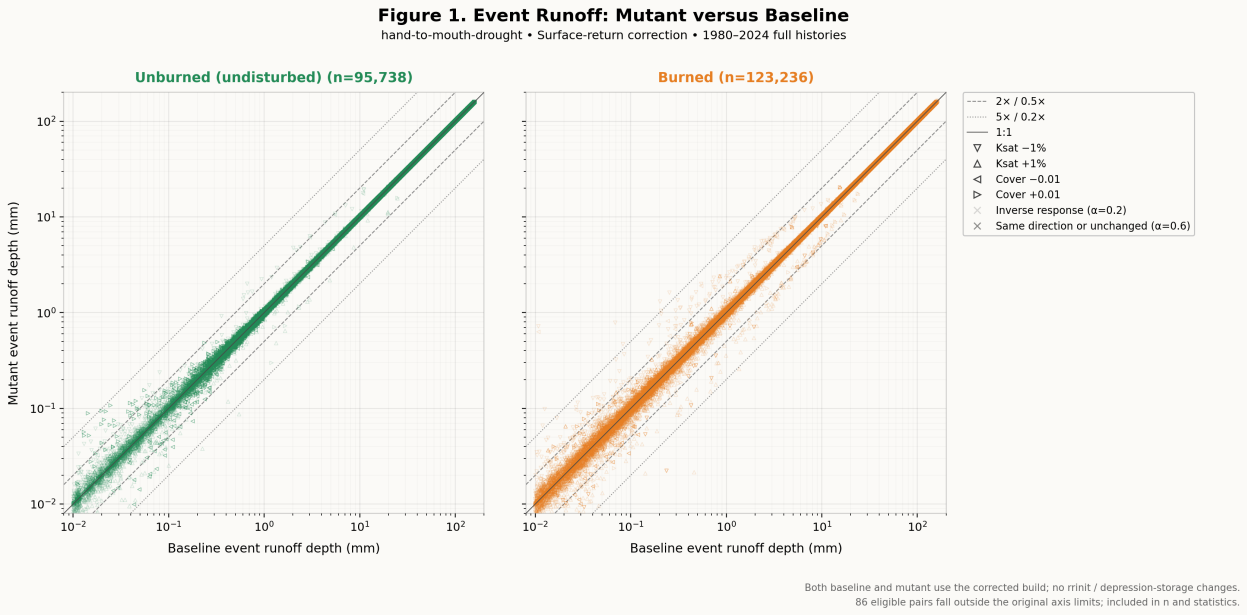


Figure 1: Candidate-build runoff response to the original small mutations. Original floors, markers and ratio guides are retained. Statistics include 86 eligible pairs outside the displayed axis bounds.

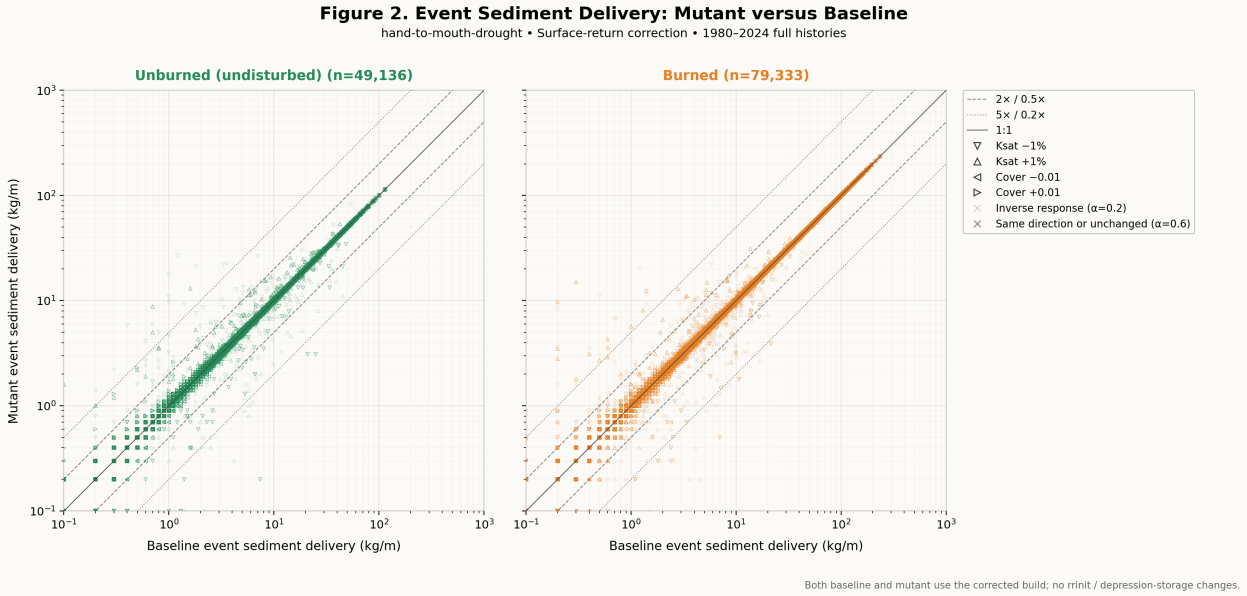


Figure 2: Candidate-build event sediment delivery. Printed sediment is quantised to 0.1 kg/m; 70,424 plotted pairs are exact ties. The ledger separately retains 549 baseline-positive-only and 619 mutant-positive-only sediment rows rather than hiding event changes.

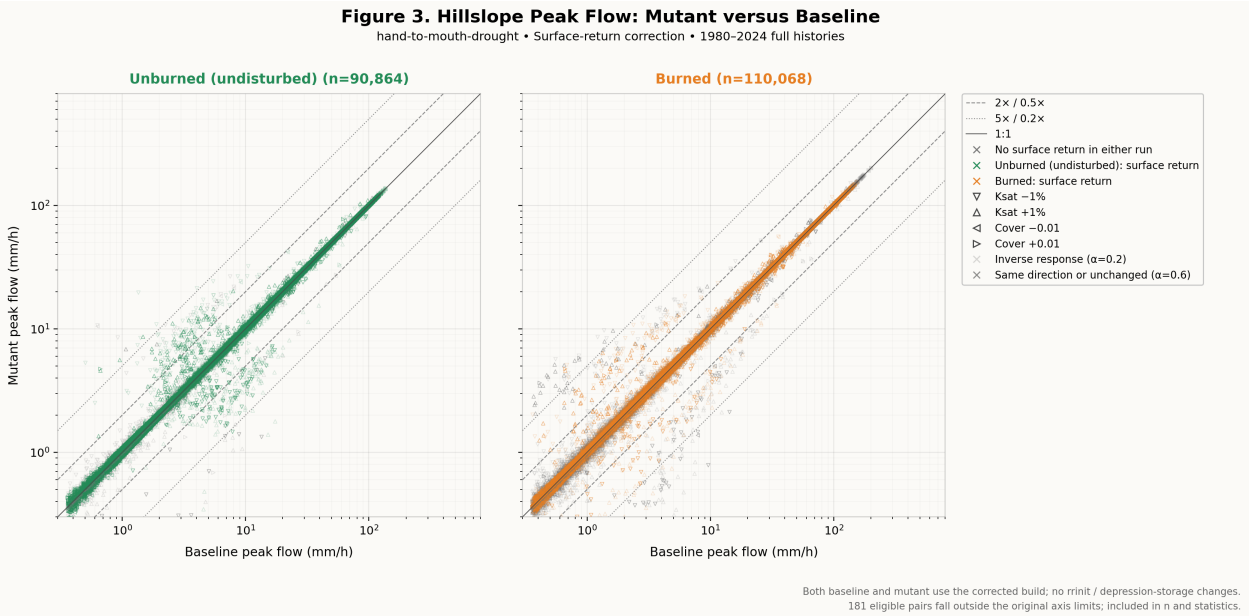


Figure 3: Candidate-build hillslope peak response. The severe high-rate tail is reduced, not eliminated. Statistics include 181 eligible pairs outside the original axis bounds. Both members of every pair use the candidate; this is not a direct original-versus-candidate scatter.

3.4 Cedar Creek watershed and roughness experiment

The `warming-championship` run on forest is titled Cedar Creek and contains 864 hillslopes, 1,904 OFEs, 371 channels and outlet element 1235. All 24 years, 1980–2003, including initialisation, were rerun. Only the 1,885 initial-condition roughness records originally at 10 cm were changed to 17 or 60 cm. Two 6 cm and seventeen 0.8 cm records were preserved. Soil, climate, cover, hourly water balance and channel parameters were held fixed. This is a roughness intervention, not a burned–unburned comparison.

Each build comparison comprises 2,592 hillslope executions, three watershed executions and one channel-output-mode control, all successful. The observer requests full 600-second discharge without changing the routing timestep; seven canonical watershed outputs are byte-identical between observer and original output mode. All 14 untouched hillslope controls pass. Each main lane contains 8,766 days and 1,262,304 discharge samples. Original project inputs remain unchanged. The legacy experiment uses verified `wepp_260803` and `wepp_260803_hill` binaries with inputs matching each candidate lane exactly [2, 3].

Table 8: Daily outlet-peak departures from each build's own 10 cm case. These ranges are not changes in the single largest simulation peak.

Build	Roughness	Minimum	Maximum	Days with $\lvert\Delta\rvert > 1\%$
Original	17 cm	−18.2345%	+4.8352%	26
Candidate	17 cm	−1.5996%	+0.4515%	2
Original	60 cm	−18.2345%	+4.7757%	61
Candidate	60 cm	−2.5049%	+0.5672%	3

The original build is sensitive to roughness: on November 6, 1983 its daily peak falls from 1.36176 to 1.11345 m^3/s at both 17 and 60 cm. That shared response is consistent with crossing a threshold, but is not an event-level trace of its cause. Some physical roughness sensitivity should remain after repair. The candidate reduces the extreme outlet sensitivity here, not all hillslope relative sensitivities or their frequency. Small denominators and different screened populations matter at hillslope scale.

Whole-record outlet-volume reductions are almost identical across builds: approximately 0.01098% at 17 cm and 0.02106% at 60 cm. Candidate maximum daily peaks are 43.06409, 42.77490 and 42.67082 m^3/s for 10/17/60 cm, all on April 27, 1990. Original maxima are 50.50288, 50.47683 and 50.53258 m^3/s: increasing roughness does not uniformly lower the maximum.

3.5 Totalwatsed yield is not the routed outlet hydrograph

Freshly rebuilt `totalwatsed` streamflow is PASS surface runoff plus bottom-OFE lateral flow plus postprocessed groundwater baseflow. It is a daily, pre-channel yield measure, not an instantaneous routed outlet discharge. The retained groundwater settings use zero initial storage, a 0.04/day baseflow coefficient, and zero deep seepage. The watershed area is 19,258,143.09 m^2.

Table 9: Candidate-build totalwatsed streamflow. Daily mean rates cannot resolve within-day peak or travel-time changes.

Roughness	Total volume (million m^3)	Mean annual depth (mm)	Max daily mean (m^3/s)
10 cm	1267.7357	2742.8563	28.7701
17 cm	1267.5960	2742.5542	28.5946
60 cm	1267.4677	2742.2766	28.1919

At 10 cm, annual component depths are approximately 416.36 mm surface runoff, 874.85 mm lateral flow and 1451.65 mm baseflow. About 85% of this streamflow therefore comes from the latter two components. Combined with the separation between depression storage and soil-water return, this helps explain why 967 mm of nominal depression depth does not remove a comparable depth of streamflow. It does not demonstrate negligible roughness effects in every hillslope or watershed.

