

Article

Development and Evaluation of a WRF-Based Mesoscale Numerical Weather Prediction System in Northwestern China

Tiejun Zhang ^{1,2} , Yaohui Li ^{1,2}, Haixia Duan ^{1,2,*}, Yuanpu Liu ^{1,2}, Dingwen Zeng ^{1,2}, Cailing Zhao ^{1,2}, Chongshui Gong ^{1,2}, Ganlin Zhou ^{1,2}, Linlin Song ^{1,2} and Pengcheng Yan ^{1,2}

¹ Institute of Arid Meteorology, China Meteorological Administration; Lanzhou 730020, China; kwke@163.com (T.Z.); li-yaohui@163.com (Y.L.); liuyp@iamcma.cn (Y.L.); zengdw@iamcma.cn (D.Z.); zhaocl@iamcma.cn (C.Z.); cs.gong@outlook.com (C.G.); zhougl@iamcma.cn (G.Z.); linlin@lzb.ac.cn (L.S.); yanpc@iamcma.cn (P.Y.)

² Key Laboratory of Arid Climatic Change and Reducing Disaster of Gansu Province, Key Laboratory of Arid Climatic Change and Disaster Reduction, China Meteorological Administration, Lanzhou 730020, China

* Correspondence: duanhx@iamcma.cn

Received: 28 May 2019; Accepted: 20 June 2019; Published: 25 June 2019



Abstract: Based on the U.S. Weather Research and Forecasting (WRF) numerical model, this study has developed the Northwest Mesoscale Numerical Prediction Service and Experimental System (NW-MNPS). Surface and sounding data assimilation has been introduced for this system. Effects of model vertical layers and land-use data replacement have been assessed. A year-long forecast validation and analysis have been performed. The following results have been obtained: (1) Data assimilation can improve the performance of regional numerical forecasting. (2) Compared to simulations with 40 vertical layers, simulations with 55 vertical layers are more accurate. The average absolute error and root-mean-square error of the 48 h surface element forecast decrease. The analysis of threat score (TS) and equitable threat score (ETS) shows that there are higher TS and ETS values for various precipitation intense levels, in particular for heavy rainfall when comparing a 55-vertical-layer test with a 40-vertical-layer test. (3) Updating the database to include vegetation coverage can more accurately reflect actual surface conditions. The updated land-use data reduce prediction errors in all domains of the NW-MNPS.

Keywords: numerical prediction; land-use update; vertical layers; data assimilation

1. Introduction

Northwestern China is located in the north and northeast of the Qinghai–Tibetan Plateau in central Asia. It accounts for 24.5% of the total area of China [1]. This region has complex and varied topography, including plateaus, plains, and mountains. Deserts are the predominant feature, with an arid climate and sparse vegetation. This arid, ecologically fragile zone plays an important role in climate change in East Asia as well as globally. It is also sensitive to climate change. Its weather and climate show significant regional characteristics and complexity, including plateau, westerly flow, and monsoon climate zones. Thus, the problems of weather and climate in this region are very complex [2,3].

Northwestern China is upstream of the synoptic and circulation system in mainland China, which is one of the key areas where cold air excited by arctic oscillation (AO) enters mainland China, and the strong wind generated by the narrow tube effect under special terrain over northwestern China can affect the weather development of the North China Plain. Accurate weather forecast systems in this

region are important for downstream forecasts in the rest of China. In addition to high-standard requirements for refinement and grids in weather models, the northwestern prediction service must also consider the needs of disaster relief services in different regions. The arid northwestern region is located inland, and is easily affected by floods and debris flows [4]. Thus, the same precipitation events that do not cause disasters in eastern China can cause varying degrees of destruction in the northwest. Therefore, there is a high demand for catastrophic strong-convection weather forecasts in the northwestern region.

Regional numerical modeling in northwestern China has changed over the past several decades, alongside developments in modeling capabilities in other nations. The earliest was a comprehensive mesoscale numerical forecast experiment model system developed by Chen et al. in the 1990s (MONFEMS), which used the improved MM4 by Pennsylvania State University–National Center for Atmospheric Research (PSU/NCAR) as its core and consisted of four parts including real-time meteorological data preprocessing, model initialization, mesoscale numerical prediction, and post-processing [3,5,6]. With the gradual development of numerical models, the Lanzhou Central Meteorological Observatory and the Chinese Academy of Meteorological Sciences jointly introduced and developed a limited regional mesoscale numerical prediction service system based on MM5 V3.6 in 2004, which has played an important role in the forecasting of catastrophic storms and flooding [7].

In spring 2005, to meet the demand of the northwestern dust prediction service, the Lanzhou Institute of Arid Meteorology of the China Meteorological Administration introduced the Global/Regional Assimilation and Prediction System sandstorm forecast model (GRAPES-SDM), which was a preliminary version of the new generation of the dust prediction system called GRAPES_CUACE/Dust [8]. It couples with the weather model GRAPES developed by the Numerical Model Center of Chinese Meteorological Administration and a Canadian dust aerosol model. It was able to simulate and predict the sand streak of sandstorms and atmospheric dust concentration to provide guidelines for sandstorm forecasts in the northwest [9].

Since the continental characteristics of northwestern China are more complicated than those in other regions of China [10], models have more limitations here than in eastern regions. For example, the significant difference between the actual and modeled vegetation cover in the northwestern region leads to prediction errors [11]. The complex land-use conditions in the northwestern region leads to greater errors of prediction. Sparse observation stations and poor observation data quality also affect forecasting in northwestern China. Thus, it is imperative to develop a Numerical Weather Prediction (NWP) system with the land–atmosphere interaction characteristics of the northwestern region in order to improve and enhance the prediction ability of numerical models in northwestern China.

Numerical models can generate different results when weather/climate conditions and natural geographic environment (e.g., land-use and cover) are changed. Due to the different dynamical core, grid design, spatial and temporal resolution, various models have different regional applicability [12], and the applicability of parameterizations in different regions varies significantly [13,14]. The Northwest Mesoscale Numerical Prediction Service and Experimental Systems (NW-MNPS) was developed in 2014 and focuses on the development and application of numerical weather predictions in northwestern China. NW-MNPS fully considers the actual demand of service prediction, optimizes the physical processes, assimilates surface and air sounding data, and makes predictions within a timeframe from one hour to one week.

Specifically, Section 3 highlights the updates of the vegetation coverage data and compares the effects on prediction of two separate months (July and December 2015); Section 4 discusses the impact of model vertical resolution on prediction by using three precipitation processes; Section 5 takes the single precipitation process as an example to analyze the impact of data assimilation in the precipitation process. Full-year results are compared and validated in Section 6. Finally, Section 7 summarizes and discusses the results.

2. Model System Design

NW-MNPS uses the mesoscale Weather Research and Forecasting (WRF) model as the basic framework. WRF is a new-generation model for mesoscale forecasting developed by multiple research institutes and universities in the United States. It has the advantage of easy transplant and maintenance as well as high efficiency [13]. The WRF model, WPS (WRF Pre-processing System) and WRFDA (WRF Data assimilation) systems used in this study are version 3.6.1. The WRF version for this study was chosen to maintain consistency with the operational system in use by the NW-MNPS.

This study uses a three-layer nesting scheme. The simulated region is given in Figure 1, with the center at 39° N, 100° E. Detailed domain parameters are given in Table 1. The model uses the terrain-following-mass η coordinate with 40 vertical layers and then upgrades to 55 layers. Based on Chou's [15] conclusion, the larger upper-level interpolation error in the low vertical resolution case produces an analysis with warmer low-level temperatures and larger surface pressure compared to the high-resolution case. The top layer of the model is 50 hPa.

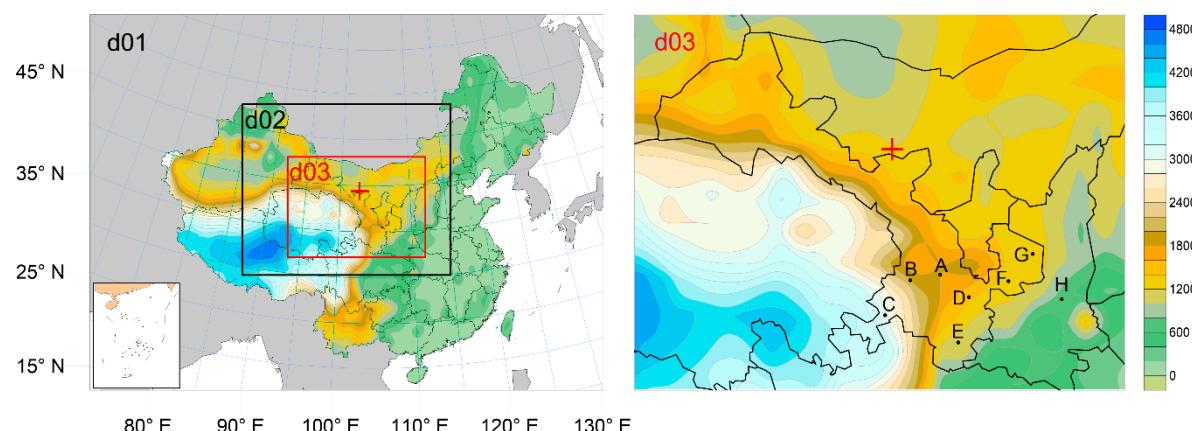


Figure 1. Map of the nested domains in the study. The grid spacings for D01, D02 and D03 are 27 km, 9 km, and 3 km, respectively. The dots in D03 represent the station names mentioned in this article. A: Dingxi, B: Linxia, C: Gannan, D: Tianshui, E: Longnan, F: Pingliang, G: Qingsyang, H: Shanxi.

Table 1. Parameters of the nested grid in three domains.

Domain	Grid Numbers	Spatial Resolution (km)	Temporal Resolution (s)
D01	220 × 173	27	162
D02	274 × 214	9	54
D03	562 × 376	3	18

Figure 2 is a schematic diagram of the operation flow of NW-MNPS; the main functional modules are WPS, WRFDA, REAL, NDOWN and WRF. The WPS program imports Global Forecast System (GFS) data, performs data re-projection, and coordinates the time-space interpolation of surface parameters and other observations. The REAL program will create the initial and boundary fields based on the WPS results. The NDOWN program can create input files from higher resolution terrestrial fields, and masked surface fields for a finer-grid run. During the simulation process, the modeling timelines of D01 and D02 are 172 h and 72 h, respectively. The interval of the model output is 3 h. The time line and model output interval of D03 are 48 h and 1 h, respectively. D01 and D02 contain 12 hours of spin-up time. The cold start times of D01 and D02 are 00:00 UTC and 12:00 UTC, respectively. We use the Global Forecast System (GFS) data from National Centers for Environmental Prediction (NCEP) as the model initial values and boundary conditions and update the data every three hours with WPS. The initial values of D03 are the nested predictions of D02 with NDOWN. We use the one-way nesting method to obtain forecast products. The flow chart is shown in Figure 1. Considering the impact of land-use on evaporation and infiltration, and since the land-use in the model was outdated, we first

chose the land-use parameter in the upgrade system to correctly describe the surface parameters [16]. Soil temperature and moisture observations will further change the local circulation and affect the weather [17,18]. So the first step was to update the land-use to improve the simulation accuracy.

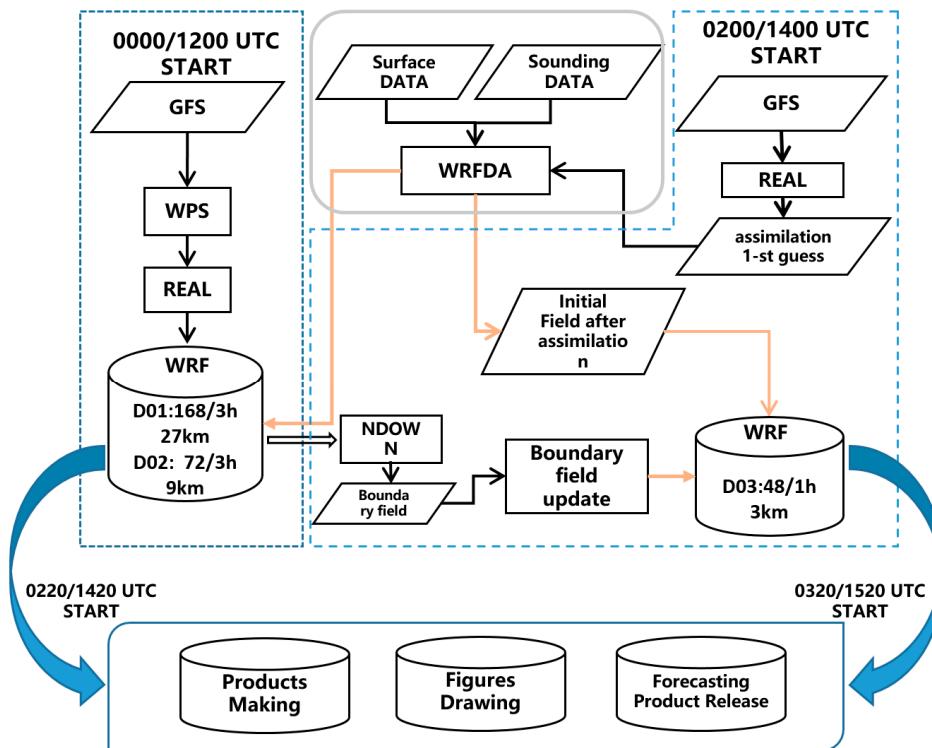


Figure 2. Northwest Mesoscale Numerical Prediction Service and Experimental System (NW-MNPS) running flow chart.

