Response (Referee #3 comment)

Ms. Ref. No.: hess-2022-379

Title: Improving predictions of land-atmosphere interactions based on a hybrid data assimilation and machine learning method

This paper takes advantage of the opportunity provided by the abundance of data in the Heihe River basin to illustrate the importance of accurate soil moisture and LAI information for climate modeling in regions with highly heterogeneous land surfaces. The spatial and temporal variations of soil moisture and LAI in the WRF are realistically expressed by data assimilation and machin learning (DA+ML). After assimilating the state variables from observations or satellite remote sensing, both soil moisture content LAI values are increased, which then increases evapotranspriation in the model and futher reduces the air warming bias and dry bias in the simulation. The improved simulation shows more realistic oasis-desert boundary and the wind shield effect within the oasis. Overall, this is an excellent study in terms of capacity building that improves climate modelling through implementing detailed information of land characteristics in a basin with very complex underlying surfaces. Nevertheless, I think this paper can be organized better and some moderate revisions are required.
We gratefully thank the reviewers for their review, which we believe has led to significant improvement on the original manuscript. The original reviewers' comments are reproduced below in black text and the corresponding response is shown in blue text.

1. The scientific question to be answered is unclear. If the authors intend to answer a question of general interest, applying satellite data as an input to SM and LAI is understandable because they are globally accessible. However, the application of in situ soil moisture as an input, as done in this study, has no way to expand spatially. If the authors are trying to answer a scientific question specific to the Heihe Basin, the challenges of climate modeling in this basin should be addressed. In either case, it should be stated in the INTRODUCTION in the form of motivation.
Thank you for your good comment. This study highlights the coupling of a hybrid data assimilation and machine learning approach with the WRF model and evaluates its performance in regional climate simulations. Compared with the direct assimilation of coarse-resolution remotely sensed soil moisture, this method can improve the estimation of soil moisture and ET in the heterogeneous land surface by utilizing soil moisture profile observations. Therefore, it is necessary to incorporate this hybrid approach in regional climate models to implement detailed land characterization information in basins with complex underlying surfaces and improve climate modeling. In addition, machine learning methods have been widely used for soil moisture estimation over larger regions based on soil moisture observation networks (Li et al., 2022). Therefore, this approach can also be applied in other regions and globally.

The objective of this study is revised as: "Previous studies have also demonstrated the importance of the hybrid DA and ML method when estimating LAI, SM, and ET in typical arid/semi-arid regions of HRB. However, the advantages of improving the representation of soil and vegetation processes in affecting regional climate via the coupled DA and ML framework have not been fully exploited, especially in basins with complex underlying surfaces. Therefore, this study aims to investigate the improvement of the hybrid DA and ML framework for regional climate and land-atmosphere interactions in the HRB based on the WRF model and to further reveal its physical mechanisms."

The advantages regarding the hybrid model will be added to section 3.2: "Compared with the direct assimilation of coarse-resolution remotely sensed SM, this method can improve the estimation of SM and ET on the heterogeneous land surface. This is because *in situ* SM profile observations are used to construct an ML-based surrogate model to improve SM and ET estimation on complex underlying surfaces."

2. The structure and presentation of the paper could be imporved. (1) 4.1 should be verification of data assimilation rather than validation of model simulations of LAI and SM. LAI and SM are essentially an input (through ML+DA). Their realistic representaiton in the model verifies the correctness of the implementation in the model, but it does not mean the model's simulation capability. (2) 4.2 For land-air interaction simulations, the key linkage is sensible heat, latent heat (or evapotranspiration), and momentum fluxes, and there is a lack of description of sensible heat and momentum fluxes. (3) 4.3 Specific humidity and air temperature, which has been presented properly. (4) 4.4 Wind speed and precipitation. This part currently lacks quantitative assessment and the results are not convincing. I am not surprised by this, because the improved representations to SM and LAI are local and there are other errors in the model itself. Therefore, it is very difficult to improve wind speed and precipitation quantitatively. I think it is acceptable to consider this part as sensitivity analysis rather than evaluation. If so, this subsection can be much shorten, e.g., deleting Figures 13b-c and Figures 14-15.
Thanks for your comment. The structure of the manuscript was reorganized as: 4.1 Validation of the hybrid model; 4.2 Sensible and latent heat fluxes; 4.3 Air temperature and specific humidity; 4.4 Wind speed and precipitation.

