Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2015-538-AC1, 2016

Hydrology and Earth System Sciences Discussions

Open Access

HESSD

Interactive comment

Interactive comment on "Flood risk reduction and flow buffering as ecosystem services: a flow persistence indicator for watershed health" *by* M. van Noordwijk et al.

M. van Noordwijk et al.

M.vanNoordwijk@cgiar.org

Received and published: 24 May 2016

Interactive comment on "Flood risk reduction and ïňĆow buffering as ecosystem services: a ïňĆow persistence indicator for watershed health" by M. van Noordwijk et al.

With both reviews showing interest in the line of argument, but indicating incomplete understanding of the various steps in the analysis, we will follow the suggestions of both reviewers to split the manuscript into two, the first part describing the theory (recursive flow models and its parameters), the second applications to a number of watersheds of contrasting characteristics. Working titles would be: "Flood risk reduction and ïňĆow buffering as ecosystem services: I. Theory on a ïňĆow persistence indicator for watershed health, II. Applications in four contrasting watersheds in Southeast Asia". We will benefit from the many specific comments by both reviewers in clarifying the overall flow of the argument and the details of its presentation.

1. Anonymous Referee #1 Received and published: 29 February 2016

R1.1 The paper presents an interesting attempt to develop a simple measure of ïňĆow persistence and address the issue of the ability to detect and attribute variations in daily riverïňĆows to the effects of land-cover change in large catchments (riverbasins), a well-known issue in hydrology. Author's response AR1.1: Thank you for the positive suggestions and interest. We realize, however, that we may need to be more clear on the primary aim: exploring the 'information content' of an empirically derived indicator of watershed health that does not require data other than the temporal pattern of river flow at a point of interest. R1.2 The authors then use a model for estimating ïňĆow persistence to try to demonstrate how difïňĄcult it can be to identify the effects of land-use change in four tropical catchments. Their illustration of the sample sizes needed to identify effects is also interesting but difïňĄcult to interpret as too little information is given on the catchments used as examples. AR1.2 We will have to provide further detail on the catchments, and the possible reasons for their differences in response to land cover change – see below where Tables 2-4 are discussed. R1.3 I am in full agreement that the issues of detectability and attribution are important for hydrologists to investigate because decision makers need evidence that catchment restoration can reduce ïňĆood risks and increase ïňĆow persistence by redirecting water from ïňĆow paths with rapid responses to rainfall to those with slow paths. This evidence can then be used to demonstrate that those beneïňĄts are being realised. AR1.3 Thanks, this is indeed the overall direction of the argument. R1.4 The authors also note that these effects are well documented in hillslope and small catchment studies but there is little evidence of such effects at in large catchments (river basins). AR1.4 Agreed R1.5 I have no problem with their argument that some form of measure of hydrological change is important, but (a) their method of deriving the persistence indicator and (b) of applying their model to the four catchments is not adequately explained. AR1.5 We will have to provide further detail then – as discussed below R1.6 Overall the paper needs substantial work and possibly reworking in to two papers. I have divided my review into two sections based on the two main components of the paper: (a) ïňĆow persistence and (b) ïňĆow change detection and attribution. There are some typographic errors but I have not gone into these as the paper needs rework. AR1.6 Thanks for the suggestion – as the method is not easily explained without practical examples, we would prefer to keep the full story in a single paper, but the suggested structure can help in the rewrite. R1.7 A) The one part of the paper addresses both ïňĆood risk and ïňĆow buffering by measuring aspects of the ïňĆow responsiveness of a catchment using a simple index (Fp) of the ïňĆow persistence. Flow persistence is deïňĄned as identical to the 'recession constant' (pg 7 line 7). However, I would argue that ïňĆow persistence is only half the picture, what is needed is actually a measure of ïňĆow responsiveness to rainfall because ïňĆooding can depend on how rapidly the ïňĆow increases versus factors that constrain that ïňĆow and cause water levels to rise. Flow responsiveness is also directly inïňĆuenced by antecedent wetness, rainfall event depth and duration, and other factors which they do discuss but do not seem to incorporate in their approach. AR1.7 We agree with the reviewer that flow persistence is focused on half the story (peak flows and subsequent decay), and that the speed at which peak flows are attained matters beyond what the peak itself is. Details on this speed will depend heavily on the specific space-time pattern of rainfall, and will require flow measurements at least hourly time-scale. In many cases daily records are the only thing that exists empirically, and we can't say much about this first part. Detailed models that use space-time pattern analysis can probably –provide reasonable inference, but will require many parameters, beyond what our 'parsimonious' targets allow. We will expand the discussion on these aspects.

