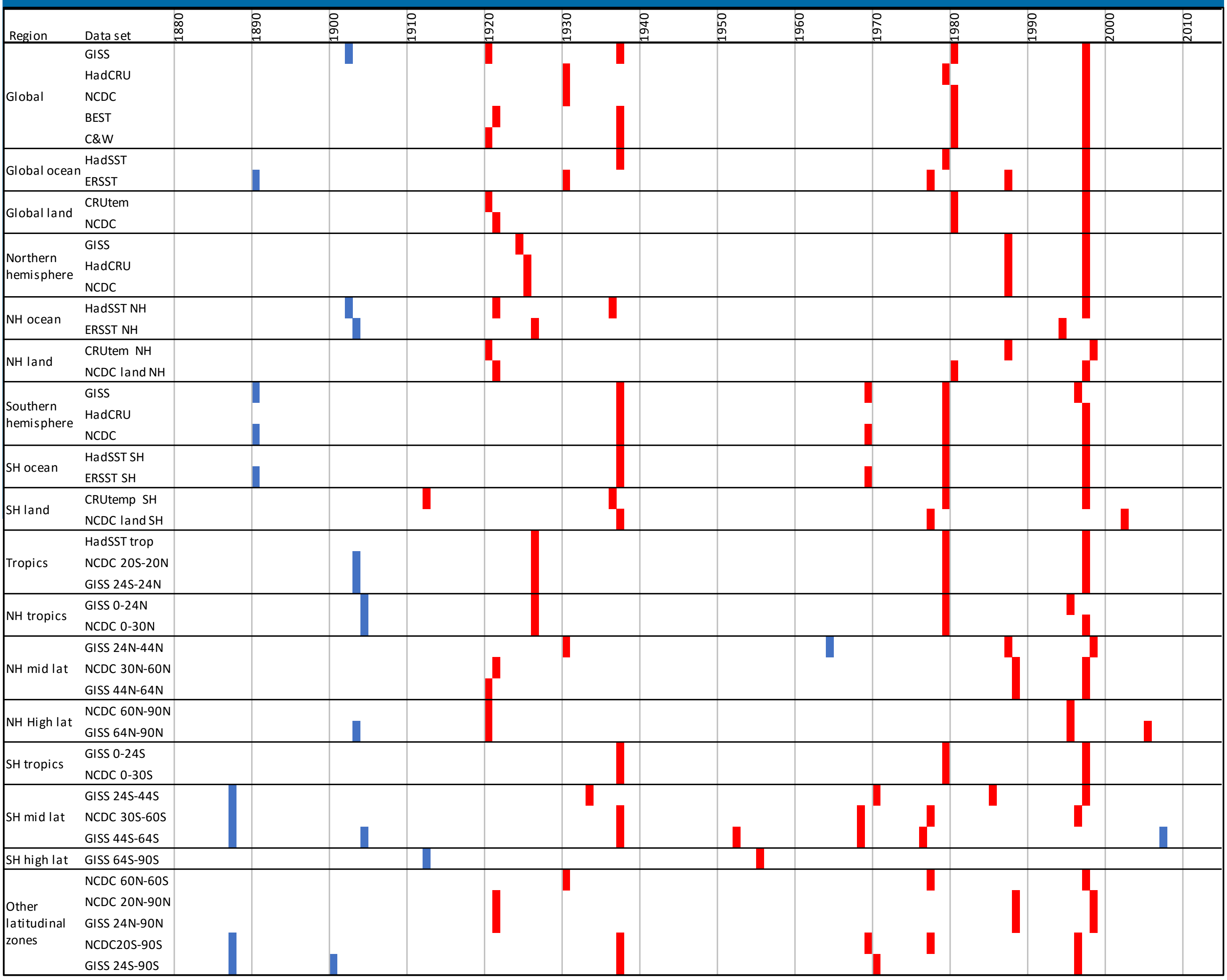
Analysing steps in global and regional observed air temperature

Roger N Jones and James H Ricketts

Climate Change Working Paper 34

Victoria Institute of Strategic Economic Studies



Key words: air temperature, global warming, decadal variability, climate change, regime change, step change, nonlinear dynamics

Acknowledgements

JR is the holder of a Victoria University postgraduate research scholarship. Data sources include the Met Office Hadley Centre, National Aeronautics and Space Administration Goddard Institute for Space Studies and United States National Climatic Data Center, Berkeley Earth, Cowtan and Way and the Australian Bureau of Meteorology. D Kelly O'Day for the programming and macros adapted to conduct multiple trend analysis and graphics. Useful discussions were held with Celeste Young, Peter Sheehan, Roger Bodman, Penny Whetton, Kevin Hennessy, Kris Ebi, Ben Preston, Jim Bowler, Rod Marsh and the big shift group. Roger Bodman reviewed the manuscript.

Suggested citation

Jones, R.N. and Ricketts, J.H. (2015) Analysing steps in global and regional observed air temperature. Climate Change Working Paper No. 34, Victoria Institute of Strategic Economic Studies, Victoria University, Melbourne, Australia.

Publishing Notes

This paper and a companion paper have been merged and added to and contribute to Reconciling the signal and noise of atmospheric warming on decadal timescales, Climate Change Working Paper. No. 38.

ISBN: 978-1-86272-722-9

Victoria Institute of Strategic Economic Studies
Victoria University
Po Box 14428 Melbourne
Victoria Australia 8001
Phone +613 9919 1340
Fax +613 9919 1350
vises@vu.edu.au

Abstract

If externally forced and internally generated climate variability interact on decadal timescales, it will most likely manifest as rapid shifts in warming. Here, the multi-step bivariate test is applied to observed global and regional surface and satellite temperatures to determine the relative roles of shifts and trends. Statistically significant shifts dominate. Since the mid-20th century, most of the observed warming has taken place in four events: in 1979/80 and 1997/98 at the global scale, 1987/88 in the northern hemisphere and 1968/70 in the southern hemisphere. Warming from internal trends is less than 40% of the total for four of five global records 1880–2013/14. Surface and tropospheric satellite temperature records (1979–2014), undergo similar step changes in 1987/88 and 1997, with limited internal trends between steps. Analyses for Texas, central England and south-eastern Australia show step-like warming that can be attributed to external forcing and coincides with these larger-scale changes. Such evidence that the climate system is exhibiting complex system behaviour on decadal timescales is important information for characterising and managing future climate risk.

Introduction

The climate research community's preferred tool for analysing the changing climate is ordinary least-squares linear trend analysis (IPCC, 2001, 2007, 2013). This approach has been highly successful in measuring long-term climate trends, for detection and attribution and for developing and communicating climate projections. Its simplicity satisfies the principle of Occam's razor, but this is simplicity of method and not necessarily theory. Trend analysis is a suitable tool for showing how <u>much</u> climate has changed but does not necessarily show <u>how</u> climate will change (Jones et al., 2013). This distinction is extremely important for decision making.
A major gap in understanding the physical process of climate change, is how climate changes over decadal timescales (Solomon et al., 2011). The scientific literature contains two competing hypotheses that link anthropogenic climate change and variability (Corti et al., 1999;Hasselmann, 2002):

1. Anthropogenic climate change occurs independently of climate variability (*H1*).
2. Anthropogenic climate change interacts with climate variability (*H2*).

H1 is generally interpreted as a monotonic trend driven by gradual climate forcing that is mediated by climate variability (Swanson et al., 2009;Rahmstorf et al., 2012;Zhou and Tung, 2013), producing a straight line or curve with fluctuations around the trend. Physically, this describes climate change as a gradual process, which for temperature, results in gradual warming. Decadal climate variability that manifests as specific regime shifts will imprint on long-term trends as changes in trend. *H1* is generally measured by various forms of time series analysis of which linear trend analysis is the most prominent. The extracted trend is interpreted as the signal, and the residual values as climate variability or 'noise' (North et al., 1995;Hegerl and Zwiers, 2011;Santer et al., 2011).

H2 is usually conceived as interactions of climate change and variability that produce significant non-linear responses (Corti et al., 1999;Solomon et al., 2011;Kirtman et al., 2013). In this formulation, the signal, expressed in terms of key variables such as temperature, would be non-linear. The main problem in detecting and attributing such a signal has been a robust method for doing so (Rodionov, 2005;Reeves et al., 2007;Overland et al., 2008).

Here we apply the Maronna-Yohai (1978) bivariate test within a rule-based procedure designed to detect multiple step changes in historical time series of air temperature – the multi-step bivariate test. The bivariate test calculates the step change with the implicit assumption that the input data is serially independent with no significant trend in the period either side. Under a stationary climate, this will be close to the mean difference between the two periods. If there is an underlying step and trend process going on (or trend with steps), then the test will register the full step but some of that difference may be due to internal trends. The presence of 'red noise' – autocorrelated data – may also register as a shift (see Supplementary Information).

Time series tested here are mean annual global air temperature anomalies from five groups (GISS, HadCRU, NCDC, C&W and BEST), hemispheric temperatures from three groups (HadCRU, NCDC and GISS) and zonal temperatures from two groups (NCDC and GISS) to see how prevalent step changes are, whether they coincide across different records and to investigate the relationship between step changes and trends. Tropospheric satellite temperatures from two groups (RSS and UAH) are also tested. The specific records used are detailed in the Supplementary Information.

These tests deal mainly with detection. Attribution of step changes is carried out for three regions using the bivariate test combined with linear inverse analysis to attribute step-wise anthropogenic warming using a method applied by Jones (2012).

Method

A multi-step application of the Maronna-Yohai (1978) bivariate test has been constructed to detect step changes in climate time series (see Supplementary Information for details). Previously, the bivariate test has been used to detect inhomogeneities in climate variables, decadal regime shifts in climate-related data and step changes in a wide range of climatic time series (Jones et al., 2013). Here, the test has been modified to account for multiple step changes. The main purpose of automating the test is to improve its objectivity and robustness by using a predefined set of rules.

