Response to Reviewer 2

Patrick Pieper, André Düsterhus, and Johanna Baehr

June 6, 2020

We thank Gabriel Blain for the effort of reviewing our work. Your comments have been very helpful in improving our manuscript. Below we answer point-by-point to each of your comments and explain how the respective comment helped us to improve the manuscript. Your comments are printed in black and our responses are printed in blue. Line numbers in our response refer to the initially submitted manuscript.

One of your comments concerning the complexity punishment by our employed information criterion caused us to reconsider our storyline. This reconsideration does not alter our conclusions. Yet, it simplifies for us to conclude which helps readers to follow our conclusions.

General comments

The manuscript "Global and regional performances of SPI candidate distribution functions in observations and simulations" proposes and new methodology to select candidates distributions for calculating the SPI; a widely used standardized drought index. The study is interesting and adds important information to the SPI literature because it evaluates the advantages and shortcomings of previous methodologies designed for the same purpose. It is also well written. So, it should be considered for publication.

Thank you for these kind comments and your endorsement.

Specific comments

L.105 The Shapiro-Wilk [...] "is unreliable to evaluate SPI normality (Naresh Kumar et al., 2009)". This is a very important statement, which now I tend to agree with. Please, provide further information regarding it.

First, we are pleased that we were able to convince you. Second, we thank you for pointing out this lack of depth in our introduction.

Goodness-of-Fit (GoF) tests are ill-suited to assess the normality of SPI time-series, because of their spatial aggregation in combination with their binary convention. To fully understand this interplay we start with SPI's calculation

procedure: (i) fit a candidate PDF onto precipitation, then (ii) Z-transform the fitted probabilities to SPI values. Because the choice of an appropriate candidate PDF is the key decision in this process, the initial fit of the candidate PDF onto precipitation should be scrutinized. GoF tests, however, measure the normality of the resulting SPI values. In theory, this switch of focus in the analysis only complicates its structure but should not impact its outcome: if the candidate PDF's fit is appropriate, then its estimated probabilities are appropriate. Thus, their exact equiprobability transformations to the standard normal variable Z are also appropriate.

Anyhow, this complicated structure blurs the view on the measure of interest: the fit of the candidate PDF (onto precipitation). Therefore, the following caveat easily arises unnoticed and is, thus, not properly dealt with. After losing sight of the actual measure of interest (the fit of the candidate PDF), the focus lays on the normality of SPI time-series. The intuitive tool to assess normality leads to GoF tests. The drawback of GoF test is the biased discrimination between the tails and the center of the distribution. GoF tests equally evaluate each value that contributes to the distribution. Such an evaluation assigns more weight to the center and almost no weight to the tails of the distribution. Yet, appropriately fitting the tails of precipitation distributions should logically be of paramount importance to any sensible candidate PDF employed in SPI's calculation algorithm (see also our argument against weighting deviations from $\mathcal{N}_{0,1}$ by the theoretical occurrence probability of the respective class in lines 244 to 251). But the complicated structure blurs the view from this consideration. Instead, GoF tests conveniently present an allegedly easy solution.

As seen in Fig. 1, deviations from $\mathcal{N}_{0,1}$ are smallest in the center and largest in the tails of the distribution. Candidate PDFs typically fit precipitation better for the center than for the tails of the distribution: its center counts more samples which translates to more weight in the optimization (e.g. by the maximum likelihood estimation). This behavior deludes GoF tests in the analysis of SPI normality. That delusion obscures the tails of the distribution from GoF tests. Nevertheless this delusion, despite this obscurity surfacing skepticism about the proper depiction of the tails of the distribution can still aggregate over many grid-points. This aggregated skepticism can still lead to a robust analysis if evaluated relative to the similarly obscured performance of other candidate PDFs (as shown by metrics such as AIC-D, and BIC-D). Anyhow, additionally aggravating for GoF tests is their convention to be interpreted binarily. As a consequence of this convention, SPI literature typically aggregates results of GoF tests over domains by counting rejections. This typical aggregation prevents surfaced skepticism to fully aggregate over many grid-points. The interplay of both caveats, the blurred tails of the distribution and the prevention of remaining skepticism to fully aggregate, leads to the conclusion that GoF tests are ill-suited to assess SPI normality. I.e. it is (admittedly more obvious but) similarly inept to round normally distributed ($\mathcal{N}_{0.1\pm\epsilon,0.1}$) variables to their nearest integer before calculating their mean to estimate ϵ.