R1.8 Maybe I missed it, but I did not ïňĄnd a clear statement that Fp is only being calculated for the descending limb of a hydrograph. Yet this must be so because the range for Fp is constrained to the range 0-1. I would argue that to understand ïňĆood risk you also need to measure the ascending limb of the ïňĆows (the rapidity of the rise in response to rainfall patterns). I think would be probably be necessary to have indices for both the ascending and the descending limb as they are rarely symmetrical. Zz AR1.8 On further thought, as may need to start from a more general case where Fp varies along the hydrological year, and only the dry season Fp equals the recession constant, as normally defined. We agree that the Fp parameter (at daily scale) alone is not sufficient to predict flood dynamics at shorter timespans (e.g. hours), as details of spatial and temporal storm patterns interact with characteristics of the streambed, beyond what Fp captures. However, the finding that the proportion of fresh rainfall (minus soil water storage capacity linked to preceding Et) that comes down as riverflow is (1-Fp), allows us to infer an important component of flood predictions: the peak daily flow volume (given rainfall). We agree that a further empirical parameter at hourly (or similar timescale) can add further value – but for the empirical data sets we used only daily records are available. Indeed we don't assume that the ascending and descending limbs are symmetrical. R1.9 Attempting to parameterise Fp for the various ïňĆow pathways seems to me to add complexity but not much insight given their focus on large catchments where the relative importance of different ïňĆow paths can vary considerably across space and temporally. AR1.9 This part is presented as an aid in the interpretation of the aggregate Fp, not as a way of empirically deriving it. This probably will have to be more clearly stated. R1.10 I did not ïňĄnd Figure 1 helpful as an illustration of the causal pathway or for placing the wide range of factors into their context. It has too much detail presented at the same level, is not adequately explained, and has too many terms that are not explained in the caption. A ïňĄgure should be able to stand on its own with its caption, but in this case it does not, even with an extensive caption. A part of the problem is the use of two sets of arrows with different meanings in ways that are confusing. A more common way of representing this kind of diagram is a ïňĆow chart where the factors that inïňĆuence (solid arrows) the (relative) magnitude of the water ïňĆows (hollow blue arrows) are represented in ways which make their role much clearer. I do not understand why rainfall is not made the start of the diagram and why the caption ends with number 0 rather than beginning there as one would intuitively expect. It is also not clear that land cover has various inïňĆuences at both the “plot” level (whatever that may be) AND at the hillslope/landscape level. It is also not clear to me why there is a blue water ïňĆow arrow directly from rainfall to #2 and to #3 without passing through the landscape (and why #2 is in brackets). Why are some of the arrows broad and others not? Is the “triangle” to the left of human population density an arrow? Why are human population density and topography (subsidence) linked? What is the relevance of subsidence? Why is topography placed here and not within the sets #0 and #1 given its importance as a factor in the generation and ïňĆow of surface and subsurface runoff at both plot and hillslope levels? AR1.10 We accept the limitations of Fig 1 in its current form and will improve it based on the comments made. It is meant at conceptual level, rather than as specification of a quantifiable model. Subsidence due to groundwater extraction in urban areas of high population density is a specific problem for a number of cities built on floodplains (such as Jakarta and Bangkok). It is a rather specific situation, but economically important flooding risks that were at first attributed to changes in upland land use are now understood to be generated in the urban areas. We’ll add a brief explanation and references to this point of detail. R1.11 For a paper that attempts to explain how land cover changes affects catchment ïňĆow responses I ïňĄnd it inexplicable that there is almost no reference to: (a) the very extensive body of hillslope hydrology research into ïňĆow pathways and the temporal effects of different water partitioning and surface/subsurface on ïňĆow response to rainfall inputs; and (b) how hillslope responses might scale up to larger parts of catchments and large catchments. Even the brief mention of the different ways in which overland ïňĆow can be generated (e.g. Hortonian versus variable (saturated) source areas) fails to cite the original research papers and the insights they provide in the catchment responsiveness. AR1.11 Our primary purpose with the paper is to evaluate the information content of a metric that is derived form observations of river flow alone – we are indeed aware of the large body of work (and various models) at hillslope scale, and will add some references, but the paper is not meant to be a review of all we know about ‘floods’, but to evaluate a very parsimonious model, that can in a single

