UAB
Universitat Autònoma de Barcelona

Master's Degree Dissertation

MSc in Modelling for Science and Engineering

Understanding recent changes in the Northern Hemisphere summer circulation and its relation to ocean temperature variations

Gerard Marcet Carbonell

Supervisors
Markus G. Donat & Carlos Delgado Torres

Course
2023/24

Call
July

Abstract

The summer climate on the Northern Hemisphere in recent decades has shown distinct trend patterns, with warming hotspots that spatially match with the ridges of a circumpolar atmospheric wave pattern. The drivers behind this wave-like trend and warming pattern are not yet well understood. By applying constraints of internal climate variability to large ensembles of climate model simulations from the Coupled Model Intercomparison Project phase 6 (CMIP6), we explore to what extent sea surface temperatures in different regions contribute to recent trends towards more stationary Rossby waves, e.g. with wave number 5, during summer in the Northern hemisphere. Our results found no indication of internal variability of sea surface temperatures being the main driver behind the observed trend.

To all of you that make my day to day a more lively place.
And a profound thank you to Markus G. Donat and Carlos Delgado Torres for your relentless guidance and help through this project.

Contents

1 Introduction

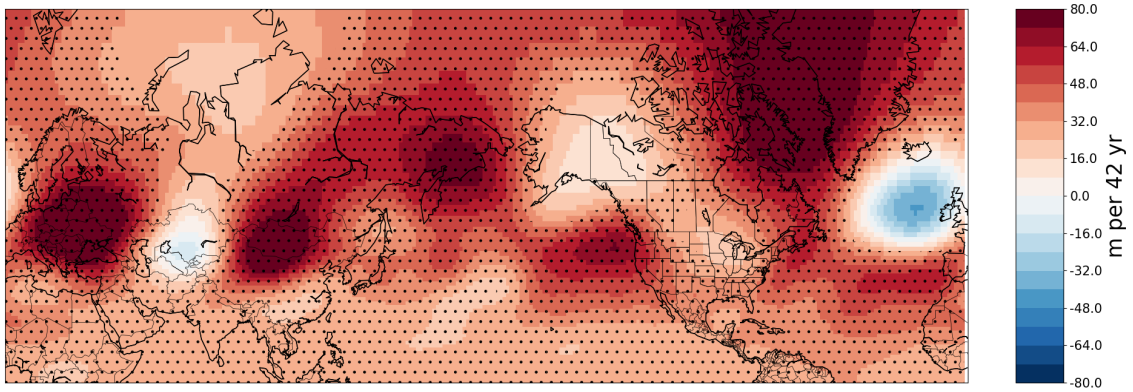
The climate of the Earth has experienced drastic changes during recent decades. The human influence on the climate, primarily by burning fossil fuels, is now unequivocal (IPCC [2021]). This causes an overall global warming which affects almost all regions globally, and also causes changes in the hydrological cycle and other climate variables including including changes in the frequency and intensity of extreme events. In addition to the long-term warming caused by human activities, the climate is also affected and modulated by natural variations caused by the interactions between the different components of the climate system. Disentangling and quantifying the specific roles of anthropogenic influences and natural climate variations in observed climate features still remains a major challenge.

Differences in radiation imbalances from incoming radiation from the Sun (which is more intense along the equator and fainter in the poles) and outgoing radiation from the Earth (which is more uniformly distributed) result in heat excess at lower latitudes and a deficit at high latitudes. The large-scale air movement associated to this imbalance is what is known as atmospheric circulation. This drives convective circulation cells as warm air rises and flows poleward where it is cooled, sinks and is driven towards lower latitudes. These cells are known as Hadley or direct cells and do not span from the equator to the poles as initially theorized since conservation of angular momentum would require very high winds at mid-latitudes to satisfy. Instead what is observed is that there are two direct convection cells, at the equator (0^{o}-30^{o}) and at the poles (60^{o}-90^{o}). However, at mid-latitudes (30^{o}-60^{o}) convection cells moving in opposite direction (known as Ferrel or indirect cells) are observed instead. In these mid-latitude regions heat transfer is not driven by meridional circulation and instead a series of coherent vortex systems, known as cyclones and anticyclones, drive heat vertically by the vortex transferring hot air to the poles on one side and cold air to the equator on the other (Cushman-Roisin and Beckers [2011]).

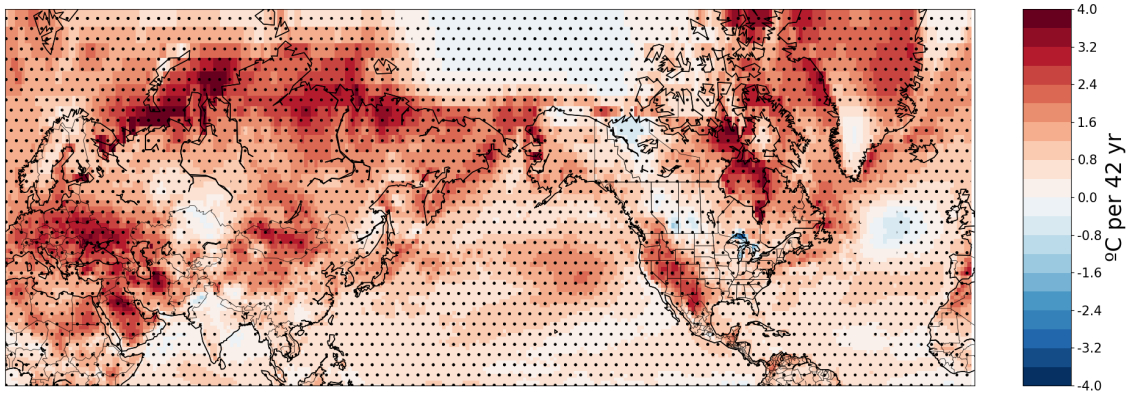
Understanding atmospheric circulation is important as it is a key driver of surface climate and is associated with regional weather conditions and extreme events. For instance, the "Dust Bowl", a devastating drought affecting the U.S., Mexico and part of Canada, was driven by a change in atmospheric circulation caused by anomalies in sea-surface temperatures (SST) (Schubert et al. [2004]). Increased hurricane activity in the Atlantic during 1995 to 2000 is linked to an increase in tropical Atlantic SST related to overlying tropospheric circulation (Goldenberg et al. [2001]). Also, atmospheric circulation has been identified as the primary influence of UK surface climate (Fereday and Knight [2023]). Despite its importance, atmospheric circulation is difficult to predict and is a major source of uncertainty in climate projections due to its chaotic nature and errors in modeling the circulation response from climate forcing (Shepherd [2014]).

