SCIENCE CHINA
Earth Sciences

• RESEARCH PAPER •

doi: 10.1007/s11430-014-4914-4

Numerical simulation and data assimilation of the water-energy cycle over semiarid northeastern China

WEN XiaoHang[1*], LIAO XiaoHan[2], YUAN WenPing[1], YAN XiaoDong[1], WEI ZhiGang[1], LIU HuiZhi[3], FENG JinMing[3], LU ShiHua[4] & DONG WenJie[1]

[1] *State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China;*
[2] *National Remote Sensing Center of China, Beijing 100036, China;*
[3] *Key Laboratory of Regional Climate-Environment Research for Temperate East Asia, Chinese Academy of Sciences, Beijing 100029, China;*
[4] *Key Laboratory of Land Surface Process and Climate Change in Cold and Arid Regions, Cold and Arid Regions Environment and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou 730000, China*

Received November 14, 2013; accepted January 24, 2014

The default fractional vegetation cover and terrain height were replaced by the estimated fractional vegetation cover, which was calculated by the Normalized Difference Vegetation Index (NDVI) of Earth Observing System Moderate-Resolution Imaging Spectroradiometer (EOS-MODIS) and the Digital Elevation Model of the Shuttle Radar Topography Mission (SRTM) system. The near-surface meteorological elements over northeastern China were assimilated into the three-dimensional variational data assimilation system (3DVar) module in the Weather Research and Forecasting (WRF) model. The structure and daily variations of air temperature, humidity, wind and energy fields over northeastern China were simulated using the WRF model. Four groups of numerical experiments were performed, and the simulation results were analyzed of latent heat flux, sensible heat flux, and their relationships with changes in the surface energy flux due to soil moisture and precipitation over different surfaces. The simulations were compared with observations of the stations Tongyu, Naiman, Jinzhou, and Miyun from June to August, 2009. The results showed that the WRF model achieves high-quality simulations of the diurnal characteristics of the surface layer temperature, wind direction, net radiation, sensible heat flux, and latent heat flux over semiarid northeastern China in the summer. The simulated near-surface temperature, relative humidity, and wind speed were improved in the data assimilation case (Case 2) compared with control case (Case 1). The simulated sensible heat fluxes and surface heat fluxes were improved by the land surface parameterization case (Case 3) and the combined case (Case 4). The simulated temporal variations in soil moisture over the northeastern arid areas agree well with observations in Case 4, but the simulated precipitation should be improved in the WRF model. This study could improve the land surface parameters by utilizing remote sensing data and could further improve atmospheric elements with a data assimilation system. This work provides an effective attempt at combining multi-source data with different spatial and temporal scales into numerical simulations. The assimilation datasets generated by this work can be applied to research on climate change and environmental monitoring of arid lands, as well as research on the formation and stability of climate over semiarid areas.

WRF model, data assimilation, water-energy cycle, semiarid northeastern China

Citation: Wen X H, Liao X H, Yuan W P, et al. 2014. Numerical simulation and data assimilation of the water-energy cycle over semiarid northeastern China. Science China: Earth Sciences, doi: 10.1007/s11430-014-4914-4

*Corresponding author (email: wenzerg@126.com)

The impacts of human activities and global climate changes over semiarid northern China are important in the research of global change and regional climate feedbacks (Tu et al., 2006). Research has shown that the aridification of northern China may be due largely to human activities, which may have accelerated aridification (Fu et al., 2002). Living conditions have been seriously threatened by lack of water resources, eco-environmental degradation, and desertification during the last 30 years in northern China (Fu et al., 2008). Therefore, it is important to study the drying trends and their causes in northern China. Many field experiments have been done over northern China to study land-atmosphere interactions during the entire year, rather than only observing turbulence flux during the Intensive Observation Period in the summer (Cheng et al., 2010; Huang et al., 2008).

Numerical modeling has become an important tool and integrates land surface observation experiments, remote sensing technology, and data assimilation methods to study land-atmosphere interactions over complex surfaces (Chu et al., 2005; Li et al., 2007; Liu et al., 2004; Liu et al., 2006; Wen et al., 2012). Numerical modeling, such as the Advanced Regional Prediction System (ARPS) (Xue et al., 2000), Pennsylvania State University's fifth-generation National Center for Atmospheric Research Mesoscale Model (MM5) (Grell et al., 1994) and the Weather Research and Forecasting (WRF) model (Janjic et al., 1998), can already be used to simulate local-scale land surface processes and large eddy simulations with high spatial resolution (Meng et al., 2009). Many researchers have used the data assimilation method to assimilate land surface observations, satellite data, and Doppler radar data to improve the simulations with respect to soil moisture, soil temperature, and surface energy flux (Li et al., 2007; Pan et al., 2012b; Wen et al., 2012). The data assimilation method is used to integrate observations with numerical simulations. The observations can be both direct and indirect data from different sources with variable temporal and spatial resolutions. The output from assimilation datasets includes land surface variables with spatiotemporal and physical consistency (Li et al., 2007). The Three or Four Dimensional Variational Assimilation (3/4 D-Var), Optimal Interpolation (OI) method and Ensemble Kalman Filter (EnKF) algorithm are used in the data assimilation program, and great improvements have been made in assimilation theory research and the assimilation system in recent years (Barrett et al., 2009; Chen et al., 2007; Irmak et al., 2009; Kumar et al., 2009; Li et al., 2010; Yang et al., 2009). The Doppler radar radial velocity data and reflectance data were assimilated in the WRF model using the Proper Orthogonal Decomposition (POD)-based Ensemble Three-Dimensional Variational data assimilation method (PODEn-3DVar) to improve the modeling of precipitation (Pan et al., 2012b). The Land Data Assimilation System (LDAS) of west China (Li et al., 2007) uses the Common Land Model (CoLM) coupled with a microwave radiation transfer model, which includes the soil freeze-thaw process, snow cover, and different land-surface characteristics. Microwave observations, such as the Special Sensor Microwave/Image (SSM/I), and the brightness temperature data from the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) were assimilated into the LDAS systems. Finally, the high temporal-spatial resolution data, such as soil moisture, soil temperature, snow, frozen soil, sensible heat flux, latent heat flux, and evapotranspiration data, over northwestern China were generated in these systems. The Jinta high-resolution data assimilation reanalysis dataset was produced by WRF-3DVar system, which includes pressure, wind, temperature, and humidity at 19 levels, soil temperature and moisture at 4 levels, and radiation and heat fluxes at ground level from June to August 2008 with 1-km horizontal resolution over the Heihe River Basin near the Jinta Oasis (Wen, 2011). This reanalysis dataset could be used in further research on the long-period climate effects and characteristics of the atmospheric boundary layer over the Oasis-Gobi system.

In this study, the Green Vegetation Fraction (GVF) data, which are derived by satellite remote sensing from the Earth Observing System (EOS) MODIS Normalized Differential Vegetation Index (NDVI) (http://ladsweb.nascom.nasa.gov/data/search.html) data from June to August 2009, and the Digital Elevation Model (DEM) data from the Shuttle Radar Topography Mission (SRTM) (http://datamirror.csdb.cn), are used to replace the default data in the WRF model. These factors improve the simulation of the land-atmosphere interactions and water-energy cycling processes over semiarid northeastern China. Meanwhile, the near-surface wind speed, wind direction, pressure, air temperature, relative humidity, and dew point temperature data, which were obtained from 47 automatic weather stations, were also assimilated into the WRF model. The heat flux data from Tongyu, Naiman, Jingzhou, and Miyun stations are selected for comparison with the simulations. The purpose of this work is to provide a new way to better understand the land surface processes of the northeastern semiarid region of China, thereby promoting a sustainable development of agriculture and forestry.

1 Brief description of data

1.1 Description of the WRF model and the numerical experiment

The WRF model is a next-generation mesoscale forecast model and data assimilation system that will advance the understanding and prediction of mesoscale weather and accelerate the transformation of research advances into practical applications. It was developed collaboratively by the National Center for Atmospheric Research (NCAR) and the National Centers for Environmental Prediction (NCEP) (Janjic et al., 1998). The nonhydrostatic WRF version 3.4.1 modeling system, released in August 2012, is used in this

study.

A double-nested grid system centered at 42.78°N, 120.41°E is used in this study. The model configurations of the nested domains are listed in Table 1. The simulation time began at 08:00 (Beijing Time) June 1 and ended at 23:00 (Beijing Time) August 31, 2009. The 6-h, 1°×1° NCEP Final Analysis (FNL) from the Global Forecast System (GFS) provided initial and lateral boundary conditions for the WRF model. In the vertical dimension, 27 unevenly spaced full Eta levels were defined. Based on a series of tests, a set of physical options within the WRF model were selected. The planetary boundary layer processes were resolved with the Mellor-Yamada-Janjic scheme (Janjic, 1990; 1996; 2002). The microphysics were described with Lin's scheme (Lin et al., 1983), and the cumulus clouds were simulated with the Kain-Fritsch convection scheme (Kain et al., 1993) (only for domain 1 and no convection scheme for domain 2). The Rapid Radiative Transfer Model (RRTM) (Mlawer et al., 1997) and the Dudhia shortwave radiation scheme (Dudhia, 1989) were used to calculate longwave and shortwave radiation and their transfer within the atmosphere. The Noah land surface model with four soil layers was used to resolve land surface processes within the WRF model (Chen et al., 2001a; 2001b). The schematic diagram of the simulated nested domains and distribution of observation sites is shown in Figure 1.

To discover the effects of changes in the underlying surface conditions and assimilation effect of the water-energy cycle over this area, four sets of simulations were performed:

Case 1: Control experiment. A control experiment was performed in which all parameters and land use categories were the default values in the WRF model. The Land use/Vegetation type data was derived from the MODIS data in the simulated area, and the initial fields was obtained from the NCEP-FNL 1°×1° gridded analysis data.

Case 2: Data assimilation experiment. The near-surface wind speed (m s^{-1}), wind direction (°), pressure (hPa), 2-m air temperature (K), relative humidity (%), and dew-point temperature (K) at 14:00 local time were obtained every day by 47 local automatic weather stations and were assimilated into the WRF model. With the WRF model running 24 h a day, the newly observed data were continually assimilated and the boundary conditions were updated. The purpose of this experiment is to provide global optimum values for surface atmospheric conditions, and the simulated results were adjusted by observations. The assimilation process was as follows:

According to three-dimensional variational assimilation theory, which minimizes the following cost function (Courtier, 1997; Lorenc, 1986):

$$J = \frac{1}{2}\left[H(\boldsymbol{x}) - \boldsymbol{y}^0\right]^T \boldsymbol{R}^{(-1)}\left[H(\boldsymbol{x}) - \boldsymbol{y}^0\right] + \left(\boldsymbol{x} - \boldsymbol{x}^b\right)^T B^{(-1)}\left(\boldsymbol{x} - \boldsymbol{x}^b\right), \quad (1)$$

where $\boldsymbol{x}$ is the analysis vector (on model grid points), $\boldsymbol{x}^b$ is the background vector (on model grid points), $\boldsymbol{y}^0$ is the observation vector (on observation sites), and $\boldsymbol{H}$ is the (nonlinear) observation operator that converts model variables to observations. $\boldsymbol{B}$ is the background error covariance matrix, which decides how the observation information can be distributed in the analysis domain. $\boldsymbol{R}$ is the observation error covariance matrix that includes the effects of both instruments and representativeness, $(\cdot)^T$ indicates transposition and $(\cdot)^{-1}$ indicates inversion. Further details about the components and real-time applications of the 3DVAR system have been reported in Barker et al. (2004) and Gu et al. (2005). The schematic diagram of cycle assimilation is shown in Figure 2.

