The Attribution of February Extremes over North America: A Forecast-Based Storyline Study

DONGHYUN LEE,[a] SARAH SPARROW,[b] NICHOLAS LEACH,[c] SCOTT OSPREY,[c] JINAH LEE,[a,d] AND MYLES ALLEN[a,c]

[a] *Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, United Kingdom*
[b] *Oxford e-Research Centre, Department of Engineering Science, University of Oxford, Oxford, United Kingdom*
[c] *Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, United Kingdom*
[d] *Korea Meteorological Administration, Daejeon, South Korea*

(Manuscript received 30 January 2024, in final form 11 May 2024, accepted 20 June 2024)

ABSTRACT: The importance of extreme event attribution rises as climate change causes severe damage to populations resulting from unprecedented events. In February 2019, a planetary wave shifted along the U.S.–Canadian border, simultaneously leading to troughing with anomalous cold events and ridging over Alaska and northern Canada with abnormal warm events. Also, a dry-stabilized anticyclonic circulation over low latitudes induced warm extreme events over Mexico and Florida. Most attribution studies compare the climate model simulations under natural or actual forcing conditions and assess probabilistically from a climatological point of view. However, in this study, we use multiple ensembles from an operational forecast model, promising statistical as well as dynamically constrained attribution assessment, often referred to as the storyline approach to extreme event attribution. In the globally averaged results, increasing CO_2 concentrations lead to distinct warming signals at the surface, resulting mainly from diabatic heating. Our study finds that CO_2-induced warming eventually affects the possibility of extreme events in North America, quantifying the impact of anthropogenic forcing over less than a week's forecast simulation. Our study assesses the validity of the storyline approach conditional on the forecast lead times, which is hindered by rising noise in CO_2 signals and the declining performance of the forecast model. The forecast-based storyline approach is valid for at least half of the land area within a 6-day lead time before the target extreme occurrence. Our attribution results highlight the importance of achieving net-zero emissions ahead of schedule to reduce the occurrence of severe heatwaves.

KEYWORDS: North America; Extreme events; Anthropogenic effects/forcing; Surface temperature; Statistical techniques; Operational forecasting

1. Introduction

As more record-breaking extreme events have occurred in recent years, the importance of attribution studies has risen dramatically (IPCC 2013, 2021). The increased risk of extreme events has raised the importance of relevant mechanisms. The pronounced media exposure of extreme events has led the public to accept impacts from climate change as a fact and illuminate the topic's significance more for a specific extreme case attribution than risk-based general detection, demanding more complex knowledge combining statistical analysis and physical understanding (Shepherd 2016; Lloyd and Shepherd 2023). Conceptually, most attribution studies can be categorized by two frameworks: storyline and probabilistic (Shepherd et al. 2018). The storyline framework examines the human-induced impacts on the main drivers of extreme events, conditioning them to the physical process. In contrast, the probabilistic framework quantifies the anthropogenic attribution of the risks by defining the specific event as one of a class, regardless of the particular weather processes.

Due to the nonlinearity and rarity of target extreme events, current attribution studies conventionally use the probabilistic framework, which is often called the risk-based approach. These studies compare the likelihood of risks or the observed event intensity based on the estimated probability distribution of extremes between factual and counterfactual results from climate model simulations contrasting anthropogenic forcing effects. These simulations provide ensembles of modeled states interacting with different physical parameterization details (e.g., Pall et al. 2011; Seong et al. 2022; Min et al. 2022). These simulations are conditioned on a clear thermodynamic signal seen with the target extreme events rather than atmospheric circulation because the changes in the latter from anthropogenic forcings are usually minor compared to natural variability (Deser et al. 2012, 2014).

However, this probabilistic framework cannot preselect those dynamical processes involved in observed extreme events, which raises questions of process fidelity within unconditional statistical approaches (Bellprat and Doblas-Reyes 2016; Bellprat et al. 2019; Palmer and Weisheimer 2018). Contrary to the probabilistic framework, with a highly constrained experimental design, some studies employ a storyline approach to understand the role of physical processes in extreme events (e.g., Wang et al. 2021; van Garderen and Mindlin 2022; Terray 2023; Blanchard-Wrigglesworth et al. 2023). Storyline approaches can use AMIP-style simulations, or nudging to inherit a physically conditioned environment (e.g., sea surface temperature

Denotes content that is immediately available upon publication as open access.

Corresponding author: Donghyun Lee, donghyun.lee@ouce.ox.ac.uk, donghyun.lee@imperial.ac.uk

DOI: 10.1175/JCLI-D-24-0074.1

and sea ice conditions or large-scale flow) corresponding to the main driver of the observed target extreme events.

A new numerical model framework for extreme attribution is proposed, utilizing operational medium-range weather forecasts (Leach et al. 2021). Unlike the climate model, this forecast model has been widely verified for particular weather events at process levels. The predictability of the operational forecast model ensures that the constrained physical process matches the storyline framework. Furthermore, the free-running coupled ocean–atmosphere integrations with large ensembles comprising different initial conditions allow for a spread in synoptic conditions over time, allowing a risk-based assessment with good statistics. Since this operational forecast approach has the advantages of both frameworks, previous studies using these have successfully assessed the human-induced CO_2 attribution to the record-shattering extreme heatwave events over Europe in 2019 (Leach et al. 2021), the Pacific Northwest in 2021 (Leach et al. 2024), and also the midlatitude windstorms over the United Kingdom in 2022 (Ermis et al. 2024).

In this study, we introduce the extreme events that occurred in February 2019 in North America and quantitatively compare the severity of these within the historical records. We examine the performance of the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble prediction system in reproducing these extremes. Then, we quantify the size of the CO_2 signal and the probabilistic risk of extremes by comparing the multiensemble simulations under perturbed CO_2 conditions. Finally, we summarize and discuss our storyline attribution findings.

2. Data and methods

a. Forecast model

We use the ECMWF Integrated Forecasting System (IFS) model cycle CY45R1, which was the version in operation at the time of the events. The ECMWF IFS comprise an atmospheric model with 91 vertical layers and about 18 km horizontal resolution up to 15 days, increasing to 36 km from 16 to 30 days. This atmospheric model is coupled to an ocean model, Nucleus for European Modelling of the Ocean (NEMO), version 3.4, which has 75 vertical levels and 0.25° horizontal resolution (about 27 km on the equator). Notably, ECMWF IFS is one of the best models with outstanding performance in predicting the development of 2019 winter sudden stratospheric warming (Rao et al. 2019).

b. Perturbed CO_2 experiment

We used the perturbed CO_2 forecast ensemble simulations implemented by Leach et al. (2021). Four different initialization dates (4, 11, 17, and 23 February 2019) of operational ensembles (ENS) are selected to simulate until the 27 February with different CO_2 concentration levels. These two counterfactual worlds are preindustrial levels of 285 ppm (PIC) and increased to 600 ppm (INC), where that level is approximately equal to the doubled global radiative forcing from the PIC to ENS. These opposite directions of CO_2 perturbation experiments isolate only the atmospheric CO_2 impacts on the extreme events, which is more comprehensible without estimating the human-caused effects on ocean states (Stone and Pall 2021). The experimental design details are depicted in Leach et al. (2021).

c. Reanalysis

We use the fifth major global reanalysis produced by ECMWF (ERA5) from 1950 to 2019 (Hersbach et al. 2020). Climatological values are calculated over the years 1999–2018. The horizontal resolution of the data in our analysis is 0.25°, except for the calculation of globally averaged results, which is at a 2.5° horizontal resolution.

d. Extreme events

Our study focuses on the daily mean air temperature at 2 m from the surface (T_{2m}). The extreme variable TX (TN) indicates the maximum (minimum) T_{2m} in February in each grid scale of reanalysis dataset. After finding some regions experiencing the highly ranked TX (TN) in February 2019, we checked the extreme event validity of the heatwaves (cold spells) as the target of our storyline attribution study. The definition of heatwaves (cold spells) follows two conditions in regionally averaged T_{2m}: 1) when T_{2m} exceeds (below) the climatological threshold and 2) the first conditions last at least three consecutive days. The threshold is 90th (10th) percentile values from the February climatological period (e.g., Puvvula et al. 2022; Thomas et al. 2023). We found highly ranked TX and TN events in the four heatwaves and a cold spell, respectively. As the definitions of TX (TN) and heatwave (cold spell) are different, we test consistency in the extreme occurrence date between the extreme event definitions of 1) regionally averaged T_{2m} in peak date of heatwave (cold spell) and 2) TX (TN), which may have different dates on each grid scale. Notably, the regionally averaged temperature on a peak day (solid red or blue horizontal lines in Fig. 5) is similar to TX (TN) during the heatwave (cold spell) days (dotted lines in Fig. 5), supporting the fact that the extreme event in most grids over the target region coincides with the local peak date.

e. Temperature tendency and thermodynamic decomposition

Atmospheric temperature changes in time (tendency) can be explained by the thermodynamic equation below:

$$\frac{\Delta T_{\mathrm{Atm}}}{\Delta t} = -\mathbf{v}\cdot\nabla_p T_{\mathrm{Atm}} - \omega\frac{T_{\mathrm{Atm}}}{\theta}\frac{\partial\theta}{\partial p} + Q, \tag{1}$$

where T_{Atm} is the atmospheric temperature, Δt is a day, $\mathbf{v}$ is the horizontal wind velocity, p is the pressure, ω is the vertical velocity in pressure coordinates, and θ is the potential temperature. From left to right, temperature tendency (Tend, $\Delta T_{\mathrm{Atm}}/\Delta t$) is comprised of contributions from horizontal advection (Adv, $-\mathbf{v}\cdot\nabla_p T_{\mathrm{Atm}}$), adiabatic heating [Adiab, $-\omega(T_{\mathrm{Atm}}/\theta)(\partial\theta/\partial p)$], and diabatic heating (Diab, Q). The three terms are computed at each grid point. Notably, the diabatic heating is a combination of shortwave and

longwave radiative processes and heating fluxes (e.g., sensible and latent). We estimated it as the residuals according to previous studies (e.g., Yanai and Johnson 1993; Lai and Gong 2017; Yao et al. 2017; González-Herrero et al. 2022) because of lack in archived variables in ECMWF IFS experiments. This indirect calculation is expected to be negligibly different compared with direct values (e.g., Kim et al. 2021; Kim and Lee 2022). We applied Eq. (1) to the first and second nearest levels to the surface in the archived IFS experiment datasets (1000 and 925 hPa). We used their vertically weighted values as the temperature tendency results for the near-surface. To minimize the complex terrain effects on the near-surface atmospheric variables at 0.25° horizontal resolution, the temperature tendency and its components over the North American regions are slightly filtered using a nine-point local smoothing before analysis, following the method of Kim et al. (2021). We do not apply smoothing for globally averaged results because they are calculated from coarse resolution (2.5°).

Now, the temperature tendency can be rewritten in different formats as follows:

$$T_{\mathrm{Atm}}(d) - T_{\mathrm{Atm}}(d-1) = \int_{d-1}^{d} \mathrm{Tend}\, dt = \mathrm{Tend}(d), \tag{2}$$

$$T_{\mathrm{Atm}}(d) = \mathrm{Tend}(d) + T_{\mathrm{Atm}}(d-1), \tag{3}$$

$$T_{\mathrm{Atm}}(d) = \mathrm{Tend}(d) + \mathrm{Tend}(d-1) + T_{\mathrm{Atm}}(d-2), \tag{4}$$

where letter d in parentheses indicates a day for each variable. As we cannot calculate the first-day value from our forecast ensembles due to a lack of variables ahead of forecast initialization date, Eqs. (2)–(4) can be replaced by the accumulation of tendency as follows:

$$T_{\mathrm{Atm}}(d_2) - T_{\mathrm{Atm}}(d_1) = \sum_{t=d_1+1}^{t=d_2} \mathrm{Tend}(t), \quad \text{where } d_2 > d_1 \geq 1. \tag{5}$$

The sum of the tendency can express the temperature between two different days. We simplify Eq. (5) as follows:

$$\Delta T_{\mathrm{Atm}} = \sum \mathrm{Tend} = \sum \mathrm{Adv} + \sum \mathrm{Adiab} + \sum \mathrm{Diab}. \tag{6}$$

As ΔT_{Atm} is calculated from the near-surface atmosphere (1000–925 hPa), we expect change in it is similar to $\Delta T_{2\mathrm{m}}$ (e.g., Kim et al. 2021; González-Herrero et al. 2022) unless land–atmosphere interaction has a significant role (e.g., Miralles et al. 2014; Barriopedro et al. 2023; García-García et al. 2023). To convert ΔT_{Atm} to $\Delta T_{2\mathrm{m}}$, we add residuals (RES) to indirectly reflect the land interaction effects on $\Delta T_{2\mathrm{m}}$ between two periods (d_1, d_2). Although near-surface Diab includes surface turbulent heat fluxes (e.g., Dirmeyer et al. 2021), RES still implies the strength of land-interaction effects because those are amplified due to a difference in coordination: near-surface pressure levels (1000–925 hPa) versus 2-m height.

$$\Delta T_{2\mathrm{m}} = \sum \mathrm{Tend} + \mathrm{RES}. \tag{7}$$

f. Temperature anomaly decomposition

$$\dot{T}_{2\mathrm{m}} = (T_{2\mathrm{m}} - \overline{T}_{2\mathrm{m}}) + (\overline{T}_{2\mathrm{m}} - \overline{\overline{T}}_{2\mathrm{m}}), \tag{8}$$

$$\dot{T}_{2\mathrm{m}} = T'_{2\mathrm{m}} + \overline{T}'_{2\mathrm{m}}. \tag{9}$$

The $T_{2\mathrm{m}}$ with an upper dot denotes an anomaly relative to all $T_{2\mathrm{m}}$ of the February climatological period (1999–2018). The $T_{2\mathrm{m}}$ with single and double overlines are monthly averaged $T_{2\mathrm{m}}$ values for February 2019 and those of all days of February in climatological periods, respectively. Therefore, two terms in parentheses indicate 1) transient eddy relative to the $T_{2\mathrm{m}}$ monthly mean of 2019 ($T'_{2\mathrm{m}}$) and 2) $T_{2\mathrm{m}}$ monthly mean anomaly of 2019 relative to all February in climatological periods ($\overline{T}'_{2\mathrm{m}}$), respectively. We interpret these two terms as the $T_{2\mathrm{m}}$ anomaly contributions from the submonthly and the interannual scale, respectively.

