HadISD: Creating a next-generation sub-daily data-product for studying extreme events

R.J.H.Dunn(1), K.M.Willett(1), P.W.Thorne(2), D.E.Parker(1)

(1) Met Office Hadley Centre, FitzRoy Road, Exeter, Devon, EX1 3PB, UK
(2) CICS-NC, NOAA's National Climatic Data, Patton Avenue, Asheville, NC, 28801, USA

To meet the requirements of climate science in the 21st Century we need high-resolution, high-quality, comprehensive global datasets which are updated in near-real time. This demands a meticulous approach to data quality-control, conducted in an objective, reproducible and consistent manner that enables the quantification of uncertainty. We present the quality control strategy and validation over a subset of recent extreme events for HadISD, a sub-daily near-surface database based on NOAA's Integrated Surface Database including temperature, dewpoint temperature, sea-level pressure, wind speed and cloud cover.

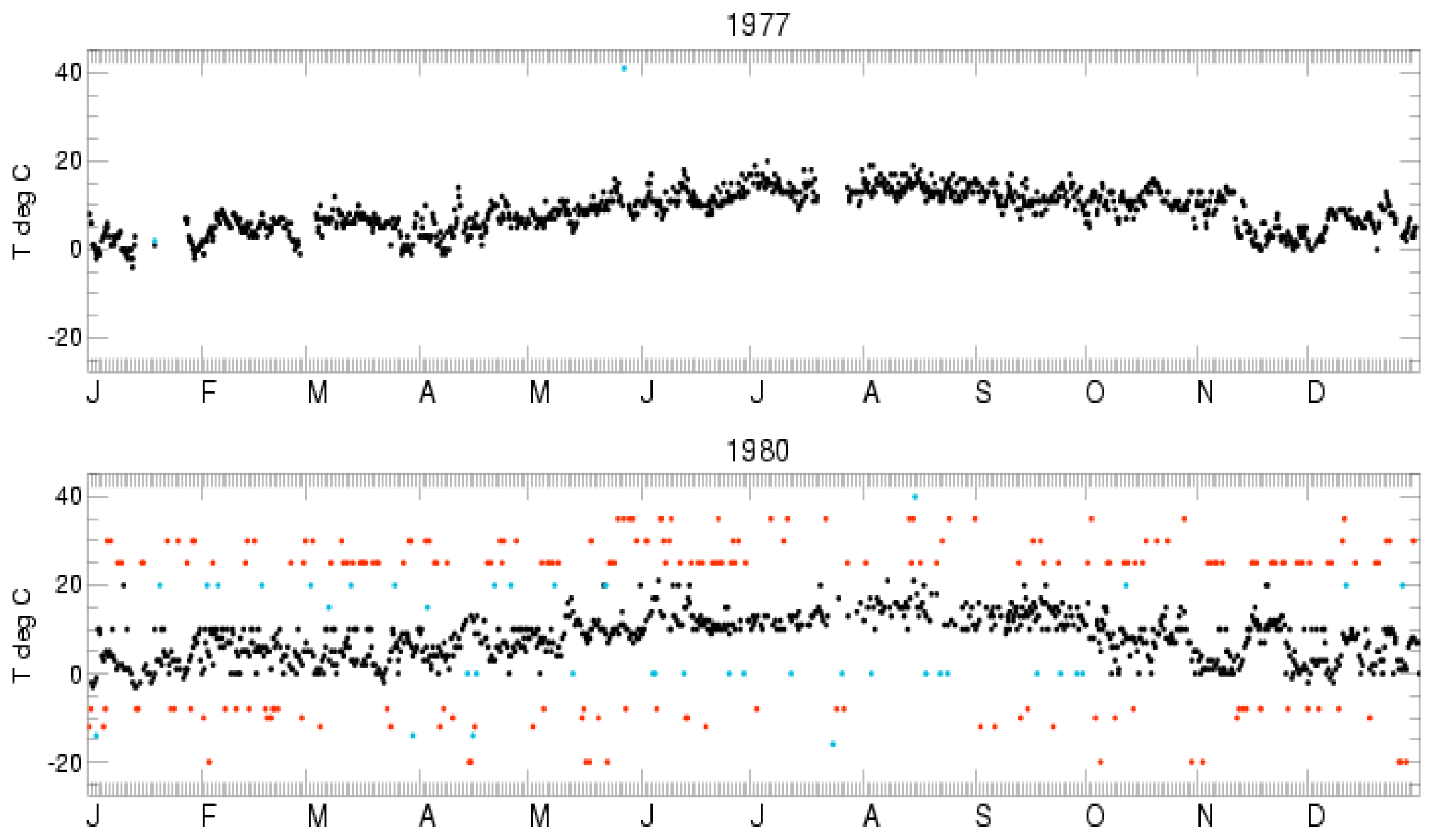

Fig 1. Frequent Value Check. 037930-99999, Anvil Green, Kent, UK. Temperature for 1977 and 1980. Red points have been removed by this test, blue points by other tests. The repeating values at 10°C and 0°C are in the main distribution and so are not detected by this test.

