Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-167-AC2, 2020

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Uncertainty analysis of the rate of change of quantile due to global warming using uncertainty analysis of non-stationary frequency model of peak-over-threshold series" *by* Okjeong Lee et al.

Okjeong Lee et al.

lover1804@nate.com

Received and published: 22 June 2020

Comment #1 :

The manuscript is overall well-written but needs to be checked for English thoroughly to increase readability. The reference list is appropriate and includes recent works on the topic. However, some improvements in the structure of the manuscript are needed, since some descriptions of the methodology are found in the discussion section. About the content of the manuscript, the analysis rely on just 2 stations in Korea, and only con-

siders one covariate (daily dew point temperature), consequently the whole reads like a technical report rather than a scientific contribution to HESS. The interesting aspect is the model evaluation that is not based on model fit but on the uncertainties on rainfall quantiles using a Bayesian framework. Overall the "reference DPT" concept used in the paper is interesting but not well defined and deserves a better explanation in the method section and discussion since it has practical interest for non-stationary analyses. I would recommend adding more stations (obviously there are enough stations in Korea for such an analysis = https://doi.org/10.1002/joc.2068) and more importantly compare different covariates commonly used in non-stationary frequency analysis of extreme rainfall in the context of Korea (see my comment below about the lack of description on how the covariate is selected and used). These two recommendations would increase the representativeness of the results but also provide regional insights for Korea. At the current state of the manuscript, the reader cannot know if these results are only valid for these two stations and with this covariate.

Response #1 :

Your detailed comments were very helpful in making a better manuscript. The authors would like to express great gratitude for this. First, let me tell you that some of the content in the discussion section has been moved to the methodology section. You can see this in the methodology section of the revised manuscript. Data from 11 sites, which began to be observed in 1961, were further analyzed. That is, a total of 13 sites were used in this study, including 2 sites that were previously applied. As a covariate, analysis was performed by adding surface air temperature in addition to the dew point temperature. The results of applying the added sites and an added covariate were prepared in the form of Supplementary Material and included in the revised manuscript. Also, as a figure showing the final result, Figure 7 of the revised manuscript was newly added. This further analysis may dispel concerns about whether the method proposed in this study applies only to two sites or is not valid only for dew point temperature. In addition, further analysis results will increase the representativeness of the results derived from this study and provide local insights into Korea. As you mentioned, because of the high level of practical interest in non-stationary frequency analysis, the concept of “reference covariate” has been described in more detail in the methodology and discussion sections. You can find out about this in the Methodology section and the Discussion section. More specific details of how and where the manuscript has been revised are described in response to the comments presented below.

==

Figure 7(a) shows the values of the negative log likelihood function of the stationary model and the non-stationary models at 13 sites. The stationary model, the SAT-based non-stationary model, and the DAT-based non-stationary model were found to have no significant difference in the fit performance with the observed POT excesses. Figure 7(b) shows the h-factor of rainfall quantile corresponding to the return level of 100-year. When all the values of covariate observed on the day of POT excesses are considered ("DPT" and "SAT" in Figure 7(b)), at all sites except Mokpo site, the non-stationary h-factor is greater than the stationary h-factor. However, when the reference covariate is applied, the non-stationary h-factor is smaller than the stationary h-factor. Results from 13 sites and most of the non-stationary models using SAT or DPT as a covariate indicate that how to determine the appropriate value of the covariate corresponding to the rainfall quantile plays an important role in securing the reliability of the non-stationary frequency analysis.

==

Comment #2 :

My first comment is about the title, quite long and not very informative of the main scientific results of the paper. Something like this title might be better: Uncertainty of non-stationary frequency analysis applied to extreme rainfall in Korea.

Response #2 :

Discussions with the authors are underway. The title so far determined is “Uncertainty in non-stationary frequency analysis of Korea’s daily rainfall POT excesses associated with covariates”. The results reflect the authors’ opinion that the title is appropriate, including keywords such as POT, coviariate, and non-stationary. The current title can be revised through communication with researchers.

Comment #3 :

Abstract, line 28, this sentence it not clear, maybe too general: “However, since the parameters of the estimated probability distribution contain a lot of uncertainty”. Abstract, line 40-42, this whole section below is quite trivial. Of course when a wrong covariate is selected in the POT model there is a stronger uncertainty. I don’t see a major finding here. Overall the abstract needs a major upgrading to better present the main findings of the study.