The test adapts the formulation of Bücher and Dessens (1991) and tests a single serially-independent variate (x_i) against a reference variate (y_i) using a random time series following Vivès and Jones (2005). The important outputs of the test in a time series of length N are, (1) The T_i statistic which is defined for times $i < N$, (2) the T_{i0} value which is the maximum T_i value, (3) i_0, the time associated with T_{i0}, (4) shift at that time, and (5) *Pr*, the probability of zero shift. Note that i_0 is the last year prior to the change.

A run of an analysis of a single time series consists of a *screening pass*, followed by a *convergent pass*, which incorporates some rules. In both passes, since the reference variate is a flat random sequence, we apply a *resampling test*. The resampling test is applied to every segment tested and undertakes 100 trials of the bivariate test. The screening pass starts from the most significant shift in a time series, determined using the resampling test and, if $Pr<0.01$, the series is divided into shorter time series either side of the shift and retested until all shifts have been detected. The convergent pass then groups segments to remove shorter term perturbations, if any, and to determine the most robust shift dates in the segments selected. The rules are described in the Supplementary Information. The convergent process is repeated until a stable set of step changes is produced.

The above analysis is run 100 times. A stable set of results will produce 100 identical solutions and less stable results will produce two or more alternatives (see Supplementary Information for more detail). The most frequent configuration is selected as the most stable. Most historical temperature records analysed contain one or two stable configurations for surface temperature and zero or one for satellite temperature.
Mean annual data is considered serially independent. Deseasonalised quarterly and monthly data can be used to locate a shift within a year, but is not serially independent, so is used here in combination with the t-test associated with the identified shift date to assess significance.

Local attribution of step changes uses a technique described in Jones (Jones, 2012). The basic methodology is suitable for continental mid-latitude areas where annual average maximum temperature (*Tmax*) is correlated with total rainfall (*P*), and minimum temperature (*Tmin*) is correlated with *Tmax* (Power et al., 1998;Nicholls et al., 2004;Karoly and Braganza, 2005). For Central England Temperature, a largely maritime climate, diurnal temperature range (DTR) is assessed against precipitation instead of *Tmax*. The method uses the following steps:

1. Homogenous regional average data is obtained for *Tmax*, *Tmin* and *P*.

2. A period of stationary climate is calculated by testing when the relationship between *Tmin* and *Tmax* undergoes a statistically significant step change. The relationship between *Tmax* and *P* will change at the same, or later date.
3. Linear regressions are calculated between each pair (*Tmax*/*P* and *Tmin*/*Tmax*) for the stationary period.
4. Anthropogenic regional warming is estimated for the non-stationary period using these regressions.
5. The results are tested for step changes.

Results

Global and zonal temperatures

Statistically significant step changes in global and zonal temperatures show a great deal of structure over the 1880–2014 time period. All series were tested from their earliest recorded date (1850 and 1880) and results from 1880–2014 are shown. Collectively, downward steps occur in the late 19th and early 20th century, upward steps between 1912 and 1938 with one downward step in 1964. From 1968, upward steps dominate, with one exception in the high southern hemisphere (SH) latitudes in 2007 (Figure 1).

Global step changes form largely unbroken vertical lines in Figure 1. The 1997 step change is global, with some regional variation in 1996 and 1998. A step change occurs in in 1979/80 globally and in many regions except the northern hemisphere mid and high latitudes. All other step changes occur across more limited regions, with some being confined solely to land or to ocean. These show that the mean global record is an integration of more regional records, combining regional extent and the size of the shift.

Upward step changes occur in 1997 and 1979/80 in all five global records. The 1997 step is the largest at 0.31±0.01°C. The 1979/80 is the next largest at 0.22±0.03°C. The greater variation in size is subject to the timing and size of previous steps and trends. In the first half of the 20th century, three records show positive steps in 1920/21 and in 1937, and two in 1930 (Figure 1). The GISS record also shows a downward step in 1903, coinciding with the northern hemisphere (NH) ocean, tropics and perhaps southern hemisphere. The global records fall into two groups based on their early 20th century analysis: GISS, BEST, C&W in one group and HadCRU and NCDC in the other. The anomaly averaged from all five records shows upward step changes in 1930, 1979 and 1997.

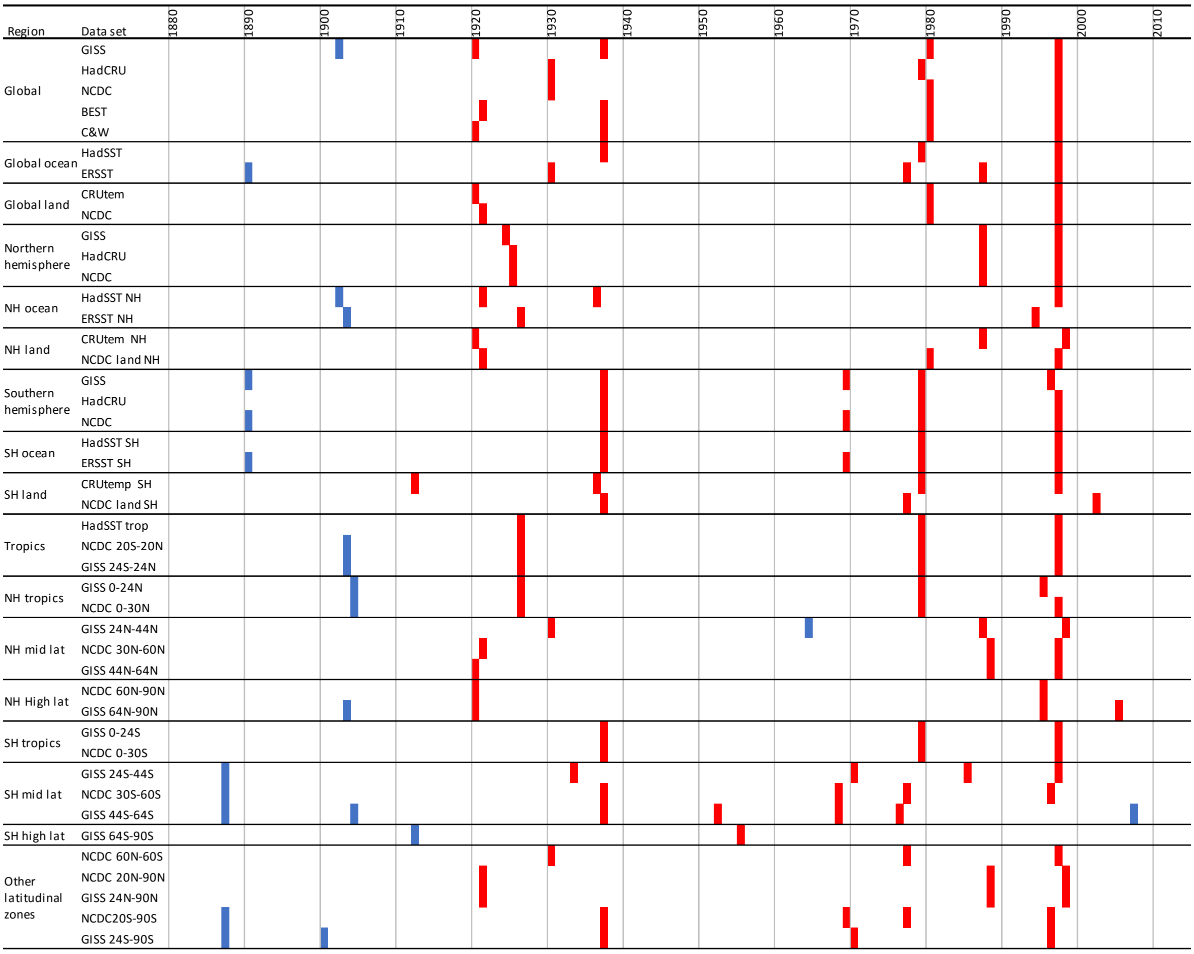
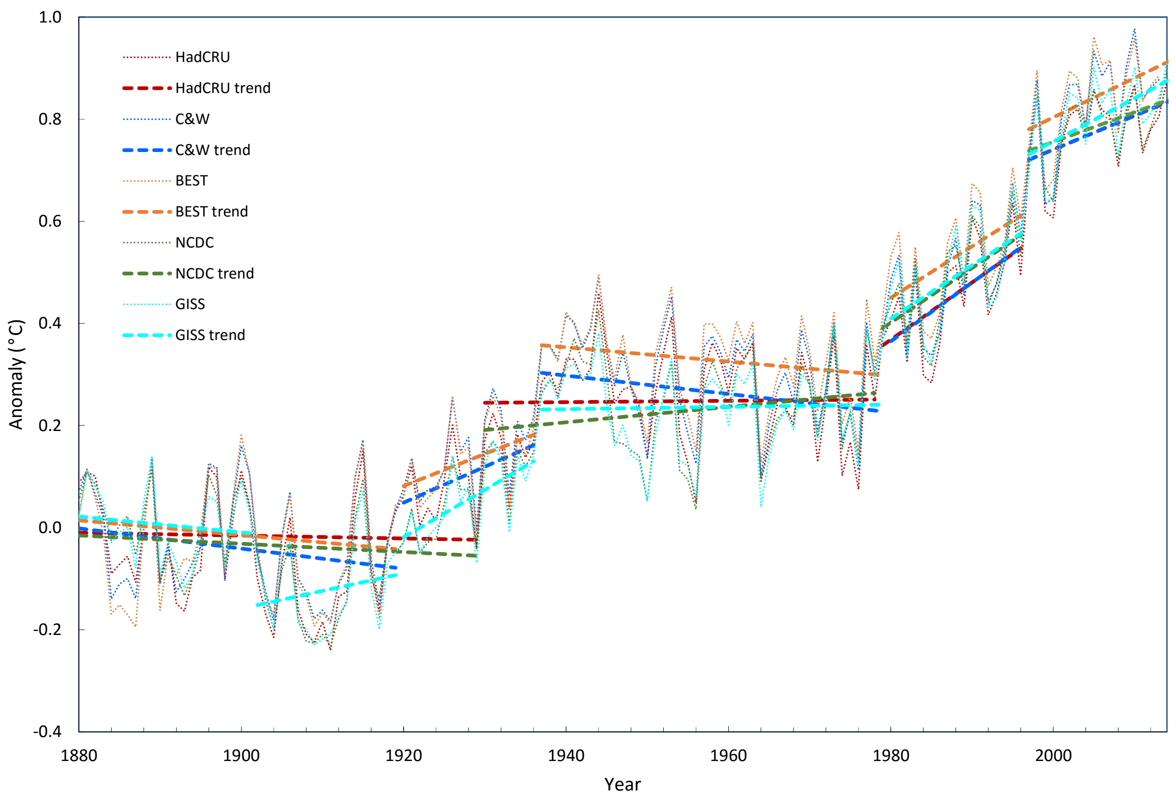


