Improving regional climate simulations based on a hybrid data assimilation and machine learning method

Xinlei He[1], Yanping Li[2], Shaomin Liu[1*], Tongren Xu[1], Fei Chen[3], Zhenhua Li[2], Zhe Zhang[2], Rui Liu[4], Lisheng Song[5], Ziwei Xu[1*], Zhixing Peng[1], Chen Zheng[6]

[1]State Key Laboratory of Earth Surface Processes and Resource Ecology, School of Natural Resources, Faculty of Geographical Science, Beijing Normal University, Beijing, China
[2]School of Environment and Sustainability, University of Saskatchewan, Saskatoon, SK, Canada
[3]National Center for Atmospheric Research, Boulder, CO, USA
[4]Institute of Urban Study, School of Environmental and Geographical Sciences (SEGS), Shanghai Normal University, Shanghai, China
[5]School of Geography and Tourism, Anhui Normal University, Wuhu, China
[6]Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

Correspondence to: Shaomin Liu (smliu@bnu.edu.cn), Ziwei Xu (xuzw@bnu.edu.cn)

Abstract. The energy and water vapor exchange between the land surface and atmospheric boundary layer plays a critical role in regional climate simulations. This paper implemented a hybrid data assimilation and machine learning framework (DA-ML method) into the Weather Research and Forecasting (WRF) model to optimize surface soil and vegetation conditions. The hybrid method can integrate remotely sensed leaf area index (LAI), multi-source soil moisture (SM) observations, and land surface models (LSMs) to accurately describe regional climate and land-atmosphere interactions. The performance of the hybrid method on the regional climate was evaluated in the Heihe River Basin (HRB), the second largest endorheic river basin in Northwest China. The results show that the estimated sensible (H) and latent heat (LE) fluxes from the WRF (DA-ML) model agree well with the large aperture scintillometer (LAS) observations. Compared to the WRF (OL), the WRF (DA-ML) model improved the estimation of evapotranspiration (ET) and generated a spatial distribution consistent with the ML-based watershed ET (ETMap). The proposed WRF (DA-ML) method effectively reduces air warming and drying biases in simulations, particularly in the oasis region. The estimated air temperature and specific humidity from WRF (DA-ML) agree well with the observations. In addition, this method can simulate more realistic oasis-desert boundaries, including wetting and cooling effects and wind shield effects within the oasis. The oasis-desert interactions can transfer water vapor to the surrounding desert in the lower atmosphere. In contrast, the dry and hot air over the desert is transferred to the oasis from the upper atmosphere. The results show that the integration of LAI and SM will induce water vapor intensification and promote precipitation in the upstream of the HRB, particularly on windward slopes. In general, the proposed WRF (DA-ML) model can improve climate modeling by implementing detailed land characterization information in basins with complex underlying surfaces.

1 Introduction

35   Land-atmosphere interactions are an essential component of the hydrological cycle and factors that influence climate change (Nelli et al., 2020; Zhou et al., 2022). Terrestrial components, such as soil and vegetation, play a crucial role in atmospheric processes, such as changes in evapotranspiration (ET), which affect the water vapor content in the atmosphere (Gentine et al., 2019; Sawada et al., 2015; Wu et al., 2023). Soil and vegetation processes directly affect surface water vapor transport and energy circulation, particularly in the arid vegetated area (Erlandsen et al., 2017; Gao et al., 2008; Liu et al.,

40   2018a; Zhang et al., 2017a, 2019; Zhao et al., 2021). Such surface variability affects the available energy distribution at the land surface and has additional effects on the sensible and latent heat fluxes (H and LE), surface temperature, and water vapor (Sawada et al., 2015; Wen et al., 2012). Although land surface models (LSMs) have been improved incrementally in the past decades, it is still challenging to effectively couple LSMs with atmospheric models to improve the description of land-atmosphere interactions (Chen and Dudhia, 2001; Liu et al., 2021). The success of the coupling depends not only on the

45   sophisticated physical processes of the LSMs, but also on soil and vegetation characteristics and the accurate characterization of the water vapor fluxes at the land-atmosphere interface (Gentine et al., 2019; Zhang et al., 2021b, 2022).

      The development of earth observation technology has provided important opportunities to study land-atmosphere interactions using the data assimilation (DA) method (Liang et al., 2021). The Land Data Assimilation System (LDAS) has been widely developed and applied in recent years under various hydrological and vegetation conditions (Wu et al., 2022;

50   Xia et al., 2019). It uses remotely sensed observations to constrain model physical processes and empirical parameters to improve water-energy-carbon flux simulations (Tian et al., 2022; Zhao and Yang, 2018). A series of studies have assimilated satellite-retrieved leaf area index (LAI), land surface temperature (LST), soil moisture (SM), and microwave brightness temperature observations into LSMs and improved simulations of ET, runoff, and gross primary productivity (GPP) (Ahmad et al., 2022; He et al., 2020b, 2021; Ling et al., 2019; Seo et al., 2021; Xie et al., 2017; Xu et al., 2019, 2021). In addition,

55   DA can improve the initial conditions of regional climate models (RCMs) and enhance the capability of the models in simulating land-atmosphere interactions (Pan et al., 2017; Yi et al., 2021). Several studies have also shown that the assimilation of air pressure, air temperature, humidity, wind speed, and lightning observations into the Weather Research and Forecasting (WRF) model can improve the simulation of atmospheric state variables and the accuracy of weather prediction (Campo et al., 2009; Cazes Boezio and Ortelli, 2019; Comellas Prat et al., 2021; Grzeschik et al., 2008; Pilguj et al., 2019).

60   Machine learning (ML) algorithms have been increasingly applied in earth and environmental modeling studies to predict land surface variables at various spatial and temporal scales (Jung et al., 2020; Reichstein et al., 2019; Xu et al., 2018). Compared to physical models, ML technology can fluently and accurately establish non-linear and complex relationships between diverse independent variables (Koppa et al., 2022; Nearing et al., 2018). Thus, ML-based approaches can create beneficial pathways for knowledge discovery in process models based on extensive data (Moosavi et al., 2021;

65   Reichstein et al., 2019). The main improvements are focused on model approximation, parameterization, bias correction, and hybrid modeling (Brajard et al., 2020; He et al., 2022; Jia et al., 2021; Xu et al., 2014; Zhao et al., 2019). Several studies

have shown that the integration of the DA and ML methods can enhance the reliability of predictions and reduce simulation errors by including physical information in observed data (Brajard et al., 2020; Buizza et al., 2022; Forman and Xue, 2017; Gottwald and Reich, 2021; He et al., 2022). Forman and Xue (2017) integrated a ML model (as a measurement operator)

70 into a DA system to improve the estimation of snow water equivalent. Zhao et al. (2019) used a physics-constrained ML method to improve the LE estimates. He et al. (2022) proposed a hybrid model that can integrate remotely sensed LAI and multi-source SM observations to improve the estimation of ET within the coupled DA and ML framework. In general, the hybrid DA and ML approach was used in land surface modeling, but it is neglected in regional climate simulations.

The Heihe River Basin (HRB) is a typical endorheic river basin in the arid and semi-arid regions of Northwest China

75 (Li et al., 2013). The upstream mountain region is mainly covered by alpine meadow and Qinghai spruce, has a complex topography, and receives abundant precipitation. The midstream oasis of the HRB is mainly characterized by irrigated croplands, while the downstream oasis is characterized by riparian forests, and at the periphery of the oasis is vast desert (Xu et al., 2020). In the HRB, precipitation is the main water resource input in mountainous areas and determines the growth of vegetation in the oasis region, as well as supporting urban and population development (Li et al., 2018, 2021). Precipitation,

80 snow, and permafrost changes in the upstream mountains can affect mountain runoff and SM, evaporation, groundwater table in the mid- and downstream oases. Strong land-atmosphere interactions in the HRB affect the water and energy exchange between the surface and atmosphere and influence the sustainability of the oasis (Gao et al., 2008; Pan et al., 2021b). The oasis-desert local circulation in the HRB can lead to the microclimate features in oasis-desert areas, which include the cooling and wetting effect and wind shield effect of the oasis, and the humidity inversion effect within the

85 surrounding desert (Liu et al., 2020).

In recent decades, several comprehensive experiments have been implemented over the HRB to study land-atmosphere interactions, including the Heihe River Basin Field Experiment (Hu et al., 1994), Watershed Allied Telemetry Experimental Research (WATER) (Li et al., 2009), and Heihe Watershed Allied Telemetry Experimental Research (HiWATER) (Li et al., 2013). In recent years, many mesoscale climate models and high-resolution computational fluid dynamics (CFD) models

90 have been used to analyze the effects of land-atmosphere interactions on the regional climate (Liu et al., 2018a, 2020; Xie et al., 2018; Zhang et al., 2017a, 2021a). Zhang et al. (2017a) added an irrigation scheme to the WRF model and identified strong cooling and wetting effects on irrigated cropland in the midstream of the HRB. Liu et al. (2020) investigated the oasis-desert microclimate effects based on an improved CFD model and found that the oasis had a cold and wet island effect and wind shield effect. Zhang et al. (2021b) applied the WRF-Hydro model in the HRB and emphasized the role of lateral

95 flow in the regional precipitation circulation. These studies illustrate that mesoscale climate models can be used as essential tools to better understand regional climate and land-atmosphere interactions in the 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

100 in the HRB based on the WRF model and to further reveal its physical mechanisms.

The goals of this study were to (1) couple the hybrid DA and ML (DA-ML) framework to the WRF model and improve the estimation of LAI and SM; (2) validate the H and LE with the large aperture scintillometer (LAS) observations and compare the ET estimates with the ML-based watershed ET; (3) investigate the performance of the air temperature and specific humidity estimates and evaluate the effects of the hybrid framework on near-surface atmospheric conditions in the mid- and downstream oasis regions; and (4) discuss the effects of the hybrid framework on weed speed and precipitation in the HRB.

2 Study area and dataset

The HRB (37.7°–42.7 °N, 97.1°–102.0 °E) is the second largest endorheic river basin in Northwest China and has an area of approximately 143,000 km^2, and the elevation ranges from 800 to 5000 m (Figure 1). The annual precipitation is approximately 400 mm, and it gradually decreases from the upstream region of the HRB (south) to the downstream region (north). Land cover types exhibit spatial zonation in the HRB. The upstream region is a typical mountainous environment, including extensive alpine meadows, a few Qinghai spruces, glaciers, snow, and permafrost. The midstream region is spatially composed of oasis-desert ecosystems, and irrigated cropland is the main component of the oasis in this area. The downstream region of the HRB is covered mainly by desert and riparian ecosystems (Populus euphratica and tamarisks). Water vapor transport in this region is predominantly controlled by mid-latitude westerly and polar northerly winds (Pan et al., 2021b). As a result, precipitation over the HRB shows strong spatial variability, with more than 70% occurring in the upstream mountains (Wang et al., 2018a, b). Nine fluxes and meteorological stations selected in the Heihe integrated observatory network were used for comparison with the model results (Figure 1). Among them, the Arou, Dashalong, and Hulugou stations in the upstream region of the HRB are covered by alpine meadows, while the Daman, Wetland, and Huazhaizi stations in the midstream region are covered by irrigated cropland, wetlands, and desert, respectively. The Sidaoqiao, Mixed forest, and Desert stations in the downstream regions are covered by Populus euphratica, tamarisk, and desert, respectively. More details on the *in-situ* information and measurement instruments can be found in Chen et al. (2014), Li et al. (2013), and Liu et al. (2018b).

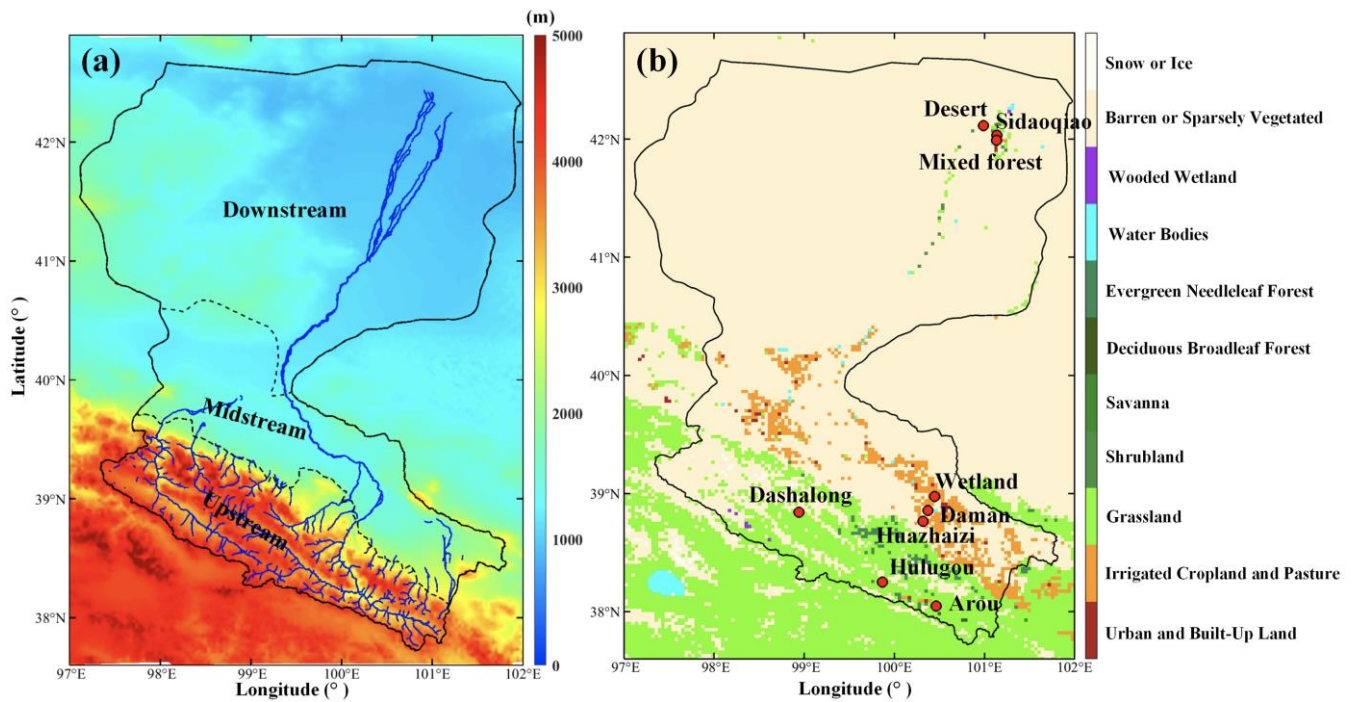



Figure 1: (a) Land surface elevation, (b) meteorological stations and land cover types in the study area. The blue and black solid lines represent the river and the boundary of the HRB, respectively.

To provide a high-resolution land cover and soil texture dataset that matched the WRF simulation period, the regional land cover and soil texture product generated by Zhong et al. (2014) and Song et al. (2016) with a spatial resolution of 30 m was employed. These datasets were downloaded from the National Tibetan Plateau Data Center (TPDC) (Pan et al., 2021a) (https://data.tpdc.ac.cn/en/). The elevation data were generated by NASA's Shuttle Radar Topography Mission (SRTM3, 90 m), which was obtained from the Geospatial Data Cloud (http://www.gscloud.cn/). The assimilated LAI data were retrieved from the Global Land Surface Satellite (GLASS) product with a spatial resolution of 1 km (Xiao et al., 2014; http://www.glass.umd.edu/). Daily LAI observations were generated by linearly interpolating the original 8-day GLASS LAI product. The GLASS product has been demonstrated to have better accuracy than Moderate Resolution Imaging Spectroradiometer (MODIS) and Advanced Very High Resolution Radiometer (AVHRR) and provides time-space continuous LAI estimation (Xiao et al., 2014). The daily Soil Moisture Active Passive (SMAP) SM product (https://appeears.earthdatacloud.nasa.gov/), with a spatial resolution of 9 km, was integrated into the hybrid framework.