Response #3 :

In order to avoid describing general contents and to better express the main results of the study, the abstract has been greatly upgraded as follows:

==

Several methods have been proposed to analyze the frequency of non-stationary anomalies. The applicability of the non-stationary frequency analysis has been mainly evaluated based on the agreement between the time series data and the applied probability distribution. However, since the uncertainty in the parameter estimate of the probability distribution is the main source of uncertainty in frequency analysis, the uncertainty in the correspondence between samples and probability distribution is inevitably large. In this study, an extreme rainfall frequency analysis is performed that fits the Peak-over-threshold series to the covariate-based non-stationary Generalized Pareto distribution. By quantitatively evaluating the uncertainty of daily rainfall quantile estimates at 13 sites of the Korea Meteorological Administration using the Bayesian approach, we tried to evaluate the applicability of the non-stationary frequency analysis with a focus on uncertainty. The results indicated that the inclusion of dew-point temperature (DPT) or surface air temperature (SAT) generally improved the goodness of fit of the model for the observed samples. The uncertainty of the estimated rainfall quantiles was evaluated by the confidence interval of the ensemble generated by the Markov chain Monte Carlo. The results showed that the width of the confidence interval of quantiles could be greatly amplified due to extreme values of the covariate. In order to compensate for the weakness of the non-stationary model exposed by uncertainty, a method of specifying a reference value of a covariate corresponding to a non-exceedance probability has been proposed. The results of the study revealed that the reference co-variate plays an important role in the reliability of the non-stationary model. In addition, when the reference co-variate was given, it was confirmed that the uncertainty reduction of quantile estimates for the increase in the sample size was more pronounced in the non-stationary model. Finally, it was discussed how information on global temperature rise could be integrated with DPT or SAT-based non-stationary frequency analysis. It has been formulated how to quantify the uncertainty of the rate of change in future quantile due to global warming using rainfall quantile ensembles obtained in the uncertainty analysis process.

==

Comment #4 :

Page 3, line 74 : change “in many documents” by “by many authors”

Response #4 :

In response to your comments, we changed “in many documents” to “by many authors” as shown below:

==

However, natural environmental changes, such as global warming, have a serious im-

pact on the assumptions of the stationarity of the observations. Non-stationarity is an important issue that can never be ignored in areas related to drainage system design, as it can alter the design flood volume obtained using the stationary frequency analysis of observed rainfall extremes. The probability of occurrence of extreme rainfall events is expected to change due to global warming (Lee et al., 2016), and this change is called non-stationarity by many authors (Alexander et al., 2006; Gregersen et al., 2013).

==

Comment #5 :

Page 3, lines 99-102: these sentences are not very clear.

Response #5 :

The sentences have been revised as follows to clearly describe the sentences.

==

Efforts to develop and apply a non-stationary model for frequency analysis to reflect changing environmental conditions can be frustrated by the additional uncertainty associated with the model’s complexity, working with sampling uncertainty. In other words, the reliability of rainfall quantiles estimated by a complex non-stationary model may not be substantially improved, or when various environmental conditions are reflected, insufficient model reliability can easily lead to physically inconsistent results (Serinaldi and Kilsby, 2015). From this point of view, investigating which model has less uncertainty in rainfall quantile as a result of frequency analysis can be an important determinant in selecting an optimal model. This is because a model with a relatively smaller uncertainty in the estimated rainfall quantile can be regarded as a more reliable model.

==

Comment #6 :

Page 4, line 111, Why the authors are considering dew point temperature as a covariate for daily rainfall extremes for their stations in Korea ? Do previous works justify this choice ? No justification is given.

Response #6 :

Considering the dew point temperature as a covariate for daily rainfall extremes has been suggested in previous studies. We would like to establish the justification for selecting DPT or SAT as a covariate for rainfall extremes by briefly introducing two representative prior studies as follows:

==

In this study, a non-stationary frequency analysis using dew point temperature (DPT) or surface air temperature (SAT) as a covariate is performed. To obtain a necessary understanding of the relationship between daily rainfall and DPT and daily rainfall and SAT in Korea, two prior studies have been conducted (Sim et al., 2019; Lee et al., 2020). Sim et al. (2019) analyzed the effects of DPT and SAT on daily rainfall extremes. Their results indicated that even if there was some cooling effect in the event of summer rainfall (Ali and Mishra, 2017), daily rainfall extremes in Korea were very sensitive to DPT and SAT. Lee et al. (2020) presented a procedure for performing non-stationary frequency analysis using DPT or SAT as a covariate. They revealed that non-stationary frequency analysis using future DPT or SAT may yield more reasonable and persuasive projections of future rainfall extremes. The purpose of this study is to focus on the uncertainty of covariate-based non-stationary frequency analysis using DPT or SAT.