The challenges to understand changes in atmospheric circulation are also highlighted in a recent publication by Teng et al. [2022] where a non-homogeneous warming trend in the Northern Hemisphere boreal summer months (JJA) coinciding with a similar trend in 200 hPa geopotential height (Z200) is reported as shown in figure 1. Geopotential height is defined as the work needed to move a mass of 1 kg from sea level to a given point divided by the globally averaged gravity acceleration at surface. It is used as vertical coordinate for any application where energy plays a role, such as atmospheric circulation (Wallace and Hobbs [2006]). With Z200 we refer to the geopotential height of a given pressure level (in this case 200 hPa), from which the horizontal pressure gradient force is inferred. The pattern observed in the trend resembles that of a circumglobal wavenumber-5 Rossby wave. Rossby (or planetary) waves are a type of wave present in stationary geostrophic flow (that is, fluids where the pressure gradient and Coriolis force are in equilibrium) characterized by low frequencies and wavenumbers (Cushman-Roisin and Beckers [2011]). Stationary Rossby waves in summer have been linked to heatwaves and droughts in Eurasia (Schubert et al. [2014]) as well as North America and the Caspian sea region (Kornhuber et al. [2019]) and may threaten global food production due to simultaneous heat extremes in different regions accounting for a large fraction of global food production (Kornhuber et al. [2020]). Teng et al. [2022] has also linked the recent trend in geopotential height during JJA with several extreme events during summer such as the Russian heatwaves and floods in Pakistan in July 2010 (Dole et al. [2011], Trenberth and Fasullo [2012]) and the intense heatwaves over Europe and the U. S. during August 2017.

The observed trend is distinct from the results of the multi-model experiments of phase 5 (Taylor [2012]) and 6 (Eyring et al. [2016]) of the Coupled Model Intercomparison Project (CMIP) as



(a) Z200



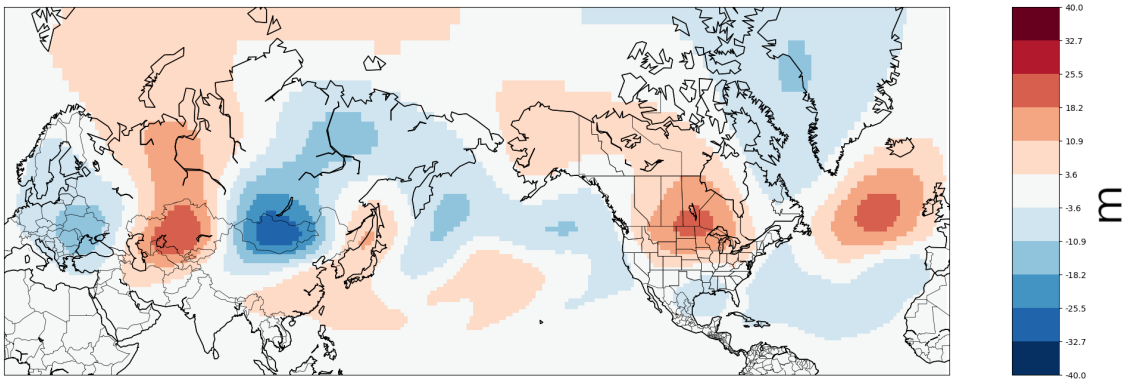
(b) TAS

Figure 1: Z200 and near-surface air temperature (TAS) summer trends from 1979 to 2021 from the ERA5 reanalysis. Stippling indicates values statistically significant at the 90% confidence level using a Student-t test.

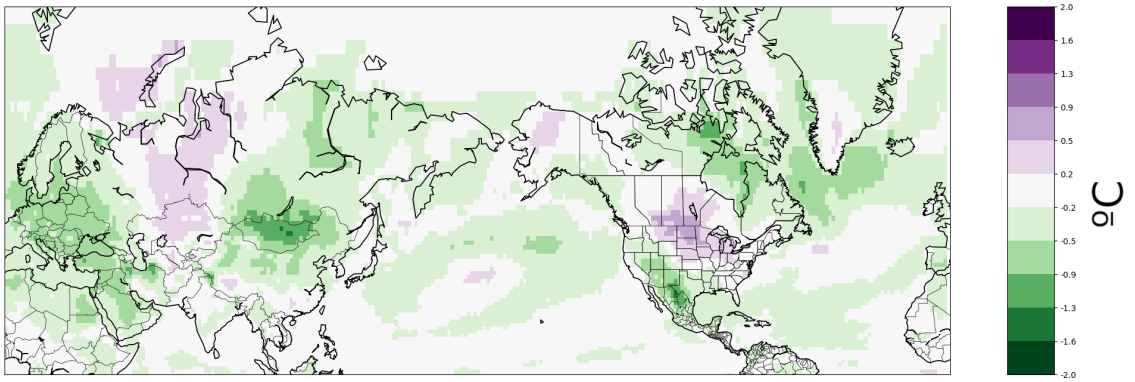
reported in Teng et al. [2022]. This could represent flaws in our atmospheric circulation modelling efforts and understanding of the underlying physical drivers of this specific trend. In Teng et al. [2022] an association with modes of low-frequency internal variability was reported, specifically Interdecadal Pacific oscillation (IPO)Power et al. [1999], and Atlantic multidecadal variability (AMV) Knight et al. [2005]. However, this finding was based on statistical regression analysis, which is undermined by the low frequency of these events, with only two AMV cycles in instrumental record. Additionally, the limitations of the models in replicating these modes of variability as well as large model-to-model variations in AMV teleconnection patterns pose a challenge in determining the significance of these associations.

In Teng et al. [2022] the Atmospheric Model Intercomparison Project (AMIP) Gates et al. [1999] runs were able to better capture the observed trend than the CMIP6 historical simulations. AMIP runs are simulations forced by observed SST and sea-ice variations as well as natural and anthropogenic external forcings which could suggest that SST plays a role in the formation of the trend as pointed out by Teng et al. [2022]. However, AMIP simulations do not represent coupling between ocean and atmosphere while CMIP6 historical simulations (Eyring et al. [2016]) are fully coupled ocean-atmosphere runs, which could undermine the association. Further evidence of a link between surface temperatures and the observed wave-5 pattern is the pattern correlation between Z200 and near-surface air temperature (TAS) anomalies when the wave-5 pattern is strongest as observed in figure 2.

