On the risk of obtaining misleading results by pooling streamflow data for trend analyses

D. Viviroli,[1,2] B. Schädler,[1,2] P. Schmocker-Fackel,[3] M. Weiler,[4] and J. Seibert[5,6,7]

Received 25 November 2011; revised 4 April 2012; accepted 17 April 2012; published 19 May 2012.

[1] Floods have broad impacts on nature, society, and the economy. The frequency and intensity of flood events are generally believed to increase with the anticipated changes in temperature and precipitation. Trend analyses are important tools to quantify these changes, but often, they provide inconclusive results, partly because of the limited data availability. One way to overcome this limitation is to pool data from different gauging stations. However, pooling data from different stations may lead to misleading results. For example, using pooled flood data Allamano et al. (2009a) found a considerable increase of flooding risks for Switzerland. Here we demonstrate that the previous finding of increased flooding risks was an artifact of the pooling of stations and the fact that the longer time series came from larger catchments, which tend to have lower values for specific peak flows than smaller catchments. Our results demonstrate the risk of obtaining incorrect statistical conclusions when statistical analyses and data selection are not considered with due care.

Citation: Viviroli, D., B. Schädler, P. Schmocker-Fackel, M. Weiler, and J. Seibert (2012), On the risk of obtaining misleading results by pooling streamflow data for trend analyses, *Water Resour. Res.*, *48*, W05601, doi:10.1029/2011WR011690.

1. Introduction

[2] The influence of an anticipated change in temperature and precipitation on the recurrence interval of large floods is a challenging and highly relevant question in hydrology. In view of the high societal and economic impacts a change in flood frequencies would have [Kundzewicz et al., 2007; Organe consultatif sur les changements climatiques, 2007; Stern, 2007], it is of paramount importance to extend our knowledge in this field. To date, there is still a lack of reliable projections for anticipated changes in flood behavior, particularly in mountain areas [Intergovernmental Panel on Climate Change, 2007], and even trend detection for the past may lead to inconclusive results, as illustrated by Kundzewicz et al. [2005] and Svensson et al. [2005], who analyzed a worldwide set of 195 long-term daily streamflow records and did not find a clear pattern of increase or decrease in annual maximum peak flow or in numbers of floods and their magnitudes. Mudelsee et al. [2003] looked at long records of two large European rivers (Elbe and Oder) and could also not detect any trends. Hirsch and Ryberg [2012] studied the relationship between global CO_2 levels and long-term (□100 years) flood data from 200 stream gauges in the United States, which they grouped into four regions. They found that there was no strong statistical evidence in any of the four regions for increasing flood magnitudes with increasing CO_2 levels, but the southwest region showed a statistically significant decrease of flood magnitudes.

[3] One approach to increase the generally limited data basis for detecting trends of extreme values and in particular the observation length for statistical analyses is to pool data from different gauging stations. One kind of pooling approach has recently been used by Allamano et al. [2009a], who applied quantile regression to a pooled set of runoff series from 27 stations in the Swiss Alps and found a significant increase in flood frequencies. In this technical note we address the problem that misleading results may be obtained when using pooled data series in particular when using discharge observation from watersheds with varying catchment areas. One common characteristic of many runoff databases is the tendency that the longest time series are available for larger catchments, whereas measurements in smaller catchments tend to have started later. This can be observed at both the global and national level using the databases of the Global Runoff Data Centre (GRDC) and the Swedish Meteorological and Hydrological Institute (SMHI) as examples (Figures 1a and 1b). This is also the case for the data used in the study by Allamano et al. [2009a], which we here use as an example of the potential problems using pooled data (Figure 1c). Among the 27 catchments used in the study (P. Allamano, personal communication, March 2010) (Table 1) there is no measured

[1]Institute of Geography, University of Bern, Bern, Switzerland.
[2]Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland.
[3]Hydrology Division, Federal Office for the Environment, Bern, Switzerland.
[4]Institute of Hydrology, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany.
[5]Department of Geography, University of Zurich, Zurich, Switzerland.
[6]Department of Physical Geography and Quaternary Geology, Stockholm University, Stockholm, Sweden.
[7]Department of Earth Sciences, Uppsala University, Uppsala, Sweden.
Corresponding author: J. Seibert, Department of Geography, University of Zurich, Winterthurerstr. 190, CH-8057 Zurich, Switzerland. (jan.seibert@geo.uzh.ch)

