Originally published as:

Hydrology and Earth System Sciences

HESS Opinions
"More efforts and scientific rigour are needed to attribute trends in flood time series"

B. Merz, S. Vorogushyn, S. Uhlemann, J. Delgado, and Y. Hundecha

GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany

Correspondence to: B. Merz (bmerz@gfz-potsdam.de)

Abstract. The question whether the magnitude and frequency of floods have changed due to climate change or other drivers of change is of high interest. The number of flood trend studies is rapidly rising. When changes are detected, many studies link the identified change to the underlying causes, i.e. they attribute the changes in flood behaviour to certain drivers of change. We propose a hypothesis testing framework for trend attribution which consists of essential ingredients for a sound attribution: evidence of consistency, evidence of inconsistency, and provision of confidence statement. Further, we evaluate the current state-of-the-art of flood trend attribution. We assess how selected recent studies approach the attribution problem, and to which extent their attribution statements seem defendable. In our opinion, the current state of flood trend attribution is poor. Attribution statements are mostly based on qualitative reasoning or even speculation. Typically, the focus of flood trend studies is the detection of change, i.e. the statistical analysis of time series, and attribution is regarded as an appendix: (1) flood time series are analysed by means of trend tests, (2) if a significant change is detected, a hypothesis on the cause of change is given, and (3) explanations or published studies are sought which support the hypothesis. We believe that we need a change in perspective and more scientific rigour: detection should be seen as an integral part of the more challenging attribution problem, and detection and attribution should be placed in a sound hypothesis testing framework.

1 Detection and attribution of changes in flood hazard trends

Flood trend studies have become a topic of high interest. Recently, many studies investigating trends in flood time series have been published (e.g. Douglas et al., 2000; Zhang et al., 2001; McCabe and Wolock, 2002; Milly et al., 2002; Robson et al., 1998; Mudelsee et al., 2003, 2004, 2006; Cunderlik and Burn, 2004; Lindström and Bergström, 2004; Kundzewicz et al., 2005; Pinter et al., 2006; Svensson et al., 2006; Novotny and Stefan, 2007; Hamlet and Lettenmaier, 2007; Cunderlik and Ouarda, 2009; Petrow and Merz, 2009; Petrow et al., 2009; Villarini et al., 2009, 2011; Delgado et al., 2010; Bormann et al., 2011). Many of these papers identify changes in flood hazard, discuss potential causes for the identified changes and attempt to attribute them, i.e. link detected changes to certain drivers.

We understand flood hazard as a chance phenomenon capable of causing inundation. Fluvial flood hazard is usually characterized by the probability and intensity of high river flows, for example in terms of flood frequency, flood magnitude or flood quantiles. The impact of high flows and inundation on elements at risk, i.e. aspects of vulnerability and damage, are not considered here. Furthermore, changes in time series can be of different types; for example, short-term versus long-term change, gradual versus abrupt change, periodic versus episodic change. For simplicity we use the terms change and trend synonymously, i.e. a trend is not necessarily a gradual change.

In this opinion paper we focus on attribution of flood hazard trends. We look at studies that analyse observational time series of flood indicators, and – in case of detecting change –

relate this change to potential drivers. The question to which extent the probability of occurrence of a single past flood event has been influenced by a certain agent of change is not in the focus of our study. However, since there are methodological similarities, the most recent developments in attribution of changes in hazard of a single natural event to external drivers (Stott et al., 2004; Kay et al., 2011; Pall et al., 2011) will be touched. Our restriction to observed time series also implies that we do not discuss the approaches that attempt to estimate future changes in flood behaviour, for example as consequence of climate change, or modelling studies that investigate the consequences of certain (hypothetical or real) agents of change. An example for the latter would be a study which analyses the change in the 100-yr flood for different scenarios of urbanisation in the catchment.

The terms detection and attribution are used in this paper as they are given by IPCC (2007): "detection" is demonstrating that a change has been observed that is significantly different (in a statistical sense) from what can be explained by natural internal variability. An observed change is said to be detected if its likelihood of occurrence by chance due to natural variability alone is small. Hence, detection is primarily a statistical argument, without explaining the causes for change. "Attribution" is the process of establishing the most likely causes for the detected change with some defined level of confidence. Attribution is understood to demonstrate that the detected change is consistent with the responses of the system to the given drivers, and not consistent with alternative, physically plausible explanations.

When aiming at detection and attribution of changes in flood hazard, we are faced with several problems:

– Flood time series often show high natural variability and a low signal-to-noise ratio, given the usually available series of observations.

– Flood time series may show complex behaviour: they may vary at a range of time scales and floods may cluster in time; different moments of the time series may behave differently, e.g. decrease in the mean flood and, at the same time, an increase in variability (Delgado et al., 2010).

