Thank you for the opportunity to review "Wildfire-induced disruptions to evapotranspiration, runoff, and water-balance closure across California's water supply watersheds". In this paper the authors present a multi-decadal synthesis of wildfire, ET, precipitation, and full natural flow data across key California basins. The results are consistent with growing body of post-wildfire hydrologic response literature, making the topic timely and likely of interest to a broad audience. I have several broader conceptual concerns that I believe should be addressed to strengthen the interpretation of the results.

We thank the reviewer for the careful and thoughtful evaluation of our manuscript, as well as the constructive comments and conceptual suggestions, which have helped us improve the clarity and interpretation of the manuscript.

1. First, I am uncertain about the physical interpretation and motivation behind the parameter adjustments (scaling + intercept) applied to precipitation and/or FNF to improve water-balance closure. While these adjustments are presented as diagnostic tools, they risk appearing ad hoc without clearer linkage to quantified uncertainty or process understanding. I think it would strengthen the manuscript to clarify whether these adjustments are intended to identify systematic dataset biases, to approximate unaccounted storage fluxes, or to represent measurement error.

Response:

We thank the reviewer for highlighting this important point. In the revision, we will clarify that the scaling and intercept parameters are intended solely as diagnostic indicators, not as physical corrections. In other words, we introduce these parameters to characterize the structure of the residual (P–ET – FNF) rather than to retroactively adjust the data. Concretely, the scaling parameter is interpreted as indicating a multiplicative bias (for example, if precipitation or outflow is systematically under- or over-estimated). The intercept parameter is interpreted as an additive offset, which could represent an unresolved flux (such as a non-zero ΔS or other constant error). We explicitly state that neither parameter is applied to the original P, ET, or FNF time series; instead, they help us determine whether the residual behaves like a simple bias (multiplicative) or an offset (additive).

To ensure consistency, we will revise the text accordingly. In the Methods section (Sec. 2.3), we will replace the previous "steady-state $\Delta S \approx 0$" statement with a description of an apparent non-closure residual $R = P - ET - FNF$. We will then then explain that we introduce "diagnostic reconciliation parameters (multiplicative scaling and... an additive offset)" to characterize the residual structure. Throughout the Results and Discussion, we updated wording to use terms like "diagnostic term" instead of "correction," and "residual" instead of "imbalance." For example, we now say "applying a multiplicative diagnostic term" rather than "proportional scaling." We will also change the section title to "Diagnosing postfire water-balance non-closure and residual structure."

These revisions will make it clear that the parameters are meant to identify systematic biases (multiplicative term) or offsets (additive term), rather than to "fix" the water balance ad hoc. The physical implications of each parameter are now explicitly described in the text. We believe this addresses the reviewer's concern by linking the adjustments to data uncertainty and storage, and removing any implication of arbitrary calibration.

Revisions:

Lines 186–189 (Methods, Sec. 2.3)*:* Original: "Under steady-state conditions ($\Delta S \approx 0$), $P - ET$ should closely match FNF. Deviations from the 1:1 relationship were therefore used to identify potential imbalance and diagnose hydrologic processes..." will be revised into: "At annual (water-year) timescales, changes in basin storage (ΔS) are often smaller than the dominant flux terms, but they are not necessarily negligible, particularly following major disturbance. Accordingly, deviations between P−ET and FNF are interpreted as an 'apparent non-closure residual', $R = P - ET - FNF$, which may reflect a combination of ΔS and uncertainties or structural mismatches in P, ET, and/or FNF."

Lines 190–196 (Methods, Sec. 2.3)*:* Original: "Where persistent discrepancies existed, we applied proportional adjustments to either P or FNF, depending on which correction produced improved alignment... These adjustments were used solely as diagnostic tools rather than for calibration or data fitting." will be revised into: "Where persistent discrepancies existed, we introduced simple diagnostic reconciliation parameters (multiplicative scaling and, where needed, an additive offset) to characterize whether the apparent residual structure was more consistent with a multiplicative bias (e.g., systematic under- or over-estimation) or an additive structural offset (e.g., drainage-area mismatch, persistent exchange fluxes, or non-zero ΔS aggregated at annual scale). Watersheds were grouped into categories such as: basins exhibiting minimal apparent non-closure; basins where a multiplicative term reduced the residual; and basins where both scaling and an intercept term were required. These parameters are used only for diagnostic interpretation of residual structure, not for calibration or modification of the underlying datasets."

Lines 295–296: Original: "To diagnose the sources of imbalance, we applied basin-specific adjustments to P or FNF." Will be revised to "To diagnose the sources of apparent non-closure, we evaluated basin-specific diagnostic scaling and intercept terms applied to P or FNF to summarize the residual structure."

Lines 303–305 Original: "P − ET > FNF, suggesting incomplete runoff measurement. Applying a proportional scaling to FNF reduced these discrepancies; for example, in FTO, a small adjustment achieved a post-correction residual of 132 mm..." will be revised into: "P − ET > FNF, indicating a persistent positive residual. Applying a multiplicative diagnostic term to FNF reduced this residual; for example, in FTO, a small scaling reduced the residual to 132 mm..."

Lines 306–310: Original: "For several basins (e.g., PSH, MSS, SIS), large residuals persisted even after proportional adjustment. The intersection of regression lines indicated structural bias rather than simple scaling errors. Introducing a second parameter (intercept) to adjust FNF produced near-perfect alignment between P − ET and FNF, reducing residuals to within 50 mm and confirming that multi-parameter corrections better captured hydrologic consistency (Figure 5d,e; Figure S6)." will be revised into: "For several basins (e.g., PSH, MSS, SIS), large residuals persisted even after applying a multiplicative term, and crossing regression patterns indicated that an additive offset was also needed. Introducing an intercept term provides a simple diagnostic representation of an additive residual component (e.g., structural mismatch, persistent exchange fluxes, or aggregated non-zero ΔS at annual scale). With both scaling and an intercept term, residuals were reduced to within 50 mm (Figure 5d,e; Figure S6), indicating that a two-parameter form better summarizes the observed residual structure in these basins."