==

Comment #7 :

Page 5: line 130, “Daily rainfall depth of 0.1 mm or more was applied to the analysis,” is not very clear. Does this mean you consider daily rainfall lower to 0.1 mm =0 ? Is this

related to rain gauges uncertainties ? How the results can be impacted by this choice?

Response #7 :

The Korea Meteorological Administration considers precipitation as more than 0.1 mm per day to be considered an official precipitation day. This is related to rainfall measurement equipment of the applied sites. In fact, including records with daily precipitation of 0.1 mm or less has no significant effect on results. A description of this has been added to the revised manuscript as follows:

==

Since the Korea Meteorological Administration only recognizes precipitation recorded at 0.1 mm or more per day as official precipitation, daily rainfall depth of 0.1 mm or more was applied to the analysis in this study. An example of this wet threshold can also be found in Chan et al. (2016). In fact, the application of a wet threshold does not significantly affect the results of quantile regression.

==

Comment #8 :

Page 7, line 170: I guess you mean here instead the “scale” parameter.

Response #8 :

Your comment is correct. The scale parameter is correct. Thank you for fixing it right away. You can see the correction as below:

==

Although studies considering the non-stationarity of the threshold of the POT series have been conducted (Tramblay et al., 2012), in this study, the non-stationarity was given only to the scale parameters of the GP distribution as follows (Um et al., 2017):

==

Comment #9 :

There is no indication on how the covariate is included in the model; is it the Dew point temperature the same day of the extreme rainfall event? at the starting day of a rainfall event or its peak ? On the opposite, is it computed for the week, or the months before the event ? No information is provided here.

Response #9 :

Covariate is defined as the daily average DPT or SAT on the day POT excesses occur. This information was included in the revised manuscript and described in the lower part of equation (4) as follows:

==

Eq. (4) tells how the covariate DPT or SAT is included in the model. The daily averaged DPT or SAT observed on the day of occurrence of each POT excess is included in the scale parameter of the GP distribution as shown in Eq. (4) to construct the non-stationary GP distribution. That is, when α_2>0, the larger the DPT or SAT, the larger the scale parameter.

==

Comment #10 :

Page 12, line 327, “However, under the condition that DPT is not given in advance” not clear to me. Do you mean you draw randomly DPT values instead of the values corresponding to the days with extreme rainfall?

Response #10 :

In “under the condition that DPT is not given in advance,” our intention was to refer to the condition of extracting samples of the scale parameter α using all DPTs observed on the day the POT excess occurred. The manuscript has been revised as follows to convey clear meaning.

==

The h-factor of rainfall quantile corresponding to the return level of 100-year was calculated in two ways. First, under the condition that the reference DPT is given (i.e., when the reference value of DPT is applied), the h-factor of the non-stationary model is reduced by 37 % (at Busan site) and 28 % (at Seoul site) than that of the stationary model. However, under the condition that all observed DPTs corresponding to POT excesses are applied, the uncertainty from parameter estimation and the effects from extreme values of the covariate overlap, and the h-factor of the non-stationary model exceeds the h-factor of the stationary model. That is, if samples of the scale parameter (i.e., α) is made by combining all samples of the coefficients of the scale parameter (i.e., α_1 and α_2) and samples of all observed DPTs corresponding to each POT excess, the uncertainty of rainfall quantiles in the non-stationary model is greater than the uncertainty of rainfall quantiles in the stationary model. The amplification of the uncertainty in the non-stationary model is because, as can be seen from Eq. (4), samples of some extreme DPTs significantly dissipate the samples of the scale parameter of the non-stationary GP distribution. This can also be confirmed through the lower right figure of Figure 4(a) and (b). The width of the 95PPU of the scale parameter of the non-stationary model corresponding to the value of the individual DPT is not significantly different from the width of the scale parameter of the stationary model. However, when all observed DPTs corresponding to the POT excesses are involved in sampling of the scale parameter, it can be recognized that the range of the 995 PPU of the scale parameter of the non-stationary model is very wide.

