Recommendations: reject.

This article is too long, very difficult to read, and has almost convinced me that Simple Scaling does work: e.g. Figure 3i and Figure 7, the scaling coefficient H is all over the places, whereas one would expected it to be constant over different durations, being a scale-invariant parameter. An other example is my last comment (just citing your text). Also, it is not clear to me why the authors chose the Bukovsky regions for their regional analysis, when they have a clerar clustering in their Figures 4 and 5 (and S1) which suggests a different regionalization. Moreover, the physical interpretation in Section 5 is not clear (how do you link the behaviour of H across different durations with the topography, as an example). Often the exposition does not follow a logical order. I doubt several results in Section 6. My overall comment is that the authors dwelve too much into the details, but they fail to convey the main message (at least to me …). I regret to say that my recommendation is a rejection: I leave however the decision to the Editor, since I am not sure I have fully understood the study.

My numerous comments:

Equation 2 (and through out the article): I suggest changing the notation by replacing above the equality sign (=) the letter “d” with “pdf”. This avoid confusion with “d” denoting the duration.

Page 5, the sentence at line 12-13-14 should be stated earlier, after line 9.

Page 5, lines 24-25: The equalities at line 24 pertains a dilatation of the (values of the) GEV distributions by a factor lambda, whereas Equation 2 pertains to a change of distribution due to a sampling on different durations, with a dilatation factor = lambda. The implication stated at line 25 (the GEV family satisfies Equation 2) is not a direct consequence of the equality at line 24. Rephrase (you might have to add a specific reference, or show explicitely the implication).

Equation 7: I would still prefer you replace d with lambda (since the protagonist here is the dilatation factor)! Here and throughout the article, wherever consistency is required.

Page 6, line 19: I appreciate your clarification wrt my previous comment. In light of that explanation, I suggest adding at the end of this sentence “ .. from May to October in the south, and from June to September in the north. Specifically, the 'year' from which the annual maxima was sampled, was limited to the observed summer season, and was defined as ... ”.
Page 7, line 4: write “ … at least 85% of valid observations for each summer season, otherwise … ”

Page 6, line 23: substitute 'cheked' with 'quality controlled'.
Page 7, line 13: typo: “finstance”.
Page 7, line 30: use the new notation 1h30 rather than 1.5h …

Page 7, line 23-24: you need to better state what MSA does. Line 23 is fine, “for intensity AMS” is too concise: you could write something like “for inferring annual maxima / extremes / GEV parameters for durations not sampled by using SS on sampled durations” or something similar.
Page 7, line 27: write *“scaling intervals”* in italic, so that the reader understand this is a definition (valid from hereafter throughout all the article).

Section 4:
I appreciate the efforts of the authors in better explaining this section (both in the text and responses to the previous revisions) and for providing very specific references. The section is indeed more clear. I

do have however still some difficulties in understanding, and I strongly believe that each article should be self-contained (to a certain degree), so here are my suggestions:

a) I do still believe that the article will gain in adding a figure showing i) the linear regression between log(duration) and log(moments) and ii) the (previously obtained) regession coefficients Kq and q (as Figure 4a in Panthou et al, 2014). These figures were fundamental for me to understand page 8, lines 8 to 16 (i.e. what you do in the MSA regression and in the slope test).
b) In the text regarding the MSA regression, you need to add a sentence which explains that the slope Kq that you are evaluating is equal to -Hq (it took me a while to figure this out), and that this is from Equation 4. Also, at line 8 write “for each q … ”, so that it is clear that you have a Kq for each individual separate q.
c) In the slope test text (page 8, lines 12-16) you need first to say that you aim to find the scaling coefficient H, which is the slope of the line you obtain while regressing Kq versus q. Then you say that you use a OLS regression and estimate H, and you call such estimate Beta_1 (I would call it \hat{H}). Finally you can say that you perform a t-test on the regression. I suggest stating that the null hypothesis was \hat{H}=K_1, and that if this null hypothesis is not rejected then you set H = \hat{H} = K_1. I would avoid whiting the relation Kq=Beta_0+Beta_1 q (otherwise the reader will ask where Beta_0 is gone, for the equality of K_1= Beta_1).
d) The Goodness of Fit test (page 8 lines 19-26) can be explained more clearly, following the cronologiocal order of the calculations you perform: first, you consider the AMS for each duration dj, AMSdj=$\{x_{dj,1},x_{dj,2}, \dots x_{dj,n}\}$, and you rescale it as dj^H \cdot AMSdj, which can be considered as a sample of the reference duration (the choice of notation x'_{dj} is confusing, since one would think it is a sample for the duration dj, whereas it is a sample for the reference duration). You then pool together these rescales samples for all dj. Then you invert equation 2 and obtain, from this large sample, a sample for each duration dj, under SS hypohesis.
e) I suggest performing the GOF test also in a cross validation way (e.g. for the ID dataset, when you test the duraton of 3h, you rescale to the duration of 1h all durations excluded the 3h; then you invert Eq.2 and you apply the test to the obtained -slightly smaller- samples).
f) Page 9, lines 21-23: this is not entirely clear, do you repeat all steps associated to the GOF part only (page 8, line 7 onwards), or also for the MSA regression and slope test? Only after reading page 11 line 1 I understood that you repeat all (points 1,2,3 at page 8). Make the sentence clearer.
g) Text from page 9 line 27 to page 10, line 4: similar to what suggested for the GOF test, describe what you calculate in order of calculation, i.e. define first the normalized RMSE for each station (Eq 11) and after the average over all stations (Eq 10).

Section 4.1:

Figure 1 shows results of slope test and GOF test together. However a reader is intrigued in disentageling the two. You have actually looked at the results seperately, and decided to put them together, with the knowledge that solely the GOF test has a signal. The reader, however, does not know this and remains in the doubt up unti reading at the end of page 10 (lines 25-27). I suggest you to move the sentence at lines 25-27 at the very beginning of the description of these results, after line 10. After this sentence, you should state that the differences in station rejections for the separate durations as shown in Figure 1 are due to the results of GOF test only. And then you keep describing Figure 1 as in your lines 11-25 (which I assume refers solely to the GOF test: this should be made more explicit). Eliminate the [not shown] at line 13 (I think you show this, with the SD dataset)

Figure 2: Page 11, lines 2-4: rewite this as “On the other hand, the extrapolation under SS of the Xd distribution is generally less accurate for durations at the boundaries of the scaling intervals (especially

for the short durations)". You do not show/perform extrapolations of Xd by estimating H with durations outside the scaling interval (as far as I can see), and I believe you meant to rewrite the sentence deleted at line 5.

Figure 2: I would be curious to see the slope and GOF test results also for the cross validation experiment (as in Figure 1, to be able to compare them). This actually is related also to my previous comments e) and f) for Section 4. In alternative, you could reproduce Figure 2 for the non-cross-validation calculation.

Section 4.2

After your introductory sentence (page 11, lines 10-12) I suggest describing first the results pertaining to ΔH (Figure 3 ii, iii, iv) and after the results pertaining to the spatial variability of H (Figure 3 i, and Figures 4 and 5). My suggestions for improvement are:

- join the paragraph at lines 13-16 with the paragraph at lines 20-25 (page 11).
- Eliminate (move) the sentence at page 11 lines 26-27 to page 12 line 11, where you will start the description of the spatial variability of H.
- The text at lines 29-33 is difficult to follow, give (for each sentence) a precise reference to the figure panels.
- You might want to discuss first the results at page 12 lines 3-10 (which are positive, showing near zero ΔH) and after the results at page 11, line 29 to page 12 line 2 (which are more detailed and negative).
- Eliminate lines 13-17 and the related sentence at lines 30-31 of page 12 (this is too technical and does not add meat to the article, but rather distracts the reader)