Figure 1: Dates of statistically significant step changes ($p<0.01$) 1880–2014, for a range of mean annual temperature records. Downward steps are blue and upward red. Records are sourced from Goddard Institute of Space Studies (GISS), the Hadley Centre and Climate Research Unit (HadCRU (land and ocean), HadSST (ocean), CRUtem (land)), National Climatic Data Center (NCDC (land, land and ocean), ERSST (ocean)), Berkeley Earth Surface Temperature (BEST) and Cowtan and Way (C&W). See Supplementary Information for details.

Differences emerge in ocean and land records. The global HadSST (HadCRU) record shifts in 1937, 1979 and 1997, whereas the ERSST (NCDC) record shifts in 1930, 1977, 1987 and 1997. Global land records from both CRU and NCDC shift in 1920/21, 1980 and 1997. Northern hemisphere land and ocean step changes are consistent across three records: in 1924/25, 1987 and 1997. The NH ocean shows a downward step in 1902/03 and is less consistent between the two records tested for subsequent upward steps. The SH is consistent across 1937, 1979 and 1997, with two records showing a downward step in 1890 and an upward step in 1969.

The tropics show a downward step in 1902/03, and upward steps in 1926, 1979 and 1997. Three NH mid-latitude records step upwards in 1920, 1921 or 1930, in 1978/89 and 1997/98. One zonal record also shows a downward step in 1964. The two NH high latitude records show a single downward step in 1902 and in 2005, both step upwards in 1921 and 1994 and a single step upwards in 2005. The three SH mid-latitude records show a downward step in 1887 and one in 1903, and upward steps in 1933 or 1937, 1968 or 1970, 1977/1978 or 1984, and 1997 or 1998. SH high latitude data is not very reliable, being absent for NCDC 60°S–90°S. The GISS 64°S–90°S average anomaly steps downward in 1912 and an upward in 1955. The combination of alternate positive and negative step changes in some records indicates the climate is responding to both internal and external drivers.



HadCRU			C&W			BEST			NCDC			GISS		
Period	Change	Trend Sig.	Period	Change	Trend Sig.	Period	Change	Trend Sig.	Period	Change	Trend Sig.	Period	Change	Trend Sig.
												1880–1901	-0.03	NS
1850–1929	-0.02	NS	1850–1919	-0.06	NS	1850–1920	-0.05	NS	1880–1929	-0.04	NS	1902	-0.13	
1930	0.26		1920	0.16		1921	0.15		1930	0.26		1902–1919	0.06	NS
												1920	0.18	
			1920–1936	0.12	NS	1920–1936	0.09	NS				1920–1936	0.15	**
			1937	0.16		1937	0.19					1937	0.18	
1930–1979	0.01	NS	1937–1979	-0.10	*	1937–1979	-0.06	NS	1930–1978	0.07	NS	1937–1979	0.01	NS
1980	0.20		1980	0.19		1980	0.20		1979	0.26		1980	0.26	
1980–1996	0.19	**	1980–1996	0.17	*	1980–1996	0.16	*	1979–1996	0.18	**	1980–1996	0.17	*
1997	0.32		1997	0.33		1997	0.31		1997	0.31		1997	0.31	
1997–2014	0.11	NS	1997–2013	0.17	*	1997–2013	0.14	*	1997–2014	0.10	NS	1997–2014	0.14	*
	0.29			0.31			0.27			0.31			0.49	
	0.78			0.84			0.85			0.83			0.80	

Figure 2. Mean global anomalies of surface temperature with internal trends. The annual anomalies (dotted lines) from five records (HadCRU, C&W, BEST, NCDC, GISS) are taken from a 1880–1899 baseline. Internal trends (dashed lines) are separated by step changes detected by the bivariate test at the $p<0.01$ level of statistical significance. The size of each step (in red) and change in temperature of each internal trend (in black) is shown in the figure table along with its significance, where NS is $p>0.05$, * is $p<0.05$ and >0.01, ** is $p<0.01$. The shift total is inclusive of internal trends.

Figure 2 illustrates the five records of global mean temperature analysed, showing internal trends and their significance. Steps and trends are consistent for the last two periods 1979/80 to 1996 and 1997 to 2013/14, but diverge somewhat in the middle of the record, seen as differences in the timing and magnitude of steps and accompanying internal trends. Data quality may be an issue in the earlier parts of the record. For example, the latest version of GISS data shows five shifts in 1902, 1920, 1937, 1980 and 1997, whereas a previous version to 2013 stabilised on shifts in 1930, 1979 and 1997, consistent with the average anomaly of all five records. This indicates that the timing and magnitude of shifts in the early 20th century can be influenced by adjustments made to improve data quality. However, all shift dates globally coincide with regional shifts, showing that while the relative importance of dates associated with step changes may be different, the dates themselves are quite stable. This gives us added confidence we are not detecting false positives.

Internal trends are mainly non-significant in the early record, the exception being the GISS 1920–37 period. The 1979/80 to 1996 trend is significant at the $p<0.01$ level in two records (HadCRU and NCDC) and $p<0.05$ in the other three records. The NH step change in 1987 strongly influences this trend, which is examined further in the next section. The post 1997 period is non-significant in two records and trends at $p<0.05$ in three records.

There is no objective way to partition shift and trend in these analyses. However, if the total change produced by internal trend is calculated, it ranges from 32% to 38% of the total change produced. The GISS record is an exception at 62% because of the downward shift in 1903 and subsequent recovery.

Satellite era records

Satellite records of lower troposphere temperatures sourced from the RSS (Mears and Wentz, 2009) and UAH records beginning in December 1978, were analysed for step changes (1979–2014). Annual mean global and zonal temperatures show 1995 and 1998 as the two main shift dates, with 1995 more prominent at the global scale (Table 1). Quarterly anomalies were investigated individually and as a continuous time series. For individual seasons, the shifts in 1995 are dominated by the NH JJA and SON periods, especially on land. This can be traced back to warm El Niño conditions in 1994/5. When the continuous quarterly time series is analysed (4 seasons x 36 years), the JJA and SON quarters of 1997 dominate the UAH global record, less so for the RSS record.

Table 1. Dates of step changes for lower tropospheric satellite temperature anomalies, with annual time series and quarterly breakdowns in parentheses (DJF, MAM, JJA, SON), and quarterly time series. Data sources are Remote Sensing Systems (RSS) and University of Alabama, Huntsville (UAH).

Region	Annual time series (quarterly breakdown)		Quarterly time series	
	RSS	UAH	RSS	UAH
Global land & ocean	1995 (98,98,95,95)	1995 (97,98,94,95)	JJA 1997	SON 1997
Global land	1995 (95,98,95,95)	1998 (98,98,94,95)	SON 1994	SON 1997
Global ocean	1998 (98, - ,97,95)	1995 (97, - , - ,95)	JJA 1997	SON 1997
NH land & ocean	1995 (98,98,94,94)	1998 (98,98,94,94)	JJA1997	SON 1997
NH land	N/A	1998 (98,98,98,98)	N/A	JJA 1997
NH ocean	N/A	1994 (- , - , - ,94)	N/A	JJA1997
SH land & ocean	1995 (98, - , - ,95)	1995 (97, - ,87,95)	SON 1997	SON 1997
SH land	N/A	1995 (95, - ,91,95)	N/A	MAM 2002
SH ocean	N/A	1995 (97, - , - ,95)	N/A	DJF 1998
Tropics land & ocean	1995 (- , - , - ,93)	- (- , - , - ,95)	JJA1997	JJA1997
Tropics land	1995 (- , - , - ,87)	1995 (98, - ,95,95)	SON 1997	JJA1997
Tropics ocean	1995 (- , - , - ,95)	-	JJA 1997	-
NH ex-trop land & ocean	1998 (95,98,98,94)	1998 (98,98,98,94)	SON 1997	DJF 1998
NH ex-trop land	1998 (- ,98,94,94)	1998 (- ,98,98,98)	MAM 1994	DJF 1998
NH ex-trop ocean	1998 (99,98,98,94)	1994 (02,98, - ,94)	SON 1997	MAM 1998
SH ex-trop land & ocean	1998 (96, - , - ,95)	1996 (97, - , - ,95)	DJF 1998	DJF 2001
SH ex-trop land	1995 (- , - , - , -)	2001 (03, - , - ,02)	JJA 1995	MAM 2002
SH ex-trop ocean	1998 (96, - , - , -)	1996 (97, - , - ,95)	DJF 1998	DJF 1998
N polar land & ocean	1995 (03,95,98,95)	1995 (05,95,98,95)	DJF 2000	MAM 1998
N polar land	1995 (- ,94,98,95)	1995 (- ,89,98, -)	DJF 2005	MAM 2000
N polar ocean	1995 (03,05,98,95)	1995 (05,95,98,95)	MAM 2002	MAM 1998
S polar land & ocean	-	-	-	-
S polar land	-	-	-	-

We investigated the quarterly anomalies for the RSS and UAH satellite and HadCRU and GISS surface mean global temperature to compare similarities and differences. Quarterly time series are affected by autocorrelation due to the El Niño-Southern Oscillation (ENSO), so for the bivariate test, the results are robust for timing but not significance. Student's t-test (two sided, unequal variance), was therefore also used to assess significance.

