ROBUST MEASUREMENT OF THE DURATION OF THE GLOBAL WARMING HIATUS

Ross McKitrick
Department of Economics
University of Guelph

Revised version, July 3, 2014

Abstract: The IPCC has drawn attention to an apparent leveling-off of globally-averaged temperatures over the past 15 years or so. Measuring the duration of the hiatus has implications for determining if the underlying trend has changed, and for evaluating climate models. Here I propose a method for estimating the duration of the hiatus that is robust to heteroskedasticity and autocorrelation in the temperature series and to cherry-picking of endpoints. For the specific case of global warming I also add the requirement of spatial consistency between hemispheres. The method makes use of a HAC-robust trend variance estimator which is valid under autocorrelation of unknown form as long as the underlying series is trend stationary, which is the case for the data used herein. Application of the method shows that there is now a trendless interval of 19 years duration at the end of the HadCRUT4 surface temperature series, and of 16-26 years in the lower troposphere. Use of a simple AR1 trend model suggests a shorter hiatus of 14-20 years but is likely unreliable.

ROBUST MEASUREMENT OF THE DURATION OF THE GLOBAL WARMING HIATUS

1 INTRODUCTION

The 2013 report of the Intergovernmental Panel on Climate Change drew attention to a pause, or "hiatus" in what appears otherwise to be a long upward trend in globally-averaged temperatures (IPCC 2013, Ch. 9). Figure 1 shows the HadCRUT4 combined land and ocean surface temperature series over the 1850-2014 interval (Morice et al. 2012). The data are shown in the form of $^{\circ}$C "anomalies" or deviations from local averages. The underlying trend is clearly upward over the 20th century, but a leveling off at the end is visible. Figure 2 shows the lower troposphere (LT) series from the University of Alabama-Huntsville method (UAH, Spencer and Christy 1990) and Figure 3 shows the LT series using the RSS method of Mears and Wentz (2005). These observations run from January 1979 to April 2014 and show a similar leveling-off over the ending segment.

The IPCC does not estimate the duration of the hiatus, but it is typically regarded as having extended for 15 to 20 years. While the HadCRUT4 record clearly shows numerous pauses and dips amid the overall upward trend, the ending hiatus is of particular note because climate models project continuing warming over the period. Since 1990, atmospheric carbon dioxide levels rose from 354 ppm to just under 400 ppm, a 13% increase.[1] IPCC (2013) reports that of the 114 model simulations over the 15-year interval 1998 to 2012, 111 predicted warming. Fyfe et al. (2013) showed a similar mismatch in comparisons over a twenty

---

[1] Data from ftp://aftp.cmdl.noaa.gov/products/trends/co2/co2_mm_mlo.txt.

year time scale, with most models predicting $0.2 - 0.4$ °C/decade warming. Hence there is a need to address two questions: (i) how should the duration of the hiatus be measured? and (ii) does it indicate a potential inconsistency between observations and models? This paper focuses solely on the first question.

Several considerations need to be addressed in order for a duration measure to be robust. Consider a basic trend model

$$y_t = a + bt + e_t \qquad (1)$$

in which y_t is assumed to be trend-stationary (in other words, does not contain a unit root), a and b are coefficients to be estimated, t is a linear trend spanning $1,..,T$ and e_t is a covariance-stationary residual which may be described as autocorrelated or persistent to unknown dimensions. A series with a positive trend is one for which

$$0 < c_\alpha(\hat{b}|y_1, \dots, y_T), \qquad (2)$$

where c_α is the $(1 - 2\alpha)$ confidence interval around $\hat{b}$. An end-of-sample pause in the upward trend is defined using the duration parameter J, defined such that

$$J = \max(J \mid 0 \in c_\alpha(\hat{b}|y_J, \dots, y_T)) \qquad (3)$$

in other words the maximum length subset of the final observations of the series y_t such that the $(1 - 2\alpha)$ confidence interval around the trend term contains zero. Throughout this paper we use $\alpha = 0.025$ and hence a 95% confidence interval. Framed this way, two criteria can be set down to ensure robustness.

(a) The confidence interval must be heteroskedasticity and autocorrelation (HAC) robust. While the IPCC itself typically uses a simple AR1 process to model residuals, there is empirical evidence the that error processes exhibit higher-order autocorrelation and long term persistence (Rybski et

[3]

al. 2006, Lennartz and Bunde 2009). Tests of the series used herein all showed evidence of higher than first-order autocorrelation. Hence the estimator must be valid for longer memory structures.

(b) The longer is T, the more likely it is possible to find a single spurious value J that satisfies equation (3) even if, for all other values of J, the trend is significant. To avoid cherry-picking J, if it really marks the beginning of a pause in the trend, rather than being simply an artefact of the pseudo-cyclical behavior of persistent residuals, we require that condition (3) be observed at every point in the interval $J, \dots, T$.