In this study, SM observations from the eco-hydrological wireless sensor networks (WSN) in the up- and midstream of the HRB are used as an independent validation to evaluate the SM estimates from the WRF (DA-ML). The validation SM dataset in the upstream regions was mainly covered by grassland and obtained by averaging SM observations from 40 nodes. There are 9 network nodes installed in the LAS source area at Daman station that measured SM at the depths of 10 cm every 5 minutes (Che et al., 2019; Liu et al., 2018) (https://data.tpdc.ac.cn/en/). The half-hourly H was measured by the LAS

instrument in Arou, Daman, and Sidaoqiao sites. LE in these sites was obtained as the residual method of the energy balance equation (Liu et al., 2016a). Compared to eddy covariance (EC) observations, scintillometer provided kilometer-scale H and

145 LE and is widely used for the validation of remote sensing products and model simulations (Zheng et al., 2023). ETMap is a ML-based watershed ET product (daily/1 km) based on EC observations, remote sensing data, meteorological data, and the random forest method (Xu et al., 2018). The retrieved ET from the ETMap agrees well with the LAS observations. The multi-site averaged R^2, root mean square error (RMSE), and mean absolute percentage error (MAPE) values are 0.68, 0.85 mm day^{-1}, and 20.27%, respectively. These results confirm that the ETMap can effectively validate watershed ET

150 simulations. In this study, nine automatic weather station (AWS) observations were used to validate the WRF meteorological simulations. The AWS variables at each station include the wind speed/direction, air temperature/humidity, precipitation, air pressure, four-component radiation, soil heat flux, and soil temperature/moisture profile, etc. (Liu et al., 2018b). The data recorded at these stations were obtained from the Heihe integrated observatory network released by the TPDC (Liu et al., 2018b). The locations of the stations are shown in Figure 1. Among them, the Hulugou weather station observations are

155 produced by Chen et al. (2014). Precipitation datasets from the China Meteorological Forcing Dataset (CMFD) (He et al., 2020a) and atmospheric forcing data (AFD) in the HRB (Pan et al., 2014) were compared with the WRF (DA-ML) simulations. These datasets were obtained from the TPDC. The CMFD generates a gridded meteorological dataset with a spatial resolution of 0.1° and a temporal resolution of 3 h by fusing observations from 740 operational stations of the China Meteorological Administration. Pan et al. (2014) generated gridded atmospheric forcing data using the WRF model over the

160 HRB at an hourly 0.05° resolution. These datasets have been widely used as input data for various models as well as for environmental and climate change analyses (Xu et al., 2019; Zhang et al., 2016).

3 Methodology

3.1 WRF model setup

The advanced research WRF model version 4.0.3 (Skamarock et al., 2019) was used in this study. The WRF is a state-

165 of-the-art numerical weather and climate model designed by the National Center for Atmospheric Research (NCAR) for meteorological research and numerical weather predictions (Wang et al., 2021). The model source code is available at the official repository for WRF (https://github.com/wrf-model/WRF). The model domain covering the HRB consisted of two-way nested domains with 9 km and 3 km grid spacing. Only the simulation results for the 3 km grid were used in this study (Figure 1). This high-resolution setting excludes the uncertainty of the cumulus parameterization and thus simulates the soil-

170 precipitation feedback more realistically (Prein et al., 2015). In the vertical direction, 28 vertical sigma levels from the surface to 50 hPa were used. The atmospheric lateral boundary conditions in the WRF model were provided by the ERA5 reanalysis data set with a 0.25° spatial resolution and hourly temporal resolution (https://cds.climate.copernicus.eu/cdsapp#!/search?type=dataset). It is widely used in WRF simulations and provides boundary and initial conditions (Liu et al., 2021; Ma et al., 2022). The land cover, soil texture, elevation, and GLASS LAI

175   dataset were resampled to 3 km to be consistent with the model simulation resolution. The physical parameterization schemes selected in this study included the Rapid Radiative Transfer Model longwave and shortwave radiation scheme (Mlawer et al. 1997), Thompson microphysics scheme (Thompson et al., 2008), Mellor–Yamada–Janjic planetary boundary layer scheme (Janjic, 1994), and Noah-MP land surface scheme (Yang et al., 2011). The time step was 30 s, and the time resolution of the model output was hourly. Further details regarding the WRF model setup are presented in Table 1. The

180   dynamic vegetation parameterization scheme was turned on in the Noah-MP model to generate dynamic LAI simulations. We resampled the spatial resolution of the WRF simulations to 1 km with the bilinear interpolation method for comparison with the station observations.

Table 1: WRF model setup.

Simulation period	May to September, 2015
Horizontal grid spacing	9 km (100×181 grid points), 3 km (181×220 grid points)
Vertical levels	28 sigma-levels
Forcing data	ERA5 ($0.25°$, hourly)
Spin-up time	April (one month)
Longwave radiation	RRTM scheme (Mlawer et al., 1997)
Shortwave radiation	RRTM scheme (Mlawer et al., 1997)
Microphysics (MP)	Thompson scheme (Thompson et al., 2008)
Planetary boundary layer (PBL)	Mellor–Yamada–Janjic scheme (Janjic, 1994)
Surface layer	Eta similarity scheme (Janjic, 1994)
Land surface model (LSM)	Noah-MP land surface model (Yang et al., 2011)

3.2 Hybrid model

185   In this study, the hybrid model proposed by He et al. (2022) based on the DA and ML methods was incorporated into the WRF model to improve the LAI, SM, and ET simulations. The hybrid approach relies on the DA method to update the vegetation dynamics of the Noah-MP model and the ML method to construct a three-layer SM surrogate model. Compared with the direct assimilation of coarse-resolution remotely sensed SM, the hybrid model 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

190   surrogate model to improve SM and ET estimation on complex underlying surfaces.

    In the DA part, the remotely sensed LAI was assimilated using the ensemble Kalman filter (EnKF) method to update the leaf biomass (LFMASS) and optimize the specific leaf area (SLA) in the Noah-MP model. LAI is estimated as the product of leaf biomass predictions and SLA (LAI = LFMASS $\times$ SLA) in the Noah-MP model. Model ensembles were generated by adding normally distributed random errors to the model states (LFMASS) and parameters (SLA). The ensemble

195   size is set as 40 to ensure an accurate approximation of the error covariances while maintaining computational efficiency (Seo et al., 2021). Normally distributed errors with a mean of zero and a standard deviation of 10 g m^{-2} were added to the

LFMASS (Ahmad et al., 2022; Xu et al., 2021). The standard deviation of the SLA was set to 10% of the default parameter (Xu et al., 2021). Further, a uniform observation error standard deviation of 0.1 (-) was added to the remotely sensed LAI (He et al., 2022; Xu et al., 2021). These relevant statistical values have been widely used in previous LAI DA studies (Ahmad et al., 2022; Ling et al., 2019; Rahman et al., 2022).

In the ML part, the normalized soil texture (ST), land cover (LC), air temperature and humidity (Ta and RH), wind speed (U), precipitation (P), solar radiation (Rs), LAI, and SM observations were used to construct the SM surrogate model. ST, LC, Ta, RH, U, P, Rs, and LAI are the predictor variables. The *in-situ* SM profile observations (from 19 automatic weather stations) and SMAP SM products in the HRB are used as target variables to train and test the SM surrogate model. The extreme gradient boosting (XGBoost) method was chosen in the SM surrogate model to improve multi-layer SM simulations. The first layer (the top 0.1 m) of *in-situ* SM observations and SMAP SM were trained to establish the surface layer ML model. The averaged second (0.1–0.4 m) and third (0.4–1.0 m) layers of *in-situ* SM observations were used to construct the root zone ML models. The SM observations at different depths were averaged to be consistent with the Noah-MP model soil layer. The number of SM training samples in the first, second, and third layers are 9824, 7804, and 7793, respectively. A ten-fold testing method is employed to examine the performance of each ML method. In each fold, 90% of the training samples are used to train the model, and the remaining 10% of the data is used to test the model. The SM surrogate model can consider the effects of midstream irrigation events and downstream shallow groundwater tables on SM and improve Noah-MP ET estimates. More details regarding this method can be found in He et al. (2022).

The coupled land-atmosphere DA-ML system consists of two steps. In the first step, the meteorological forcing data were generated from the WRF model at time t. Then, the meteorological forcing data and initial states were input into the Noah-MP model to simulate the LAI, SM, and ET. In the second step, the hybrid DA-ML method was used to update the LAI and SM estimates in the Noah-MP model at time $t + 1$. In this step, the LAI is updated by the DA method, and the SM is updated via the ML-based surrogate model. The updated soil and vegetation conditions were re-fed into the WRF model and affected the atmospheric structure and state. Eventually, the WRF model and the DA-ML method are coupled at the daily scale through the cycles of step one and two. The flowchart of the hybrid DA and ML method coupled with the WRF model is shown in Figure 2.

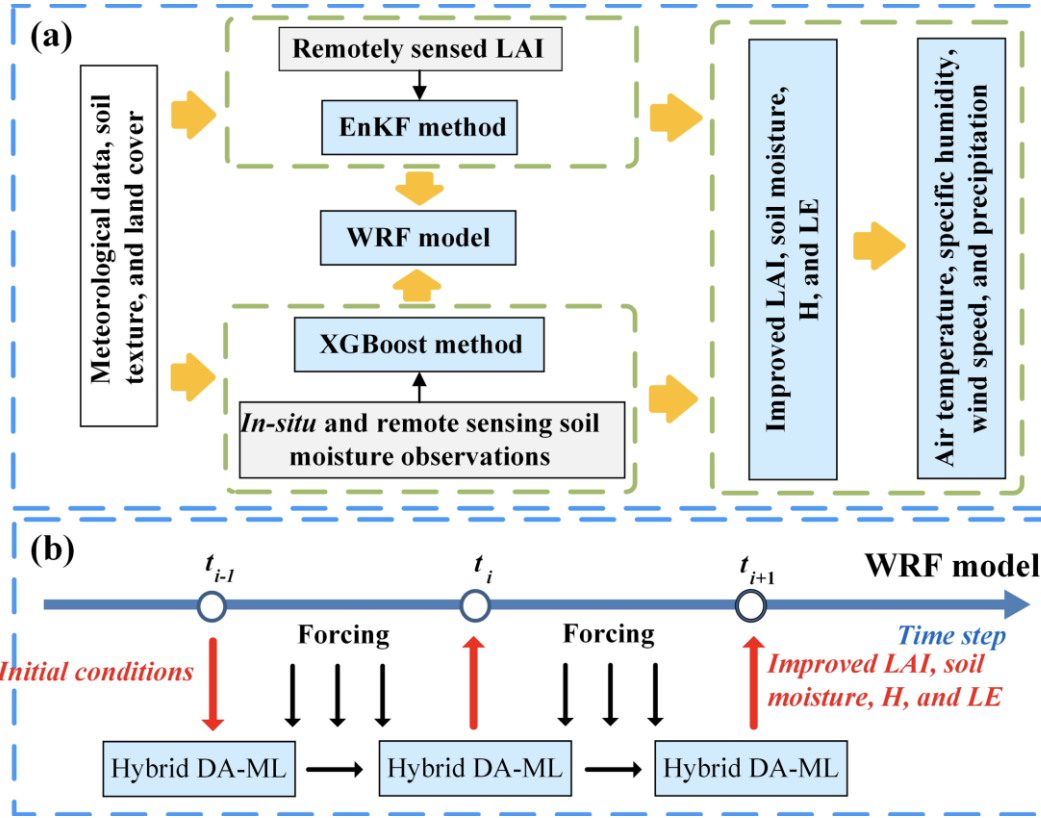



Figure 2: (a) Details of the hybrid DA and ML method, and (b) flowchart of the coupling with the WRF model.

Two experiments were conducted using the WRF model to investigate the effects of the DA-ML method. These two
225 experiments consisted of an experiment under natural conditions without DA and ML [WRF (OL)], and another experiment
implementing DA and ML [WRF (DA-ML)]. The simulation covered the period of vegetation growing season (from April 1
to September 30, 2015), and the first month was used for the spin-up. The computation details about the WRF (OL) and
WRF (DA-ML) are shown in Table A1. The differences between the WRF (DA-ML) and WRF (OL) simulations were used
to investigate the effects of LAI and SM integration. The root mean square deviation (RMSD) and coefficient of
230 determination (R^2) statistical metrics were used to evaluate the performance of the WRF (DA-ML) model,

$$\text{RMSD} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{1}$$

$$R^2 = \frac{[\sum_{i=1}^{n}(P_i - \overline{P})(O_i - \overline{O})]^2}{\sum_{i=1}^{n}(P_i - \overline{P})^2 \sum_{i=1}^{n}(O_i - \overline{O})^2} \tag{2}$$

where P_i and O_i are the predicted and observed values at time step i, respectively. $\overline{P}$ and $\overline{O}$ represents the mean values of P_i and O_i.

4 Results and discussion

4.1 Validation of the hybrid model

Figure 3 shows the monthly averaged LAI estimates from the WRF (OL), WRF (DA-ML), and GLASS products. As indicated, the WRF model failed to capture the magnitude and seasonality of the LAI. This is because the simulation of LAI dynamics in Noah-MP is controlled by the planting date, harvest date, and growing degree days in the cropland (Liu et al., 2016). In addition, inaccurate specification of the SM saturation, Vcmax25, and Clapp-Hornberger b parameter affects photosynthesis and biomass accumulation in vegetation (Cuntz et al., 2016; Levis et al., 2012). All these parameters are site-specific and empirical, and cannot be easily applied across regions. The assimilation systematically increased LAI during the growing season, and a significant increment in LAI was observed in mid-summer (June-August). The seasonal pattern of the WRF (DA-ML) was more consistent with that of the GLASS LAI than the WRF model, which indicates that the WRF (DA-ML) provides essential information for modeling vegetation dynamics. The simulated LAIs from WRF (OL) in the cropland, grassland, forest, and shrubland areas were 1.12, 1.05, 1.49, and 0.33 m^2 m^{-2}, respectively, all of which were lower than that of the GLASS LAI. After assimilation, the simulated bias of the LAI from WRF (DA-ML) in the HRB can be reduced from 0.94 to 0.11 m^2 m^{-2}. The results also show that the WRF (DA-ML) systematically overestimates the LAI, especially in the cropland. This is because, in addition to LAI assimilation, the integration of multi-source SM observations also affects the LAI dynamics.

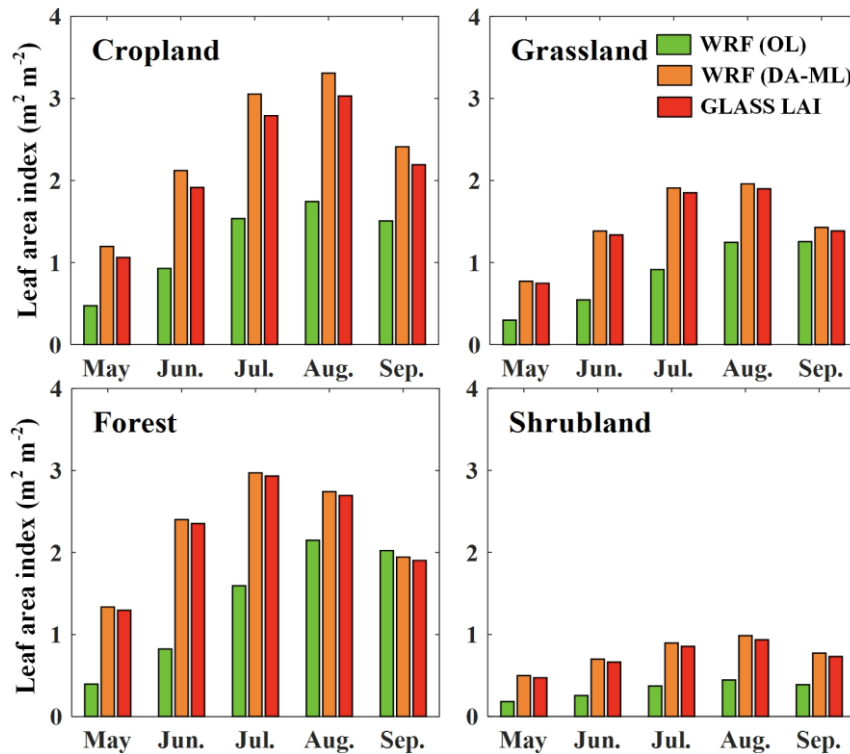



Figure 3: Seasonal variations of the LAI estimates for cropland, grassland, forest, and shrubland in the HRB.

