

Review

To predict weather radar image sequences, the authors propose the model RAP-Net that add attention modules (RAB and RAM) in a ConvRNN model, in order to improve forecasting in the area with heavy rainfall (RAB) and improve the long-term spatiotemporal representation ability (RAM).

1. In general, rainfall intensity is classified as light, moderate and heavy (AMS glossary "Rain": <https://glossary.ametsoc.org/wiki/Rain>), not as "strong" (line 7) or "middle" (lines 10, 44, 53 and 243). Please fix this throughout the text.
2. In line 22, what are the limitations of the traditional approach you are referring to?
3. In line 39, the sentence "where the similar semantics gathered in the same tensor." sounds incomplete.
4. I suggest reviewing all equations to simplify the notation and unify the letters of all equations (sections 3.2, 3.3 and 3.4).
 - a. The notation adopted by the terms of equations is confusing (sections 3.2, 3.3 and 3.4):
 1. In line 89, what are the letters "l" and "h" in X_h^l ?
 2. Is $X_h^l = X_t$?
 3. Etc.
 - b. The equations in lines 109, 100 and 110 do not reflect Fig.2, and vice-versa. The scheme shown in Fig. 2 should also be reviewed with respect to connections.
 - c. In line 157, when you say "RAP-Cell" do you mean "RAB"? Same in Figure 5, how does "RAM" use "RAP"? (Conflict with Figure 2 and the following equations).
5. At the beginning of section 3.3, part of the text is missing, perhaps comments on Figures 3a and 3b.
6. In section 4.1, please describe the dataset: data type, variable, instrument; rain or reflectivity; time period, etc.?
7. The equation 8 (page 10) is not a Z-R relation; it looks like a scale conversion. The Z-R relation is in the form of a power law $z = aR^b$.
8. In section 4.2, you should inform the range of the evaluation metrics.
9. In line 190, with "SST-LSTM" do you mean "RAP-Net"?
10. Are the results in Tables 1 to 4 calculated over the entire test set?
11. You should consider joining Table 1 with 2 and 3 with 4.
12. In line 195 "Besides, the proposed model has significant superiority especially for the nowcasting in heavy rainfall regions. Because..." you should join these sentences.
13. In line 199, what is the execution time of each compared model: PredRNN, PredRNN++ and RAP-Net?
14. In line 215, please give more arguments to state that your model is better in high thresholds.
15. In Figures 7 and 9, what is being shown, rain or reflectivity? Please fill in the caption.
16. In the conclusion (section 5), in future work, could you provide more details on how you intend to use more layers? What do you think about computational resources and execution time?

Other revisions:

1. The text is well written and comprehensible, but some sentences are short and break. The authors should join a few subsequent sentences together, using connectors and punctuation. Besides, there are some typological errors, such as: lines 52, 70, 93, 133, 148, 180, 195, Figure 3.
2. Separate citations from the text with parentheses. For example: line 16 (Shi et al., 2017); lines 22 and 23 (Shi et al., 2017; Ayzel et al., 2020; Li et al., 2021).
3. Use the acronyms RAB and RAM throughout the text in place of "Region Attention" (ex. line 218)
4. Review punctuation in lines 73, 74 and 75 "Our proposed method is different ..."
5. Separate the terms in Figure 1 with commas, it looks like they are multiplying.