Case 3: Land surface parameterization sensitivity experiment. The default vegetation coverage was recalculated and replaced by 1-km resolution MODIS NDVI data from June to August 2009. The default terrain height data were replaced by DEM data, which were produced by the SRTM system and resampled using the fine domain. The pixel aggregate method is used to resize and scale up the DEM data to match the corresponding grid. The new terrain height values are more accurate than the default (Figure 3). The GVF data in the default model are derived from the monthly global NDVI data, which use the dimidiate pixel method by Gutman and Ignatov (1998), for the period from 1985 to 1990, but the spatial resolution of default GVF in the fine domain is lower than the MODIS NDVI data. The new GVF data are calculated by the Terra-MODIS vegetation indices products (MOD13A3 Level 3 Monthly Vegetation Indices) and interpolated to the corresponding model grid using the pixel aggregate method. We used the 25-km spatial resolution GVF data for domain 1 and the 5-km spatial resolution GVF data for domain 2 (Figure 4). The purpose of this experiment is to provide more accurate land surface conditions in the WRF model and improve energy transfer processes between the land and the atmosphere.

Because the near-surface atmospheric elements, such as 2-m temperature and pressure, are affected by terrain height,

Table 1 Model configuration of nesting structure of two grids

Grid ID	Center lon-lat	Domain dimension	Model resolution (km)	Time step (s)	Data resolution
1	103°E, 38°N	230×160	25	120	10 min
2	120.41°E, 42.78°N	136×136	5	24	30 s

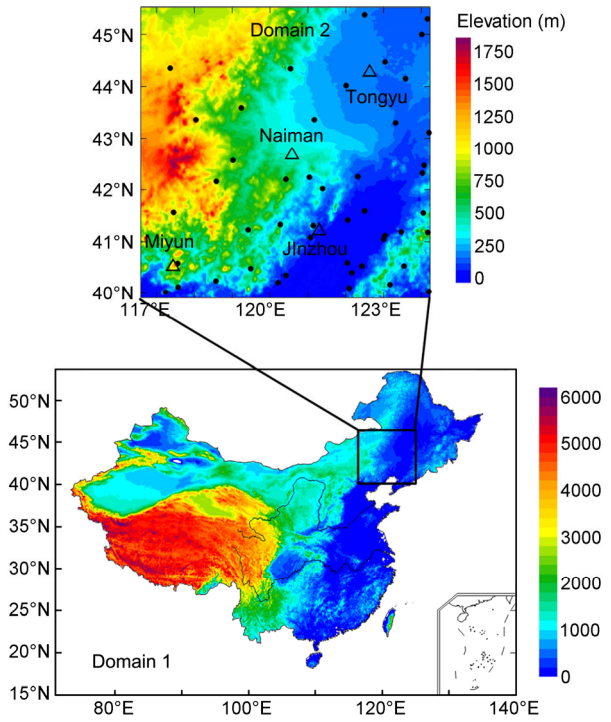

Figure 1 The schematic diagram of simulated nested domain. △, Flux stations; ● automatic weather stations.

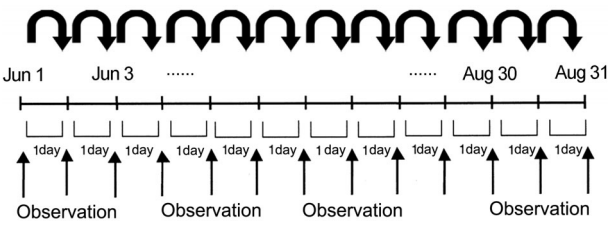

Figure 2 The schematic diagram of assimilate observation data.

it is necessary to use the more accurate terrain height information in the WRF model. Liu et al. (2007) and Pan et al. (2012a) used the mesoscale atmospheric model to research the local climate over the Heihe area in northwestern China. They showed that the simulation is improved by using more accurate land surface conditions, including GVF data and terrain height.

The GVF is obtained using the relationship by Gutman and Ignatov (1998):

$$\mathrm{GVF} = \frac{\mathrm{NDVI} - \mathrm{NDVI}_{\min}}{\mathrm{NDVI}_{\max} - \mathrm{NDVI}_{\min}}, \quad (2)$$

where GVF is Green Vegetation Fraction (%), $\mathrm{NDVI}_{\min}$ and $\mathrm{NDVI}_{\max}$ are the signals from bare soil (Leaf Area Index (LAI)→0) and dense green vegetation (LAI→∞), respectively, which are the minimum and maximum values over the research area.

The default GVF is clearly different from the new GVF, which is derived from MODIS NDVI data in domain 2 from June to August 2009. The new GVF not only is more accurate than the default value, but also can reflect variation trends over a simulated area. The GVF was lower than default value by 10%–20% over the northwestern area of the Inner Mongolian Plateau and higher by 10%–20% in the central and northeastern areas (Liu et al., 2012). These changes indicate that the default GVF uses outdated monthly surface vegetation coverage from over 20 years ago. However, the GVF has changed rapidly in recent years, and these changes may reflect the effects of human activity in northeastern China. The new research has also indicated that the LAI has decreased by approximately 1 m^2 m^{-2} compared with the averaged value of the past 30 years across the eastern area of Inner Mongolian Plateau (http://www.csi.gov.cn/).

Case 4: Integration experiment. Case 2 and Case 3 are integrated in this experiment. The default terrain height and GVF has been replaced by the new data in the initial fields and the observations from auto weather stations have been assimilated every day. The purpose of this experiment is to provide more accurate land surface conditions and global optimum values of atmospheric fields, and the simulation

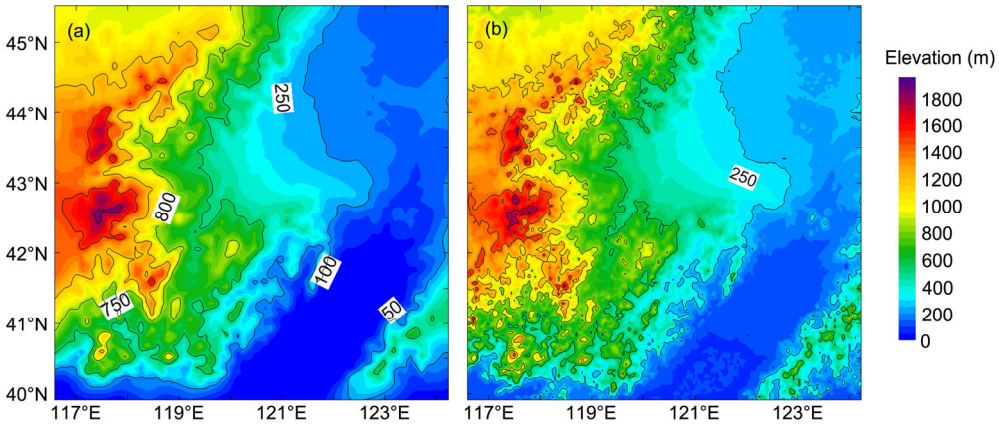

Figure 3 The elevation maps for the default data Case 1 (a) and the SRTM system data Case 3 and Case 4 (b).

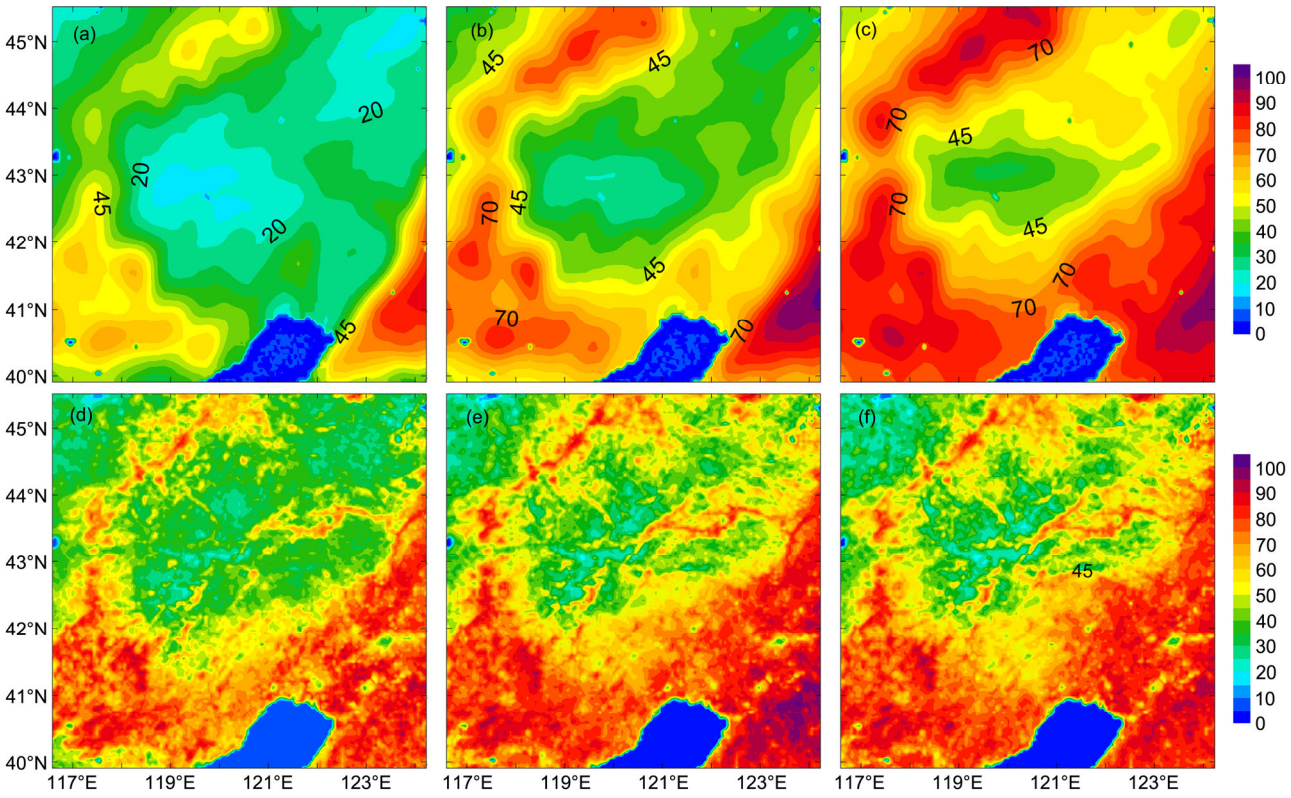

Figure 4 Green Vegetation Fraction (%) maps for the domain 2 from the default data. Interpolated from AVHRR data (a)–(c), and MODIS data (d)–(f). (a), (d) for June; (b), (e) for July; (c), (f) for August 2009.

results could be improved conspicuously.