Line 403: Original: "4.3 Diagnosing and correcting postfire water-balance closure errors" will be revised into: "4.3 Diagnosing postfire water-balance non-closure and residual structure"

Lines 404–410 Original: "Wildfire disturbance emerged as an important driver of deviations from expected water-balance closure... fire years consistently exhibited more negative residuals (P − ET − FNF), indicating that observed basin outflow (FNF) systematically exceeded modeled P − ET... This divergence demonstrates that wildfires can create short-term hydrologic imbalances..." will be revised into: "Wildfire disturbance emerged as an important correlate of deviations from apparent annual water-balance closure across California's major watersheds. Fire years consistently exhibited more negative residuals (P−ET−FNF), indicating that observed basin outflow (FNF) exceeded P−ET to a greater degree during and immediately after disturbance. Because the annual residual reflects the combined influence of ΔS and uncertainties or structural mismatches in P, ET, and FNF, these results indicate that wildfire years are associated with systematic shifts in the residual structure rather than implying a single underlying mechanism."

Lines 435–437 Original: "Applying proportional corrections reduced residuals to within 100–150 mm, indicating that flow underestimation was the dominant error in these cases." will be revised into: "Applying a multiplicative diagnostic term reduced residuals to within 100–150 mm, consistent with the presence of systematic discrepancies in the FNF–P−ET relationship in these basins."

Lines 441–444 Original: "Here, we applied a dual-parameter correction (scaling + intercept), which improved alignment and reduced residuals to within 50 mm. This suggests that in basins with mixed error sources, flexible calibration models that reflect physical measurement processes are necessary for reliable water balance closure." will be revised into: "Here, we applied a dual-parameter diagnostic form (scaling + intercept), which reduced residuals to within 50 mm. This indicates that in basins where residuals include both multiplicative and additive components, a two-parameter form provides a more informative summary of the apparent non-closure structure, although it does not uniquely identify the underlying physical processes."

2. Second, the water-balance framework assumes that annual changes in storage (ΔS) are negligible when comparing P−ET to FNF. While this generally seems OK at the scales considered here, deviations between P−ET and FNF during fire years could plausibly reflect changes in storage rather than solely disturbance-induced hydrologic "imbalance" or observational uncertainty. I encourage the authors to more explicitly justify the ΔS ≈ 0 assumption and discuss its limitations.

Response:

We thank the reviewer for this helpful comment. We agree that deviations between $P - ET$ and FNF at annual scales may partly reflect changes in basin storage (ΔS), particularly during disturbance years. Our intention was not to assume that $\Delta S = 0$, but rather that annual storage changes are generally small relative to the dominant flux terms at basin scale. We therefore interpret the residual $R = P - ET - FNF$ as an apparent non-closure residual that may include contributions from storage change as well as uncertainties in precipitation, evapotranspiration, or flow observations.

In the Methods (Sec. 2.3) we now state that at annual (water-year) scales, ΔS is often smaller than the main flux terms but "not necessarily negligible, particularly following major disturbance." We clarified that P−ET "is expected to approximate FNF under broadly balanced conditions," and that any deviation (the residual R) may include storage changes. In the Results and Discussion, we interpret deviations accordingly: for example, we added a sentence noting that deviations may come from "transient storage variation" as well as uncertainty or disturbance effects. We also discuss how fire can change infiltration and groundwater recharge. These additions ensure the reader knows that our analysis allows for $\Delta S \neq 0$ and that this is part of the interpretation of residuals.

Revisions:

Lines 289–292: Original: "Under ideal conditions, modeled P and ET should closely reproduce observed runoff, resulting in well-aligned P−ET and FNF curves. In basins where ET remains stable and unaffected by disturbance, a near-linear relationship between P and FNF is expected, with data points clustering along the 1:1 line (reflecting annual storage change) and a slope approaching unity." will be revised into "Under ideal conditions, modeled $P$and ETshould broadly reproduce observed runoff, resulting in similar temporal behavior between $P - ET$and FNF. In basins where ET remains stable and disturbance effects are limited, a near-linear relationship between $P$and FNF is expected, with data points clustering near the 1:1 line and a slope approaching unity. Deviations from this relationship represent the combined effects of transient storage variation, dataset uncertainty, and disturbance-related hydrologic responses."

Lines 300–304: Original: "the imbalance likely arose from underestimated precipitation. Multiplying P by 1.03 reduced residuals to near zero, aligning with previous evidence of precipitation underestimation in mountainous terrain." Will be revised into "the residual likely reflects a combination factors, including precipitation uncertainty, potential groundwater or baseflow changes, and other subsurface flow processes. Multiplying $P$by 1.03 reduced the apparent residual to near zero, consistent with previous evidence that gridded precipitation products may underestimate totals in complex mountainous terrain."

Lines 471–473: Original: "Persistent negative P–ET–FNF residuals reflect not only process-level shifts but also precipitation underestimation, gauge bias, and heterogeneous subsurface pathways." will be revised into "Persistent negative $P - ET - FNF$residuals reflect a combination of disturbance-driven hydrologic responses, transient storage variations, and observational uncertainties such as precipitation underestimation, gauge bias, and heterogeneous subsurface pathways.”

3. Finally, burn severity and vegetation controls are repeatedly invoked as key determinants of ET suppression and recovery, yet it is not fully clear how burn severity metrics are incorporated into the analysis. The methods primarily describe burned perimeters, and additional detail on whether severity (e.g., dNBR or similar indices) was used would improve clarity. Similarly, because vegetation structure and regrowth are central to much of the discussion of ET recovery trajectories, incorporating quantitative vegetation indices would help connect the mechanistic discussion to the results.

Response 1:

We thank the reviewer for highlighting the need to clarify how burn severity and vegetation structure are represented in the analysis. In the original manuscript, wildfire disturbance was identified using burned perimeters derived from the FRAP fire history dataset, and ET responses were evaluated within those burned areas. Burn severity was discussed conceptually to interpret differences in ET suppression and recovery, but severity metrics were not explicitly quantified in the analysis, which may have created ambiguity.