The three candidate daily-yield series almost overlap. Total yields change by −0.01102% and −0.02113% at 17 and 60 cm. A sensitivity-focused selection of smaller totalwatsed events (lower-half local peaks, separated by at least seven days, prominence at least 0.1 m^3/s) found example daily departures of roughly −0.14% to −0.23%. This small *within-candidate roughness* response should not be confused with the large *between-build* smaller-event peak changes below.

3.6 Matched 10 cm original versus candidate

At unchanged 10 cm inputs, candidate totalwatsed yield differs by only +0.305 m^3 out of 1.2677 billion m^3. Only 106 of 7,573,824 hillslope daily runoff-volume records differ at printed precision, by at most 0.1 m^3; 392,667 hillslope peak records change. H352 on January 31, 1995 retains 4,077 m^3 while its peak changes from 4.0447 to 0.10625 m^3/s. This confirms separation of volume from rate assignment, not physical accuracy of either rate. Other peaks increase.

At the outlet, 2,356 daily peaks differ by more than 1%, with departures from −46.97% to +156.00%. There are 1,930 lower, 2,417 higher and 4,419 unchanged daily peaks at printed precision. The original record maximum is 50.50288 m^3/s on October 27, 1994; the candidate is 35.45196 m^3/s that day. The candidate's own record maximum, 43.06409 m^3/s, occurs on a different date. The outlet-volume ledger changes by just −4.25 m^3 over 24 years, but individual daily volumes are not invariant: 44 days change by more than 1%, spanning −8.62% to +2.57%.

For smaller outlet events, the three most responsive non-overlapping examples increase by 111.1%, 98.9% and 86.7%. These use the lower half of original local peaks (at most 4.087 m^3/s), with seven-day separation and 0.1 m^3/s prominence, then rank absolute relative build differences. They are sensitivity examples, not representative or typical-event estimates. Their narrower, higher curves illustrate why daily yield agreement cannot validate subdaily routing.



Figure 4: Within-build roughness sensitivity at Cedar Creek. Shared vertical scales distinguish the original and candidate responses without changing the reference case between roughness lanes.

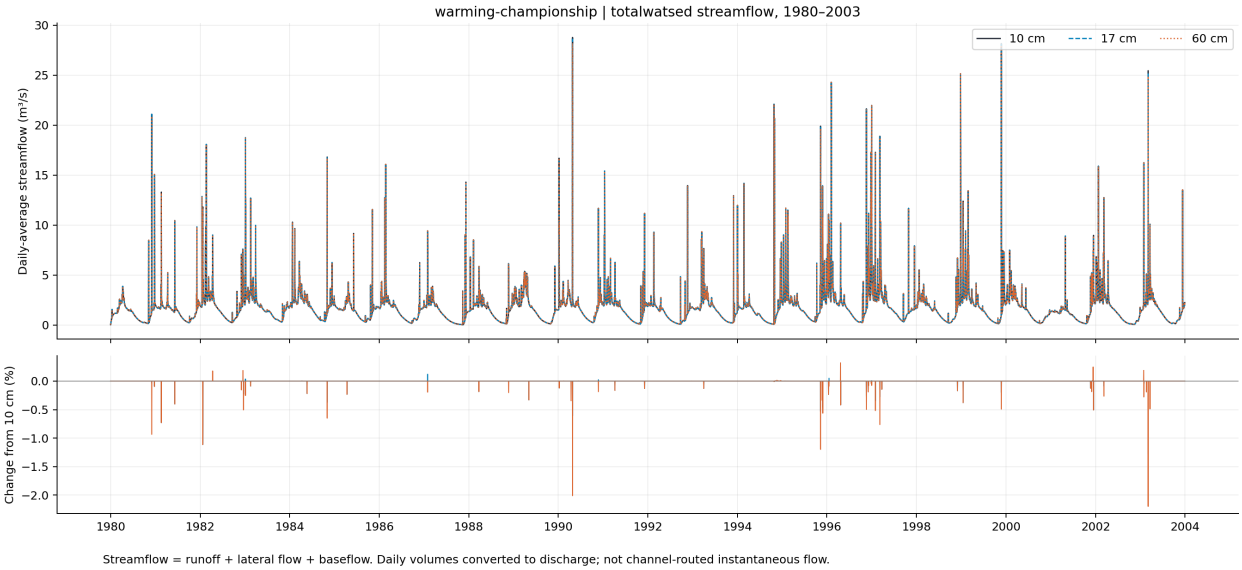


Figure 5: Candidate-build totalwatsed streamflow for all three roughness values, expressed as daily mean flow. The lower panel exposes differences hidden by near-overlapping series; these are not routed subdaily hydrographs.

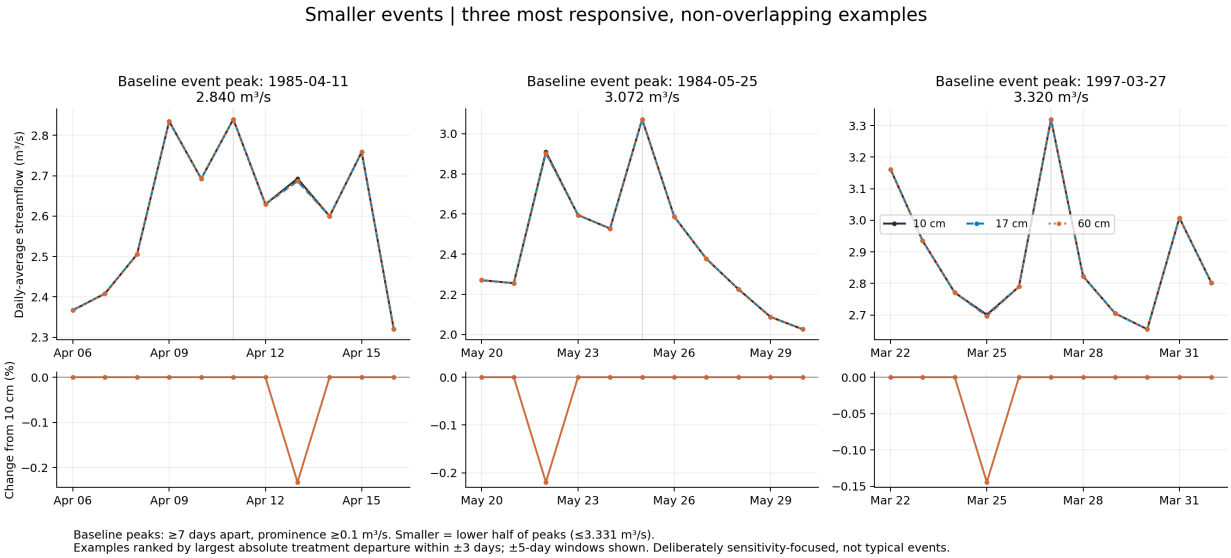


Figure 6: Sensitivity-focused smaller totalwatsed events under the candidate. The daily-yield peak cutoff is 3.3311 m^3/s. Selection ranks departures within three days of eligible peaks and displays five days on either side; it differs from the outlet-peak selection used in Figure 8.

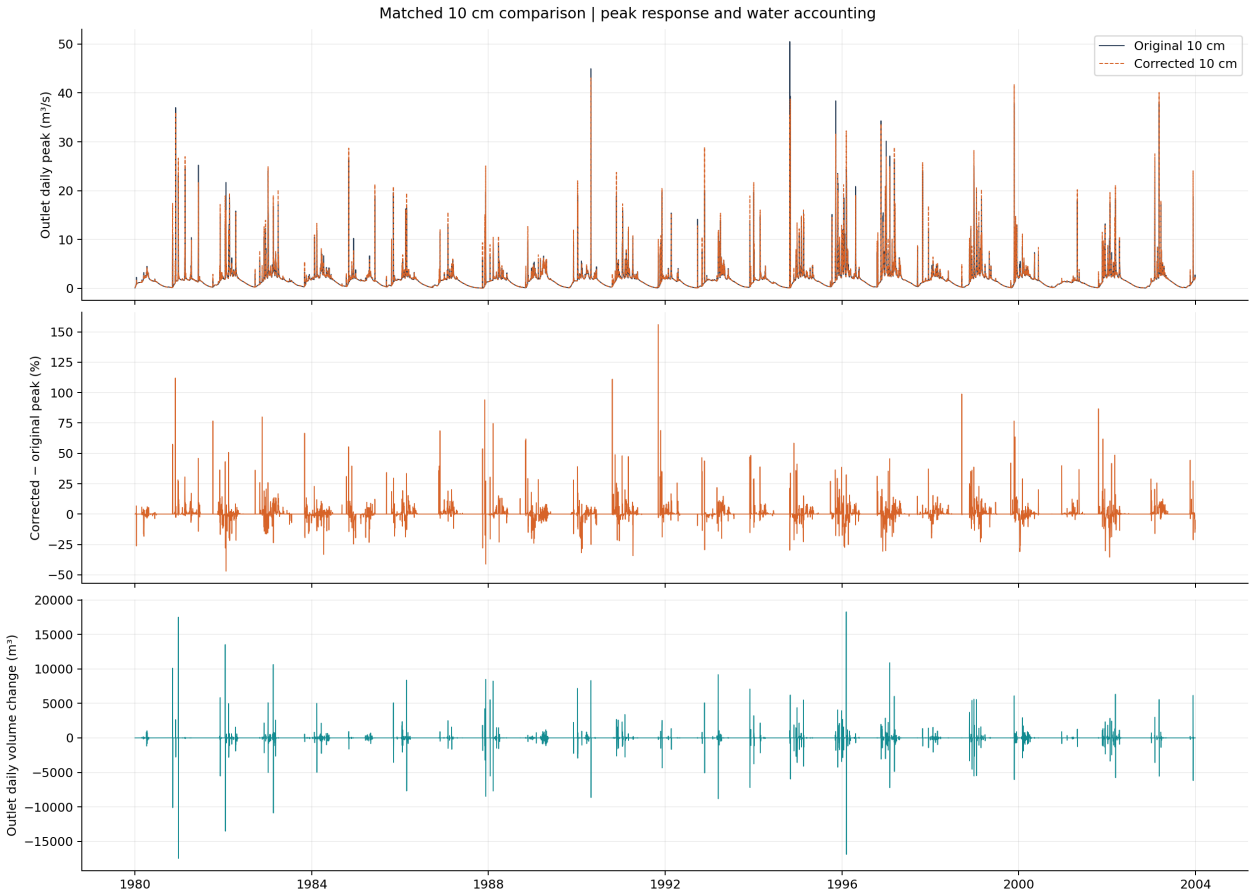


Figure 7: Matched 10 cm original-versus-candidate outlet response. Near-equal long-term yield does not imply unchanged daily volumes, peak rates or timing.