Understanding what drives this recent multi-decadal trend in geopotential height is important in order to assess if we expect a continuation of this trend, as if the main driver is anthropogenic forcing an increase in extreme weather events in the following years would be expected. Otherwise,



(a) Z200



(b) TAS

Figure 2: 10-year averages of TAS and Z200 anomalies from 1982 to 1991 with respect to climatology (1980-2010).

if the trends are the result of a multi-decadal variability, a reversal of the current trend might happen in the future. This in turn can allow us to better predict and prepare for the consequences of the wave-5 circulation pattern.

With this work we aim to learn about the mechanisms underlying the observed circulation pattern. More specifically we are interested in determining whether the historical forcings or historical SST variations have driven or contributed to the observed changes in the Northern Hemisphere atmospheric circulation during summer.

The rest of the thesis is structured as follows: In section 2 we explain how the forcing response of the models is obtained and the employed method to constrain internal variability. The datasets and models used are also shown in this section, as well as the tools used to determine the ability of the models in reproducing the observed trends. In section 3 we show the results of constraining internal variability and discuss if there is a link between SST variability and the observed trend. Finally, in section 4 we summarize the findings of our research.

2 Data and Methods

To understand and properly attribute the drivers behind the observed trend in atmospheric circulation, methods to disentangle the contributions from different sources are needed. This study is concerned with understanding the the roles of historical forcings (e.g. greenhouse gases and aerosols) on the one hand, and of ocean temperature variability on the other hand, in the formation of the observed wave-5 pattern in Z200. We do so by studying the models' response to external forcings as well as their response to SST internal variability combined with forcing. Large model ensembles provide a way to study the response to given forcings in the climate, as by having an ensemble of uninitialised members produced with different global climate models, by taking the

mean, the internal variations of the system are averaged out and only the forced response remains (Milinski et al. [2020]).

2.1 Constraining internal variability

In order to study whether trends in geopotential height (or atmospheric circulation) are driven by the variability of Sea Surface Temperatures we constrained the global variability patterns in the climate system. Our approach is based on the "poor man's" initialization scheme proposed by Mahmood et al. [2022]. The original work developed a method to constrain internal variability based on member selection of a large ensemble instead of initializing the climate state of the system by phasing the model simulations with the observed variability, which can lead to issues such as initialization shocks or drifts (Smith et al. [2013]).

The "poor man's" initialization protocol works in as follows. Given a large ensemble of members from different models, such members are ranked every year based on their similarity to the observed climate state. Specifically, the ranking is based on the pattern correlation between the simulated and observed SST anomalies averaged over 9 years. This ranking can be performed on the entire globe or can target specific oceans or regions, for example, in the original publication global, North Atlantic and Pacific constraints are studied. To do so, they used an ensemble of 212 members from 32 models of the Coupled Model Intercomparison Project phase 6 simulations and selected the best 30 members for each year. In their work they perform a more thorough breakdown on the effect of number of members selected and years averaged but different selection year averages showcase similar patterns with longer periods (9 years) being more statistically significant. Regarding the number of members used, the results appeared robust with 10, 30 and 50 member ensembles showing added skill in similar regions. The constrained ensembles showed increased skill in forecasting SST, near-surface air temperature, sea level pressure and annual precipitation.

Since our work is unconcerned with forecasting but rather studying the physical drivers behind the observed trend in geopotential height, our implementation of the scheme is the following. For a given year, the best 30 members according to a given constrain region are selected based on the spatial agreement between the simulated and observed SST anomalies over the last 9 years (including the given year). We decided to use 9 years as the selection period because of the added statistical significance and 30 member ensembles as it should be significant enough to capture the effects of forcings on atmospheric circulation in mid-latitudes (Bittner et al. [2016]). To account for the fact that the selection is done taking into account a 9-year average, a 9-year moving average is applied to the simulations and observations, matching the years used in the selection criterion. This selection implies that in each year a set of different 30 members (out of the total ensemble of 212 members) is selected. To avoid potential biases caused by different model climatologies, we subtract the zonal average at each latitude (the resulting variable will be referred to as the azonal component of Z200) for each year and remove the climatology from 1980 to 2009 of the model Z200 results. To average the different members all data is regridded to a common spatial resolution of 1.75^{o}x1.75^{o} by means of bi-linear interpolation.

2.2 Data used

Our work employs the results shown in Mahmood et al. [2022] to constrain internal variability, which used models from CMIP6. The 6th phase of the Coupled Model Intercomparison Project (Eyring et al. [2016]) is a multi-model research collaboration that aims to better understand climate change driven by both natural variability and thermodynamic forcing factors. The historical CMIP6 runs are simulations spanning from 1850 to 2014 that utilize the prescribed historical forcing records as boundary conditions for the simulations. More specifically, emissions of short-lived species and long-lived green-house gases (GHGs), GHG concentrations, land-use forcing, solar forcing, stratospheric aerosol (volcanoes), aerosol forcing, ozone concentrations and nitrogen deposition are taken into account (Eyring et al. [2016]).

For this reason we used CMIP6 historical members as basis for our study. The ensemble we used is the same as the one shown in the supplementary material of Mahmood et al. [2022] with 3 members removed (BCC-CSM2-MR r1i1p1f1", "CAS-ESM2-0 r3i1p1f1", "CNRM-ESM2-1 r3i1p1f2") due to unavailability in our database. This ensemble is not homogeneous in terms of

members per model, with some models such as MIROC6 containing up to 50 members while some models contain only one. This could derive in the dynamics of certain models being over-represented in the constrained ensembles. In figures 3 the 30 member selection for each of the constraints is shown, as can be seen consistently more members are selected from some models, specially CanESM5 (Swart et al. [2019]) in global and Pacific constrained ensembles, with up to 17 members out of the 30 pertaining to it a given year. While this is a significant portion, CanESM5 models simulate the previously listed forcings and all models are represented at least once throughout the entire period spanning the study.

(a) Global constrains

(b) North Atlantic constrains

(c) Pacific Ocean constrains

Figure 3: The

The ERA5 reanalaysis (Hersbach et al. [2020]) has been used as the observation-based reference dataset. Reanalysis are simulations with an atmospheric model that assimilates a dense network of observations, and therefore provides an estimate of the observed climate evolution that is complete in contrast with regular observations which are not homogeneously distributed and can present gaps or errors.