Printer-friendly version

Discussion paper

index capture important first-order predictive value of the influence the watershed has, as modifier of stochastic rainfall. R1.12 Despite reading section 2 a few times I am still not entirely clear on the logic of the various deductions that are made about low ïňĆow, seasonality and the inïňĆuence of varying Fp on the form of the hydrograph. Perhaps this is because the text is not always very clear. For example, the authors present the following (page 7 line 19 onwards): "If we consider the sum of river ïňĆow over a sufïňĄciently long period, we can expect ΣQt to closely approximate ΣQt-1, and thus ΣQt =Fp ΣQt-1 + $\Sigma\varepsilon$ (equation 2) From this relationship we obtain a ïňĄrst way of estimating the Fp value if a complete hydrograph is available: Fp = 1 – $\Sigma\varepsilon$/ ΣQt (equation 3)" The only way I can derive equation 3 from 2 is to assume that ΣQt-1 = ΣQt and so ΣQt can be substituted for ΣQt-1. However, if this is so, then the only way equation 2 can hold is if Fp = 1 and $\Sigma\varepsilon$ = 0. If this is so, how can this relationship then be used to estimate Fp? Or am I missing something here? In section 2.4 I assume that a model with a set input of daily rainfall and ïňĆow responses to that rainfall was used to create the ascending ïňĆow limbs so that Fp values could be used to generate the descending ïňĆow limb? And so that Fp could be varied? I also had similar difïňĄculty in following parts of the methods section. For example, on pg 12 line 9 the term Qadd is abruptly introduced without an adequate explanation of its meaning. This is followed by the 'apparent Qadd' and Fp,try, again with no proper explanation. I should not have to go and ïňĄnd the paper cited (in fact a user manual) or to download the spreadsheet for an adequate explanation of the terms or to ïňĄnd a proper explanation of the FlowPer algorithm. AR1.12 This is a crucial point in the derivation,

If we indeed assume ΣQt = ΣQt-1, we obtain ΣQt =Fp ΣQt + $\Sigma\varepsilon$ and hence $\Sigma\varepsilon$ = (1 – Fp)(ΣP – ΣE). The easiest way to obtain this relationship at the level of annual sums, is to have it hold true at event level: ε = (1 – Fp)(P – Ex) with the caveat that the way ΣE is partitioned over terms calculated at each day with rain (ΣEx) may require some nuance (where not all antecedent ET is compensated in a single storm on soils that don't easily rewet, for example).

The key point, however, is that the stochastic (P – Ex) term is to be multiplied with (1 – Fp), which leads to direct influences on peak flows (if Fp does not vary with Q).

We will have to be more explicit in the way Qadd is calculated for each Fp,try value: Every pair (Qt, Qt+1) yields an estimate of Qadd derived from Qadd = Qt+1 – Fp,try Qt Each Fp,try value thus generates a frequency distribution of inferred Qadd values (some of which will be negative for relatively high Fp,try values), and the algorithm retains the value that minimizes Var(Qadd).