The modeling accuracy is greatly affected by errors in the initial conditions [19]. WRFDA is conducted using three-dimensional variational data assimilation (3DVAR), which is physically homogeneous and dynamically harmonious [3]. This study uses 3DVAR to assimilate numerous encrypted observational data in real-time, including surface data such as conventional weather observations and automatic meteorological stations.

The NW-MNPS developed in this study includes three parts: data assimilation, model execution, and model post-processing. Data assimilation includes the assimilation of background data and observational data. Model post-processing consists of three parts: product processing and production, product release, and result validation. Parameterizations are set up based on a large amount of background research in this region, and are given in Table 2 [20–22].

Table 2. Parameterization scheme selection.

Cumulus	Cumulus 3 km	Land Surface	Boundary Layer
Kain–Fritsch(KF)	NONE	NOAH	Asymmetric Convection Model 2 (ACM2)
Long wave radiation RRTM	Short wave radiation Dudhia	Surface Monin–Obuklov	Microphysical Thompson Graupel

Precipitation is defined by the following levels: light (0–10 mm/day), moderate (10–25 mm/day), heavy (25–50 mm/day), storm (50–100 mm/day), h_storm (>100 mm/day), which is consistent with the China Meteorological Administration (CMA).

Model evaluation tools (MET) is an objective assessment tool for numerical prediction, which has been jointly developed by NCEP and NCAR with the support of the U.S. Air Force and NOAA. It integrates worldwide advanced model validation techniques and is applicable to validating the outputs of the WRF model and other numerical models. The point_stat tool in MET is used mainly to validate and assess site gridded data. This study uses observed data as site observational data; thus, to validate the NW-MNPS model, we first need to interpolate the model grid results to the observation sites and then perform the site validation and assessment.

The following are some variables we used for comparison tests. The mean error (ME) of each high-altitude element with a prediction timeline of 24 h, is expressed as

$$ME = \frac{\sum (\varphi_{im} - \bar{\varphi}_{io})}{N} \quad (1)$$

We examined the mean absolute errors (MAE) of land surface elements, which is defined as

$$MAE = \frac{\sum |\varphi_{im} - \varphi_{io}|}{N} \quad (2)$$

The root-mean-square error (RMSE) can be used to compare forecasting errors of different models for the observation dataset, which is defined as

$$RMSE = \left[\frac{\sum (\varphi_{im} - \varphi_{io})^2}{N} \right]^{1/2} \quad (3)$$

The spatial correlation coefficient can be used to indicate the simulation ability of the model for the observed spatial distribution, which is expressed as

$$CORR = \frac{\sum (\varphi_{im} - \bar{\varphi}_m)(\varphi_{io} - \bar{\varphi}_o)}{\left[\sum (\varphi_{im} - \bar{\varphi}_m)^2 \sum (\varphi_{io} - \bar{\varphi}_o)^2 \right]^{\frac{1}{2}}} \quad (4)$$

The precipitation assessment uses the TS score method, also known as the threat score

$$TS = NA / (NA + NB + NC) \quad (5)$$

where NA is the number of correct reports, NB is the number of false reports, and NC is the number of missed reports.

The precipitation ETS score, also known as the equitable TS score, is expressed as

$$ETS = (NA - r) / (NA + NB + NC - r) \quad (6)$$

where

$$r = (NA + NB)(NA + NC) / (NA + NB + NC + ND) \quad (7)$$

ND is the correct reports number with no precipitation event. The ETS score can effectively remove the probability effects of random precipitation on the scores, as well as avoid meteorological probability. According to the definition, the ideal score of ETS is 1, with a range of -0.3 to 1. An ETS score in the area with more precipitation is lower than the TS score.

3. The Effects of Vegetation Coverage Data on Prediction

3.1. Description of Vegetation Coverage Data

The vegetation coverage data (land-use) that comes with the WRF model was generated relatively early. As time passes, its ability to represent the actual landscape gradually worsens. This study refers to the satellite remote-sensing inversion of land-use data based on China land survey data from 2000 to 2005, using GIS (Geographic Information System) spatial analysis and identification algorithms to merge high-resolution Moderate Resolution Imaging Spectroradiometer (MODIS) data and International Geosphere Biosphere Programme (IGBP) land coverage data, and process the merged land classification data with the MODIS water mask product to produce a China land cover product with a spatial resolution of 500 m [23].

Figure 3 shows a comparison of different land-use data in the simulated area, the United States Geological Survey (USGS) dataset is reclassified using the IGBP land-use dataset classification method. As shown, compared with the land-use data that comes with the model, changes in the new land-use data include mainly the following: shrub area decreases and broad-leaved forest, tropical savanna, farmland, vegetation mosaic, and water body areas all decrease. Additionally, mixed forest, grassland, urban, snow-covered, and desert areas all increase.

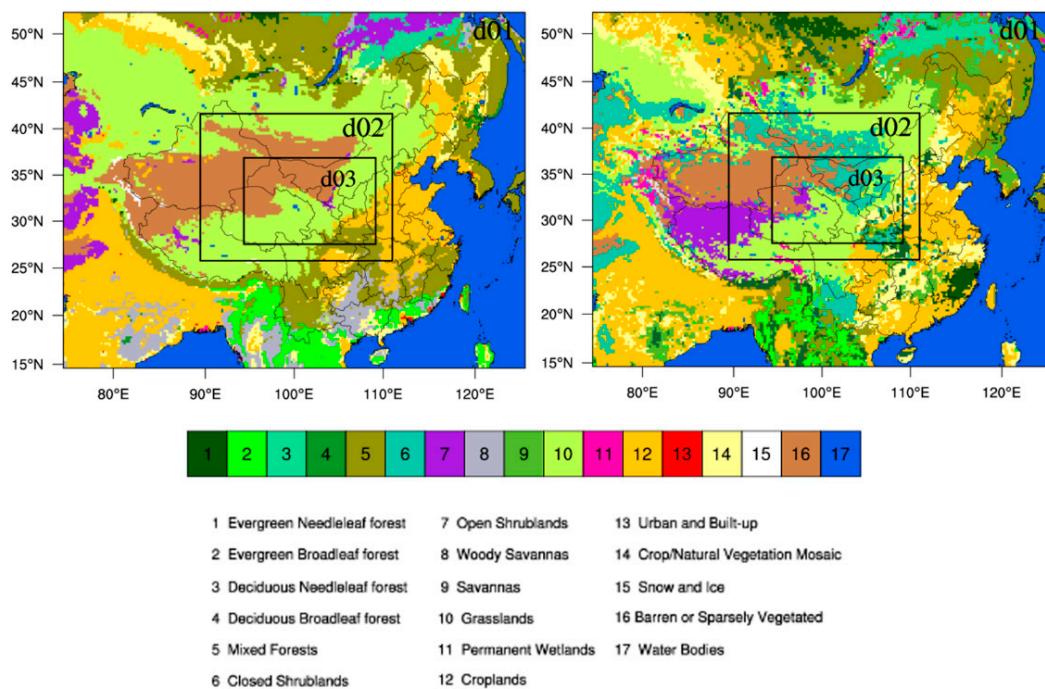


Figure 3. Comparison of different land-use data distributions in the simulation area USGS (**left**) and new land-use (**right**).

3.2. Experimental Design

To study the effects of different land-use data on the model 2m temperature, this study performs two simulations (Table 3): (1) Control test, using the land-use data (USGS data) that comes with the model (24 types); and (2) a new land cover run, using land-use data (17 types) obtained from the 2009 satellite inversion based on the IGBP (International Geosphere-Biosphere Programme) classification. The simulated model times are July and December 2015.

Table 3. Schemes of the simulation test.

Test	Test Period	Land-Use Data Used in the WRF Model
Control test	1–31 July 2015/1–31 December 2015	USGS
New land cover run	1–31 July 2015/1–31 December 2015	International Geosphere-Biosphere Programme (IGBP)

3.3. Temperature Prediction Validation

Figure 4 shows a comparison of the error (ME, mean error) and absolute error (MAE, mean absolute error) of the simulated near-surface temperature. The comparison of the area with three nested layers shows that with an increase in the forecast lead time, the temperature prediction error increases, which is similar before and after the update of land-use data. The comparison with the replaced land-use data shows that the new land-use data can effectively improve the model's prediction ability of near-surface temperature in D01 and D02. In D03, the forecast effect with new land-use has reduced.

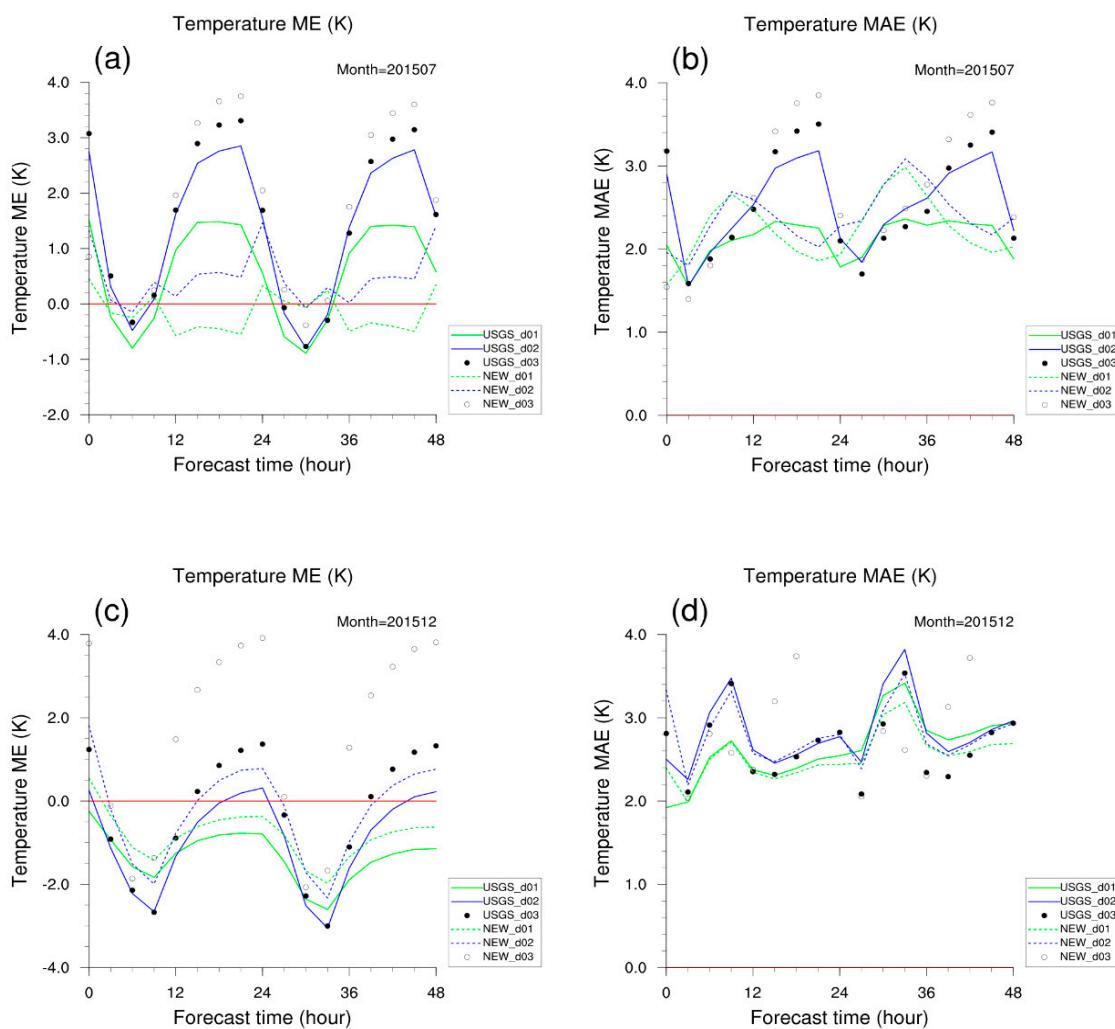


Figure 4. The error (a,c) and the absolute error (b,d) of the temperature prediction for different nested regions changed with the forecast.

