Hydrology and
Earth System
Sciences

Flood risk reduction and flow buffering as ecosystem services – Part 2: Land use and rainfall intensity effects in Southeast Asia

Meine van Noordwijk[1,2], Lisa Tanika[1], and Betha Lusiana[1]

[1]World Agroforestry Centre, Bogor, Indonesia
[2]Plant Production System, Wageningen University, Wageningen, the Netherlands

Correspondence to: Meine van Noordwijk (m.vannoordwijk@cgiar.org)

Abstract. Watersheds buffer the temporal pattern of river flow relative to the temporal pattern of rainfall. This "ecosystem service" is inherent to geology and climate, but buffering also responds to human use and misuse of the landscape. Buffering can be part of management feedback loops if salient, credible and legitimate indicators are used. The flow persistence parameter F_p in a parsimonious recursive model of river flow (Part 1, van Noordwijk et al., 2017) couples the transmission of extreme rainfall events $(1 - F_p)$, to the annual base-flow fraction of a watershed (F_p). Here we compare F_p estimates from four meso-scale watersheds in Indonesia (Cidanau, Way Besai and Bialo) and Thailand (Mae Chaem), with varying climate, geology and land cover history, at a decadal timescale. The likely response in each of these four to variation in rainfall properties (including the maximum hourly rainfall intensity) and land cover (comparing scenarios with either more or less forest and tree cover than the current situation) was explored through a basic daily water-balance model, GenRiver. This model was calibrated for each site on existing data, before being used for alternative land cover and rainfall parameter settings. In both data and model runs, the wet-season (3-monthly) F_p values were consistently lower than dry-season values for all four sites. Across the four catchments F_p values decreased with increasing annual rainfall, but specific aspects of watersheds, such as the riparian swamp (peat soils) in Cidanau reduced effects of land use change in the upper watershed. Increasing the mean rainfall intensity (at constant monthly totals for rainfall) around the values considered typical for each landscape was predicted to cause a decrease in F_p values by between 0.047 (Bialo) and 0.261 (Mae Chaem). Sensitivity of F_p to changes in land use change plus changes in rainfall intensity depends on other characteristics of the watersheds, and generalisations made on the basis of one or two case studies may not hold, even within the same climatic zone. A wet-season F_p value above 0.7 was achievable in forest–agroforestry mosaic case studies. Inter-annual variability in F_p is large relative to effects of land cover change. Multiple (5–10) years of paired-plot data would generally be needed to reject no-change null hypotheses on the effects of land use change (degradation and restoration). F_p trends over time serve as a holistic scale-dependent performance indicator of degrading/recovering watershed health and can be tested for acceptability and acceptance in a wider social-ecological context.

1 Introduction

Inherent properties (geology, geomorphology) interact with climate and human modification of vegetation, soils, drainage and riparian wetlands in effectuating the degree of buffering that watersheds provide (Andréassian, 2004; Bruijnzeel, 2004). Buffering of river flow relative to the space–time dynamics of rainfall is an ecosystem service, reducing the exposure of people living on geomorphological flood-plains to high-flow events, and increasing predictability and river flow in dry periods (Joshi et al., 2004; Leimona et al., 2015). In the absence of any vegetation and with a sealed surface, river flow will directly respond to the spatial distribution of rainfall, with only the travel time to any point of specific interest influencing the temporal pattern of river flow. Any persistence or predictability of river flow in such a situation will reflect temporal autocorrelation of rainfall, beyond statistical predictability in seasonal rainfall patterns. On

the other side of the spectrum, river flow can be constant every day, beyond the theoretical condition of constant rainfall, in a watershed that provides perfect buffering, by passing all water through groundwater pools that have sufficient storage capacity at any time during the year. Both infiltration-limited (Hortonian) and saturation-induced use of more rapid flow pathways (inter and overland flows) will reduce the flow persistence and make it, at least in part, dependent on rainfall events. Separating the effects of land cover (land use), engineering and rainfall on the actual flow patterns of rivers remains a considerable challenge (Ma et al., 2014; Verbist et al., 2010). It requires data, models and concepts that can serve as an effective boundary object in communication with stakeholders (Leimona et al., 2015; van Noordwijk et al., 2012, 2016). There is a long tradition in using forest cover as such a boundary object, but there is only a small amount of evidence supporting this (Tan-Soo et al., 2016; van Dijk et al., 2009; van Noordwijk et al., 2015a).

In Part 1 (van Noordwijk et al., 2017), we introduced a flow persistence parameter (F_p) that links the two, asymmetrical aspects of flow dynamics: translating rainfall excess into river flow, and gradually releasing water stored in the landscape. The direct link between these two aspects can be seen from Eq. (4) in Part 1 (van Noordwijk et al., 2017):

$$Q_t = F_p Q_{t-1} + \left(1 - F_p\right) \left(P_t - E_{tx}\right), \tag{1}$$

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

Commonly used rainfall–runoff models (including the curve number approach and SWAT models) only focus on the second term of the above equation (Ponce et al., 1996; Gassman et al., 2007), without link to the first. Various empirical methods for deriving "base flow" are in use, but details of the calculation procedure matter. Results in Part 1 (van Noordwijk et al., 2017) for a number of contrasting meso-scale watersheds in Southeast Asia suggested that inter-annual variation in F_p within a given watershed correlates with both the Richards–Baker (R–B) flashiness index (Baker et al., 2004) and the base-flow fraction of annual river flow. However, the slope of these relationships varied between watersheds. Here, in Part 2 we will further analyse the F_p results for these watersheds that were selected to represent variation in rainfall and land cover, and test the internal

consistency of results based on historical data: two located in the humid and one in the subhumid tropics of Indonesia, and one in the unimodal subhumid tropics of northern Thailand.

After exploring the patterns of variation in F_p estimates derived from actual river flow records, we will quantify the sensitivity of the F_p metric to variations in rainfall intensity and its response, on a longer timescale to land cover change. To do so, we will use a model that uses basic water-balance concepts: rainfall interception, infiltration, water use by vegetation, overland flow, interflow and groundwater release, to a spatially structured watershed where travel time from sub-watersheds to any point of interest modifies the predicted river flow. In the specific model used, land cover effects on soil conditions, interception and seasonal water use have been included. After testing whether F_p values derived from model outputs match those based on empirical data where these exist, we rely on the basic logic of the model to make inference on the relative importance of modifying rainfall and land cover inputs. With the resulting temporal variation in calculated F_p values, we consider the time frame at which observed shifts in F_p can be attributed to factors other than chance (i.e. null hypotheses of random effects can be rejected with accepted chance of Type I errors).

