Supplemental Material

Supplemental Material for

A global, continental and regional analysis of changes in extreme precipitation

Qiaohong Sun[1], Xuebin Zhang[2], Francis Zwiers[1, 3], Seth Westra[4], Lisa V. Alexander[5, 6]

[1] Pacific Climate Impacts Consortium, University of Victoria, Victoria, British Columbia V8W 2Y2, Canada.

[2] Climate Research Division, Environment and Climate Change Canada, Toronto, Ontario M3H 5T4, Canada

[3] Nanjing University of Information Science and Technology, Nanjing, China

[4] School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, South Australia, Australia

[5] Climate Change Research Centre, University of New South Wales, Sydney, NSW 2052, Australia

[6] Australian Research Council Centre of Excellence for Climate Extremes, Australia

Corresponding author: Qiaohong Sun (sunqh@uvic.ca)

1 The influences of different profiles of missing years

Our estimated trends and assessments of whether they are significantly different from zero might have been affected by missing annual index values. This can happen in several ways – for example, because annual values are missing randomly during the period of interest, or because values are missing more frequently during some parts of the period than others (e.g., at the ends of the period). To investigate these possibilities, we have performed a series of experiments using the 855 stations with no missing index values during the 1950–2018 period.

We investigated the first possibility by randomly sampling the years within the periods as missing years for each station and comparing the MK test results for these 855 stations when data are 50%, 60%, 70%, 90% and 100% complete. We call this experiment E1. In E1, for each station, we randomly selected N% (N=50, 30, 30, 20, 0) of years that are to be made missing at this station, we calculated and summarized statistics, and repeat this entire process 1000 times. Results show that the percentage of stations with significant increasing trends rises from about 6.0% to 10.0% as the data becomes more complete (Figure S1a, b), which is likely to be because of the enhanced statistical power of trend test for longer records. On the other hand, the sensitivity that is inferred from the incomplete records is not significantly different from that obtained from the fully complete records, but is more uncertain when more years are missing, as would be expected (Figure S1e, f).

We investigated the second possibility by performing three additional experiments in which we imposed observed profiles of missing years on the stations with complete records. We labelled these three experiments E2a, E2b, and E2c.

Experiment E2a consisted of the following steps:

1) For each of the 855 complete stations we first selected a station at random from the incomplete stations with data for 70% of years or more for 1950–2018, thereby producing 855 pairs of stations.
2) For each pair, we set the complete station values to missing when the corresponding paired station has a missing value.
3) Trends and sensitivities were then calculated for these "masked" versions of the complete stations.

This process was repeated 1000 times. Results, which are shown in Figure S1 as box and whisker plots labelled E2a. Results show that the percentage of stations with significant trends are consistent with results obtained that would be obtained when about 15% of annual values are missing at random, indicating that inferences based on typical missing data profiles are consistent with inferences that are made when no annual values are missing. Note that the average missing data rate amongst all stations with 70% or more complete records between 1950 and 2018 is approximately 15%. The median sensitivity difference relative to the fully complete records is small for both Rx1day (~0.3%) and Rx5day (~0.6%).

In experiment E2b, we examined what would happen if all stations had the same missing data profile (that is if all stations are missing at the same times), again using observed missing data occurrence sequences. This experiment was performed as follows:

1) We selected a single station at random from the incomplete stations with data for 70% of years or more for 1950–2018.
2) We used this station to set a missing year mask that was then applied to all 855 complete stations.

3) Trends and sensitivities were then calculated for the resulting “masked” versions of the complete stations.

This process was also repeated 1000 times. Results, which are shown in Figure S1 as box and whisker plots labelled E2b are similar to those obtained from experiment E2a, except that uncertainties in the fractions of stations with significant trends and in the difference from the sensitivities calculated from the full records is somewhat larger than in the case of experiment E2a.

In experiment E2c, we examine if the E2b results are sensitive to the particular 1000 stations that were chosen to form missing data masks. We performed this experiment as E2b but used the collection of incomplete stations with data for 70% of years or more for 1950–2018 and have at least 1 year of record during 1950–1959 and 2010–2018. This narrows the box plots of the summary statistics relative to those obtained in E2b, indicating that the absence of data in the first and last decades makes trend and sensitivity estimates more uncertain. Median differences relative to complete records are nevertheless, not strongly affected.

