Pervasive Warming Bias in CMIP6 Tropospheric Layers

forthcoming: *Earth and Space Science*

R. McKitrick[1], J. Christy[2]

[1]Department of Economics and Finance, University of Guelph, Guelph ON Canada

[2]Earth System Science Center, University of Alabama in Huntsville, Huntsville AL USA.

Corresponding author: first and last name (ross.mckitrick@uoguelph.ca)

Key Points:

- We compare post-1979 lower- and mid-troposphere hindcasts in 38 CMIP6 model runs to satellite, balloon and reanalysis observations
- All model runs warmed faster than observations both globally and in the tropics, in most cases significantly
- Models can be grouped by ECS value, but even low-ECS models exhibit too much tropospheric warming post-1979

Index terms
1610 Atmosphere
1620 Climate Dynamics
1626 Global Climate Models
3333 Model calibration
9820 Techniques applicable in three or more fields

Key words
Troposphere
Global warming
Climate model testing
Trend estimation
Climate sensitivity

Abstract

The tendency of climate models to overstate warming in the tropical troposphere has long been noted. Here we examine individual runs from 38 newly-released Coupled Model Intercomparison Project version 6 (CMIP6) models and show that the warm bias is now observable globally as well. We compare CMIP6 runs against observational series drawn from satellites, weather balloons and reanalysis products. We focus on the 1979-2014 interval, the maximum span for which all observational products are available and for which models were run using historically-observed forcings. For lower- and mid-troposphere layers both globally and in the tropics, all 38 models overpredict warming in every target observational analogue, in most cases significantly so, and the average differences between models and observations are statistically significant. We present evidence that consistency with observed warming would require lower model Equilibrium Climate Sensitivity (ECS) values.

Plain Language Summary

It has long been known that previous generations of climate models exhibit excessive warming rates in the tropical troposphere. With the release of the CMIP6 (Coupled Model Intercomparison Project version 6) climate model archive we can now update the comparison. We examined historical (hindcast) runs from 38 CMIP6 models in which the models were run using historically-observed forcings. We focus on the 1979-2014 interval, the maximum for which all models and observational data are available and for which the models were run with historical forcings. What was previously a tropical bias is now global. All model runs warmed faster than observations in the lower and mid-troposphere, in the tropics and globally. On average, and in most individual cases, the trend difference is significant. Warming trends in models tend to rise with the model Equilibrium Climate Sensitivity (ECS), and we present evidence that the distribution of ECS values across the model is unrealistically high.

1 Introduction

Numerous studies have pointed to a tendency across climate models to project too much contemporary warming in the tropical troposphere (Karl et al. 2006, Douglass et al. 2007, McKitrick et al. 2010, Fu et al. 2011, Bengtsson and Hodges 2011, Thorne et al. 2011, Po-Chedley and Fu 2012, McKitrick and Vogelsang 2014) with additional evidence pointing to a global tropospheric bias as well (Christy and McNider 2017). Here we present an updated comparison using the first 38 models made available in the newly-released 6th generation Coupled Model Intercomparison Project (CMIP6) archive comparing model reconstructions of historical layer-averaged lower- and mid-troposphere temperature series against observational analogues from satellites, balloon-borne radiosondes and reanalysis products. We compare trends over 1979-2014, the longest interval for which all three observational systems are available and for which models were run with historically-observed forcings. None of our conclusions would be different if we extended the end date to 2018. We examine four atmospheric regions: the global lower troposphere and mid-troposphere, and the tropical lower- and mid-troposphere layers.

In previous studies, although a warm bias was typically present, over large atmospheric regions the model spread at least partly overlapped the observational analogues, especially at the global level. This is no longer the case. Every model overpredicts warming in both the lower- and mid-

troposphere layers, in the tropics and globally. On average the discrepancies are statistically very significant and the majority of individual model discrepancies are statistically significant as well.

2 Data and Methods

2.1 Data

2.1.1 Observations

We use the temperature data collected from three general categories. (1) Radiosonde (or sonde) data are measured by thermistors carried aloft by balloons at stations around the world which radio the information down to a ground station. Sondes report temperatures at many levels, and we use here annual averages at the standard pressure-levels: 1000 (if above the launch site), 850, 700, 500, 400 300, 200 150, 100, 70, 50, 30 and 20 hPa. As noted in Table 1 there are four datasets available: NOAA (RATPAC, Free et al. 2005), U WIEN, Austria (RAOBCORE and RICH, Haimberger et al. 2012) and the University of New South Wales, Australia (UNSW, Sherwood and Nishant 2015). Note that the commercial software used to process sonde data was revised in 2011 with the result that inferred humidity levels increased after 2009 by several percent (Jauhiainen et al. 2011). This induced a slight warming step which is not observed in other other systems and may be an artifact (Christy et al. 2018.)