==

Comment #11 :

Page 13: the whole page/paragraph is quite long and not very clear to me, it could be shortened to the main findings.

Response #11 :

This section has been rewritten as follows, short and clearly centered on the main results.

==

We want to note here the condition in which the value of the covariate is given. In the upper left figure of Figure 4(a) and (b), the stationary quantile has a single value, while the ensemble average of the non-stationary quantile shows various values depending on the value of DPT. In addition, the 95 PPU of the stationary quantile has a constant range regardless of the value of the covariate, whereas the 95 PPU of the non-stationary quantile has a relatively wider range depending on the value of the covariate (see upper right figure in Figure 4(a) and (b)). This is due to the covariate dependence inherent in the scale parameter of the non-stationary GP distribution, as mentioned before. That is, since the range of the ensemble of the non-stationary rainfall quantile is a result of additionally reflecting the extreme values of the covariate in addition to the parameter uncertainty, it is more likely to be formed relatively wider than the range of the ensemble of the stationary rainfall quantile. It should be noted, however, that the width of the non-stationary 95PPU for a particular covariate value is less than the width of the stationary 95PPU. In fact, since the covariate corresponding to each POT excess is a known value, the h-factor of the rainfall quantile corresponding to each POT excess can be obtained (see lower left figure in Figure 4(a) and (b)). Given the value of covariate, it can be recognized that the non-stationary h-factor is smaller than the stationary h-factor. That is, if the value of the covariate of the non-stationary model can be determined, there is a room to say that the non-stationary frequency analysis is better in terms of reliability than the stationary frequency analysis.

==

Comment #12 :

Page 14, line 366, this equation should be in the Method section.

Response #12 :

The given equation and its description have been moved to the methodology section. It can be confirmed from the part where Eq. (5) of the revised manuscript is located.

Comment #13 :

The beginning of section 4.1 is obviously not a discussion and should be in the results section

Response #13 :

The beginning of section 4.1 has been moved to the results section. You can see the shifted content after Figure 4 of the revised manuscript.

Comment #14 :

Page 14, lines 386-390, this information should be in the method section, we should not discover in the discussion section how the covariate was implemented in the model. see my comment above.

Response #14 :

The section you mentioned has moved to the Methodology section. How the covariate was implemented in our model is described in the description of Eq. (4) of the revised manuscript. Specific modifications are already included in the answers to previous queries.

Comment #15 :

The issue of setting a reference covariate to a given return level is an interesting aspect. I believe it is necessary to first analyze the response of extreme rainfall to different values taken by the covariate, and as mentioned here it is difficult to identify a single value of the covariate related to a high risk of extreme rainfall. Yet this aspect needs more discussion, I don't see an added value of randomly selecting a covariate from a pre-defined distribution (of the covariate).

Response #15 :

We also believe that the uncertainty analysis of randomly selecting a covariate from a predefined distribution of covariate is not feasible. The method of randomly selecting the covariate is implemented under the condition that all observed covariate samples corresponding to POT excesses are applied in this study. Therefore, this study investigated the response of covariate and rainfall quantile, and introduced the concept of reference covariate as an alternative. The first half of Section 4.1 of the revised manuscript was amended as follows:

======================================

As described above, when performing the uncertainty analysis of the non-stationary frequency analysis, an undesired disturbance in which the ensemble of rainfall quantile is excessively dispersed due to some extreme covariate values appears. Since the value of the covariate is the data observed on the day that the POT excess occurred (i.e., a deterministic variable), analyzing the uncertainty in rainfall quantile by randomly sampling the value of DPT or SAT from a predefined probability distribution of covariate is likely to result in overestimating uncertainty. We thought that the uncertainty analysis of randomly sampling the values of covariate from a predefined distribution of covariate was not feasible. The method of randomly sampling the value of the covariate in this study is implemented under the condition that all observed covariate samples corresponding to POT excesses are applied. Therefore, this study investigated the relationship between the value of covariate and rainfall quantile. From Eq. (5), the DPT value (i.e., reference DPT) of the non-stationary GP distribution that returns the rainfall quantile equal to the stationary GP distribution can be calculated (reference SAT can be calculated in the same way). Figure 6 shows an example of determining a reference DPT. The results of calculating the reference DPT at Busan and Seoul sites indicate that the reference DPT increases as the return level increases. The right figure in