– Different drivers frequently act in parallel in a catchment, and there may be interactions between them. Table 1 gives examples of drivers of change and associated variables. Changes in flood behaviour are the integral response of the catchment to these different drivers and to their interactions.

– Some driver-effect mechanisms are well understood and quantified. However, the knowledge about many other driver-effect mechanisms is still limited (Blöschl et al., 2007).

There are two basically different methods for quantitatively relating detected changes in flood hazard to an assumed driver; for simplicity, we term them data-based and simulation-based attribution. The data-based approach compares flood time series or their statistics with those of the assumed driver, for example, by evaluating the correlation between the time series of the potential cause and effect variables. This approach has frequently been used when the link between flood trends and climate drivers has been investigated. It assumes, however, that other conditions, such as land use, have remained stationary or have been of minor importance over the period of investigation. An example is the study of Cunderlik and Burn (2004) who compare the similarity of the regional trend in monthly maximum flows of southern British Columbia and that of climate variables. To measure the plausibility of the link between the flow and climatic data, they perform cross-correlation analysis on residuals of the original series after subtracting all serially dependent components. Another example is given in Pinter et al. (2006) who calculate correlation coefficients between time series of flood peaks and cumulative basin precipitation for different intervals prior to the peak (1–30 days) for the Rhine basin at the gauge Cologne.

The simulation-based attribution approach uses simulation models to identify the causal link between observed trend and the assumed driver. The observed change in flood behaviour is compared to the change simulated with a hydrological model which considers in its forcing and/or parameterisation the assumed drivers of change (such as land use change or climate change). One of the advantages of this approach is that one is able to identify the relative importance of the different drivers if suitable models for these drivers exist. This advantage comes, however, at the expense of the uncertainty which is involved in the simulations. To our knowledge, there are no published studies that try to attribute observed trends in flood hazard to the possible drivers via the simulation based approach. The study of Hundecha and Merz (2012) seems to be the first step into simulation-based flood trend attribution. The authors drive a hydrological model with a large number of realisations of stationary and non-stationary meteorological time series, respectively, to support or falsify that observed flood trends are climate-driven. Similarly, Hamlet and Lettenmaier (2007) investigate changes in flood quantiles in western US, resulting from hydrologic model simulations driven by two synthetic perturbed spatial temperature patterns: one de-trended, and another one corresponding to the present temperature conditions. Hamlet et al. (2007) investigate among others the trends in simulated runoff timing and intra-annual redistribution of runoff volume and try to attribute them to changes in temperature and precipitation by fixing drivers one-at-a-time to the monthly climatological value. Although the latter two studies are conceptually very close to the study of Hundecha and Merz (2012) and hence very valuable in order to understand climate-related flood changes, they do not fit into our definition of flood trend attribution studies, since they do

Table 1. Examples of drivers of change in flood hazard and associated variables.

Compartment	Processes	Variables	Drivers of change
Atmosphere	Meteorological forcing of catchment water fluxes.	Temperature, total precipitation, precipitation intensity, snow cover, snowmelt, seasonal distribution of climatic variables, seasonality of floods.	Natural climate variability at different time scales, anthropogenic climate change.
Catchments	Runoff generation and concentration.	Infiltration capacity, runoff coefficient, water storage capacity.	Urbanization, deforestation, wildfires, agricultural management practices, drainage of wetlands and agricultural areas, construction of flood retention basins.
Rivers	Flood wave propagation, superposition of flood waves.	River morphology, conveyance, roughness, water level, discharge, inundated area.	River training, reduction in river length, construction of dikes, groynes and weirs, operation of hydropower plants and reservoirs.

not compare the simulated change in flood behaviour with the observed change.

In this opinion paper we appraise the state-of-the-art of flood trend attribution. In chapter 2 we discuss the necessary ingredients of flood attribution studies: what is necessary to attribute flood trends? Then we evaluate selected recent studies on flood trend attribution and assess to which extent they consider these ingredients, and to which extent their attribution conclusions are defendable (chapter 3). We finally present our thoughts towards the improvement of the scientific practice in flood hazard attribution (chapter 4). We would like to note that our focus in this paper is the attribution problem in flood hazard studies. This focus does not imply that the closely related topic of change detection is of less importance (see e.g. Merz et al., 2012), or that detection studies without attribution are meaningless. Robust detection, based on a number of different statistical tests and on reliable and long observational time series, may provide useful information even without attribution. But in our opinion, the attribution problem has not received the necessary attention.

2 Hypothesis testing framework for attribution

Reviewing the definition of attribution, we find three ingredients of attribution: evidence of consistency, evidence of inconsistency, and provision of confidence level.

