

1 **HESS-2015 -538.** Author's response to the third reviewer, 31 Jan 2017

2 Dear editor

3 We have taken the advice and restructured the manuscript to strengthen the overall storyline, which  
4 is focused on how flood risk reduction and flow buffering as ecosystem service can be quantified in  
5 data-scarce environments, by comparison with flashiness index and base flow indicators. We have  
6 revised the figures and captions throughout.

| Reviewer  | Authors   |
|---|---|
| I find the concept of this paper to be really interesting, but feel that the papers are still poorly organized, figures and tables need better labeling and updating, and that the authors still need to make a better case for what $F_p$ adds to a given analysis (both in the text and through benchmarking in Part II).   | Thanks for the interest – on re-reading the manuscript we realize that by elaborating points in response to various comments, the main storyline has indeed been lost at a number of places. We have overhauled the text, revised figures and captions and added a table that compares $F_p$ and flashiness index in part I, and an overall comparison table in the discussion of part II.  |
| <b>Major points are summarized as follows:</b> <ul style="list-style-type: none"><li>The connection from health, to floods, to <math>F_p</math> I think gets lost amongst both papers. I would almost change all references from health to refer to <math>F_p</math> as an indicator of alteration. The threat of flooding differs based on your location within a catchment (upstream or downstream) as well as many different watershed characteristics, including human impact. Also, as stated in a comment below, floods are healthy, so attributing their increase solely to human impact is perhaps missing some points. I think reframing this as an indicator of alteration could be useful.</li></ul> | We have further downplayed the 'health' aspect and refocused on floods as the primary 'salient' point of attention. The word 'health' is still found in the text, in reference to existing literature.<br>We fully agree with emphasis of change in $F_p$ rather than $F_p$ itself as primary indicator -- this was indicated before, but is now the primary conclusion.<br>There was some reference to the relevance of floods for downstream biota, and there is now a full paragraph in the discussion on this. We don't quite understand what you mean by "so attributing their increase solely to human impact".   |
| <ul style="list-style-type: none"><li>I find it very striking that two large bodies of literature are missing from the paper – the work of Leroy Poff (Natural Flow Regime, etc), and reference to many of the (small) benchmark catchment studies in the US that linked forest harvest to streamflow responses – these would be interesting candidates for further testing of <math>F_p</math>. I understand that you are focusing on floods at the larger scale, but given the connections you draw between forest harvest/recovery and watershed response, it would be remiss to not reference these landmark papers.</li></ul>  | There luckily are many more large bodies of literature missing from the paper... The concept of 'natural flow regimes' gets some mention. The older work in the US and Europe on changes in forest catchments gets mentioned primarily through the excellent reviews that have summarized the results. It would of course be interesting if somebody can test the $F_p$ metric on the existing US data, but we don't have access to these data. Indeed, it is the larger scale floods that are more contested (and potential involve larger values). Meanwhile, a major difference between temperate and tropical watersheds in the absence of snowpacks and snowmelt in the latter now gets mentioned several times. |

|   |   |
|---|---|
| <ul style="list-style-type: none"> <li>Floods are a natural part of the flow regime – while Part II demonstrates that the deviation from these natural flow regimes through time is really what you are looking for, I think this theoretical approach would be worth stating and referencing upfront throughout Part I. Just like most of the indicators of hydrologic alteration literature, you are interested in a deviation from average.</li> </ul>   | <p>We have done so – it is explicit in the terms ‘degradation’ and ‘restoration’ that a change over time is the core interest. Whether it is ‘deviation from the average’ or ‘deviation from what has been the past distribution’ is an issue for further debate.</p>   |
| <ul style="list-style-type: none"> <li>I am left to think that the <math>F_p</math> metric may be worthwhile for comparing within a catchment through time, but may not be effective at comparing across catchments, due to their heterogeneity in all of the effects summarized in Figure 1. It would be nice to include some discussion of the possible limitations of <math>F_p</math>.</li> </ul>   | <p>We have further clarified that the interannual variation in <math>F_p</math> versus flashiness index as well as ‘base flow’ has a pattern that differs between catchments, even with the small set (four) of examples discussed here.</p> <p>We have added a table 1 that provides strengths and weaknesses of both indicators.</p>  |
| <ul style="list-style-type: none"> <li>From Figure 2 (Part I), I am really left thinking that this indicator is a measure of flashiness</li> </ul>  | <p>In a sense yes, but it also is a metric for base flow and it does correlate with the R-B Flashiness index, but it is not equal to it</p>   |
| <ul style="list-style-type: none"> <li>Part of the utility of this indicator over flashiness could be your ability to partition it between wet and dry seasons and different flow pathways. At first, the description in Part I was lost on me, but I understood it after reading Part 2. I would better organize this description to frame <math>F_p</math> as a flexible indicator that spans the empirical and modeling realms. Currently, this is somewhat described in Part I, but highlighting the different ways it could be used by a Figure, or by organizing the text better would be really useful.</li> </ul> | <p>Thanks for the suggestion, we hope that the current text makes these points more clear.</p> <p>We have added a new Figure 1 that spells out the criteria that a ‘metric’ must meet in order to find its place in the applied field of discussion between natural resource managers, the wider public and local stakeholders. These lead to the 7 questions framed at the end of the introduction, and used to structure the discussion of part I.</p>            |
| <ul style="list-style-type: none"> <li>As a new reviewer coming to this paper with fresh eyes, I found that many of the comments from previous reviewers were not addressed, especially points of ambiguity in the text .e.g, sufficiently long period (line 224), and the wording throughout section 2, which I found difficult to follow. It is unclear if changes were made to Figure 1, despite comments from reviewers to this effect. I also find this figure difficult to dissect.</li> </ul>  | <p>Where you have been non-ambiguous in these comments we have addressed them. We don’t quite understand what your issue is in section 2. This is indeed a technical account with rather precise wording.</p> <p>Figure 1 has certainly been changed from the earlier version in the HESS_D manuscript and the changes were appreciated by the one reviewer who provided suggestions. We have made further changes now, to connect it with the terms in Fig. 1.</p> |

| Minor points:  |  |
|--|--|
| -I found interpreting the figures in part I of the paper based on their captions alone to be very difficult, especially Figures 2, 3 and 8 (and their formatting). Please revise these captions.   | We have revised the captions – but the sequence of showing Fig 3 (in the ‘methods’) before figures 4 and 5 (in the results) is suboptimal for ease of understanding. Hopefully readers will refer to the M&M section once they are interested in details, and will first look at ‘results’   |
| -All figures would benefit from some revising, especially using subscripts for the p in Fp, sizing font and axes text to the same size, labeling x and y axes, better labeling of Figures (Figure 4, Part II especially), and construction of figures (Figure 4, Part II – lines should not be used to connect different values – this implies a continuity, but these are different catchments) -I found the results/discussion of part I to be haphazard – if data is presented, whether constructed or real, it should be introduced and discussed prior to being included in the discussion. Furthermore, while I like the comparison of Fp to the flashiness index, this was introduced so late in the paper, and not touched on in the methods, and then not truly analyzed. The authors missed an opportunity to give some thought to how this flashiness index compares to Fp – if the flashiness index describes Fp, then why do we need Fp? Does it say the same thing as Fp? Constructive analysis would certainly make the case for Fp here. | Thanks, we have indeed taken a critical look at all and harmonized them.<br><br>We have brought the description of the algorithm used for former figures 2 and 3 into the ‘methods’ section, and described the graphs generated in the ‘results’. The cost of this may be that the explanation of what was Figure 4 (now Fig. 3) will be harder to follow, as no numerical examples have been presented at this stage, but no linear text representation can work for all readers, and once in print readers can switch forward and back and skip the technical sections until they have some general idea of what is being done here.<br><br>Indeed the flashiness index comparison was an ‘add on’ in the discussion, and is now fully integrated in the text. |
| Line 45: 30-50% of what?   | Rainfall, corrected  |
| Line 48: This is true at large scales, but differences in catchments of similar size would break this relationship down  | Not sure we understand what you mean here. All statements in this part have references to the literature in which they are based.  |
| Line 109: there are several historical papers out of the US that were the first to perform the paired catchment study, highlighting the effect of land cover change on streamflow across longer time periods – work at Coweeta, HJ Andrews, Hubbard Brook, and Fernow Forests would be relevant  | We state that there is indeed ‘ample proof’ at this scale and in this type of condition. As we are not writing a textbook with full historical perspective, we maintain that use of the review papers that summarized the studies and conclusions is appropriate. There is similarly an extensive literature on forest and floods in Indonesia in the 1920’s and 1930’s that will be relevant (and is probably lesser known than the well-cited US examples).  |
| Line 127: I’m not sure I fully agree – what is the alternative? Mechanistic models?  | We elaborated the text a bit here  |

|  |  |
|--|--|
| Line 130 on: also soils  | Added  |
| Line 136: the rational method and curve number approach were developed to do just this   | Yes, thanks – we have adjusted the text here and give more reference to CN and SWAT here.  |
| Line 150: I would recommend including more citations here – there are many papers that have demonstrated this  | The Hachrowitz paper is a multi-authored review of the PUB effort – we have added a few further references here.   |
| Line 310: I think there is an incorrectly placed word on this line!  | We hope it has disappeared by now...   |
| Line 469: wording  | Adjusted   |
| Line 491: For Figures 6 and 7 – If the data is used in this paper, you should describe where it came from – I think these implications may be better explored in paper 2, or the data source should be described in paper 1.   | We have indeed more fully incorporated this into the paper and use the actual flow data for the four catchments in Paper I, leaving the (model-based_) scenarios for paper II.   |
| Line 508: This paper made several assumptions that also should be acknowledged – doing hydrology backward only works well under certain cases  | We use this language as in the paper that is referenced – there is acknowledgement in the text surrounding this statement that there is further discussion on where and when it ‘works’  |
| I find Figure 1 to be too busy to follow! Consider reducing color, size, changing font, and adding arrows in such a way to better show the “flow”  | We have reduced colour and harmonized lines  |
| Part II:<br><br>Given that you wish to relate health to $F_p$ , and its change through time to land cover, it seems as though Part II would almost benefit more with a comparison to flashiness indicators. If flashiness tells the same story, then what is the value of $F_p$ ? I still think the argument could be made that it allows some simple process-based analysis. The contribution of $F_p$ in Part II needs to be more clear, and should be benchmarked against another indicator of hydrologic alteration, to show the clear value of using $F_p$ over other indicators. | Thanks for these suggestions. We have added further comparisons of $F_p$ and $F_I$ , for the LU change scenarios in the four catchments (while in Paper I this comparison is made for actual flow data). We have also added a new table (Table 6 in new numbering) that tries to summarize responses to the 7 initial questions for a range of ‘indicators’  |
| Lines 960 – 64 – these are relatively relaxed targets, and may miss peaks. Given that the emphasis of this paper is on high flows, you should include a hydrograph of your model to demonstrate that the model adequately matches high flows. As this would affect your $F_p$ values, it may be why you end up with wide scatter in Figure 3.  | We could add examples as supplementary material, but examples (especially for the Wai Besai watershed with the best data) are already available in <a href="http://www.worldagroforestry.org/output/genriver/download">http://www.worldagroforestry.org/output/genriver/download</a><br><br>The primary reason, we believe, for the scatter are limitations in the rainfall data, with an insufficient number of measurement stations for the given spatial heterogeneity of rainfall. |

|   |  |
|---|--|
| Lines 1011 on: please use words for scenarios instead of abbreviations  | Modified   |
| Line 1097: wording is confusing   | Thanks, we modified the sentence                     |
| Table 5, please use more descriptive titles or cite your abbreviations – I cannot tell what these titles mean | We have used full words instead of abbreviation now. |

7

8 We have highlighted the major changes made in the text in  
 9 yellow

10

11 Flood risk reduction and flow buffering as ecosystem  
12 services: I. Theory on flow persistence, flashiness and base  
13 flow

14 Meine van Noordwijk<sup>1,2</sup>, Lisa Tanika<sup>1</sup>, Betha Lusiana<sup>1</sup>

15 [1]{World Agroforestry Centre (ICRAF), SE Asia program, Bogor, Indonesia}

16 [2]{Wageningen University, Plant Production Systems, Wageningen, the Netherlands}

17 Correspondence to: Meine van Noordwijk ([m.vannoordwijk@cgiar.org](mailto:m.vannoordwijk@cgiar.org))

18 **Abstract 1 (284 words...)**

19 Flood damage reflects insufficient adaptation of human presence and activity to location and  
20 variability of river flow in a given climate. Flood risk increases when landscapes degrade,  
21 counteracted or aggravated by engineering solutions. Efforts to maintain and restore  
22 buffering as ecosystem function may help adaptation to climate change, but require  
23 quantification of effectiveness in their specific social-ecological context. However, the  
24 specific role of forests, trees, soil and drainage pathways in flow buffering, given geology,  
25 land form and climate, remains controversial. Complementing the scarce heavily  
26 instrumented catchments with reliable long-term data, especially in the tropics, there is a  
27 need for metrics for data-sparse conditions. We present and discuss a flow persistence  
28 metric that relates transmission to river flow of peak rainfall events, to the base flow  
29 component of the water balance. The dimensionless flow persistence parameter  $F_p$  is  
30 defined in a recursive flow model and can be estimated from limited time series of observed  
31 daily flow, without requiring knowledge of spatially distributed rainfall upstream. The  $F_p$   
32 metric (or its change over time from what appears to be the local norm) matches local  
33 knowledge concepts. Inter-annual variation in the  $F_p$  metric in sample watersheds correlates  
34 with variation in the 'flashiness index' used in existing watershed health monitoring  
35 programs, but the relationship between these metrics varies with context. Inter-annual  
36 variation in  $F_p$  also correlates with common base-flow indicators, but again in a way that  
37 varies between watersheds. Further exploration of the responsiveness of  $F_p$  in watersheds  
38 with different characteristics to the interaction of land cover and the specific realization of  
39 space-time patterns of rainfall in a limited observation period is needed to evaluate  
40 interpretation of  $F_p$  as indicator of anthropogenic changes in watershed condition.

41 **1 Introduction**

42 Floods can be the direct result of reservoir dams, log jams or protective dykes breaking, with water  
43 derived from unexpected heavy rainfall, rapid snow melt, tsunamis or coastal storm surges. We  
44 focus here on floods that are associated, at least in the public eye, with watershed degradation.  
45 Degradation of watersheds and its consequences for river flow regime and flooding intensity and  
46 frequency are a widespread concern (Brauman et al., 2007; Bishop and Pagiola, 2012; Winsemius et  
47 al., 2013). Engineering measures (dams, reservoirs, canalization, dykes, and flow regulation) can  
48 significantly alter the flow regime of rivers, and reduce the direct relationship with landscape  
49 conditions in the (upper) catchment (Poff et al., 1997). The life expectancy of such structures  
50 depends, however, on the sediment load of incoming rivers and thus on upper watershed conditions  
51 (Graf et al., 2010). Where 'flow regulation' has been included in efforts to assess an economic value

52 of ecosystem services, it can emerge as a major component of overall value; the economic damage  
53 of floods to cities built on floodplains can be huge and the benefits of avoiding disasters thus large  
54 (Farber et al., 2002; Turner and Daily, 2002; Brauman et al., 2007). The 'counterfactual' part of any  
55 avoided damage argument, however, depends on metrics that are transparent in their basic concept  
56 and relationship with observables. Basic requirements for a metric to be used in managing issues of  
57 public concern in a complex multistakeholder environment are that it i) has a direct relationship with  
58 a problem that needs to be solved ('salience'), ii) is aligned with current science-based  
59 understanding of how the underpinning systems function and can be managed ('credibility') and iii)  
60 can be understood from local and public/policy perspectives ('legitimacy') (Clark et al. 2011). Figure  
61 1 summarizes these requirements, building on van Noordwijk et al. (2016).

62  $\Rightarrow$  Figure 1

63 In the popular discussion on floods, especially in the tropics, a direct relationship with deforestation  
64 and reforestation is still commonly perceived to dominate, and forest cover is seen as salient and  
65 legitimate metric of watershed quality (or of urgency of restoration where it is low). A requirement  
66 for 30% forest cover, is for example included in the spatial planning law in Indonesia in this context  
67 (Galudra and Sirait, 2009). Yet, rivers are probably dominated by the other 70% of the landscape.  
68 There is a problem with the credibility of assumed deforestation-flood relations (van Noordwijk et  
69 al., 2007; Verbist et al., 2010), beyond the local scales ( $< 10 \text{ km}^2$ ) of paired catchments where ample  
70 direct empirical proof exists, especially in non-tropical climate zones (Bruijnzeel, 1990, 2004).  
71 Current watershed rehabilitation programs that focus on increasing tree cover in upper watersheds  
72 are only partly aligned with current scientific evidence of effects of large-scale tree planting on  
73 streamflow (Ghimire et al., 2014; Malmer et al., 2010; Palmer, 2009; van Noordwijk et al., 2015a).  
74 The relationship between floods and change in forest quality and quantity, and the availability of  
75 evidence for such a relationship at various scales has been widely discussed over the past decades  
76 (Andréassian, 2004; Bruijnzeel, 2004; Bradshaw et al., 2007; van Dijk et al., 2009). Measurements in  
77 Côte d'Ivoire, for example, showed strong scale dependence of runoff from 30-50% of rainfall at 1  
78  $\text{m}^2$  point scale, to 4% at 130 ha watershed scale, linked to spatial variability of soil properties plus  
79 variations in rainfall patterns (Van de Giesen et al., 2000). The ratio between peak and average flow  
80 decreases from headwater streams to main rivers in a predictable manner; while mean annual  
81 discharge scales with  $(\text{area})^{1.0}$ , maximum river flow was found to scale with  $(\text{area})^{0.4}$  to  $(\text{area})^{0.7}$  on  
82 average (Rodríguez-Iturbe and Rinaldo, 2001; van Noordwijk et al., 1998; Herschy, 2002), with even  
83 lower powers for area in flash floods that are linked to an extreme rainfall event over a restricted  
84 area (Marchi et al., 2010). The determinants of peak flow are thus scale-dependent, with space-time  
85 correlations in rainfall interacting with subcatchment-level flow buffering at any point along the  
86 river. Whether and where peak flows lead to flooding depends on the capacity of the rivers to pass  
87 on peak flows towards downstream lakes or the sea, assisted by riparian buffer areas with sufficient  
88 storage capacity (Baldasarte et al., 2013). Reducing local flooding risk by increased drainage  
89 increases flooding risk downstream, challenging the nested-scales management of watersheds to  
90 find an optimal spatial distribution, rather than minimization, of flooding probabilities. Well-studied  
91 effects of forest conversion on peak flows in small upper stream catchments (Bruijnzeel, 2004;  
92 Change, 2006; Alila et al., 2009) do not necessarily translate to flooding downstream. With most of  
93 the published studies still referring to the temperate zone, the situation in the tropics (generally in  
94 the absence of snow) is contested (Bonell and Bruijnzeel, 2005). As summarized by Beck et al. (2013)  
95 meso- to macroscale catchment studies ( $> 1$  and  $> 10,000 \text{ km}^2$ , respectively) in the tropics, subtropics,  
96 and warm temperate regions have mostly failed to demonstrate a clear relationship between river  
97 flow and change in forest area. Lack of evidence cannot be firmly interpreted as evidence for lack of  
98 effect, however. Detectability of effects depends on their relative size, the accuracy of the

99 measurement devices, length of observation period, and background variability of the signal. A  
100 recent econometric study for Peninsular Malaysia by Tan-Soo et al. (2014) concluded that, after  
101 appropriate corrections for space-time correlates in the data-set for 31 meso- and macroscale basins  
102 (554-28,643 km<sup>2</sup>), conversion of inland rain forest to monocultural plantations of oil palm or rubber  
103 increased the number of flooding days reported, but not the number of flood events, while  
104 conversion of wetland forests to urban areas reduced downstream flood duration. This Malaysian  
105 study may be the first credible empirical evidence at this scale. The difference between results for  
106 flood duration and flood frequency and the result for draining wetland forests warrant further  
107 scrutiny. Consistency of these findings with river flow models based on a water balance and likely  
108 pathways of water under the influence of change in land cover and land use has yet to be shown.  
109 Two recent studies for Southern China confirm the conventional perspective that deforestation  
110 increases high flows, but are contrasting in effects of Reforestation. Zhou et al. (2010) analysed a 50-  
111 year data set for Guangdong Province in China and concluded that forest recovery had not changed  
112 the annual water yield (or its underpinning water balance terms precipitation and  
113 evapotranspiration), but had a statistically significant positive effect on dry season (low) flows. Liu  
114 et al. (2015), however, found for the Meijiang watershed (6983 km<sup>2</sup>) in subtropical China that while  
115 historical deforestation had decreased the magnitudes of low flows (daily flows  $\leq$  Q95%) by 30.1%,  
116 low flows were not significantly improved by Reforestation. They concluded that recovery of low  
117 flows by Reforestation may take much longer time than expected probably because of severe soil  
118 erosion and resultant loss of soil infiltration capacity after deforestation. Changes in river flow  
119 patterns over a limited period of time can be the combined and interactive effects of variations in  
120 the local rainfall regime, land cover effects on soil structure and engineering modifications of water  
121 flow that can be teased apart with modelling tools (Ma et al., 2014).

122 Lacombe et al. (2015) documented that the hydrological effects of natural regeneration differ from  
123 those of plantation forestry, while forest statistics do not normally differentiate between these  
124 different land covers. In a regression study of the high and low flow regimes in the Volta and  
125 Mekong river basins Lacombe and McCartney (2016) found that in the variation among tributaries  
126 various aspects of land cover and land cover change had explanatory power. Between the two  
127 basins, however, these aspects differed. In the Mekong basin variation in forest cover had no direct  
128 effect on flows, but extending paddy areas resulted in a decrease in downstream low flows, probably  
129 by increasing evapotranspiration in the dry season. In the Volta River Basin, the conversion of forests  
130 to crops (or a reduction of tree cover in the existing parkland system) induced greater downstream  
131 flood flows. This observation is aligned with the experimental identification of an optimal,  
132 intermediate tree cover from the perspective of groundwater recharge in parklands in Burkina Faso  
133 (Ilstedt et al., 2016).

134 The statistical challenges of attribution of cause and effect in such data-sets are considerable with  
135 land use/land cover effects interacting with spatially and temporally variable rainfall, geological  
136 configuration and the fact that land use is not changing in random fashion or following any pre-  
137 randomized design (Alila et al., 2009; Rudel et al., 2005). Hydrological analysis across 12 catchments  
138 in Puerto Rico by Beck et al. (2013) did not find significant relationships between the change in  
139 forest cover or urban area, and change in various flow characteristics, despite indications that  
140 regrowing forests increased evapotranspiration.

141 These observations imply that percent tree cover (or other forest related indicators) is probably not  
142 a good metric for judging the ecosystem services provided by a watershed (of different levels of  
143 'health'), and that a metric more directly reflecting changes in river flow may be needed. Here we  
144 will explore a simple recursive model of river flow (van Noordwijk et al., 2011) that (i) is focused on

145 (loss of) flow predictability, (ii) can account for the types of results obtained by the cited recent  
146 Malaysian study (Tan-Soo et al., 2014), and (iii) may constitute a suitable performance indicator to  
147 monitor watershed 'health' through time.

148 Before discussing the credibility dimension of river flow metrics, the way these relate to the salience  
149 and legitimacy issues around 'flood damage' as policy issue need attention. The salient issue of  
150 'flood damage' is compatible with a common dissection of risk as the product of exposure, hazard  
151 and vulnerability (steps 1, 2 and 3 in Figure 2). Many aspects beyond forests and tree cover play a  
152 role; in fact these factors are multiple steps away (step 7A) from the direct river flow dynamics that  
153 determine floods. Extreme discharge events plus river-level engineering (steps 4 and 5) co-  
154 determine hazard (step 2), while exposure (step 1) depends on topographic position interacting with  
155 human presence, and vulnerability can be modified by engineering at a finer scale and be further  
156 reduced by advice to leave an area in high-risk periods. A recent study (Jongman et al., 2015) found  
157 that human fatalities and material losses between 1980 and 2010 expressed as a share of the  
158 exposed population and gross domestic product were decreasing with rising income. The planning  
159 needed to avoid extensive damage requires quantification of the risk of higher than usual  
160 discharges, especially at the upper tail end of the flow frequency distribution.

161 ⇒ Figure 2

162 The statistical scarcity, per definition, of 'extreme events' and the challenge of data collection where  
163 they do occur, make it hard to rely on site-specific empirical data as such. Inference of risks needs  
164 some trust in extrapolation methods, as is often provided by use of trusted underlying mechanisms  
165 and/or data obtained in a geographical proximity. Existing data on flood frequency and duration, as  
166 well as human and economic damage are influenced by topography, soils, human population density  
167 and economic activity, responding to engineered infrastructure (step 5 in Figure 2), as well as the  
168 extreme rainfall events that are their proximate cause (step 6). Subsidence due to groundwater  
169 extraction in urban areas of high population density is a specific problem for a number of cities built  
170 on floodplains (such as Jakarta and Bangkok), but subsidence of drained peat areas has also been  
171 found to increase flooding risks elsewhere (Sumarga et al., 2016). Common hydrological analysis of  
172 flood frequency (called 1 in 10-, 1 in 100-, 1 in 1000-year flood events, for example) relies on direct  
173 observations at step 4 in Fig. 2, but typically requires spatial extrapolation beyond points of data  
174 collection through river flow models that combine at least steps 5 and 6. Relatively simple ways of  
175 including the conditions in the watershed (step 7) in such models rely on the runoff curve number  
176 method (Ponce et al., 1996) and the SWAT (Soil water assessment tool) model that was built on its  
177 foundation (Gassman et al. 2007). Applications on tropical soils have had mixed success (Oliveira et  
178 al. 2016). Describing peak flows as a proportion of the rainfall event that triggered them has a long  
179 history, but where the proportionality factors are estimated for ungauged catchments results may  
180 be unreliable (Efstratiodis et al., 2014). More refined descriptions of the infiltration process (step  
181 7B) are available, using recursive models as filters on empirical data (Grimaldi et al., 2013), but data  
182 for this approach may not be generally available. According to van der Putte et al. (2013) the Green-  
183 Ampt infiltration equation can be fitted to data for dry conditions when soil crusts limit infiltration,  
184 but not in wet winter conditions. These authors argued that simpler models may be better.