This full explanation is too extensive for the scope of the introduction of

our publication. However, we do admit that only indicating problems with the binary nature of GoF tests and hinting at issues with their spatial aggregation might cut the story too short. To rectify this shortcoming, we split the paragraph (lines 106 -117). This allows us to elaborate on GoF tests (in)ability to evaluate SPI candidate distribution functions: "(...) which in turn is unreliable to evaluate SPI normality [Naresh Kumar et al., 2009].

The above-mentioned goodness-of-fit tests equally evaluate each value of SPI's distribution. Such an evaluation focuses on the center of the distribution because the center of any distribution contains per definition more samples than the tails. In contrast, SPI usually analyzes (and thus depends on a proper depiction of) the distribution's tails. Therefore, a blurred focus manifests in these goodness-of-fit tests. Moreover, the convention to binarily interpret the above-mentioned goodness-of-fit tests aggravates this blurred focus. Because of this convention, these goodness-of-fit tests are unable to produce any relative ranking of the performance of distribution functions for a specific location (and accumulation period). This inability prevents any reasonable aggregation of limitations that surface despite the blurred focus. Thus, they are ill-suited to discriminate the best performing PDF out of a set of PDFs [Blain et al., 2018]. For SPI distributions the question is not whether they are (or should be) normally distributed (for which goodness-of-fit tests are well suited to provide the answer). The crucial question is rather which PDF maximizes the normality of the resulting SPI distribution. Because of the ill-fitting focus and the ill-suited convention of these goodness-of-fit tests, they are inept to identify SPI's best-performing candidate distribution function out of a set of PDFs.

In agreement with this insight, those studies, that rigorously analyzed candidate distribution functions, or investigate an appropriate test methodology for evaluating SPI candidate PDFs, consequently advocate the use of relative assessments: (...)"

While elaborating on the methodology to test the normality of SPI time-series, we realized a missing differentiation between the analysis of AIC-D frequencies and the analysis of deviations from $\mathcal{N}_{0,1}$ in the initial submission. The fact that both analyses complement each other comes a bit too short. Thus, we also rectified this shortcoming through the following changes to the manuscript:

We substituted a sentence from the abstract in lines 6 to 7 by: "Our normality comparison bases on a complementary evaluation. Actual against theoretical occurrence probabilities of SPI categories evaluate the absolute performance of candidate distribution functions. In contrast, Akaike's information criterion evaluates candidate distribution functions relative to each other while analytically punishing complexity. SPI time-series (...)"

We added another paragraph at the end of section 2.5 in between lines 293 and 294 which reads as follows: "The analysis of deviations from $\mathcal{N}_{0,1}$ assesses performances of candidate PDFs in absolute terms irrespective of the candidate PDF's complexity. In contrast, the AIC-D analysis evaluates the performance of candidate PDFs relative to each other while analytically punishing complexity.

Consequently, the AIC-D analysis cannot evaluate whether the best-performing candidate distribution function also performs adequately in absolute terms. In opposition, deviations from $\mathcal{N}_{0,1}$ encounter difficulties when evaluating whether an increased complexity from one PDF to another justifies any given improvement. Both analyses together, however, augment each other complementary. This enables us to conclusively investigate: (i) which candidate PDF performs best while (ii) ensuring adequate absolute performance and while (iii) constraining the risk of over-fitting."