Evidence of consistency is showing that the detected change in flood characteristics is consistent with the assumed drivers of change. For example, if we detect a change in flood magnitude in a given catchment and hypothesize that a climate signal is responsible for this change, then we need to show that the effect of this climate signal translates into the observed change in flood hazard. Hence, the strength of our evidence of consistency is based on our ability to demonstrate the relationship between cause (e.g. climate signal) and effect (e.g. change in flood behaviour).

Evidence of inconsistency is showing that the detected change in flood characteristics is inconsistent with changes due to alternative possible drivers. If more than one driver of change is acting in a catchment, and since we observe only the integral response of the catchment to all the acting drivers, attribution requires evidence that the observed change has not been caused by alternative drivers. Evidence of inconsistency is necessary to avoid confirmation bias, i.e. the tendency to favour information which confirms existing preconceptions or hypotheses (e.g. Nickerson, 1998). Confirmation bias results, among others, from biased search, i.e. searching for information consistent with one's hypothesis. Another source of confirmation bias is biased interpretation, i.e. hypotheses that do not meet one's expectations are confronted with higher standards of evidence. Biased search in flood hazard trend studies would mean, for instance, that we preconceive a climate signal to be responsible for the detected change in flood behaviour, and that we are looking only for evidences that confirm our expectation, such as changes in flood-related meteorological variables in the catchment during the study period.

The requirement for a *provision of confidence* level associated with our attribution statement results from finite observational data, limitations in our knowledge of the system as well as from the uncertainties of our modelling tools used in the attribution. A confidence or uncertainty level acknowledges these limitations and attempts to quantify the strength of our attribution statement in terms of a likelihood statement, expressing how likely it is, given the available data, that a certain driver or set of drivers caused the observed change in flood characteristics.

In their discussion of climate change impacts, Blöschl and Montanari (2009) propose to distinguish between hard and soft facts. They consider future changes in air temperature as a hard fact, whereas future changes in mean precipitation at continental scale are considered as soft facts, and changes in extreme precipitation as speculation. Similarly, we propose to distinguish between hard and soft attribution. Attribution studies that show both consistency and inconsistency

with a decent amount of reliability qualify as hard attribution, whereas all other studies may be seen as soft attribution.

3   How is attribution approached in flood trend studies?

We review recent flood trend studies published in the scientific literature. We focus on those attempting to interpret the detected changes in terms of the effects of possible drivers. Since our focus is trend attribution, we do not consider studies which are limited to trend detection. To be included in our review, a paper must start from a detected change in flood behaviour, and must include an attribution statement, i.e. it must state that the observed flood trend is caused by certain drivers. Given this perspective, the number of papers decreases rapidly, although there is a wealth of papers on changes in flood behaviour. Frequently, no clear trends are detected, or there is no attribution statement associated with the detected changes.

Table S1 (in the Supplement) summarises published and peer-reviewed studies which detect changes in flood hazard and include attribution statements. In order to evaluate the state-of-the-art in flood trend attribution, we limited ourselves to ten studies. Although this selection contains a subjective element, we attempted to select publications which present the spectrum of attribution approaches that are in use in the hydrological community and which are, therefore, typical. The table contains in a very condensed form the essential characteristics of each study (investigated flood behaviour, data and methods used, main results and final attribution statements). In particular, it contains an assessment of the ingredients of attribution, i.e. we appraise if and how the studies attempt to provide evidence of consistency, of inconsistency and assess the reliability of their attribution statement.

The studies considered in our review approach the attribution of detected flood trends mainly through an attempt to show consistency. The attempts range from referring to literature that provides certain indications of expected changes (e.g. Bormann et al., 2011; Villarini et al., 2011) to own consistency analyses (e.g. Cunderlik and Burn, 2004; Pinter et al., 2006). Classifying the attempts to show consistency into quantitative and qualitative reasoning, most of the papers restrict themselves to qualitative reasoning, i.e. the authors explain the change in flood behaviour by some change in another variable (precipitation, circulation patterns, agricultural practises etc.) without quantitatively linking changes in the assumed driver to the change in flood behaviour. Evidence of inconsistency was in several cases based on choosing undisturbed catchments (e.g. Cunderlik and Burn, 2004; Hannaford and Marsch, 2008). Only Mudelsee et al. (2003) explicitly checked the inconsistency of trends in flood hazard with a local anthropogenic forcing, namely the increase of reservoir volume over time. The specific gauge analysis (SGA) carried out by, for instance, Pinter et al. (2006) and

Bormann et al. (2011) attempt to prove the inconsistency of flood hazard changes with river training measures undertaken at the point of analysis. It, however, does not allow a general statement about the inconsistency with anthropogenic river changes which may have been undertaken upstream of an investigated gauge.