(2) Since late 1978, several polar-orbiting satellites carried some form of a microwave sensor to monitor atmospheric temperatures. These spacecraft would circle the globe roughly pole-to-pole making a complete orbit in about 100 minutes. They were (and are) sun-synchronous so the Earth would essentially rotate on its axis underneath as the spacecraft orbited pole to pole so that essentially the entire planet is observed in a single Earth-rotation (or day). The intensity of microwave emissions from atmospheric oxygen are directly proportional to temperature, thus allowing a conversion of these measurements to temperature. Since the emissions come from the most of the atmosphere, they represent a deep layer-average temperature. For our purposes we shall focus on two deep layers, the lower troposphere (LT, surface to ~ 9 km) and the mid-troposphere (MT, surface to ~ 15 km). The University of Alabama in Huntsville (UAH) and Remote Sensing Systems (RSS) produce averages every month of both products (Spencer et al. 2017, Mears and Wentz 2016). NOAA provides values for MT globally and the University of Washington (UW) produces tropical value of MT (Po-Chedley et al. 2015). There are differences in all of the products discussed here and the reader may want to consult the listed publications for more information.

(3) The third category of these datasets are known as Reanalyses. In this category, a global weather model with many atmospheric layers ingests as much data as possible, from surface observations, sondes and satellites, to generate a global depiction of the surface and atmosphere that is made globally consistent through the model equations. We will access the temperature data from these datasets at 17 pressure levels from the surface to 10 hPa and will be able to calculate the deep-layer averages that match those of the satellite measurements. Four such datasets are available to us, two from the European Centre For Medium Range Forecasts (ERA-I and ERA5, Dee et al. 2011, Hersbach et al. 2018), and one each from the Japanese

Meteorological Agency (JRA55, Kobayashi et al. 2015) and NASA (MERRA2, Gelaro et al. 2017).

2.2.2 Climate models

The climate model simulations utilized here are those accepted for analysis in CMIP6 for which the models are executed in standardized simulations so they may be intercompared properly. We obtained the model runs from the Lawrence Livermore National Laboratory archive https://pcmdi.llnl.gov/CMIP6/. For this study we used the period 1979-2014 from the simulation set that represents 1850-2014 in which the models were provided with "historical" forcings. These time-varying forcings are estimates of the amount of energy deviations that occurred in the real world and are applied to the models through time. These include variations in factors such as volcanic aerosols, solar input, dust and other aerosols, important gases like carbon dioxide, ozone and methane, land-surface brightness and so on. With all models applying the same forcing as believed to have occurred for the actual Earth, the direct comparison between models and observations is appropriate. The models and runs are identified in Table 2. We also list the estimated Equilibrium Climate Sensitivity (ECS) values for the 31 models for which we were able to find values, usually through unpublished online documentation (sources available on request).

Global LT and MT data are presented in Figures 1 and 2. Individual model runs are shown as gray lines, the model average is the thick black line, and the observational mean is the thick blue line.

2.2 Methods

Linear trends were estimated on annual observations over the 1979-2014 interval, which is the maximum-length interval for which all observational series are available and for which the models were run using observed forcings. We pre-test the temperature series for unit roots, which if present imply nonstationarity of a form that makes conventional trend regressions invalid (Wooldridge 2019). We use the form of the test derived in Elliott et al (1996), allowing for a trend stationary alternative and an autoregressive lag. The null hypothesis of the test is that the series contains a unit root. Such tests can exhibit a tendency to under-reject in the presence of autocorrelation due to low power so we expanded the time interval to 1959-2014, which means the sonde record, specifically the mean of the RAOBCORE, RICH, RATPAC and UNSW products, serves as the observational series. We reject the null hypothesis for all individual model runs and the sonde mean series, thus indicating that the data can be treated as trend stationary. An appropriate method in this case for constructing confidence intervals and

hypothesis tests of trend equivalence is the autocorrelation-robust method of Vogelsang and Franses (2005). See McKitrick et al. (2010) for details on implementation.

3 Results

Figure 3 shows the trends and 95% confidence intervals in degrees C per decade in the 38 individual climate models (red), the climate model ensemble mean (thick red) and the three mean observational series (respectively, radiosondes, reanalysis and satellites, thick blue). The dashed blue line shows the satellite trend level. Differing data availability leads to somewhat different observational series combinations. For the sonde data, the average includes RAOBCORE, RICH, and RATPAC in all specifications, and additionally includes UNSW in the MT layers (global and tropics). The mean of the reanalysis data uses ERA-I, ERA5, JRA55 and MERRA2 for the global LT and the topical LT and MT layers, uses ERA5, JRA55 and MERRA2 for the global MT layer. The mean of the satellite data uses UAH and RSS for global LT and MT and for topical LT, and additionally uses NOAA and UW for tropical MT.

The top row of Figure 3 shows the MT layer results for the global (left) and tropical (right) samples. The bottom row shows the same for the LT layer. It is immediately apparent that every model run in every regional and layer average has a mean trend that exceeds the corresponding observed trends regardless of how they are measured.

Tables 3 and 4 show the trend coefficients and symmetric 95% confidence interval widths (in degrees C/decade) for all individual models, for the average of all models, and for the three observational system averages. For example, the global LT trend in the ACCESS model (top row of Table 1) is 0.250±0.103 degrees C/decade. Table 3 shows the Vogelsang-Franses test scores on the null hypothesis of trend equivalence for each test region. A value greater than 41.53 is significant at 5%. The first row shows the results of testing whether the average model trend exceeds the average sonde trend. The second row shows the corresponding result for reanalysis data and the third row shows the results for satellite data. The fourth row shows the number of individual model runs in which the trend significantly exceeds the satellite average. In the first three rows we see that all 12 tests reject, meaning the average model significantly exceeds the average observed series regardless of region or atmospheric layer, and regardless of observational measurement system. The final row shows that a majority of models also reject individually against the satellite data except in the global LT case, in which 18 of 38 models reject. If we were to extend the data sample to a 2018 end date, the sum would still be 24 and 26, respectively, for the global LT and MT layers, and would increase to 22 and 23 in the tropical LT and MT layers.

