RESEARCH ARTICLE

# WRF-based dynamical downscaling of ERA5 reanalysis data for High Mountain Asia: Towards a new version of the High Asia Refined analysis

Xun Wang [ORCID]    |    Vanessa Tolksdorf    |    Marco Otto [ORCID]    |    Dieter Scherer [ORCID]

Chair of Climatology, Institute of Ecology, Technische Universität Berlin, Berlin, Germany

**Correspondence**
Xun Wang, Technische Universität Berlin, Institute of Ecology, Chair of Climatology, Rothenburgstraße 12, Berlin 12165, Germany.
Email: xun.wang@tu-berlin.de

**Funding information**
German Federal Ministry of Education and Research (BMBF), Grant/Award Number: FKZ 03G0878G

## Abstract

The High Asia Refined analysis (HAR) is a regional atmospheric data set generated by dynamical downscaling of the Final operational global analysis (FNL) using the Weather Research and Forecasting (WRF) model. It has been successfully and widely utilized. A new version (HAR v2) with longer temporal coverage and extended domains is currently under development. ERA5 reanalysis data is used as forcing data. This study aims to find the optimal set-up for the production of the HAR v2 to provide similar or even better accuracy as the HAR. First, we conducted a sensitivity study, in which different cumulus, microphysics, planetary boundary layer, and land surface model schemes were compared and validated against in situ observations. The technique for order preference by similarity to the ideal solution (TOPSIS) method was applied to identify the best schemes. Snow depth in ERA5 is overestimated in High Mountain Asia (HMA) and causes a cold bias in the WRF output. Therefore, we used Japanese 55-year Reanalysis (JRA-55) to correct snow depth initialized from ERA5 based on the linear scaling approach. After applying the best schemes identified by the TOPSIS method and correcting the initial snow depth, the model performance improves. Finally, we applied the improved set-up for the HAR v2 and computed a one-year run for 2011. Compared to the HAR, the HAR v2 has a better representation of air temperature at 2 m. It produces slightly higher precipitation amounts, but the spatial distribution of seasonal mean precipitation is closer to observations.

**KEYWORDS**
cold bias, dynamical downscaling, ERA5, HAR, High Mountain Asia, snow depth, WRF

# 1 | INTRODUCTION

High Mountain Asia (HMA) is a geographic region that includes the Tibetan Plateau (TP) and its surrounding mountain ranges, such as the Himalayas, the Karakoram, the Tian Shan, and so on. Climate-triggered natural hazards pose a threat to human lives in HMA, for example, big landslides regularly occurring in the Fergana basin along the foothills of Tian Shan (Roessner et al., 2005). Landslides are predetermined by static factors that can be derived from surface characteristics but are triggered by dynamic factors, which are mainly extreme and prolonged rainfall, as well as earthquakes (Dai and Lee 2002; Hong et al. 2007; Kirschbaum et al., 2012). Another example of climate-triggered natural hazards is the Glacier Lake Outburst Flood (GLOF). Glacier thinning and retreat in the Himalayas, caused by rising air temperature, have resulted in the formation of new glacier lakes and the enlargement of existing ones (ICIMOD, 2011). The sudden discharge of water from these lakes is known as GLOF and leads to extensive damage in downstream villages (Kropáček et al., 2015).

Availability of climate data with a high spatial and temporal resolution is crucial for a better understanding of climatic triggering mechanisms of these localized hazards. However, in HMA, in situ meteorological observations are sparsely and unevenly distributed (Hasson et al., 2016), for example, only a few stations are available in the western TP due to the harsh environment and complex terrain. Moreover, the existing stations are commonly situated at lower altitudes close to valley-based settlements or airports. Thus, our knowledge of the climate at high elevations, that is, where most of the hazards mentioned above occur, is still limited. Global reanalysis data can provide evenly distributed climate data, but they are still too coarse to resolve fundamental processes over complex terrains, such as orographically-induced precipitation, and therefore, they are not suitable for applications at regional to local scales (Leung et al., 2003; Lo et al., 2008; Feser et al., 2011). Here, regional climate models (RCMs) applying dynamical downscaling method have great potential to overcome this problem.

The High Asia Refined analysis (HAR, Maussion et al., 2011; 2014) is a regional atmospheric data set generated by dynamical downscaling using the Weather Research and Forecasting (WRF) model version 3.3.1 (Skamarock and Klemp, 2008) as RCM. The Final operational global analysis (FNL) from the National Centres of Environmental Prediction (NCEP) was used as forcing data. The HAR covers the period from October 2000 to October 2014 and is available in 30 km (3-hourly interval) and 10 km (hourly interval) resolution. The HAR provides detailed and accurate gridded climate data for HMA region. It has been comprehensively analysed, especially in terms of precipitation and atmospheric water transport (Maussion et al., 2014; Curio et al., 2015; Pritchard et al., 2019; Li et al., 2020) and has been successfully applied in many research fields, such as glacier mass balance modelling (Mölg et al., 2014), snow and energy balance modelling (Huintjes et al., 2015), and so forth.

However, the short temporal coverage of the HAR makes it unsuitable for long-term and climatological studies. Moreover, its 10 km domain does not cover the whole TP and Tian Shan, which further limits its application within these two regions. Therefore, a new version (HAR v2) with extended temporal coverage and a larger 10 km domain is developed. The state-of-the-art ERA5 reanalysis data set (Copernicus Climate Change Service (C3S), 2017) from the European Centre for Medium-Range Weather Forecast (ECMWF) is used as forcing data. We switch to ERA5 because it will eventually cover the period from 1950 to near real time (currently available from 1979 to near real time), which is much longer than FNL (available from 1999 to near real time).

The overall goal of this study is to find an optimal model set-up for the HAR v2 to provide similar or even better accuracy as the HAR. During the development of the HAR, the sensitivity of simulated precipitation to different physical parameterization schemes (PPSs) was already thoroughly tested (Maussion et al., 2011). However, changes in forcing data and domain configuration may have a significant impact on model output (Miguez-Macho et al., 2004; Leduc and Laprise, 2009; Kala et al., 2015; Huang and Gao, 2018). Thus, the PPSs used in the HAR might not be suitable for the HAR v2, and therefore, the first objective of the current study is to investigate the sensitivity of simulated total precipitation (Prcp) and air temperature at 2 m above ground (T2) to different cumulus (CU), microphysics (MP), planetary boundary layer (PBL) and land surface model (LSM) schemes. The technique for order preference by similarity to the ideal solution (TOPSIS) method is applied to determine the best PPSs.

Snow depth in ERA5 over the TP is reported to be largely overestimated (Orsolini et al. 2019). Snow depth is an important quantity in the initial condition, which later on determines surface albedo, alters surface energy balance, and influences T2. We assume that the overestimated snow depth in ERA5 leads to an underestimation of T2, and by correcting the bias in snow depth, the cold bias might be reduced. The second objective is to validate this assumption and to examine the model's sensitivity to initial snow conditions. Snow depth from the Japanese 55-year Reanalysis (JRA-55) data set is used to correct snow depth initialized from ERA5. The

final objective is to apply the best PPSs identified by the TOPSIS method and the snow correction approach as the final set-up for the HAR v2, and to compare the two versions of the HAR.

## 2 | METHODOLOGY

### 2.1 | WRF model set-up of the reference experiment (Ref)

WRF version 4.0.3 (Skamarock *et al.*, 2019) was employed as RCM for our sensitivity studies. We chose January and July 2011 as simulation periods to consider atmospheric conditions in summer and winter, which are under different influence by monsoon and mid-latitude westerlies. Initial and boundary conditions were derived from ERA5 reanalysis data set with 0.25° spatial resolution and hourly temporal resolution. The domain setup (Figure 1) consisted of two-way nested domains with 30 km and 10 km grid spacing (hereinafter: d30km and d10km). Only the output from d10km was used in this study. One might argue that ERA5 already has a high resolution of ~32 km, so the parent domain (d30km) might not be necessary. However, a one-day experiment, which directly downscaled ERA5 to 10 km resolution (see Supporting Information), shows that the large-scale circulation patterns are distorted in the direct downscaling approach, and the 500 hPa wind field from two-way nesting approach is closer to the forcing data ERA5

(Figure S1). Thus, we kept the d30km as parent domain. In the vertical direction, 28 Eta-levels were used. Lake surface temperature was substituted by daily mean surface air temperature using the *avg_tsfc.exe* module in WRF. The forcing strategy was daily re-initialization adopted from the HAR. Each run started at 12:00 UTC and contained 36 hr, with the first 12 hr as spin-up time. This strategy avoids the model from deviating too far from the forcing data and provides computational flexibility since daily runs are totally independent of each other and can be computed in parallel and in any sequence. PPSs used in Ref are the same as in the HAR. The model set-up for Ref are summarized in Table 1.

### 2.2 | Design of sensitivity experiments and evaluation methods

At first, we conducted four sets of experiments to examine the performance of different CU, MP, PBL and LSM schemes (Table 2, hereinafter, PPS experiments). In each experiment set, except for the reference scheme adopted from the HAR, we chose two additional schemes. The selected schemes fulfil at least one of the following criteria: (a) they are commonly used in the WRF community; (b) they have excellent performance according to previous studies and (c) they were not tested in Maussion *et al.* (2011). Except for the corresponding PPS, all the other set-ups are the same as for Ref. At the end of the PPS experiments, statistical measures for model
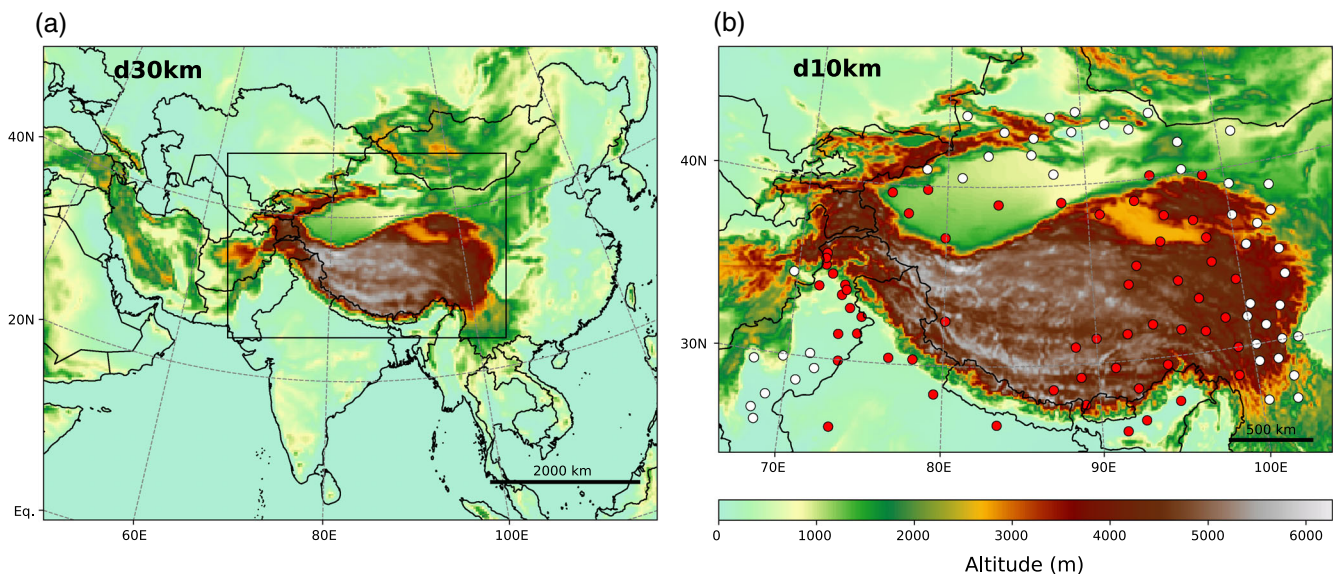


**FIGURE 1** Maps of (a) 30 km resolution domain (d30km, 281 × 217 grid points); (b) 10 km resolution domain (d10km, 382 × 253 grid points). The position of stations from global surface summary of the day (GSOD) are marked by points. For the comparison of simulation results with the HAR, only the stations marked by red points are used, because the stations marked by white points are outside the 10 km domain of the HAR [Colour figure can be viewed at wileyonlinelibrary.com]