Among the attempts to support the attribution statements made, there is surprisingly little effort to reliably quantify the assumed cause-effect relationship, and there is even less effort to falsify alternative candidate drivers. A very popular approach is to cite literature which is interpreted as support for the own claim. Those cases where we see the strongest link between drivers and change in flood behaviour are based on correlation analysis between driver time series and flood time series. However, even in these cases there are questions concerning the validity of the attribution. For example, often precipitation is correlated with flood variables. Unfortunately, the relationship between precipitation and flood magnitude or frequency is not well-defined. Precipitation is spatially and temporally highly heterogeneous, and it is not obvious if and how a certain precipitation variable, integrated over a certain time period and a certain spatial extent, is related to flood magnitude or frequency at a given gauge.

Recalling our proposed framework for flood trend attribution, we find that all papers belong to the group of soft attribution; there is not a single paper which gives convincing evidence of consistency, convincing evidence of inconsistency, and which provides a measure of confidence. In some cases, attribution statements are made which are not even further discussed; or the attribution is supported by reference to literature but when we looked into this literature we could not find this support. A further observation is that the studies tend to attribute the detected flood trends to a single driver or a very limited selection of drivers (e.g. climate and river training) disregarding the remaining drivers or assuming their effect to be negligible.

In our opinion, attribution of flood trends has not received much attention. Although authors frequently give the cause of a detected change, they do not spend much effort on supporting their claim. Among the ten papers, there is not a single one which considers all three ingredients of attribution. To our knowledge, this statement is also valid for other studies that are not contained in our overview. We have the impression that flood trend attribution is not handled according to its complexity and relevance. The lion's share of the work is devoted to detection by statistically analysing flood time series. If changes are detected, a hypothesis on the cause of change is given, and explanations or published studies are sought which support the hypothesis. Hence, attribution is treated as an appendix to detection studies.

4 Some thoughts on ways forward

4.1 Single-driver versus multiple-driver approach

Table 1 illustrates that there are many potential drivers of change affecting river systems and catchments. Although we suspect that several drivers act in parallel in most catchments, frequently flood trend attribution statements in the hydrological literature are limited to a single driver. This single-driver approach may be adequate in case of one driver dominating change in flood behaviour. However, the question "Is a single-driver explanation adequate for the given catchment, the time period studied and the investigated flood characteristics?" is rarely discussed.

We propose that any attribution study should give account of the drivers that may have impacted the flood behaviour. At least, an attempt should be made to list the drivers that may have played a role. When there are several candidates for drivers of change, a further question is: is it possible to distinguish between their impacts on flooding? In some cases a differentiation may be possible, since drivers of change act on different time, space and severity scales, and they have different consequences for different flood characteristics (Merz et al., 2012). For example, dam construction and implementation of river training measures have rather short time scales of a few years, hence, their effect on flood characteristics should be visible as step change or short-term change. Most land use changes develop with time scales of decades and centuries. They have a long-term effect on flood hazard and the change in hazard might develop slowly.

Many drivers of change are also associated with spatial scales. For instance, flood retention basins have a certain area of influence. Their effect on the flood hydrograph is greatest immediately downstream of the basin and, depending on the characteristics of the retention basin and of the flood, it may not be seen much further downstream. Other drivers of change, such as climate change and climate variability, do not have such an obvious relation to spatial scales. It is even hypothesized that these drivers may be insensitive to the spatial scale (Blöschl et al., 2007). However, an indirect link may be given by the changing dominance of flood types at different spatial scales. High-intensity, short-duration rainfall events frequently dominate the flood behaviour in small catchments, whereas other processes such as long-lasting synoptic rainfalls or snow accumulation and snow melt are of major importance in many large basins. Since climate change may affect different flood process types differently, also climate change and climate variability may be related to spatial scales.

Finally, drivers of change may be associated with flood severity scales. Some drivers of change may influence only small floods, whereas others may affect large floods. For example, an important effect of urbanization may be an increase in the runoff coefficient. This may have significant influence on flood peaks for smaller floods. For very large floods the increase in the runoff coefficient may be considerably less pronounced and even practically negligible, since it may already be high, for instance due to antecedent rainfall which saturated the soils. Similarly, the influence of flood retention basins may become negligible for floods which significantly exceed the design flood. If the candidate drivers in a given catchment act on very different temporal, spatial or severity scales, this behaviour may help to support or falsify certain hypotheses on causes of change.

Multiple-driver attribution requires much more effort and data than single-driver attribution. A particular problem is that in many catchments information on past interventions is not easily available. However, this additional effort should not stop us from trying to assess the effects of different drivers on observed changes in flood behaviour. Savenije (2009) gives a number of examples for cases where hydrologists have limited their analysis to certain domains and have failed to see important interactions and feedbacks: a too limited view on hydrology led to apparent anomalies that could only be explained by looking beyond the limits of sub-disciplines. Similarly, we argue that we should not prematurely narrow our view by limiting ourselves to the candidate driver which we perceive as dominant cause for change.