To address this point, we will clarify that the disturbance representation in the Methods and incorporated a quantitative burn-severity metric derived from the Monitoring Trends in Burn Severity (MTBS) dataset. For each fire perimeter, we calculated the fraction of high-severity burn area based on MTBS severity classifications. We then evaluated the relationship between this severity metric and the magnitude of first-year ET reduction across the analyzed fires. The results confirm that larger ET declines are generally associated with higher proportions of high-severity burn area, consistent with expectations from canopy loss and vegetation mortality.

Revisions 1:

New paragraph added after line 167: “We also explored the impacts of burn severity on the ET reduction. Time series burn severity layers across the years when the wildfires occurred were obtained from the Monitoring Trends in Burn Severity (MTBS) dataset. For each fire perimeter, we calculated the fraction of high-severity burn area based on MTBS severity classifications. We then evaluated the relationship between this severity metric and the magnitude of first-year ET reduction across the analyzed fires.”

Other text in the Results and Discussion section regarding burn severity will be revised accordingly, based on the results shown below.

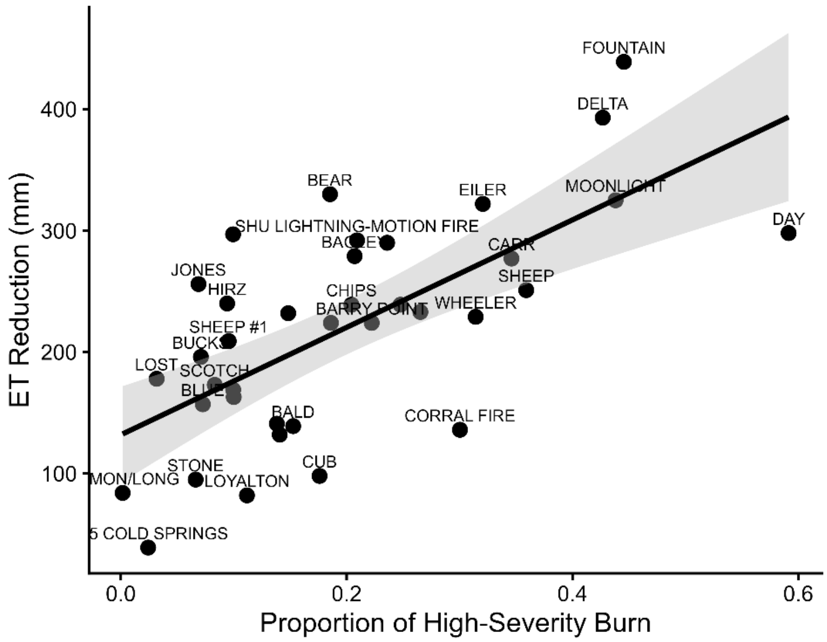

Figure S. Relationship between burn severity and first-year evapotranspiration (ET) reduction across fires. First-year ET reduction (mm yr^{-1}) is plotted against the proportion of high-severity burn area within each fire perimeter. Each point represents an individual fire. The solid line shows the linear regression fit, and the shaded region indicates the 95% confidence interval.

Line 214-217: Original 'These examples show that both burn severity and burned area govern the magnitude and duration of ET reductions. Cumulative losses ranged from tens to hundreds of millions of cubic meters before partial recovery was achieved.' will be revised into "Cumulative losses ranged from tens to hundreds of millions of cubic meters before partial recovery was achieved. Across fires, the magnitude of first-year ET reduction increased with the fraction of high-severity burn area (Fig. S)"

Add the reference to this Figure in the discussion section (Line 327, 330) to show the origin of the burn-severity analysis.

Response 2:

We agree that quantitative vegetation information strengthens interpretation of heterogeneous ET recovery. In the revision, we explicitly use the CECS vegetation composition fractions (tree/shrub/herbaceous/bare) shown in Figs. S1–S3 as quantitative proxies for postfire structural change and regrowth pathways, and we reference them directly when discussing recovery trajectories. These examples show that prolonged ET suppression coincides with sustained low tree fraction and increased shrub/herbaceous/bare fractions, consistent with delayed canopy re-establishment and/or repeated disturbance.

Revisions 2:

Line 22-24: the original 'High-severity 22 fires consistently suppressed ET by 100–250 mm in the first postfire year, with recovery strongly modulated by vegetation traits, moisture availability, and disturbance recurrence.' will be revised into 'High-severity fires consistently suppressed ET by 100–250 mm in the first postfire year, with recovery influenced by moisture availability and disturbance recurrence, and coinciding with shifts in postfire vegetation composition".

Methods: Line 167: Add sentence "To provide a quantitative context for vegetation structure during the ET recovery process, we also examined the proportional composition of vegetation within the fire perimeter. Vegetation composition data were derived from the CECS dataset, comprising the proportions of trees, shrubs, herbaceous vegetation, and bare ground."

Results 3.1, Line 199-200: We will delete the 'depending on fire severity and vegetation condition' as we will explain in detail in the following sentences.

Line 217: add at the end:" We assessed the vegetation composition and its changes within the perimeter of each fire, before and after the fire occurred. (Figs. S1–S3). Fires with prolonged ET suppression show sustained low tree fraction and elevated shrub/herbaceous/bare fractions over multiple years, whereas faster ET recovery coincides with more rapid increases in tree cover and declines in bare ground."

Discussion 4.1. Line 336: original 'Recovery trajectories varied across basins and reflected vegetation structure, climatic water availability, and disturbance history' will be revised into 'Recovery trajectories varied across basins and were influenced by climatic water availability and disturbance history, and coinciding with shifts in postfire vegetation composition (Figures S1–S3)".

L344–345. "underscoring the importance of species traits." Will be revised into "underscoring the coincidence with postfire vegetation structural recovery pathways and composition shifts (Figures S1–S3)."