We substituted three sentences in a paragraph of section 3.1.1 (lines 401 to 405) by: "It is noteworthy, that investigating deviations from $\mathcal{N}_{0,1}$ over the entire globe contains the risk of encountering deviations that balance each other in different grid-points with unrelated climatic characteristics. Until dealing with this risk, our analysis of deviations from $\mathcal{N}_{0,1}$ only indicates that three candidate PDFs (GD2, GGD3, and EWD3) display an adequate absolute performance. On the one hand, we can reduce that risk by analyzing deviations from $\mathcal{N}_{0,1}$ only over specific regions. This analysis safeguards our investigation by ensuring (rather than just indicating) an adequate absolute performance around the globe and is performed later. On the other hand, we first completely eliminate this risk by examining AIC-D frequencies: aggregating AIC-D values over the entire globe evaluates the performance of PDFs in each grid-point and normalizes these evaluations by (rather than adding them over) the total number of grid-points of the entire globe. We investigate AIC-D frequencies first to evaluate whether GGD3 and/or EWD3 perform sufficiently better than GD2 to justify their increased complexities."

We added another paragraph at the end of section 3.1.1 (in between lines 475 and 476): "Among our candidate PDFs, EWD3 is obviously the best-suited PDF for SPI. Yet, we still need to confirm whether also EWD3's absolute performance is adequate. While the global analysis indicated EWD3's adequateness, the ultimate validation of this claim is incumbent upon the regional analysis."

We added another paragraph at the end of section 3.1.2 (in between lines 514 and 515): "The analysis of AIC-D frequencies proves that EWD3 is SPI's best distribution function among our candidate PDFs. Additionally, the regional investigation confirms the global analysis: the absolute performance of EWD3 is at minimum adequate in observations and ensemble simulations."

The Bayesian information criterion (BIC) is similar to the AIC. However, the BIC uses a different penalty for the number of parameters [ln(n) k]. Can the authors verify if the BIC leads to similar results as those of the AIC.

We thank you for this exciting idea. Whether we use AIC or BIC to punish candidate PDFs for their complexity does not change our conclusions. Most of our drawn conclusions from AIC-D frequencies bases on the behavior of candidate PDFs' coverages for AIC-D_{max} values larger than 10 (right edge of Figure 5). These conclusions are then substantiated by candidate PDFs' coverages for AIC-D_{max} values larger than 7. These coverages (for AIC-D_{max}/BIC-D_{max}

Table I. Complexity penalty of candidate PDFs assessed with AIC and BIC.

Information Criterion	**AIC**		**BIC**		**Difference** BIC-AIC	
Realization	Obs. (N=31)	Sim. (N=310)	Obs. (N=31)	Sim. (N=310)	Obs. (N=31)	Sim. (N=310)
2-param. PDFs	4.43	4.04	6.87	11.47	2.44	7.43
3-param. PDFs	6.89	6.08	10.3	17.21	3.41	11.13
Difference 3-2 param.	**2.46**	**2.04**	**3.43**	**5.74**	**0.97**	**3.7**

values ≤ 7) are insensitive to the magnitude of changes caused by altered complexity penalties (Table I)

What impacts our analysis is not the absolute, parameter- and sample size-dependent punishment of candidate PDFs (values in the center of Table I). Instead, only the penalty difference between 2- and 3-parameter PDFs that base on the same sample size matters (evaluate observations and simulations isolated in the last row of Table I).

Similar, altering the information criterion (from AIC to BIC) impacts our analysis through the penalty difference between BIC and AIC (last column of Table I). Here, the difference between 2- and 3-parameter PDFs that base on the same sample size matters again (evaluate observations and simulations isolated in the bottom- and rightmost cell of Table I). I.e. the additional margin by which 3-parameter PDFs need to further outperform 2-parameter PDFs in order to still be considered as better by the new information criterion. This margin (bottom- and rightmost cell in Table I) increases in observations (simulations) by 0.97 (3.7) when using BIC instead of AIC.