An increasingly common form of model diagnostic involves examining what are called “emergent constraints” (Caldwell et al. 2018). ECS values across models vary widely but the correct value cannot be directly determined by measurement. The emergent constraint concept involves looking for observable features of the climate that have measurable counterparts in models that are correlated with the model ECS. The observed measurement of the correlate will then indicate which model ECS values are more likely to be true. Various metrics have been proposed, such as the difference between tropical and Southern Hemisphere midlatitude total

cloud fraction, Tropical zonal-average lower-troposphere relative humidity in the moist-convective region, model error in total cloud amount between 60 degrees N/S and the fraction of tropical clouds with tops below 850 mb whose tops are also below 950 mb (see list in Caldwell et al. 2018 Table 1). The correlations between the proposed metrics and ECS vary widely, and as noted in Caldwell et al., many do not have a valid physical underpinning. Since we are here analysing model warming rates, which is directly connected to ECS, it is worthwhile examining if an emergent constraint interpretation can be applied to our results.

The correlations between ECS and trend terms are as follows: LT-global 0.67, MT-global 0.60, LT-tropics 0.50 and MT-tropics 0.50. Hence the models with low ECS values tend to have lower tropospheric trends, thus closer to observed values, and therefore are more likely to be realistic. Figure 4 provides more insight into the data. The models cluster into two distinct groups based on whether the ECS is above (red squares) or below (blue circles) 3.4K. A solid square or circle indicates the trend is from the LT and an open shape indicates MT. The mean values in each cluster for both the LT and MT layers are indicated by + signs, and the layer averages are joined by the gray lines (dashed-MT, solid-LT) which represent the emergent constraint.

Within clusters, ECS and warming trend values are not correlated, but as is indicated by the gray lines the correlation emerges when comparing between low and high clusters. In the high group the overall mean trend is 0.28 C/decade and the mean ECS is 4.67K. In the low group the overall mean trend is 0.21 C/decade and the mean ECS is 2.76K. The mean observed trends in the LT and MT layers across all measurement types are indicated by the arrows along the horizontal axis (LT solid 0.15 C/decade, MT open 0.09 C/decade). Since the mean trends even in the low ECS model group are still too high compared to the observed trends the emergent constraint implies a need to extrapolate into even lower ECS levels to approximately match observations. Examining where the dotted lines cross the arrows informally indicates how far such extrapolation would need to go, however as drawn this would imply ECS values well below 1.0K. Since a curve of any shape can be fitted between two points one could equally use concave lines which would still imply ECS values below 2.0K in order to have associated warming trends consistent with observations.

5 Conclusions

The literature drawing attention to an upward bias in climate model warming responses in the tropical troposphere extends back at least 15 years now (Karl et al. 2006). Rather than being resolved the problem has become worse, since now every member of the CMIP6 generation of climate models exhibits an upward bias in the entire global troposphere as well as in the tropics. The models with lower ECS values have warming rates somewhat closer to observed, but are still significantly biased upwards and do not overlap observations. Models with higher ECS values also have higher tropospheric warming rates, and applying the emergent constraint concept implies that an ensemble of models with warming rates consistent with observations would likely have to have ECS values at or below the bottom of the CMIP6 range. Our findings mirror recent evidence from inspection of CMIP6 Equilibrium Climate Sensitivities (Vosen

2019) and paleoclimate simulations (Zhu et al. 2020) which also reveal a systematic warm bias in the latest generation of climate models.

Acknowledgments, Samples, and Data

- Christy was supported by DOE Grant DE-SC0019296. The data used in this study are available at https://data.mendeley.com/datasets/sd97vh79v8/1. First author additional affiliation: Senior Fellow, Fraser Institute.

References

Bengtsson, L. and K. Hodges, "Evaluating temperature trends in the tropical troposphere." *Climate Dynamics* 2009, doi: 10.1007/s00382-009-0680-y.

Caldwell, P. M., M. D. Zelinka, and S. A. Klein, (2018) Evaluating Emergent Constraints on Equilibrium Climate Sensitivity. *Journal of Climate* 31, 3921–3942, https://doi.org/10.1175/JCLI-D-17-0631.1.

Christy, J.R. and McNider, R.T. (2017). Satellite bulk tropospheric temperatures as a metric for climate sensitivity. *Asia-Pacific Journal of Atmospheric Science* 53(4) 511-518 DOI:10.1007/s13143-017-0070-z

Christy, J.R., R.W. Spencer, W.D. Braswell and R. Junod, 2018: Examination of space-based bulk atmospheric temperatures used in climate research.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.

Douglass, D. H., J. R. Christy, B. D. Pearson, and S. F. Singer. 2007. A comparison of tropical temperature trends with model predictions. Intl J Climatology Vol 28(13) 1693—1701 DOI 10.1002/joc.1651

Elliott, G., Rothenberg, T., & Stock, J. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, 64(4), 813-836. doi:10.2307/2171846

Free, M., D. J. Seidel, J. K. Angell, J. Lanzante, I. Durre, and T. C. Peterson, 2005: Radiosonde atmospheric temperature products for assessing climate (RATPAC): A new data set of large-area anomaly time series. *J. Geophys. Res.*, **110**, D22101, doi:10.1029/2005JD006169.