1.2 Description of automatic weather station data and flux data

The 6-h near-surface meteorological data of 47 automatic weather stations were collected from the National Meteorological Information Center (NMIC) of the China Meteorological Administration (CMA). The data include the near-surface temperature, pressure, relative humidity, wind speed, and wind direction data from June to August 2009, and the locations of these stations are shown in Figure 1. These data are subjected to quality control and inspection by the NMIC prior to use.

Currently, flux data have been collected at the Tongyu, Naiman, Jinzhou, and Miyun flux stations from June to August 2009 as a part of the "Northern China Observation Cooperation and Interaction Research in Semiarid and Arid Regions Project". The observation data include wind speed and direction at 10 m, air temperature and humidity at 2 m, surface pressure, net radiation, sensible heat flux, latent heat flux, and soil moisture, temperature and soil heat flux at 0.05 m. The instruments of these stations are similar and set at a uniform height. These instruments include a CO_2/H_2O open-path gas analyzer (Li-Cor, Li-7500), a three-dimensional sonic anemometer (Campbell, CSAT3), a temperature and relative humidity probe (Campbell, HMP45C), a solar radiation instrument (Kipp & Zonen, CNR-1 or Eppley, PSP and PIR), a rain gauge (Campbell, TE525 or RMYoung 52202/52203), a soil heat flux plate (REBS, HFT-3, HFP01 probe), and a water content reflectometer (Campbell, CS616). The sensors are set horizontally. The terrains surrounding the stations are open and level, within a radius of several hundred meters. The instruments are advanced and precise. All instruments are under the same calibration and maintenance. The observation data are subjected to uniform quality control: the flux data are monitored via remote control and the regulated check method. The EdiRe tool is used to remove the outliers, rotate coordinates, and calculate turbulent flow stabilities and lag correlations of CO_2 and H_2O, which allows the final, more accurate flux data to be generated (Zeng et al., 2012a).

Parallel comparison tests for some instruments indicate that the differences between stations in sensible heat data, latent heat data, and momentum fluxes measured by the eddy correlation method do not exceed 8 W m^{-2}, 8 W m^{-2}, and 0.02 kg m^{-1} s^{-2}, respectively. The differences in soil heat fluxes do not exceed 11 W m^{-2}, and the differences in the surface net radiation data and the radiation components are smaller than 10 W m^{-2}. These instruments have high consistency and their comparative errors are small, which meet the technical requirements (Zhang et al., 2012). The geographic information on the four flux stations is shown in Table 2.

Table 2 Information about four flux stations over northeast semiarid region [a)]

Number	Stations	Latitude & longitude	Elevation (m)	Area (km^2)	Observation height (m)	Underlying surface
1	Tongyu	44.56°N, 122.88°E	184	>0.5	3	Cropland/grassland
2	Naiman	42.92°N, 120.70°E	361	0.003	2	Desert/oasis
3	Jinzhou	41.14°N, 121.20°E	22	0.5	3.3	Cropland
4	Miyun	40.63°N, 117.32°E	350	>1.2	26.7	Cropland

a) The observation heigh indicate the height of eddy correlation system, and the area include observation site-centred, flat and open terrain area (Zeng et al., 2012a).

2 Model results

2.1 Near surface temperature and relative humidity

The observations and simulations are collected and compared from Tongyu, Naiman, Jinzhou, and Miyun flux stations from June 20 to 30, 2009. Figure 5 is the time series comparison of 2-m temperature and relative humidity. Clearly, the characteristics of the diurnal variations in the temperature and the relative humidity can be simulated well in all cases. The simulated temperature values are higher than the observations at Tongyu and Naiman from June 28 to 30, and lower than the observations at Jinzhou in Case 1 and Case 3 from June 20 to 22. The simulated temperatures in Case 2 and Case 4 are closer to the observations. The simulations of the relative humidity are lower than observations in all cases at night, but the simulated values in Case 2 and Case 4 are higher than those in Case 1 and Case 3 at night and are closer to the observations. The averaged temperature is 21.3°C at Tongyu from June to August, and the simulated average temperature of the simulations is 22.9°C, 22.8°C, 22.8°C, and 22.7°C in each of the four cases, respectively. The simulated average temperatures in Case 2 and Case 4 are also close to the observations of the other flux stations (table omitted).

The root mean square error (RMSE) represents the systematic and non-systematic (random) error and consists of the general error of the model. The bias error is the mean variation between the observations and simulations, which represents the systematic error portion (Case et al., 2008). This error is usually related to the physical processes, parameterization schemes, and numerical calculations in the model. The error statistics (Table 3) show that the RMSE is

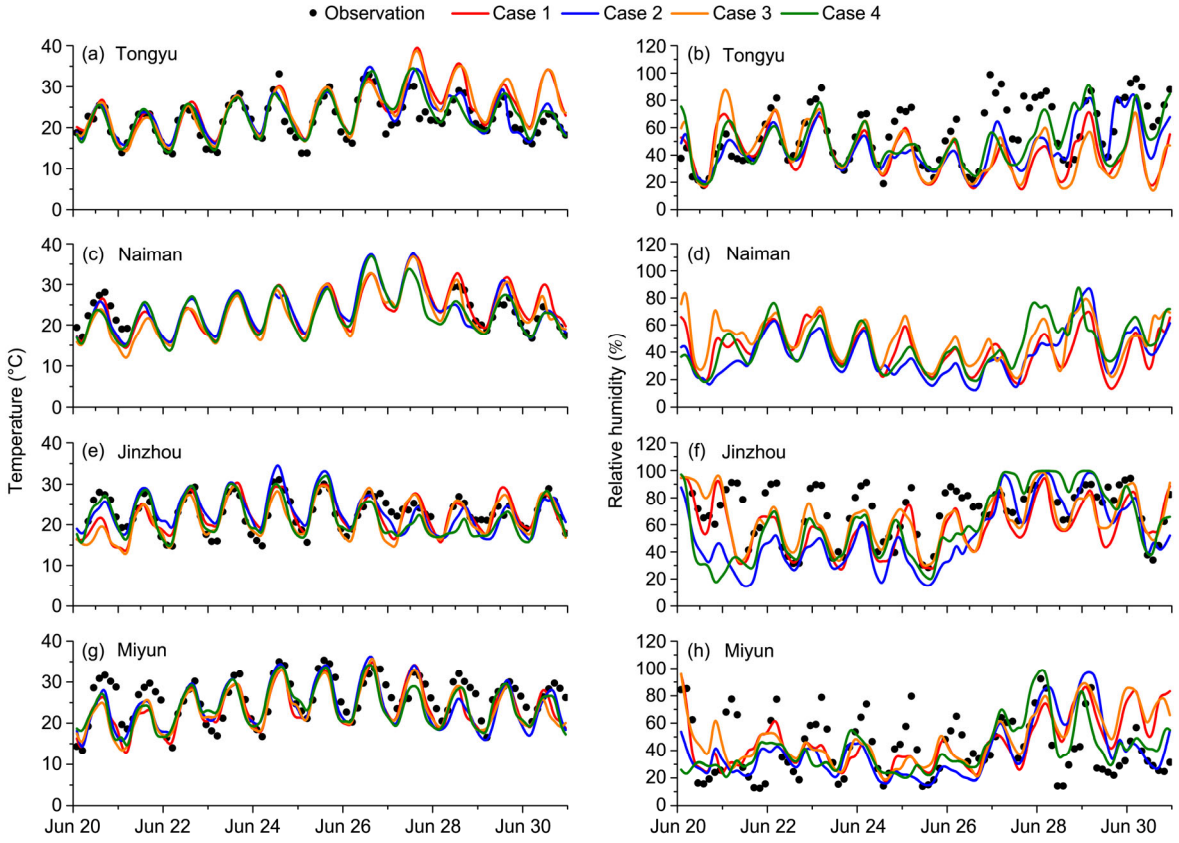

Figure 5 Comparison of observations and simulations of near surface temperature (°C) and relative humidity (%) at Tongyu, Naiman, Jinzhou, and Miyun.

Table 3 RMSE, bias and correlation coefficient (R) of air temperature, relative humidity, and wind speed in all cases[a)]

		Temperature (°C)				Relative humidity (%)				Wind speed (m s^{-1})			
		Tongyu	Naiman	Jinzhou	Miyun	Tongyu	Naiman	Jinzhou	Miyun	Tongyu	Naiman	Jinzhou	Miyun
RMSE	Case 1	3.46	3.14	2.90	3.75	21.56	–	20.32	20.53	2.36	3.12	3.03	2.58
	Case 2	2.83	2.77	2.94	3.77	19.18	–	21.94	21.78	2.08	2.81	2.99	2.52
	Case 3	3.27	2.95	3.36	3.79	19.85		19.58	21.72	2.30	3.08	2.97	2.52
	Case 4	2.96	2.69	3.19	3.72	18.12		19.84	20.09	2.23	2.73	2.72	2.38
Bias	Case 1	1.06	–0.12	0.32	–1.83	–9.51	–	–11.32	–4.96	–0.20	1.63	1.92	1.62
	Case 2	1.08	0.15	0.73	–1.17	–9.04	–	–11.19	–5.02	–0.25	1.36	1.85	1.79
	Case 3	1.16	–0.52	–0.28	–1.26	–10.0		–7.26	–5.71	–0.33	1.52	1.57	1.63
	Case 4	0.65	0.00	0.53	–0.97	–6.69		–8.92	–4.64	–0.13	1.13	1.31	1.54
R	Case 1	0.81	0.83	0.81	0.73	0.55	–	0.64	0.62	0.52	0.07	0.14	0.14
	Case 2	0.87	0.86	0.81	0.70	0.65	–	0.59	0.57	0.58	0.21	0.21	0.31
	Case 3	0.84	0.85	0.76	0.70	0.64		0.57	0.56	0.44	0.07	0.05	0.20
	Case 4	0.85	0.88	0.81	0.70	0.66		0.60	0.62	0.53	0.15	0.13	0.27

a) The gray values indicate the lowest values of RMSE and bias, and the highest values of R in all cases.

lower at Jinzhou in Case 1, and lower at Tongyu in Case 2, and has the lowest values at Miyun and Naiman in Case 4. Although the bias of Case 3 is lower than that of other cases at Jinzhou, the bias errors of the other stations are at the lowest values in Case 4, which shows that the deviation value between the observations and simulations is at a minimum in Case 4.