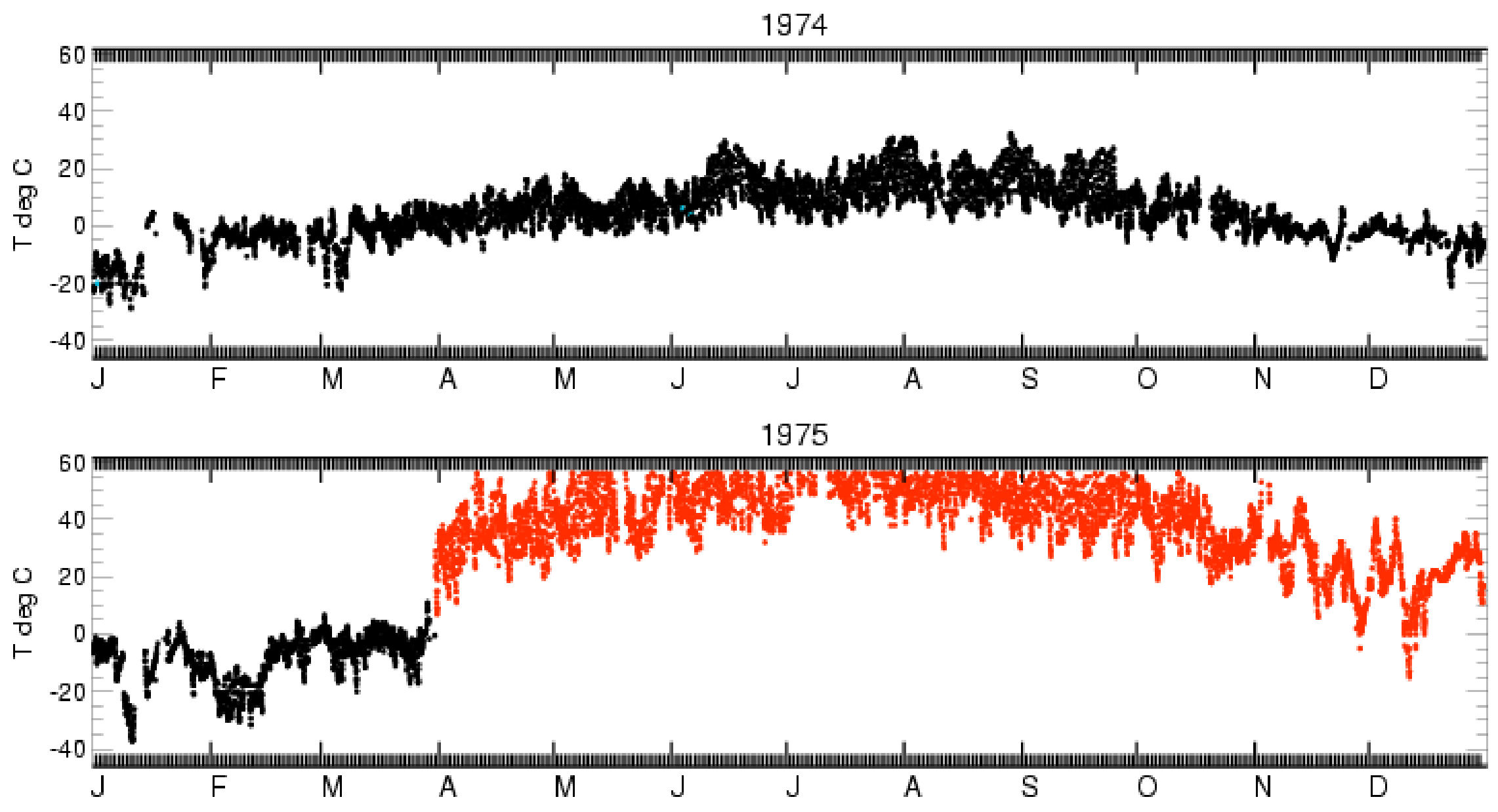

Fig 2. Distributional Gap Check. 714740-99999, Clinton, BC, Canada. Temperature for 1974 and 1975. Red points have been removed by this test, blue points by other tests.