185 Analysis of likely change in flood frequencies in the context of climate change adaptation has been  
186 challenging (Milly et al., 2002; Ma et al., 2014). There is a lack of simple performance indicators for  
187 watershed health at its point of relating precipitation P and river flow Q (step 4 in Figure 2) that align  
188 with local observations of river behaviour and concerns about its change and that can reconcile  
189 local, public/policy and scientific knowledge, thereby helping negotiated change in watershed  
190 management (Leimona et al., 2015). The behaviour of rivers depends on many climatic (step 6 in

191 Figure 2) and terrain factors (step 7A-D in Figure 2) that make it a challenge to differentiate between  
192 human induced ecosystem structural change and soil degradation (step 7B) on one hand and  
193 intrinsic variability on the other. Step 8 in Figure 2 represents the direct influence of climate on  
194 vegetation, but also a possible reverse influence (van Noordwijk et al., 2015b). Hydrological models  
195 tend to focus on predicting hydrographs at one or more temporal scales, and are usually tested on  
196 data-sets from limited locations. Despite many decades (if not centuries) of hydrological modelling,  
197 current hydrologic theory, models and empirical methods have been found to be largely inadequate  
198 for sound predictions in ungauged basins (Hrachowitz et al., 2013). Efforts to resolve this through  
199 harmonization of modelling strategies have so far failed. Existing models differ in the number of  
200 explanatory variables and parameters they use, but are generally dependent on empirical data of  
201 rainfall that are available for specific measurement points but not at the spatial resolution that is  
202 required for a close match between measured and modelled river flow. Spatially explicit models  
203 have conceptual appeal (Ma et al., 2010) but have too many degrees of freedom and too many  
204 opportunities for getting right answers for wrong reasons if used for empirical calibration (Beven,  
205 2011). Parsimonious, parameter-sparse models are appropriate for the level of evidence available to  
206 constrain them, but these parameters are themselves implicitly influenced by many aspects of  
207 existing and changing features of the watershed, making it hard to use such models for scenario  
208 studies of changing land use and change in climate forcing. Here we present a more direct approach  
209 deriving a metric of flow predictability that can bridge local concerns and concepts to quantified  
210 hydrologic function: the ‘flow persistence’ parameter as directly observable characteristic (step 4 in  
211 Figure 2), that can be logically linked to the primary points of intervention in watershed  
212 management, interacting with climate and engineering-based change.

213 In this contribution to the debate we will first define the metric ‘flow persistence’ in the context of  
214 temporal autocorrelation of river flow and then derive a way to estimate its numerical value. In part  
215 II we will apply the algorithm to river flow data for a number of contrasting meso-scale watersheds.  
216 In the discussion of this paper we will consider the new flow persistence metric in terms of three  
217 groups of criteria for usable knowledge (Fig. 1; Clark et al., 2011; Lusiana et al., 2011; Leimona et al.,  
218 2015) based on salience (I,II), credibility (III, IV) and legitimacy (V-VII):

- 219 I. Does flow persistence relate to important aspects of watershed behaviour, complementing  
220 existing metrics such as the ‘flashiness index’ and ‘base flow separation’ techniques?
- 221 II. Does its quantification help to select management actions?
- 222 III. Is there consistency of numerical results?
- 223 IV. How sensitive is it to bias and random error in data sources?
- 224 V. Does it match local knowledge?
- 225 VI. Can it be used to empower local stakeholders of watershed management?
- 226 VII. Can it inform local risk management?

227 **2 Flow persistence in water balance equations**

228 **2.1 Recursive model**

229 One of the easiest-to-observe aspects of a river is its day-to-day fluctuation in water level, related to  
230 the volumetric flow (discharge) via rating curves (Maidment, 1992). Without knowing details of  
231 upstream rainfall and the pathways the rain takes to reach the river, observation of the daily  
232 fluctuations in water level allows important inferences to be made. It is also of direct utility: sudden  
233 rises can lead to floods without sufficient warning, while rapid decline makes water utilization  
234 difficult. Indeed, a common local description of watershed degradation is that rivers become more  
235 'flashy' and less predictable, having lost a buffer or 'sponge' effect (Joshi et al., 2004; Ranieri et al.,  
236 2004; Rahayu et al., 2013). A simple model of river flow at time  $t$ ,  $Q_t$ , is that it is similar to that of the  
237 day before ( $Q_{t-1}$ ), multiplied with  $F_p$ , a dimensionless parameter called 'flow persistence' (van  
238 Noordwijk et al., 2011) plus an additional stochastic term  $Q_{a,t}$ :

239 
$$Q_t = F_p Q_{t-1} + Q_{a,t} \quad [1].$$

240  $Q_t$  is for this analysis expressed in  $\text{mm d}^{-1}$ , which means that measurements in  $\text{m}^3 \text{s}^{-1}$  need to be  
241 divided by the relevant catchment area, with appropriate unit conversion. If river flow were  
242 constant, it would be perfectly predictable, i.e.  $F_p$  would be 1.0 and  $Q_{a,t}$  zero; in contrast, an  $F_p$ -value  
243 equal to zero and  $Q_{a,t}$  directly reflecting erratic rainfall represents the lowest possible level of  
244 predictability.

245 The  $F_p$  parameter is conceptually identical to the 'recession constant' commonly used in hydrological  
246 models, typically assessed during an extended dry period when the  $Q_{a,t}$  term is negligible and  
247 streamflow consists of base flow only (Tallaksen, 1995); empirical deviations from a straight line in a  
248 plot of the logarithm of  $Q$  against time are common and point to multiple rather than a single  
249 groundwater pool that contributes to base flow. The larger catchment area has a possibility to get  
250 additional flow from multiple independent groundwater contribution.

251 As we will demonstrate in a next section, it is possible to derive  $F_p$  even when  $Q_{a,t}$  is not negligible. In  
252 climates without distinct dry season this is essential; elsewhere it allows a comparison of apparent  $F_p$   
253 between wet and dry parts of the hydrologic year. A possible interpretation, to be further explored,  
254 is that decrease over the years of  $F_p$  indicates 'watershed degradation' (i.e. greater contrast between  
255 high and low flows), and an increase 'improvement' or 'rehabilitation' (i.e. more stable flows).

256 If we consider the sum of river flow over a period of time (from 1 to  $T$ ) we obtain

257 
$$\sum_1^T Q_t = F_p \sum_1^T Q_{t-1} + \sum_1^T Q_{a,t} \quad [2].$$

258 If the period is sufficiently long period for  $Q_T$  minus  $Q_0$  (the values of  $Q_t$  for  $t=T$  and  $t=0$ , respectively)  
259 to be negligibly small relative to the sum over all  $t$ 's, we may equate  $\sum_1^T Q_t$  with  $\sum_1^T Q_{t-1}$  and obtain a  
260 first way of estimating the  $F_p$  value:

261 
$$F_p = 1 - \sum_1^T Q_{a,t} / \sum_1^T Q_t \quad [3].$$

262 The stochastic  $Q_{a,t}$  can be interpreted in terms of what hydrologists call 'effective rainfall' (i.e. rainfall  
263 minus on-site evapotranspiration, assessed over a preceding time period  $t_x$  since previous rain  
264 event):

265 
$$Q_t = F_p Q_{t-1} + (1-F_p)(P_{tx} - E_{tx}) \quad [4].$$

266 Where  $P_{tx}$  is the (spatially weighted) precipitation on day  $t$  (or preceding precipitation released as  
267 snowmelt on day  $t$ ) in  $\text{mm d}^{-1}$ ;  $E_{tx}$ , also in  $\text{mm d}^{-1}$ , is the preceding evapotranspiration that allowed  
268 for infiltration during this rainfall event (*i.e.* evapotranspiration since the previous soil-replenishing  
269 rainfall that induced empty pore space in the soil for infiltration and retention), or replenishment of  
270 a water film on aboveground biomass that will subsequently evaporate. More complex attributions  
271 are possible, aligning with the groundwater replenishing bypass flow and the water isotopic  
272 fractionation involved in evaporation (Evaristo et al., 2015).

273 The consistency of multiplying effective rainfall with  $(1-F_p)$  can be checked by considering the  
274 geometric series  $(1-F_p)$ ,  $(1-F_p) F_p$ ,  $(1-F_p) F_p^2$ , ...,  $(1-F_p) F_p^n$  which adds up to  $(1-F_p)(1 - F_p^n)/(1-F_p)$  or  $1 - F_p^n$ . This approaches 1 for large  $n$ , suggesting that all of the water attributed to time  $t$ , *i.e.*  $P_t - E_{tx}$ ,  
275 will eventually emerge as river flow. For  $F_p = 0$  all of  $(P_t - E_{tx})$  emerges on the first day, and river flow  
276 is as unpredictable as precipitation itself. For  $F_p = 1$  all of  $(P_t - E_{tx})$  contributes to the stable daily flow  
277 rate, and it takes an infinitely long period of time for the last drop of water to get to the river. For  
278 declining  $F_p$ ,  $(1 > F_p > 0)$ , river flow gradually becomes less predictable, because a greater part of the  
279 stochastic precipitation term contributes to variable rather than evened-out river flow.  
280

281 Taking long term summations of the right- and left- hand sides of Eq.(4) we obtain:

$$282 \Sigma Q_t = \Sigma (F_p Q_{t-1} + (1-F_p)(P_t - E_{tx})) = F_p \Sigma Q_{t-1} + (1-F_p)(\Sigma P_t - \Sigma E_{tx}) \quad [5].$$

283 Which is consistent with the basic water budget,  $\Sigma Q = \Sigma P - \Sigma E$ , at time scales long enough for  
284 changes in soil water buffer stocks to be ignored. As such the total annual, and hence the mean daily  
285 river flow are independent of  $F_p$ . This does not preclude that processes of watershed degradation or  
286 restoration that affect the partitioning of  $P$  over  $Q$  and  $E$  also affect  $F_p$ .

## 287 **2.2 Base flow**

288 Clarifying the  $Q_a$  contribution is equivalent with one of several ways to separate base flow from peak  
289 flows. Rearranging Eq.(3) we obtain

$$290 \Sigma_1^T Q_{a,t} = (1 - F_p) \Sigma_1^T Q_t \quad [6].$$

291 The  $\Sigma Q_{a,t}$  term reflects the sum of peak flows in mm. Its complement,  $F_p \Sigma Q_t$ , reflects the sum of base  
292 flow, also in mm. For  $F_p = 1$  (the theoretical maximum) we conclude that all  $Q_{a,t}$  must be zero, and all  
293 flow is 'base flow'.

## 294 **2.3 Low flows**

295 The lowest flow expected in an annual cycle is  $Q_x F_p^{N_{\max}}$  where  $Q_x$  is flow on the first day without rain  
296 and  $N_{\max}$  the longest series of dry days. Taken at face value, a decrease in  $F_p$  has a strong effect on  
297 low-flows, with a flow of 10% of  $Q_x$  reached after 45, 22, 14, 10, 8 and 6 days for  $F_p = 0.95, 0.9, 0.85,$   
298 0.8, 0.75 and 0.7, respectively. However, the groundwater reservoir that is drained, equalling the  
299 cumulative dry season flow if the dry period is sufficiently long, is  $Q_x/(1-F_p)$ . If  $F_p$  decreases to  $F_{px}$  but  
300 the groundwater reservoir ( $\text{Res} = Q_x/(1-F_p)$ ) is not affected, initial flows in the dry period will be  
301 higher ( $Q_x F_{px}^i (1-F_{px}) \text{ Res} > Q_x F_p^i (1-F_p) \text{ Res}$  for  $i < \log((1-F_{px})/(1-F_p))/\log(F_p/F_{px})$ ). It thus matters how  
302 low flows are evaluated: from the perspective of the lowest level reached, or as cumulative flow.  
303 The combination of climate, geology and land form are the primary determinants of cumulative low  
304 flows, but if land cover reduces the recharge of groundwater there may be impacts on dry season  
305 flow, that are not directly reflected in  $F_p$ .

306 If a single  $F_p$  value would account for both dry and wet season, the effects of changing  $F_p$  on low  
307 flows may well be more pronounced than those on flood risk. Empirical tests are needed of the  
308 dependence of  $F_p$  on  $Q$  (see below). Analysis of the way an aggregate  $F_p$  depends on the dominant  
309 flow pathways provides a basis for differentiating  $F_p$  within a hydrologic year.  
310

## 311 **2.4 Flow-pathway dependence of flow persistence**

312 The patch-level partitioning of water between infiltration and overland flow is further modified at  
313 hillslope level, with a common distinction between three pathways that reach streams: overland  
314 flow, interflow and groundwater flow (Band et al., 1993; Weiler and McDonnell, 2004). An additional  
315 interpretation of Eq.(1), potentially adding to our understanding of results but not needed for  
316 analysis of empirical data, can be that three pathways of water through a landscape contribute to  
317 river flow (Barnes, 1939): groundwater release with  $F_{p,g}$  values close to 1.0, overland flow with  $F_{p,o}$   
318 values close to 0, and interflow with intermediate  $F_{p,i}$  values.

319  $Q_t = F_{p,g} Q_{t-1,g} + F_{p,i} Q_{t-1,i} + F_{p,o} Q_{t-1,o} + Q_{a,t}$  [7],

320  $F_p = (F_{p,g} Q_{t-1,g} + F_{p,i} Q_{t-1,i} + F_{p,o} Q_{t-1,o}) / Q_{t-1}$  [8].

321 On this basis a decline or increase in overall weighted average  $F_p$  can be interpreted as indicator of a  
322 shift of dominant runoff pathways through time within the watershed. Dry season flows are  
323 dominated by  $F_{p,g}$ . The effective  $F_p$  in the rainy season can be interpreted as indicating the relative  
324 importance of the other two flow pathways.  $F_p$  reflects the fractions of total river flow that are based  
325 on groundwater, overland flow and interflow pathways:

326  $F_p = F_{p,g} (\Sigma Q_{t,g} / \Sigma Q_t) + F_{p,o} (\Sigma Q_{t,o} / \Sigma Q_t) + F_{p,i} (\Sigma Q_{t,i} / \Sigma Q_t)$  [9].

327 Beyond the type of degradation of the watershed that, mostly through soil compaction, leads to  
328 enhanced infiltration-excess (or Hortonian) overland flow (Delfs et al., 2009), saturated conditions  
329 throughout the soil profile may also induce overland flow, especially near valley bottoms (Bonell,  
330 1993; Bruijnzeel, 2004). Thus, the value of  $F_{p,o}$  can be substantially above zero if the rainfall has a  
331 significant temporal autocorrelation, with heavy rainfall on subsequent days being more likely than  
332 would be expected from general rainfall frequencies. If rainfall following a wet day is more likely to  
333 occur than following a dry day, as is commonly observed in Markov chain analysis of rainfall patterns  
334 (Jones and Thornton, 1997; Bardossy and Plate, 1991), the overland flow component of total flow  
335 will also have a partial temporal autocorrelation, adding to the overall predictability of river flow. In  
336 a hypothetical climate with evenly distributed rainfall, we can expect  $F_p$  to be 1.0 even if there is no  
337 infiltration and the only pathway available is overland flow. Even with rainfall that is variable at any  
338 point of observation but has low spatial correlation it is possible to obtain  $F_p$  values of (close to) 1.0  
339 in a situation with (mostly) overland flow (Ranieri et al., 2004).

## 340 **2.5 Relationship between flow persistence and flashiness index**

341 The Richards-Baker 'R-B Flashiness index' (Baker et al. 2004) is defined as

342  $FI = \sum_t |\Delta Q_t| / \sum_t Q_t = \sum_{ti} (Q_t - Q_{t-1}) + \sum_{td} (Q_{t-1} - Q_t)$  [10]

343 with  $ti$  indicating all times  $t$  that  $Q_t > Q_{t-1}$  and  $td$  indicating all times  $t$  that  $Q_t < Q_{t-1}$ . Over a  
344 timeframe that flow has no net trend, the sum of increments ( $\sum_{ti} (Q_t - Q_{t-1})$ ) is equal to the sum of  
345 declines ( $\sum_{td} (Q_{t-1} - Q_t)$ ).

346 Substituting equation [5] in [10] we obtain:

347  $FI = 2(1-F_p)(0.5 \Delta S + \sum_{ti} (P_t - E_{tx} - Q_t)) / \sum_t Q_t = 2(1-F_p)(-0.5 \Delta S + \sum_{td} (-P_t + E_{tx} + Q_t)) / \sum_t Q_t$  [11]

348 With  $\Delta S$  representing change in catchment storage;  $\Delta S = (1-F_p)(-\sum_{ti} (P_t - E_{tx} - Q_t) + \sum_{td} (-P_t + E_{tx} + Q_t))$ .

349 This suggests that  $FI = 2(1-F_p)$  is a first approximation and becomes zero for  $F_p = 1$ . These  
350 approximations require that changes in the catchment have no influence on  $P_t$  or  $E_{tx}$  values. If  $E_{tx}$  is  
351 negatively affected (either by a change in vegetation or by insufficient buffering, reducing water  
352 availability on non-rainfall days) flashiness will increase, beyond the main effects on  $F_p$ .

353 The rainfall term, counted positive for all days with flow increase and negatively for days with  
354 declining flow, hints at one of the major reasons why the flashiness index tends to get smaller when  
355 larger catchment areas are involved: rainfall will tend to get more evenly distributed over time,  
356 unless the spatial correlation of rainfall is (close to) 1 and all rainfall derives from fronts passing over  
357 the area uniformly. Where (part of) precipitation occurs as snow, the timing of snow melt defines  $P_t$   
358 as used here. Where vegetation influences timing and synchrony of snowmelt, this will be reflected  
359 in the flashiness index. It may not directly influence flow persistence, but will be accounted for in the  
360 flow description that uses flow persistence as key parameter.

361 **3. Methods**

362 **3.1 River flow data for four tropical watersheds**

363 To test the applicability of the  $F_p$  metric and explore its properties, data from four Southeast Asian  
364 watersheds were used, that will be described and further analyzed in part II. The first watershed  
365 data set is the Way Besai ( $414.4 \text{ km}^2$ ) in Lampung province, Sumatra, Indonesia (Verbist et al., 2010).  
366 With an elevation between 720–1831 m a.s.l., the Way Besai is dominated by various coffee  
367 production systems (64%), with remaining forest (18%), horticulture and crops (12%) and other land  
368 uses (6%). Daily rainfall data from 1976 – 2007, was generated by interpolation of eight rainfall  
369 stations using Thiessen polygons; data were obtained from BMKG (*Agency on Meteorology,*  
370 *Climatology and Geophysics*), PU (Public Work Agency) and PLN (*National Electricity Company*). The  
371 average of annual rainfall was 2474 mm, with observed values in the range 1216 – 3277 mm. River  
372 flow data at the outflow of the Way Besai was also obtained from PU and PUSAIR (*Centre for*  
373 *Research and Development on Water Resources*), with an average of river flow of  $16.7 \text{ m}^3/\text{s}$ .

374 Data from three other watersheds were used to explore the variation of  $F_p$  across multiple years and  
375 its relationship with the Flashiness Index: Bialo ( $111.7 \text{ km}^2$ ) in South Sulawesi, Indonesia with  
376 Agroforestry as the dominant land cover type, Cidanau ( $241.6 \text{ km}^2$ ) in West Java, Indonesia,  
377 dominated by mixed Agroforestry land uses but with a peat swamp before the final outlet and Mae  
378 Chaem ( $3892 \text{ km}^2$ ) in Northern Thailand, part of the upper Ping Basin, and dominated by evergreen,  
379 deciduous and pine forest. Detailed information on these watersheds and the data sources is  
380 provided in Paper II.

381 **3.2 Numerical examples**

382 For 'Monte Carlo' simulations a river flow model representing equation [1] was implemented in a  
383 spreadsheet model that is available from the authors on request. Fixed values for  $F_p$  were used in  
384 combination with a stochastic  $Q_{a,t}$  value. The latter was obtained from a random generator (rand)

385 with two settings for a (truncated) sinus-based daily rainfall probability: A) one for situations that  
386 have approximately 120 rainy days, and an annual Q of around 1600 mm, and B) one that leads to  
387 around 45 rainy days and an annual total around 600 mm. Maximum daily  $Q_{a,t}$  was chosen as 60 mm  
388 in both cases. For the figures, realizations for various  $F_p$  values were retained that were within 10%  
389 of this number of rainy days and annual flow total, to focus on the effects of  $F_p$  as such.

### 390 **3.3 Flow persistence as a simple flood risk indicator**

391 For numerical examples (implemented in a spreadsheet model) flow on each day can be derived as:

392  $Q_t = \sum_j^t F_p^{t-j} (1-F_p) p_j P_j$  [12].

393 Where  $p_j$  reflects the occurrence of rain on day  $j$  (reflecting a truncated sine distribution for seasonal  
394 trends) and  $P_j$  is the rain depth (drawn from a uniform distribution). From this model the effects of  $F_p$   
395 (and hence of changes in  $F_p$ ) on maximum daily flow rates, plus maximum flow totals assessed over a  
396 2-5 d period, was obtained in a Monte Carlo process (without Markov autocorrelation of rainfall in  
397 the default case – see below). Relative flood protection was calculated as the difference between  
398 peak flows (assessed for 1-5 d duration after a 1 year ‘warm-up’ period) for a given  $F_p$  versus those  
399 for  $F_p = 0$ , relative to those at  $F_p = 0$ .

### 400 **3.4 An algorithm for deriving $F_p$ from a time series of stream flow data**

401 Equation (3) provides a first method to derive  $F_p$  from empirical data if these cover a full hydrologic  
402 year. In situations where there is no complete hydrograph and/or in situations where we want to  
403 quantify  $F_p$  for shorter time periods (e.g. to characterise intraseasonal flow patterns) and the change  
404 in the storage term of the water budget equation cannot be ignored, we need an algorithm for  
405 estimating  $F_p$  from a series of daily  $Q_t$  observations.

406 Where rainfall has clear seasonality, it is attractive and indeed common practice to derive a  
407 groundwater recession rate from a semi-logarithmic plot of  $Q$  against time (Tallaksen, 1995). As we  
408 can assume for such periods that  $Q_{a,t} = 0$ , we obtain  $F_p = Q_t / Q_{t-1}$ , under these circumstances. We  
409 cannot be sure, however, that this  $F_{p,g}$  estimate also applies in the rainy season, because overall wet-  
410 season  $F_p$  will include contributions by  $F_{p,o}$  and  $F_{p,i}$  as well (compare Eq. 9). In locations without a  
411 distinct dry season, we need an alternative method.

412 A biplot of  $Q_t$  against  $Q_{t-1}$  will lead to a scatter of points above a line with slope  $F_p$ , with points above  
413 the line reflecting the contributions of  $Q_{a,t} > 0$ , while the points that plot on the  $F_p$  line itself  
414 represent  $Q_{a,t} = 0 \text{ mm d}^{-1}$ . There is no independent source of information on the frequency at which  
415  $Q_{a,t} = 0$ , nor what the statistical distribution of  $Q_{a,t}$  values is if it is non-zero. Calculating back from the  
416  $Q_t$  series we can obtain an estimate ( $Q_{a,Fptry}$ ) of  $Q_{a,t}$  for any given estimate ( $F_{p,try}$ ) of  $F_p$ , and select the  
417 most plausible  $F_p$  value. For high  $F_{p,try}$  estimates there will be many negative  $Q_{a,Fptry}$  values, for low  
418  $F_{p,try}$  estimates all  $Q_{a,Fptry}$  values will be larger. An algorithm to derive a plausible  $F_p$  estimate can thus  
419 make use of the corresponding distribution of ‘apparent  $Q_a$ ’ values as estimates of  $F_{p,try}$ , calculated  
420 as  $Q_{a,try} = Q_t - F_{p,try} Q_{t-1}$ . While  $Q_{a,t}$  cannot be negative in theory, small negative  $Q_a$  estimates are likely  
421 when using real-world data with their inherent errors. The FlowPer  $F_p$  algorithm (van Noordwijk et  
422 al., 2011) derives the distribution of  $Q_{a,try}$  estimates for a range of  $F_{p,try}$  values (Figure 3B) and selects  
423 the value  $F_{p,try}$  that minimizes the variance  $\text{Var}(Q_{a,Fptry})$  (or its standard deviation) (Figure 3C). It is  
424 implemented in a spreadsheet workbook that can be downloaded from the ICRAF website  
425 (<http://www.worldAgroforestry.org/output/flowper-flow-persistence-model>)

426 ➔ Figure 3

427 A consistency test is needed that the high-end  $Q_t$  values relate to  $Q_{t+1}$  in the same was as do low or  
428 medium  $Q_t$  values. Visual inspection of  $Q_{t+1}$  versus  $Q_t$ , with the derived  $F_p$  value, provides a  
429 qualitative view of the validity of this assumption. The  $F_p$  algorithm can be applied to any population  
430 of  $(Q_{t-1}, Q_t)$  pairs, e.g. selected from a multiyear data set on the basis of 3-month periods within the  
431 hydrological year.

432 **3.5 Flashiness and flow separation**

433 Hydrographs analysed for  $F_p$  were also used for calculating the Richards-Baker or R-B Flashiness  
434 index (Baker et al. 2004) by summing the absolute values of all daily changes in flow. Two common  
435 flow separation algorithms (fixed and sliding interval methods, Furey and Gupta, 2001) were used to  
436 estimate the base flow fraction at an annual basis. The average of the two was compared to  $F_p$ .

437 **4 Results**

438 **4.1 Numerical examples**

439 Figure 4 provides two examples, for annual river flows of around 1600 and 600 mm  $y^{-1}$ , of the way a  
440 change in  $F_p$  values (based on Eq. 1) influences the pattern of river flow for a unimodal rainfall  
441 regime with a well-developed dry season. The increasing 'spikiness' of the graph as  $F_p$  is lowered,  
442 regardless of annual flow, indicates reduced predictability of flow on any given day during the wet  
443 season on the basis of the flow on the preceding day.

444 ⇒ Figure 4

445 A bi-plot of river flow on subsequent days for the same simulations (Figure 5) shows two main  
446 effects of reducing the  $F_p$  value: the scatter increases, and the slope of the lower envelope  
447 containing the swarm of points is lowered (as it equals  $F_p$ ). Both of these changes can provide entry  
448 points for an algorithm to estimate  $F_p$  from empirical time series, provided the basic assumptions of  
449 the simple model apply and the data are of acceptable quality.

450 ⇒ Figure 5

451 For the numerical examples shown in Figure 4, the relative increase of the maximum daily flow when  
452 the  $F_p$  value decreased from a value close to 1 (0.98) to nearly 0 depended on the rainfall regime;  
453 with lower annual rainfall but the same maximum daily rainfall, the response of peak flows to  
454 decrease in  $F_p$  became stronger.

455 **4.2 Flood intensity and duration**

456 Figure 6 shows the effect of  $F_p$  values in the range 0 to 1 on the maximum flows obtained with a  
457 random time series of 'effective rainfall', compared to results for  $F_p = 0$ . Maximum flows were  
458 considered at time scales of 1 to 5 days, in a moving average routine. This way a relative flood  
459 protection, expressed as reduction of peak flow, could be related to  $F_p$  (Figure 6A).