Though the default GVFs at Tongyu station grid are 16.5%, 30.3% and 49.7% from June to August, the GVFs calculated by MODIS-ENVI data are 27.6%, 40.8%, and 47.1% in Case 3 and Case 4. The GVF is higher than the default value in June and July, and therefore the simulated value of relative humidity is higher than that in Case 1 and is closer to observations in June and July (figure omitted). The observed values of relative humidity are almost 100% at night from June to July, (the maximum value of relative humidity in the near-surface) due to evaporation from vegetation. The simulated values of relative humidity are much closer to the observations in Case 3 and Case 4. The error statistics and correlation coefficients between the simulations and observations from the four flux stations are listed in Table 3. There is a great improvement in the near-surface relative humidity simulations in Case 4, and the RMSE and bias errors are the lowest and the correlation coefficients (R) the highest at Tongyu and Miyun.

Comparisons between observations and simulations of near-surface wind speed and direction of the four flux stations from June 20 to 30, 2009 are shown in Figure 6. Because of uncertainty and randomness in atmospheric turbulence, the simulated wind speed value is currently not very good in all mesoscale models (Hanna et al., 2001; Wen et al., 2012). However, the trends of the wind direction can be simulated well, especially the strong background wind conditions. The simulated values of wind speed are larger than the observations in all cases, except at Tongyu. The correlation coefficients of simulated and observed wind speeds are between 0.44 and 0.58 at Tongyu, and the bias error is negative, which indicates that the simulated values are lower than the observed values. However, the correlation coefficients are between 0.05 and 0.31 at the other stations, and the bias errors are between 1.36 and 1.92. The model may underestimate the surface roughness length of the underlying surface, which may therefore lead to higher wind speeds in the simulations.

The error statistics are listed in Table 3. The RMSE of the wind speed simulations in Case 4 is the lowest of all the cases, with the exception of Tongyu, but the bias error is the lowest of all, including Tongyu. These errors indicate that there is a great improvement in the simulation of surface wind speeds in Case 4. The correlation coefficient in Case 2 is the highest of all the cases, which indicates that the variation trends are closer between observations and simulations in the data assimilation case (Case 2). The bias error of the wind speed simulation in Case 2 is lower than Dong's result (2011) from the summer of 2006, and the correlation coefficient increases from 0.26 to 0.58.

In general, due to the near-surface meteorological elements assimilated into the WRF model and the higher GVF in the flux site grids, the simulated peak value of air temperature is lower and the peak value of relative humidity is higher at night. Therefore, the simulation results in Case 2 and Case 4 are better than those in Case 1. Observations and simulations of near-surface temperature are in good agreement in Case 4. The simulated relative humidities in Case 3 and Case 4 are better than those in Case 1. Because the satellite data are integrated into the initial field of the WRF model, the GVF of the underlying surface is closer to the true situation, and there is a significant improvement in the

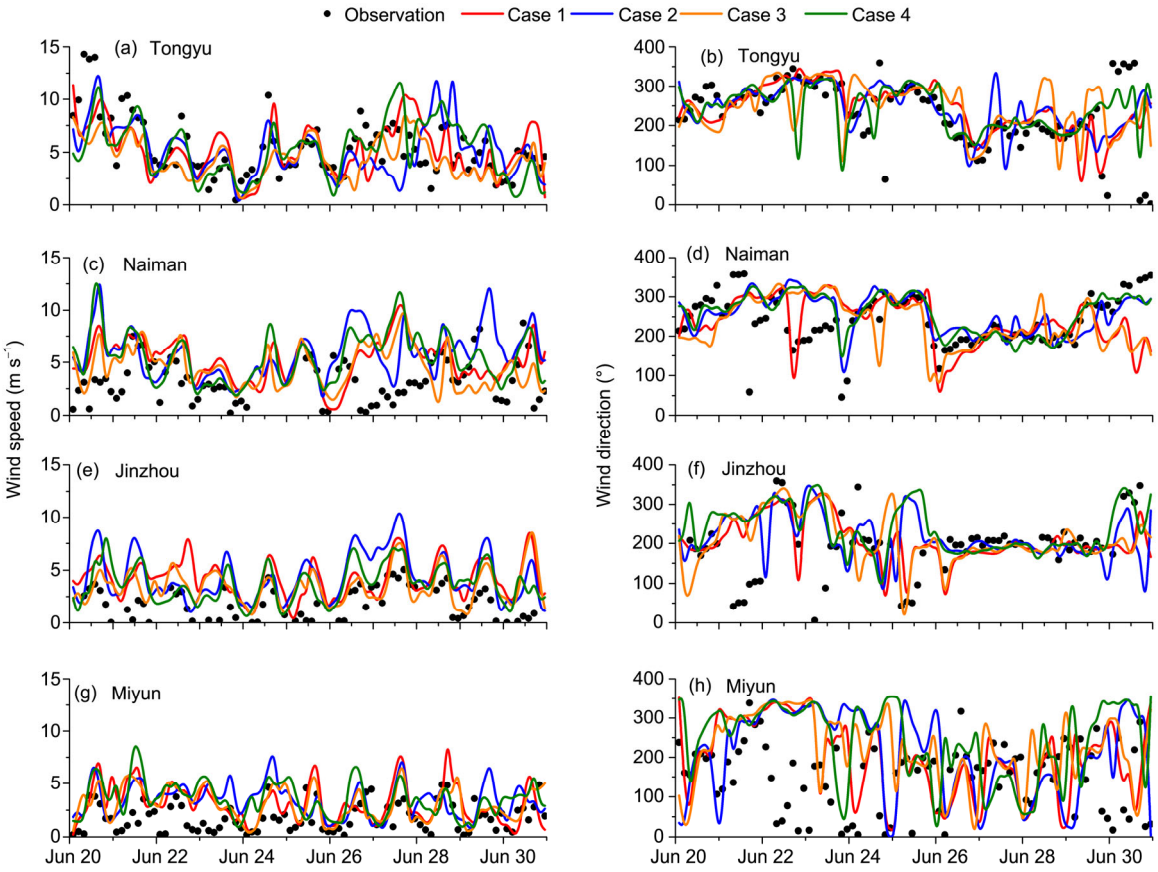

Figure 6 Comparison of observations and simulations of near surface wind speed (m s^{-1}) and wind direction (°) at Tongyu, Naiman, Jinzhou, and Miyun.

simulations of relative humidity, which agree much better with the observations. The simulated relative humidity is closer to the peak value at night than that in Case 1. However, the current simulations of wind speed in mesoscale model are generally too large, which may be due to the lower roughness length value of the land surface model. Despite this, wind direction trends and wind speeds can be simulated well (Figure 6).

2.2 Surface-energy fluxes

The radiative transfer and energy exchange processes between the near-surface soil-vegetation-atmosphere impact the climate, environment, and atmospheric circulation. Energy in the soil is received from solar shortwave radiation and atmospheric longwave radiation. At the same time, shortwave radiation and longwave radiation are reflected by the ground and atmosphere. Therefore, the surface energy balance is affected by the simulated radiation flux and affects the distribution of the sensible and latent heat fluxes. Because the ground surface heat flux is very difficult to measure directly, the ground heat flux is calculated by using the soil heat flux data, soil temperature, and soil volumetric water content at 5 cm. The equation is as follows (Tanaka et al., 2001; Yang et al., 2008):

$$G_{\mathrm{sfc}} = G_z + \int_0^z C_{\mathrm{s}} \frac{\partial T}{\partial t} \mathrm{d}z, \tag{3}$$

$$C_{\mathrm{s}} = (1-\eta_{\mathrm{sat}})\rho_{\mathrm{d}} C_{\mathrm{d}} + \eta_w \rho_w C_w, \tag{4}$$

where G_{sfc} is the soil heat flux at 0 cm, G_z is the soil heat flux at 5 cm, z is 5 cm depth, C_{s} is the soil volumetric heat capacity, T is the averaged temperature between 0 to 5 cm, and $(1-\eta_{\mathrm{sat}})\rho_{\mathrm{d}} C_{\mathrm{d}}$ represents the volumetric heat capacity of dry soil. Due to the lack of the soil texture classification data, soil textures are classified according to United States Department of Agriculture (USDA), and the soil types used in the model grid are CLAY LOAM, SANDY LOAM, CLAY LOAM, and LOAM at Tongyu, Naiman, Jinzhou, and Miyun stations, respectively. The volumetric heat capacity value is 1.23, 1.34, 1.23, and 1.21 MJ m^{-3} K^{-1}, respectively (Tanaka et al., 2001). The variables η_{w}, ρ_{w}, and C_{w} represent soil volumetric water content, liquid water density (1.00×10^3 kg m^{-3}), and the specific heat capacity of liquid water (4.18×10^3 J kg^{-1} K^{-1}), respectively.

Figure 7 shows the comparison between observations and simulations of net radiation, latent heat flux, sensible heat flux, and ground heat flux from June 20 to 30, 2009 at the Tongyu station. It can be observed that the net radiation peak value can reach 600 W m^{-2} at Tongyu. This result is

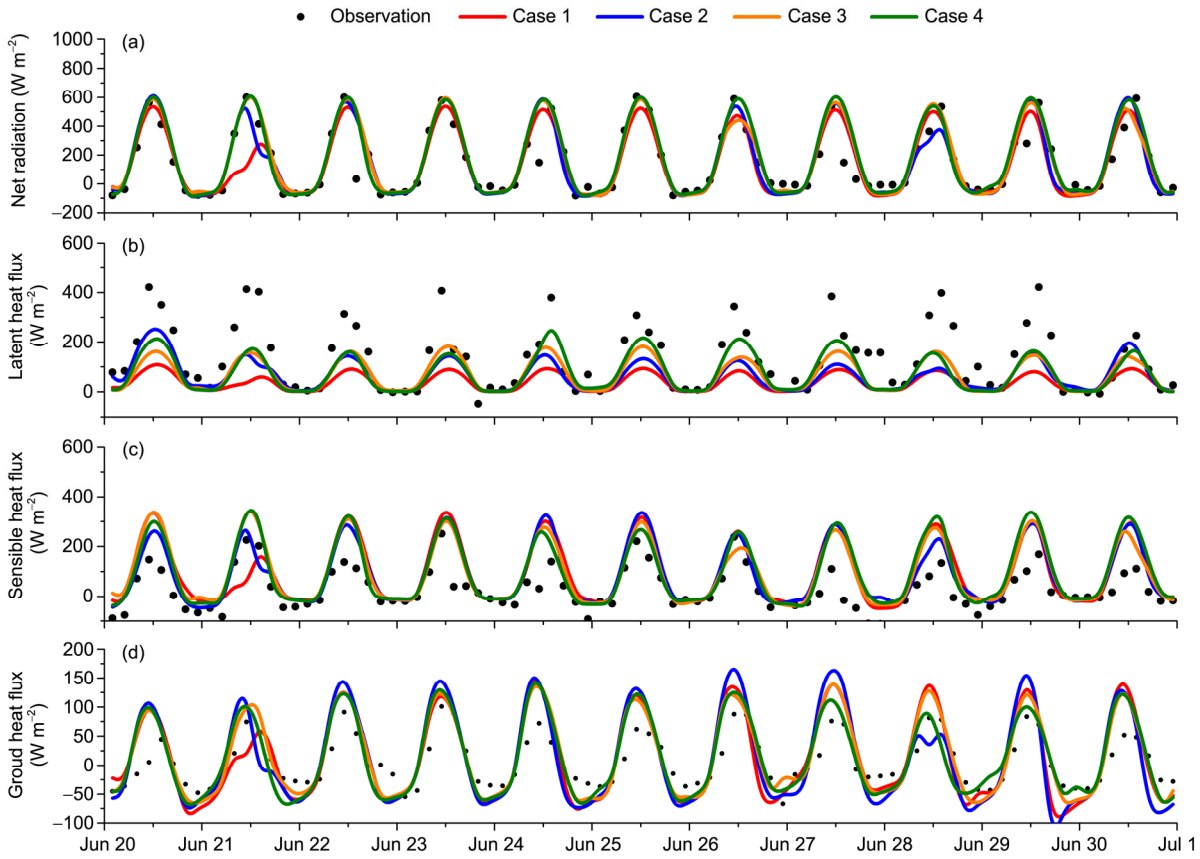

Figure 7 Comparison of observations and simulations of net radiation, latent heat flux, sensible heat flux, and ground heat flux from June to August 2009 at Tongyu.