Regarding the variables used in the study the dataset used consisted of monthly mean observations of geopotential height at 200 hPa averaged during the months of June, July and August (JJA) for both CMIP6 and ERA5 as was used in Teng et al. [2022].

2.3 Evaluation

For a model to faithfully simulate an observable we expect it to capture both the amplitude of a signal and the location of the signal. To this effect in order to assess the extent to which the climate model simulations reproduce the observed trend patterns in Z200, we computed the pattern correlation of a given map to that observed in the ERA5 reanalysis dataset. The pattern correlations were computed by an area-weighted Pearson's correlation coefficient. We calculate the pattern correlations for the spatial domain corresponding to the Northern hemisphere mid-latitudes (30^{o}N to 60^{o}N) and computed longitudinal sub-domains to assess the ability of the models in reproducing the observed trend for different regions . The selected the sub-domains used in this study are the following:

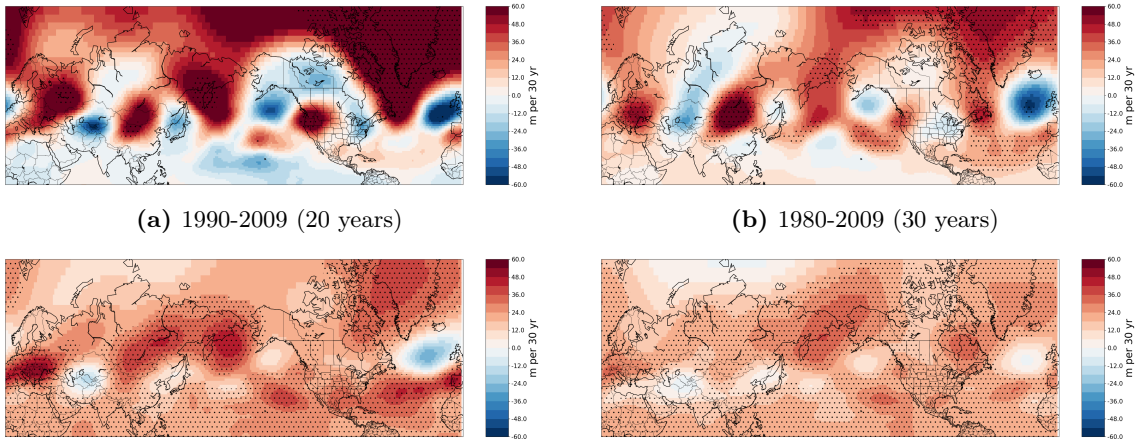
- Atlantic-Europe (60W - 60E)
- Eurasia (10W - 150E)
- East Asia - Pacific (90E - 120W)
- Pacific - North America (150E - 80W)
- North America - Atlantic (120W - 0E)

Given the magnitude of the Z200 trends in all of the ensembles is much smaller compared to the observed trends, we focus our analysis on quantifying the similarity of the spatial patterns, without taking the actual trend magnitudes into account. The significance of the trends was computed using a two-sided Student-t test with a confidence level of 95%. Student-t test is used because we expect monthly means of geopotential height to be normally distributed.

3 Results and discussion

3.1 Effect of constraints in the signal

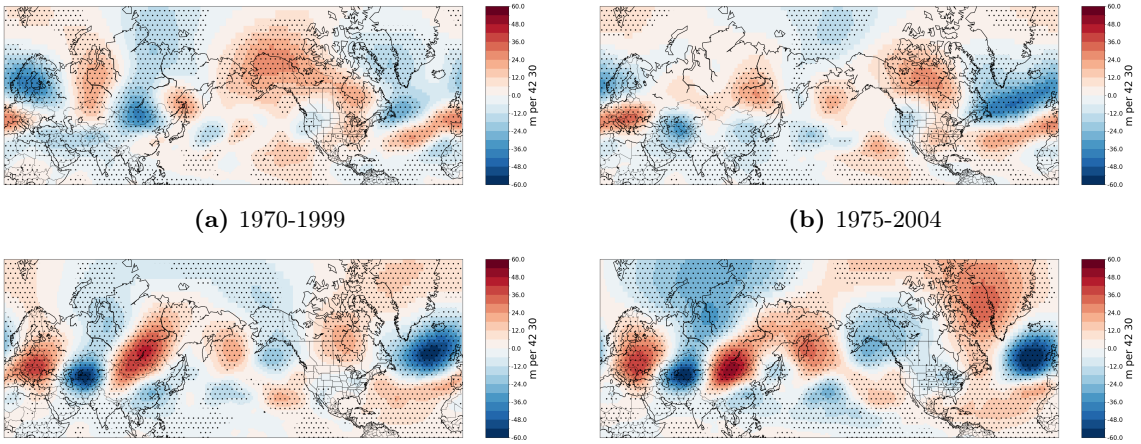
Since the historical runs of the CMIP6 ensemble end at 2014, we evaluate the changes during the last few decades of the historical forcing simulations ending in 2014. We also decided to plot the trends in Z200 for different lengths of time intervals to find the time interval with the clearest wave-5 pattern trend to base our study on. The periods analysed were 20, 30, 40 and 50-year periods spanning from 2014 backwards. We decided to work with 30-year time intervals, as it allows us to discuss temporal variations of the trend and, as can be seen, in figure 4b the interval from 1980-2009 is able to capture a very clear wave-5 pattern with a similar significance as the pattern documented in Teng et al. [2022]; note that this decision is somewhat arbitrary and we believe other time lengths could be equally suitable for the study. As was previously explained, we study trends of the azonal component of Z200 geopotential height. To do so, we computed the 30-year trends of 4 different periods (1970-1999, 1975-2004, 1980-2009 and 1985-2014). There is some correlation in the trends, but the patterns they exhibit, as seen in figure 5, show a different range of states, with the period from 1970 to 1999 and 1975 to 2004 not exhibiting the characteristic wave-5 pattern and having a fainter amplitude. While the periods from 1980 to 2009 and 1985 to 2014 exhibit a very clear pattern with some differences over the American continent among the two. Given that the observed trend is not constant in time, this could indicate that the driver of the pattern changes too, be it internal variability or external forcing.