460 ⇒ Figure 6

461 Relative flood protection rapidly decreased from its theoretical value of 100% at  $F_p = 1$  (when there  
462 was no variation in river flow), to less than 10% at  $F_p$  values of around 0.5. Relative flood protection  
463 was slightly lower when the assessment period was increased from 1 to 5 days (between 1 and 3 d it  
464 decreased by 6.2%, from 3 to 5 d by a further 1.3%). Two counteracting effects are at play here: a

465 lower  $F_p$  means that a larger fraction ( $1-F_p$ ) of the effective rainfall contributes to river flow, but the  
466 increased flow is less persistent. In the example the flood protection in situations where the rainfall  
467 during 1 or 2 d causes the peak is slightly stronger than where the cumulative rainfall over 3-5 d  
468 causes floods, as typically occurs downstream.

469 As we expect from equation 5 that peak flow is to  $(1-F_p)$  times peak rainfall amounts, the effect of a  
470 change in  $F_p$  not only depends on the change in  $F_p$  that we are considering, but also on its initial  
471 value. Higher initial  $F_p$  values will lead to more rapid increases in high flows for the same reduction in  
472  $F_p$  (Figure 6B). However, flood duration rather responds to changes in  $F_p$  in a curvilinear manner, as  
473 flow persistence implies flood persistence (once flooding occurs), but the greater the flow  
474 persistence the less likely such a flooding threshold is passed (Figure 6C). The combined effect may  
475 be restricted to about 3 d of increase in flood duration for the parameter values used in the default  
476 example, but for different parametrization of the stochastic  $\epsilon$  other results might be obtained.

#### 477 **4.3 Algorithm for $F_p$ estimates from river flow time series**

478 The algorithm has so far returned non-ambiguous  $F_p$  estimates on any modelled time series data of  
479 river flow, as well as for all empirical data set we tested (including all examples tested in part II),  
480 although there probably are data sets on which it can breakdown. Visual inspection of  $Q_{t-1}/Q_t$  biplots  
481 (as in Figure 4) can provide clues to non-homogenous data sets, to potential situations where  
482 effective  $F_p$  depends on flow level  $Q_t$  and where data are not consistent with a straight-line lower  
483 envelope. Where river flow estimates were derived from a model with random elements, however,  
484 variation in  $F_p$  estimates was observed, that suggests that specific aspects of actual rainfall, beyond  
485 the basic characteristics of a watershed and its vegetation, do have at least some effect. Such effects  
486 deserve to be further explored for a set of case studies, as their strength probably depends on  
487 context.

#### 488 **4.4 Flow persistence compared to base flow and flashiness index**

489 Figure 7 compares results for a hydrograph of a single year for the Way Besai catchment, described  
490 in more detail in paper II. While there is agreement on most of what is indicated as baseflow, the  
491 short term response to peaks in the flow differ, with baseflow in the  $F_p$  method more rapidly  
492 increasing after peak events.

493  $\Rightarrow$  Figure 7

494 When compared across multiple years for four Southeast Asian catchments (figure 8), there is partial  
495 agreement in the way interannual variation is described in each catchment, while numerical values  
496 are similar. However, the ratio of what is indicated as baseflow according to the  $F_p$  method and  
497 according to standard hydrograph separation varies from 1.05 to 0.86.

498  $\Rightarrow$  Figure 8

499 Figure 9 compares numerical results for the R-B Flashiness Index with  $F_p$  for the four test catchments  
500 and for a number of hydrographs constructed as in Fig. 3A. The two concepts are inversely related,  
501 as expected from equation [11], but where  $F_p$  is constrained to the 0-1 interval, the R-B Flashiness  
502 Index can attain values up to 2.0, with the value for  $F_p = 0$  depending on properties of the local  
503 rainfall regime. Where hydrographs were generated with a simple flow model with  $F_p$  parameter as  
504 key variable, the flashiness index is more tightly related to, especially for higher  $F_p$  values, than  
505 where both flashiness index and  $F_p$  were derived from existing flow data (Figure 9B versus 9A). The

506 difference in slope between the four watersheds in Fig. 9A appears to be primarily related to aspects  
507 of the local rainfall pattern that deserve further analysis in larger data sets of this nature.

508  $\Rightarrow$  Figure 9

509

## 510 5 Discussion

511 We will discuss the flow persistence metric based on the seven questions raised from the  
512 perspectives of salience, credibility and legitimacy and refer back to figure 2 that clarified how  
513 ecosystem structure, ecosystem function and human land use interact in causal loops that can lead  
514 to flood damage, its control and/or prevention.

### 515 5.1 Salience

516 Key *salience* aspects are “Does flow persistence relate to important aspects of watershed  
517 behaviour?” and “Does it help to select management actions?”. A major finding in the derivation of  
518  $F_p$  was that the flow persistence measured at daily time scale can be logically linked to the long-term  
519 water balance under the assumption that the watershed is defined on the basis of actual  
520 groundwater flows, and that the proportion of peak rainfall that translates to peak river flow equals  
521 the complement of flow persistence. This feature links effects on floods of changes in watershed  
522 quality, as commonly expressed in curve numbers and flashiness indices, to effects on low flows, as  
523 commonly expressed in base flow metrics. The  $F_p$  parameter as such does not predict when and  
524 where flooding will occur, but it does help to assess to what extent another condition of the  
525 watershed, with either higher or lower  $F_p$  would translate the same rainfall into larger or small peak  
526 water flows. This is salient, especially if the relative contributions of (anthropogenic) land cover and  
527 the (exogenous, probabilistic) specifics of the rainfall pattern can be further teased apart (see part  
528 II). Where  $F_p$  may describe the descending branch of hydrographs at a relevant time scale, details of  
529 the ascending branch beyond the maximum daily flow reached may be relevant for reducing flood  
530 damage, and may require more detailed study at higher temporal resolution.

531 Figures 3 and 6 show that most of the effects of a decreasing  $F_p$  value on peak discharge (which is  
532 the basis for downstream flooding) occur between  $F_p$  values of 1 and 0.7, with the relative flood  
533 protection value reduced to 10% when  $F_p$  reaches 0.5. As indicated in Figure 2, peak discharge is only  
534 one of the factors contributing to flood risk in terms of human casualties and physical damage. Flood  
535 risks are themselves nonlinearly and in strongly topography-specific ways related to the volume of  
536 river flow after extreme rainfall events. While the expected fraction of rainfall that contributes to  
537 direct flow is linearly related to rainfall via  $(1-F_p)$ , flooding risk as such will have a non-linear  
538 relationship with rainfall, that depends on topography and antecedent rainfall. Catchment changes,  
539 such as increases or decreases in percentage tree cover, will generally have a non-linear relationship  
540 with  $F_p$  as well as with flooding risks. The  $F_p$  value has an inverse effect on the fraction of recent  
541 rainfall that becomes river flow, but the effect on peak flows is less, as higher  $F_p$  values imply higher  
542 base flow. The way these counteracting effects balance out depends on details of the local rainfall  
543 pattern (including its Markov chain temporal autocorrelation), as well as the downstream  
544 topography and risk of people being at the wrong time at a given place, but the  $F_p$  value is an  
545 efficient way of summarizing complex land use mosaics and upstream topography in its effect on  
546 river flow. The difference between wet-season and dry-season  $F_p$  deserves further analysis. In  
547 climates with a real rainless dry-season, dry season  $F_p$  is dominated by the groundwater release  
548 fraction of the watershed, regardless of land cover, while in wet season it depends on the mix

549 (weighted average) of flow pathways. The degree to which  $F_p$  can be influenced by land cover needs  
550 to be assessed for each landscape and land cover combination, including the locally relevant forest  
551 and forest derived land classes, with their effects on interception, soil infiltration and time pattern of  
552 transpiration. The  $F_p$  value can summarize results of models that explore land use change scenarios  
553 in local context. To select the specific management actions that will maintain or increase  $F_p$  a locally  
554 calibrated land use/hydrology model is needed, such as GenRiver (part II), DHV (Bergström, 1995) or  
555 SWAT (Yen et al., 2015).

556 The “health” wording has been used as a comprehensive concept of the way a) climate forcing, b)  
557 watershed vegetation and soil conditions and c) engineering interventions interact on functional  
558 aspects of river flow. Ma et al (2014) described a method to separate these three influences on river  
559 flow. In the four catchments we used as example there have been no major dams or reservoirs  
560 installed upstream of the points of measurement. Where these do exist the specific operating rules  
561 of reservoirs need to be included in any model and these can have a major influence on downstream  
562 flow, depending on the primary use for power generation, dry season irrigation or stabilizing river  
563 flow for riverine transport. Although a higher  $F_p$  value will in most cases be desirable (and a decrease  
564 in  $F_p$  undesirable), we may expect that In an ecological perspective on watershed health, the change  
565 in low flows that can occur in the flow regime of degrading and intensively managed watersheds  
566 alike, depending on the management rules for reservoirs, is at least as relevant as changes in flood  
567 risks, as many aquatic organisms thrive during floods (Pahl-Wostl et al., 2013; Poff et al., 2010).  
568 Downstream biota can be expected to have adapted to the pre-human flow conditions, inherent  $F_p$   
569 and variability. Decreased variability of flow achieved by engineering interventions (e.g. a reservoir  
570 with constant release of water to generate hydropower) may have negative consequences for fish  
571 and other biota (Richter et al., 2003; McCluney et al., 2014). In an extensive literature review Poff  
572 and Zimmerman (2010) found no general, transferable quantitative relationships between flow  
573 alteration and ecological response, but the risk of ecological change increases with increasing  
574 magnitude of flow alteration.

575 Various geographically defined watershed health concepts are in use (see for example  
576 <https://www.epa.gov/hwp/healthy-watersheds-projects-region-5>; City of Fort Collins, 2015,  
577 employing a range of specific indicators, including the ‘R-B flashiness index’ (Baker et al. 2004). The  
578 definition of watershed health, like that of human health has evolved over time. Human health was  
579 seen as a state of normal function that could be disrupted from time to time by disease. In 1948 the  
580 World Health Organization (1958) proposed a definition that aimed higher, linking health to well-  
581 being, in terms of physical, mental, and social aspects, and not merely the absence of disease and  
582 infirmity. Health became seen as the ability to maintain homeostasis and recover from injury, but  
583 remained embedded in the environment in which humans function.

## 584 **5.2 Credibility**

585 Key credibility questions are “Consistency of numerical results?” and “How sensitive are results to  
586 bias and random error in data sources?”. A key strength of our flow persistence parameter, that it  
587 can be derived from a limited number of observations of river flow at a single point along the river,  
588 without knowledge of rainfall events and catchment conditions, is also its major weakness. If rainfall  
589 data exist, and especially rainfall data that apply to each subcatchment, the  $Q_a$  term doesn’t have to  
590 be treated as a random variable and event-specific information on the flow pathways may be  
591 inferred for a more precise account of the hydrograph. But for the vast majority of rivers in the  
592 tropics, advances in remotely sensed rainfall data are needed to achieve that situation and  $F_p$  may be

593 all that is available to inform public debates on the **location-specific** relation between forests and  
594 floods.

595 The main conclusions from the numerical examples analysed so far are that intra-annual variability  
596 of  $F_p$  values between wet and dry seasons was around 0.2, interannual variability in either annual or  
597 seasonal  $F_p$  was generally in the 0.1 range, while the difference between observed and simulated  
598 flow data as basis for  $F_p$  calculations was mostly less than 0.1. With current methods, it seems that  
599 effects of land cover change on flow persistence that shift the  $F_p$  value by about 0.1 are the limit of  
600 what can be asserted from empirical data (with shifts of that order in a single year a warning sign  
601 rather than a firmly established change). When derived from observed river flow data  $F_p$  is suitable  
602 for monitoring change (degradation, restoration) and can be a serious candidate for monitoring  
603 performance in outcome-based ecosystem service management contracts. In interpreting changes in  
604  $F_p$  as caused by changes in the condition in the watershed, however, changes in specific properties of  
605 the rainfall regime must be excluded. At the scale of paired catchment studies this assumption may  
606 be reasonable, but in temporal change (or using specific events as starting point for analysis), it is  
607 not easy to disentangle interacting effects (Ma et al., 2014). Recent evidence that vegetation not  
608 only responds to, but also influences rainfall (arrow 10 in Figure 2; van Noordwijk et al., 2015b)  
609 further complicates the analysis across scales.

610 As indicated, the  $F_p$  method is related to earlier methods used in streamflow hydrograph separation  
611 of base flow and quick flow. While textbooks (Ward and Robinson, 2000; Hornberger et al 2014)  
612 tend to be critical of the lack of objectivity of graphical methods, algorithms are used for deriving the  
613 minimum flow in a fixed or sliding period of reference as base flow (Sloto and Crouse, 1996; Furey  
614 and Gupta, 2001). The time interval used for deriving the minimum flow depends on catchment size.

615 Recursive models that describe flow in a next time interval on the basis of a fraction of that in the  
616 preceding time interval with a term for additional flow due to additional rainfall have been used in  
617 analysis of peak flow event before, with time intervals as short as 1 minute rather than the 1 day we  
618 use here (Rose, 2004). Through reference to an overall mass balance a relationship similar to what  
619 we found here ( $F_p$  times preceding flow plus  $1 - F_p$  times recent inputs) was also used in such  
620 models. To our knowledge, the method we describe here at daily timescales has not been used  
621 before.

622 The idea that the form of the storage-discharge function can be estimated from analysis of  
623 streamflow fluctuations has been explored before for a class of catchments in which discharge is  
624 determined by the volume of water in storage (Kirchner, 2009). Such catchments behave as simple  
625 first-order nonlinear dynamical systems and can be characterized in a single-equation rainfall-runoff  
626 model that predicted streamflow, in a test catchment in Wales, as accurately as other models that  
627 are much more highly parameterized. This model of the  $dQ/dt$  versus  $Q$  relationship can also be  
628 analytically inverted; thus, it can, according to Kirchner (2009), be used to “do hydrology backward,”  
629 that is, to infer time series of whole-catchment precipitation directly from fluctuations in  
630 streamflow. The slope of the log-log relationship between flow recession ( $dQ/dt$ ) and  $Q$  that  
631 Kirchner (2009) used is conceptually similar to the  $F_p$  metric we derived here, but the specific  
632 algorithm to derive the parameter from empirical data differs. **Further exploration of the underlying**  
633 **assumptions is needed.** Estimates of  $dQ/dt$  are sensitive to noise in the measurement of  $Q$  and the  
634 possibly frequent and small increases in  $Q$  can be separated from the expected flow recession in the  
635 algorithm we presented here.

636 **Table 1 compares a number of properties (Salience and Legitimacy in properties 1-4, Credibility**  
637 **dimensions in 5-10) for the R-B Flashiness Index (Baker et al. 2004) and flow persistence. The main**

638 advantage of continuing with the flashiness index is that there is an empirical basis for comparisons  
639 and the index has been included in existing 'watershed health' monitoring programs, especially in  
640 the USA. The main advantage of including  $F_p$  is that it can be estimated from incomplete flow  
641 records, has a clear link to peak flow events and has a more direct relationship with underlying flow  
642 pathways, changes in rainfall (or snowmelt) and evapotranspiration, reflecting land cover change.

643 ➔ Table 1

644 Seifert and Beven (2009) discussed the increase in predictive skill of models depending on the  
645 amount of location-specific data that can be used to constrain them. They found that the ensemble  
646 prediction of multiple models for a single location clearly outperformed the predictions using single  
647 parameter sets and that surprisingly little runoff data was necessary to identify model  
648 parameterizations that provided good results for 'ungauged' test periods in cases where actual  
649 measurements were available. Their results indicated that a few runoff measurements can contain  
650 much of the information content of continuous runoff time series. The way these conclusions might  
651 be modified if continuous measurements for limited time periods, rather than separated single data  
652 points on river flow could be used, remains to be explored. Their study indicated that results may  
653 differ significantly between catchments and critical tests of  $F_p$  across multiple situations are  
654 obviously needed, as paper II will provide.

655 In discussions and models of temperate zone hydrology (Bergström, 1995; Seifert, 1999) snowmelt is  
656 a major component of river flow and effects of forest cover on spring temperatures are important to  
657 the buffering of the annual peaks in flow that tend to occur in this season. Application of the  $F_p$   
658 method to data describing such events has yet to be done.

659 **5.3 Legitimacy**

660 *Legitimacy* aspects are "Does it match local knowledge?" and "Can it be used to empower local  
661 stakeholders of watershed management?" and "Can it inform risk management?". As the  $F_p$   
662 parameter captures the predictability of river flow that is a key aspect of degradation according to  
663 local knowledge systems, its results are much easier to convey than full hydrographs or exceedance  
664 probabilities of flood levels. By focusing on observable effects at river level, rather than prescriptive  
665 recipes for land cover ("Reforestation"), the  $F_p$  parameter can be used to more effectively compare  
666 the combined effects of land cover change, changes in the riparian wetlands and engineered water  
667 storage reservoirs, in their effect on flow buffering. It is a candidate for shifting environmental  
668 service reward contracts from input to outcome based monitoring (van Noordwijk et al., 2012). As  
669 such it can be used as part of a negotiation support approach to natural resources management in  
670 which levelling off on knowledge and joint fact finding in blame attribution are key steps to  
671 negotiated solutions that are legitimate and seen to be so (van Noordwijk et al., 2013; Leimona et  
672 al., 2015). Quantification of  $F_p$  can help assess tactical management options (Burt et al., 2014) as in a  
673 recent suggestion to minimize negative downstream impacts of forestry operations on stream flow  
674 by avoiding land clearing and planting operations in locally wet La Niña years. But the most  
675 challenging aspect of the management of flood, as any other environmental risk, is that the  
676 frequency of disasters is too low to intuitively influence human behaviour where short-term risk  
677 taking benefits are attractive. Wider social pressure is needed for investment in watershed health  
678 (as a type of insurance premium) to be mainstreamed, as individuals waiting to see evidence of  
679 necessity are too late to respond. In terms of flooding risk, actions to restore or retain watershed  
680 health can be similarly justified as insurance premium. It remains to be seen whether or not the  
681 transparency of the  $F_p$  metric and its intuitive appeal are sufficient to make the case in public debate

682 when opportunity costs of foregoing reductions in flow buffering by profitable land use are to be  
683 compensated and shared (Burt et al., 2014).

#### 684 **5.4 Conclusions and specific questions for a set of case studies**

685 In conclusion, the  $F_p$  metric appears to allow an efficient way of summarizing complex landscape  
686 processes into a single parameter that reflects the effects of landscape management within the  
687 context of the local climate. If rainfall patterns change but the landscape does not, the resultant flow  
688 patterns may reflect a change in watershed health (van Noordwijk et al., 2016). Flow persistence is  
689 the result of rainfall persistence and the temporal delay provided by the pathway water takes  
690 through the soil and the river system. High flow persistence indicates a reliable water supply, while  
691 minimizing peak flow events. Wider tests of the  $F_p$  metric as boundary object in science-practice-  
692 policy boundary chains (Kirchhoff et al., 2015; Leimona et al., 2015) are needed. Further tests for  
693 specific case studies can clarify how changes in tree cover (deforestation, reforestation,  
694 agroforestation) in different contexts influence river flow dynamics and  $F_p$  values. Sensitivity to  
695 specific realizations of underlying time-space rainfall patterns needs to be quantified, before  
696 changes in  $F_p$  can be attributed to changed 'watershed health', rather than chance events.

#### 697 **Data availability**

698 The algorithm used is freely available. Specific data used in the case studies are explained and  
699 accounted for in Part II.

#### 700 **Author contributions**

701 Meine van Noordwijk designed method and paper, Lisa Tanika refined the empirical algorithm and  
702 handled the case study data and modelling for part II, and Betha Lusiana contributed statistical  
703 analysis; all contributed and approved the final manuscript

#### 704 **Acknowledgements**

705 This research is part of the Forests, Trees and Agroforestry research program of the CGIAR. Several  
706 colleagues contributed to the development and early tests of the  $F_p$  method. Thanks are due to Eike  
707 Luedeling, Sonya Dewi, Sampurno Bruijnzeel and three anonymous reviewers for comments on an  
708 earlier version of the manuscript.

#### 709 **References**

710 Alila, Y., Kura, P.K., Schnorbus, M., and Hudson, R.: Forests and floods: A new paradigm sheds light  
711 on age-old controversies, *Water Resour. Res.*, 45, W08416, 2009.

712 Andréassian, V.: Waters and forests: from historical controversy to scientific debate, *Journal of  
713 Hydrology*, 291, 1–27, 2004.

714 Baker, D.B., Richards, R.P., Loftus, T.T. and Kramer, J.W.: A new flashiness index: Characteristics and  
715 applications to midwestern rivers and streams. *Journal of the American Water Resources  
716 Association*, Paper No. 03095, 2004.

717 Baldassarre, G.D., Kooy, M., Kemerink, J.S., and Brandimarte, L.: Towards understanding the dynamic  
718 behaviour of floodplains as human-water systems, *Hydrology and Earth System Sciences*, 17(8),  
719 3235–3244, 2013.

720 Bardossy, A. and Plate, E.J.: Modeling daily rainfall using a semi-Markov representation of  
721 circulation pattern occurrence, *Journal of Hydrology*, 122(1), 33-47, 1991.

722 Band, L.E., Patterson, P., Nemani, R., and Running, S.W.: Forest ecosystem processes at the  
723 watershed scale: incorporating hillslope hydrology, *Agricultural and Forest Meteorology*, 63(1-2),  
724 93-126, 1993.

725 Barnes, B.S.: The structure of discharge-recession curves, *Eos, Transactions American Geophysical  
726 Union*, 20(4), 721-725, 1939.

727 Beck, H. E., Bruijnzeel, L. A., Van Dijk, A. I. J. M., McVicar, T. R., Scatena, F. N., and Schellekens, J.: The  
728 impact of forest regeneration on streamflow in 12 mesoscale humid tropical  
729 catchments, *Hydrology and Earth System Sciences*, 17(7), 2613-2635, 2013.

730 Bergström, S.: The HBV model. In: Singh, V.P., (Ed.), *Computer Models of Watershed Hydrology*, Ch.  
731 13, pp. 443-476, Water Resources Publications, Highlands Ranch, Colorado, USA, 1130 pp., 1995.

732 Beven, K.J.: Rainfall-runoff modelling: the primer, John Wiley & Sons, 2011

733 Bishop, J. and Pagiola, S. (Eds.): *Selling forest environmental services: market-based mechanisms for  
734 conservation and development*, Taylor & Francis, Abingdon (UK), 2012.

735 Bonell, M.: Progress in the understanding of runoff generation dynamics in forests, *Journal of  
736 Hydrology*, 150, 217-275, 1993.

737 Bonell, M. and Bruijnzeel, L.A. (eds.): *Forests, water and people in the humid tropics: past, present  
738 and future hydrological research for integrated land and water management*. Cambridge  
739 University Press, Cambridge (UK), 2005.

740 Bradshaw, C.J.A., Sodhi, N.S., Peh, K.S.H., and Brook, B.W.: Global evidence that deforestation  
741 amplifies flood risk and severity in the developing world, *Global Change Biol*, 13, 2379–2395,  
742 2007.

743 Brauman, K.A., Daily, G.C., Duarte, T.K.E., and Mooney, H.A.: The nature and value of ecosystem  
744 services: an overview highlighting hydrologic services, *Annu. Rev. Environ. Resour.*, 32, 67-98,  
745 2007.

746 Bruijnzeel, L.A.: *Hydrology of Moist Tropical Forests and Effects of Conversion: a State of Knowledge  
747 Review*. IHP-UNESCO Humid Tropical Programme, Paris, 224 pp, 1990.

748 Bruijnzeel, L.A.: Hydrological functions of tropical forests: not seeing the soil for the trees, *Agr.  
749 Ecosyst. Environ.*, 104, 185–228, 2004.

750 Burt, T.P., Howden, N.J.K., McDonnell, J.J., Jones, J.A., and Hancock, G.R.: Seeing the climate through  
751 the trees: observing climate and forestry impacts on streamflow using a 60-year record,  
752 *Hydrological Processes*, doi: 10.1002/hyp.10406, 2014.

753 City of Fort Collins, River Health Assessment Framework la Poudre River.  
754 <http://www.fcgov.com/naturalareas/pdf/river-health-report-final-appendix.pdf>, 2015.

755 Clark, W. C., Tomich, T. P., van Noordwijk, M., Guston, D., Catacutan, D., Dickson, N. M., and McNie,  
756 E.: Boundary work for sustainable development: natural resource management at the  
757 Consultative Group on International Agricultural Research (CGIAR), *Proc. Nat. Acad. Sci.*,  
758 doi:10.1073/pnas.0900231108, 2011.

759 Delfs, J.O., Park, C.H., and Kolditz, O.: A sensitivity analysis of Hortonian flow, *Advances in Water*  
760 *Resources*, 32(9), 1386-1395, 2009.

761 Efstratiadis, A., Koussis, A.D., Koutsoyiannis, D. and Mamassis, N.: Flood design recipes vs. reality: can  
762 predictions for ungauged basins be trusted? *Natural Hazards and Earth System Sciences*, 14(6), 1417-1428,  
763 2014.

764 Evaristo, J., Jasechko, S., and McDonnell, J.J.: Global separation of plant transpiration from  
765 groundwater and streamflow, *Nature*, 525(7567), 91-94, 2015.

766 Farber, S.C., Costanza, R. and Wilson, M.A.: Economic and ecological concepts for valuing ecosystem  
767 services. *Ecological economics*, 41(3), 375-392, 2002.

768 Furey, P.R. and Gupta, V.K.: A physically based filter for separating base flow from streamflow time  
769 series. *Water Resources Research*, 37(11), 2709-2722, 2001.

770 Galudra, G. and Sirait, M.: A discourse on Dutch colonial forest policy and science in Indonesia at the  
771 beginning of the 20th century. *International Forestry Review* 11(4), 524-533, 2009

772 Gassman, P.W., Reyes, M.R., Green, C.H. and Arnold, J.G.: The soil and water assessment tool: historical  
773 development, applications, and future research directions. *Transactions of the ASABE*, 50(4), 1211-1250,  
774 2007.

775 Graf, W.L., Wohl, E., Sinha, T. and Sabo, J.L., 2010. Sedimentation and sustainability of western American  
776 reservoirs. *Water Resources Research*, 46(12), 2010.

777 Grimaldi, S., Petroselli, A. and Romano, N.: Green-Ampt Curve-Number mixed procedure as an empirical tool  
778 for rainfall-runoff modelling in small and ungauged basins. *Hydrological Processes*, 27(8), 1253-1264,  
779 2013.

780 Ghimire, C.P., Bruijnzeel, L.A., Lubczynski, M.W., and Bonell, M.: Negative trade-off between changes  
781 in vegetation water use and infiltration recovery after reforesting degraded pasture land in the  
782 Nepalese Lesser Himalaya, *Hydrology and Earth System Sciences*, 18(12), 4933-4949, 2014.

