Exceptional Warmth in the Northern Hemisphere during January through March of 2020: The Roles of Unforced and Forced Modes of Atmospheric Variability

Siegfried D. Schubert[12], Yehui Chang[13], Anthony M. DeAngelis[2], Randal D. Koster[1], Young-Kwon Lim[14], and Hailan Wang[5]

[1]Global Modeling and Assimilation Office, NASA GSFC

[2]Science Systems and Applications, Inc.

[3]Morgan State University

[4]Goddard Earth Sciences Technology and Research / Universities Space Research Association
[5]Climate Prediction Center, NCEP/NWS/NOAA

Revised 09 January 2022

Address correspondence to:

Siegfried Schubert

Science Systems and Applications, Inc.

Lanham, MD USA

301-938-8674; siegschu2002@yahoo.com

Submitted to the *Journal of Climate*

Abstract

Much of northern Eurasia experienced record high temperatures during the first three months of 2020, and the eastern U.S. experienced a significant heat wave during March. In this study, we show that the above episodes of extraordinary warmth reflect to a large extent the unusual persistence and large amplitude of three well-known modes of atmospheric variability: the Arctic Oscillation (AO), the North Atlantic Oscillation (NAO), and the Pacific North American (PNA) pattern. We employ a "replay" approach in which simulations with the NASA GEOS AGCM are constrained to remain close to MERRA-2 over specified regions of the globe in order to identify the underlying forcings and regions that acted to maintain these modes well beyond their typical submonthly time scales.

We show that an extreme positive AO played a major role in the surface warming over Eurasia, with forcing from the tropical Pacific and Indian Ocean regions acting to maintain its positive phase. Forcing from the tropical Indian Ocean and Atlantic regions produced positive NAO-like responses, contributing to the warming over eastern North America and Europe. The strong heat wave that developed over eastern North America during March was primarily associated with an extreme negative PNA that developed as an instability of the North Pacific jet, with tropical forcing providing support for a prolonged negative phase. A diagnosis of the zonally symmetric circulation shows that the above extratropical surface warming occurred underneath a deep layer of tropospheric warming, driven by stationary eddy-induced changes in the mean meridional circulation.

1. **Introduction**

According to the World Meteorological Organization (WMO 2021), 2020 was one of the three warmest years on record. The January through March average temperature anomalies over Europe (+3.20 °C) and Asia (+3.66 °C) were the highest in the 111-year record (NOAA, 2020). In fact, the large temperature anomalies persisted throughout the three-month period. Over Asia, January ranked 2nd (+3.46 °C), February ranked 1st (+4.09 °C), and March ranked 4th (+3.44 °C) in the 111-year temperature record (NOAA, 2020). Over Europe, January (+3.16 °C) and February (+3.86 °C) temperatures ranked second warmest. During March, most of the Northern Hemisphere (NH) land areas were unusually warm, with the aforementioned record warmth over much of Eurasia and with large positive temperature anomalies over most of the eastern United States[1].

In this study we examine the underlying causes of the record or near-record warmth that occurred over much of the NH during the first three months of 2020. As we shall see, this is a period when extremes occurred in a number of leading modes of atmospheric variability, namely, the Arctic Oscillation (AO, Thompson and Wallace 1998), the North Atlantic Oscillation (NAO, Hurrel 1995), and the Pacific North American (PNA, Wallace and Gutzler 1981) pattern. These modes have well-known associations with surface temperature (e.g., Hurrell 1995; Hurrell et al. 2003; Thompson and Wallace 2001; Leathers et. al. 1991).

---

[1] https://www.ncdc.noaa.gov/temp-and-precip

Our focus is on uncovering what drove these modes, believed to be fundamentally internal subseasonal modes of atmospheric variability (e.g., Feldstein 2000, Riviere and Drouard 2015), to exhibit unusual persistence and large amplitude. While we address the causes of a number of apparently disparate temperature extremes that occurred in the NH during the first three months of 2020, the unifying goal of this study is to better understand the mechanisms leading to the persistence of these extremes and, in particular, to elucidate the role played by external forcing from the tropics.

Our approach involves the use of regional replay (Schubert et al. 2019b; 2020), in which large ensembles of simulations with the NASA GEOS Atmospheric General Circulation Model (AGCM, Molod et al. 2015) are constrained to remain close to the NASA Modern Era Retrospective-Analysis for Research and Applications, Version 2 (MERRA-2, Gelaro *et al*. 2017) over a number of specified regions of the globe. That is, in a regional replay experiment, we force the weather patterns in a specified region to match what actually happened there (according to the reanalysis), and we allow the atmosphere outside the region to evolve freely. If the prescribed weather (including all associated heating patterns) inside the specified region has an impact on the weather (e.g., on heat wave generation) in remote regions through some teleconnection mechanism, we will see this remote impact in the ensemble average. A limitation of this approach is of course that it does not in itself address the nature of the forcing – a subject we however do address here with a diagnosis of the zonal mean circulation.

The details of how replay forces a model to track an existing reanalysis can be found in Chang et al. (2019). Briefly, the model equations governing replay have the general form: $\frac{\partial x}{\partial t} = f(x) +$

Δx, where $\Delta x = (\text{analysis} - \text{forecast})/6\text{hrs}$, and f(x) consists of all the dynamics and physics terms of the model (basically the uncorrected model). Here, Δx is recomputed every 6 hours, but applied (appropriately scaled) at each time step. We note that this differs from the traditional nudging approach in that the increments are introduced as a state-independent forcing term, whereas in the case of nudging the entire model state is relaxed toward an analysis, and the relaxation time (which can be a function of space and variable) must be specified. Details of how these differences between replay and nudging impact the response of an assimilation system can be found in Bloom et al. (1996). Nudging has been used as a diagnostic tool for understanding seasonal variability and, of particular relevance here, for examining the remote impacts of the tropics. For example, Douville et al. (2011) examined (via a nudging technique) the influence of boreal summer monsoons on northern extratropical variability and found significant tropical forcing of boreal summer midlatitude stationary waves.

The use of regional replay to help understand what drives extreme events requires some justification. One might argue that these extreme events do not require any explanation other than that they represent the tails of a distribution representing what is largely internally-generated atmospheric variability. There is in fact evidence that projected increases in heat wave probability are largely the result of rigid shifts in the probability distribution associated with a warming world, shifts that can have a large impact on the tails, and thus extremes, of the distribution (e.g., Guirguis et. al. 2018). We argue here, however, that while the most extreme events are indeed associated with internal modes of atmospheric variability, their unusual persistence and amplitude often lie outside the bounds of what one can expect from internal variability alone. Their persistence and amplitude may instead have been enhanced to the observed levels by weather conditions in remote areas through some atmospheric teleconnection mechanism. As such, replay allows us to examine the following possibilities: the episodes of extraordinary warmth of early 2020 were either exceedingly rare manifestations of internal variability or not-so-rare manifestations of internal variability in conjunction with the impact of remote teleconnections. The replayed regions found to be relevant can be deemed potential sources of the forcing of the extreme event. We refer to the identification of potentially important remote teleconnections in the attribution of extreme events as Likelihood Amplification Through Regional Replay (LATRR).

Section 2 describes the data sets, the GEOS AGCM, and the various simulations performed for this study. Results are shown in Section 3, and the summary and conclusions are presented in Section 4. The Appendix addresses the robustness of the regional replay results.

2. Data and Model Simulations

This study makes extensive use of MERRA-2, a state-of-the-art atmospheric reanalysis. The MERRA-2 precipitation used for verification is an observationally corrected product (Reichle et al. 2011). As described in Schubert et. al. (2020), the GEOS AGCM used here is essentially the same as that used to produce MERRA-2 but differs from it in two important ways. First, it is run here at a coarser (approximately 1°, compared to 1/2° for MERRA-2) horizontal resolution.

Second, it includes a tendency bias correction (TBC) that helps alleviate some of the model's long-term biases (Chang et al. 2019)[2].

The runs performed for this study (see Table 1) include simulations in which the AGCM is forced with observed SST for the time period 1981-2019. These simulations (NORPL_CLIM) are used to define the model's climatology in the absence of replay. Furthermore, it turns out that tropical replay acts to correct some of the model's extratropical climatological biases (including, for example, a cold bias over Eurasia that remains even with TBC applied, not shown). Thus, it is necessary to remove separate climatologies (RPL_TR_CLIM, RPL_PAC_CLIM, etc.) for each of the replay runs for 2020 to compute anomalies. In that way, the computed anomalies do not partly reflect a climatological bias correction. The various sets of replay simulations for 2020 (each with 90 ensemble members) are also forced by observed SST but differ in the specified areas over which they are constrained to remain close to MERRA-2 (Table 1). As a cautionary note, we point out that forcing an uncoupled model with prescribed SST (as we do here) in regions where local coupled air-sea feedback is important can lead to unrealistic surface fluxes and deep convection (e.g., Wu and Kirtman, 2004), though whether this is a problem in a replay setting where the atmosphere is also prescribed is unclear.

Finally, we note that our choice of 90 ensemble members is somewhat arbitrary, and we have found that this number is more than enough to obtain accurate estimates of the ensemble mean anomalies. The primary reason for employing such a relatively large ensemble size is that we

---

[2] As described in Chang et al. (2019), the TBCs consist of time-averaged (over several decades) 6-hourly analysis increments (first guess forecast minus analysis) obtained from MERRA-2 reanalysis data which are added with opposite sign as additional forcing terms to the model equations.

are interested here in addressing the changes in the characteristics of the ensemble, including the ensemble spread and extremes, and for that we believe something on the order of 100 is necessary, though we have not done a statistical analysis to verify that (see, however, our assessment of the statistical robustness of the estimated probabilities in Tables 5 and 7).

Figure 1 outlines the four regions over which we carry out the replay experiments (see Table 1). Note that in the case of RPL_IND, the northern latitude of the replayed region extends further north (to 40°N) than for the other runs with replay in the tropical regions (RPL_TR, RPL_PAC, and RPL_ATL, in which the northern boundary is at 25°N). We found based on our prior work that extending the IND replay region farther north appears to improve the ability of the AGCM to simulate the development of the observed downstream circulation anomalies (Schubert et al. 2020; see also Bader and Latif 2005). In the NORPL simulations the model is free running (forced only by the observed SST), having no designated area of replay. Figure 1 also shows the global distribution of the SST anomalies averaged over the first three months of 2020. We discuss the SST anomalies and their possible role in forcing the various anomalies in Section 3 in the context of our various replay experiments.

3. Results

After providing some definitions in Section 3a, we focus in Section 3b on January-March (JFM), a period during which, as seen below, the AO plays an important role in the warming that extends across much of Eurasia, with the NAO contributing to the warming over Europe and North America. Section 3c focuses on March, a month during which the PNA plays an important role in the unusual warming that occurred over much of the eastern US. Section 3d provides a synthesis of the proposed physical mechanisms by which tropical forcing produced such exceptional warmth throughout the NH extratropics, with a focus on the zonally-symmetric component of the circulation.

a) Definitions

We focus here on the 2020 anomalies computed with respect to the 1981-2019 climatology. In particular, the MERRA-2 (or observed[3]) anomaly of a quantity (Y_o) is defined as

$$Y_o' = Y_o - \overline{Y_o}, \tag{1}$$

where the prime denotes an anomaly and the overbar is the long term average (1981-2019) based on MERRA-2/observations. In the case of the model simulations (Y_m), each ensemble member will have the analogous anomaly

$$Y_m' = Y_m - \overline{Y_m} = \langle Y_m' \rangle + Y_m'^{*} \ , \tag{2}$$

where the overbar now indicates a long term average (1981-2019) computed from the appropriate climatology runs (NORPL_CLIM, RPL_TR_CLIM, etc., see Table 1). Equation (2) shows that we can decompose the full anomaly (Y_m') into an ensemble mean ($\langle Y_m' \rangle$, an average over the ensemble members of a replay simulation) and intra-ensemble anomaly ($Y_m'^{*}$). Here the angle brackets denote the ensemble mean and the star denotes the deviation from the ensemble mean.

---

[3] In the following we will use the terminology "MERRA-2" and "observations" interchangeably, recognizing that reanalysis fields are not strictly observations.

A key aspect of our analysis is to employ a rotated[4] empirical orthogonal function (REOF, Richman 1986) analysis of the 2020 intra-ensemble anomalies (Y'^{*}_{m}) in order to isolate the leading modes of atmospheric variability that were active during that time. Associated with each of those REOFs, we compute two different principal components (PCs). One set consists of the traditional PCs computed as the projections of the Y'^{*}_{m} values onto the REOFs. We refer to those PCs as **PCI**s to indicate they are based on the intra-ensemble anomalies. The PCIs have the properties of having a zero mean and being uncorrelated with each other. Additionally, we compute another set of PCs by projecting the full anomalies ($Y'_{m} = \langle Y'_{m} \rangle + Y'^{*}_{m}$) onto those same REOFs. Those PCs therefore reflect both the intra-ensemble and ensemble mean contributions to the variability associated with that REOF. We shall refer to those PCs as **PCF**s, to indicate those are based on the full anomalies as defined by (2). Furthermore, we also compute the observed PCs by projecting the MERRA-2/observed anomalies (1) onto those same REOFs. We shall refer to those PCs as **PCOs**.