4.2 A look over the rim of a tea cup: fingerprinting and fraction of attributable risk

Unlike flood trend attribution, the problem of attributing observed climate change signals to potential drivers has been approached in a more systematic way over the past three decades. The state-of-the-art method for attribution used in the climate research community is optimal fingerprinting (Hasselmann, 1997, 1993; Allen and Stott, 2003). The approach is based on multivariate regression in which a field of an observed climate indicator is represented as a linear combination of signal patterns (fingerprints) that are simulated by a climate model under external forcings plus a noise field, which represents a realisation of the internal climate variability. The scaling factors that are used to weigh the signal patterns are meant to adjust the amplitudes of the signal patterns to achieve best matching with that of the observation. The approach is "optimal" in that it allows estimation of the scaling factors that maximise the signal-to-noise ratio, thereby increasing detectability of the signal due to a forced climate change. It has been applied in the detection of changes in climate variables and attributing them to external forcing (Hegerl et al., 1996; Min et al., 2011). Detection of the external forcing signal in the observations is achieved by testing the null hypothesis that the scaling factor corresponding to a given forcing is zero. If it is significantly different from zero, a signal is detected. Attribution of the detected change to the given forcing is then performed using an attribution consistency test, which involves testing the hypothesis that its scaling factor is unity. This basically means that no scaling of the amplitude of the response signal

is needed so that it best matches the pattern of change contained in the observation field. The climate indicator used in the analysis can be a field that represents a spatial pattern or may additionally contain a time dimension. Estimates of the natural internal variability are obtained from a long control simulation. The approach can be implemented using a single response pattern (Hegerl et al., 1996), where detection of a signal of a single forcing or combination of different forcings in the observation is performed separately, or multiple response patterns (Hasselmann, 1997), which allow simultaneous dectection of signals of different forcings and quantification of the relative contributions of the different forcings to the changes in the observations. The implementation using multiple response patterns has been applied to distinguish the effect of greenhouse gas, greenhouse gas-plus-aerosols and solar forced climate change (Hegerl et al., 1997).

Provided that one can identify the most important drivers of changes in extreme flows and estimate them with reasonable confidence levels, the optimum fingerprinting approach offers an attractive analytical framework for attributing changes in extreme flows to these drivers. The approach would even enable us to identify the relative importance of each of the drivers in the detected changes. In a work, which is conceptually loosely related to this approach, Hundecha and Merz (2012) attempted to detect change signals in model simulated extreme flows driven by non-stationary meteorological drivers that are similar to the changes detected in the observations. The work was limited to detecting the change signal due either to precipitation or temperature changes and identifying the relative importance of each of the two variables in the change pattern contained in the observations. Using the optimal fingerprint approach the work can potentially be extended to include other drivers such as changes in land use and implementation of river training works.

One of the premises of the optimal fingerprinting approach is that the variable whose change is to be detected and attributed should be estimated with high signal-to-noise ratio and that it should be consistently estimated using different models under a given driver. This necessitates carefully selecting a measure of extreme flow that is not unduly sensitive to both the uncertainty in the drivers and parameterization of the hydrological model implemented in the analysis. The effect of uncertainty of drivers can be assessed through using ensembles of drivers in the estimation of the response signal, as has been done in Hundecha and Merz (2012). Similarly, the sensitivity of the response to the hydrological modelling can be assessed through sensitivity or uncertainty analysis of the hydrological model.

A special metric, which can be derived from the optimal fingerprinting methodology by spatial averaging of the trend patterns and estimating their distribution, is the fraction of attributable risk (FAR), originally developed in the epidemiological sciences and recently established in the climate research community (Allen, 2003; Stone and Allen, 2005; Pall et al., 2011). FAR is defined as $\mathrm{FAR} = 1 - P0/P1$, where $P0$

and $P1$ are the probabilities of a variable in a reference system state and in a perturbed state given external forcing, respectively. FAR represents the portion of risk (in our notation: hazard) that cannot be explained by the probability of occurrence in the reference system state (Stone and Allen, 2005). Pall et al. (2011) use this idea in order to assess the anthropogenic greenhouse gas contribution to the flood hazard in England and Wales in autumn 2000. They generate several thousand climate model simulations of autumn 2000 weather, both under realistic conditions and under conditions had these greenhouse gas emissions and the resulting warming not occurred. The results are fed into a hydrological model to simulate daily river runoff events. Finally, $P0$ and $P1$ are estimated as the fraction of runoff events that exceed a certain flood threshold under realistic weather and under no-warming weather, respectively. $P0$ and $P1$ need not necessarily be viewed as crisp probabilities but can be treated as probability distributions representing sampling or modelling uncertainties. Like optimal fingerprinting, FAR is a consistent quantitative approach to the attribution problem, and therefore, it can potentially be applied to the problem of attributing changes in the hydrological system, including changes in flood hazard due to changes in climate, land use, river training or other drivers.