The SM estimates from the WRF (OL) and WRF (DA-ML) are validated over the up- and midstream WSN in Figure 4. The SM estimates from the WRF model were markedly lower than WSN observations for cropland because the impacts of irrigation events on SM estimates are not fully considered in the Noah-MP model (He et al., 2022; Zhang et al., 2020). The Noah-MP model also slightly underestimates SM in the upstream regions because it ignores the effects of dense root systems and soil organic matter on SM estimation (Chen et al., 2012; Sun et al., 2021). As anticipated, SM predictions from the WRF (DA-ML) are closer to the measurements than those of WRF. WRF (DA-ML) SM retrievals indicate a reasonable response to the precipitation and irrigation events in the midstream cropland. Similarly, the WRF (DA-ML) SM dynamics show a characteristic response to precipitation in the upstream regions. The results also indicate that the SM simulations from the WRF (DA-ML) model are hard to capture the observed peak values. This is because the prediction accuracy of ML methods is limited by the training data set. This also means that if the model is applied under extremely wet conditions with sparse training data, the performance of the hybrid model will decrease as the number of training samples decreases. In general, the WRF (DA-ML) can use the information contained in remotely sensed LAI and multi-source SM observations to improve land surface conditions.

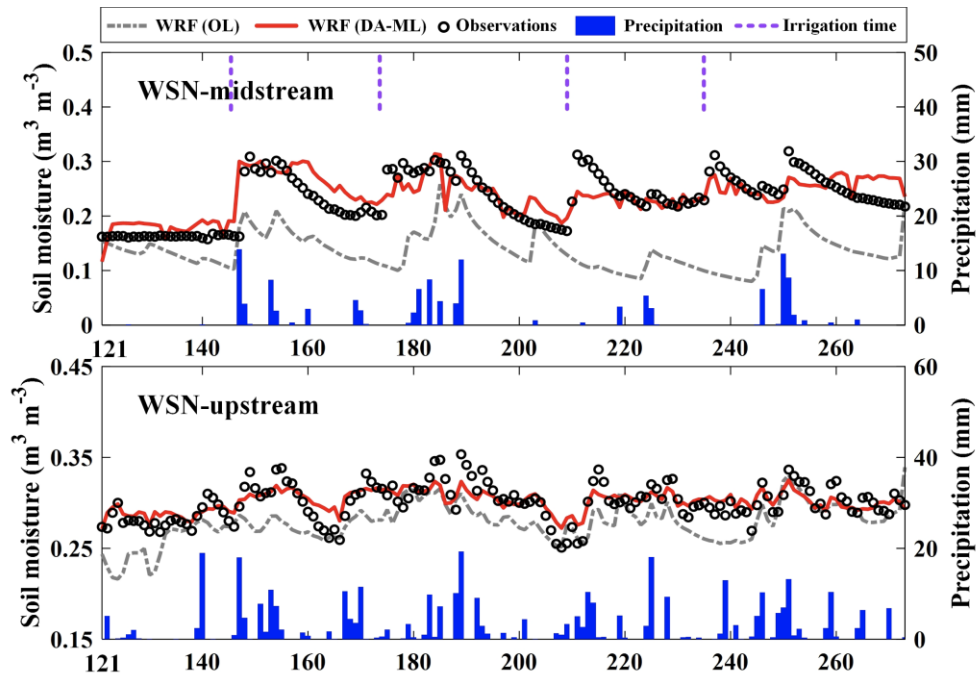



Figure 4: The time series of SM estimates from the WRF (OL) and WRF (DA-ML) models against the up- and midstream WSN observations in 2015.

Figure 5 shows the spatial patterns of the averaged LAI and SM estimates from May to September 2015. The WRF simulation significantly underestimated the LAI, particularly in the up- and midstream vegetation areas of the HRB. In addition, it underestimated the SM in the mid- and downstream vegetation regions. The integration of LAI and SM into the WRF model improved the estimation of leaf biomass and SM, and increased LAI and SM in the HRB. The maps of estimated LAI and SM from the DA-ML method consistently resembled the rainfall, vegetation cover, irrigation event, and shallow groundwater table features (Xu et al., 2018, 2020). The precipitation in the upstream mountains, irrigation in the midstream oasis, and shallow groundwater in the downstream oasis enhance SM and provide the necessary water supply for vegetation growth (Li et al., 2022). Figure 5 also shows the LAI and SM differences between the WRF (DA-ML) and WRF (OL) simulations. The maximum LAI (SM) difference from the WRF (DA-ML) and WRF (OL) reaches approximately 2.24 m^2 m^{-2} (0.16 m^3 m^{-3}) and is present in the midstream oasis of HRB. The difference between the WRF (OL) and WRF (DA-ML) was due to the effects of irrigation and crop growth. Similarly, this difference in the downstream oasis is a result of the shallow water table as well as the growth of riparian forests. In the upstream alpine meadows, the LAI simulated by the WRF (DA-ML) was greatly increased compared to that of the WRF (OL). However, the SM enhancement in the WRF (DA-ML) was not significant due to the sufficient precipitation in mountainous areas.

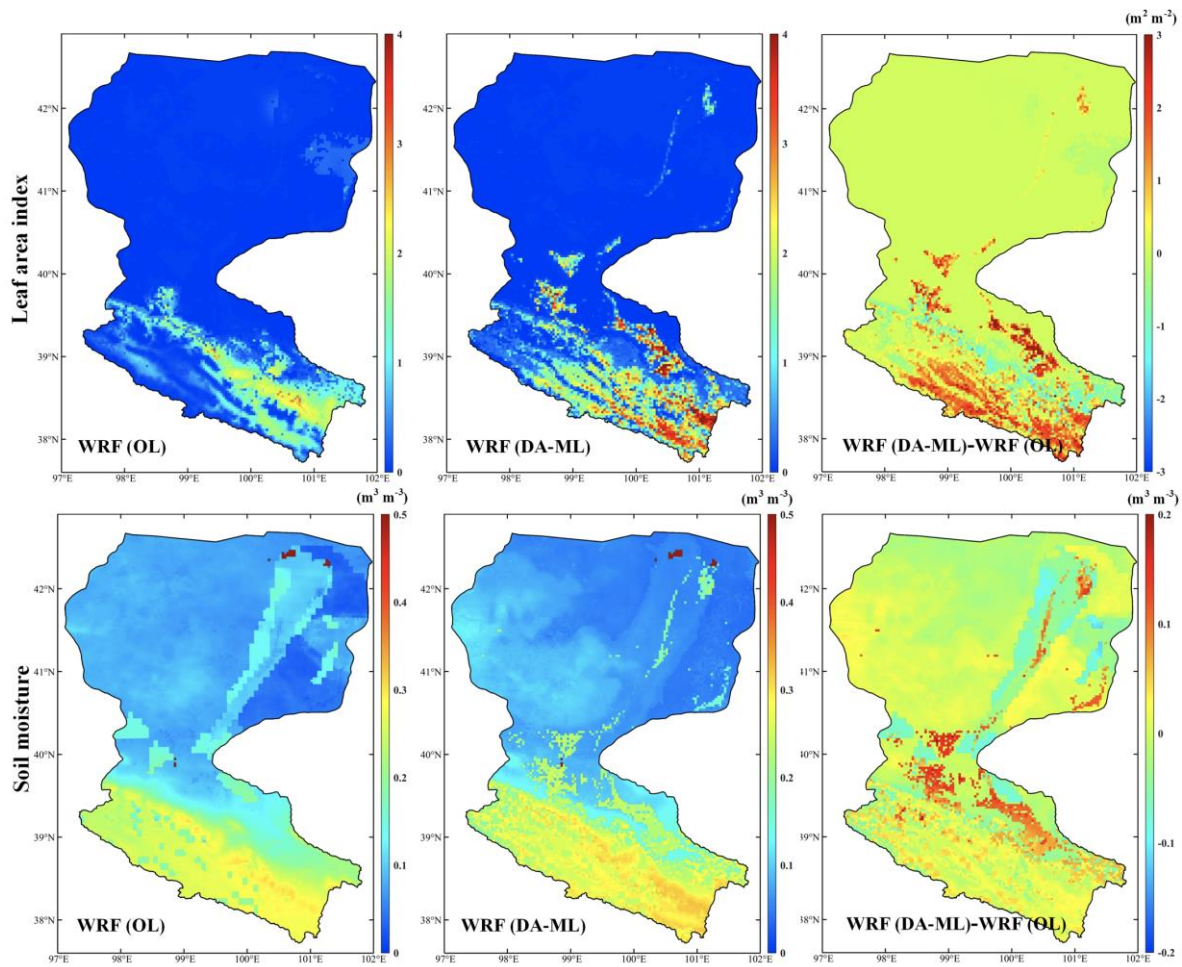



Figure 5: The LAI and SM estimates from the WRF (OL) and WRF (DA-ML) models during the growing season in 2015, and the
285 **average difference in the LAI and SM between the WRF (DA-ML) and WRF (OL) [WRF (DA-ML) minus WRF (OL)].**

4.2 Sensible and latent heat fluxes

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

290 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

295 (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). The LE measurements of the LAS instrument are obtained from the residuals of the surface energy balance equation, which may lead to uncertainties in the LE observations. 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).

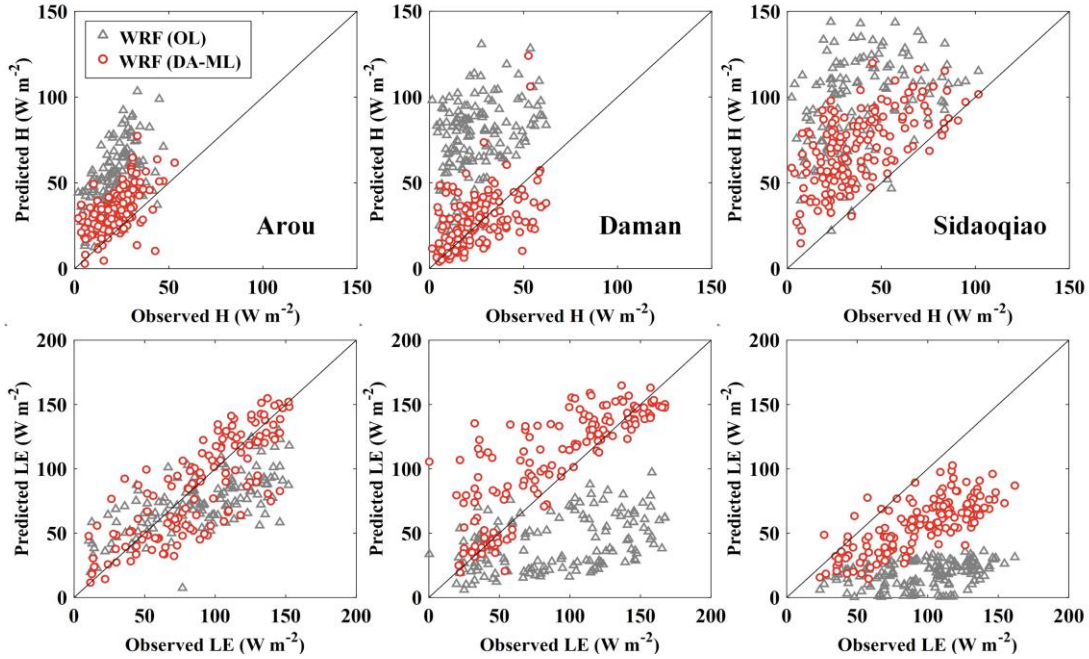



Figure 6: Scatterplot of daily sensible and latent heat flux estimates from the WRF (OL) and WRF (DA-ML) models versus measurements at the Arou, Daman, and Sidaoqiao sites.

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-site-	R^2 (-)	0.28	0.51	0.25	0.63

		average	RMSD (W m^{-2})	53.61	30.16	63.73	48.45

Figure 7 shows the spatial distribution of ET estimates from the WRF (OL), WRF (DA-ML), and ETMap over the HRB. The results indicate that the ET values from the WRF model were underestimated, especially in the midstream oasis region, which was mainly because the WRF model underestimated the SM and LAI (see Figure 3 and 4) during the growing season. Compared with the WRF (OL) model, the WRF (DA-ML) method improves the estimation of ET and the spatial distribution is consistent with that of ETMap because of the effective information contained in the remote sensing LAI and multi-source SM observations. The estimation of ET in the WRF (OL) is sensitive to SM and vegetation dynamics, especially in semi-arid regions. Therefore, the WRF (DA-ML) model will produce more improvements in the mid- and downstream oasis regions compared to the WRF (OL) model. The spatial patterns of ET from the DA-ML method showed a significant gradient from wet to dry owing to variations in precipitation and vegetation cover. In the upstream regions of the HRB, the spatial pattern of retrieved ET was mainly controlled by precipitation and vegetation cover. The ET values were higher in areas with heavier precipitation and denser vegetation. In the midstream region, the spatial pattern of ET was well aligned with the oasis caused by crop growth and irrigation. Meanwhile, the ET values were higher in the downstream oasis because of shallow water tables and transpiration from riparian forests (Xu et al., 2018, 2020). The sparsely vegetated areas covered by desert and Gobi in the mid- and downstream regions had the lowest ET values. The results show that the integration of remotely sensed LAI and multi-source SM observations is essential for studying land-atmosphere water vapor fluxes (ET) because of the realistic land surface conditions.

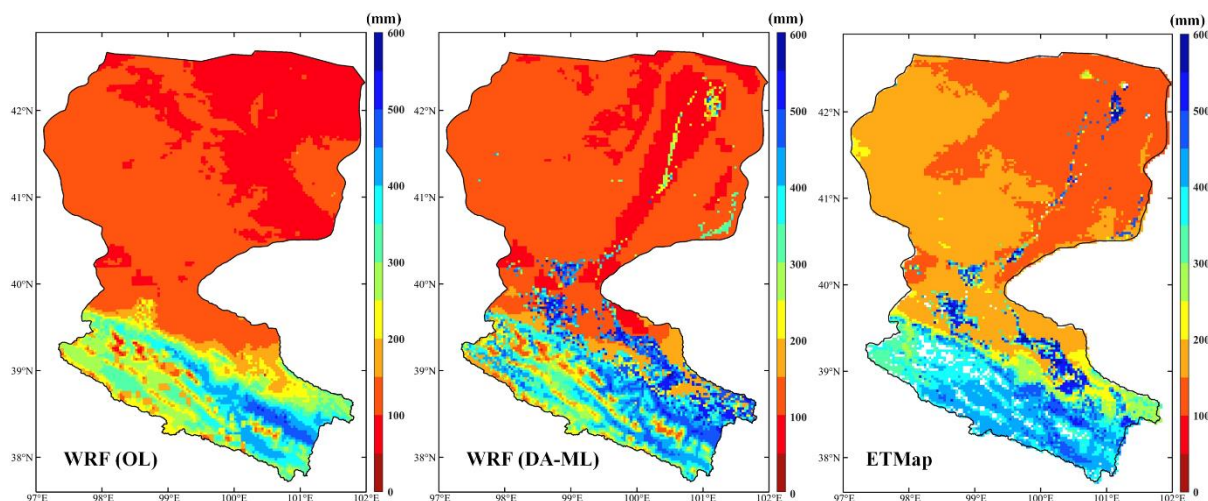



Figure 7: Spatial distribution of evapotranspiration estimates obtained from the WRF (OL), WRF (DA-ML), and ETMap during the growing season in 2015.