(a) 1990-2009 (20 years) **(b)** 1980-2009 (30 years)

(c) 1970-2009 (40 years) **(d)** 1960-2009 (50 years)

Figure 4: Trend in Z200 geopotential height from the ERA5 reanalysis for different time periods. Stippling indicates values statistically significant at the 90% confidence level using a Student-t test.



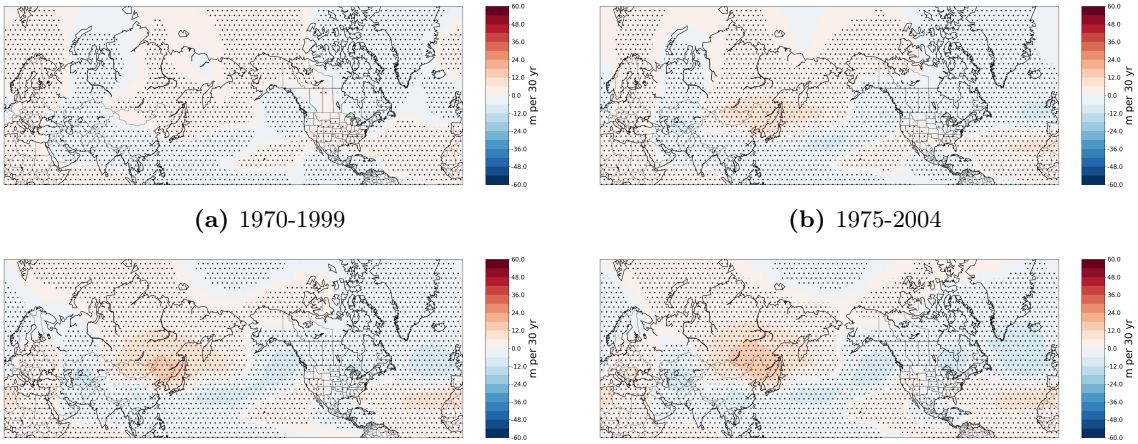
(a) 1970-1999 **(b)** 1975-2004

(c) 1980-2009 **(d)** 1985-2014

Figure 5: 30-year trends of the azonal component of Z200 geopotential height from the ERA5 reanalysis for different time periods. Stippling indicates values statistically significant at the 95% confidence level using a Student-t test.

We obtained the 30-year trends of the whole 209 members ensemble as well as the best 30 and worst 30 members of each SST constraint to study the effects of SST constraints in the emergence of the observed wave-5 train pattern. The results from the full ensemble can be seen in figure 6. As it is observed, the trends appear much fainter than in the ERA5 reanalysis. The obtained pattern remains stationary throughout all time periods with the amplitude growing in intensity over the years. This, however, is not the case in the ERA5 reanalysis, where during the first interval (figure 5a) the wave pattern over Eurasia is not present or during 1985-2014 (figure 5d) where it is lost in the Northern Atlantic and America. The resulting pattern in the unconstrained CMIP6 ensemble resembles the observed wave-5 pattern trend, specially over Europe and West Asia. However, the low pressure center in East Asia is not captured and all structure is lost over North America. This is further corroborated by table 1, where it can be seen that during the 1980-2009 period, when the wave-5 pattern is clearest, there are high pattern correlations over the Atlantic-Europe sector and lower pattern correlation over the Pacific region. During the first two periods, the ensemble is also unable to capture the observed pattern in ERA5 reanalysis, possibly indicating internal variability as the source of variability of the trends.

The constrained ensembles also fail to capture the amplitude of the trend, appearing faint as



(a) 1970-1999 **(b)** 1975-2004

(c) 1980-2009 **(d)** 1985-2014

Figure 6: 30-year trends of the azonal component of Z200 geopotential height from the unconstrained CMIP6 ensemble for different time periods. Stippling indicates values statistically significant at the 95% confidence level using a Student-t test.

	Circumference	Atlantic-Europe	Eurasia	EastAsia-Pacific	Pacific-NA	NA-Atlantic
1970-1999	0.129450	0.242364	0.052650	-0.035693	0.036759	0.233329
1975-2004	0.352705	0.568529	0.369007	0.203726	0.141692	0.473119
1980-2009	0.557536	0.780671	0.564878	0.466506	0.402919	0.654357
1985-2014	0.606628	0.791400	0.594208	0.528892	0.617177	0.662411

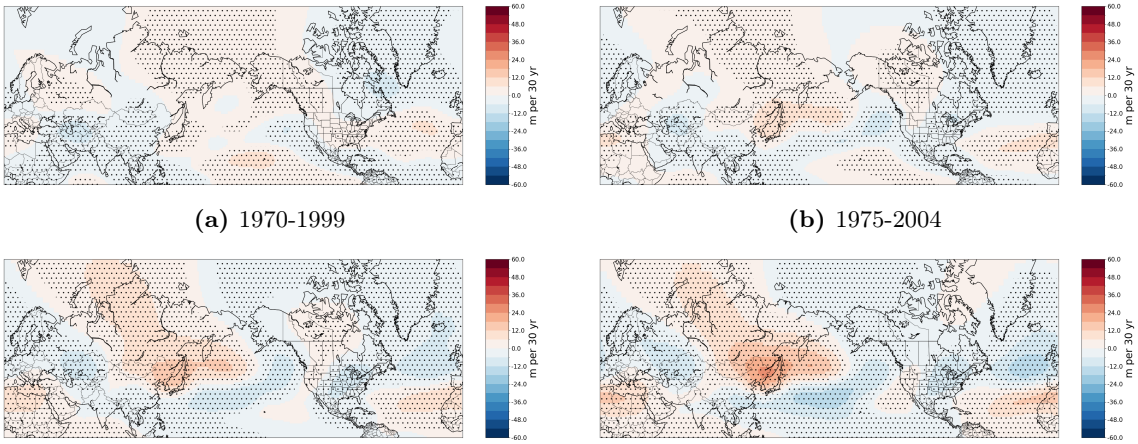
Table 1: Area-weighted spatial pattern correlation of unconstrained ensemble mean with respect to ERA5 reanalysis.

can be seen in figures 7, 8 and 9. This result is in line with the "signal-to-noise (S/N) paradox" (Scaife and Smith [2018]) where despite high correlation in variability between ensemble mean and observations and high spatial correlation, the models show small amplitudes in the signal for atmospheric circulation, specially over the Northern Atlantic. This could also indicate that the main mechanisms that induce this circulation trend are underestimated in the CMIP6 models.

