Reply to the Comments by Reviewer #1 on GMD-2024-114

" Evaluation of Dust Emission and Land Surface Schemes in Predicting a Mega Asian Dust Storm over South Korea Using WRF-Chem (v4.3.3)"

By Ji Won Yoon, Seungyeon Lee, Ebony Lee, and Seon Ki Park

This manuscript examines the performance of 20 combinations of WRF-Chem dust aerosol parameterizations and land-surface schemes in reproducing a mega Asian Dust Storm (ADS) from March 2021. The validation study is highly specific to a single numerical modeling system and individual event, so the novelty of this study and its broader impact on the wider atmospheric sciences is limited. However, I recognize that verification studies like this are sometimes important incremental steps within larger projects, and the results may nonetheless help guide the selection of appropriate dust physics settings among other researchers and practitioners in East Asia. The manuscript, though modest in its scope and potential scientific impact, is nonetheless well written and soundly conducted. I believe it could be accepted for publication pending the revisions suggested below.

→ We sincerely appreciate the reviewer's positive and valuable comments. We have carefully considered each comment and provided an item-by-item response below.

Overall Comment:

1. The manuscript computes both POD and FAR for several ACWS-relevant PM10 thresholds, and it emphasizes that these two scores need to be interpreted jointly. However, performance metrics such as the Critical Success Index (CSI) do exactly that. The manuscript would be strengthened by the addition of CSI, or a similar metric, that merges these two ideas into a single score. The CSI is easy to compute with the information already provided in the manuscript.

→ We believe the reviewer's comment is valid. According to the reviewer's comment, we have added the Critical Success Index (CSI) to Table 4. Additionally, we have included explanations of the basic concept of CSI and provided an analysis of the simulation results based on CSI in the manuscript.

(Original manuscript) Table 4

Table 4: POD and FAR values for each PM10 threshold across all scheme combinations. The bold numbers indicate POD greater than 0.5. The dashes '-' indicate POD and FAR values that cannot be calculated.

		> 80 μg m^{-3}		> 150 μg m^{-3}		≥ 300 μg m^{-3}		≥ 800 μg m^{-3}	
		POD	FAR	POD	FAR	POD	FAR	POD	FAR
GOCART	Noah	0.055	0.164	-	-	-	-	-	-
	RUC	0.097	0.221	-	-	-	-	-	-
	Noah-MP	0.114	0.120	-	-	-	-	-	-
	CLM4	0.079	0.298	0.002	0.500	-	-	-	-
AFWA	Noah	0.090	0.128	-	-	-	-	-	-
	RUC	0.223	0.256	0.027	0.063	-	-	-	-
	Noah-MP	0.126	0.103	-	-	-	-	-	-
	CLM4	0.110	0.264	0.007	0.333	-	-	-	-
UoC01	Noah	0.076	0.171	-	-	-	-	-	-
	RUC	**0.516**	0.401	0.251	0.305	0.057	0.048	-	-
	Noah-MP	0.138	0.073	-	-	-	-	-	-
	CLM4	**0.918**	0.297	**0.758**	0.351	0.448	0.325	0.034	0.727
UoC04	Noah	0.077	0.169	-	-	-	-	-	-
	RUC	**0.544**	0.407	0.282	0.331	0.075	0.037	-	-
	Noah-MP	0.152	0.093	-	-	-	-	-	-
	CLM4	**0.928**	0.310	**0.799**	0.378	**0.520**	0.320	0.069	0.714
UoC11	Noah	0.031	0.257	-	-	-	-	-	-
	RUC	0.149	0.225	-	-	-	-	-	-
	Noah-MP	0.060	0.138	-	-	-	-	-	-
	CLM4	0.071	0.298	0.002	0.500	-	-	-	-

(Revised manuscript) Table 4

Table 4: POD, FAR, and CSI values for each PM10 threshold across all scheme combinations. The bold numbers indicate that POD is greater than 0.5 and CSI is relatively higher compared to the others. The dashes '-' indicate POD, FAR, and CSI values that cannot be calculated.