**TABLE 1** WRF model setup for the reference experiment (Ref)

| Timing | |
|---|---|
| Simulation period | January and July 2011 |
| Time step | 120 s, 40 s |
| **Maps and grids** | |
| Map projection | Lambert conformal conic |
| Horizontal grid spacing | 30 km (281 × 217 grid points), 10 km (382 × 253 grid points) |
| Vertical levels | 28 eta-levels |
| Model top | 50 hPa |
| **Forcing strategy** | |
| Forcing data | ERA5 (0.25°, hourly) |
| Lake surface temperature | Substituted by daily mean surface air temperature |
| Initialization | Daily |
| Runs starting time | Daily at 12:00 UTC |
| Runs duration | 36 hr |
| Spin-up time | 12 hr |
| **Physical parameterization schemes** | |
| Longwave radiation | RRTM scheme (Mlawer et al., 1997) |
| Shortwave radiation | Dudhia scheme (Dudhia, 1989) |
| Cumulus (CU) | Grell 3D scheme (Grell, 1993; Grell and Devenyi, 2002) |
| Microphysics (MP) | Thompson scheme (Thompson et al., 2008) |
| Planetary boundary layer (PBL) | Mellor–Yamada–Janjic scheme (Janjic, 1994) |
| Land surface model (LSM) | Unified Noah land surface model (Tewari et al., 2004) |
| Surface layer | Eta similarity scheme (Janjic, 1994) |

performance were calculated, which include mean bias (MB), mean absolute error (MAE), root mean square error (RMSE) and Spearman correlation coefficient ($r_s$) for Prcp and T2, respectively. They are defined as follows (Wilks, 2006):

$$MB = \frac{1}{N}\sum_{i=1}^{N} \bar{P}_i - \bar{O}_i \qquad (1)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |\bar{P}_i - \bar{O}_i| \qquad (2)$$

$$RMSE = \frac{1}{N}\sum_{i=1}^{N} (\bar{P}_i - \bar{O}_i)^2 \qquad (3)$$

**TABLE 2** Summary of set-ups used in all the sensitivity experiments

| Experiments | Difference from Ref | Description |
|---|---|---|
| CU1 | CU | Kain-Fritsch cumulus potential scheme (Berg et al., 2013) |
| CU2 | CU | Grell-Freitas ensemble scheme (Grell and Freitas, 2014) |
| MP1 | MP | Purdue Lin scheme (Chen and Sun, 2002) |
| MP2 | MP | Morrison 2-moment scheme (Morrison et al., 2009) |
| PBL1 | PBL | Bougeault and Lacarrere scheme (BouLac, Bougeault and Lacarrere, 1989) |
| PBL2 | PBL | Yonsei University scheme (YSU, Hong et al., 2006)[a] |
| LSM1 | LSM | Unified Noah LSM with mosaic approach (Li et al., 2013) |
| LSM2 | LSM | Noah-MP land surface model (Niu et al., 2001; Yang et al., 2011)[b] |
| COMB | CU, MP, PBL | CU, MP and PBL schemes from CU1, MP2 and PBL2 |
| COMB_S | CU, MP, PBL and snow correction | PPSs same as COMB, but the initial snow depth and snow water equivalent are corrected |

[a]YSU scheme only works with revised MM5 surface layer scheme (Jiménez et al., 2012).
[b]All the options used in Noah-MP are default.

$$r_s = 1 - \frac{6\sum_{i=1}^{M} (R_{pi} - R_{oi})^2}{M(M^2 - 1)} \qquad (4)$$

Where $\bar{P}_i$ and $\bar{O}_i$ are simulated monthly averages and observed monthly averages at each station, and $\bar{P}_i - \bar{O}_i$ is the bias score at each station; $N$ is the total number of stations; $R_{pi}$ and $R_{oi}$ are the ranks of simulated and observed daily averages over all stations; $M$ is the total

number of days. MB demonstrates the systematic deviation of the model from the observations. However, MB could sometimes be misleading due to the offset of positive and negative values. MAE shows the average magnitude of the error. Since RMSE gives more weight to errors with larger absolute values and is more sensitive to outliers, it provides information about the variability of the error distribution (Chai and Draxler, 2014; Willmott and Matsuura, 2005). The distribution-free Spearman rank correlation coefficient ($r_s$) measures the extent to which simulated and observed values tend to change together. In order to identify the best scheme for each experiment set, MB, MAE, RMSE and $r_s$ were utilized as criteria in the TOPSIS method (Section 2.3). In the next step, the identified best schemes were combined to conduct two runs (Table 2): one without modifying initial snow depth (COMB) and one with bias correction of snow depth (COMB_S). We used JRA-55 to correct snow depth initialized from ERA5 (Section 2.4). Finally, the improved set-up was applied as the best set-up for the HAR v2 for a one-year run in 2011. The resulting one-year data set was then compared with observations and the HAR.

## 2.3 | Technique for order preference by similarity to the ideal solution (TOPSIS)

The TOPSIS method was applied to identify the best scheme among the sensitivity experiments. It is a multiple criteria decision-making method and aims at finding the optimal decision when the alternatives are numerous and conflicting. It was proposed by Hwang and Yoon (1981) and later applied in several studies, such as ranking general circulation models (Raju and Kumar, 2014; Jena et al., 2015; Li et al., 2019) and identifying the best PPSs in WRF (Sikder and Hossain, 2016; Stergiou et al., 2017). The basic concept of TOPSIS is to determine the best alternative that has the shortest and longest distance from the positive and negative ideal solution. We applied this method for each experiment set. The process was carried out following the description in Tzeng and Huang (2011):

1 Define the weighted normalized evaluation matrix containing $m$ alternatives and $n$ criteria:

$$r_{ij} = w_j \frac{a_{ij}}{\sqrt{\sum_{i=1}^{m} a_{ij}^2}} \qquad (5)$$

where $a_{ij}$ represents the criterion $j$ for alternative $i$; $w_j$ is the weight for each criterion. In our case, there are three

alternatives (Ref and other two schemes, $i = 1, 2, 3$) and 16 criteria (MB, MAE, RMSE and $r_s$ for Prcp and T2 in January and July, $j = 1, 2, ...16$).

Determine the positive ideal solution (PIS$_j$) and the negative ideal solution (NIS$_j$) for each criterion:

$$PIS_j = \left\{ \left( \max(r_{ij}) | j \in J_1 \right), \left( \min(r_{ij}) | j \in J_2 \right) | i = 1, 2, 3 \right\} \qquad (6)$$

$$NIS_j = \left\{ \left( \min(r_{ij}) | j \in J_1 \right), \left( \max(r_{ij}) | j \in J_2 \right) | i = 1, 2, 3 \right\} \qquad (7)$$

where $J_1$ and $J_2$ represent the benefit criteria (larger is better) and the cost criteria (smaller is better).

Calculate the Euclidean distances between each alternative to PISs ($D_i^+$) and to NISs ($D_i^-$):

$$D_i^+ = \sqrt{\sum_{j=1}^{n} \left( r_{ij} - PIS_j \right)^2} \qquad (8)$$

$$D_i^- = \sqrt{\sum_{j=1}^{n} \left( r_{ij} - NIS_j \right)^2} \qquad (9)$$

Define the closeness coefficient (CC$_i$) for each alternative as:

$$CC_i = \frac{D_i^-}{D_i^+ + D_i^-} \qquad (10)$$

Rank the alternatives by their CC in descending order. The alternative with the highest CC is chosen as the best scheme.

## 2.4 | Bias correction of snow depth

A systematic cold bias has been found over high elevated areas in WRF simulations (e.g., Gao et al., 2015; Karki et al., 2017; Bonekamp et al., 2018), including the HAR (Maussion, 2014; Pritchard et al., 2019). Several studies addressed this cold bias to snow-related processes (Tomasi et al., 2017; Meng et al., 2018). Orsolini et al. (2019) compared snow depth over the TP in five
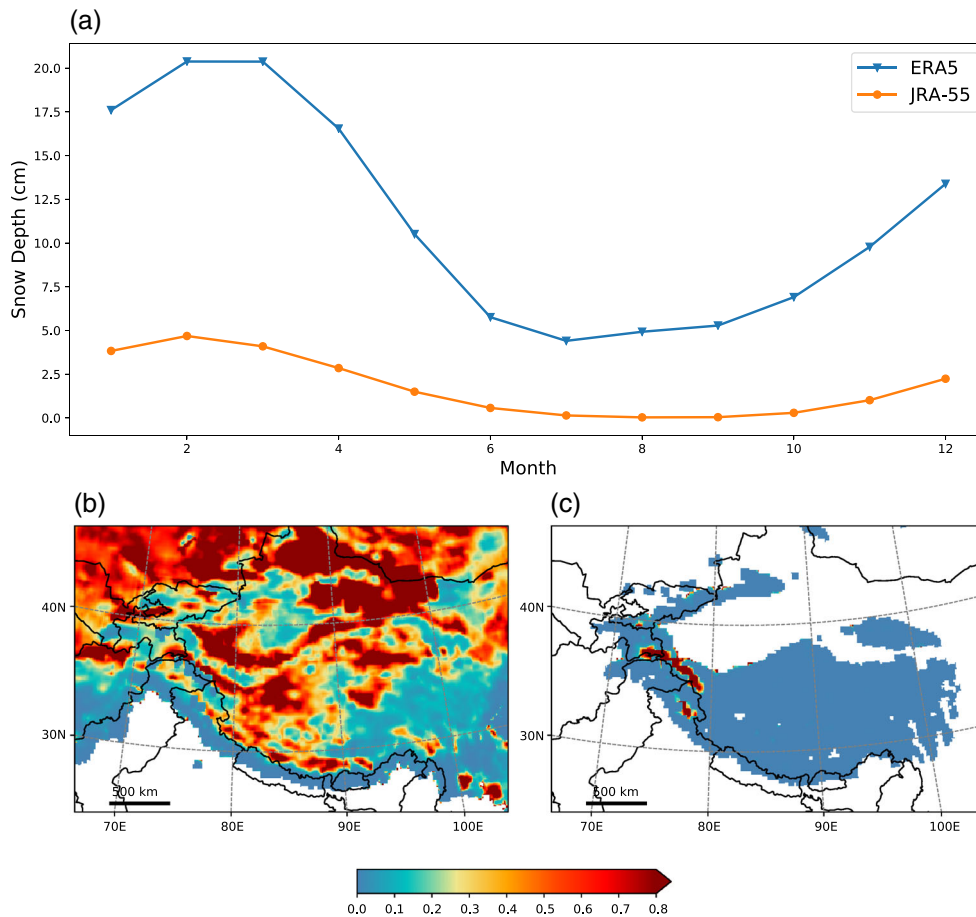
(a)



**FIGURE 2** (a) 35-year (1979–2013) mean annual cycle of snow depth from ERA5 and JRA-55 averaged over the whole area of d10km of the HAR v2. Map of scaling factor for d10km of HAR v2 in January (b) and July (c). The grid points where the value is equal to one are masked out [Colour figure can be viewed at wileyonlinelibrary.com]

(b)                                    (c)



recent global reanalyses using in-situ and satellite remote sensing observations. They found that ERA5 largely overestimates snow depth, especially during winter. Compared to its predecessor ERA-interim, snow depth in ERA5 is still much higher, because ERA5 does not assimilate snow cover from Interactive Multi-Sensor Snow and Ice Mapping System (IMS) for areas above 1,500 m (Orsolini *et al.*, 2019). A realistic representation of snow depth in the forcing data is crucial to accurately simulate snow cover, surface albedo, surface energy balance and T2. To overcome this issue, we could use snow depth from another data set to initialize WRF. But as mentioned before, we chose ERA5 as forcing data due to its long temporal coverage. If we introduce a third-party data set, the production of the HAR v2 would be dependent on it. To avoid this, we applied an alternative strategy that is already used in WRF, where land surface characteristic, such as snow-free albedo, leaf area index, and so forth, are given in the form of static data represented as 12-month climatology. Here, analogously, we used this concept of static data and introduced a gridded 12-month climatology of a scaling factor to correct snow-related variables in the initial conditions.