Figure 5 shows a comparison of the simulated average daily error and absolute error in D01, D02, and D03. The results show that the simulated surface air temperature results have overestimated except for a few days in July 2015. Replacing the default USGS land-use data with the new land-use data improves the simulation of surface air temperature in northwest China in three domains in July

2015. In December 2015, the simulated results underestimated the surface air temperature. The use of the new land-use data can effectively reduce the negative deviation. Overall, the use of the new land-use data can enhance the simulating skills of the model in surface air temperature in July and December 2015 in northwest China.

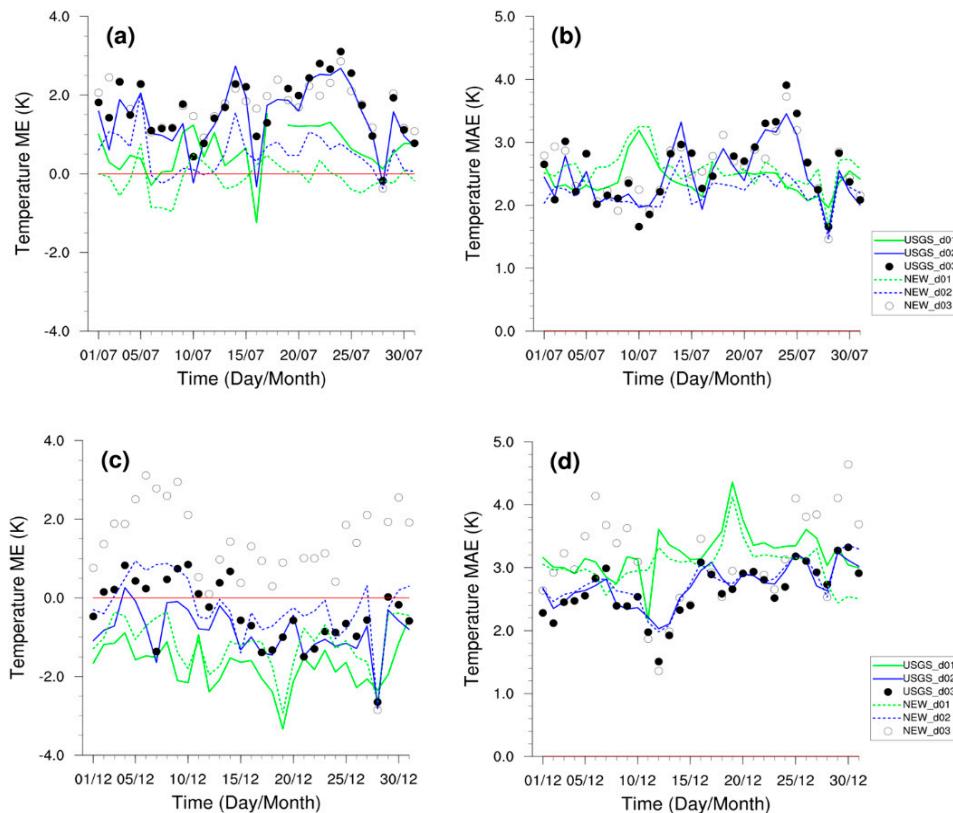


Figure 5. The error (a,c) and the absolute error (b,d) of the simulated temperature.

3.4. Wind Speed Prediction Validation

Figure 6 shows a comparison of MEs and MAEs of the simulated wind speed at 10 meters above ground level. The areas with three nested domains show that the prediction results are best in D03. With the increase in the prediction timeline, the simulated wind speed error increases. The comparison between July 2015 (Figure 6a,b) and December 2015 (Figure 6c,d) shows that the simulated error is smaller in July 2015, with a maximum simulated error (D01) of 2.1 m/s. The maximum simulated error (D01) in December 2015 is 2.5 m/s. With the increase in model resolution, the wind speed prediction error decreases. After replacing the land-use data, the wind speed prediction errors in D03 are less than 1.5 m/s in both July and December 2015. With the increase in the prediction timeline, the wind speed prediction error increases. Overall, in July 2015, the error and standard deviation of the wind speed prediction in D01 (D02 and D03) are less than 2.5 m/s (2.2 m/s and 2.2 m/s). In December 2015, the error and standard deviation of the wind speed prediction in D01 (D02 and D03) are less than 3.0 m/s (2.3 m/s and 2.0 m/s).

Figure 7 shows a comparison of MEs and MAEs for the daily average predictions in the three domains. The results show that the higher the horizontal resolution, the smaller the average error and absolute error of the wind speed prediction. Overall, the errors all decrease. The improvement from using the updated land-use data for daily prediction is similar to that of the real-time prediction. Compared with July 2015, the updated land-use data show more improvement in wind speed prediction in December. This may be because the December wind speed error is greater, and thus the improvement is bigger.

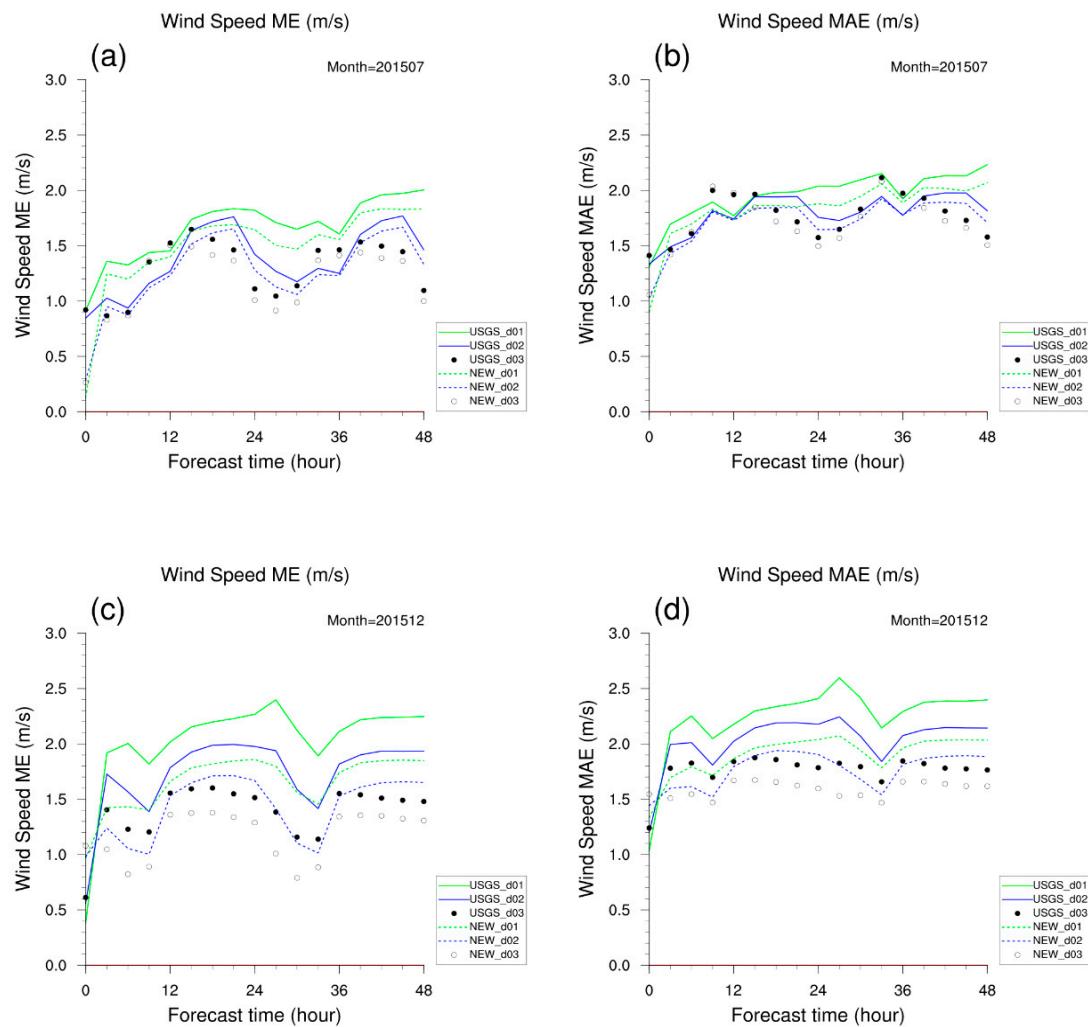


Figure 6. The error (a,c) and the absolute error (b,d) of the wind prediction for different nested regions changed with the forecast.

Table 4 shows the MEs and MAEs between the simulation results and observations in July and December 2015. The observations more clearly compare the observed and simulated surface air temperature based on the USGS and IGBP land-use data. The IGBP land-use dataset has better performance on surface air temperature in December 2015 for D01 and D02 than the USGS data. The ME of D01 (0.61) and D02 (1.35) in July 2015 and D01 (-1.76) and D02 (-0.98) in December 2015 for the simulated surface air temperature based on the USGS land-use data is greater than the ME of the simulated surface air temperature based on the IGBP land-use data. The MAE of D02 (2.61) in July 2015 and D01 (3.22) and D02 (2.91) in December 2015 based on the USGS land-use data is greater than MAE of the simulated surface air temperature based on the IGBP land-use data. However, compared with the USGS land-use data, the use of the IGBP land-use data results in higher ME and MAE for the simulated surface air temperature of D03 in July and December 2015. The factors causing this phenomenon require further analysis.

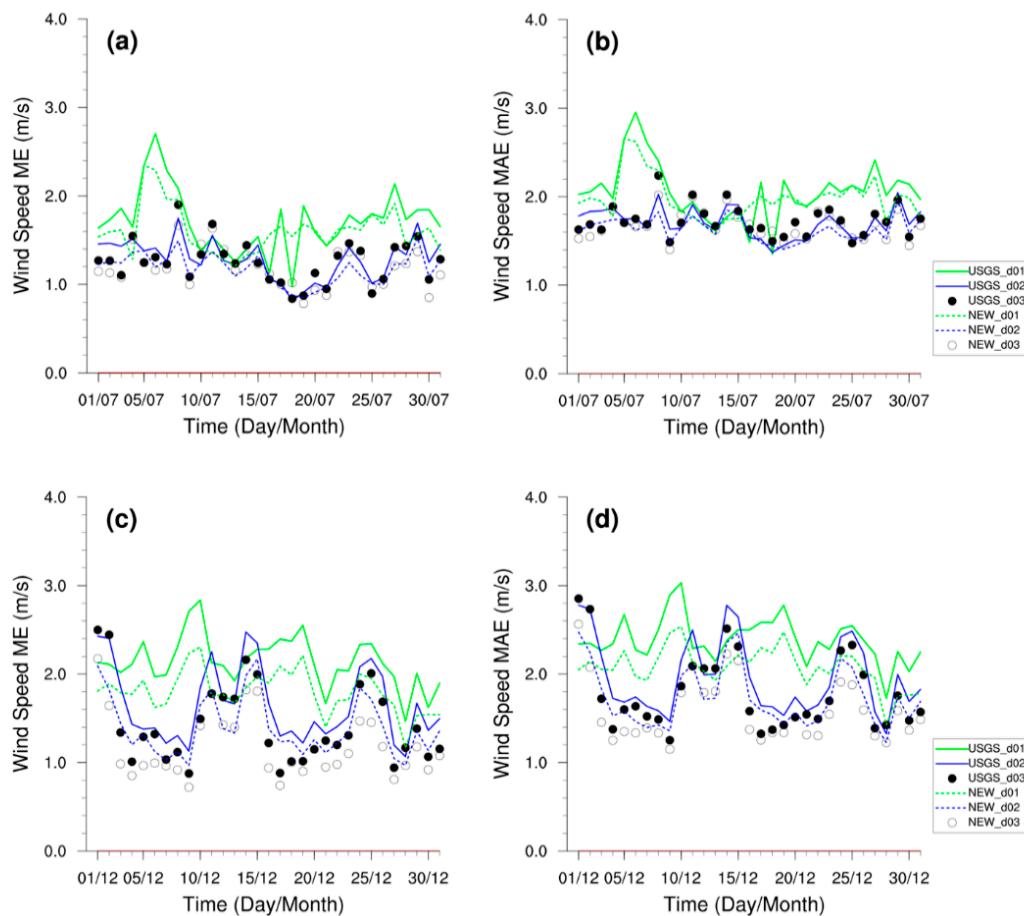


Figure 7. The error (**a,c**) and the absolute error (**b,d**) of the simulated wind speed.

Table 4. Comparison of mean error (bias) and mean absolute error (MAE) of the temperature for three regions varying from D01 to D03.