4.3 Air temperature and specific humidity

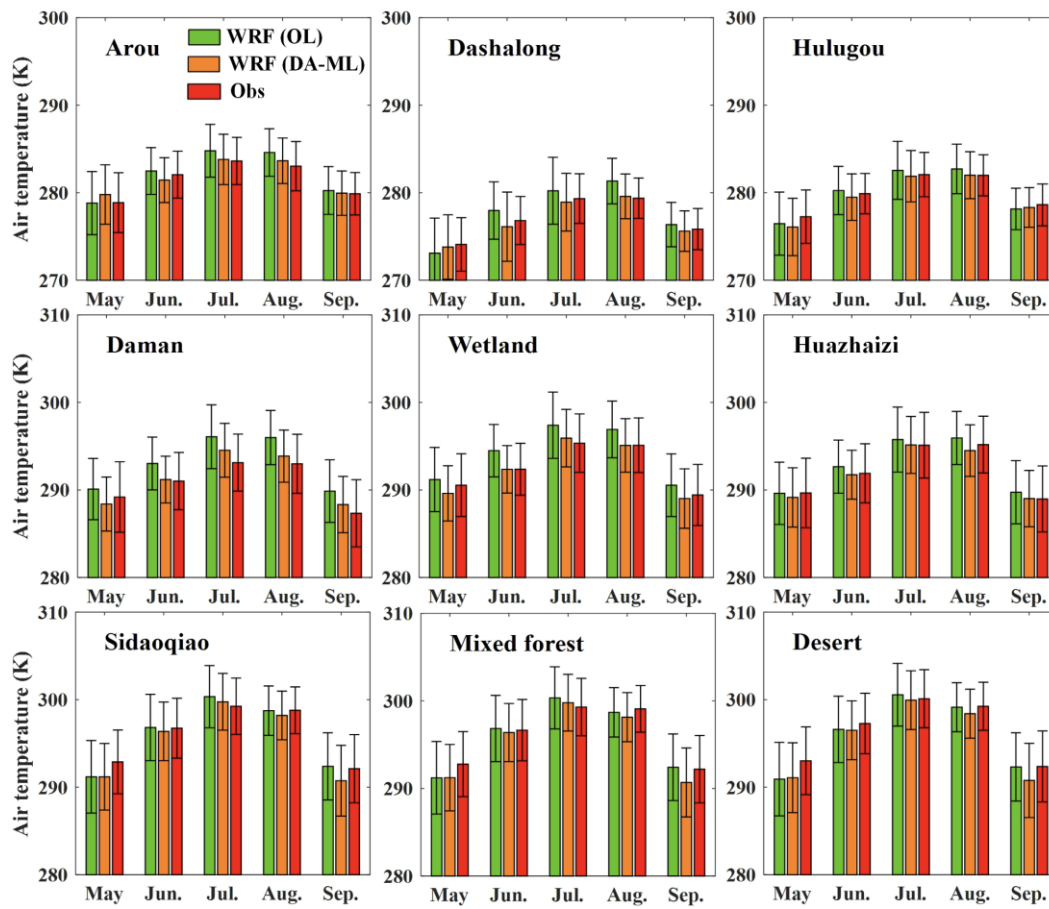
The monthly averaged 2 m air temperature and specific humidity from the WRF (OL), WRF (DA-ML), and corresponding observations at nine sites are shown in Figure 8 and 9. As indicated, the WRF model overestimated (underestimated) the air temperature (specific humidity) in the HRB, especially in the midstream oasis (Daman and Wetland stations), which was mainly because the WRF model underestimated the SM and LAI (see Figure 3 and 4) in the HRB. Compared to the WRF model, the WRF (DA-ML) simulated seasonal cycles of air temperature and specific humidity at the nine sites were closer to the measurements, which was because the integration of remotely sensed LAI and multi-source SM observations improves the estimation of vegetation dynamics and SM, decreases the air temperature, and increases the specific humidity. The increased specific humidity was due to the enhanced evaporation from the soil and stronger transpiration from the expanded vegetation cover. Simultaneously, evaporation absorbs a large amount of energy, thereby reducing the air temperature (Wen et al., 2012). The discrepancy between the WRF (OL) and WRF (DA-ML) was amplified in the middle of the growing season (June, July, and August) due to dense growing vegetation and higher SM caused by several irrigation events. After integrating LAI and SM, the simulated air temperature and specific humidity values from the Daman station decreased and increased by approximately 1.75 K and 1.86 g kg^{-1}, respectively. But for Sidaoqiao, the air temperature and specific humidity decrease and increase by about 0.59 K and 0.41 g kg^{-1}, respectively. The results show that the midstream artificial oasis exhibits a stronger cold and wet island effect than the downstream natural oasis. The estimated air temperature and specific humidity increased from May to July and decreased from August to September. The specific humidity estimated at Daman exhibited significant seasonal variations due to irrigation events and crop phenology.

345

Figure 8: Monthly averaged air temperature simulations from the WRF (OL) and WRF (DA-ML) versus the observations at nine sites in 2015 (error range denotes the standard deviation).

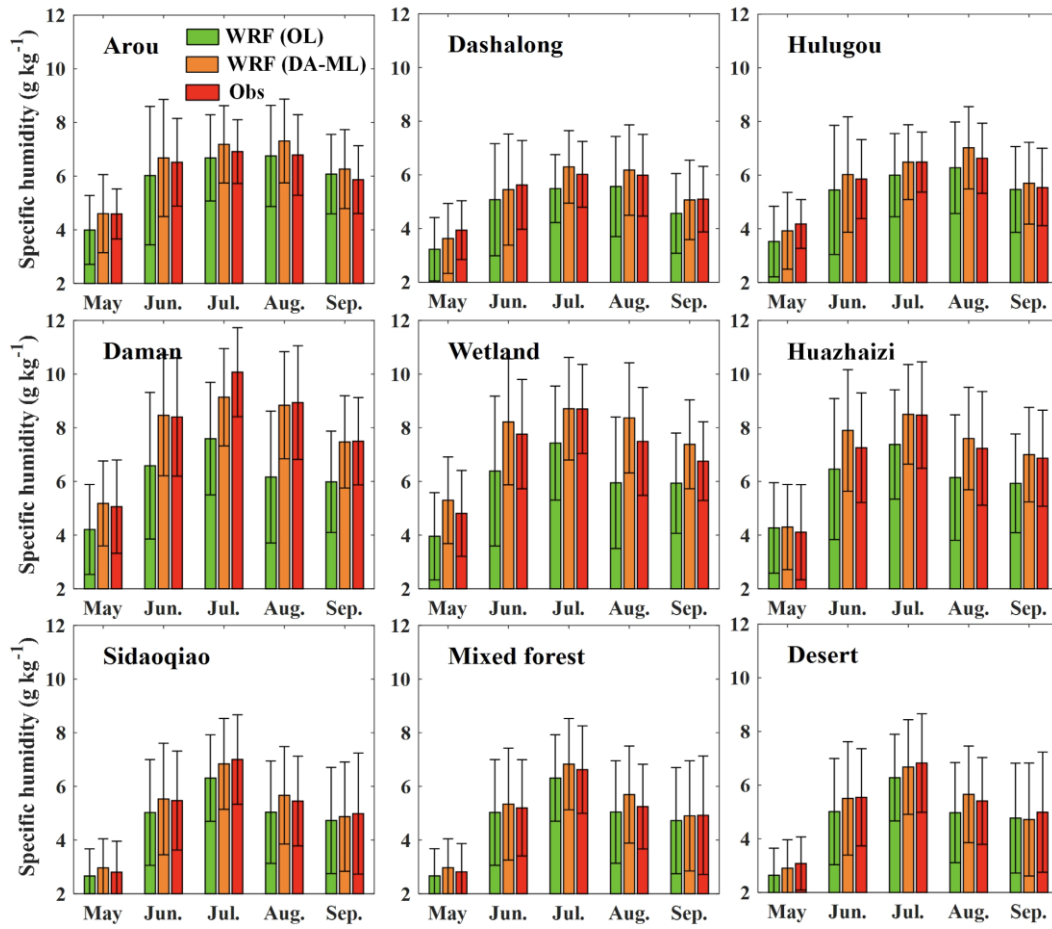



Figure 9: Monthly averaged specific humidity simulations from the WRF (OL) and WRF (DA-ML) versus the observations at nine sites in 2015 (error range denotes the standard deviation).

Table 3 and 4 further compare the simulated air temperature and specific humidity with the same variables from the station observations. The WRF (OL) results show a dry bias in the HRB region, which is reduced by the simulation of the WRF (DA-ML). The statistical metrics (i.e., R^2 and RMSD) of the daily air temperature and specific humidity estimates from the WRF (OL) and WRF (DA-ML) methods are shown in Table 3 and 4. For the nine sites, the average RMSD of the air temperature (specific humidity) estimates from the WRF (DA-ML) was 1.41 K (0.82 g kg^{-1}), which was 21.23% (24.07%) lower than the RMSD of 1.79 K (1.08 g kg^{-1}) from the WRF model.

Table 3: Averaged air temperature and R^2 and RMSD of the WRF (OL) and WRF (DA-ML) compared with the measurements at the nine sites.

Study site	Obs	WRF (OL)			WRF (DA-ML)		
		Sim	R^2 (-)	RMSD (K)	Sim	R^2 (-)	RMSD (K)
Arou	281.48	282.15	0.87	1.48	281.44	0.89	1.11
Dashalong	276.32	277.76	0.84	1.97	276.56	0.86	1.75

Hulugou	279.01	280.06	0.86	1.33	279.05	0.90	1.21
Daman	290.75	292.91	0.84	2.36	291.16	0.88	1.52
Wetland	292.49	293.99	0.90	1.98	292.29	0.92	1.31
Huazhaizi	292.09	292.64	0.91	1.33	291.94	0.94	1.12
Sidaoqiao	295.84	296.12	0.92	1.80	295.53	0.94	1.47
Mixed forest	295.68	296.31	0.92	1.83	295.32	0.93	1.56
Desert	296.30	296.79	0.93	2.04	295.77	0.95	1.61
Average	288.88	289.85	0.89	1.79	288.78	0.91	1.41

Table 4: Averaged specific humidity and R^2 and RMSD of the WRF (OL) and WRF (DA-ML) compared with the measurements at the nine sites.

Study site		WRF (OL)			WRF (DA-ML)		
	Obs	Sim	R^2 (-)	RMSD (g kg^{-1})	Sim	R^2 (-)	RMSD (g kg^{-1})
Arou	6.13	5.90	0.85	0.88	6.31	0.88	0.78
Dashalong	5.33	4.78	0.87	0.83	5.42	0.88	0.71
Hulugou	5.74	5.34	0.82	0.82	5.81	0.85	0.78
Daman	7.97	6.12	0.74	2.26	7.98	0.82	1.04
Wetland	7.09	5.94	0.81	1.55	7.57	0.84	1.03
Huazhaizi	6.78	6.05	0.79	1.25	7.13	0.84	1.10
Sidaoqiao	5.14	4.76	0.91	0.73	5.17	0.91	0.64
Mixed forest	4.96	4.75	0.90	0.68	5.09	0.91	0.63
Desert	5.17	4.73	0.91	0.75	5.09	0.91	0.65
Average	6.03	5.37	0.84	1.08	6.17	0.87	0.82

Figure 10 compares the spatial patterns of the air temperature and specific humidity maps from the WRF (OL) and WRF (DA-ML). Compared with the WRF (OL), significant differences were observed in the WRF (DA-ML). The integration of LAI and SM decreases air temperature and increases specific humidity in the vegetated area of the HRB, particularly in the midstream oasis region. The spatial distribution of specific humidity from the WRF (DA-ML) is consistent with the LAI and SM maps in Figure 5 and the ET map in Figure 7. The results show that the improved LAI and SM simulations lead to different land surface dynamic and thermal characteristics between the oasis and desert. This difference leads to oasis-desert interactions and produces microclimatic effects, including the cooling and wetting effects of the oasis. The average simulated air temperature from WRF (OL) and WRF (DA-ML) methods in the midstream oasis were 293.64 K and 291.32 K, respectively. In contrast, the near-surface air temperatures over the desert are approximately 294.13 K and 293.54 K, respectively. The difference in air temperature between the oasis and desert areas indicates that the oasis areas represent a cold and wet island compared to the surrounding desert. This difference is amplified after the implementation of the DA-ML method. The significant wetting and cooling effects propagate in desert areas to a maximum distance of approximately 5-10 km from the edge of the oasis. In the midstream oasis, the dominant vegetation is irrigated cropland, and

375 the vegetation cover was only approximately 42% in the original WRF model; however, the vegetation cover was updated to approximately 70% in the WRF (DA-ML). The different land surface dynamic and thermal characteristics between the oasis and desert can produce oasis-desert interactions and enhance local circulation. The oasis-desert interactions create a water vapor flux from the oasis to the surrounding desert. This transport process is beneficial for increasing desert water vapor and maintaining the sustainability of desert vegetation (Li et al., 2016; Liu et al., 2020). A similar pattern was observed in the

380 downstream oasis. However, because of the decreased SM and vegetation cover (Figure 5), the downstream oasis exhibited a weaker wet island effect. The results also indicated that enhanced vegetation transpiration increases specific humidity and reduces air temperature owing to increased LAI in the upstream region of the HRB.

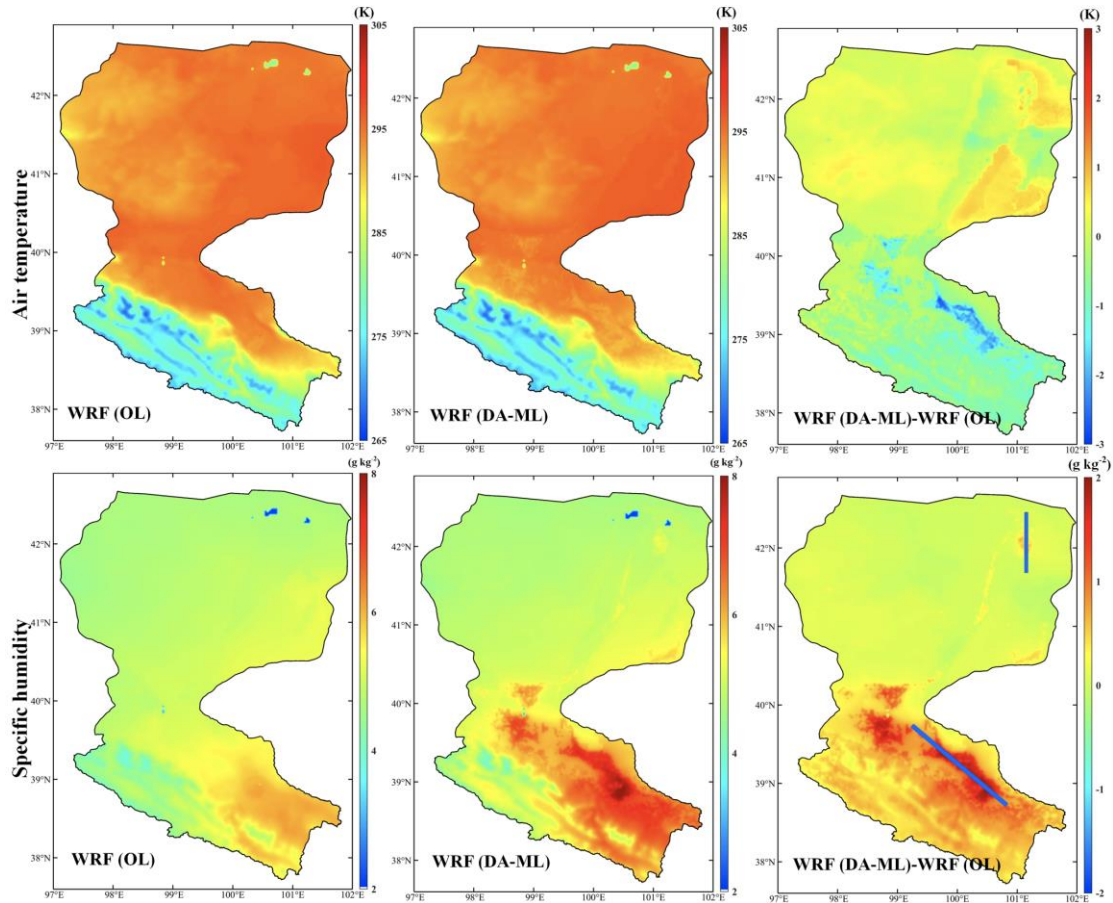



Figure 10: Spatial distribution of air temperature and specific humidity estimates from the WRF (OL) and WRF (DA-ML) during
385 **the growing season in 2015, and the average difference in air temperature and specific humidity between the WRF (DA-ML) and WRF (OL) [WRF (DA-ML) minus WRF (OL)]. The blue line indicates the mid- and downstream oasis vertical profile used in Figure 11 and 12.**

The above-mentioned findings show that the proposed WRF (DA-ML) method exhibits strong wetting and cooling effects in the mid- and downstream oasis. These wetting and cooling effect reduces the air warming bias and dry bias in the

simulation. Therefore, the WRF (DA-ML) simulation is much closer to the observations than the WRF (OL) simulation. Two vertical profiles were selected in Figure 10 to further analyze the effect of the DA-ML on the local climate in the mid- and downstream areas. The difference between the WRF (DA-ML) and WRF (OL) methods is used to represent the enhanced cooling and wetting effects after improving the LAI and SM simulations. As illustrated in Figure 11, the enhanced wetting and cooling effects of the midstream oasis were the strongest in the southern irrigated cropland and gradually decreased in the northern desert areas. The magnitudes of the surface wetting and cooling effects were consistent with the differences in LAI and SM estimates from the WRF (DA-ML) and WRF (OL). For example, the difference in LAI and SM peaks at 38.08°N and the wetting and cooling effects of midstream irrigated cropland were also stronger in this region. These results suggest that the wetting and cooling effects caused by irrigation and vegetation growth occur mainly in the oasis region and do not affect more distant non-oasis areas. 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. Similar results have been demonstrated in several previous studies (Liu et al., 2020; Wen et al., 2012; Zhang et al., 2017a, b).