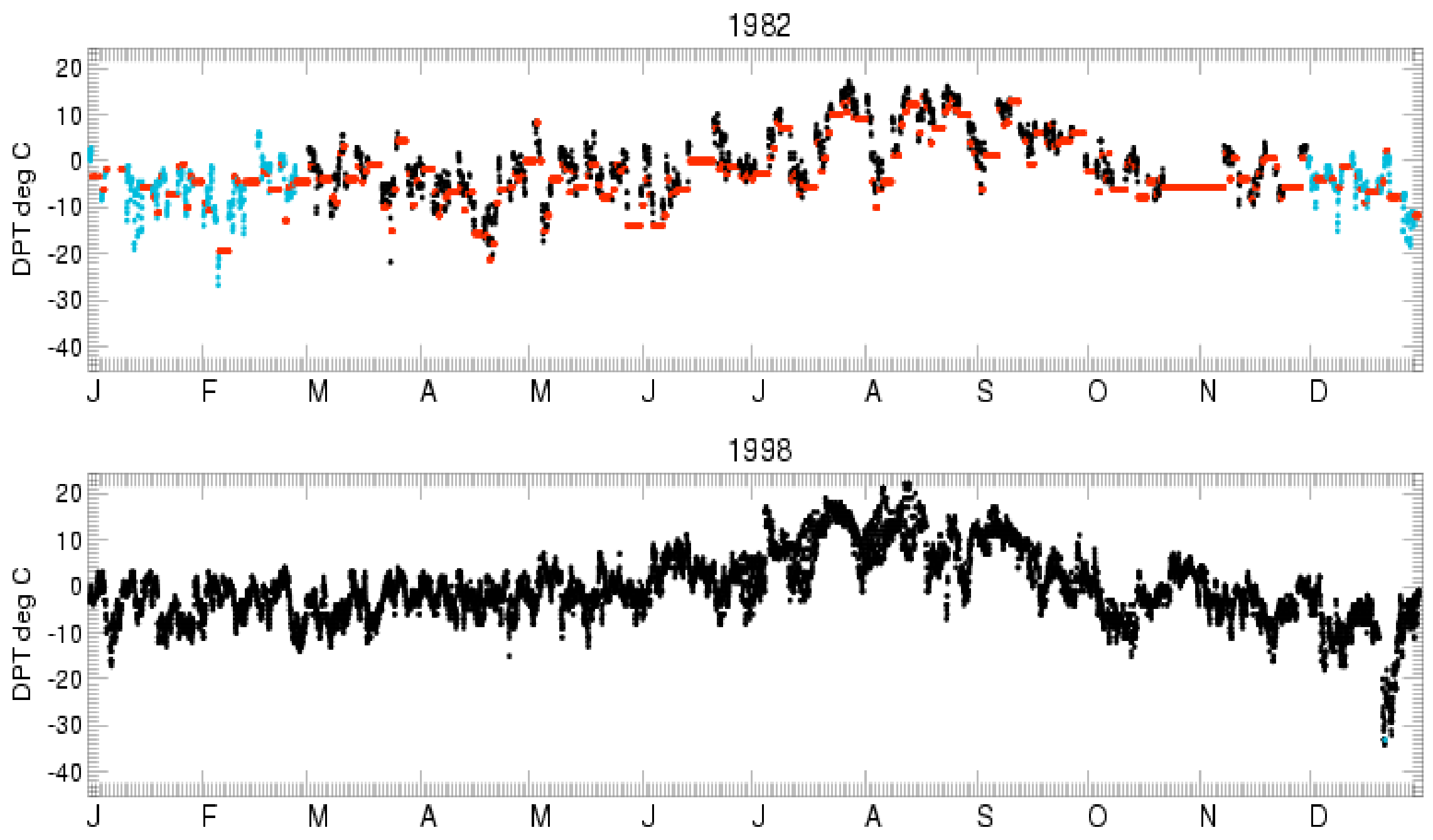

Fig 3. Streak Check. 724797-23176 Mitford, UT, USA. Dewpoint temperature for 1982 and 1998. Red points have been removed by this test, blue points by other tests.