We will adjust the text to make this clearer, while the spreadsheet version is available for anybody who wants to try it on empirical data. R1.13 B) The second part of the paper deals with the application of the GenRiver model for assessing the impacts of land cover on river ïňĆow and its attribution and detectability. The entire model and its application is only introduced in the methods but its structure and use should really be described already in the introduction. The model is said to be spatially explicit but it is not clear how that is realised in practice (i.e. it a distributed model?). AR1.13 The use of the model to explore Fp derived from hydrographs for scenarios other than current land use should indeed be more explicitly announced in the introduction – but it remains a 'tool' for exploring Fp properties of hydrographs, rather than being a focus on its own. If we were to separate the manuscript into two, the GenRiver part would probably have to come first, so that the Fp discussion can use its results. With a full specification of the model available for who wants to get into the detail, we will increase the description of key features that likely influence the results obtained. R1.14 Tables 1-4 do not provide an adequate description of the study catchments – what does dominant land cover mean? Although the authors note the importance of knowing what changes where in a catchment in relation to ïňĆow paths and times and attributing responses to changes and other factors, Table 1 does not give any indication even of how much of each land cover there is the baseline situation. Why not provide summaries or maps? Nor are we given information on where, in relation to this baseline state, the changes in land cover are made for the different scenarios. Why not provide a summary or a map? In Table 4 there is a repeated use of “some” in describing the changes made. This to me is not acceptable. AR1.14 Thanks for the suggestion – we will add a map that visualizes the baseline situation, and a table of what the scenario’s mean in local context. Supplementary material that gives all the requested details, while maintaining the overall flow of the current manuscript is probably our preference at this stage. A point of warning may be needed here, as the case study catchments have not been subject to a multi-year intensive measurement campaign of the type that reviewer probably finds necessary to fully trust results. We do provide the level. Of correspondence between Fp‘s derived from measured and modelled hydrographs, and on that basis present the further model results as illustrations of what can be expected, rather than as statements of fact. R1.15 We are not given an adequate explanation of how the single values of each of the 13 parameters of the GenRiver (Table 2) were obtained. Those parameter values are all ones that would vary a great deal spatially and with different land cover types (e.g. interception), but only a single value is given with no indication of their variability in the study catchments or how representative each value is. Providing deïňĄnitions of the terms in a user manual the reader would have to look up is simply not acceptable. Table 3 also gives values for three important parameters for each of the land cover types with no explanation of what their sources and ranges are (BTW surely interception [Table 2] differs between forest and annual crops and so island cover speciïňĄc?). Table3 also introduces the term relative drought threshold with no explanation of what it means and how the model uses it. The legitimacy, accuracy and representativeness of these values, together with the land cover changes, are critical to our conïňĄdence in the model outputs and thus in the analysis of the detectability and attribution of the changes in ïňĆows to changes in land cover. A study should be repeatable and this hypothetical modelling exercise certainly is not given the information included in the paper. In summary I am not entirely sure what to recommend overall. The idea of deriving a simple but robust measure of ïňĆow change (i.e. ïňĆow responsiveness) which can be causally related to land cover changes is sound, and necessary. AR1.15 We will provide full specification of the parameter val-

