

## Reviewer comments

The manuscript presents a potentially useful framework for multivariable WRF parameterization screening. And the authors have revised the manuscript following the comments. However, several methodological issues currently limit the robustness and comparability of the results. In particular, the mathematical interpretation of the proposed weighting rule and the comparison among weighting methods require substantial clarification. I therefore recommend major revision. Additional issues concerning terminology, figures, and presentation are listed as minor comments.

### Major comments

1. The proposed multiplicative rule is not non-compensatory

P12, L248-275; Eqs. (13)-(15); related discussion at P29, L617-633.

The manuscript describes MRE x CV weighting as a "non-compensatory joint constraint." However,  $w_1 = \text{softmax}(\text{MRE})$ ,  $w_2 = \text{softmax}(\text{CV})$ , and the normalized product is exactly  $w = \text{softmax}(\text{MRE} + \text{CV})$ . Thus, MRE and CV are additive before the final Softmax, and a high MRE can compensate for a low CV. The method should therefore not be described as strictly non-compensatory. Please revise the mathematical interpretation and related claims.

Figure 2 also suggests that the MRE component largely determines the final weights. Please add an MRE-only baseline, to the comparison in Fig. 8. This is needed to show whether the proposed method adds useful information beyond MRE alone.

2. A common variable-level comparison of all final configurations is needed

Figures 4 and 7 provide variable-level results only for the proposed and equal-weighting configurations. Entropy, CV, and CRITIC are compared only through integrated metrics.

Lines 715-719 report RMSE reductions of 0.98 degrees C for TEM, 2.47 degrees C for GST, and 1.38 mm d-1 for PRE, but the reference configuration and absolute RMSE values are not given. It is also unclear why only these variables are reported.

Please compare the final configurations selected by all five methods using the same daily station samples and the same unweighted variable-level metrics, such as RMSE, MAE or mean bias, and correlation. This would show whether the proposed method improves its highly weighted target variables without unacceptable losses in the other variables.

### Minor comments

L175 and subsequent figures and text.

The seven variables are introduced as T2, GST, PRS, PRE, RH, U10, and V10, but later figures use GST, PRE, PRS, RHU, TEM, and UV10. Please use consistent names and define all regrouping.

L175-176. The manuscript defines U10 as meridional wind and V10 as zonal wind. Under the usual WRF convention, U is zonal and V is meridional. Please check and correct these definitions.

Around L287 and throughout the manuscript. Please clearly define MIEI, cMIEI, and the median-normalized MIEI shown in Figs. 3 and 6. Their notation and roles should remain consistent throughout the text, equations, tables, and captions.

L425. The word "significantly" requires a statistical significance test. Please provide such a test or replace it with a descriptive term such as "substantially" or "noticeably."

Figs. 2, 5, and 8 and related text. The figures combine U10 and V10 as UV10, while the text often calls both components "wind speed." Please clarify whether they enter MIEI as a two-dimensional vector, are combined into wind-speed magnitude, or are treated as separate zonal and meridional components.

L672-705. The discussion mainly explains why the selected schemes are physically plausible based on their design. It does not show why these mechanisms made them outperform the competing schemes in this study. Please link the interpretation to direct variable-level results where possible, or present these explanations more cautiously as hypotheses.

L780-795. This section repeats the limitations of the stepwise search and gives a long list of fixed settings and untested uncertainty sources. Please shorten it to state that the study isolates parameterization selection under a fixed WRF configuration and does not provide a complete uncertainty decomposition.

Tables 3 and 4 would benefit from further clarification. If each cMIEI value was calculated using its corresponding method-specific weights, the values reflect different evaluation objectives and may be difficult to compare directly. To improve comparability, the authors could evaluate all selected final configurations using the same reference weights or a common set of unweighted metrics. Alternatively, a cross-evaluation matrix could be provided, in which every stepwise-selected configuration is evaluated under each weighting method and the corresponding cMIEI values are reported.

It would also be helpful to present the complete stepwise pathways for entropy, CV, and CRITIC weighting. Reporting the selected scheme and cMIEI at each stage would make the effects of different weighting methods on the selection pathways and final configurations easier to compare.