Verkerk and coauthors combine reduced-complexity volcanic aerosol (EVA_H) and climate (FaIR) models to simulate the global mean surface temperature (GMST) response to volcanic eruptions over the last 9,000 years (6755 BCE to 1900 CE).

To assess the robustness of their simulations, the authors compare their estimates for the 14 largest eruptions between 1250 CE and 1900 CE with numerous climate reconstructions (Schneider et al., 2015; Wilson et al., 2016; Guillet et al., 2017; Pages2k, 2019; King et al., 2021). The discrepancies between the new simulations and climate reconstructions are notably smaller than in previous studies.

The authors address an important topic. The paper is well-written, well-structured, and easy to follow. The figures are clear and informative. And the authors have made all their simulations publicly available.

The methodology section summarizes well the approach taken by the authors, including the forcing datasets used for the new simulations, the paleo-reconstructions and the climate simulations employed to compare the new results.

Additionally, they acknowledge the limitations of their approach, particularly the *Holocene temperature conundrum*, which is also apparent in their ensemble simulations of Holocene temperatures.

The authors emphasize the need for future products based on reduced-complexity models to include seasonal and regional outputs, which would be highly valuable for the paleo community.

I appreciated reading the manuscript and, overall, have very few comments to offer. I recommend the paper for acceptance, as I think the new product provided by the authors represents a valuable resource for the paleo community studying past volcanic eruptions. However, I do have one minor suggestion for the authors to consider.

Main text:

- **Comparing simulations with instrumental data**: Pushing the simulations beyond 1900 CE would have been a great addition. Extending the simulations into the 20th century would allow direct comparisons with instrumental data for eruptions such as the 1902 (Santa María), 1912 (Katmai/Novarupta), 1963 (Agung), and 1991 (Pinatubo) events. They could help validate the accuracy of the simulations.

Have the authors considered the possibility of comparing the accuracy of their simulations not only against climate/data assimilation reconstructions but also against instrumental datasets, such as the Berkeley Earth Surface Temperature (BEST) dataset? The BEST dataset offers two products that might be of interest: one estimating GMST since 1850 and another providing annual temperature estimates since 1750 (land-only).

Using these datasets could allow the authors to compare their simulations for the 1815 Tambora, 1831 Zavaritskii (Hutchison et al., 2024), and 1883 Krakatau events with "real" temperature observations. Additionally, the Laki eruption might also be investigated, assuming the instrumental records used by BEST are sufficiently dense to represent a reliable global average (which I am not entirely certain about).

Thank you for your very supportive comments and for suggesting these additional analyses. We have now conducted them and added them to our manuscript. Since the Berkeley Earth dataset has a very sparse spatial coverage before the 1830s (fig 1), we choose to use only the GMST dataset, starting in 1850.