JRA-55 (Kobayashi *et al.*, 2015) developed by the Japan Meteorological Agency was utilized to calculate the scaling factors. JRA-55 has an excellent performance among reanalyses regarding snow depth (Orsolini *et al.*, 2019) and also has a relatively longer temporal coverage. Monthly means of snow depth at the model resolution (∼55 km) available from 1958 to 2013 were downloaded from NCAR Research Data Archive. We used a linear approach (Lenderink *et al.* 2007) to scale the initial snow depth by the ratio of long-term monthly means of JRA-55 and ERA5. First, snow depth of ERA5 was derived from snow water equivalent and snow density. The 12-month climatological snow depth of JRA-55 and ERA5 was calculated from 35-year monthly means (1979–2013). Between 1979 and 2013, the two data sets overlap. Second, the 12-month climatological snow depth was reprojected and linearly resampled to the grid of d30km and d10km, respectively. Then the gridded 12-month climatology of the scaling factor was calculated for every domain as the ratio of reprojected climatological snow depth between JRA-55 and ERA5. The ratio of grid points where reprojected ERA5 snow depth was zero was set to one. Figure 2a depicts the 35-year mean annual cycle from ERA5 and JRA-55 averaged over the whole

area of d10km of the HAR v2. Figure 2b,c show the map of scaling factors applied for d10km in January and July, respectively. After running the *real.exe* program in WRF, files containing the initial conditions are generated. In the initial conditions, snow depth is represented by two variables: physical snow depth (SNOWH) and snow water equivalent (SNOW). We modified these two variables in the initial files by multiplying them with the scaling factor of the corresponding month.

## 2.5 | Observational data

To calculate the skill scores described in Section 2.2, in-situ observations of Prcp and T2 from Global Surface Summary of the Day (GSOD) provided by National Centres of Environmental Information (NCEI) were used. We selected stations with more than 80% of records within the simulation period. After filtering, 103 stations are available within d10km of the HAR v2 (all points in Figure 1b), and 57 stations are left within d10km of the HAR (red points in Figure 1b). We compared the station data with the nearest grid cell from the model. Because of the difference between station elevation and the elevation of grid cells in the model, we applied a constant lapse rate of $-6.5$ K·km$^{-1}$ to correct simulated T2. No correction was applied for Prcp due to its high spatial and temporal variability.

Due to the sparse and uneven distribution of GSOD stations in HMA, for example, hardly any station located in western TP and Pamir-Karakorum (Figure 1b), we also made use of a satellite-based gridded precipitation product from Tropical Rainfall Measuring Mission (TRMM). Here, Multi-Satellite Precipitation Analysis (TMPA, Huffman and Bolvin, 2015) Rainfall Estimate Product 3B42 Version 7 was applied to examine the spatial distribution of simulated Prcp in the HAR and the HAR v2. TMPA merges satellite measurement with gauge data. It has a spatial resolution of 0.25° and a quasi-global coverage between 50°N and 50°S.

## 3 | RESULTS

### 3.1 | Sensitivity to PPSs

Statistical scores for all the PPS experiments and Ref are listed in Table 3. Besides, we use box-whisker plot to visualize the spatial distribution of bias scores (WRF-observation) over 103 GSOD stations. White triangles in each box in Figure 3 (Prcp) and Figure 4 (T2) correspond to the MB scores in Table 3.

For Prcp, Ref features a wet bias, with MB of 0.14 mm·day$^{-1}$ and 0.70 mm·day$^{-1}$ in January and July,

respectively (Table 3). The HAR outperforms Ref with respect to Prcp (Table S1). Sensitivities of Prcp to PPSs vary seasonally. The fluctuation of the boxes of all experiments implies that PPSs have a stronger influence on Prcp in July than in January (Figure 3). MB of Prcp ranges from 0.27 mm·day$^{-1}$ to 1.15 mm·day$^{-1}$ in July, and only from 0.10 mm day$^{-1}$ to 0.18 mm·day$^{-1}$ in January (Table 3). The main reason for this seasonal variability is that most GSOD stations are located in areas receiving more precipitation in summer than in winter (Maussion *et al.*, 2014; Curio and Scherer, 2016). Wider boxes in July indicate a stronger spatial variation of bias score in summer (Figure 3). Compared to the other two ensemble CU schemes, CU1 (Kain-Fritsch-Cumulus Potential scheme) shows the best performance regarding MAE and MB scores. In contrary, RMSE of CU1 in July is higher than Ref, which implies that the station-wise bias in CU1 is more scattered. CU1 is a modified version of the Kain-Fritsch scheme with a better treatment of shallow cumuli (Berg *et al.*, 2013). According to Qian *et al.* (2016), CU1 tends to suppress deep convections and consequently produces lower Prcp. All three MP schemes consider five hydrometeor species: cloud, rain, ice, snow and graupel. MP1 (Purdue Lin scheme) is a single-moment scheme only predicting the mixing ratio for these hydrometeor species (Chen and Sun, 2002). Ref (Thompson scheme) uses a double-moment description for rain and ice (Thompson *et al.*, 2008), while MP2 (Morrison two-moment scheme) uses a double-moment description for rain, ice, snow and graupel (Morrison *et al.*, 2009). MP2 performs the best (Table 3), probably due to its double-moment prediction in all ice-phase particles (Orr *et al.*, 2017). The nonlocal scheme PBL2 (YSU) has the best skill compared to the other two local schemes. PBL2 largely improves summer Prcp with MB of 0.27 mm·day$^{-1}$ compared to MB of 0.70 mm·day$^{-1}$ in Ref. As for LSM schemes, LSM1 uses the same scheme as Ref (Noah LSM) but with a mosaic approach (Li *et al.*, 2013), which considers sub-grid variability of land use. However, it does not improve Prcp simulation. LSM2 is Noah LSM with multi-parameterization options (Noah-MP). It captures winter Prcp better but produces more Prcp in summer than Ref (Table 3).

The same analysis is performed for T2. The HAR has a better performance than Ref in winter (Table S1). All PPS experiments including Ref produce larger T2 bias in January than in July, with MB of T2 ranging from $-0.06$ K to 0.70 K in July and $-0.39$ K to $-2.69$ K in January (Table 3). The spatial variability of the bias score is also stronger in winter (Figure 4). Independently from PPSs applied, all experiments show an overall underestimation of T2 in winter. CU schemes hardly influence winter T2. CU1, which shows some improvement over

**TABLE 3** Statistical scores for evaluation of PPS experiments over 103 GSOD stations within d10km of the HAR v2 (red and white points in Figure 1b) for Prcp (mm·day⁻¹) and T2 (K)

| | Prcp January | | | | Prcp July | | | | T2 January | | | | T2 July | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE | MB | RMSE | $r_s$ | MAE | MB | RMSE | $r_s$ | MAE | MB | RMSE | $r_s$ | MAE | MB | RMSE | $r_s$ |
| Ref | 0.22 | 0.14 | 0.42 | 0.67 | 1.53 | 0.70 | 2.68 | 0.48 | 2.89 | −1.21 | 3.99 | 0.80 | 1.20 | 0.35 | 1.81 | 0.83 |
| CU1 | 0.21 | **0.12** | **0.34** | 0.67 | 1.48 | **0.47** | (2.99) | 0.47 | 2.89 | −1.19 | 3.98 | 0.81 | 1.27 | (0.39) | 1.87 | 0.82 |
| CU2 | 0.23 | 0.14 | 0.43 | 0.68 | (1.87) | (1.15) | (3.62) | (0.26) | 2.89 | −1.20 | 3.99 | 0.81 | 1.30 | 0.17 | 1.88 | 0.85 |
| MP1 | 0.22 | 0.14 | 0.40 | 0.71 | 1.59 | **0.56** | 2.81 | (0.38) | 2.91 | (−1.38) | 4.05 | 0.81 | 1.17 | **0.21** | 1.78 | 0.83 |
| MP2 | 0.22 | 0.13 | 0.41 | 0.64 | 1.53 | **0.45** | 2.79 | 0.50 | 2.79 | −1.12 | 3.89 | 0.81 | 1.21 | **0.20** | 1.79 | 0.83 |
| PBL1 | (0.25) | (0.18) | 0.45 | 0.65 | 1.52 | 0.71 | 2.58 | 0.45 | **2.55** | −0.39 | **3.50** | 0.85 | 1.28 | (0.70) | 1.94 | 0.82 |
| PBL2 | 0.20 | **0.12** | 0.39 | 0.68 | **1.27** | **0.27** | **2.17** | 0.42 | 3.08 | (−1.68) | 4.28 | 0.83 | 1.21 | **0.30** | 1.84 | 0.84 |
| LSM1 | 0.23 | 0.14 | 0.43 | 0.68 | 1.54 | 0.71 | 2.71 | 0.49 | 2.83 | **−1.08** | 3.89 | 0.80 | 1.20 | 0.38 | 1.82 | 0.83 |
| LSM2 | **0.18** | **0.10** | **0.37** | 0.71 | (1.89) | (1.00) | (3.45) | 0.53 | (4.11) | (−2.69) | (5.40) | 0.82 | (1.33) | −0.06 | 1.91 | 0.83 |

*Note:* The values that differ from ref of more than 10% are marked in bold and put between parentheses if they show a lower skill than Ref

Ref in simulating Prcp, produces larger summer warm bias than Ref. Both MP schemes show an improvement in summer, but MP1 produces larger winter cold bias. As for PBL schemes, PBL1 improves T2 simulation in winter but worsens it in summer. It is warmer than Ref in both seasons, which is in accordance with previous studies (Kleczek *et al.* 2014; Jänicke *et al.*, 2017; Xu *et al.*, 2019). Xu *et al.* (2019) suggests that this is because PBL1 considers turbulent exchange induced by steep terrain when calculating the turbulent diffusion coefficient. Besides, our results show that the nonlocal PBL2 scheme is colder than both local schemes, which is in contrast with some previous studies concluding that nonlocal schemes generally produce higher temperature than local schemes (Hu *et al.*, 2010; Xie *et al.*, 2012; Kleczek *et al.*, 2014). This different behaviour might due to the unique PBL characteristics over HMA compared to the plain regions (Yang *et al.*, 2004; Zou *et al.*, 2005; Chen *et al.*, 2013), which indicates that PPSs behave very differently in different regions, and the choice of PPSs always depends on the specific purpose. For LSM schemes, again, LSM1 does not show significant improvement. Noah-MP (LSM2) was reported to have better performance than Noah LSM (Ref) in previous studies (Yang *et al.*, 2011; Gao *et al.*, 2017) and has been widely applied in numerical modelling studies in HMA (Collier and Immerzeel, 2015; Karki *et al.*, 2017; Norris *et al.*, 2017). However, our results show the opposite, that is, that Noah-MP is colder than Noah LSM. This might result from our relatively short spin-up time. All the studies mentioned above use a spin-up time longer than 2 weeks. Noah-MP needs longer spin-up time than Noah LSM to reach a climatological equilibrium state (Cai *et al.*, 2014; Barlage *et al.*, 2015; Gao *et al.*, 2015).

