We thank the editor for their suggestions on our manuscript. We agree that highlighting the statistically significant biases is needed to demonstrate whether the bias-adjustment leads to a meaningful improvement in the models. We have addressed the points above as follows.

We test the null hypothesis that the biases we find arise solely from internal interannual variability, applied to Figs 3, 4 and Tables 4, 5. We model annual accumulation anomalies as a stationary AR(1) process, with lag-1 autocorrelation and innovation standard deviation estimated from the original model output. AR(1) Monte Carlo realisations are generated and cumulatively summed over each fitting period. Statistical significance is quantified as the probability that the absolute cumulative sum of the AR(1) process exceeds that of the observed bias over the same period. Biases with $p < 0.05$ are indicated in bold/hatching.

Table 5 previously showed basin-wise contributions to the total absolute bias in Gt/yr (such that percentages summed to 100%), rather than signed bias used elsewhere in the study. To ensure statistical significance is comparable across tables, we propose replacing this table with consistent signed net and relative basin-wise bias and placing it following the spatially integrated Gt/yr means for the ice-sheet, accumulation and ablation zones (Table A1) in the appendix. Statistical significance is indicated using the Monte Carlo AR(1) test (also applied to Table A1).

For Table 3, biases and RMSE are derived from direct comparison between individual observations and corresponding model grid cell values over differing time scales (monthly to multi-annual). In this context, inter-annual variability cannot be defined in a consistent or meaningful way. Instead, statistical significance of mean bias and RMSE changes is assessed using 95% confidence intervals obtained from 1,000 nonparametric bootstrap resampling. Units of RMSE have also been added to table 3.

The significance of biases has been highlighted in relevant sections of the results, discussion, conclusions and abstract. The Discussion section has also been reorganised to better contextualise the significance of the reported biases.