Fu, Q., S. Manabe and C. M. Johanson (2011) On the warming in the tropical upper troposphere: Models versus observations *Geophysical Research Letters* Vol. 38, L15704, doi:10.1029/2011GL048101, 2011.

Gelaro and co-authors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Climate, DOI:10.1175/JCLI-D-16-0758.1

Haimberger, L., C. Tavolato, and S. Sperka, 2012: Homogenization of the global radiosonde temperature dataset through combined comparison with reanalysis background series and neighboring stations. *J. Climate* 25, 8108-8131.

Hersbach, H. and co-authors, 2018: Operational global reanalysis: progress, future direcctions and synergies with NWP. ERA Report Series #27. DOI:10.21957/tkic6g3wm.

Jauhiainen, H., M. Turunen and J. Wahrn (2011) Improved Measurement Accuracy of Vaisala Radiosonde RS92 Vaisala Technical Document #185/2011, Helsinki Finland.

Karl, T. R., S. J. Hassol, C. D. Miller, and W. L. Murray (2006). *Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences.* Synthesis and Assessment Product. Climate Change Science Program and the Subcommittee on Global Change Research. http://www.climatescience.gov/Library/sap/sap1-1/finalreport/sap1-1-final-all.pdf. Accessed August 3 2010.

Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, doi:10.2151/jmsj.2015-001.

McKitrick, R. R., S. McIntyre and C. Herman (2010) Panel and Multivariate Methods for Tests of Trend Equivalence in Climate Data Sets. *Atmospheric Science Letters*, 11(4) pp. 270-277, October/December 2010 DOI: 10.1002/asl.290.

McKitrick, Ross R. and Timothy Vogelsang (2014) HAC-Robust Trend Comparisons Among Climate Series with Possible Level Shifts Environmetrics DOI: 10.1002/env.2294.

Mears, C.A. and F.J. Wentz, 2016: Sensitivity of satellite-derived tropospheric temperature trends to the diurnal cycle adjustment. J. Climate, 29, 3629-3646. DOI: 10.1175/JCLI-D-15-0744.1.

Po-Chedley, S. and Q. Fu (2012) Discrepancies in tropical upper tropospheric warming between atmospheric circulation models and satellites. *Environmental Research Letters* 7 (2012) doi:10.1088/1748-9326/7/4/044018.

Po-Chedley, S., T.J. Thorsen and Q. Fu, 2015: Removing diurnal cycle contamination in satellite-derived tropospheric temperatures. Understanding tropical tropospheric trend discrepancies. *J. Climate*, 28, 2274-2290, DOI:10.1175/JCLI-D-13-00767.1.

Sherwood, S. C., and N. Nishant, 2015: Atmospheric changes through 2012 as shown by iteratively homogenized radiosonde temperature and wind data (IUKv2). *Environ. Res. Lett.*, **10**, 054007, doi:10.1088/1748-9326/10/5/054007.

Spencer, R.W., J.R. Christy, and W.D. Braswell, 2017: UAH Version 6 global satellite temperature products: Methodology and results. *Asia-Pac. J. Atmos. Sci*, 53(1), pp 1-10. DOI:10.1007/s13143.

Thorne, P. W. , P. Brohan, H.A. Titchner, M.P. McCarthy, S. C. Sherwood, T. C. Peterson, L. Haimberger, D. E. Parker, S. F. B. Tett, B. D. Santer, D. R. Fereday, and J.J. Kennedy (2011) A quantification of uncertainties in historical tropical tropospheric temperature trends from radiosondes. *Journal of Geophysical Research-Atmospheres* Vol. 116, D12116, doi:10.1029/2010JD015487, 2011.

Vogelsang, T.J. and Franses, P. H. (2005) Testing for common deterministic trend slopes. *Journal of Econometrics* 126, 1-24.

Vosen, Paul (2019) New climate models predict a warming surge. *Science* April 26, 2019, doi:10.1126/science.aax7217

Wooldridge, J. (2020) *Introductory Econometrics: A Modern Approach*. Boston: Cengage.

Zhu, J., Poulsen, C.J. & Otto-Bliesner, B.L. (2020) High climate sensitivity in CMIP6 model not supported by paleoclimate. *Nature Climate Change* 10, 378–379. https://doi.org/10.1038/s41558-020-0764-6