A further condition can be added for the specific case studied here, namely globally-averaged surface temperatures, that there ought to be spatial consistency, namely

(c) conditions (3), (a) and (b) ought to hold jointly in both the Northern and Southern Hemispheres, since a pause in only one but not both would not be truly global, even if it caused the average of the two series to have an insignificant trend.

Combining (a)—(c), we obtain the following definition of hiatus length:

DEFINITION: A globally-averaged temperature series of length T with an upward trend coefficient b has a robust end-of-sample pause of length J_{MAX} if, for every interval beginning at $J = (J_{MAX}, \dots, T - m)$ and ending at time T, where m is the minimum-length duration of interest, $0 \in c_\alpha(\hat{b}|y_J, \dots, y_T)$, the $(1 - 2\alpha)$ confidence interval c_α is HAC-robust, and the value of J_{MAX} computed for both the Northern and Southern Hemispheres equals or exceeds that for the globe as a whole.

We apply this definition to the HadCRUT4, UAH and RSS series, obtaining hiatus durations of, respectively, 19, 16 and 26 years respectively. Were we to use an AR1 method the estimated durations would be 14, 14 and 20 years respectively.

2 APPLICATION OF HIATUS DEFINITION TO SURFACE AND TROPOSPHERIC TEMPERATURE DATA

2.1 THE VF TEST AND HAC-ROBUST CONFIDENCE INTERVALS

To address condition (a) we apply the HAC-robust trend confidence interval estimator of Vogelsang and Franses (2005, hereafter VF). HAC methods are very common in econometrics and have recently been applied in climatology (McKitrick et al. 2010). The VF approach falls in the class of non-parametric HAC estimators, but avoids the problem of bandwidth selection by setting it equal to the sample size T. The resulting estimator exhibits very little size distortion in comparison to other HAC estimators using limited bandwidth, while maintaining comparable power levels. However the estimator converges to a non-standard distribution so the critical values must either be simulated or computed using the bootstrap method. In the case of straightforward trend estimation, as in the current application, critical values are available in VF.

The trend coefficient is estimated on monthly data using OLS. The full derivation of the variance-covariance estimator, including asymptotics, is available in VF, and we will only explain the computational steps here. Since we will apply the test not only to the global series but also jointly to hemispheric trends, we will present it for the multivariate case.

A multivariate trend model is written

$$y_t^i = a^i + b^i t + e_t^i \tag{4}$$

where $i = 1, \ldots, N$, the number of time series for which trends are to be measured and tested. Denote $b = (b^1, \ldots, b^N)'$ and $e_t = (e_t^1, \ldots, e_t^N)'$. We are interested in testing null hypotheses of the form

$$H_0: Rb = r. \tag{5}$$

against alternatives $H_0: Rb \neq r$ where R and r are known restriction matrices of dimension q x N and q x 1 respectively, where q denotes the number of restrictions being tested. The matrix R is assumed to have full row rank. Robust tests of H_0 need to account for correlation across time, correlation across series, and conditional heteroskedasticity as summarized by the long run variance of e_t. This is defined as

$$\Omega = \Gamma_0 + \sum_{j=1}^{\infty}(\Gamma_0 + \Gamma_j')$$

where $\Gamma_j = E(e_t e_{t-j}')$ is the matrix autocovariance function of e_t. The VF statistic is constructed using the following estimator of Ω:

$$\widehat{\Omega}_T = \widehat{\Gamma}_0 + \sum_{j=1}^{T-1}(1 - \frac{j}{T})(\widehat{\Gamma}_j + \widehat{\Gamma}_j'), \tag{6}$$

where $\widehat{\Gamma}_j' = T^{-1}\sum_{t=j+1}^{T} \hat{e}_t \hat{e}_{t-j}'$. This is the Bartlett kernel nonparametric estimator of Ω using a bandwidth (truncation lag) equal to the sample size. The VF statistic for testing $H_0: Rb = r$ is given by

$$VF = (R\hat{b} - r)' \left[(\sum_{t=1}^{T} \tilde{d}_t^2)^{-1} R\widehat{\Omega}_T R' \right]^{-1} (Rb - 1)/q \tag{7}$$

where $\tilde{d}_t$ are the residuals from regressing the trend t on a constant. Equation (6) was originally proposed by Kiefer, Vogelsang and Bunzel (2000) although in the different but computationally identical form:

$$\widehat{\Omega}_T = 2T^{-2} \sum_{t=1}^{T-1} \hat{S}_t \hat{S}_t', \tag{8}$$

where $\hat{S}_t = \sum_{j=1}^{t} \hat{e}_j$. See Kiefer and Vogelsang (2002) for proof of the equivalence between (6) and (8).

