GISS Model E2.2: A Climate Model Optimized for the Middle Atmosphere. Part 1: Model Structure, Climatology, Variability and Climate Sensitivity

D. Rind[1], C. Orbe[1], J. Jonas[2], L. Nazarenko[2], T. Zhou[2], M. Kelley[3], A. Lacis[1], D. Shindell[4], G. Faluvegi[2], A. Romanou[1], G. Russell[1], N. Tausnev[3], M. Bauer[2] and G. Schmidt[1]

[1]NASA Goddard Institute for Space Studies, 2880 Broadway, New York, N.Y.

[2]Center for Climate Systems Research, Columbia University, New York, N.Y.

[3]SciSpace LLC, 6550 Rock Spring Drive, Suite 600, Bethesda, MD 20817

[4]Duke University, Durham, NC 27708

Corresponding author: David Rind (david.h.rind@nasa.gov)

Key Points:

- Presentation of new GISS "high-top" Climate Model
- Advantages of higher model top and increased vertical resolution
- Need for Middle Atmosphere to be a focus in climate model development

Abstract

We introduce a new climate model (GISS E2.2) that has been specially optimized for the Middle Atmosphere, and whose output is being contributed to the CMIP6 archive. The top of the model is at a geopotential altitude of 89 km, and parameterizations of moist convection and various forms of gravity wave drag based on tropospheric processes are chosen specifically for this optimization. We first evaluate the model in its configuration as a coupled atmosphere-chemistry model with respect to its simulation of the mean state of the Middle Atmosphere, from the mesosphere down through the Upper Troposphere/Lower Stratosphere. Then we assess its use as a coupled atmosphere-ocean climate model by exploring its mean ocean climatology. To evaluate its variability, we report on its simulation of the primary modes in the troposphere, stratosphere and ocean. Two climate change simulations are presented, the responses to instantaneous increases of $2xCO_2$ and $4xCO_2$, run with two different ocean models. Sensitivity studies are performed to illustrate the effect of parameterizations on the model results. We compare these results to the lower vertical resolution/top GISS Model E2.1, whose output has also been submitted to CMIP6. The different choices made for these models are explored. It is shown that important improvements in the circulation above and below the tropopause can be obtained when attention is paid to representation of Middle Atmosphere processes in climate model development.

Plain Language Summary

A new 'high-top' GISS Climate Model with results submitted to the CMIP6 archive is described and the model development is explained. The model was optimized to produce a realistic Middle Atmosphere, which involved its simultaneous assessment along with its simulation of tropospheric climate. In particular, the interplay between the parameterization of convection in the troposphere and the dynamics of the Middle Atmosphere requires appreciation of the needs of both regions.

1 Introduction

There is a growing awareness of the importance of the Middle Atmosphere for understanding and affecting climate and climate change. It has long been appreciated that trace gases such as stratospheric water vapor and ozone impact the radiative forcing at the surface and top of the atmosphere, and models have increasingly provided sufficient vertical layering and stratospheric depth to allow such calculations to be included. The potential importance of stratospheric *dynamical* forcing operating over various time scales, both in situ and of the troposphere, has now gained greater attention. On short, 'weather' (synoptic) time-scales in the extratropics there is the relationship between cold "polar vortex" intrusions into the troposphere and sudden stratospheric warmings (SSWs) (e.g., Baldwin et al., 2001, Christiansen et al. 2005); in the month after an SSW there is anomalously high pressure over the pole, hence a negative Northern Annular Mode (NAM). In the tropics there is a potential relationship between the Quasi-biennial Oscillation (QBO) and the Madden-Julian Oscillation (MJO) (Yoo and Son, 2016; Son et al., 2017; Martin and Wang, 2019). On seasonal time scales in the extratropics there is a potential influence of lower stratospheric ozone or circulation anomalies

on tropospheric regional patterns (Fletcher et al., 2007) and in the tropics on the El Niño–Southern Oscillation (ENSO) (Mantasa 2017). On interannual time-scales, the QBO may influence surface climate (Garfinkel and Hartmann, 2011), with the meridional circulation induced by the QBO extending downward into the troposphere and interacting with subtropical eddies, thus potentially extending their influence further downward to the surface.

On climate time-scales associated with climate forcings, both natural and anthropogenic, there is the effect of altered dynamical re-distribution of stratospheric trace gases on radiative forcing, particularly ozone, as well as water vapor (e.g., Banerjee et al., 2019). There is also the influence of stratospheric volcanic aerosol effects on the North Atlantic Oscillation (NAO) (Wunderlitch and Mitchell 2017); the influence of solar forcing in the stratosphere on tropospheric dynamical patterns (Rind et al, 2008) including the Indian Ocean monsoon (Kodera, 2004); and the dynamical influence initiated by stratospheric ozone reduction on the Southern Annular Mode, the position of the Southern Hemisphere midlatitude jet (e.g., Polvani et al. (2011), Son et al. (2010)) and on the Southern Ocean ventilation (Waugh et al. (2013) and sea ice (Ferreira et al., 2015), among other impacts.

With these potential influences in mind, we present a new GISS climate model for which choices were made of parameterizations and structure for the sake of the veracity of its Middle Atmosphere simulations and stratosphere-troposphere coupling, as well as atmospheric dynamics in general. For example, choices in the convective parameterization can have unintended consequences on the stratospheric circulation for parameterized non-orographic gravity wave drag (GWD) from convective sources which affects the QBO, and on tropical upwelling in the lower stratosphere. This model (E2.2) builds off the version E2.1 (Kelley et al (2019), GISS-E2.1: Configurations and climatology. J. Adv. Model. Earth Syst., submitted) whose output was recently submitted to CMIP6 by using much of the same physics, chemistry and numeric; both models, like their predecessor from CMIP5, are coupled to two different ocean models. In this paper we introduce the E2.2 model, and discuss its simulations of the mean state, variability and climate changes, with sensitivity studies to show why certain choices were made. In a companion paper (Orbe et al. in preparation; henceforth ‘Paper 2’), we discuss the effect this model configuration has on the ability to transport chemical tracers and produce trace gas distributions, including its influence on climate change 'forcing'.

GISS has long had a history of exploring climate change and the middle atmosphere (Rind et al., (1988) and numerous subsequent papers) with the focus in many of those on how climate change would affect the middle atmosphere (for both the past (Rind et al., 2001) and the future (Rind et al., 1998)) or how stratospheric climate changes can influence tropospheric dynamics (e.g., Rind et al. (2005)). Uniting the middle atmosphere modeling effort with the general climate modeling effort allows for additional exploration of climate-Middle Atmosphere model development, as well as providing a more complete atmospheric “climate change” assessment. However, certain rules set up by CMIP6, rules requiring consistency of model code among all submitted versions that include the same atmospheric processes (i.e. those run with both coupled dynamical oceans and with specified SSTs (AMIP mode)) have consequences for Middle Atmosphere modeling. This results in less than optimum performances of some versions. We

thus include in this paper a sensitivity experiment to show what improvements are possible when these rules are relaxed.

According to the CMIP6 protocol (pcmdi.llnl.govCMIP6/Guide/modelers.html), each participating model must contribute results from the four DECK experiments (piControl, AMIP, abrupt4xCO2 and 1pctCO2). Although this paper is about the E2.2 model and not specifically about the CMIP6 submissions, in the course of its description we present results from the first three of these experiments. Because of the issue described above, we choose to use the AMIP run with interactive atmospheric chemistry, as that is allowed to have a different model code than the coupled ocean version. In addition, output from one of the Model E2.2 model configurations (Model E2.2AP) is also being submitted to SPARC DynvarMip (sparcdynvar.org); the rationale and diagnostics proposed by DynVarMip are presented in Gerber and Manzini (2016). This project emerged from the Stratosphere-Troposphere Process and their Role in Climate Project (SPARC) whose emphasis was on the modeling of the dynamics and variability of the stratosphere-troposphere system. DynVarMip promotes "Coupled Global Atmosphere, Ocean and Sea ice Models with high vertical resolution and location of the model tops above the stratopause". This paper will present and assess model E2.2 in all of those domains.

2 Models and Sensitivity Experiments

2.1. E2.2 Models

There are different configurations of E2.2, following the different simulations required for CMIP6. In addition, the atmospheric models were coupled to two different dynamical ocean models, the 'GISS' model (E2.2-G), and the 'HYCOM' model (E2.2-H). For the sake of brevity, the primary comparison here will be with E2.2-G, although occasionally E2.2-H will be presented when the question of ocean involvement arises. See Schmidt et al. (2014) for a discussion of these two ocean models.

With respect to the CMIP6 designations, we have the following:

(1) "OMA" (acronym for "one-moment aerosols") is the coupled atmosphere-chemistry version of the model, run here primarily with specified SSTs (i.e., AMIP mode). It is quite time-intensive, taking (on 88 processors) around one day for a one-year simulation.

(2) "Coupled" is the version of E2.2 run coupled to either of the two ocean models (and not with interactive atmospheric chemistry). Here it is run primarily utilizing trace gases (other than CO_2 and CH_4 which are prescribed) and climate forcings saved from the "OMA" simulation; as it is without chemistry interaction it is referred to as "NINT", or "non-interactive" mode (Schmidt et al., 2014). According to CMIP6 protocol, "OMA" and "Coupled" may have different tunings, parameterizations, etc. – hence can be 'different' models, since OMA includes different atmospheric processes (i.e., chemistry). This may have an influence particularly on climate change experiments, e.g., $4xCO_2$, where the stratospheric ozone feedback may well influence the results (Chiodo and Polvani, 2017). On 88 processors, "Coupled" runs at about 5 years per day. [Note: An OMA-coupled ocean

simulation requiring extensive computer time (and expense) is currently underway but the simulation runs ~ 1 year/every 2 days].

(3) "AMIP NINT" is the version of the model run with observed sea surface temperatures and without coupling to either atmospheric chemistry or the ocean (using NINT in the title is to distinguish this run from OMA which also employs prescribed SSTs). It also utilizes trace gases and climate forcing from the "OMA" simulations. CMIP6 indicates that "AMIP NINT" and "Coupled", if both are submitted, must have the same atmospheric code, presumably to facilitate comparison of results with and without the dynamical ocean. On 88 processors, this runs at 10 years per day. [As noted, the AMIP version we are actually submitting to CMIP6 is the OMA run as in (1) above]. This version provides the tuning to put the model in radiative balance for all 'Coupled' simulations.

2.2 Sensitivity Experiments

In addition, to explore the effects of model parameterizations on the Middle Atmosphere simulation, we report in this paper on three sensitivity experiments. Several further experiments are presented in the discussion section to investigate certain issues.

The following are the sensitivity experiments that differ from the main models discussed in this paper:

(1) E2.2 "Altered physics" (AP). In this version, there was an attempt to have the physics package interact with the new vertical resolution to mimic what was done in E2.1. Simulations for the four DECK experiments in CMIP6 listed in the introduction were all done with this model version as well.

(2) E2.2 "Rain evaporation" (RE). A major difference between the physics used in E.2.2 and E2.1 revolved around the question of evaporation of rain above the cloud base generated during a (parameterized) moist convective event. In E2.1 it was allowed; in E2.2 it was not. Some results indicating what would have happened if it were allowed in E2.2 are shown from this simulation, explored in the context of both OMA AMIP and Coupled (pre-industrial) runs.

(3) "AMIP NINT Version 2" is the version of the AMIP NINT model in which some of the atmospheric code (in particular the parameterized non-orographic gravity wave drag) is re-tuned (in comparison to configuration (3) in Section 2.1). Given the importance of climate change simulations in the CMIP6 Deck, E2.2 was optimized for its performance in 'Coupled' mode, not in AMIP NINT mode. To the extent that the coupled model is not able to reproduce the observed SSTs, the ocean forcing of the troposphere and thus of the middle atmosphere will be different in the two simulations. The tunings required to offset this altered SST forcing in the coupled model had to be included in any AMIP NINT version submitted to CMIP6, even though they were no longer appropriate. Here, the appropriate tuning is restored (and, for example, the QBO now has the right period). We present this AMIP NINT simulation in this paper so that it can be used subsequently for future (non-CMIP6) research.

2.3. Model Differences from E2.1

Horizontal resolution in the atmosphere for all of the E2.2 models is 2°x2.5°, the same as E2.1. The vertical resolution, however, is very different. Shown in Table 1 for the two models are the number of layers and average thickness in different regions of the atmosphere. In the troposphere, layers in E2.2 are less than one-half the thickness of those in E2.1, and in the stratosphere, one-half to one-third the thickness. The top of the model has been raised to 0.002 hPa (~89 km) with the inclusion of six more model layers above the previous model top. The E2.2 vertical layering is similar to the GISS Global Climate Middle Atmosphere model 3 (e.g., Rind et al., 2014). The position of the model top is consistent with that required to minimize its influence on the stratosphere, i.e., more than two scales heights above (e.g., Shepherd et al., 1996), and to allow sudden stratospheric warmings (SSWs) to be followed by the re-formation of the polar stratopause between 70–80 km (elevated stratopause events) (e.g. Chandran et al., 2014). The ocean resolution is on the order of 1°, with additional refinement at the equator in the HYCOM model (see Schmidt et al. (2014), Table 3).

Building off E2.1, as noted above, parameterization choices were made specifically in order to optimize the E2.2 Middle Atmosphere model in the context of the coupled model per CMIP6 protocol. This is a departure from previous studies (e.g. Rind et al. (2014)), and is the first time the Middle Atmosphere version of ModelE has been developed fully coupled to both on-line chemistry and with an interactive ocean, so some aspects of the GWD parameterization were revisited. There was a desire to properly simulate not only the QBO but also the stratospheric mean age of air (AoA) (Hall et al., 1994) which posed still other challenges, compared with previous papers in which the AoA was less emphasized. This last point will be discussed more fully in Paper 2.

In addition to the vertical resolution and model top, the following are the primary differences in E2.2 from E2.1, or from the previous version of the Middle Atmosphere model (GCMAM Model 3, e.g., Rind et al., 2014). Except where indicated, the model parameterizations of E2.2 are similar to E2.1; see Kelley et al. (2019, submitted) for those details. The changes are in all the E2.2 model codes.

(1) As in GCMAM Model 3, E2.2 contains gravity wave drag parameterizations for both orographic and non-orographic drag, the latter consisting of gravity waves associated with convection, shear, and deformation (flow in the vicinity of fronts), all derived from model resolved or parameterized processes (see Rind et al., 1988 for the details of the parameterized gravity wave generation, propagation, and breaking). Various refinements were implemented, including the definition of the breaking flux at layer edges, applying the moist convective drag over the entirety of the model grid cell, and improving the definition of the deformation. In addition, due to changes in the tropospheric simulation in this newer version of Model E, the ‘’efficiency factors’ that relates the specific processes to gravity wave momentum fluxes have been modified somewhat. However, the biggest alteration is associated with the inability of the coupled model to reproduce the observed sea surface temperatures, values being too warm in the eastern tropical Pacific and Atlantic Oceans. This leads to a displacement of the patterns of convection, and alters the zonal winds in the tropical UTLS (upper troposphere/lower stratosphere), aspects that have a negative effect on both the Quasi-biennial Oscillation (QBO) and stratospheric AoA. The parameterized

convective phase velocity for gravity waves then had to be altered to compensate (by +5 m/s) and the convective wave "efficiency factor", the relation between convection and parameterized gravity wave momentum fluxes, re-assessed. This factor is normally tuned in models with source-driven gravity wave drag to optimize the QBO, (e.g., Rind et al., 2014; Richter et al., 2014; Richter et al., 2019) and the need to re-tune it when convection changes would be a common modeling problem. The problem does not arise in E2.1, which does not utilize the parameterized convective gravity wave drag and does not have a QBO (for that, its vertical resolution in the low-to-mid stratosphere would also need to be increased).

(2) In E2.1, atmospheric radiation is called every fifth physics time step, hence every 2.5 hours. That proved insufficient to provide reasonably smooth temperatures from one radiation call to the next in the upper stratosphere and mesosphere. In E2.2 the radiation is called every physics time step (0.5 hours). The change also had the effect of improving the diurnal cycle.

(3) A detailed comparison was made between the results from the GISS radiation code and line by line calculations for all levels of the atmosphere. Several different changes in the long wave radiation code were implemented, primarily to adjust the outgoing longwave radiation associated with CO_2 in the upper stratosphere and mesosphere. With respect to the surface, downward longwave radiation flux decreased by some 3 W/m^2 in the tropics, and increased by varying amounts at high latitudes. This may have contributed to producing cooler tropical temperatures in the model.

(4) Parameterization choices were made to provide a better dynamical and hydrologic state relative to that of E2.1. The choices of greatest consequence were made in the moist convective parameterization. As noted above, E2.1 allowed evaporation of rainfall above cloud base in moist convective updrafts, an effect that was designed to improve sub-seasonal variability like the MJO. Kim et al. (2011) found that there is often a trade-off in such choices, as they may exacerbate climatological biases. In this case, it produced a number of quite negative consequences, which will be detailed in subsequent sections (and in Paper 2). In E2.2 this evaporation was not allowed above cloud base (hence returning to the formulation in E2.0, Schmidt et al. (2014)).

(5) With two exceptions, the gas phase chemistry formulation in E2.2 is that of the E2.1 model (Kelley et al., 2019, submitted) active in all model layers. Lifting of the model top in particular had a strong effect on the dynamical transport capability of the model, and its thermodynamic structure. These necessitated a re-tuning of some aspects of the chemistry code. The photolysis code experiences a partial ozone column assumption above the top of the chemistry. In E2.1, this overhead ozone was a constant. In E2.2 it is given a spatial variation matching that of the ozone at the top layer of chemistry. More importantly, E2.2 allowed us to disable the tuning of photolysis rates at short wavelengths (<200 nm) for N_2O and O_2, which in E2.1 corrected for stratospheric circulation-induced biases in high latitude NO_x and O_3.

(6) Lightning flash rates involved in the production of nitrogen oxides and ozone were retuned in association with the changes in layering and convection.

(7) Normally, GISS models have transitioned from sigma coordinates in the troposphere (pressures greater than 150 hPa) to constant pressure coordinates above. In the vicinity of

high topography, this makes for a rather abrupt change from the thin sigma layers to thicker layers in the constant pressure regime, an inconsistency that may induce numerical inaccuracies. To minimize this possibility in E2.1, a constant pressure component was added to each of the sigma levels, in such a way as to smooth that transition. However, when investigating its impacts in E2.2, it was found to induce various errors, including carrying the 'mask' of the topography into the stratosphere when used in 'nudging mode', and so it was omitted from this model version.

The changes associated with the physics utilized in E2.2AP will be mentioned along with the results from those simulations.

3 Results

Given the plethora of results that are important to show documenting the model's potential use as both a Middle Atmosphere model *and* a climate model, it is not possible to go into great depth in any one area. The presentations in this paper may serve as starting off points for various more in-depth papers to follow, or provide background for those using its CMIP6 output. Also, as the stratospheric processes during the winter hemisphere are driven by dynamical forcing from the troposphere, this paper emphasizes the dynamical aspects of E2.2, at all levels. A discussion of the radiative characteristics important for climate considerations is provided in association with the presentation of E2.2AP (Section 3.5.1), as well as in a related discussion (Section 4.3). A more complete discussion of radiative forcing for E2.1 can be found in Miller et al. (2020) (CMIP6 historical simulations (1850-2014) with GISS ModelE2.1. J. Adv. Model. Earth Syst., submitted).

The E2.2 simulations were all run with an ensemble of 5 members. In OMA AMIP mode, for this paper they covered the time period 1974-2014 (with results shown for 1979-2014). Coupled atmosphere-ocean climate model simulations were performed in pre-industrial initial mode following a spin-up of 600 years; results shown are generally 50 year averages. E2.1 results shown for comparison had an ensemble of 10 members (necessitated by its greater internal noise from an amplified ENSO). Most of the comparisons with observations use the complete set of ensemble members, especially for the mean atmosphere and ocean fields.

3.1 Atmospheric Model Climatology

We focus here on the results from the "OMA AMIP" simulation, using interactive atmospheric chemistry and observed sea surface temperatures. This gives the best indication of the atmospheric model capability. Select meteorological comparisons for the two solstice seasons were made for the time period 1979-2014 with observations from both ERA5 and MERRA2 (both will be shown, especially when they have the potential for differing substantially). Comparisons of some of the atmospheric chemistry fields are provided in Paper 2. We will also indicate how the results differ from those of E2.1. We make no formal assessment of the significance of the model/observation differences except to note that we are using 165 years of model simulations of E2.2 (and 350 years for E2.1), the focus is generally on seasons or annual

averages, and differences are only explicitly acknowledged when they arise between the model(s) and the two reanalysis data sets.

The emphasis in this section is on the thermodynamic and dynamic mean fields. Some radiative characteristics of E2.2 as well as several hydrologic and related mean fields will be presented in Section 3.5.1 in conjunction with E2.2AP, and also in the discussion section.

3.1.1. Temperature. The mean temperature structure for the time period indicated, from the five-member ensemble, is shown in Figure 1 for the two solstice seasons for the full altitude range of the model. Starting at the top, in the upper mesosphere we note that the observed temperature range between the summer and winter pole at 0.006 hPa (nearly 83 km) is -71°C (-120°C to -49°C) in December-February (DJF), and -70°C (-121°C to -51°C) in June-August (JJA) (from CIRA (1986) as shown in Fleming et al., 1990).

Model results compare favorably, with gradients of -78°C (-126°C to -58°C) in DJF. and then -74°C° (-125°C to -51°C) in JJA. The circulation producing this difference, consisting of rising air above the summer polar stratosphere and sinking air towards the winter polar stratosphere is generated in the model by planetary wave and gravity wave breaking in the mesosphere, along with a 'near-top' frictional drag associated with instability inserted to minimize the influence of the model top. The contribution from the different terms can be assessed by their influence on reducing the west winds in winter and east winds in summer in the model, both of which lead to a meridional flow from summer to winter pole. Our analysis shows that at ~78 km in the winter hemispheres, during DJF (JJA), the resolved waves contribute 26% (17%); the parameterized gravity waves 38% (52%); and frictional drag 36% (30%). In the summer hemispheres, during DJF (JJA), the resolved waves contribute basically 0% in both hemispheres; the parameterized gravity waves 90% (85%) and the frictional drag 10% (15%) respectively. In the summer hemisphere, then, the parameterized gravity wave breaking is by far the most important, while all three processes contribute in the winter hemisphere (with planetary waves more so in the Northern Hemisphere). Of the different parameterized gravity wave drag mechanisms contributing at this level, in the extratropics those associated with shear are most important, and in the tropics, those due to convection.

In the CIRA (1986) observations, stratopause temperatures near the winter pole (80° lat.) in DJF (JJA) peak at about 56 km with values of about -13°C (-1°C). In the model the peak occurs at a similar altitude with a value of -17°C (-9°C). Near the summer pole (80° lat.), during DJF (JJA), the CIRA observations indicate peak values at ~47 km of +13°C (+9°C), while the model values, which also peak at that height, are +6°C (+13°C). Given both model and real-world variability, and scarcity of observations, no strong conclusion can be drawn but the model may be too cold over the Southern Hemisphere winter pole, and over the Northern Hemisphere summer pole. In the tropics, observed (model) stratopause temperatures are -4°C (-3°C) in DJF; -3°C (-6°C) in March-May; -8°C (-8°C) in JJA and -4°C (-8°C) in September-October. We return to an assessment of this level in the presentation of the semi-annual wind oscillation (Section 3.3.3).

Moving further down into the stratosphere, Figure 2 is a comparison of E2.2 with both MERRA2 and ERA5 temperature reconstructions up to 1 hPa (the location of the model top in both reanalyses may influence their results above that level, i.e., less than two scale heights away). Both reanalyses agree that the model is about 5°C too cold at high latitudes in the stratosphere during Northern Hemisphere winter, and too warm by a slightly greater amount at high latitudes in the stratosphere during Southern Hemisphere winter. During summer at high latitudes the stratosphere is somewhat too cool, more so in the Southern Hemisphere.

For contrast, shown in Figure 3 is a comparison of E2.1 with the two observational data sets for the same years. Double digit anomalies are the rule for the extratropical stratosphere in winter, and for the Southern Hemisphere summer. The cause(s) of the significant improvement between E2.1 and E2.2 will be addressed in the discussion section.

Continuing downward, of particular importance for climate studies is the simulation of the tropical tropopause cold point temperature (CPT), defined here as the coldest temperature in each month in the vicinity of 100 hPa (between 80 and 120 hPa) averaged from 10N to 10S (e.g., Kim and Son, 2012). This temperature is important because the air that enters the stratosphere through the tropical tropopause undergoes a freeze-drying process near the CPT, so the maximum amount of water vapor transported into the stratosphere is limited by the CPT. The results, shown in Figure 4, compared with ERA5 and MERRA2, indicate that E2.2 values are somewhat too warm each month, while E2.1 is somewhat too cold. The seasonality shown is thought to relate to the upwelling associated with the Brewer Dobson circulation, and to tropical and extratropical waves acting right near the tropical tropopause. The colder temperatures in E2.1 are consistent with what will be shown in Paper 2, that the upwelling in that model is too strong. One complicating factor is that E2.1 is considerably warmer than the reanalyses at 100 hPa; its CPT is reached at 80 hPa, and it has the largest decrease in temperature between those pressure heights of any of the models, perhaps indicative of its greater upwelling from 100 hPa to higher levels. A discussion of the stratospheric water vapor in these different runs will be part of the general tracer focus of Paper 2.

