Interactive comment on "Assessment of extreme flood events in changing climate for a long-term planning of socio-economic infrastructure in the Russian Arctic" *by* E. Shevnina et al.

E. Shevnina et al.

elena.shevnina@fmi.fi

We thank the referee, Dr. Francesco Serinaldi, for his constructive comments and suggestions. They were very useful in further improving of the manuscript.

The Referee's comments are copied below, and our responses are written after each comment.

General comments The manuscript under review uses a simple linear system of two equations (Eq. 2 in the text), which is the simplified solution of an equivalent Fokker-Plank stochastic differential equation, to update the estimates of the first two moments of the flood depth distribution by replacing historical average of annual rainfall amount

over a reference historical period () with the corresponding average over a future time window (). Since Eq. 2 provides a relationship between rainfall and runoff, the idea is to use the rainfall output of climate models as an exogenous variable to update the parameters of the flood depth distribution under climate (deterministic) change scenarios. I have to say that the paper is not very easy to read. Simple concepts are reported in a unnecessary complicated manner, while some parts are not necessary and only make the reading more difficult.

My doubts about such a type of papers are always the same: what about uncertainty? Why 50 years (data points) of annual maxima or spring flood data are not enough for standard flood frequency analysis but are enough when we deal with nonstationarity and projections, which in turn introduce further uncertainty? In this manuscript the Authors deal with time series lengths varying from 26 to 77 years with average of 51 years. For 50 observations, under iid conditions, the unbiased point estimate of the probability of exceedance of the largest observation is 1/50, but its 95% confidence interval is about (1/15, 1/2000), meaning that the return period ranges within 15 and 2000 years. Now under this condition of deep uncertainty, discussing changes of 7% in the point estimate of 1% Pearson III quantile, or changes "from 1% (calculated from the observations) to 2.5% (calculated according to the averaged climate projections)." For the exceedance probability of the largest observation is only a matter of speculation, whereby the Nadym River data consist of 37 annual values from 1955 to 1991. These changes are much lower than the large uncertainty affecting the point estimates based on historical records and easily fall within their confidence intervals. In this respect, all the moments, parameters and design quantiles reported throughout the text should be complemented by confidence intervals or something similar in order to communicate the actual uncertainty of the point estimates. The same holds for the CDFs shown in Figs. 2, 4, and 6, as well as mean values and coefficients of variation reported in Fig. 3. Simple standard bootstrap techniques can be used to accomplish this task.

Our response: To calculate the confidence intervals not only including size of sample

and level of confidence are important. Also a population variability is required. How the numbers shown in your sentence were evaluated (1/15, 1/2000)? Of course, a larger sample size normally will lead a better estimate of the population parameters. However, engineering hydrology usually operates with samples with lengths of the order of 50–80 years, and with a statistically significant autocorrelation detection in the observed time series. The pdfs of multi-year runoff (annual, maximal and minimal) have significant skewness and do not fit Pearson I type distribution. Strictly speaking, the runoff is not i.i.d. variable. However, numerous studies provide the practical aspects of application of the classical statistical methods to estimate the risk of occurrence of detrimental hydrological events, and flood frequency analysis in particular. We have added references to the revised manuscript, to put the study in general context as recommended by the Anonymous Reviewer #1. These studies also include discussion of the confidence intervals for the runoff calculations. We thank the Referee for so sharply formulating the main idea of the method: to predict the future parameters of pdf using new climatology, and to construct the pdf with a-priory defined distribution (Pearson III type), and finally to calculate the tailed values. However, in this case the confidence intervals can not be used to present the uncertainties (the credible intervals may be more suitable).

By the way, since the Bullettin 17b is mentioned in the paper, I would like to highlight that it relies on regionalization procedures aiming at 'selling space for time' in order to (try to) reduce the large uncertainty of at-site estimates (as well as allowing for estimation in ungaged sites). This is just to say that at-site methods have been recognized to provide unreliable extrapolations toward extreme (design) quantiles several decades ago under iid conditions; so, it seems to me quite anachronistic to propose at-site methods based on few tens of data to support even more uncertain 'nonstationary' or 'quasi-stationary' design procedures.

Our response: It was mentioned in the introduction that the study presents the method to evaluate a regional scale assessment of extreme flood events. The Bulletin-17B and

the Russian Guidelines (SP33-101-2003) are also dedicated to regionalization using the observed time series (and sampled statistics). The present study used the same concept, but for the projected parameters of pdf, since future time series do not exist. The regionalization concept may seems anachronistic, but this circumstance allows easy performing a regional scale evaluation of extreme detrimental events for the risk assessment purposes.

Other aspects rise further doubts. For example, the proposed system in Eq. 2 implies changes in both mean and variance, but the two-fold cross-validation is based on t-tests checking only for changes in mean. I cannot see information about possible (detected) changes in variance in the historical records. Moreover, from a modeling point of view, detected changes require attribution to avoid incorrect treatment of deterministic changes.

