Federico Gómez-Delgado
Referee comment on the article:

Hydrological spatial coupling within an abandoned Mediterranean terraced micro catchment

By Jaume Company, Francisco Cuello-Llobell, Julián García-Comendador, Josep Fortesa, Miquel Tomàs-Burguera, Miquel Mir-Gual, Laura Turnbull, John Wainwright, Mariano Moreno-de-las-Heras, Jordi Cristóbal, Adolfo Calvo-Cases, and Joan Estrany

General comments

This paper investigates the conditions under which runoff (R) reaches the outlet of a small abandoned Mediterranean terraced micro catchment. Precipitation (P), soil moisture (SM) and outlet discharge (Q) were monitored continuously over four years (2021 to 2024), with SM recorded at paired depths in positions selected to represent contrasting hillslope and terrace settings. Events with and without outlet flow are compared through P and SM metrics, disaggregated by season, and the results are interpreted in terms of the spatial organisation of hydrological coupling and connectivity within the micro catchment.

The study rests on a sustained field effort in a demanding setting, and the resulting record is, in my view, its principal asset. Several elements deserve particular recognition. The decision to treat events without outlet flow as informative rather than as an absence of data is conceptually sound and well argued (lines 68 to 71): the observation that ignoring such events may bias the identification of controlling factors and limit the interpretation of state dependent responses is an interesting contribution, and it is worth emphasising further. Among the results presented, those concerning the seasonal differentiation of the controls on outlet response are the ones most directly supported by the observations. Antecedent SM discriminates between flow and non-flow events only during the wet season, and storm duration shows the same seasonal restriction (lines 336 to 338), while metrics describing wetting during the event itself (maximum SM and the change in SM) separate the two event types consistently across all seasons (lines 486 to 493). These findings are well grounded in the data, and I believe they point to where the scientific weight of this manuscript lies. Figure 5 conveys core aspects of the experiment effectively, and the clearer separation obtained for the change in SM at the deeper sensors is an interesting and potentially important detail.

I think the central difficulty concerns the distance between what was measured and what is concluded. Hydrological coupling and hydrological connectivity are the organising concepts of the paper: they appear in the title, structure the abstract, frame the objectives, are addressed in the discussion, and carry the conclusions. However, neither is measured or at least indirectly quantified. The instrumentation consists of vertical SM profiles at discrete points (with some horizontal setting that I do not fully understand), together with P and Q at the outlet. There are no observations of surface R at the scale of individual terraces or hillslope segments (facets), no measurements at any internal point of the drainage network, no monitoring of channels, tracks or wall openings, and no direct evidence of lateral flow, whether at the surface or in the subsurface.

Point vertical profiles are informative about local wetting, but they carry little information about the horizontal transfer of water between positions, which is precisely what coupling and connectivity are meant to describe. Statements such as the dynamic activation of connectivity (lines 476 to 479), the prolongation of functional connectivity (line 450), or the demonstration that R occurrence emerges from the spatial organisation of hydrological coupling (line 606) therefore rest on inference rather than on evidence. The Sediment Connectivity Index is introduced as the one quantitative descriptor of structural connectivity, yet it is presented only by reference, its values are never clearly shown, and it is not carried into any subsequent analysis.

A related concern is terminological. The manuscript employs hydrological coupling, hydrological connectivity, structural connectivity, functional connectivity, spatial hydrological coupling and, in the conclusions only, subsurface connectivity. The relationships among these terms are not established, and several appear to overlap substantially. None is given an operational definition, that is, a definition expressed in terms of a quantity that could be measured or computed from the data presented. The consequence is felt most acutely in the abstract, where expressions such as partially coupled conditions, well connected positions, disconnected mid terrace areas and connectivity aware approaches are used before any of them has been introduced. The abstract may therefore be difficult to interpret on a first reading. I believe that simplifying this vocabulary, ideally to a single concept with an explicit metric attached, would improve the paper considerably and would make the interpretation of the results much easier to follow.

