Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-90-RC1, 2019

Geoscientific Model Development Discussions

Interactive comment on "Assessment of wavelet-based spatial verification by means of a stochastic precipitation model (wv_verif v0.1.0)" *by* Sebastian Buschow et al.

Anonymous Referee #1

Received and published: 24 May 2019

Overall comments:

In this study five newly wavelet-based scores for assessing the forecast-versus-observation scale-structure are introduced, and their discriminatory characteristics are compared with that of other established (spatial) verification scores.

The article is well structured and the analysis is clearly explained. The set up for the scores definition and testing are well motivated (construction of the synthetic fields, in Sec 2; use of the redundant discrete wavelet transform and wavelet family selection, in Sec 3). Innovative parts of the article include: the spatial aggregation by considering the histogram of the dominant central scales (Sec 4); the analysis of the sensitivity

of the wavelet spectra to the scale and smoothness parameter (Sec 5); the use of the Earth Mover's Distance to compare the spectra (Sec 6). The set-up of the forecast experiment is ingenious (Sec 7) and enable a clear analysis of the discriminating power of the different scores (Fig. 6,7). Unfortunately, there is no single score which emerges as the recommended best score. However, the analysis is very solid and the methods illustrated are extremely interesting, therefore we recommend the article for publication (after minor-to-major revisions), as a very valuable contribution towards a better understanding (and development) of new spatial (scale-separation) verification methods.

Major Revisions:

It would be nice to see the mean spectra and histogram of the dominant scales for the case study shown in Figure 3 (please add a panel): I expect a bimodal histogram and spectra, since both small scale features and a large front are present in the case study. This bi-modality (i.e. presence of of both small and large scale features) is not represented in the stochastic rain fields produced in Section 2 (the synthetic fields considered in the article have by construction uni-modal spectra), and in fact their resulting spectra and histograms are uni-modal (e.g. Figure 4). However spectra bi-modality (i.e. presence of of both small and large scale features) is bound to happen in real verification practice, and it might be badly handled by H_cd and Sp_cd. In fact, H_cd and Sp_cd are the differences of the centre of mass (of the scale histograms and of the mean spectra), and are not suitable summary statistics to compare bimodal curves (or any other non-Gaussian curve). Hemd and Spemd, on the other hand, seem more suitable statistics (to compare Gaussian or non-Gaussian curves) since based on the whole curve comparison. The authors should consider withdrawing H_cd and Sp_cd from the newly proposed wavelet-based scores. [If the authors wish to introduce a metric which measure the direction of the error, maybe they should consider a measure based on the distances along the whole curves (or the integral between the two pdf), but with a sign which accounts for the curves relative position.]

Figure 4 (Section 5) shows that both mean spectra and scale histograms are sensitive to the variation of the scale parameter b and the smoothness parameter ν, and that for both parameters, the curves shift in the expected direction (this is the main result). The histogram of the dominant scales seems slightly less sensitive (it shifts less), however it exhibits a smaller spreadÂă(hence smaller uncertaintyÂă: the signal is better defined). Because of this latter property, the scale histogram should be favoured, with respect to the mean spectra. Moreover, the smaller shifts of the histograms are probably simply related / due to their smaller spread (I have the feeling that the magnitude of the shift is proportional to the spread). These aspects should be mentioned in Section 5. (Note: the sensitivity of the spread to the parameters b and ν is secondary: be careful not to mix it up with the main result, aka the shift).

From the previous two comments, I would propose as unique new statistics H_emd.

At the end of Section 2, then authors introduce an algorithm for producing stochastic rain fields which satisfy non-stationarity and anisotropy. Some case studies are illustrated in Figure 2, and the associated verification results are discussed in Section 7.4. In my view this analysis can be removed from the article for the following reason:

a) The algorithm for producing stochastic rain fields which satisfy non-stationarity and anisotropy, despite being more sophisticated than the isotropic algorithm mainly used in the article, is still not realistic (the precipitation features of Figure 2 are still far from resembling the ones for the real case illustrated in Figure 3).

b) The article will result nicely well contained in illustrating “solely” the isotropic stochastic fields (you have already quite a lot of material! Moreover, this would provide a nice “excuse” for retaining the statistics based on the centres of mass -wink!-). In this case you need to add into the final discussion Section the need to analyze real cases, in future work ...

c) For the (future) analysis of more realistic cases, I strongly suggest to consider directly real precipitation case studies (the Spatial Verification ICP cases from Ahijevych

et al 2009 are available online), rather than using synthetic fields (you might end up spending a lot of time and implementing very complex stochastic models ... to achive the same results ...).