Printer-friendly version

Discussion paper

ues used for the calculations (and store runs in a data depository), while the model is freely downloadable, so we'll meet reasonable standards of repeatability – but of course further tests of the Fp summary statistic on hydrographs obtained with other models for other situations are likely needed before this method can be more widely accepted by the community. We hope that the discussion gives a fair assessment of the level of evidence, avoiding an oversell. R1.16 Flow persistence (Fp, recession) is an interesting measure and can be related to changes in the relative importance of different water ïňĆow paths, but it is also evident that it is not straightforward to derive and could be masked by the effects of location and catchment heterogeneity. I do think that a measure of the ïňĆow recession is not sufïňĄcient, the nature of the whole response to rainfall needs to be assessed for ïňĆood risk. The ïňĆow persistence component of the paper needs careful thought to make sure that the measure(s) are clearly and thoroughly explained. Even so, I still am left with the question of whether a simpler approach would not be to examine the slopes of the ïňĆow recession curves (in relation to rainfall event sizes and sequences [antecedent conditions]) for possible shifts due to land cover changes. Alternatively, using shifts in the ïňĆow duration curves as measures of changes in the relative importance of ïňĆow pathways, as has been done elsewhere, would be more effective and understandable. AR1.16 It seems that reviewer presents a multi-parameter description as 'simpler' than our single metric? Again, our focus is to assess the information content of a metric of flow predictability that appears to align well with the way downstream observers describe and experience changes in the conditions in upper watersheds. If our primary focus would be to assess the effects of land use change on flood risks as such, for any of the catchments studied or for the wider geographic domain in its totality, we would probably embark on further data collection... A key point here is that in many situations (historical) rainfall data are inadequate – at best a few point records are available, but no spatially weighted average, and there are many degrees of freedom in getting reasonable answers for wrong reasons in multiparameter models. R1.17 Another alternative would be to use the relationship between rainfall event sizes and sequences (e.g. antecedent wetness) and ïňĆow response to those events and sequences to infer changes. The modelling component needs a lot more information to back up the chosen parameter values for both the hydrological (Table 2) and land cover-speciïňĄc values, as well as speciïňĄc information on the extent of the land cover changes and their spatial conïňĄgurations. It also needs to provide information on how well the outputs it generates for the different land cover types compare with the ïňĄndings of other studies (i.e. how well does the model perform). Overall I need more information on the model structure and setup to interpret how well it performs in this application. This would require expanding the paper substantially. AR1.17 We agree that the current description of the GenRiver model is not sufficient to fully compare its performance with the substantial range of other models, most of which require considerably more parameters. We do believe that the exploration of how hydrographs for alternative land use scenarios can be summarized in the Fp statistic adds value to our discussion on what we can and cannot expect of this parameter, and how it can play a role as 'metric of watershed quality', as the title emphasizes. R1.18 Overall, my conclusion is that perhaps this paper attempts to cover too much ground and should be two papers: - One on the issue of catchment responsiveness to rainfall as a measure of land cover change, including ïňĆow persistence - One on modelling of the effects of land cover change on river ïňĆow responses and the difïňĄculties of detecting and attributing changes in ïňĆow responsiveness to changes in land cover (and relating this back to the changes in the relative importance of ïňĆow paths linked to the changes in land cover). AR1.18 Maybe reviewer has read more into the case studies than we intended – we agree that more detail (in an appendix) of the model results is needed, while a more comprehensive assessment of what land use change can mean in the specific locations waits further study.

Anonymous Referee #2 Received and published: 25 April 2016 R2.1. Summary Overall, I think that this study addresses a critical knowledge gap with important implications. However, the conceptual foundation regarding watershed health and ïňĆow predictability requires a closer examination. The derivation of Fp and the process used to create the GenRiver model parameters needs more discussion, clariïňĄcation, and justiïňĄcation if the reader is to have any conïňĄdence in the results. I think that if the authors were to accomplish these revisions then the paper would simply be too long and cover too much ground. Breaking the research into two separate papers is probably a better course with one focusing more on the conceptualization and creation of the Fp term and one on the application of it within the GenRiver framework.

AR2.1 Thanks for the interest, we will follow the suggestion to split the manuscript into two parts “I. Theory, II. Applications”, and appreciate the detailed suggestions and comments.

