Statistical trend analysis of annual maximum discharges of the Rhine and Meuse rivers

F.L.M. Diermanse, J.C.J. Kwadijk, J.V.L. Beckers* and J.I. Crebas

Deltares, Delft, the Netherlands **Email:Joost.Beckers@deltares.nl*

Abstract

As a result of climate change, the rainfall intensity in Northern Europe is expected to increase. In this paper, four standard statistical tests are employed to find evidence for such a trend in the available series of annual maximum discharges of the Rhine and Meuse rivers, both of which span a period of approximately 100 years. All four statistical tests fail to falsify the null hypothesis of a stationary process at the 5% significance level. This leads to the conclusion that there is no statistically significant trend in the annual maximum discharge series. However, it is also shown that the statistical tests have insufficient detection power for a relatively weak trend in the 100-year discharge series. The probability of detection of a trend by these tests will only exceed 50% for time series of more than 130 years. Furthermore, it is shown that the tests are not far from the significance level. Extension of the original data series by synthetic data demonstrates that the statistical tests will reject the null hypothesis if the presumed trend will continue over the next decades. Therefore, the fact that the discharge series of the Rhine and Meuse rivers are close to giving evidence for non-stationarity should serve as a serious warning for the possibility of a systematic increase in river discharges.

Introduction

A possible adverse effect of world-wide climate change is an increase of extreme river discharges and associated flood risk (Milly *et al.* 2002, 2005; IPCC, 2007). In Northern Europe, a trend is observed towards more intense winter precipitation, which is also supported by General Circulation Models (GCM) (EEA, 2008). This trend in winter precipitation is likely to cause an increase in the frequency of extreme river discharges and the probability of flooding in the Rhine and Meuse basins (Gellens and Roulin, 1998; Menzel *et al.,* 2006; Lenderink *et al.,* 2007; Hooijer *et al.*, 2004; Pinter *et al.,* 2006).

A climate-induced trend in the probability of extreme discharges is of major importance to water managers, because this has direct consequences for the design of flood defence systems. Firstly, the design water levels for flood protection works are usually derived from stationary extreme value theory (EVT) (Fisher and Tippett, 1928), which requires a homogeneous series of observed annual maxima (AM) or peaks over threshold (POT) (Kottegoda, 1980; Garrett and Müller, 2008). If the historic records are non-stationary, then either this trend should be removed before stationary EVT is applied or one should turn to non-stationary EVT. Secondly, a flood protection work, e.g. a levee, has a typical life span of 50 to 100 years. A confirmed trend in the design water level should be extrapolated into the future, i.e. towards the end of the life span of the levee. This will provide a design water level for a climate-proof flood protection, assuming persistence of historic rates of change.

The design and the five-yearly statutory safety assessment of the flood defenses along the Rhine and Meuse river branches in the Netherlands are based on hydrodynamic calculations that use a design discharge as boundary condition. These design discharges are associated with a return period of 1250 years. They are determined by extrapolation of the observed AM discharges at Lobith and Borgharen, where the Rhine and Meuse enter the Netherlands respectively. The AM series at these locations are important references for the Dutch flood protection system. A possible climate-induced non-stationarity of these AM series is therefore a major concern for the Dutch water managers. From the precautionary principle, a trend in river discharges is already assumed in the current design rules.

Water management in the Rhine and Meuse basins is a multi-billion euro business. A positive trend in extreme discharges requires substantial additional investments in on-going flood risk management projects, such as 'Room for the River' for the Rhine (http://www.ruimtevoorderivier.nl) and similar projects for the Meuse. To justify the additional investments, evidence for the climate-induced effect should be based on sufficient scientific grounds. One possible approach to provide evidence for a trend in the AM discharges is to apply standard statistical trend tests. These tests confirm a trend if the null hypothesis that the observations are samples from a stationary process can be rejected with sufficient confidence. It is the aim of this paper to determine whether the historical discharge AM of the Rhine and Meuse rivers provide such statistical evidence of a positive trend.

The next section describes the four statistical trend tests that are employed to test stationarity in the AM of Rhine and Meuse discharges. The results show that none of the statistical tests reject the null hypothesis. However, this does not mean that no trend exists, but rather that it cannot be confirmed by the four statistical tests, a factor that is further elaborated by investigating the detection power of the tests. The historic series are extended by adding synthetic data to determine how far the tests are from rejecting the null hypothesis. The detection power of the statistical tests is investigated by Monte Carlo sampling.