The extratropical upper troposphere and lower stratosphere is bounded in the vertical by the stratospheric overworld (440 K isentrope) and the upper troposphere (~5 km below the mean tropopause) (e.g., Gettelman et al., 2011). It couples the stratosphere to the troposphere through chemical constituent transport. The tropopause dynamical definition is viewed in terms of potential vorticity units (PVU), of 2 PVU. At high northern latitudes in E2.2 it varies from about 8 km in winter to a little over 9 km in summer; in upper mid-latitudes from about 9 km to 10 km; in lower mid-latitudes from 10 km to 12.5 km; and in the subtropics from about 13.5 km to 14.5 km. These are all reasonable numbers, regardless of definition; see, for example, the seasonal variation as a function of latitude of the CO vertical gradient in Gettelman et al., (2011, their figure 9). Transports through the tropopause will be discussed in more detail in Paper 2.

3.1.2. Zonal Wind. In the previous section we concluded that near the polar stratopause, E2.2 may be too cold in Southern Hemisphere winter and in Northern Hemisphere summer. While winds respond to temperature gradients as opposed to absolute temperatures, if these

tendencies were to show up in the gradients as well it would imply stronger west winds than observed in the extratropical mesosphere during Southern Hemisphere winter, and weaker east winds in that region during Northern Hemisphere summer. Comparison to the CIRA (1986) winds indicates that for both the model and observations during JJA there are peak west winds in the Southern Hemisphere of ~80 ms^{-1} (~0.8 hPa, 45S), and peak east winds in the Northern Hemisphere of -50 ms^{-1} (~0.08 hPa, 65N). There is obviously much uncertainty and variability associated with this comparison, but it is presented to indicate that the model's simulation of the strong seasonal winds in the mesosphere is in at least rough agreement with observations, implying the extratropical temperature gradients near the stratopause are as well.

For the stratosphere and below, the zonal wind anomalies are shown in Figure 5 for the two solstice seasons compared with ERA5, for E2.2 (upper panels) and E2.1 (lower panels). The anomalies are consistent with those for temperature via the thermal wind relationship, as E2.2 has slightly stronger winds than in ERA5 during Northern Hemisphere winter in the polar lower stratosphere, with the anomaly reaching 10 ms^{-1} above the mid-stratosphere. This is a region/season of large inter-annual variability, so the significance of this difference is uncertain. E2.1 has much larger wind anomalies, especially in the mid and upper stratosphere where the presence of the model top, or the drag utilized to minimize its influence, results in weaker than observed winter westerlies and summer easterlies in both hemispheres.

The seasonal variation of the zonal winds at 10 hPa and 1 hPa are presented in Figure 6 for E2.2, E2.1, ERA5 and MERRA2. E2.2 does a credible job in matching the observed zonal wind change as a function of season at both levels, including the approximate proper time for transition from winter westerlies to summer easterlies, the magnitude of the winter stratospheric west wind jet and its variation with month in both hemispheres, and the timing of the appearance of the maximum summer easterlies. Its main flaws are the Northern Hemisphere summer easterlies do not increase strongly enough with altitude (consistent with the summer pole being too cool), and the upper stratosphere tropical equinoctial westerlies are not sufficiently strong compared with reanalysis data. We return to this latter point in the discussion of the semi-annual oscillation (Section 3.3.3). E2.1 has both winter westerlies and summer easterlies that are too weak.

***3.1.3. Planetary Waves*.** Focusing on Northern Hemisphere winter for planetary wave structure, shown in Figure 7 are the climatological ERA5 wave 1 (left) and wave 2 (right) amplitudes, and the respective differences in the simulations of E2.2 and E2.1. The amplitudes and phases are calculated from the individual monthly geopotential heights and the seasonal average then constructed. E2.2 appears to be up to 15% deficient in stationary wave 1 energy in the mid-stratosphere (its amplitude is about 750m), while it is about 30% too abundant for stationary wave 2 energy at the same height (with an amplitude of 270m). In contrast, the wave 1 deficiencies in the mid-and upper stratosphere are much more severe in E2.1, while wave 2 is now deficient as well poleward of 60°N. Reasons for the differences between the models will be examined in the discussion section.

***3.1.4 Sea level pressure and Geopotential heights*.** The advantage of E2.2 relative to E2.1 in the stratosphere is clearly displayed in the geopotential height figures at higher levels. The sea level pressure field (Figure 8) is not noticeably better in either model, except the low pressure systems over the Southern Ocean are somewhat stronger and more accurate in E2.2. Some studies that suggest that this field is highly sensitive to the gravity wave drag in the lower stratosphere, but despite their differing vertical resolutions, their gravity wave drag in that region does not differ substantially. Both models have a stronger than observed Icelandic Low, and a weaker than observed Aleutian Low. In the Southern Hemisphere, the low pressure systems in the Southern Ocean are too weak, suggesting a more negative Southern Annular mode. This will be discussed in more detail in Section 3.3.4.

As one progresses upward in the atmosphere (Figures 9-11), one can see that E2.2 is far superior, an advantage that grows with altitude, especially in the polar regions. These differences affect ozone hole simulations, age of air, troposphere/stratosphere exchange, etc., topics to be discussed in Paper 2.

***3.1.5. Mean Meridional Circulation Cells*.** Focusing more on the troposphere, shown in Figure 12 is the mass stream function for the two solstice seasons, compared with both ERA5 and MERRA2. The Hadley Cell in model E2.2 is slightly too weak in DJF. and about 15% too strong in JJA. To the extent these errors are real (the observed value is difficult to retrieve), they can affect interhemispheric transport, discussed in Paper 2, which will also include a comparison of the model's stratospheric residual circulation with the observed, as deduced from the Age of Air and tropical upwelling. A discussion of how the Hadley Cell varies between E2.2 and E2.1, and its likely cause is presented in Sections 3.5.1.2 and 3.5.2.

3.2. Ocean Model Climatology

For these fields, we use the coupled model climatology with the G ocean model. In contrast to the OMA AMIP simulations, these results are for the 1870-1900 time period, following a spin-up of 800 years with pre-industrial (1850) forcing; to distinguish from current times, this period will henceforth be referred to as "pre-industrial" (PI). The comparisons with observations can only be approximate, as we have no observations from that time of some of the dynamic ocean fields and variability, and in those cases we must use current values. A few results with the H model will be presented here and in later sections.

***3.2.1. SSTs and Sea Ice Cover*.** The 'pre-industrial" sea surface temperatures and sea ice cover from the coupled model runs are compared with estimates of what they were like from the Hadley Centre reconstructions provided to CMIP6 in Figure 13 for SSTs and Figure 14 for sea ice over. With the G (GISS) ocean model (top row) the SSTs are somewhat too warm in the eastern tropical Pacific and Atlantic Oceans (Figure 13). These problems exist in E2.1 (not shown) except they are even greater, with positive sea surface temperature anomalies of 2°-5°C in the central to eastern Pacific and Atlantic, 2°-4C° in the Labrador Sea and widespread 2°-3°C in the Southern Ocean. The main difference in this regard appears to be that E2.2 is a colder run in general (see Section 3.5.1 and the discussion section 4.3).

Also shown in the figure are the equivalent comparisons for the HYCOM model (bottom row). The warm anomalies in the eastern tropical ocean basins are still present, as are colder regions at high northern latitudes; these problems would thus seem most likely to arise from the atmospheric forcing, as the ocean models are very different. E2.2-H is somewhat warmer than E2.2-G, and so it also has positive anomalies in the Southern Ocean, while the Northern Hemisphere high latitude ocean regions are not as cold.

With both ocean models, E2.2 generally simulates Southern Hemisphere sea ice distributions fairly well (Figure 14) but is excessive compared to this reconstruction in the Northern Hemisphere (which may be an overestimate since the reconstruction values are barely larger than current day values). On a global area basis, the models' sea ice cover is about 1.5% too large, with the greatest deficiency east of Greenland. The results can again be compared with those of E2.1 (not shown), where its warming off the coast of northeast North America and in the Labrador Sea inhibits sea ice in the western North Atlantic. That model too exhibits larger sea ice concentrations in the far northeast North Atlantic, though less so than in E2.2, and shows greater deficiencies in the Southern Ocean.

3.2.2. SSS. The model's PI sea surface annual salinity is compared with modern values in Figure 15. The model salinity is a little small in the tropics, and too large at high latitudes. In association with the higher tropical sea surface temperature shown above, the model's precipitation tends to be somewhat higher than observed in the tropics (discussed in Section 3.5.1.2), which would be consistent with the lower salinity values there. While the two different time periods for the comparison may be an issue, the historical simulations with E2.1 failed to show large differences between pre-industrial and modern times.

The higher salinity in the Indian Ocean may be due to an underestimate of Brahmaputra-Ganges river runoff (by 16%). The Amazon River runoff has increased somewhat from previous versions (both E2.0 and E2.1), to 377 km^3 $month^{-1}$, and while still less than the observed, there is no longer a positive salinity anomaly at the mouth of the river. Considering the positive Arctic anomaly, the model's 5 rivers that flow into it average 75% of the observed.

***3.2.3 Vertical profiles of temperature and salinity*.** We compare the model's annual PI values of temperature and salinity with current day observations in Figure 16. Beginning with temperature, shown are the differences for the Atlantic (top left) and Pacific (top right). The model values in the mixed layer are too cool, especially in the tropics; given that sea surface temperatures in that region are not, this implies a lack of sufficient vertical mixing, or it is an artifact of comparing modeled PI time period with modern observations. Other than that, the model is generally within ±1°C, though tending to be slightly too cool in the Pacific. The salinity profile shows the model is too fresh near the surface, which would be consistent with the precipitation being too large, especially in the tropics. Otherwise it is again in accord with current day observations.

3.2.4. Ocean overturning circulation. The annual average Atlantic and Global ocean stream functions for E2.2 are shown in Figure 17. The North Atlantic Deep Water (AMOC) circulation

peak (Figure 17, top), defined here as occurring anywhere between 20°N-80°N and below 500m, is approximately 29 Sv. In E2.1, the value was about 27 Sv (not shown). These are similar to the values found in other models when searching for maxima in this region, for example CCSM4 (Danabasoglu and Yeager, 2012). Modern day observations from the RAPID array (NCAR staff, 2017) show average values at 26.5°N ranging from around 14 Sv to more than 20 Sv. One can see from the figure that the model's maximum is further poleward of that latitude, and generally deeper than 1000 m.

Antarctic Bottom Water (AABW) simulation in models is a more complex phenomenon. Sinking of water in the Southern extratropics can be associated with two processes: one in which deep water upwells, is blown north by wind stress in an Ekman layer, and sinks at the Antarctic convergence (the so-called Deacon Cell), generally 50-60°S; and a second, deeper cell associated with bottom-water formation, sinking, and entrainment next to the Antarctic continent (close to 70°S), equatorward flow near the bottom, with the southward inflow of circumpolar deep water above (Speer et al., 2000). Models often produce sinking associated with the first or a combination of the two, but in the open waters (open ocean polynias) as opposed to the more commonly observed overturning near the ice shelves. E2.2 produces a circulation with sinking of this nature (Figure 17, bottom), with a value of some 17 Sv (13 Sv in E2.1) south of 30°S. If we restrict the AABW calculation to latitudes south of 60°S, then the model produces about 7 Sv (in E2.1 it is 3 Sv). The actual volume of AABW production is quite uncertain, though it has been estimated to be as high as (roughly) 20 Sv (Frajka-Williams et al., 2011).

One representation of this complex circulation can be seen in the model's annual salinity profile at high southern latitudes (Figure 18). One can see deeper waters with reduced salinity (<34.7 psu) at high southern latitudes, as well as the upwelling of higher salinity Circumpolar Deep Water of North Atlantic origin. This result can be compared with Figure 1 of Speer et al. (2000), which shows similar salinity values for the bottom waters (<34.7 psu) and upwelling waters (34.75 psu) except the bottom water salinity values in the observations can be seen to result from flow downward along the ice shelves, a process not obvious in the model. Lack of both ice shelves and fine enough resolution is the likely cause of this model deficiency, a common model problem that affects AABW production.

***3.2.5. Ocean currents*.** The annual average PI values for the Gulf Stream, Kuroshio Current and Antarctic Circumpolar Current are shown in Figure 19, both for E2.2 and E2.2AP. The modern observed value for the Gulf Stream varies from 30Sv in the Florida Current to 85 Sv at Cape Hatteras (3**5**°N). The model diagnostic chooses the maximum value between 24-38N°, 60-80°W; therefore, its value of ~65 Sv might be a little low compared with modern estimates. The Kuroshio Current varies from 90 Sv in winter to 40 Sv in summer and early autumn, and the value here is in approximate agreement with the annual average. The Antarctic Circumpolar Current at the Drake Passage is observed to be about 173 Sv (Donohue et al., 2016). The model results would therefore appear to be approximately correct. The CMIP5 models had a mean value of 155 Sv varying from 105 to 205 Sv (Farneti et al., 2015). The comparison between E2.2 and

E2.2AP shows that they have similar ocean currents, indicative of similar troposphere and ocean circulations.

***3.2.6. Ocean heat transports*.** Shown in Figure 20 are the annual average ocean heat transports for the global (top) and Atlantic (2nd row), Pacific (3rd row) and Indian (bottom row) ocean basins; results are provided for both E2.2 (left) and E2.1 (right). For E2.1, peak total global values are a little under 2 Pw, in agreement with estimations (e.g., Trenberth and Fasulo, 2017) while peak values in the Atlantic are about 1.35 Pw, again in approximate agreement with observations and high resolution models (e.g., Trenberth and Fasulo, 2017, Grist et al., 2018). The peak values in E2.2 are 0.1 higher globally, and similar in the Atlantic. Both models correctly show the northward transport in the Southern Hemisphere in the Atlantic; the global ocean value is southward there, peaking at about 1 Pw as observed in E2.1, and a little larger in E2.2.

Considering the contribution of the different ocean basins to the overturning circulation heat transports, the Atlantic provides a peak of about 1.5 Pw in each model, slightly higher (by about two standard deviations) than current estimates of 1.2 Pw (Trenberth and Fasulo, 2017); the subtropical overturning circulation in the Pacific contributes between 0.8 (E2.1) and 1 (E2.2) Pw, both within the range of observational and model uncertainty (Hazeleger et al., 2004); and in E2.2, there is a component from the Indian Ocean of 0.7 Pw as well, offsetting a gyre component of equal and opposite magnitude. This last aspect differs considerably from what occurs in E2.1 which has no transport of heat north of the equator by either component; it may be related to the Indian Ocean dipole (Wang et al., 2014), and its difference between the two models may be due to their differing precipitation in the Indian Ocean (see Section 3.5.1.2).

3.3. Variability

Here we discuss the leading atmospheric modes of variability in both the troposphere and stratosphere as well as the leading ocean mode. Results are from either the OMA AMIP or coupled model simulations, depending on the mode of interest.

***3.3.1. QBO*.** As it is important for the sake of predictions of future QBOs that the model be able to generate a reasonable QBO in the coupled model, with its different SSTs and convection, shown in Figure 21 are the monthly results from the PI coupled simulations. As indicated by the text in the periodogram at the bottom of the figure, the peak period is 28 -29 months, in agreement with observations (e.g., https://climatedataguide.ucar.edu/climate-data/qbo-quasi-biennial-oscillation). [The E2.2 OMA AMIP QBO amplitude is similar but the period is about 29.5 months]. The modeled peak to peak amplitude is about 15% less than observed at 30 hPa compared to data from the Free University of Berlin data and MERRA2 re-analysis, while being close to the observed at 60 hPa. For an assessment of the contributions driving the QBO from resolved and parameterized wave fluxes, and advection, see the discussion in Rind et al. (2014) (their Section 3.4 and Figure 10); this analysis from an earlier version of the model is broadly relevant here as well. The E2.2AP Coupled QBO in shown Figure 36, and the E2.2 V2 AMIP NINT QBO, properly tuned, is shown in Figure 37.

3.3.2. Sudden Stratospheric Warmings (SSW). The model's winter variability in the Northern Hemisphere is assessed in two different ways, first with respect to the WMO definition of SSWs by which the zonally averaged zonal wind reverses at 10 hPa and 60N (Butler et al. (2015), and then with regard to the standard deviation of the zonal wind at 50 hPa. The difference between the two is that the model may have sufficient variability but if its mean west wind is too strong (as shown in Figure 5), the variations may not actually bring about an east wind when zonally averaged.

The OMA AMIP and coupled versions of E2.2 (Figure 22, top, daily data) exhibit SSWs that occur at frequencies of 5.2 events/decade and 4.3 events/decade, respectively, both of which fall within the observational uncertainty (spanning 4-8 events/decade) albeit at the lower end (Charlton-Perez et al. (2013) – See their Figure 3a). Consistent with thermal wind balance, all major warmings (i.e. those in which the winds actually reverse) are associated with dramatic increases in temperatures ranging between 20 K -30 K in both models (Figure 16 bottom). Overall, this is a significant improvement relative to previous versions of ModelE, which produced warmings at only a quarter of the observed frequency (Hansen et al. (2007)). Note that while only one of the OMA AMIP ensemble members is shown we have confirmed that similar SSW frequencies occur for the other four members, with all members exhibiting frequencies ranging between 5-6 events/decade.

An analysis of the variance of the northern zonal winds reveals a similar conclusion. Specifically, the climatological mean values of the total variance of the de-seasonalized daily (November-December-January-February-March) zonal mean zonal winds at 60°N and 50 hPa are 58 ms^{-1} and 47 ms^{-1} for the OMA AMIP and coupled models, respectively (with the former value applying to the first ensemble member). Compared to Charlton-Perez et al. (2013) (their Figure 3b) these values are in good agreement with reanalyses and with the upper range of high-top models participating in CMIP5. Thus, while both OMA AMIP and coupled versions of E2.2 exhibit a positive bias in their zonal winds over northern high latitudes this bias does not significantly degrade the variability in either model, neither with respect to the number of warmings nor variance of the zonal wind.

We can also compare the monthly variation of the SSW occurrence with that in the MERRA2 data, following the procedure of Gettelman et al. (2019). The results are shown in Figure 23; final warmings are not included in this analysis. The Coupled model has the same or fewer events than the OMA AMIP version in each month. In this short sample E2.2 models seem to have more warming events very early and late in the season than in the observations, but with the inherent variability of such occurrences, no strong conclusion can be drawn.

Similarly, while the slightly lower frequency in the coupled model may be related to an overall increase in the zonal mean winds (contrast left and right panels in Figure 22) it is worth noting the large temporal variability of sudden warming occurrence within that run, wherein warmings occur much more (less) frequently during the first (second) half of the integration. Similar variability in warming frequency occurs over the course of the OMA AMIP run, most notably during the 1990s where there is a dearth of warmings as has been noted in previous studies

(e.g. Charlton and Polvani (2005)). Both examples indicate that caution should be taken when inferring statistics of variability over relatively short integrations, especially when such variability can be simply internal (i.e. unforced).

***3.3.3. SAO*.** The semi-annual wind in the tropical upper stratosphere is shown in Figure 24 for various E2.2 models, E2.1 and reanalysis data; we use 2 hPa to insure it is sufficiently away from the model top in the reanalysis data and E2.1. The different E2.2 models are shown to illustrate that the results are a common feature of all of them, regardless of their physics package (E2.2AP), gravity wave drag efficiency (E2.2 V2), or sea surface temperature patterns (E2.2 Coupled). Overall the model reproduces many features of the observed oscillation, with easterly winds during the solstices and westerly winds during the equinoxes. In addition, the winds are strongest easterly during December – February, as observed, an indication of an annual cycle effect. The model fit is obviously better in comparison to ERA5. E2.1 has only a very weak semi-annual oscillation, with no west winds at any time of the year in the upper stratosphere (it has no parameterized convective gravity wave drag).

Analysis was done of the tendency driving the tropical zonal winds right at the equator, with results for the 1 and 2 hPa levels shown in Figure 25 for E2.2 OMA AMIP. Looking first at 2 hPa, the total change was positive leading into the equinoxes, and negative leading into the solstices, more so in November than in May, consistent with the magnitude of the resulting winds in the different seasons. The individual components producing the total change are shown [they do not exactly add up to the total as they are calculated on different grids, with the total change and TEM components from -4° to 4°, and the drags and diffusion from -3° to 3°]. Advection by the residual circulation was generally associated with the same tendency as the total change leading into and during the equinoxes, i.e., providing a westerly acceleration. It also provides weak easterly acceleration leading into the summer (June – August) solstice; only leading into and during the winter solstice was its acceleration opposing the total change. The resolved waves generally forced an acceleration in opposition to the residual circulation, an effect that is commonly expected and observed in the extratropics; however in May it helps provide the easterly acceleration for the summer solstice. The parameterized gravity wave drag forced a westerly acceleration during the equinoxes, especially March, and forced an easterly acceleration throughout the region during the solstices, especially December. Momentum diffusion arising from gravity wave breaking, and a vertical diffusion associated with parameterized inertial Instability (e.g., Knox and Harvey, 2005) provided small additional forcing.

At higher levels (1 hPa) the results shown in Figure 25 indicate that the residual circulation is acting in a generally similar manner as at 2 hPa except it provides more of the easterly acceleration during the summer solstice, while the resolved waves (EP flux divergence) continue to oppose it. The big change is that the gravity wave drag more clearly acts to decelerate the prevailing zonal wind, providing westerly accelerations during the solstices, and easterly during the equinoxes. This is phase shifted by 2 to 3 months compared to what they were doing at lower levels, and may be the result of selective filtering of the upward propagating (parameterized) waves by winds below 1 hPa. Another difference is that the

parameterized vertical diffusion associated with inertial instability provides somewhat more easterly acceleration during the October through May period.

These results for the residual circulation and gravity wave drag tend to be in general agreement with those seen in various Middle Atmosphere models (Smith et al. 2019), the primary difference being that most models find a strong easterly acceleration by the residual circulation in the winter solstice when the zonal winds are strongest easterly. They are strongest easterly at that time in E2.2 as well, but it is instead the result of wave forcing (both resolved and parameterized) at 2 hPa, and by resolved waves and inertial instability at 1 hPa. As noted by Knox and Harvey (2005), inertial instability, perhaps associated with Rossby wave breaking, is most active during the winter solstice.

Depicted in Figure 6 were the zonal winds as a function of month at 1 hPa, and it was noted that in E2.2 the equinoctial west winds were not as strong as in the reanalyses. At that level, too, MERRA2 west winds at the equator during the equinoxes were stronger than in ERA5. Smith et al. (2019) analyzed SABER data to determine tropical zonal winds, and the westerlies in their results can be seen to be closer to ERA5 (i.e., on the order of 20 ms^{-1} at 1 hPa); nevertheless, at that level they are still stronger than in E2.2. At 1 hPa in the model only the residual circulation is acting to accelerate the westerlies during the equinoxes, all the other components are acting against it, so either the residual circulation is too weak, or the other components too strong. Assessing a variety of models, Smith et al. (2019) found that most had west winds that were less than in those observations.

3.3.4. Northern and Southern Annular Modes (NAM, SAM). Considering now tropospheric modes of variability, we utilize an index related to the first annular mode in each hemisphere for the OMA AMIP runs by comparing the sea level pressure, the mid-troposphere and the lower stratosphere geopotential heights, in both the OMA AMIP models and observations (Figure 26). For the Northern Hemisphere we use latitudes (30-50N) minus (60-80N) as described in Rind et al. (2005), and for the Southern Hemisphere latitude 40°S minus 65°S (Bracegriddle et al., 2008).

At the surface and in the mid-troposphere, the mean value of the index in the Southern Hemisphere winter is biased low, as the circum-Antarctic, low pressure systems are somewhat weaker than observed (see Figure 8). In the stratosphere in the Northern Hemisphere it is biased high, consistent with the somewhat stronger than observed lower stratospheric westerlies (see Figure 5).

As can be seen in Figure 26 the model has the same order of variations (i.e., high and low indices) as in observations; this was also verified by looking at storm tracks for different winters. More precisely, the standard deviation of these indices in the different models and observations are provided in the figure. E2.2AP has slightly higher variability than E2.2 and the observations in the tropophere, while both models show similar variability to observations in the lower stratosphere. The results indicate the model is capable of capturing the magnitude of the interannual winter changes in these modes, including those in the lower stratosphere, as

was also indicated by the model's frequency of stratospheric warmings and zonal wind variability (Section 3.3.2).