0043-1397/12/2011WR011690

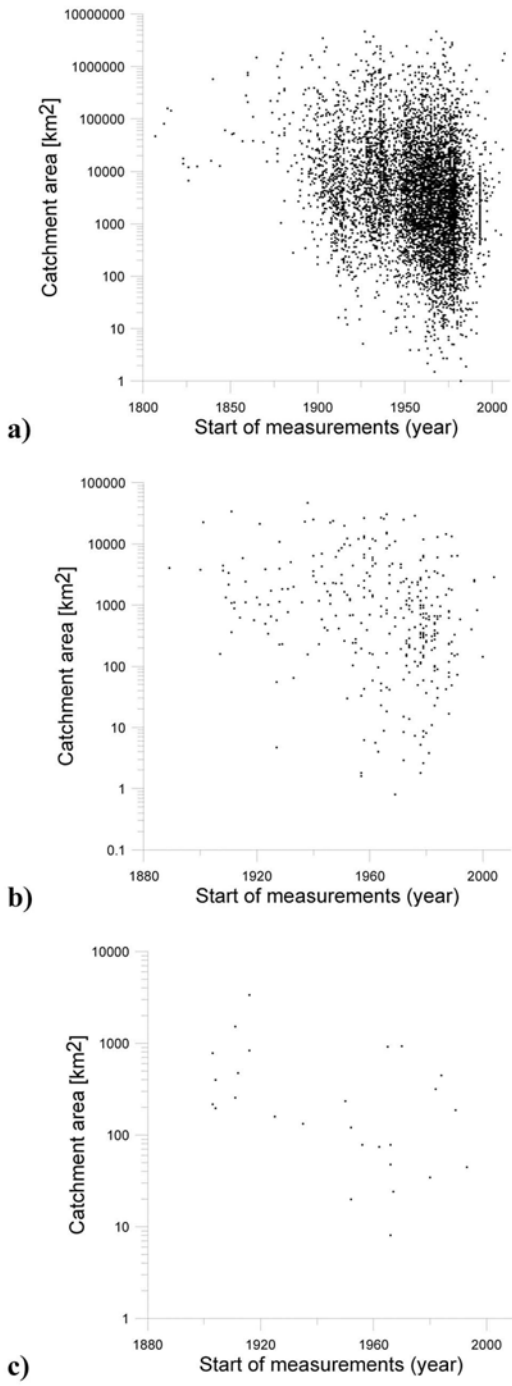

Figure 1. Catchment area versus the starting year of measurements for different sets of gauging stations: (a) stations from the Global Runoff Data Centre (GRDC), (b) stations of the Swedish Meteorological and Hydrological Institute (SMHI), and (c) stations in the southern Alps included in the study by Allamano et al. [2009a].

catchment with a drainage area of less than 100 km^2 before 1950, while 9 out of 15 catchments with records starting after 1950 have a drainage area of less than 100 km^2.

[4] Since the focus of our reanalysis was on illustrating the effect of pooling data, we used the same data and statistical analyses methods as Allamano et al. [2009a]. However, we must mention that while the paper by Allamano et al. [2009a] aims at describing changes in flood risks for all of Switzerland, catchments exclusively located south of the main alpine water divide, i.e., from the Ticino and Valais regions, were included in their study.

2. Trend Reanalysis

[5] Applying quantile regression [Koenker and Hallock, 2001] to the annual maxima of specific discharge for all gauged catchments, Allamano et al. [2009a] found a clear increasing trend (Figure 2a). The alternative way to analyze data from different runoff stations would be to determine trends for each streamflow record, which can then be interpreted to answer the question of whether there is a common trend. Figure 2b shows the 0.95 quantile regression trends that result from analyzing the annual maxima peak flow data series individually for each discharge gauging station used by Allamano et al. [2009a]. These individual quantile regressions are far from following a specific pattern, but rather show rising and falling trends, some significant, some not. This is especially true for the years after 1950, which are partly responsible for the false impression of an overall rising trend. More or less the same findings apply to the 0.5 and 0.25 quantiles, for which Allamano et al. [2009a] found significantly positive trends as well when using pooled data.