To account for the contributions from the atmosphere and the other sources to $T_{2\mathrm{m}}$ anomaly ($\dot{T}_{2\mathrm{m}}$), we combine Eqs. (7) and (8) and perturbing $\overline{T}_{2\mathrm{m}}$ term on the left-hand side [Eq. (10)].

$$T_{2\mathrm{m}}(d_2) - \overline{T}_{2\mathrm{m}} - [T_{2\mathrm{m}}(d_1) - \overline{T}_{2\mathrm{m}}] = \sum \mathrm{Tend} + \mathrm{RES}. \tag{10}$$

By reallocating terms and subtracting $\overline{\overline{T}}_{2\mathrm{m}}$ for both left- and right-hand-side terms, Eq. (10) can be transformed as

$$[T_{2\mathrm{m}}(d_2) - \overline{\overline{T}}_{2\mathrm{m}}] = \sum \mathrm{Tend} + [T_{2\mathrm{m}}(d_1) - \overline{T}_{2\mathrm{m}}] + (\overline{T}_{2\mathrm{m}} - \overline{\overline{T}}_{2\mathrm{m}}) + \mathrm{RES}, \tag{11}$$

$$\dot{T}_{2\mathrm{m}}(d_2) = \sum_{t=d_1+1}^{t=d_2} \mathrm{Tend}(t) + T'_{2\mathrm{m}}(d_1) + \overline{T}'_{2\mathrm{m}} + \mathrm{RES}. \tag{12}$$

In Eq. (12), the anomalous temperature on a specific day (d_2) is explained by contributions from the accumulated tendencies between two dates, a transient eddy on another day, monthly anomalies, and residuals. We use these decompositions to explain extreme events over five regions at specific dates experiencing different synoptic weather system. The d_1 is selected 3 days ahead of d_2 when the synoptic weather system dramatically varies the local temperature (see Fig. 2). Each term of Eq. (12) implies 3-day synoptic weather effects, the transient eddy before it, interannual variability effects on the monthly anomaly, and land-atmosphere interactions during synoptic time intervals, respectively.

g. CO_2 signal decomposition

CO_2 signal is measured from the differences between INC and PIC experiments as follows:

$$\mathrm{SIG}(T_{2\mathrm{m}}) = \mathrm{INC}(T_{2\mathrm{m}}) - \mathrm{PIC}(T_{2\mathrm{m}}), \tag{13}$$

where $\mathrm{SIG}(T_{2\mathrm{m}})$ indicates the CO_2 signal on $T_{2\mathrm{m}}$ on a specific day. By considering Eq. (7) with $d_1 = 1$, we can rearrange Eq. (13) by

$$\mathrm{SIG}[T_{2\mathrm{m}}(d)] = \mathrm{INC}\left[T_{2\mathrm{m}}(1) + \sum_{t=2}^{t=d} \mathrm{Tend}(t) + \mathrm{RES}\right] - \mathrm{PIC}\left[T_{2\mathrm{m}}(1) + \sum_{t=2}^{t=d} \mathrm{Tend}(t) + \mathrm{RES}\right]. \tag{14}$$

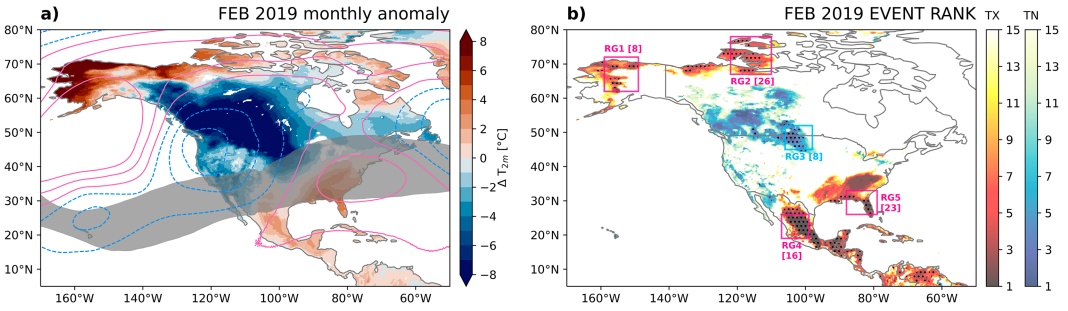

FIG. 1. (a) Monthly anomalies of temperature (shading) and 500-hPa geopotential height (contour) in February 2019 relative to the 20-yr climatology (1999–2018). The 20-, 70-, and 120-m (−20, −70, and −120 m) geopotential heights are highlighted with pink solid (sky-blue dashed) contour lines. Gray transparent shadings around 30°N denote 200-hPa zonal wind exceeding 35 m s^{-1} to illustrate the position of the subtropical jet stream. (b) The rank of TX and TN in February 2019 relative to TX and TN in February since 1950. Stipples indicate that the extreme event occurred between the 8 and 27 Feb, where the perturbed CO_2 simulations exist. The solid rectangles highlight the target regions of our study, and each date of their extreme event occurrence is displayed in brackets.

Since the CO_2 signal in our experiments originated from a perturbation in atmospheric CO_2 concentration, the size of RES in both INC and PIC should have a negligible monthly difference (see Fig. 3b). By definition of SIG [Eq. (13)], Eq. (14) can be rearranged as

$$\mathrm{SIG}[T_{2\mathrm{m}}(d)] \approx \mathrm{SIG}[T_{2\mathrm{m}}(1)] + \mathrm{SIG}\left[\sum_{t=2}^{t=d} \mathrm{Tend}(t)\right]. \tag{15}$$

We further decompose Eq. (15) by applying Eq. (6).

$$\mathrm{SIG}[T_{2\mathrm{m}}(d)] \approx \mathrm{SIG}[T_{2\mathrm{m}}(1)] + \mathrm{SIG}[\textstyle\sum \mathrm{Adv}] + \mathrm{SIG}[\textstyle\sum \mathrm{Adiab}] + \mathrm{SIG}[\textstyle\sum \mathrm{Diab}]. \tag{16}$$

Thus, SIG in $T_{2\mathrm{m}}$ can be explained by the sum of SIG for the first 24 h of forecast simulations and the cumulative sum of Adv, Adiab, and Diab. In our study, all results of SIG decomposition are first calculated from each grid and then averaged for the selected domain.

h. Statistical methods

Changes in event probability are analyzed by a risk ratio (P_1/P_0) analysis, which implies fractional changes in the likelihood of an event exceeding the level of observed extreme in ensembles under the different forcing conditions (e.g., P_1 = ENS and P_0 = PIC for $P_{\mathrm{ENS}}/P_{\mathrm{PIC}}$). To calculate the risk ratios among ensembles, we used the stationary generalized extreme value (GEV) fitting method.

$$F(x;\,\mu,\,\sigma,\,\xi) = \begin{cases} \exp\left[-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right], & \xi = 0 \\ \exp\left[-\left\{1+\xi\dfrac{x-\mu}{\sigma}\right\}^{-\xi^{-1}}\right], & \xi \neq 0,\ 1+\xi\dfrac{x-\mu}{\sigma} > 0 \end{cases}. \tag{17}$$

The cumulative distribution function, $F(x;\,\mu,\,\sigma,\,\xi)$ of the GEV, is given by Eq. (17), estimating the distribution types of Gumbel, Fréchet, and Weibull depending on the shape (ξ) parameters. The mean and spread of fitting are parameterized by the location (μ) and scale (σ), respectively. In this study, x denotes the extreme values in a forecast ensemble out of each set of initialization date. To check the sensitivity in the GEV fitting method, we examined parameters fitted by two methods: maximum likelihood estimation (MLE) and L moments. We confirmed the negligible differences in their parameters (generally less than a decimal point with the same sign). Nevertheless, we only present the results from L moments estimation because statistically larger samples are required for MLE (e.g., Kharin et al. 2007; Min et al. 2009; Lazoglou et al. 2019; Kim et al. 2020; Leach et al. 2021; Wehner et al. 2024). For the quantification of GEV uncertainty ranges, we implement bootstrapping resampling from the ensembles 100000 times and estimate the 90% confidence intervals and median values. We compute GEV for each extreme case: TX (TN) during days of heatwaves (cold spell).

To assess the statistical robustness in CO_2-induced effects on globally or regionally averaged results, we use the signal-to-noise ratios (S/N) method (e.g., IPCC 2013, 2021; Lee et al. 2023a) which divides the ensemble-averaged changing signal (S) by their spread (N). The S/N values of three, two, and one are equivalent to 99.9%, 97.7%, and 84.1% probabilities of one-sided changes in normal distribution, respectively. For grid scale, we performed a one-sample t test after calculating paired differences from ensembles in different CO_2 conditions.

3. Results

a. February 2019 extreme events in North America

The stratospheric temperature over the Arctic increased dramatically from December 2018 to January 2019. The weakening of the polar vortex continued till early February and inversely recovered in March, perturbing the latitudinal position of the jet stream in February (Lee and Butler 2020). The polar vortex, the global warming trend, and the positive phase of El Niño induce both anomalous monthly mean patterns