[6]

In the univariate case (q=1), the square root of (7) can be interpreted as a t-type statistic and critical values are tabulated in VF. By treating the denominator as the robust standard error term $se(\hat{b}^i)$, a HAC-robust 95% confidence interval can be constructed

$$\hat{b}^i \pm se(\hat{b}^i) \cdot c_{0.025} \tag{9}$$

Where the 0.025 critical value $c_{0.025}$ is, from VF, 6.482. The confidence interval given by Equation (9) is robust to all forms of autocorrelation up to but not including a unit root. All the data series used herein were tested using a Dickey-Fuller test and in every case the unit root hypothesis was rejected.

The AR1 case is handled differently, using the Bartlett approximation of Santer et al. (2000), in which the lag-one AR coefficient r is estimated using the Cochrane-Orcutt iterative method, then the standard errors are scaled up by the factor $\sqrt{(1+r)/(1-r)}$.

2.2 RESULTS FOR HADCRUT4 SERIES

Table 1 presents the surface results. The data are monthly but we only consider annual increments for J, and only for durations of more than $m = 5$ years, as failure to reject the null of no trend in the vicinity of the end of the sample could, in the case of such a short interval, arise simply due to the small size of the subsample. The bolded line at $J = 19$, for the year 1995, shows the maximum value of J for which the lower bound of the confidence interval is negative, thus indicating that equation (3) is satisfied. Looking up the column from there it can be seen that condition (b) is also satisfied. In column 7, the number shown is the VF score on the test that the separate NH and SH trends are jointly zero. The 5% critical value for the $q = 2$ case is 40.68, which is only exceeded at $J = 23$, therefore the additional condition (c) is also met in this case. Hence, for the surface HadCRUT4 data, $J_{MAX} = 19$.

[7]

Figure 4 shows these results in graphical form. It is clear that the confidence intervals widen at the end of the sample as J increases and the number of degrees of freedom decline, but not before first shrinking somewhat, indicating that at the computed value of J_{MAX} the number of degrees of freedom is not the limiting factor on the size of the confidence interval.

2.3 RESULTS FOR UAH AND RSS SERIES

Tables 2 and 3 present the UAH and RSS series, respectively. For UAH, the bolded line at $J = 16$, corresponding to the year 1998, shows the maximum value of J for which the lower bound of the confidence interval is negative. Looking up the column from there it can be seen that condition (b) is also satisfied. In column 7, the VF score on the joint NH and SH trend tests only exceed 40.68 at $J = 19$, therefore the additional condition (c) is also met in this case. Hence, for the lower troposphere UAH data, $J_{MAX} = 16$. Figure 5 shows the UAH results in graphical form. For the RSS results, comparable examination yields $J_{MAX} = 26$, corresponding to a start date at 1988. The results are shown in Figure 6.

3 CONCLUSIONS

We propose a robust definition for the length of the pause in the warming trend over the closing subsample of surface and lower tropospheric data sets. The length term J_{MAX} is defined as the maximum duration J for which a valid (HAC-robust) trend confidence interval contains zero for every subsample beginning at J and ending at $T - m$ where m is the shortest duration of interest. This definition was applied to surface and lower tropospheric temperature series, adding in the requirement that the southern and northern hemispheric data must yield an identical or larger value of J_{MAX}. In the surface data we compute a hiatus length of 19 years, and in the lower tropospheric data we compute a hiatus length of 16 years in the UAH series and 26 years in the RSS series. J_{MAX} estimates based on an AR1 estimator are

[8]

lower but likely incorrect since higher-order autocorrelation exists in the data. Overall this analysis confirms the point raised in the IPCC report (2013) regarding the existence of the hiatus and adds more precision to the understanding of its length.

4   REFERENCES

Fyfe, J.C., N.P. Gillett and F.W. Zwiers, 2013: Overestimated global warming over the past 20 years. *Nature Climate Change*, 3, 767-769, doi:10.1038/nclimate1972

Intergovernmental Panel on Climate Change (IPCC) (2013). *Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the IPCC*. Cambridge: Cambridge University Press.

Kiefer, N.M and T.J. Vogelsang (2002) "Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation," *Econometrica*, 70, 2093-2095.

Kiefer, N.M., T.J. Vogelsang & H. Bunzel (2000) "Simple robust testing of regression hypotheses," *Econometrica* 68, 695–714.

Lennartz, S., and A. Bunde (2009), Trend evaluation in records with long-term memory: Application to global warming, Geophys. Res. Lett., 36, L16706, doi:10.1029/2009GL039516.

Mears, C. A., and F. J. Wentz. (2005). The effect of diurnal correction on satellite-derived lower tropospheric temperature. *Science* 309: 1548-1551.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones (2012), Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset, J. Geophys. Res., 117, D08101, doi:10.1029/2011JD017187.