[6] Pooling runoff series from several stations can lead to misleading results because of these systematic differences in the length of the series, because this adds an artifact to any trend analysis. It is well known that small catchments generally produce significantly higher specific peak runoff values than larger catchments [e.g., Dalrymple, 1960; Gupta et al., 1994; Mimikou, 1984]. Decreasing specific peak flow values with increasing catchment size can also be found in the data used for the Swiss flood analysis (Figure 3). Because of this relation the tendency of runoff data series from smaller catchments starting later can cause an apparent increase in floods when the data is pooled. Allamano et al. [2009a] were aware of the potential influence of drainage area on peak discharge and attempted to consider this using a multiple linear quantile regression analysis. However, they still found an increasing peak flow trend, which indicates that it can be difficult to remove the effects of pooling data. We argue that the apparent increase of peak flows found by Allamano et al. [2009a] is mainly due to the addition of smaller catchments where observations started after 1950. The influence of this effect is corroborated with an additional analysis of the annual

Table 1. List of Catchments[a]

River and Discharge Gauge Site Name	Drainage Area (km^2)	Station Altitude (m asl)	Mean Catchment Altitude (m asl)	Year of First Observation	Number of Observation Years Until 2004
Breggia at Chiasso, Ponte di Polenta	47.4	255	927	1966	39
Brenno at Loderio	397	348	1820	1904	97
Calancasca at Buseno	120	746	1950	1952	51
Cassarate at Pregassona	73.9	291	990	1963	42
Drance de Bagnes at Le Châble, Villette	254	810	2630	1911	94
Drance at Martigny, Pont de Rossettan	672	474	2245	1991	14
Grande Eau at Aigle	132	414	1560	1935	70
Krummbach at Klusmatten	19.8	1795	2276	1952	51
Lonza at Blatten	77.8	1520	2630	1956	49
Maggia at Bignasco, Ponte Nuovo	315	432	1890	1929	76
Maggia at Locarno, Solduno	926	202	1530	1970	32
Magliasina at Magliaso, Ponte	34.3	295	920	1980	25
Massa at Blatten bei Naters	195	1446	2945	1904	90
Moesa at Lumino, Sassello	471	249	1662	1912	86
Rhone at Brig	913	667	2370	1916	89
Rhone at Reckingen	215	1311	2306	1903	81
Rhône at Sion	3373	484	2310	1916	89
Riale di Calneggia at Cavergno Pontit	24	890	1996	1967	38
Riale di Pincascia at Lavertezzo	44.4	536	1708	1993	12
Riale di Roggiasca at Roveredo, Bacino di compenso	8.06	980	1711	1966	39
Saltina at Brig	77.7	677	2050	1966	39
Ticino at Bellinzona	1515	220	1680	1914	91
Ticino at Piotta	158	1007	2060	1925	80
Ticino at Polleggio Campagna	444	298	1794	1987	18
Vedeggio at Agno	105	281	898	1981	24
Verzasca at Lavertezzo Campiòi	186	490	1672	1990	15
Vispa at Visp	778	659	2660	1903	93

[a]Here m asl is meters above sea level.

maximum peak flow data split by year of first observation: Figure 4a shows a quantile regression trend similar to that of Allamano et al. [2009a] for the pooled data from stations with observations before 1951 (no catchments with a drainage area of less than 100 km^2 are included; see Figure 1c). Using this subset, no trend can be detected in the 0.95 quantile regression. In other words, the trend in the overall pooled data could not be found in the data from those stations with longer records. Figure 4b shows the same analysis for the pooled data for stations with the first year of

