# Assessment of extreme flood events in changing climate for a long-term planning of socio-economic infrastructure in the Russian Arctic

**By E. Shevnina, E. Kurzeneva, V. Kovalenko, and T. Vihma**

**Submitted to Hydrol. Earth Syst. Sci. Discuss.**
***MS-NR: hess-2015-504_R1***

---

## REFEREE REPORT

### General comments

First of all I would like to reply to Authors' responses in order to better clarify some technical points and my point of view.

Uncertainty: confidence intervals for return periods I mentioned in my report are based on basic results for order statistics, which approximate confidence interval for quantiles. These issues are well-known and already discussed many years ago by Vit Klemes, among others. They can be found in his papers as well as books on applied statistics.
Saying that confidence intervals cannot be build (while credible intervals can) makes little sense. First of all, in my report I specified "confidence intervals or something similar"; so, if you prefer (Bayesian) credible intervals, or Dempster's imprecise probability or whatever else, you are free to implement it; in any case, given the limited information (data) used, uncertainty should be shown in some way because the comparison of point estimates is not enough and gives a false sense of accuracy. Second, I gave a "sharp" description of the method because the rationale is the same as whatever approach introducing dynamically-varying distributions where the parameters change according to covariates: you derived the relationships between runoff PDF parameters and rainfall moments via stochastic differential equations, while they are derived more commonly using pure empirical/statistical approaches. However, in both cases, uncertainty and CIs can be quantified irrespective of the availability of future data (which are not available by definition). In this respect, I suggested simple bootstrap techniques: leaving aside the uncertainty of future covariates (rainfall and temperature), the sampling uncertainty can easily be quantified by bootstrapping B times the observed records and re-estimating the model for the period of records, and then projecting this bundle of distributions by applying the proposed method to each of them. This way, we obtain a set of future distributions (driven by the same meteorological forcings, which are assumed to

be uncertainty-free) describing the propagation of the sampling uncertainty (which affects the parent distribution estimated in the period of records). Of course this is only one source of uncertainty but often is the most relevant. This technique, as well as more refined ones, is fully general and can be applied for whatever model; thus sentences such as “In this case, the classical estimates of the uncertainties on the PDF tailed values (i.e the confidence intervals) can not be applied since the future runoff time series do no exist.” makes little sense. As mentioned above, the only difference between this paper and others dealing with nonstationary distributions driven by covariates is the derivation of the relationship between PDF parameters and covariates (by (simplified) physical arguments rather than empirical relationships); however, in both cases the same inferential results (such as uncertainty assessment) can be applied. Unfortunately, there is a general tendency to falling in love with a particular method missing its analogies and relationships with other techniques missing the more general picture. This prevents to recognize that a method such as that proposed in this study is only a particular case of techniques already used and equipped with a set of tool which can be applied also in this case. So, please, perform some simple bootstrap exercise and show at least sampling uncertainty effects.

Bullettin 17b: The Authors’s answer does not match my comments; there should be a misunderstanding. Actually, I did not say that regionalization is anachronistic, but that the at-site analyses based on short time series are unreliable and somewhat anachronistic; the point is that when I talk about regionalization I mean techniques such as index-flow method where data from multiple sites are merged to form a unique longer sample under the hypothesis that spatial information can replace temporal information. “Regionalization” is different from “regional analysis” (i.e. visualization of at-site/local variability across an area), which is actually what is done in this study. In this respect, the study is somewhat anachronistic because overlooks the widely recognized unreliability of at-site estimates and omits a fair communication of the uncertainty (I mean the propagation of the sampling uncertainty affecting the PDF of the period of records into the future projections).

Nonstationarity and uncertainty: Again, I understand the Authors’ reply (which is quite common) but the point is different: using nonstationary models implies the identification of a deterministic trend (predictable with negligible uncertainty over a time window of interest), which in turn requires a deterministic attribution, and this is not the case in complex systems such as global climate dynamics. So, in this respect, dynamically-varying distributions provide pictures of “what if” scenarios under some given conditions (e.g. emissions) which are deterministic because they are imposed in the climate models simulations. So, my comments do not refer to nonstationary methods by themselves, but on their general use taking for granted the presence of “deterministic trends” that are often only stochastic and related to long range fluctuations of stationary processes.