Rybski, D., A. Bunde, S. Havlin, and H. von Storch (2006), Long-term persistence in climate and the detection problem, Geophys. Res. Lett., 33, L06718, doi:10.1029/2005GL025591.

Santer B.D., T.M.L. Wigley, J.S. Boyle, D.J. Gaffen, J.J. Hnilo, D. Nychka, D.E. Parker and K.E. Taylor (2000). Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *Journal of Geophysical Research* 105: 7337–7356.

Spencer, R.W. and J.C. Christy (1990), "Precise monitoring of global temperature trends from satellites," *Science* 247:1558—1562.

Vogelsang, T.J. and P. H. Franses (2005) "Testing for Common Deterministic Trend Slopes," *Journal of Econometrics* 126, 1-24.
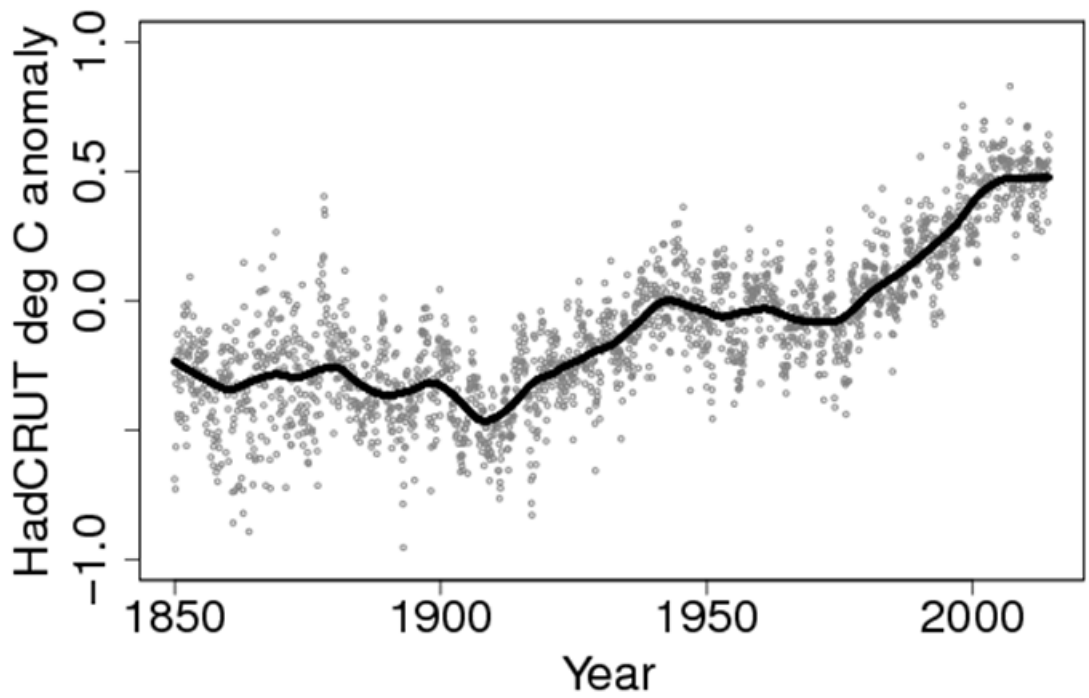