Regarding pattern correlation, the globally constrained ensemble is unable to reproduce the high pressure ridge over Europe during the 1980-2009 (figure 7c) period in contrast with the full ensemble where this is reproduced (figure 6c). This constraint is however able to better capture the initial trend (1970-1999) when the wave-5 pattern is not as prevalent as is shown in figure 5a. These results could indicate that GHG or aerosol forcing is a driver of the pattern over Europe, given that some of the expected signal is lost over Eurasia when constraining internal variability of the system. The difference in pattern correlation between the globally constrained ensemble and unconstrained ensemble is shown in table 2. These results are consistent with the resulting plots. However, this contrasts the results found in Teng et al. [2022] where internal variability was expected to play a significant role in the formation of the pattern.

The Pacific Ocean constrained ensemble appears to highly correlate with the globally constrained one as is shown in figure 9. This was the case in the original "poor man's initialization" publication from Mahmood et al. [2022], where they suggested that this hinted at the Pacific ocean being the main internal predictability source on decadal to multi-decadal timescales. This is shown in the member selection as there is a big overlap in members used for the two ensembles, as was already reported in Mahmood et al. [2022] and can be observed in figure 3. This is corroborated by table 3 where the differences in pattern correlation between global constraints and Pacific constraints are shown.

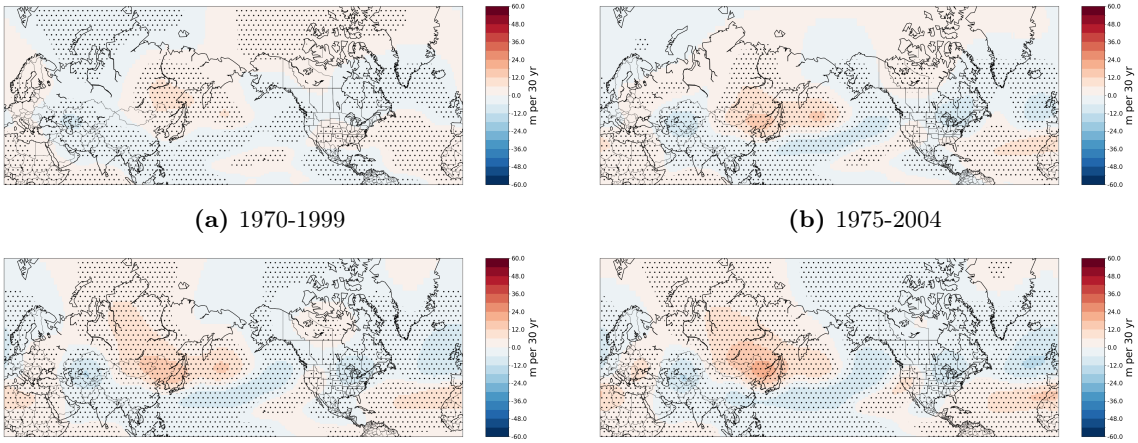
The North Atlantic constrained ensemble overall performs slightly worse than the complete



(a) 1970-1999 **(b)** 1975-2004

(c) 1980-2009 **(d)** 1985-2014

Figure 7: 30-year trends of the azonal component of Z200 geopotential height from the globally constrained CMIP6 ensemble for different time periods. Stippling indicates values statistically significant at the 95% confidence level using a Student-t test.



(a) 1970-1999 **(b)** 1975-2004

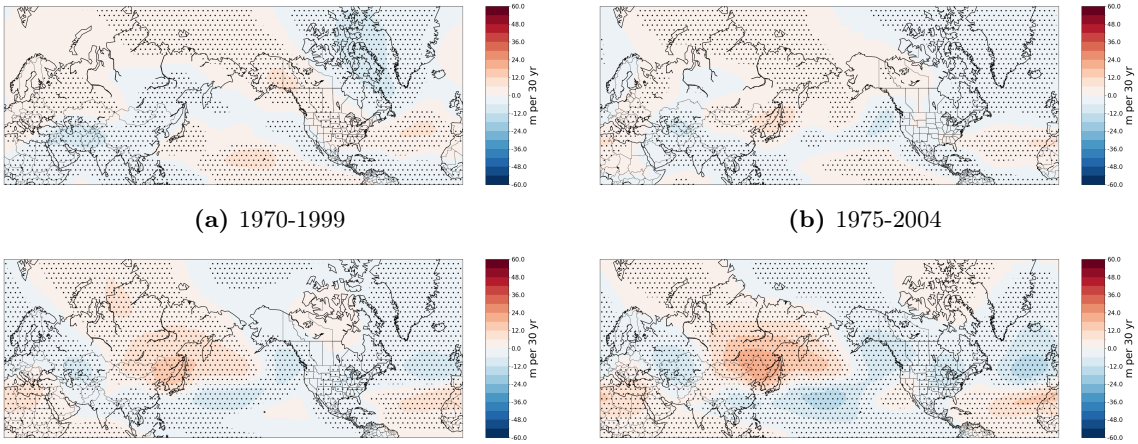
(c) 1980-2009 **(d)** 1985-2014

Figure 8: 30-year trends of the azonal component of Z200 geopotential height from the North Atlantic constrained CMIP6 ensemble for different time periods. Stippling indicates values statistically significant at the 95% confidence level using a Student-t test.

	Circumference	Atlantic-Europe	Eurasia	EastAsia-Pacific	Pacific-NA	NA-Atlantic
1970-1999	0.313085	0.182344	0.430446	0.567902	0.107392	0.066522
1975-2004	0.024213	0.060915	0.036733	-0.051899	-0.050011	0.038435
1980-2009	-0.141712	-0.240360	-0.233625	-0.159341	0.025430	-0.012963
1985-2014	-0.116812	-0.240008	-0.253612	-0.080721	0.034759	0.048927

Table 2: Difference in area-weighted spatial pattern correlation of globally constrained ensemble and the pattern correlation of the unconstrained ensemble mean (as shown in table 1) with respect to ERA5 reanalysis.

ensemble as can be seen in table 4. The only notable difference between the two ensembles is that the structure over East Asia is captured in the North Atlantic constraint during 1980-2009. This however does translate into higher values in spatial correlation for the region, as seen in



(a) 1970-1999 **(b)** 1975-2004

(c) 1980-2009 **(d)** 1985-2014

Figure 9: 30-year trends of the azonal component of Z200 geopotential height from the Pacific Ocean constrained CMIP6 ensemble for different time periods. Stippling indicates values statistically significant at the 95% confidence level using a Student-t test.