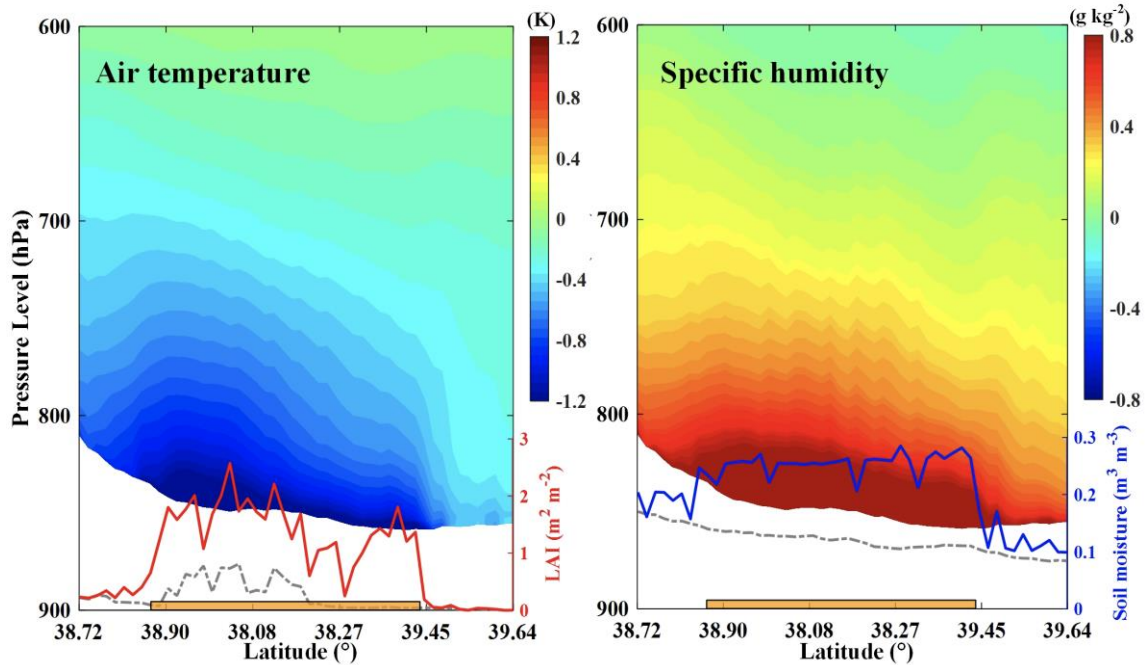



Figure 11: Mean vertical profile of differences in air temperature and specific humidity between the WRF (DA-ML) and WRF (OL) [WRF (DA-ML) minus WRF (OL)] and mean LAI and SM during the growing season in 2015 in the midstream oasis. The dashed and solid lines represent the WRF (OL) and WRF (DA-ML), respectively. The white shaded area represents the change in elevation. The orange bar represents the oasis area.

Figure 12 shows the same wetting and cooling effects in the downstream oasis. Compared to the midstream irrigated cropland, the downstream oasis wetting and cooling effects were mainly influenced by the growth of riparian forests and

410 shallow groundwater tables. The wetting and cooling effects showed maximum values at 42.01°N owing to the strong LAI and SM shifts. The results also indicate that the wetting and cooling effects of the downstream oasis were weaker than those of the midstream oasis. By integrating LAI and SM, the air temperature in mid- and downstream oasis decreases by 0.96 and 0.12 K and the specific humidity increases by 0.52 and 0.06 g kg^{-1}, respectively. In general, the integration of the LAI and SM data can produce more realistic land surface conditions in the oasis region and lead to stronger wetting and cooling 415 effects.

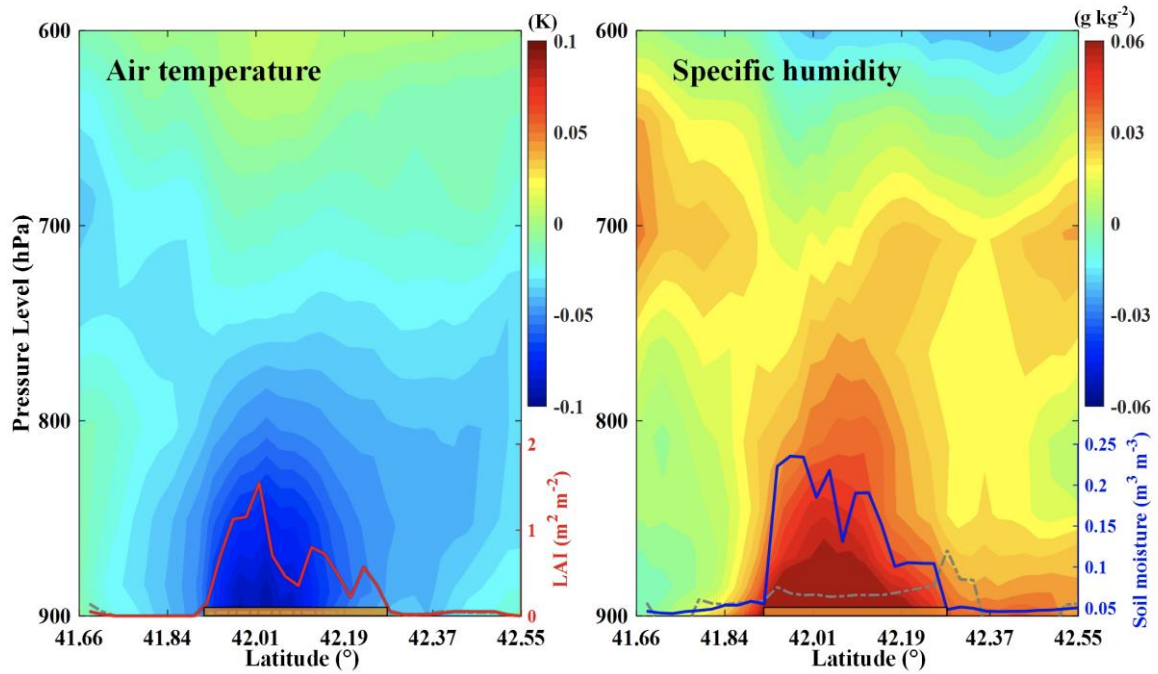



Figure 12: Mean vertical profile of differences in air temperature and specific humidity between the WRF (DA-ML) and WRF (OL) [WRF (DA-ML) minus WRF (OL)] and mean LAI and SM during the growing season in 2015 in the downstream oasis. The dashed and solid lines represent the WRF (OL) and WRF (DA-ML), respectively. The orange bar represents the oasis area.

420 **4.4 Wind speed and precipitation**

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 (Liu et al., 2020). In Figure 13, the heat transfer coefficient (C_h) from the WRF (DA-ML) was used to compare the surface roughness in the oasis with that of the surrounding desert. C_h is an important parameter for calculating the heat transfer between the land and atmosphere, and it is mainly related to the length of the surface roughness and the intensity of the stability of the atmospheric surface layer (Smedman et al., 2007). The results show that the C_h estimates are higher and reduce the wind speed in the midstream oasis compared to the surrounding

430 desert. Ozdogan et al. (2006) and Liu et al. (2020) also showed that crop growth enhanced the surface roughness and slow wind speed. Figure 13 also shows that the wind speed values from the WRF (DA-ML) scheme are slightly lower than those from the WRF (OL) scheme. The average wind speed in the midstream oasis was reduced from 1.92 m/s to 1.23 m/s by integrating the LAI and SM. As mentioned earlier, the different dynamic and thermal characteristics between the oasis and desert can produce oasis-desert interactions and generate local circulation, which drives the cold and moist airflow from the

435 oasis to the surrounding desert in the lower atmosphere. As shown in Figure 13, area to the south of the midstream irrigated cropland is the Qilian Mountains whereas the area to the north is a large desert. Therefore, the southerly airflow generated by the midstream oasis weakened the background north wind. The mean wind vectors in the downstream oasis are shown in Figure 13. The wind speed in the downstream desert regions is slightly higher than in the oasis regions. The lower wind speed in the mid- and downstream oasis is helpful to plant growth, people's survival environment, and the maintenance of

440 the oasis and desert ecosystem (Wang and Cheng, 1999). The results also indicated that the wind vector in the downstream oasis was mainly controlled by the background northerly wind and the effects of LAI and SM integration on the wind vectors were weaker.

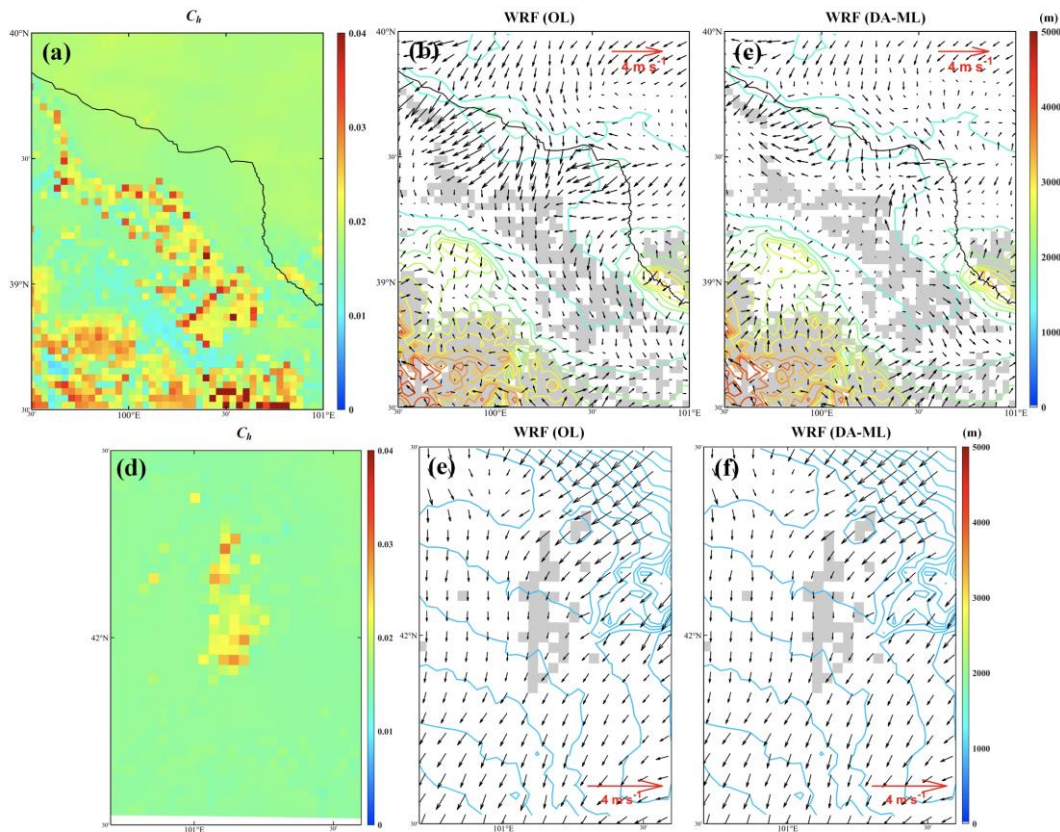



Figure 13: (a and d) Mean heat transfer coefficient (C_h) from the WRF (DA-ML) and (b, c, e, and f) wind vectors at 10 m during
445 **the growing season from the WRF (OL) and WRF (DA-ML) in the midstream (top) and downstream (bottom) oasis. Colored contours indicate elevations above ground level and shading indicates the extent of the oasis. The black line is the boundary of HRB.**

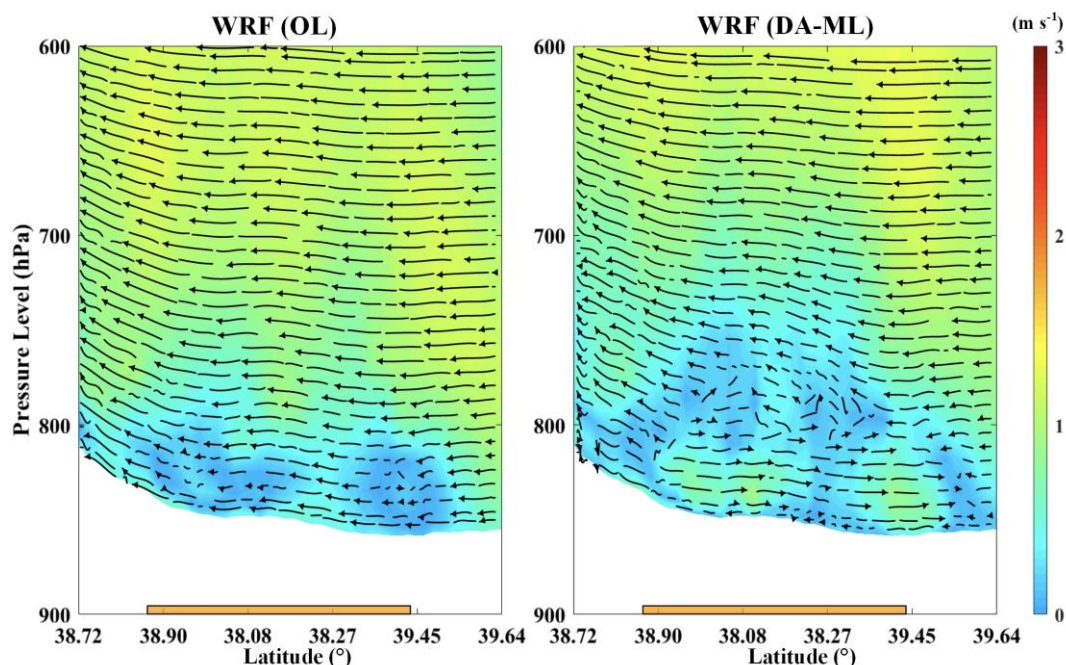
The integration of the LAI and SM affected the wind speed at the land surface and the local circulation through oasis-desert interactions. Figure 14 shows the zonal mean vertical velocity and local meridional circulation in the midstream oases
450   from the WRF (OL) and WRF (DA-ML). Compared with the flat topography of the downstream oasis, the topography of the midstream oasis generally varies from plains to mountains (from low to high altitude) from north to south. The surface dynamic and thermal characteristics of the oasis and surrounding desert differed significantly; therefore, strong horizontal temperature and humidity field gradients were observed at the intersection of the boundary layer of the mountains, oasis, and surrounding desert (Meng et al., 2015; Wen et al., 2012). The air humidity and vegetation cover in the midstream oasis were
455   enhanced by integrating the LAI and SM, which resulted in stronger evaporation from irrigated cropland than from the surrounding desert. As shown in Figure 14, the divergence of the lower atmosphere over the midstream oasis is enhanced, and the wet and cold air masses are transferred to the surrounding desert through advection, whereas the dry and hot air is transferred into the oasis from the upper atmosphere. In the upper atmosphere, the desert to oasis air masses enhance the background northerly winds, which promote atmospheric water vapor transport in the HRB. However, oasis-desert
460   interactions are weaker in the downstream region (Figure A1) than in the midstream region under actual weather or climate conditions, which is attributed to the local circulation being weakened by stronger background northerly winds. Overall, the simulation of soil and vegetation characteristics can be improved by integrating LAI and SM and enhancing land-atmosphere interactions in mid- and downstream oases.