We also calculated the temporal correlation coefficients between the lower stratosphere and the troposphere for these indices, with the results shown below:

	NH		SH	
Model	50 hPa/ 500 hPa	50 hPa/SLP	50 hPa/ 500 hPa	50 hPa/SLP
E2.2	.5541	.6663	.5132	.5226
E2.2AP	.6707	.6274	.5540	.5526
ERA5	.4890	.6038	.7248	.6924
MERRA2	.5409	.6227	.7184	.6559

All of these results are highly significant, assuming the winters of the 36 years are independent. In the Northern Hemisphere, variations in the lower stratosphere are connected to those at lower levels, with zero time lag, in a manner that is similar to or somewhat higher than in the observations. Many papers have related changes in the NAM phase in the lower stratosphere to changes in sea level pressure patterns, as noted in the Introduction, and a discussion of the various potential mechanisms is included in Rind et al. (2005). In the Southern Hemisphere, the model has somewhat less 'connectivity' than observed, which may be related to the somewhat weaker than observed low pressure systems over the Southern Ocean.

We have also investigated the relationship between the model and observed indices over these 36 years. In the Northern Hemisphere, both models show a significant relationship with the observed values in the troposphere, although the percentage of the variance explained is only 10-15%. For this OMA AMIP run, the observed sea surface temperatures are prescribed, and the atmospheric forcing calculated or prescribed, so some relationship should be expected. If these models were perfect, one could conclude that the other 85-90% is the result of natural variability. In the Northern Hemisphere stratosphere, the relationship between model and observations was not significant; either the SSTs and climate forcing are less effective at that level relative to natural variability, or the model is less accurate in its response to them. In the Southern Hemisphere the models SAM variations were not significantly related to the observed in either the troposphere or stratosphere.

3.3.5. Storm tracks. The variation in AO and SAM are related to model storm tracks, although storm tracks also relate to regions of cyclogenesis and subtropical jet stream variability. Shown in Figure 27 is the Northern Hemisphere storm track density from the OMA AMIP runs of E2.1 and E2.2 for the time period 1979-2014, along with MERRA2 observations for years 1980-2018, produced as in Bauer et al. (2016). (Note MERRA2 seems to have spurious storm densities associated with storm tracks over Greenland). Both models simulate approximate storm track patterns. E2.2 has greater (and more accurate) storm track densities in the western and central North Atlantic and northern North Pacific than E2.1, while it is deficient in storms crossing the Pacific south of 40°N. E2.1 fails to produce the 'Colorado Low', forming in the southwest United States and moving northeast to the Great Lakes; it has way too few storms over land in North America.

A further assessment of the model storms (in both E2.1 and E2.2) is that they tend to move too slowly, are deficient in both weak storms and very strong ones, and do not exhibit secondary cyclogenesis; most of these deficiencies likely relate to the models' horizontal resolution.

***3.3.6. Madden-Julian Oscillation and equatorial waves*.** The equatorial Indian Ocean, Maritime Continent, and the western Pacific Ocean have extensive deep convection, which varies from diurnal to intraseasonal time scales. The dominant mode of intraseasonal variability is the Madden-Julian oscillation (MJO), with cloud and precipitation pulses that re-occur on a time scale of 30–60 days [Madden and Julian, 1972, 1994; http://www.bom.gov.au/climate/mjo/]. It is characterized by convection spanning thousands of kilometers, originating in the Indian Ocean and propagating eastward to the Maritime Continent, largely confined to the Indo-Pacific warm pool with sea surface temperatures (SSTs) above 28°C. Numerous methods have been devised to assess the MJO; the one we use here involves a wavenumber-frequency spectral analysis of sub-daily precipitation performed for all longitudes in the domain 30°S–30°N (Wheeler and Kalidis, 1998).

These "Wheeler-Kalidis" diagrams for twenty years of E2.1 and E2.2AP are provided in Figure 28. [E2.2AP is used here, because as will be shown, it has a better (reduced) precipitation field in the region of interest]. As can be seen, the MJO power is present in both models (wave numbers 1-3, period 30−60 days), primarily in the symmetric component about the equator, with somewhat more power in E2.1.

Also shown in the diagram is the power associated with Kelvin waves, equatorial Rossby waves (ER), mixed Rossby-gravity waves (MRG), and eastward inertia-gravity waves (EIG). [Westward inertia-gravity waves, with one to two day periods, are outside the frame of the picture]. As can be seen all of these waves have somewhat similar energy in both models [with differences varying from one 20 year data set to another]. Kelvin waves and MRG are key components in helping generate the QBO, in observations and in the model (Rind et al., 2014).

The results can be compared with GPCP observations in Kim et al. (2012) (their Figure 3a). Using this data set, it would appear that the MJO in E2.2 is about 25% deficient in power, while in E2.1, it is only slightly deficient. The Kelvin wave power is slightly less than observed for both models.

3.3.7. El Nino Southern Oscillation (ENSO). The variability associated with the ENSO cycle is displayed in Figure 29 (left) with the G ocean model, and in Figure 29 (right) with the H model. We use the Niño 3.4 index which involves a 5-month running mean of temperature anomalies in the region (5N-5S, 170W-120W). El Niño or La Niña events are defined when the Niño 3.4 SSTs exceed +/- 0.4C for a period of six months or more. This is the operational definition used by NOAA.

The associated model standard deviation (+/-) is shown by the green lines in each figure, and the observed standard deviation is given in red. The E2.2 values are somewhat low, about 75%

of the observed, while it is of approximately correct amplitude with E2.2AP. In the bottom row one can see the results from E2.1, where the amplitude is some 30% too large with the G ocean, and the oscillation too regular, but appropriate amplitude with the H ocean. We return to these figures and discussion with the sensitivity experiments (Section 3.5.1.3) and in the discussion section (Section 4.3).

3.4. Climate Change Experiments

Given that this topic would really require a separate paper, we limit the presentation here to the most obvious results of the following experiments: instantaneous $2xCO_2$ and $4xCO_2$ forcing. We show results for both E2.2-G and E2.2-H. In addition we compare them to the corresponding simulations from E2.1. The $1xCO_2$ control run is by CMIP6 protocol the preindustrial (PI) run.

3.4.1. Temperature response to increased CO_2. The Northern Hemisphere, Southern Hemisphere and Global mean temperature as a function of time following an instantaneous doubling and quadrupling of CO_2 is shown in Figure 30. Solid lines refer to the $2xCO_2$ response and the dashed lines to $4xCO_2$, for E2.2 (top) and E2.1 (bottom). Results generated with the G ocean model are given on the left, and with the H model on the right.

As shown, the E2.2 runs have not gone sufficiently long to reach equilibrium (which will require a lot more than the CMIP6 requested 150 years, or the 200 years shown here) but we can estimate the equilibrium climate sensitivity (ECS) for doubled CO_2 by (1) using a "Q flux model mixed layer model" (where ocean dynamics is not allowed to change); (2) by the 'Gregory' method (Gregory et al., 2004) that relates the radiative response of the system to a change in temperature; or (3) by running the models for thousands of years. First with the Q-flux procedure, Model E2.2 has an ECS of 2.8°C for doubled CO_2, while the E2.2AP is 2.7°C; both are slightly less than the 3°C for E2.1 in NINT mode, or 2.9°C in OMA mode (Kelley et al., 2019, submitted). Note these numbers apply to both G and H models as the calculation does not involve a dynamic ocean, and is simply an atmospheric property.

With the Gregory method, the various E2.1 configurations produce an 'effective' climate sensitivity of 3.5-3.8°C, while the value for E2.2 is 3.5. Finally, when run for thousands of years in Coupled NINT mode, E2.1 produces a warming of 3.5°C. The magnitude of warming for both E2.2 and E2.1 is close to the middle of what models are so far producing for CMIP6, whose effective climate sensitivity for $2x\ CO_2$ (with the Gregory method) ranges from 1.8°C to 5.6°C (Zelinka et al., 2020), estimated by halving the warming of the actual 4xCO2 model simulations. As this procedure indicates, it is assumed $4x\ CO_2$ warming is twice that of the doubled CO_2 value, which seems more or less accurate when models are actually run long enough to see the result. [CMIP6 does not require running a $2xCO_2$ experiment]. Therefore, the GISS models' sensitivity is also close to the middle of the $4xCO_2$ CMIP6 model results.

Of interest is the hemispheric asymmetry in the modeled warming. In E2.2-G the Northern Hemisphere warms more for both doubled and quadrupled CO_2. This asymmetry effect only emerges after 100 years in the E2.1-G $2xCO_2$ run, and goes away after 50 years in its $4xCO_2$

simulation. The explanation for this response is given below. In the E2.2-H runs (Figure 30, right); the hemispheric asymmetry is roughly similar to that seen with E2.2-G, although the effect increases with greater CO_2 forcing. The asymmetry is less throughout with E2.1-H, decreasing with time in its 4xCO_2 run in contrast to the sharp reduction seen with E2.1-G.

3.4.2. Atlantic Meridional Overturning Circulation (AMOC) response. The explanation for the differences in hemispheric asymmetry of temperature change can be found in the response of the AMOC circulation in the warming climate, as shown in Figure 31. Given is the maximum value for the stream function, defined as >500m depth in the region 20-80N (although all occur at ≥40N). For the 2xCO_2 runs (top), the AMOC weakens by 50% in E2.1-G, but subsequently recovers after about 100 years (the time it took for there to be a noticeable hemispheric difference in warming to emerge). In E2.1-H it weakened even more and did not recover within the time frame of the simulation. The weakening was much less, and the recovery much quicker in E2.2, with both ocean models. Comparing these results to the global temperature changes, In both models the hemispheric asymmetry is weakened when the AMOC weakened.

This effect is even more obvious In the 4xCO_2 runs (bottom). In E2.1, the AMOC collapses with both the G and H models; after only the first decade it was already declining more rapidly than in E2.2. This collapse, and associated reduction in northward heat transport by the overturning circulation induces a relative cooling in the Northern Hemisphere in E2.1, bringing the warming of the two hemispheres closer together. In both E2.2-G and E2.2-H while the AMOC weakens, it stabilizes after the first decade, and the warming asymmetry is maintained. The additional model results shown in this figure will be discussed in Sections 3.5.2, where E2.2RE is presented, and in Section 4.3 which introduces E2.1NE. All results in Figure 31 are five year running means.

To compare these responses to that of other models for 2xCO_2, the AMOC weakening and then recovery, or then stabilization at the weaker value, are both modes that are seen in CMIP5 runs. For example, in the RCP4.5 simulations, which reach doubled carbon dioxide values about 2080, these two modes of response can be seen in simulations that extend out to 2300 (Cheng et al., 2013; their Figure 1a). In no case did the circulation completely collapse, and that was true even for the RCP8.5 simulations (which reach doubled carbon dioxide values around 2050 and 950 ppm by 2100).

For 4xCO_2 (about 1100 ppm), an assessment of the CMIP5 and CMIP6 multi-model response of the Atlantic Meridional overturning circulation in the abrupt 4xCO2 experiment was performed by Mitevski (personal communication). All models showed reductions from the control run value, with one subset producing results more or less like E2.2, and the other a collapse like in E2.1 (although none showed a reduction quite as fast or as complete). This is not directly a function of the magnitude of the warming, at least in the GISS models; the warming in E2.1 was only slightly greater than in E2.2.

Given that both ocean models are responding the same way, it is likely that the responses are the result of atmospheric forcing. While a detailed analysis of the different responses is beyond

the scope of this paper, a key element relates to the deep water in the Labrador Sea and regions just to the east in the respective models. E2.1, with warmer than observed SSTs in that region (see section 3.2.1), provides dense water via excessive salinity. Apparently, this proves extremely susceptible to the increased warmth and precipitation in the higher CO_2 climate. E2.2, being a colder run, has colder air coming off the continent, and the contrast with ocean temperatures induces strong evaporation, resulting in a locally-derived increase in salinity; combined with the colder water in that control run, this seems to make the production more stable. A full analysis of this difference will be presented elsewhere.

3.5. Sensitivity Experiments

3.5.1. E2.2AP (Altered physics). As noted in Section 2.2, parallel to the development of the standard version of the model, we also developed an 'altered physics' version; its initial goal was to have its physics associated with convection more closely resemble that in E2.1 (therefore to minimize the differences between the models), although the results ended up quite different. It has the same overall structure as E2.2, vertical layering and resolution, location of model top, horizontal resolution, but it differs in how the following processes are parameterized, often by making them less dependent on vertical resolution: detrainment, conditional instability, repartitioning of precipitation into lofted and detrained fractions, evaporating precipitation, downdrafts and updrafts. It had to be tuned differently than either E2.1 or the standard E2.2 model, for both radiative balance and gravity wave drag, but with the appropriate effort, it produced a climate that was in some ways better, and in some ways worse than E2.2. While a full depiction of the various fields produced by this model version would take up too much space, some representative fields are shown below, as well as in several of the previous figures. Many of the fields not shown are very similar to the output from E2.2. [E2.2AP has been submitted to the CMIP6 archive, run with both the G and H ocean models].

E2.2 and E2.2AP have similar climate responses to increasing CO_2 (see Section 3.4.1). However, they do differ in their response to SST and climate forcing on a seasonal time scale. Investigation of the relationship between their NAM index values in the Northern Hemisphere in the OMA AMIP runs shows that their response is not significantly correlated in either the troposphere or stratosphere; each is better correlated with observations than they are with each other (Figure 18). In the Southern Hemisphere troposphere the reverse is true, and they are significantly correlated, although with only 10% of the variance explained. Reasons for their differences will be displayed below.

3.5.1.1. Convective mass flux and humidity. Shown in Figure 32 are comparisons of the convective mass flux between E2.2, E2.2AP and E2.1 in the OMA models. Compared with E2.2, E2.2AP has smaller mass flux at altitudes below 700 hPa and larger values above 600 hPa. E2.1 has greater mass flux than either model at altitudes between 850 and 500 hPa. At altitudes above the 500 hPa level, E2.2 has the smallest values.

Associated with these fluxes are differences in specific humidity in the atmosphere, shown in Figure 33, with anomalies given compared to both observational data sets (because specific humidity is affected by the convection scheme utilized in the reanalysis model). E2.2AP with the smallest convective mass flux in the mid-troposphere has the best moisture distribution, while E2.1 with the largest mass flux there has the worst.

3.5.1.2. Precipitation and Hadley Circulation. Shown in Figure 34 is the precipitation field compared with GPCP, for E2.2, E2.2AP, and E2.1. A major difference can be seen in the western tropical Pacific, especially in June-August, where E2.2 AP has the best simulation while E2.1 is the worst (too extreme). This feature has consequences for the Hadley Circulation in that season; it was shown in Figure 12 that E2.2 has a stronger than observed value. E2.2AP, with reduced (and more accurate) precipitation is much more in line with observations. We return to this topic in Section 3.5.2. Note also that not only does E2.1 have more precipitation in the western tropical precipitation than E2.2AP, but it also has more power in the MJO (see Figure 28).

3.5.1.3. Coupled model response. The results shown above for convective mass flux and precipitation utilized specified SSTs, and highlight the differences in the atmospheric models response to them. The situation becomes more complicated when considering the coupled model. The SST anomalies, featuring warmer SSTs than observed in the equatorial eastern Pacific and Atlantic oceans, lead to spurious convection there, which alters the latitudinal distribution of convective fluxes; instead of a primary peak south of the equator, associated with convection in the western Pacific, there is now a peak of similar magnitude north of the equator. This changes the precipitation field in ways that are specific to each atmosphere/ocean model combination.

The net effect can be seen when considering again the ENSO responses shown in Figure 29. E2.2 AP also had the most accurate ENSO amplitude in the G model while E2.1 had the worst (Figure 29 left). With the H model (Figure 29 right), E2.1 no longer overestimates ENSO amplitudes, and the other models show somewhat weaker ENSOs, as the SST and convection fields are now different. This illustrates the obvious fact that coupled atmosphere-ocean phenomena depend on both models, and atmospheric model development focused on one configuration may not produce optimum results in the other.

3.5.1.4. Middle Atmosphere. E2.2AP does not have as good a Middle Atmosphere simulation as E2.2, especially in Northern Hemisphere winter, although it is still much better than that of E2.1. As an indication of that, shown in Figure 35 from its OMA AMIP simulation are the zonal wind and stationary wave #1 DJF anomalies compared with ERA5; these can be contrasted with the results shown in Figures 5 and 7, which indicates that E2.2 has reduced errors in both of these fields. The difference is associated with the increased penetrating convection in E2.2AP (see Figure 32, left hand panel), which has the effect of altering the temperature and wind field in the upper troposphere, affecting planetary wave and gravity wave propagation. This will be discussed further in Section 4.4.

Shown in Figure 36 is the QBO from the coupled model version of E2.2AP. The period is one month shorter and its amplitude slightly greater than the corresponding results for E2.2 (Figure 21). The tropical west wind component in particular is somewhat greater, while the QBO easterlies are a little smaller. This is a bit more in line with observations (see the "tropical wind" figure 4, in Rind et al. 2014), and the differences are consistent with a more westerly prevailing background wind in this model. In the OMA AMIP version of this model, the period is about 2.5 months longer than for E2.2, with a similar amplitude.

Other aspects of the Middle Atmosphere or adjoining regions have been shown in previous figures. The tropical tropopause Cold Point Temperature (Figure 4) is slightly colder than that for E2.2, and thus in better agreement with the observations. The SAO is similar to that of E2.2 (Figure 24). The NAM and SAM variations at various levels have similar magnitudes and correlations with the stratosphere (Figure 26). A discussion of the Age of Air in these two models will be given in Paper 2.

***3.5.1.5. Radiative and related properties*.** In addition, a comparison was done of various radiative and related characteristics of the atmosphere in E2.2AP with those of E2.2 and E2.1. In this subsection we examine the OMA (AMIP) versions of the models run for the 1979-2014 time period. As such they utilized observed sea surface temperatures, which means the results are indicative of the properties of the atmospheric model, rather than the coupled model. However, this mode is important because it is the way that the pre-industrial control runs are initially put in radiative balance (including the cloud cover tuning parameters), which then influences the subsequent coupled model simulations. Coupled model results are provided in the discussion, Section 4.3.

The different model values are shown in Table 2, which provides the global average annual value of the field, and an assessment of which model has the 'best' or 'worst' root-mean-square error based on the gridded distribution of the field; if none is clearly superior, or observations not available, no result is shown. The presentation shown here is modified from Table 1 of Kelley et al. (2019, submitted).

To summarize these results, E2.2 has the best high cloud cover, planetary albedo and shortwave absorbed at the surface; E2.2AP has the best precipitation but the worst cloud cover, planetary albedo and the coldest air temperature, hence the worst net radiation balance; E2.1 has the warmest air temperature, best low cloud cover and total net radiation imbalance, but its atmosphere is too moist and it has the worst precipitation distribution. The strengths and weakness are related in an obvious fashion, e.g., E2.2AP having the largest albedo can force the coldest climate, and E2.1 having the most water vapor is associated with the warmest temperatures. Interestingly, among the models, the 'veracity' of the different moisture variables simulated– atmospheric moisture, clouds, precipitation - do not correlate well with one another. The greenhouse effect (longwave energy emitted from the surface minus that leaving the top of the atmosphere) is greatest in E2.2AP despite its reduced atmospheric water vapor because of its greater high cloud cover, and smallest in E2.2 due to reduced values of both.

Also, the differences between E2.2 and E2.1 in some of their atmospheric model 'radiative' properties (albedo, clouds, moisture) do not relate in any clear fashion to the actual climate change those models calculate, as can be seen by comparing these results to those in Section 3.4. Despite their differences in control run water vapor, cloud cover, planetary albedo, etc., E2.2 and E2.2AP have similar climate sensitivities. It has been thought that colder models may have a higher climate sensitivity since there is more sea ice to 'lose' as climate warms, yet it is the warmest model (E2.1) that has the (slightly) higher sensitivity (see also the discussion Section 4.3). The large climate forcing associated with substantially larger amounts of carbon dioxide seems to result in similar responses in these models when mediated through ocean-atmosphere interaction. Our choice of E2.2 as the primary version is because it best represents the middle atmosphere climatology.

3.5.2.Rainfall evaporation above convective cloud base enabled (E2.2RE). Major differences in the troposphere, stratosphere and climate change simulations result from the choices involved in this parameterization, a process that was allowed in E2.1 but not in E2.2. Allowing this evaporation is a component in increasing the moisture in the atmosphere to excessive values in E2.1 (Figure 33), as well as increasing the low level moist convective mass fluxes (Figure 32). As noted, E2.1 also has the highest (worst) peak value of the June-August Hadley Cell associated with the most extreme rainfall in the western tropical Pacific.

The change in this parameterization from what occurred in E2.0 was one of several model changes introduced for the sake of improving the MJO for E2.1 (Del Genio et al., 2012; 2015). Kim et al. (2011) found that models with strong intraseasonal variability (of which the MJO is the dominant mode) simulate excessive rainfall over the South Asian summer monsoon and the northwestern Pacific monsoon regions during boreal summer, associated with deep convection. Here the excessive rainfall in E2.1 relative to E2.2 occurred in the western Pacific/Indian Ocean region in that same season.

Of course, E2.1 has numerous other changes, so to make a more direct comparison of the effect of this evaporation choice, we turned the evaporation back on in E2.2RE, put it in radiation balance, and ran it for 30 years (in AMIP NINT mode for pre-industrial conditions). Focusing on JJA, we can make the following comparisons:

June-August results	E2.2AP	E2.2	E2.1	E2.2RE
Convective mass flux (10^{-5} mb/s) (global value through 860 hPa)	71	115	134	203
Atmospheric moisture (mm)	24.7	26.3	27.6	28.9
Peak Hadley Cell (10^9 kg/s).	233	256	263	274

The observed Hadley Cell peak value for this season is 230 x 10^9 kgs^{-1} from ERA5 or 240 x 10^9 kgs^{-1} from MERRA2. Allowing the rainfall evaporation pushed E2.2 to be even more extreme than E2.1 in all these categories; its rainfall also increased in the tropical central and western

Pacific/Indian Ocean region. The vertical mass flux through the 300 hPa level also went up, by 33%, as deep convection was amplified. In a coupled model simulation with the G ocean model, the amplitude of ENSO events, previously 75% of observed values, now increased to equaling the observed. And when forced with $4xCO_2$, the AMOC in E2.2RE effectively disappeared, as it did in E2.1 (Figure 31).

***3.5.3. AMIP NINT Version 2*.** If both the AMIP NINT and coupled models are being submitted to CMIP6, the protocol is that they must use the same code, which includes the tunings for gravity wave drag. Given that the coupled model does not reproduce the observed sea surface temperatures completely accurately (see Figure 13), its convection and dynamics patterns differ from those in the AMIP NINT model and so require different non-orographic drag parameters, specifically the relationship between the convective mass flux and the convective gravity wave momentum flux. This in particular affects the QBO (and Age of Air); the tuning is primarily meant to provide the proper period, but it then also affects the amplitude. In this sensitivity experiment we show in Figure 37 the QBO results from a 'version 2' that, instead, uses convective gravity wave drag tuning appropriate for that model and thus in violation of CMIP6 protocol (hence it will not be provided to the CMIP6 archive). However, this model will be useful for future experiments that don't require coupling to either atmospheric chemistry or the ocean. As can be seen in the figure, the AMIP NINT (Version 2) model can produce a QBO that is in good agreement with observations (proper period and amplitude close to that observed). The results for both power and period are similar to those shown for the coupled model of E2.2 (Figure 21), and also for that of E2.2AP (Figure 36). Simulation of other stratospheric fields is also good. The SAO from this model is shown in Figure 24, and similar to some of the other E2.2 models, at 2 hPa it is a fairly close match to the ERA5 results.

4 Discussion

This manuscript has as its primary goal to introduce the new GISS 'high-top' climate model E2.2, some of whose output is being contributed to CMIP6. As discussed in Section 2.2., in the process of its formulation certain choices were made that differentiate it from E2.1, whose output has also been contributed. Here we discuss the effects of some of those choices, as well as some general topics concerned with modeling the Middle Atmosphere in a climate model.

4.1. Location of the model top

There have been various papers examining the influence of including a well-resolved stratosphere on surface climate (e.g., Sassi et al., 2010; Hardiman et al., 2010, Dall'Amico et al., 2010, and Hardiman et al., 2012 with the Hadley Centre Global Environmental Model - HADGEM). Results suggest that stratospheric variability, e.g., SSWs, QBOs, ozone variability, can affect the troposphere, and low top models with insufficient resolution cannot generate such stratospheric variations realistically. Also, certain elements of tropospheric variability, such as ENSOs and NAO variations, that interact with the stratosphere can likewise have their tropospheric expression modified if the top is too low. In addition to its influence on variability, a recent paper by Kawatani et al. (2019) showed that a well resolved Middle Atmosphere provides more realistic tropospheric mean circulation, e.g., better sea level pressure patterns; here well-resolved means inclusion of the mesosphere along with increased vertical resolution.