The estimated sensible and latent heat fluxes from the WRF (OL) and WRF (DA-ML) models were validated against large aperture scintillometer (LAS) observations at the Arou, Daman, and Sidaoqiao sites. Compared to eddy covariance (EC) observations, scintillometer provided kilometer-scale H and LE and is widely used for the validation of remote sensing products and model simulations (Zheng et al., 2023).

The following paragraph will be added to section 4.2: "Figure 6 compares the daily H and LE estimates from the WRF (OL) and WRF (DA-ML) models with the LAS at the Arou, Daman, and Sidaoqiao sites. As indicated, the retrieved H and LE from the WRF (DA-ML) model agree well with the observations, and mainly fall around the 1:1 line. The WRF (DA-ML) model performs better than the WRF (OL) model because of the improved LAI and SM simulations. The statistics of turbulent heat flux estimates at the three sites are summarized in Table 2. The three-site-averaged RMSD of daily H and LE predictions for the WRF (OL) model are 53.61 and 63.73 W m^{-2}, respectively. The WRF (DA-ML) model decreases the abovementioned RMSDs by 43.74% and 23.98%. The relatively low RMSD values indicate that the WRF (DA-ML) model can accurately estimate turbulent heat fluxes over different sites with contrasting environmental conditions. The results also show that the simulated H and LE of the WRF (DA-ML) model are still higher and lower than the observed values at the Sidaoqiao site. This is because the spatial representation of the model simulation (3 km) is inconsistent with the LAS measurements (path length: 2350 m). This mismatch will introduce uncertainty in the validation results, especially in heterogeneous land surfaces (Zhang et al., 2022). In addition, the higher surface heterogeneity and complex hydrological processes in the downstream oasis affect the training accuracy of the ML method, which further affects the performance of the WRF (DA-ML) model (He et al., 2022)."

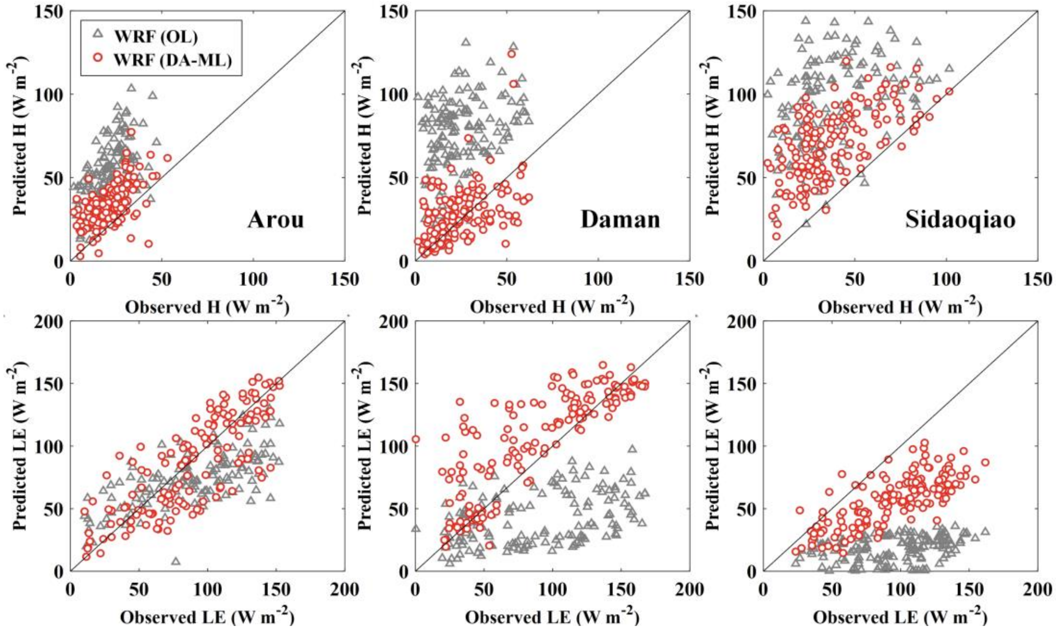

Figure 6: Scatterplot of daily sensible and latent heat flux estimates from the WRF (OL) and WRF (DA-ML) models versus measurements at the Arou, Daman, and Sidaoqiao sites.