465   **Figure 14: The zonal mean vertical velocity and meridional circulation from the WRF (OL) and WRF (DA-ML) models during the growing season in 2015 in the midstream oasis. The white shaded area represents the change in elevation. The orange bar represents the oasis area.**

Figure 15 exhibits the influence of the DA-ML on precipitation in the HRB. The results show that the integrated LAI and SM led to increased precipitation in the upstream regions of the HRB and that the spatial variation in precipitation was
470   very heterogeneous. The increase in precipitation was mainly concentrated in the southeastern part of the HRB, where it reached approximately 1.5 mm day^{-1}, which represented 32% of the simulated value of the WRF (OL) experiment. In contrast, precipitation increased insignificantly in the mid- and downstream oasis regions. The increased precipitation in the upstream region may have been due to the additional water vapor supply. Water vapor fluxes in the mountain areas and midstream oasis regions were enhanced by integrating the LAI and SM. Driven by background northerly winds (Figure 14),
475   more water vapor fluxes from the midstream oasis region were carried to the upstream region. The wind speed and precipitation estimates in the upstream region (around Babao River Basin) are shown in Figure 15b and 15c. As shown, the WRF (DA-ML) enhanced the estimation of precipitation on windward slopes compared with valleys. After integrating the LAI and SM, land-atmosphere interactions are altered. The DA-ML increased the latent heat and decreases the sensible heat flux. The water vapor carried by the air masses and lifted on the sloped surface was more likely to condense and produce
480   precipitation (Yue et al., 2021). In general, the DA-ML enhanced the precipitation estimates in the upstream mountain areas, mainly on windward slopes.

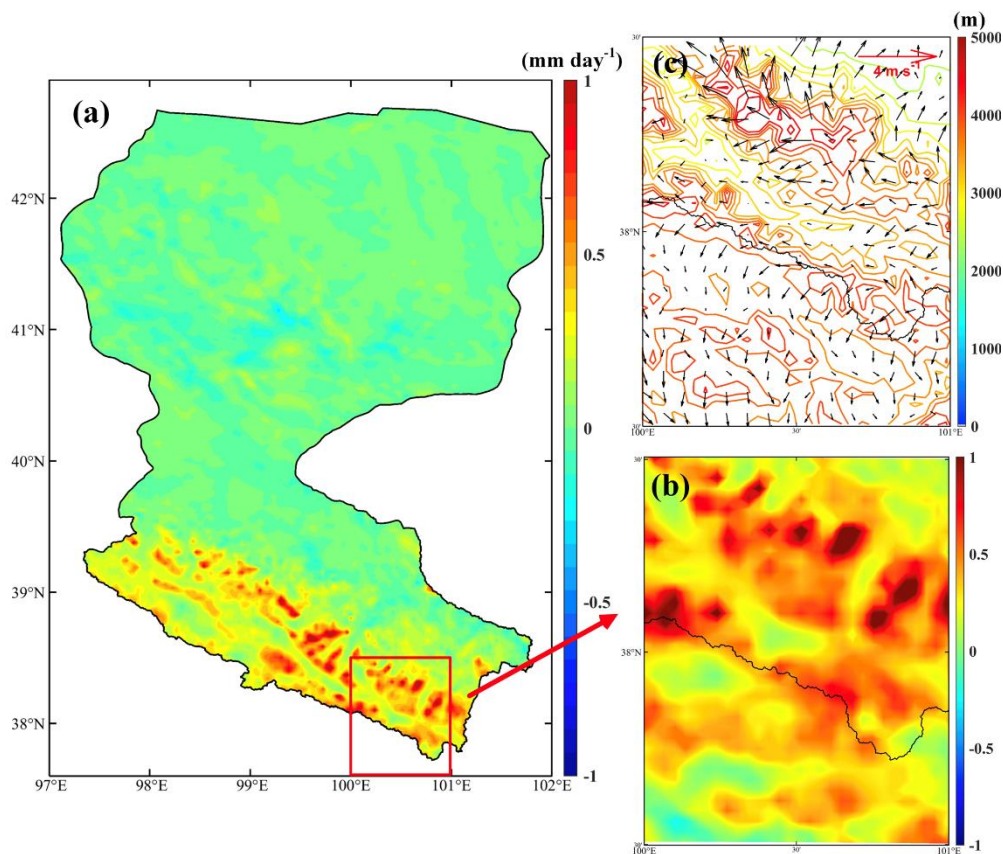



Figure 15: (a) Average difference in precipitation between the WRF (DA-ML) and WRF (OL) [WRF (DA-ML) minus WRF (OL)] in the Heihe River Basin and (b) the upstream region (around Babao River Basin) and (c) wind vectors at 10 m from the WRF
485 **(DA-ML) in the upstream region. The black line is the boundary of HRB.**

The simulated daily precipitation from the WRF (OL) and WRF (DA-ML) was compared with that of the AFD and CMFD references in Figure 16. Because the water vapor from the East Asian monsoon was blocked by the Tibetan Plateau, most of the precipitation was concentrated in the southeastern part of the Qilian Mountains. Figure 16 shows that the high precipitation zone of the HRB was mainly located in the mountainous areas below 39.5 °N due to orographic lifting and

490 convection (Zhang et al., 2021b). The main precipitation events were consistent between the WRF (OL) and WRF (DA-ML). WRF (DA-ML) had higher precipitation in the southern domain because peak precipitation was enhanced and lesser precipitation events increased (red rectangles). Both the WRF (OL) and WRF (DA-ML) captured the temporal and spatial variability of precipitation well and were consistent with the reference data, indicating that the 3-km high-resolution grid contains information on topography-related heterogeneity and accurately estimates the precipitation distribution. The

495 estimated precipitation in the upstream regions of the HRB was more consistent with the AFD reference but was overestimated by approximately 0.43 mm day^{-1} compared to the CMFD. The discrepancy between the precipitation estimated by the WRF and CMFD schemes occurred because the China Meteorological Administration sites fused by the CMFD product were mainly distributed at elevations below 3500 m (He et al., 2020a). Therefore, there are some uncertainties in the precipitation simulation of the CMFD products in high-altitude mountainous areas (Zhang et al., 2021b).

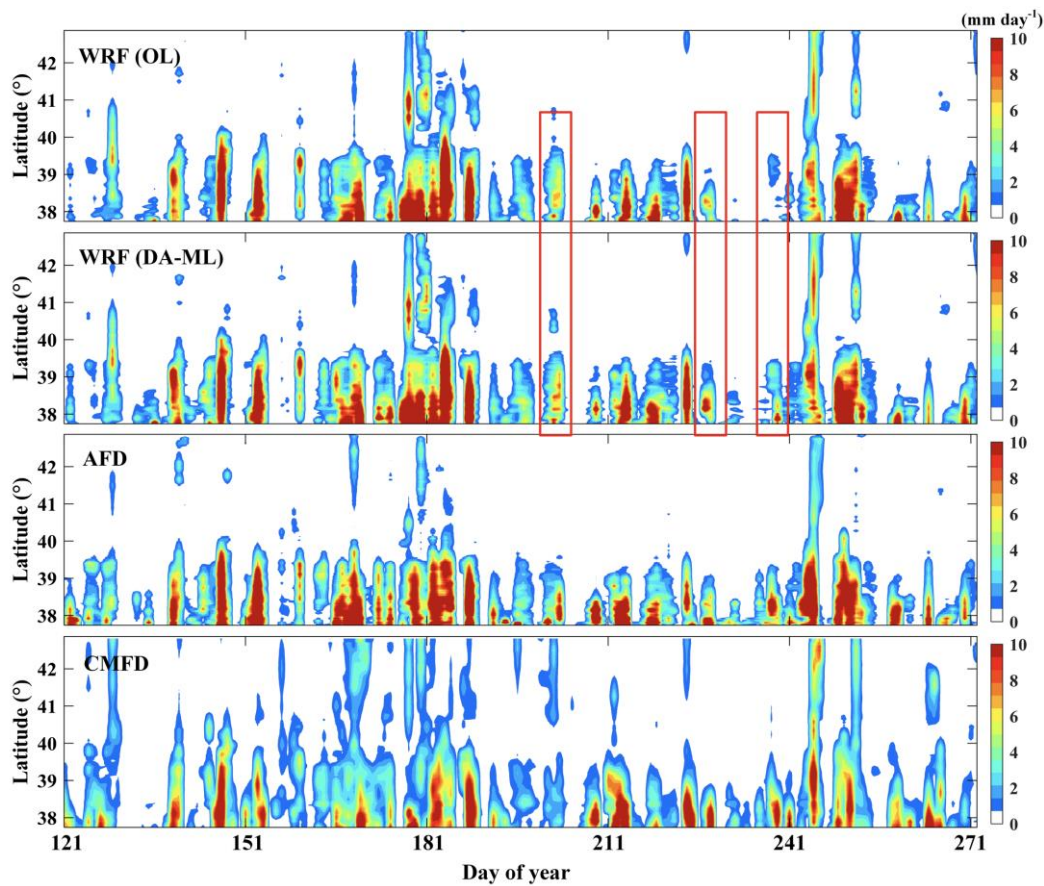



Figure 16: Comparisons of the daily precipitation estimates from the WRF (OL) and WRF (DA-ML) with the values from the AFD and CMFD references during the growing season in 2015.

5 Conclusions

In this paper, a hybrid data assimilation and machine learning framework (DA-ML approach) was proposed and implemented into the Weather Research and Forecasting (WRF) model to optimize the initialization of surface soil and vegetation variables. Remotely sensed leaf area index (LAI) and multi-source soil moisture (SM) observations (*in-situ* SM profile observations and remotely sensed SM products) were integrated into the WRF model to improve the soil and vegetation characteristics. The performance of the WRF (DA-ML) framework was tested in the Heihe River Basin (HRB) in northwestern China. The results indicated that the integration of remotely sensed LAI and multi-source SM into the WRF model improved the LAI, SM, and evapotranspiration (ET) estimates and regional climate and land-atmosphere interactions. The estimated sensible and latent heat fluxes from the WRF (OL) and WRF (DA-ML) models are validated with the large aperture scintillometer (LAS) observations at the Arou, Daman, and Sidaoqiao sites. For the WRF (DA-ML) approach, the three-site-averaged RMSDs of daily sensible and latent heat fluxes are 30.16 W m^{-2} and 48.45 W m^{-2}, respectively, which

are 43.74% and 23.98% lower than those of WRF (OL). The results indicated that the WRF (DA-ML) method improved the
515 estimation of ET and produced a spatial distribution that was consistent with the results of ETMap.

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. For the WRF model, the nine-site-averaged root mean square
deviation (RMSD) of the air temperature and specific humidity estimates was 1.79 K and 1.08 g kg^{-1}. The WRF (DA-ML)
reduces the aforementioned RMSDs by 21.23% and 24.07%. Strong wetting and cooling effects on vegetated areas were
520 observed through the integration of LAI and SM, especially in the midstream oasis. The magnitude of the surface wetting
and cooling effects corresponded well with the differences in the LAI and SM estimates from the WRF (DA-ML) and WRF
(OL). These results indicate that the wetting and cooling effects gradually decrease from the land surface upwards and are
replaced by slight warming and drying effects.

The crops, shelterbelts, and residential areas in the midstream oasis produce a wind shield effect because of the stronger
525 surface roughness. The different land surface dynamic and thermal characteristics between the oasis and desert can produce
oasis-desert interactions and generate local circulation. In the lower atmosphere, wet and cold air masses are transferred to
the surrounding desert by advection, while the dry and hot air over the desert is transferred to the oasis from the upper
atmosphere. The results show that the integration of LAI and SM will induce water vapor intensification and promote
precipitation in the upstream regions of the HRB. The WRF (DA-ML) simulation captured the temporal and spatial
530 variability of precipitation well and was consistent with the reference data. The results indicate that the 3-km high-resolution
grid can consider topographic information and produce accurate precipitation distribution estimates.

Data availability

ERA5 data for the WRF model is freely available via the European Centre for Medium-Range Weather Forecasts
(https://cds.climate.copernicus.eu/cdsapp#!/search?type=dataset). The meteorological and flux observations, ETMap, China
535 Meteorological Forcing Dataset (CMFD), and atmospheric forcing data (AFD) dataset in the HRB can be accessed from the
National Tibetan Plateau Data Center (https://data.tpdc.ac.cn/en/). The assimilated LAI data are available on the Global Land
Surface Satellite (GLASS) product (http://www.glass.umd.edu/). Soil moisture products can be downloaded from the Soil
Moisture Active Passive (SMAP) product (https://appeears.earthdatacloud.nasa.gov/). The original WRF code can be
obtained from the National Center for Atmospheric Research (NCAR) archive (https://github.com/wrf-model).

540 **Author contribution**

X.H. developed the model code and completed the original manuscript with support from all coauthors. Y.L., S.L., T.X.,
F.C., Z.L., R.L., and L.S. revised the manuscript. Z.Z. provided the original WRF code. Z.X., Z.P., and C.Z. provided the
methods for processing the observational data. All authors contributed to the synthesis of the results and key conclusions.

Competing interests

545   The authors declare that they have no conflict of interest.

Acknowledgments

This research was partially supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA20100101) and the National Natural Science Foundation of China (42171315). The computation of the WRF model was supported by sources from the Center for Geodata and Analysis, Faculty of Geographical Science, Beijing Normal

550   University (https://gda.bnu.edu.cn/).

Appendix A

Table A1: The computation details about the WRF (OL) and WRF (DA-ML).