2 Methods

2.1 GenRiver model for effects of land cover on river flow

The GenRiver model (van Noordwijk et al., 2011) is based on a simple water-balance concept with a daily time step and a flexible spatial subdivision of a watershed that influences the routing of water and employs spatially explicit rainfall. At patch level, vegetation influences interception, retention for subsequent evaporation and delayed transfer to the soil surface, as well as the seasonal demand for water. Vegetation (land cover) also influences soil porosity and infiltration, modifying the inherent soil properties. Water in the root zone is modelled separately for each land cover within a subcatchment, the groundwater stock is modelled at sub-catchment level. The spatial structure of a watershed and the routing of surface flows influences the time delays to any specified point of interest, which normally includes the outflow of the catchment. Land cover change scenarios are interpolated annually between time series (measured or modelled) data. The model may use measured rainfall data, or use a rainfall generator that involves Markov chain temporal autocorrelation (rain persistence). As our data sources are mostly restricted to daily rainfall measurements and the infiltration model compares instantaneous rainfall to infiltration capacity, a stochastic rainfall intensity was applied at subcatchment level, driven by the mean as parameter and a standard deviation for a normal distribution (truncated at 3 standard deviations from the mean) proportional to it via
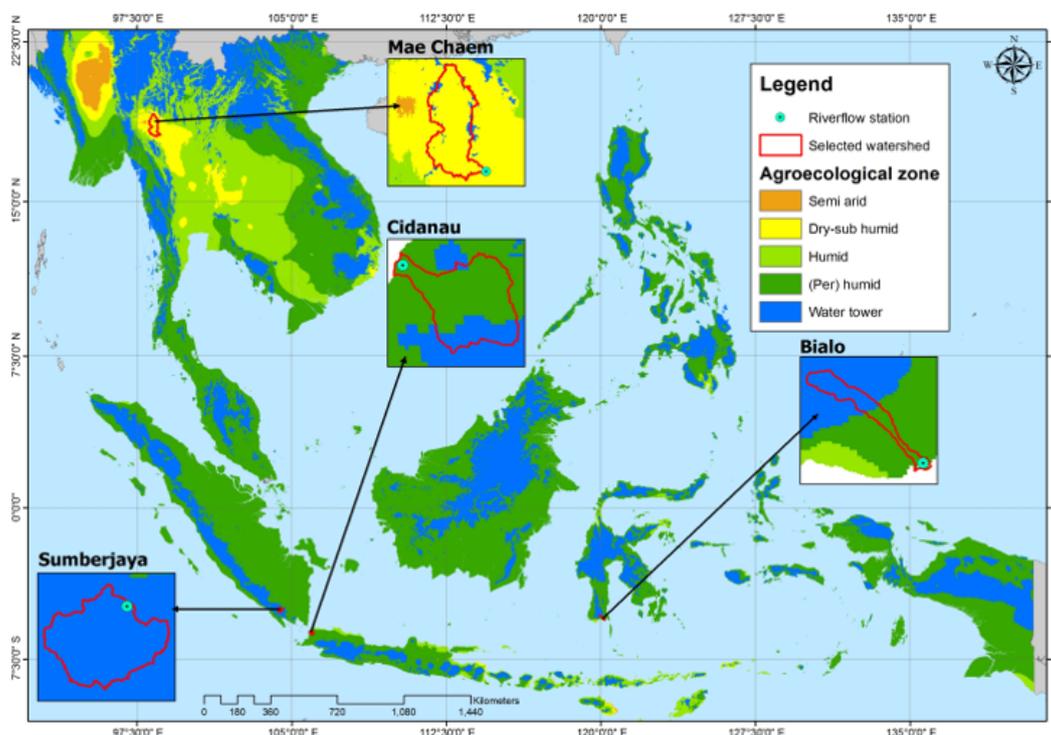
Figure 1. Location of the four watersheds in the agro-ecological zones of Southeast Asia (water towers are defined on the basis of its ability to generate river flow and being in the upper part of a watershed).

a coefficient of variation as parameter. For the Mae Chaem site in northern Thailand, data by Dairaku et al. (2004) suggested a mean of less than $3 \, \mathrm{mm} \, \mathrm{h}^{-1}$. For the three sites in Indonesia we used $30 \, \mathrm{mm} \, \mathrm{h}^{-1}$, based on Kusumastuti et al. (2016). Appendix B provides further detail on the GenRiver model. The model itself, a manual and application case studies, are freely available (http://www.worldAgroforestry.org/output/genriver-generic-river-model-river-flow; van Noordwijk et al., 2011).

2.2 Empirical data sets, model calibration

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

As major parameters for the GenRiver model were not independently measured for the respective watersheds, we tuned (calibrated) the model by modifying parameters within a predetermined plausible range, and used correspondence with measured hydrograph as test criterion (Kobold et al., 2008). We used the Nash–Sutcliffe efficiency (NSE) parameter (target above 0.5) and bias (less than 25 %) as test criteria

and targets. Meeting these performance targets (Moriasi et al., 2007), we accepted the adjusted models as basis for describing current conditions and exploring model sensitivity. The main site-specific parameter values are listed in Table 2 and (generic) land-cover-specific default parameters in Table 3.

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

2.3 Bootstrapping to estimate the minimum observation

The bootstrap method (Efron and Tibshirani, 1986) is a resampling methods that is commonly used to generate "surrogate population" for the purpose of approximating the sampling distribution of a statistic. In this study, the bootstrap approach was used to estimate the minimum number of observation (or yearly data) required for a pair-wise comparison test between two time series of streamflow or discharge data (representing two scenarios of land use distributions) to be distinguishable from a null hypothesis of no effect. The pair-wise comparison test used was Kolmogorov–Smirnov test that is commonly used to test the distribution of dis-

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
Coordinate	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^2)	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), agroforest (59), crops (22), others (6)	Forest (20), agroforest (32), crops (33), others (11), swamp (4)	forest (evergreen, deciduous and pine) (84), crops (15), others (1)	Forest (18), coffee (monoculture and multi-strata) (64), crop and horticulture (12), 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	Inceptisols	Inceptisols

charge data (Zhang et al., 2006). We built a simple macro in R (R Core Team, 2017) that entails the following steps:

i. bootstrap or resample with replacement 1000 times from both time series discharge data with sample size n;

ii. apply the Kolmogorov–Smirnov test to each of the 1000 generated pair-wise discharge data, and record the p value;

iii. perform (i) and (ii) for different size of n, ranging from 5 to 50;

iv. tabulate the p value from the different sample size n, and determine the value of n when the p value reached equal to or less than 0.025 (or equal to the significance level of 5 %); the associated n represents the minimum number of observations required.

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

3 Results

3.1 Empirical data of flow persistence as basis for model parameterisation

Inter-annual variability of F_p estimates derived for the four catchments (Fig. 2) was of the order of 0.1 units, while the intra-annual variability between dry and rainy seasons was 0.1–0.2. For all years and locations, rainy-season F_p values, with mixed flow pathways, were consistently below dry-season values, dominated by groundwater flows. If we can expect $F_{p,i}$ and $F_{p,0}$ (see Eq. 8 in Part 1, van Noordwijk et al., 2017) to be approximately 0.5 and 0, this difference between wet and dry periods implies a 40 % contribution of interflow in the wet season, a 20 % contribution of overland flow or any combination of the two effects.

