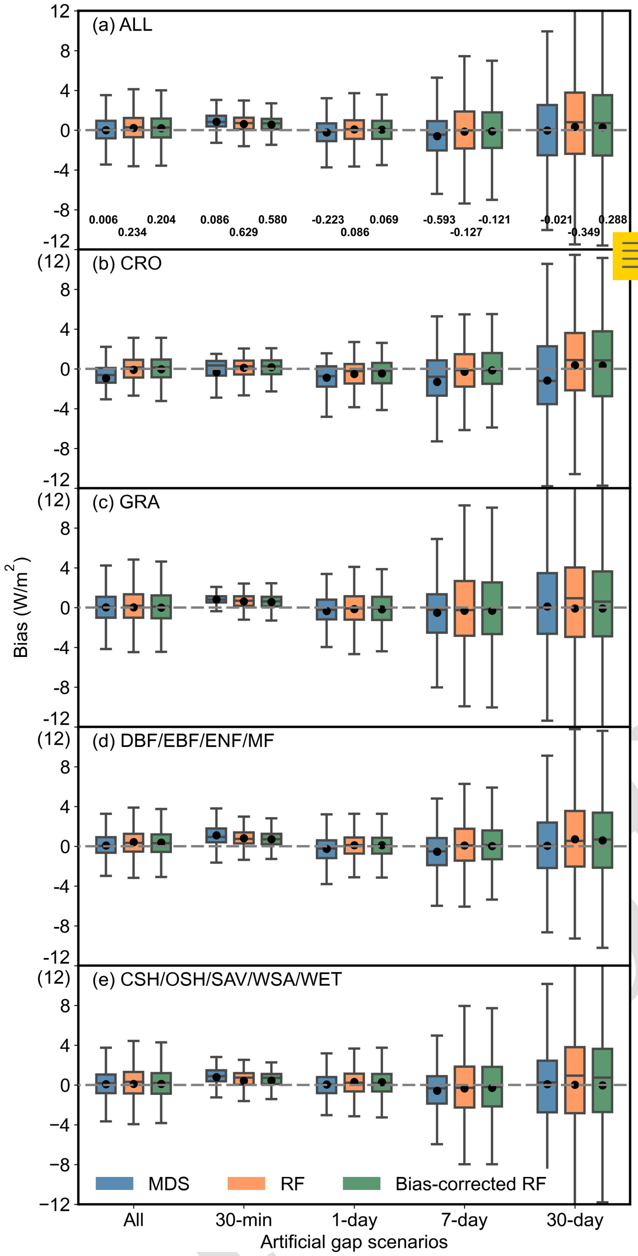

Figure 5. The bias between gap-filled values and observations of three methods under different gap-length scenarios. Different rows of this figure indicate different land cover types. The three horizontal lines of the boxes indicate the first quartile, median, and third quartile, and the black dots indicate the means. Data labels in this figure are the mean value of bias. MDS: marginal distribution sampling. RF: random forest. TS1

performance trends are consistent across land cover types. For short gap lengths the bias-corrected RF demonstrates performance similar to MDS, and both the RF and bias-corrected RF significantly outperform the MDS for longer gap lengths. Given that long gaps comprise 44 % of the FLUXNET2015 dataset, the bias-corrected RF can serve as a more reliable alternative to MDS for hourly-scale data gap-filling, yielding more robust results than those produced by MDS. Overall, the bias-corrected RF algorithm combines the superior performance of the original RF algorithm in long-gap-length scenarios, while providing corrections in cases where the original RF underperforms.

4.1.2 Examples of gap-filled data under artificial 30 d gap-length scenario

For the 30 d gap scenario, the bias-corrected RF algorithm performs better than the MDS algorithm in characterizing time series. As illustrated in Fig. 6, the bias-corrected RF demonstrates strong performance across all land cover types and provides a more accurate representation of daily periodic variations. Although minor biases persist in predicting certain extreme values, these are generally smaller compared to those produced by MDS. In contrast, MDS exhibits significant gap-filling biases across different land cover types, resulting in abnormal overestimations and underestimations (Fig. 6a, b, and i). In some cases it even fails to capture the daily variations of LE (Fig. 6e), while also distorting irregular LE changes (Fig. 6c).

4.2 Evaluation of daily prolonged LE

4.2.1 Consistency between forward and backward prolongation

As shown in Fig. 7a and b, the prolongation performance in both forward and backward directions exhibits high consistency. The results have good accuracy, with RMSE (CC) values of 16.58 W m^{-2} (0.91) for forward and 17.35 W m^{-2} (0.90) for backward. The slight difference may be mainly due to a higher volume of missing data in the first two-thirds of the data compared to the last two-thirds for sites of these land cover types (see Sect. 5.1). There are slight variations in prolongation results for different land cover types (Fig. 7c and d). Performance of CRO and DBF/EBF/ENF/MF is almost the same in both directions. Similar to the half-hourly data gap-filling, our results also demonstrate excellent performance in cropland, with a CC of 0.93 in both directions. GRA and CSH/OSH/SAV/WSA/WET perform slightly worse (2.46 and 3.74 W m^{-2} higher) in the backward direction.

Figure 2b indicates that the need for forward prolongation is significantly greater than that for backward prolongation from 2000 to 2022. Therefore, the validation in the following sections will only focus on the forward direction.

4.2.2 Temporal stability of the prolongation

We used data from the first 3 years and the first 8 years for training, and evaluated the prolongation performance for each subsequent year. Three years of data represents an extreme case of the minimum training data volume in