The robustness of our conclusions stems from the robustness of the candidate's coverages for large AIC-D_{max}/BIC-D_{max} values (≥ 7). In this AIC-D_{max} regime, the candidate PDFs' coverages are sufficiently robust concerning changes caused by altered complexity penalties (Fig. I). Comparing 2- against 3-parameter in Fig. 5 with AIC-D or BIC-D does not substantially change the evaluation of large AIC-D_{max}/BIC-D_{max} values (≥ 7). As a first-order approximation, we can compare in observations the coverages of 2-parameter PDFs at the AIC-D_{max} value of 7 against the coverages of 3-parameter PDFs at the AIC-D_{max} value 7.97 (we shift the line indicating the coverages of 3-parameter PDFs by 0.97 units to the right). Since the slope of that line is sufficiently flat, this shift does not impact the conclusions for large AIC-D_{max} values (≥ 7).

In observations (simulations), coverages of 3-parameter PDFs are highly sensitive to the change of the information criterion at AIC-D_{max}/BIC-D_{max} values smaller than approximately 4 (6) (compare in Fig I the top row against the bottom row). The first-order approximation outlined before (shifting the coverages of 3-parameter PDFs by 0.97 (3.7) units to the right in observations (simulations)), describes the changes caused by using BIC (instead of AIC) quite well. The shifted coverages of 3-parameter PDFs exhibit slope-dependent

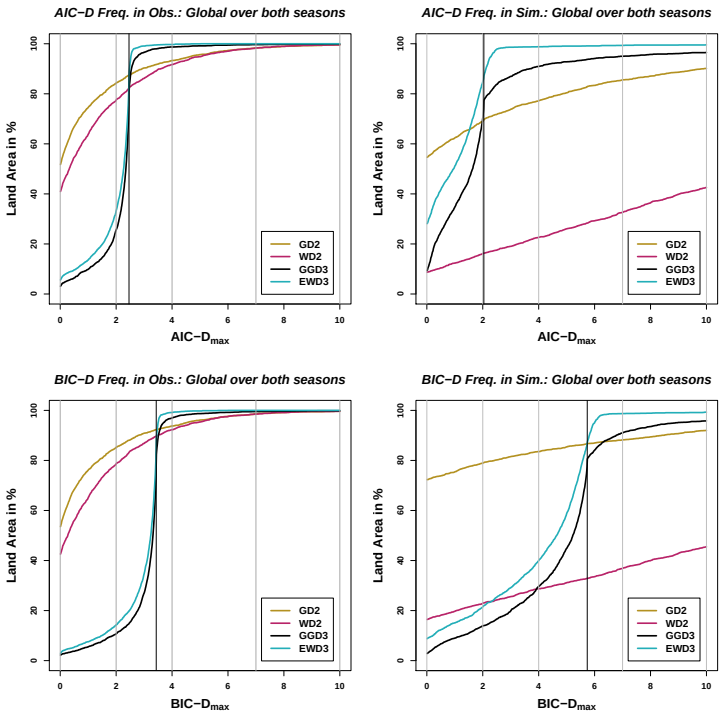

Figure I. AIC-D (**top**) and BIC-D (**bottom**) frequencies: percentage of global grid-points during both seasons in which each PDF yields AIC-D/BIC-D values that are smaller than or equal to a given AIC-D$_{max}$/BIC-D$_{max}$ value. The vertical black line indicates the different complexity penalties between 3- and 2-parameter PDFs (see bottom row of Table I). AIC-D/BIC-D frequencies are displayed for each candidate PDF for observations (**left**) and simulations (**right**).

changes at all AIC-D$_{max}$/BIC-D$_{max}$ values. That causes 3-parameter PDFs to be best-suited (AIC-D$_{max}$/BIC-D$_{max}$ value of 0) in fewer grid-points. In each grid-point, a single PDF must still be best-suited. In a second-order approximation, the coverages of 2-parameter PDFs, thus, slightly adjust for small AIC-D$_{max}$/BIC-D$_{max}$ values to the changes of 3-parameter PDFs' coverages at the AIC-D$_{max}$/BIC-D$_{max}$ value of 0. Consequently, the coverages of 2-parameter PDFs are overall fairly insensitive to the change of the information criterion because they only adjust slightly. The coverages of 3-parameter PDFs are more sensitive to the changed information criterion because they universally exhibit a horizontal shift.