		ME		MAE	
		July 2015	December 2015	July 2015	December 2015
D01	Control and observation	0.61 *	-1.76 *	2.41 *	3.22 *
	New land cover and observation	-0.12	-1.18 *	2.57	3.01 *
D02	Control and observation	1.35	-0.98 *	2.61	2.91 *
	New land cover and observation	0.44 *	-0.32	2.49 *	2.83
D03	Control and observation	1.54 *	-0.37	2.58 *	2.66
	New land cover and observation	1.76 *	1.46 *	2.72 *	3.13 *

* Passed t-test with a significance (α) of 0.05.

4. The Effect of Model Vertical Resolution on Precipitation Prediction

4.1. Experimental Design

We selected one precipitation event for each month in the summer of 2015 including June, July, and August. The initial reporting times of the three events are 00:00 UTC 22 June 2015; 00:00 UTC 8 August 2015; and 00:00 UTC 11 August 2015. The first precipitation area is composed of Longnan, Tianshui, Pingliang, and Qingyang. The second precipitation area includes Dingxi, Tianshui, and Pingliang. The third precipitation area includes Tianshui, Pingliang, and its junction with Shanxi

(relevant stations are shown in Figure 1). The precipitation levels in all three events reach storm (50–100 mm/day) and h_storm (>100 mm/day).

The control experiment was performed using three nested layers: D01 and D02 are two-way nested, and D03 is one-way nested in D02. The results of D02 are used as the control experiment. We use 40 vertical layers in the control experiment, and 55 layers in the sensitivity experiment. In these experiments, all model parameters are kept the same except for the number of model vertical layers. As shown in Figure 8, there are more near-surface levels in the 55-layer setup.

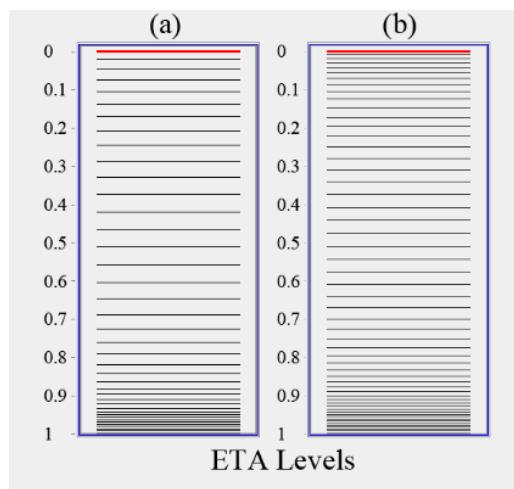


Figure 8. Forty layers (a) and 55 layers (b) of vertical stratification.

4.2. Analysis of Results

4.2.1. Land Surface Elements

We mainly compare the 48 h MAE of the three events (Figure 9). The comparison analysis shows that there are smaller 3 h MAEs when the sensitivity experiment is compared with the control experiment, suggesting that increasing vertical layers indeed improves the performance of the forecast system. As the prediction timeline increases, the error differences gradually decrease. As for relative humidity, the errors at 0 h of the three events with 55 layers are all below 10%, whereas those with 40 layers are all above 17%. The August precipitation error even exceeds 20%. The MAE of absolute humidity also shows a similar change with time. Thus, in general, the 3 h MAEs with 55 layers are all smaller than those with 40 layers. The MAE difference is largest at the initial time and decreases in the later stages of prediction. The MAEs at the initial time are all smaller than 0.0012 kg/kg with 40 layers and greater than 0.0015 kg/kg (even reaching 0.002 kg/kg) with 55 layers. With 55 layers, the daily change in absolute humidity MAE is small.

In Figure 10, after increasing the underlying layer, the correlation coefficient of the observation and simulation of the 55 layers of meteorological elements is higher than that of the 40 layers. The advantage of the 55 layers is more obvious than the prediction of the three elements of humidity, relative humidity, and temperature. The advantage can last for 18 to 24 h, and the 55-layer advantage of the wind forecast is not obvious, which is related to the poor predictability of the wind speed itself.

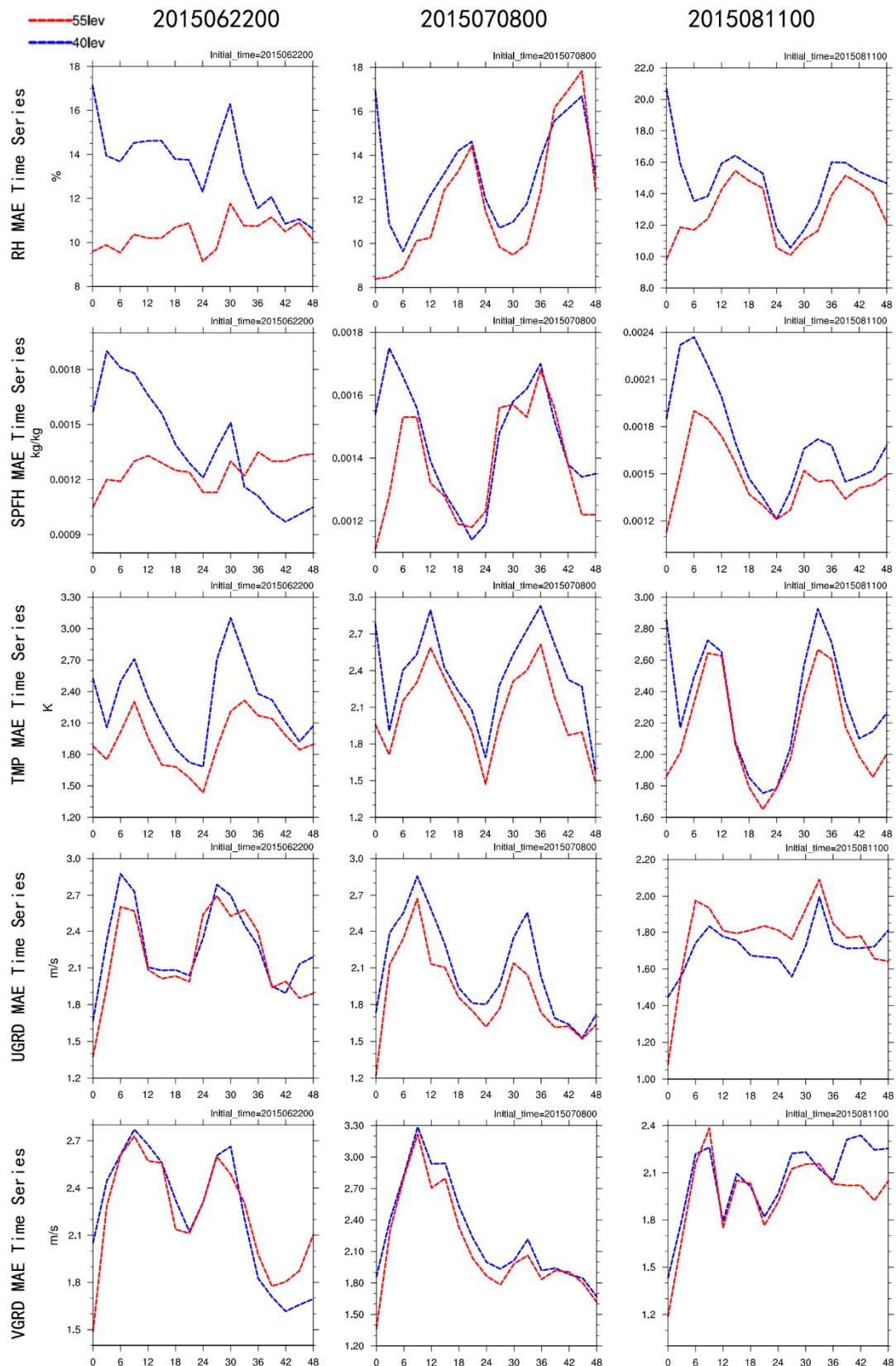


Figure 9. The mean absolute error (MAE) of the ground elements in three rain-event case studies. (RH: relative humidity; SPFH: specific humidity; TMP: temperature at 2 m; UGRD: Wind component in the latitudinal direction at 10 m (ground); VGRD: Wind component in the meridional direction at 10 m (ground)).

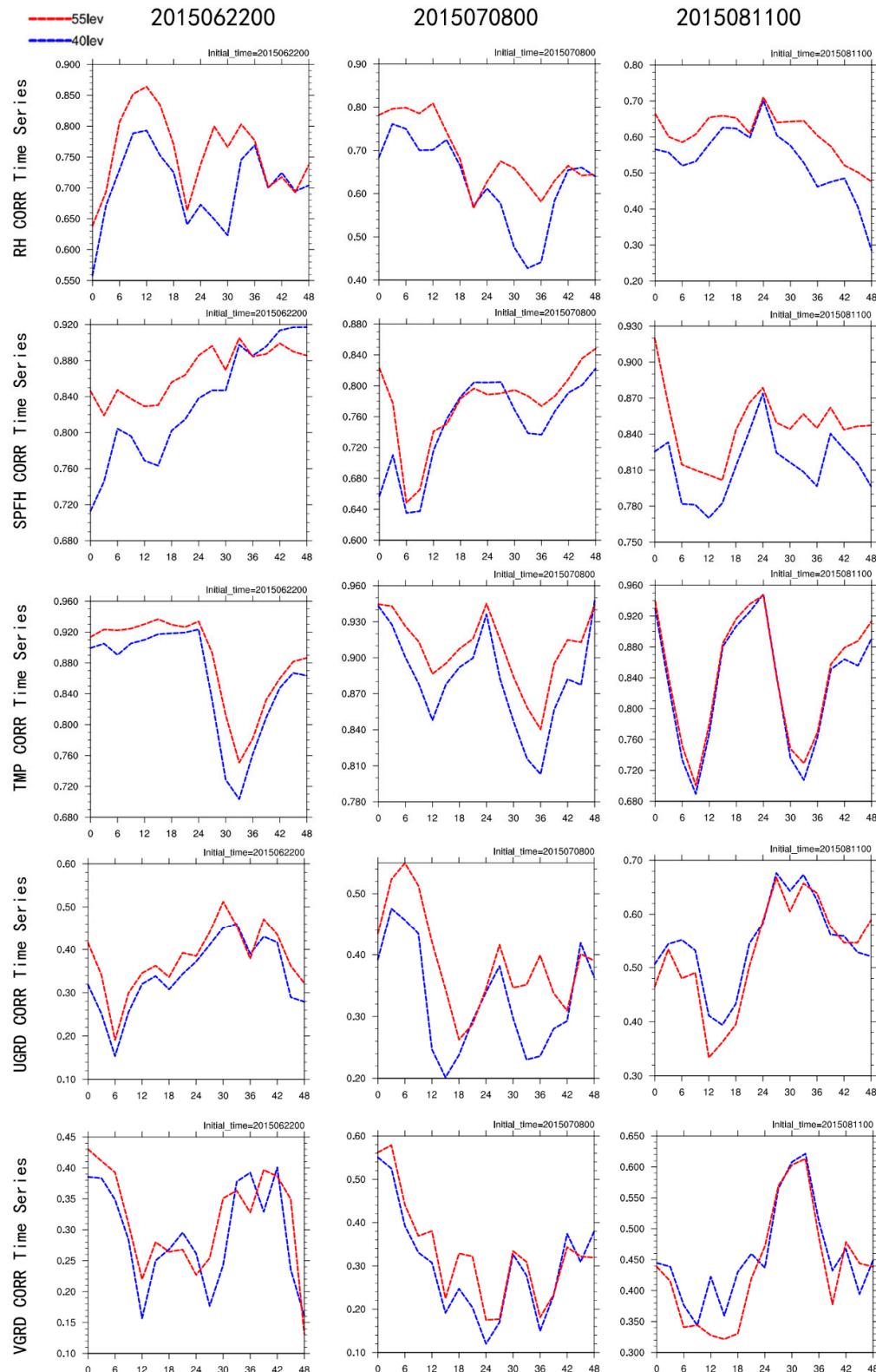


Figure 10. The correlation coefficient of ground elements in three rain-event case studies.

The MAEs of wind speed (zonal, meridional) are different from those of humidity and temperature. During the three precipitation events, the MAEs with 55 layers are smaller than those with 40 layers within 24 h, whereas they are comparable between 24 h and 48 h.

In the validation of surface meteorological elements, a small RMSE is used to indicate that the degree of prediction dispersion relative to the corresponding observation is small. The relative humidity in the three precipitation events is different in the two experiments. First, the RMSE with 55 layers is smaller than that with 40 layers. Using the June event as an example, the initial values are 12% with 55 layers and 20.5% with 40 layers, which is a difference of almost two times. Second, the daily RMSE change with 55 layers is smaller. The change in the June event is 14–21% with 40 layers and 12.5–15% with 55 layers. The difference characteristics of absolute humidity are similar.