It should be noted that none of these experiments are fully representative of how missing data would actually occur. Experiment E1, with randomly missing years, is unrealistic because data are more likely to be missing near the beginning and end of the 1950-2018 period. Experiment E2a uses observed missing data occurrences, but supposes that the occurrence of a missing year at one location is independent of whether the same year is missing at other locations, which also might not be the case. Experiments E2b and E2c also use observed missing data occurrences, but in these cases, it is supposed that all stations within a given domain have missing in identically the same years, which would not be the case except in circumstances where a jurisdiction started or stopped making data available for all of its stations abruptly at a

specific date. While it is difficult to conceive a comprehensive experimental strategy to test the sensitivity of results to missing data covering all of the plausible ways data might be missing, it seems likely that the effect of missing data relative to full records lies somewhere between the results from experiments E2a and E2b.

The completeness criterion obviously involves a tradeoff between spatial and temporal coverage. Substantially higher representativeness of available observing locations can be achieved by requiring a minimum of 30-years within the 1950-2018 period (12323 stations) than by requiring 100% complete records (855 stations, 93% fewer than for the 30-year criterion). Based on the 12323 stations that meet a minimum 30-year requirement, the percentage of stations with significant trends changes only modestly as the completeness requirement is made more stringent (Figures S2a–S2d), and sensitivity estimates (Figure S2e, S2f) remain relatively unaffected, except confidence intervals (5–95%) are larger when the number of stations are smaller. There is, however, a concern about spatial representativeness. To avoid excess sensitivity in this regard, and to allow the evaluation of trends and sensitivities in IPCC regions, we judged 70% temporal completeness to be a reasonable tradeoff between spatial and temporal coverage, which allows 7293 stations for 1950–2018 and 1974 stations for 1900–2018.