		> 80 μg m^{-3}			> 150 μg m^{-3}			≥ 300 μg m^{-3}			≥ 800 μg m^{-3}		
		POD	FAR	CSI	POD	FAR	CSI	POD	FAR	CSI	POD	FAR	CSI
GO-CART	Noah	0.055	0.164	0.055	-	-	-	-	-	-	-	-	-
	RUC	0.097	0.221	0.095	-	-	-	-	-	-	-	-	-
	Noah-MP	0.114	0.120	0.112	-	-	-	-	-	-	-	-	-

	CLM4	0.079	0.298	0.077	0.002	0.500	0.002	-	-	-	-	-	-
AFWA	Noah	0.090	0.128	0.089	-	-	-	-	-	-	-	-	-
	RUC	0.223	0.256	0.207	0.027	0.063	0.027	-	-	-	-	-	-
	Noah-MP	0.126	0.103	0.124	-	-	-	-	-	-	-	-	-
	CLM4	0.110	0.264	0.106	0.007	0.333	0.007	-	-	-	-	-	-
UoC01	Noah	0.076	0.171	0.074	-	-	-	-	-	-	-	-	-
	RUC	**0.516**	0.401	0.383	0.251	0.305	0.226	0.057	0.048	0.057	-	-	-
	Noah-MP	0.138	0.073	0.136	-	-	-	-	-	-	-	-	-
	CLM4	**0.918**	0.297	**0.661**	**0.758**	0.351	**0.537**	0.448	0.325	**0.369**	0.034	0.727	0.032
UoC04	Noah	0.077	0.169	0.076	-	-	-	-	-	-	-	-	-
	RUC	**0.544**	0.407	0.396	0.282	0.331	0.247	0.075	0.037	0.074	-	-	-
	Noah-MP	0.152	0.093	0.150	-	-	-	-	-	-	-	-	-
	CLM4	**0.928**	0.310	**0.655**	**0.799**	0.378	**0.538**	**0.520**	0.320	**0.418**	0.069	0.714	0.059
UoC11	Noah	0.031	0.257	0.031	-	-	-	-	-	-	-	-	-
	RUC	0.149	0.225	0.143	-	-	-	-	-	-	-	-	-
	Noah-MP	0.060	0.138	0.059	-	-	-	-	-	-	-	-	-
	CLM4	0.071	0.298	0.069	0.002	0.500	0.002	-	-	-	-	-	-

(Original manuscript) Lines 296~301

Therefore, it is essential to consider FAR with POD to address these limitations. The formulas for POD and FAR are as follows:

$$POD = \frac{a}{a+c} \quad (10)$$

$$FAR = \frac{b}{b+d} \quad (11)$$

(Revised manuscript) Addition

Therefore, it is essential to consider FAR with POD to address these limitations. Additionally, the Critical Success Index (CSI) is an important metric used to evaluate the overall accuracy of forecasts. It measures the ratio of correctly forecast events to the total number of observed and forecast events, accounting for both ‘False Alarm’ and ‘Miss’ events. In other words, CSI addresses the limitations of POD and FAR by integrating both metrics, providing a clearer assessment of overall forecast performance. The CSI value ranges from 0 to 1, with values closer to 1 indicating higher forecast skill. CSI is particularly useful because it considers both over-forecasting and under-forecasting, showing how accurate the forecast is. The formulas for POD, FAR, and CSI are as follows:

$$POD = \frac{a}{a+c} \quad (10)$$

$$FAR = \frac{b}{b+d} \quad (11)$$

$$CSI = \frac{a}{a+b+c} \quad (12)$$

(Original manuscript) Lines 413 ~ 414

Therefore, considering both POD and FAR, UoC04-CLM4 demonstrated the best performance, followed by UoC01-CLM4.

(Revised manuscript) Addition

Therefore, considering both POD and FAR, UoC04-CLM4 demonstrated the best performance, followed by UoC01-CLM4.