Overall the estimates from modelled and observed data are related with 16 % deviating more than 0.1 and 3 % more than 0.15 (Fig. 3). As the Moriasi et al. (2007) performance criteria for the hydrographs were met by the calibrated models for each site, we tentatively accept the model to be a basis for a sensitivity study of F_p to modifications to land cover and/or rainfall.
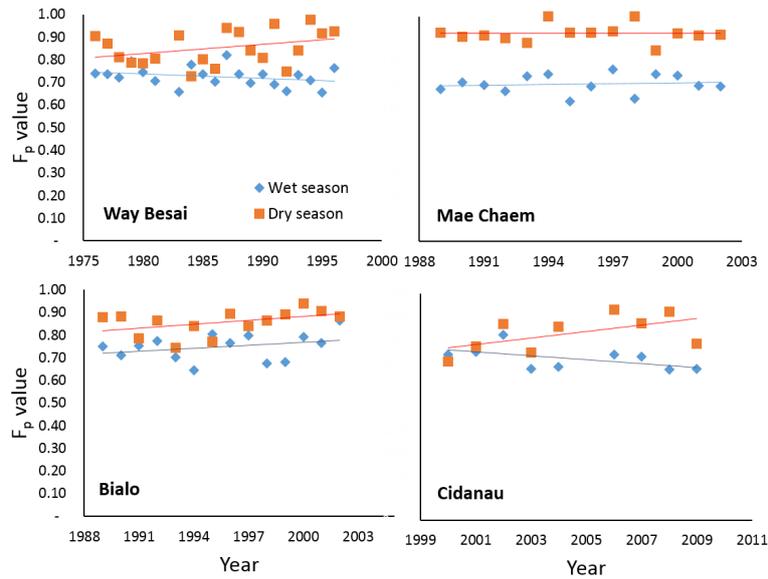
Figure 2. Flow persistence (F_p) estimates derived from measurements in four Southeast Asian watersheds, separately for the wettest and driest 3-month periods of the year.
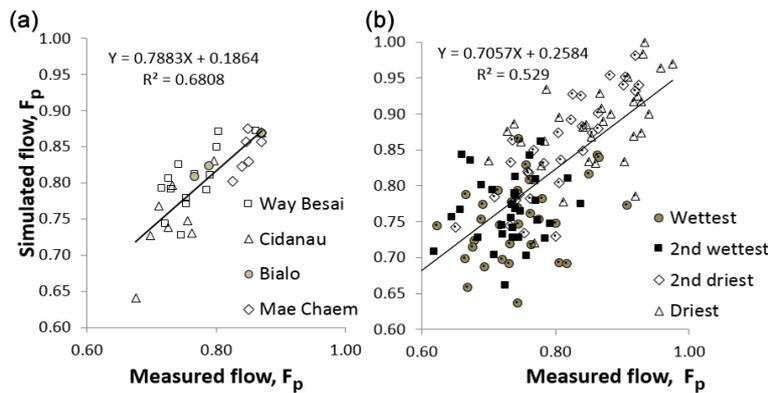

Figure 3. Inter- (**a**) and intra-annual (**b**) variation in the F_p parameter derived from empirical vs. modelled flow: for the four test sites on annual basis (**a**) or 3-monthly basis (**b**).

3.2 Comparing F_p effect of rainfall intensity and land cover change

A direct comparison of model sensitivity to changes in mean rainfall intensity and land use change scenarios is provided in Fig. 4. Varying the mean rainfall intensity over a factor 7 shifted the F_p value by only 0.047 and 0.059 in the case of Bialo and Cidanau, respectively, but by 0.128 in Way Besai and 0.261 in Mae Chaem (Fig. 4a). The impact of the land use change scenarios on F_p was the smallest in Cidanau (0.026), intermediate in Way Besai (0.048) and relatively large in Bialo and Mae Chaem, at 0.080 and 0.084, respectively (Fig. 4b). The order of F_p across the land use change scenarios was mostly consistent between the watersheds, but the contrast between the reforestation and natural forest scenario was the largest in Mae Chaem and the small-

est in Way Besai. In Cidanau, Way Besai and Mae Chaem, variations in rainfall were 2.2 to 3.1 times more effective than land use change in shifting F_p, whereas in Bialo its relative effect was only 58 %. Apparently, the sensitivity to changes in land use change plus changes in rainfall intensity depends on other characteristics of the watersheds, and generalisations made on the basis of one or two case studies may not hold, even within the same climatic zone.

3.3 Further analysis of F_p effects for scenarios of land cover change

Among the four watersheds there is consistency in that the "forest" scenario has the highest, and the "degraded lands" the lowest, F_p value (Fig. 5), but there are remarkable differences as well; in Cidanau the inter-annual variation in F_p