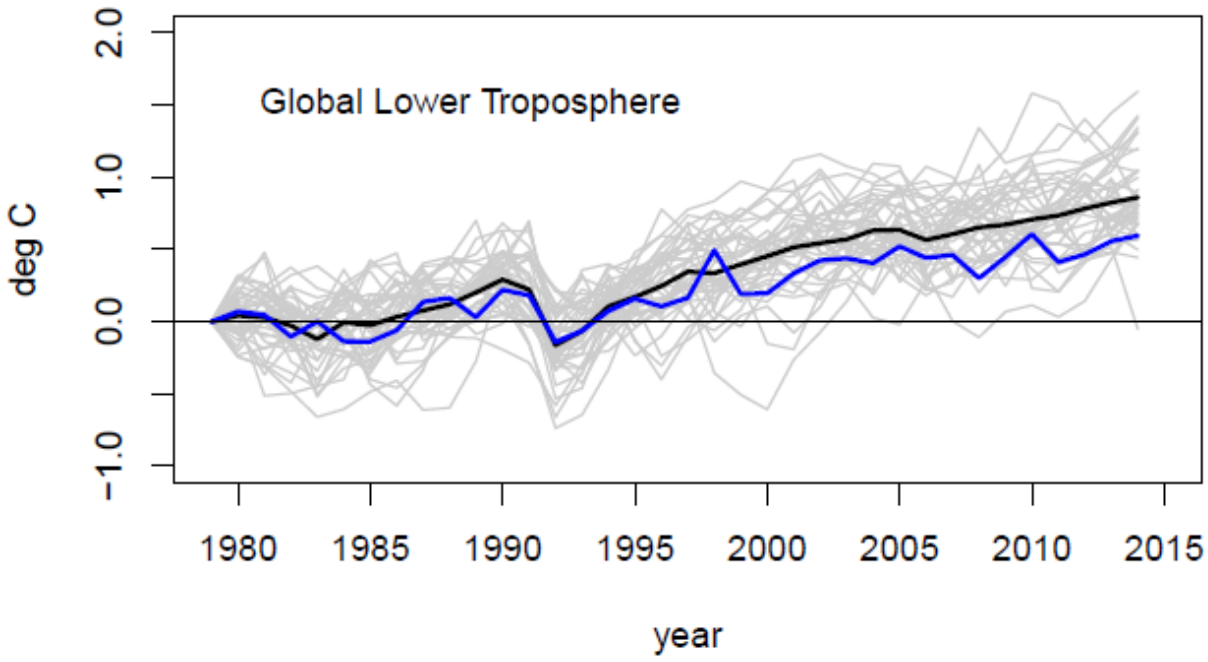

Figure 1. Time series of model and observation temperature anomalies, global lower troposphere. Individual model runs (gray lines), model mean (black line), observational mean (blue line). All series shifted to begin at 0 in 1979.

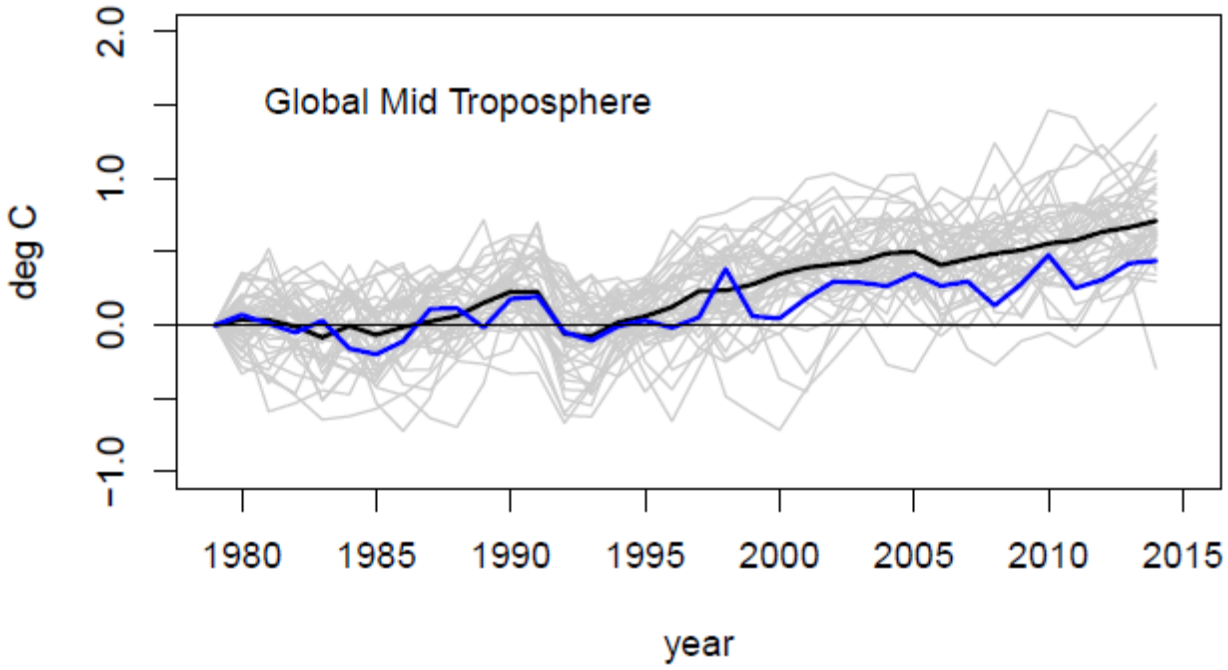

Figure 2. Time series of model and observation temperature anomalies, global mid-troposphere. Individual model runs (gray lines), model mean (black line), observational mean (blue line). All series shifted to begin at 0 in 1979.