The temporal RMSE changes for the 2 m temperature are identical between 40 layers and 55 layers, whereas the dispersion degree with 55 layers is smaller than that with 40 layers during the entire prediction timeline. Additionally, the daily RMSE temperature change with 55 layers is smaller; thus, the prediction results are better. The RMSE of wind speed (zonal and meridional) with 55 layers is smaller at the early stage of prediction and comparable at later stages of prediction.

In the three precipitation events, the spatial correlation coefficients of surface relative humidity, absolute humidity, and 2 m temperature with 55 vertical layers are all higher than those with 40 layers. This suggests that the higher the vertical resolution, the more accurate the model simulation for horizontal distributions of surface temperature and humidity. As for the simulation of spatial distribution for 10 m wind speed, the results with 55 layers are also better than those with 40 layers. However, the spatial correlation coefficient of wind speed is generally lower than that of temperature and humidity. The model's ability to simulate the spatial distribution of each element is best for temperature, followed by absolute humidity and relative humidity, and is worst for wind speed.

4.2.2. High-Altitude Elements

The sounding data were used for the upper level validation in Figure 13a and the stations in D02 were selected. The MEs of temperature prediction are smaller in the middle to lower layers of the troposphere; thus, the results with 55 layers are even better. The analysis of full wind speed ME shows that the model-predicted wind speed with 55 layers is higher than that with 40 layers.

In all three events, the TS scores of 24 h precipitation with 55 layers are all higher than those with 40 layers from light rain to heavy rain, and the heavier the precipitation, the more obvious the advantage with 55 layers, especially on 8 July and 11 August (Figures 11 and 12). Heavy rain scores with 55 layers are, respectively, 0.2 and 0.3, whereas those with 40 layers are, respectively, 0.05 and 0.1, smaller than those with 55 layers. The rainstorm score on 11 August is 0.1 with 55 layers and 0 with 40 layers. Thus, the model predictions for heavy rain and rainstorm are more accurate with 55 layers. The 48 h TS scores also show that the predictions with 55 layers are better than those with 40 layers.

In the three events, ETS scores are all lower than TS scores, and the 24 h ETS scores with 55 layers are all higher than those with 40 layers. Similar to TS scores, ETS scores also show that the heavier the precipitation, the higher the advantage of the prediction with 55 layers. The ETS scores of heavy rain on 8 July and 11 August are, respectively, 0.2 and 0.25 with 55 layers and, respectively, 0.05 and 0.1 with 40 layers. The ETS score of the rainstorm on 11 August is 0.1 with 55 layers and 0 with 40 layers, indicating the lack of a prediction technique. In addition, we compare the precipitation difference (figure not shown). Overall, the model-predicted precipitation is higher than the observed precipitation, and the prediction with 55 layers is higher than that with 40 layers.

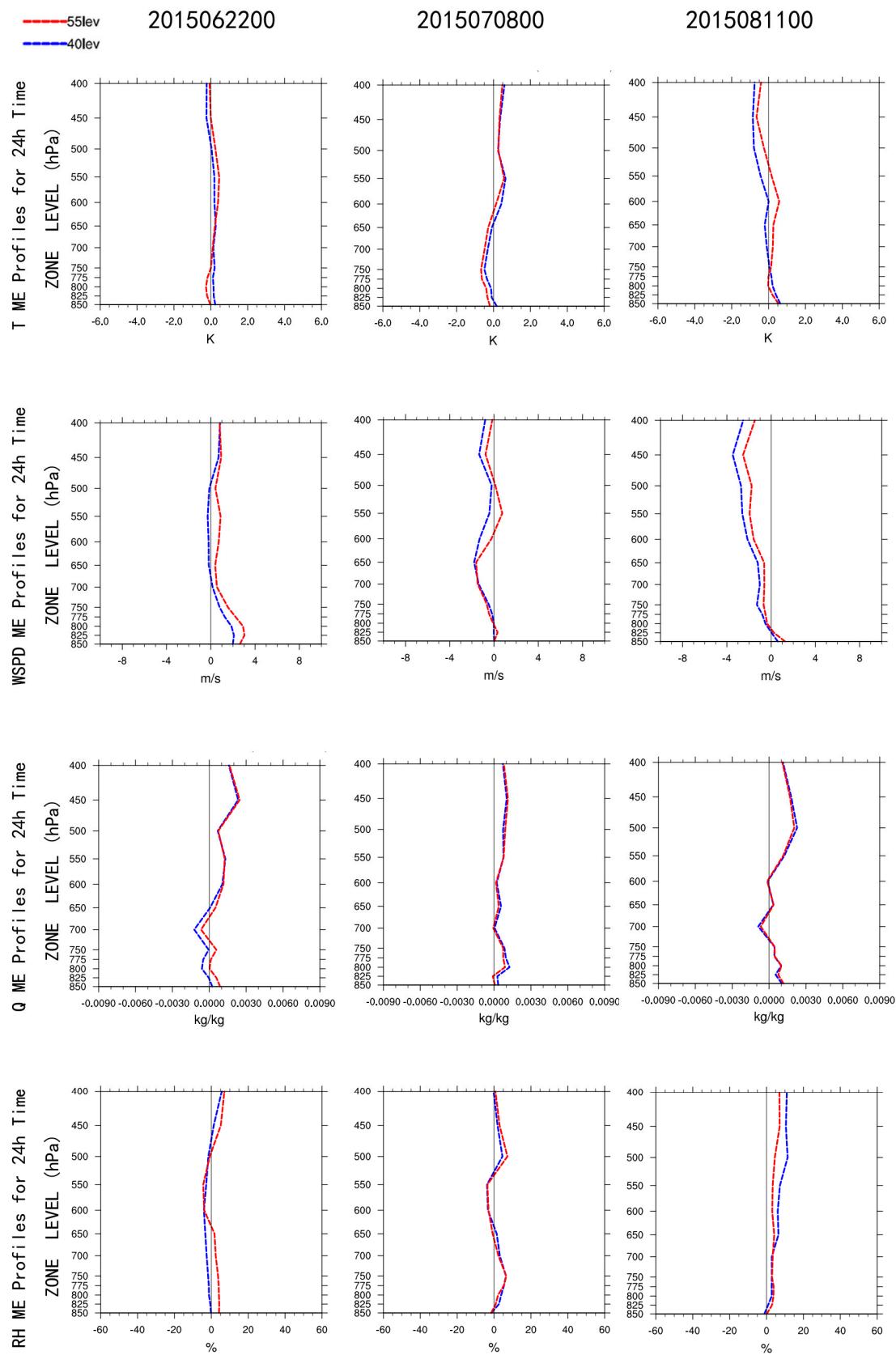


Figure 11. The mean error (ME) of high-altitude elements in three weather processes (T: temperature; WSPD: wind speed; Q: specific humidity; RH: relative humidity).

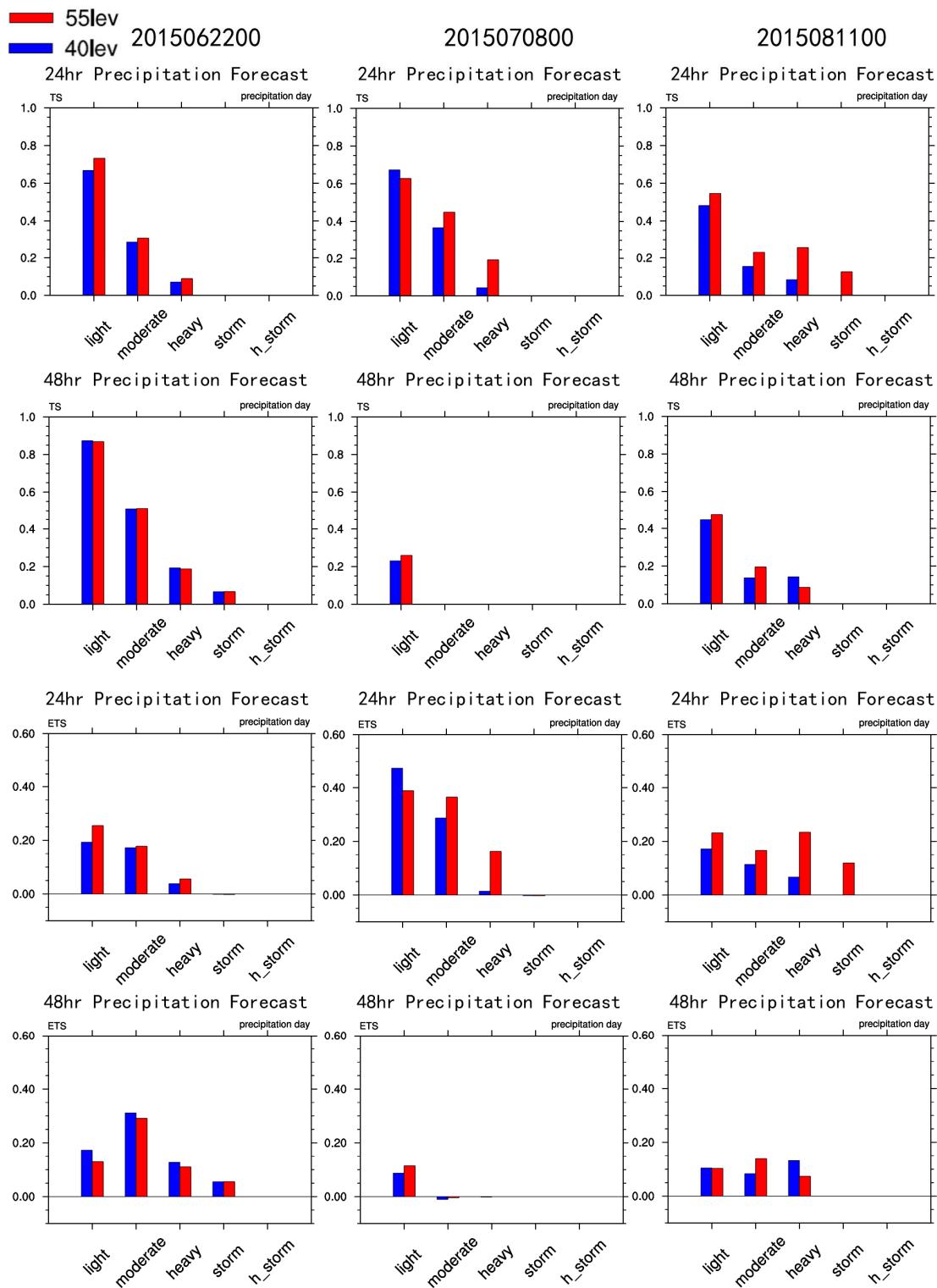


Figure 12. Equitable threat score (ETS) scores for three precipitation events.

5. The Assimilation of Conventional Data and Its Impact on Precipitation

5.1. The Design of Data Assimilation Schemes

NW-MNPS uses the modular 3DVAR included in the WRF package to assimilate conventional surface observations and air sounding data. Assimilation occurs daily at 00:00 UTC and 12:00 UTC to update model initial values. Using the NMC (National Meteorological Center) method [24], this study

conducts a statistical analysis of the 12 h and 24 h predictions for 1–31 August 2015, to generate a new background error covariance matrix (B matrix) that is suitable for northwestern China.

The assimilation experiments are four contrasting experiments including NODA without the assimilation of conventional surface observations or air sounding data; AWS, with the assimilation of only conventional surface observations; SOUNDING, with the assimilation of only air sounding data; and A + S (AWS + SOUNDING), with the assimilation of both conventional surface observations and air sounding data. Detailed station distributions are given in Figure 13. The distribution of air sounding stations is rather homogeneous, whereas the distribution density of surface observations in the eastern and southern areas is higher than that in the northwestern, northeastern, and Qinghai–Tibet areas.

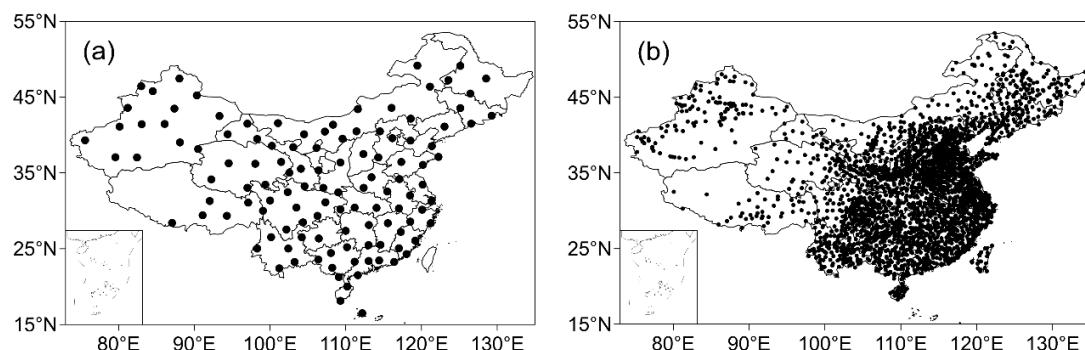


Figure 13. Sounding stations (a) and ground stations distribution (b) in China.