Line 490: Original “with recovery rates governed by vegetation traits, moisture availability, and disturbance recurrence.” will be revised into “with recovery rates influenced by moisture availability and disturbance recurrence, and coinciding with shifts in postfire vegetation composition”.

I’ve included some line-by-line comments below.

4. Abstract: Line 26: Should “Interannual P” be “annual P”?

Response: Agreed, and the wording will be revised at Line 26.

5. Introduction: *L 49: What are the compounding disturbances? Repeated wildfires? Wildfire and drought?*

Response: We agree the term should be defined. Here we mean wildfire interacting with drought disturbance recurrence, which jointly shape vegetation recovery and hydrologic partitioning. We clarified this explicitly in the Introduction.

Revisions: Line 49: Revise “These compound disturbances…” to “These compound disturbances including wildfire interacting with drought and with disturbance recurrence”.

6. *L57: I’m not sure what “vegetation mortality surplus” means? Here and throughout it might be helpful to add a little more detail about the expected change to individual ET fluxes. The post-fire signal is most likely driven by decreasing transpiration, but there can be compensatory increases in ground evaporation. It might be worth referencing some of the papers by Collar et. al.*

- Collar, N. M., Ebel, B. A., Saxe, S., Rust, A. J., & Hogue, T. S. (2023). Implications of fire-induced evapotranspiration shifts for recharge-runoff generation and vegetation conversion in the western United States. Journal of Hydrology, 621, 129646. https://doi.org/10.1016/j.jhydrol.2023.129646
- Collar, N. M., Saxe, S., Rust, A. J., & Hogue, T. S. (2021). A CONUS-scale study of wildfire and evapotranspiration: Spatial and temporal response and controlling

factors. Journal of Hydrology, 603, 127162. https://doi.org/10.1016/j.jhydrol.2021.127162

Response: We agree this phrase is unclear. We will replace it with more direct language describing wildfire-driven vegetation mortality and canopy loss. We will also add a short clarification that total ET changes can reflect reduced transpiration partially offset by increased soil/ground evaporation, consistent with prior post-fire ET partitioning studies.

Revisions: Line 57-58: we will revise "ET typically declines due to vegetation mortality surplus" in to "ET typically declines due to wildfire-driven vegetation mortality and canopy loss", and add a sentence following line 58 with "While reduced transpiration is expected to dominate the first-year ET signal, compensatory increases in soil/ground evaporation can partially offset this decline in some settings (Collar et al., 2023; Collar et al., 2021)."

7. *L 80: and moisture availability? Especially in California, which is likely a moisture limited system?*

Response: We thank the reviewer for this important clarification. We agree that moisture availability is a primary control on ET dynamics in California's seasonally water-limited climate.
Revisions: We will revise the text (L79–80) to explicitly acknowledge soil moisture availability as a key control on ET alongside vegetation structure and energy availability.

8. *L 88: While I generally agree that that this gap exists, there are quite a few papers in this space (some of which the authors already reference) that do address this gap. For example:*

- Goeking, Sara A.;Tarboton, David G. 2020. Forests and water yield: A synthesis of disturbance effects on streamflow and snowpack in western coniferous forests. Journal of Forestry. 2020: 172-192.
- Hallema, D.W., Sun, G., Caldwell, P.V. *et al.* Burned forests impact water supplies. *NatCommun* **9**, 1307 (2018). https://doi.org/10.1038/s41467-018-03735-6
- Beyene MT, Leibowitz SG, Pennino MJ. Parsing Weather Variability and Wildfire Effects on the Post-Fire Changes in Daily Stream Flows :A Quantile-Based Statistical Approach and its Application. Water Resour Res. 2021 Aug 31;57(10):1-20. PMID: 34898727; PMCID: PMC8654146.

Response:

We appreciate the reviewer's comment and agree that several important studies have substantially advanced understanding of post-wildfire hydrologic responses, including the synthesis by Goeking & Tarboton (2020), the large-scale analysis by Hallema et al. (2018), and the statistical attribution framework of Beyene et al. (2021). We have revised the Introduction to better acknowledge this growing body of literature.

Our intent was not to suggest that post-fire hydrologic effects remain unexplored, but rather to highlight the relative scarcity of basin-scale analyses that directly integrate long-term satellite-derived ET, gridded precipitation, full natural flow (FNF) records, and spatially explicit fire history within a unified annual mass-balance framework. In particular, few studies have evaluated water-balance closure (P − ET vs. FNF) across multiple water-supply basins over multi-decadal time scales while explicitly diagnosing disturbance-related deviations and dataset biases.

Revisions:

We will revise the text to clarify this distinction and to better position our contribution within the existing literature in line 86-93. "Although substantial progress has been made in understanding post-wildfire hydrologic responses (e.g., Goeking & Tarboton, 2020; Hallema et al., 2018; Beyene et al., 2021), comparatively fewer studies have integrated long-term satellite ET, precipitation, and full natural flow records within a unified basin-scale mass-balance framework across multiple major water-supply basins."

9. *L118 – Some recent work has identified aridity as an important driver of post-wildfire hydrologic response. Is there an aridity gradient across your watersheds? It might be worth including in Table 1*

- Baudena M, Santana VM, Baeza MJ, Bautista S, Eppinga MB, Hemerik L, Garcia Mayor A, Rodriguez F, Valdecantos A, Vallejo VR, Vasques A, Rietkerk M. Increased aridity drives post-fire recovery of Mediterranean forests towards open shrublands. New Phytol. 2020 Feb;225(4):1500-1515. doi: 10.1111/nph.

Response: We thank the reviewer for this helpful suggestion. We agree that aridity is an important control on post-wildfire hydrologic response and vegetation recovery, and that explicitly characterizing the climatic gradient across basins strengthens the interpretation of our results.