In addition to POD and FAR, CSI provides a comprehensive evaluation of forecast accuracy by accounting for correct predictions, false alarms, and missed events. The CSI values for all scheme combinations for each threshold are as follows: 1) At 80 $\mu g\ m^{-3}$, UoC04-CLM4 and UoC01-CLM4 exhibited the highest CSI values of 0.655 and 0.661, respectively, while the other schemes had significantly lower values, mostly below 0.2, except for UoC04-RUC (0.396) and UoC01-RUC (0.383); 2) At 150 $\mu g\ m^{-3}$, UoC04-CLM4 and UoC01-CLM4 demonstrated CSI values of 0.538 and 0.537, respectively, indicating higher prediction accuracy compared to other scheme combinations; 3) At 300 $\mu g\ m^{-3}$, UoC04-CLM4 outperformed the other schemes with a CSI of 0.418. Although this was a comparatively lower value, it still demonstrated better performance compared to the other schemes, most of which showed poor or non-existent forecast skill; 4) At 800 $\mu g\ m^{-3}$, only UoC01-CLM4 and UoC04-CLM4 were calculated only for CSI, but both showed very low values of 0.059 and 0.032, respectively. Overall, UoC04-CLM4 consistently maintained CSI values above 0.5 up to 300 $\mu g\ m^{-3}$, showing the highest performance among all experiments.

2. The study references PM10 PCC values (Figure 6) and scatterplot relationships as “good” for some scheme combinations. However, visually, the observed-vs-simulated PM10 relationships appear quite weak. The PCCs for even the most skillful LSM-dust scheme combinations still only explain a relatively small fraction of the variance (~30% at most) if thought of as R^2 rather than R. The manuscript should clarify how the PCCs, even low ones, are indicating “good” performance.

→ We appreciate the reviewer giving us these valuable comments, which further improved the quality of the manuscript. Following the reviewer's comment, we acknowledge that the explanation suggesting good performance despite the low correlation for PM10 lacked clarity. The PCC value of the most skillful LSM-dust scheme combination for PM10 is lower compared to those of meteorological variables, such as 2m temperature and relative humidity. Therefore, for PM10, we think that the term "better" is more appropriate than "good" when comparing with other scheme combinations. Based on this, we have revised the manuscript as follows.

(Original manuscript) Lines 354~360

Fig. 6 shows PCC, RMSE, and MBE for all scheme combinations. Overall, UoC04-CLM4 showed the best performance, followed by UoC01-CLM4. The UoC04-RUC and UoC01-RUC also showed good performance compared to other scheme combinations. Conversely, the combinations of UoC01 and UoC04 with Noah and Noah-MP, as well as the combinations of GOCART, AFWA, and UoC11 with all land surface schemes, performed poorly. The detailed descriptions of the verification results are as follows: 1) For PCC (Fig. 6a), UoC04-CLM4 showed the highest value (0.61), indicating the best performance, followed by UoC01-CLM4 (0.60), UoC04-RUC (0.47), and UoC01-RUC (0.44). In all scheme combinations except for combinations of UoC04 and UoC01 with CLM4 and RUC, PCC was below 0.4, indicating very weak or almost no correlation; ~

(Revised manuscript)

Fig. 6 shows PCC, RMSE, and MBE for all scheme combinations. Overall, UoC04-CLM4 showed the best performance, followed by UoC01-CLM4. The UoC04-RUC and UoC01-RUC also demonstrated relatively better performance compared to other scheme combinations. Conversely, the combinations of UoC01 and UoC04 with Noah and Noah-MP, as well as the combinations of GOCART, AFWA, and UoC11 with all land surface schemes, showed poor performance. The detailed descriptions of the verification results are as follows: 1) For PCC (Fig. 6a), UoC04-CLM4 showed the highest value (0.61), indicating a moderate correlation, followed by UoC01-CLM4 (0.60), UoC04-RUC (0.47), and UoC01-

RUC (0.44) which also showed moderate correlations. In all other scheme combinations except for UoC04 and UoC01 with CLM4 and RUC, PCC was below 0.3, indicating a weak or almost no correlation; ~

※ Please refer to "Point 6" for the correlation strength criteria of the PCC.