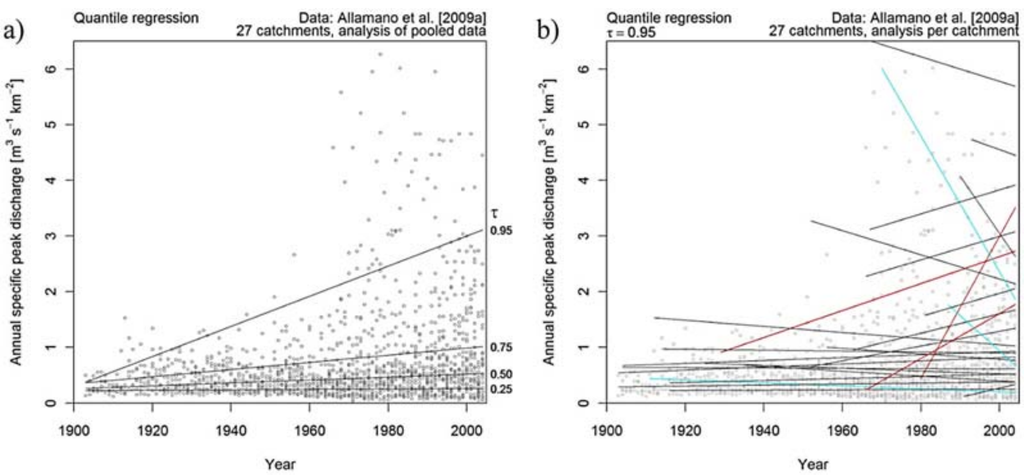

Figure 2. (a) Overall quantile regression for pooled data from 27 discharge gauging stations as reported by Allamano et al. [2009a, Figure 2a]. (b) Our analysis of the same data, showing individual quantile regressions for the time series of each of the 27 discharge gauging stations at $\tau = 0.95$; each line refers to the time series of one station used by Allamano et al. [2009a]. Series with significantly rising or falling trend at $\alpha < 0.05$ are drawn in red and cyan, respectively.

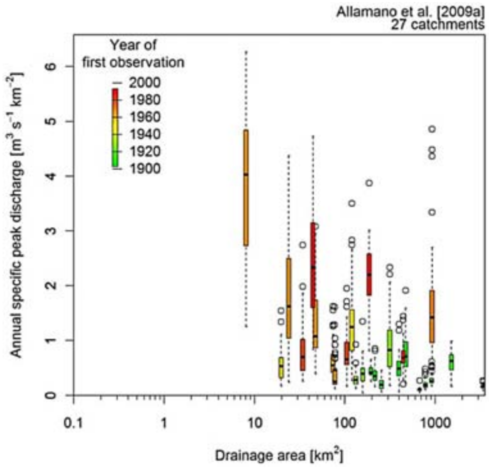

Figure 3. Peak flow data as used by Allamano et al. [2009a] drawn against drainage area, with one box plot per station. The box plot color indicates the year of the first observation.

observation after 1950 (60% of these catchments have a drainage area of less than 100 km^2, see Figure 1c) and the trend in the 0.95 quantile is even reversed, which further illustrates that the detected overall trend is misleading.

3. Discussion and Conclusion

[7] Deriving false trends when using pooled data can lead to serious consequences. In the case of the study by Allamano et al. [2009a] the statistical analysis was further used to confirm the results of a simple model, which had been described by Allamano et al. [2009b]. Simple models may serve as powerful tools for studying the impact of changes in key driving variables on hydrological processes. However, the application of a simple model can lead to misleading results if representativeness of data and validity of statistical analyses are not considered with due care. It is of concern that the simple model used by Allamano et al. [2009a] actually reproduces an apparent trend in the observed data that does not seem valid at closer examination. We argue, therefore, that the conclusions drawn are severely limited in their validity, if not even misleading.

[8] There is strong need for reliable estimations of anticipated changes in flood frequencies especially in mountain areas. In view of the high societal and economic challenges