This shift, however, does not result in a universally uniform sensitivity. The sensitivity of the coverages of 3-parameter PDFs depends on their slope. Because their slope is in both realizations flat for AIC-D$_{max}$ values beyond 2.5, the coverages of 3-parameter PDFs are insensitive beyond AIC-D$_{max}$/BIC-

D_{max} values of 2.5 plus 0.97 (3.7) in observations (simulations). Therefore, the coverages of the AIC-D/BIC-D category "no skill" (AIC-D/BIC-D > 10) and "sufficient" (AIC-D/BIC-D ≤ 7) are robust concerning a change of the information criterion from AIC to BIC in both realizations. In observations, the AIC-D/BIC-D category "well" (AIC-D/BIC-D ≤ 4) is also robust to the change of the information criterion (because $2.5 + 0.97 \leq 4$). Further, the slope of coverages of both 3-parameter PDFs is rather flat between AIC-D_{max} values of 1 and 2, in observations. In observations, the AIC-D/BIC-D category "ideal" (AIC-D/BIC-D ≤ 2) is, therefore, also rather robust to the change of the information criterion. Ergo, all AIC-D/BIC-D categories are in observations sufficiently robust to the change of the information criterion. We identify sensitive performances to the change of the information criterion only in simulations for the AIC-D/BIC-D categories "ideal" and "well". This sensitivity does not affect the main argument against GD2 in simulations. GD2 displays a worthless (insufficient) performance in 12% (18%) of grid-points. Also for BIC-D frequencies, GD2 displays a worthless (insufficient) performance in more than 10% (14%) of grid-points in simulations. In contrast, EWD3 displays, irrespective of the employed information criterion, a worthless or insufficient performance only in 1% of grid-points – EWD3 reduces the count of grid-point characterized by this highly undesirable performance by over one magnitude.

We extensively draw our conclusion from erroneous performances of our candidate PDFs. Irrespective of the information criterion, erroneous performances are for EWD3 virtually non-existent, but manifest for GD2 in a non-negligible percentage of grid-points in both realizations. Thus, as also discussed in the initial submission (e.g. when introducing AIC-D in the results, and when elaborating on them in the discussion), the risk of underfitting by using 2-parameter PDFs seems larger than the risk of overfitting by using 3-parameter PDFs. Consequently, once the need for 3-parameter candidate PDFs is established, their remaining punishment relative to 2-parameter PDFs biases the analysis; particularly for small AIC-D values. Because of the complexity penalty in the information criterion, our 3-parameter candidate PDFs outperform our 2-parameter candidate PDF only for AIC-D_{max} values beyond their increased complexity penalty (black vertical line in Fig I). We argue that maintaining the complexity penalty (beyond the proven inability of 2-parameter distributions) causes an artificial disadvantage for 3-parameter PDFs for small AIC-D values. Therefore, the complexity penalty biases and obscures our analysis for small AIC-D_{max} values. We interpret the results from this BIC-D analysis as another confirmation of our line of argumentation. Anyhow, this discussion (and our interpretation of a confirmation of our line of argumentation) only underlines our conclusion that EWD3 is better suited than GD2. In contrast, we draw that conclusion from erroneous performances of GD2 that manifest irrespective of the employed information criterion.

The above-conducted analysis helped us to streamline our reasoning. In consequence, we slightly altered several lines of the manuscript to simplify our

line of argumentation. This helps us to convey, and readers to intuitively understand our conclusions. In this process, we conducted two different types of changes. Firstly, changes concerning the proper communication of AIC's punishment (including the above-mentioned bias). Secondly, changes that focus our analysis on GD2 and EWD3, instead of highlighting all four candidate PDFs almost equally prominent.