Method	Computation time	Code	Processor	Parallelization
WRF (OL)	76 h	Fortran	24-core 2.6 GHz	Intel MPI
WRF (DA-ML)	112 h	Python and Fortran	Intel Xeon Gold	(100 cores)

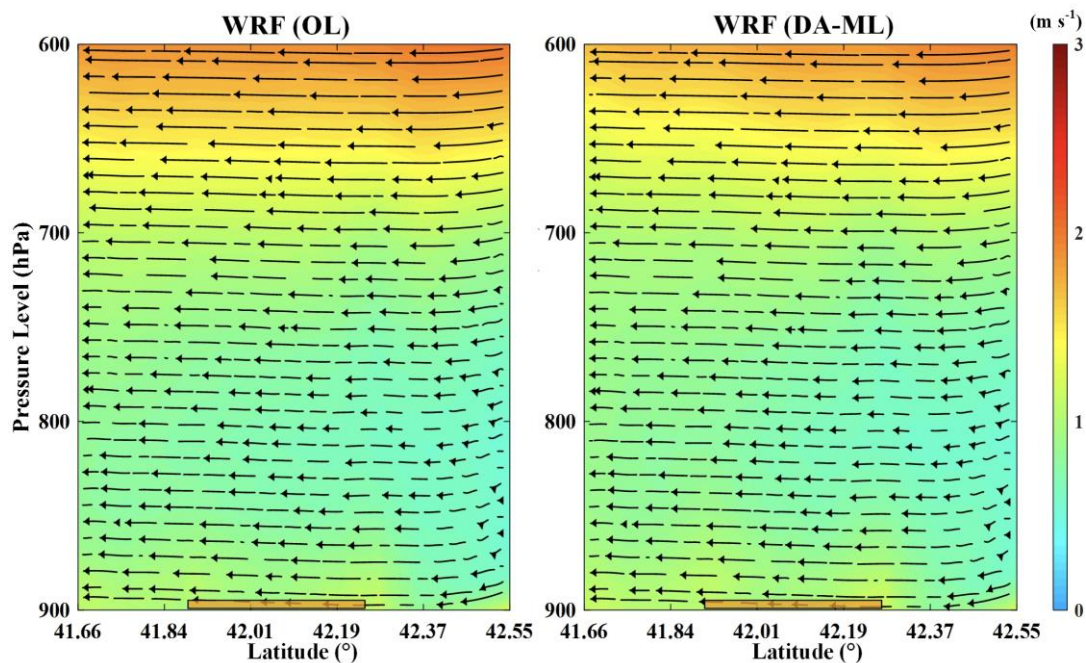



Figure A1: The zonal mean vertical velocity and meridional circulation from the WRF (OL) and WRF (DA-ML) models during
555   **the growing season in 2015 in the downstream oasis. The orange bar represents the oasis area.**

References

Ahmad, S.K., Kumar, S.V., Lahmers, T.M., Wang, S., Liu, P., Wrzesien, M.L., Bindlish, R., Getirana, A., Locke, K.A., Holmes, T.R., and Otkin, J.A.: Flash Drought Onset and Development Mechanisms Captured with Soil Moisture and Vegetation Data Assimilation. Water Resour. Res., https://doi.org/10.1029/2022WR032894, 2022.

560 Brajard, J., Carrassi, A., Bocquet, M., and Bertino, L.: Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the Lorenz 96 model. J. Comput. Sci., 44, 101171. https://doi.org/10.1016/j.jocs.2020.101171, 2020.

Buizza, C., Quilodrán Casas, C., Nadler, P., Mack, J., Marrone, S., Titus, Z., Le Cornec, C., Heylen, E., Dur, T., Baca Ruiz, L., Heaney, C., Díaz Lopez, J.A., Kumar, K.S.S., and Arcucci, R.: Data Learning: Integrating Data Assimilation and

565 Machine Learning. J. Comput. Syst. Sci., 58, 101525. https://doi.org/10.1016/j.jocs.2021.101525, 2022.

Campo, L., Castelli, F., Entekhabi, D., and Caparrini, F.: Land-atmosphere interactions in an high resolution atmospheric simulation coupled with a surface data assimilation scheme. Nat. Hazards Earth Syst. Sci., 9, 1613–1624. https://doi.org/10.5194/nhess-9-1613-2009, 2009.

Cazes Boezio, G. and Ortelli, S.: Use of the WRF-DA 3D-Var Data Assimilation System to Obtain Wind Speed Estimates in

570 Regular Grids from Measurements at Wind Farms in Uruguay. Data, 4, 142. https://doi.org/10.3390/data4040142, 2019.

Chen, F. and Dudhia, J.: Coupling an Advanced Land Surface–Hydrology Model with the Penn State–NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity. Mon. Wea. Rev., 129, 569–585. https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2, 2001.

Chen, R., Song, Y., Kang, E., Han, C., Liu, J., Yang, Y., Qing, W.W., and Liu, Z.W.: A cryosphere-hydrology observation

575 system in a small alpine watershed in the Qilian mountains of China and its meteorological gradient. Arct. Antarct. Alp. Res., 46 (2), 505–523. https://doi.org/10.1657/1938-4246-46.2.505, 2014.

Che, T., Li, X., Liu, S., Li, H., Xu, Z., Tan, J., Zhang, Y., Ren, Z., Xiao, L., Deng, J., Jin, R., Ma, M., Wang, J., and Yang, X.: Integrated hydrometeorological, snow and frozen-ground observations in the alpine region of the Heihe River Basin, China. Earth Syst. Sci. Data, 11, 1483–1499. https://doi.org/10.5194/essd-11-1483-2019, 2019.

580 Chen, Y.Y., Yang, K., Tang, W.J., Qin, J., and Zhao, L.: Parameterizing soil organic carbon's impacts on soil porosity and thermal parameters for Eastern Tibet grasslands. Sci. China Earth Sci., 55, 1001–1011. https://doi.org/10.1007/s11430-012-4433-0, 2012.

Comellas Prat, A., Federico, S., Torcasio, R.C., Fierro, A.O., and Dietrich, S.: Lightning data assimilation in the WRF-ARW model for short-term rainfall forecasts of three severe storm cases in Italy. Atmos. Res., 247, 105246.

585 https://doi.org/10.1016/j.atmosres.2020.105246, 2021.

Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., Attinger, S., and Thober, S.: The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model. J. Geophys. Res. Atmos., 121, 10,676–10,700. https://doi.org/10.1002/2016JD025097, 2016.

Erlandsen, H.B., Haddeland, I., Tallaksen, L.M., and Kristiansen, J.: The Sensitivity of the Terrestrial Surface Energy and Water Balance Estimates in the WRF Model to Lower Surface Boundary Representations: A South Norway Case Study. J. Hydrometeor., 18, 265–284. https://doi.org/10.1175/JHM-D-15-0146.1, 2017.

Forman, B.A., and Xue, Y.: Machine learning predictions of passive microwave brightness temperature over snow-covered land using the special sensor microwave imager (SSM/I). Phys. Geog., 38, 176–196. https://doi.org/10.1080/02723646.2016.1236606, 2017.

Gao, Y., Chen, F., Barlage, M., Liu, W., Cheng, G., Li, X., Yu, Y., Ran, Y., Li, H., Peng, H., and Ma, M.: Enhancement of land surface information and its impact on atmospheric modeling in the Heihe River Basin, northwest China. J. Geophys. Res., 113, D20S90. https://doi.org/10.1029/2008JD010359, 2008.

Gentine, P., Massmann, A., Lintner, B.R., Hamed Alemohammad, S., Fu, R., Green, J.K., Kennedy, D., and Vilà-Guerau de Arellano, J.: Land–atmosphere interactions in the tropics-a review. Hydrol. Earth Syst. Sci., 23, 4171–4197. https://doi.org/10.5194/hess-23-4171-2019, 2019.

Gottwald, G.A. and Reich, S.: Combining machine learning and data assimilation to forecast dynamical systems from noisy partial observations. Chaos, 31, 101103. https://doi.org/10.1063/5.0066080, 2021.

Grzeschik, M., Bauer, H.-S., Wulfmeyer, V., Engelbart, D., Wandinger, U., Mattis, I., Althausen, D., Engelmann, R., Tesche, M., and Riede, A.: Four-dimensional variational analysis of water-vapor Raman lidar data and their impact on mesoscale forecasts. J. Atmos. Ocean. Tech., 25, 1437–1453. https://doi.org/10.1175/2007JTECHA974.1, 2008.

He, J., Yang, K., Tang, W. Lu, H., Qin, J., Chen, Y.Y., and Li, X.: The first high-resolution meteorological forcing dataset for land process studies over China. Sci. Data, 7, 25, https://doi.org/10.1038/s41597-020-0369-y, 2020a.

He, X., Xu, T., Bateni, S.M., Ek, M., Liu, S., and Chen, F.: Mapping Regional Evapotranspiration in Cloudy Skies via Variational Assimilation of All-Weather Land Surface Temperature Observations. J. Hydrol., 585, 124790. https://doi.org/10.1016/j.jhydrol.2020.124790, 2020b.

He, Xinlei, Xu, T., Bateni, S.M., Ki, S.J., Xiao, J., Liu, S., Song, L., and He, Xiangping: Estimation of Turbulent Heat Fluxes and Gross Primary Productivity by Assimilating Land Surface Temperature and Leaf Area Index. Water Resour. Res., 57. https://doi.org/10.1029/2020WR028224, 2021.

He, X., Liu, S., Xu, T., Yu, K., Gentine, P., Zhang, Z., Xu, Z., Jiao, D., and Wu, D.: Improving predictions of evapotranspiration by integrating multi-source observations and land surface model. Agric. Water Manag., 272, 107827. https://doi.org/10.1016/j.agwat.2022.107827, 2022.

Hu, Y.Q., Gao, Y.X., Wang, J.M., Ji, G.L., Shen, Z.B., Cheng, L.S., Cheng, J.Y., and Li, S.Q.: Some achievements in scientific research during HEIFE. (In Chinese, with English abstract.) Plateau Meteorol., 13:225–236, 1994.

Janjic, Z.: The Step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. Mon. Weather Rev., 122(5), 927–945. https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2, 1994.

Jia, X., Willard, J., Karpatne, A., Read, J.S., Zwart, J.A., Steinbach, M., and Kumar, V.: Physics-Guided Machine Learning for Scientific Discovery: An Application in Simulating Lake Temperature Profiles. ACM/IMS Trans. Data Sci., 2, 1–26. https://doi.org/10.1145/3447814, 2021.

625   Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D.S., Haverd, V., Köhler, P., Ichii, K., Jain, A.K., Liu, J., Lombardozzi, D., Nabel, J.E.M.S., Nelson, J.A., O'Sullivan, M., Pallandt, M., Papale, D., Peters, W., Pongratz, J., Rödenbeck, C., Sitch, S., Tramontana, G., Walker, A., Weber, U., and Reichstein, M.: Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach. Biogeosciences, 17, 1343–1365. https://doi.org/10.5194/bg-17-1343-630   2020, 2020.

Koppa, A., Rains, D., Hulsman, P., Poyatos, R., and Miralles, D.G.: A deep learning-based hybrid model of global terrestrial evaporation. Nat. Commun., 13, 1912. https://doi.org/10.1038/s41467-022-29543-7, 2022.

Levis, S., Bonan, G.B., Kluzek, E., Thornton, P.E., Jones, A., Sacks, W.J., and Kucharik, C.J.: Interactive Crop Management in the Community Earth System Model (CESM1): Seasonal Influences on Land–Atmosphere Fluxes. J. Clim., 25(14), 4839–635   4859. https://doi.org/10.1175/jcli-d-11-00446.1, 2012.

Li, Xin, Li, Xiaowen, Li, Z., Ma, M., Wang, Jian, Xiao, Q., Liu, Qiang, Che, T., Chen, E., Yan, G., Hu, Z., Zhang, L., Chu, R., Su, P., Liu, Qinhuo, Liu, S., Wang, Jindi, Niu, Z., Chen, Y., Jin, R., Wang, W., Ran, Y., Xin, X., and Ren, H.: Watershed Allied Telemetry Experimental Research. J. Geophys. Res., 114, D22103. https://doi.org/10.1029/2008JD011590, 2009.

Li, X., Cheng, G., and Liu, S.: Heihe watershed allied telemetry experimental research (HiWATER): scientifc objectives and 640   experimental design. Bull. Am. Meteorol. Soc., 94 (8), 1145–1160. https://doi.org/10.1175/BAMS-D-12-00154.1, 2013.

Li, X., Yang, K., and Zhou, Y.: Progress in the study of oasis-desert interactions. Agric. For. Meteorol., 230, 1–7. https://doi.org/10.1016/j.agrformet.2016.08.022, 2016.

Li, X., Cheng, G., Lin, H., Cai, X., Fang, M., Ge, Y., Hu, X., Chen, M., and Li, W.: Watershed System Model: The Essentials to Model Complex Human-Nature System at the River Basin Scale. J. Geophys. Res. Atmos., 123, 3019–3034. 645   https://doi.org/10.1002/2017JD028154, 2018.

Li, X., Zhang, L., Zheng, Y., Yang, D., Wu, F., Tian, Y., Han, F., Gao, B., Li, H., Zhang, Y., Ge, Y., Cheng, G., Fu, B., Xia, J., Song, C., and Zheng, C.: Novel hybrid coupling of ecohydrology and socioeconomy at river basin scale: A watershed system model for the Heihe River basin. Environ. Model. Softw., 141, 105058. https://doi.org/10.1016/j.envsoft.2021.105058, 2021.

650   Li, Xin, Cheng, G., Fu, B., Xia, J., Zhang, L., Yang, D., Zheng, C., Liu, S., Li, Xiubin, Song, C., Kang, S., Li, Xiaoyan, Che, T., Zheng, Y., Zhou, Y., Wang, H., and Ran, Y.: Linking Critical Zone With Watershed Science: The Example of the Heihe River Basin. Earth's Future, 10. https://doi.org/10.1029/2022EF002966, 2022.

Liang, S., Cheng, J., Jia, K., Jiang, B., Liu, Q., Xiao, Z., Yao, Y., Yuan, W., Zhang, X., Zhao, X., and Zhou, J.: The Global Land Surface Satellite (GLASS) Product Suite. Bull. Am. Meteorol. Soc., 102, E323–E337. https://doi.org/10.1175/BAMS-655   D-18-0341.1, 2021.

Ling, X.L., Fu, C.B., Guo, W.D., and Yang, Z.L.: Assimilation of remotely sensed LAI into CLM4CN using DART. J. Adv. Model. Earth Syst., 11, 2768–2786. https://doi.org/10.1029/2019MS001634, 2019.

Liu, R., Liu, S., Yang, X., Lu, H., Pan, X., Xu, Z., Ma, Y., and Xu, T.: Wind Dynamics Over a Highly Heterogeneous Oasis Area: An Experimental and Numerical Study. J. Geophys. Res. Atmos., 123, 8418–8440. https://doi.org/10.1029/2018JD028397, 2018a.

Liu, R., Sogachev, A., Yang, X., Liu, S., Xu, T., and Zhang, J.: Investigating microclimate effects in an oasis-desert interaction zone. Agric. For. Meteorol., 290, 107992. https://doi.org/10.1016/j.agrformet.2020.107992, 2020.

Liu, S., Xu, Z., Song, L., Zhao, Q., Ge, Y., Xu, T., Ma, Y., Zhu, Z., Jia, Z., and Zhang, F.: Upscaling evapotranspiration measurements from multi-site to the satellite pixel scale over heterogeneous land surfaces. Agric. For. Meteorol., 230–231, 97–113. https://doi.org/10.1016/j.agrformet.2016.04.008, 2016a.

Liu, S., Li, X., Xu, Z., Che, T., Xiao, Q., Ma, M., Liu, Q., Jin, R., Guo, J., Wang, L., Wang, W., Qi, Y., Li, H., Xu, T., Ran, Y., Hu, X., Shi, S., Zhu, Z., Tan, J., Zhang, Y., and Ren, Z.: The Heihe Integrated Observatory Network: A Basin-Scale Land Surface Processes Observatory in China. Vadose Zone J., 17, 180072. https://doi.org/10.2136/vzj2018.04.0072, 2018b.

Liu, L., Ma, Y., Menenti, M., Su, R., Yao, N., and Ma, W.: Improved parameterization of snow albedo in Noah coupled with Weather Research and Forecasting: applicability to snow estimates for the Tibetan Plateau. Hydrol. Earth Syst. Sci., 25, 4967–4981. https://doi.org/10.5194/hess-25-4967-2021, 2021.

Liu, X., Chen, F., Barlage, M., Zhou, G., and Niyogi, D.: Noah-MP-Crop: Introducing dynamic crop growth in the Noah-MP land surface model. J. Geophys. Res. Atmos., 121(23), 13,953–13,972. https://doi.org/10.1002/2016jd025597, 2016b.