In a more general sense, planetary waves propagating up from the troposphere may reflect off an artificial model top in the stratosphere and thus alter the troposphere unrealistically. And models with full stratospheres may employ different gravity wave drag parameterizations or altered radiation calculations that in themselves can alter tropospheric climate (Sigmond et al., 2008; Hardiman et al., 2010). These are all reasons why lifting the top of the model may produce somewhat different surface climate responses.

Those model studies that compared a 'high top' with a 'low top' have made extreme changes. In the case of the HADGEM model, it was raised from 38 km to 84 km. Similarly, in April, 2015, ECMWF raised the top of their medium range forecast model from 5 hPa (~35 km) to 0.01 hPa (~80km) (https://confluence.ecmwf.int/display/UDOC/MARS+content) and this is now the standard model top for their various model configurations (https://www.ecmwf.int/en/forecasts/documentation-and-support). In the case of E2.2 vs. E2.1, the change is not as extreme: E2.1 has a top at 0.1 hPa (65km) and E2.2 at 0.002 hPa (89km). We find that the major differences are not in the mean surface climate but in the stratosphere itself, with effects that may then influence surface climate changes as the previous references suggest. The large differences highlighted in Figures 5 and 9-11 especially result primarily from the raising of the model top. Both E2.1 and E2.2 have a stability-dependent drag in the layers near the top to minimize its influence (above 1 hPa in E2.1, above 0.02 hPa in E2.2), functionally similar to the turbulent drag near the surface. It is possible that altering this drag might have improved the simulation in E2.1 somewhat; different approaches are used in different models, such as Rayleigh friction, diffusion of different orders, etc., (Jablonowski and Williamson, 2011), with the results most important at levels within two scale heights of the top. From the climate forcing perspective, understanding the impact of solar forcing or increased CO_2 on the stratopause region of ozone production is more feasible if the top of the model is far removed (more than 2 scale heights above the stratopause).

4.2. Increased vertical resolution

The increased vertical resolution in the lower stratosphere allowed the model to produce its own QBO, as discussed in Rind et al. (2014). Increased vertical resolution in the upper stratosphere allows for a more delineated vertical temperature gradient around the stratopause, where temperature and ozone production are closely related. Increased vertical resolution in the UTLS can improve stratosphere - troposphere exchange by more properly defining the tropopause and tropopause folding events.

In E2.2, however, we also increased the vertical resolution in the troposphere. Using the Rossby ratio between vertical and horizontal scales in quasi-geostrophic flow to estimate the required vertical resolution in the extratropics (Lindzen and Fox-Rabinowitz, 1989) suggests that the model at 2x2.5° resolution would need a tropospheric vertical resolution of around 1 km. Both E2.1 and E2.2 fulfill that requirement (Table 1), and indeed, comparison shows that there is no significant difference between them in overall baroclinic generation of eddy energy or tropospheric eddy energy itself.

In the tropics, considering the dispersion relationship for internal gravity waves (Lindzen and Fox-Rabinowitz, 1989), a resolution of 100m or less is necessary. While neither model has that, E2.2 comes much closer. We might therefore expect differences to arise in the tropics between the two models, in particular associated with the ability to resolve waves of short-wavelength and their velocity. Nevertheless, a spectral analysis of the eddy energy in the tropics failed to reveal any significant difference between E2.2 and E2.1, nor did a wave power spectral analysis as a function of wave number and period. Either the vertical resolution is still sufficiently coarse even in E2.2 that one cannot resolve a greater portion of the spectrum, or the other differences between the models overwhelm any effect approximately halving the vertical resolution might produce.

To examine this further, we performed another test in which we doubled the vertical resolution in E2.2 between 800 and 100 hPa (now 70-170 m). The eddy energy in the tropics increased noticeably for the faster moving waves at a variety of wavelengths. While it has no immediate relevance for E2.2, the result, if verified by more systematic experiments, could lead to a modification of that model for use in certain studies.

Vertical resolution can have other effects in the model. For example, with finer vertical resolution, the vertical profile of moist static energy can be better resolved, influencing the height to which convection occurs. In a comparison with ISCCP data, the root-mean square errors for cloud top temperature and cloud top pressure are considerably reduced in E2.2 compared with E2.1. This result may not strictly be a convective phenomenon, as a better delineation of altitude could be beneficial for a wide variety of cloud forming processes. It can be of great importance in climate change assessments, given the sensitivity of the cirrus cloud greenhouse effect to their change in temperature and altitude with climate.

With greater vertical resolution the boundary layer can be better resolved. This could allow for more adequate resolution of shallow boundary layers during stable situations. Finer vertical resolution will also change the degree of instability possible between layers, and that may then affect the maximum height of the boundary layer. Comparing the diurnal variation of the boundary layer depth for E2.2 and E2.1 in 34 different locations, we found that in about 2/3 of the cases, the minimum depth had decreased and the maximum increased in E2.2, consistent with expectations. No comparison with observations has been performed as of yet.

4.3. Rainfall evaporation above cloud base during convective events

The particular choice of parameterization associated with the evaporation of rainfall above cloud base in convective events is highly consequential in the GISS models. As noted earlier, this was allowed in E2.1, and not in E2.2, E2.2 AP, or, for that matter, in the CMIP5 model E2.0. It was implemented originally to help in generation of the MJO; ironically, recent tests have shown that in the E2.1 configuration, the MJO is equally as strong without this evaporation.

When this evaporation is allowed, the particulars of the GISS convection scheme and dynamics result in excessive rainfall in the western Pacific, especially in June-August, and the most accurate precipitation is simulated when none is allowed. Tests show that intermediate values

of allowed evaporation produce intermediate values of rainfall accuracy. This excessive rainfall produces an exaggerated Hadley Circulation in that season, and also an exaggerated Walker Circulation that leads to larger than observed ENSO amplitudes in E2.1 using the G model. This last feature is avoided entirely in E2.2 and E2.2AP (Figure 29) and the others are minimized, more so in E2.2AP as other parameterization choices are also involved.

As might be expected, one other aspect that allowing this evaporation produces is increased moisture loading in the atmosphere. Comparison to both ERA5 and MERRA2 shows that the resulting moisture is excessive; with this evaporation disallowed in E2.2, the moisture loading and distribution is more accurate (Figure 33, Table 2). Yet with this added moisture in E2.1, the atmospheric model is warmer, and a better match to observations (Table 2 and the associated discussion in Section 3.5.1.5). The warmer air would allow for more evaporation and moisture, so the result is interactive. In the CMIP5 version of the model, E2.0, both the temperature and the moisture loading were in accordance with observations (without allowing this evaporation) (Schmidt et al., 2014, their Table 4) so in this one respect that model was more realistic.

This choice has perhaps even great consequences in the coupled versions of these models. Shown in Table 3 are the annual average pre-industrial values for temperature and the relevant variables that influence it, for E2.2, E2.2AP, E2.2 RE and E2.1. To further evaluate the effect of this parameterization independent of model configuration, we reversed the procedure and ran E2.1 with the evaporation disallowed above cloud base in convective events (E2.1, no evaporation (NE)), and this result is also presented. All of these runs are put in radiation balance by retuning their cloud cover before the coupled simulations start.

The global air temperature in E2.2 and E2.2AP is some 2°C colder than E2.1, due to a reduced greenhouse capacity especially in E2.2, and a higher planetary albedo in E2.2AP. With evaporation turned on in E2.2RE, its water vapor increased as did its greenhouse capacity, while cloud cover, sea ice and planetary albedo decreased; the air temperature is now about the same as in E2.1. With the evaporation turned off in E2.1NE, the atmospheric water vapor and greenhouse capacity decreased, while cloud cover, sea ice, and planetary albedo increased; the air temperature decreased by 1.5°C. As can be seen in the table, E2.2 RE and E2.1NE have a number of their climate parameters (surface air temperature, sea ice and snow cover) quite similar, implying that this evaporation difference could be primarily responsible for the differences between E2.2 and E2.1 in those fields. The ones that still differ – cloud cover, planetary albedo, atmospheric water vapor – are presumably associated with other parameterizations that differ (see Section 2.3), or the vertical resolution (e.g., the greater number of layers might be increasing the probability of low cloud cover). Note also that E2.2 has an improved longwave radiation calculation for downwelling to the surface that by itself can produce cooling in the tropics, independent of any other changes.

It might appear to be a concern for climate change investigations that E2.2 and E2.2AP are colder than E2.1, whose temperature is more representative of actual conditions, both PI and current. Climate change assessments generally utilize differences between experiments and control, with the absolute value of some control run climate variables of less importance unless

they lead to unrealistic feedbacks or impacts. In terms of climate feedbacks, the E2.2 models had less water vapor (Figure 33) and more sea ice than E2.1 (Section 3.2.1). In terms of climate impacts, its precipitation distribution was better than E2.1 (Figure 34). Ultimately, the temperature difference does not seem to affect the climate sensitivity, which is only slightly different among the models (Section 3.4.1) (implying somewhat similar cloud feedbacks).

This parameterization had a direct effect on the ENSO amplitude with the G model. With the evaporation allowed, the ENSO amplitude in E2.1 is some 30% excessive; when disallowed in E2.1NE, it dropped back to near the observed. In E2.2, without the evaporation, the ENSO is 75% of observed; with it enabled in E2.2RE, it rose to equal the observed.

The parameterization has another major consequence in the GISS model: it affects the sensitivity of North Atlantic Deep Water in warming climates. We show in Figure 31 the results for E2.1NE along with the other model configurations. In the colder climate runs (E2.2, E2.2AP, E2.1NE), with $4xCO_2$ the AMOC weakens but does not collapse, and (especially with E2.1NE) eventually recovers, while in the warmer climates of E2.1 and E2.2RE, the circulation collapses quickly and does not recover (Figure 31). (E2.1 has been run for several thousand years, and it has still not recovered).

A number of the CMIP6 models are showing a collapse of the AMOC in their $4xCO_2$ runs, although none as rapidly as in E2.1. There were collapses in CMIP5 models as well, but not as many, and the CMIP6 models are generally showing greater warming. It is important to note that during the first decade after the instantaneous $4xCO_2$ forcing, when a distinction is already apparent in their AMOC responses, E2.1 is not warming globally any more than E2.2 (both being around 3.5°C, Figure 30), and closer investigation shows this is also true of the North Atlantic region. The response difference is salinity-driven.

Even the initial warmth of the model may not be the deciding factor. The CMIP5 control run (E2.0) has even warmer temperatures than E2.1, although, as indicated above, achieved without allowing precipitation evaporation above cloud base in convective events (Schmidt et al., 2014). With E2.0, the NINT coupled simulation produced a global current day temperature of 14.9°C, while that of E2.1 is 14.1°C (Kelley et al., 2019, submitted, Table 1). However, during the first decade, both models had roughly the similar initial warming in response to instantaneous $4xCO_2$ forcing, and that E2.0 configuration had no AMOC collapse (Rind et al., 2018). Hence, while there are other differences between them, so far in all the GISS E2 NINT models, this parameterization is correlated with the modeled fate of the Atlantic Meridional Overturning Circulation in the $4xCO_2$ climate (and even, less dramatically, in the $2xCO_2$ climate in terms of weakening).

Finally, how does this relate to a main theme of this paper, optimizing a Middle Atmosphere model in a climate model? We found that with the rainfall evaporation allowed, the resulting additional deep convection and effect on winds in the UTLS made it difficult to generate a proper QBO and simultaneously a proper Age of Air in the stratosphere, as will be discussed in Paper 2. That is in fact why we initially focused on this particular parameterization choice in the

convection routine; there are many others, such as entrainment rates, downdraft details, etc. to which the Middle Atmosphere and the model in general could be sensitive. With the basic GISS ModelE structure, the different choices made for this parameterization in E2.1 and E2.2 influence many different simulation results; the same may not be true in other GCMs.

4.4 Building a Middle Atmosphere model in a climate model

While E2.2AP has a Middle Atmosphere simulation that is less 'accurate' than E2.2 for Northern Hemisphere winter, the differences are not very large (compare Figure 35 with Figures 5 and 7), and the QBO is properly developed. However, there were times in the development process when E2.2AP was much worse, and these had to do with the revised elements of the convective parameterization introduced in that model. Earlier choices led to greater values of tropical convective mass flux to high altitudes than can be seen in the current version (Figure 32) where E2.2AP is now only somewhat larger than in E2.2 above 500 hPa. With greater convective fluxes, the associated heat transports and circulation altered the temperature distribution in the tropical upper troposphere, and thus the subtropical zonal wind field in the UTLS. It led to greater wave refraction toward lower latitudes, not only reducing wave-mean flow interaction at higher northern latitudes but also increasing the stratospheric residual circulation to unacceptable magnitudes (as will be discussed in Paper 2). The changes we made to E2.2AP convection/clouds to improve the Middle Atmosphere helped drive it away from the convection and clouds of E2.1 that it was attempting to reproduce.

Convective parameterizations for the tropics generally focus on simulating heating profiles from the few observational campaigns available, and producing acceptable observed features such as the MJO or cloud cover. Researchers in that community do not usually investigate what their choices do to the Middle Atmosphere depiction in the models. In the GISS modelE simulations, the model version (E2.2) that had the least penetrative convection above 500 hPa had the best Middle Atmosphere simulation. Our experience suggests that the two communities need to work together if the goal is to have both a good troposphere and good Middle Atmosphere for climate and climate change experiments.

4.5. CMIP6 constraints

The CMIP6 requirements are that if both AMIP NINT and Coupled atmosphere-ocean model results are submitted to CMIP6, they must have the same 'model code', which includes parameterizations and their tunings. Given the current state of atmosphere-ocean modeling, the observed sea surface temperature field cannot be reproduced without significant errors, so the two models are being forced from below differently. With the uncertainties in the physics underlying clouds and convection, among other things, modelers have a range of choices available, and would conceivably make different choices given different underlying forcing patterns. But CMIP6 requirements, if both are to be submitted, do not allow them to do that, with the result that it only allows one of those models to be 'optimized', for both the parameterizations and their tunings.

The effect is noticeable in the models run here due to the gravity wave drag parameterization being associated with the model's convection. With different SSTs in the AMIP NINT versus the

Coupled atmosphere-ocean model, the model's convection and thus parameterized gravity wave momentum fluxes are different, and so is the tropical wind field through which they must 'propagate'. The same choices relating convection to gravity wave momentum flux and associated gravity wave phase velocities did not work to produce an acceptable QBO in both models simultaneously. We chose to optimize the coupled model, for its use in assessing how the QBO may change with climate, and so in particular had to alter the assigned phase velocities and the efficiency factor away from what had normally been used. Consequentially, that made the QBO in the (CMIP6-approved) AMIP NINT version less accurate (this also was true for the Age of Air, discussed in Paper 2). When we 'retuned' the convection/momentum flux scaling, the AMIP NINT version 2 could now produce a good QBO (Figure 37), although this cannot be submitted to CMIP6. Instead we submit only the OMA AMIP version, which is allowed to have a differing model code as it is considered a 'different model' given the interactive chemistry.

While CMIP6 does not have the Middle Atmosphere simulation as one of its prime foci, this 'conflict' would not be isolated to Middle Atmosphere concerns, as many other model features would probably be improved in one or the other of the models if more flexibility were allowed for joint submissions.

5 Conclusions

In this paper we present the atmospheric and oceanic mean fields, variability, climate sensitivity and sensitivity to parameterizations for GISS model E2.2. An over-riding theme is to discuss choices that have been made to optimize the simulation of the Middle Atmosphere within a climate model context. The output of one of the versions of this model, coupled to two ocean models, is being provided to the CMIP6 archive.

The results show that it is possible to build a climate model with an appropriate simulation of the Middle Atmosphere if one has that intention during the climate model development process. Of particular importance is raising the model top far outside of the stratosphere so it no longer influences stratospheric dynamics. In addition, changes in tropospheric physics, as in the case of moist convection, must be regarded with a broader perspective that includes these other levels; this point will be emphasized further in Paper 2.

Acknowledgments and Data

Climate modeling at GISS is supported by the NASA Modeling, Analysis and Prediction program, and resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center. We thank Jan Perlwitz for examining the dust parameterization, Lee Murray for the lightning parameterization and Suzanne Bauer for assessing E2.2 OMA forcings. We thank Ivan Mitevski, a doctoral candidate in the Department of Applied Mathematics and Applied Physics at Columbia University, for sharing his analysis of the CMIP5 and CMIP6 multi-model response of the Atlantic Meridional overturning circulation in the abrupt 4xCO2 experiment. We also thank all the authors who contributed to creating GISS model E2.1, on which these models are built. Two anonymous reviewers provided suggestions that have helped to improve this paper.

E2.2 AP Model data is available through the Coupled Model Intercomparison Project (CMIP) data archive of the World Climate Research Programme (WCRP) (http://pcmdi9.llnl.gov) [under the generic title E2.2]. E2.2 Model data, following the CMIP6 protocol and naming, is available from NCCS portal (under the title E2.2.1) at https://portal.nccs.nasa.gov/datashare/giss_cmip6/CMIP/NASA-GISS/GISS-E2.2.1-G/. Observational data is available from https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/data_access/ for MERRA2, from https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5 for ERA5 and https://www.nodc.noaa.gov/OC5/woa13/ for WORLD OCEAN ATLAS 2013 version 2. Other observational data sources used for comparison are given in the text.

References

Baldwin, M. P., & Dunkerton, T. J. (2001), Stratospheric harbingers of anomalous weather regimes. *Science*, 294, 581-584. doi: 10.1126/science.1063315

Banerjee, A., Chiodo, G., Previdi, M., Ponater, M., Conley, A. J., & Polvani, L. M. (2019), Stratospheric water vapor: an important climate feedback. *Climate Dynamics*, 53,1697–1710. doi: 10.1007/s00382-019-04721-4

Bauer, M., Tselioudis, G., & Rossow, W. B. (2016), A new climatology for investigating storm Influences in and on the extratropics. *J. App. Met. and Clim*., 55, 1287-1303. doi: 10.1175/JAMC-d-15-0245.1

Bracegridele, T. J., Connolley, W. M., & Turner, J. (2008), Antarctic climate change over the twenty first century. *Climate and Dynamics,* 113, 13pp. doi: 10.1029/2007JD008933

Butler, A. H., Seidel, D. J., Hardiman, S. C., Butchart, N., Birner, T., & Match, A. (2015), Defining sudden stratospheric warmings. *Bull. Amer. Meteor. Soc.,* 96, 1913-1928. doi: 10.1175/BAMS-D-13-00173.1

Chandran, A., Collins, R. L., & Harvey, V. L. (2014), Stratosphere-mesosphere coupling during stratospheric sudden warming events. *Advances in Space Research,* 53, 1265–1289. doi: 10.1016/j.asr.2014.02.005

Charlton, A. J., & Polvani, L. M. (2007), A new look at stratospheric sudden warmings. Part I: Climatology and modeling benchmarks. *J. Clim.,* 20, 449-469. doi: 10.1175/JCLI3994.1

Charlton-Perez, A. J., Baldwin, M. P., Birner, T., Black, R. X., Butler, A. H., Calvo, N., Davis , N. A., et al. (2013), On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models. *J. Geophys. Res. Atmospheres,* **118**, 2494-2505. doi: 10.1002/jgrd.50125

Cheng, W., Chiang, J. C. H. & Zhang, D. (2013), Atlantic meridional overturning circulation (AMOC) in CMIP5 models: RCP and historical simulations. *J. Clim*., 26, 7187-7197. doi:10.1175/JCLI-D-12-00496.1

Chiodo, Gabriel, & Polvani, L. M. (2017), Reduced Southern Hemispheric circulation response to quadrupled CO2 due to stratospheric ozone feedback. *Geophys. Res. Lett.* 44, 465-474. doi: 10.1002/2016GL071011

Christiansen, B. (2005), Downward propagation and statistical forecast of the near-surface weather. *J. Geophys. Res.,* 110, D14104. doi:10.1029/ 2004JD005431.

Dall'Amico, M., Stott, P. A., Scaife, A. A., Gray, L. J., Rosenlof, K. H. & Karpechko, A. Y. (2010),

Impact of stratospheric variability on tropospheric climate change. *Clim Dyn*, 34, 399–417. doi: 10.1007/s00382-009-0580-1

Danabasoglu, G. & Yeager, S. G. (2012), Variability of the Atlantic Meridional Overturning Circulation in CCSM4. *J. Clim*., **25**, 5153-5172. doi: 10.1175/JCLI-D-11-00463.1

Del Genio, A. D., Chen, D. Y., Kim, D. & Yao, M.-S. (2012), The MJO transition from shallow to deep convection inCloudSat/CALIPSO data and GISS GCM simulations. *Journal of Climate* , 25, 3755–3770, doi:10.1175/jcli-d-11-00384.1.

Del Genio, A. D., Wu, J., Wolf, A. B., Chen, Y., Yao, M.-S & Kim, D. (2015), Constraints on cumulus parameterization from simulations of observed MJO events, *Journal of Climate,* 28, 6419–6442, doi:10.1175/jcli-d-14-00832.1.

Donohue, K. A., Tracey, K. L., Watts, D. R., Chidichimo, M. P. & Chereskin, T. K. (2016), Mean Antarctic Circumpolar Current transport measured in Drake Passage. *Geophys. Res. Lett.,* 43 (22), 11,760-11,767. doi: 10.1002/2016GL070319

Farneti, R., Downes, S. M., Griffies, Marslandd, S. J., Behrens, E., Bentsenf, M., & Bi, D. (2015) et al. (2015), An assessment of Antarctic Circumpolar Current and Southern Ocean meridional overturning circulation during 1958–2007 in a suite of interannual CORE-II simulations. *Ocean Modelling,* 93, 84–120. doi: 10.1016/j.ocemod.2015.07.009 (see http://www.cgd.ucar.edu/staff/mclong/pubs/Farneti-Downes-etal-2015.pdf)

Ferreira, D., Marshall, J. Bitz, C. M., Solomon, S. & A. Plumb (2015), Antarctic Ocean and Sea Ice Response to Ozone Depletion: A Two-Time-Scale Problem. *J. Climate*, 28, 1206-1226. doi: 10.1175/JCLI-D-14-00313.1

Fleming, E.L., Chandra, S., Barnett, J. J., & Corney, M. (1990), Zonal mean temperature, pressure, zonal wind, and geopotential height as functions of latitude, COSPAR International Reference Atmosphere: 1986, Part II: Middle Atmosphere Models, *Adv. Space Res.,* 10, 11-59. doi: 10.1016/0273-1177(90)90386-E

Fletcher, C.G., Kushner, P. J., & Cohen, J. (2007), Stratospheric control of the extratropical circulation response to surface forcing. *Geophys. Res. Lett.,* 34, L21802. doi:10.1029/2007GL031626

Frajka-Williams, E., Cunningham, S. A., Bryden, H., & King, B. A. (2011), Variability of Antarctic Bottom Water at 24.5°N in the Atlantic. *Journal of Geophysical Research Oceans,* 116, (C11). doi: 10.1029/2011JC007168

Garfinkel, C. I., & Hartmann, D. L. (2011), The influence of the quasi-biennial oscillation on the troposphere in winter in a hierarchy of models. Part I: Simplified dry GCMs. *J. Atmos. Sci.,* 68, 1273–1289. doi: 10.1175/2011JAS3665.1

Gerber, E. P., & Manzini, E. (2016), The Dynamics and Variability Model Intercomparison Project (DynVarMIP) for CMIP6: assessing the stratosphere-troposphere system. *Geoscientific Model Development,* 9, 3413-3425. doi: 10.5194/gmd-9-3413-2016

Gettelman, A., Hoor, P., Pan, L. L., Randel, W. J., Hegglin, M. I. & Birner, T. (2011), The extratropical upper troposphere and lower stratosphere. *Rev. Geophys*., 49, RG3003, doi:10.1029/2011RG000355.

Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., & Marsh, D. R. et al. (2019), The Whole Atmosphere Community Climate Model Version 6 (WACCM6). *J. Geophys. Res. Atmos*., 124, doi: 10.1029/2019JD030943

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., & Thorpe, R. P. et al. (2004), A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett*., 31, doi: 10.1029/2003GL018747

Grist, J. P., Josey, S. A., New, A. L., Roberts, M., Koenig, T., & Iovino, D. (2018), Increasing Atlantic Ocean Heat Transport in the Latest Generation Coupled Ocean-Atmosphere Models: The Role of Air-Sea Interaction (2018). *J. Geophys. Res. Oceans*, 123, 8624-8637. doi: 10.1029/2018JC014387

Hall, T. M., & Plumb, R. A. (1994), Age as a diagnostic of stratospheric transport. *J. Geophys. Res. Atmos.* 99, 1059-1070. doi: 10.1029/93JD03192

Hansen, J., Sato, M., Ruedy, R., Kharecha, P., Lacis, A., Miller, R., Nazarenko, L. et al., (2007), Climate simulations for 1880–2003 with GISS modelE. *Climate Dynamics* 29, 661-696. doi: 10.1007/s00382-007-0255-8

Hardiman, S.C., Butchart, N., Osprey, S. M., Gray, L. J., Bushell, A.C. & Hinton, T. J. (2010), The Climatology of the Middle Atmosphere in a vertically extended version of the Met Office's Climate Model. Part I: Mean State. *J. Atmos. Sci.*, 67, 1509-1525. doi: 10.1175/2009JAS3337

Hardiman, S.C., Butchart, N., Hinton, T. J., Osprey, S. M. & Gray, L. J. (2012), The effect of a well-resolved stratosphere on surface climate: differences between CMIP5 simulations with high and low top versions of the Met Office Climate Model. *J of Climate*, 25, 7083-7099. doi: 10.1175/JCLI-D-11-00579.1

Hazeleger, W., Seager, R., Cane, M. A., & Naik, N. H. (2004), How can tropical Pacific Ocean heat transport vary? *J. Physical Ocean*., 34, 320-333. doi: 10.1175/1520-0485(2004)034<0320:HCTPOH>2.0.CO;2.