BHS Third International Symposium, Managing Consequences of a Changing Global Environment, Newcastle 2010

Statistical tests

There are several statistical tests available for testing stationarity of time series (Hirsch *et al.,* 1993). These tests start from a null hypothesis that the observations are samples from a stationary process. The likelihood of this hypothesis is evaluated based on the value of a test statistic, a property of the data set. A large deviation of the test statistic from the stationary value is unlikely to be coincidental. The P-value is the probability that the deviation of the test statistic from the homogeneous case is coincidental. If this probability is sufficiently small, then the null hypothesis is rejected and the alternative hypothesis is believed to be true — that the process is non-stationary. A P-value of 5% is a common critical value for accepting statistical significance, but there is no reason why other values cannot be used. Rejection of the null hypothesis at the 5% significance level, means that we are 95% confident of non-stationarity.

If the alternative hypothesis specifies no direction for the trend, the test is called two-sided. In a one-sided test the alternative hypothesis does specify a direction. In this study, the interest lies only in positive trends. A negative trend is considered part of the null hypothesis. This effectively increases the probability of accepting a positive deviation as evidence for non-stationarity. In this study, the significance levels are set for a one-sided test, because the alternative hypothesis is a positive trend in the discharges.

The four statistical trend tests that have been employed are described briefly below. For a more detailed description, readers may refer to generally available texts on statistics.

Pearson t-test (linear trend test)

The classical Student's t-test evaluates the significance of the correlation between the values of the AM discharges and their years of observation. Pearson's correlation coefficient ρ is calculated from the covariance and standard deviation of both variables. Student's t-test is then used to test the P-value of the test statistic ρ. This is a parametric test, because it assumes that ρ follows a Student's t distribution. If this assumption is not true the conclusions may be invalidated.

Spearman's rank correlation test

This test is the non-parametric analog of the Pearson t-test. The test statistic is Spearman's rank correlation coefficient r_s, which is the correlation between the *ranks* of the AM discharges and their years of observation. Because of the use of ranks instead of the absolute values, the sampling distribution of r_s for a stationary process can be calculated without the assumption of a distribution function. For short series, the P-value can be calculated analytically. For longer series the P-value can be approximated by an Edgeworth series (Best and Roberts, 1975).

Mann-Kendall test

The Mann-Kendall test is another non-parametric significance test for a monotonic trend in a time series based on the Kendall's τ (Mann, 1945; Kendall, 1975). This test compares the ranks for all pairs of AM discharges. This amounts to N*(N-1)/2 combinations. The test statistic τ is the difference between the number of pairs that support a positive trend and the number of pairs that support a negative trend, divided by the standard deviation. The null hypothesis is that the data are independent and thus randomly ordered. For a substantial number of observations the test statistic τ will then be normally distributed.

Wilcoxon-Mann-Whitney test

The Wilcoxon-Mann-Whitney test (Wilcoxon, 1945), also known as the Wilcoxon ranks sum test or the Mann-Whitney U test, is a non-parametric test for two independent samples. The WMW test can be used for trend analysis by splitting the time series into a first and second half and testing the null hypothesis that the two sub-sets are taken from the same distribution. The test statistic is the sum of the ranks of the elements in each sub-set. The P-value can be calculated exactly by considering all possible combinations or approximated by a normal distribution for large sample sizes. The two sub-sets need not have identical lengths, as was pointed out by Mann and Whitney (1947). If a crossover is expected at, for example, two-thirds of the time series, then the test should be performed on the two sub-sets before and after this moment. The only further requirement is that the variables are ordinal.

Application of the tests to Rhine and Meuse data series

The statistical tests were applied to the AM series of river discharges at Lobith for the hydrological years 1901–2003. The Meuse discharges recorded at Borgharen run from 1911 to 2003. The AM data are corrected for changes in the measurement method and river stage relationship due to man- made water works. The AM series for both locations are displayed in Figures 1 and 2. The plotted linear trends indicate an increase in mean annual maximum of 8 $m^3 s^{-1}$ per year (Rhine) and 3.4 $m^3 s^{-1}$ per year (Meuse), as determined by