	Circumference	Atlantic-Europe	Eurasia	EastAsia-Pacific	Pacific-NA	NA-Atlantic
1970-1999	-0.022666	0.056583	-0.120360	-0.261943	0.052820	0.151869
1975-2004	0.023789	-0.010016	-0.016895	-0.042844	0.087477	0.040824
1980-2009	0.007691	-0.006736	0.022340	0.028206	-0.000661	0.001089
1985-2014	0.083858	0.066055	0.113826	0.107791	0.113329	0.051639

Table 3: Difference in area-weighted spatial pattern correlation of Pacific Ocean constrained ensemble and the pattern correlation of the globally constrained ensemble mean with respect to ERA5 reanalysis.

aforementioned table.

	Circumference	Atlantic-Europe	Eurasia	EastAsia-Pacific	Pacific-NA	NA-Atlantic
1970-1999	-0.048432	-0.043145	-0.044641	0.011763	0.066662	-0.099952
1975-2004	-0.101943	-0.117726	-0.095595	-0.103928	-0.092831	-0.142835
1980-2009	-0.045898	-0.069292	-0.034281	-0.092845	-0.078794	-0.059169
1985-2014	-0.001110	-0.008738	-0.033809	-0.045621	-0.001781	0.061267

Table 4: Difference in area-weighted spatial pattern correlation of North Atlantic constrained ensemble and the pattern correlation of the unconstrained ensemble mean (as shown in table 1) with respect to ERA5 reanalysis.

We have also evaluated the constrained ensembles containing the worst 30 ranked members of a given constraint to further explore the effects of not aligning with the modes of internal variability of the system. The pattern correlations of these ensembles further solidifies the previous results, as it is seen in table 5 during the period from 1980 to 2009 there is no significant change in skill between the best and worst 30 members for any of the globally constrained ensemble. However, for the initial periods (1970-1999 and 1975-2004), when the wave-5 pattern trend is not observed, there is a significant difference in pattern correlation, indicating that the earlier changes in atmospheric circulation may be influenced by SST variability.

	Circumference	Atlantic-Europe	Eurasia	EastAsia-Pacific	Pacific-NA	NA-Atlantic
1970-1999	-0.520387	-0.501648	-0.624735	-0.624347	-0.215900	-0.144872
1975-2004	-0.154057	-0.221259	-0.345147	0.010739	0.147013	-0.029299
1980-2009	0.054846	0.189865	0.144831	0.179603	-0.061039	-0.138788
1985-2014	-0.039656	0.028607	0.287658	0.034197	-0.509942	-0.408799

Table 5: Difference in area-weighted spatial pattern correlation between the worst 30 ranked and best 30 ranked members mean of the global constrain with respect to ERA5 reanalysis.

3.2 Further discussion

The observed results point toward internal variability of SST not being the main driver behind the observed Z200 trend over the Northern hemisphere, as constraining SST variability has not translated into better skill of the ensemble in reproducing the wave-5 pattern documented in Teng et al. [2022]. Additionally, although with very weak magnitude, a similar pattern to the originally documented seems to occur as a response to historical forcings. This differs from the findings in Teng et al. [2022], where AMV and IPO multidecadal variability seemed to be highly correlated with the appearance of the wave-5 pattern.

There are interpretations, however, that are consistent with both results. In Teng et al. [2022] the model analysis showed a weak signal and high variability in the model's AMV/IPO teleconnections and more robust results were found by applying a linear planetary wave model with synoptic eddy forcing, which is a mechanism in which SST could affect atmospheric circulation. This could imply that, despite constraining SST variability, the coupling mechanisms and teleconnections between atmospheric circulation and SST are not captured well enough in the models and neither is the atmospheric circulation response as a result.

Another possible interpretation would be that, despite better correlating to the SST anomalies than the whole ensemble, the constrained ensembles are not able to accurately reproduce the AMV and IPO modes of variability. This would be consistent with findings in the literature pointing that the models underestimate multidecadal variability (Kravtsov [2017]). In order to explore this possibility, the AMV/IPO indices of the constrained ensembles could be studied in further studies.

The contribution from forcing is likewise unclear, while the lack of improvement in pattern correlation of the constrained ensemble could be evidence of forcing playing an important role in the formation of the observed trend, the whole ensemble mean is still not able to completely capture the amplitude and location the wave-5 pattern structure. Thus, if forcing is the main driver, why are the models not reproducing the expected signal? The signal-to-noise paradox could explain the weak signal compared to the ERA5 reanalysis, but the ensemble signal fails to exhibit the low pressure center over east Asia and the high pressure ridges east and west of the American continent. However, the fact that the pattern over Eurasia is well captured is an indication that forcing could play a role in triggering this circulation pattern.

4 Conclusions

Despite previous literature linking multidecadal modes of variability of SST fluctuations to the 1979-2020 summer trend in Z200 (Teng et al. [2022]), our results show no evidence of SST variations being the driver of the reported trend. While the average response to forcing indicates a spatial structure of changes with features resembling the observed wave-5 pattern at a much fainter magnitude, by aligning (or disaligning) the ensemble mean of a large multi-model ensemble to the observed state of the climate by selecting the members with best (worst) spatial agreement with SST, no improvements (declines) in our ability to reproduce the pattern were shown during the periods with stronger signal (1980-2009 and 1985-2014) with respect to the unconstrained ensemble. Conversely, for the 1970-1999 and 1975-2004 phasing in internal variability results in higher pattern correlation despite the trends not showing the characteristic wave-5 pattern.

While this points towards radiative forcings playing a larger than previously expected role in the

formation of the pattern, to what extent forcings drive the phenomenon is hard to assess, specially because the unconstrained ensemble is unable to reproduce the main characteristics of the pattern over North America and parts of the Pacific Ocean. Additionally, the amplitude of the signal appears significantly fainter than it is in the reanalysis dataset, and while this result is in line with the signal-to-noise paradox (Scaife and Smith [2018]), this highlights that current models fail to properly capture the correct intensity of changes in atmospheric circulation and could explain why models are not able to reproduce the observed trend.

5 References

M. Bittner, H. Schmidt, C. Timmreck, and F. Sienz. Using a large ensemble of simulations to assess the northern hemisphere stratospheric dynamical response to tropical volcanic eruptions and its uncertainty. *Geophysical Research Letters*, 43(17):9324–9332, 2016.