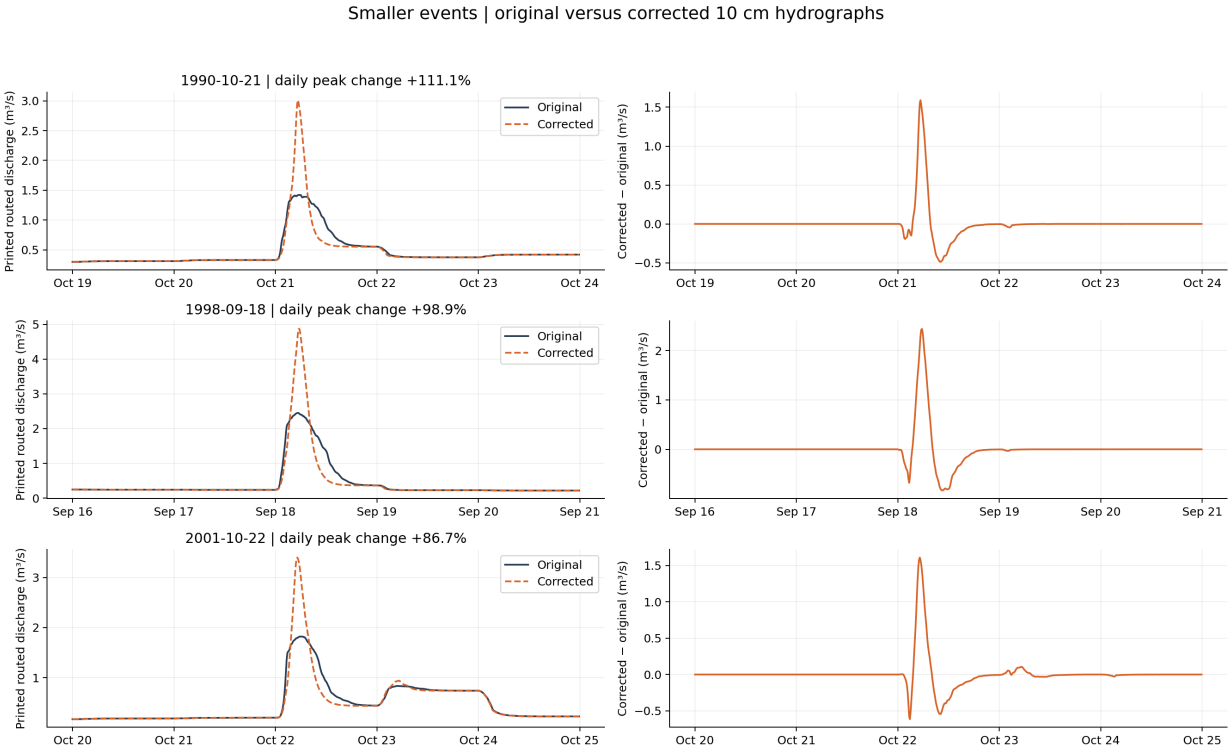


Figure 8: Selected smaller-event hydrographs at matched 10 cm inputs. The candidate can increase peaks. Selection is deliberately sensitivity-focused; the unresolved hydrograph-volume discrepancy also applies to interpretation.

3.7 Unresolved hydrograph-volume gate

Agreement between the event report and channel water-balance outflow is exact at printed precision, but these are ledger measures. In `WSHCHR`, `chvol` derives from inflow plus change in storage, separately from the sampled routed-discharge array `q1`. Integrating the printed 600-second hydrograph exposes a distinct inconsistency:

Table 10: Matched 10 cm hydrograph-integral shortfall relative to the outlet-volume ledger. Printing bounds assume rounding to three significant digits.

Build	Shortfall (million m^3)	Fraction of ledger	Rounding bound (million m^3)
Original	5.889	0.4628%	2.879
Candidate	16.810	1.3212%	2.879

Both shortfalls exceed the conservative cumulative discharge-printing bound. The candidate increases the discrepancy by about 10.92 million m^3. The five-day October 25–29, 1994 window is more concerning: ledger volumes are 2,539,873.62 and 2,539,873.33 m^3, but the original and candidate curves integrate to 2,518,796.22 and 2,075,129.58 m^3. The deficit increases from 0.83% to 18.30%. The lower candidate peak on October 27 therefore cannot yet be counted as a validated hydrograph improvement. Other selected events have smaller deficits or a small surplus. Curves have not been normalised to force closure.

This is an observed routed-output/ledger inconsistency, not proof that the soil-water balance lost that volume. Its mechanism requires tracing discharge, inflow, storage and termination bookkeeping. The shared H670 events on November 20–21, 1985 also retain positive runoff (22.775 and 19.581 m^3) with zero reported peak in both builds; those are pre-existing anomalies, not newly introduced regressions.

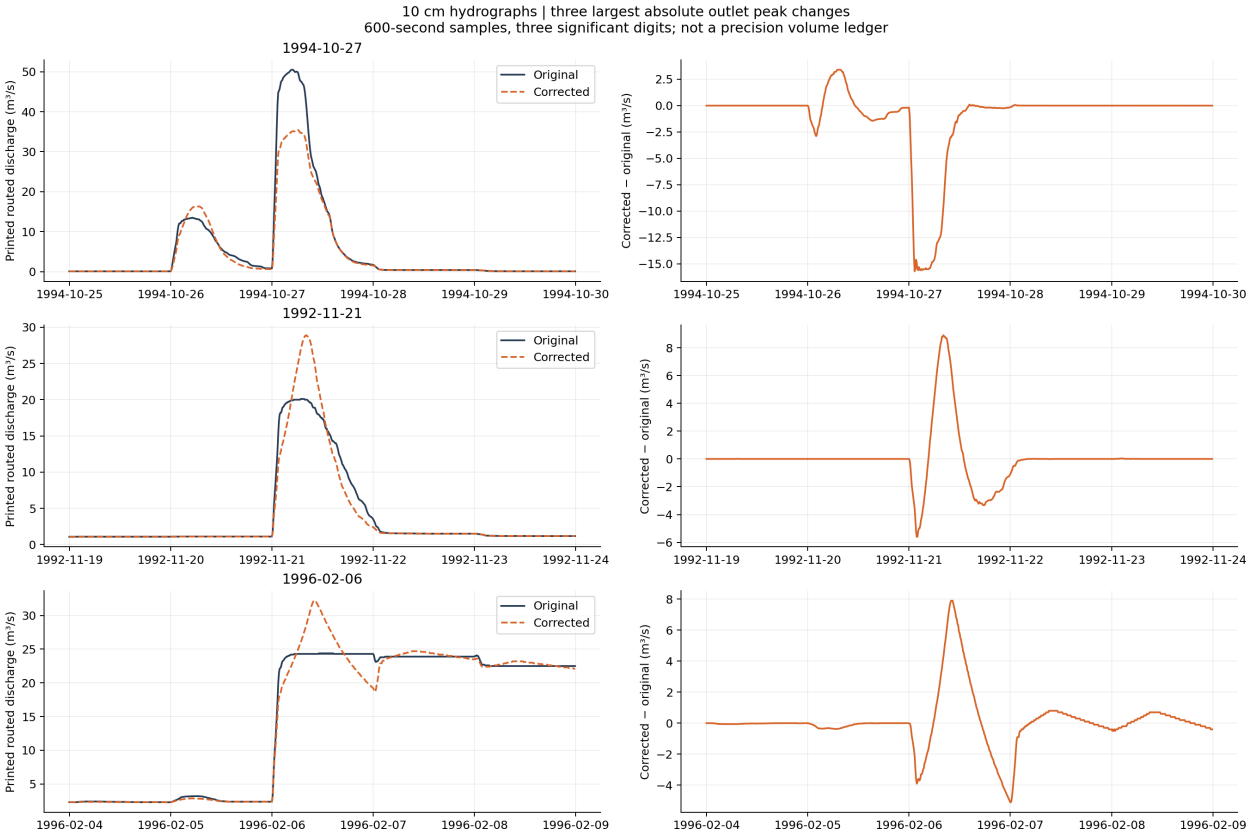


Figure 9: Three largest absolute matched outlet-peak changes, separated by more than seven days. Selection retains both increases and decreases. The October 1994 window has an 18.30% candidate hydrograph-volume deficit despite nearly unchanged ledger volume; lower peaks alone do not validate the repair.

Disposition and stakeholder disclosure. Retain physically justified roughness rather than inflate it to suppress a numerical symptom. The candidate is implemented and has completed these research comparisons, including input isolation, negative controls and output readback. It is not a full kinematic-wave repair, does not remove every mutation sensitivity, and has not passed an unconditional watershed-hydrograph release gate. Reconcile the sampled-discharge/ledger discrepancy and repeat the affected acceptance checks before declaring it validated for that use. Any stakeholder summary should disclose both the mutation improvement and the adverse closure finding, distinguish daily yield from instantaneous peak, and state deployment status separately. This report update neither changes production binaries nor sends an external stakeholder message.

4 Original August Census Study Design

4.1 Scenario strata and experimental unit

The study used two frozen Topanga strata:

- **Burned**: the spatially resolved burned base scenario.
- **Unburned**: the spatially uniform undisturbed Omni scenario.

Each stratum contains 140 hillslopes. Burned and unburned are descriptive strata, not the causal contrast in this experiment. The causal comparison is within a stratum: a hillslope's baseline full-history run versus the same hillslope and history with exactly one controlled mutation.

The experimental unit is a scenario–hillslope–mutation combination. This gives 280 scenario–hillslope combinations before mutation direction is added. No mutation outcome was inspected when the population and mutation plan were frozen.

4.2 WEPPcloud source run and climate provenance

Both strata were derived from the archived WEPPcloud run `hand-to-mouth-drought`, originally hosted at https://wepp.cloud/weppcloud/runs/hand-to-mouth-drought/cfg and copied to the local run archive on August 7, 2026. The burned authority is the source run's `wepp` tree. The unburned authority is its Omni `scenarios/undisturbed/wepp` tree. The trial plan froze those two input trees by SHA-256 before any census execution. Thus, "burned" and "unburned" refer to two preserved WEPPcloud scenario products, not to inputs reconstructed for this report.

The climate is the WEPPcloud `observed_gridmet` product for calendar years 1980–2024, giving 45 complete simulation years. Climate mode was `GridMetPRISM` with spatial mode `Multiple`: WEPPcloud generated a separate climate file for each spatial element rather than applying one basin-wide file. Daily precipitation and maximum and minimum air temperature came from GRIDMET. WEPPcloud used CLIGEN to construct the breakpoint climate sequence, then populated solar radiation, dew-point temperature, wind speed, and wind direction from GRIDMET. The CLIGEN parameter anchor was station `ca041484`, Canoga Park Pierce, California (34.18°N, 118.57°W; 240 m in the generated climate-file header), using the 2015 station database and CLIGEN 5.323. No precipitation scaling or maximum 30-minute-intensity adjustment was enabled. Although the generated climate header retains a nominal "100 years simulated" CLIGEN field, its dated records end in 2024 and the WEPP management and event outputs contain 45 simulation years; 1980–2024 is the forcing period used here.

Precipitation has a strongly Mediterranean seasonal distribution. Calculated directly from the 140 frozen hillslope climate files, the unweighted hillslope-file mean is 607 mm/year. The mean

monthly totals for January through December are 145, 146, 100, 24, 10, 2, 1, 3, 5, 25, 47, and 98 mm, respectively. October through March supplies 561 mm, or 92.4% of the annual mean. The 45-year mean varies spatially among hillslope files from 485 to 688 mm/year. Interannual variability is larger: after averaging each year across the 140 hillslope files, annual precipitation ranges from 150 mm in 2013 to 1,505 mm in 1983, with a median of 496 mm. These are descriptive summaries of the model forcing, not an area-weighted watershed precipitation product. They show that the full-history probes include very dry and very wet years and retain the antecedent soil-water histories produced by that sequence.