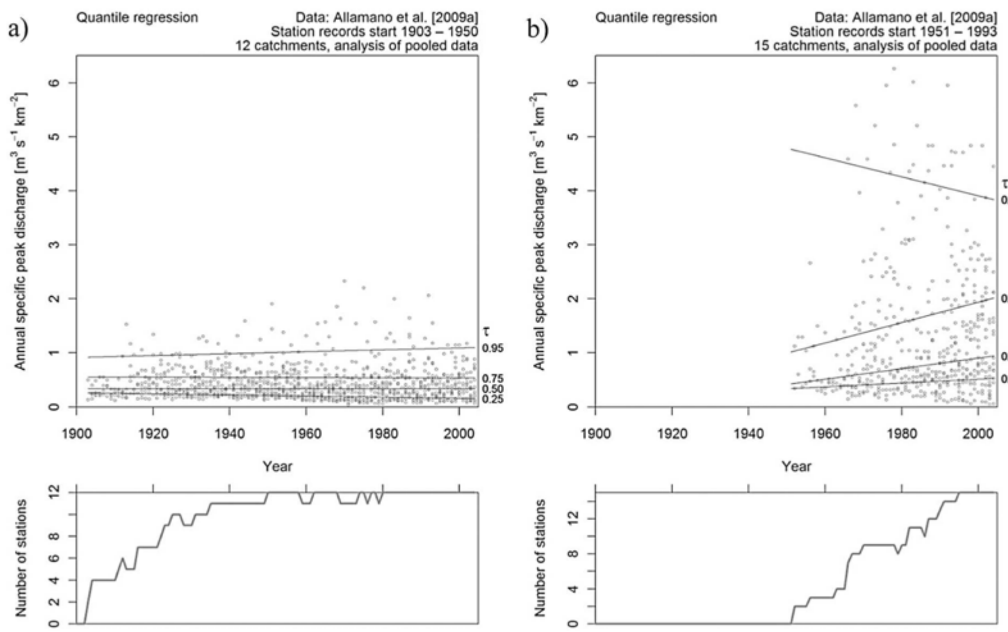

Figure 4. (a) Quantile regression for pooled data from discharge gauging stations in the set used by Allamano et al. [2009a] that have their first year of observation before 1951. (b) Quantile regression for pooled data from discharge gauging stations in the set used by Allamano et al. [2009a] that have their first year of observation after 1950. The bottom plots show the number of stations with data for a particular year.

related to climate change and floods, and the credibility of climate and hydrology change research, we wish to stress urgently the need to be very careful when analyzing data records. As illustrated in this note, it is not advisable to pool flow data from different gauging stations with varying start times and record duration if the sample of stations is not uniform over time regarding catchment characteristics or other system properties. In particular, it is important to recognize that gauging stations with records starting in the early 20th century generally refer to larger drainage areas than those with a starting date of 1960 or later.

[9] Acknowledgments. We thank Caroline Kan from the Federal Office for Environment (FOEN) for communication of peak flow series and further auxiliary data and Jose D. Salas and an anonymous reviewer for valuable comments.

References

Allamano, P., P. Claps, and F. Laio (2009a), Global warming increases flood risk in mountainous areas, Geophys. Res. Lett., 36, L24404, doi:10.1029/2009GL041395.

Allamano, P., P. Claps, and F. Laio (2009b), An analytical model of the effects of catchment elevation on the flood frequency distribution, Water. Resour. Res., 45, W01402.

Dalrymple, T. (1960), Flood-frequency analyses, Manual of hydrology: Part 3. Flood-flow techniques, U.S. Geol. Surv. Water Supply Pap., 1543-A, 80 pp.

Gupta, V. K., O. J. Mesa, and D. R. Dawdy (1994), Multiscaling theory of flood peaks: Regional quantile analysis, Water. Resour. Res., 30(12), 3405–3421.

Hirsch, R. M., and K. R. Ryberg (2012), Has the magnitude of floods across the USA changed with global CO_2 levels? Hydrol. Sci. J., 57, 1–9.

Intergovernmental Panel on Climate Change (2007), Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by M. L. Parry et al., Cambridge Univ. Press, Cambridge, U. K.

Koenker, R., and K. F. Hallock (2001), Quantile regression, J. Econ. Perspect., 15(4), 143–156.

Kundzewicz, Z. W., D. Graczyk, T. Maurer, I. Pinskwar, M. Radziejewski, C. Svensson, and M. Szwed (2005), Trend detection in river flow series: 1. Annual maximum flow, Hydrol. Sci. J., 50(5), 797–810.

Kundzewicz, Z. W., L. J. Mata, N. W. Arnell, P. Döll, P. Kabat, B. Jiménez, K. A. Miller, T. Oki, Z. Sen, and I. A. Shiklomanov (2007), Freshwater resources and their management, in Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by M. L. Parry et al., pp. 173–210, Cambridge Univ. Press, Cambridge, U. K.

Mimikou, M. (1984), Regional relationships between basin size and runoff characteristics, Hydrol. Sci. J., 29(1), 63–73.

Mudelsee, M., M. Borngen, G. Tetzlaff, and U. Grunewald (2003), No upward trends in the occurrence of extreme floods in central Europe, Nature, 425(6954), 166–169.

Organe consultatif sur les changements climatiques (2007), Climate Change and Switzerland 2050: Expected Impacts on Environment, Society and Economy, 168 pp., Bern.

Stern, N. H. (2007), The Economics of Climate Change: The Stern Review, 692 pp., Cambridge Univ. Press, Cambridge, U. K.

Svensson, C., Z. W. Kundzewicz, and T. Maurer (2005), Trend detection in river flow series: 2. Flood and low-flow index series, Hydrol. Sci. J., 50(5), 811–824.