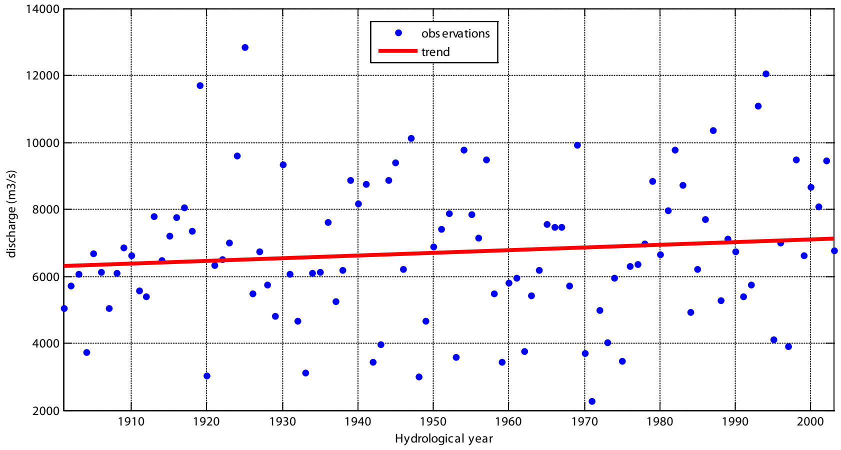

Figure 1 *Observed AM discharges of the Rhine river at Lobith and the least squares fit of a linear trend.*

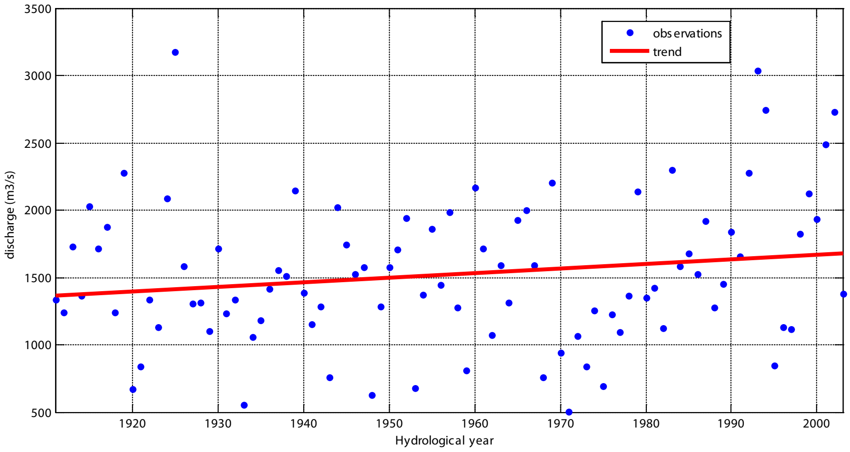

Figure 2 *Observed AM discharges of the Meuse river at Borgharen and the least squares fit of a linear trend.*

least squares linear regression. This amounts to respectively a 13% and 23 % increase over the observation periods.

The four statistical tests were applied to test the significance of the positive trend. The results are given in Table 1. Based on these results, the null hypothesis of a homogeneous series cannot be rejected by any of the four statistical tests. None of these significance levels are lower than the defined critical level of 5%. So, based on these tests, we cannot state with (more than) 95% certainty that there is indeed a positive trend. The Meuse discharge series is closer to rejection of the null hypothesis than the Rhine series, but the P-values are still far from the significance level of 5%. This may seem to conflict with the earlier observation of 13% and 23% increase in discharges over the past decade. On the other hand, natural variation of annual maximum discharges of these rivers is much larger than these percentages. One or two extremes can therefore strongly influence the slope of the trend line. Therefore, coincidence cannot be ruled out as the cause of the observed positive trend in AM for the Rhine and Meuse.

Table 1 *P-values for the four statistical tests applied to the AM series of the Rhine and Meuse rivers.*

Statistical test	*Rhine*	*Meuse*
Pearson t-test (linear trend test)	25.7%	9.8%
Spearman	18.2%	14.7%
Mann-Kendall	15.4%	11.3%
Mann-Whitney-Wilcoxon	31.3%	21.4%

Sensitivity analysis

To further interpret the results from the previous section, we now investigate how far off the average trends in the observed AM discharge series are from being statistically significant. This is investigated by:

- Analysis of the detection power of the four statistical trends tests by means of Monte Carlo simulation of data series of variable trend and length;
- Extension of the observed time series by repeating the last years of observed AM data ;
- Extension of the observed time series by adding one extreme event.

The results of these investigations for the Rhine discharge series are described below. The conclusions also hold for the Meuse river data.