The Omni unburned scenario links to the parent run's climate directory and climate state. Corresponding burned and unburned hillslope climate files are therefore byte-identical; scenario differences arise from soil and management parameterisation, not different meteorological histories. Table 11 summarises the WEPPcloud settings most relevant to interpretation.

Table 11: Frozen WEPPcloud source parameters shared by the two scenario strata unless otherwise noted.

Parameter	Frozen value
WEPPcloud configuration	`disturbed9002.cfg`
Climate catalog and period	Observed GRIDMET, 1980–2024 (45 years)
Climate spatial mode	Multiple; element-specific climate files
CLIGEN parameter station	`ca041484`, Canoga Park Pierce, CA
Precipitation scaling	None
DEM and model cell size	NED 1 arc-second (2024); 30 m
TOPAZ delineation controls	CSA 4 ha; minimum channel length 60 m
Initial soil saturation	0.75
PMET crop coefficient	Enabled; $K_{cb} = 1.20$, `rawp`=0.80
Restrictive-layer conductivity	`kslast`=0.0 (disabled)
WEPP options	snow and baseflow on; frost and TCR off
Hillslope representation	one OFE per hillslope

These are provenance descriptors, not parameters varied by the census. Each mutation retained the target scenario's climate file, slope, management, remaining soil properties, initial condition, and complete 45-year antecedent history. Only first-horizon K_{sat} or the paired cover values changed.

4.3 Land-use and soil provenance

WEPPcloud built both strata on the same 30-m watershed discretisation and the same underlying SSURGO/gNATSGO 2025 soil geography. The dominant undisturbed land-use assignment is shrub/scrub on 131 of 140 hillslopes, with mixed forest on two and developed moderate-intensity management on seven. By mapped area, the unburned Omni scenario is 92.67% shrub/scrub, 0.38% mixed forest, and 6.95% developed. Its standard management cover values are 0.70 canopy and 0.90 interrill/rill ground cover for shrub, 0.55 canopy and 1.00 ground cover for mixed forest, and 0.30 canopy and 1.00 ground cover for developed land.

The burned stratum was generated from the spatial burn-severity raster `palisades_sbs_20250115.tif`. Dominant hillslope management comprises 63 low-severity shrub hillslopes, 64 moderate-severity shrub hillslopes, two high-severity shrub hillslopes, two undisturbed shrub hillslopes, two moderate-severity forest hillslopes, and seven developed hillslopes. The main burned management classes use canopy/interrill/rill cover values of 0.33/0.80/0.80 for low-severity shrub,

0.27/0.55/0.55 for moderate-severity shrub, and 0.05/0.30/0.30 for high-severity shrub. These are frozen WEPPcloud management inputs; the census cover probe subsequently changes only the paired interrill and rill values by ±0.01 within one target hillslope.

The soil inputs derive primarily from six mapped soil components (map-unit keys 469823, 469827, 469952, 469954, 469992, and 470003), plus the developed soil template. Dominant surface textures are loam on 93 hillslopes and sandy loam on 40; the remaining seven use the developed template. In the unburned stratum, WEPPcloud produced eight dominant soil–land-use variants: 133 shrub or forest hillslopes retain their undisturbed soil treatments and seven use the developed template. In the burned stratum, the same soil geography expands to 15 dominant variants because WEPPcloud applies land-use- and severity-specific erosion and vegetation parameters: 63 low-severity, 66 moderate-severity, two high-severity, two undisturbed, and seven developed hillslope assignments.

The source SSURGO horizon geometry and texture remain the basis of each soil file, while the disturbed-9002 builder supplies scenario-specific erodibility, effective conductivity, rooting, leaf-area, and cover-related parameters. All soil files were initialised at 0.75 saturation. The source run disabled the artificial terminal restrictive-layer conductivity (`kslast`=0.0). The present experiment does not compare alternative soil maps or burn-severity products: it holds each frozen soil file fixed except for the explicit first-horizon K_{sat} probe.

4.4 Small-mutation matrix

Each scenario–hillslope combination requested four trials:

Table 12: Frozen Topanga small-mutation matrix.

Family	Minus probe	Plus probe
First-horizon K_{sat}	baseline ×0.99	baseline ×1.01
Paired ground cover	`inrcov`, `rilcov` −0.01	`inrcov`, `rilcov` +0.01

The cover probe changes interrill and rill cover together. It is an absolute one-percentage-point change, not a one-percent relative change. A cover pair was excluded before execution if either direction would require clipping. This symmetry requirement excluded 32 requested cover trials. No K_{sat} trial was excluded.

These perturbations are intentionally small. They are not proposed alternative calibration values. A smooth model response should change locally and proportionally under such probes, except where a genuine physical threshold or an implementation branch intervenes.

4.5 Execution and provenance controls

The study used a diagnostic build of the WEPP hillslope model developed from the pinned WEPP-Forest source at commit `ea25ad79`. This is the ordinary model with observational logging added at the existing peak-flow calculation; it does not replace either peak solver or change the runoff calculation. When enabled by a marker file, the build writes a record for every peak-solver call, including the event identifiers, runoff and surface-return depths, the duration used to distribute returned water, the rate added to the subdaily forcing, forcing mode, selected solver, and calculated peak. The observer was compiled with GNU Fortran 13.3.0, and the executable used for the census has SHA-256 `2ec15778df957f...c0731f`. Before the census, both Hill 106 K_{sat} lanes were run with diagnostic logging active and inactive. All seven ordinary WEPP output files were byte-for-byte

identical within each lane. Standalone replay of the recorded solver inputs also reproduced the selected production peaks exactly. These tests establish that the diagnostic build observes the legacy calculation without changing its reported results.

Every eligible trial executed the complete antecedent history for the target hillslope. The execution system validated the frozen plan and selection, executable and observer hashes, input snapshot, sole changed input, requested and realised mutation values, terminal identity, return code, trace hash, and hillslope-pass hash. Each trial received an explicit complete, failed, or stopped disposition.

The final population used plan identifier `b575fde4a28cf85f...617080de` and selection identifier `2e450acbaa131c...e3e3d`. All 1,088 selected trials completed; none failed or stopped. Repeated aggregation produced identical ledger hashes.

4.6 Event pairing

For each mutation trial, every recorded event in the baseline run was matched to the corresponding event in the mutant run using scenario, hillslope, model year, model day, overland flow element (OFE), and the event's order among peak-solver calls on that day. The combined record keeps every event observed in either run. This distinguishes three cases:

1. an event is present in both lanes;
2. an event is present only in the baseline; or
3. an event is present only in the mutant.

If no corresponding event exists in one run, that side is recorded as absent rather than assigned a peak or runoff value of zero. This matters because the appearance or disappearance of an event is itself a candidate response, and a made-up zero would create an artificial ratio.

4.7 Preregistered screening rules

The candidate screen used fixed SI-unit floors before evaluating ratios:

- runoff denominator floor: 10^{-5} m (0.01 mm);
- peak magnitude and difference floor: 10^{-7} m/s (0.36 mm/h); and
- surface-return-rate denominator floor: 10^{-8} m/s (0.036 mm/h).

An event row became a candidate if any of the following was true:

1. event presence changed between baseline and mutant;
2. the selected peak solver changed;
3. peak changed by more than 25% while runoff changed by less than 5%;
4. the mutant-to-baseline peak ratio was at least 2 or at most 0.5;
5. the imposed surface-return rate changed by more than twofold; or
6. peak moved opposite the ordinary inverse response expected for the mutation direction.

The screen is intentionally inclusive. A candidate flag is evidence for review, not a declaration that an implementation defect has been confirmed. Confirmation requires local bracketing, frozen-event replay, and mechanism tracing.

4.8 Descriptive magnitude and direction analysis

The magnitude scatter uses paired-positive rows satisfying baseline peak $\geq 10^{-7}$ m/s and mutant peak > 0. Peaks are converted to mm/h. "Congruent" and "incongruent" describe whether the sign of the modelled peak response agrees with the ordinary first-order expectation for surface runoff; they do not describe whether the size of that response is reasonable.

Table 13: Complete directional logic used for congruent and incongruent labels.

Mutation	Congruent peak response	Incongruent peak response
K_{sat} +1%	mutant peak lower	mutant peak higher
K_{sat} −1%	mutant peak higher	mutant peak lower
Cover +0.01	mutant peak lower	mutant peak higher
Cover −0.01	mutant peak higher	mutant peak lower

Thus, both parameter families use an inverse expectation: increasing either K_{sat} or cover should lower surface-runoff peak flow, while decreasing either parameter should raise it. An exactly unchanged peak provides no directional support for the expected response and is placed in the incongruent class for the opacity display and counts. Only eight of the 199,086 plotted rows are exact ties, so this convention has no material effect on the reported proportions. The classification is descriptive and independent of the candidate screen.

The figure separates unburned and burned rows onto left and right axes with identical scales. Green and orange identify positive surface return in the unburned and burned panels, respectively; gray identifies pairs with no surface return in either run. Hollow triangle orientation shows the two-by-two mutation matrix, and opacity shows directional agreement. Faint markers ($\alpha = 0.2$) are congruent; dark markers ($\alpha = 0.6$) are incongruent. Reference lines mark ratios of 1, 0.5 and 2, and 0.2 and 5.

5 Original August Topanga Results

5.1 Complete execution and event inventory

Table 14: Requested and executed mutation trials.

Family	Requested	Executed	Excluded before outcomes
K_{sat}	560	560	0
Cover	560	528	32
Total	1,120	1,088	32

The combined event record contains 225,654 rows: 225,036 matched baseline–mutant pairs, 320 baseline-only events, and 298 mutant-only events. The 618 events observed in only one run are retained as event-presence changes rather than converted to numerical zero.

5.2 Candidate prevalence

The screen identified 11,506 candidate event rows across 1,027 eligible trials. Table 15 separates two estimands that should not be confused. Event-row prevalence describes how often a recorded

solver-call event is flagged. Trial prevalence describes whether a full-history mutation trial contains at least one flagged event. Because a trial contains many events, the trial metric saturates easily and does not measure anomaly severity.

Table 15: Preregistered candidate-screen prevalence.

Population	Candidate events	Event rows	Candidate trials	Trials
Overall	11,506 (5.10%)	225,654	1,027 (94.4%)	1,088
Burned	6,928 (5.42%)	127,750	516 (94.5%)	546
Unburned	4,578 (4.68%)	97,904	511 (94.3%)	542
K_{sat}	10,055 (8.62%)	116,651	560 (100%)	560
Cover	1,451 (1.33%)	109,003	467 (88.4%)	528

The family contrast is substantial. K_{sat} produces about seven times the candidate event-row rate of cover, and every eligible K_{sat} trial contains at least one candidate. This does not mean every K_{sat} trial is defective; it means the inclusive screen is highly sensitive to K_{sat} somewhere in each full history.

5.3 Event runoff remains tightly constrained

Figure 10 compares mutant and baseline event runoff depth using the same scenario colours, mutation markers, ratio guides, and directional opacity rules as the peak-flow figure. The runoff population is more tightly concentrated around 1:1 over the full plotted range.