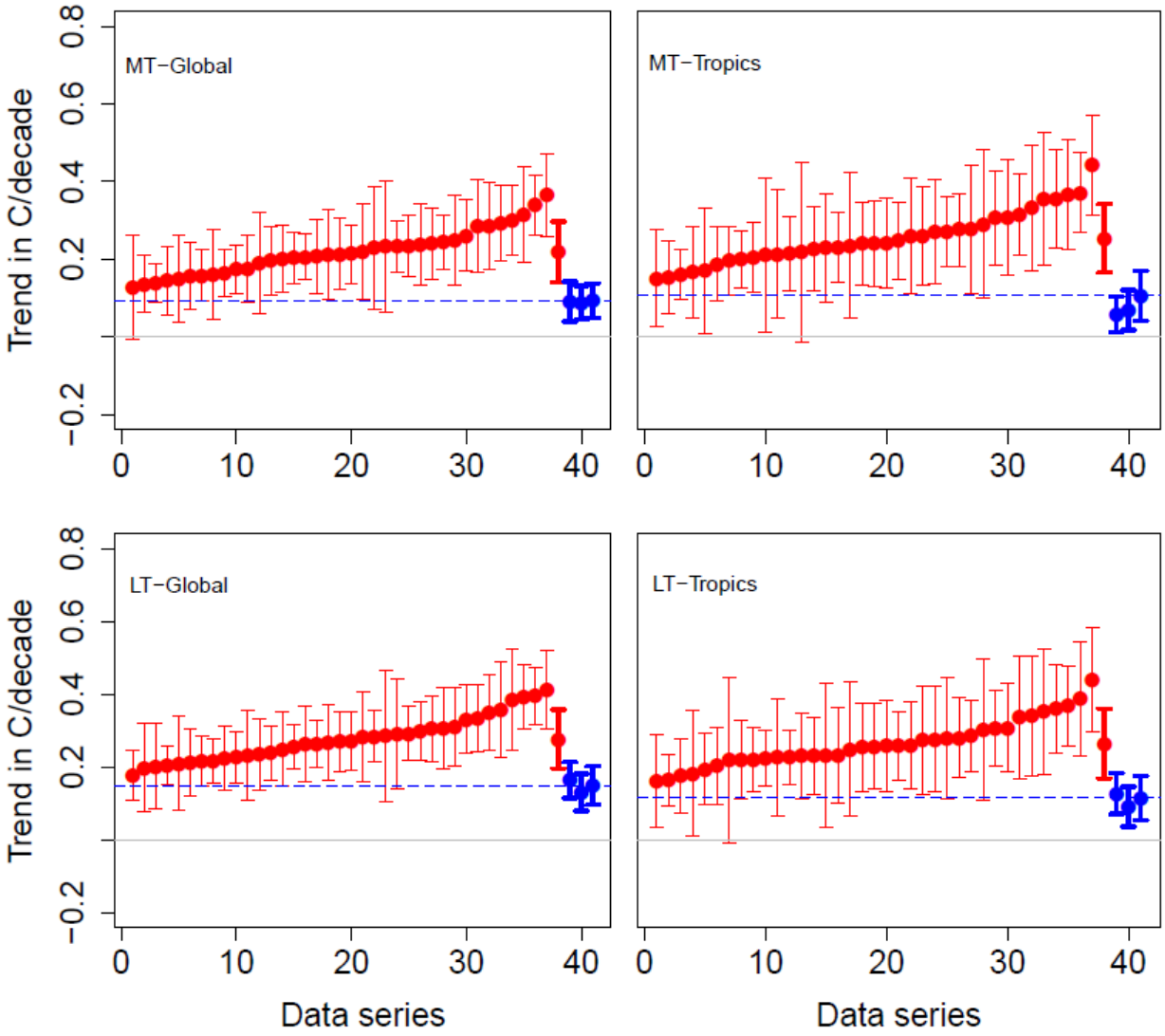

Figure 3. Trends and 95% CI's for individual models (red dots and thin bars), CMIP6 mean (red dot and thick bar) and observational series (blue). Horizontal dashed line shows mean satellite trend.