Detection power analysis

The detection power of the four statistical tests was studied by means of Monte Carlo simulation. Synthetic time series of 100 annual maxima are generated by sampling (with replacement) from the empirical probability distribution function of the Rhine discharge peaks. This distribution is a fitted piecewise log-linear function, with different coefficients a_i and b_i for different values for the return period T (Diermanse, 2004a,b):

$$Q = a_i \ln(T) + b_i$$

A linear trend was introduced by multiplying all values of Q by a factor that increases linearly with time. The four statistical trend tests were applied to the resulting time series, in order to evaluate the probability of detection (POD) of the trend. The POD is defined as the number of data series that yielded a P-value lower than the significance level of 5%, divided by the total number of generated series (10 000). The results for different trends are shown in Table 2. The POD for a zero trend is 5%, which is the expected rate of false positives or type I errors (rejecting a null hypothesis that is actually true). The detection powers for the first three tests are identical, except for the WMW test, which has a slightly lower POD. These results show that even in the case of an existing positive trend in the high discharges, there may not be enough statistical evidence for this trend in the observed series of annual maximum discharges. For instance, even for the hypothetical situation of a rather strong increase in average annual maximum discharges of 15% over a period of 100 years, according to Table 2 there is only 40% chance that the resulting series of annual maximum discharges provides enough statistical evidence for this trend.

In the following test, the POD as a function of the length of the series was investigated. In this test, a fixed positive trend in the mean annual maximum discharge of 13% per 100 years was induced, which is the derived linear trend in the observed Rhine data (see Figure 1). The results are shown in Table 3 and Figure 3. For a time series of 100 years (the length of the Rhine discharge data), all trend tests have a POD of 30%, except for the WMW test, which has a POD of 25%.

The POD analysis shows that the detection power of the trend tests is small (30%) for the current sample size of around 100 years and an assumed trend of 13% increase per century. In other words, the fact that the observed trend in the Rhine and Meuse data series is not found significant is not surprising. For longer data sets of 125 and 150 years, the POD increases to 42% and 58% respectively. The POD

Table 2 *Probability of detection (POD) at 5% significance level for a 100 year data series with induced trends.*

	Induced increase in the mean annual maximum discharge (% per 100 years)						
	0	5	10	15	20	25	30
Pearson t-test:	5%	12%	24%	40%	57%	72%	83%
Spearman:	5%	12%	24%	40%	57%	72%	83%
Mann-Kendall:	5%	12%	24%	40%	57%	72%	83%
Wilcoxon-Mann-Whitney:	5%	11%	20%	32%	47%	61%	73%

Table 3 *Probability of detection (POD) for varying length of the time series and a trend of 8 m^3/s per year.*

	Trend factor on the last data point in the time series						
	50	75	100	125	150	175	200
Pearson t-test:	11%	18%	30%	46%	64%	80%	91%
Spearman:	11%	18%	30%	46%	64%	80%	91%
Mann-Kendall:	11%	18%	30%	46%	64%	80%	91%
Wilcoxon-Mann-Whitney:	9.5%	16%	25%	38%	54%	70%	83%

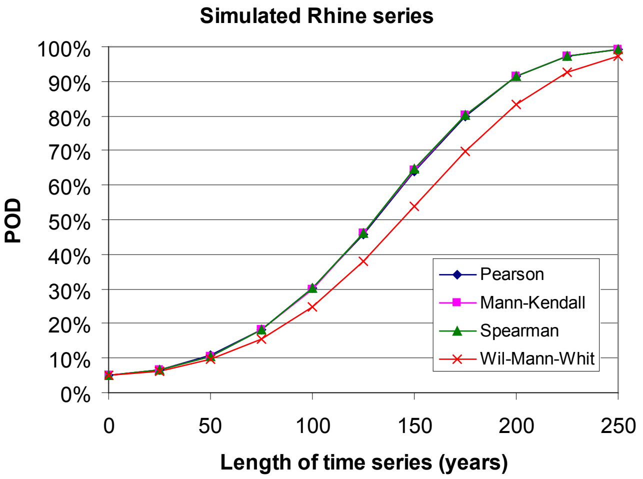

Figure 3 *POD of an induced trend of 13% per 100 years at the 5% confidence level, for different lengths of the time series.*

reaches 50% at 130 years for the Pearson, Mann-Kendall and Spearman tests and 145 years for the WMW test. This leads to the conclusion that if the presumed trends in the discharge data will persist over the next 25 to 50 years, the chance of this trend becoming statistically significant becomes more likely than not.

Extension of the observed data series

To investigate this further, the observed data series was extended by simply repeating the most recent observations at the end of the series. The 5, 10, 15, etc. final years of observed AM discharges were repeated at the end of the 100 year time series, to yield 105, 110, 115 etc. year time series. The results in Table 4 show that for the Rhine data, the extended series start to provide evidence for non-stationarity after 25 additional years of observations. The Meuse series already shows a significant trend after repeating the last five years of observations. This is due to the fact that four of the five most recent Meuse AM discharges were significantly above average. This shows that with a repetition of the observed extremes of the past few years there will be enough evidence for non-stationarity.