Minor Revisions:

Abstract and Introduction

Page 1 line 7: replace ’spatial correlation’ with ’spatial structure’ (or ’scale structure’). Page 1, line 23: please quote (also) Dorninger et al (2018): “The set-up of the Mesoscale Verification Inter-Comparison over Complex Terrain project”. Bull. Amer. Meteorol. Soc., 99 (9), 1887 – 1906. Page 1, line 23: replace ’avoid’ with ’deal with’. Page 1, lines 16-19: rephrase . . . (this is a bit weak, as first sentence of the article). Page 2, line 5: I suggest adding in this paragraph one sentence introducing the fourth class of spatial verification methods, the scale-separation techniques (with the key references). Then you start the new paragraph by stating that the technique introduced in your article belongs to this latter class. Then you describe the most recent literature on variograms etc. (as from line 8 onwards). Here you need to state that the variogram-based techniques are a sub-set of the scale-separation techniques. Page 2, the paragraph ending at line 22 can be joined with the one starting at line 23.

Section 2

Page 3 line 25: write ’The threshold T determines the percentage of the field which has non-zero values’. You need to state (here) that T is the base rate. Page 5, line 15: When introducing the scale auto-correlation parameter b, and when discussing Figure 1, you need to mention explicitly that smaller b are associated with larger scales, and vice-versa larger b are associated with smaller scales (this is counter-intuitive, therefore it needs to be reminded here and there in the article). Page 5, lines 11-13: it is not clear where this statement lead to: in the article, are you imposing $\nu > 1$? Are you using random Gaussian distributions to create / perturb you parameters? Please state. Page 5, line 26: define the rotation angle.

Section 3

Page 7, line14 - Page 8 line 1: this is not “loosely speaking”, please redefine (in easier words) the concept of local stationarity: does it mean that locally your auto-correlation is zero? You can also decide to remain with mathematical strict definitions ... in the rest of the paragraph, you are quite technical ... however my preference is always to accompany the mathematical explanation with a sentence which explain / vulgarize the mathematical content. You might need to summarize the findings of Eckley et al (2010), Kapp et al. (2018).

Section 4

Re-title section 4 as 'Wavelet spectra spatial aggregation'. Page 9, line 10: for the case study add the reference to Ahijevych, D., E. Gilleland, B.G. Brown, and E.E. Ebert, 2009: Application of spatial verification methods to idealized and NWP-gridded precipitation forecasts. Weather Forecast., 24 (6), 1485 – 1497. [From the major comment: please, add a panel in Figure 3, with the mean spectra and histogram of the central scales for the shown case study. I expect a bimodal histogram and spectra, since both small scale features and a large front are present in the case study.]

Section 5

Re-title Section 5 as “Wavelet Spectra Sensitivity Analysis”. Page 9, line 22: please remind here that larger (smaller) b is associated with smaller (larger) scales. Page 9, lines 26-27: eliminate the sentence “Simultaneously ... observed scales” (I do not see this in the Figure; moreover the sentence distracts from the main point).

Paragraph starting at page 9, line 32 and ending at page 10, line 2 (describing the major findings of Figure 4): In this paragraph you have one main result and a secondary result. The main result is that both mean spectra and scale histogram are sensitive to the variation in the parameters b and ν, and that for both parameters they shift in the expected direction. The sensitivity of the spread as you vary b or ν is a secondary

results (which is actually neither too visible, nor to important for your study). In the paragraph these are mixed up in the discussion, so that the latter takes away the focus from the former. Rephrase the paragraph. E.g. at page 10, line 2, I suggest writing: ' ... only affected by b: larger scales (smaller b) lead to a greater variance (panel b) whereas changes in smoothness (parameter ν) do not substantially change the histogram shape' (avoid mentioning the shift here). [From the major comment, you should also state that: 1. the scale histogram exhibits less spread, the dominant scales are better defined, and hence it is favoured wrt the mean spectra. 2. the smaller shift of the scale histogram is possibly proportional / due to its smaller spread, and not to a lack of sensitivity.]

Page 10, line 6: the lack of sensitivity of both the mean spectra and the scale histogram on the base rate (parameter T) is a very welcome property in a verification scoring rule (it implies that the score cannot be edged, e.g. by over-forecasting, and that the performance does not depend on the underlying climatology). This should be mentioned.

Section 6

Page 11: [From the major comment: real precipitation fields might generate bi-modal spectra (whereas the synthetic fields considered in the article have by construction uni-modal spectra). H_cd and Sp_cd (page 11), are not suitable statistics for comparing bi-modal (or non-Gaussian) spectra, because they compare the centre of mass of the curves: this limitation ought to be (at least) mentioned. H_emd and Sp_emd, on the other hand, seem more suitable statistics (to compare Gaussian or non-Gaussian curves) since based on the whole curve comparison. If the authors wish to introduce a metric which measure the direction of the error (such as H_cd and Sp_cd), maybe they should consider a measure based on the distances along the whole curves (or the integral between the two pdf), but with a sign which accounts for the curves relative position.]