783 Herschy, R.W.: The world's maximum observed floods. *Flow Measurement and Instrumentation* 13  
784 (5-6), 231-235, 2002.

785 Hornberger, G.M., Wiberg, P.L., D'Odorico, P. and Raffensperger, J.P.: *Elements of physical*  
786 *hydrology*. John Hopkins University Press, 2014.

787 Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W.,  
788 Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H.V.,  
789 Hughes, D. A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S.,  
790 Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E., and Cudennec, C.: A decade of Predictions  
791 in Ungauged Basins (PUB)—a review, *Hydrological sciences journal*, 58(6), 1198-1255, 2013.

792 Ilstedt, U., Tobella, A.B., Bazié, H.R., Bayala, J., Verbeeten, E., Nyberg, G., Sanou, J., Benegas, L.,  
793 Murdiyarsa, D., Laudon, H., and Sheil, D.: Intermediate tree cover can maximize groundwater  
794 recharge in the seasonally dry tropics, *Scientific reports*, 6, 2016

795 Jones, P.G. and Thornton, P.K.: Spatial and temporal variability of rainfall related to a third-order  
796 Markov model, *Agricultural and Forest Meteorology*, 86(1), 127-138, 1997.

797 Jongman, B., Winsemius, H. C., Aerts, J. C., de Perez, E. C., van Aalst, M. K., Kron, W., and Ward, P. J.:  
798 Declining vulnerability to river floods and the global benefits of adaptation, In: *Proceedings of*  
799 *the National Academy of Sciences*, 112(18), E2271-E2280, 2015.

800 Joshi, L., Schalenbourg, W., Johansson, L., Khasanah, N., Stefanus, E., Fagerström, M.H., and van  
801 Noordwijk, M.: Soil and water movement: combining local ecological knowledge with that of  
802 modellers when scaling up from plot to landscape level, In: van Noordwijk, M., Cadisch, G. and  
803 Ong, C.K. (Eds.) *Belowground Interactions in Tropical Agroecosystems*, CAB International,  
804 Wallingford (UK), 349-364, 2004.

805 Kirchhoff, C.J., Esselman, R., and Brown, D.: Boundary Organizations to Boundary Chains: Prospects  
806 for Advancing Climate Science Application, Climate Risk Management,  
807 doi:10.1016/j.crm.2015.04.001, 2015.

808 Kirchner, J.W.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff  
809 modeling, and doing hydrology backward. *Water Resources Research*, 45(2).  
810 DOI: 10.1029/2008WR006912, 2009.

811 Lacombe, G. and McCartney, M.: Evaluating the flow regulating effects of ecosystems in the Mekong  
812 and Volta river basins, Colombo, Sri Lanka, International Water Management Institute (IWMI)  
813 Research Report 166, doi: 10.5337/2016.202, 2016.

814 Lacombe, G., Ribolzi, O., de Rouw, A., Pierret, A., Latsachak, K., Silvera, N., Pham Dinh, R., Orange, D.,  
815 Janeau, J.-L., Soulileuth, B., Robain, H., Taccoen, A., Sengphaathith, P., Mouche, E.,  
816 Sengtaheuanghoun, O., Tran Duc, T., and Valentin, C.: Afforestation by natural regeneration or  
817 by tree planting: Examples of opposite hydrological impacts evidenced by long-term field  
818 monitoring in the humid tropics, *Hydrology and Earth System Sciences Discussions*, 12, 12615-  
819 12648, 2015.

820 Leimona, B., Lusiana, B., van Noordwijk, M., Mulyoutami, E., Ekadinata, A., and Amaruzama, S.:  
821 Boundary work: knowledge co-production for negotiating payment for watershed services in  
822 Indonesia, *Ecosystems Services*, 15, 45–62, 2015.

823 Liu, W., Wei, X., Fan, H., Guo, X., Liu, Y., Zhang, M., and Li, Q.: Response of flow regimes to  
824 deforestation and reforestation in a rain-dominated large watershed of subtropical  
825 China, *Hydrological Processes*, 29, 5003-5015 , 2015.

826 Lusiana, B., van Noordwijk, M., Suyamto, D., Joshi, L., and Cadisch, G.: Users' perspectives on validity  
827 of a simulation model for natural resource management, *International Journal of Agricultural  
828 Sustainability*, 9(2), 364-378, 2011.

829 Ma, X., Xu, J. and van Noordwijk, M.: Sensitivity of streamflow from a Himalayan catchment to  
830 plausible changes in land cover and climate, *Hydrological Processes*, 24, 1379–1390, 2010.

831 Ma, X., Lu, X., van Noordwijk, M., Li, J.T., and Xu, J.C.: Attribution of climate change, vegetation  
832 restoration, and engineering measures to the reduction of suspended sediment in the Kejie  
833 catchment, southwest China, *Hydrol. Earth Syst. Sci.*, 18, 1979–1994, 2014.

834 Maidment, D.R.: *Handbook of hydrology*, McGraw-Hill Inc., 1992.

835 Malmer, A., Murdiyarso, D., Bruijnzeel L.A., and Ilstedt, U.: Carbon sequestration in tropical forests  
836 and water: a critical look at the basis for commonly used generalizations, *Global Change  
837 Biology*, 16(2), 599-604, 2010.

838 Marchi, L., Borga, M., Preciso, E. and Gaume, E.: Characterisation of selected extreme flash floods in Europe  
839 and implications for flood risk management. *Journal of Hydrology*, 394(1), 118-133, 2010.

840 McCluney, K.E., Poff, N.L., Palmer, M.A., Thorp, J.H., Poole, G.C., Williams, B.S., Williams, M.R., and  
841 Baron, J.S.: Riverine macrosystems ecology: sensitivity, resistance, and resilience of whole river  
842 basins with human alterations, *Frontiers in Ecology and the Environment*, 12(1), 48-58, 2014.

843 Milly, P.C.D., Wetherald, R., Dunne, K.A., and Delworth, T.L.: Increasing risk of great floods in a  
844 changing climate, *Nature*, 415(6871), 514-517, 2002.

845 Oliveira, P.T.S., Nearing, M.A., Hawkins, R.H., Stone, J.J., Rodrigues, D.B.B., Panachuki, E. and Wendland, E.:  
846 Curve number estimation from Brazilian Cerrado rainfall and runoff data. *Journal of Soil and Water  
847 Conservation*, 71(5), 420-429, 2016

848 Pahl-Wostl, C., Arthington, A., Bogardi, J., Bunn, S.E., Hoff, H., Lebel, L., Nikitina, E., Palmer, M., Poff, L.N.,  
849 Richards, K. and Schlüter, M.: Environmental flows and water governance: managing sustainable water  
850 uses. *Current Opinion in Environmental Sustainability*, 5(3), pp.341-351, 2013.

851 Palmer, M.A.: Reforming watershed restoration: science in need of application and applications in  
852 need of science, *Estuaries and coasts*, 32(1), 1-17, 2009.

853 Poff, N.L. and Zimmerman, J.K.: Ecological responses to altered flow regimes: a literature review to inform the  
854 science and management of environmental flows. *Freshwater Biology*, 55(1), 194-205, 2010.

855 Poff, N.L., Allan, J.D., Bain, M.B., Karr, J.R., Prestegaard, K.L., Richter, B.D., Sparks, R.E. and Stromberg, J.C.: The  
856 natural flow regime. *BioScience*, 47(11), 769-784, 1997.

857 Poff, N.L., Richter, B.D., Arthington, A.H., Bunn, S.E., Naiman, R.J., Kendy, E., Acreman, M., Apse, C., Bledsoe,  
858 B.P., Freeman, M.C. and Henriksen, J.: The ecological limits of hydrologic alteration (ELOHA): a new  
859 framework for developing regional environmental flow standards. *Freshwater Biology*, 55(1), 147-170,  
860 2010

861 Ponce, V.M. and Hawkins, R.H.: Runoff curve number: Has it reached maturity? *Journal of Hydrologic  
862 Engineering*, 1(1), 11-19, 1996

863 Rahayu, S., Widodo, R.H., van Noordwijk, M., Suryadi, I., and Verbist, B.: Water monitoring in  
864 watersheds. Bogor, Indonesia, World Agroforestry Centre (ICRAF) SEA Regional Program,, 2013

865 Ranieri, S.B.L., Stirzaker, R., Suprayogo, D., Purwanto, E., de Willigen, P., and van Noordwijk, M.:  
866 Managing movements of water, solutes and soil: from plot to landscape scale. In: van Noordwijk,  
867 M., Cadisch, G. and Ong, C.K. (Eds.) *Belowground Interactions in Tropical Agroecosystems*, CAB  
868 International, Wallingford (UK), 329-347, 2004.

869 Richter, B.D., Mathews, R., Harrison, D.L., and Wigington, R.: Ecologically sustainable water  
870 management: managing river flows for ecological integrity, *Ecological applications*, 13(1), 206-  
871 224, 2003.

872 Rodríguez-Iturbe, I. and Rinaldo, A.: *Fractal river basins: chance and self-organization*, Cambridge  
873 University Press, Cambridge, 2001.

874 Rose, C.W.: *An introduction to the environmental physics of soil, water and watersheds*. Cambridge  
875 University Press, Cambridge (UK) 2004.

876 Rudel, T. K., Coomes, O. T., Moran, E., Achard, F., Angelsen, A., Xu, J., and Lambin, E.: Forest  
877 transitions: towards a global understanding of land use change, *Global Environmental  
878 Change*, 15(1), 23-31, 2005.

879 Seibert, J.: Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and  
880 Forest Meteorology* 98-99: 279-293, 1999.

881 Seibert, J., Beven, K.J.: Gauging the ungauged basin: how many discharge measurements are  
882 needed? *Hydrol. Earth Syst. Sci.*, 13, 883–892, 2009

883 Sloto, R.A. and Crouse, M.Y.: *HYSEP, a computer program for streamflow hydrograph separation and*  
884 *analysis*. US Department of the Interior, US Geological Survey, 1996.

885 Sumarga, E., Hein, L., Hooijer, A., and Vernimmen, R.: Hydrological and economic effects of oil palm  
886 cultivation in Indonesian peatlands, *Ecology and Society*, 21(2), 2016.

887 Tallaksen, L.M.: A review of baseflow recession analysis, *J Hydrol.*, 165, 349-370, 1995.

888 Tan-Soo, J.S., Adnan, N., Ahmad, I., Pattanayak, S.K., and Vincent, J.R.: Econometric Evidence on  
889 Forest Ecosystem Services: Deforestation and Flooding in Malaysia. *Environmental and Resource*  
890 *Economics*, on-line: <http://link.springer.com/article/10.1007/s10640-014-9834-4>, 2014.

891 Turner, R.K. and Daily, G.C.: The ecosystem services framework and natural capital  
892 conservation. *Environmental and Resource Economics*, 39(1), 25-35, 2008

893 Van de Giesen, N.C., Stomph, T.J., and De Ridder, N.: Scale effects of Hortonian overland flow and  
894 rainfall-runoff dynamics in a West African catena landscape, *Hydrological Processes*, 14, 165-  
895 175, 2000.

896 Van den Putte, A., Govers, G., Leys, A., Langhans, C., Clymans, W. and Diels, J.: Estimating the parameters of  
897 the Green-Ampt infiltration equation from rainfall simulation data: Why simpler is better. *Journal of*  
898 *Hydrology*, 476, 332-344, 2013.

899 van Dijk, A.I., van Noordwijk, M., Calder, I.R., Bruijnzeel, L.A., Schellekens, J., and Chappell, N.A.:  
900 Forest-flood relation still tenuous – comment on ‘Global evidence that deforestation amplifies  
901 flood risk and severity in the developing world’, *Global Change Biology*, 15, 110-115, 2009.

902 van Noordwijk, M., van Roode, M., McCallie, E.L., and Lusiana, B.: Erosion and sedimentation as  
903 multiscale, fractal processes: implications for models, experiments and the real world, In: F.  
904 Penning de Vries, F. Agus and J. Kerr (Eds.) *Soil Erosion at Multiple Scales, Principles and*  
905 *Methods for Assessing Causes and Impacts*. CAB International, Wallingford, 223-253, 1998.

906 van Noordwijk, M., Agus, F., Verbist, B., Hairiah, K., and Tomich, T.P.: Managing Watershed Services,  
907 In: Scherr, S.J. and McNeely, J.A. (Eds) *Ecoagriculture Landscapes. Farming with Nature: The*  
908 *Science and Practice of Ecoagriculture*, Island Press, Washington DC, 191 – 212, 2007.

909 van Noordwijk, M., Widodo, R.H., Farida, A., Suyamto, D., Lusiana, B., Tanika, L., and Khasanah, N.:  
910 GenRiver and FlowPer: Generic River and Flow Persistence Models. User Manual Version 2.0.  
911 Bogor, Indonesia, World Agroforestry Centre (ICRAF) Southeast Asia Regional Program, 2011.

912 van Noordwijk, M., Leimona, B., Jindal, R., Villamor, G.B., Vardhan, M., Namirembe, S., Catacutan, D.,  
913 Kerr, J., Minang, P.A., and Tomich, T.P.: Payments for Environmental Services: evolution towards  
914 efficient and fair incentives for multifunctional landscapes, *Annu. Rev. Environ. Resour.*, 37, 389-  
915 420, 2012.

916 van Noordwijk, M., Lusiana, B., Leimona, B., Dewi, S., and Wulandari, D.: Negotiation-support toolkit  
917 for learning landscapes, Bogor, Indonesia, World Agroforestry Centre (ICRAF) Southeast Asia  
918 Regional Program, 2013.

919 van Noordwijk, M., Leimona, B., Xing, M., Tanika, L., Namirembe, S., and Suprayogo, D.: Water-  
920 focused landscape management. Climate-Smart Landscapes: Multifunctionality In Practice. eds  
921 Minang PA et al.. Nairobi, Kenya, World Agroforestry Centre (ICRAF), 179-192, 2015a.

922 van Noordwijk, M., Bruijnzeel, S., Ellison, D., Sheil, D., Morris, C., Gutierrez, V., Cohen, J., Sullivan, C.,  
923 Verbist, B., and Muys, B.: Ecological rainfall infrastructure: investment in trees for sustainable  
924 development, ASB Brief no 47. Nairobi. ASB Partnership for the Tropical Forest Margins, 2015b.

925 van Noordwijk M, Kim Y-S, Leimona B, Hairiah K, Fisher LA,: Metrics of water security, adaptive  
926 capacity and Agroforestry in Indonesia. Current Opinion on Environmental Sustainability (in  
927 press: <http://dx.doi.org/10.1016/j.cosust.2016.10.004>), 2016.

928 Verbist, B., Poesen, J., van Noordwijk, M. Widianto, Suprayogo, D., Agus, F., and Deckers, J.: Factors  
929 affecting soil loss at plot scale and sediment yield at catchment scale in a tropical volcanic  
930 Agroforestry landscape, Catena, 80, 34-46, 2010.

931 Ward, R.C. and Robinson, M.: *Principles of hydrology*. 4<sup>th</sup> edition. New York: McGraw-Hill, 2000.

932 Weiler, M. and McDonnell, J.: Virtual experiments: a new approach for improving process  
933 conceptualization in hillslope hydrology, Journal of Hydrology, 285(1), 3-18, 2004

934 Winsemius, H.C., van Beek, L.P.H., Jongman, B., Ward, P.J., and Bouwman, A.: A framework for global  
935 river flood risk assessments, Hydrol Earth Syst Sci, 17, 1871–1892, 2013.

936 World Health Organization: The first ten years of the World Health Organization. Geneva: WHO,  
937 1958

938 Yen, H., White, M.J., Jeong, J., Arabi, M. and Arnold, J.G.: Evaluation of alternative surface runoff  
939 accounting procedures using the SWAT model, International Journal of Agricultural and  
940 Biological Engineering, 8(3), 54-68, 2015.

941 Zhou, G., Wei, X., Luo, Y., Zhang, M., Li, Y., Qiao, Y., Liu, H., and Wang, C.: Forest recovery and river  
942 discharge at the regional scale of Guangdong Province, China, Water Resources Research, 46(9),  
943 W09503, doi:10.1029/2009WR00829, 2010.

944

## A. Interests↔Understanding↔Metrics

### multistakeholder resource management processes

→ Monitoring → Diagnosis → Tradeoff analysis → Innovation → Scenarios → Negotiations →

#### Basis of current land use policies:

Deforestation → increased flood risk  
Reforestation → reduced flood risk

*Forestry perspective*

#### Ecohydrology perspective

**Relationship between land cover & river flow**  
depends on complex interactions, non-linearities, partial reversibility, climate variability

**Engineering of river storage and flow** can control all relevant risks, once these are quantified

*Engineering perspective*

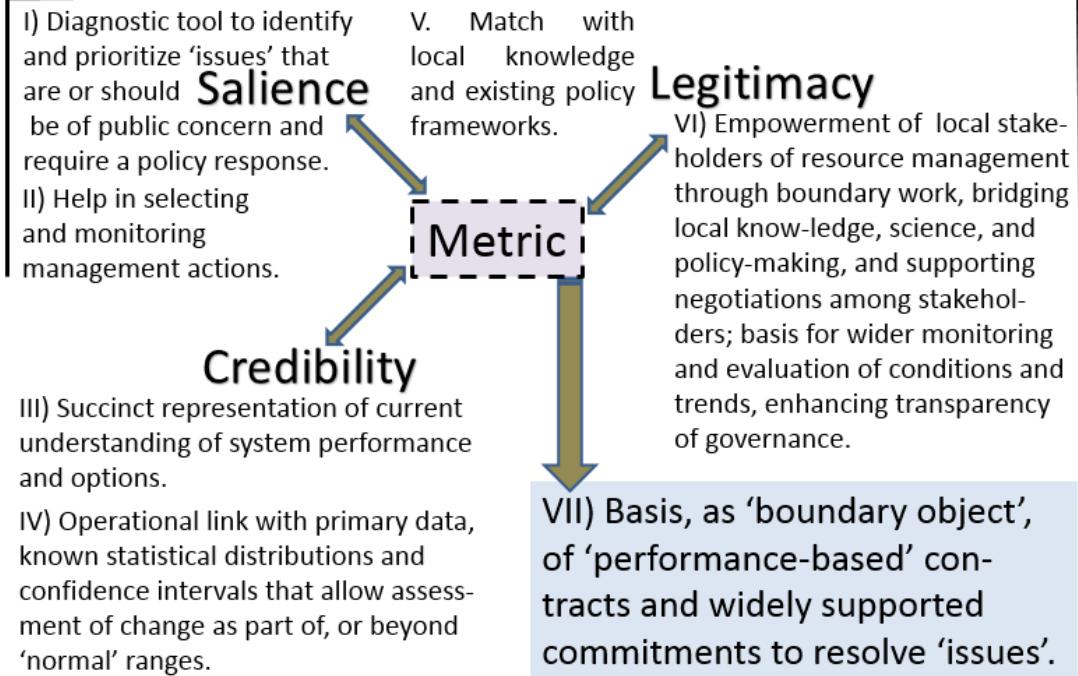
#### Climate Change adaptation view

**Climate change** creates new challenges, requiring costly adaptation measures, but climate policy and finance needs clear attribution, cause & effect links

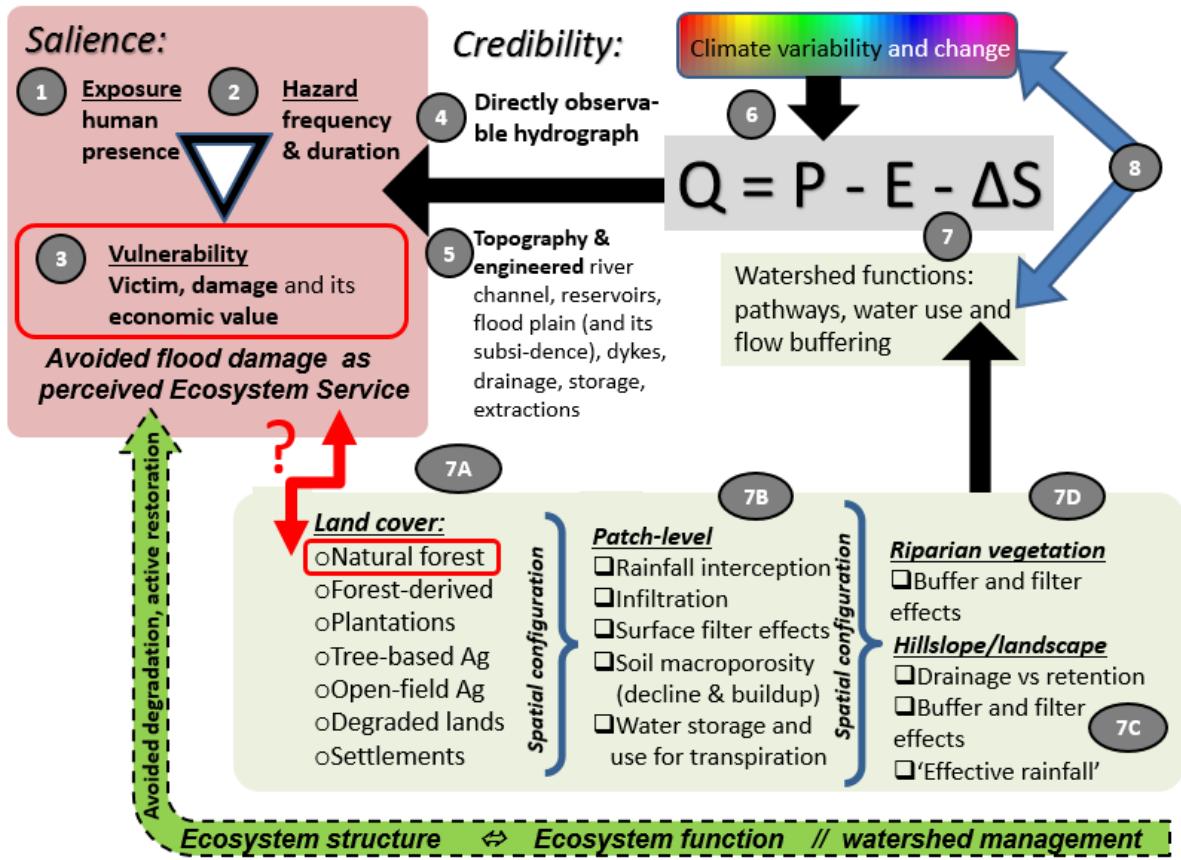
**Local land users** want river flow to be **predictable** but also like to have flexibility in how land use is regulated as part of ecosystem services management

*Local landscape stakeholders*

## B.

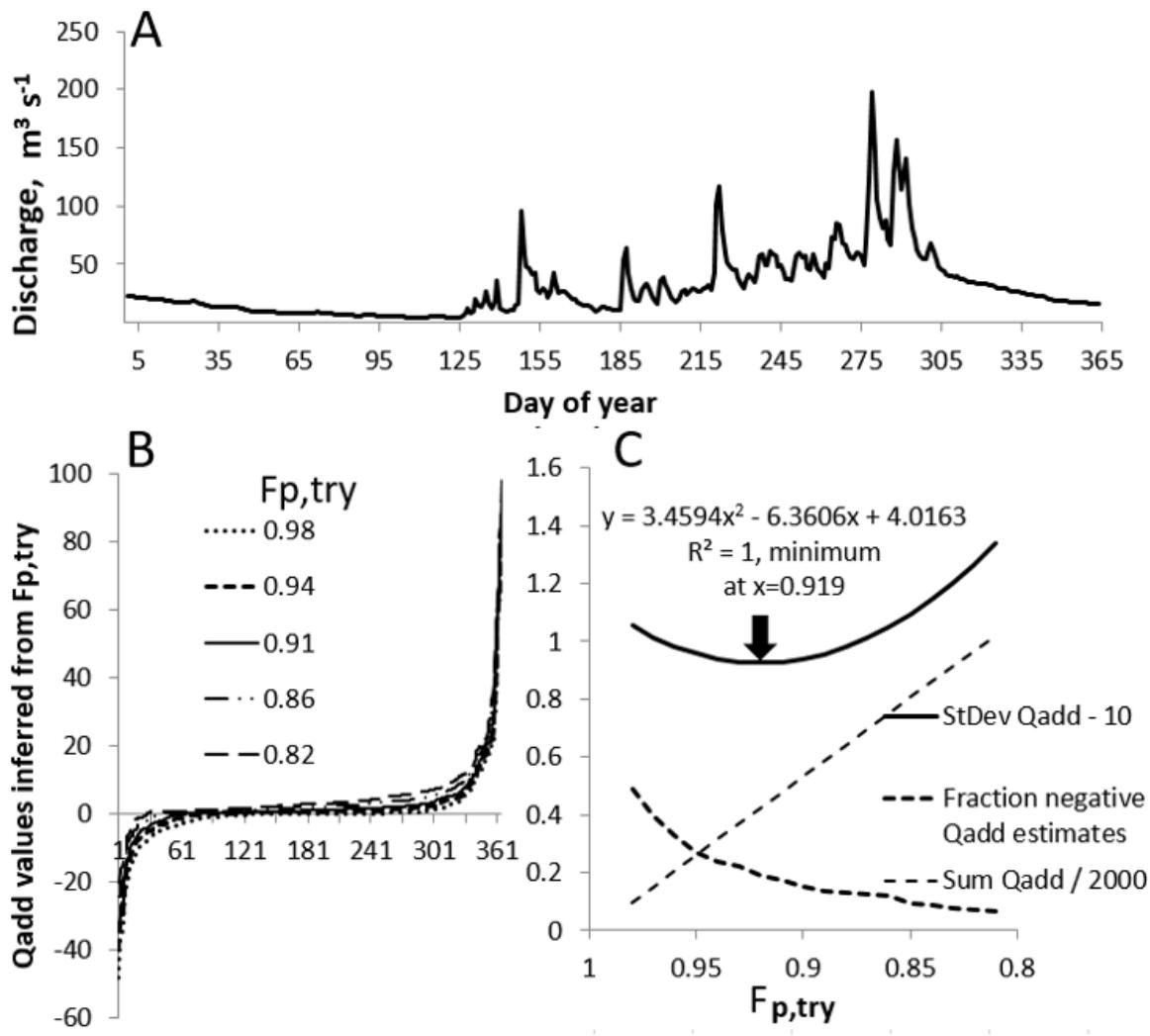


948 Figure 1. A. Multiple perspectives on the way flood risk is to be understood, monitored and handled  
949 according to different knowledge systems; B. Basic requirements for a 'metric' to be used in public  
950 discussions of natural resource management issues that deserve to be resolved and acted upon  
951 (modified from van Noordwijk et al., 2016)