4.3   The role of uncertainty in flood hazard attribution

Our analysis of the state-of-the-art of flood trend attribution has shown that the attribution statements come without a statement on the confidence of the attribution. The importance of providing uncertainty statements in hydrology has often been discussed. For instance, Beven (2008) proposes to referees, editors and hydrological journals that they insist on the presentation of uncertainty estimates in modelling papers. In the context of flood trend attribution we would like to highlight in particular the problem of natural climate variability which usually translates into high natural variability of flood time series. The observed flood peak discharges have to be seen as a single realisation of a stochastic process, and since we are interested in flood flows, the available data base to derive the statistical characteristics of this process is sparse, and hence, the sampling uncertainty is large. The generation of many synthetic realisations of flood time series is a way to assess the effect of natural variability on sampling uncertainty. In the context of flood attribution this approach has been used by Hundecha and Merz (2012) by coupling a weather generator to a time-continuous hydrological model. In climate change attribution studies, natural climate variability is usually estimated from long control simulations and/or several climate model realisations with perturbed initial conditions. By using their output as forcing to a hydrological model, natural climate variability can be transformed to natural variability in flood time series.

There are many other sources of uncertainty in flood trend attribution studies. In this respect flood trend attribution is

similarly plagued with uncertainty as other quantitative assessments in hydrology. Besides the sampling uncertainty related to natural variability, the simulation-based approach to the attribution problem is inherently confronted with uncertainties in model formulation, parameterisation and input data. Consideration of these uncertainties is essential and desirable, if data availability and computational resources allow, in order to reliably discern the effects of drivers. For example, Monte-Carlo based assessments of parameter, initial and boundary condition uncertainties are an already well-established approach to support the reliability assessment of flood attribution studies.

4.4 Hypothesis testing framework and scientific rigour

The current mainstream of flood trend studies focuses on detection. The far more difficult problem of trend attribution is addressed, if at all, rather sloppily. We see the goal of trend studies not just in the detection of changes in observed time series and in the discussion of possible causes. We suggest to look at trend detection from another perspective, namely to regard it as a tool for testing hypotheses about the influences of drivers on changes in flood characteristics. The emphasis on attribution studies in a hypothesis testing perspective would force us to more thoroughly discuss candidate drivers, their possible effects on flood characteristics and the time periods we expect them to manifest themselves and become discernible from alternative forcing. In our opinion, flood trend attribution should more closely follow the concept of fallibilism which is based on the idea that final certainty is unattainable. Along this line, Karl Popper's Critical Rationalism (Popper, 1998) stresses the importance of falsification: a hypothesis is tested by performing carefully designed experiments with the aim of falsifying the hypothesis or finding weaknesses. This perspective would help to better understand the link between potential drivers and changes in flood behaviour, and would successively lead to new and improved hypotheses. The experiment should be set up utilizing the available process knowledge in order to encompass relevant mechanisms which relate the drivers to the state variables, e.g. hydraulic models should be appropriately selected in order to reflect the effect of past river training measures on possible changes in flood hazard. Taking into account model structure and parameter uncertainty, uncertainty assessments may be used to understand the robustness of the attribution statements. Additionally, corroborating lines of evidence can be provided to support the attribution hypothesis. To stay with the example of river training, an analysis of changes in arrival time of flood peaks in a main stream and tributaries, i.e. illustration of a superposition of flood waves in main stream and tributaries, may provide additional evidence for the influence of river training onto the flood flows. Similarly, the inclusion of a variety of flood characteristics – beyond flood magnitude – in detection and attribution studies could provide additional evidence. Variables such as flood

seasonality and flashiness of the runoff regime may allow valuable insights into changing flood processes (Parajka et al., 2010; Holko et al., 2011).

Finally, we call for more scientific rigour in flood detection and attribution studies. The quite often used approach to support own detection results and attribution statements by references to related works should be viewed with caution and does not substitute own attempts to show consistency or inconsistency, unless the cited works establish an unambiguous link (related to the responsible variables, consistent time periods etc.) to the changes under study. The mere discussion of possible causes of detected changes is helpful to formulate hypotheses and to identify the candidate drivers and their possible impact mechanisms. However, this should not substitute the attempt to show consistency, show inconsistency with alternative hypotheses, and provide a statement on the reliability of the attribution. One has to be very careful, since such discussions may easily end up in speculation.

Supplementary material related to this article is available online at: http://www.hydrol-earth-syst-sci.net/16/1379/2012/hess-16-1379-2012-supplement.pdf.

References

Allen, M.: Liability for climate change, Nature, 421, 891–892, 2003.

Allen, M. R. and Stott, P. A.: Estimating signal amplitudes in optimal fingerprinting, part I: theory, Clim. Dynam., 21, 477–491, doi:10.1007/s00382-003-0313-9, 2003.