R2.2. As I understand the article, the authors are attempting to develop a single measure of watershed health called ‘ïňĆow persistence’ (Fp). This Fp parameter measures the volatility of daily river ïňĆow in response to land cover change within large catchments. AR2.2 Maybe the manuscript needs to be more clear in the steps involved: Fp is a measure of ‘volatility’ (or its complement), that can be used to quantify one aspect of the impacts, at multiple scales, of land cover change. R2.3. One of the key objectives of the study is to determine the value of speciïňĄc land cover types in terms of ïňĆood mitigation. The study itself is broken down into two phases: (1) the derivation of a river ïňĆow algorithm, and (2) the application of the algorithm within four watersheds with different rainfall and land cover characteristics. AR2.3 In the current manuscript the Fp definition in terms of a recursive river flow model is presented as step 0, before the empirical steps on how actual (or simulated) flow data can be used to derive an estimate of Fp (step 1), illustrated with applications in four case studies (step 2). R2.4. The key points that need to be addressed include: 1. A better justiïňĄcation that ïňĆow predictability does in fact correspond with watershed health; 2. A better explanation of the ïňĆow persistence derivation; and 3. A much more thorough explanation of the Fp algorithm application within the four catchments. AR2.4 We thank reviewer for the positive suggestions and will try to clarify these points in a resubmission. “does in fact correspond with watershed health”, is however a complex question, as “watershed health” hasn’t been satisfactorily quantified in absolute terms anywhere. What we claim is that changes in Fp, towards lower values match “degradation” and towards higher values match “restoration”, from a downstream perspective. The metric matches a common way of describing degradation as loss of predictability. In unpacking the concept, however, we find that the specific pattern of rainfall in a given year interacts with the condition of the watershed in generating a river flow pattern, as captured in Fp. Our conclusion that multiple years of “paired catchment” type data (different watershed conditions, same rainfall pattern) are needed to be reject null-hypotheses that land cover use effects are neutral. So – we qualify our claim that Fp is an “indicator”, not a “measure” of watershed health. But, so far we don’t know of a better simple indicator. We accept that steps 2 and 3 need improvement in terms of clarity and documentation. Splitting the manuscript in two parts, as suggested by reviewer 1 will help us do so.