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 h^{-1}	30	30	3	30
RainIntensCoefVar	Coefficient of variation of rainfall intensity	mm h^{-1}	0.8	0.3	0.5	0.3
RainInterceptDripRt	Maximum drip rate of intercepted rain	mm h^{-1}	80	10	10	10
RainMaxIntDripDur	Maximum dripping duration of intercepted rain	h	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 day^{-1}	580	800	150	720
MaxInfSubsoil	Maximum infiltration capacity of the sub-soil	mm day^{-1}	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
DispersalFactor	Drainage density	–	0.3	0.4	0.3	0.45
RiverVelocity	River flow velocity	m s^{-1}	0.4	0.7	0.35	0.5
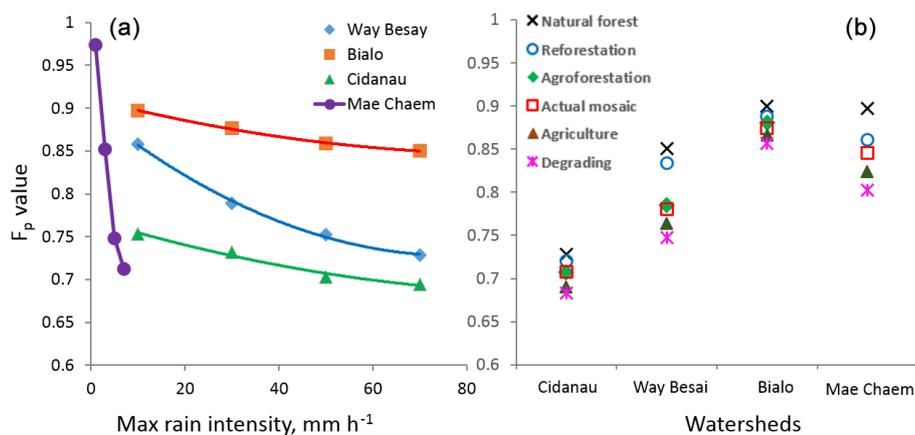
Figure 4. Inter- **(a)** and intra-annual **(b)** variation in the F_p parameter derived from empirical vs. modelled flow: for the four test sites on annual basis **(a)** or 3-monthly basis **(b)**.
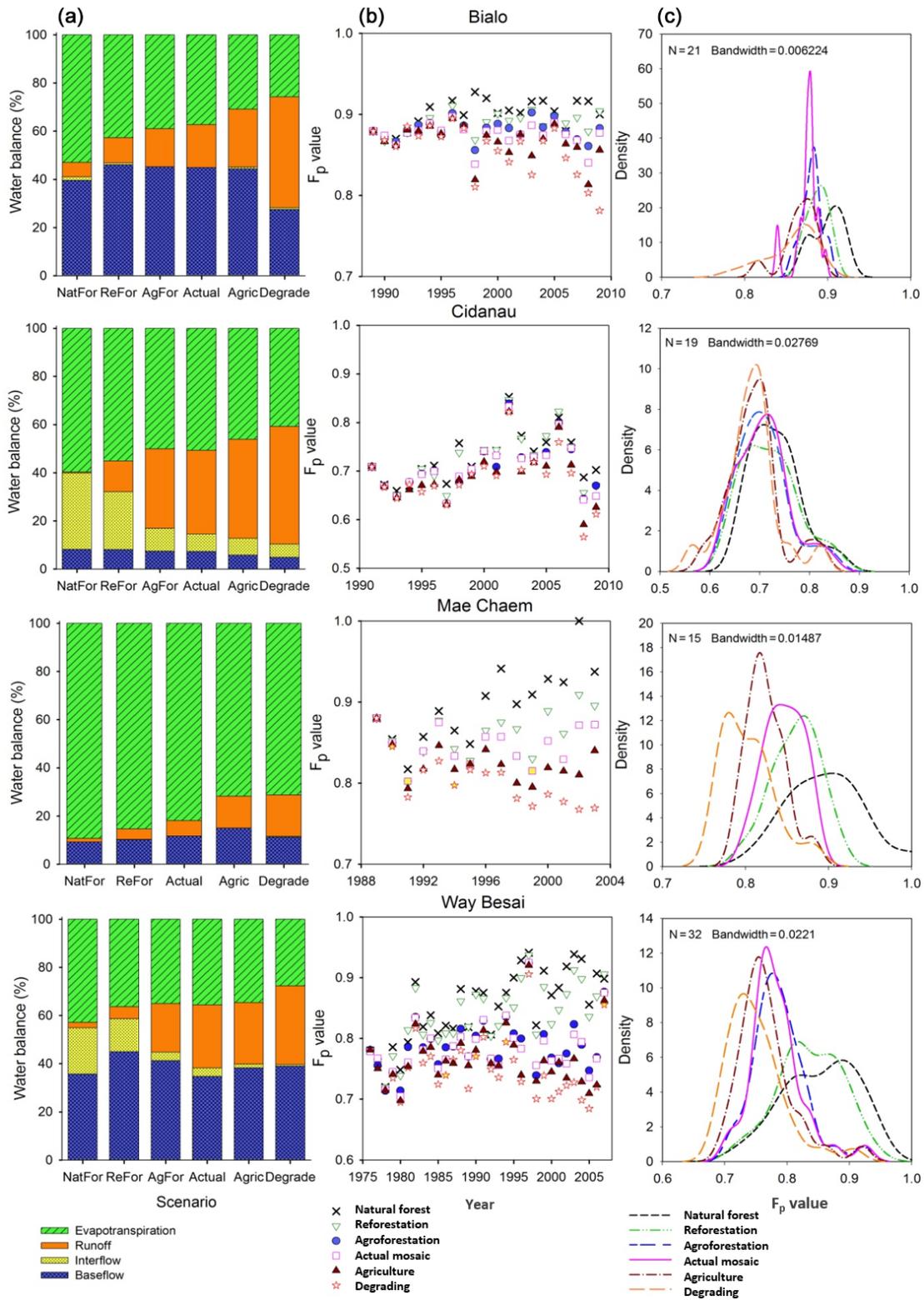
Figure 5. Effects of land cover change scenarios (Table 4) on the flow persistence value in four watersheds, modelled in GenRiver over a 20-year time period, based on actual rainfall records; **(a)** the average water balance for each land cover scenario, **(b)** the F_p values per year and land use, **(c)** the derived frequency distributions (best-fitting Weibull distribution).
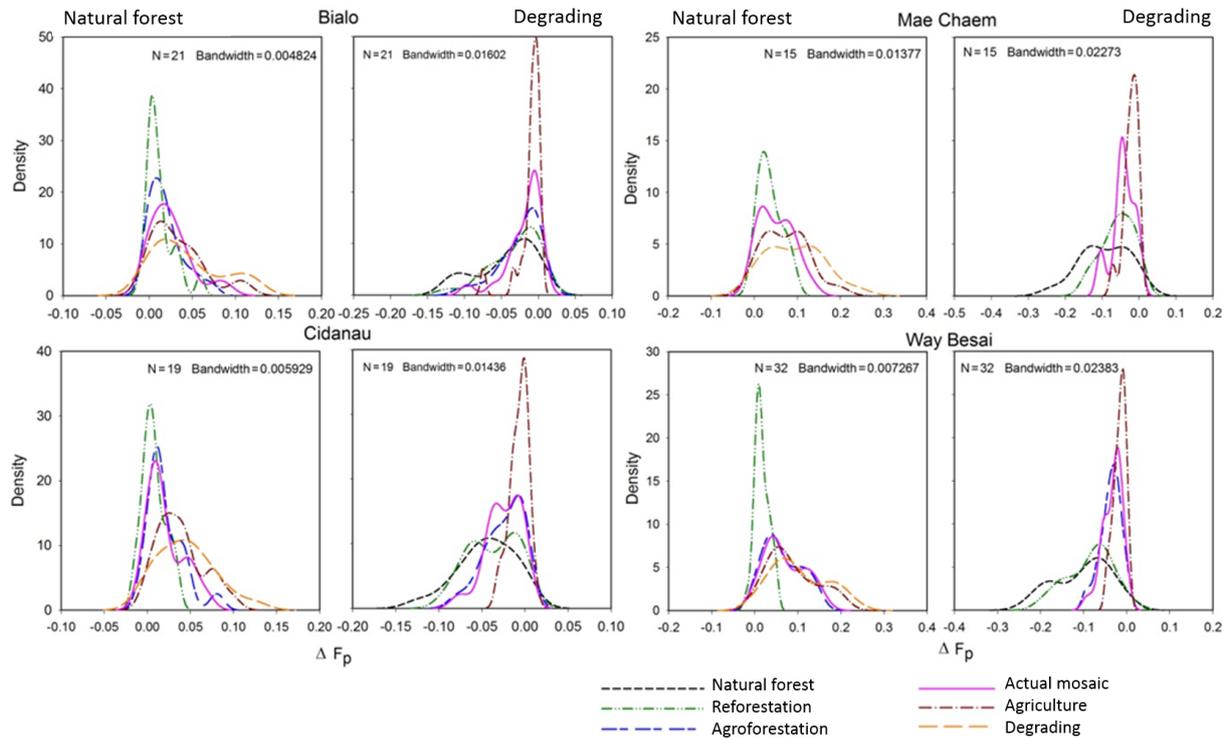
Figure 6. Frequency distribution of expected difference in F_p in "paired-plot" comparisons where land cover is the only variable; left panels: all scenarios compared to "reforestation"; right panel: all scenarios compared to degradation; graphs are based on a kernel density estimation (smoothing) approach.

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

Land cover type	Potential interception (mm day^{-1})	Relative drought threshold	Bd/BDref
Forest[1]	3.0–4.0	0.4–0.5	0.8–1.1
Agroforestry[2]	2.0–3.0	0.5–0.6	0.95–1.05
Monoculture tree[3]	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[4]	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

Note [1] forest: primary forest, secondary forest, swamp forest, evergreen forest, deciduous forest. [2] agroforest: mixed garden, clove, coffee, cocoa. [3] monoculture: coffee; [4] rice field: irrigation and rainfed.

is clearly larger than land cover effects, while in the Way Besai the spread in land use scenarios is larger than inter-annual variability. In Cidanau a peat swamp between most of the catchment and the measuring point buffers most of

land-cover-related variation in flow, but not the inter-annual variability. Considering the frequency distributions of F_p values over a 20-year period, we see one watershed (Way Besai) where the forest stands out from all others, and one (Bialo) where the degraded lands are separate from the others. Given the degree of overlap of the frequency distributions, it is clear that multiple years of empirical observations will be needed before a change can be affirmed.