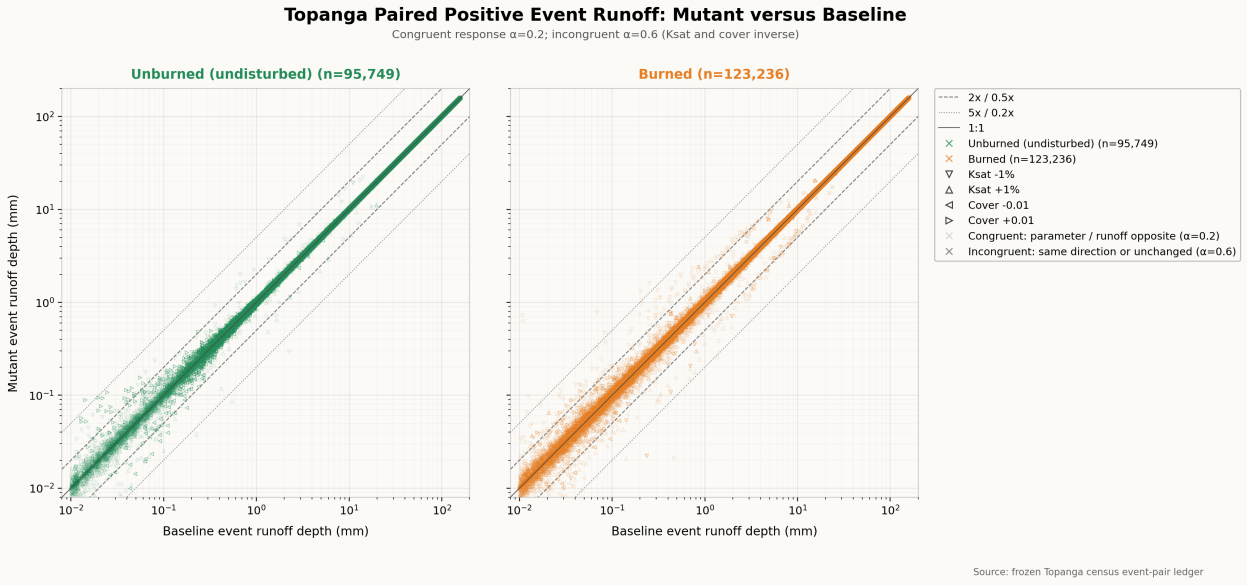


Figure 10: Mutant versus baseline event runoff depth for 218,985 paired-positive Topanga event rows. The population requires baseline runoff of at least 0.01 mm and positive mutant runoff. Unburned and burned strata occupy the left and right axes, respectively; triangle orientation identifies mutation family and direction. Faint markers are directionally congruent and dark markers are incongruent.

Only 451 runoff rows (0.21%) lie outside the 0.5–2 ratio band, and only 36 (0.016%) lie outside the 0.2–5 band. K_{sat} accounts for 366 of the 451 twofold-tail rows (81.2%) and 26 of the 36

fivefold-tail rows (72.2%). These tails exist, but they are both absolutely and proportionally smaller than the corresponding peak-flow tails.

The separated panels confirm stable runoff in both strata. The twofold-tail rate is 0.127% for unburned rows and 0.267% for burned rows; fivefold departures remain rare at 0.014% and 0.019%, respectively.

Runoff direction is also strongly coherent for K_{sat}. Overall, 146,131 of 218,985 runoff rows (66.7%) are congruent. The K_{sat} share is 111,359 of 113,242 rows (98.3%), compared with 34,772 of 105,743 cover rows (32.9%). Exactly unchanged runoff is assigned to the incongruent display class; only 110 rows are exact ties.

This volume–peak contrast is central to the diagnosis. Under the same small K_{sat} probes, total event runoff is overwhelmingly stable and moves in the ordinary hydrologic direction, while peak-flow magnitude contains the larger, more structured extreme population. The evidence therefore points toward the construction of subdaily runoff timing and peak rather than a comparable instability in total event runoff depth.

5.4 Event sediment delivery is mostly controlled but noisy

Figure 11 compares mutant and baseline WEPP event sediment delivery (`Sed.Del`). This quantity is sediment delivered at the hillslope outlet per unit width, not gross detachment across the hillslope. The figure pairs records by scenario, hillslope, mutation, calendar year, and day of year using the frozen Phase 2A baseline and census mutant event files.

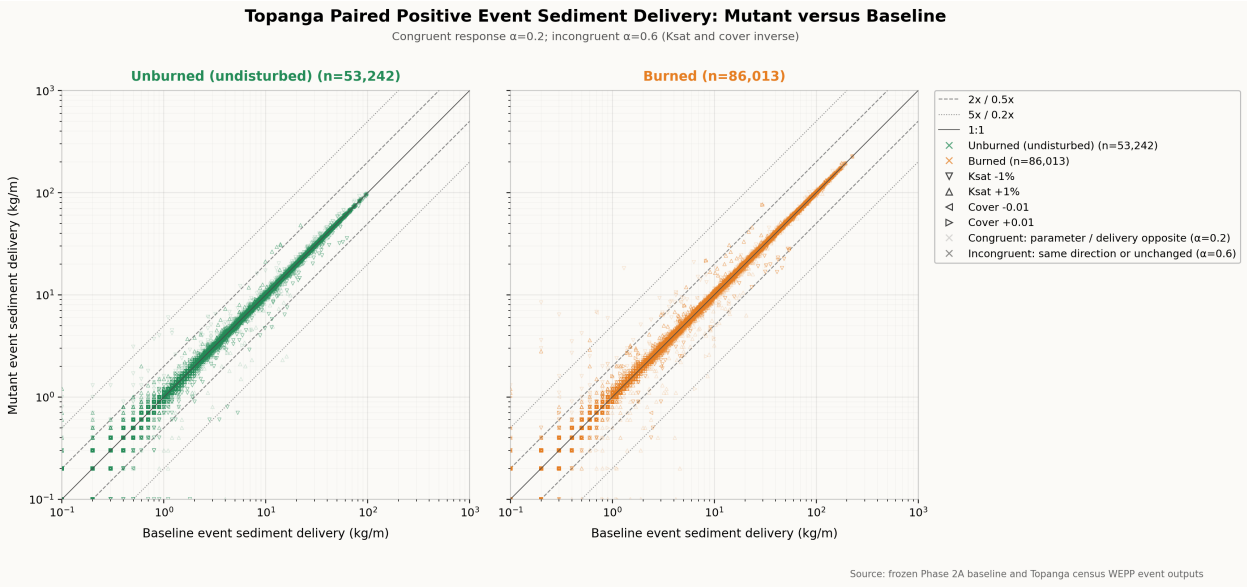


Figure 11: Mutant versus baseline event sediment delivery for 139,255 paired-positive Topanga event rows. Both axes use kg/m. Unburned and burned strata occupy the left and right axes, respectively; triangle orientation identifies mutation family and direction. Faint markers are directionally congruent and dark markers are incongruent. The stepped pattern at small magnitudes reflects the 0.1 kg/m precision of the retained WEPP event output.

Only 271 sediment-delivery rows (0.195%) lie outside the 0.5–2 ratio band, and only 55 (0.039%) lie outside the 0.2–5 band. K_{sat} accounts for 245 of the 271 twofold-tail rows (90.4%) and 49 of the 55 fivefold-tail rows (89.1%). The erosion response is therefore mostly controlled, and its material

tail is small and again concentrated in K_{sat} mutations.

Twofold-tail prevalence is nearly identical after separating the strata: 0.203% for unburned rows and 0.190% for burned rows. Neither stratum's material sediment-delivery tail was hidden by the other.

The central band should not be mistaken for precision. WEPP writes `Sed.Del` to these retained files at 0.1 kg/m precision, producing 76,398 exact printed ties (54.9% of paired-positive rows) and conspicuous steps at small delivery. Relative differences near zero are consequently noisy and less material in absolute terms. At moderate and large delivery, the population remains concentrated near 1:1, but visible dispersion persists within the 0.5–2 band. Event sediment-delivery values should therefore be interpreted with an empirical factor-of-two sensitivity envelope in mind. This envelope is descriptive of the Topanga small-mutation experiment; it is not a formal statistical confidence interval or a universal WEPP uncertainty bound.

Before the positive-delivery filter, event-date pairing produced 225,654 rows: 139,255 were positive in both runs, 505 were positive only in the baseline, 516 were positive only in the mutant, and 85,378 were positive in neither. The one-sided and zero-delivery rows remain in the accounting but cannot be displayed on logarithmic axes.

5.5 Peak-flow magnitude distribution

Figure 12 shows the 199,086 paired-positive event rows. The dense diagonal is the ordinary response population. Most events remain close to baseline despite the high inclusive trial-screen prevalence. A sparse set of points forms large upper and lower departures.

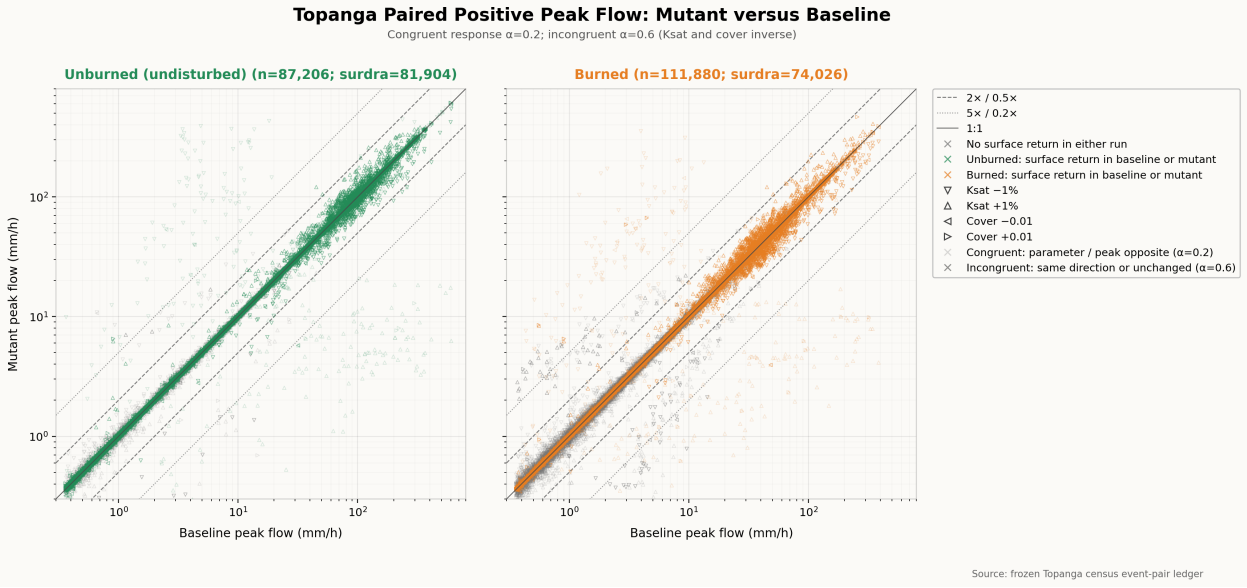


Figure 12: Mutant versus baseline hillslope peak flow for paired-positive Topanga event rows. Unburned and burned strata occupy the left and right axes. Green unburned and orange burned markers identify pairs with positive surface return in the baseline or mutant run; gray markers have no surface return in either run. Triangle orientation identifies mutation family and direction. Opacity identifies agreement with the ordinary inverse hydrologic response. The thin 1:1 line is drawn above the markers; dashed and dotted lines mark twofold and fivefold departures.