To address this point, we calculated a basin-scale aridity indicator based on long-term averages (1991–2020). Specifically, we defined an aridity index as the ratio of the Maximum Annual AET possible (AET_m) to mean annual precipitation (P). AET_m was derived from the CECS dataset at 30m resolution and stacked over the 30 year for the mean value.

The results show a clear aridity gradient across the study basins, with wetter basins in the Klamath–Cascade region (e.g., Trinity, Upper Sacramento, and McCloud; $AET_m/P \approx 0.53$–0.59) and comparatively drier conditions in the Pit basin ($AET_m/P \approx 0.94$), with Feather exhibiting intermediate conditions. This gradient is consistent with regional hydroclimatic patterns and provides additional context for interpreting differences in ET response and recovery across basins.

Revisions: We will add this aridity indicator to Table 1 and revise the study-area description to explicitly note this climatic gradient. In line 134, we will add sentences: “The study basins span a clear hydroclimatic gradient, which we quantify using a long-term aridity indicator (AET_m/P) range from approximately 0.53–0.59 in wetter basins (Trinity, Upper Sacramento, and McCloud) to ~0.94 in the relatively drier Pit basin, with Feather showing intermediate conditions (Table 1).”

Watershed	AET_m, mm	AET_m/P
Trinity	798	0.57
Upper Sacramento	831	0.53
McCloud	830	0.59
Pit	625	0.94
Feather	770	0.71

10. *Data section.*

As a reader who is not familiar with these particular hydrometeorological datasets it would be helpful to include a litter more detail describing how they are calculated. For example from the spatial resolution I assume that the ET data is derived from an energy balance model built on Landsat imagery, but maybe that's incorrect? Does the ET model make assumptions over burned areas that could influence the ET estimates? Is P derived from the same model? Does it incorporate any other precipitation datasets?

Response:

We thank the reviewer for this helpful suggestion. We agree that additional clarification of the hydrometeorological datasets would improve transparency for readers who may be unfamiliar with these products.

Evapotranspiration (ET) in this study was obtained from the California Ecosystem Climate Solutions (CECS) dataset, which provides annual ET estimates at 30-m spatial resolution. These ET estimates are derived from a Landsat-based empirical model that combines satellite-derived vegetation indices with climate variables and eddy-covariance flux-tower measurements across California. Specifically, the model relates annual ET to Landsat-derived normalized difference vegetation index (NDVI) using empirical relationships calibrated with flux-tower observations (Goulden & Bales, 2019; Chung et al., 2024).

Importantly, the ET product is not derived from a surface energy-balance model. Instead, it relies on empirically calibrated relationships between remotely sensed vegetation activity and measured ET. Consequently, wildfire impacts on ET are represented indirectly through changes in vegetation structure and greenness captured by Landsat NDVI.

Precipitation (P) is derived independently from the PRISM climate dataset, which provides gridded daily precipitation at approximately 800-m spatial resolution. Annual precipitation totals were calculated by summing daily PRISM values and then resampled to the 30-m ET grid to maintain spatial alignment. (Roche et al., 2020).

We will add this clarification to the Methods section to better explain the derivation of the hydrometeorological datasets used in this study.

Revisions:

Line 142-146: “CECS provides annual evapotranspiration (ET) and precipitation (P) raster products from 1985 to 2024 at 30-m resolution derived from its internal water-balance calculations” will be revised into “The Center for Ecosystem Climate Solutions (CECS) provides annual evapotranspiration (ET) products at 30-m spatial resolution across California. These ET estimates are derived from a Landsat-based empirical model that combines remotely sensed vegetation indices with climate variables and eddy-covariance flux-tower observations. Specifically, annual ET is estimated from relationships between Landsat-derived normalized difference vegetation index (NDVI) and measured ET at flux-tower sites distributed across California forests (Goulden & Bales, 2019; Guo., et al, 2023; Chung et al., 2024).

Precipitation (P) was obtained from the PRISM climate dataset, which provides gridded daily precipitation at approximately 800-m spatial resolution. Annual precipitation totals were aggregated from daily PRISM data and resampled to match the 30-m ET grid (Roche et al., 2020)".

11. *Similarly with the FNF data. Do these data account for diversions and reservoir operations? It might be helpful to include a little more detail since you reference reservoir effects a few times in the discussion.*

Response:

We agree that the manuscript did not sufficiently describe what CDEC full natural flow (FNF) represents and how it relates to reservoir operations and diversions. CDEC FNF is not a raw gauge measurement; it is a "naturalized/unimpaired" flow estimate intended to approximate basin water production in the absence of upstream diversions, storage regulation, and net basin imports/exports. FNF is computed by adjusting gauged flows at the reporting location to account for upstream operations, based on available operator/agency calculations and records. We have revised Section 2.2 (Data) to (i) define FNF explicitly as an unimpaired-flow estimate, (ii) note that daily FNF values are preliminary and may differ from monthly values due to data availability and lagged operational effects, and (iii) acknowledge that uncertainty in the naturalization procedure and management records can contribute to apparent water-balance residual structure.

Revisions:

Lines 147–152: Revise the sentence that begins "Additionally, we obtained monthly full natural flow…" to"Additionally, we obtained monthly full natural flow (FNF) estimates from the California Data Exchange Center (CDEC). CDEC defines FNF (also termed 'unimpaired runoff') as the natural water production of a basin absent upstream diversions, storage regulation, and net basin imports/exports; FNF is calculated by adjusting gauged flows at the reporting location to account for upstream operations. We use CDEC FNF as the best available long-term unimpaired outflow estimate for these basins, while recognizing that uncertainty in the naturalization procedure and in upstream operation/diversion records can contribute to apparent water-balance residual structure."

In Discussion 4.3, revise the sentence at Lines 418–419 “Climatic gradients, lithologic differences, and variable human regulation contribute…” to: “Climatic gradients, lithologic differences, and differences in the degree of upstream regulation (which affects uncertainty in FNF naturalization) contribute…”

Line 456-460, revise the sentence “In addition, many watersheds may be affected by data gaps or inaccurately reported upstream diversions, reservoir operations, and inter-basin transfers, complicating attribution of observed runoff deficits” to “In addition, many watersheds may be affected by data gaps or uncertainties in the naturalization procedure used to estimate full natural flow (FNF), including incomplete records of reservoir operations, diversions, and inter-basin transfers. Such uncertainties can propagate into the reconstructed unimpaired flow series and complicate attribution of observed water-balance residuals”.