Huffman, G. J., Adler, R. F., Arkin, A., Chang, A., Ferraro, R., Gruber, A., Janowiak, J., Joyce, R. J., McNab, A., Rudolf, B., Schneider, U. & Xie, P. (1997), The Global Precipitation Climatology Project (GPCP) combined precipitation data set. *Bull. Amer. Meteor. Soc.,* 78,5–20. doi: 10.1175/1520-0477(1997)078%3C0005:TGPCPG%3E2.0.CO;2

Jablonowski, C. & Williamson, D.L. (2011), The Pros and Cons of Diffusion, Filters and Fixers in Atmospheric General Circulation Models. In "Numerical Techniques for Global Atmospheric Models", Chapter 11, 381-493. P.H. Lauritzen et al. (eds.), Lecture Notes in Computational Science and Engineering 80. doi: 10.1007/978-3-642-11640-7_13, c Springer-Verlag Berlin Heidelberg 2011

Jones, P. D., New, M., Parker, D. E., Martin, S. & Rigor, I. G. (1999), Surface air temperature and its variations over the last 150 years. *Revs. Geophys*., 37, 173–199. doi: 10.1029/1999RG900002

Kawatani, Y., Hamilton, K., Gray, L. J., Osprey, S. M., Watanabe, S. & Yamashita, Y. (2019). The effects of a well-resolved stratosphere on the simulated boreal winter circulation in a climate model. *J. Atmos. Sci.,* 76, 1203-1226. doi.org/10.1175/JAS-D-18-0206.1

Kim, D., Sobel, A. H., Maloney, E. D., Frierson, D. M., & Kang, I. S. (2011), A systematic relationship between intraseasonal variability and mean state bias in AGCM simulations. *J. Clim.*, 24, 5506-5520. doi: 10.1175/2011JCLI4177.1

Kim, D., Sobel, A. H., Del Genio, A. D., Chen, Y., Camargo, S. J., Yao, M.-S. e al. (2012), The tropical subseasonal variability simulated in the NASA GISS general circulation model. *J. Clim*., 25, 4641-4659. doi: 10.1175/JCLI-D-11-00447.1

Kim, J. & Son, S.W. (2012), Tropical cold-point tropopause: climatology, seasonal cycle, and intraseasonal variability derived from COSMIC GPS radio occultation measurements. *J. Clim*., 25, 5343-5360. doi: 10.1175/JCLI-D-11-00554.1

Knox, J. A. & Harvey, V. L. (2005), Global climatology of inertial instability and Rossby wave breaking in the stratosphere. *J. Geophys. Res. Atmos*., 110, doi: 10.1029/2004JD005068

Kodera, K. (2004), Solar influence on the Indian Ocean Monsoon through dynamical processes, *Geophys. Res. Lett.,* 31, L24209. doi: 10.1029/2004GL020928.

Lindzen, R. S. & Fox-Rabinowitz, M. (1989), Consistent vertical and horizontal Resolution. *Mon. Wea. Rev*., 117, 2575-2583, doi: 10.1175/1520-0493(1989)117<2575:CVAHR>2.0.CO;2

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, G. et al. (2018), Clouds and the Earth's Radiant Energy System (CERES) energy balanced and filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 Data Product. *J of Climate*, 31, 895-918. doi: 10.1175/JCLI-D-17-0208.1

Loeb, N. G., Rose, F. G., Kato, S., Rutan, D. A., Su, W., Wang, H., Doelling, D. R., Smith, W. L. & Gettelman, A. (2019), Towards a consistent definition between satellite and model clear-sky radiative fluxes. *Journal of Climate*, doi:10.1175/jcli-d-19-0381.1.

Madden, R. A. & Julian, P. R. (1972), Description of global-scale circulation cells in the tropics with a 40–50 day period, *J. Atmos. Sci.,* 29, 1109– 1123, doi:10.1175/1520-0469(1972)029<1109:DOGSCC>2.0.CO;2.

Madden, R. A. & Julian, P. R. (1994), Observations of the 40–50-day tropical oscillation—A review. *Mon. Wea. Rev*., 122, 814–837. doi: 10.1175/1520-0493(1994)122%3C0814:OOTDTO%3E2.0.CO;2

Mantasa, D., 2017: A connection from stratospheric ozone to El Niño-Southern Oscillation. *Scientific Reports,* 7, Article number: 5558. doi: 10.1038/s41598-017-05111-8

Martin, Z. & Wang, S. (2019), The Impact of the QBO on MJO Convection in Cloud-Resolving Simulations. *J. Atmos. Sci.,* 76, 669-688. doi:10.1175/JAS-D-18-0179.1

National Center for Atmospheric Research Staff (Eds), 2017: "The Climate Data Guide: RAPID: monitoring the Atlantic Meridional Overturning Circulation at 26.5°N." Retrieved from https://climatedataguide.ucar.edu/climate-data/rapid-monitoring-atlantic-meridional-overturning-circulation-265%C2%B0n.

Polvani, L., Previdi, M., and Deser, C. (2011), Large cancellation, due to ozone recovery, of future Southern Hemisphere atmospheric circulation trends, *Geophys. Res. Lett*., doi: 10.1029/2011GL046712

Richter, J. H., Solomon, A. & Bacmeister (2014), J. T., On the simulation of the quasi-biennial oscillation in the Community Atmosphere Model, version 5. *J. Geophys. Res. Atmos,* 119, 3045–3062. doi:10.1002/2013/D021122.

Richter, J. H., Chen, C-C., Tang, Q., Shaocheng, X. & Rasch, P.J., 2019, Improved Simulation of the QBO in E3SMv1, *J. Geophys. Res. Atmos.,* 124. doi: 10.1029/2019MS001763

Rind, D., Suozzo, R., Balachandran, N. K., Lacis, A., & Russell, G. (1988), The GISS Global Climate-Middle Atmosphere Model. Part I: Model structure and climatology. *J. Atmos. Sci.*, 45, 329-370. doi: 10.1175/1520-0469(1988) 045<0329:TGGCMA>2.0.CO;2.

Rind, D., Shindell, D., Lonergan, P., & Balachandran, N. K. (1998), Climate change and the middle atmosphere. Part III: The doubled CO2 climate revisited. *J. Climate*, 11, 876-894. doi: 10.1175/1520-0442(1998)011<0876:CCATMA>2.0.CO;2.

Rind, D., Chandler, M., Lonergan, P., & Lerner, J. (2001), Climate change and the middle atmosphere: 5. Paleostratosphere in cold and warm climates. *J. Geophys. Res.,* 106, 20195-20212. doi:10.1029/2000JD900548.

Rind, D., Perlwitz, J. & Lonergan, P. (2005), AO/NAO response to climate change: 1. Respective influences of stratospheric and tropospheric climate changes. *J. Geophys. Res.,* 110, D12107. doi:10.1029/2004JD005103.

Rind, D., Lean, J., Lerner, J., Lonergan, P., & Leboissetier, A. (2008), Exploring the stratospheric/tropospheric response to solar forcing. *J. Geophys. Res.,* 113, D24103. doi:10.1029/2008JD010114.

Rind, D., Jonas, J., Balachandran, N. K., Schmidt, G. A., & Lean, J. (2014), The QBO in two GISS global climate models: Part 1: Generation of the QBO. *J. Geophys. Res. Atmos.,* 119, 8798-8824. doi:10.1002/2014JD021678.

Rind, D., Schmidt, G. A., Jonas, J., Miller, R., Nazarenko, L., Kelley, M. & Romanski, J. (2018), Multicentury instability of the Atlantic Meridional Circulation in rapid warming simulations with GISS ModelE2. J. Geophys. Res. Atmos., 123, doi: 10.1029/2017JD027149

Sassi, F., Garcia, R. R., Marsh, D., & Hoppel, K. W. (2010), The role of the middle atmosphere in simulations of the troposphere during Northern Hemisphere winter: Differences between high- and low-top models. *J. Atmos. Sci*., 67, 3048– 3064. doi: 10.1175/2010JAS3255.1

Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I. et al. (2014), Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *J. Adv. Model. Earth Syst.,* 6, 141–184. doi:10.1002/2013MS000265.

Shepherd, T.G., Semeniuk, K. & Koshyk, J.N. (1996), Sponge-layer feedbacks in middle atmosphere models. *J. Geophys. Res.,* 101, 23447-23464. doi: 10.1029/96JD01994

Sigmond, M., Scinocca, J. F., & Kushner, P. J. (2008), Impact of the stratosphere on tropospheric climate change. *Geophys. Res. Lett.,* 35, L12706. doi:10.1029/2008GL033573.

Smith, A. K., Holt, L. A., Garcia, R. R., Antsey, J. A., Serva, F., Butchart, N. et al. (2019), The equatorial stratospheric semiannual oscillation and time-mean winds in QBOi models. Accepted for publication. in the *Roy. Meteor. Soc*. doi: 10.1002/qj3690

Son, S-W., Gerber, E. P., Perlwitz, J., Polvani, L. M., Gillett, N. P., Seo, K-H., Eyring, V. et al. (2010), Impact of stratospheric ozone on Southern Hemisphere circulation change: A multimodel assessment. *J. Geophys. Res. Atmospheres*, 115, doi:10.1029/2010JD014271

Son, S-W, Lim, Y., Yoo, C., Hendon, H. H., & Kim, J. (2017). Stratospheric control of the Madden–Julian oscillation. *Journal of Climate,* 30, 1909-1922. doi: 10.1175/JCLI-D-16-0620.1

Speer, K., Rintoul, A. E., & Sloyan, B. (2000), The diabatic Deacon Cell. *J. Phys. Ocean*., 30, 3212-3222. doi: 10.1175/1520-0485(2000)030%3C3212:TDDC%3E2.0.CO;2

Stephens, G. L., Li, J., Wild, M., Clayson, C. A., Loeb, N., Kato, S., L'Ecuyer et al. (2012), An update on Earth's energy balance in light of the latest global observations. *Nat. Geosci.,* 5, 691–696, doi:10.1038/ngeo1580

Stubenrauch, C. J., Rossow, W. B., Kinne, S., Ackerman, S., Cesana, G. Chepfer, H. et al. (2013), Assessment of global cloud datasets from satellites: Project and database initiated by the GEWEX radiation panel, *Bull. Amer. Meteor. Soc*., 94,

1031—1049. doi:10.1175/BAMS-D-12-00117.1.

Trenberth, K. E., Fasullo, J. T. & Kiehl, J. (2009), Earth's global energy budget, *Bull. Amer. Meteor. Soc*., 90, 311–324. doi:10.1175/2008BAMS2634.1.

Trenberth, K. E. & Fasulo, J. T. (2017), Atlantic meridional heat transports computed from balancing Earth's energy locally. *Geophys. Res. Lett*., 44, doi: 10.1002/2016GL072475

Waugh, D., Primeau, W. F., DeVries, T. & Holzer, M. (2013), Recent changes in the ventilation of the southern oceans. *Science,* 339, 568-570. doi: 10.1126/science.1225411

Wang, W., Zhu, X., Wang, C., & Kohl, A. (2014), Deep meridional overturning circulation in the Indian Ocean and its relation to Indian Ocean Dipole. *J. Clim.,* 27, 4508-4520. doi: 0.1175/JCLI-D-13-00472.1

Wheeler, M. & Kiladis, G. N. (1999), Convectively Coupled Equatorial Waves: Analysis of Clouds and Temperature in the Wavenumber-Frequency Domain. *J. Atmos. Sci.,* 56, 374-399. doi: 10.1175/1520-0469(1999)056%3C0374:CCEWAO%3E2.0.CO;2

Wunderlich, F. & Mitchell, D. M. (2017), Revisiting the observed surface climate response to large volcanic eruptions. *Atmos. Chem. Phys.*, 17, 485–499, doi:10.5194/acp-17-485-2017

Yoo, C., & Son, S. W. (2016), Modulation of the boreal wintertime Madden-Julian oscillation by the stratospheric quasi-biennial oscillation. *Geophys. Res. Lett.,* 43, 1392-1398. doi: 10.1002/2016GL06776

Zelinka, M. D., Myers, T. A., McCoy, D. T., Pro-Chedley, S., Caldwell, P. M. & Ceppi, P. et al. (2020), Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett*., 47, doi.:10.1029/2019GL085782

Table 1. Vertical resolution and layering in model versions E2.2 and E2.1.

REGION	PRESSURE LEVELS (hPa)	E2.1 (40 Layers)		E2.2 (102 Layers)	
		# of Layers	Average Thickness (m)	# of Layers	Average Thickness (m)
TROPOSPHERE					
Lower	1000-800	7	300	12	140
Middle	800-400	8	650	21	265
Upper	400-100	10	830	25	355
STRATOSPHERE					
Lower	100-50	3	1000	7	530
Middle	50-30	2	2500	5	700
	30-10	2	3500	7	1020
Upper	10-5	1	4000	4	1170
	5-0.7	4	4200	11	1350
MESOSPHERE					
Lower	0.7-0.1	3	4500	4	3375
Middle	0.1-0.01			4	3350
Upper	0.01-0.002			2	4500

Table 2. Global annual mean model features over the period 1979-2014 for the E2.2 and E2.1 models run in OMA (AMIP) mode and key diagnostics compared to observations or best estimates. Cloud cover is estimated based on clouds with optical thickness >0.1 . [J] Jones et al. [1999] with updates, [C] CERES EBAF Ed4.1 Loeb et al. (2019), [T] Trenberth et al. (2009) and updates, [G]GPCP V2.3/TRMM TMPA V7 Huffman et al. (1997), [O]Obs4MIPs, [NO]NOAA NCEI https://www.nodc.noaa.gov/OC5/3M_HEAT_CONTENT/, [L]Loeb et al. (2018), [SEA]Stephens et al. [2012], [SRK]Stubenrauch et al. [2013]. "Best" and "Worst" categorized by comparing gridded model data with observations to produce root-mean-square errors (RMSE).

Field	E2.2	E2.2AP	E2.1	Observations	Best RMSE	Worst RMS
Surface Air Temp. °C	13.82	14.05	14.29	14.3±0.5[J]		
Planetary Albedo (%)	29.20	31.43	30.18	29.4[SEA] 29.1[L]	E2.2	E2.2AP
Cloud cover (%)	61.64	69.37	60.14	68 [SRK]	E2.2	E2.2AP
High cloud cover (%)	29.3	40.5	32.2		E2.2	E2.2AP
Low cloud cover (%)	45.2	45.8	40.5		E2.1	E2.2AP
Precip. (mm day^{-1})	2.92	2.88	2.98	2.9[G]	E2.2AP	E2.1
Atmos. water (mm)	24.75	23.48	26.14	24.9[O]		E2.1
Temp of Air °C	-22.85	-23.23	-22.29		E2.1	
Energy Fluxes (Wm^{-2})						
TOA Abs. SW	240.97	233.43	237.61	240.2[SEA]/239.4[T]		E2.2AP
TOA Outgoing LW	240.89	233.64	237.13	239.7[SEA]/238.5[T]	E2.2	E2.2AP
Surf. Abs. SW	164.86	158.52	161.89	165[SEA]/169[T]	E2.2	
Surf. Down. LW	343.2	344.4	345.7	345.6[SEA/]343[T]		
Surf. Net LW (up)	52.0	50.9	50.6	52.4[SEA]/57[T]		
Sensible heat flux	27.38	23.62	23.69	24[SEA]/17[T]		E2.2
Latent heat flux	84.52	83.34	86.20	88[SEA]/82[T]		E2.1
TOA cld. forcing	-15.9	-11.8	-15.3	-18.5 [C]		
TOA Net. Rad. Imb.	0.08	-0.21	0.48	0.40±.03[NO]		
Greenhouse Effect	154.37	161.63	159.23			

Table 3. Global, annual average model values from their pre-industrial coupled model simulations averaged over 30 years.

Field	E2.2	E2.2AP	E2.2RE	E2.1	E2.1NE
Surf Air Temp. °C	11.30	11.75	12.85	13.78	12.69
Temp of Air °C	-24.54	-24.28	-22.60	-22.47	-23.92
Planetary Albedo (%)	30.11	31.09	28.74	29.79	30.55
Cloud Cover (%)	65.52	68.83	57.46	58.49	62.52
Low Cloud Cover (%)	47.16	46.24	45.33	38.69	38.72
High Cloud Cover (%)	33.05	38.63	23.30	32.13	37.67
Sea Ice Cover (%)	6.43	6.41	5.43	4.76	5.47
Snow Cover (%)	13.77	13.72	12.18	11.40	12.37
Atmos. water (mm)	21.01	20.85	26.13	25.97	22.08
TOA Absorb. SW (Wm^{-2})	237.82	234.47	242.47	238.89	236.33
TOA Outgo LW(Wm^{-2})	-238.21	-234.77	-242.11	-238.85	-236.43
Greenhouse Effect (Wm^{-2})	145.20	149.90	150.18	155.58	152.55

FIGURE LEGENDS

Figure 1. E2.2 temperatures in December-February (left), and June-August (right), averaged over 1979-2014 in the five ensemble runs of the OMA NINT model.

Figure 2. E2.2 Temperature comparisons with ERA5 (left) and MERRA2 (right) for December-February (top) and June-August (bottom).

Figure 3. As in Figure 2 but for E2.1.

Figure 4. Tropical tropopause cold point temperature averaged from 10N-10S in different E2.2 model simulations, E2.1, and reanalysis data.

Figure 5. Zonal wind comparisons with ERA5 for E2.2 (top) and E2.1 (bottom), December-February (left) and June-August (right).

Figure 6. Zonal winds at 1 hPa (top) and 10 hPa (bottom) from the two reanalysis data sets, E2.2 and E2.1. Numbers in the upper right hand corner indicate maximum and minimum values.

Figure 7. Planetary Wave 1 (left) and Wave 2 (right) amplitude (top row) in ERA5 (note the different scales). Anomalies relative to ERA5 for E2.2 (second row) and E2.1 (third row).

Figure 8. E2.2 Sea level pressure (-1000 hPa) for the annual average (1979-2014), and two solstice seasons (top row), and anomalies compared to ERA5 for E2.2 (middle row) and E2.1 (bottom row). The numbers in the upper right hand corner represent the global average value for the image below, i.e., sea level pressure in the top row, E2.2 and E2.1 global average anomalies in the middle and bottom rows.

Figure 9. As in Figure 8 but for geopotential height – 16400m at 100 hPa.

Figure 10. As in Figure 8 but for geopotential height – 20000m at 50 hPa.

Figure 11. As in figure 8 but for geopotential height – 46000m at 1 hPa.

Figure 12. E2.2 Stream function for December-February (left) and June – August (right) (top row). Anomalies compared to ERA5 (middle row) and MERRA2 (bottom row).

Figure 13. Anomalies in annual, DJF and JJA sea surface temperature for E2.2 coupled to the G model (top) and H model (bottom). Model results are compared with the Hadley Centre reconstruction (from the HADSST-PI) for 1870-1900. Global average anomalies are shown in the upper right hand corner of each panel.

Figure 14. As in Figure 9a but for sea ice cover compared with HadISST-PI (which may be an underestimate for the N.H. as it is only slightly higher than current day values).

Figure 15. Model PI annual sea surface salinity compared with modern values. Observational data from the WORLD OCEAN ATLAS 2013 version 2 (WOA13). Localized extremes were -6.4 and 10.7 psu, respectively.

Figure 16. Comparison of model PI simulation with current observations (WOA13) for annual temperature (top) and salinity (bottom) for the Atlantic (left) and the Pacific (right).

Figure 17. The model stream function for the Atlantic (top) and Global (bottom). Positive values indicate clockwise circulation within the frame of the figure.

Figure 18. The zonally averaged ocean salinity in the Southern Hemisphere extratropics in model version E2.2, PI conditions.

Figure 19. Ocean currents in E2.2 (Gulf Stream (top), Kuroshio (middle), Antarctic Circumpolar (bottom).

Figure 20. Ocean heat transports in E2.2 (left) and E2.1 (right) for the global average (top), the Atlantic (2nd row), Pacific (3rd row) and Indian (bottom row). The values are broken down into the contributions by the overturning circulation, the gyre, the Gent McWilliams (GMW) parameterization for the effect of mesoscale eddies, and the total.

Figure 21. E2.2 coupled model tropical zonal wind (4N-4S) profile (top), at ~30 hPa (blue) and ~60 hPa (red) as a function of time in the PI control (middle) and periodogram of 30 years of those values (bottom). Peak period and month of the periodogram is indicated as text on the figure.

Figure 22. November through March zonally averaged zonal winds (top) and temperatures (bottom) evaluated at 60^{o}N and 10 hPa for one member of the E2-2 OMA AMIP ensemble (left) and the E2.2 NINT Coupled pre-industrial simulation (right). Years correspond to actual model years in the left panels and to years relative to "branching" for the abrupt CO_2 experiments within the underlying control in the right panels. Only forty years are shown in the latter (150-year-long) simulation per the daily data saving protocol for the CMIP6-endorsed Dynamics and Variability Model Intercomparison Project (DynVarMIP) (*Gerber and Manzini (2016)*).

Figure 23. SSW variation with month in the Northern Hemisphere for the E2.2 OMA AMIP and Coupled runs compared with MERRA2 observations.

Figure 24. Zonal winds at 2 hPa between 4N and 4S as a function of month, showing the semi-annual wind oscillation (SAO) for the E2.2 models discussed in this paper, as well as E2.1 and reanalysis data.

Figure 25. Tropical zonal wind changes induced by various mechanisms at 1 hPa (top) and 2 hPa (bottom) for E2.2 OMA. See text for the explanation of the different contributions.

Figure 26. December – February Index related to the Northern Annular Mode (left), defined as (30-50N) minus (60-80N), for sea level pressure (top), 500 hPa geopotential heights (middle) and 50 hPa geopotential height (bottom) for E2.2, E2.2AP, ERA5 and MERRA2. The standard deviations for each model and observational data set are given in the lower left hand corner of each panel. Similarly, an index related to the Southern Annular Mode is shown (right), 40°S minus 65°S averaged for June-August, with the standard deviations given in the lower right-hand corner.

Figure 27. Storm track density for November-March, ~1980-2014 in the OMA runs of E2.1 and E2.2 as well as observations from MERRA2 (1980-2018).

Figure 28. Zonal wave-number frequency power spectra about the equator of precipitation. Shown are the results for E2.2AP (left) and E2.1 (right) indicating the MJO and power in various tropical waves. The horizontal line depicts the 60 day period; power above that line in wave numbers 1-3 is indicative of the MJO. Contours are unitless, i.e., the tropical mean symmetric or antisymetric component of precipitation divided by the background spectrum.

FIgure 29. NINO3.4 amplitude in E2.2, E2.2AP and E2.1 with the G ocean model (left) and H ocean model (right). Model standard deviation is given in green, observed in red.

Figure 30. Surface air temperature change with time for instantaneous $2xCO_2$ (solid lines) and $4xCO_2$ (dashed lines) forcing with the G model (left) and H model (right) for E2.2 (top) and E2.1 (bottom).

Figure 31. Maximum AMOC value in response to instantaneous $2xCO_2$ (top) and $4xCO_2$ (bottom) forcings for E2.2 and E2.1 experiments with the G and H ocean models. Results also shown for additional experiments (E2.2RE and E2.1NE) described later in the text. All values are five year running means.

Figure 32. Difference in convective mass flux for E2.2AP - E2.2 (left), E2.2 - E2.1 (middle) and E2.2AP – E2.1 (right) for the 36 year annual averages in the OMA AMIP runs.

Figure 33. Annual specific humidity anomalies for E2.2 (top), E2.2AP (middle) and E2.1 (bottom) compared with ERA5 (left) and MERRA2 (right).

Figure 34. Precipitation for the E2.2 OMA AMIP model (top row) on the annual average (left), December-February (middle) and June-August (right). The following three rows show the precipitation anomaly relative to GPCP for E2.2, E2.2AP and E2.1. Global average values are given in the upper right hand corner, for either precipitation or the anomaly.

Figure 35. Anomalies of zonal wind and Wave 1 amplitude in E2.2AP relative to ERA5 in the AMIP NINT simulation.