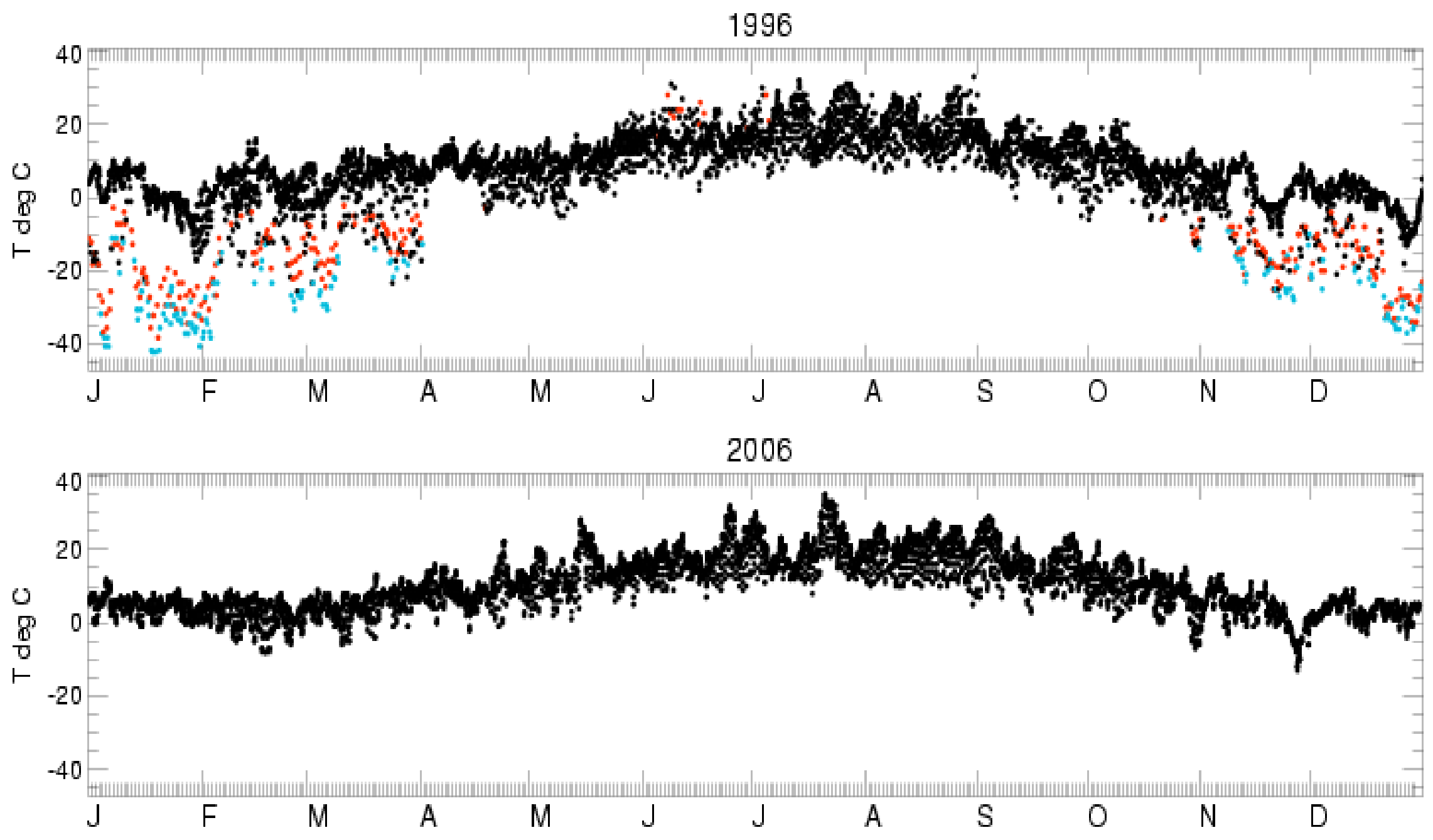

Fig 4. Spike Check. 718936-99999, Campbell River, BC, Canada. Temperature for 1996 and 2006. Red points have been removed by this test, blue points by other tests.

An initial version of the QC suite was constructed, based on the methods outlined in Durre *et al.* (2010, Journal of Applied Meteorology and Climatology. 49, 1615-1633). Station extreme values for the UK during documented extreme events were used to validate this version of the QC suite. A number of true extremes were removed or not clearly identified and erroneous values were being retained. Consequently the QC tests were overhauled, creating the current version which retains these extreme values. A sample of the QC tests applied are highlighted below. Full details will be published in Dunn *et al.* (in prep).

Example Quality Control Tests

Frequent Value Check: Unusually common values are identified using a histogram-based method on the entire series. Each year is then assessed and unusually frequent values removed (Fig. 1). This removes the most obvious and pervasive frequent values.

Distributional Gap Check: On a per-calendar month basis, whole months are removed if their anomalised median value is sufficiently far from the median of all months and there is a large asymmetry in the shape of the distribution (Fig. 2). An extension of this test looks at the distribution of all observations within a calendar month across the entire data series. Any secondary populations are removed.

Streak Check: Repetitions of the same observation value over a number of consecutive time periods, the same value at the same hour over a number of days, and whole day repetitions over a series of days are removed (Fig. 3).

Spike Check: Time series spikes of up to three values are removed using the threshold values (calculated using only data passing previous tests) on the first differences into and out of the spike.