## 3.2 | Combination run and sensitivity to initial snow depth

The results from Section 3.1 indicate that there is a trade-off in model performance between Prcp and T2, as well as between January and July. No single PPS performs ideally in both seasons and for both quantities. For instance, PBL2 has a better skill in predicting Prcp, but it worsens the winter cold bias. PBL1 shows superior performance in terms of T2 during winter with the lowest MAE, MB and RMSE among all experiments, but it also features the highest warm bias in summer. Therefore, we used the TOPSIS method to find the optimal solution for each experiment set, giving the same weight to each criterion (MAE, MB, RMSE and $r_s$) for each quantity (Prcp and T2). Table 4 lists the closeness coefficient (CC) and TOPSIS ranking for each experiment set. We then used
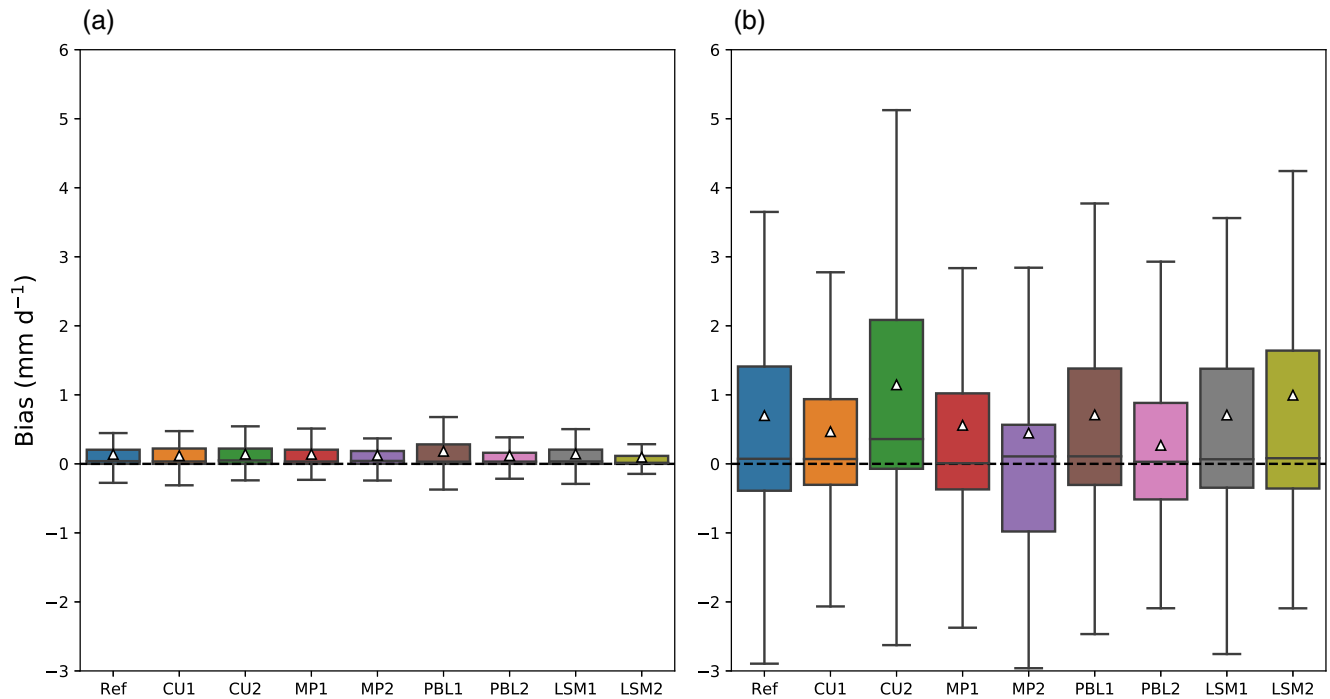
**FIGURE 3** Box-whisker plot of Prcp bias in January (a) and July (b) for all the 103 GSOD stations (red and white points in Figure 1b). White triangles represent the mean value [Colour figure can be viewed at wileyonlinelibrary.com]
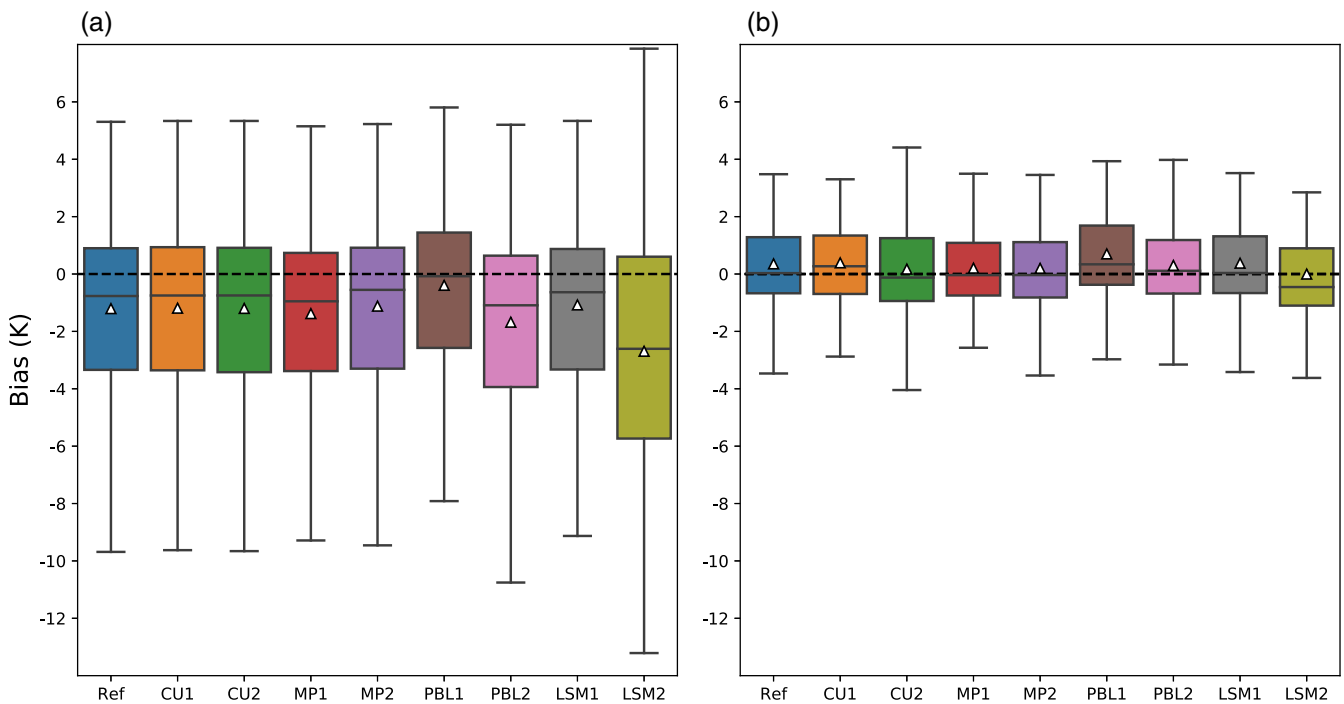


**FIGURE 4** Box-whisker plot of T2 bias in January (a) and July (b) for all the 103 GSOD stations (red and white points in Figure 1b). White triangles represent the mean value [Colour figure can be viewed at wileyonlinelibrary.com]

**TABLE 4** Closeness coefficient (CC) and TOPSIS ranking for each experiment set

|  | Ref | CU1 | CU2 | Ref | MP1 | MP2 | Ref | PBL1 | PBL2 | Ref | LSM1 | LSM2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CC | 0.56 | **0.61** | 0.38 | 0.27 | 0.60 | **0.88** | 0.46 | 0.47 | **0.53** | **0.51** | 0.50 | 0.50 |
| Ranking | 2 | **1** | 3 | 3 | 2 | **1** | 3 | 2 | **1** | **1** | 3 | 2 |

*Note:* The best schemes under each set are marked in bold.

**TABLE 5** Statistical scores for ref, COMB and COMB_S over 103 GSOD stations within d10km of the HAR v2 (red and white points in Figure 1b) for Prcp (mm·day⁻¹) and T2 (K)

| | Prcp January | | | | Prcp July | | | | T2 January | | | | T2 July | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE | MB | RMSE | $r_s$ | MAE | MB | RMSE | $r_s$ | MAE | MB | RMSE | $r_s$ | MAE | MB | RMSE | $r_s$ |
| Ref | 0.22 | 0.14 | 0.42 | 0.67 | 1.53 | 0.70 | 2.68 | 0.48 | 2.89 | −1.21 | 3.99 | 0.80 | 1.20 | 0.35 | 1.81 | 0.83 |
| COMB | **0.18** | **0.10** | **0.30** | 0.67 | **1.34** | **0.12** | 2.58 | 0.44 | 2.96 | (**−1.54**) | 4.15 | 0.82 | 1.31 | **0.29** | 1.91 | 0.80 |
| COMB_S | 0.22 | 0.14 | **0.35** | (**0.59**) | 1.41 | **0.22** | 2.80 | 0.48 | **1.65** | **0.17** | **2.68** | 0.83 | 1.28 | 0.33 | 1.89 | 0.79 |

*Note:* The values that differ from ref of more than 10% are marked in bold and put between parentheses if they show a lower skill than Ref.

the first ranking schemes, namely Kain-Fritsch cumulus potential scheme (CU1), Morrison 2-moment scheme (MP2), Yonsei University scheme (PBL2) and Noah LSM (Ref) to conduct a combination run (COMB). Scores in Table 5 show that COMB inherits from these first-ranking schemes. COMB has a better skill in predicting Prcp than Ref because CU1, MP2 and PBL2 all perform better than Ref in Prcp simulation. The wet bias is significantly reduced in COMB, especially in summer. Figures 5a–d shows the bias of monthly Prcp at each GSOD stations for Ref and COMB. The stations with high bias are located along the foothills of the Himalayas, where the actual precipitation amount is also large (Figure S2). Compared to Ref, COMB produces lower Prcp in India and Pakistan. COMB captures Prcp over the TP better, especially for stations in Southern TP, where Ref exhibits a dry bias. However, COMB does not improve the simulation of winter T2 (Figures 6a–d, Table 5), which is mainly due to the larger winter cold bias induced by PBL2 (Table 3).

To validate the assumption that the large winter cold bias is partly induced by overestimation of snow depth in the initial conditions, we conducted a further simulation (COMB_S). The PPSs used in this experiment are the same as in COMB, but initial snow depth and snow water equivalent were corrected using the method described in Section 2.4.

After the correction, the prediction of winter T2 is largely improved with MB of −1.54 K in COMB and MB of 0.17 K in COMB_S (Table 5). The impact of snow correction on summer T2 is minor compared to winter. Comparing the station-wise bias scores (Figure 6c,e), the most affected stations are located in the TP and in Northern Pakistan, where snow depth differs the most between ERA5 and JRA-55 (Section 4.2). This snow correction approach leads to slightly higher Prcp in both months (Table 5), but COMB_S still performs better than Ref.

## 3.3 | Comparison of HAR v2 with HAR

COMB_S achieves a significantly better performance than Ref (Table 5). Therefore, the set-up of COMB_S was applied to the HAR v2. We switched to a newer version of WRF (V4.1) for the generation of the HAR v2 and firstly produced HAR v2 d10km data for the whole year of 2011. Figure S3 shows that the change of model version only has a minor impact on the output.

We compared daily Prcp from the HAR v2 with the HAR and GSOD data (Figure 7). Here, only 57 GSOD stations within the d10km of the HAR are used (red points in Figure 1b). The HAR shows an overall lower MB and is better in explaining the variance in observations than
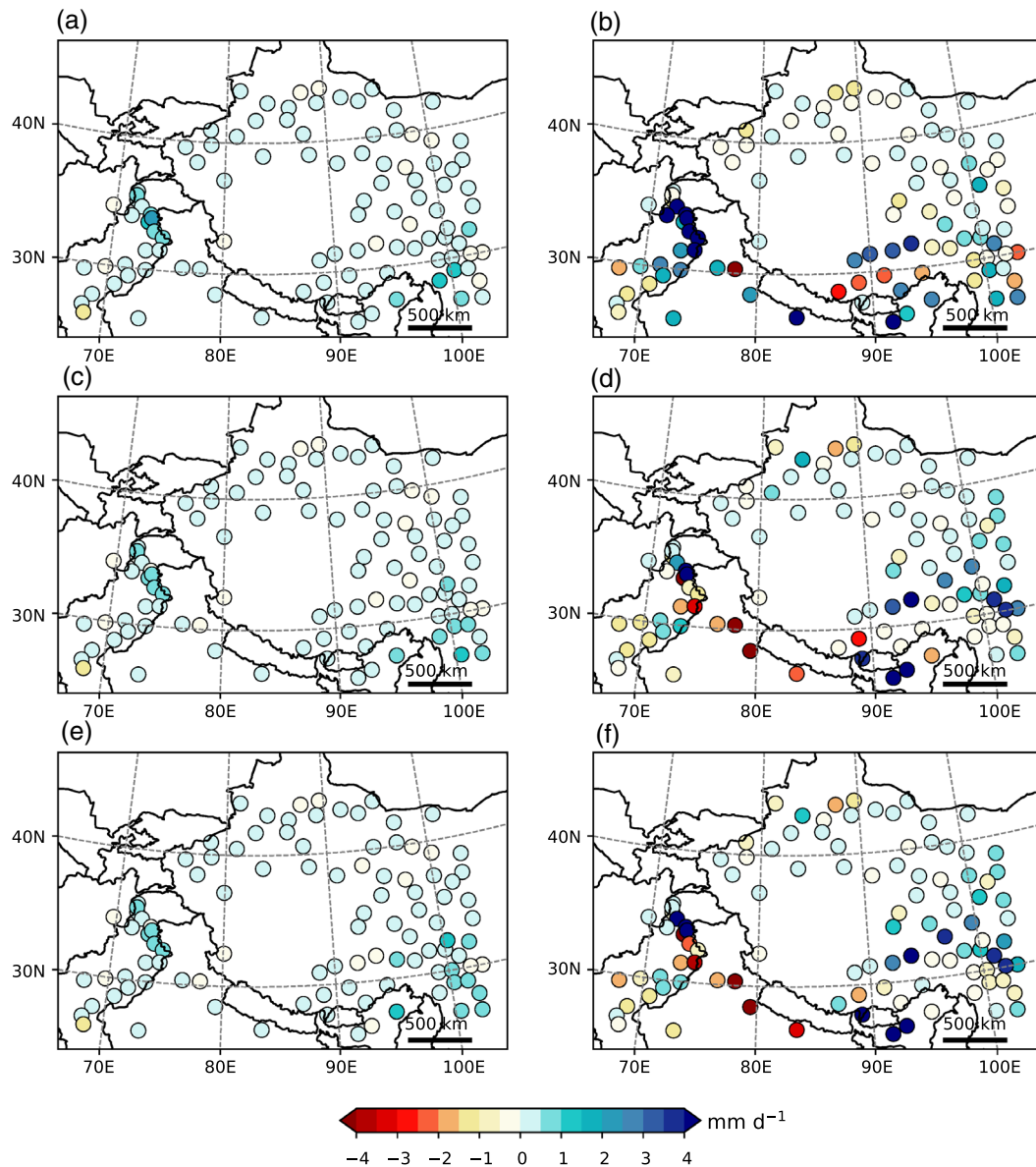
**FIGURE 5** Bias scores of monthly mean Prcp for ref (a and b), COMB (c and d) and COMB_S (e and f) in January (a, c and e) and July (b, d and f) at all 103 GSOD stations [Colour figure can be viewed at wileyonlinelibrary.com]