Figure 1. Globally-averaged HadCRUT4 surface temperature anomalies, January 1850 to April 2014. Dark line is lowess smoothed with bandwidth parameter = 0.09.
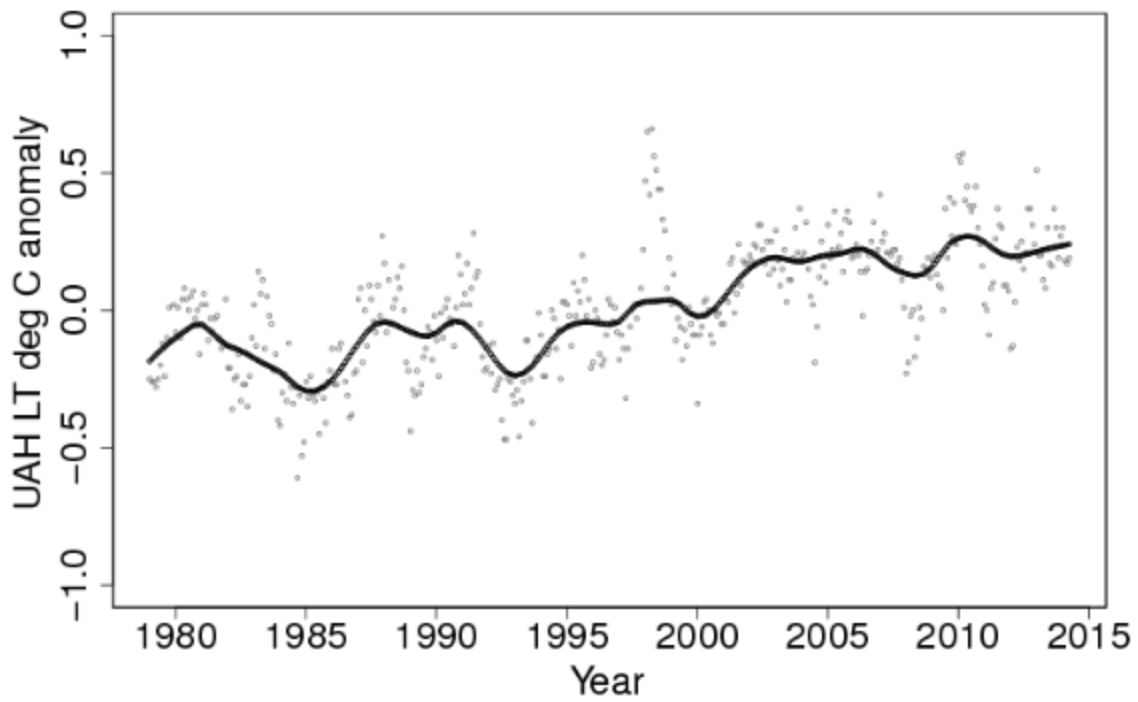
Figure 2. Globally-averaged UAH lower troposphere temperature anomalies, January 1979 to April 2014. Dark line is lowess smoothed with bandwidth parameter = 0.11.
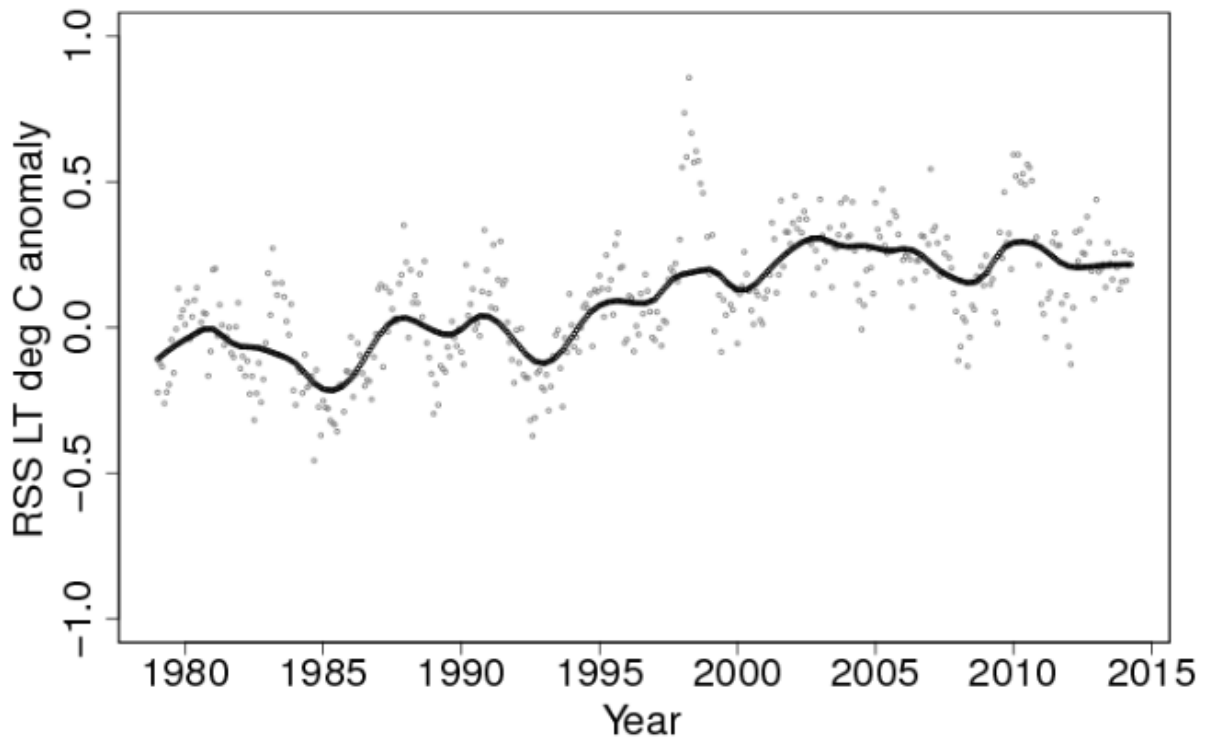
Figure 3. Globally-averaged RSS lower troposphere temperature anomalies, January 1979 to April 2014. Dark line is lowess smoothed with bandwidth parameter = 0.11.

Figure 4. Trend magnitudes (black dots) and 95% robust confidence intervals (gray lines) around trend in HadCRUT4 global series from start date indicated to April 2014.