Figure 6(a) and (b) shows the histogram of DPT corresponding to POT excesses. The distribution of DPT is slightly distorted to the left. It can be found that the reference DPT corresponding to various return levels at Busan and Seoul sites is similar to the location of the mode of the DPT distribution. This fact reveals that covariate values that deviate significantly from the reference covariate (i.e., some extreme values of the covariate) amplify the uncertainty of rainfall quantile from the non-stationary frequency analysis. From the results of regression analysis of rainfall quantile for various return levels and the corresponding reference DPT, the relationship of DPT = 18.8589RLˆ0.01555 (where RL is the return level in year and the unit of DPT is °C) was obtained at Busan site. At Seoul site, a relationship of DPT = 19.8540RLˆ0.01728 was obtained. The coefficient of determination of the regression analysis was 0.99 or higher at Busan and Seoul sites. From these results, the reference DPT corresponding to the return level of 100-year at Busan site could be applied to 20.2567 °C and Seoul site to 21.4958 °C. As shown in Figure 6 and Figures S3 and S4 of supplementary material, the value of the reference covariate is almost completely dependent on the return level. It should be noted that the return level and the reference covariate are proportional to each other at some sites, and are inversely proportional to other sites. This means that it is not easy to identify a single covariate value corresponding to a rainfall quantile. In this study, we tried to overcome the problem of random sampling of covariates by introducing the concept of reference covariate when estimating rainfall quantile estimation and its uncertainty from non-stationary frequency analysis based on covariate. From a practical point of view, how to set the value of the reference covariate may be an important research topic in the covariate-based non-stationary frequency analysis.

======================================

Comment #16 :

Page 16, line 423; is this not totally expected?

Response #16 :

A more detailed description has been added to the revised manuscript as below:

==

The data period required to achieve a certain level of h-factor can play an important role in optimal model selection. As in other regions, the observed data period in Korea varies widely from site to site. If the data period is short and there is no significant difference in performance (both in terms of goodness of fit and uncertainty) between the stationary model and the non-stationary model, it can be said that it is better to apply a stationary model with a relatively well-established methodology. However, in terms of uncertainty, if the value of the reference covariate can be well defined, the results in Figure 8 show that the non-stationary model can estimate rainfall quantile with the same level of uncertainty even with relatively shorter data periods. That is, when frequency analysis is performed using samples of the same data period at a site, if the appropriate covariate is applied and the reference value of the covariate is appropriately determined, it can be said that the rainfall quantile estimated from the non-stationary model is more reliable than the rainfall quantile estimated from the stationary model.

==

Comment #17 :

Conclusions, lines 498-501, these sentences are elements of context and should not appear here but rather in the introduction.

Response #17 :

In the revised manuscript, the sentences were moved to the beginning of the introduction.

Comment #18 :

Table 2: units?

Response #18 :

The numbers in Table 2 refer to m-factor and h-factor defined in Eqs. (9) and (10), respectively. The units of m-factor and h-factor are dimensionless. It was mentioned that it was dimensionless in the introduction of Eqs. (9) and (10). Also, what the numbers in Table 2 mean was more clearly described in the text referring to Table 2.

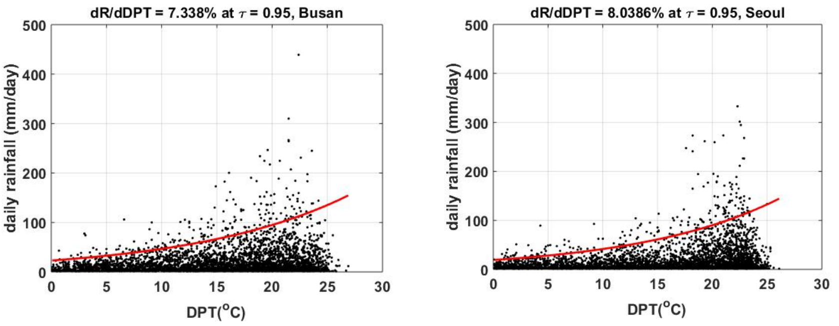
==

In this study, the following dimensionless quantitation factors were defined to quantify the uncertainty between the stationary and non-stationary models: Table 2 shows the results at Busan and Seoul sites. For reference, the results of applying DPT or SAT as a covariate at other sites are shown in Table S2 of the supplementary material.