Extension of the data series by a single additional event

Another way to investigate how far the observed AM discharge series are from providing evidence for a non-stationarity is to add one extreme event at the end of the series. The result for the Rhine series is shown in Figure 4, where the magnitude of this event is expressed in terms of its return period. Figure 4 shows that up to T=1000 years, a single additional event does not lead to a P-value below the significance level. The parametric Pearson t-test will eventually reject the null hypothesis, but the return period of

Table 4 *P-values for extended data series by repeating the last years of observations at the end of the series.*

add	**Rhine**				**Meuse**			
years	Pearson	Mann-Kend	Spearman	WMW	Pearson	Mann-Kend	Spearman	WMW
0	26%	15%	18%	31%	10%	11%	15%	21%
1	26%	15%	17%	33%	11%	12%	16%	25%
2	19%	10%	12%	24%	5%	7%	10%	18%
3	16%	8%	9%	17%	3%	4%	6%	13%
4	12%	6%	6%	18%	2%	3%	4%	9%
5	13%	6%	6%	22%	2%	2%	2%	5%
6	9%	4%	4%	21%	1%	1%	2%	6%
7	14%	6%	7%	27%	2%	2%	3%	10%
8	13%	6%	6%	27%	3%	4%	5%	13%
9	19%	9%	10%	34%	4%	6%	8%	17%
10	11%	6%	6%	33%	0.2%	0.3%	0.3%	3%
11	7%	4%	4%	32%	1%	2%	3%	7%
12	8%	5%	5%	31%	1%	1%	2%	6%
13	9%	6%	7%	32%	1%	1%	2%	6%
14	9%	6%	7%	26%	1%	1%	1%	4%
15	9%	6%	6%	21%	1%	1%	1%	4%
20	9%	7%	7%	13%	1%	1%	1%	8%
25	4%	3%	3%	3%	0.5%	1%	1%	7%
30	7%	4%	5%	9%	1%	2%	3%	5%

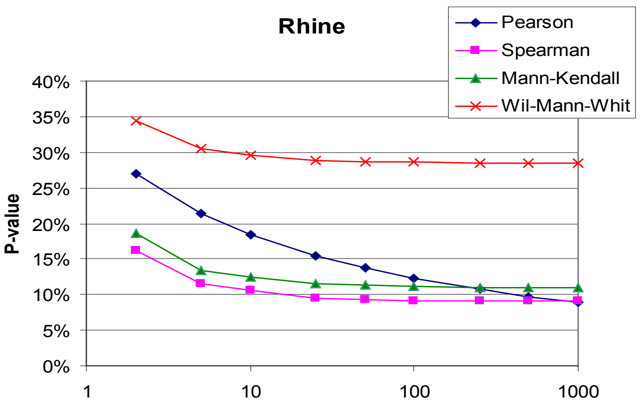

Figure 4 *P-values of extended Rhine series by adding one additional event at the end of the series.*

the event that is required for this is extremely long. Therefore, we conclude that a single event will not provide sufficient proof for a significant trend in the Rhine data. A more sustained continuation of the positive trend is required for this. The same conclusion can be drawn for the Meuse series (data not shown).

Conclusion and discussion

Although the AM discharge of the Rhine and Meuse series indicate a positive trend over the 100-year observational period, this trend is not found to be statistically significant. Significance levels are between 10% and 30% for the four statistical tests applied, which is above the significance level of 5%. There is insufficient statistical ground to reject the possibility that the observed trends are coincidental. This does not mean that there is no physical cause for a trend, but rather that it cannot be confirmed with sufficient confidence.

On the other hand, extension of the observed data series by repeating the observations from the last period demonstrates that, if the trend continues, the observations are likely to give evidence for non-stationarity within 5 to 25 years. This should serve as a warning to water managers to account for the possibility of a climate-induced trend in the Rhine and Meuse maximum annual discharges becoming statistically significant in a few decades. A flood risk protection strategy that is both safe and cost-effective would not yet take a positive trend in AM discharges for granted, since the evidence for non-stationarity is still lacking. However, in order to be safe, this strategy would take into account the possibility of a trend becoming evident in the near future. A flexible design strategy that leaves room for measures that may be needed in 25 years therefore seems optimal. This will have, for instance, consequences for spatial planning, i.e. no houses should be built near dikes that may need expansion in 25 years.