The evidential base for the central comparison is thinner than the manuscript acknowledges. Over four years, R at the outlet was rare: 2023 yielded no R at all and suffered data loss, 2022 was dominated by essentially a single large event, and much of the total is attributable to the anomalously wet November 2021. Establishing a robust contrast between flow and non-flow events, and characterising the thresholds that separate them, is difficult with so few positive cases. Two further methodological issues compound this. First, the water level sensor was replaced in January 2023 with no reported overlap period, so no relationship between the two instruments was established. The 2021 and 2024 hydrographs behave very differently despite comparable cumulative P, and the runoff response in 2024 corresponds to the precipitation record more consistently than in 2021, where the accumulated curve shows an abrupt step without a clear counterpart in precipitation. On the evidence presented, the possibility that part of this difference is instrumental cannot be excluded. Second, the SM sensors appear to have been used with factory calibration and without site specific verification. Under those conditions the record supports relative comparisons and relative thresholds, but it does not support the identification of absolute SM thresholds. Consistency of sensor type and installation protocol across locations, as argued in lines 197 to 198, secures comparability among positions; it does not secure accuracy in absolute terms. The statistical strategy would also benefit from fuller reporting: the normality and variance tests invoked in lines 240 to 241 are never reported, nor is the false discovery rate correction, so it is not possible to verify which procedure was applied in each comparison.

Two further points concern the representativeness of the monitored positions and the terminology used for events. The instrumented locations are clustered in the southern portion of the catchment, so that most of the catchment area lacks observations. The manuscript acknowledges this only in part, and only in the Discussion (lines 613 to 620), by referring to the unmonitored western sector. This makes it difficult to sustain the claim that these positions capture structural variability across the hillslope and terrace system as a whole, or that they support inferences about connectivity at catchment scale. Separately, a comment on terminology. The manuscript refers throughout to flood events. This usage is consistent with the formal definition of flood in the International Glossary of Hydrology (WMO/UNESCO, 2011), which admits a relatively high flow measured by stage or discharge. In practice, however, the term is widely associated with inundation processes, including riverine flooding, flash flooding and urban flooding, and a reader may therefore expect something other than outlet discharge in a micro catchment of this size. Storm flow events, or alternatively runoff generating events, might convey the intended meaning with less ambiguity, with non-flow events retained as the complementary category. This is offered as a suggestion for clarity and is left to the authors' judgement.

Some aspects of the presentation would benefit from further attention. Lettering in essentially all figures is too small to be read comfortably; enlarging it would be particularly valuable in this manuscript, since the figures carry much of the burden of conveying the experimental design, and they are well placed to do so. Figure 1 and Figure 2 overlap somewhat in content, and the reader would be helped by a figure devoted specifically to the experimental design. Photographs of the instrumented positions, together with a scheme illustrating what is meant by disconnected, mid terrace and high connectivity settings, would be a valuable addition in this respect. The Discussion is long, and some material is revisited in more than one place, for instance in the overlap between Sections 4.1.1 and 4.2.3 on evapotranspiration and vegetation, and in the restatement at lines 510 to 512 of arguments made earlier. The authors may wish to consider a nested organisation, temporal scale followed by spatial scale, and to explore whether some discursive passages could be conveyed through tables or schematic summaries; either change would make this section easier to follow.

Finally, in my view, there are passages whose scope reaches somewhat further than the observations reported here can support. The land management discussion (lines 613 to 620) addresses terrace abandonment, degradation, afforestation and wildfire. These are important themes, but none of them was measured or tested in this study, and the discussion of them necessarily draws on the wider literature rather than on the results presented. The acknowledgement at line 623 that vegetation effects were not explicitly assessed is welcome, although the difficulty remains, since the surrounding text still attributes hydrological consequences to processes that were not observed. One option would be to set these themes aside; another would be to retain them within a clearly labelled section on directions for future research, where their status would be unambiguous. The manuscript would also benefit from a consolidated statement of its limitations, covering the small number of R events, the missing 2023 record, the sensor replacement without a reported overlap period, the calibration status of the SM probes, and the fact that observations are confined to a small number of points clustered in the southern part of the catchment, so that most of sectors and connectivity structures in the catchment area are unmonitored.