Figure 36. QBO results as in Figure 21 but for E2.2AP. Peak month and power of the periodogram is indicated on the figure.

Figure 37. QBO generated with AMIP NINT VERSION 2, amplitude of tropical zonal wind (top) (4N-4S) and periodogram (bottom). Text on the figure indicates period and magnitude of peak power.

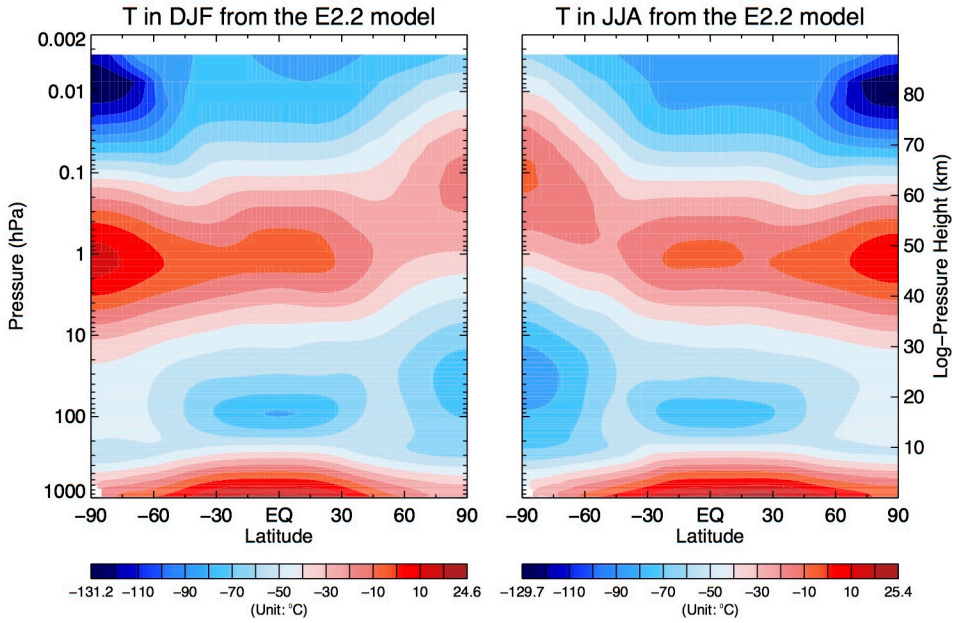

Figure 1. E2.2 temperatures in December-February (left), and June-August (right), averaged over 1979-2014 in the five ensemble runs of the OMA NINT model.

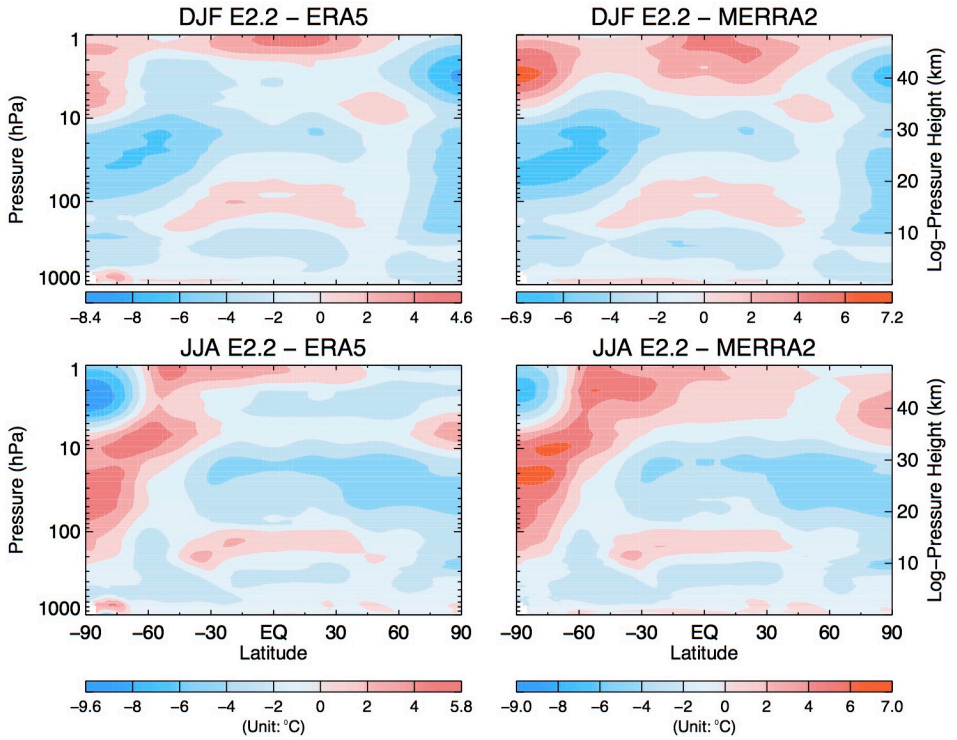

Figure 2. E2.2 Temperature comparisons with ERA5 (left) and MERRA2 (right) for December-February (top) and June-August (bottom).

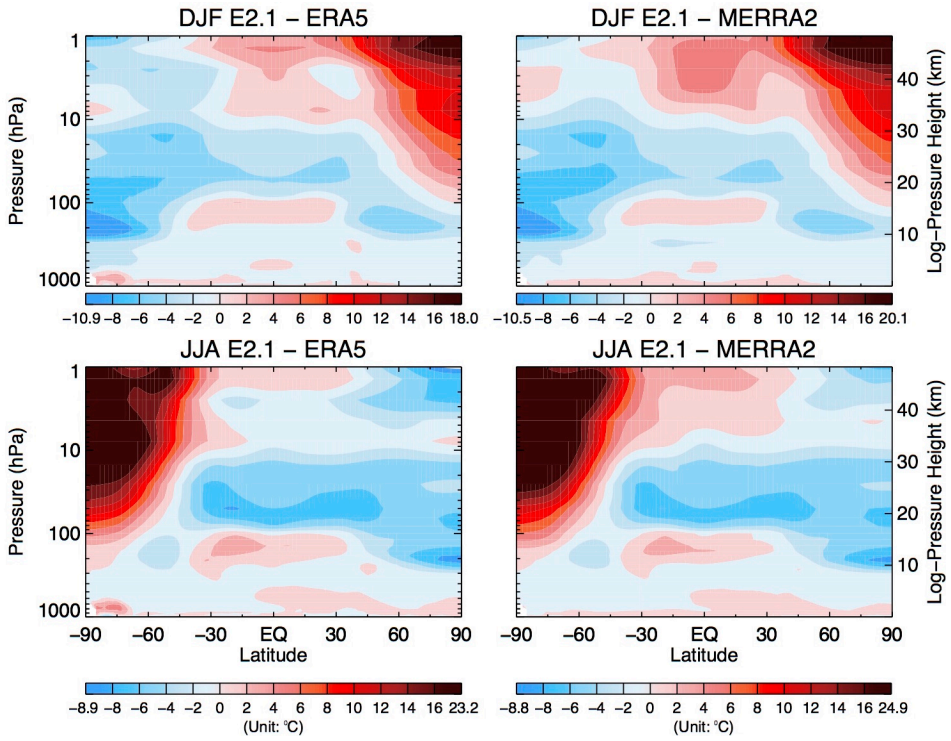

Figure 3. As in Figure 2 but for E2.1.

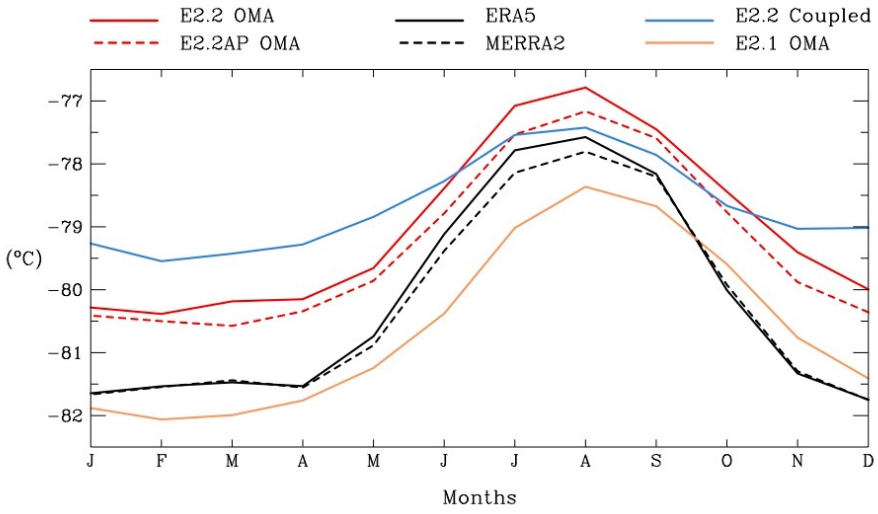

Figure 4. Tropical tropopause cold point temperature averaged from 10N-10S in different E2.2 model simulations, E2.1, and reanalysis data.

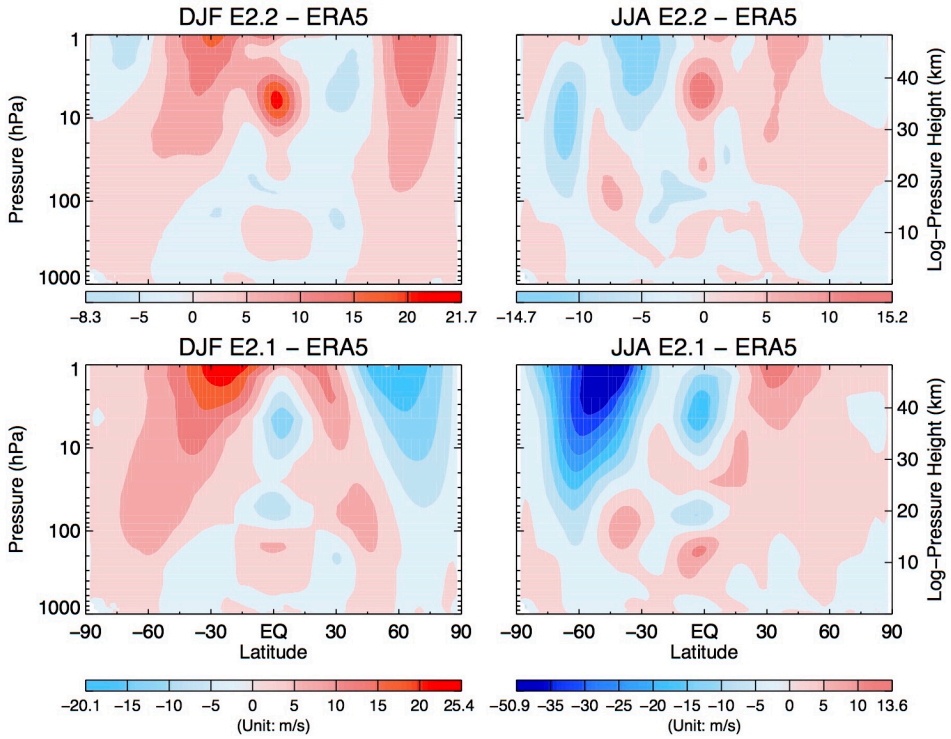

Figure 5. Zonal wind comparisons with ERA5 for E2.2 (top) and E2.1 (bottom), December-February (left) and June-August (right).

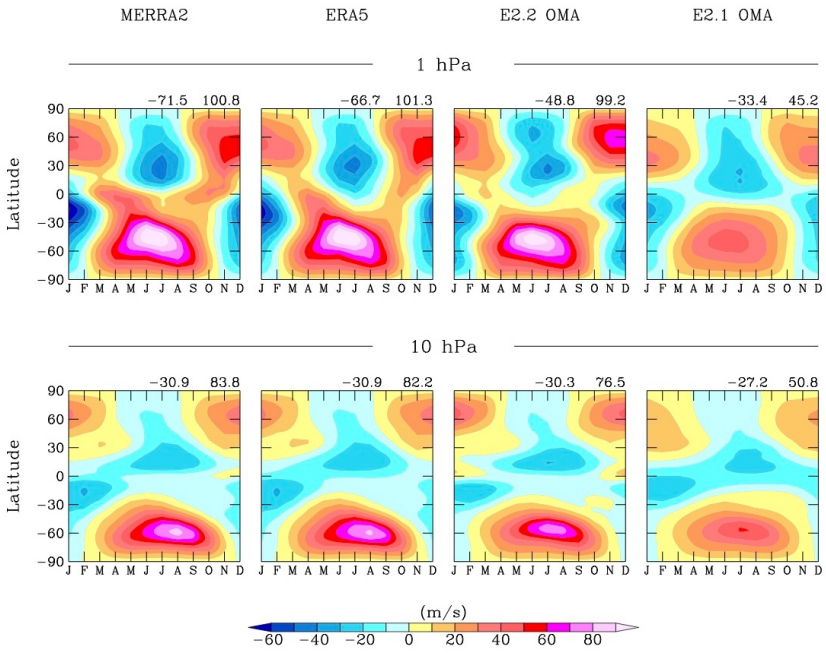

Figure 6. Zonal winds at 1 hPa (top) and 10 hPa (bottom) from the two reanalysis data sets, E2.2 and E2.1. Numbers in the upper right hand corner indicate maximum and minimum values.

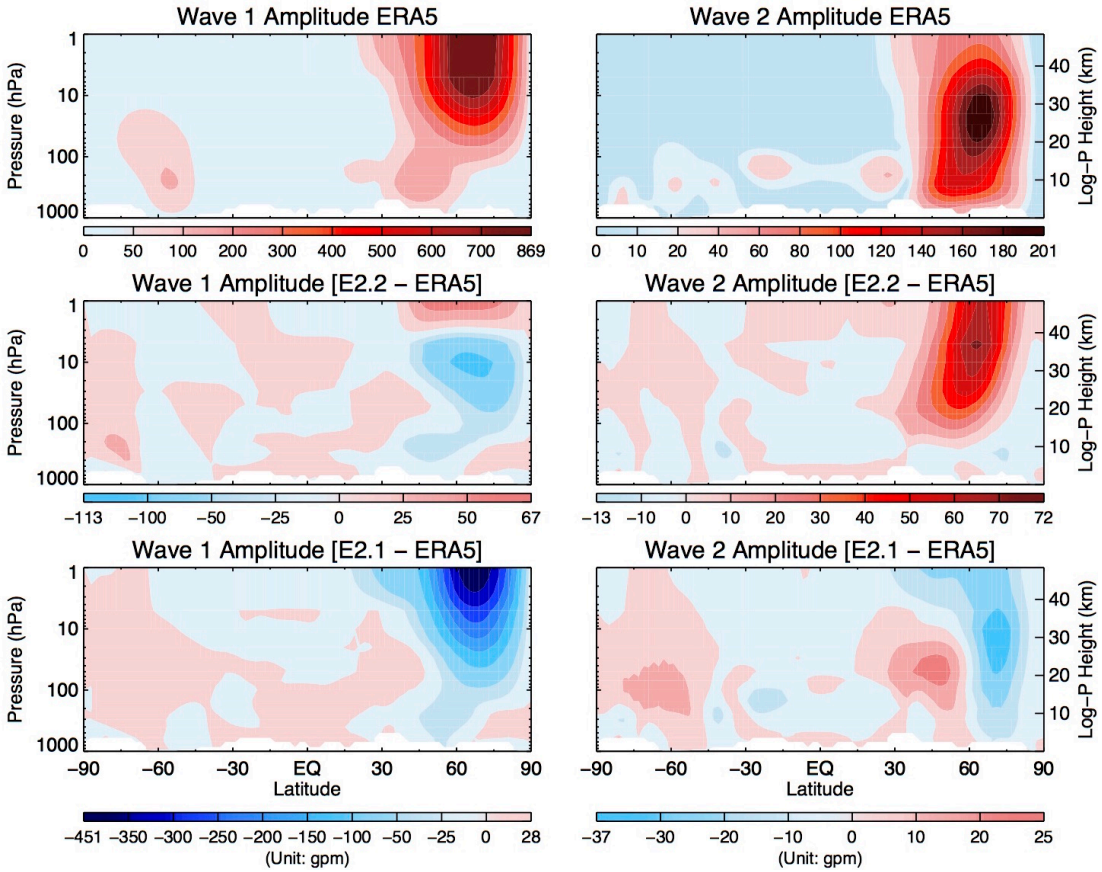

Figure 7. Planetary Wave 1 (left) and Wave 2 (right) amplitude (top row) in ERA5 (note the different scales). Anomalies relative to ERA5 for E2.2 (second row) and E2.1 (third row).

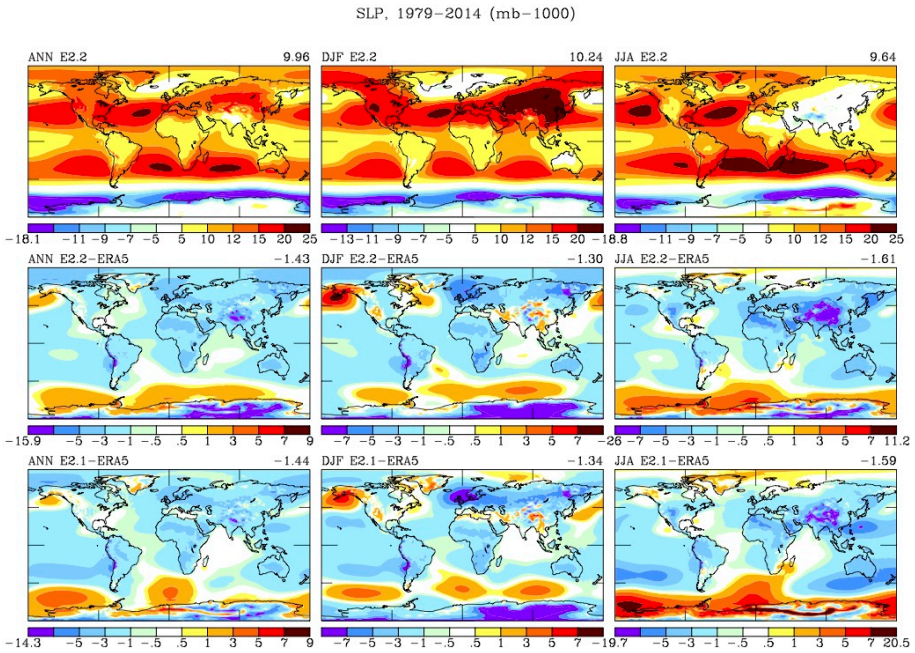

Figure 8. E2.2 Sea level pressure (-1000 hPa) for the annual average (1979-2014), and two solstice seasons (top row), and anomalies compared to ERA5 for E2.2 (middle row) and E2.1 (bottom row). The numbers in the upper right hand corner represent the global average value for the image below, i.e., sea level pressure in the top row, E2.2 and E2.1 global average anomalies in the middle and bottom rows.

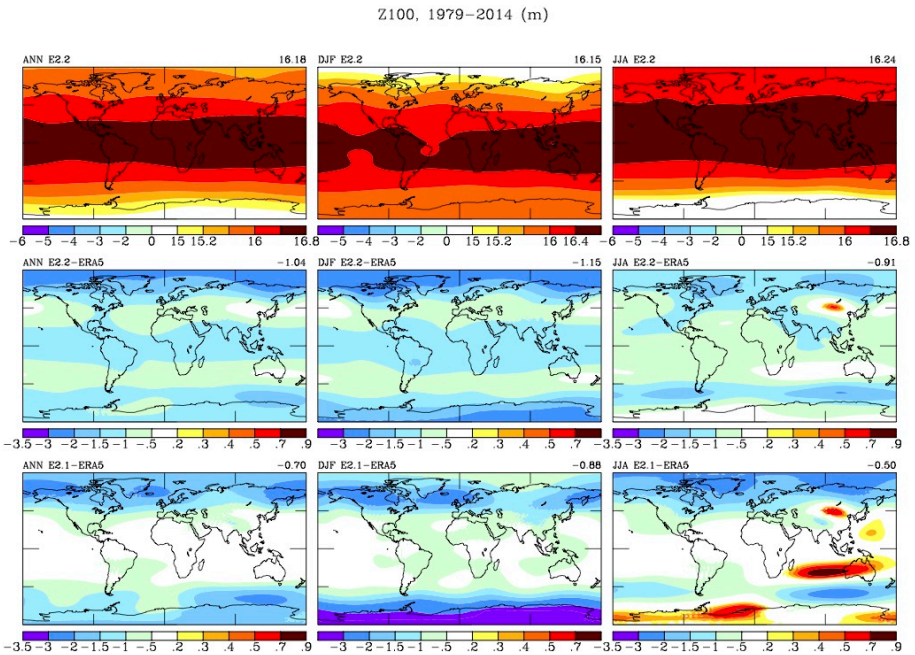

Figure 9. As in Figure 8 but for geopotential height – 16400m at 100 hPa.

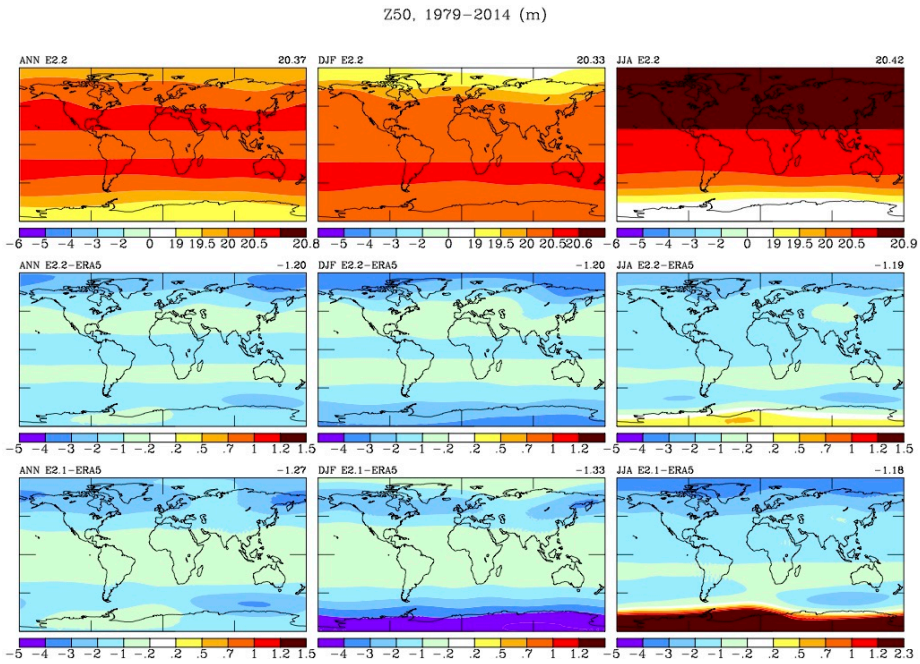

Figure 10. As in Figure 8 but for geopotential height – 20000m at 50 hPa.

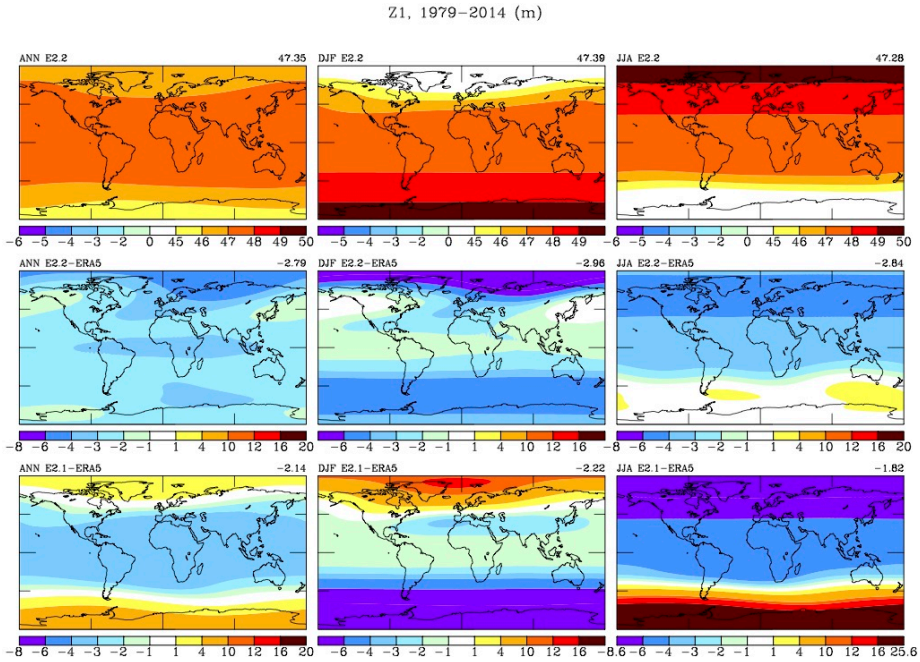

Figure 11. As in figure 8 but for geopotential height – 46000m at 1 hPa.