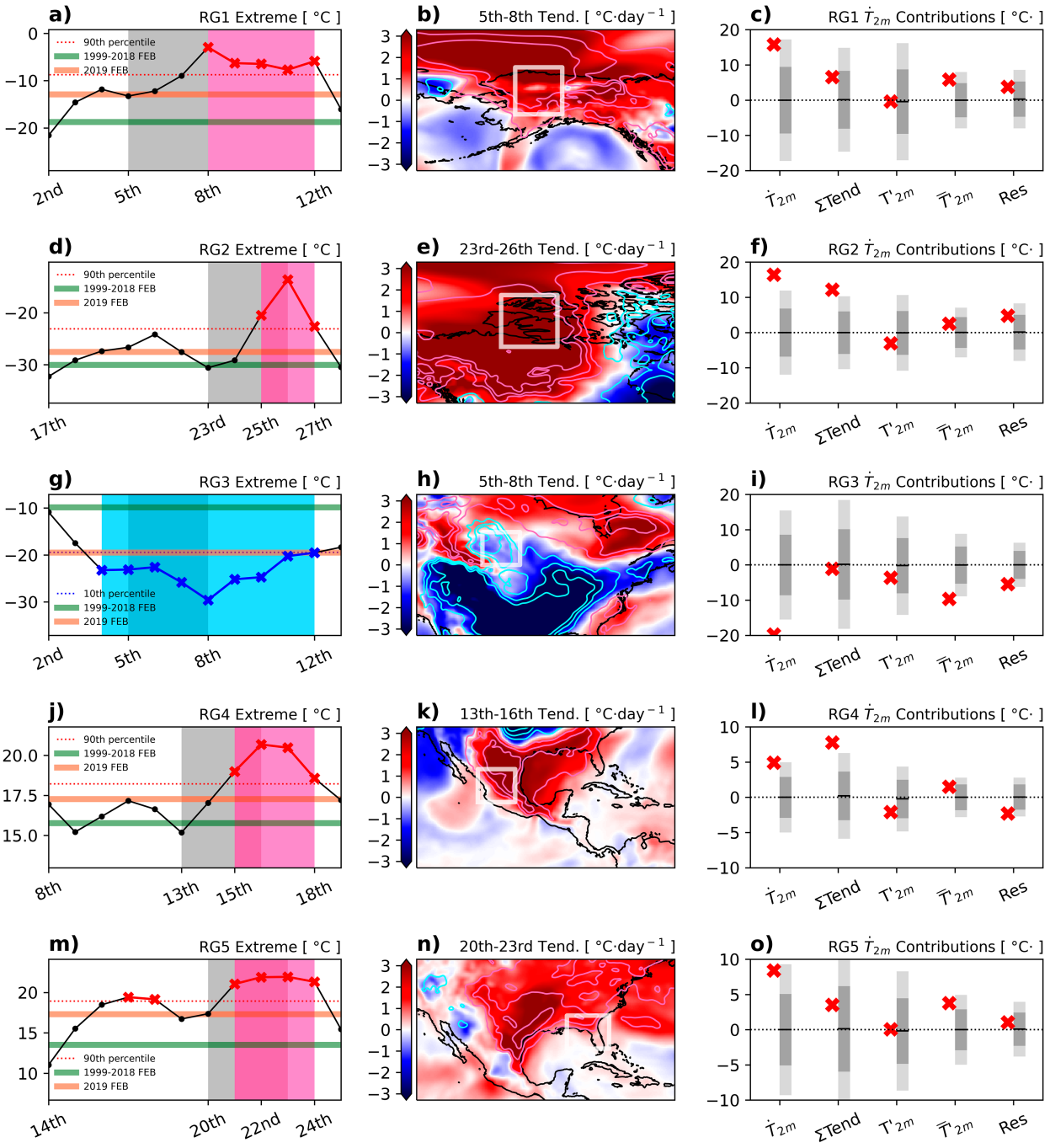

FIG. 2. (a) The 2-m surface air temperature (T_{2m}) time series before and after record warm extreme events over RG1. Climatology (1999–2018) and 2019 February mean are shown together with horizontal solid lines. The red dotted line denotes the heatwave definition threshold (90th percentile). Heatwave events are highlighted with multiplication marks. Vertically filled areas with pink and gray represent the heatwave event periods including the peak of extreme event (TX) and 3 days before TX. (b) Spatial patterns of temperature Tend in the atmosphere and surface are illustrated by shading and contours. Contour levels are +1°, +2°, and +3°C (−1°, −2°, and −3°C). (c) Inland-averaged T_{2m} anomaly and contributions from the sum of Tend during the 3 days, transient eddy, monthly mean anomaly, and residual, respectively [see Eq. (12)]. Gray and dark gray bars represent $\pm\sigma$ and $\pm 2\sigma$ ranges from climate period, respectively. Multiplication marks denote the values at the peak date (8 Feb) over RG1. [(d)–(f) and (j)–(o)] As in (a)–(c), except for the different regions (RG2–RG5) experiencing different extreme events at different dates (check Fig. 1). (g)–(i) As in (a)–(c), except for the cold extreme event. (g) The threshold for cold spell events is the 10th percentile. Instead of red and pink colors, blue and light blue are used for contrast to the warm extremes in the other regions.

TABLE 1. The decomposition of 2019 February regionally averaged T_{2m} anomalies (°C) in the peak date of the extreme events [$\dot{T}_{2m}$, see Eq. (12)]. The percentile of each value relative to climatology (1999–2018) is shown together in parentheses.

Region (date)	ΣTend	$\overline{T}'_{2m}$	$\overline{T}'_{2m}$	Res	$\dot{T}_{2m}$
RG1 (8)	+6.5 (84.3)	−0.4 (49.0)	+5.9 (100.0)	+3.8 (87.6)	+15.8 (99.8)
RG2 (26)	+12.2 (99.8)	−3.1 (25.2)	+2.5 (85.0)	+4.8 (90.4)	+16.4 (99.8)
RG3 (8)	−1.1 (42.9)	−3.6 (28.1)	−9.6 (0.0)	−5.4 (1.1)	−19.7 (0.0)
RG4 (16)	+7.8 (100.0)	−2.1 (16.7)	+1.5 (100.0)	−2.3 (0.5)	+4.9 (100.0)
RG5 (23)	+3.5 (72.8)	+0.1 (50.7)	+3.8 (95.0)	+1.0 (73.3)	+8.4 (99.1)

(Fig. 1a) and several warm and cold extreme events in North America (Fig. 1b). With very active upper-level circulation during February, the most severe events happened in various regions at different phases of wave circulation. Compared to the T_{2m} February historical record (1950–2018), many regions experienced extreme weather in 2019 ranging from the 1st to the 15th warmest or coldest rank (Fig. 1b).

Out of those regions, five regions were selected as the target of this study because they experienced highly ranked severe extreme events between 8 and 27 February 2019 within the simulation periods of perturbed CO_2 experiments (Fig. 1b). The anomalous characteristics of each extreme event are analyzed, given the growth and decay processes of the weather system (Fig. 2). The T_{2m} in five regions exceeded (fell) the 90th (10th) percentile of 20 years climate (Figs. 2a,d,g,j,m). The T_{2m} raised (dropped) from approximately their climatological or monthly mean levels to the highest (lowest) temperature within 3 days before the peak record occurred (gray shaded area in Figs. 2a,d,g,j,m). We check the similarity in tendency between T_{2m} and T_{Atm} to proclaim the validity of our assumptions in Eqs. (6) and (7) (Figs. 2b,e,h,k,n). The tendency patterns of T_{2m} (contour) and T_{Atm} (shading) were similar, having statistically significant correlation between them (1% significance level).

We decompose the anomalous T_{2m} value [Eq. (12)] and compare each term with its climatological ranges (Figs. 2c,f,i,l,o). We assess each contribution as typical or atypical by comparing its values with climatological spread (S/N, see section 2). Three regions (RG1, RG3, and RG5) show that the synoptic weather contribution (ΣTend) is within a standard deviation of climatology, indicating the usual synoptic weather system. Their extreme events are primarily driven by spatiotemporarily larger-scale variations affecting monthly anomalies, such as the positive phase of El Niño–Southern Oscillation (e.g., S. Lee et al. 2023), the negative phase of the Pacific–North America teleconnection (e.g., Manthos et al. 2022), and global warming (e.g., Zhang and Boos 2023). These are illustrated in monthly anomaly contributions ($\overline{T}'_{2m}$) where these values of the events are located outside of a standard deviation. On the other hand, RG2 and RG4 display ΣTend before extremes are larger than a standard deviation. This group highlights the anomalous synoptic weather system, dramatically raising the temperature. Each contribution size and its percentile in climatology are summarized in Table 1. The typical synoptic weather group members (RG1, RG3, and RG5) exhibit the first or second largest contributions from monthly anomalies, while the atypical synoptic weather group members (RG2 and RG4) exhibit the largest contributions from the transient eddy to anomaly in T_{2m} with its percentile beyond 99.8. The quantified sizes and percentile values in Table 1 reaffirm the different atmospheric mechanisms behind each regional extreme event (e.g., Fig. 2).

In addition to this, we summarize the 2019 February extreme events in five regions from two perspectives: grid and regionally averaged sense (Table 2). Higher-latitude regions (RG1, RG2, and RG3) denote about half of the grids experiencing extreme rank levels between the 4th and 15th, while those of low-latitude regions (RG4 and RG5) show between the 1st and 3rd. These latitudinal extreme rank contrasts in the grid scale correspond to the previous reports based on the observed situation datasets (Comisión Nacional del Agua 2019; NOAA 2019). Although portions of the extreme ranks in grids vary, ranging from 1st to 15th, in a regionally averaged sense, all five regions are within the top 5 ranks of extreme temperature records.

b. CO_2 perturbation effects in globe

To understand how CO_2 perturbation modifies the operation forecast simulations, we first check the global averaged T_{2m} by comparing the ensembles of INC and PIC (Fig. 3), which have less uncertainty than on a regional scale. Although different initial conditions are applied, the increased amount of CO_2 raises

TABLE 2. The percentage of the area experiencing different ranks of TX (TN) in grid scale (check Fig. 1b) and the rank of regionally averaged T_{2m} during the 2019 heatwave (cold spell for RG3) relative to the highest (lowest) regionally averaged T_{2m} in February of previous years (1950–2018). The grids outside of the land are excluded in this analysis.

Rank	First	Second	Third	Fourth	Fifth	Sixth –fifteenth	Regional extreme
RG1	0.8%	2.8%	7.8%	6.9%	6.6%	39.2%	Fourth
RG2	0%	0.1%	3.7%	21.4%	25.6%	46.9%	Fourth
RG3	6.2%	5.7%	12.0%	12.5%	10.4%	32.7%	Fifth
RG4	40.8%	12.1%	6.0%	5.6%	4.8%	22.2%	First
RG5	10.2%	20.7%	17.0%	12.4%	3.7%	28.6%	Fifth

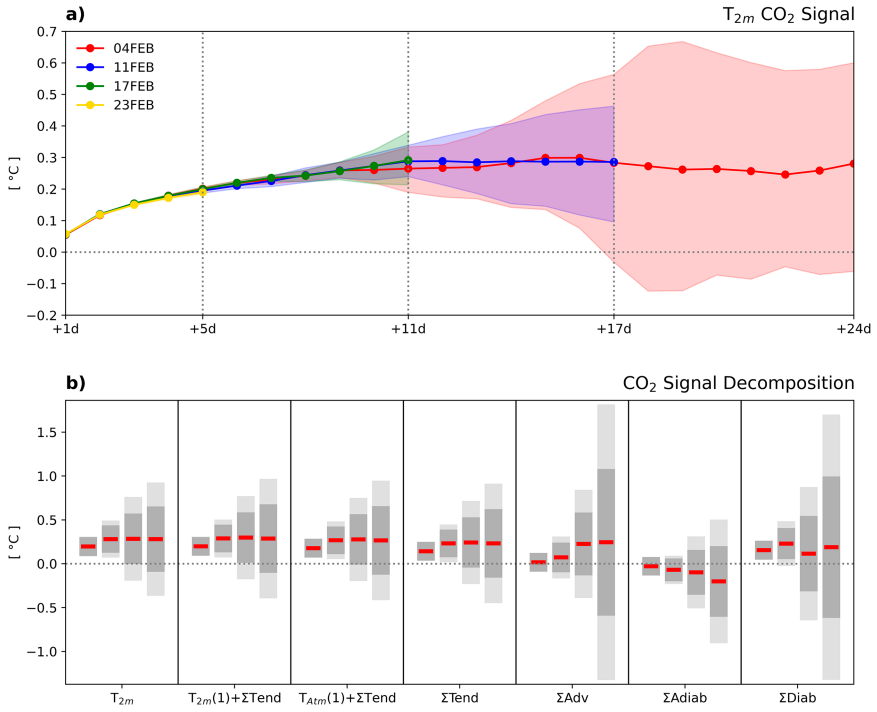

FIG. 3. (a) Time-evolving CO_2 signal in global mean T_{2m}, by subtracting results of piControl experiment from those of increased CO_2. Four different colors are applied for the different initiate dates. Solid lines and shaded areas represent 51 ENS' mean and spread (10th–90th percentile). (b) CO_2 signal in T_{2m} and decomposed contributions from Eq. (16). Three evolution times are examined [(left to right) CO_2 decomposition at +5d, +11d, +17d, and +23d]. Horizontal red lines represent the grand ensemble averages from 204, 153, 102, and 51 available ensemble members with different initiate dates. Shaded areas of dark and light gray indicate $\pm 1\sigma$ and $\pm 2\sigma$ ranges of ENS, respectively.

T_{2m} gradually until the 11 simulation days with good agreement. After the stabilization around 0.3°C warming, the ensemble spread gradually increases, eventually making the CO_2 signal ambiguous, located within the S/N ratio $< \pm 1$ after 17 simulation days (Fig. 3a). To understand the source of rising temperature and uncertainty details, we investigate T_{2m} signal decomposition as introduced in Eq. (16). The CO_2 signal in T_{2m} excellently matches T_{Atm} responses because the simulation time scale in forecast experiments is too short to induce distinct changes in land–atmosphere interactions (e.g., Meehl et al. 2021). Noteworthy, as the simulation time increases, the CO_2 signal in T_{2m} is more similar to that in ΣTend (Fig. 3b), implying that the decomposition of ΣTend can spotlight the major factor changing the temperature. In +5d and +11d cases, the increases in T_{2m} are mainly explained by diabatic heating illustrating S/N larger than one in these cases (Fig. 3b). The main contribution to the CO_2 signal shifts from diabatic heating for +5d and +11d to horizontal advections for +17d and +24d and the ensemble standard deviation range of each term across the zero line, suggesting that the perturbed CO_2 eventually grows uncertainty in weather noises later than 10-day forecast.