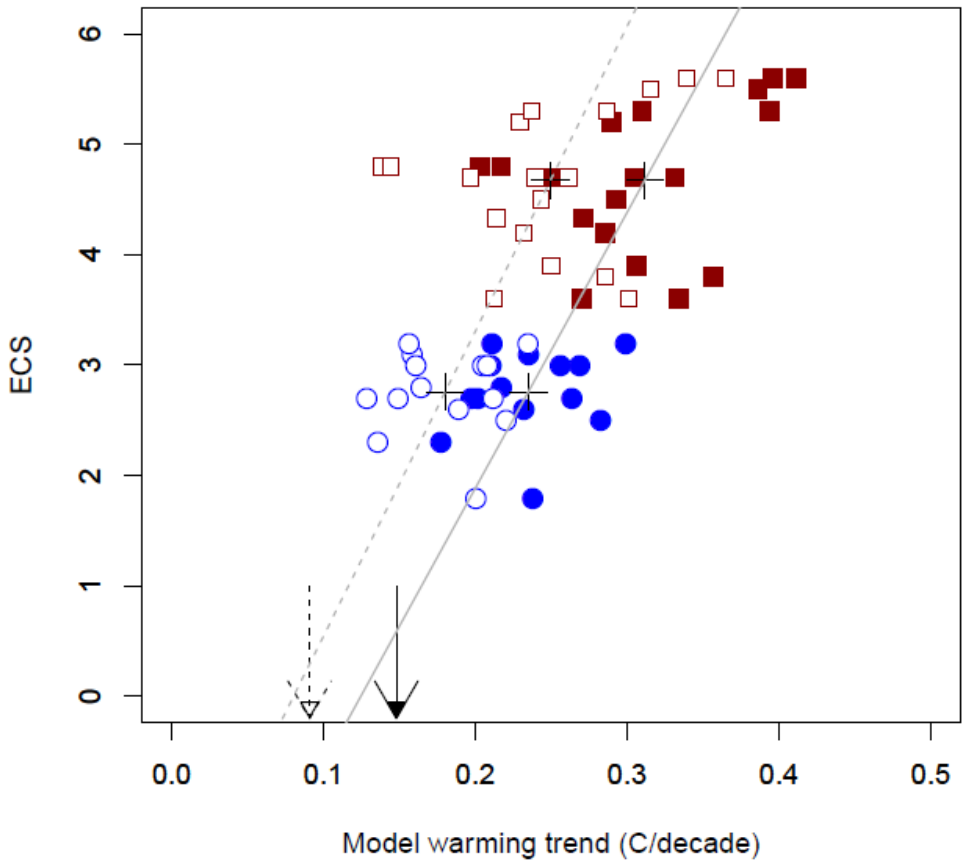

Figure 4. Model ECS values plotted against model warming trends. Red squares: high ECS group, Blue circles: low ECS group. Open shape: MT trend, closed shape: LT trend. Inverted triangles: mean observed LT trend (solid), mean observed MT trend (open).

	Dataset	Citation
Radiosonde	NOAA/RATPACvA2	Free et al. 2005
	RAOBCOREv1.7	Haimberger et al. 2012
	RICHv1.7	Haimberger et al. 2012
	UNSWv1.0	Sherwood and Nishant 2015
Satellite	RSSv4.0	Mears and Wentz 2016
	UAHv6.0	Spencer et al. 2017
	NOAA/STARv4.1	Zou and Wang 2011
	UWv1.0	Po-Chedley et al. 2015
Reanalyses	ERA-I	Dee et a. 2011
	ERA5	Hersbach, H. et al. 2018
	JRA-55	Kobayashi et al. 2015
	NASA/MERRA-2	Gelaro et al. 2017

Table 1. Listing of observational datasets utilized in this study.

Model Name	Run	Origin	ECS
ACCESS-CM2	r1i1p1f1_gn	Australia	4.7
ACCESS-ESM1-5	r1i1p1f1_gn	Australia	3.8
AWI-CM-1-1-MR	r1i1p1f1_gn	Germany	3.2
BCC-CSM2-MR	r1i1p1f1_gn	China	3.1
CAMS-CSM1-0	r1i1p1f1_gn	China	2.3
CanESM5	r1i1p1f1_gn	Canada	5.6
CanESM5-CanOE	r1i1p2f1_gn	Canada	5.6
CESM2	r3i1p1f1_gn	US NCAR	5.2
CESM2-WACCM	r1i1p1f1_gn	US NCAR	4.7
CIESM	r1i1p1f1_gr	China	
CNRM-CM6-1	r5i1p1f2_gr	France	4.8
CNRM-ESM2-1	r5i1p1f2_gr	France	4.8
E3SM-1-0	r1i1p1f1_gr	US DOE	5.3
EC-Earth3	r24i1p1f1_gr	Europe	4.2
EC-Earth3-Veg	r1i1p1f1_gr	Europe	4.3
FGOALS-f3-L	r1i1p1f1_gr	China	3.0
FGOALS-g3	r1i1p1f1_gn	China	3.0
FIO-ESM-2-0	r1i1p1f1_gn	China	
GFDL-CM4	r1i1p1f1_gr1	US NOAA	3.9
GFDL-ESM4	r1i1p1f1_gr1	US NOAA	2.7
GISS-E2-1-G	r1i1p1f1_gn	US NASA	2.7
HadGEM3-GC31-LL	r1i1p1f3_gn	UK	5.5
INM-CM4-8	r1i1p1f1_gr1	Russia	1.8
INM-CM5-0	r1i1p1f1_gr1	Russia	
IPSL-CM6A-LR	r1i1p1f1_gr	France	4.5
KACE-1-0-G	r1i1p1f1_gr	So. Korea	
MCM-UA-1-0	r1i1p1f2_gn	US U-AZ	3.6
MIROC6	r1i1p1f1_gn	Japan	2.6
MIROC-ES2L	r1i1p1f2_gn	Japan	2.7
MPI-ESM1-2-HR	r1i1p1f1_gn	Germany	3.0
MPI-ESM1-2-LR	r1i1p1f1_gn	Germany	2.8
MPI-ESM-1-2-HAM	r1i1p1f1_gn	Europe	
MRI-ESM2-0	r1i1p1f1_gn	Japan	3.2
NESM3	r1i1p1f1_gn	China	4.7
NorESM2-LM	r1i1p1f1_gn	Norway	2.5
NorESM2-MM	r1i1p1f1_gn	Norway	
SAM0-UNICON	r1i1p1f1_gn	So. Korea	3.6
UKESM1-0-LL	r1i1p1f2_gn	UK	5.3

Table 2: Models and runs used in this study. ECS denotes model Equilibrium Climate Sensitivity.