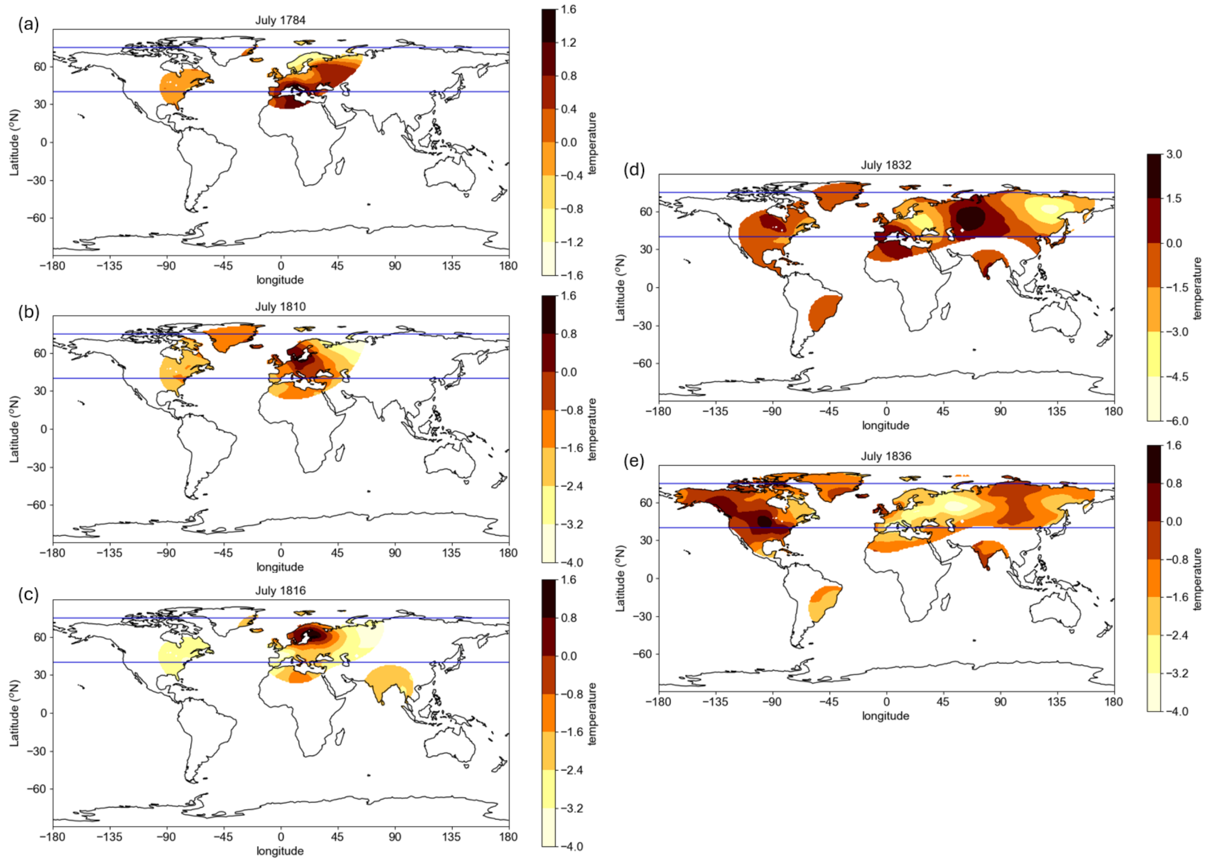

Figure 1: coverage of the Berkeley Earth dataset in July of the year following a major eruption (a. Laki 1783, b. 1809, c. Tambora 1815, d. Zavaritskii 1831, e. Cosiguina 1835). Horizontal lines delimit the 40-75°N latitudinal band. Temperature are expressed as anomaly with respect to the 1951-1980 mean.

For comparison, we also included the Cowtan and Way dataset. We generate a 1000-member ensemble for the period 1850-2021 following the same methodology as our 6755 BCE – 1900 CE ensemble. The results are summarized in the text below that has been added to the manuscript.

Sect. 2.1:
"In addition to this 6755 BCE – 1900 CE volcanic emission dataset, we also use the CMIP7 volcanic emission dataset to construct an ensemble for the period 1850 – 2021 CE. The eVolv2k dataset provides emissions for the period 1850-1900, emissions

between 1901 and 1978 come from the bipolar ice core record of Sigl et al. (2015). Between 1979 and 2021, we use the satellite record MSVOLSO2L4 (Carn, 2024). For unidentified eruptions or when the injection height is unknown, we apply the same principles as for the 6755 BCE – 1900 CE emission dataset."

Sect. 2.2:

"Forcings for the 1850-2021 CE ensemble

For the simulations covering exclusively the historical period, we include a larger range of anthropogenic forcings. These include short lived climate forcer (seven species, e.g. black carbon, carbon monoxide), halogen gases (18 species, e.g. chlorofluorocarbons), fluorinated greenhouse gases (23 species, e.g. Hydrofluorocarbons), CO2, N2O and CH4 emissions, and solar and land-use forcings. Ozone and anthropogenic aerosol forcings are calculated from their precursor emissions."

Sect. 2.5: Historical observations

"For the historical period (i.e. after 1850), we compare our 1850-2021 CE ensemble to instrumental observations. We use two observation datasets, the 1850 – present Berkely Earth temperature record (Rohde and Hausfather, 2020) and the 1850 – 2017 Cowtan and Way record (Cowtan and Way, 2014). Both datasets contain monthly temperatures, with a coarser spatial resolution for Cowtan and Way (5° by 5° grid, whereas Berkeley Earth has a spatial resolution of 1° by 1°). The surface temperature is interpolated in region with no station coverage, with Berkeley Earth using a larger number of land stations than Cowtan and Way (around four times more). Here, we use the global annual mean from these spatially resolved datasets to compare it to our simulations."

Sect 4.2.4: Historical period variability (1850-2021 CE)

"At a multi-decadal timescale, we observe that the global warming trend in our 1850-2021 ensemble of simulations follow closely the trend in the observations (+0.99 K in Cowtan and Way for the 2010-2016 period relative to the 1850-1900 mean, +1.07 K in Berkeley Earth, +0.97 ± 0.19 for our simulations). Most of the observed annual mean temperatures are within the 95% confidence interval of our simulations (92% of the Berkeley Earth dataset, 95% of the Cowtan and Way dataset). We note that the mid-1930s to early 2000s temperatures appear warmer in the observations.

The observation datasets show a strong interannual variability, with an amplitude similar to the response to volcanic eruptions (e.g. -0.22 K between 1991 and 1992 following the Pinatubo eruption and -0.24 K between 1998 and 1999). To compare the response to volcanic forcing between observations and simulations, we perform a superposed epoch analysis over 6 eruptions that injected more than 7 Tg of SO2 in the stratosphere (Kie Besi 1861, Krakatau 1883, Novarupta 1912, Agung 1963, El Chichon

1982, Pinatubo 1991). We obtain a peak cooling of 0.15 ± 0.06 K on average in our simulations, against 0.12 K in the two observation datasets (Fig. 8.c)."