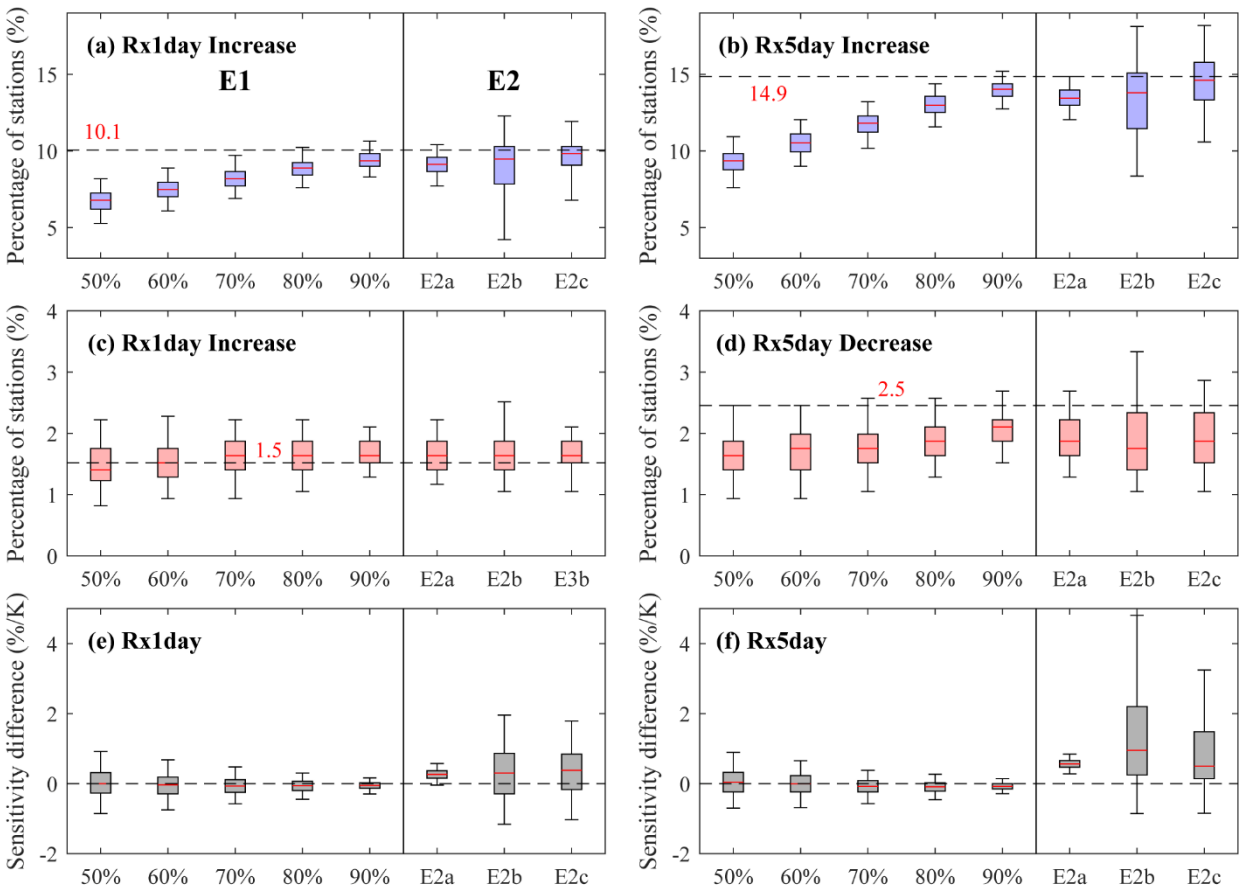

Figure S1. Percentage of stations with statistically significant increasing (a, b) and decreasing (c, d) trends, and median of the sensitivity differences relative to fully complete records (e, f) experiment E1 and experiment E2. (a, c, e) are for Rx1day and (b, d, f) for Rx5day. The black dashed lines in (a, b, c, d) and red numbers indicate the results from the 855 stations with full record. The box-and-whisker plots represent the distribution of percentage of stations with significant trends or median of the sensitivity differences relative to fully complete records from 1000 realizations by sampling the missing years. The upper and lower edges of the boxes show the 75th and the 25th percentile while red lines mark the median. The upper and lower whiskers show the 97.5th and the 2.5th percentiles.