Figure 5 shows the frequency distributions of expected effect sizes on F_p of a comparison of any land cover with either forest or degraded lands. Table 5 translates this information to the number of years that a paired plot (in the absence of measurement error) would have to be maintained to reject a null hypothesis of no effect, at 5 % probability (p). As the frequency distributions of F_p differences of paired catchments do not match a normal distribution, a Kolmogorov–Smirnov test can be used to assess the probability that a no-difference null hypothesis can yield the difference found. By bootstrapping within the years where simulations supported by observed rainfall data exist, we found for the Way Besai catchment, for example, that 20 years of data would be needed to assert (at $p = 0.05$) that the reforestation scenario differs from agroforestation, and 16 years that it differs from actual and 11 years that it differs from degrade. In practice, that means that empirical evidence that survives statis-

Table 4. Land use scenarios explored for four watersheds.

Scenario	Description
Natural forest	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
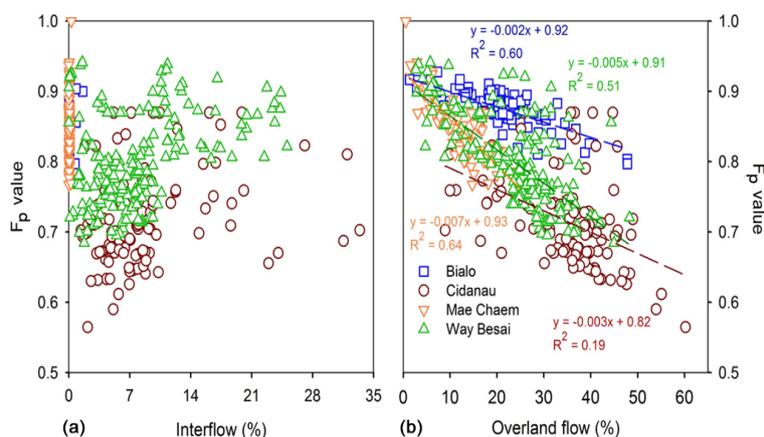
Figure 7. Correlations of F_p with fractions of rainfall that take overland flow and interflow pathways through the watershed, across all years and land use scenarios of Fig. App2.

tical tests will not emerge, even though effects on watershed health are real.

At process level the increase in "overland flow" in response to soil compaction due to land cover change has a clear and statistically significant relationship with decreasing F_p values in all catchments (Fig. 6), but both year-to-year variation within a catchment and differences between catchments influence the results as well, leading to considerable spread in the bi-plot. Contrary to expectations, the disappearance of "interflow" by soil compaction is not reflected in measurable change in F_p value. The temporal difference between overland and interflow (1 or a few days) gets easily blurred in the river response that integrates over multiple streams with variation in delivery times; the difference between overland- or interflow and base flow is much more pronounced. Apparently, according to our model, the high macroporosity of forest soils that allows for interflow, and may be the "sponge" effect attributed to forest, delays delivery to rivers by 1 or a few days, with little effect on the

flow volumes at locations downstream where flow of multiple days accumulates. The difference between overland- or interflow and base flow in time-to-river of rainfall peaks is much more pronounced.

Tree cover has two contradicting effects on base flow; it reduces the surplus of rainfall over evapotranspiration (annual water yield) by increased evapotranspiration (especially where evergreen trees or trees with a large canopy interception are involved), but it potentially increases soil macroporosity that supports infiltration and interflow, with relatively little effect on water holding capacity measured as "field capacity" (after runoff and interflow have removed excess water). Figure 7 shows that the total volume of base flow differs more between sites and their rainfall pattern than it varies with tree cover. Between years total evapotranspiration and base flow totals are positively correlated, but for a given rainfall there is a trade-off. Overall these results support the conclusion that generic effects of deforestation on decreased flow persistence, and of (agro)/(re)-forestation on increased

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

(a) Natural forest as reference				
Way Besai ($N = 32$) Reforestation	Agroforestation	Actual	Agricultural	Degrading
Reforestation –	20 (0.0035)	16 (0.037)	13 (0.046)	11 (0.023)
Agroforestation –	–	n.s	n.s.	n.s.
Actual –	–	–	n.s.	n.s.
Agricultural –	–	–	–	n.s.
Degrading –	–	–	–	–
Bialo ($N = 18$) Reforestation	Agroforestation	Actual	Agricultural	Degrading
Reforestation –	n.s.	n.s.	37 (0.04)	27 (0.040)
Agroforestation –	–	n.s	n.s.	n.s.
Actual –	–	–	n.s.	n.s.
Agricultural –	–	–	–	n.s.
Degrading –	–	–	–	–
Cidanau ($N = 20$) Reforestation	Agroforestation	Actual	Agricultural	Degrading
Reforestation –	n.s.	n.s.	32 (0.037)	48 (0.043)
Agroforestation –	–	n.s	n.s.	n.s.
Actual –	–	–	n.s.	n.s.
Agricultural –	–	–	–	n.s.
Degrading –	–	–	–	–

Mae Chaem ($N = 15$) Reforestation		Actual	Agricultural	Degrading
Reforestation –		n.s.	23 (0.049)	18 (0.050)
Actual –		–	45 (0.037)	33 (0.041)
Agricultural –		–	–	33 (0.041)
Degrading –		–	–	–

(b) Degrading scenario as reference				
Way Besai ($N = 32$) Natural forest	Reforestation	Agroforestation	Actual	Agricultural
Natural forest –	n.s.	17 (0.042)	13 (0.046)	7 (0.023)
Reforestation –	–	21 (0.037)	19 (0.026)	7 (0.023)
Agroforestation –	–	–	n.s.	28 (0.046)
Actual –	–	–	–	30 (0.029)
Agricultural –	–	–	–	–
Bialo ($N = 18$) Natural forest	Reforestation	Agroforestation	Actual	Agricultural
Natural forest –	n.s.	n.s.	41 (0.047)	19 (0.026)
Reforestation –	–	n.s.	n.s.	32 (0.037)
Agroforestation –	–	–	n.s.	n.s.
Actual –	–	–	–	n.s.
Agricultural –	–	–	–	–
Cidanau ($N = 20$) Natural forest	Reforestation	Agroforestation	Actual	Agricultural
Natural forest –	n.s.	n.s.	33 (0.041)	8 (0.034)
Reforestation –	–	n.s.	n.s.	15 (0.028)
Agroforestation –	–	–	n.s.	n.s.
Actual –	–	–	–	25 (0.031)
Agricultural –	–	–	–	–

Mae Chaem ($N = 20$) Natural forest	Reforestation		Actual	Agricultural
Natural forest –	n.s.		25 (0.031)	12 (0.037)
Reforestation –	–		n.s.	18 (0.050)
Actual –	–		–	18 (0.050)
Agricultural –	–		–	–

flow persistence are small relative to inter-annual variability due to specific rainfall patterns, and that it will be hard for any empirical data process to pickup such effects, even if they are qualitatively aligned with valid process-based models.