the HAR v2 (MB of 0.30 mm·day$^{-1}$ and 0.36 mm·day$^{-1}$, $r^2$ of .61 and .57). The HAR v2 tends to produce more Prcp than the HAR in April and from September to December. However, this kind of comparison only demonstrates the ability of the two versions of the HAR to reproduce the overall Prcp amount. In addition, we utilized gridded Prcp products to examine their ability to reproduce the spatial distribution of Prcp. Figure 8 shows seasonal Prcp from the HAR, the HAR v2 and the TMPA in 2011. In winter, the HAR and the HAR v2 produce similar spatial patterns of Prcp with maximum Prcp over Pamir-Karakoram-western-Himalayas (PKwH) region. However, this maximum centre is not detected in the TMPA. In MAM, two Prcp maxima exist in both versions

of the HAR: one in PKwH and one over the region near Brahmaputra River. Only the latter one is visible in the TMPA, but with lower Prcp amount. The HAR and the HAR v2 show the largest differences in JJA. The HAR v2 shares more spatial similarity with the TMPA, while the HAR produces higher Prcp amount in Indian, Pakistan and eastern TP. In SON, the HAR still shows higher Prcp amount in Indian. Same as in winter and spring, both versions of the HAR detect more Prcp in PKwH than the TMPA.

Daily T2 from the HAR and the HAR v2 are compared with each other and with GSOD (Figure 9). Both data sets reproduce T2 seasonality well ($r^2$ = .99). The HAR v2 simulates T2 better with MB of −0.58 K
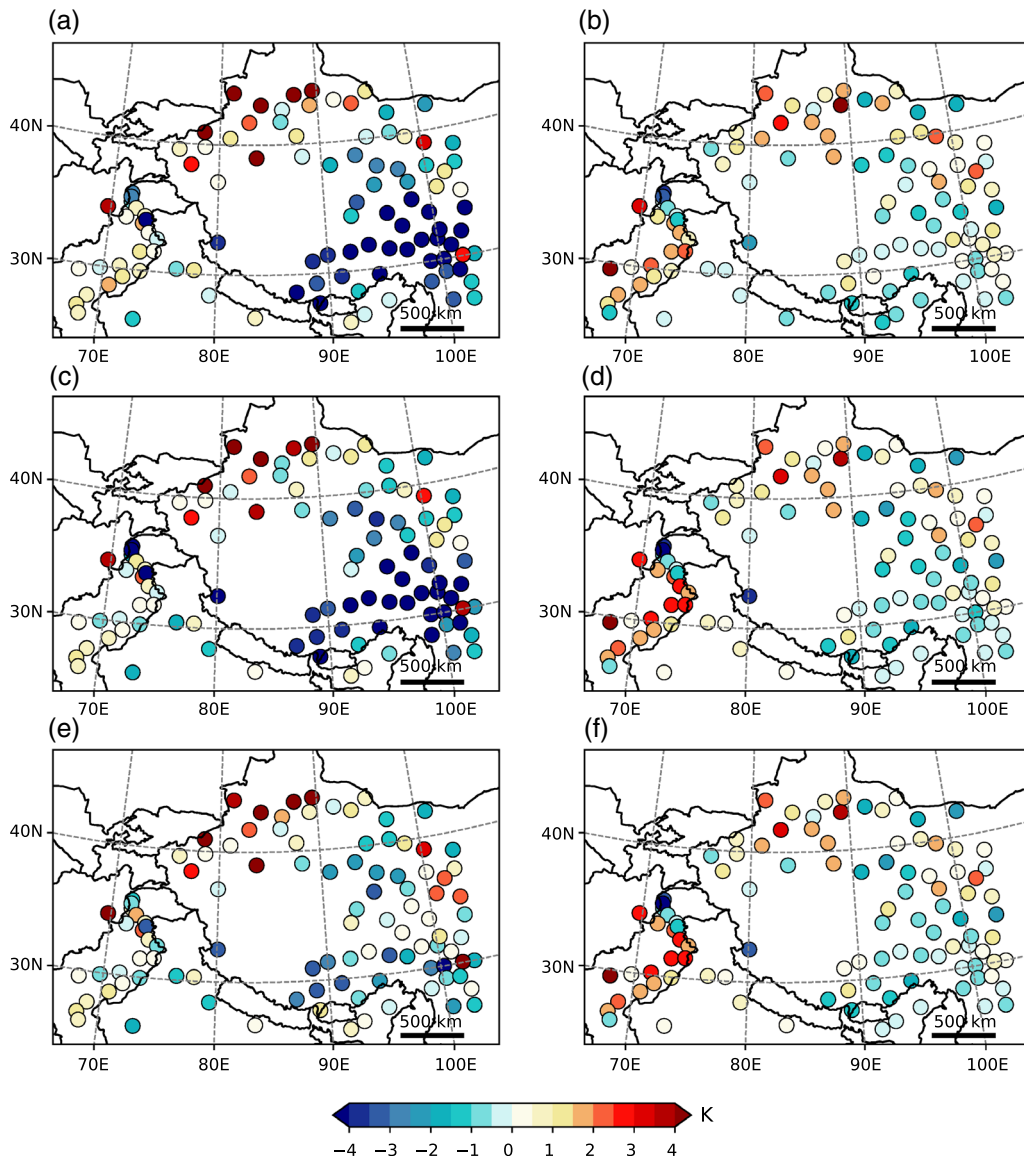
**FIGURE 6** Bias scores of monthly mean T2 for ref (a and b), COMB (c and d) and COMB_S (e and f) in January (a, c and e) and July (b, d and f) at all 103 GSOD stations [Colour figure can be viewed at wileyonlinelibrary.com]

compared to MB of −0.86 K for the HAR. The HAR has a cold bias in spring, autumn and winter but a warm bias in summer. The HAR v2 has the same pattern, but the magnitudes of these biases are smaller.

# 4 | DISCUSSION

## 4.1 | Sources of uncertainties

One should always keep in mind that the accuracy of observational data and the method used for validation could lead to uncertainty of the result. In this study, we use freely available observations from GSOD as the ground truth to calculate statistical metrics. Station data

from this data set underwent quality control before being published (NOAA, 2019). However, some GSOD stations seem to be outliers, for example, the station in southeastern TP featuring winter warm bias, while all the surrounding stations show cold bias (Figure 6a,c,e). With regards to Prcp, it is well known that rain gauges undercatch Prcp under strong wind conditions (Duchon and Essenberg, 2001; WMO, 2008). We used T2 and Prcp from the nearest grid point to compare them with observations. With a resolution of 10 km, the distance between the actual location of stations and the associated grid point could reach a few kilometres, which leads to a large difference in elevation over complex terrain. A constant lapse rate of −6.5 K·km$^{-1}$ is applied to correct the simulated T2 values. For Prcp, no correction is applied.
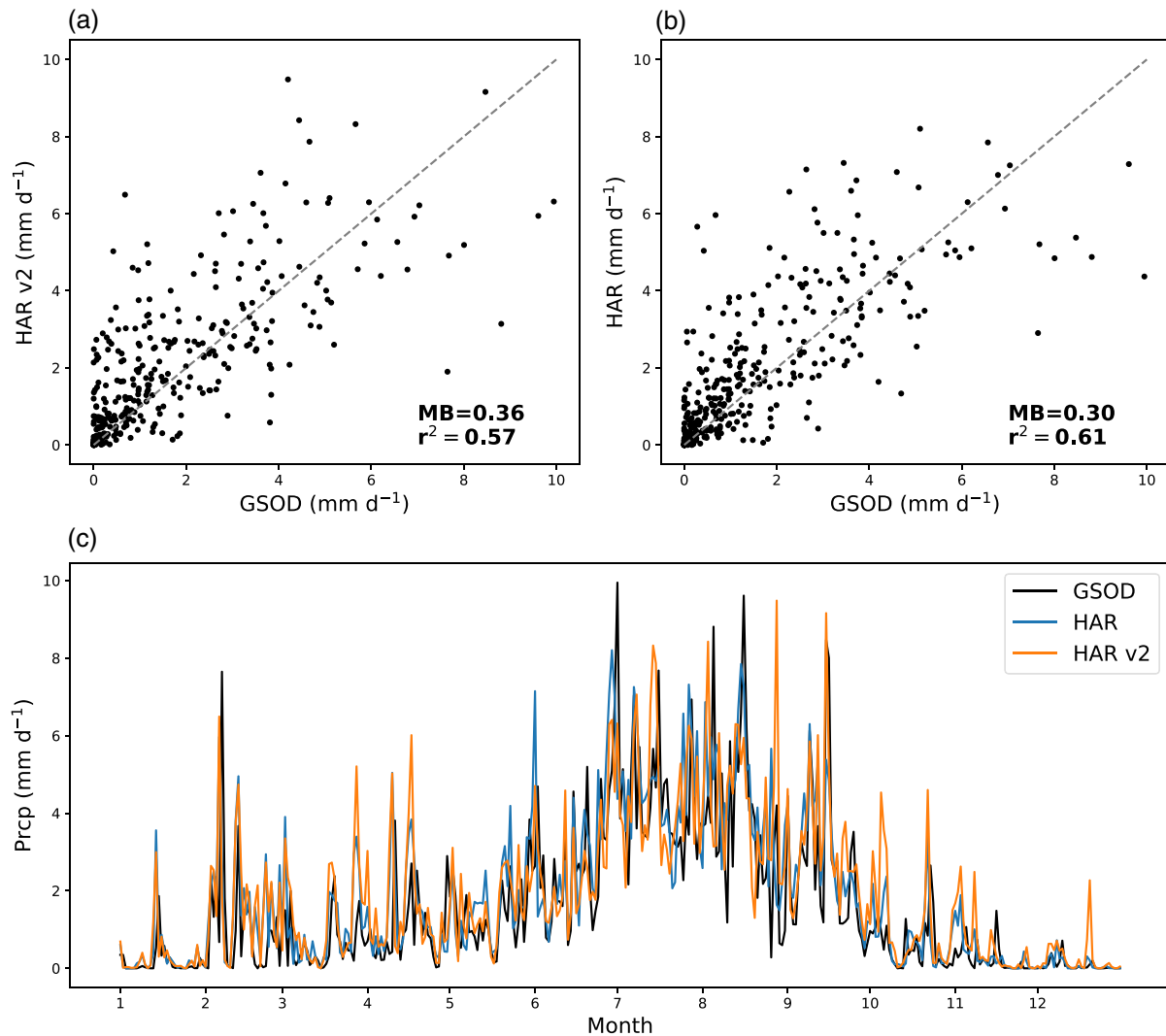
**FIGURE 7**  Comparison of daily Prcp from HAR, HAR v2 and GSOD averaged over 57 GSOD stations (red points in Figure 1b) in 2011. Scatter plots of HAR v2 (a) and HAR (b) with statistical scores MB in mm day$^{-1}$ and r$^2$. Daily Prcp time series (c) [Colour figure can be viewed at wileyonlinelibrary.com]

Therefore, the point-to-point comparison between modelled and observed Prcp always depicts discrepancies. Another source of uncertainty comes from the fact that we only conducted sensitivity experiments for the year 2011 due to the high computational costs.