	Glob LT	CI	Glob MT	CI
ACCESS	0.250	0.103	0.197	0.089
ACCESS_E	0.357	0.132	0.286	0.119
AWI	0.299	0.079	0.235	0.078
BCC	0.235	0.098	0.158	0.066
CAMS	0.177	0.069	0.136	0.074
Can5	0.411	0.107	0.365	0.108
Can5OE	0.396	0.079	0.339	0.078
CE2r3	0.290	0.152	0.229	0.158
CE2_WAC	0.305	0.091	0.240	0.093
CIESM	0.351	0.103	0.294	0.098
CNRM_C61r5	0.203	0.053	0.139	0.049
CNRM_E2	0.217	0.068	0.144	0.089
E3SM	0.310	0.107	0.237	0.104
EC_E3	0.285	0.180	0.232	0.170
EC_E3V	0.271	0.082	0.214	0.075
FGOALS_f3	0.256	0.060	0.205	0.066
FGOALS_g3	0.269	0.104	0.208	0.095
FIO	0.264	0.064	0.206	0.059
GFDL-CM4	0.306	0.111	0.250	0.116
GFDL-ESM4	0.263	0.104	0.212	0.116
GISSE21G	0.197	0.121	0.129	0.135
HadGEM	0.386	0.139	0.316	0.123
INM48	0.238	0.075	0.200	0.086
INM50	0.225	0.088	0.175	0.087
IPSL6A	0.293	0.075	0.243	0.069
KACE	0.285	0.071	0.232	0.066
MCM_UA	0.334	0.093	0.301	0.091
MIROC	0.232	0.123	0.189	0.131
MIROC_2L	0.202	0.117	0.149	0.113
MPI_H	0.210	0.130	0.161	0.116
MPI_L	0.217	0.062	0.164	0.062
MPI_HAM	0.228	0.070	0.173	0.061
MRI_E2	0.211	0.092	0.156	0.087
NESM	0.331	0.093	0.261	0.091
NOR_LM	0.283	0.123	0.220	0.124
NOR_MM	0.224	0.118	0.171	0.123
SAM0	0.270	0.081	0.212	0.092
UK10LL	0.394	0.089	0.286	0.113
Model Avg	0.276	0.080	0.218	0.078
SONDE Avg	0.164	0.049	0.091	0.051
REANAL Avg	0.130	0.051	0.088	0.044
SAT Avg	0.150	0.053	0.093	0.044

Table 3: Trend coefficients and symmetric 95% CI widths for all model runs and average observations from each observing system, global LT and MT layers. Data span 1979-2014.

	Trop LT	CI	Trop MT	CI
ACCESS	0.231	0.106	0.214	0.096
ACCESS_E	0.388	0.156	0.367	0.142
AWI	0.281	0.110	0.272	0.091
BCC	0.221	0.109	0.196	0.090
CAMS	0.176	0.103	0.154	0.095
Can5	0.439	0.143	0.442	0.130
Can5OE	0.367	0.108	0.372	0.101
CE2r3	0.220	0.228	0.219	0.231
CE2_WAC	0.232	0.132	0.229	0.141
CIESM	0.352	0.173	0.355	0.172
CNRM_C61r5	0.224	0.078	0.200	0.074
CNRM_E2	0.195	0.099	0.166	0.117
E3SM	0.285	0.098	0.276	0.094
EC_E3	0.302	0.194	0.290	0.192
EC_E3V	0.254	0.121	0.240	0.110
FGOALS_f3	0.257	0.117	0.241	0.116
FGOALS_g3	0.230	0.117	0.227	0.109
FIO	0.258	0.093	0.247	0.099
GFDL-CM4	0.276	0.145	0.271	0.135
GFDL-ESM4	0.274	0.150	0.259	0.149
GISSE21G	0.232	0.199	0.211	0.199
HadGEM	0.340	0.166	0.332	0.163
INM48	0.228	0.074	0.230	0.089
INM50	0.221	0.088	0.205	0.090
IPSL6A	0.308	0.121	0.306	0.121
KACE	0.259	0.119	0.240	0.108
MCM_UA	0.361	0.122	0.356	0.128
MIROC	0.250	0.183	0.235	0.189
MIROC_2L	0.182	0.172	0.170	0.163
MPI_H	0.227	0.160	0.214	0.166
MPI_L	0.203	0.105	0.187	0.095
MPI_HAM	0.163	0.071	0.160	0.065
MRI_E2	0.162	0.127	0.151	0.125
NESM	0.306	0.104	0.314	0.105
NOR_LM	0.279	0.166	0.277	0.167
NOR_MM	0.211	0.221	0.196	0.226
SAM0	0.258	0.124	0.262	0.127
UK10LL	0.336	0.169	0.307	0.149
Model Avg	0.263	0.095	0.252	0.088
SONDE Avg	0.127	0.056	0.058	0.046
REANAL Avg	0.091	0.055	0.069	0.051
SAT Avg	0.115	0.061	0.106	0.065

Table 4: Trend coefficients and symmetric 95% CI widths for all model runs and average observations from each observing system, tropical LT and MT layers. Data span 1979-2014.

	Glob LT	**Glob MT**	**Trop LT**	**Trop MT**
>SONDE Avg	227.3	362.1	136.3	248.1
>REANAL Avg	262.9	200.2	129.0	147.1
>SAT Avg	97.1	118.7	68.5	70.1
Num > SAT Avg	24	26	18	20

Table 5: Vogelsang-Franses (2005) test scores for test of trend equivalence.