Surface return occurs in 81,904 of 87,206 unburned plotted rows and 74,026 of 111,880 burned

rows. It is present in 567 of the 989 twofold-tail rows and 420 of the 604 fivefold-tail rows. Surface return therefore accompanies much of the anomalous population, but it does not explain the entire tail: 422 twofold-tail rows and 184 fivefold-tail rows have no positive surface return in either paired run. Most of that no-return tail is in the burned stratum. The partition supports the documented surface-return timing mechanism for a large part of the population while showing that additional peak-response mechanisms must also be present.

Only 989 rows (0.50% of the plotted population, approximately one in 201) lie outside the 0.5–2 band, and 604 (0.30%, approximately one in 330) lie outside the 0.2–5 band. These tails are small but strongly structured:

Table 16: Composition of the peak-ratio tails.

Ratio region	Rows	K_{sat} rows	K_{sat} share	Congruent share
Outside 0.5–2	989	907	91.7%	79.0%
Outside 0.2–5	604	555	91.9%	83.8%

The dominant extreme response is therefore not a generic reaction to any small input change. It is overwhelmingly associated with first-horizon K_{sat}. Most of these extreme K_{sat} responses have the expected sign. The surprising feature is their magnitude under a one-percent perturbation.

The separated panels show that the extreme branches occur in both strata. The twofold-tail rate is 0.393% for unburned rows and 0.577% for burned rows; the fivefold-tail rates are 0.280% and 0.322%, respectively. Burned conditions have the higher tail prevalence, but the unburned extreme population is not an overplotting artefact.

The split-stratum figure also locates the K_{sat} response clusters at different absolute peak magnitudes. Among directionally incongruent K_{sat} rows inside the 0.5–2 band, the median baseline peak is 74 mm/h in the unburned stratum (interquartile range 50–103 mm/h) and 39 mm/h in the burned stratum (26–51 mm/h). A separate magnitude criterion gives the same ordering: among rows departing from baseline by more than 25%, the median baseline peak is 37 mm/h unburned and 10 mm/h burned. This displacement is not a climate difference because corresponding burned and unburned hillslopes use identical meteorological histories.

Pairing the same hillslope, mutation, year, day, OFE, and solver-call order across strata further describes the shift. Of matched central-band K_{sat} rows, 1,164 are incongruent in both strata, 2,998 only in the unburned stratum, and 7,204 only in the burned stratum. For the unburned-only group, median baseline peaks are 72 mm/h unburned and 44 mm/h burned. For the burned-only group, they are 38 mm/h burned and 9 mm/h unburned. The response regime is not tied to one universal peak threshold. Soil, vegetation, cover, and burn-severity parameterisation shift the modelled hydrologic state at which the unstable peak response is expressed. These are descriptive event-row statistics; they do not identify the responsible state variable or solver boundary.

5.6 Directional agreement

Across all plotted paired-positive rows, 64.1% are directionally congruent. The family-specific result is more informative: 87.4% of K_{sat} rows are congruent, while only 39.2% of cover rows are congruent.

Table 17: Directional classification of paired-positive event rows.

Population	Congruent	Incongruent	Congruent share
Overall	127,665	71,421	64.1%
K_{sat}	90,033	13,012	87.4%
Cover	37,632	58,409	39.2%
Burned	79,336	32,544	70.9%
Unburned	48,329	38,877	55.4%

This correction is important for interpretation. The K_{sat} tail does not primarily show higher conductivity producing a higher surface-runoff peak. Rather, small conductivity changes usually move peak flow in the expected direction but sometimes produce an implausibly large change. Directional congruence does not validate the magnitude.

5.7 Incongruence increases with baseline peak within the central band

The opacity pattern in Figure 12 suggests more dark markers at higher baseline peaks even within the 0.5–2 band. Direct tabulation confirms a real conditional trend, especially for K_{sat} (Table 18). The overall trend is not perfectly monotonic, but the high-flow bins are more often incongruent than the low-flow bins.

Table 18: Incongruent share within the 0.5–2 peak-ratio band, by baseline peak magnitude. The final bin contains only 145 rows and should be treated cautiously.

Baseline peak (mm/h)	All families	K_{sat}	Cover
0.3–1	29.7%	1.2%	60.6%
1–3	34.7%	0.8%	72.1%
3–10	35.8%	1.7%	73.8%
10–30	31.6%	15.1%	48.3%
30–100	39.8%	25.4%	54.7%
100–300	43.6%	30.2%	57.4%
300–800	66.9%	52.9%	79.2%

Below 10 mm/h, K_{sat} responses within the central band are almost always directionally congruent. The incongruent share rises to 15.1% at 10–30, 25.4% at 30–100, and 30.2% at 100–300 mm/h. This is not explained by cover, which is frequently incongruent throughout the range and lacks the same clean onset. It suggests that the local response regime changes with event magnitude or with hydrologic states correlated with larger peaks.

The binning is descriptive. Event rows are clustered within trials and years, and no independence-based significance test is claimed. Candidate mechanism work should test whether this gradient aligns with surface-return depth, positive-excess duration, forcing-mode changes, solver changes, or antecedent profile state.

5.8 Excluding mixed runoff sharply reduces directional incongruence

Figure 13 repeats Figure 12 after excluding pairs in which either the baseline or mutant uses forcing mode 1: surface return is added to intervals already having positive infiltration excess. In the census observer source, this branch requires realised surface return (`surpls`) greater than 10^{-6} m

and positive pre-addition rainfall-excess duration (`durre`). The filter therefore isolates the mixed Green–Ampt runoff and surface-return pathway implicated by the Hill 106 trace. It retains events without surface return and events with surface return but no positive-excess duration; it is not a filter for all positive `surdra`. The peak floor, pairing rules, axes, markers, and opacity convention are unchanged. Of 199,086 plotted rows, 115,932 remain: 56,356 unburned and 59,576 burned.

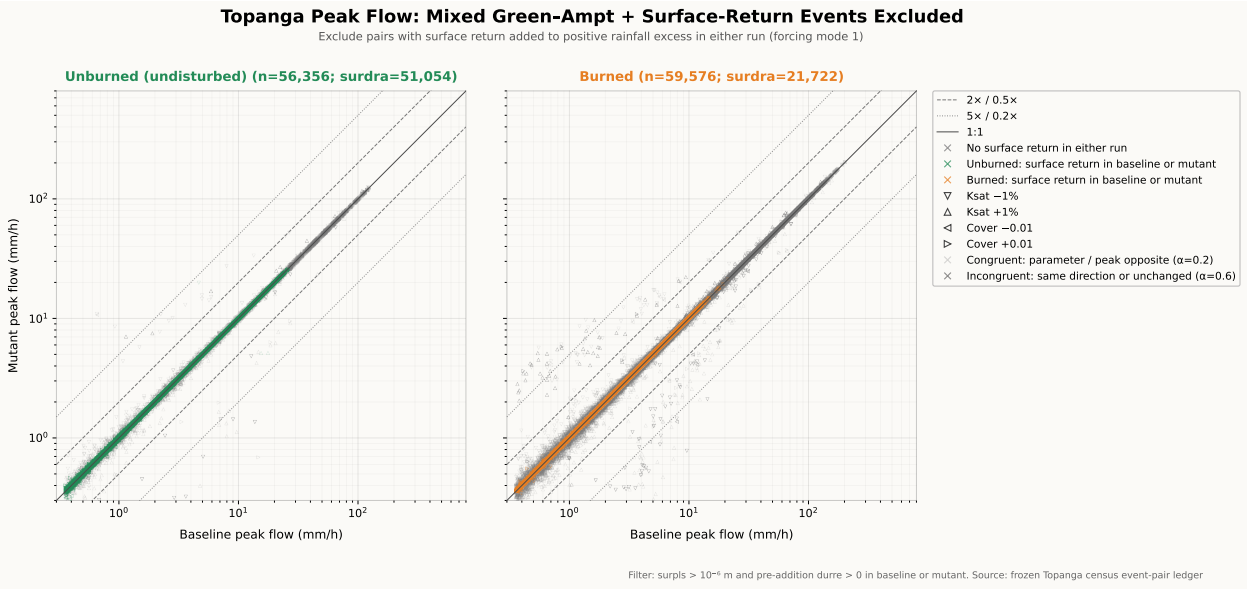


Figure 13: Mutant versus baseline peak flow after excluding mixed infiltration-excess and surface-return forcing in either paired run. Green and orange markers remain because surface-return-only events are retained. Faint markers are directionally congruent; dark markers are incongruent. The filtered population is much more tightly aligned with the expected conductivity response, including at higher peak intensities.

The principal result is the sharp fall in the incongruent share of K_{sat} responses within the central 0.5–2 peak-ratio band (Table 19). At baseline peaks of 30–100 mm/h, the share falls from 24.1% to 1.8% in the unburned stratum and from 26.0% to 1.9% in the burned stratum. At 10–30 mm/h, it falls from 7.4% to 0.7% unburned and from 18.5% to 2.7% burned. Thus, the high-intensity directional instability in the full census is concentrated in the mixed-forcing population. Among retained rows, the steep rise in wrong-direction conductivity responses is largely absent.

Table 19: Incongruent share of central-band K_{sat} rows before and after mixed-forcing exclusion. Bins include their lower bound and exclude their upper bound; n is the number of event rows in each denominator.

Stratum	Baseline peak (mm/h)	Original n	Incongruent	Retained n	Incongruent
Unburned	10–30	5,540	7.4%	3,740	0.7%
	30–100	10,399	24.1%	652	1.8%
	100–300	3,254	33.4%	42	2.4%
Burned	10–30	12,596	18.5%	5,831	2.7%
	30–100	21,149	26.0%	4,199	1.9%
	100–300	1,466	23.1%	368	1.4%

This pattern supports the report's timing-based interpretation. Conductivity changes affect both infiltration excess and profile return. In the mixed branch, a change in positive-excess duration also changes the rate imposed by dividing daily return depth by that duration. The filtered comparison locates most high-intensity directional incongruence in the population exposed to that coupling. It does not establish that duration compression causes every excluded reversal; antecedent state and solver selection also vary within this population and require frozen-event replay.

The extreme-ratio tail also contracts, but remains distinct from directional incongruence. Exclusion removes 41.8% of plotted rows and 69.5% of rows outside the 0.2–5 band. The retained population contains 425 rows outside 0.5–2 (0.367%) and 184 outside 0.2–5 (0.159%). All 184 remaining fivefold departures have no positive surface return in either run; 144 are burned and 169 are K_{sat} mutations. The burned twofold-tail rate is nearly unchanged (0.577% before versus 0.574% after filtering). Mixed-forcing exclusion therefore explains the disappearance of much of the directional pattern without resolving the separate no-return extreme population.