==

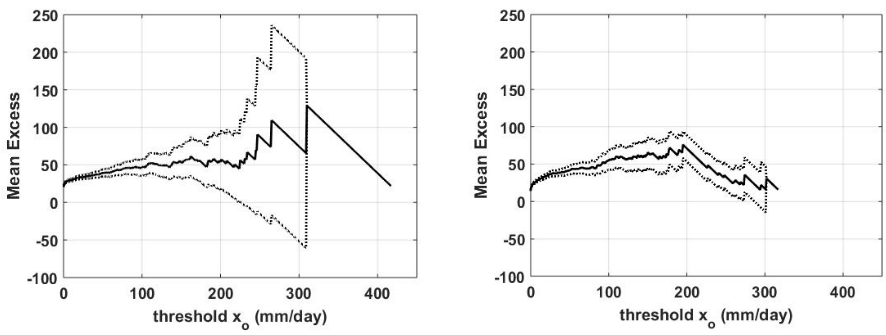
Please also note the supplement to this comment:
https://www.hydrol-earth-syst-sci-discuss.net/hess-2020-167/hess-2020-167-AC2-supplement.zip

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-167, 2020.



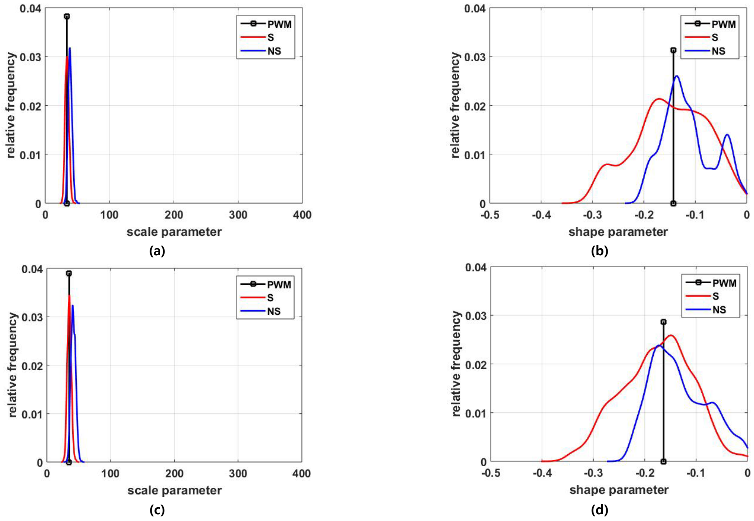
(a) (b)

Fig. 1. Sensitivity of 95 % daily rainfall depth to dew-point temperature at (a) Busan and (b) Seoul sites



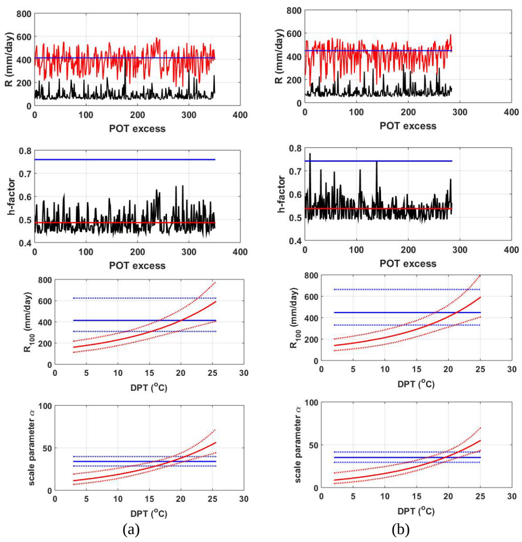
(a) (b)

Fig. 2. Mean residual life plot at (a) Busan and (b) Seoul sites.