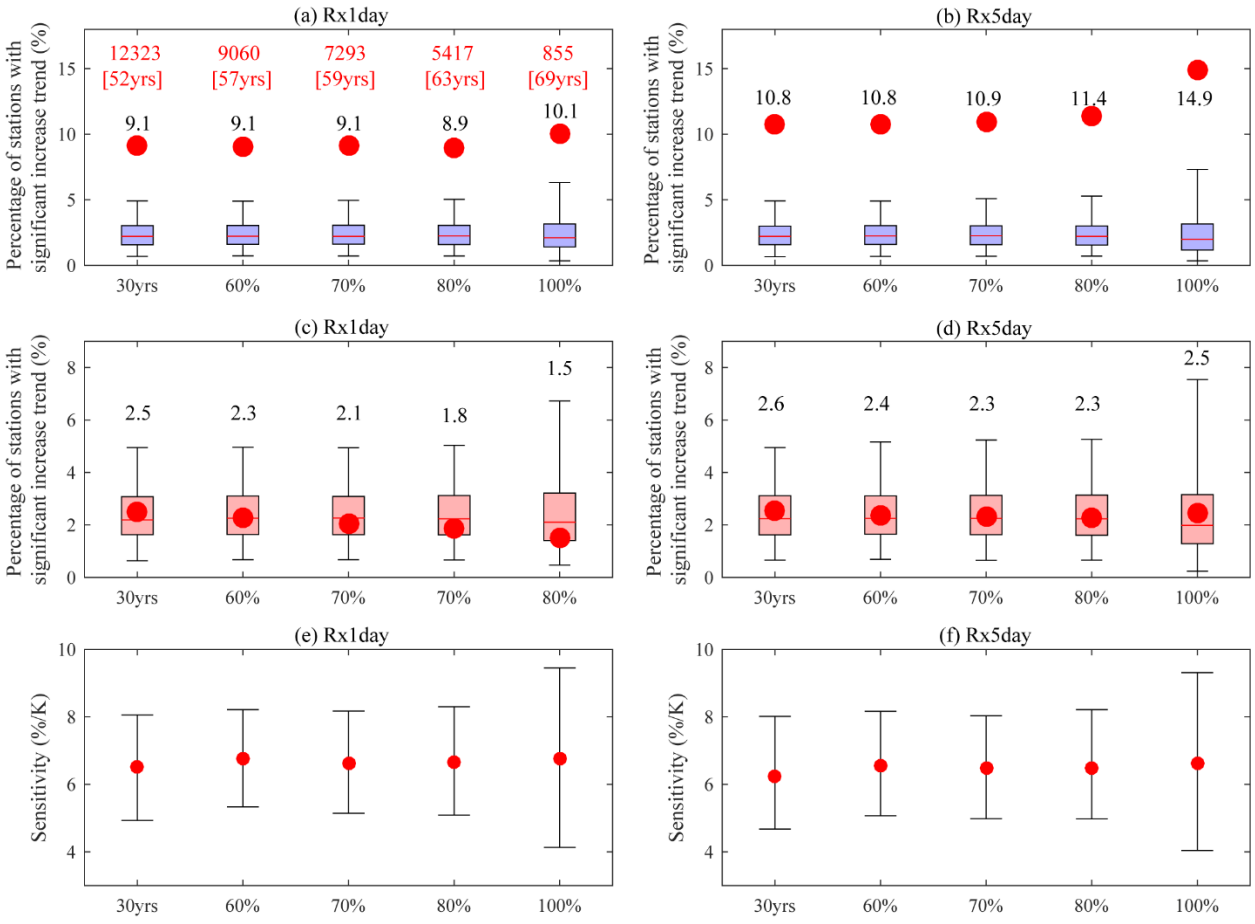

Figure S2. Percentage of stations with statistically significant increasing (a, b) and decreasing (c, d) trends, and global median sensitivity values with global mean surface temperature change (e, f) for different completeness criteria of records within the 1950–2018 period. (a, c, e) are for Rx1day and (b, d, f) for Rx5day. The red numbers in (a) indicate the number of stations satisfying the completeness criteria of records and the median record length. The box-and-whisker plots in (a, b, c, d) summarize the breadth of the distribution from 1000 bootstrap realizations under the no trend null hypothesis; The upper and lower edges of the boxes show the 75th and 25th percentiles while red lines mark the median; The upper and lower whiskers show the 97.5th and the 2.5th percentiles; Red dots indicate the observed results. The plots in (e, f)

represent confidence interval (5–95%) for the sensitivity value; red dots indicate the median value from 1000 bootstrap realizations.