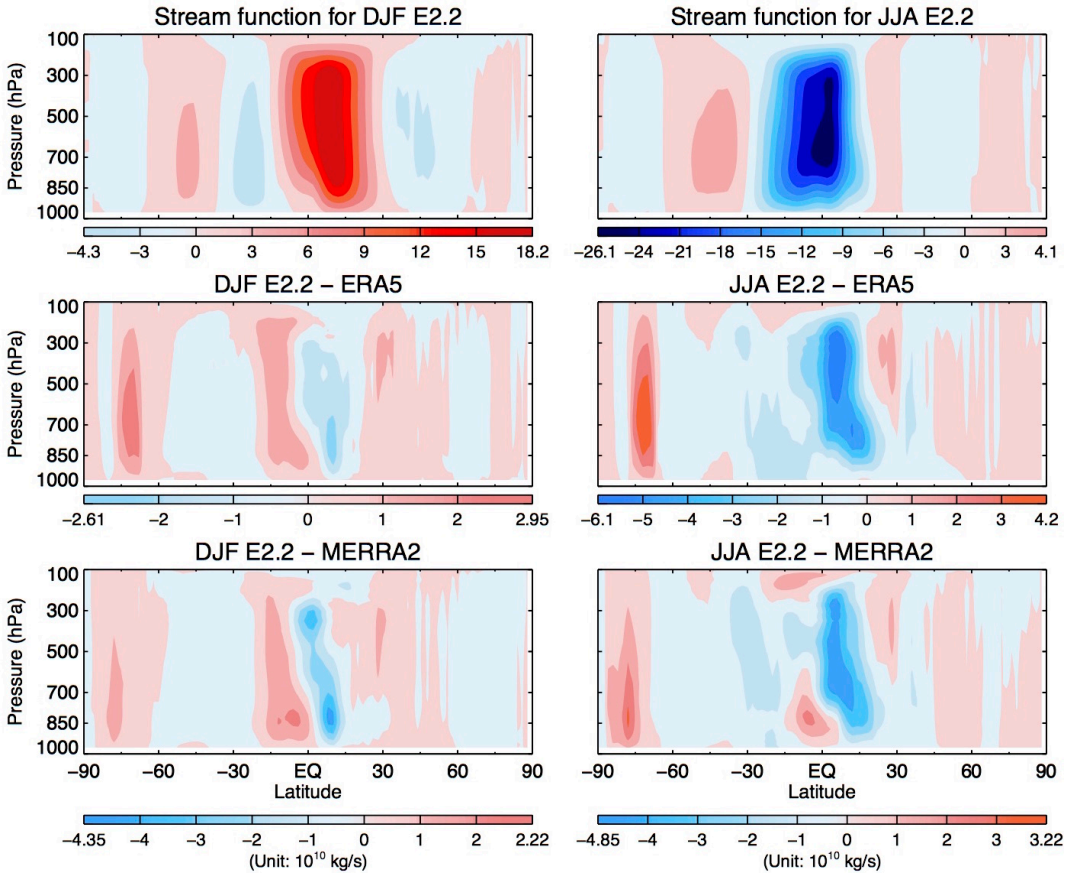

Figure 12. E2.2 Stream function for December-February (left) and June – August (right) (top row). Anomalies compared to ERA5 (middle row) and MERRA2 (bottom row).

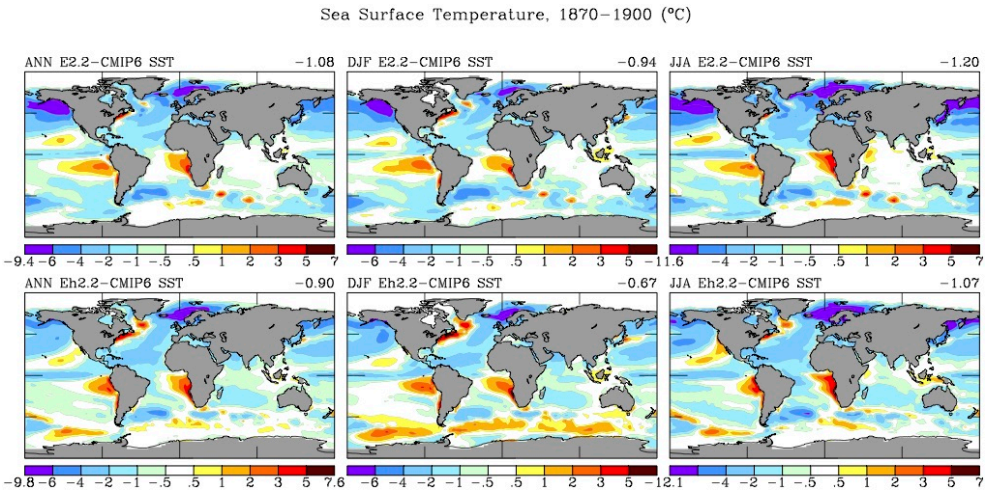

Figure 13. Anomalies in annual, DJF and JJA sea surface temperature for E2.2 coupled to the G model (top) and H model (bottom). Model results are compared with the Hadley Centre reconstruction (from the HADSST-PI) for 1870-1900. Global average anomalies are shown in the upper right hand corner of each panel.

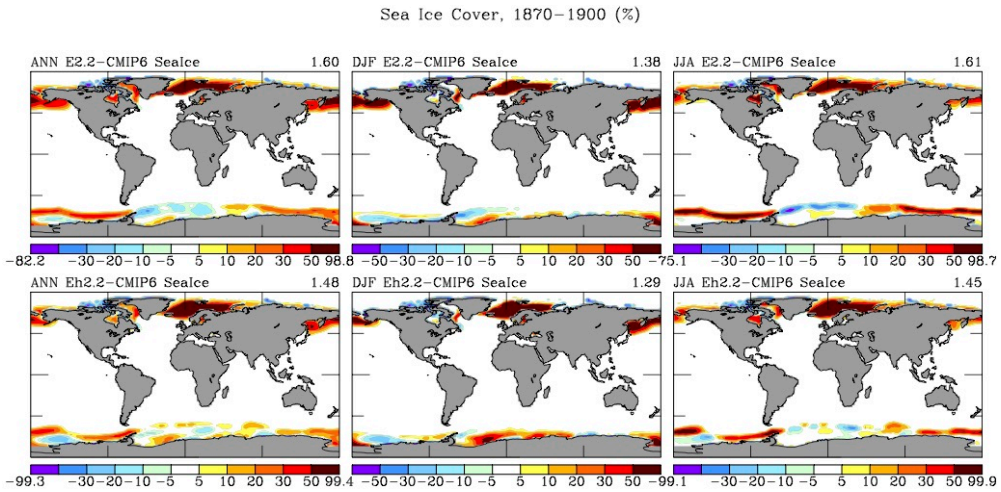

Figure 14. As in Figure 9a but for sea ice cover compared with HadISST-PI (which may be an underestimate for the N.H. as it is only slightly higher than current day values).

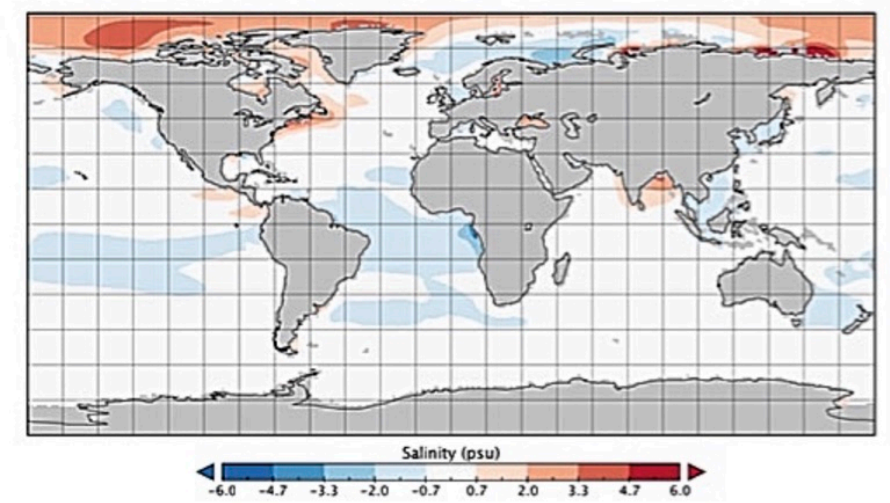

Figure 15. Model PI annual sea surface salinity compared with modern values. Observational data from the WORLD OCEAN ATLAS 2013 version 2 (WOA13). Localized extremes were -6.4 and 10.7 psu, respectively.

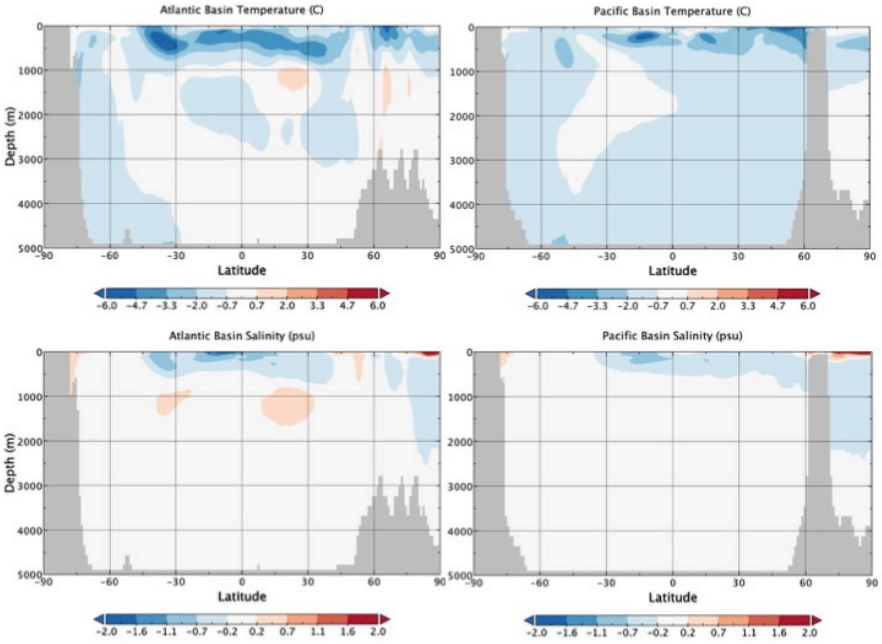

Figure 16. Comparison of model PI simulation with current observations (WOA13) for annual temperature (top) and salinity (bottom) for the Atlantic (left) and the Pacific (right).

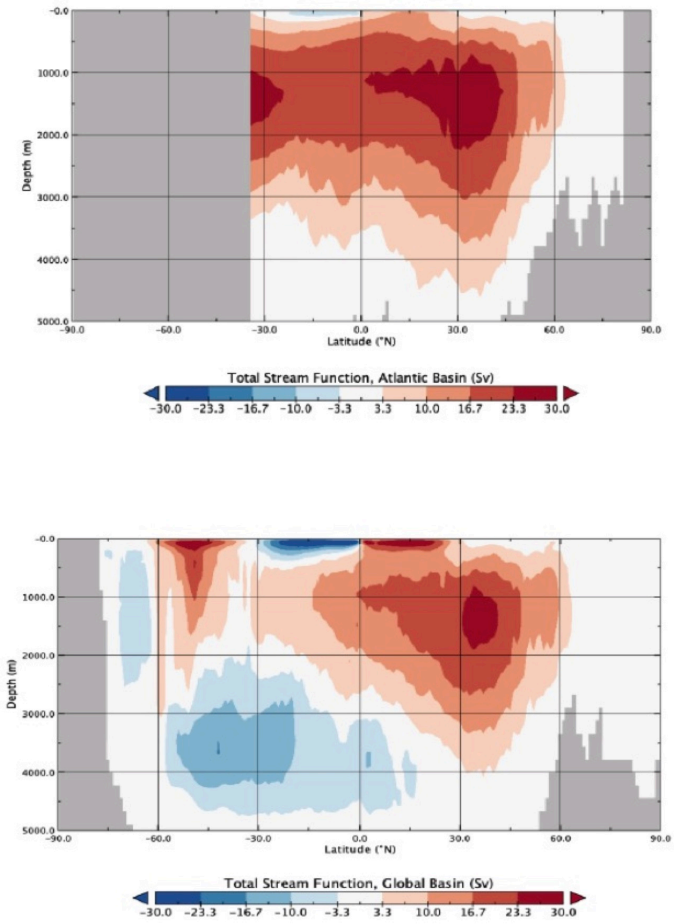

Figure 17. The model stream function for the Atlantic (top) and Global (bottom). Positive values indicate clockwise circulation within the frame of the figure.

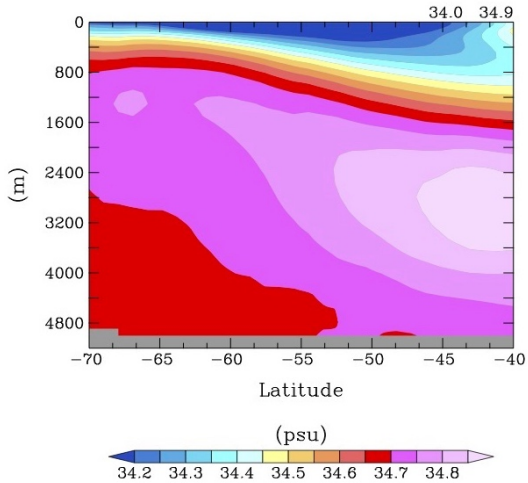

Figure 18. The zonally averaged ocean salinity in the Southern Hemisphere extratropics in model version E2.2, PI conditions.

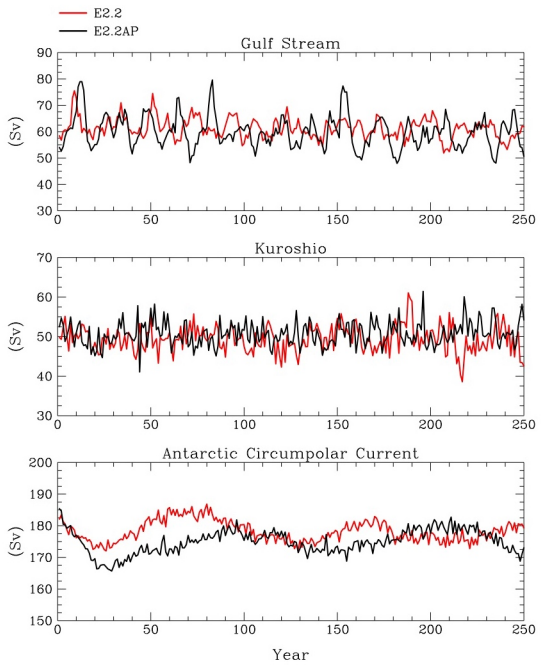

Figure 19. Ocean currents in E2.2 (Gulf Stream (top), Kuroshio (middle), Antarctic Circumpolar (bottom).

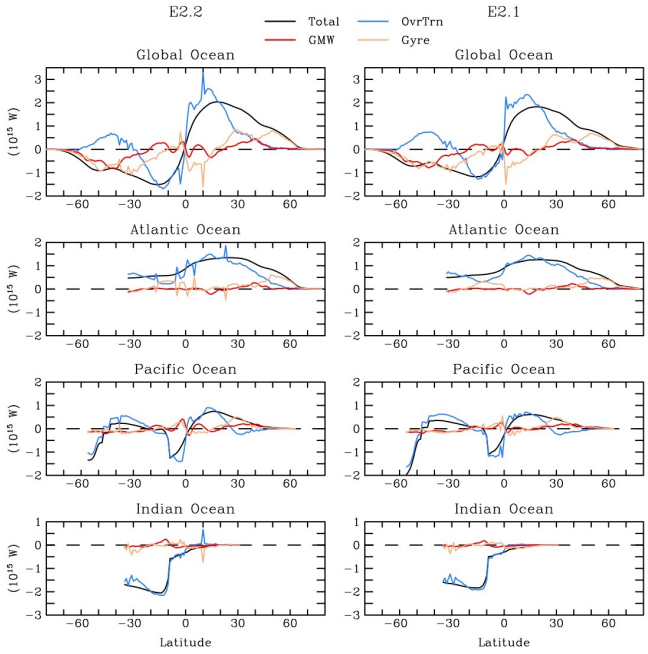

Figure 20. Ocean heat transports in E2.2 (left) and E2.1 (right) for the global average (top), the Atlantic (2nd row), Pacific (3rd row) and Indian (bottom row). The values are broken down into the contributions by the overturning circulation, the gyre, the Gent McWilliams (GMW) parameterization for the effect of mesoscale eddies, and the total.

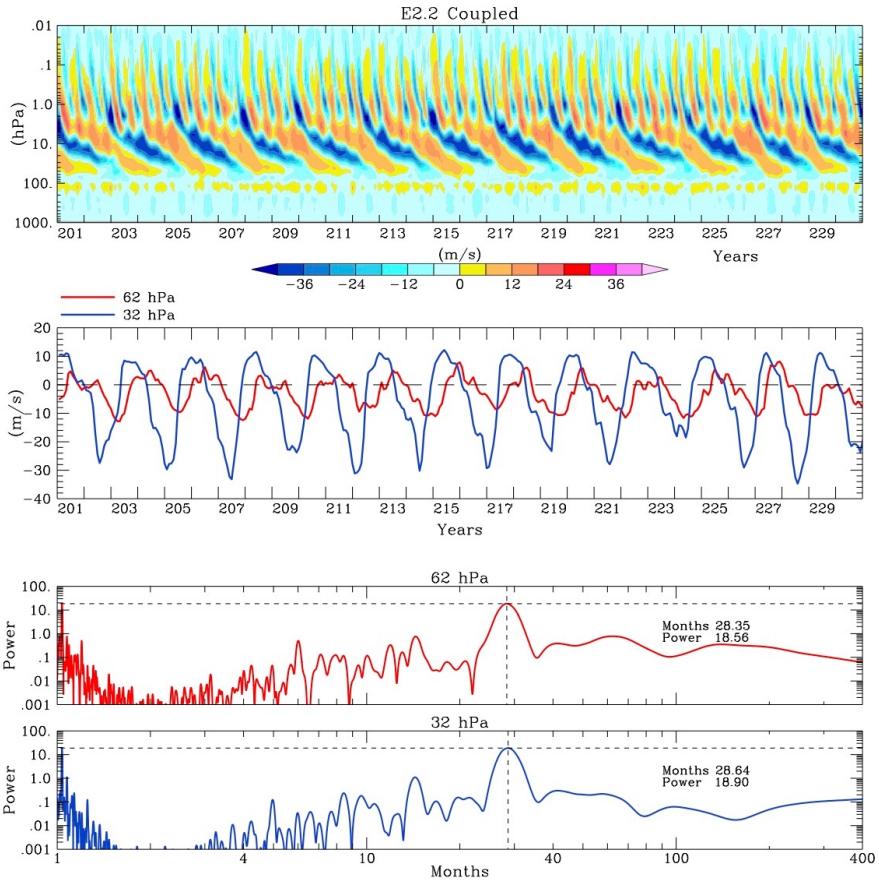

Figure 21. E2.2 coupled model tropical zonal wind (4N-4S) profile (top), at ~30 hPa (blue) and ~60 hPa (red) as a function of time in the PI control (middle) and periodogram of 30 years of those values (bottom). Peak period and month of the periodogram is indicated as text on the figure.

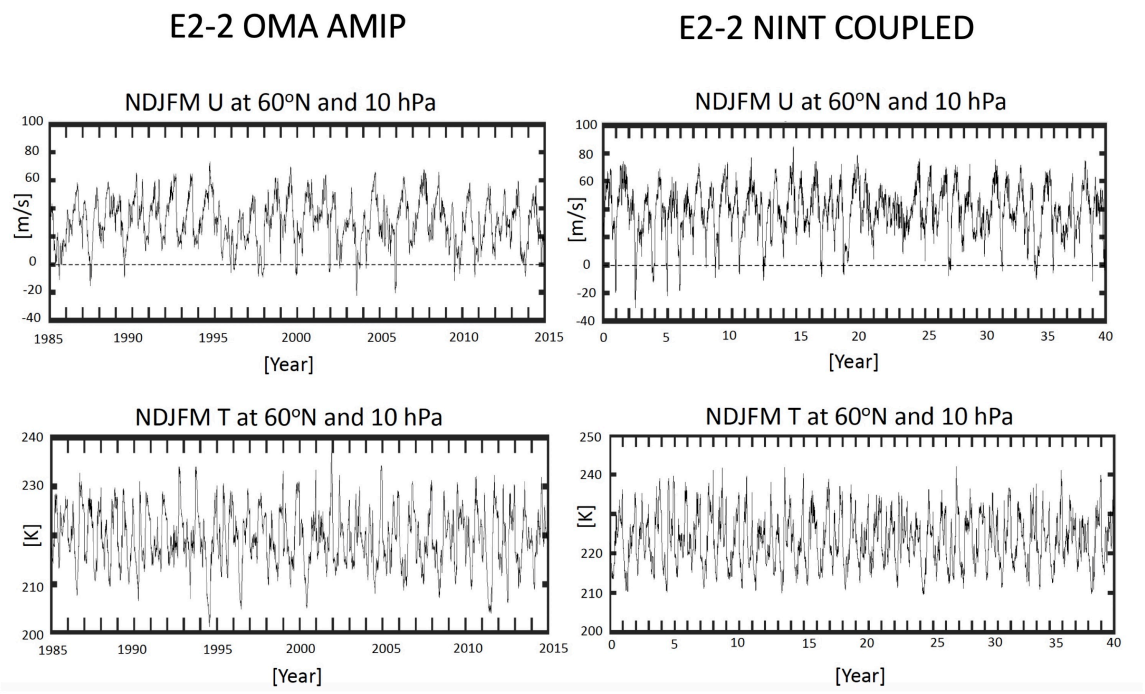

Figure 22. November through March zonally averaged zonal winds (top) and temperatures (bottom) evaluated at 60°N and 10 hPa for one member of the E2-2 OMA AMIP ensemble (left) and the E2.2 NINT Coupled pre-industrial simulation (right). Years correspond to actual model years in the left panels and to years relative to "branching" for the abrupt CO_2 experiments within the underlying control in the right panels. Only forty years are shown in the latter (150-year-long) simulation per the daily data saving protocol for the CMIP6-endorsed Dynamics and Variability Model Intercomparison Project (DynVarMIP) (*Gerber and Manzini (2016)*).

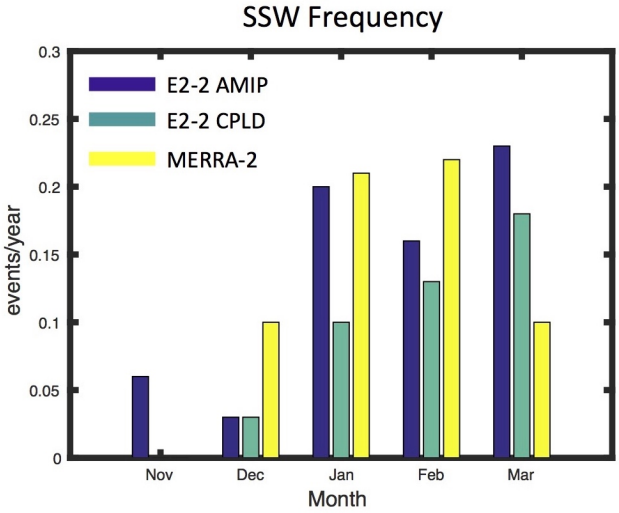

Figure 23. SSW variation with month in the Northern Hemisphere for the E2.2 OMA AMIP and Coupled runs compared with MERRA2 observations.

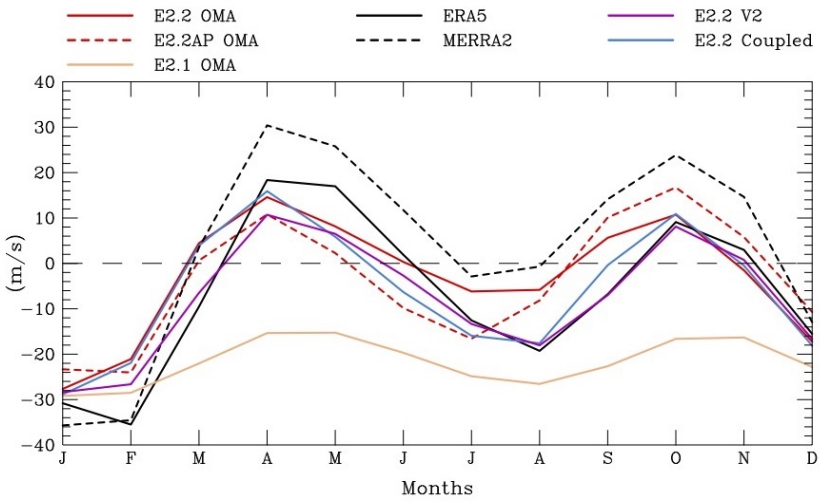

Figure 24. Zonal winds at 2 hPa between 4N and 4S as a function of month, showing the semi-annual wind oscillation (SAO) for the E2.2 models discussed in this paper, as well as E2.1 and reanalysis data.