952

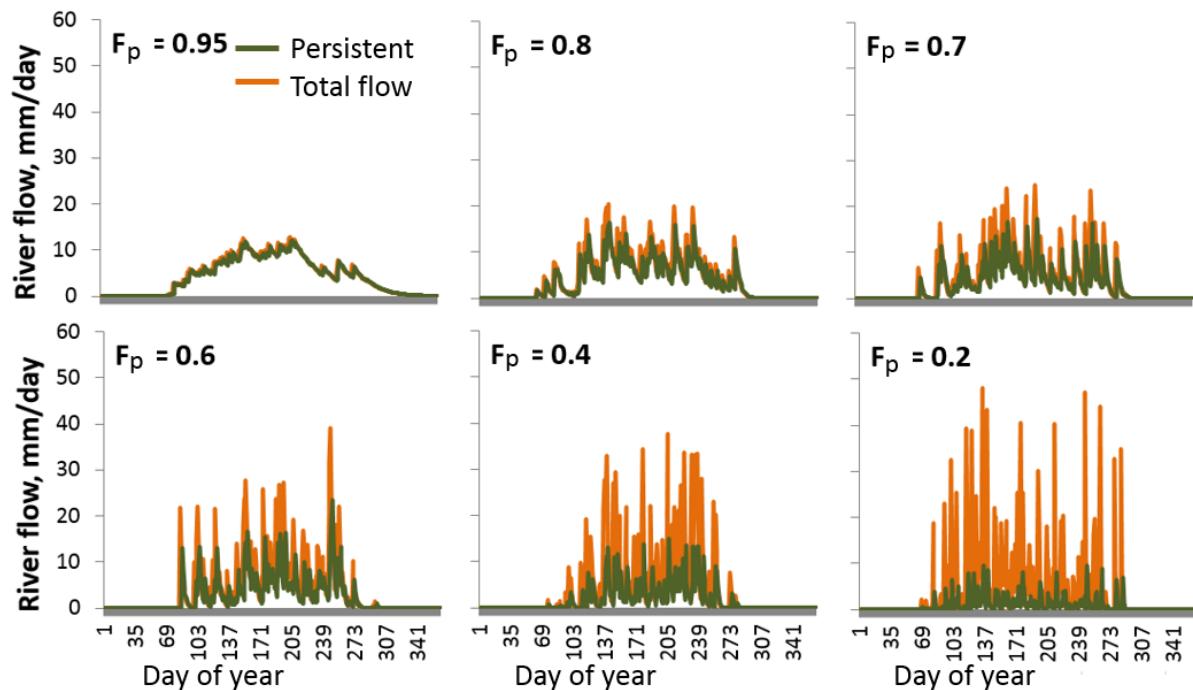
953 Figure 2. Steps in a causal pathway that relates the salience of 'avoided flood damage as  
 954 ecosystem service' to the interaction of exposure (1; being in the wrong place at critical  
 955 times), hazard (2; spatially explicit flood frequency and duration) and human determinants  
 956 of vulnerability (3); the hazard component depends, in common scientific analysis, on the  
 957 pattern of river flow described in a hydrograph (4), which in turn is understood to be  
 958 influenced by conditions along the river channel (5), precipitation and potential  
 959 evapotranspiration ( $E_{pot}$ ) as climatic factors (6) and the condition in the watershed (7)  
 960 determining evapotranspiration ( $E_{act}$ ), temporary water storage ( $\Delta S$ ) and water partitioning  
 961 over overland flow and infiltration; these watershed functions in turn depend on the  
 962 interaction of terrain (topography, soils, geology), vegetation and human land use; current  
 963 understanding of a two-way interaction between vegetation and rainfall adds further  
 964 complexity (8)



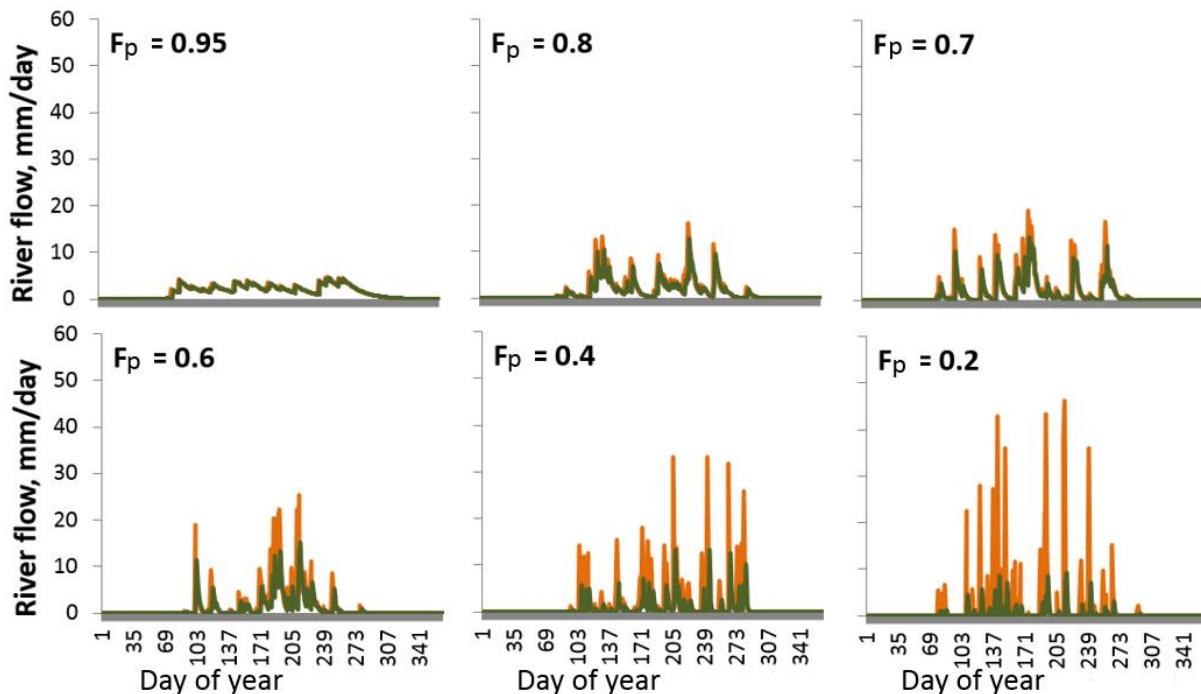
965

966 Figure 3. Example of the derivation of best fitting  $F_{p,try}$  value for an example hydrograph (A) on the  
 967 basis of the inferred  $Q_a$  distribution (cumulative frequency in B), and three properties of this  
 968 distribution (C): its sum, frequency of negative values and standard deviation; the  $F_{p,try}$  minimum  
 969 of the latter is derived from the parameters of a fitted quadratic equation

970

**A. 120 rainy days,  $Q \sim 1600 \text{ mm/yr}$** 

971

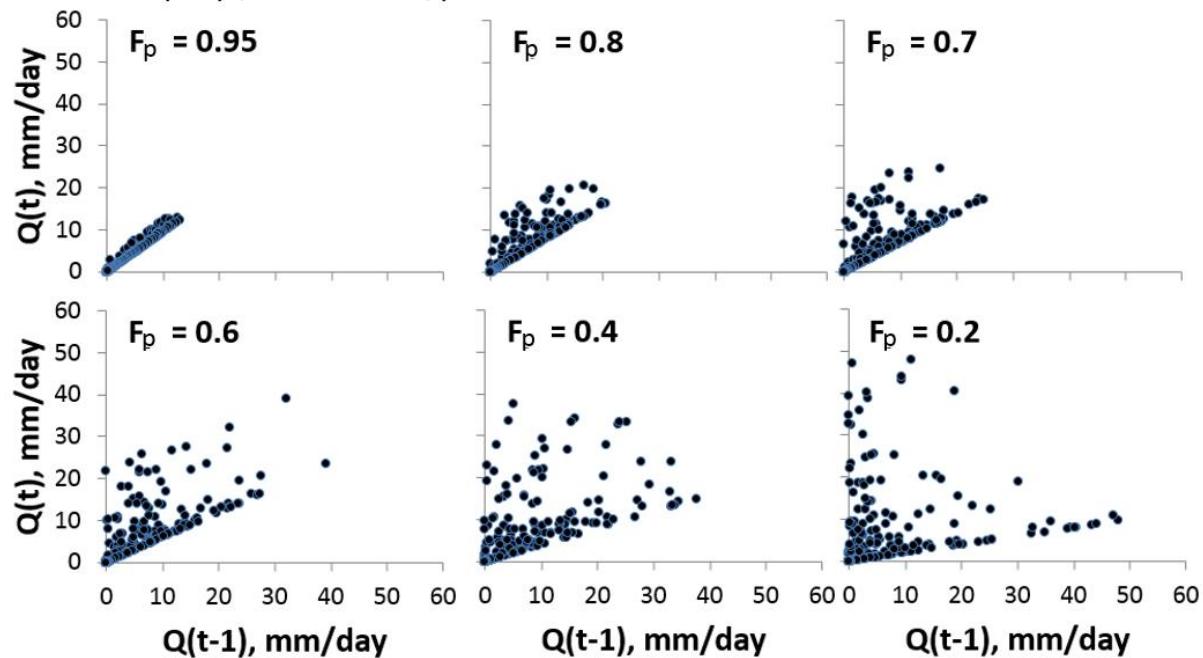
**B. 45 rainy days,  $Q \sim 600 \text{ mm/yr}$** 

972

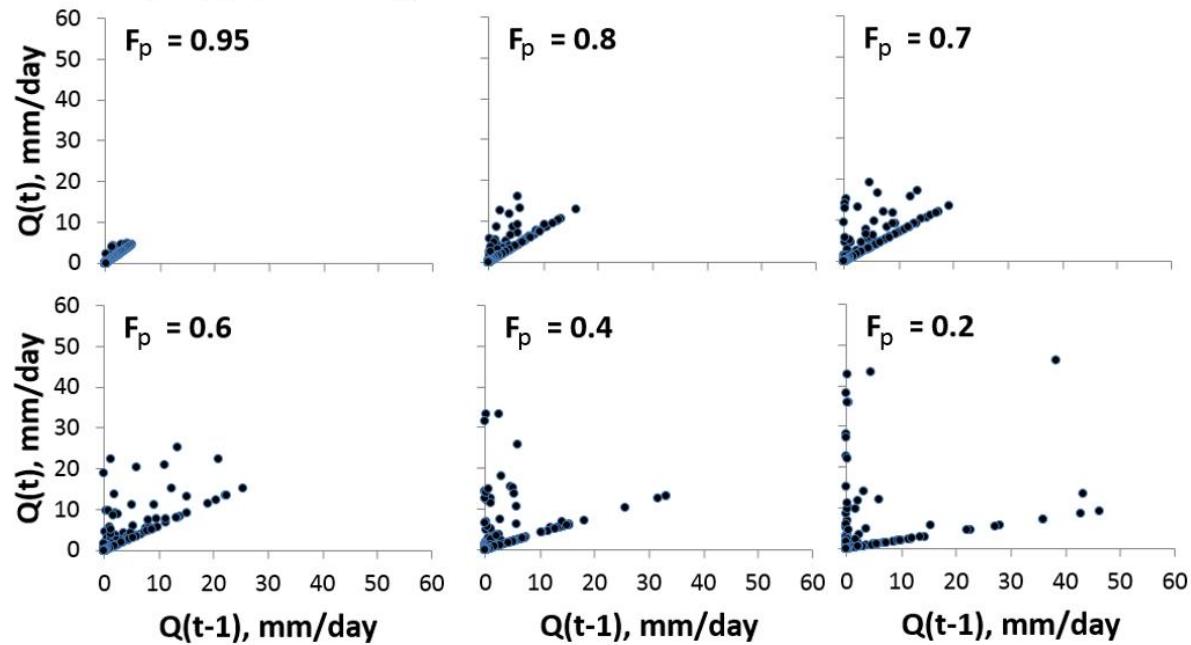
Figure 4. Effects of the  $F_p$  parameter on hydrographs of daily river flow generated by a random rainfall generator, with persistent and additional flow components indicated, for two settings with total rainfall of approximately 1600 and 600 mm/yr (NB river flow is here expressed as  $\text{mm d}^{-1}$  rather than as  $\text{m}^3 \text{s}^{-1}$  as in figure 3)

977

978

**A.** 120 rainy days,  $Q \sim 1600 \text{ mm/yr}$ 

979

**B.** 45 rainy days,  $Q \sim 600 \text{ mm/yr}$ 

980

981 Figure 5A and B. **Temporal autocorrelation of river flow** for the same simulations as Figure 4; the  
 982 lower envelope of the points indicated slope  $F_p$ , the points above this line the effect of fresh  
 983 additions to river flow

984

985

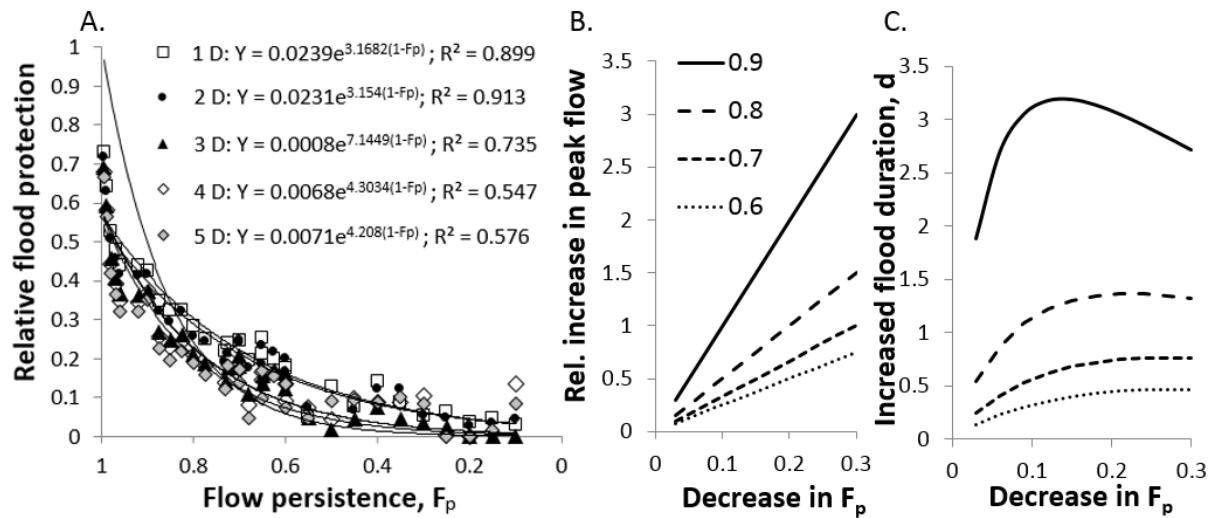
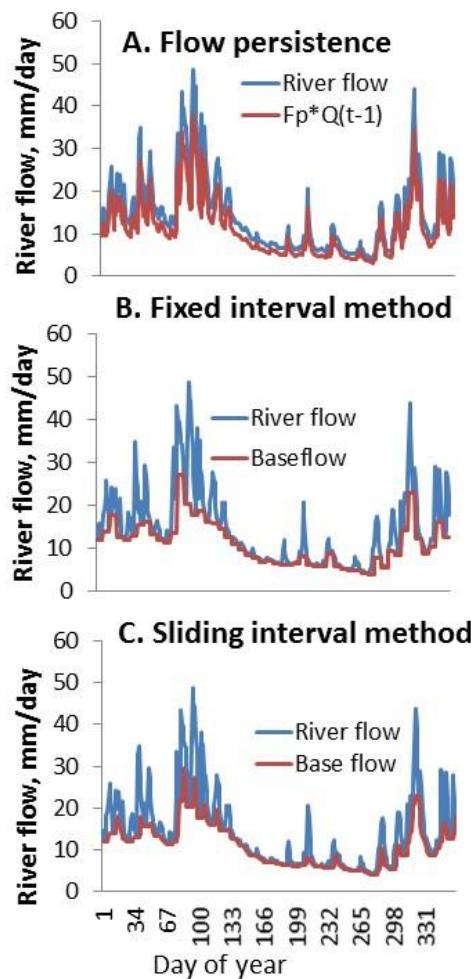
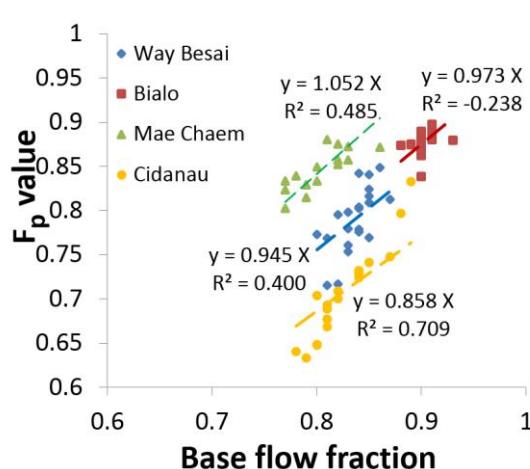


Figure 6. A. Effects of flow persistence on the relative flood protection (decrease in maximum flow measured over a 1 – 5 d period relative to a case with  $F_p = 0$  (a few small negative points were replaced by small positive values to allow the exponential fit); B and C. effects of a decrease in flow persistence on the volume of water involved in peak flows (B; relative to the volume at  $F_p$  is 0.6 – 0.9) and in the duration (in d) of floods (C)



995  
996 Figure 7. Comparison of base flow separation of a hydrograph according to the flow  
997 persistence method (A) and two common flow separation methods, respectively with  
998 fixed (B) and sliding intervals (C)  
999



1000  
1001  
1002 Figure 8. Comparison of yearly data for four Southeast Asian watersheds analysed with  
1003 common flow separation methods (average of results in Fig. 7) and the flow persistence  
1004 method  
1005  
1006

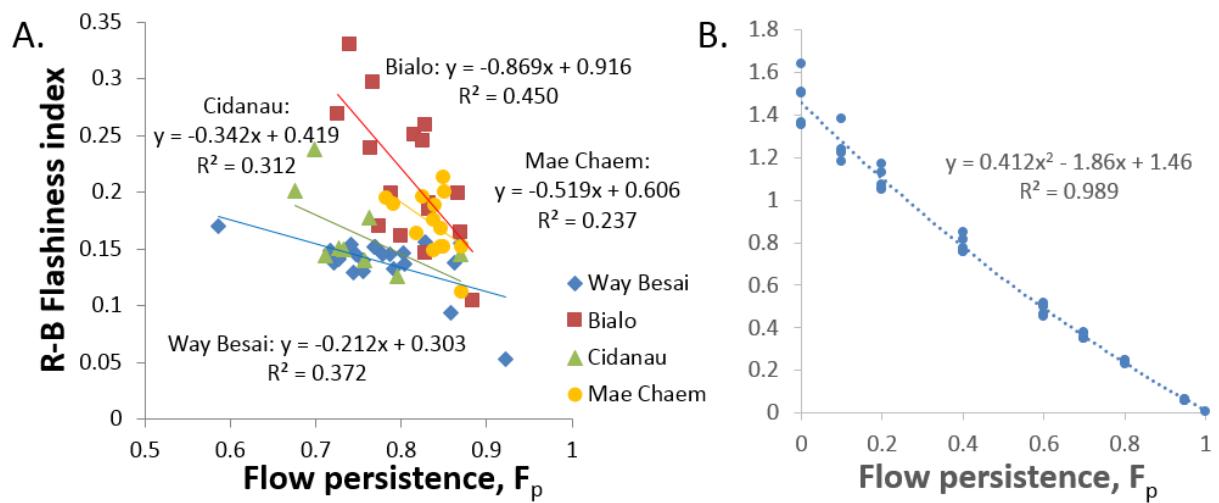


Figure 9. Comparison of the Richards-Baker Flashiness Index (Baker et al., 2004) and the flow persistence metric  $F_p$  for A) four Southeast Asian watersheds, B) a series of hydrographs as in Fig. 4A, with 5 replicates per  $F_p$  value

1007  
1008  
1009  
1010  
1011

Table 1. Comparison of properties of the Flashiness Index and Flow persistence  $F_p$ 

| Flashiness Index (Baker et al. 2004)   | Flow persistence (as defined here)   |
|--|--|
| 1. Has direct appeal to non-technical audiences  | Potentially similar  |
| 2. Where reservoir management rules imply major changes in $\Delta S$ , flashiness still describes implications for flow regimes                                     | Is focused on the effects of changes in (upper) catchment land cover, not where reservoir management determines flow                           |
| 3. Values depend on the scale of evaluating river flow; no absolute criteria for what is 'healthy'   | Similar  |
| 4. Increase generally not desirable  | Decrease generally not desirable   |
| 5. Varies in range [0-2], may need normalizing by division by 2  | Varies in range [0-1]  |
| 6. Requires full year flow record to be calculated   | Can be estimated from any set of sequential flow observations  |
| 7. Empirical metric, no direct link to underlying process understanding  | Overall $F_p$ can be understood as weighted average of the $F_p$ 's of contributing flow pathways (overland, subsurface and groundwater-based) |
| 8. No directly visible relationship between peak and low flow characteristics  | The $F_p$ term low flows and the $(1 - F_p)$ term for peak flows show the water balance logic of a link between peak and low flows             |
| 9. Aggregates changes in flow regime; no directly visible link between the performance metric, rainfall (or snow melt) and (vegetation dependent) evapotranspiration | The main water balance terms are directly reflected in the flow descriptions based on $F_p$  |
| 10. Substantial empirical data bases available for comparison and meta studies   | Not yet  |

1014 Flood risk reduction and flow buffering as ecosystem  
1015 services: II. Land use and rainfall intensity effects in  
1016 Southeast Asia

1017 Meine van Noordwijk<sup>1,2</sup>, Lisa Tanika<sup>1</sup>, Betha Lusiana<sup>1</sup>

1018 [1]{World Agroforestry Centre (ICRAF), SE Asia program, Bogor, Indonesia}

1019 [2]{Wageningen University, Plant Production Systems, Wageningen, the Netherlands}

1020 Correspondence to: Meine van Noordwijk ([m.vannoordwijk@cgiar.org](mailto:m.vannoordwijk@cgiar.org))

1021 **Abstract** (currently 399 words...)

1022 Watersheds buffer the temporal pattern of river flow relative to the temporal pattern of  
1023 rainfall. This 'ecosystem service' is inherent to geology and climate, but buffering also  
1024 responds to human use and misuse of the landscape. Buffering can be part of management  
1025 feedback loops if salient, credible and legitimate indicators are used. The flow persistence  
1026 parameter  $F_p$  in a parsimonious recursive model of river flow (Part I) couples the  
1027 transmission of extreme rainfall events ( $1 - F_p$ ), to the annual base flow fraction of a  
1028 watershed ( $F_p$ ). Here we compare  $F_p$  estimates from four meso-scale watersheds in  
1029 Indonesia (Cidanau, Way Besai, and Bialo) and Thailand (Mae Chaem), with varying climate,  
1030 geology and land cover history, at a decadal time scale. The likely response in each of these  
1031 four to variation in rainfall properties (incl. the maximum hourly rainfall intensity) and land  
1032 cover (comparing scenarios with either more or less forest and tree cover than the current  
1033 situation) was explored through a basic daily water balance model, GenRiver. This model  
1034 was calibrated for each site on existing data, before being used for alternative land cover  
1035 and rainfall parameter settings. In both data and model runs, the wet-season (3-monthly)  $F_p$   
1036 values were consistently lower than dry-season values for all four sites. Across the four  
1037 catchments  $F_p$  values decreased with increasing annual rainfall, but specific aspects of  
1038 watersheds, such as the riparian swamp (peat soils) in Cidanau reduced effects of land use  
1039 change in the upper watershed. Increasing the mean rainfall intensity (at constant monthly  
1040 totals for rainfall) around the values considered typical for each landscape was predicted to  
1041 cause a decrease in  $F_p$  values by between 0.047 (Bialo) and 0.261 (Mae Chaem). Sensitivity of  
1042  $F_p$  to changes in land use change plus changes in rainfall intensity depends on other  
1043 characteristics of the watersheds, and generalizations made on the basis of one or two case  
1044 studies may not hold, even within the same climatic zone. A wet-season  $F_p$  value above 0.7  
1045 was achievable in forest-Agroforestry mosaic case studies. Interannual variability in  $F_p$  is  
1046 large relative to effects of land cover change. Multiple (5-10) years of paired-plot data would  
1047 generally be needed to reject no-change null-hypotheses on the effects of land use change  
1048 (degradation and restoration).  $F_p$  trends over time serve as a holistic scale-dependent  
1049 performance indicator of degrading/recovering watershed health and can be tested for  
1050 acceptability and acceptance in a wider social-ecological context.

1051 **Introduction**

1052 Inherent properties (geology, geomorphology) interact with climate and human modification of  
1053 vegetation, soils, drainage and riparian wetlands in effectuating the degree of buffering that  
1054 watersheds provide (Andréassian 2004; Bruijnzeel, 2004). Buffering of river flow relative to the

1055 space-time dynamics of rainfall is an ecosystem service, reducing the exposure of people living on  
1056 geomorphological floodplains to high-flow events, and increasing predictability and river flow in dry  
1057 periods (Joshi et al., 2004; Leimona et al., 2015; Part I). In the absence of any vegetation and with a  
1058 sealed surface, river flow will directly respond to the spatial distribution of rainfall, with only the  
1059 travel time to any point of specific interest influencing the temporal pattern of river flow. Any  
1060 persistence or predictability of river flow in such a situation will reflect temporal autocorrelation of  
1061 rainfall, beyond statistical predictability in seasonal rainfall patterns. On the other side of the  
1062 spectrum, river flow can be constant every day, beyond the theoretical condition of constant rainfall,  
1063 in a watershed that provides perfect buffering, by passing all water through groundwater pools that  
1064 have sufficient storage capacity at any time during the year. Both infiltration-limited (Hortonian) and  
1065 saturation-induced use of more rapid flow pathways (inter and overland flows) will reduce the flow  
1066 persistence and make it, at least in part, dependent on rainfall events. Separating the effects of land  
1067 cover (land use), engineering and rainfall on the actual flow patterns of rivers remains a considerable  
1068 challenge (Ma et al., 2014; Verbist et al., 2019). It requires data, models and concepts that can serve  
1069 as effective boundary object in communication with stakeholders (Leimona et al. 2015; van  
1070 Noordwijk et al. 2012, 2016). There is a long tradition in using forest cover as such a boundary  
1071 object, but there is only a small amount of evidence supporting this (Tan-Soo et al., 2014; van Dijk et  
1072 al., 2009; van Noordwijk et al. 2015a; part I).