R2.5. The study addresses a signiïňĄcant point of contention in the literature: the in-ïňĆuence of land cover (particularly forests) and ïňĆooding at the watershed scale. If the Fp model is properly justiïňĄed and performs adequately, then it would undoubtedly increase our understanding of the linkages between land cover and ïňĆood risk. The beneïňĄts of such an approach are clear as it would make for a much more parsimonious model of river ïňĆow that would greatly enhance the monitoring and prioritization of speciïňĄc landscape management decisions. AR2.5 We thank reviewer for the interest – indeed our target is a key parameter for a parsimonious model that can communicate key functional property of the way a watershed functions, given variable rainfall. R2.6. That said, the paper requires substantial work to adequately address the points above and may need to be split into two separate papers. I will address the three points I mentioned above in greater detail below. There are quite a few of typographical and grammatical errors in the paper, but I will leave these alone for now as the paper requires substantial work. AR2.6 As reviewer 1 also suggested splitting the paper along similar lines, we will follow this suggestion. R2.7. Point 1 In the paper the authors use persistence, predictability, and watershed health interchangeably. AR2.7 We aim to use “persistence” as a first descriptor of what Fp measures (the part of todays flow that can be counted on for tomorrow, regardless of additional precipitation); 1-Fp controls the part of new rainfall that adds to the stream, and as such controls the predictability of overall flow; the way that Fp is an indicator of watershed health is at a further level of interpretation. We will scrutinize which words are used where in the process. R2.8. One of the key assumptions of the paper and previous watershed rehabilitation efforts is that increasing the presence of natural land covers (particularly forests and wetlands) will restore the natural ïňĆood regime with lower peak ïňĆows and less damaging ïňĆood events. AR2.8 Actually we take this as testable hypothesis rather than a priori assumption R2.9. The authors do a good job of documenting previous studies that have illustrated the complexity of the linkage between reforestation and river ïňĆows. Moreover, the ability of wetlands and riparian forests to absorb rainfall, slow streamïňĆow, and attenuate peakïňĆows is supported by many studies and is fairly well understood. However, these types of stream corridor ecosystems also require a particular type of disturbance regime that creates opportunities for species speciïňĄc recruitment processes and establishes landscape and topographic heterogeneity that are critical components of watershed health. These disturbance regimes are often characterized by variable ïňĆow patterns with various ïňĆood magnitudes required for speciïňĄc types of ecosystem level processes. Most efforts to create a stable and predictable ïňĆood regime have been anthropogenic in origin through engineered based interventions like dams, and retention and detention ponds, which are also some of the primary drivers behind the degradation of watershed health. A perfectly stable ïňĆow regime could, theoretically, be established by a highly integrated system of engineered solutions (albeit until they are either overwhelmed by a storm or undermined by system failure) within a very ecologically degraded watershed. Likewise, there could, theoretically, be examples of ecologically well connected and healthy watersheds with fairly volatile and unpredictable daily ïňĆow regimes. To overcome this, the authors need to discuss what exactly watershed health means and whether or not a predictable ïňĆow regime is the product of an ecosystem service. I could see an argument in which the shape of a storm speciïňĄc hydrograph within a healthy watershed should be fairly predictable, however, to study this would preclude the advantages proposed by this paper (i.e. the application of the Fp algorithm in data sparse regions). I also agree that more human development and less natural systems generally leads to more ïňĆashy river ïňĆows as a result of decreased buffering capacity, however this study examines ïňĆow rates at daily intervals which washes out the ability to assess this linkage. Or maybe I'm missing something. The authors do point out later in the paper that Fp of zero (i.e. low predictability) would be the result of erratic rainfall (page 7 ïňĄrst paragraph). This is somewhat confusing because most of the introduction and discussion is focused on using Fp as a way to summarize "complex land use mosaics". Two paragraphs later the authors state "a decrease of Fp indicates watershed degradation." So how much of the decrease in Fp is explained by watershed degradation as opposed to just more erratic rainfall? I know the authors say that the GenRiver model is spatially explicit, but this is a little vague. Does this mean that spatial autocorrelation in precipitation is controlled for or that the model is spatially distributed? I understand that to have an Fp equal to 0 would require erratic rainfall, but the authors need to be consistent when describing what proxy measurements that Fp is suitable for. AR2.9 Thanks for these thoughts. We agree that stabilizing riverflow beyond what the "natural condition Fp" indicates is not without risk for the ecological functions of the river and its biota, and will add some comments to this effect. Quantitative estimates of Fp derived from a limited sampling period do depend on specifics of the actual rainfall. As rainfall is not generally known at the required spatial and temporal resolution to disentangle these relationships, we have to accept that Fp does not only respond to the land use mosaic, but also to rainfall. We'll check whether this can be stated more clearly upfront. R2.10. Figure 1 gets at the interconnection between many different elements that inïňĆuence the hydrological cycle, and the authors break up the components into ecosystem structure, function, and human land use/perceived ecosystem service. However, I ïňĄnd the ïňĄgure difïňĄcult to navigate and poorly described in the study. The different color arrows with different shades and outlines is one of main culprits of the confusion. The graphic needs to be simpliïňĄed, it should probably start with rainfall, and terms like "plot-level" should either be deïňĄned or removed. AR2.10 Thanks, we will try to simplify the graphic R2.11. Point2. An Fp value ranging from 0-1 essentially represents the buffering capacity of the watershed, but there are also characteristics that inïňĆuence how rapidly water reaches the stream. In this sense, Fp is only represents half of the picture. AR2.11 We don't quite understand the reviewer here: Fp quantifies a property of flow dynamics, and integrates over multiple aspects of the subsystems along the rainfall-vegetation-soil-streams pathway. The Fp concept is close to 'buffering' – but this is itself a scale-dependent concept (variability of outflow relative to variability of inflow, depending on time sacale and system boundaries). R2.12. The authors go on to create separate Fp's for each ïňĆow pathway, which is probably necessary for large catchments as each ïňĆow pathway likely do have large inïňĆuences over space and time. However, this seems to be overcomplicating a model that was originally being created out of need for greater parsimony. If these pathway speciïňĄc indices are necessary, then more discussion and justiïňĄcation is required in the text. AR2.12 We offer the weighted average of pathway-specific Fp values as an additional way of interpreting results, hoping that for some readers this will help understand what Fp is. It is not an essential component of the main argument, and we will state this more clearly. R2.13. The authors use vague language that needs some more clariïňĄcation. Line 19 on page 7 contains "ïňĆow over a sufïňĄciently long period". What is a sufïňĄciently long period? Wouldn't a sufïňĄciently long period wash out the "ïňĆashy" ïňĆuctuations that the authors are trying to explain with changing land cover/watershed degradation? If the 'sufïňĄciently long period' is preventing what the study is attempting to explain, then I do not see how equation three could be derived. Maybe I'm missing something, but wouldn't the stochastic term represent all unexplained variations in the predicted river ïňĆow? Line 28 on page 7 explains that the stochastic term is equal to the sum of peak ïňĆows. Couldn't other unpredicted river ïňĆows have other anthropogenic origins that contribute to the river ïňĆow stochasticity (e.g. dam operations/failure, irrigation, urban water use, etc.)? The authors also mention new variables like Qadd and Fp,try without adequately discussing what they actually represent. AR2.13 We obviously created some confusion here and will try to more clearly separate the recursive model of river flow (and its associated Fp) as a “principle”, before getting into issues of data and empirical estimates of Fp in defined data sets.