4 Discussion

In the discussion of Part 1 (van Noordwijk et al., 2017), the credibility questions on replicability of the F_p metric and its sensitivity to details of rainfall pattern vs. land cover as potential causes of variation were seen as requiring case studies in a range of contexts. Although the four case studies in Southeast Asia presented here cannot be claimed to represent the global variation in catchment behaviour (with absence of a snowpack and its dynamics as an obvious element of flow buffering not included), the diversity of responses among these four already point to challenges for any generic interpretation of the degree of flow persistence that can be achieved under natural forest cover, as well as its response to land cover change.

Where Fig. 8 in Part 1 (van Noordwijk et al., 2017) explored the relationship in inter-annual variation between flashiness index and F_p in the actual data for the four watersheds, we can now repeat the analysis for the modelled results for each scenario. Figure 8 presents two examples with, again, evidence that the flashiness index and F_p are related, but with considerable variation between the watersheds and a lower slope for the Cidanau watershed with its downstream flow buffering.

The empirical data summarised here for (sub)humid tropical sites in Indonesia and Thailand show that values of F_p above 0.9 are scarce in the case studies provided, but values above 0.8 were found, or inferred by the model, for forested landscapes. Agroforestry landscapes generally presented F_p values above 0.7, while open-field agriculture or degraded soils led to F_p values of 0.5 or lower. Due to differences in local context, it may not be feasible to relate typical F_p values to the overall condition of a watershed, but temporal change in F_p can indicate degradation or restoration if a location-specific reference can be found. The difference between wet- and dry-season F_p can be further explored in this context. The dry-season F_p value primarily reflects the underlying geology, with potential modification by engineering and operating rules of reservoirs, the wet-season F_p is generally lower due to partial shifts to overland and interflow pathways. Where further uncertainty is introduced by the use of modelled rather than measured river flow, the lack of fit of models similar to the ones we used here would mean that scenario results are indicative of directions of change rather than a precision tool for fine-tuning combinations of engineering and land cover change as part of integrated watershed management.

The differences in relative response of the watersheds to changes in mean rainfall intensity and land cover change suggest that generalisations derived from one or a few case studies are to be interpreted cautiously. If land cover change would influence details of the rainfall generation process (arrow 10 in Fig. 1 of Part 1, (van Noordwijk et al., 2017); e.g. through release of ice-nucleating bacteria, (Morris et al., 2014; van Noordwijk et al., 2015b; Ellison et al., 2017) this can easily dominate over effects via interception, transpiration and soil changes.

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

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

In summarising findings on the F_p metric, we can compare it with existing ones across the seven questions raised in Fig. 1 of Part 1 (van Noordwijk et al., 2017). Comparator metrics can be derived from various data sources, including the amount (and/or quality) of forest cover upstream, the fraction of flows that is technically controlled, direct records of river flow (over a short or longer time period), records of rainfall and/or models that combine landscape properties, climate and land cover. Tentative scoring for these metrics (Table 6) suggests that the F_p metric is an efficient tool for data-scarce environments, as it indicates aspects of hydrographs that so far required multi-annual records of river flow.

5 Conclusions

Overall, our analysis suggests that the level of flow buffering achieved depends on both land cover (including its spatial configuration and effects on soil properties) and space–time patterns of rainfall (including maximum rainfall intensity as determinant of overland flow). Generalisations on dominant influence of either, derived from one or a few case studies are to be interpreted cautiously. If land cover change would
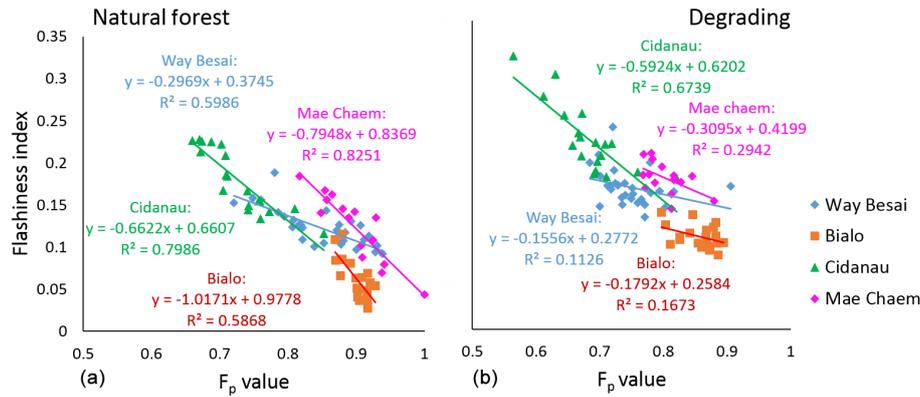
Figure 8. Relationship between F_p value and R–B Flashiness index across years in four Southeast Asian watersheds under a "natural forest" and "degrading" scenario, simulated with the GenRiver model.

Table 6. Comparison of metrics at various points in the causal network (Fig. 2 of Part 1, van Noordwijk et al., 2017) that can support watershed management and prevention of flood damage on the list of seven issues (I–VII) introduced in Fig. 1 Part 1*.

| Issues* | Terrain-based (7A and 5 in Fig. 2 of Part 1) | | Based on river flow characteristics (4 in Fig. 2 of Part 1) | | | | | | Integrated (5–7) terrain + climate + land use + river flow models | |
	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%	1–100%	1ω	0–2		1–100	0–100 %	0–1		
IA	No	Yes	No	Yes	Yes	Yes	No	Yes	Partially	Yes
IB	No	Yes	No	No	Yes	No	Yes	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

(I) Does the indicator relate to important aspects of watershed behaviour (A. flood damage prevention; B. low flow water availability)? (II) Does its quantification help to select management actions (A. risk assessment, insurance design; B. Sspatial planning, engineering interventions; C. fine-tuning land use)? (III) Is it consistent with current understanding of key processes? (IV) Are data requirements feasible (A. lowest temporal resolution for estimates (years); B. consistency of numerical results and sensitivity to bias and random error in data sources)? (V) Does it match local knowledge and concerns? (VI) Can it be used to empower local stakeholders of watershed management through performance-based (outcome) contracts? (VII) Can it inform local risk management?

influence details of the rainfall generation process this can easily dominate over effects via interception, transpiration and soil changes. Multi-year data will generally be needed to attribute observed changes in flow buffering to degradation/restoration of watersheds, rather than specific rainfall events. With current methods, it seems that effects of land cover change on flow persistence that shift the F_p value by about 0.1 are the limit of what can be asserted from empirical data, with shifts of that order in a single year a warning sign rather than a firmly established change. When derived from observed river flow data, F_p is suitable for monitoring change (degradation, restoration) and can be a serious candi-

date for monitoring performance in outcome-based ecosystem service management contracts. Watershed health is here characterised through the flow pattern it generates, leaving the attribution to land cover, rainfall pattern and engineering of that pattern and of changes in pattern to further location-specific analysis, in the same way a symptom of a high body temperature can indicate health, but not diagnose the specific illness causing it.

The data sets analysed so far did not indicate that the flow persistence at high flows differed from that at lower flows within the same season, but in other circumstances this may not be the case and further care may be needed to use F_p val-

ues beyond the measurement period in which they were derived. While a major strength of the F_p method over existing procedures for parameterising curve number estimates, for example, is that the latter depends on scarce observations during extreme events and F_p can be estimated for any part of the flow record, the reliability of F_p estimates will still increase with the length of the observation period.