In summary, this is a carefully instrumented study containing several results that are genuinely supported by the data, particularly the seasonal differentiation of the controls on outlet response and the treatment of non-flow events as informative. Those results are, however, embedded in a conceptual framework that the measurements cannot sustain, and the conclusions are correspondingly overextended.

Considering the above, I would suggest the following:

1. Define the central concepts operationally. Reduce the number of overlapping terms, and for each retained term provide a definition expressed as a measurable or computable quantity. If the Sediment Connectivity Index is to serve as the descriptor of structural connectivity, introduce it earlier and with a proper description, present its values, and use it explicitly in the analysis. If it is not used analytically, its role should be reconsidered.
2. State explicit working hypotheses. The abstract and the Introduction do not currently present testable hypotheses, and their absence is felt throughout the Discussion, which is organised around themes rather than around propositions being tested.
3. Realign the claims with the evidence. Where coupling, connectivity or lateral flow activation are invoked, either present the supporting measurement or reformulate the statement as a hypothesis consistent with the observations. I would suggest concentrating the scientific contribution on the seasonal and event scale results that the data do support.
4. Address the data limitations explicitly. Report any comparison between the two water level sensors, or state clearly that no overlap exists and discuss the consequences for the homogeneity of the series. Clarify the calibration status of the SM sensors and, if no site-specific calibration was performed, reframe all thresholds as relative. Report the outcomes of the normality, variance and multiple comparison procedures.
5. Consider an alternative designation for the events analysed. The use of flood events is consistent with the definition of flood in the International Glossary of Hydrology (WMO/UNESCO, 2011), so this is a suggestion for clarity rather than a correction: since the term is widely associated in practice with inundation processes, “storm flow events” or “runoff generating events” might convey the intended meaning with less ambiguity, with non-flow events retained as the complementary category.
6. Restructure the Discussion, removing duplication and perhaps adopting a nested temporal and spatial organisation, and remove or clearly delimit the passages on vegetation, abandonment, afforestation and wildfire that are not supported by the observations.
7. Rework the figures. Increase lettering size throughout, merge or differentiate Figure 1 and Figure 2, add a dedicated figure of the experimental design, add photographs of the instrumented positions, and add a scheme clarifying the connectivity settings referred to in the text.

Specific comments

Abstract

L.22 to 23: in the statement that legacy features such as stone walls modify storage and flow pathways, the verb may understate the role you attribute to these features elsewhere. Walls appear to do more than modify storage: they determine, modulate, define or configure it, and they impose the geometry of the flow pathways. A stronger verb might better convey the active control you describe.

L.24: after the word "repellence" the construction becomes somewhat ambiguous. You might consider replacing the comma with a subordinating link, for example: "[…] and water repellence, hence lowering infiltration capacity, enhancing R generation […]". In the same sentence you refer to reorganising connectivity during a post fire window. What do you mean by connectivity here, and how is it defined? This is the first appearance of a term that carries much of the argument.

L.28: you introduce Q and use R as a separate quantity throughout. What exactly distinguishes the two in this study? R is not measured anywhere as a flux in its own right, and it is not defined, for example, as surface R generated within a plot. Short definitions, even in the abstract, would be welcome.

L.32: what are partially coupled conditions? The expression appears before any definition of coupling has been given.

L.33: in the reference to weak coupling between local SM and outlet Q, what does local SM denote? The SM at one sensor, at one position, an average across positions within a plot or similar land units?

L.34: similarly, what is meant by well-connected positions and by disconnected mid-terrace areas? A reader encountering these expressions for the first time may find lines 34 to 35 difficult to follow.

L.35: you state that R emerges from the interaction between P forcing, evolving SM and spatially variable structural connectivity. What of Q in this formulation? The relationship between R and Q remains unclear at this point.

L.36: the call for connectivity-aware, event-scale approaches builds on terminology that has not yet been introduced, which compounds the difficulty.

General point on the abstract: no working hypotheses are stated. Given that the paper is framed around the conditions under which coupling is or is not activated, explicit hypotheses would help the reader considerably. It would also help to introduce connectivity by relating it to flow paths or water pathways before the term is used analytically.