Figure 5. Trend magnitudes (black dots) and 95% robust confidence intervals (gray lines) around trend in UAH lower troposphere series from start date indicated to April 2014.

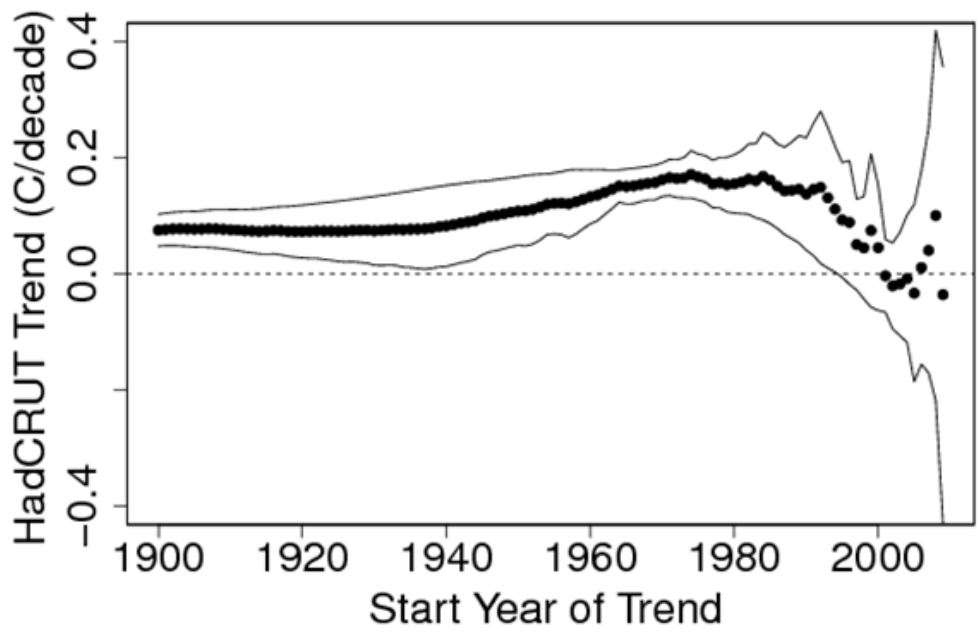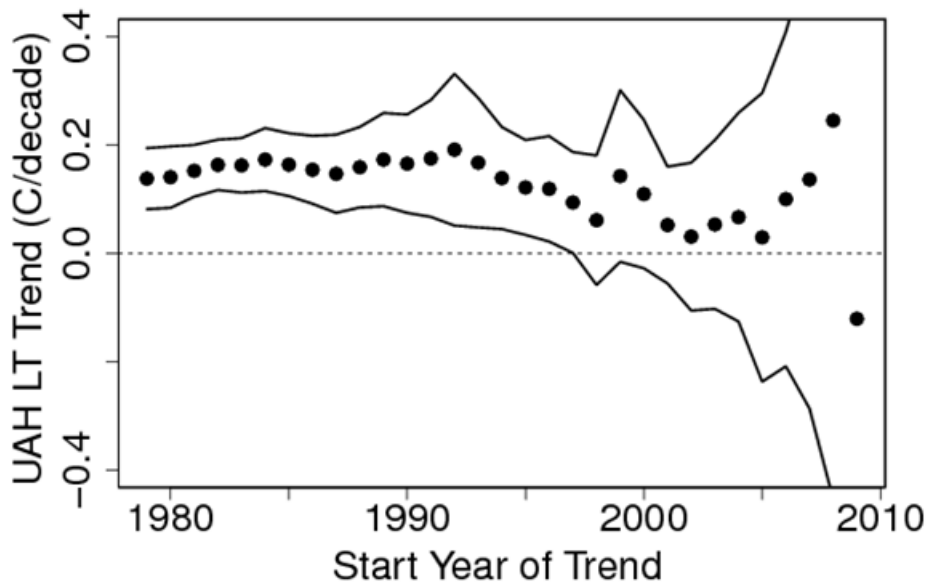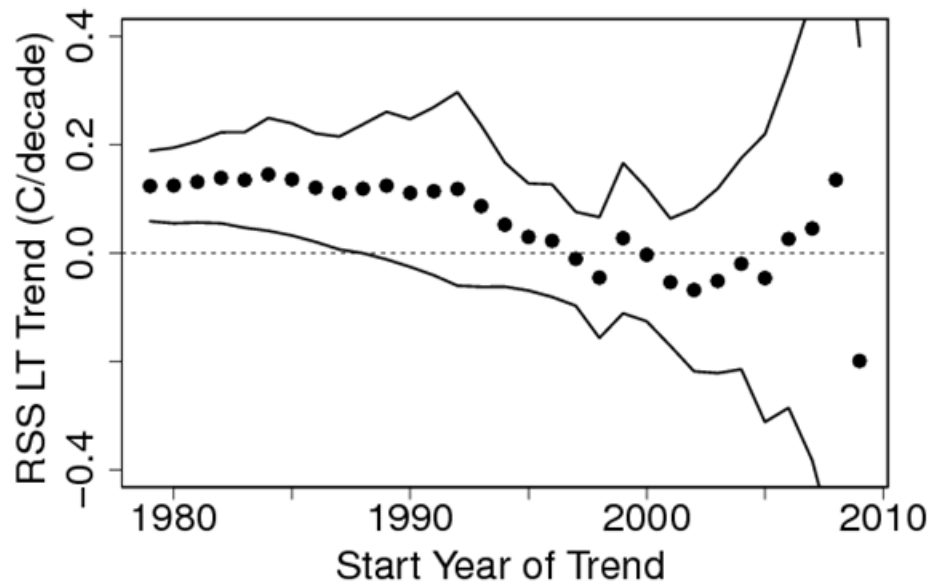Figure 6. Trend magnitudes (black dots) and 95% robust confidence intervals (gray lines) around trend in RSS lower troposphere series from start date indicated to April 2014.