In the thorough analysis of AIC's and BIC's complexity penalties, we identified an intuitive way to visualize the penalty difference between 2- and 3-parameter PDFs. The black vertical line in Fig. I. Including this black line also in Fig. 5 enables us to elaborate more precise on the impact of that penalty difference. Therefore, we adapted Fig. 5 and discuss the adaptation in the text. This simplifies our line of argumentation.

We changed a paragraph in Section 3.1.1 (lines 458 to 470) to: "It seems worth elaborating on the insufficient (only average) confidence in EWD3 to perform ideally in observations (ensemble simulations) around the globe. The complexity penalty of AIC correctly punishes EWD3 stronger than GD2 because AIC evaluates whether EWD3's increased complexity (relative to GD2) is necessary. However, the results justify the necessity for this increased complexity – GD2 performs erroneously in 26% (6%), insufficiently in 18% (2%), and without any skill in 12% (1%) of the global land area in ensemble simulations (observations). The risk of underfitting by using 2-parameter PDFs seems larger than the risk of overfitting by using 3-parameter PDFs. Once the need for 3-parameter candidate PDFs is established, their remaining punishment relative to 2-parameter PDFs biases the analysis; particularly for the ideal AIC-D category. EWD3's increased complexity penalty relative to 2-parameter candidate PDFs depends on the sample size and amounts to 2.46 in observations and 2.04 in ensemble simulations (see black vertical lines in Fig. 5 (a)–(d)). The AIC-D_{max} value beyond which EWD3 reaches coverages close to 100% approximately amounts to EWD3's increased penalty (see Fig. 5 (a)–(d)). Correcting EWD3's coverages for this bias would affect our evaluation of EWD3's performance only for the ideal AIC-D category. To illustrate this effect, we only consider AIC's estimated likelihood (without its penalty). Such a consideration corrects this complexity bias in EWD3's performance. While we analytically analyzed this consideration, a first-order approximation suffices for the scope of this publication. In that first-order approximation of this consideration, we simply shift the curve of EWD3 by 2.46 units leftwards in observations (Fig. 5 (a) and (b))) and by 2.04 units leftwards in ensemble simulations (Fig. 5 (c) and (d)). After this shift, EWD3 would also perform ideal with substantial confidence."

We substituted a sentence in Section 3.3 (lines 579 to 580) by the following elaboration: "(...) higher AIC-penalty compared to GD2. As a reminder, AIC punishes EWD3 stronger than GD2. Nevertheless this complexity punishment, it is obvious by now that our 2-parameter PDFs are inept to universally deliver normal distributed SPI time-series; particularly if one considers all depicted dimensions of the task at hand. As it turns out, this punishment is the sole

reason for both performance limitations that EWD3 displays in Table 6: (i) for the ideal AIC-D category and (ii) EWD3's tied performance with GD2 for an accumulation period of 12-months in ensemble simulations. As shown before, AIC's punishment is particularly noticeable in the ideal category. Further, this punishment also affects the tied performance ranking for the accumulation period of 12-months. To illustrate this effect, we again consider AIC's estimated likelihood (without its penalty) to correct EWD3's performance for the complexity punishment. While we again analytically analyzed this consideration, for the scope of this publication a first-order approximation suffices also here. In that first-order approximation of this consideration, EWD3's coverages of Table 6 shift again by 2.46 (2.04) AIC units in observations (ensemble simulations). Since neighboring AIC-D categories differ by 2-3 AIC units, this approximation shifts EWD3's coverages of Table 6 by roughly one category. Such a shift would solve EWD3's limitation in the ideal AIC-D category. Further, EWD3 would also perform best across all AIC-D categories in ensemble simulations; including the accumulation period of 12-months.

Despite the inclusion of the complexity penalty, EWD3 performs (...)"