Introduction

L.64: where you refer to the transient filling of soil storage, I think "soil storage capacity" would be the more precise expression.

L.68 to 70: you argue that focusing on flow events risks overlooking the hydrological relevance of non-flow events, and that non-flow events inform the conditions under which spatial coupling fails to develop. Relevance in what sense? The point would repay being made explicit because, in my view, this is one of the stronger arguments in the paper.

L.70 to 71: the observation that ignoring these events may bias the identification of controlling factors and limit the interpretation of state dependent responses is well made, and I would encourage you to give it greater prominence.

L.73 to 74: you state that addressing this limitation is essential for advancing a probabilistic understanding of R occurrence. Why probabilistic specifically, and on what basis? An example or an indication of the intended application would clarify the ambition.

L.75 to 79: hydrological connectivity is presented as a useful framework, and structural and functional connectivity are then distinguished, although connectivity itself is not defined in objective, measurable terms. The relationship between hydrological connectivity and hydrological coupling is also left open: as used here they appear closely related, if not equivalent. It may be worth considering whether both are needed, and whether the set of terms could be simplified. Whichever terms are retained would benefit from being defined by reference to a metric that can be evaluated with the data available.

L.92 to 93: I would suggest "[…] and induced significant and time dependent changes […]", rather than "[…] and induced significant but time dependent changes […]". Note also that infiltration capacity is itself one of the soil properties listed, so the enumeration is not strictly parallel.

L.96 to 98: the statement that the interaction of legacy effects with structural features and with seasonally varying SM demands further investigation is well founded, and I agree with it.

Study area and methods

L.129 to 137: the description of the different land units would benefit greatly from photographs of each, and from a schematic representation of the whole system. As it stands the reader must reconstruct the setting from text alone.

L.136 to 137: you describe terraces as exerting a persistent ecogeomorphological and hydrological control on surface R, sediment redistribution and SM patterns, despite variable preservation. What is meant by preservation in this context, and what is the variability being referred to? The sense of the term is not entirely clear here.

L.140: the statement that these soils exhibit poorly developed profiles may not be readily meaningful to readers outside soil science. A short explanation in terms of observable or measurable attributes would help.

L.143 to 145: you report that terraced soils have higher carbonate concentrations and alkalinity but lower organic matter, cation exchange capacity and biological activity than hillslope soils. Why is this so? Some explanation of the process would be valuable.

L.152: you state that terraces support denser and more continuous vegetation. How is this reconciled with the lower organic matter content reported at lines 143 to 145? The two statements appear to be in tension, and a brief clarification would resolve the apparent contradiction.

Figure 1: the lettering throughout is small and not easy to read; Figure 1(b) is particularly demanding, since the panels are small and the labels smaller still. In Figure 1(c) there are marks that are not explained in the caption. In Figure 1(d) you identify the AEMET reference meteorological station, although your own experimental meteorological station does not appear anywhere in the figure, and its location is important. In the caption to Figure 1(b) the meteorological station is again referred to without specifying which one is meant: the AEMET reference station or your own. Showing both, and distinguishing them explicitly, would remove the ambiguity.

L.180: you report a total of eleven SM sensors, whereas Figure 2 shows eight probe locations. It would help to reconcile the two.

L.181 to 182: you state that pairs of sensors were installed horizontally at 10 and 50 cm depth in order to capture vertical variability. What does installed horizontally mean in this context? Depth cannot be interpreted horizontally, and the description is difficult to follow. A diagram of the installation would resolve this. The same wording recurs at lines 188 to 189.

L.193 to 196: you state that the monitoring locations were selected to capture structural variability within the hillslope and terrace system and the associated contrasts in structural hydrological connectivity. Two difficulties arise. First, the monitored positions are a small number of points clustered in the southern part of the catchment, so that the greater part of the catchment area (of which the western sector is one portion) carries no observations at all; these positions are therefore unlikely to represent the catchment as a whole. Second, and more fundamentally, these are vertical SM profiles at points. They provide no information on horizontal transfer of water, whether at the surface or in the subsurface, and no surface R, channel flow or flow along tracks is measured anywhere within the catchment. On what basis, then, can hydrological connectivity or coupling be assessed from these observations? If you consider that they can, the justification would be worth setting out explicitly in the manuscript.