(Original manuscript) Lines 376~386

Figure 7 shows a scatter diagram for CLM4-based combination---the land surface scheme that showed the best prediction performance when combined with UoC04 and UoC01 in the verification (see Fig. 6). The x-axis represents PM10 observations, while the y-axis indicated the predicted values of PM10 for each experiment. The red circles represent the predicted PM10 values corresponding to observations. The scheme combinations UoC04-CLM4 (Fig. 7c) and UoC01-CLM4 (Fig. 7d) showed similarly good performances while the other three combinations showed no correlations between observations and forecasts (Fig. 7a, b, and e): UoC04-CLM4---the best performance in verification---primarily showed overestimation for values below approximately 180 $\mu g\ m^{-3}$ and wider dispersion with underestimation tendencies for values above 180 $\mu g\ m^{-3}$.

(Revised manuscript)

Figure 7 shows a scatter diagram for the CLM4-based combination---the land surface scheme that demonstrated the best prediction performance when combined with UoC04 and UoC01 in the verification (see Fig. 6). The x-axis represents PM10 observations, while the y-axis indicates the simulated PM10 values for each experiment. The red circles represent the simulated PM10 values corresponding to the observations. In UoC04-CLM4 (Fig. 7c) and UoC01-CLM4 (Fig. 7d), the blue solid line shows that the trend between observed and simulated values generally increases positively compared to other scheme combinations. However, in UoC04-CLM4---which showed the best performance in the verification---the model primarily overestimates values below approximately 180 $\mu g\ m^{-3}$ and exhibits wider dispersion with underestimation tendencies for values above 180 $\mu g\ m^{-3}$. In contrast, the other three combinations (Fig. 7a, b, and e) showed little to no correlation between observations and simulations, with a wider spread of data. Therefore, UoC04-CLM4 showed relatively better performance compared to the other scheme combinations.

(Original manuscript) Supplement

Fig. S3 shows a scatter diagram for UoC04-based combination---the dust emission scheme that showed the best prediction performance when combined with CLM4 in the verification (see Fig. 6). As mentioned earlier, the UoC04-CLM4 combination exhibited the highest correlation, followed by UoC04-RUC. In contrast, the UoC04-Noah and UoC04-Noah-MP showed no linear correlation (Figs. S3a, and c).

(Revised manuscript)

Fig. S3 shows a scatter diagram for the UoC04-based combinations---the dust emission scheme that showed the best prediction performance when combined with CLM4 in the verification (see Fig. 6). The UoC04-CLM4 combination showed the highest correlation between observed and simulated values among the UoC04-based combinations. In contrast, UoC04-Noah and UoC04-Noah-MP demonstrated little to no correlation, suggesting very low prediction reliability.

Specific Comments:

1. Does this mean you wrote the output at 1-hr intervals? The integration timestep had to be much shorter than this.

→ We appreciate the reviewer for pointing this out. The previous explanation was unclear. We have revised the manuscript to clarify this point accordingly.

(Original manuscript) Line 166

We run WRF-Chem with a 1-hour interval from the occurrence of ADSs in the source region to their complete disappearance in South Korea, including a spin-up time of 72 hours; ~

(Revised manuscript)

We ran WRF-Chem, including a 72-hour spin-up time, from the occurrence of ADSs in the source region until their complete disappearance in South Korea, and the model output was saved at 1-hour intervals; ~

2. MAE is referenced as MBE throughout the rest of the manuscript. Please revise for consistency.

→ We sincerely appreciate your thoughtful review. We have carefully checked the entire manuscript and corrected mention of MAE to MBE.

(Original manuscript) Line 281

;mean bias error (MAE) is the arithmetic average of ~

(Revised manuscript)

;mean bias error (MBE) is the arithmetic average of ~

3. 1.0 does not necessarily indicate a “perfect forecast.” It just indicates that all true events were successfully identified. The manuscript clarifies this in the following sentence, but “skillful” is more appropriate phrasing than “perfect.”

→ We appreciate your insightful comment. We agree that 1.0 indicates the successful identification of all true events, not a "perfect forecast". Based on your suggestion, I have revised "perfect" with "skillful".