Table 2: Statistical indices of daily H and LE estimates from the WRF (OL) and WRF (DA-ML) models at the Arou, Daman, and Sidaoqiao sites.

Site		H		LE	
		WRF (OL)	WRF (DA-ML)	WRF (OL)	WRF (DA-ML)
Arou	R^2 (-)	0.49	0.53	0.46	0.67
	RMSD (W m^{-2})	39.57	29.72	47.21	43.52
Daman	R^2 (-)	0.16	0.52	0.18	0.65
	RMSD (W m^{-2})	59.74	23.18	61.06	49.69
Sidaoqiao	R^2 (-)	0.19	0.48	0.12	0.56
	RMSD (W m^{-2})	61.53	37.59	82.93	52.13
Three-sites-average	R^2 (-)	0.28	0.51	0.25	0.63
	RMSD (W m^{-2})	53.61	30.16	63.73	48.45

Figures 13b-c provide more details of precipitation and are retained in the manuscript. To address the reviewer's comment, the original Figure 14 and 15 was removed from the manuscript.

3. Suggest to revise the title. The work of this paper is not a prediction but a simulation; land-air interactions are not presented: it shows the response of the atmosphere to the change of the surface state, but does not present the influence of the atmosphere on the land.

Thanks for your comment. The title is revised as "Improving regional climate simulations based on a hybrid data assimilation and machine learning method."

4. L329-332: How mountain winds affect the climate in the oasis and how the cooling/wetting affects the air temperature and humidity aloft should be further clarified. Particularly, the height of 600hPa was

chosen too arbitrarily. In the oasis and desert region, the influenced height is far lower than 600hPa. Later, the authors explained the phenomenon of warmer and drier aloft through horizontal advection between oasis and desert, but I guess the enhanced subsidence over the oasis in the WRF(DA-ML) is the cause.
Thanks for your comment. It is difficult to discuss the effect of mountain winds on oases from the current results. We need additional experiments, but it is beyond the scope of the paper. Therefore, the description of the mountain winds has been removed from the manuscript.

We removed "the height of 600hPa" in the manuscript. The sentence will be revised in section 4.3: "Moreover, the wetting and cooling effects of the oasis were mainly concentrated in the boundary layer, gradually decreased from the land surface upward, and were replaced by slightly warming and drying effects. Such warming and drying effects may be related to the enhanced subsidence over the oasis. The downward motion may result in increased temperatures and bring dry air from the upper atmosphere."

5. In section 4.3, about the simulated wind speed difference between the two cases: when you update LAI, do you update the vegetation height (or roughness length and zero-plain displacement) in the WRF?
Thanks for your comment. The roughness length and zero plane displacement are constant in the WRF model based on the different land cover types. It does not change with the update of LAI.

To avoid ambiguity, we have revised the relevant sentences as: "The mean wind vectors at 10 m during the growing season from the WRF (OL) and WRF (DA-ML) in the mid- and downstream oases are shown in Figure 13. By comparing the simulated wind speeds in the oasis and the surrounding desert, we found that crops, shelterbelts, and residential areas in the midstream oasis produced a wind-shield effect. The wind speed within the oasis is less than that of the surrounding desert because the drag force of crops, shelterbelts, and residential areas reduces the wind speed and also changes the wind direction."

Minor comments:
In relevant figures, please indicate where is desert and where is oasis; otherwise, it is hard to understand what you are describing.
Thanks for your good suggestion. The indicator about the oasis area was added to Figure 9, 10, and 12.

L179: what is "the standardized soil texture"
Thank you for your comment. It refers to the normalized soil texture (ST).

L195-196: "the WRF model and DA-ML method were coupled and run dynamically and consistently through the cycles of steps one and two." What is the time interval of the cycles? This is critical information for applications.
Thanks for your comment. The following paragraph is revised in section 3.2: "Eventually, the WRF model and the DA-ML method are coupled at the daily scale through the cycles of step one and two."

