Many thanks for the comments on our initial responses. We have addressed them below, and we believe these sections of the paper are clearer to the reader as a result.

Robustness of linear regressions

Apologies for a lack of clarity here. The target and predictor variables are all taken from the IPCC AR6 database, with the predictor data for each scenario used to make the predictions (blue) of the target data (orange), using the regressions we calculated historically (and use in FRIDA-Clim). This is therefore internally consistent at the scenario level, allowing us to test the predictive power of the regression across the scenario database.

We have therefore added to the figure S1 caption to clarify this, and to briefly contextualise the errors relative to the absolute responses:

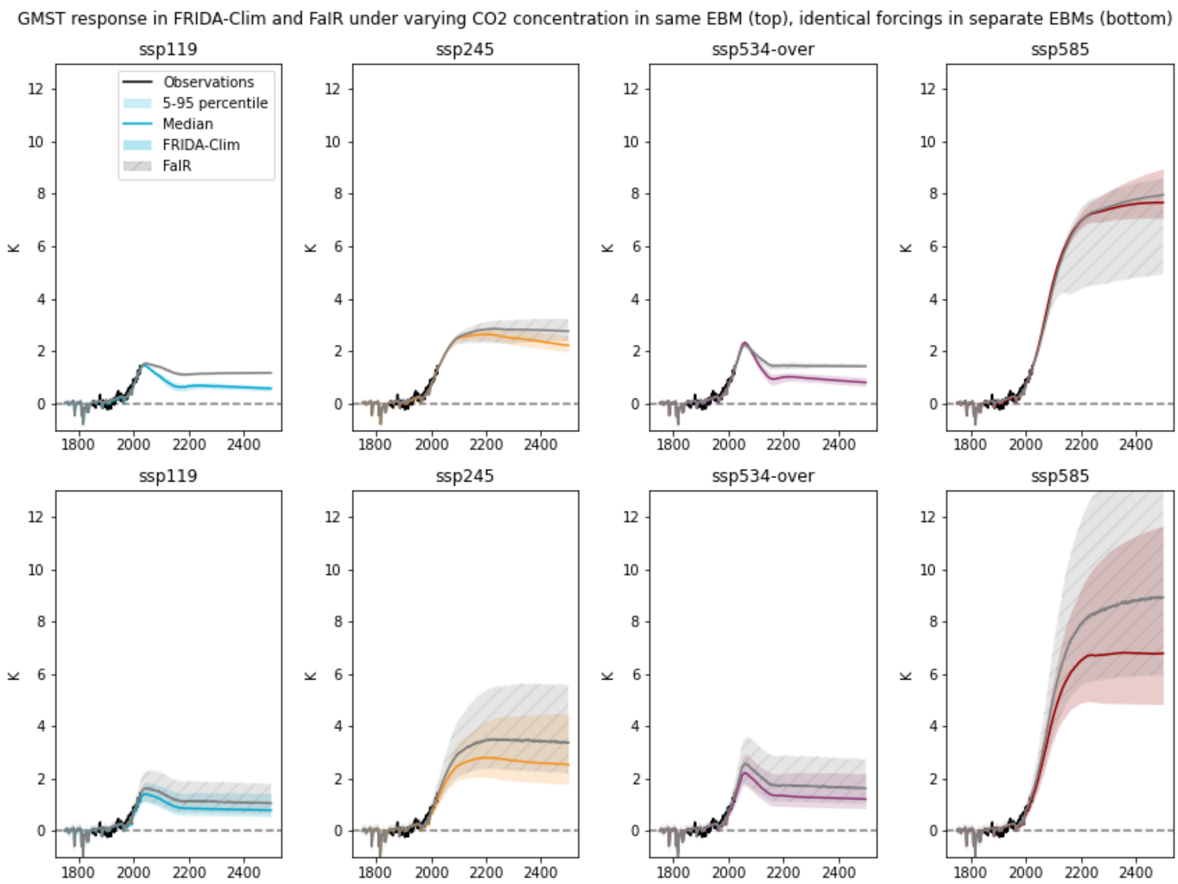
Figure S1: Regression predictions of NOx, VOC, and CO emissions, and the BC Snow forcing, and the effect of using these fits on temperature projections; see Section 2.1 for details on the targets and choice of predictors. Left: historical values of both the target variable (blue) and the emulation estimate when building a regression based on the predictor(s) (orange) are shown to 2015; this historical dataset is used in the regression. These are extended to 2100 for the 1703 scenarios in the IPCC AR6 database (Byers et al., 2022) which include all species input to the FaIR simple climate model, with IAM-reported emissions and their approximation using the regression shown. Data for the predictor variables are taken for each of the 1703 scenarios, and the historically-derived regression parameters used with these to predict the target variables for the scenario. The full ensemble of these scenario timeseries is shown on the left, with the error then calculated at the scenario level. The 5th, 10th, 50th, 90th, and 95th percentiles of this emulation error are then applied to the default single-run SSP245 scenario in FaIR, with the resultant impacts on surface temperature shown on the right, using the same scale for each. Errors are less than a few hundredths of a degree throughout the period, representing an error of around 1% on the absolute temperature response, which are several degrees.

We have also modified the main text on L176 to note the errors are slight compared to the absolute response:

errors are typically under a few hundredths of a degree, or around 1% of the total temperature response

Unexplained Temperature Divergence

We have adjusted figure S6 to feature the scenario names as headers instead, consistent with figure 6:



We have also clarified the figure S6 caption to clarify the ensemble members, including referring back to section 3.1, so it now reads:

Figure S6: investigation of the relative contribution of differences in carbon cycle and EBM parameters to the difference between FRIDA-Clim and FaIR in the SSPs. Top row: GMST when passing the CO2 concentrations in the tow row of Figure S5 to the same (default FaIR single-member) EBM with all other species as before; differences between the models here are therefore due to the effect of CO2 concentration variation. Uncertainty within these ensembles is then due to the variation in CO2 concentration in the ensembles in Figure S5. Bottom row: GMST when running the same single forcing for each scenario (from RFMIP) in each posterior ensemble (i.e. just the EBM parameters, from the broader ensemble in Section 3.1 for FRIDA-Clim, and from the full FaIR EBM ensemble as used in Figure 6); differences here are due to the EBM variations. Overall, comparing to Figure 6, it appears the stronger carbon sinks in FRIDA-Clim drive the cooler response in overshoot scenarios, while a more sensitive EBM causes FaIR's higher temperatures in high emissions scenarios.