Our response: It is not easy to find the time series with two periods with statistically significant differences in two moments (in Eq.2). The main problem was already mentioned by the Referee: the length of time series does not allow to perform the sub-division based of variance analysis (F-test) even used the statistical estimators "adapted" to engineering calculations. Then, we used the assumption, that for the runoff time series, the difference in the second statistical moment is proved by the statistical significance of the differences in the first statistical moment.

Temperature is mentioned in the text but I cannot see where it is used in Eqs. 1 and 2 or elsewhere.

Our response: The mean values of air temperature were used to perform the regional-oriented parameterization scheme of the model. This part of the study is presented in details within other paper (Shevnina, 2012), and was published unfortunately only in Russian (as numerous papers described the basic assumptions used by the method (Kovalenko, 1993, Kovalenko et al. 2006, 2010)). This circumstance makes difficulties during reading of the paper (as was mentioned by the Referee in his general com-

ments) since the study tries to reconcile statistic and physical principles both used in this study.

Finally, a minor remark about the introduction. This the n-th time I read this sentence in an introduction 'However, the frequency and magnitude of extreme flood events based on historical data do not provide correct estimations for a future under changing climate (Milly et al., 2008)'. Now, we can discuss for hours about the meaning of 'changing' in this context, but the main point is that this sentence is more or less copied and pasted as is from paper to paper. I would like to stress that Milly et al. (2008) is a one page opinion paper with no diagrams, proofs, equations, and only few references. It is perfectly legitimate but also debatable. I suggest thinking more critically before reporting sentences that are taken for granted as a truth (aletheia), when they are debatable opinions (doxa). In this respect, it is fairer to report different points of view, so that the readers can build their own. See e.g.: Koutsoyiannis D, Montanari A, Negligent killing of scientific concepts: the stationarity case Hydrol Sci J (2014) http://dx.doi.org/10.1080/02626667.2014.959959, Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, R. J. Stouffer, M. D. Dettinger, and V. Krysanova (2015), On Critiques of "Stationarity is Dead: Whither Water Management?," Water Resour. Res., 51, 7785- 7789, doi:10.1002/2015WR017408. Lins Harry F.,Timothy A. Cohn (2011) Stationarity: Wanted Dead or Alive? Journal of the American Water Resources Association 47(3), 475-480 Montanari A, Koutsoyiannis D, Modeling and mitigating natural hazards: Stationarity is immortal! Water Resour Res, 50 (12) (2014), pp. 9748-9756 Stedinger Jery R., Veronica W. Griffis (2011) Here to Where? Flood Frequency Analysis and Climate 47(3), 506-513

Our response: Yes, we agree that it is necessary to emphasize the doxa-status of the hypothesis of the non-stationarity in the context of hydrological applications using observed datasets. There are numerous studies which proof and deny both stationary and non-stationary hypotheses (we added discussion in the introduction of the revised manuscript). We consider it improbable that changes in meteorological variables would

remain unnoticed in runoff, which is an element of general water balance. From a practical point of view, the method allowing to evaluate the regional scale assessment of detrimental hydrological events is required besides of discussions of reality of the changes in climate, that why the studies similar to presented are usually supported in national and international levels.

It might also be useful to recall Russell's principles ('On the value of scepticism' from 'The will to doubt'): 'The scepticism that I advocate amounts only to this: âĂć that when the experts are agreed, the opposite opinion cannot be held to be certain; âĂć that when they are not agreed, no opinion can be regarded as certain by a nonexpert; and âĂć that when they all hold that no sufficient grounds for a positive opinion exist, the ordinary man would do well to suspend his judgment.'

Our response: Yes, skepticism is very useful, especially in science. But even in science, we need something (hypothesis, axioms, etc.) to rely on, to have a starting point to move forward. Otherwise, we will stay in a certain point, having no courage to move forward.

To summarize, even though I understand the requirement of developing simple methods for practitioners, in my opinion, the method suggested in this study is too simple and overlooks the large uncertainty involved in this type of studies as well as several other aspects. I think that a better approach is to provide fair stationary analyses based on observations, quantify the uncertainty, understand if possible changes due to climate projections are significant (taking the projection uncertainty/reliability into account), and provide a set of possible values to be shrunk at a second stage via e.g. ex post economic/financial analyses.

Our response: Yes, a fair stationary analysis based on observations is very important but not sufficient alone, since it does not provide a forecast, which is important for practical applications. In this study we try to fill the gap between engineering hydrology (with statistical methods) and physical hydrology (water balance) to provide fore-

casts of extreme flood events using climate projections. Again we note that the way of inter-comparison of the method used in this study with physically-based hydrological modelling is of a high interest.

C7