Skepticism: it is related to the arguments above. Skepticism comes before hypotheses and axioms, as it has an epistemological role, meaning that it provides a rule to assess the suitability of hypotheses and axioms in order to retain sound hypotheses and discard

nonsense/wrong assumptions. Skepticism is exactly the rule that helps quicker advances avoiding nonsense theories and starting points leading forward but in the wrong direction.

Concerning my last remark in the previous report, the answer is vague and no very convincing. Every method can provide a forecast: stationary methods provide stationary forecasts, nonstationary methods nonstationary forecasts. For design purposes, the point is which one (if any) is the most credible and reliable based on the available information, in order to make a decision. Every method is useful in principle for practical applications, but it depends on whether it fits the problem at hand or it is used under wrong hypotheses. Again, in the present context, for practical applications, uncertainty and reliability should be assessed.

**Specific comments**

Although the Authors state that "the language issues were also checked", actually it seems that they did not, as the revised text shows several syntax/grammar errors as well as the use of inappropriate terminology, even in the abstract (e.g. "probability dencity functions", "correction shold be applied"). In general, almost all the paragraphs introduced in the new version show some error. In the following, I report only some examples as the manuscript requires a professional proofreading.

L3-13: This new paragraph does not describe correctly the actual situation. The point is not if changes in climate drive changes in runoff, as this is obvious. The point is that nonstationary models require that we are sure about the future changes or, in other words, that the changes are deterministic (predictable). If the deterministic evolution (nonstationarity) of the process is only hypothetical, it is evident that we have no idea of how it will evolve, and supposed trends or regime shifts can simply be local fluctuations of perfectly stationary processes. In other words, we have hypothetical projections, but we have no idea of the actual evolution of climate and so runoff. Therefore, tools for nonstationarity are surely interesting, but their suitability in this context is highly questionable because they involve an additional source of uncertainty related to the unknown evolution pattern. This stresses once again the importance of assessing the uncertainty and reliability of design values.
"We consider it improbable that changes"… do you mean something like "We consider that it is improbable that changes"?? Please reword.

P3L4: please consider "physically-based model described by dynamic equations"

P3L4: please consider "the meteorological signal could be simulated by random generators." Do you mean "weather generators"? By the way, as I can see "a-priory" several time throughout the text, it is worth recalling that it should be "a priori": it is Latin and means "in (a) advance (priori)". If you are not familiar with such a kind of expressions please avoid them and use plain English.
P3L14-15: "the parameters of PDF are directly *simulated* from the meteorological mean values" please use correct terms: simulated means generate by e.g. a MC procedure. In

this case, parameters are estimated as functions of covariates (meteorological mean values). Estimation and simulation have a different meaning and should be used in the correct context. “These parameters are further used to model PDF with theoretical distribution”??

P3L17: please replace “detrimental” with “extreme”

P3L34: “classical assumption used *behing* the *engeenering* applications” please consider “classical assumption used in engineering applications”

P4L34: “is calculated according method from”… do you mean “is calculated according to the method proposed (discussed) by (in)”

P4L36: Please consider “which reflects the water income to the catchment (due to melting) that affects the shape of hydrograph”

P5L23: “based on”

P5L28-30: what do you mean? Please reword in a more readable way

P6L21-22: “different statistically significant PDF parameter values” do you mean that the difference between the parameter values in the two periods is statistically significant?

P6L25: “which allows performing the model ability to reproduce the measurements. In simplest case…” -> “which allows one to assess the model ability to reproduce the measurements. In the simplest case…”

P6L27-29: “Then, the training set is used to evaluate the model parameters, which are further used to calculate the modelling (or nominally predicted) dataset to compare with the testing/control set using chosen measure (the statistical goodness-of-fit tests in our case).” What does it mean “to calculate the modelling dataset”? Please reword using appropriate terms to describe corresponding concepts.

P9L20: Chylek et al. (2011) was not accepted for final publication. Please provide a published reference.

P9L22: “the the”

P10L13: “do no”

P10L26: do you mean “especially”? Please reword the sentence.

P10L26: “to calculate the maximal runoff from tailed values with required probability of exceedance” -> “to calculate the yearly maximum runoff with required probability of exceedance”. Please use homogenous terms to describe the same object, avoiding e.g. “maximal runoff” for what is previously defined as “yearly maximum runoff” and brand

new terms like “tailed values” or “maximal extremes” to denote “extreme quantiles” or simply “extreme values”.

P13L18: “toevaluate”

Sincerely,

Francesco Serinaldi