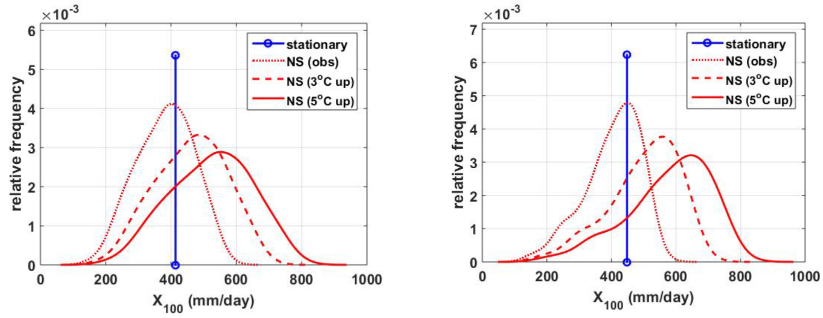
. (a) Scale and (b) shape parameters at Busan site, and (c) scale and (d) shape parameters at Seoul site. The black vertical lines are a parameter calculated by PWM, which is expressed as a single value. The posterior distribution of parameters for the stationary GP distribution sampled using the MH algorithm is indicated by red lines. The posterior distribution of parameters for the non-stationary GP distribution is indicated by blue lines. The scale parameter of the non-stationary GP distribution using covariate is defined as a function of DPT. Therefore, the posterior distribution of the scale parameters were derived on the assumption that DPT was given at 20.2567 ℃ (Busan site) and 21.4958 ℃ (Seoul site), respectively.

Fig. 3. Posterior distribution of parameters of stationary and non-stationary GP distribution



The upper left figures in Figure 4(a) and (b) show the POT series (blank line), and the ensemble average of stationary (blue line) and non-stationary (red line) rainfall quantile corresponding to the return level of 100-year. In the upper right figures, the ensemble average (blue line for stationary model, and red line for non-stationary model), and 95PPU of the stationary (blue dotted line) and non-stationary (red dotted line) rainfall quantile for the return level of 100-year are shown. The lower left figures show the h-factor of the stationary (blue line) and non-stationary (black line) rainfall quantile corresponding to the return level of 100-year. Red lines mean the average of black line. The lower right figures show the ensemble average (blue line for stationary model, and red line for non-stationary model), and 95PPU of the stationary (blue dotted line) and non-stationary (red dotted line) scale parameter.

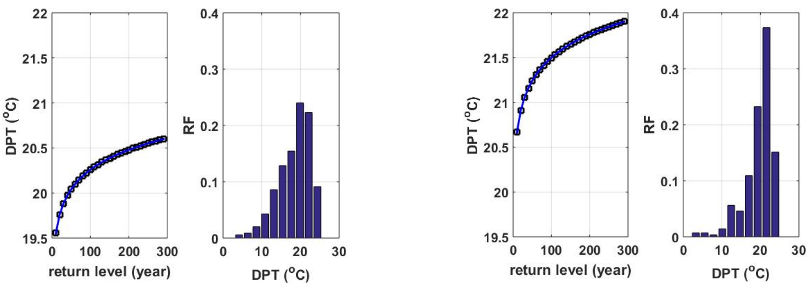
Fig. 4. Changes in uncertainty for co-variate at (a) Busan and (b) Seoul sites.



(a) (b)

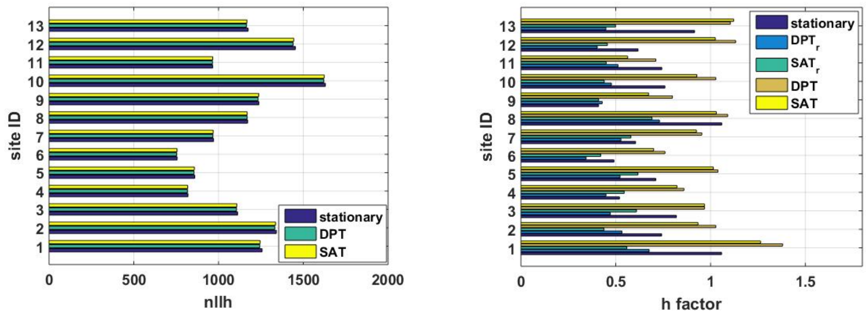
The stationary rainfall quantile is indicated as a blue vertical line since it is a single value. The non-stationary rainfall quantiles were calculated using the average of the parameter ensemble sampled by MCMC and the DPT observed on the day of POT excesses (red dotted line). In this figures, 'NS (3 ℃ up)' is an empirical distribution of rainfall quantile derived using DPTs that add 3 ℃ to DPTs observed on the day of POT excesses. Likewise, 'NS (5 ℃ up)' is an empirical distribution of rainfall quantile under the scenario condition where DPT has risen 5 ℃ due to global warming.