L.194 to 196: the Sediment Connectivity Index is introduced here by citation alone, without description. Given that it is the only quantitative descriptor of structural connectivity in the paper, it would benefit from being introduced earlier and explained. It would also help to state what it is used for, since it does not appear to enter any subsequent analysis.

L.197 to 198: you state that the setup is appropriate for identifying SM thresholds associated with R generation. If the sensors were used with factory calibration and were not verified in the field (local soil types), the record would capture relative variability of SM, and substantial bias may be present in absolute terms. On that basis I would hesitate to describe the arrangement as appropriate for identifying absolute thresholds. Relative thresholds, understood as levels of the measured signal rather than as actual SM values, would be a defensible formulation, and making the distinction explicit would strengthen this passage. The argument that thresholds are derived consistently across locations using the same sensor type and installation protocol establishes comparability among positions. It does not, however, establish that the values correspond to actual SM, and so does not underwrite absolute thresholds.

L.204: you report that the sensor was replaced in January 2023. Was there an overlap period during which both instruments operated? Since the series combines measurements from the earlier probe and from the new pressure transducer, a relationship between the two would be valuable, and if no such comparison is available this would be worth stating as a limitation.

Figure 2: this figure overlaps substantially with Figure 1, and the two might either be merged or differentiated more clearly. In Figure 2(c), what do the orange lines or delineations represent? The lettering is again very small. The caption refers to disconnected, mid terrace and high connectivity settings, although these are not schematised anywhere, so the reader is left without a clear sense of what they denote. A dedicated diagram would help considerably.

L.72 and throughout: the use of flood event is consistent with the definition of flood given in the International Glossary of Hydrology (WMO/UNESCO, 2011). I would nonetheless invite you to consider an alternative, since in practice the term is generally associated with inundation processes, whereas what is observed here is outlet Q in a micro catchment. Storm flow event, runoff generating event or flow producing event might reduce the potential for ambiguity, with non-flow event retained for the complementary case. This is offered simply as a suggestion for your consideration.

L.240 to 241: you state that t tests or Mann Whitney U tests were applied depending on the results of the preliminary checks. You might consider adding Welch t tests for cases that satisfy normality but not homogeneity of variance, which would complete the strategy. More substantively, the results of the Shapiro Wilk and Levene tests are not reported, nor is the false discovery rate correction, so it is not possible to verify which procedure was applied in each comparison or how multiplicity was handled.

L.246 to 247: you restrict the analysis to timestamps at which Q was observed at the outlet. What motivated this choice? The evolution of SM prior to the response is precisely what would inform the analysis of antecedent conditions, and its exclusion appears to set aside information central to the stated objectives.

L.248: you refer to the seasonal evolution of SM conditions locally associated with ongoing Q. Q at any instant reflects processes and conditions that occurred earlier in time and elsewhere in the catchment, so a strictly contemporaneous association may not capture the relationship of interest.

Results

L.257 to 258: the total of 186 P events is restated here, having already been given in Section 2.4. Several such repetitions occur through the manuscript and could usefully be trimmed.

L.261: you report a contribution of 72.9 % of total R. Please give the corresponding depth in millimetres and clarify whether total R refers to that year or to the whole study period.

L.262: you indicate that the station shown in Figure 1(b) is the AEMET station. Where, then, is your own experimental station located, and could it be displayed as well? This information is important for interpreting the P record.

L.264: where you refer to data loss at the gauging station, do you mean the Q gauging station? A brief clarification would help.

L.265: for the figure of 8.2 % of total R, please clarify whether this refers to the entire measurement period. In the same line, “a R coefficient” should read “an R coefficient”, and I believe you mean an annual R coefficient of 0.1 %.

