1. Sensitivity experiments: ERA-Interim as forcing dataset

The ERA experiment is designed as follows:

- **Time period**: four years (2002/03 and 2010/11).
- **Domain**: HAR30
- **Forcing data**: ERA-Interim (Dee et al., 2011), 0.75° resolution, 6-Hourly.
- **Model physics**: unchanged

The validation is realized for seasonal means (DJF and JJA) using all NCDC stations available in the HAR30 domain.

### 1.1 Comparison with weather stations: precipitation

![Figure S1](image)

**Figure S1**: **Upper panel**: 4-yr averaged precipitation in winter (DJF) for FNL and ERA and their difference over the HAR30 domain. **Lower panel**: mean deviations (bias, %) at each station location for FNL and ERA, and closest dataset at each station (the closest dataset is indicated when one experiment outperforms the other with at least 10 % difference). In total, 628 stations are available for the validation. FNL has the smallest MD at 190 stations, ERA at 118. The patterns of precipitation (and therefore of biases) are very similar (e.g. south to north band of positive and then negative band over China), indicating that most of the differences are related to either WRF or the same errors in both global datasets. On the TP in winter, ERA is wetter than FNL, with the latter being closer to observations for the majority of stations.
Supplemental information to Maussion et al., DOI: 10.1175/JCLI-D-13-00282.1

Figure S2: Same as Fig. S1 but for summer (JJA). FNL (ERA) has the smallest MD at 196 (185) stations. As for DJF, the patterns of biases in JJA are very similar, indicating that most of the differences are related to either WRF or the same errors in both global datasets. On the TP the performance of both forcing datasets is similar.

Figure S3: Box plots of Annual Mean (Absolute) Deviation for all stations over the HAR30 domain (628 stations) in % for FNL (blue) and ERA (red). The two periods (2002/03 and 2010/11) are plotted separately. The whiskers are drawn for the 05%, 25%, 50%, 75% and 95% percentiles. ERA has a marked positive bias on average, but in absolute value both datasets have comparable accuracies (FNL is slightly better).
1.2 Comparison with weather stations: temperature

Figure S4: **Upper panel**: 4-yr averaged 2 m air temperature in winter (DJF) for FNL and ERA and their difference over the HAR30 domain. **Lower panel**: mean deviations (bias) at each station location for FNL and ERA, and closest dataset at each station (the closest dataset is indicated when one experiment outperforms the other with at least 0.5 K difference). In total, 747 stations are available for the validation. FNL (ERA) has the smallest MD at 150 (188) stations. The patterns of temperature biases are very similar (e.g. too cold over the TP and most parts of the Indochina peninsula, too warm in the north), indicating again that most of the differences are related to either WRF or the same errors in both global datasets. On the TP in winter, ERA is significantly warmer than FNL, and closer to observations.
Supplemental information to Maussion et al., DOI: 10.1175/JCLI-D-13-00282.1

Figure S5: Same as Fig. S4 but for summer (JJA). FNL (ERA) has the smallest MD at 84 (93) stations. The differences between the datasets are smaller in summer than in winter. There is an obvious artefact in ERA over the Karakoram, which is due to the snow cover initialization in ERA land surface model for gridpoints with more than 50% glacier area (Collier et al, 2013). Another (unrelated) cold bias is visible in Pakistan and over the TP (approx 0.5 to 1 K).

Figure S6: Box plots of Mean Deviation for all stations over the HAR30 domain (747 stations) in winter (DJF, left) and summer (JJA, right). The whiskers are drawn for the 05%, 25%, 50%, 75% and 95% percentiles. Temperatures are much closer to observations in summer. Again, the performances of both datasets are similar.
2. Supplementary validation of HAR data with TRMM

Figure S7: HSS of HAR30 for different months/thresholds using TRMM3B43 as reference in the full HAR30 domain (N=39990). The upper plot represents the spatially averaged monthly precipitation timeseries from HAR30 and TRMM3B43, together with the maximum HSS for all thresholds at a given month. The right plot represents the HSS evolution with varying thresholds for all pairs. Grey color occurs when no precipitation event was detected by TRMM3B43.

The objective of this domain-wide analysis is to detect potential systematic divergences or inconsistencies in the HAR dataset during the decade. TRMM3B43 and HAR30 are in good agreement for the spatial distribution of precipitation, as indicated by constant high scores for small thresholds. The best scores are reached during the summer months, when most of the precipitation is occurring. HAR30 predicts more precipitation than TRMM3B43, but the two datasets agree well on precipitation seasonality and variability. Seasonality aside, the scores are regular and constant throughout the decade and no singularities can be detected. There is no evident outlier or discrepancy, indicating that the HAR dataset is of constant quality.
3. Supplementary validation of HAR data with NCDC

Figure S8: Same as Fig. 3 but with all stations in the HAR10 domain (71 stations, 8356 valid months)
Figure S9: Mean decadal seasonal cycles of NCDC stations, TRMM3B43 and HAR10 (X-axis units: months, Y-axis units: mm d\(^{-1}\)). The map shows the station locations, with grey points when HAR10 is closer to observations, white points when TRMM3B43 is closer.

The majority of the stations exhibits a characterized summer precipitation regime but there are large variations in shape and in magnitude. The westernmost station, Murree (01), has two pronounced precipitation peaks. The second westernmost, Shiquanhe (10), has a less-pronounced winter precipitation mode that is captured by both TRMM3B43 and HAR10. The stations in the Qaidam Basin (02 to 08) are the driest, with substantial variations in shape. The stations in central TP have generally dry winters, but are different with respect to the length of the precipitation period or to the precipitation magnitude (see for example 16-Tingri and 21-Qumarleb). For 16 of the 31 stations, HAR10 is closer to observations than TRMM3B43.
Figure S10: Snowfall as a function of air temperature at the station WUDAOLIANG (see Figure S9 for the station location). The elevation difference between the HAR10 grid elevation and the station is 47m.

(a) Daily NCDC frozen precipitation flag (1 or 0) for precipitation days (> 1 mm d\(^{-1}\)) as a function of daily temperature measured at the station (grey crosses). 1°C sized binned modes (blue squares)

(b) Daily HAR10 frozen precipitation (%) for precipitation days (> 1 mm d\(^{-1}\)) as a function of daily temperature measured at the station (grey crosses). 1°C sized binned means (red circles) and modes (blue squares)

(c) Daily HAR10 frozen precipitation (%) for precipitation days (> 1 mm d\(^{-1}\)) as a function of daily HAR10 air temperature (grey crosses). 1°C sized binned means (red circles) and modes (blue squares)

Frozen precipitation is reported at the station for a wide range of temperatures and the mode (majority of reports) changes at 3°C. Snowfall from the HAR is given as percentage of total amounts and the mode (where the majority of days with > 50% snowfall) also changes at 3°C. The results are similar when plotted against NCDC temperature (b) or HAR10 Temperature (c).