12. *L 160-161: Can you clarify that this is mean annual ET? And is the post-fire ET just a single year average?*

Response: We agree and clarified that all ET and P quantities are water-year aggregated (mean annual) values, and that the post-fire ET refers to the first complete water year following the fire year.

Revisions: We will add a sentence after “ETburned, postfire … denote ET after the fire” with “All ET and P values are water-year aggregated (mean annual) quantities; the ‘first postfire year’ refers to the first complete water year after the fire occurrence.”

13. *L 186-196: I appreciate the discussion around uncertainty and challenges it could pose for interpretation. It seems like these are real challenges when considering the interacting effects of P and ET in a post-fire setting. I’m not quite sure I follow the motivation of your proportional adjustments approach though. Perhaps this an attempt to uncover systematic biases in the datasets? Is the assumption that d_S is really zero? Doesn’t a land cover disturbance like wildfire kind of inherently violate the assumption of steady state?*

Response:

We clarify that yes, our adjustments were indeed motivated by the need to reveal potential biases and offsets, and we have now made this explicit. Moreover, we do not assume that ΔS is strictly zero, especially after disturbance. Instead, we interpret the residual R = P−ET − FNF as an apparent non-closure, which can include ΔS as one component. In the manuscript we revised the Methods to state that on annual timescales ΔS is often smaller than major fluxes but "not necessarily negligible" after a disturbance. We also mention in the Results and Discussion that deviations may reflect transient storage changes.

Thus, our approach is not purely about dataset biases: it is designed to capture any systematic structure in the residual, whether from biases or from real storage change. For example, if a wildfire causes extra infiltration and temporary groundwater storage, that would appear as a residual that might require an additive (intercept) term. In practice, we use proportional (scaling) terms to test for multiplicative biases, and intercept terms to account for any persistent offset. We have reframed the text to make this motivation clear, and removed any language implying an assumed steady state.

Revisions: This concern overlaps with the changes already described above. In particular, the revised description in Section 2.3 now explicitly mentions that ΔS may vary post-disturbance (see the replacement in lines 186–190 above).

14. *L 203 – 217: Interesting to see the range of recovery trajectories. Are there and vegetation indices that could help explain why some burned areas seem to take so long to recover? You mention vegetation structure as a control on ET recovery a few times in the discussion, but I didn't see it really built into the results.*

Response:

We thank the reviewer for this helpful comment. To clarify this point, we will revise the manuscript to more explicitly describe how vegetation recovery is represented in the analysis, as described above in comment 3. Specifically, vegetation composition changes derived from CECS datasets (including tree, shrub, and herbaceous cover fractions) were already included in the Supplementary Materials (Figs. S1–S3). We will revise the Results and Discussion sections to clarify that differences in ET recovery trajectories are consistent with the vegetation composition changes shown in Figs. S1–S3. In particular, areas exhibiting slower ET recovery tend to correspond to locations where tree canopy recovery is slower and shrub or herbaceous cover increases following fire. The corresponding revisions include (shown in the response to comment 3):

Revisions:

Methods: Line 167: Add sentence "To provide a quantitative context for vegetation structure during the ET recovery process, we also examined the proportional composition of vegetation within the fire perimeter. Vegetation composition data were derived from the CECS dataset, comprising the proportions of trees, shrubs, herbaceous vegetation, and bare ground."

Results 3.1. Line 217: add at the end:" We assessed the vegetation composition and its changes within the perimeter of each fire, before and after the fire occurred. (Figs. S1–S3). Fires with prolonged ET suppression show sustained low tree fraction and elevated shrub/herbaceous/bare fractions over multiple years, whereas faster ET recovery coincides with more rapid increases in tree cover and declines in bare ground."

15. *L 215: Here and elsewhere you reference burn severity, but it seems like your only considering burned perimeters? Is burn severity built into the analysis? If so, more details would be helpful*

Response: We thank the reviewer for noting that burn severity was referenced without sufficient methodological explanation. We have now clarified in the Methods that burn severity was quantified using MTBS burn severity classifications and incorporated into the analysis as the fraction of high-severity burn area within each fire perimeter. Corresponding clarification will be added in the Results section where burn severity is first referenced (as mentioned above in comment 3, for line 167 and 202).

16. *L 218-219: Maybe consider moving this sentence in the methods?*

Response: We thank the reviewer for this helpful suggestion. This comment also relates to the comment below. Net ET reductions at the watershed scale in the analysis were actually derived by aggregating the ET responses from individual fires using area-weighted averages across all fires occurring within each basin. This aggregation summarizes the average disturbance signal of burned areas within each watershed rather than representing recovery of the entire basin to a single baseline condition.

Revisions: We will move the description to Section 2.3 (Line 167), as "Net ET reductions at the watershed scale were derived as area-weighted aggregates of ET responses from individual fires within each basin" and remove the sentence from the Results (Line 218-219) to avoid duplication.

17. *L239-240: Im not sure I understand the ET recovery at the watershed scale. Don't the repeated fires in the basin (even if they aren't spatially overlapping) continually move the "baseline"? There were likely fires prior to the start of your analysis as well. Over the larger basins, there might not really be such a thing a "prefile ET"?*

Response: We thank the reviewer for raising this important point regarding the interpretation of ET recovery at the watershed scale. As mentioned in the last comment. In this study, the recovery trajectories were first quantified at the individual fire level, where prefire ET was defined relative to conditions immediately preceding each fire event. The basin-level recovery metrics shown in Figure 3 therefore do not represent recovery of the entire watershed to a single static "prefire" baseline. Instead, they reflect the area-weighted aggregation of recovery trajectories across multiple fires occurring within the basin.