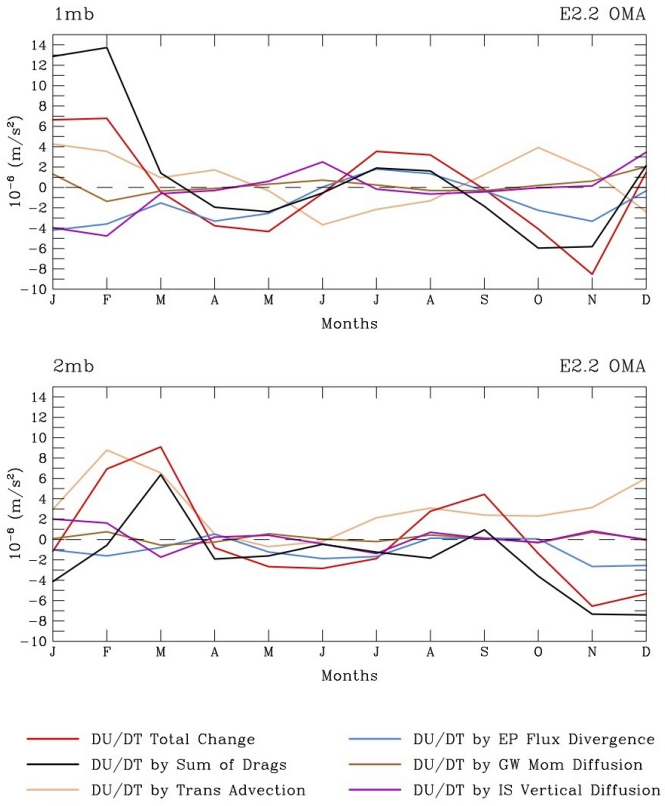

Figure 25. Tropical zonal wind changes induced by various mechanisms at 1 hPa (top) and 2 hPa (bottom) for E2.2 OMA. See text for the explanation of the different contributions.

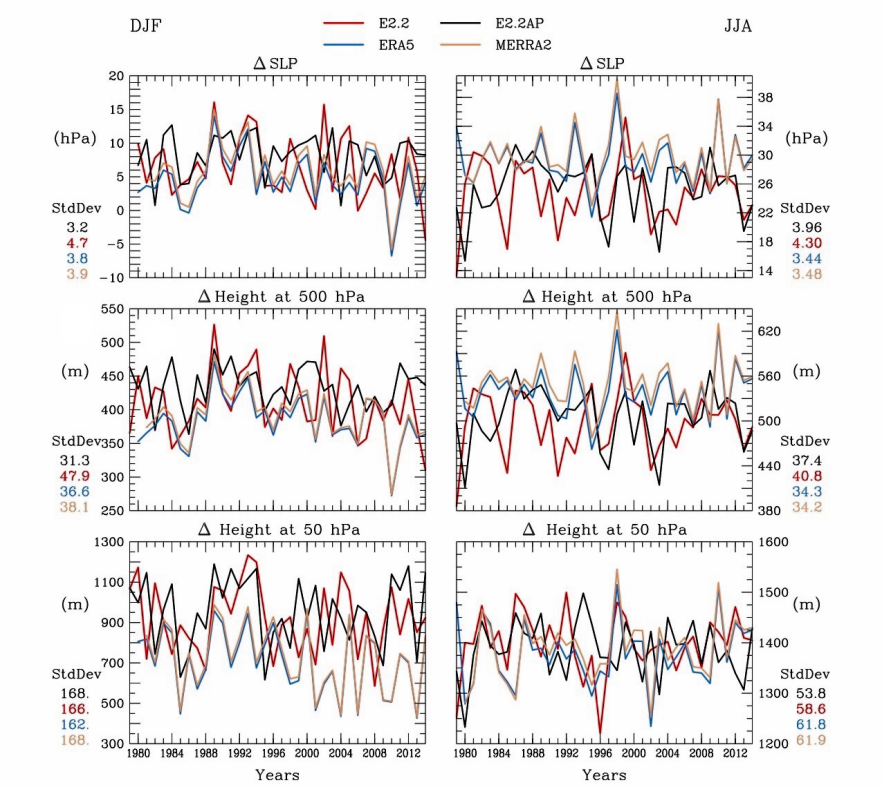

Figure 26. December – February Index related to the Northern Annular Mode (left), defined as (30-50N) minus (60-80N), for sea level pressure (top), 500 hPa geopotential heights (middle) and 50 hPa geopotential height (bottom) for E2.2, E2.2AP, ERA5 and MERRA2. The standard deviations for each model and observational data set are given in the lower left hand corner of each panel. Similarly, an index related to the Southern Annular Mode is shown (right), 40°S minus 65°S averaged for June-August, with the standard deviations given in the lower right-hand corner.

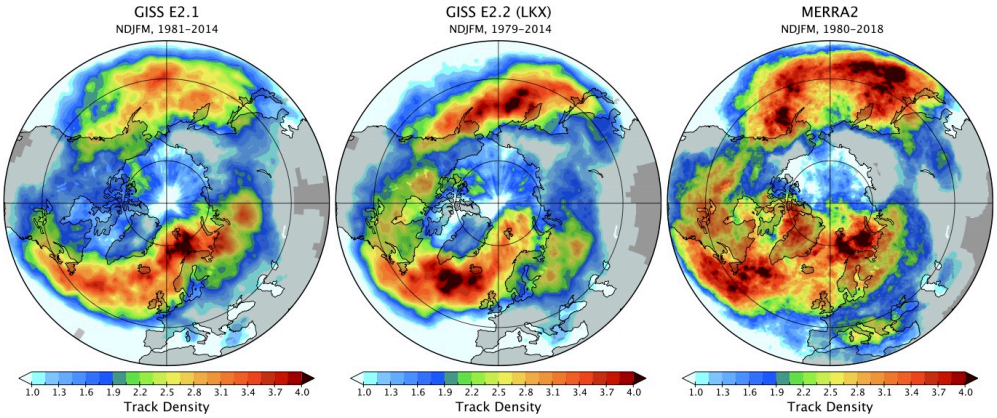

Figure 27. Storm track density for November-March, ~1980-2014 in the OMA runs of E2.1 and E2.2 as well as observations from MERRA2 (1980-2018).

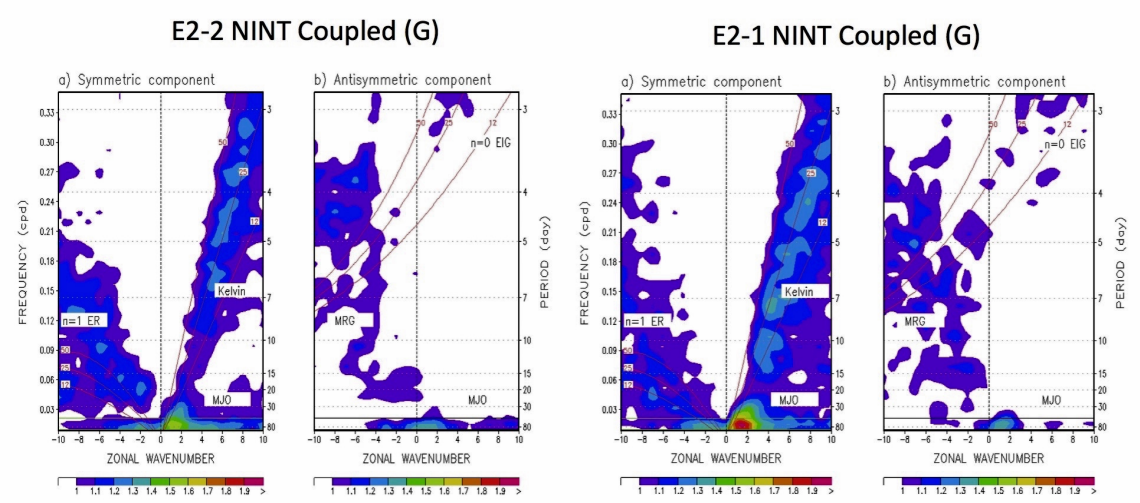

Figure 28. Zonal wave-number frequency power spectra about the equator of precipitation. Shown are the results for E2.2AP (left) and E2.1 (right) indicating the MJO and power in various tropical waves. The horizontal line depicts the 60 day period; power above that line in wave numbers 1-3 is indicative of the MJO. Contours are unitless, i.e., the tropical mean symmetric or antisymmetric component of precipitation divided by the background spectrum.

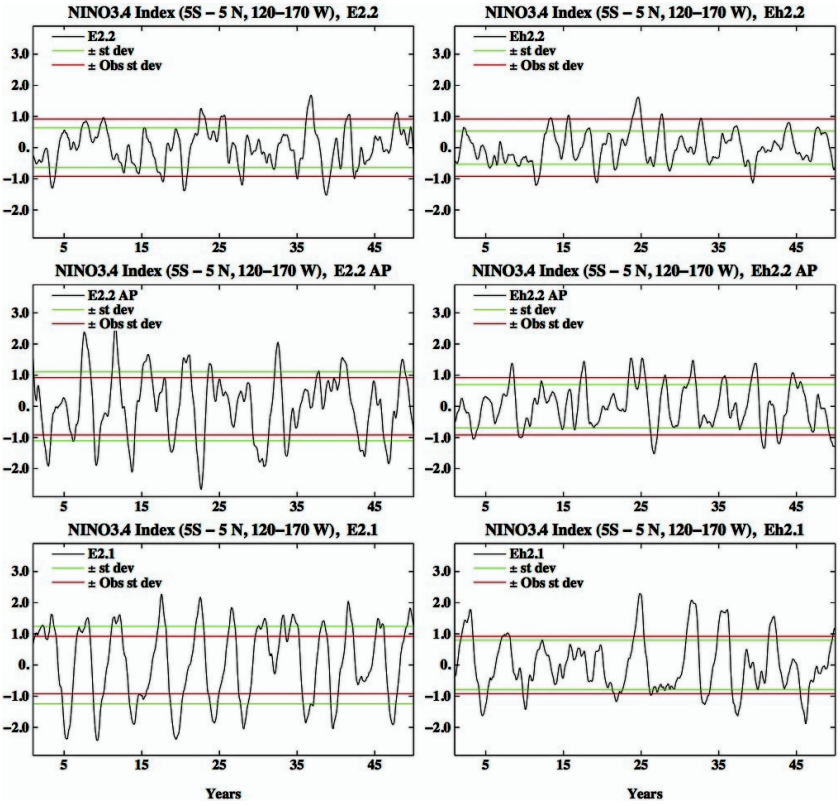

FIgure 29. NINO3.4 amplitude in E2.2, E2.2AP and E2.1 with the G ocean model (left) and H ocean model (right). Model standard deviation is given in green, observed in red.

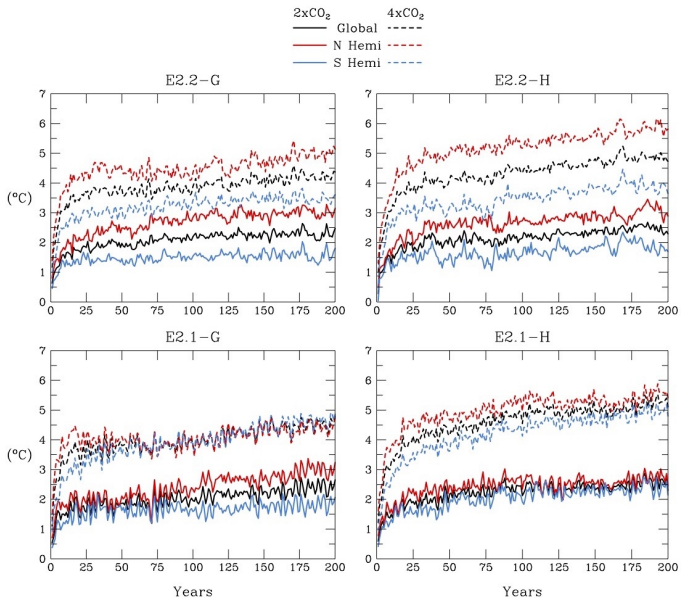

Figure 30. Surface air temperature change with time for instantaneous $2xCO_2$ (solid lines) and $4xCO_2$ (dashed lines) forcing with the G model (left) and H model (right) for E2.2 (top) and E2.1 (bottom).

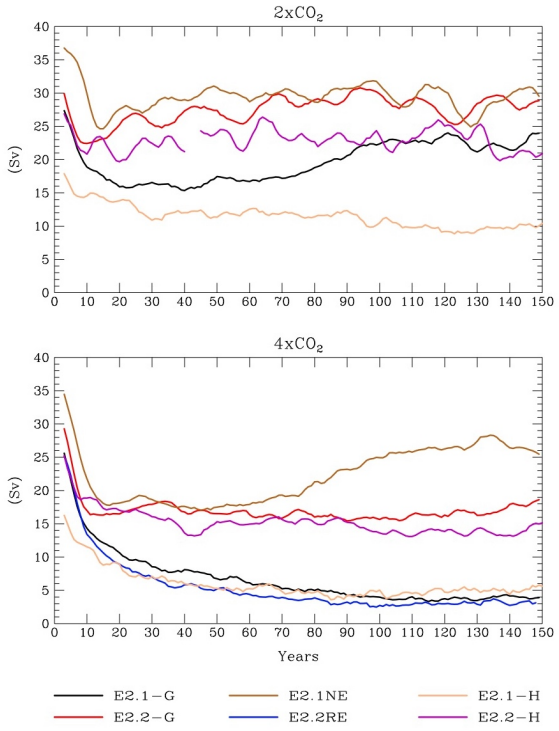

Figure 31. Maximum AMOC value in response to instantaneous $2xCO_2$ (top) and $4xCO_2$ (bottom) forcings for E2.2 and E2.1 experiments with the G and H ocean models. Results also shown for additional experiments (E2.2RE and E2.1NE) described later in the text. All values are five year running means.

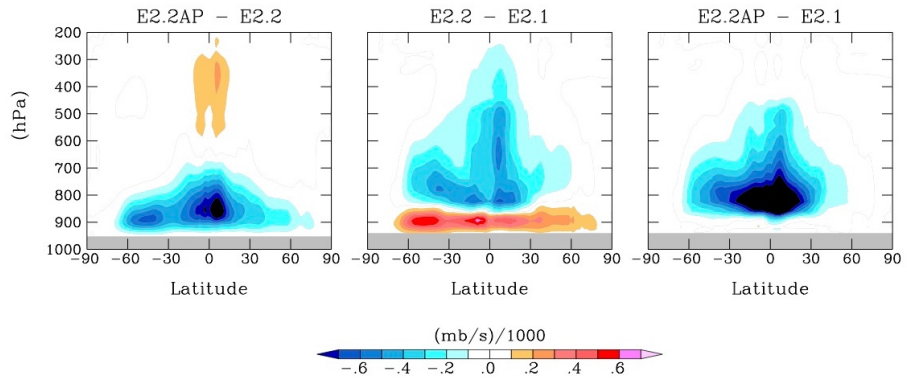

Figure 32. Difference in convective mass flux for E2.2AP - E2.2 (left), E2.2 - E2.1 (middle) and E2.2AP – E2.1 (right) for the 36 year annual averages in the OMA AMIP runs.

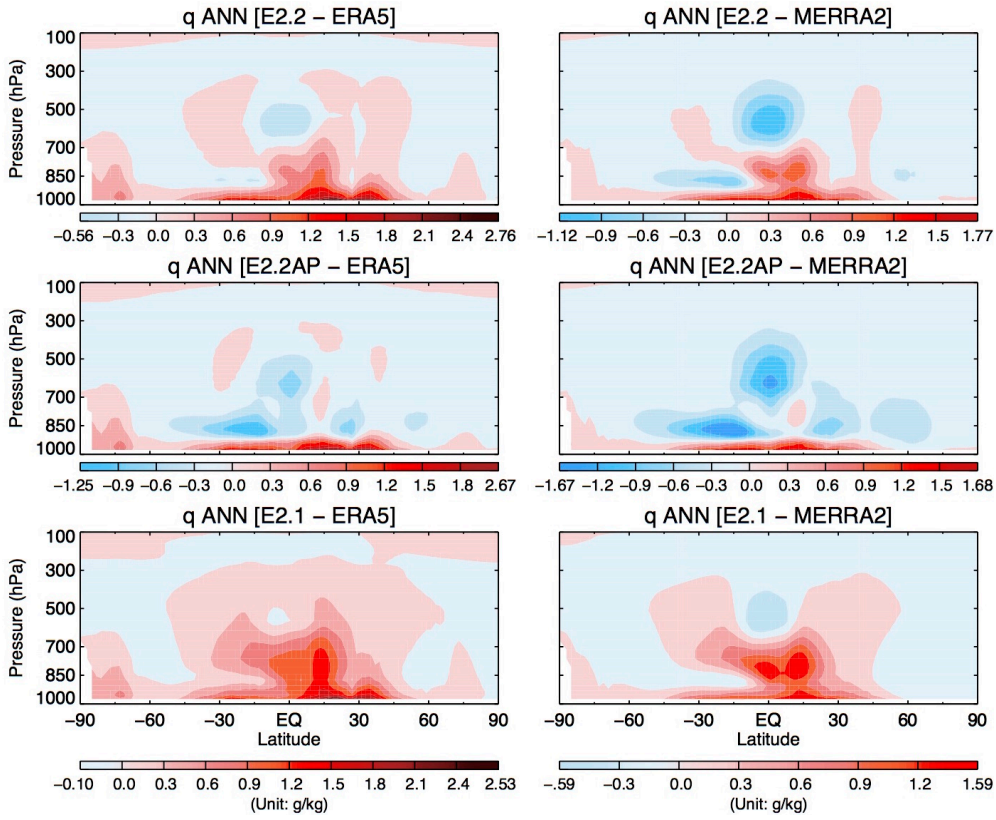

Figure 33. Annual specific humidity anomalies for E2.2 (top), E2.2AP (middle) and E2.1 (bottom) compared with ERA5 (left) and MERRA2 (right).

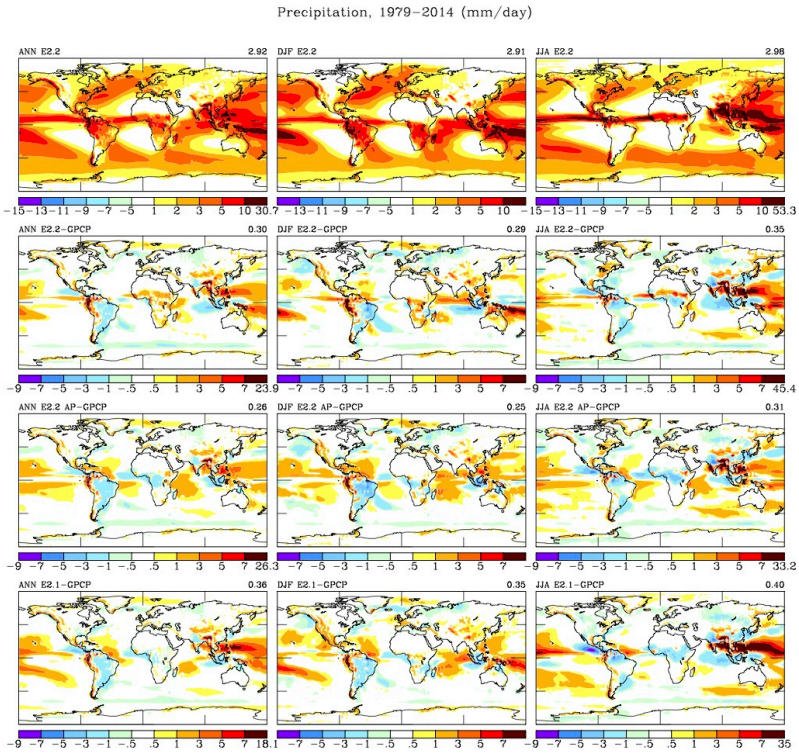

Figure 34. Precipitation for the E2.2 OMA AMIP model (top row) on the annual average (left), December-February (middle) and June-August (right). The following three rows show the precipitation anomaly relative to GPCP for E2.2, E2.2AP and E2.1. Global average values are given in the upper right hand corner, for either precipitation or the anomaly.

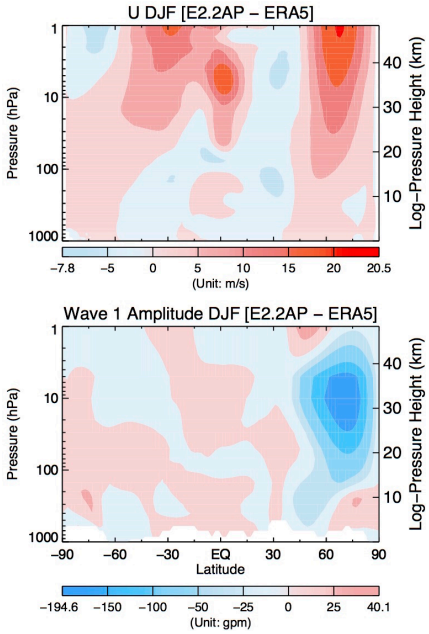

Figure 35. Anomalies of zonal wind and Wave 1 amplitude in E2.2AP relative to ERA5 in the AMIP NINT simulation.

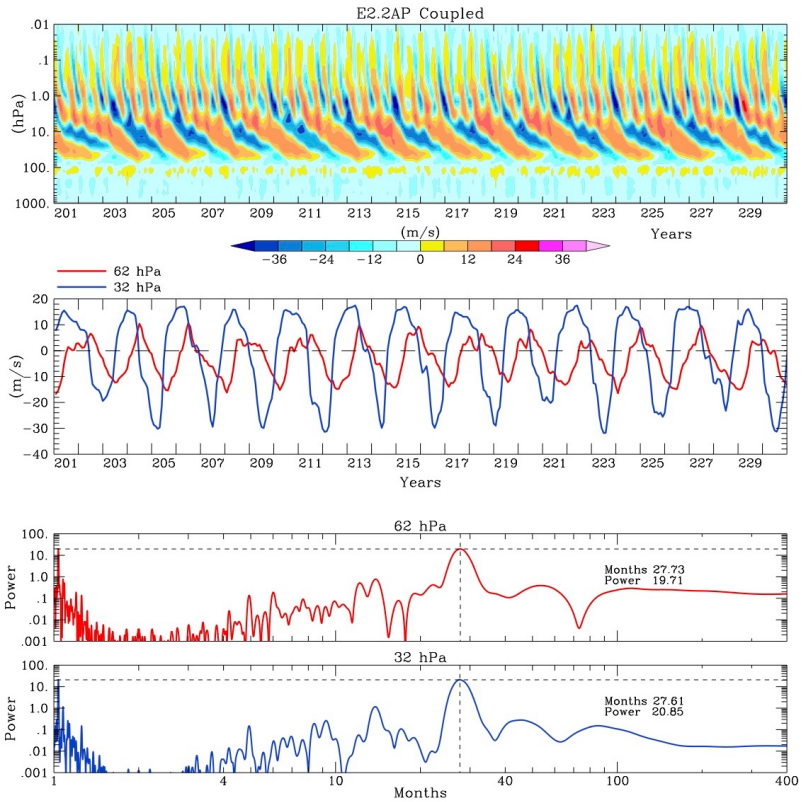

Figure 36. QBO results as in Figure 21 but for E2.2AP. Peak month and power of the periodogram is indicated on the figure.

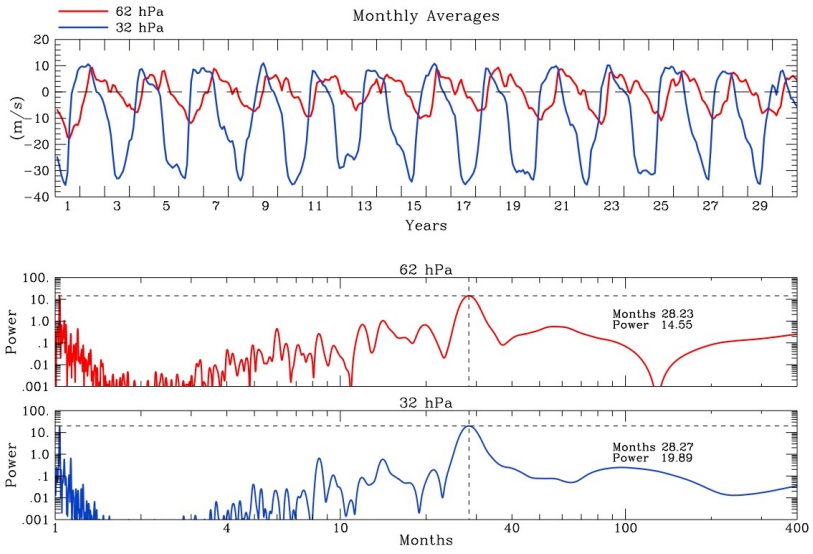

Figure 37. QBO generated with AMIP NINT VERSION 2, amplitude of tropical zonal wind (top) (4N-4S) and periodogram (bottom). Text on the figure indicates period and magnitude of peak power.