J	Start Year	CI Lower Bound	Estimated Global Trend	CI Upper Bound	VF on Global Trend	VF on test SH=NH=0
5	2009	-0.4290	-0.0362	0.3566	0.3561	4.4514
6	2008	-0.2185	0.1002	0.4189	4.1514	4.1568
7	2007	-0.1710	0.0400	0.2509	1.5084	3.6456
8	2006	-0.1560	0.0105	0.1770	0.1659	3.8931
9	2005	-0.1862	-0.0337	0.1189	2.0477	7.7247
10	2004	-0.1183	-0.0086	0.1010	0.2601	5.8458
11	2003	-0.1065	-0.0179	0.0708	1.7087	2.2853
12	2002	-0.0951	-0.0211	0.0528	3.4276	1.6879
13	2001	-0.0664	-0.0037	0.0589	0.1499	0.6314
14	2000	-0.0632	0.0448	0.1528	7.2274	3.6809
15	1999	-0.0578	0.0743	0.2064	13.2871	6.7020
16	1998	-0.0435	0.0445	0.1326	10.7528	6.9820
17	1997	-0.0282	0.0501	0.1284	17.1944	10.7270
18	1996	-0.0183	0.0883	0.1949	28.8502	15.3172
19	**1995**	**-0.0063**	**0.0925**	**0.1913**	**36.8666**	**19.8734**
20	1994	0.0026	0.1111	0.2195	44.0516	25.4490
21	1993	0.0096	0.1306	0.2517	48.9188	29.5653
22	1992	0.0175	0.1485	0.2795	54.0100	33.1299
23	1991	0.0312	0.1450	0.2589	68.1903	42.0591
24	1990	0.0405	0.1372	0.2339	84.5609	48.9278
25	1989	0.0527	0.1453	0.2378	103.4453	61.4013
26	1988	0.0604	0.1436	0.2268	125.1039	73.7079
27	1987	0.0672	0.1426	0.2179	150.4521	92.7571
28	1986	0.0777	0.1505	0.2234	179.2456	110.6976
29	1985	0.0862	0.1611	0.2360	194.2570	112.6455
30	1984	0.0925	0.1677	0.2429	208.8947	117.1244

Table 1. Results for sequential tests of surface trend hiatus. J: length of pause being considered. Start year: implied beginning year of pause. CI Lower Bound and Upper Bound: 95% confidence interval limits from equation (8). Estimated Global Trend: $\hat{b}_1$. *VF* on Global Trend: *VF* from equation (9) applied to HadCRUT4 series. The 5% critical value is 41.53. *VF* on test SH=NH=0: VF test score from equation (8) on null hypothesis $H_0: b_1^1 = b_1^2 = 0$, where the two series are the NH and SH averages from HadCRUT4. The 5% critical value is 40.68.