Figure 3(a): the column headed “Study period” for the R coefficient might be labelled “Annual mean for the study period”, since it is an average. No uncertainty is reported for the R coefficient; uncertainties can be estimated for percentages and would be informative here. The 2023 column would also benefit from a note indicating the data loss for that year. Finally, in the rows where you refer to average SM at 10 and 50 cm, “Catchment-average SM” would distinguish it more clearly from the local values discussed elsewhere.

Figure 3: lettering and numerals in panels (a), (b), (c) and (d) are very small and hard to read.

Figure 3(c) and Figure 3(f): cumulative P in 2021 and 2024 is similar in amount and comparable in the shape of its distribution, yet the R response differs drastically. The R response in 2024 corresponds to the P record more consistently than in 2021, where the accumulated curve shows an abrupt step without a clear counterpart in P. Given that the water level sensor was replaced between these years and that no overlap period is reported, an instrumental origin for part of this contrast is difficult to exclude on the evidence presented. Some indication of the homogeneity of the combined series would help before the difference is interpreted hydrologically.

Related to the above: 2023 records no R at all and 2022 is dominated by essentially a single event. With so few R events across the four years, establishing a robust contrast between flow and non-flow events, which is the central aim of the paper, becomes difficult. Stating this limitation openly, and considering its consequences for the conclusions, would strengthen the paper.

L.278: “high atmospheric demand” would read more precisely as “high atmospheric water demand”.

L.287 to 290: you report that daily reference evapotranspiration peaked during the drying down and dry seasons and was lowest during wet periods. Since reference evapotranspiration is computed from meteorological variables through a fixed formulation, this is the behaviour expected from a well measured meteorological series. Stating this explicitly would present the result as a consistency check rather than as a finding.

L.287 to 290: there appears to be a causality issue in the statement that total P was highest during wet periods and lowest during dry periods. If seasons are delimited on the basis of measured SM levels, then high P produces the high SM that defines the wet period, so the relationship as written inverts the direction of the reasoning.

L.290 to 292: the result that R occurrence was strongly seasonal, with the highest totals and coefficients in wet and wetting up seasons and minimal R in drying down and dry seasons, is as expected but nonetheless worth reporting, and it is soundly supported.

L.311 and L.317 to 318: in my view, this is one of the most interesting passages in the paper. The finding that events with and without outlet flow differ significantly in all P and SM metrics, yet overlap in antecedent SM at both depths, so that initial wetness alone does not determine outlet response, is a valuable result. The greater separation obtained for average SM, maximum SM and the change in SM is equally interesting. My reservation concerns the interpretation rather than the result: these are point observations of vertical wetting, and they do not in themselves demonstrate the activation of coupling or the transmission of water to the outlet. It would help either to justify the representativeness of the monitored positions explicitly, or to frame the interpretation as consistent with, rather than demonstrative of, the proposed mechanism.

Figure 5: the lettering is small, but the figure itself is very effective and summarises the core of the experiment well. It is notable that the change in SM discriminates more clearly at the deeper sensors than at the shallow ones, which is an interesting result and worth highlighting in the text. In the caption you note that symbol colour represents maximum 15-minute P intensity (IPmax15) for each event; this appears to be applied only to events with outlet flow. Extending it to non-flow events would be informative and would make the comparison more complete.

L.336 to 338: this result is among those most directly supported by the observations. Antecedent SM discriminates between event types only during the wet season, while storm duration shows the same seasonal restriction. This is the kind of finding that the observational design is well placed to support, and I would encourage you to give it a more central place in the scope of the paper, in preference to claims about coupling that the measurements cannot readily underwrite. One way to retain the concept of coupling on a firm evidential footing would be to define it as a binary metric at catchment scale: whether or not R is aggregated and routed, and reaches the outlet as Q. That is directly observable with your data, and it would allow the seasonal analysis to be framed in terms of coupling without requiring inferences about internal flow pathways that were not measured.

L.339: you might consider opening the sentence with a phrase such as "Over the observation period, only two flood events […]".

L.339 to 340: for the two events described here, it would be worth examining whether the response is consistent with infiltration-excess (Hortonian) overland flow, and stating the evidence either way.

L.342: I would write “based on the consistent discrimination” rather than “based on their consistent discrimination”. In the same line, you retain maximum SM and ΔSM: what led you to set average SM aside, given that at lines 317 to 318 it was also reported to separate the two event types?