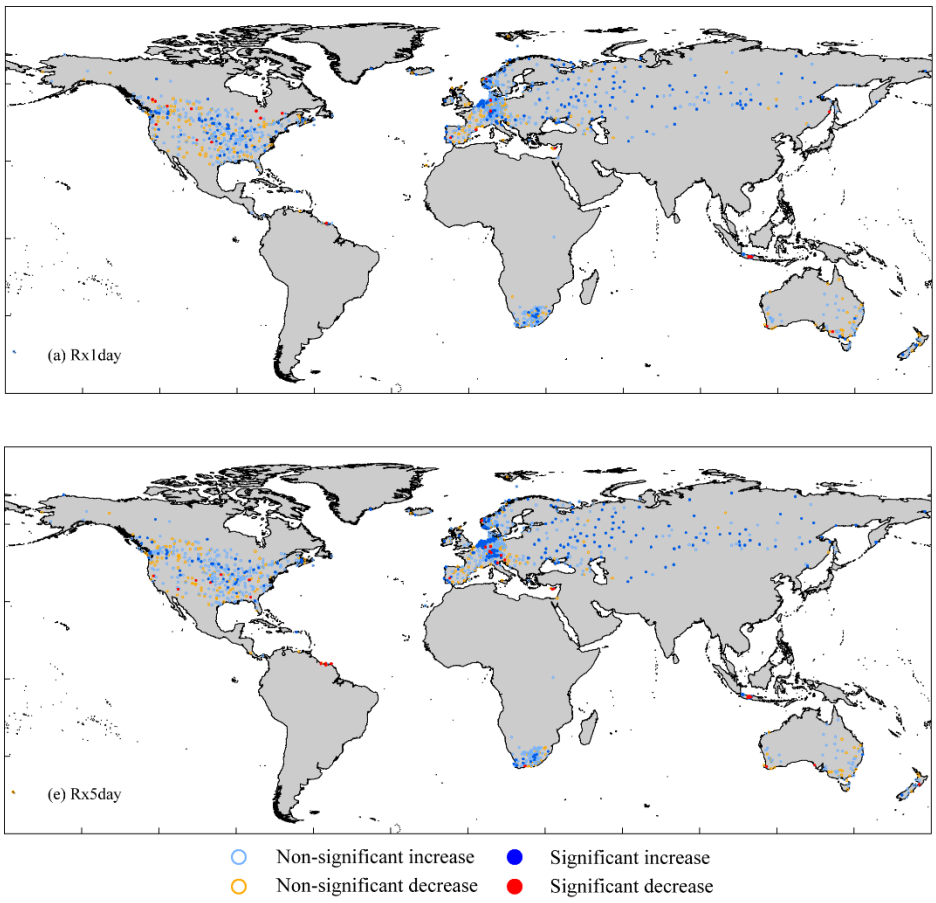

Figure S3. Maps of locations of stations and Mann-Kendal trend analyses for 1974 stations during the period 1950–2018. (a) Rx1day, (b) Rx5day. Light blue open dots indicate increasing trends and light red open dots mark decreasing trends. Dark blue and red filled dots indicate trends that are significant at the 5% level based on a two-sided test.

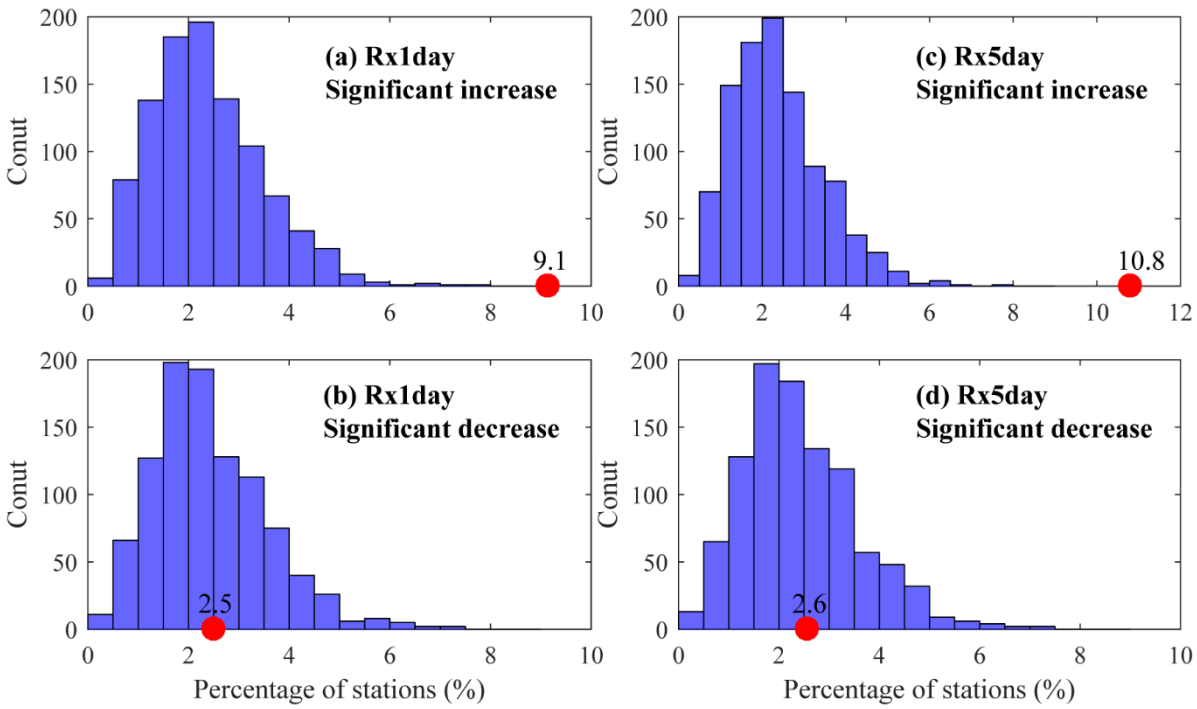

Figure S4. Summary of Mann-Kendal trend analyses for the period 1950-2018 for 12323 stations with no less 30 years' records. (a), (b), (c) and (d), percentage of stations with statistically significant increasing and decreasing trends. The histogram represents the distribution of percentage of stations with significant trends from 1000 bootstrap realizations when the connection between time and the occurrence of annual extremes is broken. The red dots represent the values from the observational data.