5.2. Comparison of Assimilation Experiment Results

We compare the improvements provided by the four assimilation experiments for model prediction. Figures 14 and 15 show the comparison of 24 h accumulated precipitation on 8–9 July 2015 between the four assimilation methods and the actual observations. The precipitation zone in the A + S group is closer to the observation, and the precipitation zone in NODA shifts westward in Gannan and Linxia. The model predicts more precipitation than the observation, suggesting a biased performance of the forecast system. The 24 h TS score shows that compared to the NODA experiment, predicted technical scores improve for all precipitation intensity levels when data assimilation procedure is involved. Furthermore, the improvement is more significant for heavier precipitation cases.

Analysis of surface fields show that the errors in A + S, AWS, and SOUNDING are smaller than those in NODA at the initial time. This suggests that the assimilation of surface observations and upper-air sounding data has a positive impact for model initial values.

Figure 16 shows the evolution of ME for each variable with time. The four schemes all capture the diurnal variation in the bias in 2 m temperature. The temperature bias predicted by the AWS scheme is lower. The forecast deviations of U (Wind component in the latitudinal direction), specific humidity (Q), and relative humidity (RH) of the four scenarios are negative throughout the forecast period, while the forecast deviations for V (Wind component in the meridional direction) and wind speed are positive for the entire forecast period. The 10 m U wind speed, Q, and RH deviations predicted by the AWS scheme are smaller than other schemes, and the predicted V wind speed and full wind speed is slightly larger than other schemes.

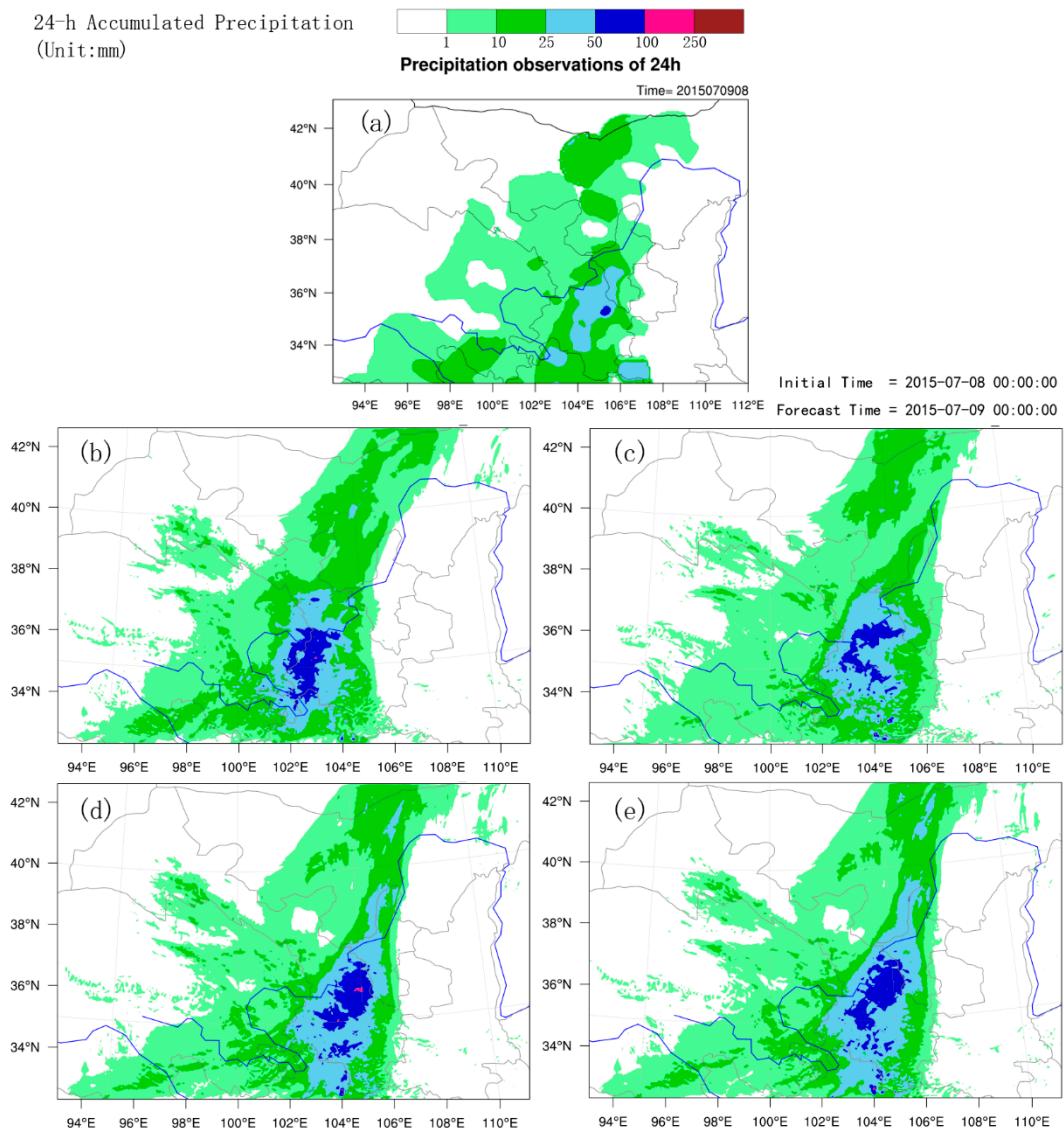


Figure 14. Precipitation on 8–9 July 2015 for 24 h (a) Observation, (b) NODA, (c) AWS, (d) SOUNDING, (e) AWS + SOUNDING.

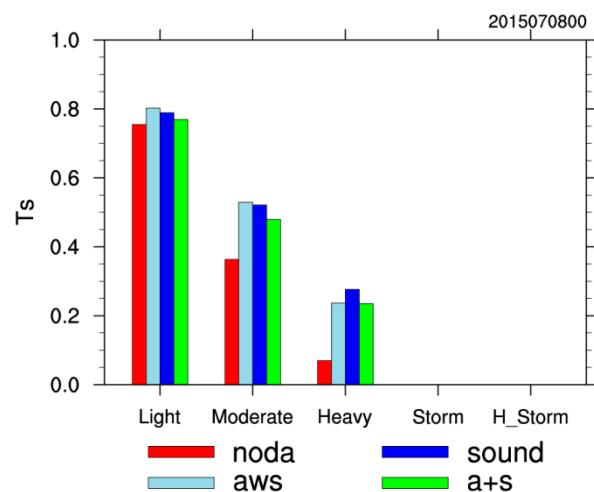


Figure 15. Threat score (TS) score on 8–9 July 2015.

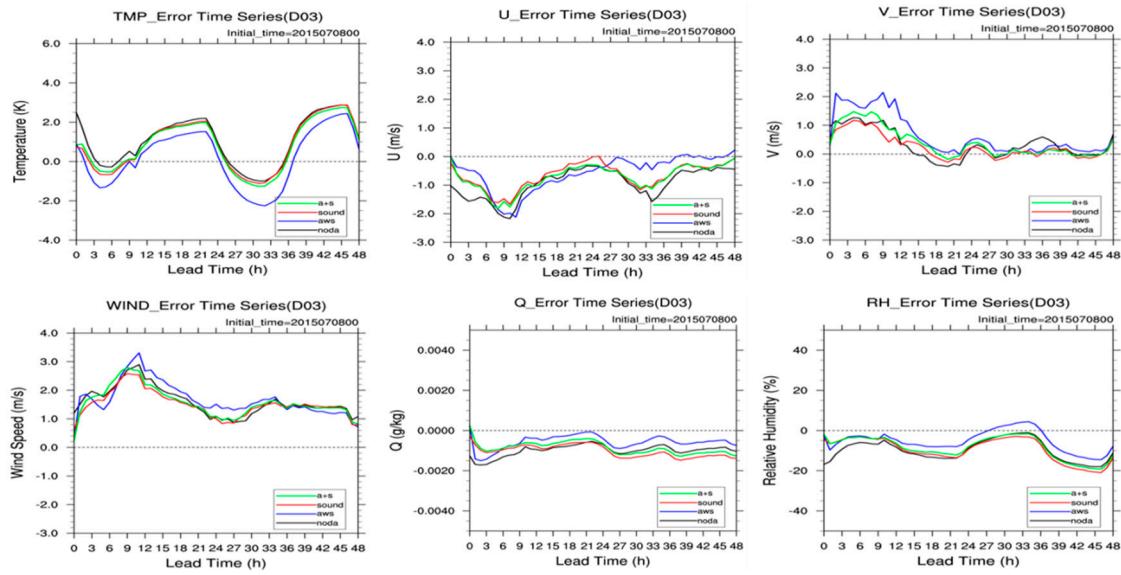


Figure 16. Time series for each forecasted variable using different data assimilation methods, on 8–9 July 2015.

Figure 17 shows the evolution of MAE for each element. The absolute errors of 2 m T (temperature), 10 m U, V, and wind speed of AWS forecast are larger than other schemes in the first 36 h, while the A + S and SOUNDING schemes are absolute for these factors. The error is the smallest in the forecast before 36 hours. The absolute error of the A + S and SOUNDING schemes for Q and RH is quite small in the first 12 h. The absolute error after 12 h is slightly higher than other schemes, indicating that the initial field is in addition to the ground data and sounding data. The hourly forecast gets some improvement. In addition, it can be seen from the figure that the error of the initial moment of the NODA scheme is the largest among all the elements, indicating that the addition of ground data and sounding data at the initial time can effectively improve the accuracy of the initial value at the beginning.

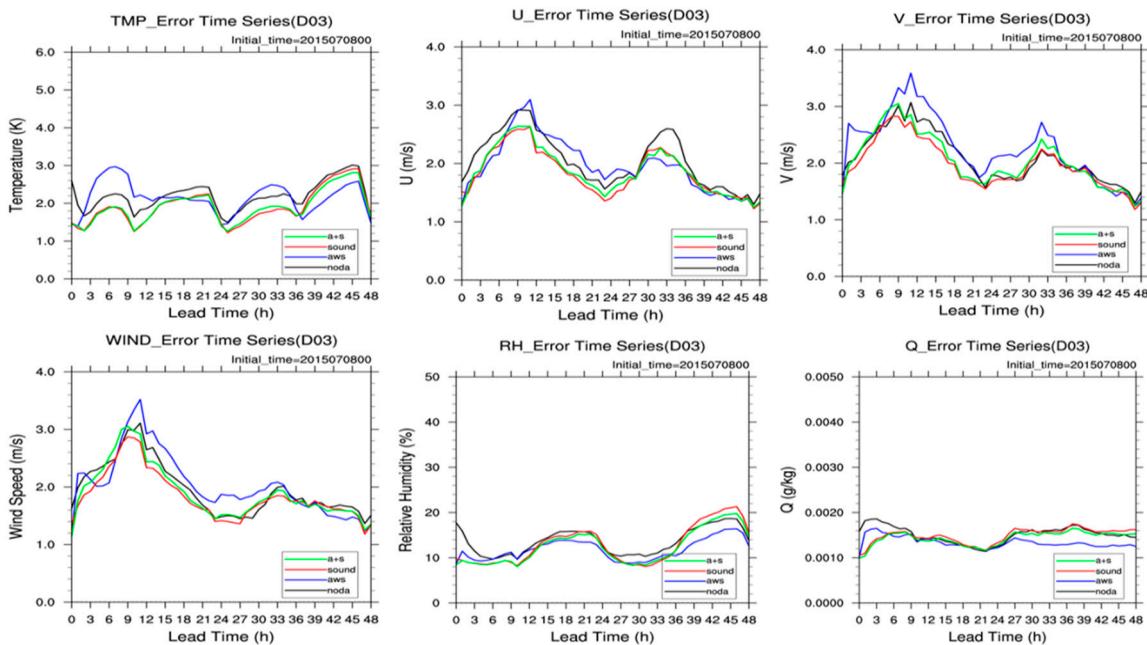


Figure 17. Absolute error of each element on 8–9 July 2015.

6. Model Prediction Validation

This section uses mainly the point_stat tool in MET to analyze the evaluation results of the 72 h prediction for NW-MNPS in D02. The duration is one year, from 1 May 2015 to 30 April 2016.