As a way of quantifying the results in the context of LATRR (as defined in the Introduction), we compute the probabilities of the PCFs exceeding the MERRA-2 (PCO) values (hereafter PEM) for the various replay runs: this is computed simply as the fraction of the 90 ensemble members that exceed the MERRA-2 value.

It is helpful to clarify the factors that drive the circulation anomalies. In particular, we can consider the full forcing (F) to be composed of

---

[4] The rotation criterion is varimax, which in our experience produces atmospheric modes that are easier to interpret physically, compared with unrotated EOFs. The rotation retains the mutual orthogonality in time, but the patterns are no longer orthogonal in space.

$$F = F_{SST} + F_{other}. \tag{3}$$

Here, F_{SST} is the forcing due to the global SST anomalies as determined from our NORPL runs, while F_{other} represents the additional forcing from replay–so that part of the total forcing in the replay region that is not directly tied to SST anomalies. Accordingly, it is F_{other} that differentiates (acts to increase) the probabilities seen in the replay runs over those seen in the NORPL runs.

b. JFM

i. The ensemble mean

As mentioned in the Introduction, excessive heat encompassed much of northern Eurasia during the first three months of 2020. This is illustrated in Fig. 2e, where we show that the average MERRA-2 T2m anomalies ($Y_o^{'}$) over that period have large positive (greater than 3°C) values extending eastward from eastern Europe across much of Russia. Exceptional warmth also occurred over much of eastern North America. Figure 2a shows that in the NH the 250mb stream function is characterized by negative anomalies north of about 60°, with a tendency for positive anomalies to occur throughout the middle latitudes, characteristic of a strong positive AO. In fact, the AO index for JFM achieved a record positive value in 2020 (based on values tabulated since 1950)[5]. Also prominent is a negative PNA-like structure over the North Pacific/North American region and a positive NAO-like pattern over the North Atlantic and Europe associated with a

---

[5] https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/season.JFM.ao.gif

positive NAO index that was exceeded only twice in the last 70 years[6]. The T2m anomalies are generally positive (though weak, < 1°C) throughout the tropics, with locally enhanced (> 1°C) anomalies over some regions, such as at the dateline (Fig. 2e). In fact, the Oceanic Niño Index (ONI) was 0.5, just at the threshold of a weak El Niño[7] (see also Fig. 1). The tropical precipitation (Fig. 2i) shows enhanced positive anomalies in the central Pacific with negative anomalies to the west and east. Positive anomalies also occur over the western Indian Ocean. In the NH extratropics, the storm tracks show negative precipitation anomalies, with positive anomalies occurring over the southeast U.S. and small portions of northern and southern Eurasia.

The tropical replay runs (RPL_TR: Figs. 2b, f and j) reproduce a substantial component of the MERRA-2 NH extratropical circulation, T2m and precipitation anomalies, though with weaker amplitude (a result that is not unexpected given these are the averages of 90 ensemble members, $\langle Y'_m \rangle$). The similarity to the MERRA-2 T2m anomalies over northern Eurasia is especially striking. While RPL_TR does reproduce the positive AO-like structure of the stream function anomalies (cf. Figs. 2a and 2b), it does less well in reproducing the pronounced negative PNA-like structure, particularly the large positive anomaly in the North Pacific, and the positive anomalies over the southeast U.S.. The lack of a strong negative PNA structure in the ensemble mean likely impacts the simulation of the T2m anomalies over North Amercia. We will return to this issue (the lack of a PNA structure) when discussing the results for March (Section 3c). RPL_TR, to a large extent, reproduces the MERRA-2 tropical precipitation anomalies, including the positive anomalies over the central Pacific and the western Indian Ocean (cf. Figs. 2i and 2j). This is not guaranteed, since

---

[6] https://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/season.JFM.nao.gif
[7] https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php

we do not directly constrain the precipitation during replay (only the prognostic fields), and the MERRA-2 precipitation shown here is an observationally corrected product (see Section 2).

The results for NORPL (F_{SST}) are shown in Figs. 2c, g, and k. The stream function anomalies (Fig. 2c), while weak, do have some imprint of a positive AO and a negative PNA. In fact, the circulation anomalies in the Pacific are symmetric about the equator (Fig. 2k). As shown later, it appears the anomalies reflect a local enhancement of the zonal mean zonal wind response to the tropical warming, where the enhanced warming is maintained by the precipitation anomalies in the tropical Pacific just north of the equator (see Section 3d). The T2m anomalies over Eurasia in NORPL are for the most part a weaker version of the RPL_TR results (cf. Figs. 2f and 2g). Fig. 2d shows that the impact of F_{other} (RPL_TR-NORPL) is to enhance the positive AO, primarily by producing positive anomalies throughout much of the middle latitudes including a positive NAO-like signal in the eastern North Atlantic. Over Eurasia, F_{other} enhances the warming especially over western Eurasia including much of eastern Europe (Fig. 2h). In contrast with NORPL, for which the main positive precipitation anomalies (associated with F_{SST}) occur in the Pacific and straddle the equator, the main tropical precipitation anomalies of RPL_TR-NORPL (associated with F_{other}) exhibit an east/west structure with positive anomalies over the western Indian Ocean, negative anomalies over the maritime continent, and again positive anomalies to the east of that (cf. Figs.2k and 2l).

Fig. 3 shows the impacts of F_{other} from each of the subregions PAC, ATL and IND. RPL_PAC contributes only weakly to the warming over Eurasia (mostly less than 0.5°C), acting primarily to warm Alaska, Canada and Greenland, and cool the western United States (Fig. 3d). The main circulation anomalies in the NH (Fig. 3a) consist of positive stream function anomalies in the midlatitude North Pacific with negative anomalies to the north, extending over North America. RPL_ATL produces warming over the northern fringe of Eurasia (Fig. 3e), apparently as part of an anomalous wave pattern that extends northeastward from the tropical Atlantic across Europe and Eurasia (Fig. 3b) – a pattern that projects onto the NAO. RPL_ATL also produces warming over much of North America. RPL_IND produces the most widespread warming over northern Eurasia with local maxima over Europe and eastern China (Fig. 3f). This is associated with generally positive stream function anomalies across much of Eurasia and the eastern North Pacific, including an NAO-like stream function response in the North Atlantic (Fig. 3c) that is especially prominent during February (not shown). This is consistent with previous studies linking forcing in the Indian Ocean region to the NAO (e.g., Hoerling et al. 2001). Furthermore, Bader and Latif (2005) conclude that the mechanism linking the NAO with heating/cooling in the Indian Ocean involves the South Asian jet (via its waveguiding effect) carrying anomalies that originate in the Indian Ocean region into the North Atlantic sector.

Interpretation of the above results implicitly assumes linearity in the sense that the sum of results from replaying the different tropical subregions should be roughly equal to the results obtained from replaying the full tropical region (something that we have found from our previous work is not guaranteed, Schubert et al. 2019b, Schubert 2020). We address this issue further in the Appendix, where we show that the circulation (stream function) responses do in fact exhibit substantial linearity, while that is less true for the T2m responses.

ii. The intra-ensemble variability

Referring back to equation (2), we have so far looked only at the model's three month (JFM) ensemble mean results, $\langle Y_m^{'} \rangle$. We next examine the intra-ensemble variability ($Y_m^{\prime *}$), focusing in particular on the leading modes of monthly intra-ensemble variability that characterized the 2020 January through March variability. To do this we decompose the monthly mean intra-ensemble 250mb stream function variability into REOFs (see Section 3a). The covariance matrix used in the REOF calculation is obtained by pooling together the intra-ensemble anomalies ($Y_m^{\prime *}$) from the 2020 NORPL and replay runs (see Table 1)[8]. As such, we are assuming that to first order the intra-ensemble variability outside the tropical replay regions is not strongly impacted by replay (though, as discussed later, this is not always the case). The region used in the covariance calculation spans the entire NH.

Rather than showing the stream function REOFs, we show instead the correlations between the associated PCIs and the stream function everywhere over the globe. The patterns of the correlations are, not surprisingly, very similar to the REOFs themselves over the NH (not shown). The correlations with the 250mb stream function associated with the three leading REOFs (Fig. 4 a, d and g) consist of patterns that resemble the PNA (accounting for 8.9% of the intra-ensemble stream function variance over the NH), the NAO (8.3%) and the AO (6.8%). In fact, the patterns are quite similar to the leading REOFs obtained from monthly (JFM) 250mb height fields based on MERRA data for the period 1981 – 2009 (Schubert and Lim 2013). One difference is that for MERRA, ENSO is the first mode (since that analysis included interannual variability) and that the

---

[8] This results in 1350 monthly mean states used in the covariance calculation: 5 experiments x 3 months/experiment x 90 ensemble members/month.

second, third, and fourth modes are the AO, the PNA, and the NAO, respectively. We should also note that the order of the three leading modes can change by including fewer experiments in the pooled results (not shown), though the spatial structure of the three leading modes is very robust.

The correlations of the PNA PCI with T2m (Fig. 4b) and precipitation (Fig. 4c) over North America are generally weak, with a slight tendency for cooling over the northwest coast, and enhanced precipitation on the west coast of the U.S. during its negative phase. The negative phase also produces a tendency for warming over the central U.S., and reduced precipitation along the U.S. southeast coast. The T2m correlations for the NAO (Fig. 4e) show that in its positive phase it is associated with warming over the eastern U.S. and northern Europe, and cooling over much of northeastern Canada and Greenland. The positive phase of the NAO REOF is associated with enhanced precipitation over the eastern U.S., reduced precipitation over northeast Canada, and enhanced precipitation east of Greenland (Fig. 4f). The positive phase of the AO REOF is associated with enhanced warming over much of northern Eurasia and cooling at the pole (Fig. 4h). The correlations with precipitation (Fig. 4i) show the positive phase of the AO is associated with enhanced precipitation along much of the northern edge of the Eurasian continent, with a tendency for reduced precipitation to the south. Overall these monthly correlations with temperature and precipitation are consistent with documented observed impacts of the PNA, NAO and AO (e.g., Hurrell 1995; Hurrell et al. 2003; Thompson and Wallace 2001; Leathers et. al. 1991; Schubert and Lim 2013). The main difference with observations of concern here is the lack of a clear impact of the PNA on T2m over the southeastern US. We will come back to this issue in the next section, where we focus on the March results.

Up to this point we have examined separately the ensemble mean (Figs 2-3) and intra-ensemble (Fig. 4) statistics. We now examine the full anomalies ($Y_m^{'}$) to better quantify the contribution from the forced ($\langle Y_m^{'} \rangle$) component of the circulation. The scatter plots in Figure 5 show just that, for the three leading modes. Here, as described in Section 3a, we project the full model anomalies ($Y_m^{'}$) onto the three leading model REOFs to produce the corresponding PCFs, averaged over January-March of 2020. Furthermore, the MERRA-2 anomalies ($Y_o^{'}$) are projected onto those same REOFs to produce the analogous "observed" principal components (the PCOs), allowing a direct comparison with the PCFs from the model results.

Figure 5a shows the scatter of the JFM mean PCFs for the PNA (PCF 1) and AO (PCF 3) modes. (Each small dot represents an ensemble member, color-coded to its replay experiment; the large dots represent ensemble means.) The preponderance of the red and blue points in the bottom right quadrant indicates that both the AO and the PNA are in part forced from the tropics. All three ocean basins appear to contribute to the forcing of the PNA, while for the AO only the PAC and IND regions have projections exceeding that obtained from the NORPL results. Overall, replaying the tropics makes the observed anomalies of both the AO and PNA more likely to have occurred compared to the NORPL runs (black dots in Fig. 5a). These results are quantified in Table 2, which shows the probabilities of exceeding the MERRA-2 values (PEM) for the various replay runs. Here (and in Table 3) the values are the mean and standard deviation of 10,000 PEM values obtained by randomly subsampling 80 ensemble members from the 90 available ensemble members. We note that the mean values of PEM shown in Tables 2 and 3 (computed from the subsampled 80 ensemble members) are identical (to two significant digits) to the single PEM values obtained from the 90 ensemble members (not shown). Those results reflect the relatively small standard deviations shown in Tables 2 and 3, and indicate that the PEM estimates are statistically quite stable. The results for the PNA (second column of Table 2) show that replaying the tropics results in a doubling (from 0.18 for NORPL to 0.36) of the PEM, with the IND ocean basin alone showing an even greater increase (0.53). The results for the AO (third column of Table 2) also show a substantial impact on the PEM of replaying the tropics. In this case, the PEM value increases from 0.0 for NORPL to 0.11, with the PAC and IND regions being the main contributors. As such, tropical replay changes what is an impossible event (obtaining the observed value of the AO index) in the model ensemble in the absence of replay, to one that is possible but still a rather rare event.