Likewise, the vertical temperature responses to the perturbed CO_2 levels are explored by ΣTend and its decomposed terms (Fig. 4) instead of T_{Atm} itself [see Fig. 3b and Eqs. (15) and (16)]. According to the sudden increase in CO_2 and its greenhouse gas effects, stratospheric cooling and near-surface warming are confirmed with $S/N > 3$ for the first 11 simulation days, expecting distinct signals in changing stratospheric temperature (Fig. 4a). These clear signals originate from diabatic heating, illustrating similar levels of change (Fig. 4d). However, temperature responses in the troposphere become uncertain as weather and associated cloud characteristics vary during the adjustment period in perturbed CO_2 concentration (e.g., Turner et al. 2018). The contributions from adiabatic and advection are relatively low magnitude (Figs. 4b,c). Adiabatic contributions show cooling in low atmospheric levels except for 200 hPa, which offsets diabatic cooling, resulting in near-zero changes in 200 hPa. Horizontal advections induce warming signals in the low atmosphere after 11 days. Ensembles display a large uncertainty in horizontal advections, similar to the T_{2m} results (Fig. 3b). The vertical analysis highlights that the CO_2 signal is evident in the stratosphere for several weeks. However, it is uncertain in the upper troposphere and the near-surface after 11 days of simulations. The S/N uncertainty differences between the stratosphere and troposphere are explained mainly by the differences in S (stratosphere > troposphere, e.g., Dong et al. 2009) because the sizes of N from each pressure level are similar and smaller than 0.4°C.

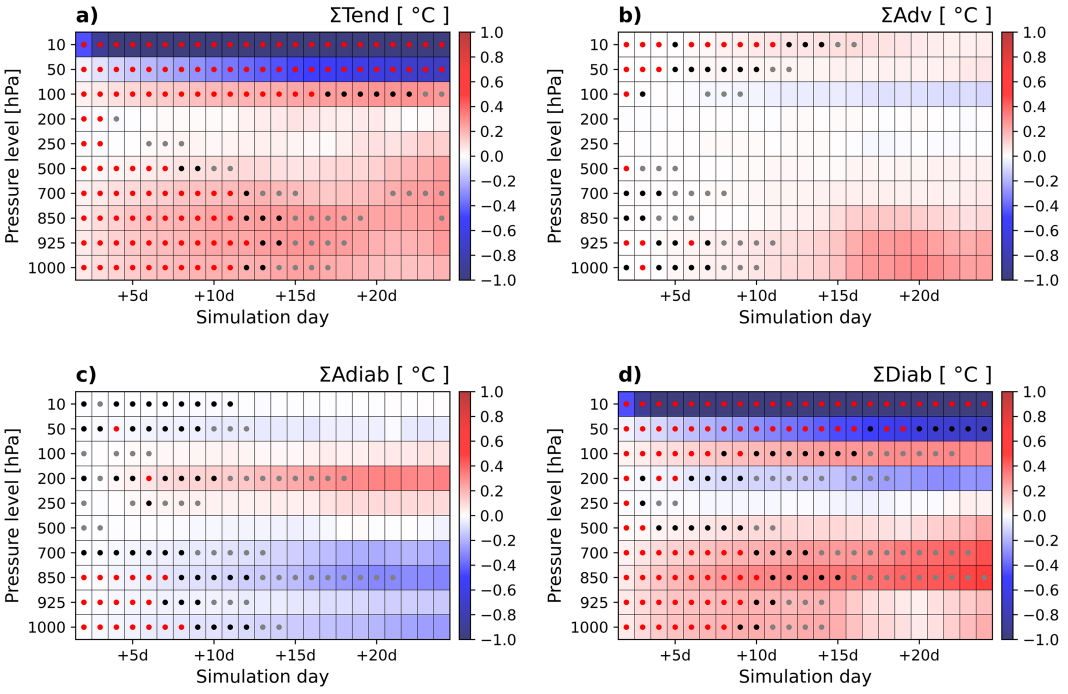

FIG. 4. (a) Time-evolving CO_2 signal in globally averaged atmospheric temperature in (ordinate) vertical levels from 51 ensemble members with the longest simulation days (initialization: 4 Feb 2019). (b)–(d) As in (a), except for the decomposed variables of (a) by temperature Tend equation. (a)–(d) All the Tend terms (°C day^{-1}) are shown in accumulated values (°C) from forecast day 2 to the day denoted in the abscissa. *S*/*N* in ENS is denoted with colored dots (red: $S/N > 3$, black: $S/N > 2$, and gray: $S/N > 1$).

c. *Extreme events in forecast ensembles*

The ECMWF IFS forecast ensembles with different initiate dates are compared to check the appropriate members applicable to the storyline analysis (Fig. 5). We also draw the CO_2 perturbed experiments to check forecast results with perturbing CO_2 concentration levels, which are conditional to the identical dynamic process. Although the forecast ensembles within a weak lead time show comparable performances in extremes, as the lead time increases, the other forecast ensembles barely catch the extremes at the heatwave or cold spell days in ERA5 results. In this case, ensembles are widely located around the monthly averaged T_{2m} of February 2019, shown by yellow horizontal lines in Fig. 5. Those results confirm that the forecast model makes it hard to predict actual extreme events if it has 2 or 3 weeks lead time. The process-constrained analysis is only capable for the ensembles having less than a week lead time prior to target events (e.g., Dai et al. 2021; Leach et al. 2021). The increased CO_2 raises T_{2m} within a week. However, these CO_2 effects are much smaller than a daily variation of T_{2m} in February, as shown in Fig. 2.

Therefore, we focus on the ensembles having a forecast lead time of less than a week to examine the storyline attribution. As Fig. 3b illustrates the importance of temperature tendency in understanding CO_2 effects, in Fig. 6, we investigate its decomposed size over each region to check the performance of the forecast model and CO_2-perturbing effects details in the process-constrained synoptic weather system. First, we check the model performance for the decomposed weather system. As the lead time is less than a week, the simulated atmospheric tendency before the extreme event has the same sign as the ERA5 result, except for the RG3 case. Over the RG3, the tendency of ECMWF IFS shows the opposite signs compared to ERA5, illuminating that the ECMWF IFS has a distinct bias in the synoptic weather system in this case, which is addressed in the later chapter (e.g., Fig. 9). Except for that, advection, adiabatic, and diabatic terms are reproduced well in forecast ensembles. The major atmospheric processes contributing to regional extremes differ depending on the planetary wave phase along the subtropical jet stream (Fig. 1). For example, the horizontal warm (cold) advection is the primary factor raising (dropping) T_{Atm} over RG2 and RG5 (RG3), while dry or moist dynamics mainly lead to warmer T_{Atm} over RG1 or RG4. CO_2-perturbing effects in atmospheric conditions are illustrated together. CO_2-induced changes in decomposed terms of tendency are much smaller than the daily variation of T_{2m} (e.g., Fig. 2). Although globally averaged responses in near-surface have clear CO_2 signals ($S/N > 1$) within a week (Fig. 4), these on a regional scale are much diverted, even in a diabatic heating response (e.g., Diab of RG3: ENS > INC, RG4: PIC > ENS). We highlight that the CO_2 signal in decomposed terms of tendency over regions is inhomogeneous, unlike the globally averaged cases (cf. Figs. 3 and 4).

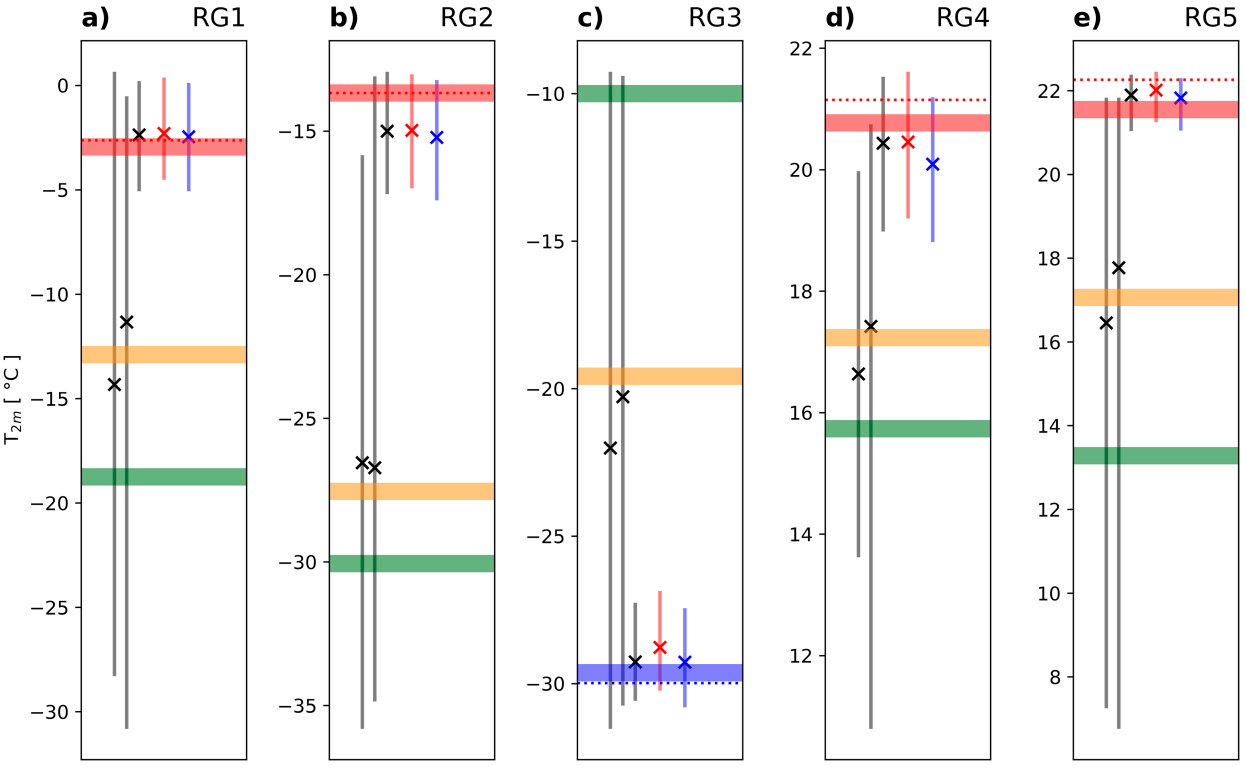

FIG. 5. ECMWF IFS ensemble results for (a),(b),(d),(e) TX during heatwaves and (c) TN during cold spell period. Multiplication marks and vertical solid lines indicate the mean and minimum and maximum ranges from 51 ENS, respectively. Different colors indicate operational (black) and perturbed CO_2 experiments (red: increased CO_2 and blue: preindustrial CO_2). Please note that the results are computed from the ENS with different initiate dates: (left to right) about 3 weeks, about 2 weeks, and within a week lead time before the peak date. For the results of perturbed CO_2 experiments, only ensemble members within a week lead time are shown. The field-averaged values of T_{2m} climatology (1999–2018), monthly mean T_{2m} for February 2019, and T_{2m} in the peak date during the extreme events are computed from ERA5 and distinguished by horizontal lines with colors (green: climatology, orange: February of 2019, red: field-averaged warmest record in a day, and blue: field-averaged coldest record in a day). Red (blue) dotted lines denote the TX (TN) events during the heatwave (cold spell) days to check coincidence over regional extreme events.

d. Probabilistic assessment in forecast ensembles

As an apparent benefit of the forecast-based attribution experiments is large ensemble members with weather systems constrained by forecast lead time, we investigate the statistical assessment for the extreme events in different forcing conditions. By applying GEV fitting to the ensembles, we estimate the likelihood of the extreme events in ENS, INC, and PIC and check the CO_2 signal in paired sets of ensembles (Fig. 7). Bootstrapping is applied to reflect the uncertainty ranges of GEV fitting. The parameters of GEV are summarized in Table 3. Since the forecast days are less than a week, GEV distributions of ENS, INC, and PIC are positioned around the levels of each regional extreme event, similar to the ensemble spread in Fig. 5, proving the constrained results in synoptic weather processes. We examine the CO_2 signals in two different sets: the differences between 1) INC and PIC (INC–PIC) as shown in the previous figures (Figs. 3 and 4) and 2) ENS and PIC (ENS–PIC) to check consistency in additional CO_2 effects in regional scales. The comparison between these two demonstrates whether the regional CO_2 signal within a week still matches with theoretical linearity (e.g., Cubasch et al. 2001) or not because INC, ENS, and PIC have different CO_2 concentration levels, which are experimentally designed to have double CO_2 radiative forcing in the first set than in the second one (see section 2).