consistent with Zeng and Zhang (2012b) and Tu et al. (2006), but the peak value is less than that on crop land in the Heihe River Basin (Ma et al., 1997; Meng et al., 2009; Wen et al., 2012; Yan et al., 2001). The diurnal variation characteristics of net radiation can be simulated well in all cases. The simulated latent heat flux is obviously less than the observed values in Case 1 because the GVF at Tongyu station is 16.5% in the default model, but 27.6% in Cases 3 and 4 in June. Thus, the increase in the GVF leads to an increase in latent heat flux. The error statistics (Table 4) clearly show that the simulations of latent heat flux have obviously improved at Tongyu and Naiman, but the Bias error has increased at Jinzhou and Miyun. One possible reason may be that variations in the latent heat flux are closely related to the soil water content over arid regions in summer. The summer precipitation has clearly shown local climate characteristics. Therefore, the bias error appears due to the limitation of the model's horizontal resolution in simulating latent heat flux. Another reason may be that the bias error is caused by the coarse grids in model due to the limited representation of the stations. The simulated sensible flux values are higher than observed values in all cases. The bias error is the lowest at Naiman, and RSMEs and Bias errors in Cases 3 and 4 are less than those in Case 1. Because the GVF at the Naiman and Tongyu stations are less than those at Miyun and Jinzhou, the surface temperatures of the former two are higher than the latter two, which leads to strong sensible heating over the ground surface. The simulated peak value of sensible heat flux at Tongyu is larger than the observed peak from June to July, but smaller in August. The characteristics of the diurnal variation of sensible heat flux have been simulated well in all cases (figure omitted). The ground heat flux can be simulated well in all cases, and the peak values of ground heat flux can reach 100 W m^{-2} during the daytime, when energy is transferred from the atmosphere to the soil and retained in the soil. The minimum values of ground heat flux are approximately –50 to –70 W m^{-2} at night, when energy is transferred from the soil to the atmosphere. Simulated ground heat flux values at Tongyu are higher than the calculated values during the day because lush crops surround the Tongyu station. This vegetation makes the GVF greater than 80% in actuality, but the GVF value in corresponding model grid is lower than the true value. Solar radiation reaches the ground surface during day and a part of this is absorbed by vegetation, which reduces the energy transferred from the surface into the deep soil layer. This, in turn, can lead to simulated ground heat flux values higher than those observed during daytime and lower than observed during nighttime. In addition, the lack of the actual soil type data for the observation sites could

Table 4 The root mean square error (RMSE), bias and correlation coefficient (R) of the surface energy flux in all cases

		Net radiation (W m^{-2})				Latent heat flux (W m^{-2})				Sensible heat flux (W m^{-2})				Ground heat flux (W m^{-2})			
		Tongyu	Naiman	Jinzhou	Miyun	Tongyu	Naiman	Jinzhou	Miyun	Tongyu	Naiman	Jinzhou	Miyun	Tongyu	Naiman	Jinzhou	Miyun
RMSE	Case 1	123.81	122.04	138.52	135.69	72.79	68.86	86.07	114.85	85.86	62.89	100.08	70.65	44.54	45.98	36.78	40.23
	Case 2	129.23	118.70	150.06	162.26	77.18	76.09	98.92	145.17	86.82	76.10	110.56	69.83	46.21	42.32	38.85	44.05
	Case 3	129.53	136.61	149.37	149.09	61.19	67.51	99.08	134.46	83.82	59.90	94.46	69.31	37.45	45.65	22.97	41.11
	Case 4	129.33	120.40	157.36	142.78	58.67	66.66	108.32	150.49	85.27	57.15	101.22	64.29	39.63	44.08	21.97	39.88
Bias	Case 1	4.47	9.23	18.92	28.84	−14.86	−17.42	−7.74	34.28	34.17	9.28	49.93	19.70	5.28	−4.08	2.76	−8.95
	Case 2	21.39	31.41	30.25	38.44	0.37	−18.80	1.30	59.21	33.69	19.31	49.47	21.06	5.75	0.55	4.78	−8.29
	Case 3	23.60	25.60	25.16	32.83	−2.24	6.16	18.15	58.14	37.71	2.11	40.47	18.40	5.82	−4.44	0.71	−9.72
	Case 4	26.16	26.76	47.85	34.79	−2.57	6.54	15.39	64.90	35.72	5.93	47.19	16.19	4.77	−4.02	1.07	−8.66
R	Case 1	0.85	0.86	0.83	0.85	0.61	0.76	0.74	0.62	0.76	0.82	0.62	0.49	0.75	0.82	0.72	0.64
	Case 2	0.86	0.89	0.82	0.80	0.61	0.70	0.67	0.59	0.74	0.78	0.51	0.53	0.76	0.85	0.72	0.57
	Case 3	0.86	0.85	0.82	0.83	0.74	0.79	0.71	0.64	0.80	0.82	0.53	0.47	0.81	0.82	0.74	0.62
	Case 4	0.86	0.88	0.81	0.84	0.76	0.80	0.63	0.58	0.76	0.84	0.51	0.50	0.77	0.84	0.77	0.64

also affect the soil heat capacity, which leads to errors in calculating the soil heat flux at 5 cm. The daily averaged integral values of ground heat flux are 0.43, 0.74, 0.55, and 0.5 MJ m^{-2} at Tongyu, Naiman, Jinzhou, and Miyun, respectively. The portion of ground heat flux in the net radiation is 4.3%, 6.8%, 3.8%, and 4.2% for each site, respectively. The daily averaged integral values of ground heat flux across the Jinta oasis in the Heihe River Basin range from 0.7 to 0.9 MJ m^{-2}, slightly higher than the values of the semiarid northeastern region (Wen, 2011). The reason may be that there are fewer rainy and cloudy days and more sunny days in the northwestern region during summer, and therefore more solar radiation is received compared with the northeastern region, leading to more energy heat reserved in the soil.

In general, the characteristics of diurnal variation of the net radiation and surface energy flux over the semiarid regions can be simulated well in Cases 2–4. The GVF and the terrain height accuracy have been improved in Case 3 and Case 4, which led to improvement in the simulations of latent heat flux, sensible heat flux, and ground heat flux at Tongyu and improvement in the simulated ground heat flux at Miyun. However, the simulation error increased slightly at Jinzhou. Because local rainfall will lead to increased soil moisture in summer and rapid, abrupt changes in surface energy, the accuracy of the rainfall simulations will affect the simulations of energy budgets over the underlying surface.

The scatter diagram and error statistics (Figure 8) clearly show that the correlation coefficients of net radiation between simulations and observations are all above 0.81 in The correlation coefficient of the ground heat flux is 0.84 at Naiman, which is the highest value, and only 0.50 at Miyun, which is the lowest. This is due to the underlying surface land use type, oasis, at Naiman. There is desert around the Naiman station, and the vegetation cover is low, so the energy transfer in summer depends mainly on sensible heat transmission. The GVF is lower in the model grid at the

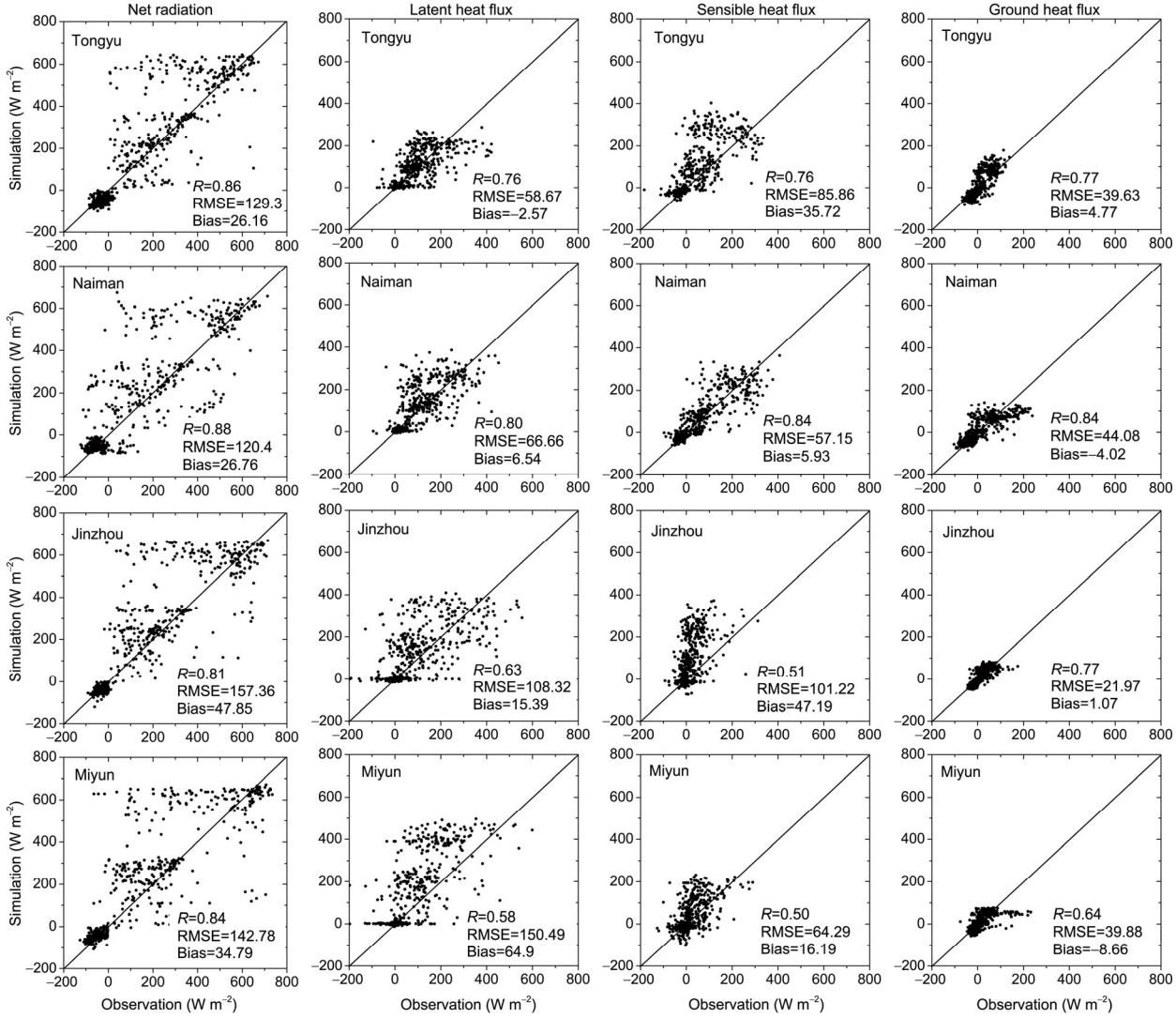

Figure 8 The scatter diagram and error statistics between observations and Case 4 results of net radiation and surface energy fluxes.