Section 5

Figure 4 and 5 (and Figure S1) show clear spatial clustering in the behaviour of H (as commented in my previous revisions): I still think that the authors should apply a cluster analysis to their own data. Climatology and extremes are different, and extremes might not follow the Bukovsky regions. You can develop a spatial model for SS and IDF estimation (as you state at page 15, lines 16-18) solely considering a regionalization based on the extreme behaviour (rather than pre-set regions).

page 13, lines 9-23: I am not sure the two statistics described here add too much to the article (unless they help the physical interpretation of Section 5.1, which is obscure to me -see following comment-): in fact, I believe that their behaviour is expected, from their definition: shorter duration have a larger number of events, which decay the longer is d1 (S4); conversely, the mean wet time per event decay as d1 grows (averaged on longer duration) … I suggest moving all this to the supplementary material (also the related text at page 14 lines 7-8, 10-13, 23-32). The implications at page 14, line 13 is not clear.

Section 5.1

From page 13 line 30 to page 14 line 3: The link between the behaviour of H and the physical characteristics of the precipitation in the region is missing / not clear. Similarly, at page 14 lines 4-7, the implication is not clear at all.

Maybe you need to explain <u>beforehand</u> what does it means (physically) when H is small and when H is large (as at page 15, lines 13-15), when H increase and when H decreases with d1.

Page 14, lines 14-22: good interpretation of the behaviour in Region D.
Page 15 lines 2-5: clear, whereas you loose me at lines 6-9.

page 15, lines 10-11: I disagree with this sentence, I have not seen any evidence that the material illustrated here supports these results (you have not proven this).

Section 6

From page 15 line 30 to page 16 line 7: since you are assessing a distribution, why don't you use a KS statistics, rather than inventing an engeneered metric which compares the quantiles? Recall Equation 11 for clarity, please.

Page 16, line 13: for the ID and LD datasets, the behaviour of the SS parameters versus non-SS parameters is opposite, they cannot be both more right skewed for the SS estimation.

Page 16, line 16: it seems to me that the discerepancies between SS and non-SS parameters for the LD dataset and long durations (the $\Delta\mu$ and $\Delta\sigma$ in the supporting material) and quite big.

Page 16, lines 14-21: there is something strange about your results: you state that the estimation of the scale parameter has a small uncertainty when the shape parameter is correctly estimated. However, from Figure 8. I can see that the shape parameter is not correctly estimated (red and black curves do not match). What is the implication of the mismatch of the shape parameter for the non-SS and SS estimation on your results? The shape parameter is usually set to zero because it is all over the places, and often not-significantly different from zero (as you find in your Figure 9c: the estimated shape parameter is quite noisy)!!!

Page 16, lines 22-23: I agree with this sentence, nice spatial coherence.
Page 16, lines 25-26: why SS shape parameter is nearer to zero and exhibit less spread? Is this an effect of the assumptions of SS?
Page 16, line 27: very difficult sentence to read (essentially the shape parameter is zero).

Page 16, lines 27-31: these results are not intuitive. Figure 10 suggests that the shape parameter for the non-SS estimates is predominantely zero, whereas for the SS estimates there is a large proportions of positive and negative shape parameters. However, the right column of Figure 8 shows exactely the opposite (SS estimate are nearer to zero than non-SS estimates). I assume the difference is due to the very different width of confidence interval associated to the estimate of the shape parameter, for non-SS and SS samples. Are the sample sizes very different (I believe so … given that x* in equation 8 is obtained from a very large sample). Then maybe these results are artificial … (as you conclude afterwards, at page 17, lines 3-5: but this link is not explicit)!!!
Page 17, lines 7: eliminate (not clear what this refer too).

Overall, the text starting at page 16 line 27 and ending at page 17 line 10 should be all reorganized ina more coherent single paragraph.

Page 17, lines 12-15: Figure S13 shows that for shape parameter equal to 0 (the majority of the stations) the error of quantiles estimated by the non-SS is smaller than that for SS estimates.