RSS shifts in DJF 1986/87 by 0.11°C (p<0.05 bv and p<0.01 t-t) and UAH shifts in DJF 1986/87 by 0.13°C (p<0.01, both tests). For surface temperature, HadCRU and GISS shift in JJA 1987 by 0.14°C and 0.15°C, respectively (p<0.01, both tests). On an annual basis the 1986/87 is significant to the p<0.05 level using the bivariate test. RSS shifts in JJA 1997 by 0.23°C, UAH shifts in DJF 1997/98 by 0.20°C, HadCRU in JJA 1997 by 0.26°C and GISS in SON 1997 by 0.25°C (all p<0.01, both tests). These four data sets show consistent shift dates in 1997 and similar shift dates in 1986/87, showing that the significant step change in the NH is present at the global scale. This suggests that the period of accelerated trend noted by many for 1976–1998 (Trenberth, 2015) is actually a period containing two step changes, one global (1978/79) and one regional (1986/87).

When four records are plotted on a common baseline of 1979–1998, the surface and satellite temperatures display similar shifts but different internal trends (Figure 3). Shown this way, the supposed substantial differences between surface and satellite trends are largely removed. The satellite data show significant negative internal trends over 1979–1986 (RSS p<0.01, UAH p<0.05), otherwise are non-significant. The surface data show significant positive internal trends over 1997–2014 (GISS p<0.01, HadCRU p<0.05), otherwise are non-significant. The decline post 1981 and lower trends in the early 1990s in the satellite data are likely due to volcanic eruptions, which amplify cooling at altitude(Free and Lanzante, 2009). The differences in internal trends post 1996 may be due to orbital decay that has not been fully allowed for in the satellite record, cooling from above affecting the satellite data and heating from below affecting the surface data, or a combination of these.

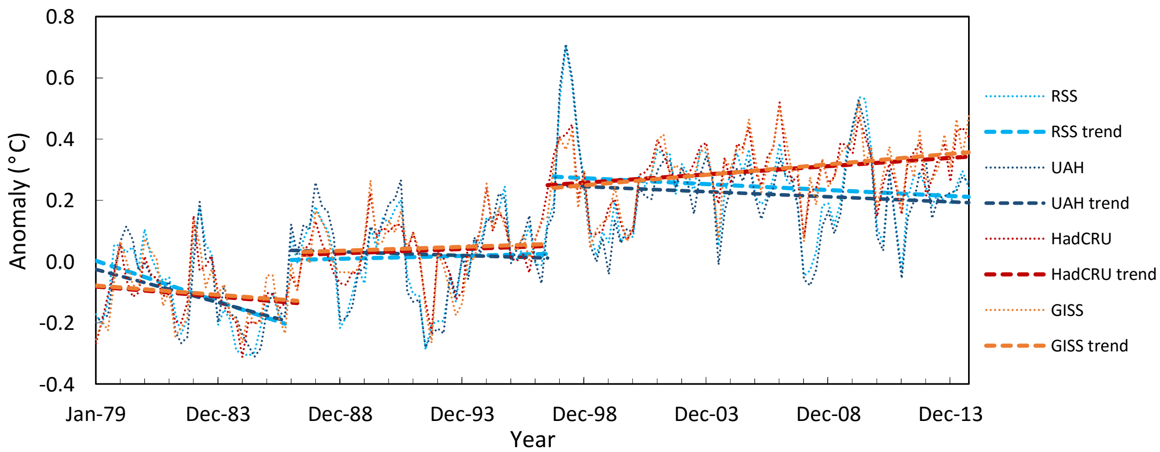


Figure 3. Quarterly mean satellite (RSS, UAH) and surface (HadCRU, GISS) temperature anomalies 1979–2014. Annual anomaly (solid lines), and internal trends from surface (dashed lines) and satellite (dotted lines) separated by step changes detected by the bivariate test are shown.

Regional attribution

Regional attribution of step changes has previously been carried for south-eastern Australia (SEA) annual mean temperature (Jones, 2012) and is repeated here for Texas and central England. The methodology is suitable for continental mid-latitude areas where annual average minimum temperature (*Tmin*) is correlated with maximum temperature (*Tmin*/*Tmax*), and *Tmax* is correlated with total annual rainfall (*Tmax*/*P*) (Power et al., 1998;Nicholls et al., 2004;Karoly and Braganza, 2005). For maritime areas such as central England, DTR is used (*DTR*/*P*) instead of *Tmax*/*P*. The method uses the bivariate method to test the dependent variable against the reference variable. A shift in the dependent variable denotes a regime change.
SEA climate was stationary until 1967 when a step change shifted minimum temperature (*Tmin*) by 0.6°C with respect to maximum temperature (*Tmax*) (Jones, 2012). Six independent climate model simulations for the

same region become non-stationary between 1964 and 2003, showing shifts of 0.4 to 0.7°C (Jones, 2012). Texas becomes non-stationary in 1990 with an increase in *Tmin*/*Tmax* of 0.5°C. *Tmax* increases by 0.8°C against *P* in 1998. For Central England, *Tmin* increases against *DTR* by 0.3 °C and *Tmax* against P by 0.9 °C in 1989. Tmax also increases against P in 1911 by 0.5°C (Table 2).

Table 2. Year of non-stationarity in regional temperature for south-eastern Australia, Texas and Central England. Data source, year of first change greater than one standard deviation for *Tmax* against *P* and *Tmin* against *Tmax*, or *DTR*/*P* using the bivariate test. The stationary period is also shown.

Data source	*Tmax/P*		*Tmin/Tmax*		*DTR/P*		Stationary Period
	Year	Change	Year	Change	Year	Change	(SEA)
SE Australia	1999	0.7	1968	0.6			1910–1967
Texas	1998	0.8	1990	0.5			1895–1990
Central UK	1989	0.9	N/S		1989	0.3	1878–1988
	1911	0.5					

The stationary period is used to established regression relationships that calculate *Tmax* and *Tmin* from *P* and *Tmax*, respectively. These regressions are used to estimate how *Tmax* and *Tmin* would have evolved during the non-stationary period. The residual is then attributed to anthropogenic regional warming and is tested using the bivariate test. Here the residuals for *Tmax* and *Tmin* are averaged to estimate the average anthropogenic regional warming (Tav_{ARW}).

In SEA, Tav_{ARW} shifts up by 0.5°C in 1973 (Figure 4). Similar patterns were found for 11 climate model simulations for SEA, undergoing a series of step changes to 2100 (Jones, 2012). For Texas, Tav_{ARW} shifts by 0.8°C in 1990. Central England temperature shifts up by 0.7°C in 1989 and by 0.5 °C in 1911. Using the full record for Central England average temperature from 1659, a significant step change was found in 1920, whereas using a starting date of 1878 identifies 1911. Given that the second mode identified in the longer test is 1911, we conclude the 1911 date is an artefact of the starting date in 1878 and a step change in 1920, consistent with NH data, would register if earlier data were available. None of the internal trends in Figure 4 are significant, suggesting that at the local level, step changes may dominate trends. This has also been found for high quality tide gauge records of sea level (Jones et al., 2013). Elsewhere, warming in both Central England and continental US has been found to be nonlinear (Franzke, 2012;Capparelli et al., 2013), supporting the analysis here.