Naiman station, which is in good agreement with the true situation. The GVF is higher at Jinzhou and Miyun and the sensible heat transmission is lower, but the simulated value is higher. Because the surface temperature decreases at night from 20∶00 to 05∶00 and atmosphere temperature rises relatively, sensible heat flows from the atmosphere to the ground. The sensible heat flux is generally negative and more than −50 W m^{-2} at night and can even reach −150 W m^{-2} sometimes. This phenomenon appears mostly in cloudy situations. The model can simulate this phenomenon with smaller values, due to simulation error.

The correlation coefficients between the simulations and the observations for the surface heat flux of each station are all above 0.64. The lowest RMSE value is 21.9 W m^{-2} at Jinzhou, and the highest is 44.1 W m^{-2} at Naiman. The bias errors are positive at Tongyu and Jinzhou, but negative at Naiman and Miyun. The transmission of sensible heat depends mainly on the vegetation's foliar heat transport and, to a lesser degree, on soil heat transport during the day. The transmission of latent heat depends mainly on soil evaporation, but if there is an inadequate water supply, the transmission of latent heat is affected by increases in stoma damping (Zhang et al., 1998). In low GVF areas, where there is insufficient rainfall and low soil moisture levels, the energy exchange from the land surface to the atmosphere is mainly dependent on the sensible heat flux during daytime when the solar energy is absorbed by the soil. Thus, the amount of soil moisture is the key factor in the energy balance process.

In general, the errors are caused by the followings. First, both observation errors and the systematic errors in model can cause the Bias error. For example, a thin layer water or water drop adhered to observation instrument on rainy days may affect the observations in some way, but how great the effect of this is in detail needs to be investigated further. Second, the accuracy of the initial field and boundary conditions can also cause systematic bias (Hanna et al., 2001), which will affect simulation results. Third, the spatial distribution of the flux stations is limited and can only represent approximately 300 m^2 to 1 km^2, but the spatial resolution of the model grid is 5 km. The mismatch of spatial resolutions can cause errors. However, if the model resolution is too small, it will cause problems, such as slow operation speeds and unstable calculations. Thus, the 5-km resolution minimizes the trade-offs and is selected for our research.

2.3 Precipitation and soil moisture

The soil moisture of the surface is the most critical factor in the surface energy-water cycle. The main factors in the model that affect soil moisture are permeability, surface and sub-surface runoff, gradient diffusion, gravity, and water absorption by plant roots. Soil moisture is balanced by two processes: soil moisture is lost by evaporation at the ground surface and is supplied by deeper soil moisture. Generally, evaporation begins to be greater than the deeper soil moisture supplement 2 h after sunrise, at which time the surface layer soil moisture has reached its peak value. As the evaporation begins to be smaller than deep soil water supplement about 2 h before sunset, the surface layer soil moisture reaches its minimum value. The loss of soil moisture is due primarily to evaporation. However, the moisture supplied by the deep soil is much less than the surface layer soil water evaporation.

Figure 9 is the comparison of precipitation and soil moisture between simulations and observations at all flux stations in Case 4 from June to August 2009. There is a clear trend of diurnal variation in the soil moisture at 10 cm depth. Because of the evaporation of surface layer soil water, the soil water content decreases during the daytime and slightly increases at nighttime. Soil moisture increases rapidly when precipitation occurs and subsequently decreases slowly. During the observation period, the averaged 10 cm depth observed soil moisture is 0.157 cm^3 cm^{-3} at Tongyu, and the soil moisture reaches values of 0.3 cm^3 cm^{-3} or higher with heavy precipitation events on June 11 and July 1, and slowly decreased following these events. There is almost no heavy precipitation from July to August in 2009, so the observed soil moisture remains at 0.15 cm^3 cm^{-3} or less. The simulations are slightly higher than the observations; the averaged simulated soil moisture is 0.21 cm^3 cm^{-3}. Several precipitation events occurred and can be simulated by the WRF model. The variation of simulated soil moisture is almost consistent with the observations at Jinzhou, but the rate of soil moisture decrease is greater than the observed. The observed averaged soil moisture value is 0.33 cm^3 cm^{-3} at Miyun, and there is little change after precipitation. The simulation, having an averaged value of 0.22 cm^3 cm^{-3}, is lower than the observed.

We can conclude from the above analysis that the main factor that affects soil moisture is precipitation. Although the energy exchanges caused by soil water phase changes are considered by Noah Land Surface Model, the soil types can also effect the water-energy cycle between the ground surface and the atmosphere. Difference in soil textures and soil types can also determine other soil parameters, such as the maximum water content, soil permeability, and diffusion coefficients, among others (Wen, 2011). At present, true soil type data are hard to acquire for the correction of the model in the simulation area, which may introduce bias errors in the soil moisture simulations. Because soil moisture is very sensitive to precipitation, the large discontinuities in precipitation both spatially and temporally over simulation area will lead to greater errors and further affect the simulated surface energy balance.

2.4 Differences in the distribution of surface meteorological elements

To investigate the differences resulting from changes in the

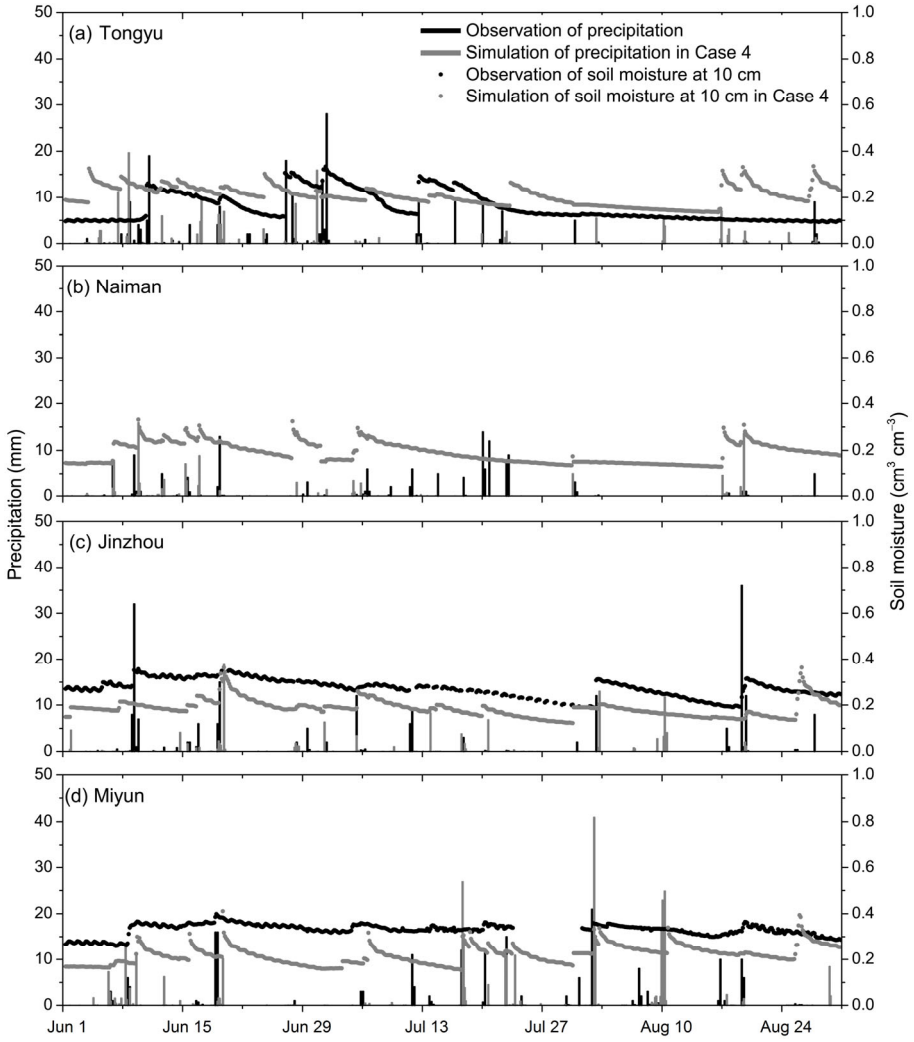

Figure 9 The variation tendency of observations and Case 4 results of precipitation and 10 cm soil moisture from June to August 2009 at Tongyu, Naiman, Jingzhou, and Miyun.

GVF, terrain height and data assimilation of near-surface meteorological parameters between Cases 1–4, the comparison diagrams for all cases are as follows. The distribution of the 2-m temperatures (Figure 10(a)) shows that the temperature ranges from 14°C to 20°C over the northwestern area of Inner Mongolian Plateau, while the temperature is higher than 20°C over the central and northeastern areas due to low elevations. The air temperature in Case 4 is higher by approximately 0.5–1.0°C compared to Case 1 over the northwestern area. Figure 10(b) shows the distribution of surface temperature. The spatial distribution pattern of surface temperature is similar to the distribution of 2-m temperature, which forms inhomogeneous "patches" in Case 3 and Case 4. The SRTM terrain height data, which are more accurate and complex than those in default model, used in Case 3 and Case 4 eventually lead to the appearance of "patches" of surface temperatures over the simulation area, which corresponds to Figure 3(b). The surface temperature over the northwest Inner Mongolian Plateau is higher by approximately 1–2°C than in Case 1. The distribution of 10-cm soil temperature is similar to the 2-m air temperature and surface temperature (Figure 10(c)). The averaged soil temperature is lower by approximately 1–2°C than the average surface temperature, but is 2°C higher than the averaged 2 m temperature. The average soil temperature decreases by approximately 1–2°C over northeastern region in Case 4, but increases by approximately 2°C over northwestern region compared to Case 1. It is clear that the GVF can affect the soil temperature. The net shortwave radiation absorbed by vegetation canopy increases, while the energy received by soil is reduced, resulting in reduced surface layer soil temperature. Figure 10(d) shows the distribution of 2-m relative humidity. The simulated relative humidity is reduced by 5%–10% over the western region in Case 4 compared with in Case 1, but is increased by 10% over the southeastern region, which corresponds with the distribution of GVF in Figure 4. The relative humidity will increase with increased GVF.