Beven, K.: On doing better hydrological science, Hydrol. Process., 22, 3549–3553, 2008.

Blöschl, G. and Montanari, A.: Climate change impacts-throwing the dice?, Hydrol. Process., 24, 374–381, doi:10.1002/hyp.7574, 2009.

Blöschl, G., Ardoin-Bardin, S., Bonell, M., Dorninger, M., Goodrich, D., Gutknecht, D., Matamoros, D., Merz, B., Shand, P., and Szolgay, J.: At what scales do climate variability and land cover change impact on flooding and low flows? Invited Commentary, Hydrol. Process., 21, 1241–1247, doi:10.1002/hyp.6669, 2007.

Bormann, H., Pinter, N., and Elfert, S.: Hydrological signatures of flood trends in German rivers: flood frequencies flood heights and specific stages, J. Hydrol., 404, 50–66, 2011.

Cunderlik, J. M. and Burn, D. H.: Linkages between Regional Trends in Monthly Maximum Flows and Selected Climatic Variables, J. Hydrol. Eng., 9, 246–256, 2004.

Cunderlik, J. M. and Ouarda, T. B. M. J.: Trends in the timing and magnitude of floods in Canada, J. Hydrol., 375, 471–480, 2009.

Delgado, J. M., Apel, H., and Merz, B.: Flood trends and variability in the Mekong river, Hydrol. Earth Syst. Sci., 14, 407–418, doi:10.5194/hess-14-407-2010, 2010.

Douglas, E. M., Vogel, R. M., and Kroll, C. N.: Trends in floods and low flows in the United States: impact of spatial correlation, J. Hydrol., 240, 90–105, 2000.

Hamlet, A. F. and Lettenmaier, D. P.: Effects of 20th century warming and climate variability on flood risk in the western U.S., Water Resour. Res., 43, W06427, doi:10.1029/2006WR005099, 2007.

Hamlet, A. F., Mote, P. W., Clark, M. P., and Lettenmaier, D. P.: Twentieth-century trends in runoff, evapotranspiration, and soil moisture in the western United States, J. Climate, 20, 1468–1486, doi:10.1175/JCLI4051.1, 2007.

Hannaford, J. and Marsh, T. J.: High-flow and flood trends in a network of undisturbed catchments in the UK, Int. J. Climatol., 28, 1325–1338, doi:10.1002/joc.1643, 2008.

Hasselmann, K.: Optimal fingerprints for the detection of time-dependent climate change, J. Climate, 6, 1957–1971, 1993.

Hasselmann, K.: Multi-pattern fingerprint method for detection and attribution of climate change, Clim. Dynam., 13, 601–611, 1997.

Hegerl, G. C., von Storch, H., Hasselmann, K., Santer, B. D., Cubasch, U., and Jones, P. D.: Detecting greenhouse-gas-induced climate change with an optimal fingerprint method, J. Climate, 9, 2281–2306, 1996.

Hegerl, G. C., Hasselmann, K., Cubasch, U., Mitchell, J. F. B., Roeckner, E., Voss, R., and Waszkewitz, J.: Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plusaerosol and solar forced climate change, Clim. Dynam., 13, 613–634, 1997.

Holko, L., Parajka, J., Kostka, Z., Škoda, P., and Blöschl, G.: Flashiness of mountain streams in Slovakia and Austria, J. Hydrol., 405, 392–401, 2011.

Hundecha, Y. and Merz, B.: Exploring the relationship between changes in climate and floods using a model-based analysis, Water Resour. Res., 48, W04512, doi:10.1029/2011WR010527, 2012.

IPCC, 2007: Summary for Policymakers, in: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Kay, A. L., Crooks, S. M., Pall, P., and Stone, D. A.: Attribution of Autumn/Winter 2000 flood risk in England to anthropogenic climate change: A catchment-based study, J. Hydrol., 406, 97–112, doi:10.1016/j.jhydrol.2011.06.006, 2011.

Kundzewicz, Z. W., Graczyk, D., Maurer, T., Pinskwa, I., Radziejewski, M., Svensson, C., and Szwed, M.: Trend detection in river flow series: 1. Annual maximum flow, Hydrol. Sci. J., 50, 797–810, 2005.

Lindström, G. and Bergström, S.: Runoff trends in Sweden 1807–2002, Hydrol. Sci. J., 49, 69–83, 2004.

McCabe, G. J. and Wolock, D. M.: A step increase in streamflow in the conterminous United States, Geophys. Res. Lett., 29, 38-1–38-4, 2002.

Merz, B., Kundzewicz, Z. W., Delgado, J. M., Hundecha, Y., and Kreibich, H.: Detection and attribution of changes in flood hazard and risk, in: Changes in flood risk in Europe, edited by: Kundzewicz, Z. W., to be published by IAHS, 2012.