It is recommended to explore alternative approaches to investigate the possible effects of climate change on extreme river discharges, for example by introducing average climate change in historical time series (Lenderink *et al.,* 2007) or by downscaling of GCM to regional climate models (Wilby *et al.,* 2000; Kay *et al.*, 2006; Te Linde *et al.,* 2010). Extension of the analysis to other rivers in Northern Europe and combination of the data might also be worthwhile.

References

Best, D.J. and Roberts, D.E. 1975. AS 89, Upper Tail Probabilities of Spearman's Rho. *Applied Statistics* **24** (3).

Diermanse, F.L.M. 2004a. HR2006-herberekening werklijn Rijn, *WL| Delft Hydraulics Tech. report Q3623-1.*

Diermanse, F.L.M. 2004b. HR2006-herberekening werklijn Maas, *WL| Delft Hydraulics Tech. report Q3623-2.*

European Environment Agency 2008. Impacts of Europe's changing climate: 2008 indicator-based assessment. *Tech. report*.

Fisher, R.A. and Tippett, L.H.C. 1928. Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proc. Camb. Phil. Soc.,* **24,** 180–190.

Garrett, C. and Müller P. 2008. Extreme events. *Bull. Amer. Meteorol. Soc*. **89** (11), 133.

Gellens, D. and Roulin E. 1998. Streamflow response of Belgian catchments to IPCC climate change scenarios, *J. Hydrol.* **210,** 242–258.

Hirsch, R. M., Helsel, D. R., Cohn, T. A. and Gilroy, E. J. 1993. Statistical analysis of hydrologic data. In: *Handbook of Hydrology* (ed. by D. R. Maidment), Ch. 17, 17.11–17.37. McGraw-Hill, New York, USA.

Hooijer, A., Klijn, F. Pedroli, G.B. M. and Van Os A.G. 2004. Towards sustainable flood risk management in the Rhine and Meuse basins: Synopsis of the findings of IRMA□SPONGE, *River Res. Appl*., **20**, 343–357.

Intergovernmental Panel on Climate Change 2007, Summary for policymakers, in Climate Change 2007: The Physical Science Basis, edited by S. Solomon *et al*., pp. 1–23, Cambridge Univ. Press, Cambridge, U. K.

Kay, A. L., Reynard, N. S. and Jones R. G. 2006, RCM rainfall for UK flood frequency estimation: I. Method and validation, *J. Hydrol*., **318,** 151–162.

Kendall, M. G. 1975. *Rank Correlation Method*s. Griffin, London, UK.

Kottegoda, N.T. 1980. *Stochastic Water Resources Technology*. The Macmillan Press Ltd., London. p. 384.

Lenderink, G., Buishand, A. and Van Deursen W. 2007. Estimates of future discharges of the river Rhine using two scenario methodologies: Direct versus delta approach, *Hydrol. Earth Syst. Sci.,* **11** 1145–1159.

Mann, H.B. 1945. Nonparametric tests against trend. *Econometrica* **13**, 245–259.

Mann, H. B., and Whitney, D. R. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Mathematical Statistics*, **18,** 50–60.

Menzel, L., Thieken, A. H. Schwandt, D. and Bürger G. 2006. Impact of climate change on the regional hydrology, Scenario-based modelling in the German Rhine catchment, *Nat. Hazards*, **38,** 45–61.

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

Milly, P.C. D., Dunne, K. A. and Vecchia, A.V. 2005. Global pattern of trends in streamflow and water availability in a changing climate, *Nature* **438**, 347–350.

Pinter, N., Van der Ploeg, R.R. Schweigert, P. and Hoefer, G. 2006. Flood magnification on the river Rhine, *Hydrol. Processes,* **20**, 147–164.

Te Linde, A. H., Aerts, J.C.J.H., Bakker, A.M.R. and Kwadijk, J.C.J. 2010. Simulating low-probability peak discharges for the Rhine basin using resampled climate modeling data, *Water Resour. Res.*, **46**, W03512.

Wilby, R.L., Hay, L.E. Gutowski Jr., W. J. Arritt, R. W. Takle, E. S. Pan, Z. Leavesley, G. H. and Clark, M. P. 2000. Hydrological responses to dynamically and statistically downscaled model output, *Geophys. Res. Lett.*, **27**(8), 1199–1202.

Wilcoxon, F. 1945. Individual comparisons by ranking methods. *Biometrics*, **1**, 80-83.