Figure 5b shows, for the NAO (PCF 2) and AO (PCF 3) modes, the scatter of the PCFs for the individual ensemble members. Here we see that replaying the tropics has little impact on the ensemble mean NAO, compared with that for the NORPL runs. In fact replaying the Pacific region acts to suppress the ensemble mean. Nevertheless, replaying the tropics does increase the PEM, though in this case that it is due to an increase in the intraensemble variability of the NAO (cf. the scatter of the black dots and those from the replay runs in Fig. 5b). Table 2 (fourth column) shows that the PEM for the NAO increases from 0.00 (in NORPL) to 0.12 for RPL_TR, with even larger values for both RPL_IND (0.17) and RPL_ATL (0.24).

Summarizing the above results within the context of LATRR, the PEM analysis highlights the key role of the tropics in forcing negative PNA-like anomalies during the first three months of 2020, with both F_{SST} (NORPL) and F_{other} (RPL_TR-NORPL) contributing. The impact of tropical forcing is weaker for the AO, yet appears critical to obtaining the observed amplitude in that

RPL_TR-NORPL (F_{other}) changes what is an impossible event in the model ensemble in NORPL (F_{SST}), to one that is possible, though still a rare event. Furthermore, the analysis highlights the importance of F_{other} in the Atlantic and Indian basins in making the observed extreme positive NAO more likely in the model ensemble, primarily by increasing its intra-ensemble variability.

c. March

We now take a closer look at March 2020 (Figure 6). The MERRA-2 stream function anomalies (Fig. 6a) show a clear signature of a strong positive AO. A key difference with respect to the previous two months (not shown) is the pronounced negative PNA pattern (the PNA index had a value of -2.64 standard deviations)[9]. The MERRA-2 T2m anomalies (Fig. 6e) show warming throughout much of Eurasia, though weaker and less spatially extensive compared with January and February (Collow et al. 2021), with some of the largest anomalies limited to western Russia, and eastern Europe. In addition, much of the eastern U.S. has positive T2m anomalies. In fact, much of the southeast US ranked in the top 5 warmest Marches[10]. The precipitation anomalies in the central Pacific consist of generally positive anomalies just north of the equator and negative anomalies just south of the equator. Also, generally positive precipitation anomalies occur in the Indian Ocean south of the equator and over the maritime continent. Comparing these anomalies with the JFM mean (Fig. 2j), we see that the strong positive anomaly in the central tropical Pacific is absent: a feature of the precipitation anomalies that was in fact limited to January and February (not shown).

---

[9] https://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/month_pna_index2.shtml

[10] https://www.ncdc.noaa.gov/temp-and-precip

The results for RPL_TR are shown in the second column of Fig. 6.   The NH extratropical stream function anomalies (Fig. 6b) are overall similar in pattern to the observed, exhibiting a tendency for positive anomalies in middle latitudes and negative anomalies in the polar region, consistent with a positive AO. The anomalies are, however, considerably weaker in amplitude. This is especially true for the PNA-related positive anomaly in the North Pacific which is not only weaker but also shifted to the west compared with the MERRA-2 results. Although not shown here, all three ocean basins (RPL_IND, RPL_PAC, and especially RPL_ATL) contribute to the NH extratropical positive stream function anomalies seen in RPL_TR. The T2m anomalies (Fig. 6f) are generally consistent with the observed over Eurasia, but over North America they are weak, reflecting a much weaker negative PNA structure in the stream function anomalies over North America (cf. Figs 6a and 6b). The NORPL results (third column of Fig. 6) show somewhat stronger negative stream function anomalies in the NH polar region compared with RPL_TR (Fig. 6c) and generally weak positive T2m anomalies over Eurasia and the United States (Fig. 6g). A weak negative PNA anomaly (Fig. 6c), is very similar to (but somewhat larger in amplitude) the NORPL results for January and February (not shown).   The NORPL precipitation anomalies in the tropical Pacific (Fig. 6k) are very similar to the JFM mean (Fig. 2k), presumably reflecting the fact that the SST anomalies change very little during the three months.

In analogy to Figure 5, Figure 7 shows the March 2020 scatterplots of the PCFs for the three leading REOF modes (the same modes described in Section 3.b.ii). Fig. 7a (PNA versus AO) shows that replaying the tropics does indeed produce some ensemble members that have PCF amplitudes (for both the PNA and AO) that exceed those of MERRA-2. Table 3 shows that for the PNA, all three regions produce PEM values more than twice those for NORPL (PEM =0.11 for NORPL versus 0.23 for RPL_TR), while for the AO, for which the PEM value is 0.0 for NORPL, replaying the tropics has for the most part a negligible impact on the PEM values with the largest impact coming from the IND region (PEM = 0.07). Fig. 7b (NAO versus AO), shows that replaying the tropics also produces ensemble members that have PCF amplitudes for the NAO that are as large or larger than the MERRA-2 (PCO2) value, especially for RPL_ATL (PEM = 0.17, Table 3). While NORPL fails to produce any ensemble members with an NAO PCF amplitude as large as that from MERRA-2 (PEM = 0.00, Table 3), it nevertheless produces an ensemble mean amplitude that is larger than those from any of the replay runs (Fig. 7b). This again indicates that the main impact of the tropical forcing from replay (F_{other}) is to increase the intra-ensemble variability of the NAO rather than to affect its ensemble mean.

Given that the large amplitude of the observed negative PNA is not reproduced with tropical replay (cf. Figs. 6a and 6b), we will next focus more on the NORPL results (for which the ensemble mean does show a weak signature of a negative PNA, Fig. 6c). We look, in particular, at the most extreme ensemble members, a number of which have PNA amplitudes that exceed the observed (Fig. 7a). The last column of Fig. 6 shows an average of the 10 NORPL ensemble members having the largest projections onto the PNA and AO REOFs (here we choose the 10 ensemble members for which the sum of the squares of the amplitudes of the PNA and AO PCFs is the largest, with the constraint that PNA PCF < 0, and the AO PCF > 0) . The results (Fig. 6d) are, not surprisingly, characterized by a much more pronounced negative PNA structure and positive AO, with a strong resemblance to the observed stream function anomalies (Fig. 6a), although over the eastern U.S., the NORPL composite shows little or no positive anomalies. We also now find substantial warm anomalies over the United States that are however centered somewhat further to the west compared with the observations (cf. Figs 6e and 6h) – a difference that could well reflect sampling errors (a single realization for the observations versus 90 for the model). What is perhaps the more interesting result is that the tropical precipitation anomalies associated with these most extreme ensemble members differ little from those produced from the full 90 ensemble member average (cf. Figs. 6k and 6l), as this indicates that the extreme amplitudes are not driven by more extreme tropical forcing. Thus we must look to other sources of energy, especially for the extreme negative PNA.

The above suggests that we should examine the barotropic conversion of kinetic energy between the ensemble mean and the intra-ensemble components of the flow in the NORPL runs. This is motivated by Simmons et al. (1983), who concluded that much of the low-frequency variability in the NH winter circulation is associated with disturbances that derive their energy from the basic state through barotropic instability. In particular, they found that structures in a global barotropic model resembling the observed PNA are related to the fastest growing modes associated with barotropic instability of the zonally-varying basic state. Following Simmons et al. (1983) and Schubert et al. (2001), the local ensemble mean barotropic conversion from the ensemble mean flow to the ensemble mean intra-ensemble kinetic energy,

$$IKE = \frac{1}{2}\langle u^{*2} + v^{*2} \rangle, \qquad (4)$$

is approximately given by

$$C1 = -\langle u^{*2} - v^{*2} \rangle \frac{\partial \langle u \rangle}{\partial x} - \langle u^{*} v^{*} \rangle \frac{\partial \langle u \rangle}{\partial y} = \vec{E} \cdot \nabla \langle u \rangle, \tag{5}$$

where $\vec{E}$ (Hoskins et al. 1983) is a vector with components

$$\vec{E} = (\langle v^{*2} - u^{*2} \rangle, -\langle u^{*} v^{*} \rangle). \tag{6}$$

Here, u is the zonal wind, v is the meridional wind, the angle brackets denote an ensemble mean, and the star indicates a deviation from the ensemble mean. It is clear from *C1* that when $\vec{E}$ points in the direction of the gradient of the mean zonal wind there is a growth of IKE. The E-vectors (Fig. 8a) and the associated local barotopic conversion (Fig. 8b) show quite clearly that the intra-ensemble variability in the North Pacific is gaining energy from the North Pacific jet. Furthermore, as discussed by Hoskins et al. (1983), the local generation of IKE associated with the first component of $\vec{E}$, $\langle v^{*2} - u^{*2} \rangle$, is associated with zonally elongated anomalies in regions of diffluence, and we see here that this is consistent with the PNA (a zonally elongated anomaly) extracting energy from the mean flow in the exit region of the North Pacific jet (Fig. 8b). As such, we argue that while the tropical forcing may have acted to produce a predilection for a negative phase of the PNA, the initiation and the subsequent growth of the PNA was the result of energy that it derived from the basic state through barotropic instability.

d. Interpretation and Synthesis of Mechanisms

To facilitate our understanding of the dynamical mechanisms by which the tropics impacted the extratropics during the first three months of 2020, we focus here on what drove the zonal mean circulation (including the AO) and on how that increased the temperature in the NH middle latitudes. Figure 9 shows the zonal mean u-wind, temperature, and mean meridional circulation (MMC) anomalies for NORPL, RPL_TR and MERRA-2. All three results show a somewhat similar pattern of zonal wind anomalies (left panels of Fig. 9), consisting of a strengthening of the winds on the equatorward flanks of both subtropical jets, negative anomalies centered at about 30°N, and positive anomalies centered at about 60N. The magnitude of the anomalies is strongest in MERRA-2 (Fig. 9g) and weakest in NORPL (Fig. 9a). In the case of RPL_TR (Fig. 9d), the MERRA-2 anomalies between 25°S and 25°N are of course reproduced by design. Nevertheless, RPL_TR also does a reasonable job of reproducing the MERRA-2 anomalies poleward of 25°N. The wind anomalies are to a large extent reflected (via the thermal wind relationship) in the temperature anomalies (middle panels of Fig. 9) with warming occurring in the tropics (+/- 25° lat), and (focusing on the NH) between about 30°N and 60°N, and with cooling occurring poleward of about 60°N, although the cooling is mostly in the stratosphere in NORPL.

We are particularly interested in understanding the cause of the NH extratropical tropospheric warming (Fig. 9h) that occurs in approximately the same latitude band as the surface warming (Fig. 2e), although the latter extends somewhat further north. One potential mechanism by which that can occur involves eddy-forced changes in the mean meridional circulation (MMC, Chang 1995, Seager et al. 2003). In fact, Seager et. al. (2003) outline a series of events in which tropical warming leads to strengthening of the equatorward sides of the subtropical jets (similar to that in Fig. 9g), which impacts the transient eddy momentum fluxes, creating eddy-induced changes in the MMC by which different latitude bands experience (via anomalous descent or ascent) either adiabatic warming or cooling. The right panels of Fig. 9 show that 2020 did indeed experience changes in the MMC in the NH: the Hadley Cell was strengthened, and the Ferrel Cell was shifted poleward and strengthened. The latter change is particularly relevant here in that the location of the anomalous descending motion (as part of the anomalous Ferrel Cell, Fig. 9i) coincides with the region of tropospheric warming between 30°N and 60°N (Fig. 9h). RPL_TR produces similar changes in the MMC (Fig. 9f) although the change in the Ferrel Cell is weaker. The changes in the MMC in NORPL (Fig. 9c) are also generally similar though substantially weaker, consistent with the much weaker middle latitude anomalous warming that occurs in NORPL (Fig. 9b).

We next examine whether it is indeed changes in the eddy fluxes that are driving the anomalous MMC in the extratropics (as outlined by Seager et al. 2003). A useful diagnostic for examining the impact of the eddies on the zonal mean flow is the Eliassen-Palm (E-P) flux (F, Eliassen and Palm 1961). As described in Edmon et al. (1980), the divergence of F has two components: the meridional divergence of eddy momentum fluxes and the vertical divergence of eddy heat fluxes. Furthermore, the average quadratic (flux) terms can be decomposed as (shown here for the eddy momentum fluxes),

$$\langle\overline{[u^e v^e]}\rangle = \langle[\bar{u}^e \bar{v}^e]\rangle + \langle\overline{[u'^e v'^e]}\rangle \tag{7}$$

where the square brackets denote a zonal mean and the superscript (*e*) denotes a deviation from the zonal mean (the eddy). Also, the overbar denotes a time (monthly) mean, the prime (‘) denotes a devation from the time mean, and the angle brackets denote an ensemble mean. As such, the first term on the RHS of (7) is the contribution from the stationary eddies, while the second term is the contribution from the submonthly transients.