B. Cushman-Roisin and J.-M. Beckers. *Introduction to geophysical fluid dynamics: physical and numerical aspects*. Academic press, 2011.

R. Dole, M. Hoerling, J. Perlwitz, J. Eischeid, P. Pegion, T. Zhang, X.-W. Quan, T. Xu, and D. Murray. Was there a basis for anticipating the 2010 russian heat wave? *Geophysical research letters*, 38(6), 2011.

V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor. Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958, 2016.

D. Fereday and J. Knight. The roles of atmospheric circulation and sea surface temperature in uk surface climate. *Atmospheric Science Letters*, 24(3):e1139, 2023.

W. L. Gates, J. S. Boyle, C. Covey, C. G. Dease, C. M. Doutriaux, R. S. Drach, M. Fiorino, P. J. Gleckler, J. J. Hnilo, S. M. Marlais, T. J. Phillips, G. L. Potter, B. D. Santer, K. R. Sperber, K. E. Taylor, and D. N. Williams. An overview of the results of the atmospheric model intercomparison project (amip i). *Bulletin of the American Meteorological Society*, 80(1):29 – 56, 1999. doi: 10.1175/1520-0477(1999)080<0029:AOOTRO>2.0.CO;2. URL `https://journals.ametsoc.org/view/journals/bams/80/1/1520-0477_1999_080_0029_aootro_2_0_co_2.xml`.

S. B. Goldenberg, C. W. Landsea, A. M. Mestas-Nuñez, and W. M. Gray. The recent increase in atlantic hurricane activity: Causes and implications. *Science*, 293(5529):474–479, 2001.

H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.

IPCC. *Human Influence on the Climate System*, page 423–552. Cambridge University Press, 2021.

J. R. Knight, R. J. Allan, C. K. Folland, M. Vellinga, and M. E. Mann. A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophysical Research Letters*, 32 (20), 2005.

K. Kornhuber, S. Osprey, D. Coumou, S. Petri, V. Petoukhov, S. Rahmstorf, and L. Gray. Extreme weather events in early summer 2018 connected by a recurrent hemispheric wave-7 pattern. *Environmental Research Letters*, 14(5):054002, 2019.

K. Kornhuber, D. Coumou, E. Vogel, C. Lesk, J. F. Donges, J. Lehmann, and R. M. Horton. Amplified rossby waves enhance risk of concurrent heatwaves in major breadbasket regions. *Nature Climate Change*, 10(1):48–53, 2020.

S. Kravtsov. Pronounced differences between observed and cmip5-simulated multidecadal climate variability in the twentieth century. *Geophysical Research Letters*, 44(11):5749–5757, 2017.

R. Mahmood, M. G. Donat, P. Ortega, F. J. Doblas-Reyes, C. Delgado-Torres, M. Samsó, and P.-A. Bretonnière. Constraining low-frequency variability in climate projections to predict climate on decadal to multi-decadal timescales–a poor man's initialized prediction system. *Earth System Dynamics*, 13(4):1437–1450, 2022.

S. Milinski, N. Maher, and D. Olonscheck. How large does a large ensemble need to be? *Earth System Dynamics*, 11(4):885–901, 2020. doi: 10.5194/esd-11-885-2020. URL `https://esd.copernicus.org/articles/11/885/2020/`.

S. Power, T. Casey, C. Folland, A. Colman, and V. Mehta. Inter-decadal modulation of the impact of enso on australia. *Climate dynamics*, 15:319–324, 1999.

A. A. Scaife and D. Smith. A signal-to-noise paradox in climate science. *npj Climate and Atmospheric Science*, 1(1):28, 2018.

S. D. Schubert, M. J. Suarez, P. J. Pegion, R. D. Koster, and J. T. Bacmeister. On the cause of the 1930s dust bowl. *Science*, 303(5665):1855–1859, 2004. doi: 10.1126/science.1095048. URL https://www.science.org/doi/abs/10.1126/science.1095048.

S. D. Schubert, H. Wang, R. D. Koster, M. J. Suarez, and P. Y. Groisman. Northern eurasian heat waves and droughts. *Journal of Climate*, 27(9):3169 – 3207, 2014. doi: 10.1175/JCLI-D-13-00360.1. URL https://journals.ametsoc.org/view/journals/clim/27/9/jcli-d-13-00360.1.xml.

T. G. Shepherd. Atmospheric circulation as a source of uncertainty in climate change projections. *Nature Geoscience*, 7(10):703–708, 2014.

D. M. Smith, R. Eade, and H. Pohlmann. A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Climate dynamics*, 41:3325–3338, 2013.

N. C. Swart, J. N. S. Cole, V. V. Kharin, M. Lazare, J. F. Scinocca, N. P. Gillett, J. Anstey, V. Arora, J. R. Christian, S. Hanna, Y. Jiao, W. G. Lee, F. Majaess, O. A. Saenko, C. Seiler, C. Seinen, A. Shao, M. Sigmond, L. Solheim, K. von Salzen, D. Yang, and B. Winter. The canadian earth system model version 5 (canesm5.0.3). *Geoscientific Model Development*, 12 (11):4823–4873, 2019. doi: 10.5194/gmd-12-4823-2019. URL https://gmd.copernicus.org/articles/12/4823/2019/.

K. E. Taylor. A summary of the cmip5 experiment design. *Bulletin of the American Meteorological Society*, 4:1, 2012.

H. Teng, R. Leung, G. Branstator, J. Lu, and Q. Ding. Warming pattern over the northern hemisphere midlatitudes in boreal summer 1979–2020. *Journal of Climate*, 35(11):3479–3494, 2022.

K. E. Trenberth and J. T. Fasullo. Climate extremes and climate change: The russian heat wave and other climate extremes of 2010. *Journal of Geophysical Research: Atmospheres*, 117(D17), 2012.

J. M. Wallace and P. V. Hobbs. 3 - atmospheric thermodynamics. In J. M. Wallace and P. V. Hobbs, editors, *Atmospheric Science (Second Edition)*, pages 63–111. Academic Press, San Diego, second edition edition, 2006. ISBN 978-0-12-732951-2. doi: https://doi.org/10.1016/B978-0-12-732951-2.50008-9. URL https://www.sciencedirect.com/science/article/pii/B9780127329512500089.