The GEV distribution of CO_2-perturbed experiments (INC and PIC) reflects the effects of the CO_2 signal on their location parameter [μ in Eq. (17)] with a positive shift (0.0° ~ +0.5°C, Table 3). If CO_2 perturbation responses in a weekly were perfectly proportional to the designed difference in CO_2 radiative forcing, the CO_2 signal of the first set would be located around two times larger than its second set, according to the linear relation between the additional radiative forcing and temperature changes on a yearly scale (e.g., Hausfather et al. 2020; Tsutsui 2020). However, the regional CO_2 signal on a weekly scale looks nonlinear (Fig. 7). Our results illustrate that a rise in CO_2 concentration affects atmospheric weather, resulting in a warmer surface. However, the CO_2 signal in the region cannot be simplified as a global mean response.

As GEV distribution changed (Table 3), we investigated the risk ratios in regional extreme events associated with

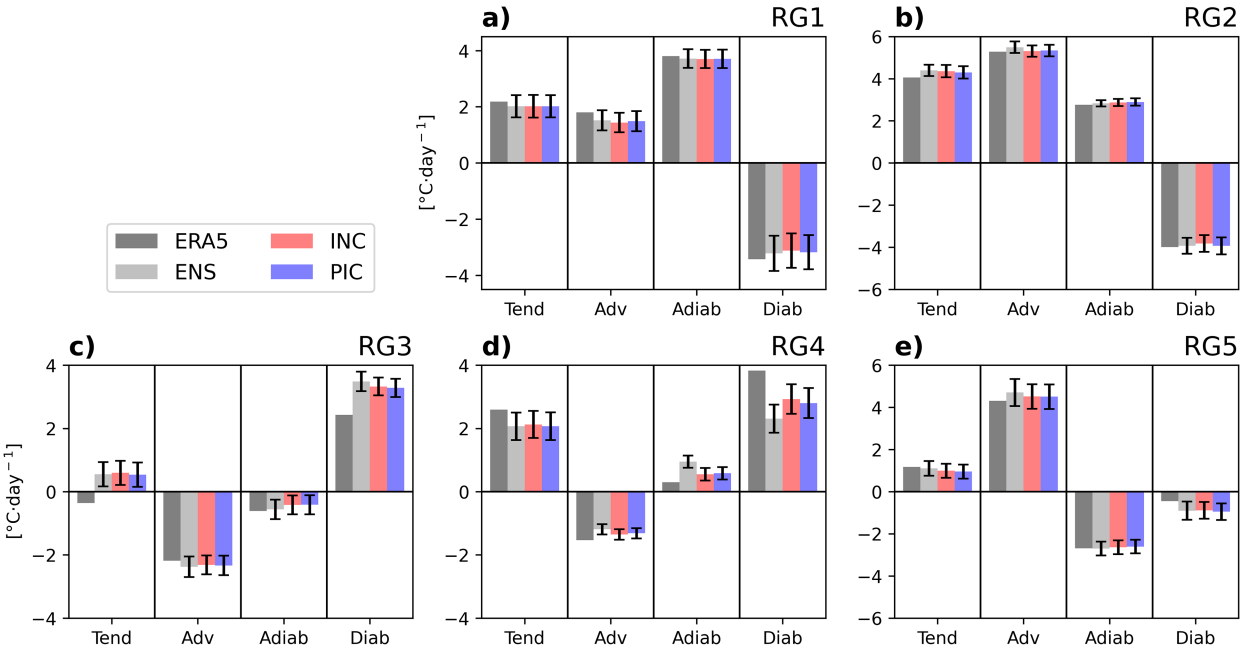

FIG. 6. (a),(b),(d),(e) Three-day averaged terms of atmospheric temperature Tend before the warmest day in the heatwave event. (c) As in (a), except for the atmospheric temperature Tend before the coldest day in the cold spell event that occurred over RG3. Range bars denote standard deviation ranges of ENS.

increased CO_2 levels (Fig. 8) under the storyline approach. Increases in CO_2 generally cause local warming, which is favorable for heatwaves (RG1, RG2, RG4, and RG5) and unfavorable for the cold spell (RG3). Most of the results correspond to these general expectations, except for the RG3 in the ENS–PIC set, where the CO_2 signal is nearly zero (Fig. 7f). According to the bootstrapped uncertainty ranges, the most significant and precise CO_2 attribution is confirmed in the heatwave over RG4. The increased CO_2 provokes about 27 times more risks for both sets of experiments. Extremes over low latitudes (RG4 and RG5) illustrate statistically more significant results than those over higher latitudes (RG1–RG3), where a 90% confidence interval ranges across a unit probability.

RG3 results are different from others, such as the forecast model performance (Fig. 6c), the unclear CO_2 signal between ENS and PIC (Fig. 7f), and the inconsistency of risk-based probability (Fig. 8c) from different sets. Hence, we conduct an additional analysis of systematic bias effects in extreme event cases over RG3. After that, we suggest the three conditions of a forecast-based attribution study, fulfilling the storyline and probabilistic frameworks coincidentally.

e. Conditions for the forecast-based storyline approaches

The statistical assessment is essential to confirm the significance of CO_2 effects under the dynamic constraint in the storyline approach as the weather noise in ensembles grows as the forecast days increase. The systematic bias in the forecasting system induces the disparity in reanalysis. Also, the unconstrained CO_2 signal impairs the findings of attribution results due to much larger uncertainty sizes. We investigate further extreme cases over RG3, where forecast ensembles fail to reproduce the atmospheric process affecting local extremes (Fig. 6c) and where the increased CO_2 effects from the ENS–PIC set are near zero (Fig. 7f). Figure 9 displays the ECMWF IFS results with an initialization date of 4 February 2019, which are used to examine the extreme case study over RG3. As shown in Fig. 2g, the regionally averaged temperature dramatically dropped between 6 and 8 February.

We check the ensemble results relative to the reanalysis dataset, hereinafter referred to as bias (Figs. 9b,c). The forecast model generally shows reasonable performances until the sixth. However, this model has a distinct cold bias on the seventh while a warm bias on the eighth when the minimum low temperature mainly occurs. These biases induce some ensembles to get TN on 7 February, not as the expected processes in reanalysis but a day earlier (Fig. 9a). The size of the bias in T_{2m} is comparable to that in T_{Atm}. This suggests that the bias in this extreme event significantly originated from the atmospheric process. We investigate the details of the temperature-varying process (Figs. 9b,c). The land–atmosphere interaction size is estimated by subtracting T_{Atm} from T_{2m} [Eq. (7)]. A slightly colder bias exists on 7 February, which makes some ensembles have TN earlier. However, the ensemble spread crosses the zero line, indicating statistically insignificant bias (Fig. 9b). The problem is severe in the eighth case (Fig. 9c). The ensemble spread shows that bias in T_{Atm} is the most prominent source affecting 3°C warmer in T_{Atm} than the values from ERA5. The ensemble mean bias on the eighth is much larger than the ensemble spread (σ), indicating significance in ensemble bias. The decomposed tendency clarifies that the CO_2 local effects are heterogeneous between INC–PIC and ENS–PIC cases. We also checked the spatial patterns of the

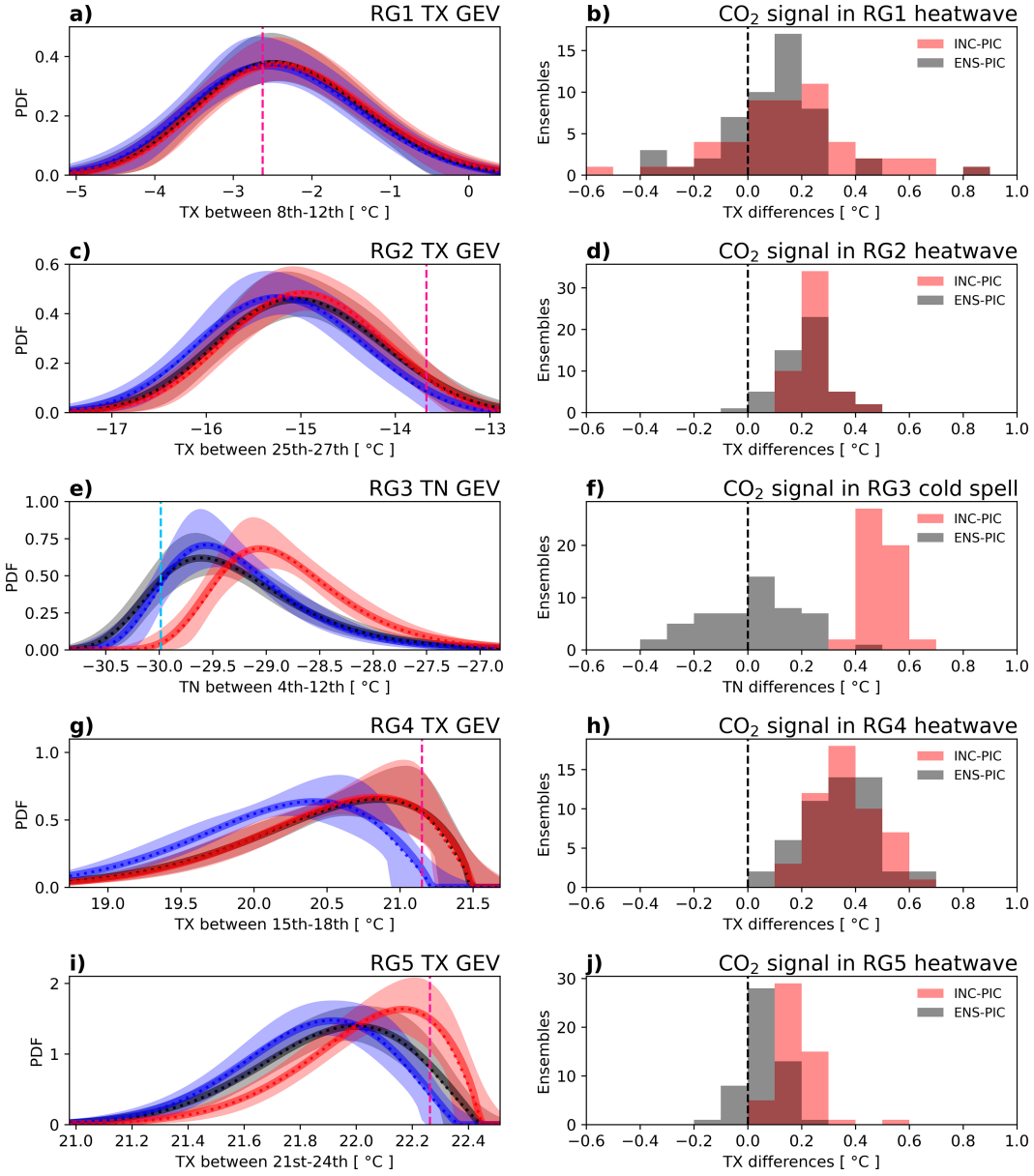

FIG. 7. (a) Solid (dashed) lines with colors indicate the GEV distribution for extreme events in RG1 based on the 51 ENS members (median values of 100000 bootstrapping samples) of the forecast (black) and CO_2-perturbed experiments (INC: red and PIC: blue). 90% confidence interval ranges of GEV distributions are estimated by the bootstrap method and shown by shading area. Vertical dashed lines indicate the target extreme event levels. (b) Histograms for the CO_2 signal in extreme events over RG1. Samples are computed from the 51 paired ensemble sets of INC and PIC (red) and ENS and PIC (black). (c)–(j) As in (a),(b), except for the different extreme events that occurred in different periods and regions (RG2–RG5, Fig. 2).

CO_2 signal from ENS–PIC (not shown). Since their differences in T_{2m} are not overall positive signals as INC–PIC case until the fifth simulation day, it confirmed that about 125 ppm CO_2 concentration differences (ENS–PIC) may not be enough to promise the statistically significant positive signals in the early period, larger than the weather noises.

The local tendency patterns on 7 (Figs. 9d,e) and 8 February (Figs. 9g,h) are investigated to understand the process differences between simulations and reanalysis. The tendency patterns are similar between ENS and ERA5 on the seventh. However, its atmospheric decomposition notably has hidden biases in advection and diabatic, offsetting each other (Fig. 9b). The errors in forecast ensembles grew on 8 February. The temperature tendency over RG3 in reanalysis is slightly negative, while it in ENS is positive. We checked which factor explains the biased patterns over tendency (Figs. 9f,i), affirming that the biases in meridional or

TABLE 3. GEV parameter sets for the regionally averaged TX and TN in February 2019 extreme events. The type of extreme event is described in parentheses. GEV is fitted from the 51 ensemble members of each initialization group within a week before the peak extreme day. The initialization date and peak date of the extreme for each event are shown in parentheses.