Figure 10g shows that the observed (stationary + transient) E-P flux is generally upward in middle latitudes (below 300mb), with equatorward flux south of about 60°N, and poleward flux north of 60°N. Flux convergence (negative values) occurs throughout the middle and upper troposphere both equatorward (between 100mb - 300mb) and poleward (between 300mb – 500mb) of the jet, acting to decelerate the westerly winds there. This overall behavior of the E-P fluxes is similar in NORPL and RPL_TR (Figs. 10a and 10d), giving us confidence that the model is producing realistic fluxes. The anomalous E-P flux and anomalous E-P flux divergence are shown in the middle column of Figure 10. The results for MERRA-2 (Fig. 10h) show that the upper troposphere (200mb-500mb) is characterized by a region of anomalous divergence (positive values, producing westerly acceleration) between about 40°N-65°N together with a region of convergence just to the south (negative values, producing easterly acceleration), thereby supporting the observed zonal wind anomalies and a positive AO in the upper troposphere. The results for RPL_TR (Fig. 10e) are similar but mostly positive in sign and weaker in amplitude. In NORPL (Fig. 10b) the divergence occurs somewhat lower in the troposphere and is even weaker than that of RPL_TR. The right panels of Fig. 10 show the results for the stationary component of the E-P flux alone. The similarity between the middle and right columns indicates that the 2020 flux divergence anomalies are dominated by anomalies in the stationary component. These results are consistent with those of Limpasuvan and Hartmann (2000), who found that the AO is primarily maintained by the stationary eddies. Furthermore, we found that both momentum and heat fluxes contribute to the anomalous divergence, with the contribution of the latter occurring primarily below 300mb, and the former's contribution centered at about 300mb (separate impacts not shown).

The collocation of the region of upper tropospheric E-P flux divergence and the upper (return) branch of the anomalous Ferrel Cell (cf. Figs. 9i and 10i) is highly suggestive of the atmosphere acting to balance the eddy-induced westerly acceleration with a coriolis torque (consistent with Seager et al. 2003) and, by continuity, producing the anomalous Ferrel Cell. We thus conclude that it was primarily the stationary eddies that drove the anomalous descending motion that produces the observed warming in the NH extratropics. Furthermore, the overall similarity between the results for MERRA-2 and RPL_TR supports the idea that forcing from the tropics played a key role.

Note that Seager et al. (2003) found (in the context of El Niño) that it was the changes in the subtropical jet and associated changes in wave refraction, that acted to alter the transient eddy fluxes. In our case, the relevant eddies are primarily stationary (composed of PNA-like and NAO-like structures, Figs. 2a and 6a) and any tropical impact on them appears to be more direct. Specifically, we have shown that the stationary eddies consist of:

1) A weak but persistent SST-forced (as determined by NORPL, F_{SST}) negative PNA-like circulation anomaly in the upper troposphere that occurred throughout the three month period. A tropically-forced enhancement of the above weak negative PNA via F_{other} in all three ocean basins, with the IND region playing a key role.

2) An extreme negative PNA during March resulting from an instability of the North Pacific jet.
3) A positive NAO, with an intra-ensemble variability (in the model environment) increased by tropical forcing (F_{other}), and with an amplitude increased by forcing from the ATL and IND regions, especially during February.

Finally, we should point out that the impact of the transient eddies is not insignificant. A comparison of Figs. 10h and 10i shows that they are the main contributors to the negative anomalies (anomalous E-P flux convergence) that occur in the middle and upper troposphere. As such, we leave open the possibility that a change in wave refractivity does play a role in altering those eddy fluxes – something we do not pursue further here.

4. Summary and Conclusions

In this study we employ a modeling approach involving regional replay (wherein the model is forced to track nature, as represented by reanalysis, over specified regions) to help us isolate the underlying physical mechanisms and regions that likely contributed to the exceptional warmth that occurred over much of the Northern Hemisphere middle and high latitude land areas during the first three months of 2020. While by definition this exceptional warmth is a rare event, regional replay allows us to identify those remote regions that (when specified) significantly increase its likelihood. The identified regions are accordingly deemed to be potential sources of forcing for the exceptional warmth. We refer to this attribution approach for extreme events as Likelihood Amplification Through Regional Replay (LATRR).

The GEOS AGCM was forced with observed SST in a number of experiments (90 ensemble members each) in which, for each experiment, the simulations were constrained to remain close to MERRA-2 over a specific region of the globe. The regional replay runs are compared to runs in which the only constraint is the observed SST. The results show that the 2020 episodes of extratrordinary warmth reflect to a large extent the unusual persistence and large amplitude of well-known modes of atmospheric variability: the Arctic Oscillation (AO), the North Atlantic Oscillation (NAO), and the Pacific North American (PNA) pattern. In particular, we found that an extreme positive AO played a major role in the warming over Eurasia, consistent with the findings of previous studies (Overland and Wang 2021, Collow et al. 2021). We further show here that forcing from the tropical Pacific and Indian Ocean regions increased the probability of such an extreme event occurring in the simulations from essentially zero (i.e., an essentially impossible event without replay) to 10%. Forcing from the tropical Indian Ocean and Atlantic regions increased the probability of the NAO achieving the observed extreme value from essentially zero (without replay) to roughly 20% percent, partly by increasing its variability, with forcing from the Indian Ocean region contributing to the excessive heat over Europe. The strong heat wave that developed over eastern North America during March was primarily associated with an extreme negative PNA that developed as an instability of the North Pacific jet, with tropical forcing providing support for a prolonged negative phase.

In an effort to better understand the underlying dynamical processes by which the tropics impacted the AO, PNA and NAO and hence acted to warm the extratropics, we diagnosed the zonally symmetric component of the flow. In particular, we showed that anomalous stationary eddy forcing acted to maintain the positive AO [consistent with Limpasuvan and Hartmann

(2000)]. Furthermore, guided by the analysis of Seager et al. (2003), we showed that those same eddies forced an anomalous Ferrel Cell, thereby creating subsidence-induced warming of a deep layer of the troposphere between 30°N and 60°N, roughly coinciding with the latitudes of surface warming. While a complete understanding of the nature of the above stationary eddy forcing and its connection to the tropics has not been established in this study, our results indicate that the PNA and NAO played key roles, with forcing from the tropics supporting the negative phase of the PNA and the positive phase of the NAO.

There are still a number of unresolved issues. For example, the reason for the increased intra-ensemble variability of the NAO when replaying the tropics is unclear. Feldstein (2003) showed that NAO growth is driven by both high-frequency (period < 10 days) and low-frequency (period > 10 days) transient eddy vorticity fluxes. Presumably, tropical replay is acting to increase the overall middle latitude transient eddy activity. A more in-depth evaluation of that including the implications for predictability of the extratropical circulation is, however, beyond the scope of this study. Also, the idea that the stationary eddy flux divergence associated with the NAO is acting to maintain the AO needs further investigation, especially given the lack of agreement as to whether these two modes represent distinct physical phenomena (e.g., Ambaum et al. 2001, Feldstein and Franzke 2006).

It is also not clear exactly which aspects of the tropical forcing distinguish 2020 from previous years, though it is very likely that the overall warming of the tropical atmosphere (as a response to the warm tropical SST) is key, so that global warming is likely playing a role. There is, in fact, a tendency for both the AO and NAO to be in their positive phases since the late 1980s, though there also appears to be a substantial variability of those modes on decadal time scales[11]. The impact of anthropogenic climate change was investigated by Ciavarella et al. (2021). Based on the results of a number of different global climate models they concluded that without human influence the unusually hot temperatures widely experienced in Siberia in the first half of 2020 would have been practically impossible.

We have also not addressed the link to the stratosphere and the fact that the strong positive AO during January through March was accompanied by an unusually strong Arctic stratospheric polar vortex (SPV). For example, Lee et al. (2020) examined the forecasts of 6 different seasonal prediction systems for January–March 2020 and demonstrated a strong interdependence between the accuracy of SPV and AO forecasts. Lawrence et al. (2020) showed that the very strong and persistent SPV during the winter of 2019/2020 was associated with low tropospheric planetary wave driving and wave-reflection in the stratosphere that supported the strong vortex. They highlight this as a particularly extreme example of two-way coupling between the troposphere and stratosphere. Overland and Wang (2021) concluded that the proximate cause for the warm extremes over Siberia from January through April was the record strength of the stratospheric polar vortex and the associated strong tropospheric westerlies that reduced the penetration of cold air from the north. Given the above studies, it seems clear that our focus on the troposphere likely provides an incomplete picture of the causes of the unusual warming in the NH during JFM 2020. Nevertheless, those studies are not inconsistent with our main results

---

[11] https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/teleconnections.shtml

pointing to the ultimate forcing being in the tropical troposphere, with the caveat that coupling with the stratosphere may play an important role in facilitating the tropospheric response.

To end on a general note, we believe that LATRR, as illustrated through replay experiments in this study, can be an important tool for advancing prediction efforts, especially the prediction of extremes of societal relevance that occur on subseasonal time scales (e.g., heat waves, floods, flash droughts). In theory, and as illustrated in this study, LATRR can be used to identify those regions and mechanisms most responsible for driving such extremes. With these regions and mechanisms identified, we can better prioritize and direct model development efforts aimed at improving our ability to predict them.

Acknowledgments

This work has been supported by the NASA MAP (NNG17HP01C) program and the NOAA Climate Program Office Modeling, Analysis, Prediction, and Projections (MAPP) program (NA14OAR4310221). MERRA-2 data were developed by the Global Modeling and Assimilation Office at NASA GSFC under funding by the NASA MAP program, and are disseminated through the Goddard Earth Science Data and Information Services Center (GES DISC). The file specifications for the MERRA-2 output are documented in Bosilovich (2015).

Data Availability Statement.

MERRA-2 data is available from the Goddard Earth Sciences Data and Information Services Center (GES DISC) at https://disc.gsfc.nasa.gov/. Output from the AGCM simulations are available by request.

APPENDIX

Robustness of the Results from Replaying the Tropical Subregions

In our design of the various tropical replay regions, RPL_TR was meant to quantify the overall impact of the tropics, while replaying the three subregions (RPL_IND, RPL_PAC, and RPL_ATL, Fig. 1) was intended to give some idea of how each of these regions contributed to that. Since our choice of the longitudes that separate the three regions (IND, PAC and ATL) is rather arbitray, it would be instructive to assess the sensitivity of the results to those longitudes. Carrying out such a series of sensitivity experiments would however be computationally prohibitive so, instead, we simply examine here whether replaying a region that encompasses two adajecent subregions produces essentially the same results as summing the results from replaying the two smaller adjacent regions. In particular, we compare in Fig. A1 the sum of the results (F_{other}) from 1) RPL_IND and RPL_PAC with the results from replaying the region that encompasses the IND and PAC regions (RPL_IND+PAC), and the 2) the sum of the results from RPL_IND and RPL_ATL with the results from replaying the region that encompasses the IND and ATL regions (RPL_IND+ATL). See Table A1 for details of the runs.

Comparing the first two columns of Fig. A1 we see that the NH circulation (250mb stream function) responses to replaying the PAC and IND regions of the anomalies are roughly linear in that the sum resembles that of RPL_IND+PAC (cf. Figs A1a and A1b). However that is not the case for T2m for which the sum underestimates the warming over much of northern Eurasia (cf. Figs. A1e and A1f), though it does reproduce the T2m anomalies over North America.

Comparing the third and fourth columns of Fig. A1 we see similar results in that the NH circulation (250mb stream function) responses to replaying the ATL and IND regions of the anomalies are roughly linear in that the sum resembles that of RPL_IND+ATL (cf. Figs A1c and A1d), though the sum tends to overestimate the anomalies indicating that there is some double counting (e.g., Schubert et al. 2019b and Schubert et al. 2020). Here again, the sum underestimates the warming over central Asia, while the warming over North America is overestimated (cf. Figs. A1g and A1h).

In summary, we conclude that despite the approximate linearity in the circulation responses to replaying the three tropical subregions, there is sufficient nonlinearity in the T2m responses to require some caution in the interpretation of the results from the individual subregions (IND, PAC and ATL). In particular, while the IND region appears to play a major role in the warming over Northern Eurasia, we conclude that fully reproducing the warming over central northern Eurasia requires that the replay region extend into the PAC region to encompass the broader warm pool region (Indian Ocean plus western Pacific). Also, it is likely that RPL_ATL is replaying tropical circulation features that are in part forced from the IND region, leading to double counting of the extratropical anomalies (e.g., a too strong combined circulation response over North America and corresponding too widespread warming over North America).

References

Ambaum, M. H. P., Hoskins, B. J., & Stephenson, D. B. (2001). Arctic Oscillation or North Atlantic Oscillation?, Journal of Climate, 14(16), 3495-3507.