6.1. Assessment of the Overall Model Prediction Ability

6.1.1. Surface Elements

For surface prediction elements, we mainly tested the 2 m temperature, 10 m wind speed, and 2 m relative humidity. The year-round ME, MAE, and RMSE (Figure 18) show daily changes in the 2 m temperature and 2 m relative humidity. The 2 m temperature prediction is low during the daytime and high during the nighttime, leading to a smaller diurnal difference, whereas the 2 m relative humidity prediction is high during the daytime and low during the nighttime, leading to a larger diurnal difference. The 10 m wind speed prediction is greater and has no daily change. The overall prediction results show that the prediction RMSEs of the 2 m temperature, 10 m wind speed and 2 m relative humidity for NW-MNPS are 3.0–4.0 K, 2.0–3.0 m/s, and 20%, respectively. The prediction errors of all three elements increase with an increased prediction timeline.

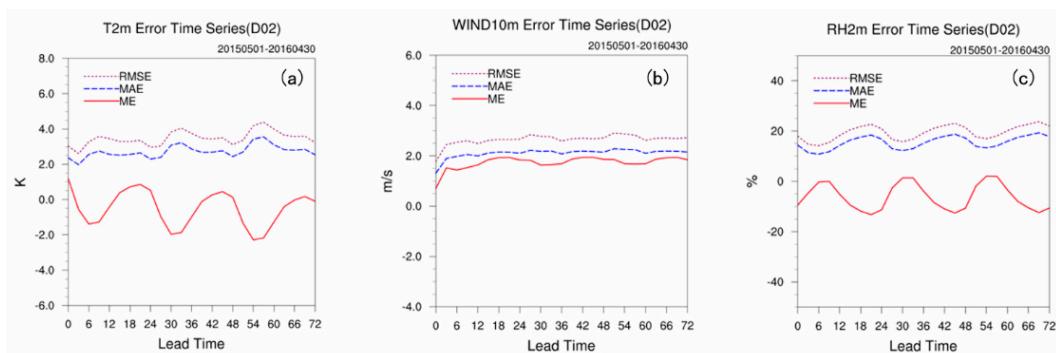


Figure 18. The ground element error, the absolute error, and the root-mean-square error with the forecast aging change, (a) 2 m temperature, (b) 10 m wind speed, (c) 2 m relative humidity.

6.1.2. High-Altitude Elements

The NW-MNPS model prediction of high-altitude elements with vertical height is given in Figure 19. The year-round average prediction shows that the model predicted temperatures are low. The 24 h RMSEs are generally within 2 K at high altitudes, except at the land surface, and those of the 48 h and 72 h predictions are within 2–4 K. The model prediction error of air temperature is greater in the lower layers than that in the higher layers. The model prediction of relative humidity is higher, with an RMSE of approximately 20%. Additionally, the model-predicted relative humidity error is greater in higher layers than that in lower layers. The model-predicted wind speed error is greater at lower levels and smaller at higher levels. The RMSEs at higher levels are greater than in lower layers by approximately 3–4 m/s.

6.1.3. Precipitation

The NW-MNPS model prediction effects of 24 h precipitation are shown by TS, ETS, and FBIAS (fractional bias) (Figure 20). The cumulative number of light rains is between 500 and 4500 per month in D02 and that number of moderate rains is between 80 and 600. Very few heavy events happened (only five times). With an increase in the model prediction timeline, the scores decrease and the prediction effect deteriorates. We assess it using the average year-round precipitation scores. The TS scores show that the model TS score is over 0.4 for the 24 h light rain, whereas it is relatively low with rains stronger than moderate, especially for downpours, for which the score is generally 0. The ETS scores show that the ETS of the model-predicted light rain is approximately 0.3, whereas it is relatively low for rains

stronger than moderate. The prediction error shows that the model over predicts precipitation for light and moderate rain, especially for light rain, which can be twice as high. However, the model prediction of precipitation stronger than heavy rain is less, especially rainstorm prediction. The precipitation probability in the northwestern region shows that this region is dominated by light and moderate rain, and heavy rains and even stronger rains are relatively rare, especially rainstorms and downpours. Thus, this model can effectively predict general precipitation events, whereas its prediction of heavy rains with a small probability is much worse.

6.2. Statistics of the Year-Round Evaluation

The following section analyzes the statistics of the NW-MNPS model prediction of surface elements, high-altitude elements, and precipitation with different timelines (24 h, 48 h, and 72 h) between 1 May 2015 and 30 April 2016 (one-year duration), and the model prediction of different elements in different seasons.

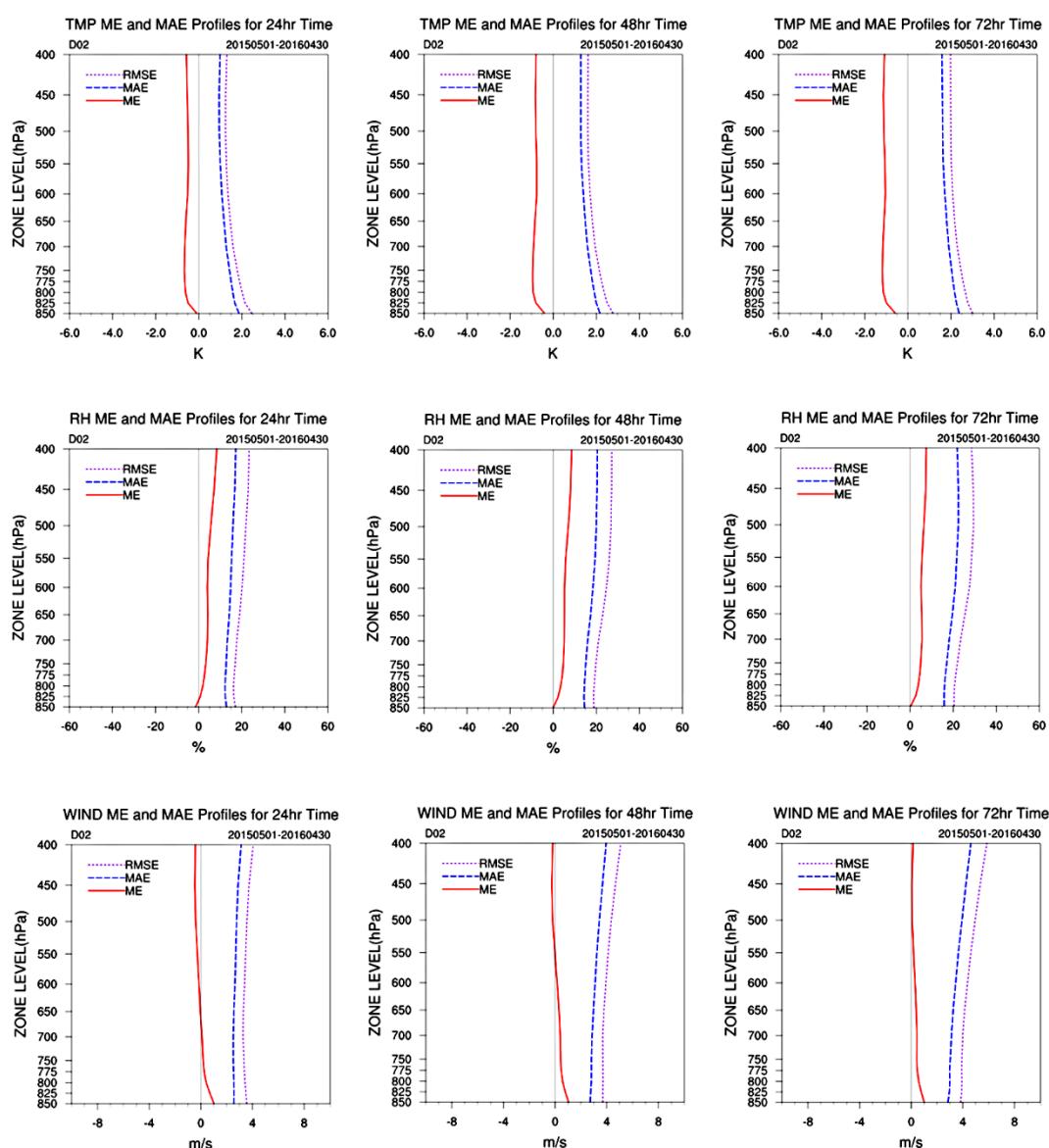


Figure 19. High-altitude element prediction average error, absolute error, and root-mean-square error with height vertical profile.

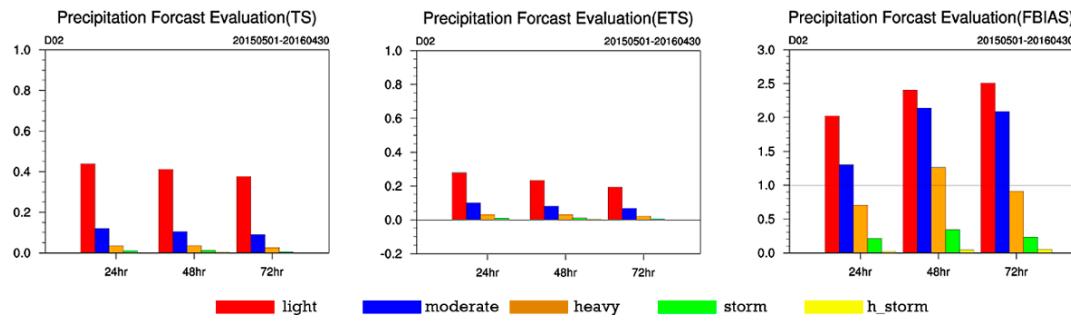


Figure 20. Precipitation forecast score.

6.2.1. Surface Variables

The evaluation of the NW-MNPS model prediction of surface elements focuses on the temporal changes of average errors and RMSEs for the 24 h 2 m temperature, 2 m relative humidity, and 10 m wind speed (Figure 21). The predictions in different seasons show that the summer 2 m temperature prediction is better than that during winter, and the fall prediction is close to summer and that of spring is close to winter. The model prediction of the 10 m wind speed is also better during summer than during winter. The seasonal RMSEs in the 2 m relative humidity model prediction are all below 20%.

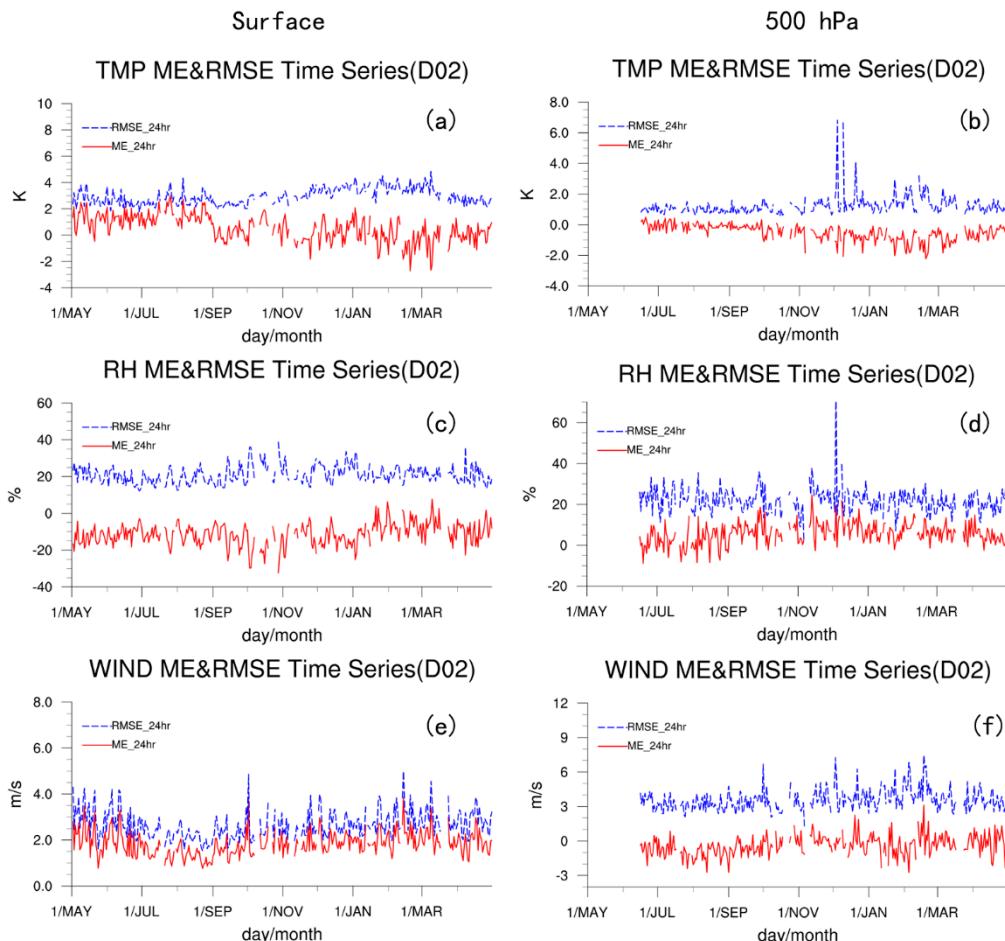


Figure 21. Average error and root-mean-square error of temperature, relative humidity, and wind speed at surface and 500 hPa.