Climatological Check: Climatologies and anomalies are created for each hour of the day within a month. Values sufficiently far from the centre of the distribution are removed. Those within a threshold are checked against neighbours and some values reinstated. To retain storm signals, this test is not applied to pressure data.

Excess Variance Check: Whole months are removed where the within-month variance of the normalised anomalies is sufficiently larger than the median variance over the full series for that calendar month. To retain storm signals, wind speed data are combined with a first difference of the pressure data to find the progressive spikes created by the passage of a low pressure core.

Other Checks: Further checks flag isolated clusters of data, periods of time with poor conversion to UT using the diurnal cycle, observations which exceed known records, excessive supersaturation and suspected wet-bulb drying. Cloud cover values are also checked for logical consistency.

External Checks: Up to ten nearest neighbour stations within 500m vertically, 300km horizontally and spread across four directional quadrants are selected. Difference series' between the station and its neighbours are used to identify suspect observations and reinstate some flagged values.

QC Sequence: After initial QC and data removal, new isolated clusters and spikes of data are identified. Then the *external checks* are repeated and finally months with exceedingly poor data are removed.

Example Extremes

Figs. 5 and 6 show the passage of the low pressure core of Hurricane Katrina as it crossed the USA in August 2005 and the end of the heat wave in southern Australia in 2009, showing the return of cooler temperatures on the 9th February.