Discussion

L.435: where you refer to “internal storage”, perhaps “subsurface storage”, “ground storage” or “soil storage” would each be more specific.

L.441 to 443: you attribute the likelihood of outlet response to an internal state shaped by SM, P and atmospheric [water] demand. The evidence presented does not seem to support atmospheric demand as a control on R occurrence. November is consistently among the months of lowest reference evapotranspiration, and November 2021 was anomalously wet and contributed disproportionately to total R. Atmospheric demand varies on a seasonal timescale, whereas R generation here responds to individual storms and, in particular, to their intensity: over the few days that separate antecedent wetting from an intense event, atmospheric demand does not change materially. That a small number of high magnitude storms dominates the R record itself suggests that atmospheric demand is not a determining factor at the event scale. Restricting its role to the seasonal conditioning of SM, would make this passage more robust.

L.450: what is meant by prolonging functional connectivity? Is the sense one of delay, or of continuity in time? The difficulty is compounded by the fact that functional connectivity is not measured; the index mentioned earlier is not presented with data or figures, so there is no quantity to which prolongation could refer.

L.466 to 469: the statements that R occurrence reflects the seasonal reorganisation of connectivity, and that structural connectivity defines the potential flow pathways, come close to the definitions that are missing earlier. They would be considerably more useful in the Introduction, expressed in terms of objective metrics, and accompanied by a statement of how structural connectivity relates to hydrological coupling.

L.476 to 479: this passage is not entirely clear to me and might benefit from reformulation. Beyond the wording, the substantive question is how a dynamic activation of connectivity can be concluded when connectivity is neither measured nor modelled in this study. What is the quantity, where is it observed, and by what method is it evaluated?

L.486: you report that total P and IPmax15 remained consistently associated with R occurrence. Why these two variables in particular, and what does this imply about the underlying process? The association with short duration intensity would be consistent with an infiltration excess (Hortonian) mechanism, whereas the emphasis elsewhere on progressive wetting and storage replenishment points towards a saturation excess (Dunne) mechanism. Clarifying which is operating, and on what evidence, would strengthen the paper considerably.

L.486 to 493: this passage contains genuinely valuable findings and is worth foregrounding. In summary: total IPmax15 minute intensity are associated with R occurrence across all seasons, identifying them as the primary triggers; maximum SM and ΔSM separate the two event types consistently throughout the year, indicating that both the peak wetness attained and the magnitude of wetting during the event are controls that operate year round; whereas antecedent SM discriminates only in the wet season, when soils approach saturation and, as you put it, relatively small P inputs could trigger outlet flow. Outside that season its predictive value declines and the response depends more strongly on P magnitude. I believe this is a coherent and well supported account, and it does not depend on the connectivity framework.

L.498: "lateral surface flow activation" would be preferable to "lateral flow activation", given that no evidence of subsurface lateral flow is presented. It is worth noting in any case that no lateral flux of either kind is measured directly in this study.

L.500 to 502: the association between large changes in SM and R occurrence during drying down and wetting up periods is interpreted as a fill and spill mechanism. This interpretation is plausible, and it would be worth developing further, including its relationship to saturation-excess (Dunne) generation, and indicating what evidence would be required to distinguish it from the alternatives.

L.504 to 505: the statement that deep soil wetting is necessary to activate lateral flow does not appear to be supported by measurement. ΔSM is a subsurface quantity, and unless conditions approach saturation, which would be an extreme case here, its relevance to the activation of lateral flow is not established by the data presented.

L.510 to 512: this material restates arguments made earlier in the Discussion. More generally, Section 4.2 is demanding to follow, and the Discussion as a whole revisits the same results and interpretations in more than one place. A nested organisation, treating the temporal scale and then the spatial scale, might be more efficient. Conveying some of the discursive text through tables or schematic representations would also help the reader considerably, given the length of this section.

L.599 to 601: the treatment of vegetation and its influence on SM depletion and recharge is hard to follow here, and it overlaps substantially with the evapotranspiration material in Section 4.1.1. It is not entirely clear how Section 4.2.3 differs in scope from Section 4.1.1, and the two might usefully be consolidated.