Milly, P. C. D., Wetherald, R. T., Dunne, K. A., and Delworth, T. L.: Increasing risk of great floods in a changing climate, Nature 415, 514–517, 2002.

Min, S.-K., Zhang, X., Zwiers, F. W., and Hegerl, G. C.: Human contribution to more-intense precipitation extremes, Nature, 470, 378–381, doi:10.1038/nature09763, 2011.

Mudelsee, M., Börngen, M., Tetzlaff, G., and Grünewald, U.: No upward trends in the occurrence of extreme floods in central Europe, Nature, 421, 166–169, 2003.

Mudelsee, M., Börngen, M., Tetzlaff, G., and Grünewald, U.: Extreme floods in central Europe over the past 500 years: role of cyclone pathway "Zugstrasse Vb", J. Geophys. Res., 109, 1–21, doi:10.1029/2004JD005034, 2004.

Mudelsee, M., Deutsch, M., Borngen, M., and Tetzlaff, G.: Trends in flood risk of the River Werra (Germany) over the past 500 years, Hydrol. Sci. J., 51, 818–833, 2006.

Nickerson, R.: Confirmation Bias; A Ubiquitous Phenomenon in Many Guises, Rev. Gen. Psychol. (Educational Publishing Foundation), 2, 175–220, 1998.

Novotny, E. V. and Stefan, H. G.: Stream flow in Minnesota: Indicator of climate change, J. Hydrol., 334, 319–333, 2007.

Pall, P., Aina, T., Stone, D., Stott, P., Nozawa, T., Hilberts, A. G. J., Lohmann, D., and Allen, M. R.: Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000, Nature, 470, 382–386, 2011.

Parajka, J., Kohnová, S., Bálint, G., Barbuc, M., Borga, M., Claps, P., Cheval, S., Dumitrescu, A., Gaume, E., Hlavčová, K., Merz, R., Pfaundler, M., Stancalie, G., Szolgay, J., and Blöschl, G.: Seasonal characteristics of flood regimes across the Alpine-Carpathian range, J. Hydrol., 394, 78–89, 2010.

Petrow, T. and Merz, B.: Trends in flood magnitude, frequency and seasonality in Germany in the period 1951–2002, J. Hydrol., 317, 129–141, 2009.

Petrow, T., Zimmer, J., and Merz, B.: Changes in the flood hazard in Germany through changing frequency and persistence of circulation patterns, Nat. Hazards Earth Syst. Sci., 9, 1409–1423, doi:10.5194/nhess-9-1409-2009, 2009.

Pinter, N. J. H., van der Ploeg, R. R., Schweigert, P., and Hoefer, G.: Flood magnification on the River Rhine, Hydrol. Process., 20, 147–164, 2006.

Popper, K. R.: Objektive Erkenntnis. Ein evolutionärer Entwurf [Objective Knowledge. An Evolutionary Approach], 414 pp., Hamburg: Hoffmann und Campe, 1998.

Robson, A. J., Jones, T. K., Reed, D. W., and Bayliss, A. C.: A study of national trend and variation in UK floods, Int. J. Climatol., 18, 165–182, 1998.

Savenije, H. H. G.: HESS Opinions "The art of hydrology"*, Hydrol. Earth Syst. Sci., 13, 157–161, doi:10.5194/hess-13-157-2009, 2009.

Stone, D. and Allen, M. R.: The end-to-end attribution problem: from emissions to impacts, Climatic Change, 71, 303–318, doi:10.1007/s10584-005-6778-2, 2005.

Stott, P. A., Stone, D. A., and Allen, M. R.: Human contribution to the European heatwave of 2003, Nature, 432, 610–614, 2004.

Svensson, C., Hannaford, J., Kundzewicz, Z. W., and Marsh, T. J.: Trends in river floods: why is there no clear signal in observations?, in: Frontiers in Flood Research, edited by: Tchiguirin-

skaia, I., Thein, K. N. N., and Hubert, P., IAHS-AISH P., 305, Wallingford, UK, 1–18, 2006.

Villarini, G., Smith, J. A., Serinaldi, F., Bales, J., Bates, P. D., and Krajewski, W. F.: Flood frequency analysis for nonstationary annual peak records in an urban drainage basin, Adv. Water Resour., 32, 1255–1266, 2009.

Villarini, G., Smith, J. A., Serinaldi, F., and Ntelekos, A. A.: Analyses of seasonal and annual maximum daily discharge records for central Europe, J. Hydrol., 399, 299–312, doi:10.1016/j.jhydrol.2011.01.007, 2011.

Zhang, X., Harvey, K. D., Hogg, W. D., and Yuzyk, T. R.: Trends in Canadian streamflow, Water Resour. Res., 37, 987–998, 2001.