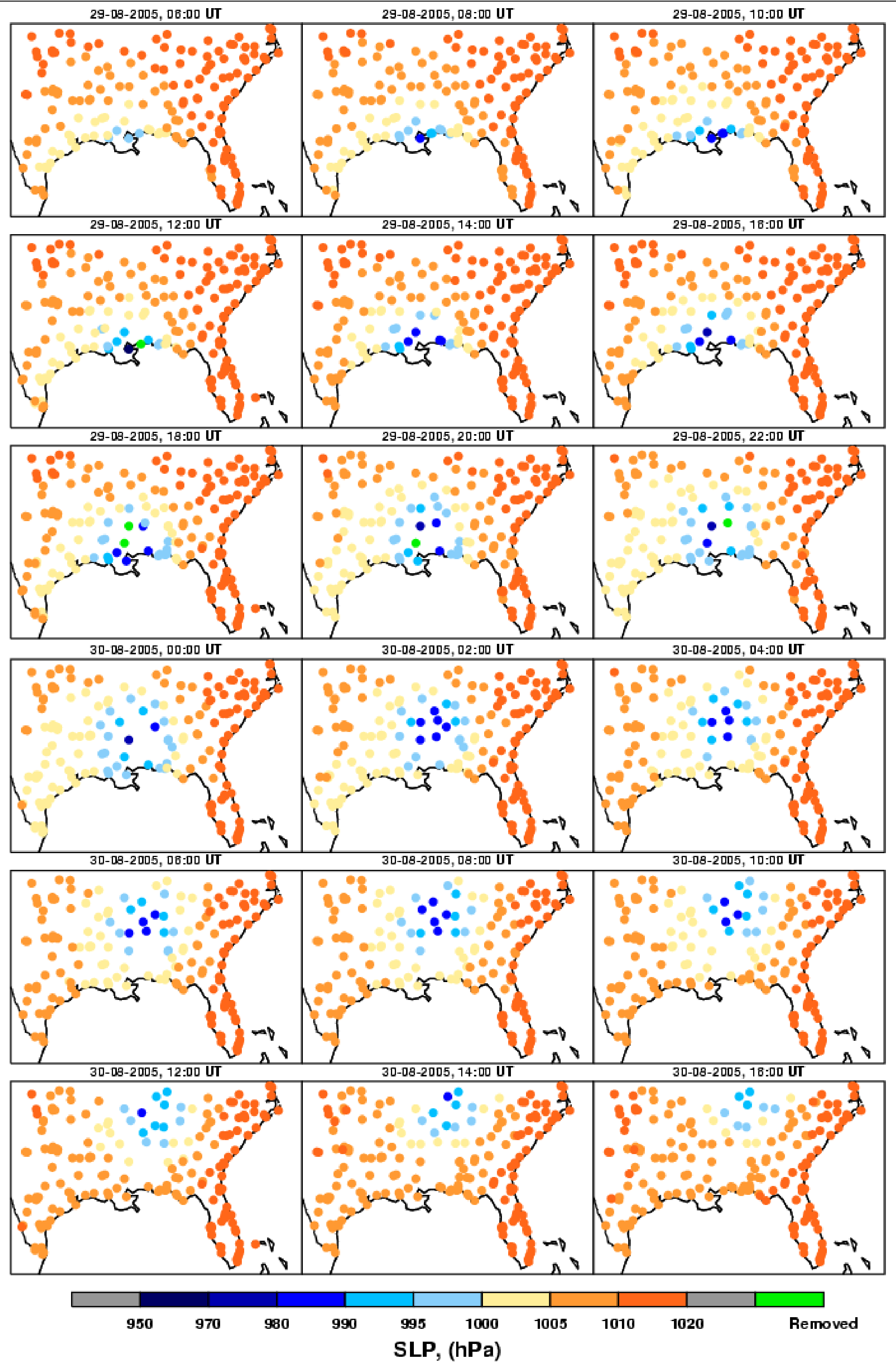

Fig 5. The low pressure core of Hurricane Katrina as it moves over the USA on 29th and 30th August 2005. Points coloured grey are beyond the values shown in the colour bar. Points coloured green are those removed by the quality control suite outlined on the left. In this case, the neighbour outlier check has erroneously removed valid points. Times are UT and only every second hour is shown.