1073 In part I, we introduced a flow persistence parameter ( $F_p$ ) that links the two, asymmetrical aspects of  
1074 flow dynamics: translating rainfall excess into river flow, and gradually releasing water stored in the  
1075 landscape. The direct link between these two aspects can be seen from equation [4] in part I:

$$1076 Q_t = F_p Q_{t-1} + (1-F_p)(P_t - E_{tx})$$

1077 Where  $Q_t$  and  $Q_{t-1}$  represent river flow on subsequent days,  $P_{tx}$  the precipitation on day  $t$  (or  
1078 preceding precipitation released as snowmelt on day  $t$ ) and  $E_{tx}$  the preceding evapotranspiration  
1079 since the previous precipitation event, creating storage space in the soils of the watershed. The first  
1080 term on the right-hand side of the equation represents the gradual release of stored water, causing  
1081 a slow decline of flow as the pools feeding this flow are gradually depleted. The second term reflects  
1082 the part of fresh additions of water are partitioned over immediate river flow and the increase of  
1083 stocks from which water can be gradually released. The derivation of the link depended on the long  
1084 term water balance, and thus assumed that all out- and inflows are accounted for in the watershed.

1085 Commonly used rainfall-runoff models (including the curve number approach and SWAT models)  
1086 only focus on the second term of the above equation (Ponce et al., 1996; Gassman et al., 2007),  
1087 without link to the first. Various empirical methods for deriving 'base flow' are in use, but details of  
1088 the calculation procedure matter. Results in part I for a number of contrasting meso-scale  
1089 watersheds in Southeast Asia suggested that interannual variation in  $F_p$  within a given watershed  
1090 correlates with both the R-B Flashiness Index (Baker et al., 2004) and the base-flow fraction of  
1091 annual river flow. However, the slope of these relationships varied between watersheds. Here, in  
1092 part II we will further analyse the  $F_p$  results for these watersheds that were selected to represent  
1093 variation in rainfall and land cover, and test the internal consistency of results based on historical  
1094 data: two located in the humid and one in the subhumid tropics of Indonesia, and one in the  
1095 unimodal subhumid tropics of northern Thailand.

1096 After exploring the patterns of variation in  $F_p$  estimates derived from actual river flow records, we  
1097 will quantify the sensitivity of the  $F_p$  metric to variations in rainfall intensity and its response, on a  
1098 longer timescale to land cover change. To do so, we will use a model that uses basic water balance  
1099 concepts: rainfall interception, infiltration, water use by vegetation, overland flow, interflow and

1100 groundwater release, to a spatially structured watershed where travel time from sub watersheds to  
1101 any point of interest modifies the predicted river flow. In the specific model used land cover effects  
1102 on soil conditions, interception and seasonal water use have been included. After testing whether  $F_p$   
1103 values derived from model outputs match those based on empirical data where these exist, we rely  
1104 on the basic logic of the model to make inference on the relative importance of modifying rainfall  
1105 and land cover inputs. With the resulting temporal variation in calculated  $F_p$  values, we consider the  
1106 time frame at which observed shifts in  $F_p$  can be attributed to factors other than chance (that means:  
1107 null-hypotheses of random effects can be rejected with accepted chance of Type I errors).

1108 **2. Methods**

1109 **2.1 GenRiver model for effects of land cover on river flow**

1110 The GenRiver model (van Noordwijk et al., 2011) is based on a simple water balance concept with a  
1111 daily time step and a flexible spatial subdivision of a watershed that influences the routing of water  
1112 and employs spatially explicit rainfall. At patch level, vegetation influences interception, retention  
1113 for subsequent evaporation and delayed transfer to the soil surface, as well as the seasonal demand  
1114 for water. Vegetation (land cover) also influences soil porosity and infiltration, modifying the  
1115 inherent soil properties. Water in the root zone is modelled separately for each land cover within a  
1116 subcatchment, the groundwater stock is modelled at subcatchment level. The spatial structure of a  
1117 watershed and the routing of surface flows influences the time delays to any specified point of  
1118 interest, which normally includes the outflow of the catchment. Land cover change scenarios are  
1119 interpolated annually between time-series (measured or modelled) data. The model may use  
1120 measured rainfall data, or use a rainfall generator that involves Markov chain temporal  
1121 autocorrelation (rain persistence). As our data sources are mostly restricted to daily rainfall  
1122 measurements and the infiltration model compares instantaneous rainfall to infiltration capacity, a  
1123 stochastic rainfall intensity was applied at subcatchment level, driven by the mean as parameter and  
1124 a standard deviation for a normal distribution (truncated at 3 standard deviations from the mean)  
1125 proportional to it via a coefficient of variation as parameter. For the Mae Chaem site in N Thailand  
1126 data by Dairaku et al. (2004) suggested a mean of less than 3 mm/hr. For the three sites in Indonesia  
1127 we used 30 mm/hr, based on Kusumastuti et al. (2016). Appendix 1 provides further detail on the  
1128 GenRiver model. The model itself, a manual and application case studies are freely available  
1129 (<http://www.worldAgroforestry.org/output/genriver-genetic-river-model-river-flow>; van Noordwijk  
1130 et al., 2011).

1131 **2.2 Empirical data-sets, model calibration**

1132 Table 1 and Figure 1 provide summary characteristics and the location of river flow data used in four  
1133 meso-scale watersheds for testing the  $F_p$  algorithm and application of the GenRiver model. Figure 1  
1134 includes a water tower category in the agro-ecological zones; this is defined on the basis of a ratio of  
1135 precipitation and potential evapotranspiration of more than 0.65, and a product of that ratio and  
1136 relative elevation exceeding 0.277.

1137     ⇒ Table 1  
1138     ⇒ Figure 1

1139 As major parameters for the GenRiver model were not independently measured for the respective  
1140 watersheds, we tuned (calibrated) the model by modifying parameters within a predetermined  
1141 plausible range, and used correspondence with measured hydrograph as test criterion (Kobolt et al.  
1142 2008). We used the Nash-Sutcliff Efficiency (NSE) parameter (target above 0.5) and bias (less than  
1143 25%) as test criteria and targets. Meeting these performance targets (Moriasi et al., 2007), we

1144 accepted the adjusted models as basis for describing current conditions and exploring model  
1145 sensitivity. The main site-specific parameter values are listed in Table 2 and (generic) land cover  
1146 specific default parameters in Table 3.

1147     ⇒ Table 2

1148     ⇒ Table 3

1149 Table 4 describes the six scenarios of land use change that were evaluated in terms of their  
1150 hydrological impacts. Further description on the associated land cover distribution for each scenario  
1151 in the four different watersheds is depicted in Appendix 2.

1152     ⇒ Table 4

### 1153 **2.3 Bootstrapping to estimate the minimum observation**

1154 The bootstrap methods (Efron and Tibshirani, 1986) is a resampling methods that is commonly used  
1155 to generate 'surrogate population' for the purpose of approximating the sampling distribution of a  
1156 statistic. In this study, the bootstrap approach was used to estimate the minimum number of  
1157 observation (or yearly data) required for a pair-wise comparison test between two time-series of  
1158 stream flow or discharge data (representing two scenarios of land use distributions) to be  
1159 distinguishable from a null-hypothesis of no effect. The pair-wise comparison test used was  
1160 Kolmogorov-Smirnov test that is commonly used to test the distribution of discharge data (Zhang et al,  
1161 2006). We built a simple macro in R (R Core Team, 2015) that entails the following steps:

1162 (i)   Bootstrap or resample with replacement 1000 times from both time-series discharge data  
1163       with sample size  $n$ ;

1164 (ii)   Apply the Kolmogorov-Smirnov test to each of the 1000 generated pair-wise discharge data,  
1165       and record the P-value;

1166 (iii)   Perform (i) and (ii) for different size of  $n$ , ranging from 5 to 50.

1167 (iv)   Tabulate the p-value from the different sample size  $n$ , and determine the value of  $n$  when the  
1168       p-value reached equal to or less than 0.025 (or equal to the significance level of 5%). The  
1169       associated  $n$  represents the minimum number of observations required.

1170 Appendix 3 provides an example of the macro in R used for this analysis.

## 1171 **3. Results**

### 1172 **3.1 Empirical data of flow persistence as basis for model parameterization**

1173 Inter-annual variability of  $F_p$  estimates derived for the four catchments (Figure 2) was of the order of  
1174 0.1 units, while the intra-annual variability between dry and rainy seasons was 0.1-0.2. For all years  
1175 and locations, rainy season  $F_p$  values, with mixed flow pathways, were consistently below dry-season  
1176 values, dominated by groundwater flows. If we can expect  $F_{p,i}$  and  $F_{p,o}$  (see equation 8 in part I) to be  
1177 approximately 0.5 and 0, this difference between wet and dry periods implies a 40% contribution of  
1178 interflow in the wet season, a 20% contribution of overland flow or any combination of the two  
1179 effects.

1180 Overall the estimates from modelled and observed data are related with 16% deviating more than  
1181 0.1 and 3% more than 0.15 (Figure 3). As the Moriasi et al. (2007) performance criteria for the

1182 hydrographs were met by the calibrated models for each site, we tentatively accept the model to be  
1183 a basis for sensitivity study of  $F_p$  to modifications to land cover and/or rainfall

1184  $\Rightarrow$  Figure 2

1185  $\Rightarrow$  Figure 3

### 1186 **3.2 Comparing $F_p$ effects of rainfall intensity and land cover change**

1187 A direct comparison of model sensitivity to changes in mean rainfall intensity and land use change  
1188 scenarios is provided in Figure 4. Varying the mean rainfall intensity over a factor 7 shifted the  $F_p$   
1189 value by only 0.047 and 0.059 in the case of Bialo and Cidanau, respectively, but by 0.128 in Way  
1190 Besai and 0.261 in Mae Chaem (Figure 4A). The impact of the land use change scenarios on  $F_p$  was  
1191 smallest in Cidanau (0.026), intermediate in Way Besai (0.048) and relatively large in Bialo and Mae  
1192 Chaem, at 0.080 and 0.084, respectively (Figure 4B). The order of  $F_p$  across the land use change  
1193 scenarios was mostly consistent between the watersheds, but the contrast between the  
1194 Reforestation and NatForest scenario was largest in Mae Chaem and smallest in Way Besai. In  
1195 Cidanau, Way Besai and Mae Chaem, variations in rainfall were 2.2 to 3.1 times more effective than  
1196 land use change in shifting  $F_p$ , in Bialo its relative effect was only 58%. Apparently, the sensitivity to  
1197 changes in land use change plus changes in rainfall intensity depends on other characteristics of the  
1198 watersheds, and generalizations made on the basis of one or two case studies may not hold, even  
1199 within the same climatic zone.

1200  $\Rightarrow$  Figure 4

### 1201 **3.3 Further analysis of $F_p$ effects for scenarios of land cover change**

1202 Among the four watersheds there is consistency in that the 'forest' scenario has the highest, and the  
1203 'degraded lands' the lowest  $F_p$  value (Figure 5), but there are remarkable differences as well: in  
1204 Cidanau the interannual variation in  $F_p$  is clearly larger than land cover effects, while in the Way  
1205 Besai the spread in land use scenarios is larger than interannual variability. In Cidanau a peat swamp  
1206 between most of the catchment and the measuring point buffers most of landcover related variation  
1207 in flow, but not the interannual variability. Considering the frequency distributions of  $F_p$  values over  
1208 a 20 year period, we see one watershed (Way Besai) where the forest stands out from all others, and  
1209 one (Bialo) where the degraded lands are separate from the others. Given the degree of overlap of  
1210 the frequency distributions, it is clear that multiple years of empirical observations will be needed  
1211 before a change can be affirmed.

1212 Figure 5 shows the frequency distributions of expected effect sizes on  $F_p$  of a comparison of any land  
1213 cover with either forest or degraded lands. Table 5 translates this information to the number of  
1214 years that a paired plot (in the absence of measurement error) would have to be maintained to  
1215 reject a null-hypothesis of no effect, at  $p=0.05$ . As the frequency distributions of  $F_p$  differences of  
1216 paired catchments do not match a normal distribution, a Kolmogorov-Smirnov test can be used to  
1217 assess the probability that a no-difference null hypothesis can yield the difference found. By  
1218 bootstrapping within the years where simulations supported by observed rainfall data exist, we  
1219 found for the Way Besai catchment, for example, that 20 years of data would be needed to assert (at  
1220  $P = 0.05$ ) that the Reforestation scenario differs from Agroforestation, and 16 years that it differs  
1221 from Actual and 11 years that it differs from Degrade. In practice, that means that empirical  
1222 evidence that survives statistical tests will not emerge, even though effects on watershed health are  
1223 real.

1224  $\Rightarrow$  Figure 5

1225  $\Rightarrow$  Table 5

1226 At process-level the increase in 'overland flow' in response to soil compaction due to land cover  
1227 change has a clear and statistically significant relationship with decreasing  $F_p$  values in all catchments  
1228 (Figure 6), but both year-to-year variation within a catchment and differences between catchments  
1229 influence the results as well, leading to considerable spread in the biplot. Contrary to expectations,  
1230 the disappearance of 'interflow' by soil compaction is not reflected in measurable change in  $F_p$  value.  
1231 The temporal difference between overland and interflow (one or a few days) gets easily blurred in  
1232 the river response that integrates over multiple streams with variation in delivery times; the  
1233 difference between overland- or interflow and baseflow is much more pronounced. Apparently,  
1234 according to our model, the high macroporosity of forest soils that allows interflow and may be the  
1235 'sponge' effect attributed to forest, delays delivery to rivers by one or a few days, with little effect on  
1236 the flow volumes at locations downstream where flow of multiple days accumulates. The difference  
1237 between overland- or interflow and baseflow in time-to-river of rainfall peaks is much more  
1238 pronounced.

1239  $\Rightarrow$  Figure 6

1240 Tree cover has two contradicting effects on baseflow: it reduces the surplus of rainfall over  
1241 evapotranspiration (annual water yield) by increased evapotranspiration (especially where  
1242 evergreen trees or trees with a large canopy interception are involved), but it potentially increases  
1243 soil macroporosity that supports infiltration and interflow, with relatively little effect on water  
1244 holding capacity measured as 'field capacity' (after runoff and interflow have removed excess  
1245 water). Figure 7 shows that the total volume of baseflow differs more between sites and their  
1246 rainfall pattern than it varies with tree cover. Between years total evapotranspiration and baseflow  
1247 totals are positively correlated, but for a given rainfall there is a trade-off. Overall these results  
1248 support the conclusion that generic effects of deforestation on decreased flow persistence, and of  
1249 (agro)/(re)-forestation on increased flow persistence are small relative to interannual variability due  
1250 to specific rainfall patterns, and that it will be hard for any empirical data process to pick-up such  
1251 effects, even if they are qualitatively aligned with valid process-based models.

1252  $\Rightarrow$  Figure 7

#### 1253 4. Discussion

1254 In the discussion of Part I the credibility questions on replicability of the  $F_p$  metric and its sensitivity  
1255 to details of rainfall pattern versus land cover as potential causes of variation were seen as requiring  
1256 case studies in a range of contexts. Although the four case studies in Southeast Asia presented here  
1257 cannot be claimed to represent the global variation in catchment behaviour (with absence of a  
1258 snowpack and its dynamics as an obvious element of flow buffering not included), the diversity of  
1259 responses among these four already point to challenges for any generic interpretation of the degree  
1260 of flow persistence that can be achieved under natural forest cover, as well as its response to land  
1261 cover change.

1262 The empirical data summarized here for (sub)humid tropical sites in Indonesia and Thailand show  
1263 that values of  $F_p$  above 0.9 are scarce in the case studies provided, but values above 0.8 were found,  
1264 or inferred by the model, for forested landscapes. Agroforestry landscapes generally presented  $F_p$   
1265 values above 0.7, while open-field agriculture or degraded soils led to  $F_p$  values of 0.5 or lower. Due  
1266 to differences in local context, it may not be feasible to relate typical  $F_p$  values to the overall  
1267 condition of a watershed, but temporal change in  $F_p$  can indicate degradation or restoration if a  
1268 location-specific reference can be found. The difference between wet and dry season  $F_p$  can be  
1269 further explored in this context. The dry season  $F_p$  value primarily reflects the underlying geology,

1270 with potential modification by engineering and operating rules of reservoirs, the wet season  $F_p$  is  
1271 generally lower due to partial shifts to overland and interflow pathways. Where further uncertainty  
1272 is introduced by the use of modelled rather than measured river flow, the lack of fit of models  
1273 similar to the ones we used here would mean that scenario results are indicative of directions of  
1274 change rather than a precision tool for fine-tuning combinations of engineering and land cover  
1275 change as part of integrated watershed management.

1276 The differences in relative response of the watersheds to changes in mean rainfall intensity and land  
1277 cover change, suggest that generalizations derived from one or a few case studies are to be  
1278 interpreted cautiously. If land cover change would influence details of the rainfall generation process  
1279 (arrow 10 in Figure 1 of part I; e.g. through release of ice-nucleating bacteria Morris et al., 2014; van  
1280 Noordwijk et al., 2015b) this can easily dominate over effects via interception, transpiration and soil  
1281 changes.

1282 Our results indicate an intra-annual variability of  $F_p$  values between wet and dry seasons of around  
1283 0.2 in the case studies, while interannual variability in either annual or seasonal  $F_p$  was generally in  
1284 the 0.1 range. The difference between observed and simulated flow data as basis for  $F_p$  calculations  
1285 was mostly less than 0.1. With current methods, it seems that effects of land cover change on flow  
1286 persistence that shift the  $F_p$  value by about 0.1 are the limit of what can be asserted from empirical  
1287 data (with shifts of that order in a single year a warning sign rather than a firmly established change).  
1288 When derived from observed river flow data  $F_p$  is suitable for monitoring change (degradation,  
1289 restoration) and can be a serious candidate for monitoring performance in outcome-based  
1290 ecosystem service management contracts. Choice of the part of the year for which  $F_p$  changes are  
1291 used as indicator may have to depend on the seasonal patterns of rainfall.

1292 In view of our results the lack of robust evidence in the literature of effects of change in forest and  
1293 tree cover on flood occurrence may not be a surprise; effects are subtle and most data sets contain  
1294 considerable variability. Yet, such effects are consistent with current process and scaling knowledge  
1295 of watersheds.

1296 In summarizing findings on the  $F_p$  metric, we can compare it with existing ones across the seven  
1297 questions raised in Fig. 1 of part I. Comparator metrics can derive from various data sources,  
1298 including the amount (and/or quality) of forest cover upstream, the fraction of flows that is  
1299 technically controlled, direct records of river flow (over a short or longer time period), records of  
1300 rainfall and/or models that combine landscape properties, climate and land cover. Tentative scoring  
1301 for these metrics (Table 6) suggest that the  $F_p$  metric is an efficient tool for data-scarce  
1302 environments, as it indicates aspects of hydrographs that so far required multi-annual records of  
1303 river flow.

1304 →Table 6

## 1305 Conclusion

1306 Overall, our analysis suggests that the level of flow buffering achieved depends on both land cover  
1307 (including its spatial configuration and effects on soil properties) and space-time patterns of rainfall  
1308 (including maximum rainfall intensity as determinant of overland flow). Generalizations on dominant  
1309 influence of either, derived from one or a few case studies are to be interpreted cautiously. If land  
1310 cover change would influence details of the rainfall generation process this can easily dominate over  
1311 effects via interception, transpiration and soil changes. Multi-year data will generally be needed to  
1312 attribute observed changes in flow buffering to degradation/restoration of watersheds, rather than  
1313 specific rainfall events. With current methods, it seems that effects of land cover change on flow

1314 persistence that shift the  $F_p$  value by about 0.1 are the limit of what can be asserted from empirical  
1315 data, with shifts of that order in a single year a warning sign rather than a firmly established change.  
1316 When derived from observed river flow data  $F_p$  is suitable for monitoring change (degradation,  
1317 restoration) and can be a serious candidate for monitoring performance in outcome-based  
1318 ecosystem service management contracts. Watershed health is here characterized through the flow  
1319 pattern it generates, leaving the attribution to land cover, rainfall pattern and engineering of that  
1320 pattern and of changes in pattern to further location-specific analysis, just as a symptom of a high  
1321 body temperature can indicate health, but not diagnose the specific illness causing it.

1322 The data sets analysed so far did not indicate that the flow persistence at high flows differed from  
1323 that at lower flows **within the same season**, but in other circumstances this may not be the case and  
1324 further care may be needed to use  $F_p$  values beyond the measurement period in which they were  
1325 derived. **While a major strength of the  $F_p$  method over existing procedures for parameterizing curve**  
1326 **number estimates, for example, is that the latter depend on scarce observations during extreme**  
1327 **events and  $F_p$  can be estimated for any part of the flow record, the reliability of  $F_p$  estimates will still**  
1328 **increase with the length of the observation period.**

1329 Further tests on the performance of the  $F_p$  metric and its standard incorporation into the output  
1330 modules of river flow and watershed management models will broaden the basis for interpreting the  
1331 value ranges that can be expected for well-functioning watersheds in various conditions of climate,  
1332 topography, soils, vegetation and engineering interventions. Such a broader empirical base could  
1333 test the possible use of  $F_p$  as performance metric for watershed rehabilitation efforts.

## 1334 **Data availability**

1335 Table 7 specifies the rainfall and river flow data we used for the four basins and specifies the links to  
1336 detailed descriptions.

1337  $\Rightarrow$  Table 7

## 1338 **Acknowledgements**

1339 This research is part of the Forests, Trees and Agroforestry research program of the CGIAR. Several  
1340 colleagues contributed to the development and early tests of the  $F_p$  method. Thanks are due to  
1341 Thoha Zulkarnain for assistance with Figure 1 and to Eike Luedeling, Sonya Dewi, Sampurno  
1342 Bruijnzeel and two anonymous reviewers for comments on an earlier version of the manuscript.

## 1343 **References**

1344 Andréassian, V.: Waters and forests: from historical controversy to scientific debate, *Journal of*  
1345 *Hydrology*, 291, 1–27, 2004.

1346 Baker, D.B., Richards, R.P., Loftus, T.T. and Kramer, J.W.: A new flashiness index: Characteristics and  
1347 applications to midwestern rivers and streams. *Journal of the American Water Resources*  
1348 *Association*, Paper No. 03095, 2004.

1349 Bruijnzeel, L.A.: Hydrological functions of tropical forests: not seeing the soil for the trees, *Agr.*  
1350 *Ecosyst. Environ*, 104, 185–228, 2004.

1351 Dairaku K., Emori, S., and Taikan, T.: Rainfall Amount, Intensity, Duration, and Frequency  
1352 Relationships in the Mae Chaem Watershed in Southeast Asia, *Journal of Hydrometeorology*, 5,  
1353 458–470, 2004.

1354 Efron, B and Tibshirani, R.: Bootstrap Methods for Standard Errors, Confidence Intervals, and Other  
1355 Measures of Statistical Accuracy. *Statistical Science* 1 (1): 54-75, 1986.

1356 Gassman, P.W., Reyes, M.R., Green, C.H. and Arnold, J.G.: The soil and water assessment tool: historical  
1357 development, applications, and future research directions. *Transactions of the ASABE*, 50(4), 1211-1250,  
1358 2007.

1359 Joshi, L., Schalenbourg, W., Johansson, L., Khasanah, N., Stefanus, E., Fagerström, M.H., and van  
1360 Noordwijk, M.: Soil and water movement: combining local ecological knowledge with that of  
1361 modellers when scaling up from plot to landscape level, In: van Noordwijk, M., Cadisch, G. and  
1362 Ong, C.K. (Eds.) *Belowground Interactions in Tropical Agroecosystems*, CAB International,  
1363 Wallingford (UK), 349-364, 2004.

1364 Kobold, M., Suselj, K., Polajnar, J., and Pogacnik, N.: Calibration Techniques Used For HBV  
1365 Hydrological Model In Savinja Catchment, In: XXIVth Conference of the Danubian Countries On  
1366 The Hydrological Forecasting And Hydrological Bases Of Water Management, 2008.

1367 Kusumastuti, D.I., Jokowinarno, D., van Rafi'i, C.H., and Yuniarti, F.: Analysis of rainfall characteristics  
1368 for flood estimation in Way Awi watershed, *Civil Engineering Dimension*, 18, 31-37, 2016

1369 Leimona, B., Lusiana, B., van Noordwijk, M., Mulyoutami, E., Ekadinata, A., and Amaruzama, S.:  
1370 Boundary work: knowledge co-production for negotiating payment for watershed services in  
1371 Indonesia, *Ecosystems Services*, 15, 45-62, 2015.

1372 Ma, X., Lu, X., van Noordwijk, M., Li, J.T., and Xu, J.C.: Attribution of climate change, vegetation  
1373 restoration, and engineering measures to the reduction of suspended sediment in the Kejie  
1374 catchment, southwest China, *Hydrol. Earth Syst. Sci.*, 18, 1979-1994, 2014.

1375 Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., and Veith, T.L.: Model  
1376 Evaluation Guidelines For Systematic Quantification Of Accuracy In Watershed Simulations,  
1377 American Society of Agricultural and Biological Engineers, 20(3), 885-900, 2001

1378 Morris, C.E., Conen, F., Huffman, A., Phillips, V., Pöschl, U., and Sands, D.C.: Bioprecipitation: a  
1379 feedback cycle linking Earth history, ecosystem dynamics and land use through biological ice  
1380 nucleators in the atmosphere, *Glob Change Biol*, 20, 341-351. 2014.

1381 Ponce, V.M. and Hawkins, R.H.: Runoff curve number: Has it reached maturity? *Journal of Hydrologic  
1382 Engineering*, 1(1), 11-19, 1996

1383 R Core Team R: A language and environment for statistical computing. R Foundation for Statistical  
1384 Computing, Vienna, Austria, URL <http://www.R-project.org/>, 2015

1385 Tan-Soo, J.S., Adnan, N., Ahmad, I., Pattanayak, S.K., and Vincent, J.R.: Econometric Evidence on  
1386 Forest Ecosystem Services: Deforestation and Flooding in Malaysia. *Environmental and Resource  
1387 Economics*, on-line: <http://link.springer.com/article/10.1007/s10640-014-9834-4>, 2014.

1388 van Dijk, A.I., van Noordwijk, M., Calder, I.R., Bruijnzeel, L.A., Schellekens, J., and Chappell, N.A.:  
1389 Forest-flood relation still tenuous – comment on ‘Global evidence that deforestation amplifies  
1390 flood risk and severity in the developing world’, *Global Change Biology*, 15, 110-115, 2009.

1391 van Noordwijk, M., Widodo, R.H., Farida, A., Suyamto, D., Lusiana, B., Tanika, L., and Khasanah, N.:  
1392 GenRiver and FlowPer: Generic River and Flow Persistence Models. User Manual Version 2.0,  
1393 Bogor, Indonesia, World Agroforestry Centre (ICRAF) Southeast Asia Regional Program, 2011.