We also add the following figure (Fig 8 in the manuscript):

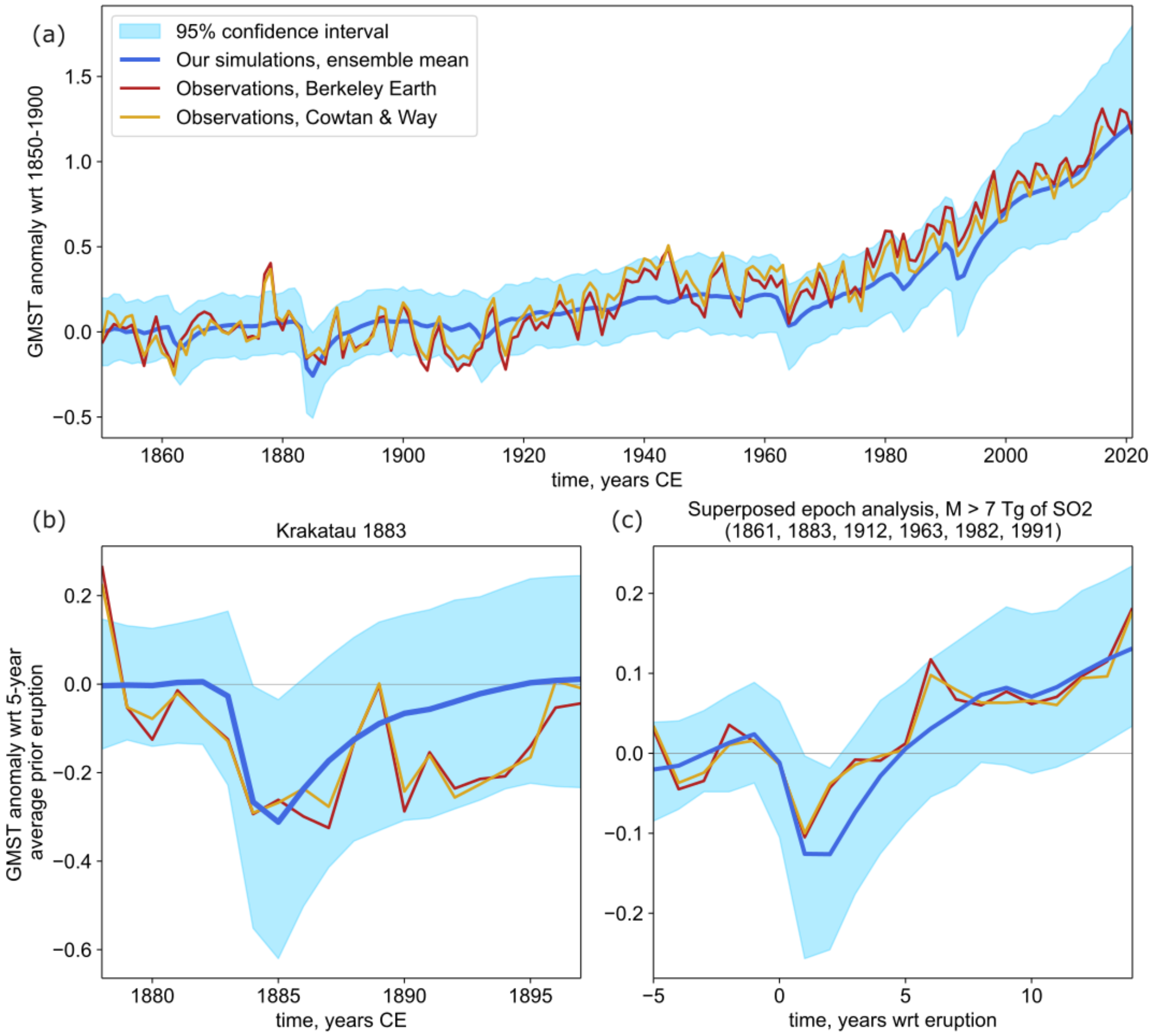

Figure 2: a. GMST for 1850-2021 from our simulations, compared to observations from Berkeley Earth and Cowtan and Way datasets; b. GMST response to the Krakatau 1883 eruption; c. Superposed epoch analysis of the GMST response to 6 eruptions injecting more than 7 Tg of SO2 that occurred between 1850 and 2021 (Kie Besi 1861, Krakatau 1883, Novarupta 1912, Agung 1963, El Chichon 1982, Pinatubo 1991).

- Line 130: Change Hutchison et al., in review to Hutchison et al., 2024

Thank you, we have updated it.

Supplementary Material

- Line 60: "Table S3: Integrated response of the superposed epoch analysis (Error! Reference source not found.d)." There appears to be a reference issue here that should be corrected.

Thanks for pointing that, it now refers to Fig 7 in the main text.