(Original manuscript) Line 290

It ranges from 0 to 1, with 1 indicating a perfect forecast and ~

(Revised manuscript)

It ranges from 0 to 1, with 1 indicating a skillful forecast and ~

4. What is the basis of using 0.4 as the threshold for a “weak” correlation?

→ We appreciate the reviewer’s comment. After careful consideration, we realize that using 0.4 as a threshold for PCC was generalized without sufficient basis. Generally, interpretations of relationship strength may vary across disciplines (Turney et al., 2022). Therefore, we referred to the criteria used in air quality-related study (Cha et al, 2018).

Relationship strength:

- **Strong:** Values between ±0.7 and ±1 indicate a strong correlation.
- **Moderate:** Values between ±0.3 and ±0.7 indicate a moderate correlation.
- **Weak:** Values between ±0.1 and ±0.3 indicate a weak correlation.
- **No Correlation:** Values between 0 and ±0.1 indicate negligible correlation.

※ Reference: Analysis of fine dust correlation between air quality and meteorological factors using SPSS (Cha et al, 2018)

Therefore, based on the above, we have revised the manuscript as follows.

(Revised manuscript) Figure 4

→ We adjusted the minimum value of the y-axis in the figure to 0.3 and modified the caption accordingly.

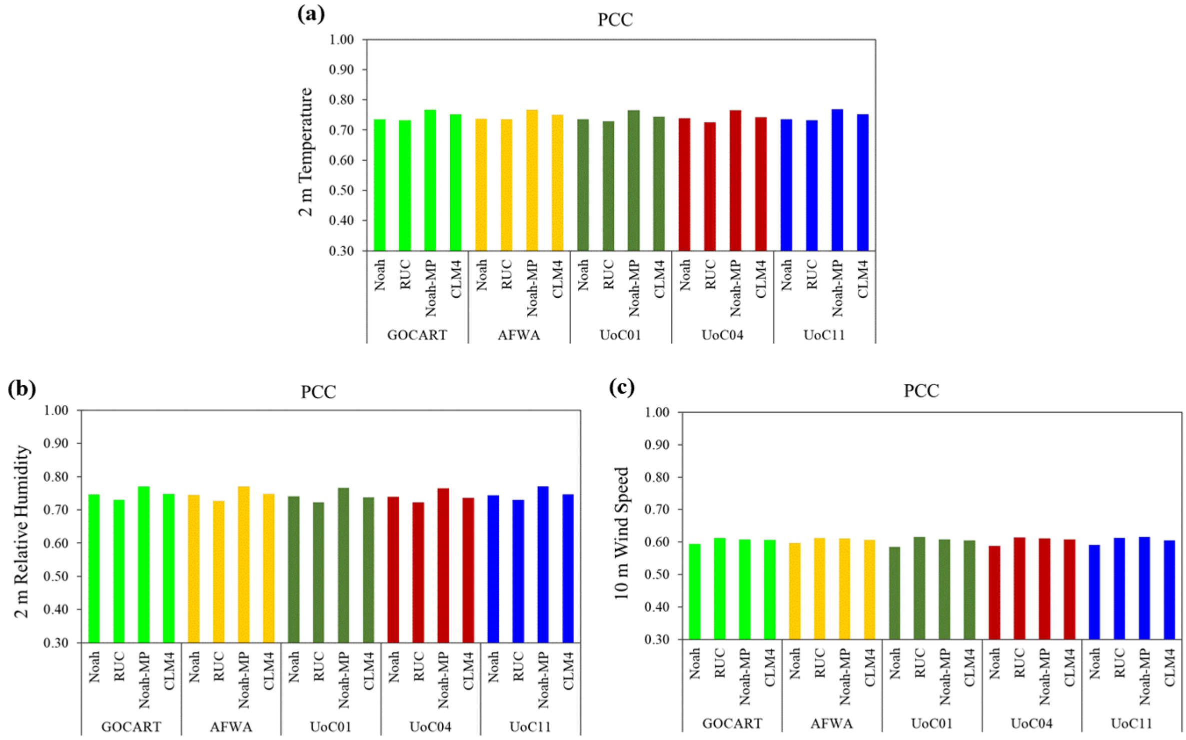

Figure 4: Pearson's correlation coefficient (PCC) of all scheme combinations for (a) T2m, (b) RH2m, and (c) WS10m, respectively, using the ASOS data. The y-axis represents values greater than 0.3, indicating the minimum threshold for a weak correlation. The values are averaged over the stations (see Fig. 3c).