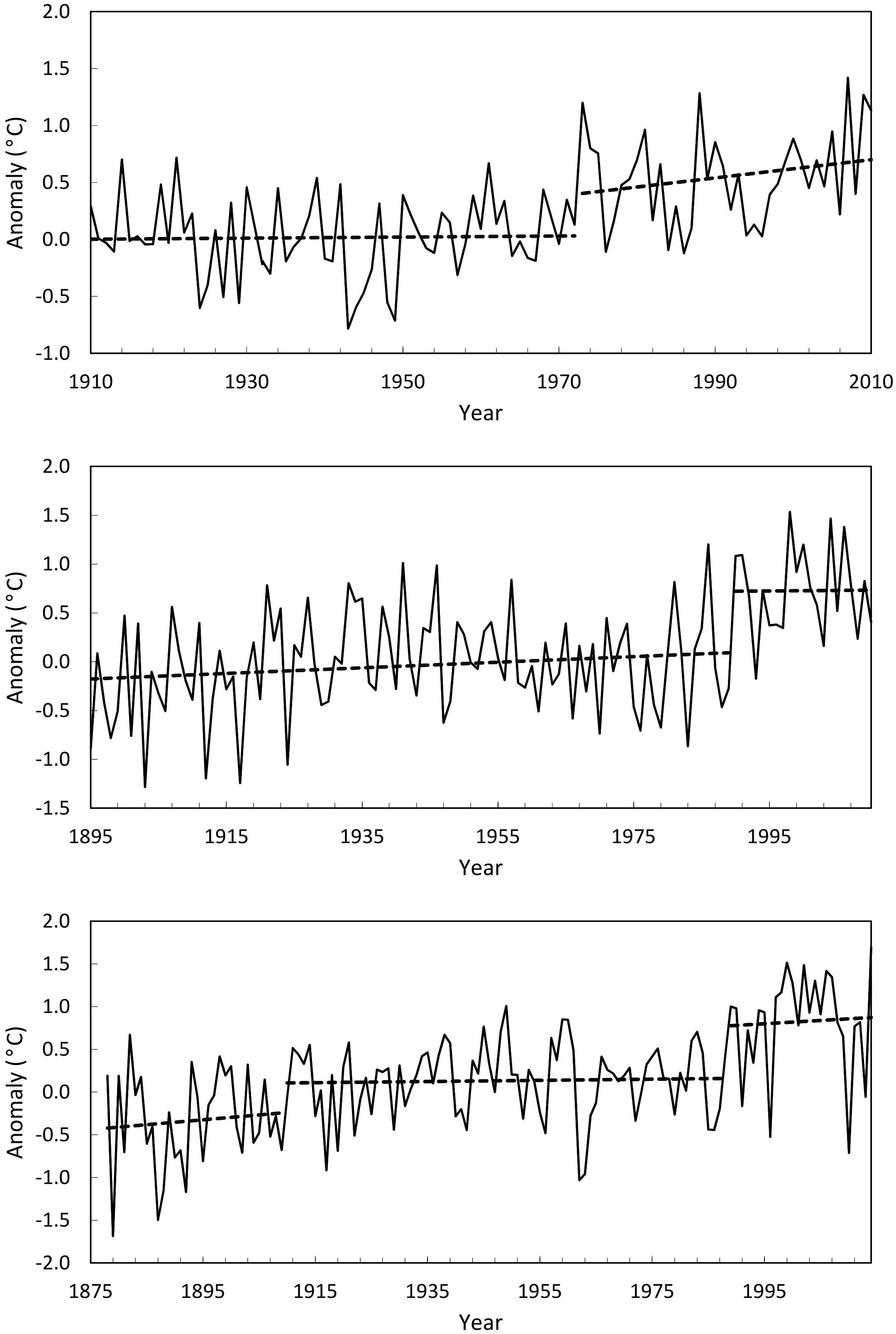


Figure 4. Anomalies of Tav_{ARW} for Central England (top), Texas (centre) and south-eastern Australia (bottom). Internal trends (dashed lines) are separated by statistically significant step changes to the $p<0.01$ level.

Discussion

A growing body of evidence describes regime changes that match key dates identified in this study (Trenberth, 1990;Trenberth and Hurrell, 1994;Hare and Mantua, 2000;McFarlane et al., 2000;Mantua, 2004). Menberg et al. (2014) identify shifts in 1977, 1987/88 and 1997 in global mean temperature and 90–24°N zonal temperature using the STARS method (Rodionov, 2006), noting the last two dates also occur in German records. Significant change-points (1986/87, 1997/98, 1968/69, and 2003/04) are detected in the time series of temperature and rainfall in the Zhujiang River Basin with 1986/87 being most significant there (Fischer et al., 2012) and on the Tibetan Plateau (Chen et al., 2014).

Dates associated with regime changes in decadal variability include the north Pacific in 1997 and 1989 (Hare and Mantua, 2000), and in 1996–7 (Gedalof and Smith, 2001;Chikamoto et al., 2012), the Mediterranean (Conversi et al., 2010) and the North Sea in 1987 (Beaugrand, 2004) and in the NH high latitudes in the early 1920s (Drinkwater, 2006). Many studies note the presence of rapid shifts and step-wise behaviour in temperature associated with such changes but these tend to be interpreted as the influence of decadal variability on a secular trend (Swanson et al., 2009;Wu et al., 2011). Alternatively, studies that force trends onto the data to dispel the recent so-called hiatus in global mean temperature produce change points unrelated to known climate phenomena. For example, Cahill et al. (2015) produce trend change points in 1912, 1940 and 1970, which they maintain relates back to forcing.

There is a great deal of theoretical evidence to support climate change over decadal timescales as a nonlinear process. The latest Intergovernmental Panel on Climate Change (IPCC) report does not select between linear and nonlinear interactivity between internally generated and externally forced near-term climate change. Both are treated as linearly additive with the following caveat: *This separation of T, and other climate variables, into components is useful when analyzing climate behaviour but does not, of course, mean that the climate system is linear or that externally forced and internally generated components do not interact* (Kirtman et al., 2013). Following Ozawa et al. (2003) we see the conversion from radiative forcing to heat as being a linear process, but view the distribution of that heat within the climate system as being nonlinear, consistent with H2. The analysis of attributable step changes in regional warming in three locations that coincide with larger-scale changes is consistent with nonlinear behaviour. We conclude that climate is more usefully seen as a turbulent hydrodynamic system exhibiting complex system behaviour over multiple temporal and spatial scales (Ozawa et al., 2003;Ghil, 2012), not a Newtonian mechanistic system following simple trends.

Conclusions

When global and regional mean temperature records to 2014 are analysed for step changes, steps dominate over trends, even if the internal trends are summed first. An accompanying study investigates simulated global mean surface warming by climate models, concluding that step changes dominate the warming process through to 2100 (Jones and Ricketts, 2015). This is counter to the prevailing narrative where rising temperatures are seen as an ongoing trend mediated by decadal variability (Santer et al., 2011;Kirtman et al., 2013;Meehl et al., 2013;England et al., 2014;Schmidt et al., 2014). This feeds into a gradualist narrative that, when applied to climate impacts, understates risk (Jones et al., 2013;Jones, 2015).

Every statistical model carries a set of assumptions about how a system or process works, whether explicit or implicit (von Storch, 1999;von Storch and Zwiers, 2001). The implicit assumption in trend analysis is that the atmosphere is warming gradually in a trend-like manner with the rate of warming being modulated by climate variability. For example, empirically-based studies of climate model output suggest that a period of at least 17 years, should be a sufficiently reliable measure for gauging trends in lower tropospheric temperature (Santer et al., 2011), whereas here we have shown that the historical record is dominated by steps. Based on the evidence here, trend analysis will not provide a reliable guide for decision making over decadal timescales. Warming that is not gradual but stepwise, has significant implications for how climate risk is managed.

Supplemental information

Statistical method

This formulation from Bücher and Dessens (1991) is based on normalised data with no trend. It takes advantage of the normalisation step to allow simpler structure. According to the authors, it was obtained from Potter, although the method published in Potter (1981) is essentially that of Maronna and Yohai (1978).

In the following, primes reference un-normalised data and functions, normalised data in step 2 is denoted by removal of primes. This usage has been slightly modified from Bücher and Dessens (1991). Additionally, the second part of (4) corrects an inconsistency in that paper.

Let $x_i',\ i = 1...n$ be a stationary reference time series and $y_i',\ i = 1...n$ be a test time-series which is assumed to correlate to x' except for a single shift at some time i_0.

Step 1. Standardize series.

$$\overline{X}' = \frac{\sum_{j=1}^{n} x_j'}{n},\ \overline{Y}' = \frac{\sum_{j=1}^{n} y_j'}{n},\ S_x' = \left(\frac{\sum_{j=1}^{n}\left(x_j' - \overline{X'}\right)^2}{n}\right)^{1/2},\ S_y' = \left(\frac{\sum_{j=1}^{n}\left(y_j' - \overline{Y'}\right)^2}{n}\right)^{1/2} \quad (1)$$

$$x_j = \frac{\left(x_j' - \overline{X'}\right)}{S_x'},\ y_j = \frac{\left(y_j' - \overline{Y'}\right)}{S_y'} \text{ for all } j \leq n. \quad (2)$$

Step 2. Compute test statistics.

$$S_{xy} = \sum_{j=1}^{n} x_j y_j \quad (3)$$

$$X_i = \frac{\sum_{j=1}^{i} x_j}{i},\ Y_i = \frac{\sum_{j=1}^{i} y_j}{i} \text{ for all } i < n \quad (4)$$

$$F_i = n - \frac{{X_i}^2 ni}{(n-i)} \text{ for all } i < n \quad (5)$$

$$D_i = \frac{\left(S_{xy}X_i - nY_i\right)n}{(n-i)F_i} \text{ for all } i < n \quad (6)$$

$$T_i = \frac{\left[i(n-i)D_i^2 F_i\right]}{\left(n^2 - S_{xy}^2\right)} \text{ for all } i < n. \quad (7)$$

$$T_{i0} = \max(T_i),\ i_0 = i \text{ when } T_{i0} = \max(T_i)\ . \quad (8)$$

The time associated with T_{i0} represents the time period during which a change occurred and its successor is the first time of the new regime. A mean shift can be computed. For the null trend case, critical values of T_i are provided by Maronna and Yohai (1978) for two-tailed, alpha levels of (0.1, 0.05, and 0.01) for the null hypothesis of no change, given time series lengths of 15, 20, 75 and Potter (1981) provides these for 100. An interpolating function is used to generalize these results for time series of varying length.

Multi-step bivariate test: application

The purpose of constructing a rule-based process for analysing multiple step changes in a time series is to remove the need to make individual decisions that utilise the experimenter's judgement. It also allows multiple sampling and the addition of randomness, which increases the robustness of the results.
The bivariate test, rule-based process and diagnostics are coded in Python 2.7.6, developed with the Spyder environment (© 2009-2012 Pierre Raybaut), running in 32-bit Windows 7 and 64-bit Windows 8.1

environments. Moba-Xterm PE v7.2, a windows based Unix emulator, was used to support some collation of results, and data acquisition.

Description

The method is a technique for segmenting time series with zero to many step changes in the mean. The test returns a list of break-points that divide a time series into segments bounded by statistically significant step changes, except for the start and the end. The routine consists of a *screening pass*, which produces a first approximation break-list. This break-list is iteratively refined by a *convergent pass*. Both passes are described below. Each application of the test is subject to a *resampling test*, which determines the resilience of a step-point determination to noise. One hundred iterations sample a test series against a randomly selected reference time series.

The method is probabilistic. Each iteration returns a list comprising a set of shift points, their timing and magnitude and null probability against a serially independent reference, along with a variety of diagnostic variables. A time series with distinct step changes will return the same list for a set of iterations, whereas others may yield several variations. This is especially the case for areal averages that integrate local changes from two or more regions, data with quality issues, or where autocorrelation due to trending behaviour or other processes is present.