1394 van Noordwijk, M., Leimona, B., Jindal, R., Villamor, G.B., Vardhan, M., Namirembe, S., Catacutan, D.,  
1395 Kerr, J., Minang, P.A., and Tomich, T.P.: Payments for Environmental Services: evolution towards  
1396 efficient and fair incentives for multifunctional landscapes, *Annu. Rev. Environ. Resour.*, 37, 389-  
1397 420, 2012.

1398 van Noordwijk, M., Leimona, B., Xing, M., Tanika, L., Namirembe, S., and Suprayogo, D.: Water-  
1399 focused landscape management. *Climate-Smart Landscapes: Multifunctionality In Practice*, eds  
1400 Minang PA et al.. Nairobi, Kenya: World Agroforestry Centre (ICRAF), 179-192, 2015a.

1401 van Noordwijk, M., Bruijnzeel, S., Ellison, D., Sheil, D., Morris, C., Gutierrez, V., Cohen, J., Sullivan, C.,  
1402 Verbist, B., and Muys, B.: Ecological rainfall infrastructure: investment in trees for sustainable  
1403 development, ASB Brief no 47, Nairobi, ASB Partnership for the Tropical Forest Margins, 2015b.

1404 van Noordwijk, M., Kim, Y-S., Leimona, B., Hairiah, K., Fisher, L.A.: Metrics of water security,  
1405 adaptive capacity and Agroforestry in Indonesia. *Current Opinion on Environmental Sustainability*  
1406 (in press: <http://dx.doi.org/10.1016/j.cosust.2016.10.004>). 2016

1407 Verbist, B., Poesen, J., van Noordwijk, M. Widianto, Suprayogo, D., Agus, F., and Deckers, J.: Factors  
1408 affecting soil loss at plot scale and sediment yield at catchment scale in a tropical volcanic  
1409 Agroforestry landscape, *Catena*, 80, 34-46, 2010.

1410 Zhang, Q., Liu, C., Xu, C., Xu, and Jiang T.: Observed trends of annual maximum water level and  
1411 streamflow during past 130 years in the Yangtze River basin, China, *Journal of Hydrology*, 324,  
1412 255-265, 2006.

1413

1414 Table 1. Basic physiographic characteristics of the four study watersheds

| Parameter                | Bialo  | Cidanau  | Mae Chaem  | Way Besai  |
|--------------------------|--|--|--|--|
| Location                 | South Sulawesi, Indonesia                                      | West Java, Indonesia   | Northern Thailand  | Lampung, Sumatera, Indonesia   |
| Coordinates              | 5.43 S, 120.01 E   | 6.21 S, 105.97 E   | 18.57 N, 98.35 E   | 5.01 S, 104.43 E   |
| Area (km <sup>2</sup> )  | 111.7  | 241.6  | 3892   | 414.4  |
| Elevation (m a.s.l.)     | 0 – 2874   | 30 – 1778  | 475-2560   | 720-1831   |
| Flow pattern             | Parallel   | Parallel (with two main river flow that meet in the downstream area)         | Parallel   | Radial   |
| Land cover type          | Forest (13%)<br>Agroforest (59%)<br>Crops (22%)<br>Others (6%) | Forest (20%)<br>Agroforest (32%)<br>Crops (33%)<br>Others (11%)<br>Swamp(4%) | Forest (evergreen, deciduous and pine) (84%)<br>Crops (15%)<br>Others (1%) | Forest (18%)<br>Coffee (monoculture and multistrata) (64%)<br>Crop and Horticulture (12%)<br>Others (6%) |
| Mean annual rainfall, mm | 1695   | 2573   | 1027   | 2474   |
| Wet season               | April – June   | January - March  | July - September   | January - March  |
| Dry season               | July - September   | July - September   | January - March  | July - September   |
| Mean annual runoff, mm   | 947  | 917  | 259  | 1673   |
| Major soils              | Inceptisols  | Inceptisols  | Ultisols, Entisols   | Andisols   |

1415

1416 Table 2. Parameters of the GenRiver model used for the four site specific simulations (van Noordwijk et al., 2011 for definitions of terms; sequence of parameters follows the pathway of water)

| Parameter         | Definition                                     | Unit                | Bialo | Cidanau | Mae Chaem | Way Besai |
|-------------------|--|---------------------|-------|---------|-----------|-----------|
| RainIntensMean    | Average rainfall intensity                     | mm hr <sup>-1</sup> | 30    | 30      | 3         | 30        |
| RainIntensCoefVar | Coefficient of variation of rainfall intensity | mm hr <sup>-1</sup> | 0.8   | 0.3     | 0.5       | 0.3       |

|                        |   |                     |      |      |      |      |
|------------------------|---|---------------------|------|------|------|------|
| RainInterceptDripRt    | Maximum drip rate of intercepted rain                                 | mm hr <sup>-1</sup> | 80   | 10   | 10   | 10   |
| RainMaxIntDripDur      | Maximum dripping duration of intercepted rain                         | hr                  | 0.8  | 0.5  | 0.5  | 0.5  |
| InterceptEffectontrans | Rain interception effect on transpiration                             | -                   | 0.35 | 0.8  | 0.3  | 0.8  |
| MaxInfRate             | Maximum infiltration capacity   | mm d <sup>-1</sup>  | 580  | 800  | 150  | 720  |
| MaxInfSubsoil          | Maximum infiltration capacity of the sub soil                         | mm d <sup>-1</sup>  | 80   | 120  | 150  | 120  |
| PerFracMultiplier      | Daily soil water drainage as fraction of groundwater release fraction | -                   | 0.35 | 0.13 | 0.1  | 0.1  |
| MaxDynGrWatStore       | Dynamic groundwater storage capacity                                  | mm                  | 100  | 100  | 300  | 300  |
| GWReleaseFracVar       | Groundwater release fraction, applied to all subcatchments            | -                   | 0.15 | 0.03 | 0.05 | 0.1  |
| Tortuosity             | Stream shape factor   | -                   | 0.4  | 0.4  | 0.6  | 0.45 |
| Dispersal Factor       | Drainage density  | -                   | 0.3  | 0.4  | 0.3  | 0.45 |
| River Velocity         | River flow velocity   | m s <sup>-1</sup>   | 0.4  | 0.7  | 0.35 | 0.5  |

1419 Table 3. GenRiver defaults for land use specific parameter values, used for all four watersheds  
 1420 (BD/BDref indicates the bulk density relative to that for an agricultural soil pedotransfer function;  
 1421 see van Noordwijk et al., 2011)

| Land cover Type               | Potential<br>interception<br>(mm/d) | Relative drought<br>threshold | BD/BDref    |
|-------------------------------|-------------------------------------|-------------------------------|-------------|
| Forest <sup>1</sup>           | 3.0 - 4.0                           | 0.4 - 0.5                     | 0.8 - 1.1   |
| Agroforestry <sup>2</sup>     | 2.0 - 3.0                           | 0.5 - 0.6                     | 0.95 - 1.05 |
| Monoculture tree <sup>3</sup> | 1.0                                 | 0.55                          | 1.08        |
| Annual crops                  | 1.0 - 3.0                           | 0.6 - 0.7                     | 1.1 - 1.5   |
| Horticulture                  | 1.0                                 | 0.7                           | 1.07        |
| Rice field <sup>4</sup>       | 1.0 - 3.0                           | 0.9                           | 1.1 - 1.2   |
| Settlement                    | 0.05                                | 0.01                          | 1.3         |
| Shrub and grass               | 2.0 - 3.0                           | 0.6                           | 1.0 - 1.07  |
| Cleared land                  | 1.0 - 1.5                           | 0.3 - 0.4                     | 1.1 - 1.2   |

1422 Note: 1. Forest: primary forest, secondary forest, swamp forest, evergreen forest, deciduous forest  
 1423 2. Agroforestry: mixed garden, coffee, cocoa, clove  
 1424 3. Monoculture : coffee  
 1425 4. Rice field: irrigation and rainfed  
 1426

1427 Table 4. Land use scenarios explored for four watersheds

| Scenario        | Description  |
|-----------------|--|
| NatForest       | Full natural forest, hypothetical reference scenario   |
| Reforestation   | Reforestation, replanting shrub, cleared land, grass land and some agricultural area with forest   |
| Agroforestation | Agroforestry scenario, maintaining Agroforestry areas and converting shrub, cleared land, grass land and some of agricultural area into Agroforestry                         |
| Actual          | Baseline scenario, based on the actual condition of land cover change during the modelled time period  |
| Agriculture     | Agriculture scenario, converting some of tree based plantations, cleared land, shrub and grass land into rice fields or dry land agriculture, while maintain existing forest |
| Degrading       | No change in already degraded areas, while converting most of forest and Agroforestry area into rice fields and dry land agriculture   |

1428

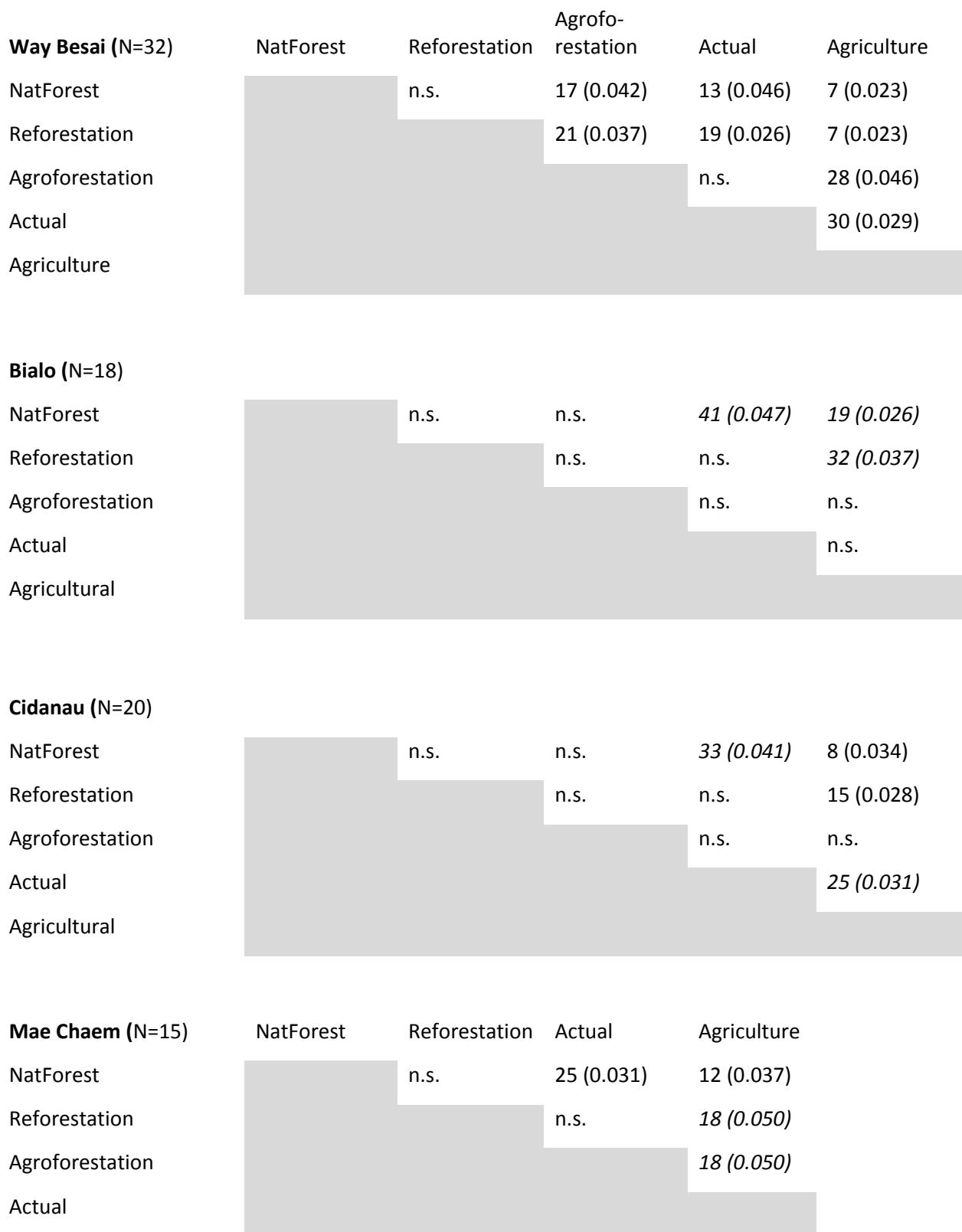
1429

1430 Table 5. Number of years of observations required to estimate flow persistence to reject the null-  
 1431 hypothesis of 'no land use effect', at p-value = 0.05 using Kolmogorov-Smirnov test. The probability  
 1432 of the test statistic in the first significant number is provided between brackets and where the  
 1433 number of observations exceeds the time series available, results are given in *italics*

A. Natural Forest as reference

|                             | Reforestation | Agroforestation | Actual     | Agricultural |
|-----------------------------|---------------|-----------------|------------|--------------|
| <b>Way Besai (N=32)</b>     |               |                 |            |              |
| Reforestation               |               | 20 (0.035)      | 16 (0.037) | 13 (0.046)   |
| Agroforestation             |               |                 | n.s.       | n.s.         |
| Actual                      |               |                 |            | n.s.         |
| Agricultural                |               |                 |            |              |
| Degrading                   |               |                 |            |              |
| <br><b>Bialo (N=18)</b>     |               |                 |            |              |
| Reforestation               |               | n.s.            | n.s.       | 37 (0.04)    |
| Agroforestation             |               |                 | n.s.       | n.s.         |
| Actual                      |               |                 |            | n.s.         |
| Agricultural                |               |                 |            |              |
| Degrading                   |               |                 |            |              |
| <br><b>Cidanau (N=20)</b>   |               |                 |            |              |
| Reforestation               |               | n.s.            | n.s.       | 32 (0.037)   |
| Agroforestation             |               |                 | n.s.       | n.s.         |
| Actual                      |               |                 |            | n.s.         |
| Agricultural                |               |                 |            |              |
| Degrading                   |               |                 |            |              |
| <br><b>Mae Chaem (N=15)</b> |               |                 |            |              |
| Reforestation               |               | n.s.            | 23 (0.049) | 18 (0.050)   |
| Agroforestation             |               |                 | 45 (0.037) | 33 (0.041)   |
| Actual                      |               |                 |            | 33 (0.041)   |
| Agricultural                |               |                 |            |              |

### B. Degrading scenario as reference



1435 Table 6. Comparison of metrics at various points in the causal network (Fig. 2 of Paper I) that can  
 1436 support watershed management and prevention of flood damage on the list of seven issues (I – VII)  
 1437 introduced in Fig. 1 Paper I\*.

| Terrain-based (7A and 5 in Fig. 2 of part I) |              |  | Based on river flow characteristics (4 in Fig. 2 of part I) |                  |                         |                                |           |                         | Integrated (5-7) terrain + climate + land use + river flow models |                          |
|--|--------------|--|---|------------------|-------------------------|--------------------------------|-----------|-------------------------|---|--------------------------|
| Is-<br>sues*                                 | Forest cover | Fraction of flow technically regulated | $Q_{\max} / Q_{\min}$                                       | Flashiness index | Flow frequency analysis | Curve-number (rainfall-runoff) | Base-flow | Flow persistence, $F_p$ | Spatial analysis  | Spatial water flow model |
| Range  | 0-100%       | 0-100%                                 | 1 - $\omega$  | 0 - 2            |                         | 1 - 100                        | 0-100%    | 0 - 1                   |   |                          |
| IA   | No           | Yes                                    | No  | Yes              | Yes                     | No                             | Yes       | Partially               | Yes   |                          |
| IB   | No           | Yes                                    | No  | No               | Yes                     | No                             | Yes       | Partially               | Yes   |                          |
| IIA  | Not          | Partially                              | Not   | Not              | Yes                     | Partially                      | Partially | Partially               | Partially   | Partially                |
| IIB  | Partially    | Yes                                    | Not   | Not              | Not                     | Partially                      | Partially | Partially               | Partially   | Yes                      |
| IIC  | Not          | Partially                              | Not   | Partially        | Partially               | Not                            | Partially | Partially               | Partially   | Yes                      |
| III  | Partially    | Partially                              | Not   | Partially        | Yes                     | Partially                      | Partially | Partially               | Partially   | Yes                      |
| IVA  | Single       | -                                      | Single  | Single           | Multi                   | Multi                          | Single    | Single                  | Single  | Single                   |
| IVB  | Robust       | Robust                                 | Sensitive   | Sensitive        | Sensitive               | Sensitive                      | Robust    | Robust                  | Robust  | Robust                   |
| V  | Partially    | Not                                    | Not   | Yes              | No                      | No                             | Partially | Yes                     | Partially   | Partially                |
| VI   | Not          | Not                                    | Not   | Partially        | Not                     | Not                            | Not       | Yes                     | Partially   | Partially                |
| VII  | Not          | Neutral                                | Not   | Yes              | Yes                     | Neutral                        | Neutral   | Yes                     | Yes   | Yes                      |

1438

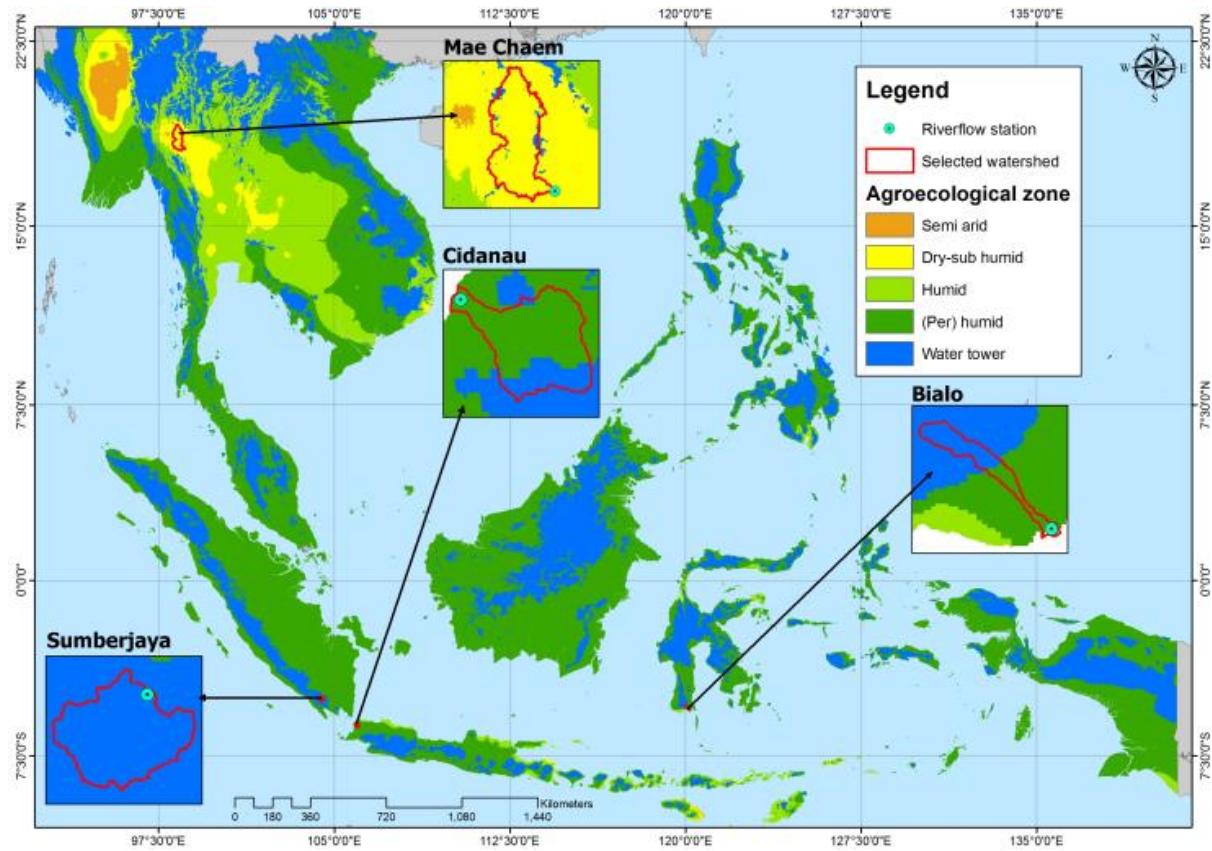
- 1439 I. Does the indicator relate to important aspects of watershed behaviour (A. Flood damage  
 1440 prevention; B. Low flow water availability)?
- 1441 II. Does its quantification help to select management actions? (A. Risk assessment, insurance  
 1442 design; B. Spatial planning, engineering interventions; C. Fine-tuning land use)
- 1443 III. Is it consistent with current understanding of key processes
- 1444 IV. Are data requirements feasible (A. Lowest temporal resolution for estimates (years); B.  
 1445 Consistency of numerical results and sensitivity to bias and random error in data sources?)
- 1446 V. Does it match local knowledge and concerns?
- 1447 VI. Can it be used to empower local stakeholders of watershed management through  
 1448 performance (outcome) based contracts?
- 1449 VII. Can it inform local risk management?

1451 Table 7. Data availability

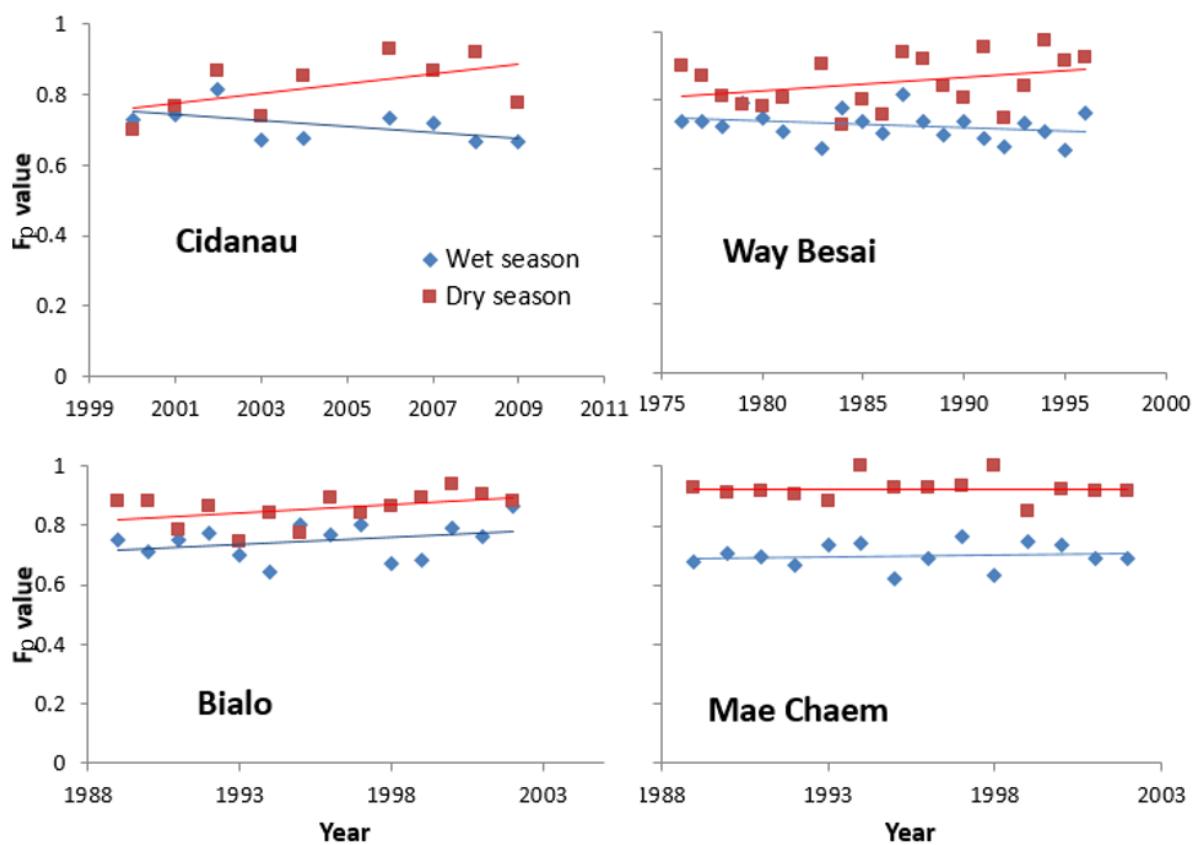
|                              | Bialo   | Cidanau   | Mae Chaem   | Way Besai   |
|------------------------------|---|---|---|---|
| Rainfall data                | 1989-2009, Source: BWS Sulawesi <sup>a</sup> and PUSAIR <sup>b</sup> ; Average rainfall data from the stations Moti, Bulo-bulo, Seka and Onto   | 1998-2008, source: BMKG <sup>c</sup>  | 1998-2002, source: WRD55, MTD22, RYP48, GMT13, WRD 52   | 1976-2007, Source: BMKG, PU <sup>d</sup> and PLN <sup>e</sup> (interpolation of 8 rainfall stations using Thiessen polygon)   |
| River flow data              | 1993-2010, source; BWS Sulawesi and PUSAIR  | 2000-2009, source: KTI <sup>f</sup>   | 1954-2003, source: ICHARM <sup>g</sup>  | 1976-1998, source: PU and PUSAIR  |
| Reference of detailed report | <a href="http://old.icraf.org/regions/southeast_asia/publications?do=view_pub_detail&amp;pub_no=PP0343-14">http://old.icraf.org/regions/southeast_asia/publications?do=view_pub_detail&amp;pub_no=PP0343-14</a> | <a href="http://worldAgroforestry.org/regions/southeast_asia/publications?do=view_pub_detail&amp;pub_no=PO0292-13">http://worldAgroforestry.org/regions/southeast_asia/publications?do=view_pub_detail&amp;pub_no=PO0292-13</a> | <a href="http://worldAgroforestry.org/regions/southeast_asia/publications?do=view_pub_detail&amp;pub_no=MN0048-11">http://worldAgroforestry.org/regions/southeast_asia/publications?do=view_pub_detail&amp;pub_no=MN0048-11</a> | <a href="http://worldAgroforestry.org/regions/southeast_asia/publications?do=view_pub_detail&amp;pub_no=MN0048-11">http://worldAgroforestry.org/regions/southeast_asia/publications?do=view_pub_detail&amp;pub_no=MN0048-11</a> |