Due to the sparse distribution of GSOD stations, we also compared the spatial distribution of seasonal Prcp from two versions of the HRA with the TMPA. The results (Figure 8) indicate that the TMPA fails to reproduce Prcp over PKwH in winter, spring and autumn. This region is dominated by mid-latitude westerlies and experiences its maximum Prcp amount in winter (Archer and Fowler, 2004; Bookhagen and Burbank, 2010; Curio and Scherer, 2016; Mölg et al., 2018). The majority of Prcp in this region falls as snow (Maussion et al., 2014). Satellite-based Prcp has a deficiency in detecting snowfall (Behrangi et al., 2014; Immerzeel et al., 2015). The Global

Precipitation Measurement (GPM) mission, launched in February 2014, is a successor of TRMM and a next generation of global observations of precipitation from space. The onboard Dual-frequency Precipitation Radar is more sensitive than TRMM in detecting snowfall and light rainfall (NASA, 2019). GPM-based products are expected to be a better validation tool for modelled results after 2014.

## 4.2 | Snow correction approach

In this study, we propose a bias correction method for initial snow depth and snow water equivalent based on the concept of linear scaling approach. This approach has the advantage that only monthly climatological information is required. Figure 6 and Table 5 show that bias-
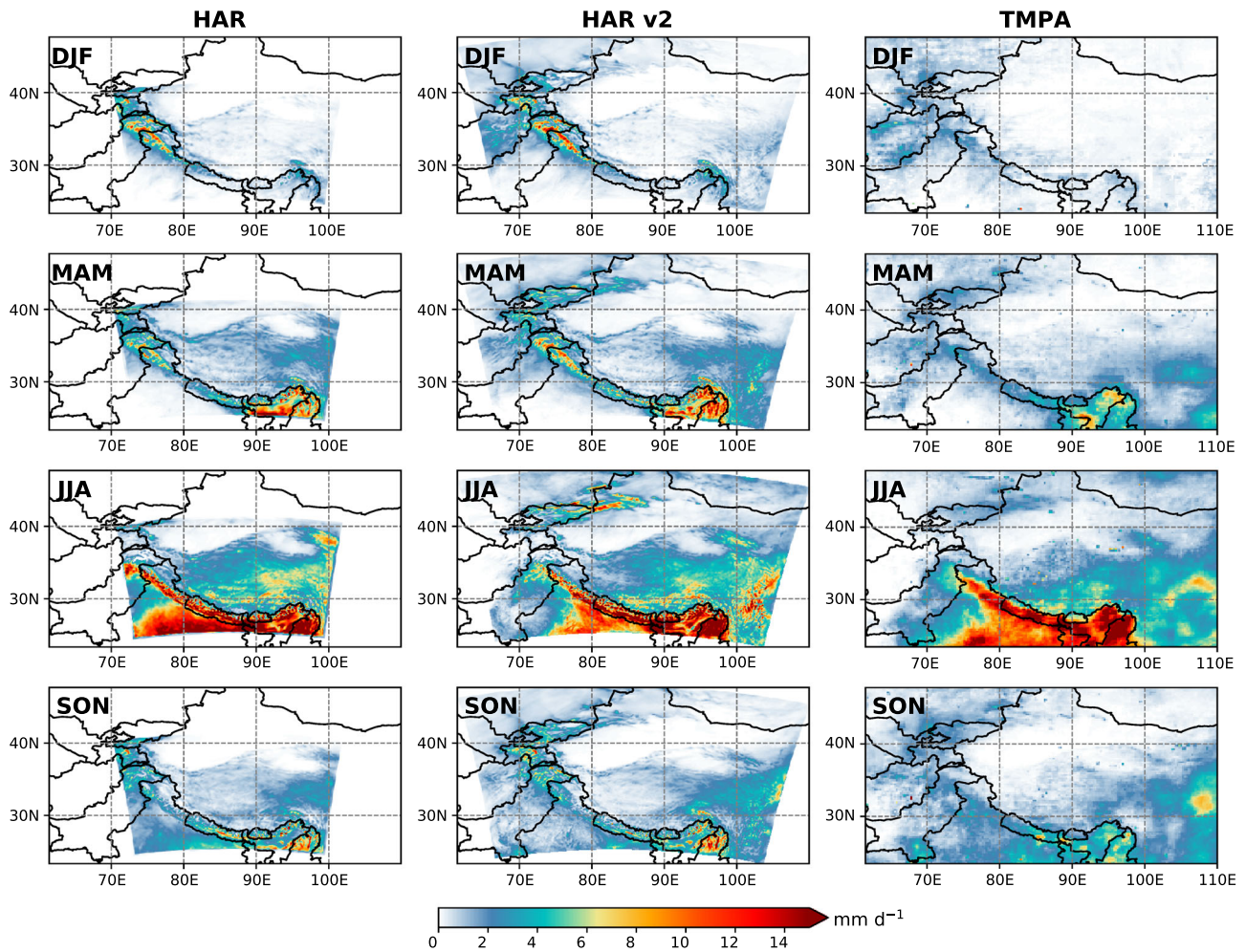
**FIGURE 8**    Seasonal mean Prcp from HAR, HAR v2 and TMPA in 2011 [Colour figure can be viewed at wileyonlinelibrary.com]

corrected initial conditions improve winter T2 simulations.

To further explore the influence of snow depth on T2, monthly mean difference of T2, surface albedo and snow cover fraction between COMB and COMB_S in January and July are presented in Figure 10. In January, regions with the highest increase of T2 in COMB_S are Southeastern TP, the Himalayas, and the Tian Shan. These regions also feature lower albedo and lower snow cover fraction. The same relationship can be found in July. However, the increase of T2 is much weaker in July than in January, which is in accordance with the MB scores in Table 5.

In the Noah LSM, surface albedo ($\alpha$) is calculated as:

$$\alpha = \alpha_{sn,free} + f_{sn} \cdot (\alpha_{sn} - \alpha_{sn,free}) \quad (11)$$

where $\alpha_{sn,\,free}$ and $\alpha_{sn}$ refer to snow-free albedo and snow albedo, respectively. $f_{sn}$ is snow cover fraction. $\alpha$ is proportional to $f_{sn}$, which is defined as:

$$f_{sn} = \begin{cases} 1 - e^{-a_s \frac{W}{W_{cr}}} + \frac{W}{W_{cr}} e^{-a_s}, & W < W_{cr} \\ 1, & W \geq W_{cr} \end{cases} \quad (12)$$

In Equation 12, $W$ is snow depth in water equivalent; $W_{cr}$ is a land-use-dependent threshold of snow depth, over which $f_{sn}$ is set to 1; $a_s$ is a distribution shape parameter. After snow correction, $W$ decreases in most grid points in d10km (Figure 2), which leads to smaller $f_{sn}$ according to Equation 12. Smaller $f_{sn}$ results in lower surface albedo (Equation 11) and modifies the surface energy balance by reflecting less short-wave radiation, and thus, influences T2 in COMB_S. On the other hand, more absorption of short-wave radiation could lead to larger moist static energy (Meng et al., 2014), which could be a possible reason of the enhanced precipitation in COMB_S (Table 5). Note that snow-related changes affect T2 and Prcp through multiple complex processes. Further work is needed to quantify the impact of these processes on T2 and Prcp.
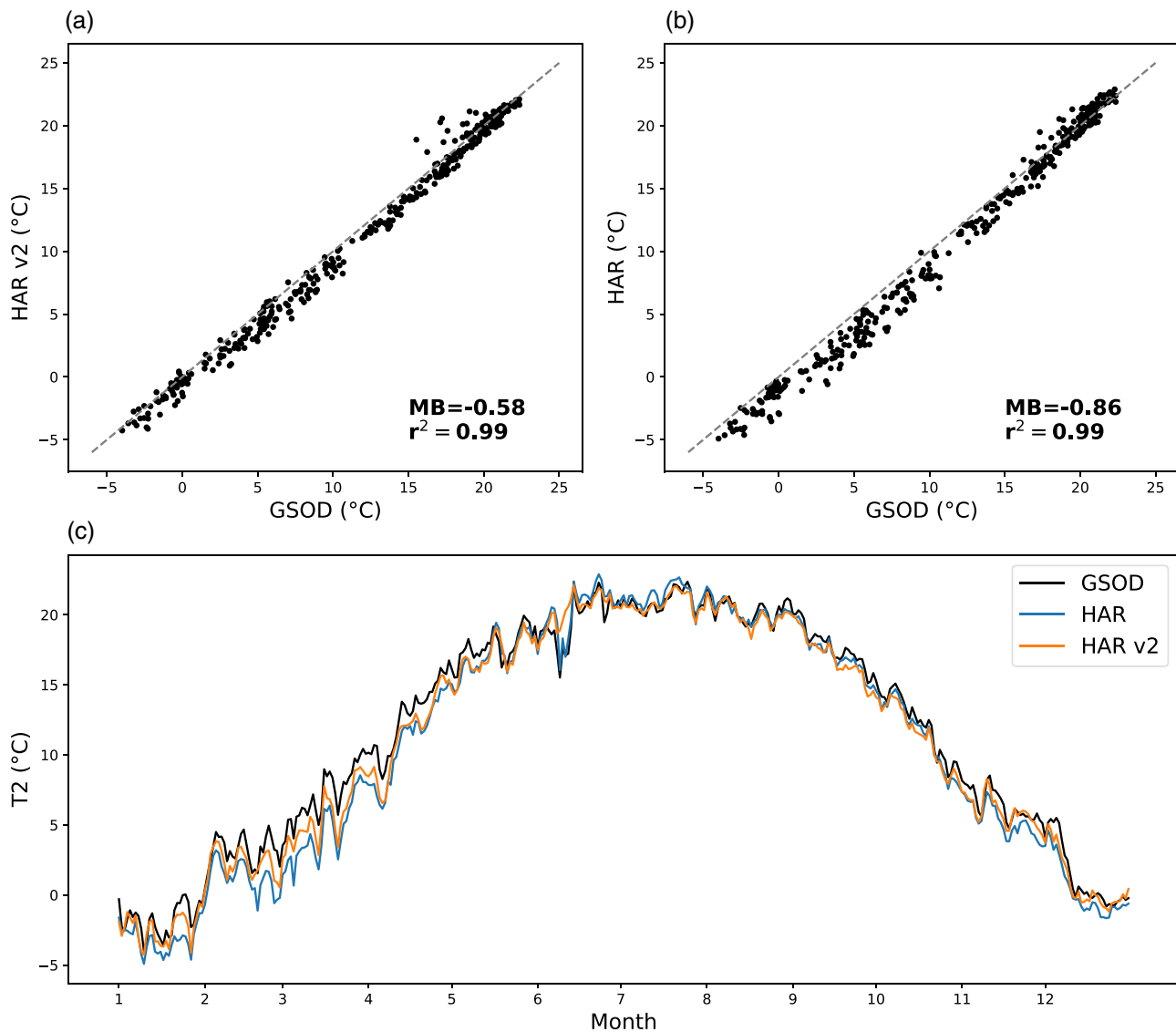
**FIGURE 9** Comparison of daily T2 from HAR, HAR v2 and GSOD averaged over 57 GSOD stations (red points in Figure 1b) in 2011. Scatter plots of HAR v2 (a) and HAR (b) with statistical scores MB in K and $r^2$. Daily T2 time series (c) [Colour figure can be viewed at wileyonlinelibrary.com]

The error sources of dynamical downscaling are two-fold: deficiencies in model dynamics and physics, as well as inaccuracies in forcing data. After snow correction, the cold bias over TP is reduced but not eliminated. The distribution of T2 bias score in COMB_S (Figure 6e,f) share some similar pattern with the HAR (figure 2.4 in Maussion, 2014): cold bias over the TP but warm bias in Pakistan and Northwestern China. This implies that these patterns of biases are related to WRF itself, or the same errors exist in both ERA5 and FNL data sets. Cold bias over the TP is reported from many other WRF studies independent of the forcing data (ERA-interim: Gao *et al.*, 2015; Karki *et al.*, 2017; Bonekamp *et al.*, 2018; Huang and Gao, 2018; FNL: Huang and Gao, 2018; Maussion, 2014; Zhou *et al.*, 2018). Meng *et al.* (2018)

pointed out that this is due to the overestimation of albedo over snow-covered areas in Noah LSM. After switching off albedo parameterization and replacing the albedo with MODIS time-varying albedo, their simulated T2 improved. According to Chen *et al.* (2014), the major weakness of LSM is snow processes, and Noah LSM needs a higher snow albedo to retain snow on ground since it only considers a single layer of snowpack.