Resampling test. The bivariate test is repeated 100 times using different random sequences and the i_0 values and associated T_{i0}, and shifts are collated by mode.

1. On the screening pass only the modal value is examined.
2. On the convergent pass additional selection rules apply. The mode and second mode are examined. There may be a single mode (e.g., 100% selection), or the two modes may be close together or well separated.

The i_0 (time i preceding the shift) returned by the resampling test is the modal value (i.e., most frequent) of each test. Similarly, T_{i0}, and shift magnitude are the mean of those values associated with i_0. A segment contains a breakpoint in position i if T_{i0} exceeds the critical T_i value for segment length with a given probability.

Screening pass. This is a binary segmentation technique, similar to that used in similar applications(Scott and Knott, 1974;Killick et al., 2012). The entire time series is analysed for a single breakpoint using the resampling test (100 iterations). If T_{i0} is significant ($p<0.01$), then the segment up to and including i_0 is analysed for an earlier break, and the segment after i_0 is analysed for a later break. This process is repeated for the sub-segments so formed until no significant breaks are found. The result is a series of breakpoints which are then refined on the convergent pass. Because breakpoints found on this pass are returned on the basis of a recursive process, end point effects caused by sampling time series of different lengths may influence the results.

The role of the convergent pass is to combine segments to determine whether the screening pass has oversampled for steps and also to ensure that the selected break points are robust within the selected segmentation.

Convergent pass. The list of n breakpoints from the screening pass breaks the original time series into $s=n+1$ segments. The algorithm then works its way from earliest to latest segments combining consecutive segments into one, and then searching within that segment using binary segmentation to produce a *candidate list* from which two most frequent i_0 values are retained (in practice there are rarely more than two and usually just one). There are two special cases, segments 1 and s which are analysed individually at either end of this process to cover the impact of end point adjustments. This procedure will sometimes reduce the number of step changes from the screening pass.
The convergent pass is reiterated until it produces the same list for a second time and this is returned as the final result.

This pass incorporates **some decision rules**.:

1. A prohibition period of seven years is applied at the start and the end of the time series, and after a break point before another point will be accepted. This is because the bivariate is sensitive to end effects.
2. If the modal year is within the prohibition period from the previous break point, then the two are compared. A resample test is conducted by extending the segment backwards to the start date of the previous segment. If it is a valid break, this then replaces the previous break, otherwise the previous break is retained, and a small "safety margin" is added for one iteration to the low bound of the first segment next time around to prevent the point being re-selected.
3. If the mode is equal to or >90%, or the first mode is >50% and the second mode is > 20% then the modal year is accepted, else it is dropped.
4. If after this, a segment contains a single breakpoint, it is retained.
5. If the candidate list is empty, that is, a segment no longer contains a breakpoint then the two segments are merged and treated as a single segment on the next iteration.
6. If the candidate list contains more than one point, then the earliest two are retained and the rest discarded. The two points are then trialled using a resampling test to determine if the interval up to the later of the two still contains a break, and if this is still present, it is retained. If not, then the second candidate is similarly tested.

Considerations about the multi-step bivariate test

The role of time

This analysis treats time unidirectional. The bivariate test itself selects the last time *before* a change so is asymmetric. The convergent pass operates from earliest to latest, revising provisional breakpoints and then using them to delineate later breakpoints. On each pass, as soon as a breakpoint is provisionally established, it is used, and no other information is preserved. Therefore, every breakpoint is complete unto itself, and the segment within which it is embedded has its own statistics. *This means that the final set of segments, broken up by the final set of breakpoints,* consists of a series of segments, each of which has its own variance, mean and shape parameters, and embedded trend. The analysis treats each segment as independent, but whether physical dependence (including memory) or otherwise, can be assumed, remains to be assessed. Information theory approaches to assessing statistical best fit may also break down because the system may not deliver the same information between break-points. This is certainly the case for other complex systems covering economics and ecology.

End point effects

The determination of a breakpoint in a time series is sensitive to all of the data including the first and last elements, but is less reliable near the start and end of that series (Vivès and Jones, 2005). This affects two aspects of the analysis:

1. Previously determined shift points become end points when a subsequent segment is tested, making nearby observations more sensitive to end effects. This is the principal reason for the 7-year restriction on break-points. However, temperature data can also produce peaks/troughs due to interannual variability giving a multi-modal distribution of potential break points clustered around an underlying shift. Altering segment lengths can potentially alter the distribution of these modes, therefore leading to the application of decision rules 1, 2 and 6 above, to determine the most robust outcome.
2. The choice of starting year.
    1. The quality of long-term climate data characteristically degrades backwards in time. This may produce artificial break-points, move existing ones, or simplify 'natural' variability. Autocorrelation may also be introduced by some infilling methods. Early shift points should not be regarded as being as reliable as more recent dates.
    2. If the data record starts just before a true shift point or is influenced by a truncated sequence of low or high years then the next (and to a lesser extent, consequent) dates may be affected.

On balance it is much better to start an analysis from the earliest date available, unless data quality is clearly compromised.

Assumptions

The bivariate test itself assumes two variates that are both stationary, except that the test variate may have a single point change in the mean. A time series that has multiple shift points will register only one value of T_{i0} and other points cannot be relied upon. Sometimes the original value of T_{i0} will be removed once the time series is segmented, because it is an 'average' of several others.

The assumption of serial independence is very important. A number of studies have concluded that observed annual temperature and rainfall records fulfil that stricture. Such climate data may also contain components of autocorrelation and variable trends. Some of the variability of trend may simply be redness (drift due to persistence of previous values), some may represent transient processes. However, autocorrelated data such as regional or global mean sea level, sea surface temperature or climate data divided into a monthly or quarterly time series may lead to statistical significance being over-estimated, although the timing of a shift will remain accurate. In the paper, we use the t-test to address this, but when autocorrelation is due to a sustained trend in a time series that contains both steps and trends, the t-test also will give misleading results. This is addressed further in the Supplementary Information for the companion model analysis paper (Jones and Ricketts, 2015).

Here we treat annual temperature data as a signal composed of a small but arbitrary number of linear segments delineated by step changes, and embedded in Gaussian noise. The impact of trend is on the assigned significance of the shift returned from the bivariate test, rather than its timing. Additionally, a change of trend when there is no step may cause the bivariate test to allocate a step some years after the change of trend. These can all be determined in post-processing.

A time series that contains nothing but a general trend and variation, will have two properties when analysed by our method:

1. The sum of steps will converge on zero.
2. The probabilistic test will be dominated by random sampling of the reference variate and the number of different break-point lists will increase – that is, it will return unstable sets.

Diagnostics

Every iteration of the 100 break-list runs that comprise an analysis produces a csv file of results, plus a trace of the decision process. The trace file contains the initial data as well as a summary of the break dates with some QA diagnostics. All 100 trace files are collated and the diagnostics are given for each analysis. This includes Ti, Shift, Modal Year, Modal Frequency, The Second Modal Year and its frequency, Internal Trends of the segment containing the step, the pre-step and post-step segments, Delta of pre and post segments at the shift date. For comparison there is an analysis of a 30-year segment comprising 15 years either side of a shift, the trend of the full segment, and pre and post trends, modal and second modal years and frequencies.

Terminology

The language of non-linear change is nowhere near as established as is the language for trend analysis. Here we use the following terms in the ways described:

- Break, break-point, break-year: a break denotes an abrupt change in statistical characteristics of any kind (e.g., change in trend, variance).
- Shift: in the paper a shift is the distance between the end of one internal trend and the beginning of the next across a step change.
- Step: an abrupt step-like change as measured by the test.

Calibration of the method.

The method has been calibrated against synthetic data composed with variable lag one/seven autocorrelation, variable number of shift points, varying trends and changes of trend. Its performance has also been tested for its ability to locate a randomly timed shift point in a random series to which is added varying shifts, varying trends up to those well in excess of any climate model run, and simulating a random shift month within the simulated shift year.

Data sources

Global mean surface temperature

Time series tested are mean annual global air temperature anomalies from five groups (GISS, HadCRU, NCDC, C&W and BEST), hemispheric temperatures from three groups (HadCRU, NCDC and GISS) and zonal temperatures from two groups (NCDC and GISS). Tropospheric satellite temperatures from two groups (RSS and UAH) are also tested (Table S1).

Supplementary Table 1: Source groups for 20th century observations, surface and satellite.

Name	Version	Download date	Base Period	Global	Hemi-spheric	Zonal	Land-Ocean	References
BEST		15 Jan 2015	1951–1980	Y	N	N	N	(Rohde et al., 2012)
COWTAN & WAY	2.0	15 Jan 2015	1961–1990	Y	N	N	N	(Cowtan and Way, 2014)
GISSTEMP3	V3	15 Apr 2015	1951–1980	Y	Y	N	N	(Hansen et al., 1988;GISSTemp Team, 2015)
HadCRUT4 HadSST3 CRUT4	4.3.0.0 3.1.1.0 4v	25 May 2015	1961–1990	Y	Y	Y	Y	(Jones et al., 1999;Jones et al., 2001;Brohan et al., 2006;Rayner et al., 2006;Kennedy et al., 2011;Jones et al., 2012;Morice et al., 2012;Osborn and Jones, 2014)
NCDC	v3.5.4. 201504	18 Mar 2015	"20th C Average"	Y	Y	Y	Y	(Smith et al., 2008;Vose et al., 2012)
Satellite based atmospheric temperature estimates, Lower Troposphere to Lower Stratosphere								
RSS	V03.3	7 May 2015	1979-1998	Y	Y	Y	Y	(Mears et al., 2003;Mears and Wentz, 2009)
UAH	6.0.beta	5 May 2015	1981-2010	Y	Y	Y	Y	(Christy et al., 2000)

NDCD zonal data version v3.5.4.201504.