R2.14. Point 3 Figures 2-9 were very readable and for the most part stand on their own, however the table were poorly formatted and vague. Table 1 does not provide land cover proportions by land cover type (other than forest). Percent developed land, existing ïňĆood control infrastructure, and population would all be helpful information. Not sure what ‘dominant land cover’ means. Do these watersheds have a history of damaging ïňĆoods? AR2.14 Thanks, we will improve the presentation of Table 1 and define the terms used. R2.15. Why were the parameters in table 2 chosen? The authors do not provide an adequate discussion of how these parameter values were estimated. Why were the defaults in GenRiver used for each of the land cover types in table 3? What process or methodology did GenRiver use to estimate these values? What does ‘relative drought threshold’ mean? AR2.15 As also commented on by reviewer 1, the reference to the existing manual of the GenRiver model is clearly not sufficient here, and the model will have to be summarized in its key equations and assumptions, before we can use it to illustrate how Fp can be interpreted for alternative land cover scenarios.

R2.16. The use of the word ‘some’ in table 4 is simply too vague when describing the scenarios. The reader is left wondering what the magnitude of change would be within each of these scenarios. All of the information in tables 1-4 are critical components of the GenRiver model. The legitimacy and accuracy of the model is weak without proper documentation and justiïňĄcation of the underlying model parameters. The authors must correct this if we are to have any conïňĄdence in the results. AR2.16 We agree that further detail is needed here – however, the primary target here is “sensitivity analysis” of the way Fp will respond to changes within a plausible range, not to get into detail on any of the watersheds as such. R2.17. Table 5 and its corresponding discussion regarding the sample sizes required to reject a null hypothesis is interesting, but not enough information was given to make this section clear. The methodology is clear enough, but the implications were not really discussed. The statement beginning on line 21 on page 14: "In practice, that means that empirical evidence that survives statistical tests will not emerge, even though effects on watershed health are real" is vague and needs some more clariïňĄcation. AR2.17 Thanks for the interest in these results – we will bring in some quantitative terms in these overall, qualitative conclusions R2.18. Lastly, in table 6, the authors provide broken links to the detailed reports of rainfall and river ïňĆow data. Moreover, there is very little discussion on the accuracy and metadata of each of these data sources, all of which have different origins. AR2.18 Unfortunately the website to which links are provided has been off-line for some time, we will double check all links function again in a resubmitted manuscript.

Please also note the supplement to this comment:
http://www.hydrol-earth-syst-sci-discuss.net/hess-2015-538/hess-2015-538-AC1-supplement.pdf

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2015-538, 2016.

Printer-friendly version

Discussion paper