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

Appendix A: Data availability

Table A1 specifies the rainfall and river flow data we used for
the four basins and specifies the links to detailed descriptions.

Table A1. Data availability.

	Bialo	Cidanau	Mae Chaem	Way Besai
Rainfall data	1989–2009, source: BWS Sulawesi[a] and PUSAIR[b]; average rainfall data from the stations Moti, Bulo-bulo, Seka and Onto	1998–2008, source: BMKG[c]	1998–2002, source: WRD55, MTD22, RYP48, GMT13, WRD 52	1976–2007, source: BMKG, PU[d] and PLN[e] (interpolation of 8 rainfall stations using Thiessen polygon)
River flow data	1993–2010, source; BWS Sulawesi and PUSAIR	2000–2009, source: KTI[f]	1954–2003, source: ICHARM[g]	1976–1998, source: PU and PUSAIR
Reference of detailed report	[1]	[2]	[3]	[4]

Note: [a] BWS: Balai Wilayah Sungai (Regional River Agency). [b] PUSAIR: Pusat Litbang Sumber Daya Air (Centre for Research and Development on Water Resources). [c] BMKG: Badan Meteorologi Klimatologi dan Geofisika (Agency on Meteorology, Climatology and Geophysics). [d] PU: Dinas Pekerjaan Unum (Public Work Agency). [e] PLN: Perusahaan Listrik Negara (National Electric Company). [f] KTI: Krakatau Tirta Industri, a private steel company. [g] ICHARM: The International Centre for Water Hazard and Risk Management.

[1] http://old.icraf.org/regions/southeast_asia/publications?do=view_pub_detail&pub_no=PP0343-14;

[2] http://worldAgroforestry.org/regions/southeast_asia/publications?do=view_pub_detail&pub_no=PO0292-13;

[3] http://worldAgroforestry.org/regions/southeast_asia/publications?do=view_pub_detail&pub_no=MN0048-11;

[4] http://worldAgroforestry.org/regions/southeast_asia/publications?do=view_pub_detail&pub_no=MN0048-11.
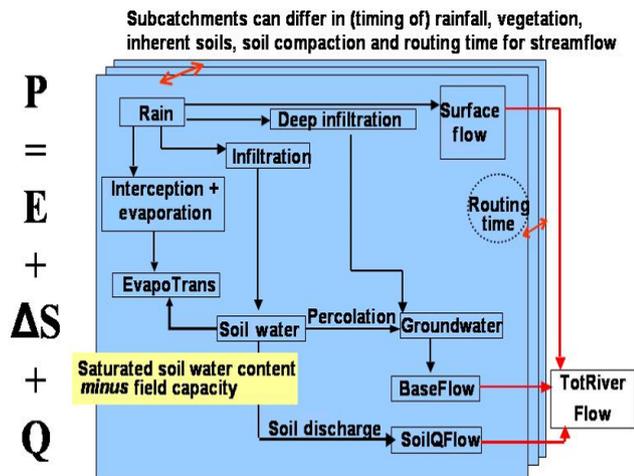
Figure B1. Overview of the GenRiver model.

Appendix B: Genriver model for effects of land cover on river flow

The Generic River flow (GenRiver) model (van Noordwijk et al., 2011) is a simple hydrological model that simulates river flow based on water-balance concept with a daily time step and a flexible spatial subdivision of a watershed that influences the routing of water. The core of the GenRiver model is a "patch" level representation of a daily water balance, driven by local rainfall and modified by the land cover and land cover change and soil properties. The model starts accounting of rainfall or precipitation (P) and traces the subsequent flows and storage in the landscape that can lead to either evapotranspiration (E), river flow (Q) or change in storage (ΔS) (Fig. B1):

$$P = Q + E + \Delta S. \tag{B1}$$

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

At patch level, vegetation influences interception, retention for subsequent evaporation and delayed transfer to the soil surface, as well as the seasonal demand for water. Vegetation (land cover) also influences soil porosity and infiltration, modifying the inherent soil properties. Groundwater pool dynamics are represented at subcatchment rather than patch level, integrating over the land cover fractions within a subcatchment. The output of the model is river flow, which is aggregated from three types of streamflow: surface flow on the day of the rainfall event, interflow on the next day and base flow gradually declining over a period of time. The multiple subcatchments that make up the catchment as a whole can differ in basic soil properties, land cover fractions that affect interception, soil structure (infiltration rate) and seasonal pattern of water use by the vegetation. The subcatchment will also typically differ in "routing time" or in the time it takes the streams and river to reach any specified observation point (with default focus on the outflow from the catchment). The model itself (currently implemented in Stella plus Excel), a manual and application case studies are freely available (http://www.worldAgroforestry. org/output/genriver-generic-river-model-river-flow; van Noordwijk et al., 2011.

Appendix C: Watershed-specific consequences of the land use change scenarios

The generically defined land use change scenarios (Table 4) led to different land cover proportions, depending on the default land cover data for each watershed, as shown in Fig. C1.
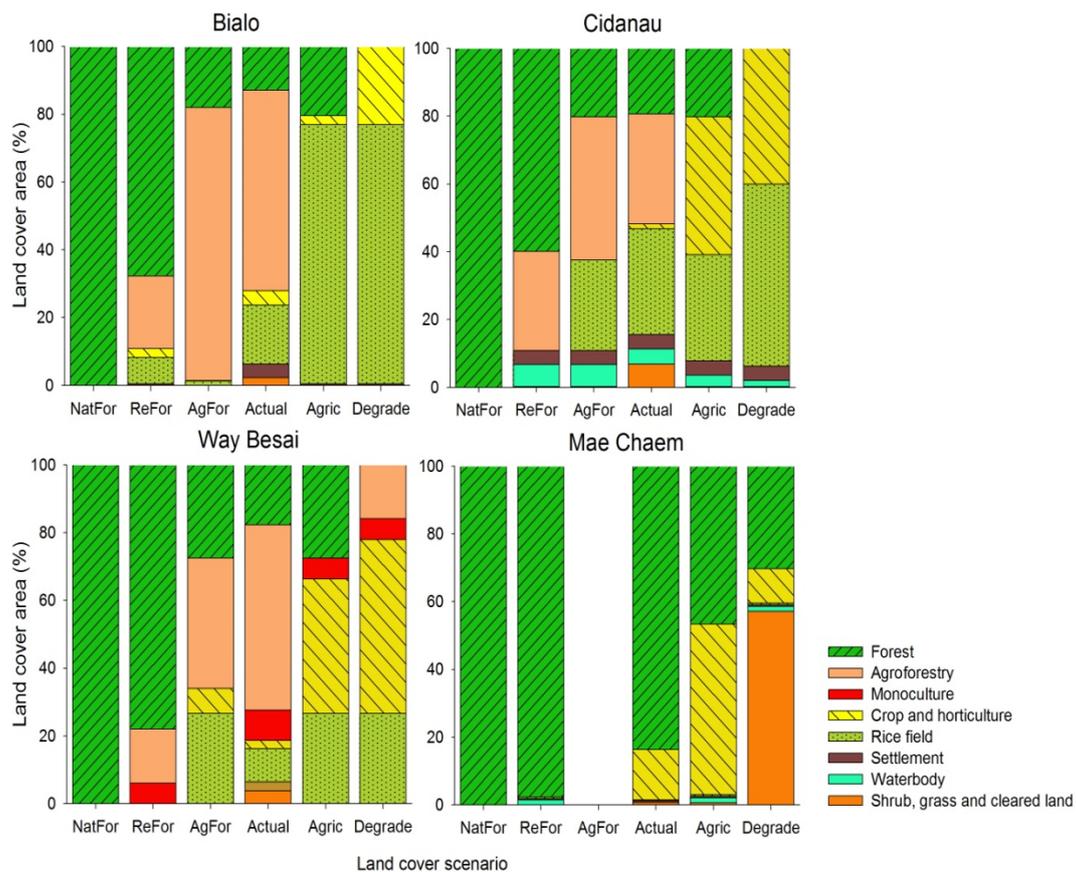
Figure C1. Land use distribution of the various land use scenarios explored for the four watersheds (see Table 4).