We will clarify this interpretation in the Section 2.3 (Line 167 as "Net ET reductions at the watershed scale were derived as area-weighted aggregates of ET responses from individual fires within each basin"

18. *L248: maybe consider replacing dry years with precipitation and temperature?*

Response: Thanks for the comment. We will replace the term 'dry year' with 'precipitation and temperature' in line 248

19. *L 250: Could you describe what you mean by ET playing a small but detectable role in driving FNF? Aren't interannual fluctuations in ET expected? Especially in a moisture limited system?*

Response: We thank the reviewer for this helpful comment. We agree that interannual fluctuations in ET are expected, particularly in moisture-limited systems where hydroclimatic variability strongly influences evapotranspiration. Our intention was not to imply that ET independently drives basin outflow. Rather, precipitation remains the dominant control on annual FNF, while variations in ET may contribute to smaller year-to-year differences in runoff through the basin water-balance relationship. To clarify this point, we revised the text to avoid the phrase "detectable response" and instead describe ET as playing a secondary role in modulating basin outflow within the water-balance framework.

Revisions: revise Line 248-250. " while annual basin outflow, represented by observed full natural flow, is primarily driven by precipitation and shows a small but detectable response to changes in ET (Figure 4)" to "Annual basin outflow, represented by observed full natural flow (FNF), remains primarily driven by precipitation, while variations in ET may contribute to smaller year-to-year differences in runoff through the basin water-balance relationship (Figure 4)."

20. *L 283: I know you mention this later, but a similar FNF exceedance seems common in the DAV basin. The sample size might be too small, but a basin by basin comparison might be interesting as well. I'm also curious how long this difference persists, is it significant only the first year following fire?*

Response: We thank the reviewer for this helpful observation. While the DAV basin does show several years in which FNF exceeds P − ET, the residual pattern varies through time and includes both positive and negative deviations (Figure 5a). This suggests that the DAV basin does not exhibit a consistently positive residual signal, but rather a variable residual structure comparable to that observed in other basins.

The comparison presented in Section 3.3 was designed as a pooled cross-basin analysis to evaluate whether years with substantial burned area tend to be associated with more negative residuals relative to non-fire years. Because the number of high-fire years identified by the basin-scale threshold is limited (n = 13) and these events are distributed across multiple basins, basin-specific statistical comparisons were not performed. To further evaluate the robustness of the classification of high-fire years, we will also conduct a brief sensitivity analysis of the burned-area threshold used to define high-fire years (3% of basin area) and report whether the main qualitative conclusions remain consistent across alternative thresholds.

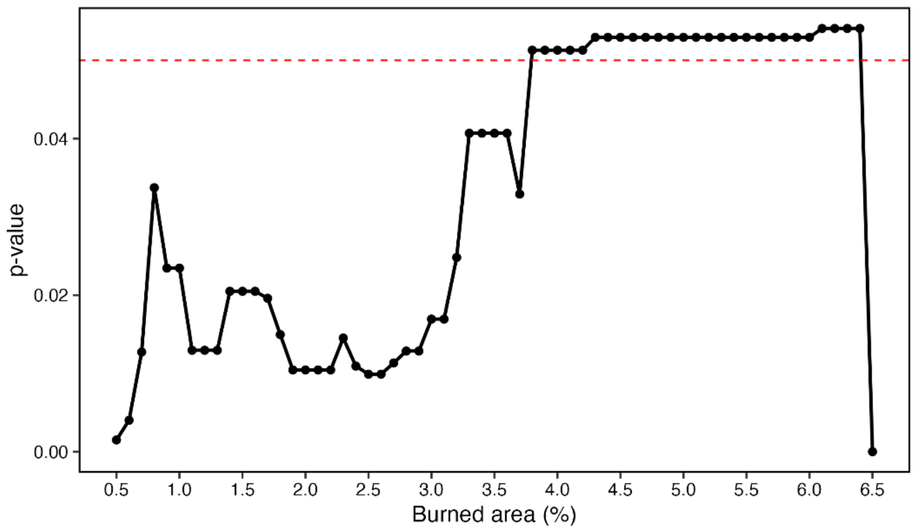

Figure S. The sensitivity analysis of the selection of the threshold for the percentage of area burned defined as high-fire years with Welch's two-sample t-test. The effect is statistically significant when the threshold ranges from 0.5% to 3.8%. However, when the threshold exceeds 3.8%, the number of remaining wildfire events becomes too small (fewer than three), resulting in insufficient sample size for robust statistical inference.

Revision: we will clarify this point in the revised manuscript (line 285-286) from "demonstrating that years with extensive fire activity systematically depart from hydrologic closure" to "indicating that years with extensive fire activity are associated with more negative residuals relative to non-fire years across basins, for the first year of fire occurrence." We will also add the text regarding the Methods and Results for the sensitivity analysis.

21. Figure 5: *Please add labels to the y axis.*

Response: Thanks for the reviewer's comment and we will refine the figure.

22. *L 300-301: Couldn't changing baseflow/groundwater contribution be another explanation?*

Response: We thank the reviewer for this helpful suggestion. We agree that changes in groundwater storage or baseflow contributions could also contribute to the observed residual structure between P − ET and FNF. Wildfire disturbance can alter infiltration, subsurface flow pathways, and groundwater recharge, which may influence the timing and magnitude of basin outflow. We will revise the text in the manuscript to acknowledge that groundwater and baseflow processes may also contribute to the apparent water-balance residuals. See the revision above for line 300-304.

23. *L 303: I understand the inclination to point to observational uncertainties, but couldn't a change in storage also explain FNF < P-ET?*

Response: Yes, the reviewer is correct that a positive residual (P−ET > FNF) can indeed reflect a storage increase. In the revised manuscript, we have made this possibility explicit. For example, when we discuss basins where P−ET > FNF, we now say this indicates a "persistent positive residual" rather than attributing it solely to "incomplete measurement." This rephrasing is intended to leave open the interpretation that some of the water might be retained in the basin (i.e., stored) instead of measured as flow. We also note in the Discussion that persistent residuals may reflect both transient storage changes and measurement biases. In short, the revised text acknowledges that storage change is a plausible explanation for FNF < P−ET, in line with the reviewer's suggestion.