6.2.2. High-Altitude Elements

The evaluation of the NW-MNPS model prediction of high-altitude elements focuses on the temporal changes in average errors and RMSEs of the 24 h 500 hPa temperature, relative humidity, and wind speed (Figure 21). The model prediction of air temperatures for different seasons shows that model prediction is best during summer, with RMSEs remaining below 2 K, and relatively worse during winter. The results are unstable and partially exceed 2 K, indicating that the model prediction of winter weather requires further improvement. The seasonal differences of model prediction for relative humidity and wind speed are clear, with worse prediction during the winter.

6.2.3. Twenty-Four Hour Precipitation

The annual changes in the NW-MNPS model prediction for the 24 h precipitation at different levels are given in Figure 22. We evaluate mainly the TS scores of light rain, moderate rain, and heavy rain in the northwestern area. Precipitation predictions at different levels show that the TS score of the light rain prediction is highest. The TS scores of the model-predicted precipitation for different seasons show that the TS score of the model-predicted 24 h precipitation is highest during summer, with a maximum of up to 0.8 and worst during winter, with values generally lower than 0.6. Since winter precipitation is sparse, the TS scores of the predicted precipitation with levels equal to, or greater, than moderate are generally 0.

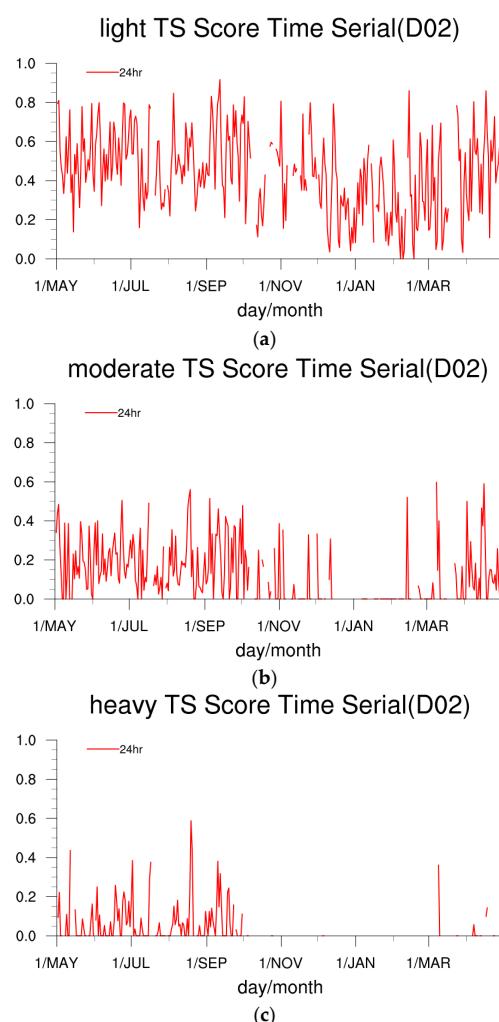


Figure 22. TS score changes over 24 h prediction of (a) light rain, (b) moderate rain, and (c) heavy rain.

7. Conclusions and Discussion

Based on the WRF model, the Northwest Mesoscale Numerical Prediction Service and Experiment System (NW-MNPS) has been developed. We tested the impact of assimilating ground and sounding data, increasing vertical layers (from 40 into 55), and changing land-use and cover data on predictability of the NW-MPS. The simulation was performed July and December 2015 (i.e., 1–31 July, 1–31 December). Predicted precipitation, 2 m air temperature, 10-m wind speed, as well as relative humidity, temperature, wind speed, and specific humidity profile are compared with the observations. The conclusions are summarized as follows:

1. Adding a data assimilation process can significantly enhance and improve the prediction ability and performance of the system. The results suggest that assimilating ground or air sounding data to the system can significantly improve the forecast skills although assimilating both data has the largest improvement. If different forecast periods are compared, data assimilation has most significant improvement for its 6–9 h forecast product.
2. Analysis of surface fields shows that the simulated spatial distributions/patterns are closer to the observations when the sensitivity with 55 vertical layers and control experiment with 40 vertical layers are compared. The sensitivity experiment has smaller averaged errors for 48 h predicted 2 m air temperature, absolute humidity, and relative humidity when compared to the control experiment, in particular at the initial time. As forecast time increases, the difference between sensitivity and control experiment decreases.
3. Comparisons of the 24 h predictions of high-altitude elements show that the full wind speed prediction with 55 layers is greater than that with 40 layers. The scores of the 24 h precipitation prediction show that the TS and ETS of different precipitation levels with 55 layers are all higher than those with 40 layers. Additionally, the higher the precipitation level is, the better the advantage of the 55 layer model.
4. Updating the vegetation coverage data in the WRF model can accurately reflect the actual conditions in some areas and improve prediction results. The updated land-use data can reduce error in all three model domains.

Evaluation of the year-round model predictions of surface and high-altitude elements shows that model prediction errors of all types are within the reliable ranges. Prediction accuracy decreases with increases in the prediction timeline. In addition, model land-use parameters can greatly impact model predictions, including topography, land cover, and soil moisture. Thus, in a later version of the model, the databases used for terrestrial processes will be updated based on the latest remote sensing data.

Author Contributions: Conceptualization, T.Z. and H.D.; Methodology, Y.L. (Yuanpu Liu), D.Z., C.Z., L.S.; Software, C.G.; Validation, Y.L. (Yuanpu Liu), P.Y.; Formal Analysis, H.D., C.G.; Data Curation, G.Z.; Writing-Original Draft Preparation, T.Z., C.G.; Writing-Review & Editing, Y.L. (Yaohui Li); Funding Acquisition, T.Z.

Funding: This research was partially supported by the Ministry of Science and Technology of the People's Republic of China (2018YFC1506804), the Key Special Scientific Research Fund of Meteorological Public Welfare Profession of China (grant GYHY201506001), the Mountain Flood Project of the China Meteorological Administration (CMA), and State Grid Corporation of China (Hebei) (SGTYHT/16-JS-198). Thanks to the Northwest Regional Numerical Forecasting Innovation Team (GSQXCXTD-2017-02).

Acknowledgments: We especially thank Zhaoxia Pu from University of Utah for her comments on this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kang, E.; Cheng, G.; Dong, Z. *Glacier-Snow Water Resources and Mountain Runoff in the Arid Area of Northwest China*; Science Press: Beijing, China, 2002.
2. Anxiang, D.; Huzhi, B.; Xiaobin, L. New development of arid climate research in northwest China from 2001 to 2005 and main scientific problems. *Arid Meteorol.* **2006**, *24*, 57–62.

3. Barker, D.; Huang, X.-Y.; Liu, Z.; Auligné, T.; Zhang, X.; Rugg, S.; Ajjaji, R.; Bourgeois, A.; Bray, J.; Chen, Y. The weather research and forecasting model's community variational/ensemble data assimilation system: *Wrfda. Bull. Am. Meteorol. Soc.* **2012**, *93*, 831–843. [[CrossRef](#)]
4. Zhang, Q.; Zhao, Y.; Zhang, C.; Li, Y.; Sun, G.; Gao, Q. Issues about hydrological cycle and water resource in arid region of northwest China. *Arid Meteorol.* **2008**, *2*, 001.
5. Chen, W.; Wang, Q. The principal and application of the mesoscale operational numerical forecasting experiment model system: Part 2 model physical process. *Gansu Meteorol.* **1995**, *2*, 6–12.
6. Chen, W.; Liu, X.; Li, R.; Wang, Q.; Shen, B. The principal and application of the mesoscale operational numerical forecasting experiment model system: Part 1 system introduction, equations and numerical treatment of the model. *Gansu Meteorol.* **1995**, *1*, 9–15.
7. Zhang, T.; Wang, S.; Wang, X.; Cheng, P.; He, X. The operational system of Lanzhou limited area mesoscale numerical prediction model and its application. *Arid Meteorol.* **2005**, *3*, 79–84.
8. Duan, H.-X.; Zhao, J.-H.; Li, Y.-H.; Huo, W. A verification on the prediction efficiency of grapes-sdm dust-storm model used in recent years. *J. Desert Res.* **2013**, *1*, 29.
9. Li, Y.; Zhao, J.; Xue, J.; Chen, D.; Shen, X.; Wang, H.; Chen, Y. Study on sand-dust numerical forecasting model coupled with grapes and its application in northwest China. *Adv. Earth Sci.* **2005**, *20*, 999–1011.
10. Zhang, Q.; Zhang, C.; Bai, H.; Li, L.; Sun, L.; Liu, D.; Wang, J.; Zhao, H. New development of climate change in northwest China and its impact on arid environment. *J. Arid Meteorol.* **2010**, *1*, 001.
11. Lin, M.-L.; Chen, C.-W.; Wang, Q.-B.; Cao, Y.; Shih, J.-Y.; Lee, Y.-T.; Chen, C.-Y.; Wang, S. Fuzzy model-based assessment and monitoring of desertification using modis satellite imagery. *Eng. Comput.* **2009**, *26*, 745–760. [[CrossRef](#)]
12. Chen, J.; Xue, D. An overview on recent progresses of the operational numerical weather prediction models. *Acta Meteorol. Sin.* **2004**, *5*, 009.
13. Done, J.; Davis, C.A.; Weisman, M. The next generation of nwp: Explicit forecasts of convection using the weather research and forecasting (WRF) model. *Atmos. Sci. Lett.* **2004**, *5*, 110–117. [[CrossRef](#)]
14. Yan, H.; Qian, Y.; Lin, G.; Leung, L.; Yang, B.; Fu, Q. Parametric sensitivity and calibration for the kain-fritsch convective parameterization scheme in the WRF model. *Clim. Res.* **2014**, *59*, 135–147. [[CrossRef](#)]
15. Chou, S.-H. An example of vertical resolution impact on WRF-VAR analysis. *Electron. J. Oper. Meteorol* **2011**, *12*, 1–20.
16. Wang, X.P.; Li, X.R.; Xiao, H.L.; Berndtsson, R.; Pan, Y.X. Effects of surface characteristics on infiltration patterns in an arid shrub desert. *Hydrol. Process. Int. J.* **2007**, *21*, 72–79. [[CrossRef](#)]
17. Kumar, A.; Chen, F.; Barlage, M.; Ek, M.B.; Niyogi, D. Assessing impacts of integrating modis vegetation data in the weather research and forecasting (WRF) model coupled to two different canopy-resistance approaches. *J. Appl. Meteorol. Climatol.* **2014**, *53*, 1362–1380. [[CrossRef](#)]
18. Wen, X.; Lu, S.; 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. Hydrometeorol.* **2012**, *13*, 573–587. [[CrossRef](#)]
19. Yang, J.; Duan, K.; Wu, J.; Qin, X.; Shi, P.; Liu, H.; Xie, X.; Zhang, X.; Sun, J. Effect of data assimilation using WRF-3dvar for heavy rain prediction on the northeastern edge of the tibetan plateau. *Adv. Meteorol.* **2015**. [[CrossRef](#)]
20. Siuta, D.; West, G.; Stull, R. WRF hub-height wind forecast sensitivity to pbl scheme, grid length, and initial condition choice in complex terrain. *Weather Forecast* **2017**, *32*, 493–509. [[CrossRef](#)]
21. Thompson, G.; Tewari, M.; Ikeda, K.; Tessendorf, S.; Weeks, C.; Otkin, J.; Kong, F. Explicitly-coupled cloud physics and radiation parameterizations and subsequent evaluation in WRF high-resolution convective forecasts. *Atmos. Res.* **2016**, *168*, 92–104. [[CrossRef](#)]
22. Yang, B.; Zhang, Y.; Qian, Y.; Huang, A.; Yan, H. Calibration of a convective parameterization scheme in the WRF model and its impact on the simulation of east asian summer monsoon precipitation. *Clim. Dyn.* **2015**, *44*, 1661–1684. [[CrossRef](#)]
23. Ran, Y.-h.; Li, X.; Lu, L. Accuracy evaluation of the four remote sensing based land cover products over China. *J. Glaciol. Geocryol.* **2009**, *3*, 011.
24. Parrish, D.F.; Derber, J.C. The national meteorological center's spectral statistical-interpolation analysis system. *Mon. Weather Rev.* **1992**, *120*, 1747–1763. [[CrossRef](#)]