Appendix D: Example of a macro in R to estimate number of observation required using bootstrap approach

```
#The bootstrap procedure is to calculate the minimum
sample size (number of observation) required
#for a significant land use effect on Fp
#bialo1 is a data set contains delta Fp values for two different
from Bialo watershed

#read data
bialo1 <- read.table("bialo1.csv", header=TRUE, sep=",")
#name each parameter
BL1 <- bialo1$ReFor
BL5 <- bialo1$Degrade

N = 1000 #number replication

n <- c(5:50) #the various sample size

J <- 46 #the number of sample size being tested (~
number of actual year observed in the data set)

P15= matrix(ncol=J, nrow=R) #variable for storing p-
value
P15Q3 <- numeric(J) #for storing p-Value at 97.5 quantile

for (j in 1:J) #estimating for different n
```

```
#bootstrap sampling
{
for (i in 1:N)
{
#sampling data
S1=sample(BL1, n[j], replace = T)
S5=sample(BL5, n[j], replace = T)

#Kolmogorov–Smirnov test for equal distribution and
get the p value
KS15 <- ks.test(S1, S5, alt = c("two.sided"), exact = F)
P15[i,j] <- KS15$p.value
}

#Confidence interval of CI
P15Q3[j] <- quantile(P15[,j], 0.975)
}

#saving P value data and CI

write.table(P15, file = "pValue15.txt") write.table(P15Q3,
file = "P15Q3.txt")v
```

References

Andréassian, V.: Waters and forests: from historical controversy to scientific debate, J. Hydrol., 291, 1–27, 2004.

Baker, D. B., Richards, R. P., Loftus, T. T., and Kramer, J. W.: A newflashiness index: Characteristics and applications to midwestern rivers and streams, J. Am. Water Resour. Assoc., 40, 503–522, 2004.

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

Dairaku, K., Emori, S., and Taikan, T.: Rainfall Amount, Intensity, Duration, and Frequency Relationships in the Mae Chaem Watershed in Southeast Asia, J. Hydrometeorol., 5, 458–470, 2004.

Efron, B and Tibshirani, R.: Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy, Stat. Sci., 1, 54–75, 1986.

Ellison, D., Morris, C. E., Locatelli, B., Sheil, D., Cohen, J., Murdiyarso, D., Gutierrez, V., van Noordwijk, M., Creed, I. F., Pokorny, J., Gaveau, D., Spracklen, D., Tobella, A. B., Ilstedt, U., Teuling, R., Gebrehiwot, S. G., Sands, D. C., Muys, B., Verbist, B., Springgay, E., Sugandi, Y., and Sullivan, C. A.: Trees, forests and water: cool insights for a hot world, Global Environ. Change, 43, 51–61, 2017.

Gassman, P. W., Reyes, M. R., Green, C. H., and Arnold, J. G.: The soil and water assessment tool: historical development, applications, and future research directions, T. ASABE, 50, 1211–1250, 2007.

Joshi, L., Schalenbourg, W., Johansson, L., Khasanah, N., Stefanus, E., Fagerström, M. H., and van Noordwijk, M.,: Soil and water movement: combining local ecological knowledge with that of modellers when scaling up from plot to landscape level, in: Belowground Interactions in Tropical Agroecosystems, edited by:

van Noordwijk, M., Cadisch, G., and Ong, C. K., CAB International, Wallingford, UK, 349–364, 2004.

Kobold, M., Suselj, K., Polajnar, J., and Pogacnik, N.: Calibration Techniques Used For HBV Hydrological Model In Savinja Catchment, in: XXIVth Conference of the Danubian Countries on the Hydrological Forecasting and Hydrological Bases of Water Managemet, 2–4 June 2008, Slovenia, 2008.

Kusumastuti, D. I., Jokowinarno, D., van Rafi'i, C. H., and Yuniarti, F.: Analysis of rainfall characteristics for flood estimation in Way Awi watershed, Civ. Eng. Dimens., 18, 31–37, 2016

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

Ma, X., Lu, X. X., van Noordwijk, M., Li, J. T., and Xu, J. C.: Attribution of climate change, vegetation restoration, and engineering measures to the reduction of suspended sediment in the Kejie catchment, southwest China, Hydrol. Earth Syst. Sci., 18, 1979–1994, doi:10.5194/hess-18-1979-2014, 2014.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model Evaluation Guidelines For Systematic Quantification Of Accuracy In Watershed Simulations, Am. Soc. Agr. Biol. Eng., 20, 885–900, 2007.

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

Ponce, V. M. and Hawkins, R. H.: Runoff curve number: Has it reached maturity?, J. Hydrol. Eng., 1, 11–19, 1996

R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/, last access: 27 April 2017.

Tan-Soo, J. S., Adnan, N., Ahmad, I., Pattanayak, S. K., and Vincent, J. R.: Econometric Evidence on Forest Ecosystem Services: Deforestation and Flooding in Malaysia, Environ. Resour. Econ., 63, 25–44, 2016.

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

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

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

van Noordwijk, M., Leimona, B., Xing, M., Tanika, L., Namirembe, S., and Suprayogo, D.: Water-focused landscape management, in: Climate-Smart Landscapes: Multifunctionality In Practice, edited by: Minang, P. A., van Noordwijk, M., Freeman, O. E., Mbow, C., Leeuw, J. D., Catacutan, D., World Agroforestry Centre (ICRAF), Nairobi, Kenya, 179–192, 2015a.

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

van Noordwijk, M., Kim, Y.-S., Leimona, B., Hairiah, K., Fisher, L. A.: Metrics of water security, adaptive capacity and Agroforestry in Indonesia, Curr. Opin. Environ. Sustain., doi:10.1016/j.cosust.2016.10.004, in press, 2016.

van Noordwijk, M., Tanika, L., and Lusiana, B.: Flood risk reduction and flow buffering as ecosystem services – Part 1: Theory on flow persistence, flashiness and base flow, Hydrol. Earth Syst. Sci., 21, 2321–2340, doi:10.5194/hess-21-2321-2017, 2017.

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

Zhang, Q., Liu, C., Xu, C., Xu, and Jiang, T.: Observed trends of annual maximum water level and streamflow during past 130 years in the Yangtze River basin, China, J. Hydrol., 324, 255–265, 2006.