Bader, J., & Latif, M. , 2005: North Atlantic Oscillation Response to Anomalous Indian Ocean SST in a Coupled GCM, Journal of Climate, 18(24), 5382-5389.

Bloom, S.C., L.L. Takacs, A.M. da Silva, and D. Ledvina, 1996: Data Assimilation Using Incremental Analysis Updates. *Mon. Wea. Rev.,* 124, 1256–1271.

Bosilovich, M. G., 2015: *GMAO Office Note No. 9 (Version 1.1)*: MERRA-2: File Specification. Available at https://gmao.gsfc.nasa.gov/pubs/

Ciavarella, A., Cotterill, D., Stott, P. et al., 2021: Prolonged Siberian heat of 2020 almost impossible without human influence. Climatic Change 166, 9 (2021).

Chang, E. K. M. (1995). The Influence of Hadley Circulation Intensity Changes on Extratropical Climate in an Idealized Model, Journal of Atmospheric Sciences, 52(11), 2006-2024.

Chang, Y., S.D. Schubert, R.D. Koster, A.M. Molod, and H. Wang, 2019: Tendency Bias Correction in Coupled and Uncoupled Global Climate Models with a Focus on Impacts over North America. *J. Climate,* **32**, 639–661.

Collow, A. B. M., N. P. Thomas, M. G. Bosilovich, Y.-K. Lim, S. D. Schubert, and R. D. Koster, 2021. Seasonal Variability in the Mechanisms Behind the 2020 Siberian Heatwaves. J. Climate, under review.

Douville,H., S. Bielli, C. Cassou, M.Dequ e, N. M. J.Hall, S. Tyteca, and A. Voldoire, 2011: Tropical influence on boreal summer mid-latitude stationary waves. Climate Dyn., 38, 1783–1798.

Edmon , H. J., Jr., Hoskins, B. J., & McIntyre, M. E. (1980). Eliassen-Palm Cross Sections for the Troposphere, Journal of Atmospheric Sciences, 37(12), 2600-2616.

Eliassen, A. and E. Palm, 1961: On the transfer of energy in stationary mountain waves. Geofys. Publ., 22, No. 3, 1-23.

Feldstein, S. B., 2000: The Timescale, Power Spectra, and Climate Noise Properties of Teleconnection Patterns, Journal of Climate, 13(24), 4430-4440.

Feldstein, S.B., 2003: The Dynamics of NAO teleconnection pattern growth and decay. Quarterly Journal of the Royal Meteorological Society. 129. 901 - 924. 10.1256/qj.02.76.

Feldstein, S. B., & Franzke, C. (2006). Are the North Atlantic Oscillation and the Northern Annular Mode Distinguishable?, Journal of the Atmospheric Sciences, 63(11), 2915-2930.

Gelaro, R., W. McCarty, M.J. Suárez, R. Todling, A. Molod, L. Takacs, C.A. Randles, A. Darmenov, M.G. Bosilovich, R. Reichle, K. Wargan, L. Coy, R. Cullather, C. Draper, S. Akella, V. Buchard, A. Conaty, A.M. da Silva, W. Gu, G. Kim, R. Koster, R. Lucchesi, D. Merkova, J.E. Nielsen, G. Partyka, S. Pawson, W. Putman, M. Rienecker, S.D. Schubert, M. Sienkiewicz, and B. Zhao, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Climate,* 30, 5419–5454.

Guirguis, K., Gershunov, A., Cayan, D.R. et al., 2018: Heat wave probability in the changing climate of the Southwest US. Clim Dyn 50, 3853–3864. https://doi.org/10.1007/s00382-017-3850-3.

Hoerling, M. P., J. W. Hurrell, and T. Y. Xu, 2001: Tropical origins for recent North Atlantic climate change. Science, 292, 90–92.

Hoskins, B. J., I. N. James, and G. H. White, 1983: The shape propagation and mean flow interaction of large-scale weather systems. J. Atmos. Sci., 40, 1595–1612.

Hurrell, J. W., 1995: Decadal trends in the North Atlantic oscillation: Regional temperatures and precipitation, Science, 269, 676–679.

Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck, 2003: An overview of the North Atlantic Oscillation, in The North Atlantic Oscillation: Climatic Significance and Environmental Impact, edited by J. W. Hurrell et al., pp. 1–35, AGU, Washington, D. C., doi:10.1029/134GM01.

Lawrence, Z. D., Perlwitz, J., Butler, A. H., Manney, G. L., Newman, P. A., Lee, S. H. and Nash, E. R., 2020: The remarkably strong Arctic stratospheric polar vortex of Winter 2020: links to record-breaking arctic oscillation and ozone loss. Journal of Geophysical Research: Atmospheres, 125 (22). e2020JD033271. ISSN 2169-8996

Leathers, D. J., Yarnal, B., & Palecki, M. A. (1991). The Pacific/North American Teleconnection Pattern and United States Climate. Part I: Regional Temperature and Precipitation Associations, Journal of Climate, 4(5), 517-528.

Lee, S. H., Lawrence, Z. D., Butler, A. H., & Karpechko, A. Y., 2020: Seasonal forecasts of the exceptional Northern Hemisphere winter of 2020. *Geophysical Research Letters*, 47, e2020GL090328.

Limpasuvan, V., & Hartmann, D. L. (2000). Wave-Maintained Annular Modes of Climate Variability, Journal of Climate, 13(24), 4414-4429.

Molod, A. M., L. Takacs, M. Suarez, and J. Bacmeister, 2015: Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2. *Geosci. Model Dev.*, 8, 1339-1356, doi:10.5194/gmd-8-1339-2015.

NOAA, 2020: Global Climate report. https://www.ncdc.noaa.gov/sotc/global-regions/2020.

Overland, JE, Wang, M, 2021: The 2020 Siberian heat wave. Int J Climatol., 41 (Suppl. 1): E2341– E2346.

Reichle, R. H., R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Toure (2011), Assessment and enhancement of MERRA land surface hydrology estimates, *Journal of Climate*, 24, 6322-6338, doi:10.1175/JCLI-D-10-05033.1.

Richman, M. B., 1986: Rotation of principal components. J. Climatol., 6, 293–335, doi:10.1002/joc.3370060305.

Rivière, G., and M. Drouard, 2015: Dynamics of the northern annular mode at weekly time scales. *J. Atmos. Sci.*, **72**, 4569–4590, doi:10.1175/JAS-D-15-0069.1.

Schubert, S. D., Suarez, M. J., Chang, Y., & Branstator, G. (2001). The Impact of ENSO on Extratropical Low-Frequency Noise in Seasonal Forecasts, Journal of Climate, 14(10), 2351-2365.

Schubert, S.D. and Y.-K. Lim, 2013: Climate Variability and Weather Extremes: Model-Simulated and Historical Data, in *Hydrologic Extremes in a Changing Climate - Detection, Analysis & Uncertainty.* Sorooshian, Soroosh, Easterling, David, AghaKouchak, Amir, Schubert, Siegfried, Hsu, Kuolin, Editors, Springer, Water Science and Technology Library, Volume 65, 2013, DOI: 10.1007/978-94-007-4479-0, pp423.

Schubert, S., A. Borovikov, Y-K. Lim, and A. Molod, 2019a: Ensemble Generation Strategies Employed in the GMAO GEOS-S2S Forecast System. *Technical Report Series on Global Modeling and Data Assimilation,* Randal D. Koster, editor. NASA/TM-2019-104606, Vol. 53, 65 pp.

Schubert, S. D., Y. Chang, H. Wang, R. D. Koster, and A. M. Molod, 2019b. A Systematic Approach to Assessing the Sources and Global Impacts of Errors in Climate Models. J. Climate, 32, 8301-8321. doi: 10.1175/JCLI-D-19-0189.1.

Schubert, S.D, Y. Chang, A. M. DeAngelis, H. Wang, and R.D. Koster, 2021: On the Development and Demise of the Fall 2019 Southeast U. S. Flash Drought: Links to an Extreme Positive IOD, *J.* Climate, 34(5), 1701-1723.

Seager, R., Harnik, N., Kushnir, Y., Robinson, W., & Miller, J., 2003: Mechanisms of Hemispherically Symmetric Climate Variability, Journal of Climate, 16(18), 2960-2978.

Simmons, A. J., J. M. Wallace, and G. Branstator, 1983: Barotropic wave propagation and instability, and atmospheric teleconnection patterns. J. Atmos. Sci., 40, 1363–1392.

Thompson, D.W.J., and J.M. Wallace, 1998: The Arctic oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.,* **25**, 1297 1300.

Thompson, D. W. J., , and J. M. Wallace, 2001: Regional climate impacts of the Northern Hemisphere annular mode. *Science*, **293**, 85–89, doi:10.1126/science.1058958.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere Winter. Mon. Wea. Rev., 109, 784-812.

WMO, 2021: State of the Global Climate 2020: Unpacking the indicators. World Meteorological Organization, January 14, 2021. https://public.wmo.int/en/our-mandate/climate/wmo-statement-state-of-global-climate.

Wu, R., and B. P. Kirtman, 2004: Impacts of the Indian Ocean on the Indian summer monsoon–ENSO relationship. J. Climate, **17**, 3037–3054.

Table 1: List of the AGCM experiments discussed in the text. All runs are forced with observed daily mean SST (see Gelaro et al. 2017 for details about the SSTs). NORPL_CLIM, is an average of two climatologies, one based on a single continuous 39-year (1981-2019) simulation starting from MERRA-2 initial conditions in January 1979, and the other is based on 39 shorter runs, each initialized from MERRA-2 on May 31 of each year (1980-2018) and run through the following year (1981-2019). The first 7 months of each run are not used. All the climatologies with replay are similarily based on 39 runs initialized on May 31 of 1980-2018. This segmented approach to computing the climatologies was done to allow running them in parallel. For all other runs, the perturbations to the initial conditions for each of the 90 ensemble members were produced by taking the differences between two MERRA-2 atmospheric states separated by one day (using the 15 days prior to 11/30/2018), scaling those differences by 1/8, and adding them to (or subtracting them from) the MERRA-2 initial state. Further information about that approach to perturbing initial conditions can be found in Schubert et al. 2019a.

Name	Time period	Replay region	Ensemble members
NORPL_CLIM	1981-2019: one continuous run, and a set of 39 shorter runs initialized May 31 of each previous year	none	2
RPL_TR_CLIM	1981-2019: a set of 39 shorter runs initialized May 31 of each previous year	Tropics: 25°S-25°N	1
RPL_PAC_CLIM	same as in RPL_TR_CLIM	Pacific Ocean region (25°S-25°N, 120E°-120°W)	1
RPL_ATL_CLIM	same as in RPL_TR_CLIM	Atlantic Ocean region (25°S-25°N, 120°W-0°)	1
RPL_IND_CLIM	same as in RPL_TR_CLIM	Indian Ocean region (25°S-40°N, 0-120°E)	1
NORPL	11/30/2018 – 06/30/2020	none	90
RPL_TR	11/30/2018 – 06/30/2020	Tropics: 25°S-25°N	90
RPL_PAC	11/30/2018 – 06/30/2020	Pacific Ocean region (25°S-25°N, 120°E-120°W)	90
RPL_ATL	11/30/2018 – 06/30/2020	Atlantic Ocean region (25°S-25°N, 120°W-0°)	90
RPL_IND	11/30/2018 – 06/30/2020	Indian Ocean region (25°S-40°N, 0-120°E)	90

Table 2: Probability of the model PCFs exceeding MERRA-2 values (PEM), for each experiment and leading REOF, for the 2020 JFM mean. MERRA-2/observed PCO values are

given in the top row for reference. PEM values are the mean and standard deviation (in parenthesis) obtained by chosing at random 10,000 subsamples of size 80 (from the 90 ensemble members).

Experiment	**PNA** (PCO = -1.62)	**AO** (PCO = 2.48)	**NAO** (PCO = 1.77)
NORPL	0.18 (0.014)	0.00 (0.000)	0.00 (0.000)
RPL_TR	0.36 (0.018)	0.11 (0.012)	0.12 (0.012)
RPL_PAC	0.31 (0.018)	0.10 (0.011)	0.03 (0.007)
RPL_IND	0.53 (0.019)	0.08 (0.010)	0.17 (0.014)
RPL_ATL	0.40 (0.019)	0.03 (0.007)	0.24 (0.016)

Table 3: Probability of the model PCFs exceeding MERRA-2 values (PEM), for each experiment and leading REOF, for March 2020. MERRA-2/observed PCO values are given in the top row for reference. PEM values are the mean and standard deviation (in parenthesis) obtained by chosing, at random, 10,000 subsamples of size 80 (from the 90 ensemble members).