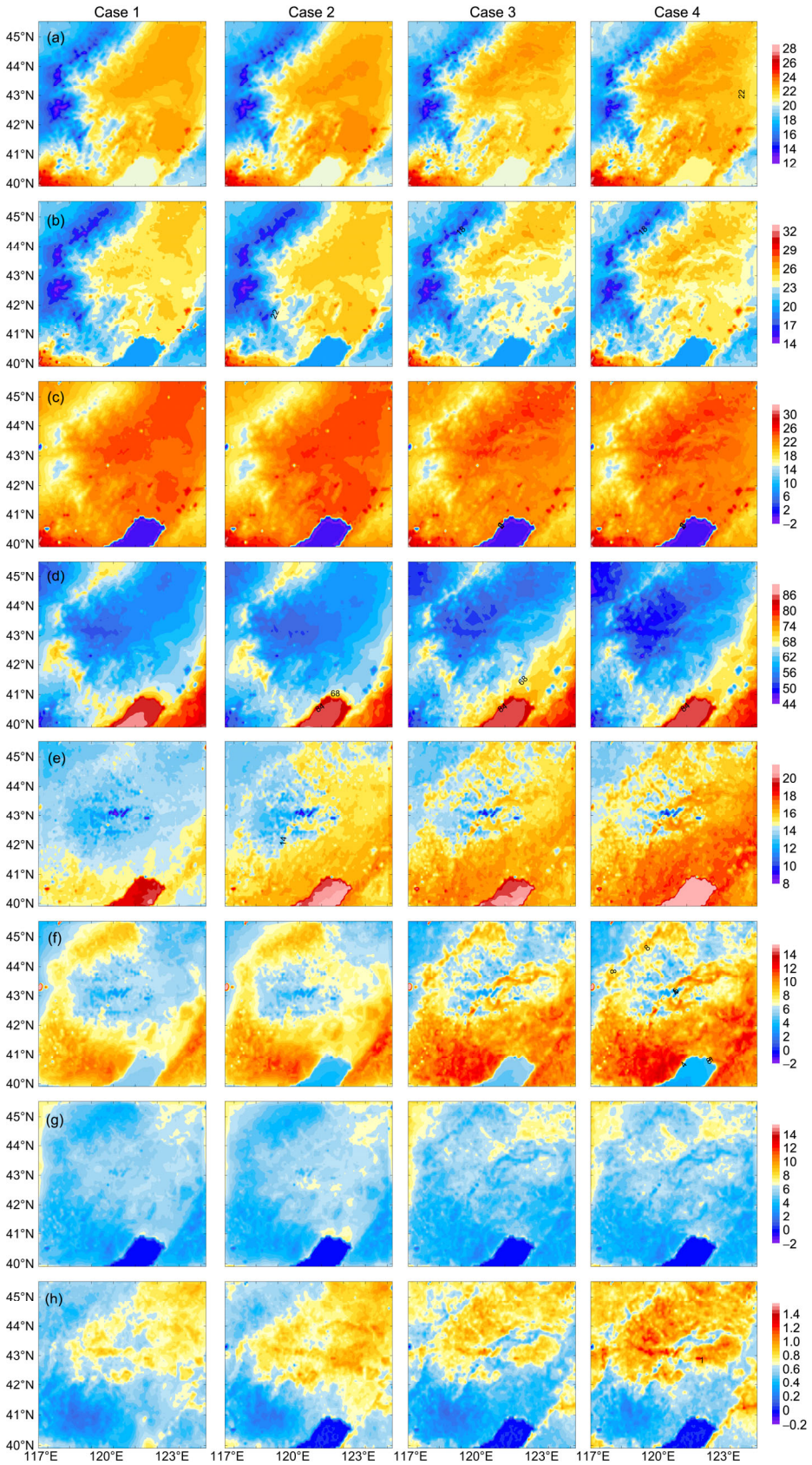

Figure 10 the comparison of averaged surface meteorological elements in all cases from June to July 2009. (a) 2-m temperature (°C); (b) surface temperature (°C); (c) 10-cm soil temperature (°C); (d) 2-m relative humidity (%); (e) net radiation (MJ m^{-2}); (f) latent heat flux (MJ m^{-2}); (g) sensible heat flux (MJ m^{-2}), and (h) ground heat flux (MJ m^{-2}).

Figure 10(e) shows the regional distribution of the daily averaged integral values of net radiation. Compared with Case 1, the net radiation is approximately 3 MJ m^{-2} d^{-1} higher in Case 4 in most parts of the simulation area. This is similar to the radiation budget characteristics of the Oasis-Gobi system in northwest China. The peak value of net radiation is approximately 150 W m^{-2} higher over oases than over the surrounding Gobi (Meng et al., 2009; Wen et al., 2012), which indicates that when the GVF is higher, the surface albedo is lower, resulting in higher surface energy and a surplus of energy. When the GVF is decreased, the albedo is increased, which leads to a reduction and loss of surface energy in this area. Compared to Case 1 and Case 2, the simulated values of latent heat flux in Case 3 and Case 4 are higher by approximately 2–4 MJ m^{-2} d^{-1} over the central and southern areas (Figure 10(f)). The "patches" pattern appears in Case 3 and Case 4 over the central areas. The sensible heat flux is increased by approximately 1–2 MJ m^{-2} d^{-1} over the north, while reduced by approximately 1–2 MJ m^{-2} d^{-1} over the central and eastern areas (Figure 10(g)). The distribution characteristics of ground heat flux are shown in Figure 10(h). The ground heat flux simulation of Case 4 is higher by 0.3–0.5 MJ m^{-2} d^{-1} over the whole northeastern region, indicating that the reduction of the GVF over the central and northern regions will lead to reduced vegetation evaporation and the weakening of latent heat transportation. The increasing surface temperatures will lead to increasing sensible heat flux and enhanced energy transmission from the soil to the atmosphere. Finally, GVF changes can change the distribution ratio of the energy transfer between the ground and the atmosphere (Chu et al., 2005; Pan et al., 2012a).

In general, soil moisture is the most critical factor in the surface energy-water cycle over arid and semiarid areas. Due to the lack of irrigation, soil moisture is very sensitive to precipitation, but precipitation has significant discontinuities both in space and time, which will cause errors in the simulation. Vegetation indices also significantly affect the sensible and latent heat fluxes between the land surface and the atmosphere. During the growing season, vegetation grows rapidly. The NDVI and LAI values are higher, the vegetation has vitality, and there is more evaporation with little impedance, so the proportion of latent heat transport increases across the surface and soil temperature decreases. The surface energy should be balanced according to the energy budget principle, which will lead to decreases in the sensible heat and surface heat fluxes. As an integrated experiment in Case 4, not only have the atmospheric fields been improved by the data assimilation system in the WRF model but also the vegetation coverage and terrain height have been improved in the initial fields, which can now reflect more realistic conditions during simulations. Therefore, the WRF model can simulate the partitioning of surface energy transmission, such as sensible heat flux, latent heat flux, and ground heat flux, more accurately and can reflect the surface phenomena that affect the atmosphere semiarid areas. The simulations in Case 4 are better than those in Case 1 on the whole.

3 Discussion and conclusions

In this paper, the near-surface meteorological elements and the processes of the energy water cycle have been simulated with the WRF model and its three-dimensional variational system from June to July 2009 over semiarid areas of northeastern China. The improvement of the simulations by using the data assimilation method and updating the land surface characteristics has been discussed and analyzed in four experiments. Realistic high spatial resolution land surface information has been incorporated into model's initial fields to improve the simulation of land surface processes. Near-surface meteorological elements have been assimilated into the model to reflect more realistic atmospheric fields in the simulations. The meso-scale characteristics of the water-energy cycle can be simulated well. The high-quality near-surface meteorological assimilation data and surface energy flux data generated by Case 4 are important basic data for monitoring climate change and droughts. The main conclusions are as follows:

(1) The surface terrain height and surface GVF data in initial fields cannot fully reflect the inhomogeneous surface conditions in high-resolution simulations in the WRF model. Higher temporal and spatial resolution remote sensing data of GVFs and DEMs are used to replace the default data in the WRF model, and the near-surface temperature, relative humidity, wind speed, and wind direction are assimilated using WRF-3DVar. The model results show that the simulations, such as surface meteorological elements and surface energy flux, are in good agreement with the observations, and the bias error is low compared with control experiment in Case 1. The simulation of near-surface temperature, wind, and humidity is improved by the data assimilation experiment (Case 2), and the simulation of the land surface energy budget components is improved by the land surface parameterization sensitivity experiment (Case 3) and integrated experiment (Case 4).

(2) The simulations of the near-surface temperature, relative humidity, and wind speed and direction in Case 4 are in good agreement with the observations at Tongyu, Naiman, Jinzhou, and Miyun. The bias errors of the near-surface temperature are lower than the control experiment (Case 1) by 0.41, 0.12, and 0.86°C at Tongyu, Naiman, and Miyun, respectively. The correlation coefficients between the simulations and the observations of temperatures are 0.85, 0.88, 0.81, and 0.70 at the four sites, respectively. The simulated wind speed is too large in the WRF model, which may be caused by the low value of surface roughness length in the land surface scheme. However, the simulated wind speeds have also improved significantly in Case 4, and the wind

direction and wind speed trend is still simulated well.

(3) The peak value of net radiation can reach 600 W m^{-2} in cropland over semiarid northeastern China during the summer. The surface energy transmission depends mainly on the latent heat flux in June and July. Energy transmission due to sensible heat flux is the second most important factor. The peak value of latent heat flux can reach 400 W m^{-2}, but this drops to 200 W m^{-2} in August. The bias error of the sensible heat flux is lower in Case 4 than in Case 1. Because the GVF over the simulation region is higher compared with Case 1, the simulated values of net radiation and latent heat flux are increased, and the sensible heat flux and ground heat flux are decreased over cropland. The altered portions of the water-energy transmission between the ground surface and the atmosphere have been simulated well over the semiarid northeastern area in summer and are reflected by Case 4.

(4) Soil moisture is the key factor in surface water-energy cycle. The time series variation of soil moisture can be simulated well in the WRF model over the northeastern semiarid region in summer. The simulation results show that there are no large differences between the simulations and the observations. The simulation results are higher than observations at Tongyu, and lower than observations at Jinzhou and Miyun. Because local rainfall leads to increased soil moisture in summer and makes abruptly changes the surface energy, creating more accurate rainfall simulations in the future will improve the overall model.

(5) The comparison of the averaged surface meteorological elements of all cases shows that the surface energy distribution ratio changes after the status of underlying surface was changed and the meteorological elements were assimilated. The GVF was 10%–20% lower than the default value over the northwestern area of the Inner Mongolian Plateau and 10%–20% higher over the central and northeastern areas. The air temperature, surface temperature, and soil temperature are lower in the higher GVF area. The net radiation is approximately 3 MJ m^{-2} d^{-1} higher with the higher GVF, leading to surplus over this area. The simulated value of latent heat flux is approximately 2–4 MJ m^{-2} d^{-1} higher over the central and southern areas. The sensible heat flux is approximately 1–2 MJ m^{-2} d^{-1} higher over the north, but approximately 1–2 MJ m^{-2} d^{-1} lower over the central and eastern areas. The transmission of latent heat and sensible heat between the surface and the atmosphere is significantly affected by vegetation coverage.