GEV parameters (extreme events)	Location (μ, °C)			Scale (σ, °C)			Shape (ξ)		
	INC	ENS	PIC	INC	ENS	PIC	INC	ENS	PIC
RG1 (heatwave) (4 Feb 2019 | 8 Feb 2019)	−2.70	−2.75	−2.85	1.01	1.01	1.03	0.23	0.24	0.24
RG2 (heatwave) (23 Feb 2019 | 26 Feb 2019)	−15.25	−15.31	−15.51	0.80	0.84	0.83	0.29	0.27	0.29
RG3 (cold spell) (4 Feb 2019 | 8 Feb 2019)	−29.07	−29.61	−29.57	0.54	0.59	0.52	0.02	0.00	0.01
RG4 (heatwave) (11 Feb 2019 | 16 Feb 2019)	20.33	20.31	19.95	0.70	0.73	0.70	0.61	0.63	0.56
RG5 (heatwave) (17 Feb 2019 | 23 Feb 2019)	21.97	21.82	21.76	0.28	0.31	0.29	0.60	0.49	0.47

eastward wind speeds over RG3 in ENS promote the wave propagation phase faster than in reanalysis, bringing earlier TN in some ensembles. This atmospheric weather bias limits the validity of the storyline framework application for the extreme events over RG3.

As the bias and CO_2 signal limit storyline application of forecast ensembles, we numerically evaluate the appropriate lead time for a forecast-based storyline approach. The size of both noise in CO_2 signal and bias in ENS over land area grow with lead time increases, illustrating the low compatibility of the storyline approach for the longer predictions (Fig. 10). Thus, we consider two conditions to find the appropriate dates of storyline approaches: 1) CO_2-positive signal in a grid is significant at the 10% confidence level among the ensembles and 2) bias size in grids is <2°C in an ensemble-averaged sense. Figures 10a–f illustrate the patterns of CO_2 signal and ENS bias from the ensembles initialized from 4 February. CO_2-positive signals (INC–PIC) over RG1 and RG3 are significant on the extreme event date (8 February), as confirmed in Fig. 7, while the forecast ensemble bias over RG3 limits the storyline approach applying for the TN extremes, as investigated in Fig. 9. As lead time increases, the CO_2 signals in grids grow, but contemporarily, the bias and uncertainties among ensembles are much proliferated. By comparing the ensembles with different initialization dates (Fig. 10g), we confirmed the similar characteristics in forecast ensembles regardless of their initial conditions. Although the spatial patterns vary among forecast members, CO_2 signal (bias) conditions generally quantify at least 50% over the land regions until the +10 (+6) lead time. We further compute the area where these two conditions qualified (Fig. 10h). The storyline framework of the forecast system is capable of about a half-land area within a week lead time earlier to extreme event occurrence. Although the fraction fulfilling both conditions is similar among the ensembles in different initialization dates, showing consistency, the preinspection of the model performance on the target extreme event is essential because the bias pattern varies. As much of the applicable area is limited by the bias conditions (Fig. 10g), we stress that the stabilized performances in the forecast model are essential to lead the universal application of this framework.

4. Summary and discussion

The background warming from the rising CO_2 concentration induces favorable conditions for warm extremes. However, the warming-driven melting of Arctic ice weakens the polar vortex. It perturbs the position of the polar jet, resulting in drastic variation in midlatitude winter weather (Overland et al. 2021; Cohen et al. 2021) interacting with different phases

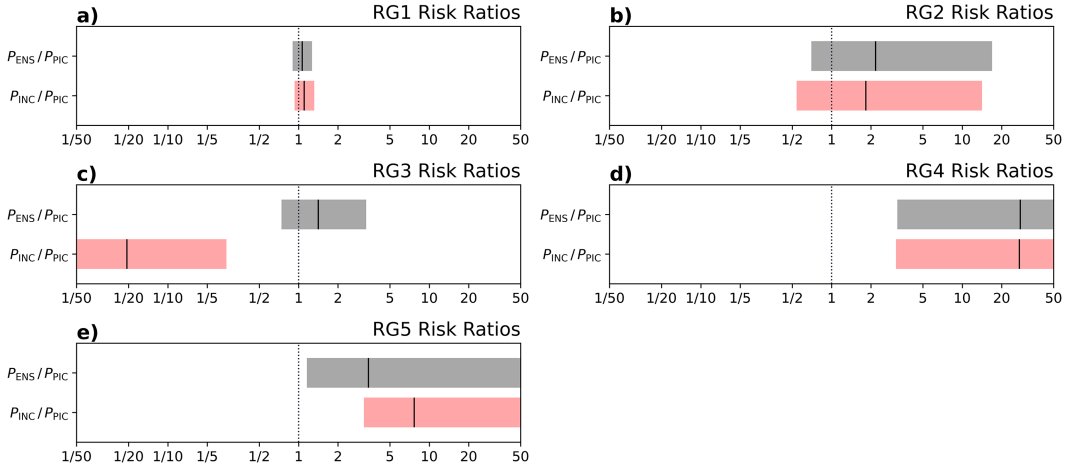

FIG. 8. (a)–(e) Boxplots of the risk ratios using two sets of attributions in probabilistic: ENS versus PIC (black) and INC versus PIC (red) as the factual versus counterfactual conditions over five extreme events. Shading (black solid lines) indicates 90% confidence interval (median) values from bootstrapped GEV samples in Fig. 7.

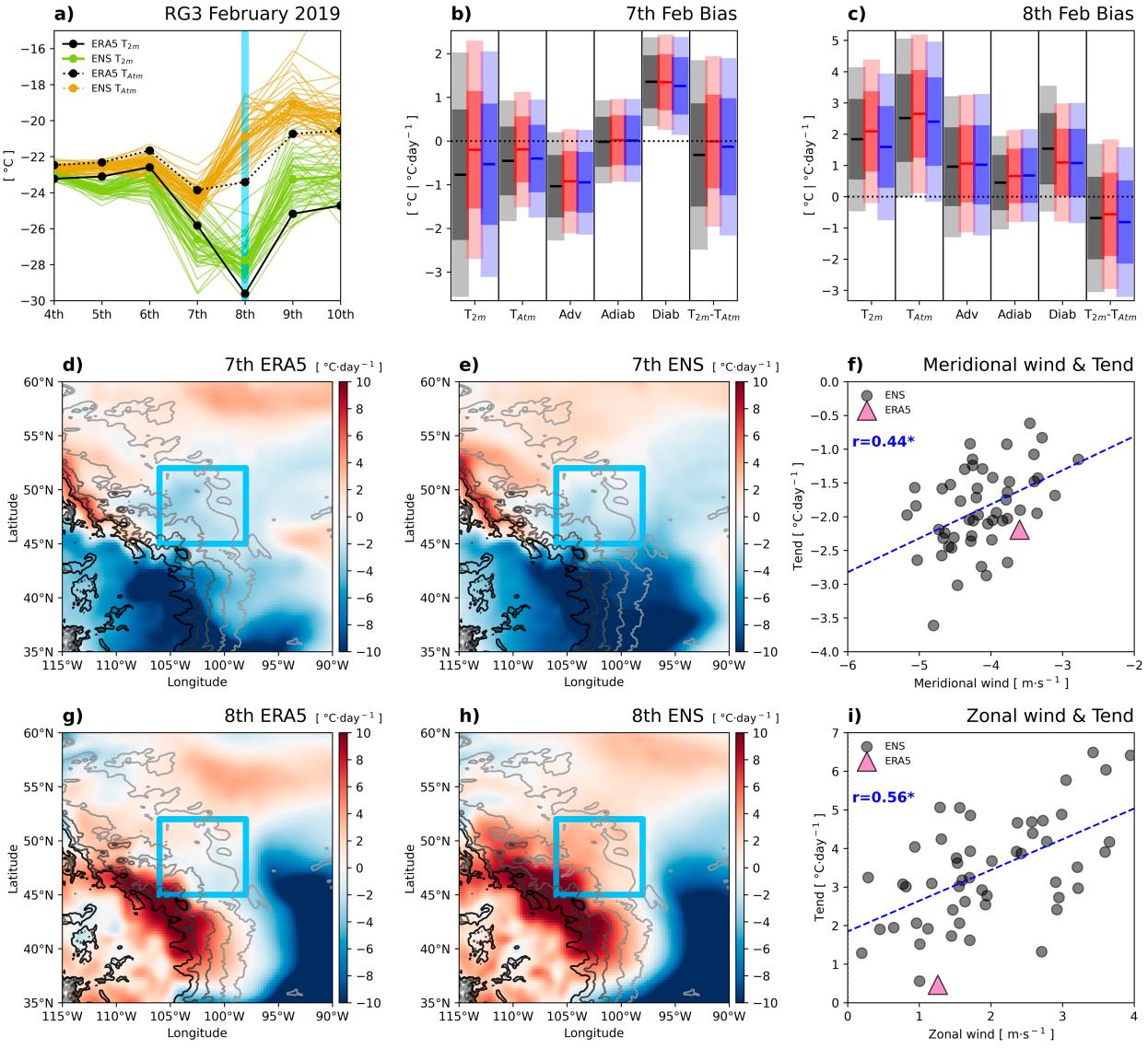

FIG. 9. (a) Time series of T_{2m} and T_{Atm} in the cold spell period over RG3 from ERA5 (black) and 51 members of ENS initiated from 4 Feb 2019 (green and khaki). The 8th is highlighted by the vertical thick line when most TN events occurred. (b) Bias in ENS, INC, and PIC results in different variables on the 7th. (c) As in (b), except for 8 Feb. (d) T_{Atm} Tend patterns for the 7th case from ERA5. Contours denote the complex topography around RG3 ranging from (light gray) 500 m to (black) 1500 m with every 250-m intervals with contrasts. (e) As in (d), except for the data from the ensemble-averaged results of ENS. (f) Scatterplots for the meridional wind and Tend for the seventh case. The correlation coefficient between them is shown, and the 1% significance level is marked by an asterisk after its number (*). (g),(h) As in (d),(e), except for the event of 8 Feb. (i) As in (f), except for the zonal wind and Tend for the eighth case.

of natural variability (Yang et al. 2018; Soulard et al. 2019). As a result, the observed cold extreme trend characteristics over Central America are insignificant for the recent periods associated with unapprehensive dynamics in Arctic amplification (Cohen et al. 2023). Accordingly, the highly ranked polarized extremes have occurred frequently: the heatwaves (e.g., over the Southeast part of North America in February 2023, Erdman 2023) as well as the cold spells (e.g., over the south-central part of North America in February 2021, Cohen et al. 2021; NOAA 2021; Hsu et al. 2022). Notably, February 2019 has the opposite phase in two natural variability indices (Niño-3.4: 0.68 and PNA: −1.97), which is an exceptional case considering the significant positive correlation between them (e.g., Soulard et al. 2019). With these anomalous backgrounds, the warm and cold opposing extreme events contemporarily happened over North America in February 2019 (Fig. 1). The core synoptic weather systems or interannual variability effectively vary the daily temperature apart from its climatological mean state ranging from 4.9° to 19.7°C (Fig. 2 and Table 1), causing the severe drop (RG3) or rise in temperature (RG1, RG2, RG4, and RG5). The analysis of the atmospheric temperature tendency quantified the size of contributions to the