Annual and monthly files in ASCII format covering land, ocean, and combined land and ocean were downloaded on 29 May 2015 from ftp://ftp.ncdc.noaa.gov/pub/data/mlost/operational/products/ using wget in recursive mode. Each file contains data for one zonal average and for one of land, ocean and combined land and ocean.

The zonal averages were over: 90°S–90°N (Global), 90°S–0°S (°Southern hemisphere), 0°N–90°N (Northern hemisphere), 90°S–20°S, 60°S–30°S, 60°S–60°N, 30°S–0°N, 0°N–30°N, 20°S–20°N, 20°N–90°N, and 60°N–90°N. Data in the files labelled as 90°S–60°S for all three subsets was clearly corrupted on receipt and was not used.

The data format is documented on-line in the file
ftp://ftp.ncdc.noaa.gov/pub/data/mlost/operational/products/readme.timeseries,
Annual averages are as provided, rather than simple averages of monthly values.

GISSTEMP_3

Data was downloaded on 15 April 2015 in ASCII from
http://data.giss.nasa.gov/gistemp/tabledata_v3/ZonAnn.Ts+dSST.txt and format converted to CSV for use.
All values are multiplied by 0.01 to produce degrees C, as per the metadata in the file.

Cowtan and Way

Data representing annually averaged was downloaded in ASCII format on 15 Jan 2015, from http://www-users.york.ac.uk/~kdc3/papers/coverage2013/had4_krig_annual_v2_0_0.txt
Both annual and monthly data were downloaded but this initial analysis was of the annual data only.
Data is described at http://www-users.york.ac.uk/~kdc3/papers/coverage2013/series.html.

Berkeley

Data representing annual averaged mean global temperature was downloaded in ASCII format on 15 Jan 2015 from http://berkeleyearth.lbl.gov/auto/Global/Land_and_Ocean_summary.txt
Two versions are present in the file. The data used in this study is from column 1, 'Annual Anomaly' computed by extrapolation of temperature in the presence of sea ice by using land-air temperature surface anomalies.

NCDC Land, Ocean, and combined Land and Ocean data

Seasonal analysis was based on data downloaded on 18 Mar 2015, as individual csv files, one per month, using the wget utility from http://www.ncdc.noaa.gov/cag/time-series/global/$extent/$set/1/1/*.csv where $extent is replaced by one of ["global","nhem", "shem"] and $set is one of ["land","ocean","land_ocean"]. Seasonal averages were computed as simple averages of the monthly values.

Annual averaged data was downloaded interactively from http://www.ncdc.noaa.gov/cag/ on 26 May 2015 (the same site) using 12 Month time scales to December for global and hemispheric extents giving a total of nine files.

Hadley/CRU Land, Ocean, Land and Ocean data

Data reported here was downloaded on 25 May 2015 as ASCII text files from http://www.metoffice.gov.uk/hadobs/ File formats are described algorithmically at http://www.metoffice.gov.uk/hadobs/hadcrut4/data/current/series_format.html
Monthly and seasonal analyses were performed using the appropriate monthly values, corresponding annual averages were drawn from the last column.

Satellite derived lower tropospheric temperature data, RSS and UAH

RSS

The front page for this organisation is at http://www.remss.com/. Information on upper air temperatures is at http://www.remss.com/measurements/upper-air-temperature.
One complex data set is provided, Temperature of Lower Troposphere (TLT), "constructed by calculating a weighted difference between measurements made at different Earth incidence angles to extrapolate MSU channel 2 and AMSU channel 5 measurements lower in the atmosphere"
Data for Land, Ocean, and Land and Ocean were downloaded in a simpler ASCII format, all bands on one line per month per year, on 7 May 2015 from ftp://ftp.remss.com/msu/data/uah_compatible_format
Data files are from Jan 1979 to present.
Anomalies are computed by subtracting the mean monthly value determined by averaging 1979 through 1998 data for each channel from the average brightness temperature for each month. The set of 12 month means for 1979 to 1998 are included in the netCDF files available on the ftp server (ftp.remss.com/msu)

UAH

These data are version 6.0.
UAH6:0 Data were downloaded on 5 May 2015 from http://vortex.nsstc.uah.edu/data/msu/v6.0beta/.
http://vortex.nsstc.uah.edu/data/msu/v6.0beta/tlt/uahncdc_lt_6.0beta1
A readme file is at http://vortex.nsstc.uah.edu/data/msu/docs/readme.msu.

References

Beaugrand, G., 2004: The North Sea regime shift: evidence, causes, mechanisms and consequences. *Progress In Oceanography*, **60**, 245-262.

Brohan, P., J.J. Kennedy, I. Harris, S.F.B. Tett and P.D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *Journal of Geophysical Research*, **111**, D12106.

Bücher, A. and J. Dessens, 1991: Secular trend of surface temperature at an elevated observatory in the Pyrenees. *Journal of Climate*, **4**, 859-868.

Cahill, N., S. Rahmstorf and A.C. Parnell, 2015: Change points of global temperature. *Environmental Research Letters*, **10**, 084002.

Capparelli, V., C. Franzke, A. Vecchio, M.P. Freeman, N.W. Watkins and V. Carbone, 2013: A spatiotemporal analysis of US station temperature trends over the last century. *Journal of Geophysical Research: Atmospheres*, **118**, 7427-7434.

Chen, Y., H. Deng, B. Li, Z. Li and C. Xu, 2014: Abrupt change of temperature and precipitation extremes in the arid region of Northwest China. *Quaternary International*, **336**, 35-43.

Chikamoto, Y., M. Kimoto, M. Ishii, M. Watanabe, T. Nozawa, T. Mochizuki, H. Tatebe, T.T. Sakamoto, Y. Komuro and H. Shiogama, 2012: Predictability of a stepwise shift in Pacific climate during the late 1990s in hindcast experiments using MIROC. *Journal of the Meteorological Society of Japan*, **90**, 1-21.

Christy, J.R., R.W. Spencer and W.D. Braswell, 2000: MSU tropospheric temperatures: Dataset construction and radiosonde comparisons. *Journal of Atmospheric and Oceanic Technology*, **17**, 1153-1170.

Conversi, A., S. Fonda Umani, T. Peluso, J.C. Molinero, A. Santojanni and M. Edwards, 2010: The Mediterranean Sea regime shift at the end of the 1980s, and intriguing parallelisms with other European basins. *PLoS ONE*, **5**, e10633.

Corti, S., F. Molteni and T.N. Palmer, 1999: Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature*, **398**, 799-802.

Cowtan, K. and R.G. Way, 2014: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society*, **140**, 1935-1944.

Drinkwater, K.F., 2006: The regime shift of the 1920s and 1930s in the North Atlantic. *Progress In Oceanography*, **68**, 134-151.

England, M.H., S. McGregor, P. Spence, G.A. Meehl, A. Timmermann, W. Cai, A.S. Gupta, M.J. McPhaden, A. Purich and A. Santoso, 2014: Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nature Climate Change*, **4**, 222-227.

Fischer, T., M. Gemmer, L. Liu and B. Su, 2012: Change-points in climate extremes in the Zhujiang River Basin, South China, 1961–2007. *Climatic Change*, **110**, 783-799.

Franzke, C., 2012: Nonlinear trends, long-range dependence, and climate noise properties of surface temperature. *Journal of Climate*, **25**, 4172-4183.

Free, M. and J. Lanzante, 2009: Effect of volcanic eruptions on the vertical temperature profile in radiosonde data and climate models. *Journal of Climate*, **22**, 2925-2939.

Gedalof, Z.e. and D.J. Smith, 2001: Interdecadal climate variability and regime-scale shifts in Pacific North America. *Geophysical Research Letters*, **28**, 1515-1518.

Ghil, M., 2012: Climate variability: nonlinear and random effects. *Encyclopedia of Atmospheric Sciences. Elsevier*, 1-6.

GISSTemp Team, 2015: *GISS Surface Temperature Analysis (GISTEMP)*. NASA Goddard Institute for Space Studies, http://data.giss.nasa.gov/gistemp/, (Last accessed 15-April-2015).

Hansen, J., I. Fung, A. Lacis, D. Rind, S. Lebedeff, R. Ruedy, G. Russell and P. Stone, 1988: Global climate changes as forecast by Goddard Institute for Space Studies three-dimensional model. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **93**, 9341-9364.

Hare, S.R. and N.J. Mantua, 2000: Empirical evidence for North Pacific regime shifts in 1977 and 1989. *Progress In Oceanography*, **47**, 103-145.

Hasselmann, K., 2002: Is Climate Predictable? In: *The Science of Disasters: Climate Disruptions, Heart Attacks, and Market Crashes* [Bunde, A., J. Kropp and H.J. Schellnhuber (eds.)] Springer, Berlin Heidelberg, 141-188.

Hegerl, G. and F. Zwiers, 2011: Use of models in detection and attribution of climate change. *Wiley Interdisciplinary Reviews: Climate Change*, **2**, 570-591.

IPCC, 2001: *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 881 pp.

IPCC, 2007: *Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, 996 pp.

IPCC, 2013: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Jones, P.D., M. New, D.E. Parker, S. Martin and I.G. Rigor, 1999: Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics*, **37**, 173-199.

Jones, P.D., T.J. Osborn, K.R. Briffa, C.K. Folland, E.B. Horton, L.V. Alexander, D.E. Parker and N.A. Rayner, 2001: Adjusting for sampling density in grid box land and ocean surface temperature time series. *Journal of Geophysical Research: Atmospheres*, **106**, 3371-3380.

Jones, P.D., D.H. Lister, T.J. Osborn, C. Harpham, M. Salmon and C.P. Morice, 2012: Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research: Atmospheres*, **117**, D05127.

Jones, R.N., 2012: Detecting and attributing nonlinear anthropogenic regional warming in southeastern Australia. *Journal of Geophysical Research*, **117**, D04105.