Experiment	**PNA** (PCO =-3.28)	**AO** (PCO = 3.49)	**NAO** (PCO = 2.24)
NORPL	0.11 (0.012)	0.00 (0.000)	0.00 (0.000)
RPL_TR	0.23 (0.016)	0.01 (0.004)	0.07 (0.007)
RPL_PAC	0.26 (0.016)	0.04 (0.008)	0.07 (0.009)
RPL_IND	0.23 (0.016)	0.07 (0.009)	0.04 (0.008)
RPL_ATL	0.26 (0.016)	0.04 (0.008)	0.17 (0.014)

Table A1: List of additional replay experiments to help assess the robustness of the regional replay experiments. All runs are forced with observed daily mean SST (see Gelaro et al. 2017 for details about the SSTs). The climatologies are based on 39 runs initialized on May 31 of the previous year. The perturbations to the initial conditions for each of the 90 ensemble members for 2020 were produced by taking the differences between two MERRA-2 atmospheric states separated by one day (using the 15 days prior to 11/30/2018), scaling those differences by 1/8, and adding them to (or subtracting them from) the MERRA-2 initial state.

Name	Time period	Replay region	Ensemble members
RPL_IND+PAC_CLIM	1981-2019: a set of 39 shorter runs initialized May 31 of each previous year	Indian Ocean region (25°S-40°N, 0-120°E) + Pacific Ocean region (25°S-25°N, 120E°-120°W)	1
RPL_IND+ATL_CLIM	same as RPL_IND+PAC_CLIM	Indian Ocean region (25°S-40°N, 0-120°E) + Atlantic Ocean region (25°S-25°N, 120°W-0°)	1
RPL_IND+PAC	11/30/2018 – 06/30/2020	Indian Ocean region (25°S-40°N, 0-120°E) + Pacific Ocean region (25°S-25°N, 120E°-120°W	90
RPL_IND+ATL	11/30/2018 – 06/30/2020	Indian Ocean region (25°S-40°N, 0-120°E) + Atlantic Ocean region (25°S-25°N, 120°W-0°)	90

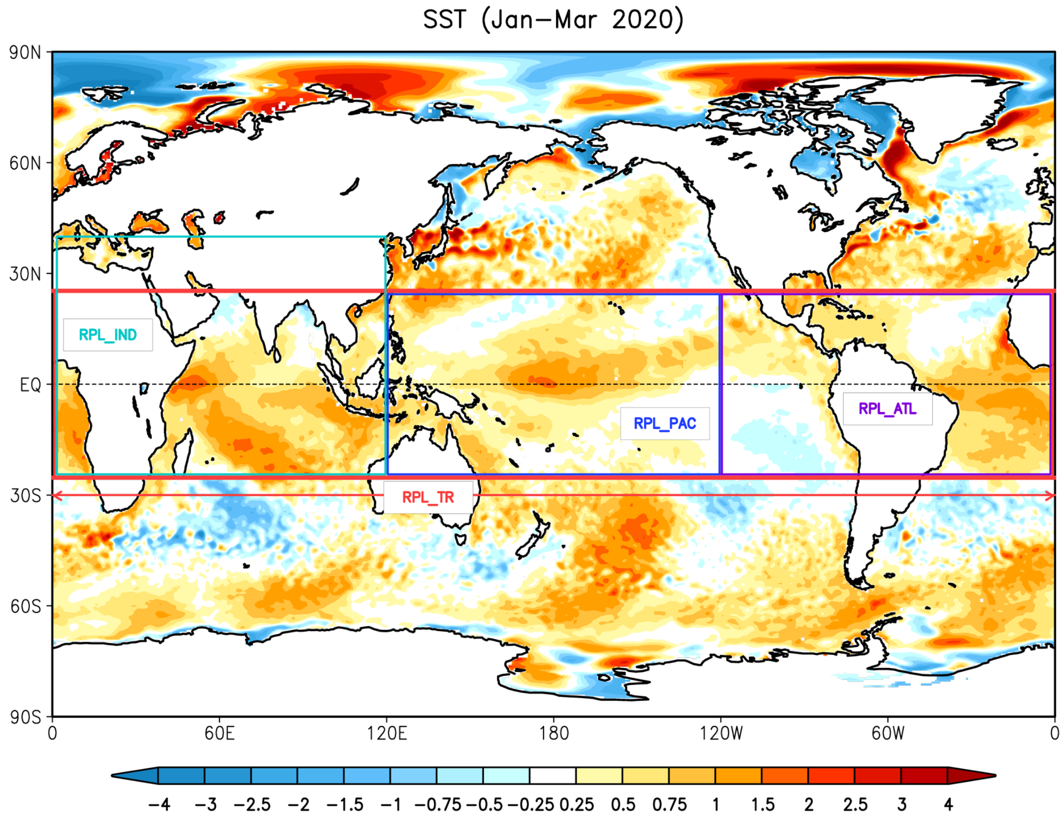

Figure 1: The various replay regions discussed in the text (outlined by the boxes, see Table 1), superimposed on the observed 2020 January - March mean SST anomalies (°C). See Gelaro et al. (2017) for a description of the observed SSTs. The regions consist of the full tropics (RPL_TR, outlined in red), a region that includes the Indian Ocean (RPL_IND, outlined in turquoise), a region covering much of the tropical Pacific (RPL_PAC, outlined in blue), and a region that encompasses the tropical Atlantic (RPL_ATL, outlined in purple). The dashed line indicates the equator.

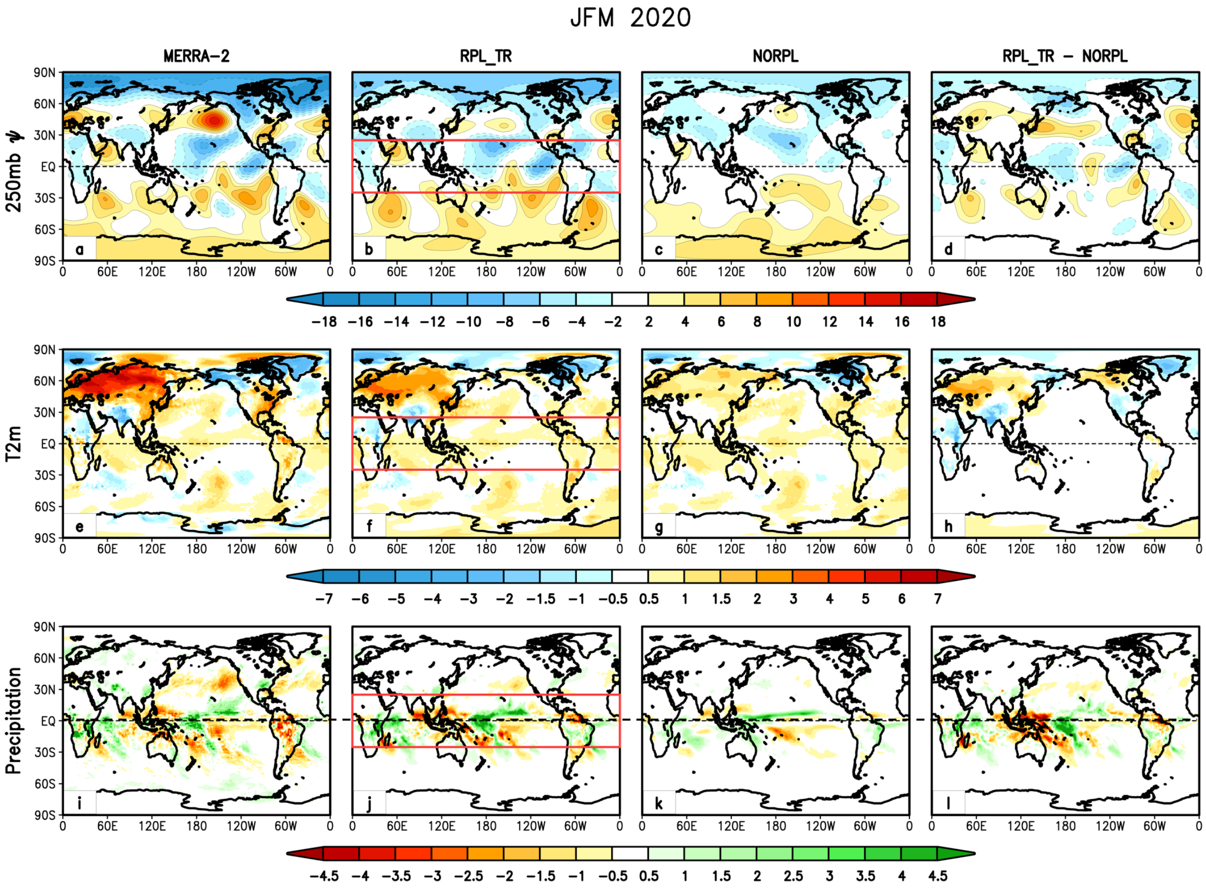

Figure 2: Top panels: 250mb stream function anomalies (Ψ, 10^6 m^2/s). Middle panels: two-meter temperature anomalies (T2M, °C). Bottom panels: precipitation anomalies (mm/day). From left to right the results are for MERRA-2, RPL_TR, NORPL and RPL_TR - NORPL. The JFM 2020 anomalies are computed with respect to the appropriate 1981-2019 climatologies (see Table 1 for details). All model results are the averages of 90 ensemble members. The red lines outline the tropical replay region. The dashed lines indicate the equator.

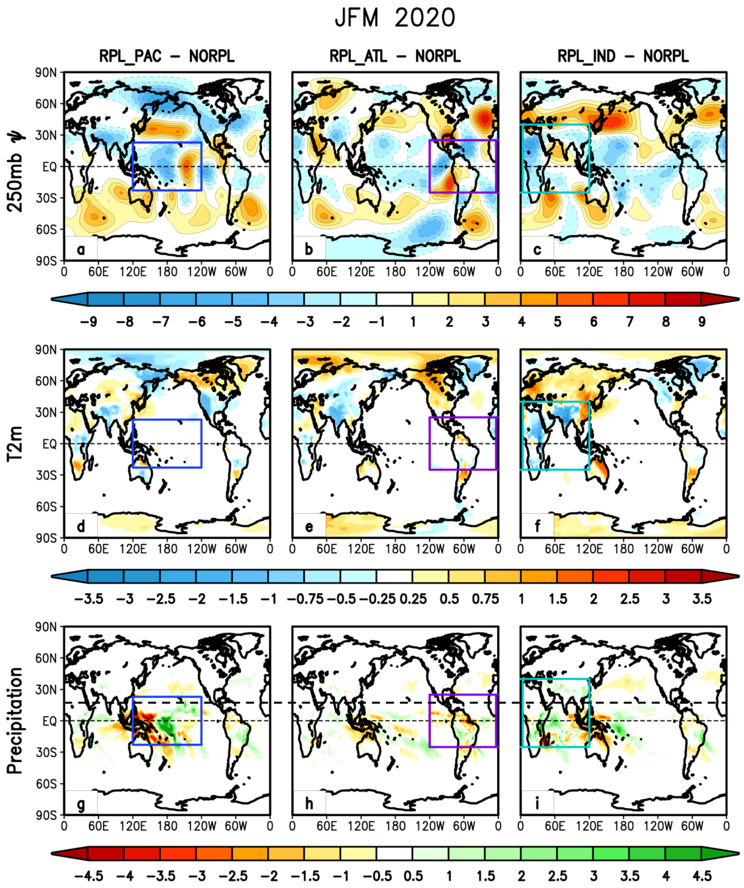

Figure 3: Top panels: 250mb stream function anomalies (Ψ, 10^6 m^2/s). Middle panels: two-meter temperature anomalies (T2M, °C). Bottom panels: precipitation anomalies (mm/day). From left to right the results are for RPL_PAC - NORPL, RPL_ATL - NORPL and RPL_IND - NORPL. The JFM 2020 anomalies are computed with respect to the appropriate 1981-2019 climatologies (see Table 1 for details). Note that in the top and middle panels the contour intervals are ½ those in Figure 2. The boxes outline the corresponding replay regions. The dashed lines indicate the equator.