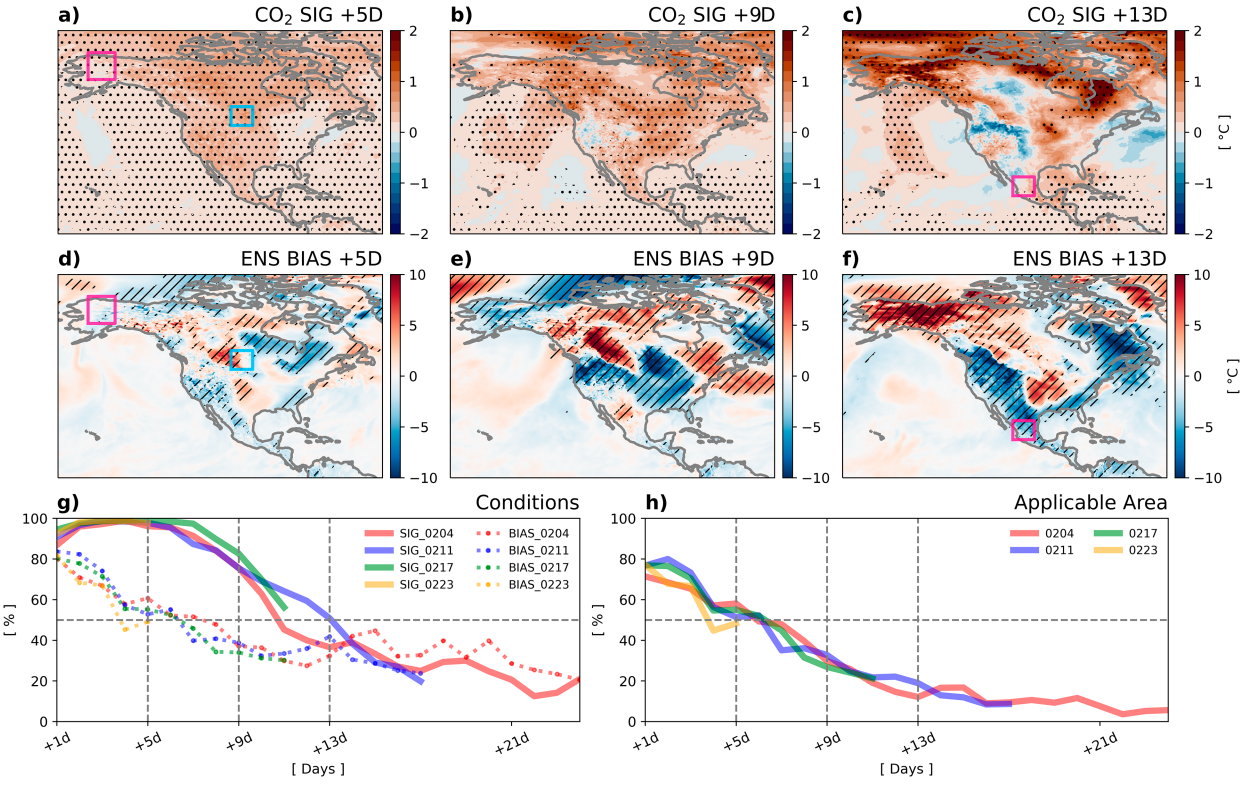

FIG. 10. (a)–(c) CO_2 signals from INC and PIC, which were initiated on 4 Feb 2019 with different lead times (+5D, +9D, and +13D). Stipples in grids indicate a statistically significant positive signal at the 10% confidence level. (d)–(f) As in (a)–(c), except for the bias in ENS relative to ERA5. Hatches indicate the absolute size of bias is larger than 2°C. (g) Land area fraction fulfilling the condition of CO_2 signals (SIG: solid lines) or bias size (BIAS: dashed lines) with different lead times. Colors denote the ENS with different initialization dates. (h) As in (g), except for the area fraction fulfilling both conditions simultaneously.

extreme events, where the regions are affected dramatically by abnormal atmospheric synoptic weather (RG2 and RG4) or not (RG1, RG3, and RG5). The relative contribution size of adiabatic, diabatic, and advection in our results in Eulerian perspective is analogous to the results in previous studies based on the Lagrangian perspective (Röthlisberger and Papritz 2023a,b).

We analyzed the atmospheric CO_2 attribution in each target extreme event based on the operational forecast model, ECWMF IFS, simulations by perturbing CO_2 concentration settings (Leach et al. 2021). As reported in the global climate model study (Dong et al. 2009), the changes in atmospheric CO_2 level lead to surface and troposphere warming but stratosphere cooling in forecast simulations. Although our experiments do not resolve any effects of human-induced changes in ocean conditions, perturbed CO_2 experiments illuminate increases in CO_2, causing a higher (lower) chance for warm (cold) extremes. The extremes in low latitudes (RG4 and RG5) show statistically significant results. In contrast, those in high latitudes show insignificance in magnitude or large uncertainties. We state two reasons for the difference in probability risk in regions. First, larger daily T_{2m} variability in high latitudes (Figs. 5 and 7 and Table 3) likely inhibits the statistical significance of the CO_2 signal, which is less than a degree Celsius. Second, the atmospheric process inducing positive CO_2 signals differs on regional scales (Figs. 7 and 10), further complicating how an increase in CO_2 concentration changes the weather in extreme days. The CO_2 signals of two different estimations (INC–PIC and ENS–PIC) are similar except for the case of RG3. We have checked the CO_2 signals from ENS–PIC as in Fig. 10 (not shown) and confirmed the ambiguous CO_2 signals in simulation days of one and four, showing significant CO_2-positive signals over less than half of the land area. Our experiment reveals that the early radiative energy disturbance may not promise linear-like clear signals (e.g., Cubasch et al. 2001) in regional patterns. INC–PIC set may show the clearer signal even because the concentration differences are much larger (315 ppm) than ENS–PIC (125 ppm). Regardless of constraints in dynamical weather systems, uncertainties remained in our results pending the discernible noise size of the CO_2 signal after seven simulation days and growing sizes in forecast errors, as confirmed by a previous study (Leach et al. 2021). Based on our findings, we suggest qualification guidelines for the storyline approach with forecast ensembles (Fig. 10). As the capable area decreases as simulation time increases, we recommend forecast-based storyline attribution studies to investigate extreme events within 6 days of forecast lead time to secure

performance in the atmospheric process and significant positive CO_2 signal together.

Our forecast-based attribution experiments have merits in perturbing forcing conditions and conditioning the dynamical noise size in the operation forecast model, which is the most reliable numerical simulation for the ongoing phenomena. However, our findings have several caveats. Since human-induced warming over the ocean can change global surface temperature much more than concentrations only (e.g., Dong et al. 2009), the additional ECMWF forecast experiments reflecting counterfactual ocean conditions are a compelling topic for the subsequent study. Spectral nudging technology is a potentially considerable option to promise better results in a much longer lead time in order to constrain the uncertainty and error sizes in forecasting (Hitchcock et al. 2022). Since CO_2 responses are determined by Earth systems' climate sensitivities (e.g., Lee et al. 2023b; Song et al. 2023), a single type of model with large ensembles may have a systematic bias to quantify the size of attribution. Like other attribution studies (e.g., Philip et al. 2020), future studies of multiforecast model–based assessment are required. We focused on the atmospheric process in our analysis, but it is noteworthy that the potential importance of ocean and soil conditions reinforces the extreme events. As Yoon et al. (2024) quantify these contribution sizes based on the series of model sensitivity simulations with different conditions, the additional forecast model experiments with the combinations of various boundary conditions are worth it as a future challenging topic.

Nonetheless of our caveats, due to the scarcity of previous studies focusing on storyline attribution, we highlight that our discoveries enhance the overall understanding of human-driven forcing effects on the extreme events with a medium-range operation forecast model. Our findings reaffirm general information on CO_2-perturbing effects on the globe and region on a monthly scale and provide the scientific guidance for application of forecast-based ensembles in storyline attribution study. As the additional CO_2 provokes a more favorable environment for the February 2019 severe heatwave occurrences over Europe (Leach et al. 2021), North America (Figs. 7 and 8), and the globe (Figs. 3 and 4), we argue that early fulfillment of CO_2 net-zero emissions eventually inhibits record-shattering warm extreme occurrences in the world.

Acknowledgments. This work was supported by the European Union's Horizon 2020 research and innovation program under Grant Agreement 821003 (project 4C). SO is supported by the United Kingdom Natural Environment Research Council (NERC) National Centre for Atmospheric Science (NCAS) and has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 101003469 (XAIDA). Acknowledgement is made for the use of ECMWF's computing and archive facilities in this research under the special project spgbleac. We thank Antje Weisheimer and Tim Palmer who contributed to the designs of CO_2-perturbed IFS experiments. We acknowledge the time and effort devoted by three anonymous reviewers and the editor.

Data availability statement. The ERA5 data used are freely available through the C3S Climate Data Store (https://cds.climate.copernicus.eu). Forecast ensemble datasets are available through the ECMWF archive (https://www.ecmwf.int/en/forecasts/datasets/ archive-datasets). Code is available upon request to the corresponding author.

REFERENCES

Barriopedro, D., R. García-Herrera, C. Ordóñez, D. G. Miralles, and S. Salcedo-Sanz, 2023: Heat waves: Physical understanding and scientific challenges. *Rev. Geophys.*, **61**, e2022RG000780, https://doi.org/10.1029/2022RG000780.

Bellprat, O., and F. Doblas-Reyes, 2016: Attribution of extreme weather and climate events overestimated by unreliable climate simulations. *Geophys. Res. Lett.*, **43**, 2158–2164, https://doi.org/10.1002/2015GL067189.

——, V. Guemas, F. Doblas-Reyes, and M. G. Donat, 2019: Towards reliable extreme weather and climate event attribution. *Nat. Commun.*, **10**, 1732, https://doi.org/10.1038/s41467-019-09729-2.

Blanchard-Wrigglesworth, E., T. Cox, Z. I. Espinosa, and A. Donohoe, 2023: The largest ever recorded heatwave—Characteristics and attribution of the Antarctic heatwave of March 2022. *Geophys. Res. Lett.*, **50**, e2023GL104910, https://doi.org/10.1029/2023GL104910.

Cohen, J., L. Agel, M. Barlow, C. I. Garfinkel, and I. White, 2021: Linking Arctic variability and change with extreme winter weather in the United States. *Science*, **373**, 1116–1121, https://doi.org/10.1126/science.abi9167.

——, ——, ——, and D. Entekhabi, 2023: No detectable trend in mid-latitude cold extremes during the recent period of Arctic amplification. *Commun. Earth Environ.*, **4**, 341, https://doi.org/10.1038/s43247-023-01008-9.

Comisión Nacional del Agua, 2019: Reporte del clima en México Febrero 2019. CONAGUA, 35 pp., https://smn.conagua.gob.mx/es/climatologia/diagnostico-climatico/reporte-del-clima-en-mexico.

Cubasch, U., and Coauthors, 2001: Projections of future climate change. *Climate Change 2001: The Scientific Basis*, J. T. Houghton et al., Eds., Cambridge University Press, 525–582, https://www.ipcc.ch/site/assets/uploads/2018/03/WGI_TAR_full_report.pdf.

Dai, G., M. Mu, C. Li, Z. Han, and L. Wang, 2021: Evaluation of the forecast performance for extreme cold events in East Asia with subseasonal-to-seasonal data sets from ECMWF. *J. Geophys. Res. Atmos.*, **126**, 2020JD033860, https://doi.org/10.1029/2020JD033860.

Deser, C., A. Phillips, V. Bourdette, and H. Teng, 2012: Uncertainty in climate change projections: The role of internal variability. *Climate Dyn.*, **38**, 527–546, https://doi.org/10.1007/s00382-010-0977-x.

——, A. S. Phillips, M. A. Alexander, and B. V. Smoliak, 2014: Projecting North American climate over the next 50 years: Uncertainty due to internal variability. *J. Climate*, **27**, 2271–2296, https://doi.org/10.1175/JCLI-D-13-00451.1.

Dirmeyer, P. A., G. Balsamo, E. M. Blyth, R. Morrison, and H. M. Cooper, 2021: Land-atmosphere interactions exacerbated the drought and heatwave over northern Europe during summer 2018. *AGU Adv.*, **2**, e2020AV000283, https://doi.org/10.1029/2020AV000283.

Dong, B., J. M. Gregory, and R. T. Sutton, 2009: Understanding land–sea warming contrast in response to increasing greenhouse gases. Part I: Transient adjustment. *J. Climate*, **22**, 3079–3097, https://doi.org/10.1175/2009JCLI2652.1.

Erdman, J., 2023: Winter heat wave smashes February records. Weather Channel, accessed 23 February 2023, https://weather.com/forecast/national/news/2023-02-20-february-heat-wave-monthly-records-southeast-cold-north-west.

Ermis, S., N. J. Leach, F. C. Lott, S. N. Sparrow, and A. Weisheimer, 2024: Event attribution of a midlatitude windstorm using ensemble weather forecasts. *Environ. Res. Climate*, **3**, 035001, https://doi.org/10.1088/2752-5295/ad4200.

García-García, A., F. J. Cuesta-Valero, D. G. Miralles, M. D. Mahecha, J. Quaas, M. Reichstein, J. Zscheischler, and J. Peng, 2023: Soil heat extremes can outpace air temperature extremes. *Nat. Climate Change*, **13**, 1237–1241, https://doi.org/10.1038/s41558-023-01812-3.

González-Herrero, S., D. Barriopedro, R. M. Trigo, J. A. López-Bustins, and M. Oliva, 2022: Climate warming amplified the 2020 record-breaking heatwave in the Antarctic Peninsula. *Commun. Earth Environ.*, **3**, 122, https://doi.org/10.1038/s43247-022-00450-5.

Hausfather, Z., H. F. Drake, T. Abbott, and G. A. Schmidt, 2020: Evaluating the performance of past climate model projections. *Geophys. Res. Lett.*, **47**, e2019GL085378, https://doi.org/10.1029/2019GL085378.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hitchcock, P., and Coauthors, 2022: Stratospheric Nudging and Predictable Surface Impacts (SNAPSI): A protocol for investigating the role of stratospheric polar vortex disturbances in subseasonal to seasonal forecasts. *Geosci. Model Dev.*, **15**, 5073–5092, https://doi.org/10.5194/gmd-15-5073-2022.