L368: what you mean by "divergent wind direction". I can understand the whole sentence neither.
Thanks for your comment. We will remove the description of the mountain winds in the manuscript.

L390: In the upper atmosphere, air masses enhanced the background northerly winds (orange areas)? It is hard to understand.

Thanks for your comment. The meridional circulation of the mid- and downstream oasis is added to the appendix. According to Figure 12 and A1, the airflow from the desert to the oasis in the upper atmosphere can enhance the background northerly winds. This can be observed in the background color of Figure 12 (orange represents enhanced wind speed).

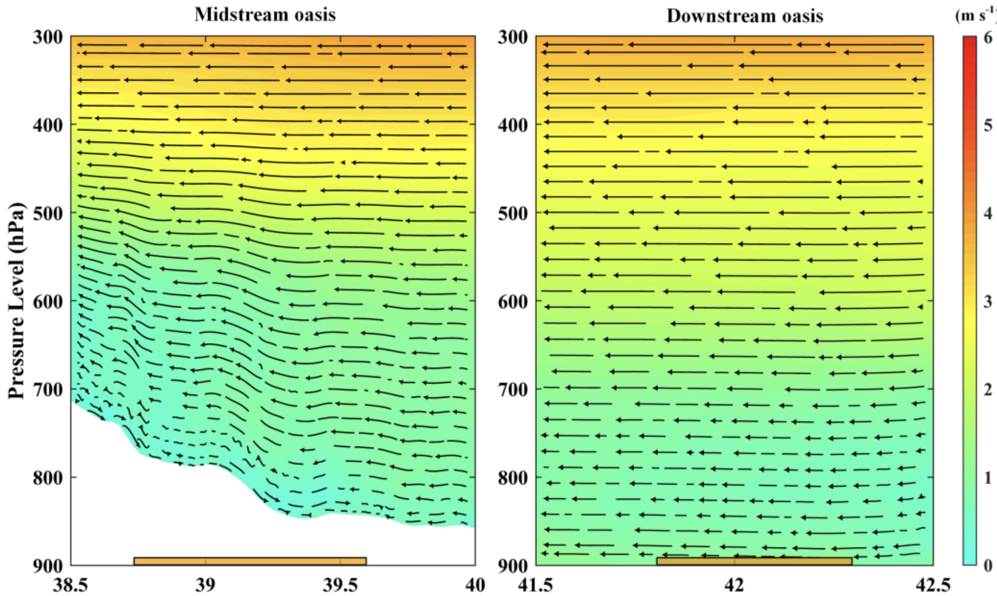

Figure A1. The zonal mean vertical velocity and meridional circulation from the WRF (DA-ML) model in the mid- and downstream oasis. The orange bar represents the oasis area.

The sentence is revised in section 4.4: "In the upper atmosphere, the desert to oasis air masses enhance the background northerly winds (Figure A1), which promote atmospheric water vapor transport in the HRB."

L391-395, and some similar sentences: it is not the focus of this work to study the ecological effect, which has been discussed in many early studies. Please delete.

Thanks for your comment. We deleted the relevant descriptions.

L396-397: You have not established the causality among these components.

Thanks for your comment. We deleted the relevant descriptions.

L415-418: Figure 12 shows downslope wind, so how could it transfer water vapor from the oasis upslope. There are some similar issues (e.g. L424-425). The authors must be more cautious to draw a conclusion.

Thanks for your comment. The meridional circulation of the mid- and downstream oasis is added to the appendix. As shown in Figure A1, the water vapor transport in the HRB is predominantly controlled by polar northerly winds. Driven by background northerly winds, more water vapor fluxes from the midstream oasis region were carried to the upstream region. This atmospheric water vapor transport can enhance precipitation in upstream mountainous regions (Zhang et al., 2017a).