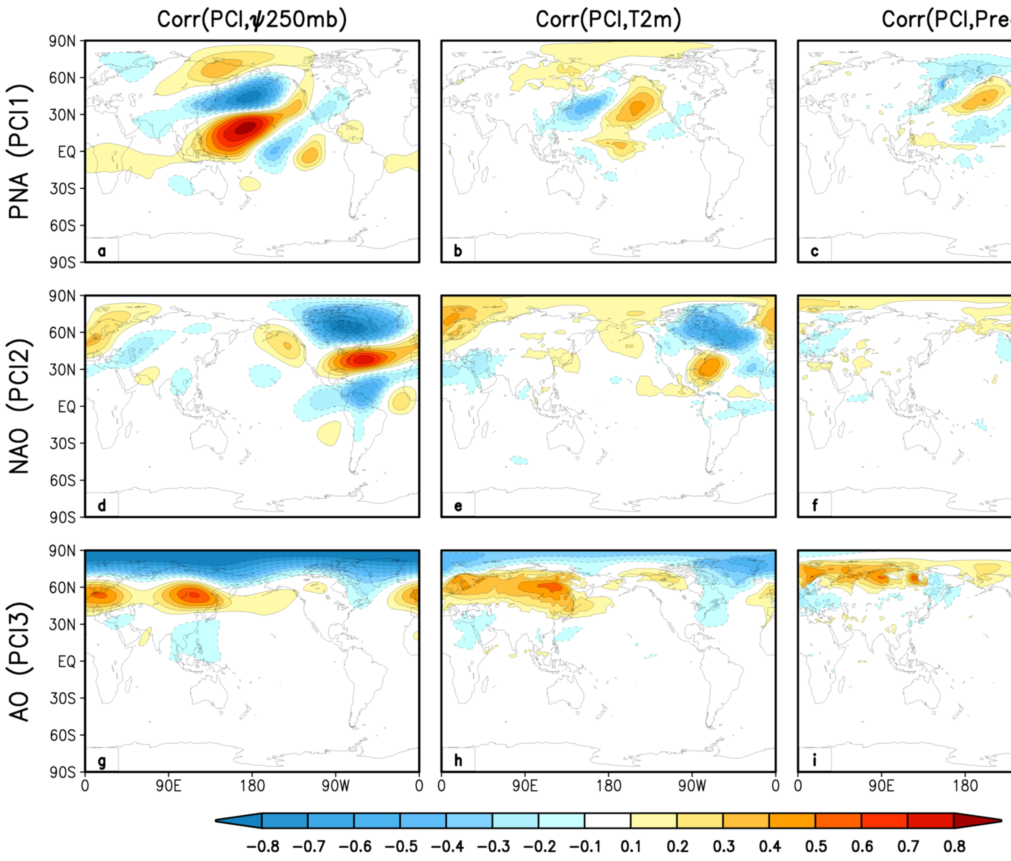

Figure 4: The correlations between the intra-ensemble monthly principal components (PCIs) associated with the three leading REOFs and, the 250mb stream function (left column), T2m (middle column) and precipitation (right column) for January - March of 2020. The correlations are computed by pooling together all the replay and no-replay simulations. The top row is for PCI 1 (PNA), the middle row is for PCI 2 (NAO), and the bottom row is for the PCI 3 (AO).

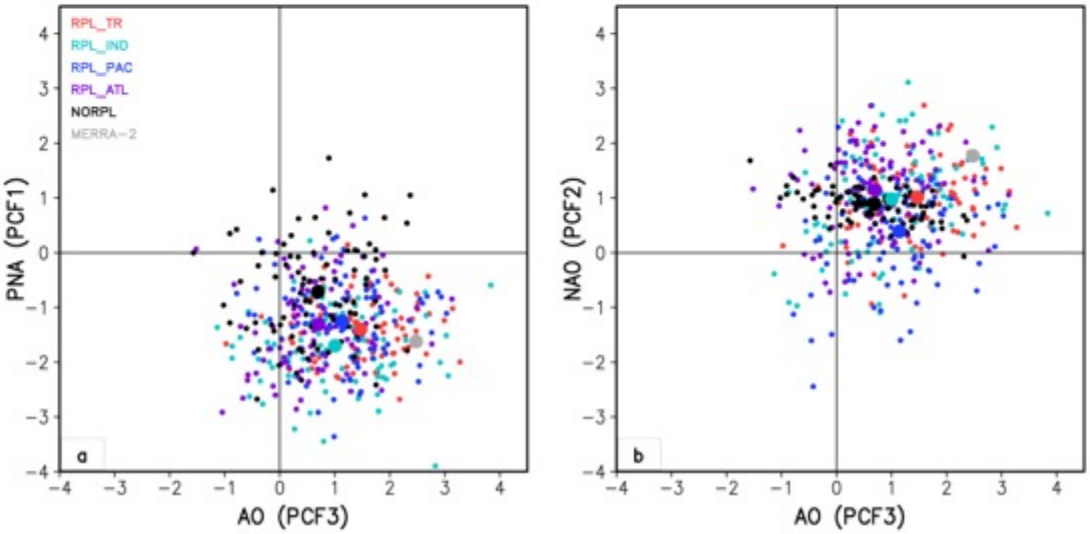

Figure 5: Scatterplots of the JFM mean PCFs for a) PNA versus AO, and b) NAO versus AO. The small dots denote the individual ensemble members, and the large dots are the ensemble mean values. The large grey dots denote the MERRA-2 anomaly projected onto the model REOFs (the PCOs). Results are color coded to show RPL_TR in red, RPL_IND in turquoise, RPL_PAC in blue, RPL_ATL in purple, and NORPL in black. Units are 10^6 m^2/s.

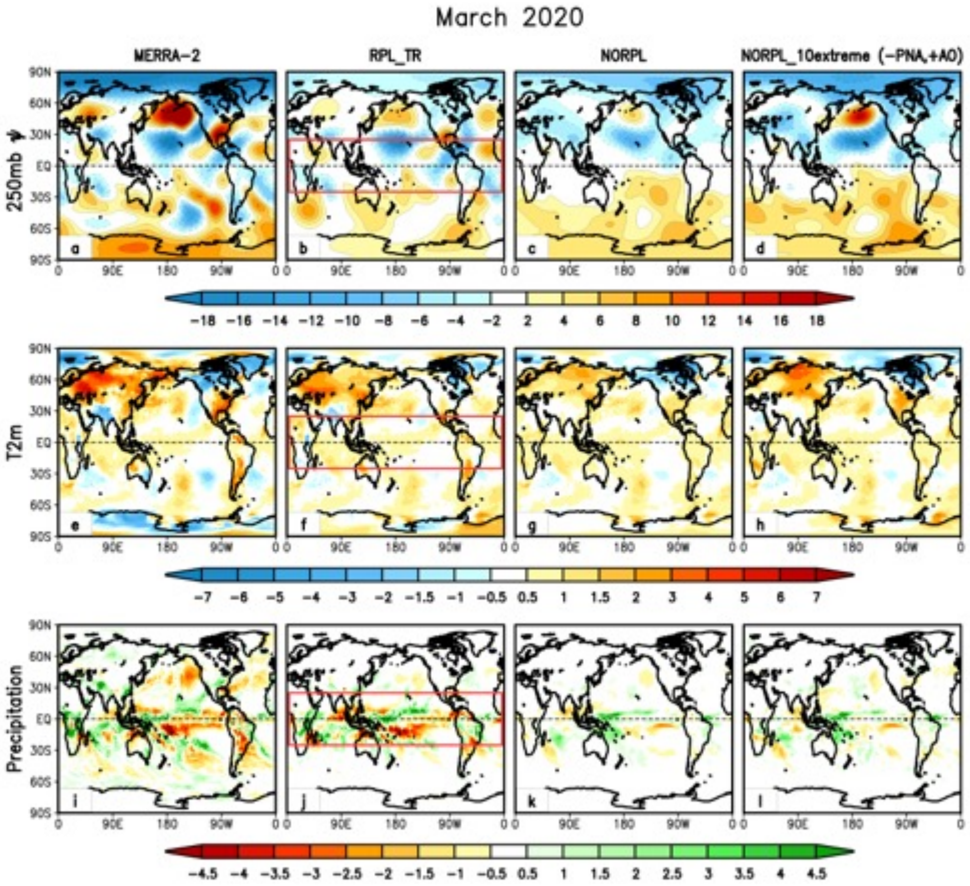

Figure 6: Top panels: 250mb stream function anomalies (Ψ, 10^6 m^2/s). Middle panels: two-meter temperature anomalies (T2M, °C). Bottom panels: precipitation anomalies (mm/day). From left to right the results are for MERRA-2, RPL_TR, NORPL, and an average of the 10 NORPL ensemble members with the largest PNA and AO amplitudes (indicated by circles in Fig. 8, see text). The March 2020 anomalies are computed with respect to the appropriate climatologies (1981-2019, see Table 1 for details). The red lines outline the tropical replay region. The dashed lines indicate the equator.

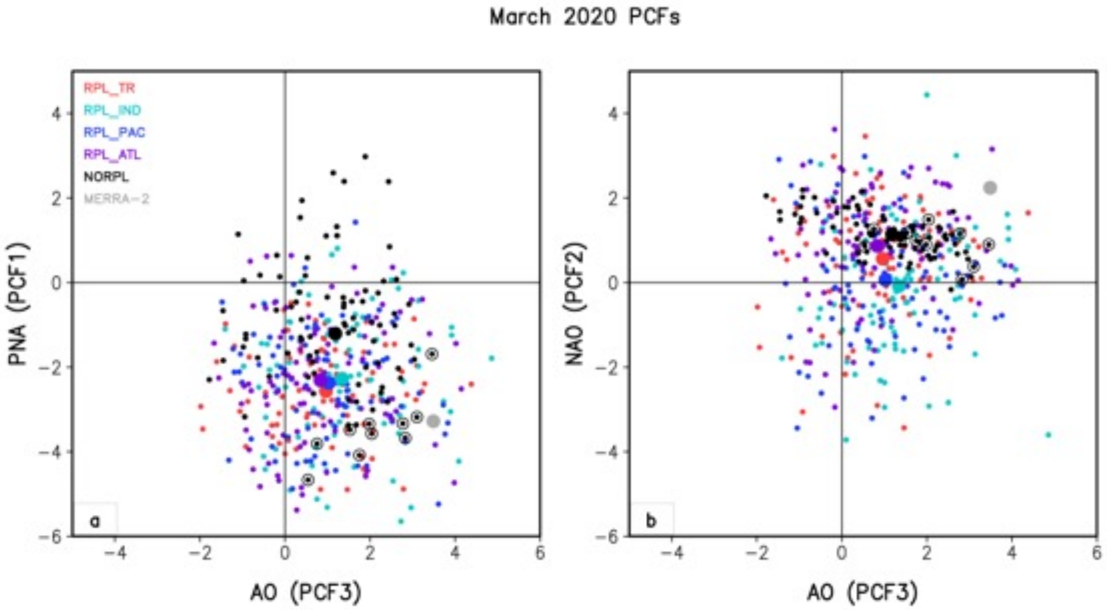

Figure 7: Scatterplots of a) PNA versus AO and b) NAO versus AO for the March 2020 PCFs. The small dots denote the individual ensemble members, and the large dots are the ensemble mean values. The large grey dots denote the MERRA-2 anomaly projected onto the model REOFs (the PCOs). Results are color coded to show RPL_TR in red, RPL_IND in turquoise, RPL_PAC in blue, RPL_ATL in purple, and NORPL in black. The encircled dots identify those NORPL ensemble members used to produce the mean values shown in the last column of Fig. 6. Units are 10^6 m^2/s.

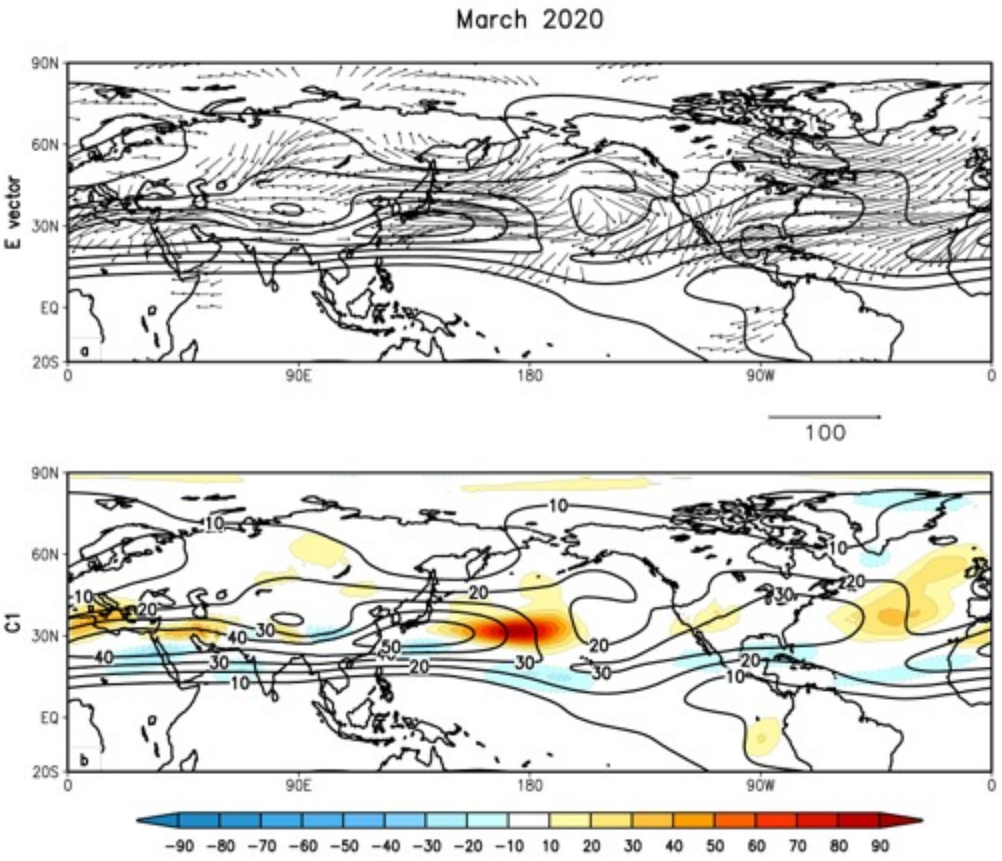

Figure 8: The a) E-vectors (m^2s^{-2}), and b) local barotropic energy conversion (C1, $m^2s^{-2}day^{-1}$) at 250mb for March 2020 for the NORPL runs. The heavy contours in both panels are the ensemble mean 250mb u-wind (m/s).