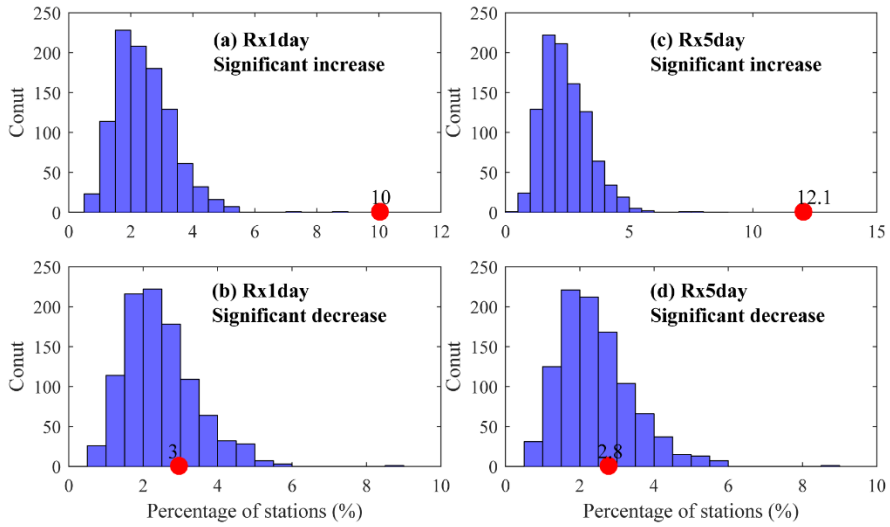

Figure. S5. Summary of Mann-Kendal trend analysis for the period 1900-2018 for 14796 stations with no less 30 years' records. (a), (b), (c) and (d), percentage of stations with statistically significant increasing and decreasing trends. The histogram represents the distribution of percentage of stations with significant trends from 1000 bootstrap realizations when the connection between time and the occurrence of annual extremes is broken. The red dots represent the values from the observational data.

TABLE S1. Percentage of stations with increasing, decreasing, statistically significant increasing, and statistically significant decreasing trends in Rx1day and Rx5day based on the Mann–Kendall test over the IPCC AR6 reference regions (Iturbide et al. 2020) during 1950–2018. The stations that have at least 70% of the annual Rx1day/Rx5day data within the 1900–2018 period. There are 14 regions selected, requiring regions to have more than 30 stations meeting our completeness criterion during the 1900–2018 period. Values higher than the upper 97.5th percentile from the 1000 bootstrap realizations under the no-trend null hypothesis are shown in bold. There are not regions in which the percentage values lie below the lower 2.5th percentile of these bootstrap distributions.

Region name	Rx1day				Rx5day			
	Increase (%)	Decrease (%)	Significant increase (%)	Significant decrease (%)	Increase (%)	Decrease (%)	Significant increase (%)	Significant decrease (%)
N.W. North-America [NWN]	37.5	62.5	5.0	5.0	50.0	50.0	2.5	0.0
W. North-America [WNA]	61.5	38.5	3.8	3.8	56.4	43.6	7.7	3.8
C. North-America [CAN]	**72.2**	27.8	**12.7**	1.6	**69.0**	31.0	**12.7**	0.0
E. North-America [ENA]	70.8	29.2	7.5	0.8	**75.0**	25.0	**11.7**	0.8
N. South-America [NSA]	45.2	54.8	1.6	4.8	58.1	41.9	1.6	6.5
N. Europe [NEU]	**80.3**	19.7	**17.8**	0.0	**87.6**	12.4	**23.9**	0.3
West & Central-Europe [WCE]	65.3	34.7	**8.4**	0.3	70.9	29.1	**12.7**	0.6
E. Europe [EEU]	**75.9**	24.1	**14.8**	1.9	**81.5**	18.5	**22.2**	1.9
Mediterranean [MED]	58.5	41.5	2.4	6.1	48.8	51.2	1.2	6.1
W. Southern-Africa [WSAF]	64.1	35.9	6.3	1.6	64.1	35.9	7.8	0.0
E. Southern-Africa [ESAF]	65.8	34.2	6.3	1.3	54.4	45.6	0.0	0.0
E. Siberia [ESB]	62.2	37.8	6.7	0.0	48.9	51.1	0.0	0.0
S.E. Asia [SEA]	57.4	42.6	**9.7**	4.3	74.2	25.8	10.6	1.8
S. Australia [SAU]	44.9	55.1	3.8	2.6	29.5	70.5	0.0	7.7