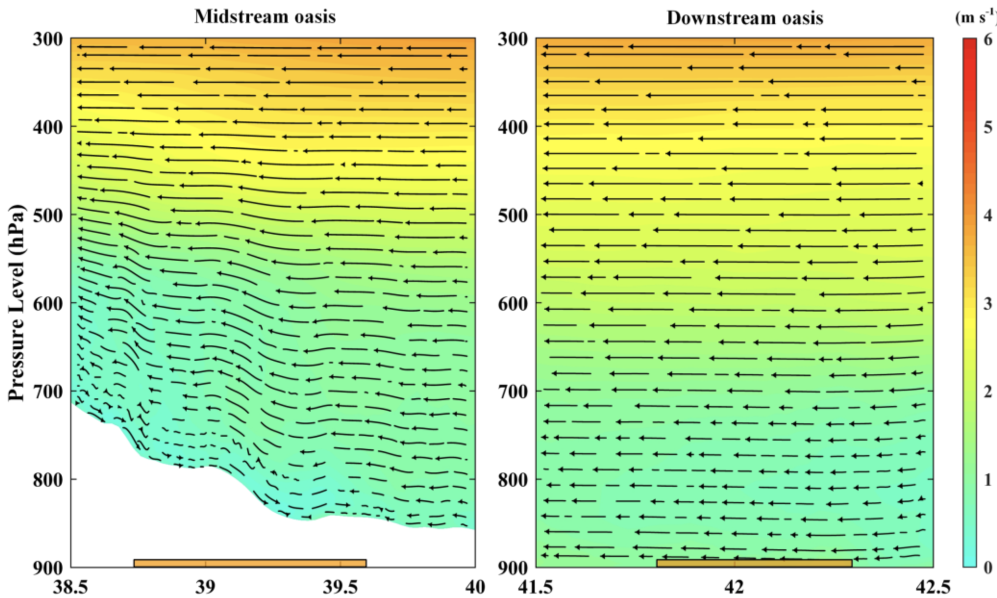

Figure A1. The zonal mean vertical velocity and meridional circulation from the WRF (DA-ML) model in the mid- and downstream oasis. The orange bar represents the oasis area.

The sentence in lines 424-425 of the manuscript was deleted.

L 474: "resembled the rainfall, vegetation cover, irrigation event, and shallow groundwater table features." Is it the conclusion of this study?
Thanks for your comment. It is not the conclusion. This sentence was removed from the conclusion section.

L476: You have not presented "the simulated seasonal cycles of air temperature". Instead, you only give the seasonal mean!
The sentence is revised in section 5: "Compared to the WRF model, the seasonal mean air temperature and specific humidity simulated by the WRF (DA-ML) at the nine sites were closer to the station measurements."

References:

Li, Q., Shi, G., Shangguan, W., Nourani, V., Li, J., Li, L., Huang, F., Zhang, Y., Wang, C., Wang, D., Qiu, J., Lu, X., Dai, Y., 2022. A 1 km daily soil moisture dataset over China using in situ measurement and machine learning. Earth Syst. Sci. Data 14, 5267–5286. https://doi.org/10.5194/essd-14-5267-2022.

Zhang, X., Xiong, Z., Tang, Q., 2017a. Modeled effects of irrigation on surface climate in the Heihe River Basin, Northwest China: Modeling Effects of Irrigation. J. Geophys. Res. Atmos. 122, 7881–7895. https://doi.org/10.1002/2017JD026732.

Zhang, M., Luo, G., Hamdi, R., Qiu, Y., Wang, X., Maeyer, P.D., Kurban, A., 2017b. Numerical Simulations of the Impacts of Mountain on Oasis Effects in Arid Central Asia. Atmosphere 8, 212. https://doi.org/10.3390/atmos8110212.

Zhang, Y., Liu, S., Song, L., Li, X., Jia, Z., Xu, T., Xu, Z., Ma, Y., Zhou, J., Yang, X., He, X., Yao, Y., Hu, G., 2022. Integrated Validation of Coarse Remotely Sensed Evapotranspiration Products over Heterogeneous Land Surfaces. Remote Sens. 14(3467). https://doi.org/10.3390/rs14143467.

Zheng, C., Liu, S., Song, L., Xu, Z., Guo, J., Ma, Y., Ju, Q., Wang, J., 2023. Comparison of sensible and latent heat fluxes from optical-microwave scintillometers and eddy covariance systems with respect to surface energy balance closure. Agric. For. Meteorol. 331, 109345. https://doi.org/10.1016/j.agrformet.2023.109345.