## 5 | CONCLUSIONS

Validation of the reference experiment (Ref) indicates that, due to the changes in forcing data and domain configuration, the PPSs used in the HAR are no longer
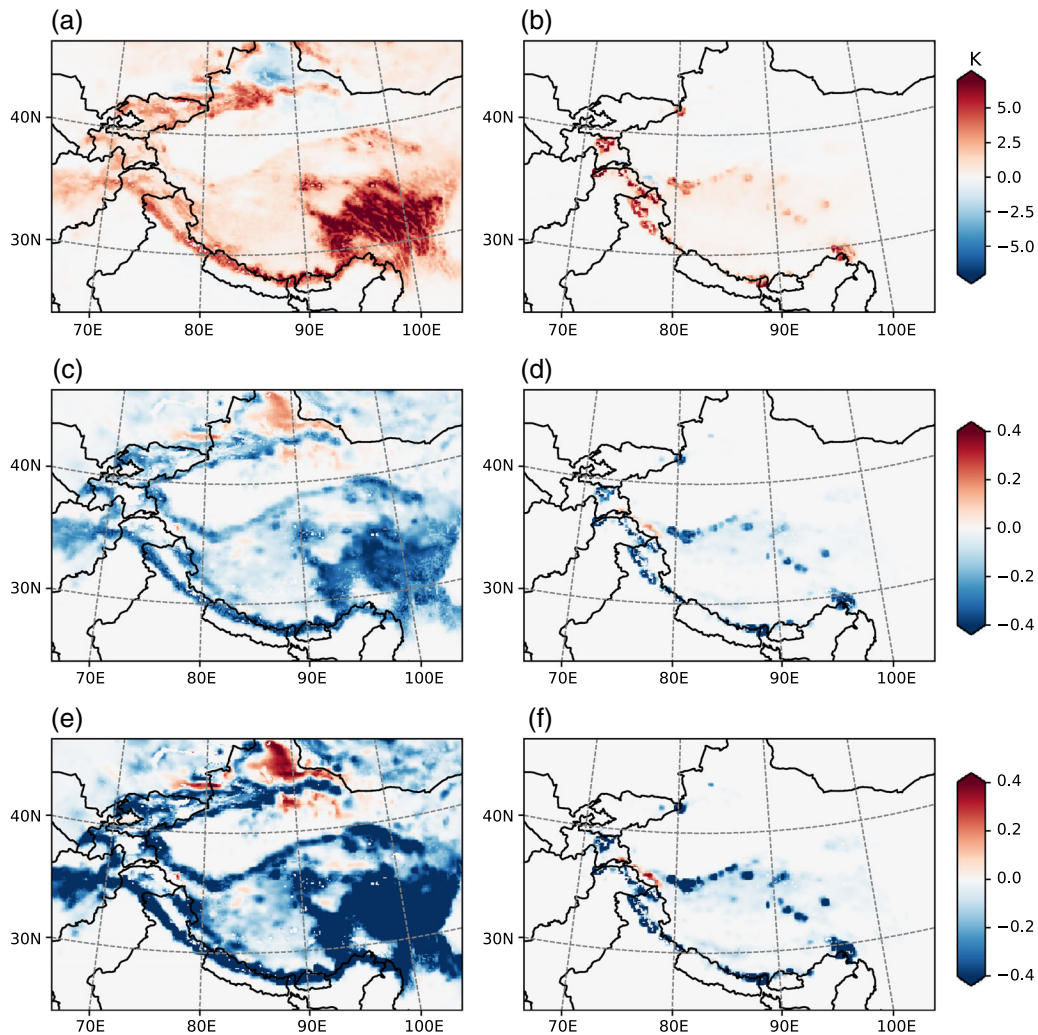
**FIGURE 10** Difference of T2 (a and b), surface albedo (c and d) and snow cover fraction (e and f) between COMB_s and COMB in January (a, c and e) and July (b, d and f) [Colour figure can be viewed at wileyonlinelibrary.com]

suitable for the HAR v2 since they produce large summer wet bias and winter cold bias. We examined model sensitivities to different PPSs and initial snow conditions to find an optimal set-up for the HAR v2. The results of PPS experiments reveal that no single PPS is optimally suited for both seasons (January and July) and both quantities (T2 and Prcp). Trade-offs always exist between quantities and between seasons.

Kain-Fritsch cumulus potential scheme, Morrison 2-moment scheme, Yonsei University scheme and Noah LSM are identified by the TOPSIS method as the best schemes. The combination run (COMB) inherits from these schemes and has a superior performance in Prcp but not in T2 when compared to Ref.

Overestimation of snow depth in ERA5 in most areas of d10km leads to higher snow cover fraction and higher surface albedo, which ultimately results in a cold bias. After correction of initial snow depth, cold bias is significantly reduced but not eliminated. This could imply a

deficiency of Noah LSM. In future LSM development, improvement of representation of snow-related processes is needed.

After careful selection of PPSs and correction of initial snow depth, model performance is improved. The improved set-up was applied to produce 1 year of HAR v2 data for 2011. Compared to the old version, the HAR v2 generally produces slightly higher Prcp amounts, but the spatial distribution of seasonal Prcp matches better to observations. Both versions of the HAR show cold bias in spring, autumn and winter, but warm bias in summer. The HAR v2 has smaller magnitudes of these biases.

The HAR v2 is planned to have a temporal coverage of at least 30 years. Comprehensive validation against observations and comparison with the HAR are scheduled once data production is finished. The HAR v2 is developed within the framework of the "Climatic and Tectonic Natural Hazards in Central Asia (CaTeNA)" project to investigate the climatic triggering mechanism

of landslides in Central Asia. However, application of this data set will not be limited to this research field. With high resolution and extended temporal coverage, the HAR v2 can contribute to a better understanding of climate-related processes in the remote and data-sparse HMA region.

## ORCID

*Xun Wang* https://orcid.org/0000-0003-3294-747X
*Marco Otto* https://orcid.org/0000-0003-4464-2673
*Dieter Scherer* https://orcid.org/0000-0002-3670-0864

## REFERENCES

Archer, D.R. and Fowler, H.J. (2004) Spatial and temporal variations in precipitation in the upper Indus Basin, global teleconnections and hydrological implications. *Hydrology and Earth System Sciences*, 8, 47–61.

Barlage, M., Tewari, M., Chen, F., Miguez-Macho, G., Yang, Z.L. and Niu, G.Y. (2015) The effect of groundwater interaction in north American regional climate simulations with WRF/Noah-MP. *Climatic Change*, 129, 485–498.

Behrangi, A., Tian, Y., Lambrigtsen, B.H. and Stephens, G.L. (2014) What does CloudSat reveal about global land precipitation detection by other spaceborne sensors? *Water Resources Research*, 50, 4893–4905.

Berg, L.K., Gustafson, W.I., Jr., Kassianov, E.I. and Deng, L. (2013) Evaluation of a modified scheme for shallow convection: implementation of CuP and case studies. *Monthly Weather Review*, 141, 134–147.

Bonekamp, P.N.J., Collier, E. and Immerzeel, W.W. (2018) The impact of spatial resolution, land use, and spinup time on resolving spatial precipitation patterns in the Himalayas. *Journal of Hydrometeorology*, 19, 1565–1581.

Bookhagen, B. and Burbank, D.W. (2010) Toward a complete Himalayan hydrological budget: spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge. *Journal of Geophysical Research: Earth Surface*, 115, F03019.

Bougeault, P. and Lacarrere, P. (1989) Parameterization of orography-induced turbulence in a mesobeta—scale model. *Monthly Weather Review*, 117, 1872–1890.

Cai, X., Yang, Z.L., David, C.H., Niu, G.Y. and Rodell, M. (2014) Hydrological evaluation of the Noah-MP land surface model for the Mississippi River basin. *Journal of Geophysical Research: Atmospheres*, 119, 23–38.

Chai, T. and Draxler, R.R. (2014) Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7, 1247–1250.

Chen, F., Barlage, M., Tewari, M., Rasmussen, R., Jin, J., Lettenmaier, D., Livneh, B., Lin, C., Miguez-Macho, G., Niu, G. Y., Wen, L. and Yang, Z.L. (2014) Modeling seasonal snowpack evolution in the complex terrain and forested Colorado headwaters region: a model intercomparison study. *Journal of Geophysical Research: Atmospheres*, 119, 13–795.

Chen, S.H. and Sun, W.Y. (2002) A one-dimensional time dependent cloud model. *Journal of the Meteorological Society of Japan*, 80, 99–118.

Chen, X., Anel, J.A., Su, Z., de la Torre, L., Kelder, H., van Peet, J. and Ma, Y. (2013) The deep atmospheric boundary layer and its significance to the stratosphere and troposphere exchange over the Tibetan plateau. *PLoS One*, 8(2).e56909.

Collier, E. and Immerzeel, W.W. (2015) High-resolution modeling of atmospheric dynamics in the Nepalese Himalaya. *Journal of Geophysical Research: Atmospheres*, 120, 9882–9896.

Copernicus Climate Change Service (C3S) (2017) *ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS)*. Available at https://cds.climate.copernicus.eu/cdsapp#!/home [Accessed 10th November 2018]

Curio, J. and Scherer, D. (2016) Seasonality and spatial variability of dynamic precipitation controls on the Tibetan plateau. *Earth System Dynamics*, 7, 767–782.

Curio, J., Maussion, F. and Scherer, D. (2015) A 12-year high-resolution climatology of atmospheric water transport over the Tibetan plateau. *Earth System Dynamics*, 6, 109–124.

Dai, F.C. and Lee, C.F. (2002) Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology*, 42, 213–228. https://doi.org/10.1016/S0169-555X(01)00087-3.

Duchon, C.E. and Essenberg, G.R. (2001) Comparative rainfall observations from pit and aboveground rain gauges with and without wind shields. *Water Resources Research*, 37, 3253–3263.

Dudhia, J. (1989) Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences*, 46, 3077–3107.

Feser, F., Rockel, B., von Storch, H., Winterfeldt, J. and Zahn, M. (2011) Regional climate models add value to global model data: a review and selected examples. *Bulletin of the American Meteorological Society*, 92, 1181–1192.

Gao, Y., Xiao, L., Chen, D., Chen, F., Xu, J. and Xu, Y. (2017) Quantification of the relative role of land-surface processes and large-scale forcing in dynamic downscaling over the Tibetan plateau. *Climate Dynamics*, 48, 1705–1721.

Gao, Y., Xu, J. and Chen, D. (2015) Evaluation of WRF mesoscale climate simulations over the Tibetan plateau during 1979–2011. *Journal of Climate*, 28, 2823–2841.

Grell, G.A. (1993) Prognostic evaluation of assumptions used by cumulus parameterizations. *Monthly Weather Review*, 121, 764–787.

Grell, G.A. and Devenyi, D. (2002) A generalized approach to parameterizing convection combining ensemble and data

assimilation techniques. *Geophysical Research Letters*, 29, 38–31.

Grell, G.A. and Freitas, S.R. (2014) A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmospheric Chemistry and Physics*, 14, 5233–5250.

Hasson, S., Gerlitz, L., Schickhoff, U., Scholten, T. and Böhner, J. (2016) Recent climate change over High Asia. In: *Climate Change, Glacier Response, and Vegetation Dynamics in the Himalaya*. Cham, Switzerland: Springer, pp. 29–48. https://doi.org/10.1007/978-3-319-28977-9_2.

Hong, S.Y., Noh, Y. and Dudhia, J. (2006) A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, 134, 2318–2341.

Hong, Y., Adler, R. and Huffman, G. (2007) Use of satellite remote sensing data in the mapping of global landslide susceptibility. *Natural Hazards*, 43, 245–256.

Hu, X.M., Nielsen-Gammon, J.W. and Zhang, F. (2010) Evaluation of three planetary boundary layer schemes in the WRF model. *Journal of Applied Meteorology and Climatology*, 49, 1831–1844.

Huang, D. and Gao, S. (2018) Impact of different reanalysis data on WRF dynamical downscaling over China. *Atmospheric Research*, 200, 25–35.

Huffman, G.J. and Bolvin, D.T. (2015) Real-time TRMM multi-satellite precipitation analysis data set documentation. *NASA TechDoc*.

Huintjes, E., Sauter, T., Schröter, B., Maussion, F., Yang, W., Kropáček, J., Buchroithner, M., Scherer, D., Kang, S. and Schneider, C. (2015) Evaluation of a coupled snow and energy balance model for Zhadang glacier, Tibetan plateau, using glaciological measurements and time-lapse photography. *Arctic, Antarctic, and Alpine Research*, 47, 573–590.

Hwang, C.L. and Yoon, K.P. (1981) Multiple attribute decision making. In: *Methods and Applications: A State-of-the-Art Survey*, Vol. 186. Berlin, Germany: Springer.