Jones, R.N., C.K. Young, J. Handmer, A. Keating, G.D. Mekala and P. Sheehan, 2013: *Valuing Adaptation under Rapid Change*. National Climate Change Adaptation Research Facility, Gold Coast, Australia, 182 pp.

Jones, R.N., 2015: *Reconciling anthropogenic climate change and variability on decadal timescales: hypotheses and scientific narratives*. Climate Change Working Paper No. 32, Victoria Institute of Strategic Economic Studies, Victoria University, Melbourne, 16 pp.

Jones, R.N. and J.H. Ricketts, 2015: *Analysing steps in modelled global surface air temperature*. Climate Change Working Paper No. 35, Victoria Institute of Strategic Economic Studies, Victoria University, Melbourne, 16 pp.

Karoly, D.J. and K. Braganza, 2005: A new approach to detection of anthropogenic temperature changes in the Australian region. *Meteorology and Atmospheric Physics*, **89**, 57-67.

Kennedy, J.J., N.A. Rayner, R.O. Smith, D.E. Parker and M. Saunby, 2011: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *Journal of Geophysical Research: Atmospheres*, **116**, D14103.

Killick, R., P. Fearnhead and I. Eckley, 2012: Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, **107**, 1590-1598.

Kirtman, B., S. Power, A.J. Adedoyin, G. Boer, R. Bojariu, I. Camilloni, F. Doblas-Reyes, A. Fiore, M. Kimoto, G. Meehl, M. Prather, A. Sarr, C. Schär, R. Sutton, G.J.v. Oldenborgh, G. Vecchi and H.-J. Wang, 2013: Near-term Climate Change: Projections and Predictability. In: *Climate Change 2013: The Physical Science Basis. Working Group I contribution to the IPCC 5th Assessment Report* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)] Cambridge University Press, Cambridge and New York, 121.

Mantua, N., 2004: Methods for detecting regime shifts in large marine ecosystems: a review with approaches applied to North Pacific data. *Progress In Oceanography*, **60**, 165-182.

Maronna, R. and V.J. Yohai, 1978: A bivariate test for the detection of a systematic change in mean. *Journal of the American Statistical Association*, **73**, 640-645.

McFarlane, G.A., J.R. King and R.J. Beamish, 2000: Have there been recent changes in climate? Ask the fish. *Progress In Oceanography*, **47**, 147-169.

Mears, C.A., M.C. Schabel and F.J. Wentz, 2003: A Reanalysis of the MSU Channel 2 Tropospheric Temperature Record. *Journal of Climate*, **16**, 3650-3664.

Mears, C.A. and F.J. Wentz, 2009: Construction of the RSS V3. 2 lower-tropospheric temperature dataset from the MSU and AMSU microwave sounders. *Journal of Atmospheric and Oceanic Technology*, **26**, 1493-1509.

Meehl, G.A., A. Hu, J.M. Arblaster, J. Fasullo and K.E. Trenberth, 2013: Externally Forced and Internally Generated Decadal Climate Variability Associated with the Interdecadal Pacific Oscillation. *Journal of Climate*, **26**, 7298-7310.

Menberg, K., P. Blum, B.L. Kurylyk and P. Bayer, 2014: Observed groundwater temperature response to recent climate change. *Hydrology and Earth System Sciences*, **18**, 4453-4466.

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 data set. *Journal of Geophysical Research: Atmospheres*, **117**, D08101.

Nicholls, N., P. Dellamarta and D. Collins, 2004: 20th century changes in temperature and rainfall in New South Wales. *Australian Meteorological Magazine*, **53**, 263-268.

North, G.R., K.-Y. Kim, S.S.P. Shen and J.W. Hardin, 1995: Detection of Forced Climate Signals. Part 1: Filter Theory. *Journal of Climate*, **8**, 401-408.

Osborn, T.J. and P.D. Jones, 2014: The CRUTEM4 land-surface air temperature data set: construction, previous versions and dissemination via Google Earth. *Earth Systems Science Data*, **6**, 61-68.

Overland, J., S. Rodionov, S. Minobe and N. Bond, 2008: North Pacific regime shifts: Definitions, issues and recent transitions. *Progress In Oceanography*, **77**, 92-102.

Ozawa, H., A. Ohmura, R.D. Lorenz and T. Pujol, 2003: The second law of thermodynamics and the global climate system: A review of the maximum entropy production principle. *Reviews of Geophysics*, **41**, 1018.

Potter, K., 1981: Illustration of a new test for detecting a shift in mean in precipitation series. *Monthly Weather Review*, **109**, 2040-2045.

Power, S., F. Tseitkin, S. Torok, B. Lavery and B. McAvaney, 1998: Australian temperature, Australian rainfall, and the Southern Oscillation, 1910-1996: coherent variability and recent changes. *Australian Meteorological Magazine*, **47**, 85-101.

Rahmstorf, S., G. Foster and A. Cazenave, 2012: Comparing climate projections to observations up to 2011. *Environmental Research Letters*, **7**, 044035.

Rayner, N.A., P. Brohan, D.E. Parker, C.K. Folland, J.J. Kennedy, M. Vanicek, T.J. Ansell and S.F.B. Tett, 2006: Improved Analyses of Changes and Uncertainties in Sea Surface Temperature Measured In Situ since the Mid-Nineteenth Century: The HadSST2 Dataset. *Journal of Climate*, **19**, 446-469.

Reeves, J., J. Chen, X.L. Wang, R. Lund and Q.Q. Lu, 2007: A Review and Comparison of Changepoint Detection Techniques for Climate Data. *Journal of Applied Meteorology and Climatology*, **46**, 900-915.

Rodionov, S.N., 2005: A brief overview of the regime shift detection methods. *Large-Scale Disturbances (Regime Shifts) and Recovery in Aquatic Ecosystems: Challenges for Management Toward Sustainability. UNESCO-ROSTE/BAS Workshop on Regime Shifts, Varna, Bulgaria*, Velikova, V. and N. Chipev, Eds., City, pp 17-24.

Rodionov, S.N., 2006: Use of prewhitening in climate regime shift detection. *Geophysical Research Letters*, **33**, L12707.

Rohde, R., R.A. Muller, R. Jacobsen, E. Muller, S. Perlmutter, A. Rosenfeld, J. Wurtele, D. Groom and C. Wickham, 2012: A new estimate of the average earth surface land temperature spanning 1753 to 2011. *Geoinformatics & Geostatistics: An Overview*, **1**, 1000101.

Santer, B.D., C. Mears, C. Doutriaux, P. Caldwell, P.J. Gleckler, T.M.L. Wigley, S. Solomon, N.P. Gillett, D. Ivanova, T.R. Karl, J.R. Lanzante, G.A. Meehl, P.A. Stott, K.E. Taylor, P.W. Thorne, M.F. Wehner and F.J. Wentz, 2011: Separating signal and noise in atmospheric temperature changes: The importance of timescale. *Journal of Geophysical Research*, **116**, D22105.

Schmidt, G.A., D.T. Shindell and K. Tsigaridis, 2014: Reconciling warming trends. *Nature Geoscience*, **7**, 158-160.

Scott, A.J. and M. Knott, 1974: A Cluster Analysis Method for Grouping Means in the Analysis of Variance. *Biometrics*, **30**, 507-512.

Smith, T.M., R.W. Reynolds, T.C. Peterson and J. Lawrimore, 2008: Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006). *Journal of Climate*, **21**, 2283-2296.

Solomon, A., L. Goddard, A. Kumar, J. Carton, C. Deser, I. Fukumori, A.M. Greene, G. Hegerl, B. Kirtman, Y. Kushnir, M. Newman, D. Smith, D. Vimont, T. Delworth, G.A. Meehl and T. Stockdale, 2011: Distinguishing the Roles of Natural and Anthropogenically Forced Decadal Climate Variability. *Bulletin of the American Meteorological Society*, **92**, 141-156.

Swanson, K.L., G. Sugihara and A.A. Tsonis, 2009: Long-term natural variability and 20th century climate change. *Proceedings of the National Academy of Sciences*, **106**, 16120-16123.

Trenberth, K.E., 1990: Recent observed interdecadal climate changes in the Northern Hemisphere. *Bulletin of the American Meteorological Society*, **71**, 988-993.

Trenberth, K.E. and J.W. Hurrell, 1994: Decadal atmosphere-ocean variations in the Pacific. *Climate Dynamics*, **9**, 303-319.

Trenberth, K.E., 2015: Has there been a hiatus? *Science*, **349**, 691-692.

Vivès, B. and R.N. Jones, 2005: *Detection of Abrupt Changes in Australian Decadal Rainfall (1890-1989)*. CSIRO Atmospheric Research Technical Paper, CSIRO Atmospheric Research, Melbourne, 54 pp.

von Storch, H., 1999: Misuses of Statistical Analysis in Climate Research. In: *Analysis of Climate Variability* [von Storch, H. and A. Navarra (eds.)] Springer Berlin Heidelberg, 11-26.

von Storch, H. and F.W. Zwiers, 2001: *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge UK, 484 pp.

Vose, R.S., D. Arndt, V.F. Banzon, D.R. Easterling, B. Gleason, B. Huang, E. Kearns, J.H. Lawrimore, M.J. Menne, T.C. Peterson, R.W. Reynolds, T.M. Smith, C.N. Williams and D.B. Wuertz, 2012: NOAA's Merged Land–Ocean Surface Temperature Analysis. *Bulletin of the American Meteorological Society*, **93**, 1677-1685.

Wu, Z., N. Huang, J. Wallace, B. Smoliak and X. Chen, 2011: On the time-varying trend in global-mean surface temperature. *Climate Dynamics*, **37**, 759-773.

Zhou, J. and K.-K. Tung, 2013: Deducing Multidecadal Anthropogenic Global Warming Trends Using Multiple Regression Analysis. *Journal of the Atmospheric Sciences*, **70**, 3-8.