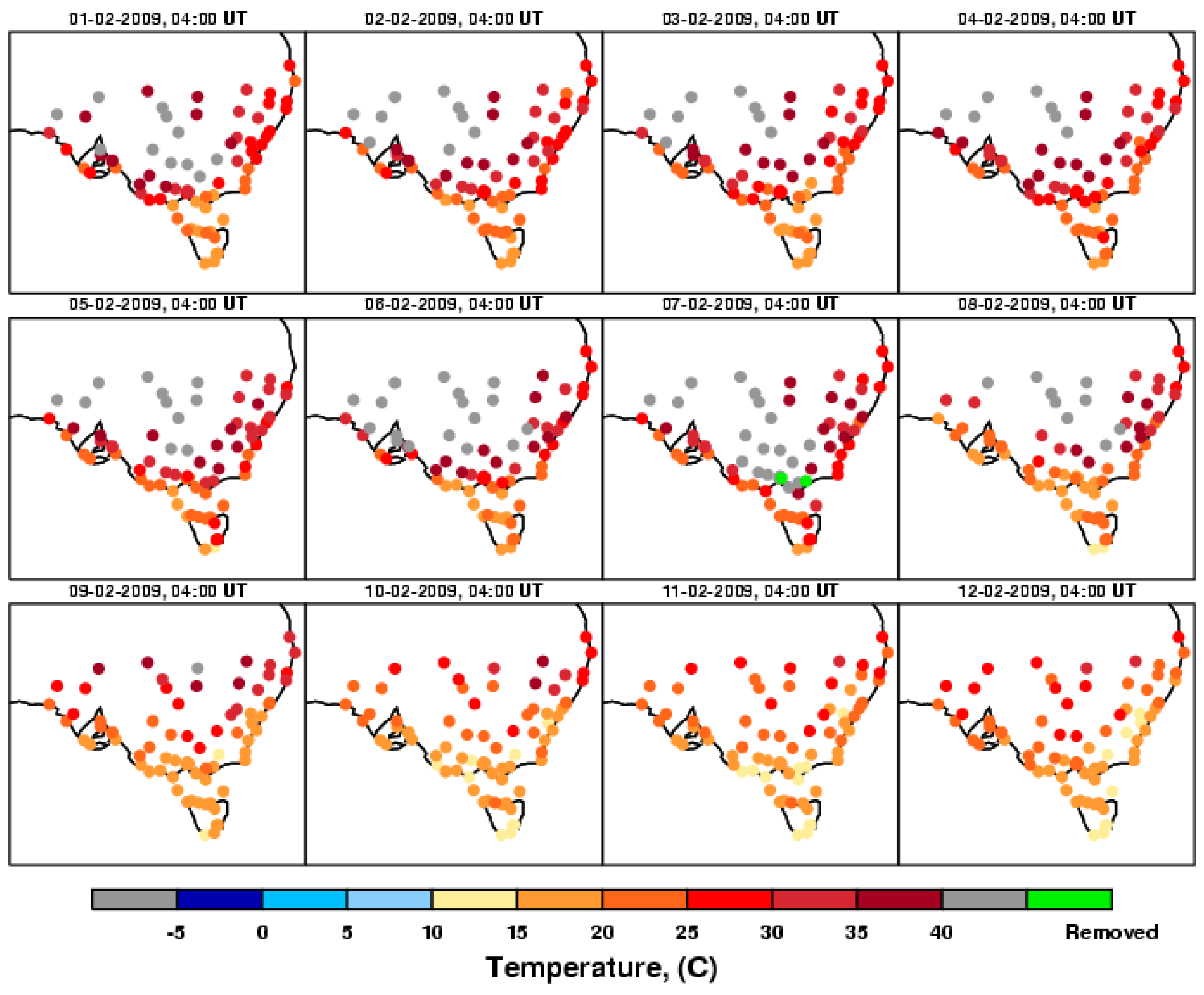

Fig 6. The end of the heat wave in southern Australia in February 2009. Points coloured grey are beyond the values shown in the colour bar. Points coloured green are those removed by the quality control suite outlined on the left. In this case, they have been removed by the gap check routine. Times are UT and only one hour is shown for every day.

Homogenisation is an essential part of data-product creation but difficult to apply to high resolution data. Attempts will be made to detect large inhomogeneities within HadISD. There will be two initial releases through www.hadobs.org: the full set of QC'd stations with detected inhomogeneities identified; and the subset believed to be free from large inhomogeneities. Later work will attempt to produce a homogenous product.

Met Office Hadley Centre, FitzRoy Road, Exeter, Devon, EX1 3PB United Kingdom
Tel: 01392 885680 Fax: 01392 885681
Email: robert.dunn@metoffice.gov.uk