Answering this question helped us to further streamline the conclusions we would like to convey. We realized that the manuscript highlights all four candidate PDFs almost equally for too long. Dismissing WD2 and GGD3 earlier helps us in telling the story. To focus our story on GD2 and EWD3, we conducted the following changes to the manuscript:

We substituted lines 342 to 345 in Section 3.1.1 by: "(...) during both seasons (Fig. 4, (b)). Aside from GD2, GGD3 and EWD3 also perform adequately in absolute terms for observations. Discriminating their deviations from $\mathcal{N}_{0,1}$ is difficult. On the one hand, GD2 represents the especially important left-hand tail of SPI_{3M} time-series' frequency distribution (D3) in JJA worse than our 3-parameter candidate PDFs (compare Fig. 4, (a) against (c) and (d)). On the other hand, GD2 displays smaller deviations from $\mathcal{N}_{0,1}$ than our 3-parameter candidate PDFs in the center of the SPI's distribution. Despite these minor differences (...)"

We substituted lines 411 to 431 in Section 3.1.1 by: "(...) considerably faster than GD2. EWD3 quickly compensates for AIC's complexity punishment (which is 2.46 units larger for EWD3 than for GD2 (indicated by the vertical black line in Fig. 5)). Beyond this vertical black line, EWD3 conclusively outperforms GD2 (the only intersection of the yellowish, and the bluish lines coincide with the intersection of that vertical black line in Fig. 5, (a) and (b)). EWD3 performs well (AIC-D_{max} ¡ 4) in virtually every global land grid-point. During DJF (JJA), EWD3 displays globally (in all land grid-points) AIC-D values of less than 5.03 (7.03). In contrast, GD2 performs erroneously (apparent by AIC-D_{max} values in excess of 4) in approximately 7% (6%) of the global land grid-points during DJF (JJA). Further, GD2 performs during both seasons insufficiently (AIC-D_{max} values beyond 7) in 2% and without skill (AIC-D_{max} values beyond 10) in 1% of the global land area. While EWD3 strictly

outperforms GGD3, GGD3 still performs similarly to EWD3 in observations. Thus, our focus on EWD3 becomes only plausible during the investigation of AIC-D frequencies in ensemble simulations."

We substituted lines 436 to 448 in Section 3.1.1 by: "We interpret EWD3's performance in ensemble simulations as ideal in approximately 85% (86%) of the global land area during DJF (JJA). For AIC-D$_{max}$ values beyond 2, EWD3 quickly approaches 100 % coverage, again, and performs erroneously or insufficiently only in 1% of the global land area during both seasons. In contrast, GD2 performs erroneously in 23% (30%) and insufficient in 14% (21%) of the global land grid-points during DJF (JJA). Yet, most telling might be the fraction (...)"

We included the following transition in between lines 453 and 454 in Section 3.1.1: "(...) over one magnitude (by a factor of roughly 20). EWD3 also universally outperforms GGD3. In view of their equal parameter-count, it seems rational to rather employ EWD3 than GGD3.

Analyzing AIC-D frequencies for both seasons (DJF and JJA) discloses no distinct season-dependent differences, similar to before in the investigation of deviations from $\mathcal{N}_{0,1}$. Therefore, we average identified land area coverages over both seasons in the summary of AIC-D frequencies. Table 3 summarizes (...)"

Aside, we inserted a sentence in the discussion. This sentence states that we also analyzed BIC-D frequencies and that they deliver similar results as shown for AIC-D frequencies. We insert this sentence at the beginning of the paragraph that starts in line 617: "We also repeated our AIC-D analysis with the Bayesian information criterion [Schwarz et al., 1978] which delivered similar results. Irrespective of the employed information criterion, the findings sketched above stay valid (...)"

References

[Blain et al., 2018] Blain, G. C., de Avila, A. M. H., and Pereira, V. R. (2018). Using the normality assumption to calculate probability-based standardized drought indices: selection criteria with emphases on typical events. *International Journal of Climatology*, 38:e418–e436.

[Naresh Kumar et al., 2009] Naresh Kumar, M., Murthy, C., Sesha Sai, M., and Roy, P. (2009). On the use of standardized precipitation index (spi) for drought intensity assessment. *Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling*, 16(3):381–389.

[Schwarz et al., 1978] Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.