(Revised manuscript) Figure 6

→ We adjusted the positions of the blue in (a) and modified the caption accordingly.

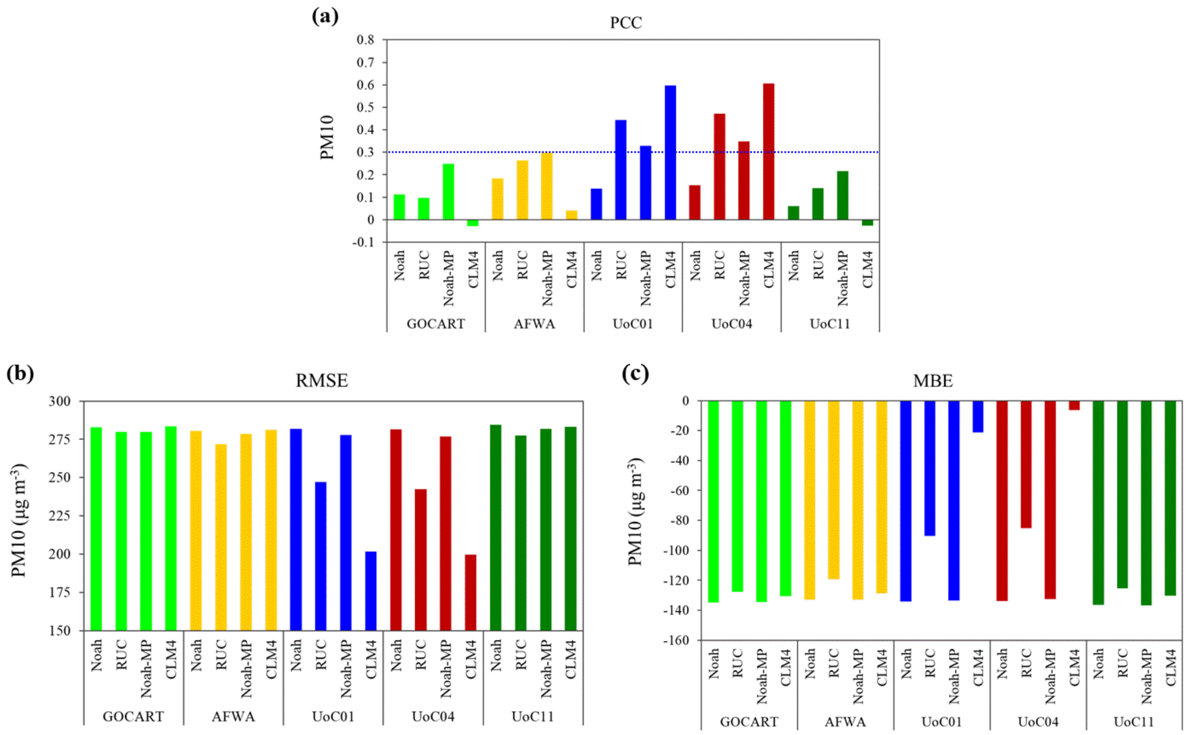

Figure 6: The verification results of all experiments for PM10 concentrations; (a) PCC, (b) RMSE, and (c) MBE, respectively, using the in-situ data. Based on PCC values, the blue dashed line represents the minimum threshold for a moderate correlation. The values are averaged over the stations (see Fig. 3c).

(Revised manuscript) Lines 360 ~ 361

In all scheme combinations except for combinations of UoC04 and UoC01 with CLM4 and RUC, PCC was below 0.3, indicating a weak or almost no correlation; ~

5. By my understanding, no forecasts were produced in this study (i.e., there was no attempt to predict the future). So, “forecasted” values is really referring to “modeled” or “simulated” values. (Line 346, 351, and elsewhere)

→ We appreciate the reviewer’s comment. We agree with the reviewer's comments and believe that the term “modeled” or “simulated” is more appropriate than “forecasted” in this study. So, we have revised all the relevant content as follows.