L.606: the statement that this study demonstrates that R occurrence in terraced Mediterranean systems emerges from the spatial organisation of hydrological coupling is, in my assessment, considerably stronger than the evidence allows. Given the small number of R events, the limitations of spatial sampling, and the absence of any direct measurement of coupling, a more cautious formulation would be better matched to the data.

L.613 to 620: the discussion of terrace abandonment, degradation, afforestation and wildfire, and of their effects on the balance between storage and transmission, addresses important themes, although none of them was measured or tested here, so the discussion necessarily draws on the wider literature rather than on the results presented. There are no data in this study against which these effects could be contrasted or tested.

L.623: the acknowledgement that vegetation effects were not explicitly assessed is welcome, although the difficulty identified above remains, since the surrounding text still attributes hydrological consequences to processes that were not observed. One option would be to set these themes aside; another would be to retain them within a clearly labelled section on directions for future research, where their status would be unambiguous.

L.624: the reference to the persistence of functional connectivity across hillslopes is presented as a finding, although functional connectivity is neither defined in objective terms nor measured or computed anywhere in the manuscript.

Conclusions

L.643 to 647: you conclude that R occurrence was strongly modulated by seasonal variations in SM, which controlled the extent of functional hydrological coupling across the micro catchment. Following your own definition of coupling, was this quantity measured in any way? If the intention is to refer to the spatial index mentioned in the methods, where is that analysis presented? As written, the conclusion asserts a result for which no supporting analysis appears in the paper.

L.650: subsurface connectivity appears here, so far as I can determine, for the first and only time in the manuscript, and it appears directly in the Conclusions. What is the definition of this term, and by what means was it verified?

L.652 to 656: the contrasts described among hillslope, high-connectivity terrace and disconnected mid-terrace positions are interesting, and the observation that similar event scale responses may arise from different underlying mechanisms is a valuable one. The difficulty is that the R generating mechanisms themselves are not clearly established by the evidence presented, so the attribution of the contrasts to distinct mechanisms remains a hypothesis rather than a conclusion.

Technical corrections

Below I recommend technical and typographical corrections to this manuscript, and some wording suggestions. In each case the wording of the manuscript is given in quotation marks, followed by the proposed change.

L.22 to 23: "legacy features such as stone walls modify storage and flow pathways". Consider replacing "modify" with "determine", "define" or "configure".

L.24: "and water repellence, and lower infiltration capacity, enhancing R generation". Consider: "and water repellence, hence lowering infiltration capacity, enhancing R generation".

L.64: "the transient filling of soil storage". Consider: "the transient filling of soil storage capacity".

L.92 to 93: "and induce significant but time dependent changes in vegetation cover, soil properties, and infiltration capacity". Consider: "and induced significant and time dependent changes in vegetation cover, soil properties, and infiltration capacity".

L.265: "with a R coefficient of 0.1 %". Should read "with an R coefficient of 0.1 %", and consider "with an annual R coefficient of 0.1 %".

L.278: "high atmospheric demand". Consider: "high atmospheric water demand".

L.339: consider opening the sentence with "Over the observation period, only two...", so that the small number of cases is stated plainly.

L.342: "based on their consistent discrimination". Consider: "based on the consistent discrimination".

L.435: "where internal storage imposes". Consider: "where subsurface storage imposes".

L.498: "critical SM conditions required for lateral flow activation". Consider: "critical SM conditions required for lateral surface flow activation".

Throughout: "Average SM at 10 and 50 cm". Where a catchment wide value is intended, consider "Catchment average SM", to distinguish it from the local values discussed elsewhere.

Figure 3(a): the column headed "study period" for the R coefficient. Consider "annual mean for the study period", since it is an average.

All figures: please consider increasing the size of lettering, axis labels and legends throughout, as several panels are not easy to read at the size at which they are reproduced.

Figure 1(c): the marks shown in the panel are not explained in the caption.

Figure 2(c): the orange lines or delineations shown in the panel are not explained in the caption.