Fig. 5. Rainfall quantile estimates at (a) Busan, and (b) Seoul sites for return level of 100-year using observed dew-point temperature and global warming scenarios



(a) (b)

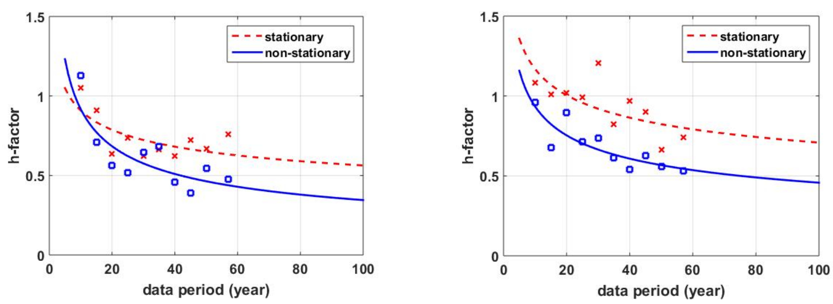
Fig. 6. Selection of reference dew-point temperature for estimating rainfall quantiles at (a) Busan and (b) Seoul sites



(a) negative log likelihood (b) h factor

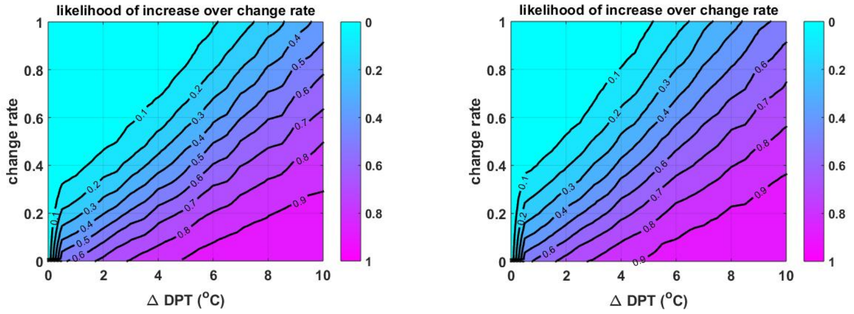
Figure 7. Performance of stationary and non−stationary frequency analysis models. At Site ID, 1: Ghangreung, 2: Seoul, 3: Incheon, 4: Chupungryeong, 5: Pohang, 6: Daegu, 7: Jeonju, 8: Ulsan, 9: Ghwangju, 10: Busan, 11: Mokpo, 12: Yeosu and 13: Jeju site.

Fig. 7. Performance of stationary and non-stationary frequency analysis models



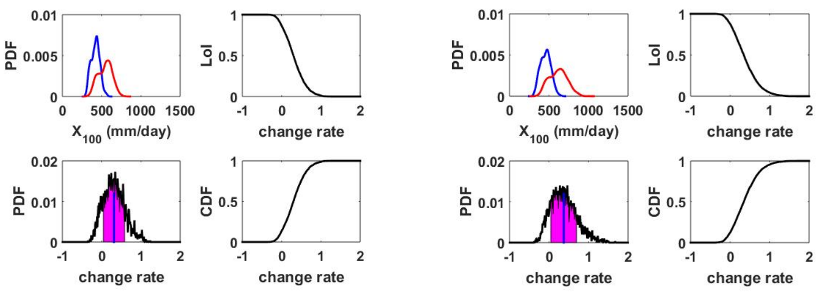
(a) Busan (b) Seoul

Fig. 8. Effect of the number of samples on the uncertainty of rainfall quantile using reference dew-point temperature



(a) Busan (b) Seoul

Fig. 9. Likelihood of increase over change rate of rainfall quantile for return level of 100-year.



(a) Busan (b) Seoul

Procedure for analyzing uncertainty in rate of change. In upper left figures, the blue line is the probability distribution of X_p^T in the present condition, and the red line is the probability distribution of X_f^T in the DPT 4 ℃ rising condition. In the lower left figures, the section of the standard deviation was colored in pink.

Fig. 10. Procedure for analyzing uncertainty in rate of change