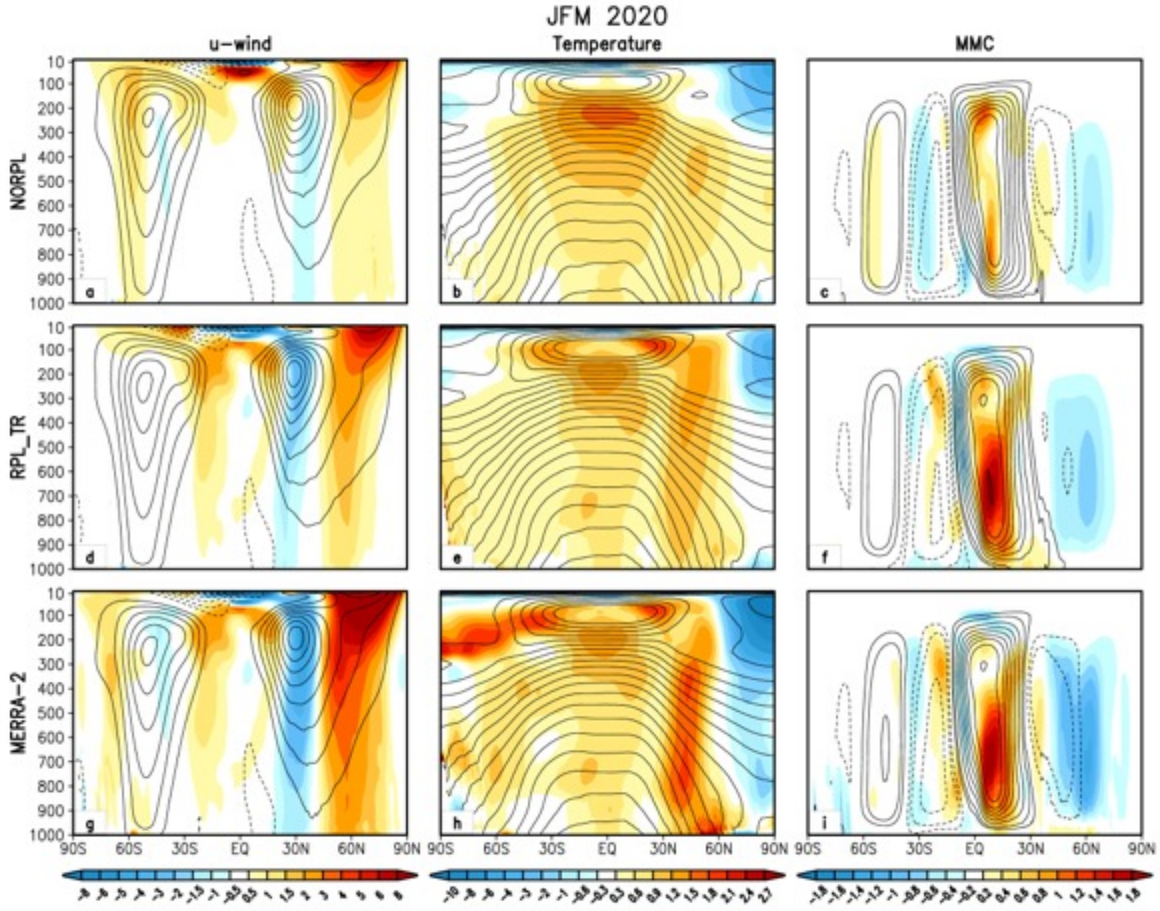

Figure 9: JFM mean 2020 zonal mean u-wind (m/s, left panels), temperature (°C, middle panels) and mean meridional circulation (10^{10} kg/sec, right panels) for NORPL (top row), RPL_TR (middle row), and MERRA-2 (bottom row). Shading indicates the 2020 anomalies, and the contours are the climatological mean (1981-2019) values. For the climatological values, the u-wind contour levels are drawn in increments of 5 m/s (negative dashed and excluding zero), the temperature contour levels are drawn from 200K to 300K in increments of 5K, and the MMC contour levels are drawn at +/- 1, 2, 4, 6, 8, 10, 12, 15, and 20 (10^{10} kg/sec, negative dashed).

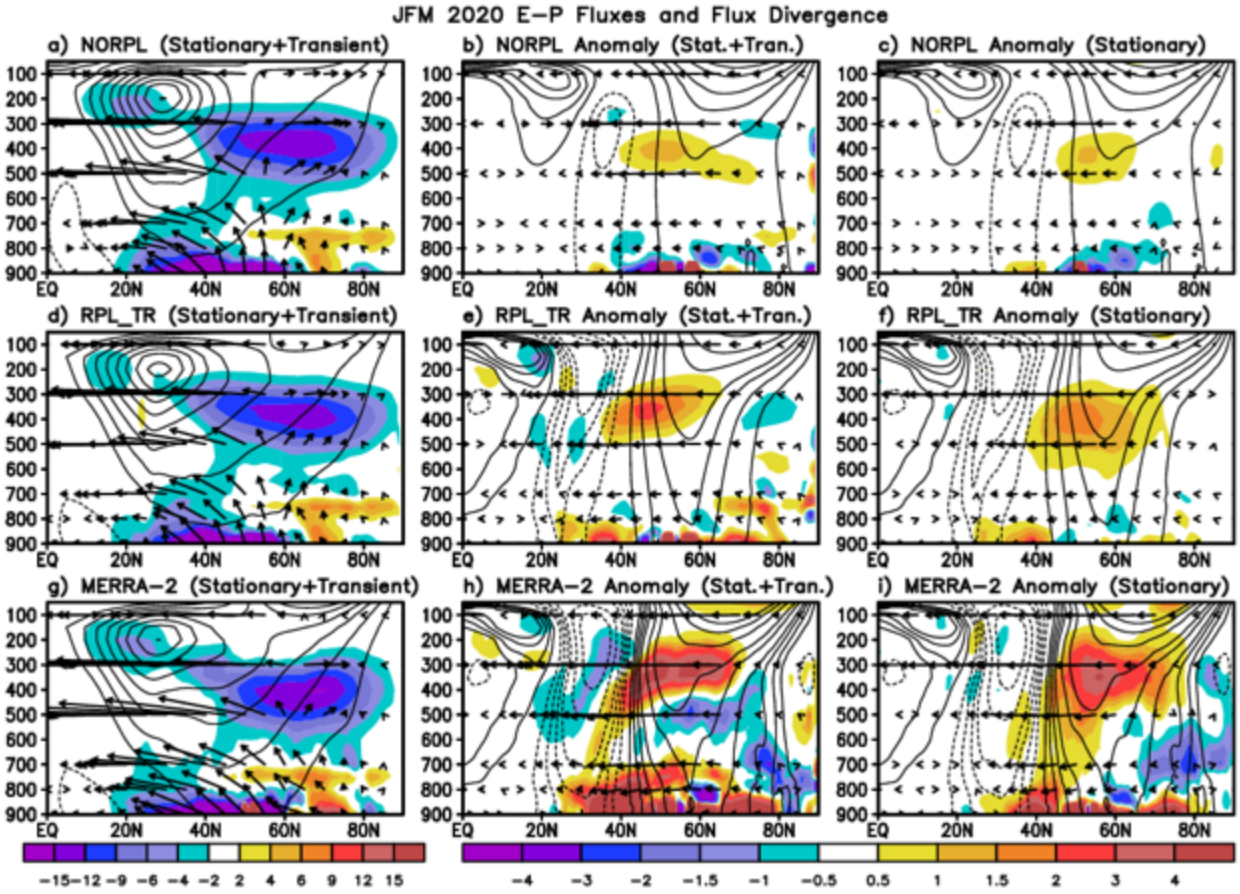

Figure 10: The 2020 January-March E-P flux (F, vectors) and divergence of F (shaded, m/s day^{-1}) computed for each month separately, and then averaged. Results are for NORPL (top panels) RPL_TR (middle panels), and MERRA-2 (bottom panels). The left panels are for the full fields (stationary +transient), the middle panels are for the anomalies (stationary +transient), and the right panels are again the anomalies but for only the stationary component. The contours are the zonal mean zonal winds with contour intervals of 5m/s for the plots in the left column, and (-8 -6 -4 -3 -2 -1.5 -1 -0.5 0.5 1 1.5 2 3 4 6 8) m/s for the plots in the middle and right columns (the anomalies). The model results are the averages of 90 ensemble members.

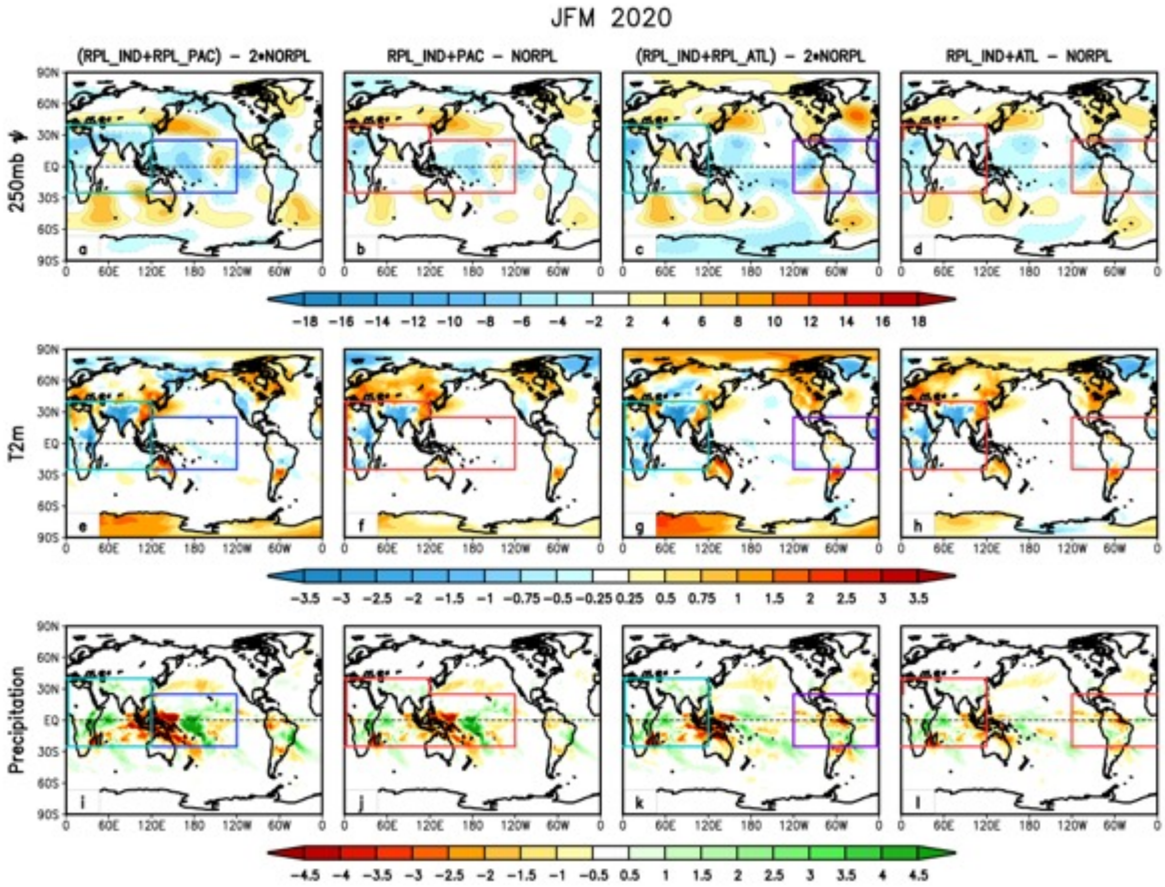

Figure A1: The JFM 2020 replay anomalies (top: Ψ, 10^6 m^2/s; middle: T2m, °C; and bottom: precipitation, mm/day), for (from left to right), the sum of RPL_IND and RPL_PAC, replaying the region that encompasses the IND and PAC regions (RPL_IND+PAC), the sum of RPL_IND and RPL_ATL, and replaying the region that encompasses the IND and ATL regions (RPL_IND+ATL). All results are the averages of 90 ensemble members and have NORPL removed. The dashed lines indicate the equator.