Reviewer 2

This research proposes a self-supervised contrastive learning-based precipitation forecast verification method. It can effectively reflect the certain degree of forecasts errors through contrastive learning, thereby successfully addressing the “double penalty” issue of traditional point-to-point verification methods and the reliance on manually set parameters in spatial verification methods. Overall, the research holds considerable value and is recommended for publication after minor revisions.

Thank you very much for your thorough review of this manuscript. Your comments and suggestions are very helpful. We have carefully revised the manuscript based on your concerns. Responses are marked as green and revisions in the revised manuscript are highlighted as blue.

Comments

1. In lines 65-66, this sentence “Therefore, a more comprehensive...still needed” is too long to read, please modify it for better understanding.

We completely agree with this comment, now this sentence is improved in lines 64-65: Therefore, a more comprehensive verification method is needed to tolerate minor errors yet penalize significant ones, thereby better reflecting the varying degrees of error.

2. In the paragraph of lines 70–75, the rationale for using deep learning to verify precipitation forecasts is not convincible enough. Please provide a more detailed explanation.

Thanks a lot for this considerable comment. We added an elaboration at the start of this paragraph in lines 66-72: The limitation of existing spatial verification methods essentially stems from their reliance on predefined parameters and rules, preventing them from truly capturing the spatial distributions of observed and forecast precipitation fields. Consequently, conducting PFV from an overall structural perspective promises more reliable results. Inspired by this, we propose a deep-learning-based PFV method that evaluates forecast performance by comparing the high-dimensional features of observed and forecasted precipitations. This approach leverages the exceptional capabilities of deep learning in simulating human cognitive processes and extracting complex features, as well as its remarkable success in image verification practices in recent years.

3. In lines 180-183, the punishment function is not correct, first, it should be p(f-), second, it should provide a specific equation.

Thank you for pointing out this issue. Since added quadratic penalty of CLPFV loss function is to reflect different degrees of precipitation augmentation (f+), the punishment function should be p(f+). But a specific equation is still needed. Thus, we added Eq. (9) and an explanatory sentence in lines 190-191: Eq. (9) shows the specific equation of punishment function \p(f+), which is the combination of the magnitude square of all kinds augmentations \m with a penalty coefficient \lambda.

$$p(f_+) = \lambda \cdot ({m_{displacement}}^2 + {m_{intensity}}^2 + {m_{area\ size}}^2) \tag{9}$$

4. In figure 7, to improve its self-explanatory ability, it is recommended that the specific deviation values used be labeled directly below each subplot (e.g., "Displacement: -10, -5, 0, +5, +10" under subplot (a)), rather than only described in the caption.

We have modified figure 7, directly added concrete deviation information under each subplot.

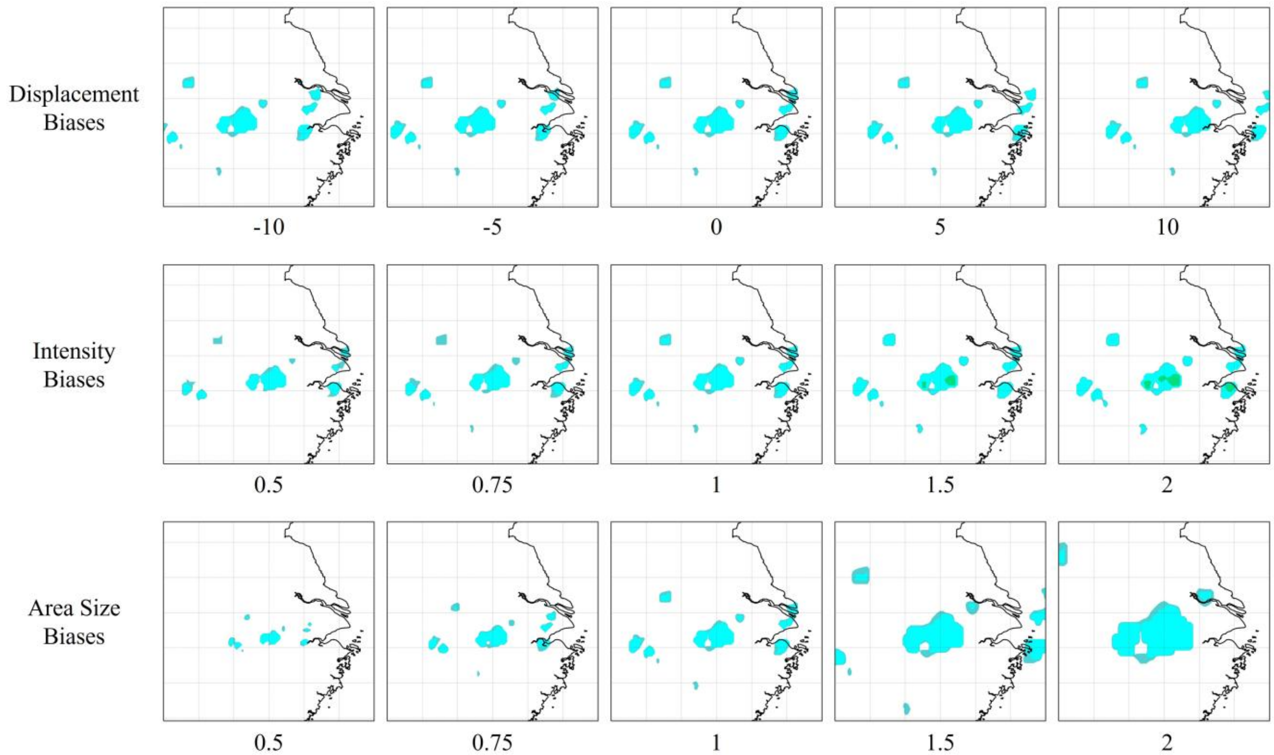

Figure 7. An example of simulated forecasted precipitations by applying gradient biases.