The land surface remote sensing data and meteorological observation data are combined in an atmospheric model in this study. The meteorological elements have been assimilated into the WRF model, leading to the improvement of atmospheric fields, and the simulation of land surface energy budget components are improved by better land surface characteristics. However, comparison of the simulations with observations shows that the model grid point is larger than the area covered by flux stations, which may cause bias errors. Therefore, we will provide simulation with a more appropriate resolution to reduce the bias error in the future. This work combines multi-source data with different spatial and temporal scales into numerical simulations. These simulations not only can make up for the limitations of using observation data but also can provide references for regional atmospheric models of inhomogeneous underlying surfaces over semiarid northeastern China. The assimilation datasets generated by this work can be applied to research of climate change and monitoring of arid environments. We will publish this dataset as soon as possible, and we look forward to its verification by scientists.

We thank the Key Laboratory of Regional Climate-Environment Research for Temperate East Asia (RCE-TEA), the Chinese Academy of Sciences and the Monsoon Asia Integrated Regional Study (MAIRS) for providing the field experimental flux data and the National Meteorological Information Center (NMIC) of China Meteorological Administration (CMA) for providing the near-surface meteorological data. We thank Jiancheng Shi from the Institute of Remote Sensing and Digital Earth (RADI), Chinese Academy of Sciences for his support of this work, as well as the suggestions from two reviewers. This work was supported by the National Basic Research Program of China (Grant No. 2010CB950504), the National High-tech R&D Program of China (Grant No. 2013AA122003), and the open funds of the Key Laboratory of Land Surface Process and Climate Change in Cold and Arid Regions, Chinese Academy of Sciences (Grant No. LPCC201101).

Barker D M, Huang W, Guo Y R, et al. 2004. A three-dimensional variational data assimilation system for mm5: Implementation and initial results. Mon Weather Rev, 132: 897–914

Barrett D J, Renzullo L J. 2009. On the efficacy of combining thermal and microwave satellite data as observational constraints for root-zone soil moisture estimation. J Hydrometeorol, 10: 1109–1127

Case J L, Crosson W L, Kumar S V, et al. 2008. Impacts of high-resolution land surface initialization on regional sensible weather forecasts from the wrf model. J Hydrometeorol, 9: 1249–1266

Chen F, Dudhia J. 2001a. Coupling an advanced land surface-hydrology model with the penn state-ncar mm5 modeling system. Part I: Model implementation and sensitivity. Mon Weather Rev, 129: 569–585

Chen F, Dudhia J. 2001b. Coupling an advanced land surface-hydrology model with the penn state-ncar mm5 modeling system. Part II: Preliminary model validation. Mon Weather Rev, 129: 587–604

Chen F, Manning K W, LeMone M A, et al. 2007. Description and evaluation of the characteristics of the ncar high-resolution land data assimilation system. J Appl Meteorol Clim, 46: 694–713

Cheng T T, Han Z W, Zhang R J, et al. 2010. Black carbon in a continental semiarid area of northeast China and its possible sources of fire emission. J Geophys Res-Atmos, 115: D23

Chu P C, Lu S H, Chen Y C. 2005. A numerical modeling study on desert oasis self-supporting mechanisms. J Hydrol, 312: 256–276

Courtier P. 1997. Variational methods. J Meteorol Soc Jpn, 75: 211–218

Dong J L, Han Z W, Zhang R J, et al. 2011. Evaluation and analysis of wrf-simulated meteorological variables in the urban and semiarid areas of China (in Chinese). Sci Meteorol Sin, 31: 484–492

Dudhia J. 1989. Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. J Atmos Sci, 46: 3077–3107

Fu C B, Wen G. 2002. Several issues on aridification in the northern China (in Chinese). Clim Environ Res, 7: 22–29

Fu C B, Ma Z G. 2008. Global change and regional aridification (in Chinese). Chin J Atmos Sci, 32: 752–760

Grell G A, Dudhia J, Stauffer D R. 1994, A description of the fifth-generation penn state/ncar mesoscale model (mm5). NCAR Technical

Note. NCAR/TN-398+STR, doi: 10.5065/D60Z716B
Gu J F, Xiao Q N, Kuo Y H, et al. 2005. Assimilation and simulation of typhoon rusa (2002) using the wrf system. Adv Atmos Sci, 22: 415–427
Gutman G, Ignatov A. 1998. The derivation of the green vegetation fraction from noaa/avhrr data for use in numerical weather prediction models. Int J Remote Sens, 19: 1533–1543
Hanna S R, Yang R X. 2001. Evaluations of mesoscale models' simulations of near-surface winds, temperature gradients, and mixing depths. J Appl Meteorol, 40: 1095–1104
Huang J P, Zhang W, Zuo J Q, et al. 2008. An overview of the semiarid climate and environment research observatory over the Loess Plateau. Adv Atmos Sci, 25: 906–921
Irmak A, Kamble B. 2009. Evapotranspiration data assimilation with genetic algorithms and swap model for on-demand irrigation. Irrigation Sci, 28: 101–112
Janjic Z I. 1990. The step-mountain coordinate-physical package. Mon Weather Rev, 118: 1429–1443
Janjic Z I. 1996. The surface layer in the ncep eta model. Eleventh Conference on Numerical Weather Prediction. 354–355
Janjic Z I. 2002, Nonsingular implementation of the mellor-yamada level 2.5 scheme in the ncep meso model. NCEP Office Note. 61
Janjic Z I, Black T L, DiMego G. 1998. Contributions toward development of a future community mesoscale model (wrf). 16th Conference on Weather Analysis and Forecasting/Symposium on the Research Foci of the US Weather Research Program. 258–261
Kain J S, Fritsch J M. 1993. Convective parameterization for mesoscale models: The Kain-Fritsch scheme. The representation of cumulus convection in numerical models. Meteor Monogr, 46: 165–170
Kumar S, Sekhar M, Bandyopadhyay S. 2009. Assimilation of remote sensing and hydrological data using adaptive filtering techniques for watershed modelling. Curr Sci India, 97: 1196–1202
Li F Q, Crow W T, Kustas W P. 2010. Towards the estimation root-zone soil moisture via the simultaneous assimilation of thermal and microwave soil moisture retrievals. Adv Water Resour, 33: 201–214
Li X, Huang C L, Che T, et al. 2007. Development of a Chinese land data assimilation system: Its progress and prospects (in Chinese). Prog Nat Sci, 17: 163–173
Lin Y L, Farley R D, Orville H D. 1983. Bulk parameterization of the snow field in a cloud model. J Clim Appl Meteorol, 22: 1065–1092
Liu H Z, Dong W J, Fu C B, et al. 2004. The long-term field experiment on aridification and the ordered human activity in semiarid area at Tongyu, northeast China (in Chinese). Clim Environ Res, 9: 378–389
Liu H Z, Tu G, Dong W J, et al. 2006. Seasonal and diurnal variations of the exchange of water vapor and CO_2 between the land surface and atmosphere in the semiarid area (in Chinese). Chin J Atmos Sci, 30: 108–118
Liu S, Gong P. 2012. Change of surface cover greenness in china between 2000 and 2010. Chin Sci Bull, 57: 2835–2845
Liu W, Gao Y H, Ran Y H, et al. 2007. Contrast analyses of simulation results in Heihe basin utilizing the different resolution dem data (in Chinese). Plateau Meteorol, 26: 525–531
Lorenc A C. 1986. Analysis-methods for numerical weather prediction. Q J Roy Meteor Soc, 112: 1177–1194
Ma Y M, Wang J M, Menenti M, et al. 1997. The study of regional surface heat fluxes over the heterogenous landscape of Heife (in Chinese). Clim Environ Res, 2: 293–301
Meng X H, Lu S H, Zhang T T, et al. 2009. Numerical simulations of the atmospheric and land conditions over the Jinta oasis in northwestern China with satellite-derived land surface parameters. J Geophys Res-Atmos, 114: 100–120
Mlawer E J, Taubman S J, Brown P D, et al. 1997. Radiative transfer for inhomogeneous atmospheres: Rrtm, a validated correlated-K model for the longwave. J Geophys Res, 102: 16663–16682
Pan X D, Li X, Ran Y H, et al. 2012a. Impact of underlying surface information on wrf modeling in Heihe River basin (in Chinese). Plateau Meteorol, 31: 657–667
Pan X D, Tian X J, Li X, et al. 2012b. Assimilating doppler radar radial velocity and reflectivity observations in the weather research and forecasting model by a proper orthogonal-decomposition-based ensemble, three-dimensional variational assimilation method. J Geophys Res-Atmos, 117: D17
Tanaka K, Ishikawa H, Hayashi T, et al. 2001. Surface energy budget at amdo on the Tibetan Plateau using game/Tibet IOP98 data. J Meteorol Soc Jpn, 79: 505–517
Tu G, Liu H Z, Dong W J, et al. 2006. The near surface layer energy budge on degraded-grassland in semiarid area at tongyu in Jilin Province (in Chinese). Clim Environ Res, 11: 723–732
Wen X H. 2011. The numerical simulation of water-energy budget and micro-climate effect on inhomogeneous underlying surface over Jinta oasis-gobi. Doctoral Dissertation. Lanzhou: Chinese Academy of Science. 131
Wen X H, Lu S H, Jin J M. 2012. 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, 13: 573–587
Xue M, Droegemeier K K, Wong V. 2000. The advanced regional prediction system (arps)—A multi-scale nonhydrostatic atmospheric simulation and prediction model. Part I: Model dynamics and verification. Meteorol Atmos Phys, 75: 161–193
Yan Y P, Wang J M, Menenti M, et al. 2001. Numerical simulation of surface energy fluxes over heterogeneous land surfaces in heife area (in Chinese). Plateau Meteorol, 20: 132–139
Yang K, Wang J M. 2008. A temperature prediction-correction method for estimating surface soil heat flux from soil temperature and moisture data. Sci China Ser D-Earth Sci, 51: 721–729
Yang K, Koike T, Kaihotsu I, et al. 2009. Validation of a dual-pass microwave land data assimilation system for estimating surface soil moisture in semiarid regions. J Hydrometeorol, 10: 780–793
Zeng J, Zhang Q. 2012a. Mean characteristics of land surface key parameters in semiarid and arid regions of China in summer of 2008 (in Chinese). Plateau Meteorol, 31: 1539–1550
Zeng J, Zhang Q. 2012b. A comparative study of the characteristics of the clear-sky land surface processes over the different underlying surfaces in the northern part of china during July–September 2008 (in Chinese). Acta Meteorol Sin, 70: 821–836
Zhang Q, Zhao M. 1998. A numerical simulation of characteristics of land surface process under interaction of oasis and desert in arid region (in Chinese). Plateau Meteorol, 17: 336–346
Zhang Q, Zeng J, Zhang L Y. 2012. Characteristics of land surface thermal-hydrologic processes for different regions over North China during prevailing summer monsoon period. Sci China Earth Sci, 55: 1872–1880