This is a descriptive exclusion analysis, not a corrected simulation. No retained peak changes value. Larger events are disproportionately removed, and only 42 unburned central-band K_{sat} rows remain at 100–300 mm/h; none remain at 300–800 mm/h in either stratum. The within-bin comparisons reduce, but do not eliminate, differences in population composition. Event rows remain clustered within trials and years. A tighter diagonal demonstrates local response stability in the retained population, not agreement with observed peaks or proof that excluded events are physically invalid.

6 Interpretation of the Original Census

6.1 What the Topanga census establishes

The result has six levels:

1. **Event runoff is stable.** The runoff scatter remains tightly concentrated around 1:1, and 98.3% of K_{sat} runoff responses have the expected direction.

2. **Peak response separates from volume response.** The peak-flow tail is larger and more structured even though the corresponding runoff population remains tightly constrained.

3. **Sediment delivery is mostly controlled but not exact.** Only 0.195% of paired-positive sediment-delivery rows depart by more than twofold, but low-magnitude quantisation and central-band dispersion require a factor-of-two sensitivity envelope for event-level interpretation.

4. **The extreme peak population is parameter-specific.** More than 91% of the twofold and fivefold tails are K_{sat} mutations.

5. **The main K_{sat} concern is magnitude, not sign.** Most extreme K_{sat} responses move in the hydrologically expected direction, but a one-percent change sometimes produces a many-fold peak change. Within the central band, wrong-sign K_{sat} responses become more frequent as baseline peak increases.

6. **Parameterisation shifts the response regime.** Under identical climate histories, incongruent K_{sat} responses cluster near substantially higher baseline peaks in the unburned stratum than in the burned stratum. The instability is conditioned on scenario-specific hydrologic state, not on a single absolute peak threshold.

These observations are consistent with the known Hill 106 mechanism. K_{sat} changes infiltration, daily profile surplus, and the duration used to assign that daily volume to a subdaily forcing. Near a duration or forcing-mode boundary, a very small input change can be amplified. A solver branch can add another discontinuity. The census does not prove that this exact sequence explains every tail event, but it shows why K_{sat} is the family in which the extreme population should concentrate.

The between-stratum displacement adds a key finding: parameterisation does not merely change runoff and peak magnitude; it changes when the local peak-response regime is entered. The known surface-return construction gives a plausible implementation pathway because soil and cover affect infiltration, profile surplus, and positive-excess duration. The census establishes the state-dependent displacement, while candidate replay must determine which duration, forcing-mode, or solver boundary explains each cluster.

6.2 What changed after Chapter 4

Chapter 4's statement that surface hydrology should perform best for large events on one OFE described the model and evidence available in 1995. The source history shows that the surface-return pathway now implicated in the Topanga anomalies was not part of that evaluated calculation.

The source history changes how that claim should be judged. In 1993, WEPP calculated excess water above the upper soil-water limit in a variable named `surdra`. The source comments state that the water was not used because runoff had already been routed and erosion had already been calculated. The 1995 documentation was therefore accurate when it said that return flow was not explicitly represented. The surface-return pathway implicated at Topanga was added later. A source comment dates S. Dun's initial `surdra` runoff addback to May 22, 2003, with a modification on January 28, 2004. Related comments date its integration with the peak-flow routine to June 6, 2003, with the no-duration handling extended on July 6, 2004. That code spreads the daily returned-water volume across selected subdaily rainfall-excess intervals before calculating the peak. On November 19, 2009, the call to the daily water balance was moved inside the peak-flow routine; the accompanying comment notes that this differed from WEPP's original sequence and might affect other global variables [6].

Table 20: Source-supported history of the surface-return pathway.

Date	Source evidence and meaning
May 1993	`surdra` calculated but explicitly unused because runoff routing and erosion had already occurred
July 1995	Chapter 4 published with return flow outside the documented formulation
May 22, 2003	Initial `surdra` runoff addback
June 6, 2003	Integration with the peak-flow routine
January 28, 2004	Surface-return addback modified
July 6, 2004	No-duration handling extended
November 19, 2009	Daily water balance moved inside the peak-flow routine
2026	Topanga population-scale mutation census

This history is supported by source-code forensics. The dated comments show what the developers calculated, what they deliberately left unused, and when the later addback and sequencing changes appeared. That evidence is not available from the model output or the 1995 documentation alone. It was possible to identify the Topanga defect because unofficially preserved WEPP source copies could be compared and instrumented. USDA has distributed WEPP primarily as an executable and has not provided users with an official, complete release of the production source. Most WEPP users therefore cannot inspect this calculation, trace a surprising peak through the program, or see that the documented performance claim predates the surface-return code.

Stone et al.'s favorable claim could therefore have been true for the 1995 model they evaluated. This study does not have a 1995 executable and does not test that historical version. What the Topanga census refutes is the continued use of the 1995 claim as assurance about the present implementation. On the current single-OFE hillslopes, the incongruent share of central-band K_{sat} responses increases rather than decreases with baseline peak magnitude, and a one-percent change can produce a many-fold peak change while runoff remains nearly fixed. The large events most relevant to frequency analysis are not the most dependable individual peak estimates in the current model.

The most likely history is straightforward. A later development change added a useful water-balance pathway that the original developers had deliberately left unresolved because of its timing within the daily calculation. That change altered the water supplied to the peak-flow calculation, but the peak method and its performance guidance were not reevaluated at population scale. The Chapter 4 statement remained in the documentation even though the model around it had changed. In that sense, the documentation became stale: it continued to describe the scope and evidence of the 1995 peak calculation, not the behaviour of the later surface-return implementation. Limited access to the source made that mismatch much harder for users to recognise.

The historical omission is understandable. The original developers evaluated the approximation using the theory, selected cases, data, and computing resources available in the early 1990s. A census of hundreds of thousands of paired event responses would have been impractical. The Topanga census applies modern computing to a validation question that appears not to have been asked after the surface-return pathway was added. Its contribution is not to show that Stone et al. were wrong in 1995. It is to show that a later model change invalidated the assumption that their original performance claim still applied.

6.3 Production-scale implication

The small tail fraction does not make an individual modelled peak dependable. The Topanga census found 604 paired-positive event rows outside the 0.2–5 ratio band, an empirical rate of approximately

one in 330. This is an event-row rate within an enriched-discovery site, not an independent-event probability and not a universal WEPP defect rate. Nevertheless, production watersheds can contain thousands of hillslopes evaluated over thousands of events. At that scale, even a sparse local tail provides many opportunities for one hillslope or a small group of hillslopes to contribute an abnormal peak. The risk is not confined to the conspicuous many-fold tail. Within the nominally central 0.5–2 ratio band, directional incongruence increases with baseline peak magnitude, especially for K_{sat}. The K_{sat} incongruent share rises from 1.7% at baseline peaks of 3–10 mm/h to 15.1% at 10–30 mm/h, 25.4% at 30–100 mm/h, 30.2% at 100–300 mm/h, and 52.9% at 300–800 mm/h. The final bin contains only 145 total event rows and is therefore uncertain, but the broader increase begins well below it. Thus, the events most capable of influencing an upper-tail frequency analysis are also the events for which the expected local response becomes less dependable.

This magnitude-conditioned incongruence is distinct from the far-tail result. Most responses outside the 0.5–2 and 0.2–5 bands are directionally congruent; their problem is a disproportionate magnitude under a one-percent probe. Within the central band, the additional concern is that larger events increasingly move in the unexpected direction. Production return-period estimates may therefore be exposed both to rare many-fold peak anomalies and to a broader, peak-dependent instability among events that remain within a factor of two of baseline.

This distinction changes the practical interpretation. The population-level response is mostly sensible, but a particular hillslope-event peak cannot be presumed reliable merely because most rows lie near 1:1. Extreme-value statistics are especially vulnerable because a few unusually large values can alter exceedance ranks, fitted tails, and estimated return periods. A plausible production mechanism is therefore that rare abnormal hillslope peaks propagate into aggregated peak-flow statistics and distort reported return periods. Topanga demonstrates the anomalous hillslope-scale source term; because the census intentionally omitted per-mutation watershed routing, the downstream propagation and return-period effect remain a specific hypothesis requiring routed replay and production-run audit.

6.4 Why larger mutations are not presently needed

The purpose of the mutation is to interrogate continuity near the operational baseline, not to span a plausible calibration range. The $\pm 1\%$ K_{sat} and ± 0.01 cover probes already provide:

- a large paired event population;
- a dense ordinary-response reference population;
- clearly separated tails;
- strong family attribution; and
- known-positive events suitable for bracketing and replay.

Increasing the mutation magnitude would make it harder to distinguish a local implementation discontinuity from an ordinary nonlinear physical response. A larger bracket is useful only after a small probe flags a candidate and the candidate is examined adaptively. For cross-site screening, the same small matrix should remain the default.

6.5 Do not assume the undisturbed result is the known result

When hydrologists compare burned and undisturbed simulations, we usually treat the undisturbed run as the control. If the burned run gives an unexpected answer, we first question the burn-severity

map or the postfire soil and cover values. We rarely begin by asking whether the undisturbed result is the one that is wrong. Topanga shows why that habit can mislead us.

The burned model is, in some ways, the more tightly constrained model. A remotely sensed soil-burn-severity map tells us where the fire effects occurred and how severe they were. The WEPP postfire soil and cover values are based on rainfall-simulation experiments. These observations do not make the burned model perfect, but they give it a spatial and experimental basis.

The undisturbed model does not have an equivalent map of the watershed's hydrologic condition. Before the fire, Topanga contained a complicated mixture of chaparral, shrubs, grass, litter, roots, and vegetation growing in and near channels. These plants intercept rain and remove water from different depths and at different times of year. The undisturbed Omni scenario reduces much of that variation to one general representation. In other words, "undisturbed" is the name of a modelled condition. It does not mean that the condition is known more accurately.

The 2020 water balance gives a reason to question that baseline [7]. With the restrictive layer removed and PMET $K_{cb} = 1.20$, the complete undisturbed watershed produced 149.68 mm of routed outlet flow. The available gauge record corresponds to about 6.2 mm. Those numbers are not directly comparable: the gauge and model boundaries differ, and the gauge includes all sources and losses. The observation is only a magnitude check. Even so, the difference is too large to support the assumption that the undisturbed run is a well-established control. Removing the restrictive layer did not solve the discrepancy. It moved water from lateral flow into deep percolation and groundwater return. Changing K_{cb} also had little effect on the event peaks.

The mutation study points in the same direction. As the computed undisturbed peak flow becomes larger, the response to a very small K_{sat} change becomes more scattered and more often moves in the wrong direction. This pattern occurs at higher peak flows in the undisturbed run than in the burned run. These results do not prove that the real undisturbed watershed is noisy. They show that the model's large undisturbed peaks are not a stable reference.

This changes how a burned-versus-undisturbed inversion should be investigated. If the burned peak is lower than the undisturbed peak, the first question should not automatically be, "What is wrong with the burned model?" The undisturbed peak may be abnormally high. Both results need to be checked, and at Topanga the undisturbed result deserves at least as much scrutiny as the burned result. The central lesson is simple: the prefire scenario is a comparison case, not ground truth.

7 Original Census Limitations and Claim Boundary

What is established. The documentation, source comparison, diagnostic execution, and census jointly establish that:

- present WEPP assigns a daily `surdra` depth new subdaily timing rather than routing the hourly return series;
- that operation caused the 1980 K_{sat} peak reversal, while the solver switch enlarged but did not create the reversal;
- controlled canopy and cover changes expose additional repeatable peak discontinuities;
- the Topanga population contains a structured, predominantly K_{sat}-associated extreme tail; and
- runoff volume is substantially more stable than calculated peak rate.