1452 Note:

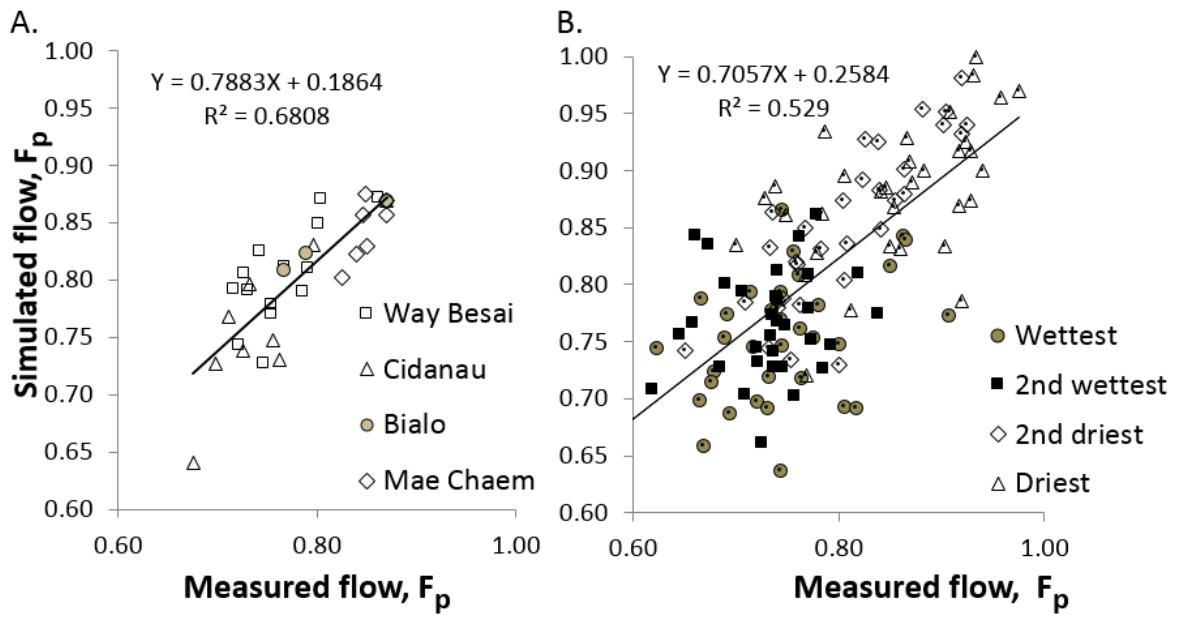
1453 <sup>a</sup> BWS: Balai Wilayah Sungai (*Regional River Agency*)1454 <sup>b</sup>PUSAIR: Pusat Litbang Sumber Daya Air (*Centre for Research and Development on Water Resources*)1455 <sup>c</sup>BMKG: Badan Meteorologi Klimatologi dan Geofisika (*Agency on Meteorology, Climatology and Geophysics*)1457 <sup>d</sup>PU: Dinas Pekerjaan Umum (*Public Work Agency*)1458 <sup>e</sup>PLN: Perusahaan Listrik Negara (*National Electric Company*)1459 <sup>f</sup>KTI: Krakatau Tirta Industri, a private steel company1460 <sup>g</sup>ICARM: The International Centre for Water Hazard and Risk Management



1463 Figure 1. Location of the four watersheds in the agroecological zones of Southeast Asia (water  
 1464 towers are defined on the basis of ability to generate river flow and being in the upper part of a  
 1465 watershed)



1469 Figure 2. Flow persistence ( $F_p$ ) estimates derived from measurements in four Southeast Asian  
 1470 watersheds, separately for the wettest and driest 3-month periods of the year

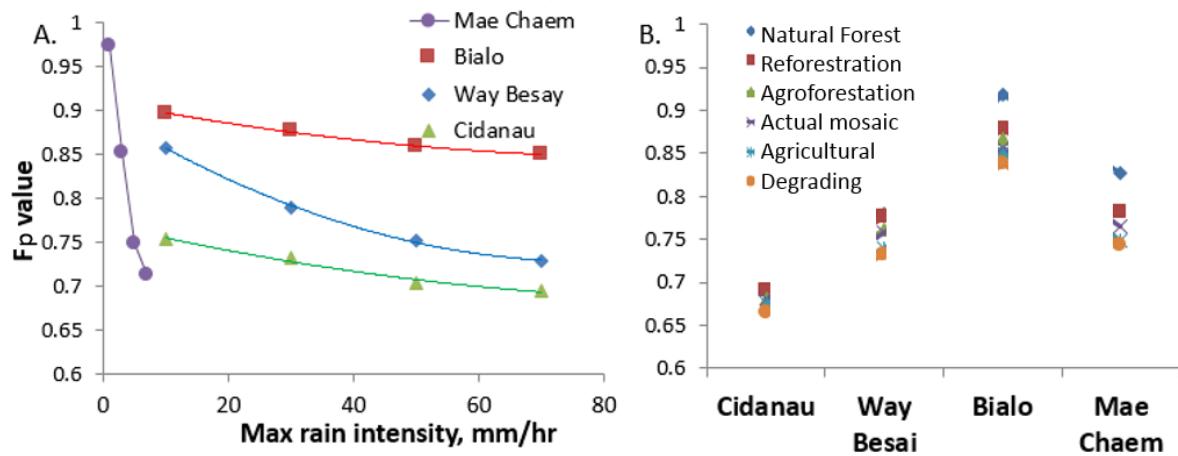


1473

1474 Figure 3. Inter- (A) and intra- (B) annual variation in the  $F_p$  parameter derived from empirical versus  
1475 modeled flow: for the four test sites on annual basis (A) or three-monthly basis (B)

1476

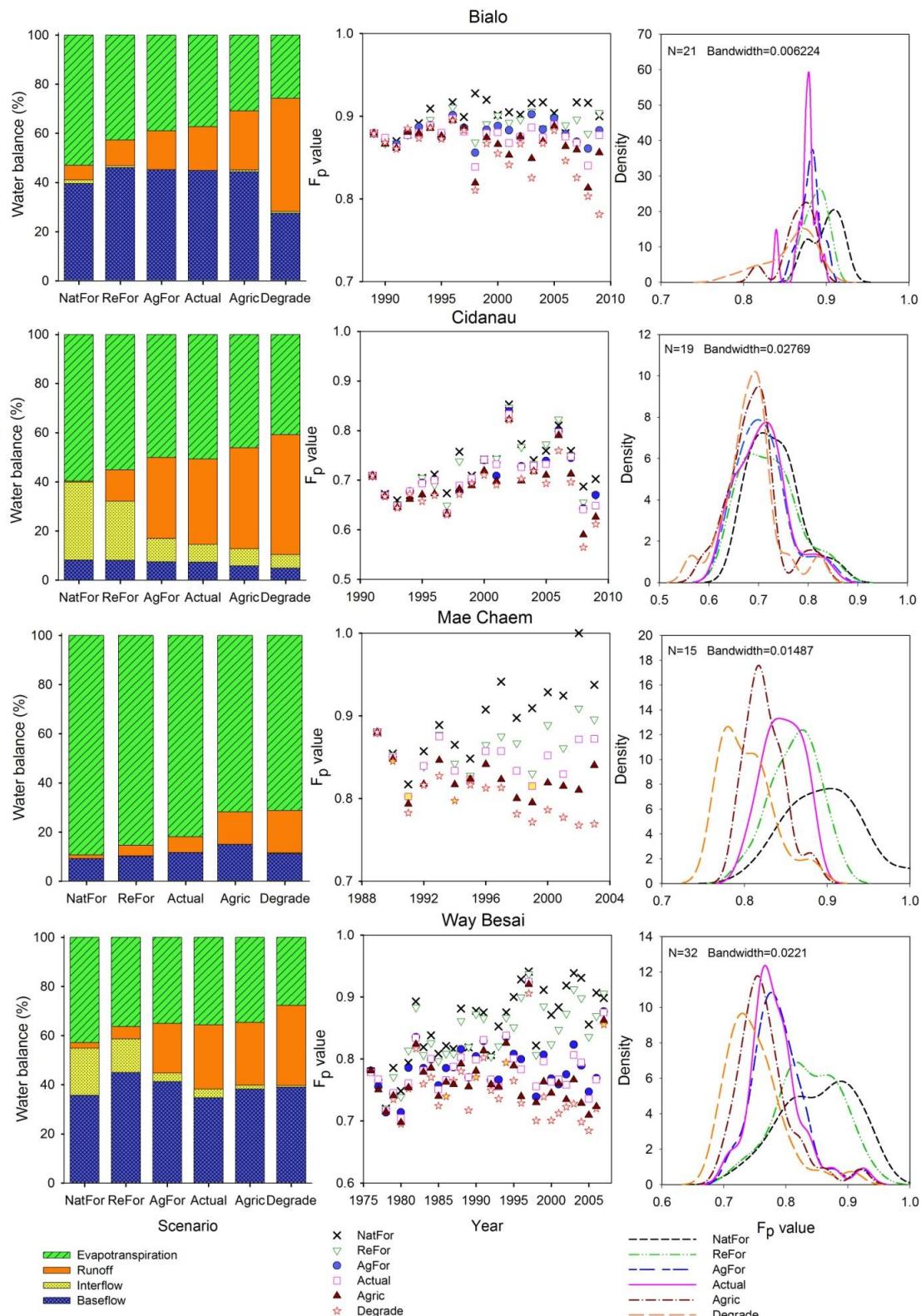
1477



1478

1479 Figure 4 Effects on flow persistence of changes in A) the mean rainfall intensity and B)  
 1480 change scenarios of Table 4 across the four watersheds

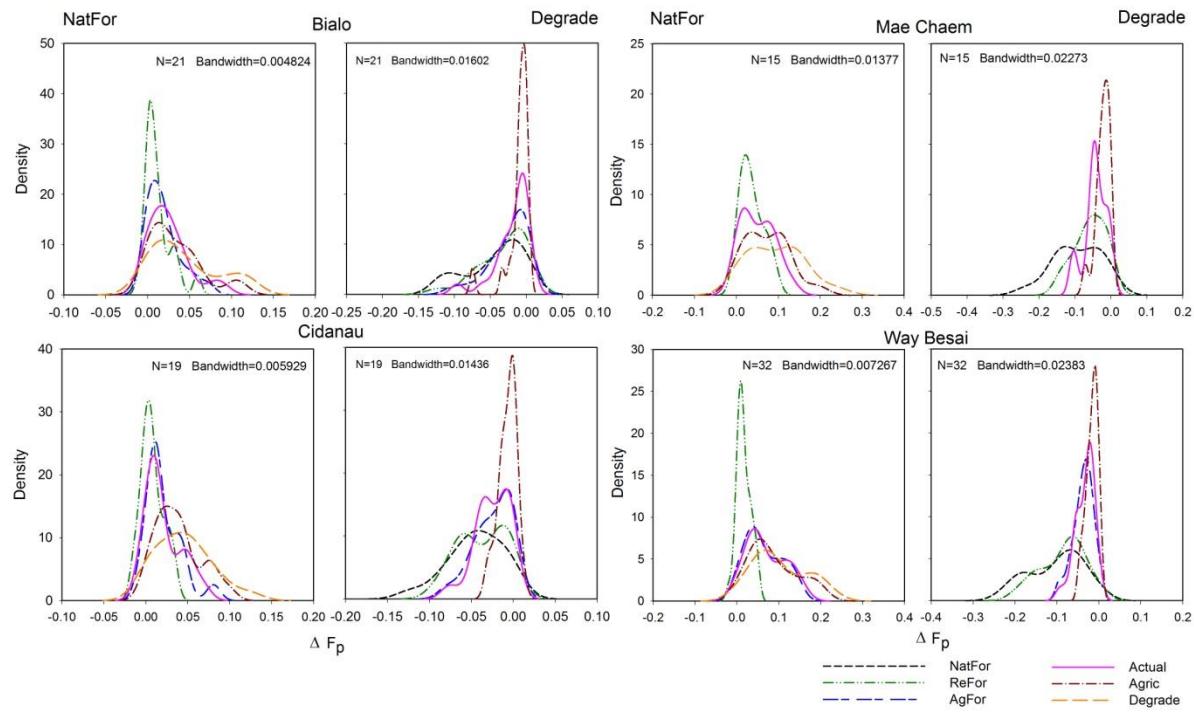
1481



1482

1483 Figure 5. Effects of land cover change scenarios (Table 4) on the flow persistence value in four  
 1484 watersheds, modelled in GenRiver over a 20-year time-period, based on actual rainfall records;  
 1485 the left side panels show average water balance for each land cover scenario, the middle panels  
 1486 the  $F_p$  values per year and land use, the right-side panels the derived frequency distributions  
 1487 (best fitting Weibull distribution)

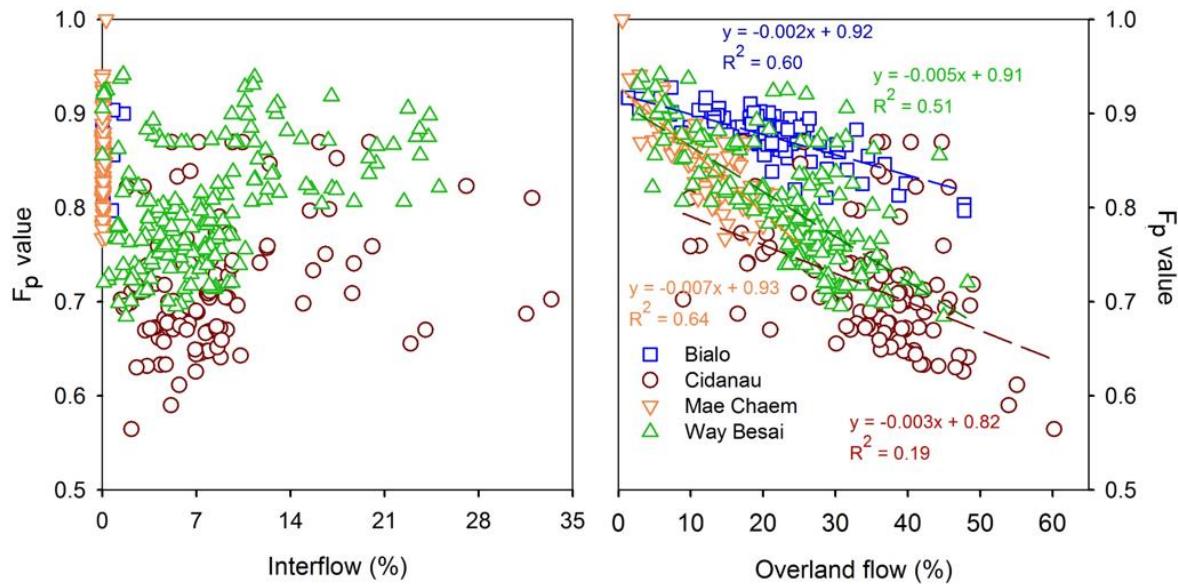
1488



1489

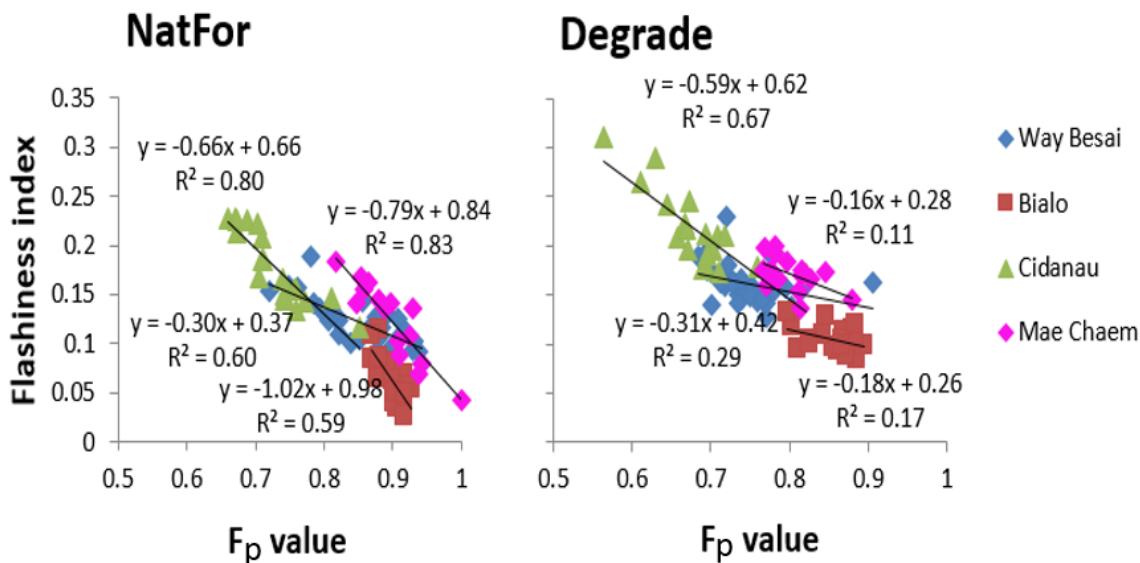
1490 Figure 6. Frequency distribution of expected difference in  $F_p$  in 'paired plot' comparisons where land  
1491 cover is the only variable; left panels: all scenarios compared to 'Reforestation', right panel: all  
1492 scenarios compared to degradation; graphs are based on a kernel density estimation (smoothing)  
1493 approach

1494



1495

1496 Figure 7. Correlations of  $F_p$  with fractions of rainfall that take overland flow and interflow pathways  
 1497 through the watershed, across all years and land use scenarios of Figure App2  
 1498



1499

1500 Figure 8. Relationship between  $F_p$  value and R-B Flashiness index across years in four Southeast Asian  
 1501 watersheds under a 'natural forest' and 'degradation' scenario, simulated with the GenRiver model

1502 Appendix 1. GenRiver model for effects of land cover on river flow

1503 The Generic River flow (GenRiver) model (van Noordwijk et al., 2011) is a simple hydrological model  
1504 that simulates river flow based on water balance concept with a daily time step and a flexible spatial  
1505 subdivision of a watershed that influences the routing of water. The core of the GenRiver model is a  
1506 “patch” level representation of a daily water balance, driven by local rainfall and modified by the  
1507 land cover and land cover change and soil properties. The model starts accounting of rainfall or  
1508 /precipitation (P) and traces the subsequent flows and storage in the landscape that can lead to  
1509 either evapotranspiration (E), river flow (Q) or change in storage ( $\Delta S$ ) (Figure App1):

1510  $P = Q + E + \Delta S$  [1]

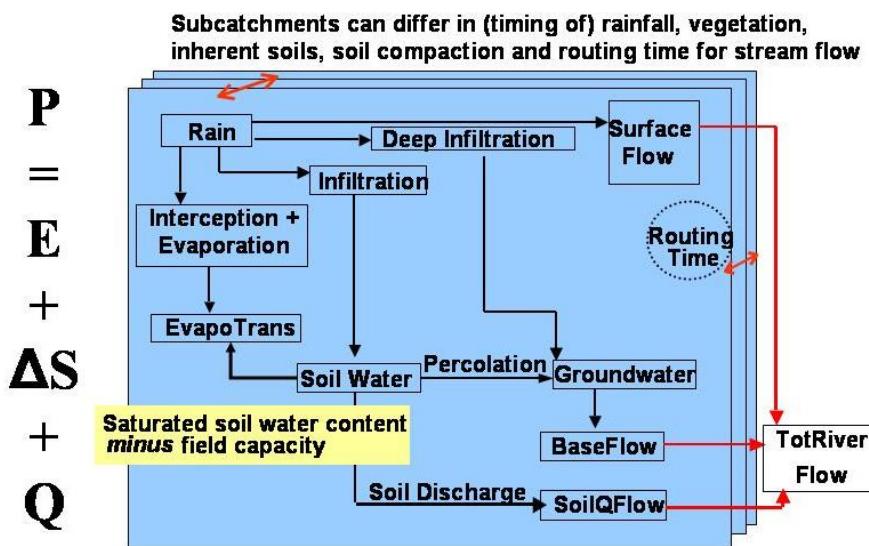


Figure App1.Overview of the GenRiver model

1511

1512 The model may use measured rainfall data, or use a rainfall generator that involves Markov chain  
1513 temporal autocorrelation (rain persistence). The model can represent spatially explicit rainfall, with  
1514 stochastic rainfall intensity (parameters RainIntensMean, RainIntensCoefVar in Table 2) and partial  
1515 spatial correlation of daily rainfall between subcatchments. Canopy interception leads to direct  
1516 evaporation of an amount of water controlled by the thickness of waterfilm on the leaf area that  
1517 depends on the land cover, and a delay of water reaching the soil surface (parameter  
1518 RainMaxIntDripDur in Table 2). The effect of evaporation of intercepted water on other components  
1519 of evapotranspiration is controlled by the InterceptEffectontrans parameter that in practice may  
1520 depend on the time of day rainfall occurs and local climatic conditions such as windspeed)

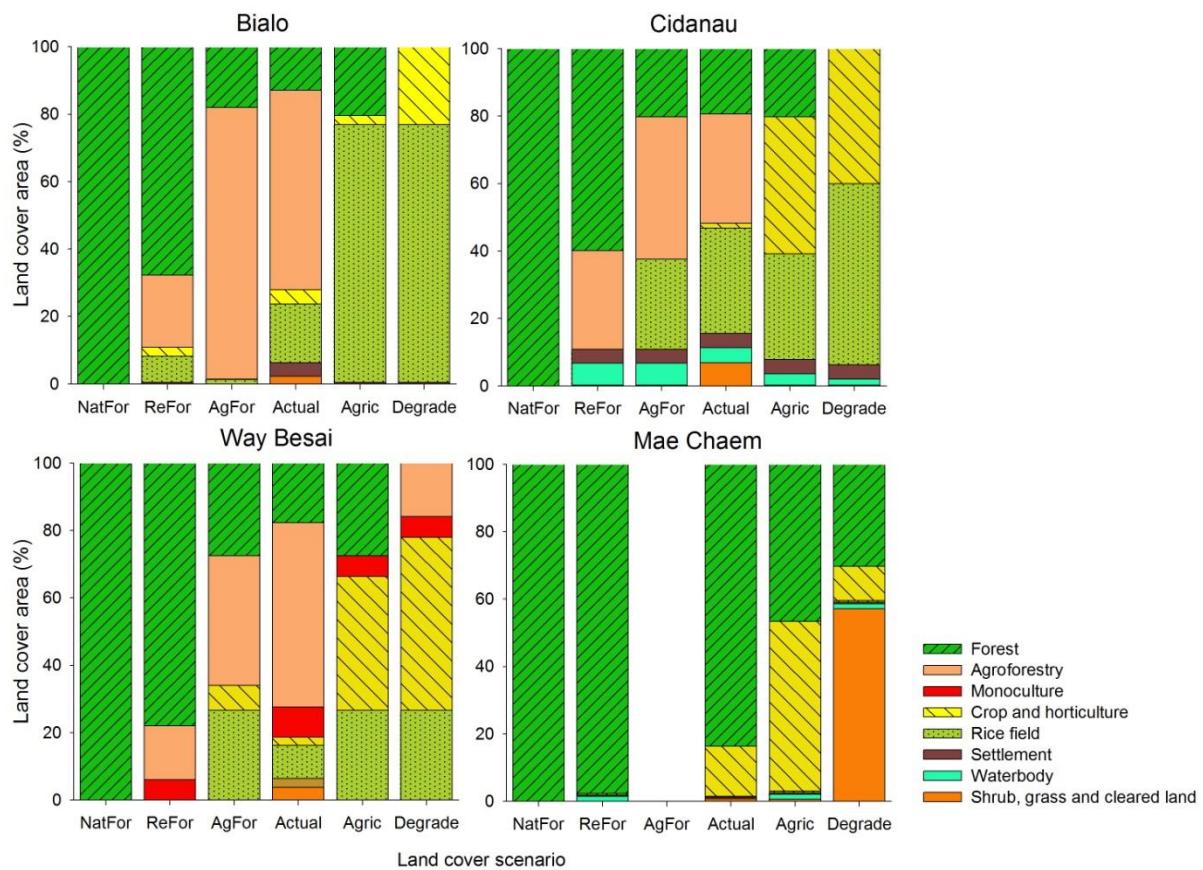
1521 At patch level, vegetation influences interception, retention for subsequent evaporation and delayed  
1522 transfer to the soil surface, as well as the seasonal demand for water. Vegetation (land cover) also  
1523 influences soil porosity and infiltration, modifying the inherent soil properties. Groundwater pool  
1524 dynamics are represented at subcatchment rather than patch level, integrating over the landcover  
1525 fractions within a subcatchment. The output of the model is river flow which is aggregated from  
1526 three types of stream flow: surface flow on the day of the rainfall event; interflow on the next day;  
1527 and base flow gradually declining over a period of time. The multiple subcatchments that make up  
1528 the catchment as a whole can differ in basic soil properties, land cover fractions that affect

1529 interception, soil structure (infiltration rate) and seasonal pattern of water use by the vegetation.  
1530 The subcatchment will also typically differ in “routing time” or in the time it takes the streams and  
1531 river to reach any specified observation point (with default focus on the outflow from the  
1532 catchment). The model itself (currently implemented in Stella plus Excel), a manual and application  
1533 case studies are freely available (<http://www.worldAgroforestry.org/output/genriver-generic-river-model-river-flow>;van Noordwijk et al., 2011).

1535

1536 Appendix 2. Watershed-specific consequences of the land use change scenarios

1537 The generically defined land use change scenarios (Table 4) led to different land cover proportions,  
1538 depending on the default land cover data for each watershed, as shown in Figure App2.



1539  
1540 Figure App2. Land use distribution of the various land use scenarios explored for the four  
1541 watersheds (see Table 4)

1542

```

1543 Appendix 3. Example of a macro in R to estimate number of observation required using bootstrap
1544 approach.

1545

1546 #The bootstrap procedure is to calculate the minimum sample size (number of observation) required
1547 #for a significant land use effect on Fp
1548 #bialo1 is a dataset contains delta Fp values for two different from Bialo watershed
1549
1550 #read data
1551 bialo1 <- read.table("bialo1.csv", header=TRUE, sep=",")
1552
1553 #name each parameter
1554 BL1 <- bialo1$ReFor
1555 BL5 <- bialo1$Degrade
1556
1557 N = 1000 #number replication
1558
1559 n <- c(5:50) #the various sample size
1560
1561 J <- 46 #the number of sample size being tested (~ number of actual year observed in the dataset)
1562
1563 P15= matrix(ncol=J, nrow=R) #variable for storing p-value
1564 P15Q3 <- numeric(J) #for storing p-Value at 97.5 quantile
1565
1566 for (j in 1:J) #estimating for different n
1567
1568 #bootstrap sampling
1569 {
1570 for (i in 1:N)
1571 {
1572 #sampling data
1573 S1=sample(BL1, n[j], replace = T)
1574 S5=sample(BL5, n[j], replace = T)
1575
1576 #Kolmogorov-Smirnov test for equal distribution and get the p-Value
1577 KS15 <- ks.test(S1, S5, alt = c("two.sided"), exact = F) P15[i,j] <- KS15$p.value
1578 }
1579
1580 #Confidence interval of CI
1581 P15Q3[j] <- quantile(P15[,j], 0.975)
1582
1583 }
1584
1585 #saving P value data and CI
1586
1587 write.table(P15, file = "pValue15.txt") write.table(P15Q3, file = "P15Q3.txt")v
1588 /

```