Page Page 11, lines 10-13: please define EMD (either write the formula or describe how it is calculated ... “moving the dirt ... work” is visually clear, but it would be better to be more precise). Page 11, lines 13-14: by normalizing the spectra to obtain a unit sum you essentially remove the bias, and concentrate solely on the pure scale structure (how the total energy is distributed across the scales). This should be mentioned. Page 12, line 5: there is an incoherence in the naming of the Energy score, in this Section it is “Sp_e”, whereas in Figure 5 it is “SpEn”. I personally prefer the latter, or “Sp_en”, to well separate it from “Sp_emd”.

Section 7

Page 14, lines 11-13 (describing the bottom panels of Figure 5, evaluating the ensembles against a RS observation): not only the RS ensemble scores best (for all scores), but also the SmS and RL exhibits the second best score and the SmL (the most dissimilar ensemble with respect to RS) exhibits always the worst score. You should mention this.

From page 14 line 15, to page 15 line 6, need to be rephrased: a) when comparing RL to SmS (Page 14, bottom 2 lines): the compensating error affect solely the location / mean value of the mean spectra ans scale histogram, or does it affect the whole mean spectra and scale histogram? I question the phrasing ’on location of the spectra and histograms along the scale axis’ (I would eliminate this part of the sentence). In the following sentence (page 15, line 1) I question ’by their centres of mass alone’. b) Page 15, lines 1-3: I think that the SmS and RL ensemble cannot be separated well for all scores (also Vw5), not only for Hcd (I won’t attribute the lack of separation to the fact that Hcd compare centres of mass). This is possibly due to the fact that the mean spectra and scale histograms for RL and SmS are similar (From Figure 1, the top-left and bottom-right panels are more similar than the top-right versus bottom-left). Nevertheless, in the top panels of Figure 5 all scores (but V20) shows a slightly larger error for the SmS ensemble than for the RL ensemble (which is encouraging), and then even larger errors for SmL and RS (it seems to me that the scores are informative ...

). c) Top panels of Figure 5: The two scores considering the sign of the error (H_cd and S) exhibit the same behaviour, not only for SmL and RS, but also for SmS (they both exhibit slightly negative values): the sentences at page 15 lines 4-5 are partially incorrect, please re-phrase them.

From Figure 5 and 6, it is clear that V20 is the less informative score: please add this comment (you can relate to your comment when introducing V20 in Section 6, . . .).

Section 7.2: The results associated to Figure 7 are very nicely discussed and very interesting! For ensembles, SpEn is the champion score followed by Vw5, whereas for deterministic Vw5 closely followed by Sp_emd are the champion scores. I am surprised of the lower performance of H_emd: why? After these results, one could be tempted to choose Vw5 as scoring rule . . . however its strong dependency on the base rate/climatology (Section 7.4) cannot be ignored. Maybe you can add some of this comment in the discussion?

Section 7.3: please specify in the caption of Table 3 (and Table 4), or write in the text, that ExP1 = D1 = Haar, and that Exp4 = D4 is the wavelet considered in the main experiment of the article.

Section 7.4, page 18 lines 5-6: given that in the original experiment T was set to 0.2 (aka 20% of the domain was precipitation, and 80% was zero values), I imagine that with this model the precipitation area is ranging in 15-25% of the domain: can you please phrase this more clearly? (rather than using the 75%-85% range, refer to your previously fix 20% base rate ...)

Discussion and conclusions

page 20, line 14: I suggest writing ’mis-representation of feature sizes (e.g. smoother representation of small-scale convective organization)’.

Page 20, lines 17-25: the findings of Figure 6 and 7 are well summarized in the conclusions (page 20, lines 21-25). I would end this paragraph at line 25. The sensitivity

of the Variogram score to p and w (lines 31-32) could also be added to this paragraph. Then (at page 20, line 26) I would start a new paragraph, discussing the results of the sensitivity analyses (sensitivity to T and to the wavelet choice).

Sensitivity to T: I suggest to phrase differently lines 25-30 (page 20): you need to remind that the 'perturbation of the data' is essentially an assessment of the sensitivity of the scores to the sample climatology. I would express more concern about the loss of discrimination of the variogram scores found in section 7.4.

Sensitivity to the wavelet choice: I would rephrase lines 32-33 (page 20) as 'We have also tested the sensitivity of the newly introduced wavelet-scores to the choice of the mother wavelet. We have performed . . . '.

As the last paragraph of the conclusion suggests, this study is still exploratory: there is no single score which has emerged as the recommended best score. This should be mentioned. Moreover, the paragraph could be re-phrased to include real case studies and scores which accounts for the direction of the error while applied to bi-modal spectra (as explained in the major comments).

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-90, 2019.