(Revised manuscript) Figure 5

→ We have changed 'Forecast' on the y-axis in Figure 5 to 'Simulation' and 'forecasted' in the caption to 'simulated'.

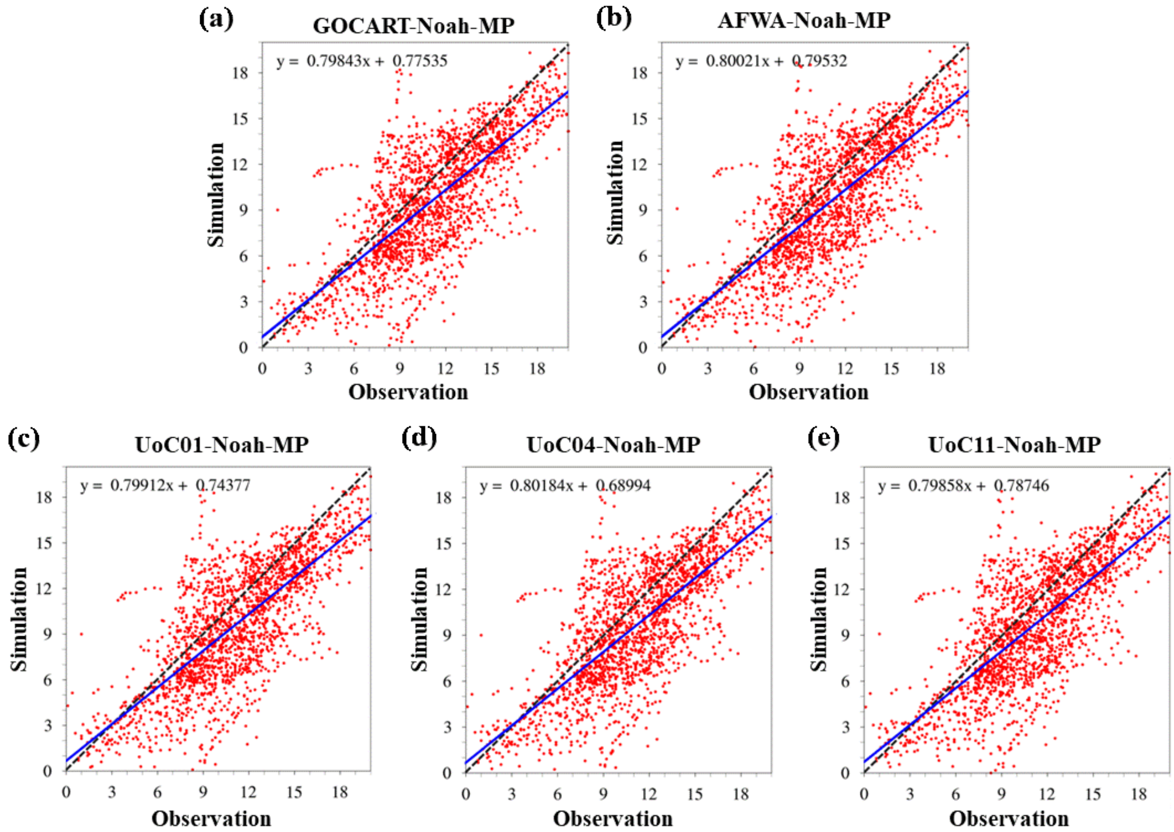

Figure 5: Scatter plots showing the relationship between observed and simulated values for T2m, using Noah-MP-based combinations. Each panel represents a different scheme combinations: (a) GOCART-Noah-MP, (b) AFWA-Noah-MP, (c) UoC01-Noah-MP, (d) UoC04-Noah-MP, and (e) UoC11-Noah-MP. The black dashed line represents that the simulation perfectly matches the observation. The blue line indicates the linear regression fit to the data, providing relationship between the observed and simulated values.

→ We have also changed 'Forecast' on the y-axis in Figure 7 and Figure S3 to 'Simulation'.