J	Start Year	CI Lower Bound	Estimated Global Trend	CI Upper Bound	VF on Global Trend	VF on test SH=NH=0
5	2009	-0.7097	-0.1211	0.4675	1.7794	1.9804
6	2008	-0.4540	0.2449	0.9437	5.1573	2.9882
7	2007	-0.2868	0.1360	0.5587	4.3457	4.7759
8	2006	-0.2087	0.0998	0.4082	4.3954	9.1960
9	2005	-0.2371	0.0292	0.2955	0.5062	9.0689
10	2004	-0.1260	0.0665	0.2591	5.0167	6.3076
11	2003	-0.1027	0.0530	0.2088	4.8697	7.4862
12	2002	-0.1058	0.0306	0.1669	2.1112	1.9953
13	2001	-0.0558	0.0519	0.1596	9.7610	6.4981
14	2000	-0.0278	0.1092	0.2462	26.7072	13.2031
15	1999	-0.0159	0.1424	0.3007	34.0090	16.9742
16	**1998**	**-0.0586**	**0.0609**	**0.1804**	**10.9176**	**5.6454**
17	1997	0.0006	0.0936	0.1866	42.5276	21.8670
18	1996	0.0215	0.1187	0.2160	62.6642	30.9915
19	1995	0.0338	0.1213	0.2088	80.8337	40.0284
20	1994	0.0445	0.1386	0.2328	91.0876	45.1304
21	1993	0.0475	0.1670	0.2865	82.0301	41.8279
22	1992	0.0509	0.1910	0.3310	78.1324	44.6887
23	1991	0.0670	0.1748	0.2826	110.5101	60.5313
24	1990	0.0744	0.1652	0.2560	139.1791	73.5291
25	1989	0.0867	0.1727	0.2587	169.5376	91.5560
26	1988	0.0844	0.1587	0.2330	191.6675	96.7039
27	1987	0.0744	0.1465	0.2186	173.4259	89.0209
28	1986	0.0912	0.1540	0.2167	253.0072	128.0247
29	1985	0.1054	0.1634	0.2213	334.1565	165.9710
30	1984	0.1145	0.1727	0.2309	370.0354	194.8356

Table 2. Results for sequential tests of UAH trend hiatus. *J*: length of pause being considered. Start year: implied beginning year of pause. CI Lower Bound and Upper Bound: 95% confidence interval limits from equation (8). Estimated Global Trend: $\hat{b}_1$. *VF* on Global Trend: *VF* from equation (9) applied to UAH series. The 5% critical value is 41.53. *VF* on test SH=NH=0: VF test score from equation (8) on null hypothesis $H_0: b_1^1 = b_1^2 = 0$, where the two series are the NH and SH averages from UAH. The 5% critical value is 40.68.

J	Start Year	CI Lower Bound	Estimated Global Trend	CI Upper Bound	VF on Global Trend	VF on test SH=NH=0
5	2009	-0.7800	-0.1992	0.3815	4.9457	10.7215
6	2008	-0.5481	0.1349	0.8179	1.6388	0.8505
7	2007	-0.3829	0.0451	0.4731	0.4669	2.1173
8	2006	-0.2858	0.0259	0.3376	0.2897	2.9083
9	2005	-0.3121	-0.0464	0.2192	1.2830	7.6291
10	2004	-0.2147	-0.0199	0.1750	0.4371	2.9538
11	2003	-0.2218	-0.0516	0.1187	3.8551	9.8512
12	2002	-0.2184	-0.0684	0.0817	8.7247	5.5605
13	2001	-0.1715	-0.0541	0.0634	8.9095	4.4892
14	2000	-0.1262	-0.0034	0.1194	0.0322	0.6432
15	1999	-0.1114	0.0273	0.1659	1.6247	0.8209
16	1998	-0.1569	-0.0455	0.0660	6.9916	3.5254
17	1997	-0.0975	-0.0109	0.0757	0.6657	1.8102
18	1996	-0.0813	0.0225	0.1263	1.9751	6.0081
19	1995	-0.0694	0.0295	0.1283	3.7330	6.4918
20	1994	-0.0623	0.0521	0.1665	8.7136	8.3488
21	1993	-0.0625	0.0865	0.2354	14.1598	12.9572
22	1992	-0.0605	0.1181	0.2966	18.3772	15.4799
23	1991	-0.0413	0.1139	0.2691	22.6313	19.4612
24	1990	-0.0258	0.1106	0.2469	27.6416	23.7481
25	1989	-0.0120	0.1242	0.2605	34.9387	29.3777
26	**1988**	**-0.0005**	**0.1184**	**0.2373**	**41.6345**	**35.4573**
27	1987	0.0064	0.1106	0.2147	47.3827	42.9785
28	1986	0.0203	0.1204	0.2204	60.8263	49.2883
29	1985	0.0322	0.1357	0.2391	72.2039	52.1000
30	1984	0.0408	0.1449	0.2490	81.3904	55.7195

Table 3. Results for sequential tests of RSS trend hiatus. J: length of pause being considered. Start year: implied beginning year of pause. CI Lower Bound and Upper Bound: 95% confidence interval limits from equation (8). Estimated Global Trend: $\hat{b}_1$. VF on Global Trend: VF from equation (9) applied to RSS series. The 5% critical value is 41.53. VF on test SH=NH=0: VF test score from equation (8) on null hypothesis $H_0: b_1^1 = b_1^2 = 0$, where the two series are the NH and SH averages from RSS. The 5% critical value is 40.68.