What remains unresolved. The evidence does not establish that every extreme row has the same cause. Positive surface return occurs in 567 of 989 twofold-tail rows and 420 of 604 fivefold-tail rows, leaving a material no-return tail, especially in the burned stratum. The immediate branch behind the 1986 cover boundary is not yet known; neither side of an extreme pair has been established as the physically correct peak; and the Windows source comparison is static rather than a fully instrumented execution of every Topanga fixture. Watershed routing and effects at the outlet also remain outside this census.

Study-scope limitations.

- **One enriched-discovery site.** Topanga was selected because it contained known anomalies. These rates are not cross-site prevalence estimates.
- **Event-row clustering.** A trial contributes many events. Event-row percentages are not percentages of independent hillslopes, storms, or watersheds.
- **Inclusive screening.** A candidate can be triggered by event appearance, solver change, directional reversal, magnitude change, or surface-return-rate change. Candidate prevalence is not confirmed-defect prevalence.
- **Local hillslope scope.** Per-mutation watershed routing was intentionally removed after the pilot exposed unresolved off-path changes and incompatible channel-volume authorities. No attenuation, synchronisation, or outlet-impact claim is made.
- **Positive-event magnitude figures.** The peak scatter excludes rows below the baseline peak floor or without a positive mutant peak. The runoff scatter similarly applies the 0.01 mm baseline runoff floor and requires positive mutant runoff. The sediment-delivery scatter requires positive delivery on both sides; its zero and one-sided rows are reported separately. Event-presence changes remain in the complete ledger and candidate screen.
- **Sediment-output precision.** The retained event files report sediment delivery to 0.1 kg/m. Exact ties and low-magnitude relative noise therefore include output quantisation, not solely model response.
- **Mechanisms not exhaustively adjudicated.** Hill 106 establishes implementation mechanisms that can produce discontinuity. The full tail population still requires trial-level grouping, adaptive bracketing, and frozen-event replay.

8 Cross-Site Replication Next Steps

8.1 Additional sites

The next sites should retain the exact small-mutation design, execution contract, floors, and event-pairing keys. Site selection should proceed under the two preregistered portfolios already defined by the audit:

- **Enriched discovery**: known-positive rain-dominated sites such as Palisades and Stevens Canyon, used to test recurrence of known mechanisms.
- **Blind audit**: sites frozen by hydrologic and provenance criteria before their peak responses are examined.

The portfolio should add rain-dominated forest, transient-snow, and snow-dominated forest conditions. Results must be reported separately for enriched and blind portfolios. A shared small-probe design makes the site comparison interpretable; increasing mutation size at later sites would change the estimand.

9 Original Census Conclusion

The Topanga census demonstrates why small-mutation screening is valuable. A large perturbation was not required to reveal the issue. Under ±1% K_{sat} and ±0.01 paired-cover probes, most paired events remain near baseline, providing a clear ordinary-response reference. Event runoff is especially stable: 98.3% of K_{sat} runoff responses have the expected direction, and only 0.21% of plotted runoff rows depart by more than twofold. Event sediment delivery is also mostly controlled: 99.805% of paired-positive rows remain within a factor of two. Its low-magnitude values are quantised and noisy, and the visible central-band dispersion supports interpreting individual erosion outputs with an empirical factor-of-two sensitivity envelope rather than as precise event predictions.

Against the stable runoff-volume response, a small but conspicuous peak-flow extreme population appears. It is overwhelmingly associated with K_{sat}, usually moves in the expected hydrologic direction, and is too large to summarise as ordinary local sensitivity to a one-percent change. The 604 rows outside the 0.2–5 band represent an empirical Topanga rate of approximately one in 330 paired-positive event rows. Thus, the peak calculation is sensible in aggregate but cannot be trusted without qualification for an arbitrary hillslope event. Topanga definitively demonstrates that the diagnostic peak-flow calculation can generate anomalous peaks of meaningful magnitude at hillslope scale.

The burned–unburned separation supplies an additional key result. Because the two strata share the same climate sequence, the shift from an unburned K_{sat} incongruence cluster near 74 mm/h to a burned cluster near 39 mm/h reflects the effect of soil, vegetation, cover, and burn-severity parameterisation on the modelled hydrologic state. Parameterisation changes not only runoff magnitude but the peak scale at which the unstable response regime is encountered. This state shift provides a credible link to previously observed burned–unburned peak inversions: scenario-specific events can enter the amplified peak regime at different points even under the same storm forcing. The association is demonstrated here; attribution to a particular duration or solver transition still requires frozen-event replay.

The filtered comparison further locates the high-intensity directional instability in mixed infiltration-excess and surface-return events. Excluding that pathway in either paired run lowers central-band K_{sat} incongruence at 30–100 mm/h from approximately one in four rows to fewer than one in 50 in both strata. This supports the surface-return timing interpretation, while the 184 retained fivefold departures without surface return demonstrate that it does not explain every extreme response. Filtering selects a different population; it does not repair the excluded peaks.

In production, a single hillslope or a small set of hillslopes may contribute abnormal events that distort aggregated extremes and return-period estimates. That pathway is operationally credible given the empirical tail rate and the large hillslope-by-event population of a production run. It remains an inference, however, until routed replays show how the hillslope anomalies propagate to watershed outputs and affect fitted return periods.

The appropriate next step is replication, not expansion of mutation magnitude. Apply the same small design at additional preregistered sites, preserve the Topanga claim boundary, and mechanism-trace selected candidates. Until that work is complete, the report supports a local Topanga finding: WEPP hillslope peak flow contains a structured K_{sat}-sensitive response whose

magnitude and state dependence warrant implementation-level investigation.

10 Current Remediation Conclusion

Increasing `rrinit` is a parameter intervention that can mask one defective timing pathway, not a repair of it. Neither 17 cm nor 60 cm is a universal threshold or a justified replacement default. The implemented surface-return estimator is a more targeted response: it avoids compressing daily return into a potentially tiny positive-excess duration without adding user flags or altering roughness. Its repeated mutation results support a substantial reduction of the most extreme peak departures and high-rate directional instability, not elimination of all sensitivity.

The watershed comparison supports nearly preserved long-term yield and shows that peaks can change substantially in either direction, including smaller events. It also exposes a larger routed-discharge/volume-ledger discrepancy with the candidate. Reconciling that discrepancy is the next release gate; lower peaks alone are not evidence of a more accurate hydrograph. Stakeholder disclosure should distinguish diagnosed defect, implemented candidate, completed research, unresolved validation, and any later deployment. No production deployment or external stakeholder notification is implied here.

Evidence and Reproducibility

The principal local sources for this report are:

- the Topanga investigation handoff, including the Windows reconstruction and Hill 106 mechanism narrative;
- the stakeholder report on WEPP peak-flow documentation and Topanga implementation evidence;
- the multi-site audit protocol and its local-census amendment;
- the frozen Topanga execution disposition and prevalence summary; and
- the figures, sidecars, plotting scripts, and repeatable stratum-cluster analysis under the Topanga execution work package;
- the completed openWEPP hourly peak-authority closure and repeated Topanga mutation census; and
- the closed WEPP-Forest feasibility package documenting why direct hourly saturation-return insertion was abandoned.

Two diagnostic stages should not be confused. The original Hill 106 mechanism trace and forced-solver tests used WEPP-Forest source commit `2f65506d239b449b...fa55158` and executable `wepp_260803`. The later census observer used commit `ea25ad79` and recorded every peak-solver call without changing ordinary WEPP outputs. The public Windows comparison used Daily Erosion Project commit `e2609d9e67757f...ef657c`, whose `src/wepp20240930` directory is recorded as a verbatim September 2024 WEPP import. Public implementation links, complete hashes, and file comparisons are retained in the stakeholder evidence report [6].

The source-code evidence is essential to the defect identification. Ordinary WEPP output does not expose the daily return depth, selected duration, imposed rate, ponding-time sentinel, or solver

handoff. USDA has distributed WEPP primarily as an executable and has not provided an official complete release of the production source. The analysis was possible because preserved source copies could be compared, instrumented, and checked against unchanged model outputs.

The figure source is `docs/work-packages/20260809_peakflow_topanga_census_execution/artifacts/`. The immutable event ledger is inventoried by the execution package's external evidence manifest. The report deliberately cites compact repository evidence rather than embedding the large Parquet ledger.

October evidence is retained in three immutable work packages, including methods, binary identities, input/output manifests, machine-readable summaries, plotting scripts, and validation reviews [1, 2, 3]. The revised Figures 1–3 are candidate-build mutation figures; the original August scatter plots are retained later with their original population and interpretation. Figure numbering is updated through LaTeX cross-references.

References

[1] Surface-Return Mutation Repeat (2026). `docs/work-packages/20261006_surface_return_mutation/artifacts/results.md`.

[2] Warming-championship Initial Roughness Experiment (2026). `docs/work-packages/20261006_warming_rrinit/artifacts/results.md`.

[3] Original WEPP Roughness Sensitivity and Surface Return Comparison (2026). `docs/work-packages/20261006_warming_rrinit_legacy/artifacts/results.md`.

[4] Stone, J. J., Lane, L. J., and Shirley, E. D. (1995). "Hillslope Surface Hydrology." In *USDA–Water Erosion Prediction Project: Hillslope Profile and Watershed Model Documentation*, Chapter 4, NSERL Report No. 10. https://www.ars.usda.gov/ARSUserFiles/50201000/WEPP/chap4.pdf.

[5] Savabi, M. R., Skaggs, R. W., and Onstad, C. A. (1995). "Subsurface Hydrology." In *USDA–Water Erosion Prediction Project: Hillslope Profile and Watershed Model Documentation*, Chapter 6, NSERL Report No. 10. https://www.ars.usda.gov/ARSUserFiles/50201000/WEPP/chap6.pdf.

[6] Lew, R. and agentic collaborators (2026). "WEPP Peak-Flow Estimation: Long-Standing Discontinuities and Their Implementation Basis." `docs/investigations/2026-08-07-topanga-2025-fire-peak-flow-analysis/artifacts/wepp-peak-flow-solver-documentation-and-topanga-evidence.md`.

[7] Lew, R. and agentic collaborators (2026). "Topanga Investigation Handoff." `docs/investigations/2026-08-07-topanga-2025-fire-peak-flow-analysis/artifacts/bill-windows-replication-handoff.md`.

[8] WEPP Peak-Flow Discontinuity Multi-Site Audit (2026). `docs/investigations/2026-08-08-wepp-peak-flow-discontinuity-multi-site-audit/README.md`.

[9] Frozen Topanga Local Census Execution Disposition (2026). `docs/work-packages/20260809_peakflow_topanga_census_execution/artifacts/execution-disposition.md`.

[10] openWEPP Hourly Peak-Runoff Authority Closure (2026). `/workdir/openWEPP/docs/work-packages/20260809-hourly-peak-runoff-authority-closure-001/`.

[11] WEPP-Forest PF01 `surdra` Kinematic-Wave Partitioning Work Package (2026). `/workdir/wepp-forest_260430_baseline/docs/work-packages/20260810-pf01-surdra-kinematic-wave-partitioning/`.