Revisions: As shown above, in Results we will change line 303–305 to call P−ET > FNF a "persistent positive residual." In the Discussion (lines 471–473) we list "transient storage variations" as a contributor to residuals. These changes clarify that storage change is included in our interpretation of a positive residual

24. *L 307: Again I wonder if considering change in storage is appropriate? Or maybe even lagged response?*

Response: We agree that post-fire changes in infiltration and groundwater flow can introduce lags or transient storage effects. To address this, we have explicitly added mention of such processes in the manuscript. In the Discussion, we now point out that residuals may reflect "transient storage responses such as altered infiltration, groundwater recharge, or delayed subsurface drainage." This wording highlights that water may be temporarily stored or delayed in the system after a fire. These additions ensure that our interpretation of the annual residual includes both immediate imbalances and any delayed hydrologic response.

Revisions: We will insert language about delayed flow in the Discussion. For example, in Discussion (lines 471–473) we expanded the list of factors contributing to negative residuals to include "transient storage variations." We will also add the phrase "or delayed subsurface drainage" to indicate lagged runoff release.

25. L 308: The intercept parameter wasn't mentioned in the methods section. Can you help the reader interpret the physical implications of this parameter?

Response: We agree that the intercept parameter needed a clear introduction. In response, we have explicitly described the intercept as an additive offset. Specifically, in the revised Methods we state that the intercept can represent an unresolved flux component (such as an aggregated storage change ΔS or a constant measurement bias). In the context of our diagnostic framework, the intercept simply shifts the modeled runoff up or down, capturing any persistent difference that scaling alone cannot address. We mention this in the replacement text for lines 190–196 (see above), where we list examples of an additive structural offset. In the Discussion we further note that basins requiring an intercept likely have mixed error sources. Thus, the revised manuscript now consistently explains the intercept in both Methods and Results, addressing the reviewer's point.

Revisions: The new text in Sec. 2.3 (lines 190–196) explicitly mentions an additive offset and provides examples. This text now introduces the intercept parameter as part of our "diagnostic reconciliation parameters," fulfilling the need to define it in the Methods. The revisions for lines (306-310) mentioned above also correspond to this revision.

26. Discussion: *L 327: How was burn severity incorporated?*

Response: We have clarified in the Methods and Results that burn severity was quantified using MTBS burn severity classifications and summarized as the fraction of high-severity area within each fire perimeter. As the results show, the magnitude of first-year ET reduction increased with the fraction of high-severity burn area. This metric provides a quantitative indicator of disturbance intensity and is now explicitly referenced when discussing drivers of ET suppression and recovery.

27. L 336-347: This is where I think some further analysis of vegetation indices could help inform your recovery discussions. If they are in the SI it might be worth bringing them into the main ms.

Response: We agree that quantitative vegetation information strengthens interpretation of heterogeneous ET recovery. In the revision, we explicitly use the CECS vegetation composition fractions (tree/shrub/herbaceous/bare) shown in Figs. S1–S3 as quantitative proxies for postfire structural change and regrowth pathways, and we reference them directly when discussing recovery trajectories. These examples show that prolonged ET suppression coincides with sustained low tree fraction and increased shrub/herbaceous/bare fractions.

We will revise the corresponding text in the Methods and Results section, as described in the response to comment 3.

Revisions: For Discussion 4.1. Line 336: original 'Recovery trajectories varied across basins and reflected vegetation structure, climatic water availability, and disturbance history' will be revised into 'Recovery trajectories varied across basins and were influenced by climatic water availability and disturbance history, and coinciding with shifts in postfire vegetation composition (Figures S1–S3)".

28. L 339: Are you referring to individual burned areas within Pit or the basin total?

Response: Thank you for this clarification. The recovery trajectories described here refer to individual burned areas within each basin, rather than basin-wide averages. We will revise the text to clarify this point, and this revision also corresponds to the comment for line 239-240.

29. L 434: This was the first mention of reservoir operations. As I mentioned above, some explanation of how they are accounted for in the FNF data might be helpful.

Response: We thank the reviewer for pointing out that reservoir operations were mentioned without sufficient explanation. In the revised manuscript we clarify that reservoir operations do not directly affect the FNF values used in this study; rather, FNF represents a naturalized flow estimate intended to remove the effects of upstream storage regulation and diversions. Because this naturalization relies on operational records and adjustments, uncertainties in reservoir operations can propagate into the reconstructed FNF series.

Revisions: We therefore will revise the sentence at Line 434 to explicitly link reservoir operations to the uncertainty in the FNF naturalization procedure, from "Permeable lithology, faulting, and reservoir operations further obscure true basin outflow." to "Permeable lithology, faulting, and uncertainties associated with upstream reservoir operations involved in the naturalization of full natural flow (FNF) estimates may further obscure true basin outflow."

30. L 464: There are a few recent reviews of post-wildfire hydrologic modeling, it might be worth including a reference here.

Response: We will add references to recent reviews of post-wildfire hydrologic processes and modeling to provide broader context for line 464. including but not limited to Goeking & Tarboton (2020), the large-scale analysis by Hallema et al. (2018), and the statistical attribution framework of Beyene et al. (2021).

Revisions: Line 464, original 'Wildfires disrupt the balance among precipitation, ET, and basin outflow in ways that conventional hydrologic models do not fully capture.' will be revised into 'Although substantial progress has been made in understanding post-wildfire hydrologic responses (e.g., Goeking & Tarboton, 2020; Hallema et al., 2018; Beyene et al., 2021), this study integrated long-term satellite ET, precipitation, and full natural flow records trying to full capture the disruption of wildfire to the balance among precipitation, ET, and basin.