Ma, X., Yang, K., La, Z., Lu, H., Jiang, Y., Zhou, X., Yao, X., and Li, X.: Importance of Parameterizing Lake Surface and Internal Thermal Processes in WRF for Simulating Freeze Onset of an Alpine Deep Lake. J. Geophys. Res. Atmos., 127. https://doi.org/10.1029/2022JD036759, 2022.

Meng, X., Lü, S., Gao, Y., and Guo, J.: Simulated effects of soil moisture on oasis selfmaintenance in a surrounding desert environment in Northwest China. Int. J. Climatol., 35 (14), 4116–4125. https://doi.org/10.1002/joc.4271, 2015.

Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J., and Clough, S.A.: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. J. Geophys. Res. Atmos., 102(14), 16663–16682. https://doi.org/10.1029/97JD00237, 1997.

Moosavi, A., Rao, V., and Sandu, A.: Machine learning based algorithms for uncertainty quantification in numerical weather prediction models. J. Comput. Syst. Sci., 50, 101295. https://doi.org/10.1016/j.jocs.2020.101295, 2021.

Nelli, N.R., Temimi, M., Fonseca, R.M., Weston, M.J., Thota, M.S., Valappil, V.K., Branch, O., Wulfmeyer, V., Wehbe, Y., Al Hosary, T., Shalaby, A., Al Shamsi, N., and Al Naqbi, H.: Impact of Roughness Length on WRF Simulated Land-Atmosphere Interactions Over a Hyper-Arid Region. Earth Space Sci., 7. https://doi.org/10.1029/2020EA001165, 2020.

Nearing, G.S., Ruddell, B.L., Clark, M.P., Nijssen, B., and Peters-Lidard, C.: Benchmarking and Process Diagnostics of Land Models. J. Hydrometeorol., 19, 1835–1852. https://doi.org/10.1175/JHM-D-17-0209.1, 2018.

Ozdogan, M. and Salvucci, G.D.: Irrigation-induced changes in potential evapotranspiration in Southeastern Turkey: Test and application of Bouchet's complementary hypothesis. Water Resour. Res., 40, W04301. https://doi.org/10.1029/2003WR002822, 2004.

Pan, X., Li, X., Yang, K., He, J., Zhang, Y., and Han, X.: Comparison of Downscaled Precipitation Data over a Mountainous Watershed: A Case Study in the Heihe River Basin. J. Hydrometeorol. 15, 1560–1574. https://doi.org/10.1175/JHM-D-13-0202.1, 2014.

Pan, X., Li, X., Cheng, G., and Hong, Y.: Effects of 4D-Var Data Assimilation Using Remote Sensing Precipitation Products in a WRF Model over the Complex Terrain of an Arid Region River Basin. Remote Sens., 9, 963. https://doi.org/10.3390/rs9090963, 2017.

Pan, X.D., Guo, X.J., Li, X., Niu, X.L., Yang, X.J., Feng, M., Che, T., Jin, R., Ran, Y.H., Guo, J.W., Hu, X.L., and Wu, A.D.: National Tibetan Plateau Data Center: Promoting Earth System Science on the Third Pole. Bull. Am. Meteorol. Soc., 102(11), E2062-E2078. https://doi.org/10.1175/BAMS-D-21-0004.1, 2021a.

Pan, X., Ma, W., Zhang, Y., and Li, H.: Refined Characteristics of Moisture Cycling over the Inland River Basin Using the WRF Model and the Finer Box Model: A Case Study of the Heihe River Basin. Atmosphere, 12, 399. https://doi.org/10.3390/atmos12030399, 2021b.

Pilguj, N., Taszarek, M., Pajurek, Ł., and Kryza, M.: High-resolution simulation of an isolated tornadic supercell in Poland on 20 June 2016. Atmos. Res., 218, 145–159. https://doi.org/10.1016/j.atmosres.2018.11.017, 2019.

Prein, A.F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tölle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., Lipzig, N.P.M., and Leung, R.: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. Rev. Geophys., 53, 323–361. https://doi.org/10.1002/2014RG000475, 2015.

Rahman, A., Maggioni, V., Zhang, X., Houser, P., Sauer, T., and Mocko, D.M.: The Joint Assimilation of Remotely Sensed Leaf Area Index and Surface Soil Moisture into a Land Surface Model. Remote Sens., 14, 437. https://doi.org/10.3390/rs14030437, 2022.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat.: Deep learning and process understanding for data-driven Earth system science. Nature, 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1, 2019.

Sawada, Y., Koike, T., and Walker, J.P.: A land data assimilation system for simultaneous simulation of soil moisture and vegetation dynamics: LDAS FOR ECOHYDROLOGICAL MODEL. J. Geophys. Res. Atmos., 120, 5910–5930. https://doi.org/10.1002/2014JD022895, 2015.

Seo, E., Lee, M.I., and Reichle, R.H.: Assimilation of SMAP and ASCAT soil moisture retrievals into the JULES land surface model using the Local Ensemble Transform Kalman Filter. Remote Sens. Environ., 253, 112222. https://doi.org/10.1016/j.rse.2020.112222, 2021.

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Liu, Z., Berner, J., Wang, W., Powers, J.G., Duda, M.G., Barker, D.M., and Huang, X.Y.: A Description of the Advanced Research WRF Version 4. NCAR Tech. Note NCAR/TN-556+STR, 145 pp, 2019.

Smedman, A., Högström, U., Sahlee, E., and Johansson, C.: Critical re-evaluation of the bulk transfer coefficient for heat over the ocean. Q. J. R. Meteorol. Soc., 133, 227–250. https://doi.org/10.1002/qj.7, 2007.

Song, X., Liu, F., Zhang, G., Li, D., and Zhao, Y.: Estimation of soil texture at a regional scale using local soil-landscape models. Soil Sci., 181, 435–445. https://doi.org/10.1097/SS.0000000000000180, 2016.

Thompson, G., Field, P.R., Rasmussen, R.M., and Hall, W.D.: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. Mon. Weather Rev., 136(12), 5095–5115. https://doi.org/10.1175/2008mwr2387.1, 2008.

Sun, S., Zheng, D., Liu, S., Xu, Z., Xu, T., Zheng, H., and Yang, X.: Assessment and improvement of Noah-MP for simulating water and heat exchange over alpine grassland in growing season. Sci. China Earth Sci., 64. https://doi.org/10.1007/s11430-021-9852-2, 2021.

Tian, J., Qin, J., Yang, K., Zhao, L., Chen, Y., Lu, H., Li, X., and Shi, J.: Improving surface soil moisture retrievals through a novel assimilation algorithm to estimate both model and observation errors. Remote Sens. Environ., 269, 112802. https://doi.org/10.1016/j.rse.2021.112802, 2022.

Wang, G. X., and Cheng, G.D.: Water resource development and its influence on the environment in arid areas of China-the case of the Hei River basin. J. Arid Environ., 43, 121–131, https://doi.org/10.1006/jare.1999.0563, 1999.

Wang, L., Chen, R., Song, Y., Yang, Y., Liu, J., Han, C., and Liu, Z.: Precipitation–altitude relationships on different timescales and at different precipitation magnitudes in the Qilian Mountains. Theor. Appl. Climatol., 134, 875–884. https://doi.org/10.1007/s00704-017-2316-1, 2018a.

Wang, X., Pang, G., Yang, M., Wan, G., and Liu, Z.: Precipitation changes in the Qilian Mountains associated with the shifts of regional atmospheric water vapour during 1960-2014. Int. J. Climatol., 38, 4355–4368. https://doi.org/10.1002/joc.5673, 2018b.

Wang, X., Tolksdorf, V., Otto, M., and Scherer, D.: WRF-based dynamical downscaling of ERA5 reanalysis data for High Mountain Asia: Towards a new version of the High Asia Refined analysis. Int. J. Climatol., 41, 743–762. https://doi.org/10.1002/joc.6686, 2021.

Wen, X., Lu, S., and Jin, J.: Integrating Remote Sensing Data with WRF for Improved Simulations of Oasis Effects on Local Weather Processes over an Arid Region in Northwestern China. J. Hydrometeor., 13, 573–587. https://doi.org/10.1175/JHM-D-10-05001.1, 2012.

Wu, D., Liu, S., Wu, X., Xu, T., Xu, Z., He, X., and Shi, H.: Evaluation of the intrinsic temperature sensitivity of ecosystem respiration in typical ecosystems of an endorheic river basin. Agric. For. Meteorol., 333, 109393. https://doi.org/10.1016/j.agrformet.2023.109393, 2023.

Wu, W. Y., Yang, Z. L., Zhao, L., and Lin, P.: The impact of multi-sensor land data assimilation on river discharge estimation. Remote Sens. Environ., 279, 113138. https://doi.org/10.1016/j.rse.2022.113138, 2022.

Xia, Y., Hao, Z., Shi, C., Li, Y., Meng, J., Xu, T., Wu, X., and Zhang, B.: Regional and Global Land Data Assimilation Systems: Innovations, Challenges, and Prospects. J. Meteorol. Res., 33, 159–189. https://doi.org/10.1007/s13351-019-8172-4, 2019.

Xiao, Z., Liang, S., Wang, J., Chen, P., Yin, X., Zhang, L., and Song, J.: Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance. IEEE Trans. Geosci. Remote Sens., 52(1), 209–223. https://doi.org/10.1109/TGRS.2013.2237780, 2014.

Xie, Y., Wang, P., Bai, X., Khan, J., Zhang, S., Li, L., and Wang, L.: Assimilation of the leaf area index and vegetation temperature condition index for winter wheat yield estimation using Landsat imagery and the CERES-Wheat model. Agric. For. Meteorol., 246, 194–206. https://doi.org/10.1016/j.agrformet.2017.06.015, 2017.

Xie, Z., Liu, S., Zeng, Y., Gao, J., Qin, P., Jia, B., Xie, J., Liu, B., Li, R., Wang, Y., and Wang, L.: A High-Resolution Land Model With Groundwater Lateral Flow, Water Use, and Soil Freeze-Thaw Front Dynamics and its Applications in an Endorheic Basin. J. Geophys. Res. Atmos., https://doi.org/10.1029/2018JD028369, 2018.

Xu, T., Valocchi, A.J., Choi, J., and Amir, E.: Use of Machine Learning Methods to Reduce Predictive Error of Groundwater Models. Ground water., 52, 448–460. https://doi.org/10.1111/gwat.12061, 2014.

Xu, T., Guo, Z., Liu, S., He, X., Meng, Y., Xu, Z., Xia, Y., Xiao, J., Zhang, Y., Ma, Y., and Song, L.: Evaluating Different Machine Learning Methods for Upscaling Evapotranspiration from Flux Towers to the Regional Scale. J. Geophys. Res. Atmos., 123, 8674–8690. https://doi.org/10.1029/2018JD028447, 2018.

Xu, T., He, X., Bateni, S.M., Auligne, T., Liu, S., Xu, Z., Zhou, J., and Mao, K.: Mapping regional turbulent heat fluxes via variational assimilation of land surface temperature data from polar orbiting satellites. Remote Sens. Environ., 221, 444–461. https://doi.org/10.1016/j.rse.2018.11.023, 2019.

Xu, T., Chen, F., He, Xinlei, Barlage, M., Zhang, Z., Liu, S., and He, Xiangping.: Improve the Performance of the Noah-MP-Crop Model by Jointly Assimilating Soil Moisture and Vegetation Phenology Data. J. Adv. Model. Earth Syst., 13. https://doi.org/10.1029/2020MS002394, 2021.

Xu, Z., Liu, S., Zhu, Z., Zhou, J., Shi, W., Xu, T., Yang, X., Zhang, Y., and He, X.: Exploring evapotranspiration changes in a typical endorheic basin through the integrated observatory network. Agric. For. Meteorol., 290, 108010. https://doi.org/10.1016/j.agrformet.2020.108010, 2020.

Yang, Z.L., Niu, G.Y., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Longuevergne, L., Manning, K., Niyogi, D., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. J. Geophys. Res., 116, D12110. https://doi.org/10.1029/2010JD015140, 2011.

Yi, L., Yong, B., Chen, J., Zheng, Z., and Li, L.: Impact of 4D-Var Data Assimilation on Performance of the Coupled Land–Atmosphere Model WRF–TOPX: A Case Study of a Flood Event in the Wangjiaba Watershed, China. J. Hydrometeorol., 22, 689–701. https://doi.org/10.1175/JHM-D-20-0161.1, 2021.

Yue, S., Yang, K., Lu, H., Zhou, X., Chen, D., and Guo, W.: Representation of Stony Surface-Atmosphere Interactions in WRF Reduces Cold and Wet Biases for the Southern Tibetan Plateau. Geophys. Res. Atmos., 126. https://doi.org/10.1029/2021JD035291, 2021.

Zhang, G., Chen, F., and Gan, Y.: Assessing uncertainties in the Noah-MP ensemble simulations of a cropland site during the Tibet Joint International Cooperation program field campaign: UNCERTAINTY IN NOAH-MP SIMULATIONS. J. Geophys. Res. Atmos. 121, 9576–9596. https://doi.org/10.1002/2016JD024928, 2016.

Zhang, X., Xiong, Z., and Tang, Q.: 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, 2017a.

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

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

Zhang, Z., Barlage, M., Chen, F., Li, Y., Helgason, W., Xu, X., Liu, X., and Li, Z.: Joint Modeling of Crop and Irrigation in the central United States Using the Noah-MP Land Surface Model. J. Adv. Model. Earth Syst., 12. https://doi.org/10.1029/2020MS002159, 2020.

Zhang, Z., Arnault, J., Laux, P., Ma, N., Wei, J., Shang, S., and Kunstmann, H.: Convection-permitting fully coupled WRF-Hydro ensemble simulations in high mountain environment: impact of boundary layer- and lateral flow parameterizations on land–atmosphere interactions. Clim. Dyn., https://doi.org/10.1007/s00382-021-06044-9, 2021a.

Zhang, Z., Arnault, J., Laux, P., Ma, N., Wei, J., and Kunstmann, H.: Diurnal cycle of surface energy fluxes in high mountain terrain: High-resolution fully coupled atmosphere-hydrology modelling and impact of lateral flow. Hydrol. Process., 35. https://doi.org/10.1002/hyp.14454, 2021b.

Zhang, Z., Chen, F., Barlage, M., Bortolotti, L.E., Famiglietti, J., Li, Z., Ma, X., and Li, Y.: Cooling effects revealed by modeling of wetlands and land-atmosphere interactions. Water Resour. Res., https://doi.org/10.1029/2021WR030573, 2022.

Zhao, J., Feng, H., Xu, T., Xiao, J., Guerrieri, R., Liu, S., Wu, X., He, Xinlei, and He, Xiangping.: Physiological and environmental control on ecosystem water use efficiency in response to drought across the northern hemisphere. Sci. Total Environ., 758, 143599. https://doi.org/10.1016/j.scitotenv.2020.143599, 2021.

Zhao, L. and Yang, Z.L.: Multi-sensor land data assimilation: Toward a robust global soil moisture and snow estimation. Remote Sens. Environ., 216, 13–27. https://doi.org/10.1016/j.rse.2018.06.033, 2018.

Zhao, W.L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, C., Li, X., and Qiu, G. Y.: Physics-Constrained Machine Learning of Evapotranspiration. Geophys. Res. Lett., 46, 14496–14507. https://doi.org/10.1029/2019GL085291, 2019.

Zheng, C., Liu, S., Song, L., Xu, Z., Guo, J., Ma, Y., Ju, Q., and Wang, J.: 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, 2023.

Zhong, B., Ma, P., Nie, A., Yang, A., Yao, Y., Lü, W., Zhang, H., and Liu, Q.: Land cover mapping using time series HJ-1/CCD data. Sci. China Earth Sci., 57, 1790–1799. https://doi.org/10.1007/s11430-014-4877-5, 2014.

Zhou, Y., Liao, W., and Li, X.: The contributions of individual factors to the oasis cold island effect intensity in the Heihe River Basin. Agric. For. Meteorol., 312, 108706. https://doi.org/10.1016/j.agrformet.2021.108706, 2022.