Hsu, P.-C., H.-H. Hsu, H.-J. Hong, Y.-T. Chen, Y.-L. Chen, and W.-L. Tseng, 2022: 2021 Texas cold snap: Manifestation of natural variability and a recent warming trend. *Wea. Climate Extremes*, **37**, 100476, https://doi.org/10.1016/j.wace.2022.100476.

IPCC, 2013: *Climate Change 2013: The Physical Science Basis*. T. F. Stocker et al., Eds. Cambridge University Press, 1535 pp., https://doi.org/10.1017/CBO9781107415324.

——, 2021: *Climate Change 2021: The Physical Science Basis*. V. Masson-Delmotte et al., Eds., Cambridge University Press, 2391 pp., https://doi.org/10.1017/9781009157896.

Kharin, V. V., F. W. Zwiers, X. Zhang, and G. C. Hegerl, 2007: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Climate*, **20**, 1419–1444, https://doi.org/10.1175/JCLI4066.1.

Kim, D. W., and S. Lee, 2022: Dynamical mechanism of the summer circulation trend pattern and surface high temperature anomalies over the Russian Far East. *J. Climate*, **35**, 6381–6393, https://doi.org/10.1175/JCLI-D-22-0244.1.

Kim, H.-J., S.-W. Son, W. Moon, J.-S. Kug, and J. Hwang, 2021: Subseasonal relationship between Arctic and Eurasian surface air temperature. *Sci. Rep.*, **11**, 4081, https://doi.org/10.1038/s41598-021-83486-5.

Kim, Y.-H., S.-K. Min, X. Zhang, J. Sillmann, and M. Sandstad, 2020: Evaluation of the CMIP6 multi-model ensemble for climate extreme indices. *Wea. Climate Extremes*, **29**, 100269, https://doi.org/10.1016/j.wace.2020.100269.

Lai, X., and Y. Gong, 2017: Relationship between atmospheric heat source over the Tibetan Plateau and precipitation in the Sichuan–Chongqing region during summer. *J. Meteor. Res.*, **31**, 555–566, https://doi.org/10.1007/s13351-017-6045-2.

Lazoglou, G., C. Anagnostopoulou, K. Tolika, and F. Kolyva-Machera, 2019: A review of statistical methods to analyze extreme precipitation and temperature events in the Mediterranean region. *Theor. Appl. Climatol.*, **136**, 99–117, https://doi.org/10.1007/s00704-018-2467-8.

Leach, N. J., A. Weisheimer, M. R. Allen, and T. Palmer, 2021: Forecast-based attribution of a winter heatwave within the limit of predictability. *Proc. Natl. Acad. Sci. USA*, **118**, e2112087118, https://doi.org/10.1073/pnas.2112087118.

——, and Coauthors, 2024: Heatwave attribution based on reliable operational weather forecasts. *Nat. Commun.*, **15**, 4530, https://doi.org/10.1038/s41467-024-48280-7.

Lee, D., and Coauthors, 2023a: Uncertainty analysis of future summer monsoon duration and area over East Asia using a multi-GCM/multi-RCM ensemble. *Environ. Res. Lett.*, **18**, 064026, https://doi.org/10.1088/1748-9326/acd208.

——, S. N. Sparrow, S.-K. Min, S.-W. Yeh, and M. R. Allen, 2023b: Physically based equation representing the forcing-driven precipitation in climate models. *Environ. Res. Lett.*, **18**, 094063, https://doi.org/10.1088/1748-9326/acf50f.

Lee, S., H.-S. Park, S.-Y. Song, and S.-W. Yeh, 2023: Distinct impacts of two types of El Niño events on northern winter high-latitude temperatures simulated by CMIP6 climate models. *Environ. Res. Lett.*, **18**, 034035, https://doi.org/10.1088/1748-9326/acbce9.

Lee, S. H., and A. H. Butler, 2020: The 2018–2019 Arctic stratospheric polar vortex. *Weather*, **75**, 52–57, https://doi.org/10.1002/wea.3643.

Lloyd, E. A., and T. G. Shepherd, 2023: Foundations of attribution in climate-change science. *Environ. Res. Climate*, **2**, 035014, https://doi.org/10.1088/2752-5295/aceea1.

Manthos, Z. H., K. V. Pegion, P. A. Dirmeyer, and C. Stan, 2022: The relationship between surface weather over North America and the mid-latitude seasonal oscillation. *Dyn. Atmos. Oceans*, **99**, 101314, https://doi.org/10.1016/j.dynatmoce.2022.101314.

Meehl, G. A., and Coauthors, 2021: Initialized Earth system prediction from subseasonal to decadal timescales. *Nat. Rev. Earth Environ.*, **2**, 340–357, https://doi.org/10.1038/s43017-021-00155-x.

Min, S.-K., X. Zhang, F. W. Zwiers, P. Friederichs, and A. Hense, 2009: Signal detectability in extreme precipitation changes assessed from twentieth century climate simulations. *Climate Dyn.*, **32**, 95–111, https://doi.org/10.1007/s00382-008-0376-8.

——, S.-Y. Jo, M.-G. Seong, Y.-H. Kim, S.-W. Son, Y.-H. Byun, F. C. Lott, and P. A. Stott, 2022: Human contribution to the 2020 summer successive hot-wet extremes in South Korea. *Bull. Amer. Meteor. Soc.*, **103**, S90–S97, https://doi.org/10.1175/BAMS-D-21-0144.1.

Miralles, D. G., A. J. Teuling, C. C. Van Heerwaarden, and J. Vilà-Guerau de Arellano, 2014: Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nat. Geosci.*, **7**, 345–349, https://doi.org/10.1038/ngeo2141.

NOAA, 2019: Monthly global climate report for February 2019. https://www.ncei.noaa.gov/access/monitoring/monthly-report/global/201902.

——, 2021: Monthly national climate report for February 2021. https://www.ncei.noaa.gov/access/monitoring/monthly-report/national/202102.

Overland, J. E., and Coauthors, 2021: How do intermittency and simultaneous processes obfuscate the Arctic influence on midlatitude winter extreme weather events? *Environ. Res. Lett.*, **16**, 043002, https://doi.org/10.1088/1748-9326/abdb5d.

Pall, P., T. Aina, D. A. Stone, P. A. Stott, T. Nozawa, A. G. J. Hilberts, D. Lohmann, and M. R. Allen, 2011: Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature*, **470**, 382–385, https://doi.org/10.1038/nature09762.

Palmer, T. N., and A. Weisheimer, 2018: A simple pedagogical model linking initial-value reliability with trustworthiness in the forced climate response. *Bull. Amer. Meteor. Soc.*, **99**, 605–614, https://doi.org/10.1175/BAMS-D-16-0240.1.

Philip, S., and Coauthors, 2020: A protocol for probabilistic extreme event attribution analyses. *Adv. Stat. Climatol. Meteor. Oceanogr.*, **6**, 177–203, https://doi.org/10.5194/ascmo-6-177-2020.

Puvvula, J., A. M. Abadi, K. C. Conlon, J. J. Rennie, H. Jones, and J. E. Bell, 2022: Evaluating the sensitivity of heat wave definitions among North Carolina physiographic regions. *Int. J. Environ. Res. Public Health*, **19**, 10108, https://doi.org/10.3390/ijerph191610108.

Rao, J., C. I. Garfinkel, H. Chen, and I. P. White, 2019: The 2019 New Year stratospheric sudden warming and its real-time predictions in multiple S2S models. *J. Geophys. Res. Atmos.*, **124**, 11 155–11 174, https://doi.org/10.1029/2019JD030826.

Röthlisberger, M., and L. Papritz, 2023a: Quantifying the physical processes leading to atmospheric hot extremes at a global scale. *Nat. Geosci.*, **16**, 210–216, https://doi.org/10.1038/s41561-023-01126-1.

——, and ——, 2023b: A global quantification of the physical processes leading to near-surface cold extremes. *Geophys. Res. Lett.*, **50**, e2022GL101670, https://doi.org/10.1029/2022GL101670.

Seong, M.-G., S.-K. Min, and X. Zhang, 2022: A Bayesian attribution analysis of extreme temperature changes at global and regional scales. *J. Climate*, **35**, 8189–8203, https://doi.org/10.1175/JCLI-D-22-0104.1.

Shepherd, T. G., 2016: A common framework for approaches to extreme event attribution. *Curr. Climate Change Rep.*, **2**, 28–38, https://doi.org/10.1007/s40641-016-0033-y.

——, and Coauthors, 2018: Storylines: An alternative approach to representing uncertainty in physical aspects of climate change. *Climatic Change*, **151**, 555–571, https://doi.org/10.1007/s10584-018-2317-9.

Song, S.-Y., S.-W. Yeh, R. P. Allan, S.-P. Xie, S.-I. An, and H.-S. Park, 2023: Climate sensitivity controls global precipitation hysteresis in a changing CO_2 pathway. *npj Climate Atmos. Sci.*, **6**, 156, https://doi.org/10.1038/s41612-023-00484-2.

Soulard, N., H. Lin, and B. Yu, 2019: The changing relationship between ENSO and its extratropical response patterns. *Sci. Rep.*, **9**, 6507, https://doi.org/10.1038/s41598-019-42922-3.

Stone, D. A., and P. Pall, 2021: Benchmark estimate of the effect of anthropogenic emissions on the ocean surface. *Int. J. Climatol.*, **41**, 3010–3026, https://doi.org/10.1002/joc.7002.

Terray, L., 2023: A storyline approach to the June 2021 northwestern North American heatwave. *Geophys. Res. Lett.*, **50**, e2022GL101640, https://doi.org/10.1029/2022GL101640.

Thomas, N. P., A. B. Marquardt Collow, M. G. Bosilovich, and A. Dezfuli, 2023: Effect of baseline period on quantification of climate extremes over the United States. *Geophys. Res. Lett.*, **50**, e2023GL105204, https://doi.org/10.1029/2023GL105204.

Tsutsui, J., 2020: Diagnosing transient response to CO_2 forcing in coupled atmosphere-ocean model experiments using a climate model emulator. *Geophys. Res. Lett.*, **47**, e2019GL085844, https://doi.org/10.1029/2019GL085844.

Turner, D. D., M. D. Shupe, and A. B. Zwink, 2018: Characteristic atmospheric radiative heating rate profiles in Arctic clouds as observed at Barrow, Alaska. *J. Appl. Meteor. Climatol.*, **57**, 953–968, https://doi.org/10.1175/JAMC-D-17-0252.1.

van Garderen, L., and J. Mindlin, 2022: A storyline attribution of the 2011/2012 drought in southeastern South America. *Weather*, **77**, 212–218, https://doi.org/10.1002/wea.4185.

Wang, G., P. Hope, E.-P. Lim, H. H. Hendon, and J. M. Arblaster, 2021: An initialized attribution method for extreme events on subseasonal to seasonal time scales. *J. Climate*, **34**, 1453–1465, https://doi.org/10.1175/JCLI-D-19-1021.1.

Wehner, M. F., M. L. Duffy, M. Risser, C. J. Paciorek, D. A. Stone, and P. Pall, 2024: On the uncertainty of long-period return values of extreme daily precipitation. *Front. Climate*, **6**, 1343072, https://doi.org/10.3389/fclim.2024.1343072.

Yanai, M., and R. H. Johnson, 1993: Impacts of cumulus convection on thermodynamic fields. *The Representation of Cumulus Convection in Numerical Models*, *Meteor. Monogr.*, No. 46, Amer. Meteor. Soc., 39–62, https://doi.org/10.1007/978-1-935704-13-3_4.

Yang, S., Z. Li, J.-Y. Yu, X. Hu, W. Dong, and S. He, 2018: El Niño–Southern Oscillation and its impact in the changing climate. *Natl. Sci. Rev.*, **5**, 840–857, https://doi.org/10.1093/nsr/nwy046.

Yao, Y., D. Luo, A. Dai, and I. Simmonds, 2017: Increased quasi stationarity and persistence of winter Ural blocking and Eurasian extreme cold events in response to Arctic Warming. Part I: Insights from observational analyses. *J. Climate*, **30**, 3549–3568, https://doi.org/10.1175/JCLI-D-16-0261.1.

Yoon, A., J. Kim, J. Lee, H. Min Sung, J.-W. Hong, S.-K. Min, J. Lee, and J. Hong, 2024: Factor analysis of recent major heatwaves in East Asia. *Geosci. Front.*, **15**, 101730, https://doi.org/10.1016/j.gsf.2023.101730.

Zhang, Y., and W. R. Boos, 2023: An upper bound for extreme temperatures over midlatitude land. *Proc. Natl. Acad. Sci. USA*, **120**, e2215278120, https://doi.org/10.1073/pnas.2215278120.