ICIMOD. (2011) *Glacial Lakes and Glacial Lake Outburst Floods in Nepal, Technical Report*. Kathmandu, Nepal: ICIMOD.

Immerzeel, W.W., Wanders, N., Lutz, A.F., Shea, J.M. and Bierkens, M.F.P. (2015) Reconciling high-altitude precipitation in the upper Indus basin with glacier mass balances and runoff. *Hydrology and Earth System Sciences*, 19, 4673–4687.

Jänicke, B., Meier, F., Fenner, D., Fehrenbach, U., Holtmann, A. and Scherer, D. (2017) Urban–rural differences in near-surface air temperature as resolved by the Central Europe refined analysis (CER): sensitivity to planetary boundary layer schemes and urban canopy models. *International Journal of Climatology*, 37, 2063–2079.

Janjic, Z.I. (1994) The step-mountain eta coordinate model: further developments of the convection, viscous sublayer, and turbulence closure schemes. *Monthly Weather Review*, 122, 927–945.

Jena, P., Azad, S. and Rajeevan, M. (2015) Statistical selection of the optimum models in the CMIP5 dataset for climate change projections of Indian monsoon rainfall. *Climate*, 3, 858–875.

Jiménez, P.A., Dudhia, J., González-Rouco, J.F., Navarro, J., Montávez, J.P. and García-Bustamante, E. (2012) A revised scheme for the WRF surface layer formulation. *Monthly Weather Review*, 140, 898–918.

Kala, J., Andrys, J., Lyons, T.J., Foster, I.J. and Evans, B.J. (2015) Sensitivity of WRF to driving data and physics options on a seasonal time-scale for the southwest of Western Australia. *Climate Dynamics*, 44, 633–659.

Karki, R., Gerlitz, L., Schickhoff, U., Scholten, T. and Böhner, J. (2017) Quantifying the added value of convection-permitting climate simulations in complex terrain: a systematic evaluation of WRF over the Himalayas. *Earth System Dynamics*, 8, 507–528.

Kirschbaum, D., Adler, R., Adler, D., Peters-Lidard, C. and Huffman, G. (2012) Global distribution of extreme precipitation and high-impact landslides in 2010 relative to previous years. *Journal of Hydrometeorology*, 13, 1536–1551.

Kleczek, M.A., Steeneveld, G.J. and Holtslag, A.A. (2014) Evaluation of the weather research and forecasting mesoscale model for GABLS3: impact of boundary-layer schemes, boundary conditions and spin-up. *Boundary-Layer Meteorology*, 152, 213–243.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K. and Takahashi, K. (2015) The JRA-55 reanalysis: general specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, 93, 5–48.

Kropáček, J., Neckel, N., Tyrna, B., Holzer, N., Hovden, A., Gourmelen, N., Schneider, C., Buchroithner, M. and Hochschild, V. (2015) Repeated glacial lake outburst flood threatening the oldest Buddhist monastery in North-Western Nepal. *Natural Hazards and Earth System Sciences*, 15, 2425–2437.

Leduc, M. and Laprise, R. (2009) Regional climate model sensitivity to domain size. *Climate Dynamics*, 32, 833–854.

Lenderink, G., Buishand, A. and Deursen, W.V. (2007) Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach. *Hydrology and Earth System Sciences*, 11, 1145–1159.

Leung, L.R., Qian, Y. and Bian, X. (2003) Hydroclimate of the western United States based on observations and regional climate simulation of 1981–2000. Part I: seasonal statistics. *Journal of Climate*, 16, 1892–1911.

Li, D., Bou-Zeid, E., Barlage, M., Chen, F. and Smith, J.A. (2013) Development and evaluation of a mosaic approach in the WRF-Noah framework. *Journal of Geophysical Research: Atmospheres*, 118, 11–918.

Li, D., Yang, K., Tang, W., Li, X., Zhou, X. and Guo, D. (2020) Characterizing precipitation in high altitudes of the western Tibetan plateau with a focus on major glacier areas. *International Journal of Climatology*.

Li, J., Liu, Z., Yao, Z. and Wang, R. (2019) Comprehensive assessment of coupled model Intercomparison project phase 5 global climate models using observed temperature and precipitation over mainland Southeast Asia. *International Journal of Climatology*, 39(10), 4139–4153.

Lo, J.C.F., Yang, Z.L. and Pielke, R.A., Sr. (2008) Assessment of three dynamical climate downscaling methods using the weather research and forecasting (WRF) model. *Journal of Geophysical Research: Atmospheres*, 113, D09112.

Maussion, F. (2014) *A new atmospheric dataset for High Asia: development, validation and applications in climatology and in glaciology*. PhD Thesis, Techinische Universität Berlin, Germany, http://doi.org/10.14279/depositonce-3979.

Maussion, F., Scherer, D., Finkelnburg, R., Richters, J., Yang, W. and Yao, T. (2011) WRF simulation of a precipitation event

over the Tibetan plateau, China—an assessment using remote sensing and ground observations. *Hydrology and Earth System Sciences*, 15, 1795–1817.

Maussion, F., Scherer, D., Mölg, T., Collier, E., Curio, J. and Finkelnburg, R. (2014) Precipitation seasonality and variability over the Tibetan plateau as resolved by the high Asia reanalysis. *Journal of Climate*, 27, 1910–1927.

Meng, X.H., Evans, J.P. and McCabe, M.F. (2014) The influence of inter-annually varying albedo on regional climate and drought. *Climate Dynamics*, 42, 787–803.

Meng, X., Lyu, S., Zhang, T., Zhao, L., Li, Z., Han, B., Li, S., Ma, D., Chen, H., Ao, Y., Luo, S., Shen, Y., Guo, J. and Wen, L. (2018) Simulated cold bias being improved by using MODIS time-varying albedo in the Tibetan plateau in WRF model. *Environmental Research Letters*, 13, 044028.

Miguez-Macho, G., Stenchikov, G.L. and Robock, A. (2004) Spectral nudging to eliminate the effects of domain position and geometry in regional climate model simulations. *Journal of Geophysical Research: Atmospheres*, 109, D13104.

Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J. and Clough, S.A. (1997) Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, 102, 16663–16682.

Mölg, N., Bolch, T., Rastner, P., Strozzi, T. and Paul, F. (2018) A consistent glacier inventory for Karakoram and Pamir derived from Landsat data: distribution of debris cover and mapping challenges. *Earth System Science Data*, 10, 1807–1827.

Mölg, T., Maussion, F. and Scherer, D. (2014) Mid-latitude westerlies as a driver of glacier variability in monsoonal high Asia. *Nature Climate Change*, 4, 68–73.

Morrison, H., Thompson, G. and Tatarskii, V. (2009) Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: comparison of one-and two-moment schemes. *Monthly Weather Review*, 137, 991–1007.

NASA (n.d.) *Precipitation measurement missions*. Available at https://pmm.nasa.gov/gpm/flight-project/dpr [Accessed 30th August 2019].

Niu, G.Y., Yang, Z.L., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M. and Xia, Y. (2001) The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres*, 116, D12109.

NOAA (2019) *Integrated Surface Database (ISD)*. Available at https://www.ncdc.noaa.gov/isd [Accessed 30th August 2019].

Norris, J., Carvalho, L.M., Jones, C., Cannon, F., Bookhagen, B., Palazzi, E. and Tahir, A.A. (2017) The spatiotemporal variability of precipitation over the Himalaya: evaluation of one-year WRF model simulation. *Climate Dynamics*, 49, 2179–2204.

Orr, A., Listowski, C., Couttet, M., Collier, E., Immerzeel, W., Deb, P. and Bannister, D. (2017) Sensitivity of simulated summer monsoonal precipitation in Langtang Valley, Himalaya, to cloud microphysics schemes in WRF. *Journal of Geophysical Research*, 122(12), 6298–6318.

Orsolini, Y., Wegmann, M., Dutra, E., Liu, B., Balsamo, G., Yang, K., de Rosnay, P., Zhu, C., Wang, W. and Senan, R. (2019) Evaluation of snow depth and snow-cover over the Tibetan plateau in global reanalyses using in-situ and satellite remote sensing observations. *The Cryosphere*, 13, 2221–2239.

Pritchard, D.M., Forsythe, N., Fowler, H.J., O'Donnell, G.M. and Li, X.F. (2019) Evaluation of upper Indus near-surface climate representation by WRF in the high Asia refined analysis. *Journal of Hydrometeorology*, 20, 467–487.

Qian, Y., Yan, H., Berg, L.K., Hagos, S., Feng, Z., Yang, B. and Huang, M. (2016) Assessing impacts of PBL and surface layer schemes in simulating the surface–atmosphere interactions and precipitation over the tropical ocean using observations from AMIE/DYNAMO. *Journal of Climate*, 29(22), 8191–8210.

Raju, K.S. and Kumar, D.N. (2014) Ranking general circulation models for India using TOPSIS. *Journal of Water and Climate Change*, 6, 288–299.

Roessner, S., Wetzel, H.U., Kaufmann, H. and Sarnagoev, A. (2005) Potential of satellite remote sensing and GIS for landslide hazard assessment in southern Kyrgyzstan (Central Asia). *Natural Hazards*, 35, 395–416.

Sikder, S. and Hossain, F. (2016) Assessment of the weather research and forecasting model generalized parameterization schemes for advancement of precipitation forecasting in monsoon-driven river basins. *Journal of Advances in Modeling Earth Systems*, 8, 1210–1228.

Skamarock, W.C. and Klemp, J.B. (2008) A time-split non-hydrostatic atmospheric model for weather research and forecasting applications. *Journal Computational Physics*, 227, 3465–3485.

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

Stergiou, I., Tagaris, E. and Sotiropoulou, R.E.P. (2017) Sensitivity assessment of WRF parameterizations over Europe. *Multidisciplinary Digital Publishing Institute Proceedings*, 1, 119.

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M.A., Mitchell, K., Ek, M., Gayno, G., Wegiel, J. and Cuenca, R.H. (2004) Implementation and verification of the unified NOAH land surface model in the WRF model. In: *20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction*, Vol. 1115. Seattle, WA: American Meteorological Society.

Thompson, G., Field, P.R., Rasmussen, R.M. and Hall, W.D. (2008) Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: implementation of a new snow parameterization. *Monthly Weather Review*, 136, 5095–5115.

Tomasi, E., Giovannini, L., Zardi, D. and de Franceschi, M. (2017) Optimization of Noah and Noah_MP WRF land surface schemes in snow-melting conditions over complex terrain. *Monthly Weather Review*, 145, 4727–4745.

Tzeng, G.H. and Huang, J.J. (2011) *Multiple Attribute Decision Making: Methods and Applications*. Boca Raton: Chapman and Hall/CRC.

Wilks, D.S. (2006) *Statistical Methods in the Atmospheric Sciences*, 2nd edition. Burlington: Academic press.

Willmott, C.J. and Matsuura, K. (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30, 79–82.

WMO (Ed.). (2008) Precipitation gauge errors and corrections. In: *Guide to Meteorological Instruments and Methods of Observation*, 7th edition. Geneva: WMO-8, I.6-6–I.6-7. World Meteorological Organisation.

Xie, B., Fung, J.C., Chan, A. and Lau, A. (2012) Evaluation of nonlocal and local planetary boundary layer schemes in the WRF model. *Journal of Geophysical Research: Atmospheres*, 117, D12103.

Xu, L., Liu, H., Du, Q. and Xu, X. (2019) The assessment of the planetary boundary layer schemes in WRF over the central Tibetan plateau. *Atmospheric Research*, 230, 104644.

Yang, K., Koike, T., Fujii, H., Tamura, T., Xu, X., Bian, L. and Zhou, M. (2004) The daytime evolution of the atmospheric boundary layer and convection over the Tibetan plateau: observations and simulations. *Journal of the Meteorological Society of Japan*, 82(6), 1777–1792.

Yang, Z.L., Niu, G.Y., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Longuevergne, L., Manning, K., Niyogi, D., Tewari, M. and Xia, Y. (2011) The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. *Journal of Geophysical Research: Atmospheres*, 116, D12110.

Zhou, X., Yang, K. and Wang, Y. (2018) Implementation of a turbulent orographic form drag scheme in WRF and its application to the Tibetan plateau. *Climate Dynamics*, 50, 2443–2455.

Zou, H., Hu, Y., Li, D., Lü, S. and Ma, Y. (2005) Seasonal transition and its boundary layer characteristics in Anduo area of Tibetan plateau. *Progress in Natural Science*, 15(3), 239–245.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.