5. In the discussion of figure 9, is it possible to provide a benchmark to compare all verification methods? This will make the experimental results more intuitive.

This suggestion is helpful, since benchmark makes Fig. 9 to provide a clearer comparison of different PFV methods. Therefore, we introduced a downward-opening parabola in Fig. 9 as a benchmark reference curve for the PFV experimental results. This specific parabola was selected because its shape precisely characterizes the expected behavior of an ideal PFV method: it assigns the highest verification score in the absence of errors; permits a gradual decline for minor errors to ensure fault tolerance and avoid "double penalty" problem; and enforces an accelerated descent for larger errors to significantly penalize severe forecast failures.

The figure below is the revised Fig.9.

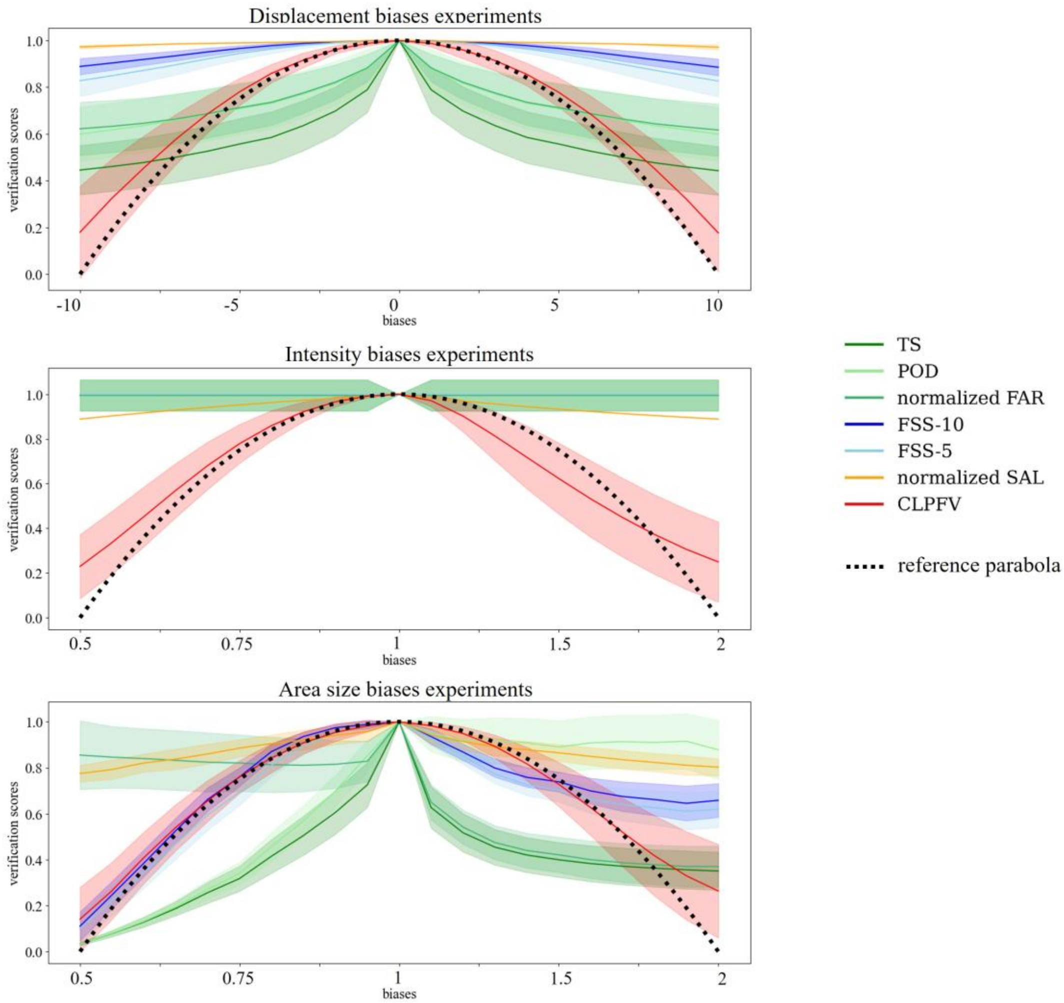

Figure 9. Results of displacement biases (top), intensity biases (middle), area size biases (bottom) experiments. The experimental results of each verification method are presented as average verification score curves with their 95% confidence intervals.

6. The content in lines 296–335 may be appropriately simplified to prevent redundancy with the information presented in figure 9.

We completely agree with this comment. Now we have improved the paragraphs in lines 311-336, reduced the words amount from 660 to 426 and did not change the content.